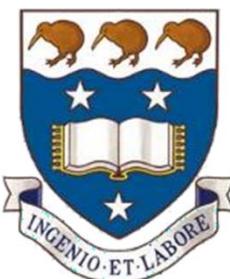


Deep Learning for Improving the Detection, Prognosis, and Treatment of Atrial Fibrillation



**THE UNIVERSITY
OF AUCKLAND**
NEW ZEALAND
Te Whare Wānanga o Tāmaki Makaurau

Author:

Zhaohan Xiong

Supervisors:

Dr. Jichao Zhao

A/Prof. Martin Stiles

Dr. Gonzalo Daniel Maso Talou

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy in Bioengineering, the University of Auckland, 2022. This thesis is for examination purposes only and is confidential to the examination process.

Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Xiong, Z., Nalar, A., Jamart, K., Stiles, M.K., and Zhao, J., 2019. Fully Automatic 3D Bi-Atria Segmentation from Late Gadolinium-Enhanced MRIs Using Double Convolutional Neural Networks. *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 63-71.

Nature of contribution by PhD candidate	Manuscript Writing, Data Preparation, Model Development, Results Analysis, Figures Creation, Manuscript Finalization
Extent of contribution by PhD candidate (%)	90

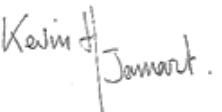
CO-AUTHORS

Name	Nature of Contribution
Aaqel Nalar	Data Preparation, Advising, Manuscript Editing
Kevin Jamart	Advising, Manuscript Editing
Martin Stiles	Advising, Manuscript Editing
Jichao Zhao	Supervision, Manuscript Writing, Reviewing

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ that the candidate wrote all or the majority of the text.

Name	Signature	Date
Aaqel Nalar		01-01-2021
Kevin Jamart		01-01-2021
Martin Stiles		01-01-2021
Jichao Zhao		01-01-2021

Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Xiong, Z., Jamart, K., Stiles, M.K., and Zhao, J., 2019. Fully Automatic 3D Whole Atria Segmentation and Reconstruction from Gadolinium-Enhanced MRIs Using Convolutional Neural Networks. *The Annual Scientific Meetings of Heart Rhythm Society*.

Nature of contribution by PhD candidate	Manuscript Writing, Data Preparation, Model Development, Results Analysis, Figures Creation, Manuscript Finalization
Extent of contribution by PhD candidate (%)	90

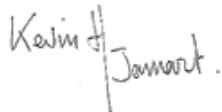
CO-AUTHORS

Name	Nature of Contribution
Kevin Jamart	Advising, Manuscript Editing
Martin Stiles	Advising, Manuscript Editing
Jichao Zhao	Supervision, Manuscript Writing, Reviewing

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ that the candidate wrote all or the majority of the text.

Name	Signature	Date
Kevin Jamart		01-01-2021
Martin Stiles		01-01-2021
Jichao Zhao		01-01-2021

Co-Authorship Form

This form is to accompany the submission of any PhD that contains published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in all copies of your thesis submitted for examination and library deposit (including digital deposit), following your thesis Acknowledgements. Co-authored works may be included in a thesis if the candidate has written all or the majority of the text and had their contribution confirmed by all co-authors as not less than 65%.

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., Heng, P.A., Ni, D., Li, C., Tong, Q., Si, W., Puybareau, E., Khoudli, Y., Géraud, T., Chen, C., Bai, W., Rueckert, D., Xu, L., Zhuang, X., Luo, X., Jia, S., Sermesant, M., Liu, Y., Wang, K., Borra, D., Masci, A., Corsi, C., de Vente, C., Veta, M., Karim, R., Preetha, C., Engelhardt, S., Qiao, M., Wang, Y., Tao, Q., Nuñez-Garcia, M., Camara, O., Savioli, N., Lamata, P., Zhao, J., 2020. A Global Benchmark of Algorithms for Segmenting the Left Atrium from Late Gadolinium-Enhanced Cardiac Magnetic Resonance Imaging. *Medical Image Analysis*, 67, 101832.

Nature of contribution by PhD candidate	Benchmarking/Competition Organizer, Manuscript Writing, Results Curation, Results Analysis, Figures Creation, Rebuttal Drafting, Manuscript Finalization
Extent of contribution by PhD candidate (%)	90

CO-AUTHORS

Name	Nature of Contribution
Qing Xia	Benchmarking/Competition Participant*
Zhiqiang Hu	Benchmarking/Competition Participant*
Ning Huang	Benchmarking/Competition Participant*
Cheng Bian	Benchmarking/Competition Participant*
Yefeng Zheng	Benchmarking/Competition Participant*
Sulaiman Vesal	Benchmarking/Competition Participant*
Nishant Ravikumar	Benchmarking/Competition Participant*
Andreas Maier	Benchmarking/Competition Participant*
Xin Yang	Benchmarking/Competition Participant*
Pheng-Ann Heng	Benchmarking/Competition Participant*
Dong Ni	Benchmarking/Competition Participant*
Caizi Li	Benchmarking/Competition Participant*
Qianqian Tong	Benchmarking/Competition Participant*
Weixin Si	Benchmarking/Competition Participant*
Elodie Puybareau	Benchmarking/Competition Participant*
Younes Khoudli	Benchmarking/Competition Participant*
Thierry Géraud	Benchmarking/Competition Participant*
Chen Chen	Benchmarking/Competition Participant*
Wenjia Bai	Benchmarking/Competition Participant*

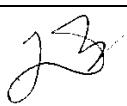
Daniel Rueckert	Benchmarking/Competition Participant*
Lingchao Xu	Benchmarking/Competition Participant*
Xiahai Zhuang	Benchmarking/Competition Participant*
Xinzhe Luo	Benchmarking/Competition Participant*
Shuman Jia	Benchmarking/Competition Participant*
Maxime Sermesant	Benchmarking/Competition Participant*
Yashu Liu	Benchmarking/Competition Participant*
Kuanquan Wang	Benchmarking/Competition Participant*
Davide Borra	Benchmarking/Competition Participant*
Alessandro Masci	Benchmarking/Competition Participant*
Cristiana Corsi	Benchmarking/Competition Participant*
Coen de Vente	Benchmarking/Competition Participant*
Mitko Veta	Benchmarking/Competition Participant*
Rashed Karim	Benchmarking/Competition Participant*
Chandrakanth Jayachandran Preetha	Benchmarking/Competition Participant*
Sandy Engelhardt	Benchmarking/Competition Participant*
Menyun Qiao	Benchmarking/Competition Participant*
Yuanyuan Wang	Benchmarking/Competition Participant*
Qian Tao	Benchmarking/Competition Participant*
Marta Nuñez-Garcia	Benchmarking/Competition Participant*
Oscar Camara	Benchmarking/Competition Participant*
Nicolo Savioli	Benchmarking/Competition Participant*
Pablo Lamata	Benchmarking/Competition Participant*
Jichao Zhao	Benchmarking/Competition Organizer, Supervision, Manuscript Writing, Reviewing

*Note these co-authors were participants of the competition and did not have any contribution to actual publication.

Certification by Co-Authors

The undersigned hereby certify that:

- ❖ the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and
- ❖ that the candidate wrote all or the majority of the text.

Name	Signature	Date
Jichao Zhao		25-01-2021

Abstract

Atrial fibrillation (AF) is the most common abnormal heart rhythm, contributing to stroke in one in five cases for people aged over 60 years and increasing the risk of heart failure three-fold. The majority of patients initially develop paroxysmal AF which is sustained for less than seven days. If left untreated, these patients can progress to persistent (and eventually permanent) AF, where the AF episodes last much longer and confer substantially higher risks of stroke. Persistent AF leads to progressive alterations in the electrical and structural properties of the atria which make restoring normal rhythm more difficult. Early detection and intervention are therefore needed to minimise the impact of AF and improve the quality of life for patients. However, the diagnosis of AF in clinics using a Holter monitor by trained physicians is difficult and error-prone, particularly due to its intermittent nature. Furthermore, current treatment performs poorly for persistent and longstanding-persistent AF, even when using the most promising ablation treatments which target localised atrial substrates via a minimally invasive procedure to treat AF. The low success rate of ablation in these groups is primarily due to the lack of effective clinical tools to detect and visualise the underlying atrial substrates sustaining AF.

It follows that there is an urgent need for more intelligent methods to address these problems. Artificial intelligence is now widely applied in medicine. The superior data-driven capabilities of convolutional neural networks (CNNs) over traditional approaches have led to their becoming the dominant driver of artificial intelligence in the past decade. The overriding objective of this thesis was to develop novel CNN-based algorithms to automate and improve AF detection and treatment.

Firstly, we addressed the lack of efficient methods for AF diagnosis by designing a convolutional recurrent network (CRN) for portable electrocardiogram (ECG) recording devices. The CRN was designed to process ECGs of varying input lengths by combining the feature learning capabilities of CNN and the recursive capacity of recurrent neural networks. A novel neural style transfer and mathematical modelling pipeline was designed to simulate synthetic ECGs and increase the training samples for the CRN. The framework was tested using the world's largest open-source ECG dataset. We demonstrate that our approach detected AF with higher accuracy than commercial devices with significantly lower rates of false AF detection than current methods.

Secondly, we addressed the lack of effective methods for detecting atrial substrates by

developing a CNN-based pipeline for automatic structural analysis from 3D late gadolinium-enhanced magnetic resonance imaging (LGE-MRI). This was capable of performing segmentation and the extraction of key biomarkers including the bi-atrial chamber measurements, anatomic structure, fibrosis distribution, and wall thickness variation. The pipeline was trained and validated on the world's largest LGE-MRI datasets. It was further tested on independent clinical data and demonstrated its efficacy by performing a range of analytical tasks with a higher degree of confidence and efficiency than traditional methods.

Thirdly, we improved on current clinical mapping techniques for 3D atrial visualisation of the atria during ablation procedures. Present commercial tools combine 3D electrical mapping data with additional imaging and manual registration to achieve accurate anatomic reconstructions. We developed a novel CNN for automatic 3D atrial chamber reconstruction directly from sparse point clouds acquired during mapping. The approach was validated using clinical MRI and computed tomography datasets, and demonstrated accuracies equal to or better than commercial software. Importantly, the CNN was capable of predicting anatomical structures in sparsely mapped regions, allowing accurate reconstruction of 3D atrial anatomy from smaller datasets. This new approach potentially accelerates and improves 3D atrial visualisation needed for ablation and could therefore reduce the time needed for these procedures.

Overall, the study presented significant contributions to improving the detection, prognosis, and treatment of AF by developing intelligent algorithms which can be integrated into the current clinical workflow. The methods and results derived from this research could facilitate clinical decision making and improve clinical outcomes for patients suffering from AF worldwide.

Acknowledgements

I would like to express my sincere gratitude to Dr. Jichao Zhao for his supervision, guidance, and leadership over the past three years as a doctoral student. Being a part of Dr. Zhao's group has undoubtedly taught me many key skills and values, opening many opportunities for the future. Dr. Zhao has been extremely hands-on with every aspect of my research. The things I learned as a result of his extensive mentorship will follow me for the rest of my life. He has also been extremely supportive of all my efforts over the years, leading me every step of the way and putting tremendous amounts of resources behind me to guarantee my success. I will forever be thankful to him for believing in me from the very start and giving me the opportunity to develop my skills and elevate me to a high level.

In addition, I would like to thank my co-supervisor, A/Prof. Martin Stiles for the support with regard to the clinical aspect of the study. I thoroughly enjoyed the clinical visits which were generously hosted by him, and gained tremendous value from a clinical perspective. Furthermore, thanks to Dr. Gonzalo Daniel Maso Talou for his co-supervision and help during the thesis preparation stages.

I would especially like to thank my girlfriend, Pearl Wu, for supporting me since the very start of my research journey and compromising the many sacrifices I had to make to pursue my studies. Her constant encouragement was pivotal to my success and I couldn't have done it otherwise.

I would also like to thank our international collaborators for providing the valuable datasets used in this thesis, particularly Prof. Rob Macleod from the University of Utah and Prof. Steffen Petersen from the Queen Mary University of London. A special thanks to Aaqel Nalar and James Kennelly for helping with the tedious data preparation.

Lastly, I am grateful to the Heart Foundation New Zealand for funding my doctoral study.

Table of Contents

Abstract.....	i
Acknowledgements	iii
Table of Contents	iv
List of Figures.....	ix
List of Tables	xiii
List of Equations	xiv
List of Abbreviations	xvi
Statement of Contribution	xvii
1. Introduction.....	1
1.1 Background	1
1.2 Motivations.....	2
1.3 Research Objectives	4
1.4 Thesis Outline.....	5
1.5 Publications	7
2. Background	10
2.1 The Heart.....	10
2.1.1 The Structure and Function of the Heart	10
2.1.2 The Structure of the Atria.....	11
2.1.3 Atrial Fibrillation.....	13
2.1.4 Electrocardiogram	14
2.2 Structural Remodelling of the Atria	16
2.2.1 Atrial Remodelling	16
2.2.2 Atrial Fibrosis.....	17
2.2.3 Late Gadolinium-Enhanced Magnetic Resonance Imaging.....	20
2.3 Treatments for Atrial Fibrillation	24
2.3.1 Overview	24
2.3.2 Catheter Ablation	25
2.3.3 Electro-Anatomical Mapping	26
3. Literature Review on Deep Learning.....	30

3.1	Background	30
3.2	Basic Principles of Neural Networks	31
3.2.1	Simple Feedforward Neural Networks.....	31
3.2.2	Convolutional Neural Networks.....	34
3.2.3	Residual Convolutional Neural Networks.....	36
3.2.4	Recurrent Neural Networks.....	37
3.3	Convolutional Neural Networks for Classification	39
3.4	Convolutional Neural Networks for Localization	41
3.5	Fully Convolutional Networks for Image Segmentation and Reconstruction	44
3.6	Neural Style Transfer Networks for Data Augmentation.....	48
3.7	Evaluation of Deep Learning Models	52
3.7.1	Technical Measures	53
3.7.2	Biological Measures for Atrial Anatomical Accuracy	54
4.	Enhancing the Detection of Atrial Fibrillation	56
4.1	Introduction	56
4.2	Methods	59
4.2.1	Clinical ECG Dataset	59
4.2.2	Synthetic ECG Generation	59
4.2.3	Atrial Fibrillation Detection Algorithm	63
4.3	Results	66
4.3.1	Synthetic ECG Data	66
4.3.2	Validation of Synthetic ECGs	66
4.3.3	Atrial Fibrillation Detection Results	68
4.4	Discussion	70
4.4.1	Methods of AF Detection	71
4.4.2	Deep Learning for ECG Classification	71
4.4.3	Limitations and Future Work	72
4.5	Conclusions	73
5.	Automatic Left Atrium Segmentation from LGE-MRI.....	74
5.1	Introduction	74
5.2	Methods	75
5.2.1	Dataset and Labels.....	75

5.2.2	Fully Convolutional Network for Segmentation	77
5.2.3	Evaluation.....	79
5.3	Results	79
5.4	Discussion	80
5.4.1	Left Atrium Segmentation from LGE-MRIs	80
5.4.2	Multi-Scale Processing in Convolutional Neural Networks	82
5.4.3	Limitations and Future Work	82
5.5	Conclusions	83

6. A Global Benchmark of Algorithms for Segmenting the Left Atrium from LGE-MRI

.....	84
-------	-------	----

6.1	Introduction	85
6.2	Methods	88
6.2.1	Dataset and Labels.....	88
6.2.2	2018 Atrial Segmentation Challenge Setup	90
6.2.3	Algorithm Evaluation	91
6.3	Results	91
6.3.1	Performance of Submitted Algorithms.....	91
6.3.2	Top Performing CNN Methods	95
6.3.3	Key Factors Influencing the Performance of Double CNN Approaches	98
6.4	Discussion	100
6.4.1	Characteristics of Top-Performing CNNs for Segmentation	100
6.4.2	Comparison of Evaluation Metrics.....	100
6.4.3	Error Analysis and Future Work	102
6.5	Conclusions	104

7. Automatic Extraction of Pathologic Biomarkers in the Bi-Atrial Chambers from Cine-MRI and LGE-MRI.....

.....	105
-------	-------	-----

7.1	Introduction	106
7.2	Methods	108
7.2.1	Multi-Centre Data for Validation	108
7.2.1.1	Ground Truth Atrial Segmentations for Cine-MRIs and LGE-MRIs	109

7.2.1.2	Ground Truth Atrial Wall Thickness for LGE-MRIs.....	110
7.2.1.3	Ground Truth Atrial Fibrosis for LGE-MRIs.....	112
7.2.2	AtriaNet for Automatic Atrial Phenotyping	113
7.2.2.1	AtriaNet Configuration for Segmentation.....	116
7.2.2.2	AtriaNet Configuration for Wall Thickness Estimation.....	117
7.2.2.3	AtriaNet Configuration for Fibrosis Quantification.....	118
7.2.3	Evaluation.....	118
7.3	Results	119
7.3.1	Accurate Atrial Chamber Segmentation.....	119
7.3.2	AtriaNet Calculates the Chamber Volume and Diameter with Clinical Level Precision	122
7.3.3	AtriaNet Estimates Accurate 3D Wall Thicknesses	125
7.3.4	AtriaNet Fibrosis Quantification and Validation with Clinical Electro- Anatomical Mapping	127
7.4	Discussion	130
7.4.1	Automatic Extraction of Bi-Atrial Biomarkers from MRIs	130
7.4.2	Importance of Accurate Segmentation of the Atrial Wall.....	131
7.4.3	Multi-Task CNN Architecture.....	132
7.4.4	Limitations and Future Work	132
7.5	Conclusions	133

8. Automatic Atrial Surface Reconstruction from Clinically Mapped Point Clouds ...134

8.1	Introduction	135
8.2	Methods	137
8.2.1	Data and Pre-Processing.....	137
8.2.1.1	Paired Training Data	138
8.2.1.2	Test #1: Clinically Paired MRI and Point Cloud Data.....	138
8.2.1.3	Test #2: Clinically Paired CT and Point Cloud Data	139
8.2.2	Method of Generating the Paired Training Data	139
8.2.2.1	Landmark Registration from Point Cloud to LGE-MRI	139
8.2.2.2	Projection from Point Cloud to LGE-MRI.....	141
8.2.2.3	Creating Partial Left Atrial Shell from Point Cloud.....	142
8.2.3	CNN for Left Atrial Reconstruction from Point Clouds	144
8.2.4	Evaluation.....	148

8.3	Results	148
8.3.1	Accuracy for Predicting the Left Atrium from Point Clouds.....	148
8.3.2	Visualization and Error Analysis	150
8.3.3	Impact of Point Cloud Coverage on Accuracy.....	152
8.4	Discussion	155
8.4.1	3D Reconstruction of Anatomical Structures Directly from Point Clouds.....	155
8.4.2	Pre-Processing and Network Tuning.....	155
8.4.3	Limitations and Future Work	156
8.5	Conclusions	157
9.	Conclusions.....	158
9.1	Summary and Contributions.....	158
9.2	Limitations and Future Work	160
9.2.1	Data for Training Deep Learning Models	160
9.2.2	Neural Network Architectures	161
9.2.3	Future Work in ECG and Point Cloud Analysis	162
9.3	Concluding Remarks	163
Appendix Text.....	164	
Appendix Figures.....	167	
Bibliography	176	

List of Figures

Figure 1.1. Schematic summary of the three main objectives investigated in this thesis.....	4
Figure 1.2. The organization overview of this thesis.	6
Figure 2.1. Illustration of the detailed anatomy of the heart.....	11
Figure 2.2. The anatomical structure and components of the left and right atrial chambers..	12
Figure 2.3. Visualization of the electrical propagation patterns present in the heart during sinus rhythm and atrial fibrillation.....	13
Figure 2.4. Illustration of a single cycle of an electrocardiogram along with the corresponding cardiac functions throughout the cardiac cycle.....	15
Figure 2.5. Illustration and comparison of typical electrocardiogram recordings of sinus rhythm, and atrial fibrillation.....	15
Figure 2.6. Light microscopy imaging with Masson's trichrome stain of the crista terminalis in the right atrium for observing the degree of fibrosis in patients with varying severities of atrial fibrillation.....	19
Figure 2.7. Illustration of the difference in imaging resolution between conventional and high-resolution late gadolinium-enhanced magnetic resonance imaging.....	21
Figure 2.8. The percentage of left atrial fibrosis calculated from late gadolinium-enhanced magnetic resonance imaging of atrial fibrillation patients prior to catheter ablation is an independent predictor of the success of ablation.	23
Figure 2.9. Illustration of the pulmonary vein isolation catheter ablation procedure in the left atrial chamber using radiofrequency ablation and cryoablation.	26
Figure 2.10. Illustration of the cardiac reconstruction algorithm from Biosense Webster....	28
Figure 2.11. Visualization of left atrial reconstruction and voltage map recorded with electro-anatomical mapping using CARTO 3 (Biosense Webster, California, United States) during catheter ablation of atrial fibrillation at Waikato Hospital.....	29
Figure 3.1. The basic concept of an artificial neural network.....	33
Figure 3.2. Illustration of the basic operations in a convolutional neural network.	35
Figure 3.3. A simple convolutional neural network with a single convolutional layer and two fully connected layers.	35
Figure 3.4. Illustration of the basic operations in residual blocks which are used in convolutional neural networks.....	36
Figure 3.5. Demonstration of the operations of a layer in a recurrent neural network with a single node.	38

Figure 3.6. Illustration of the residual convolutional neural network architecture which won the 2015 ImageNet Large Scale Visual Recognition Challenge.....	40
Figure 3.7. A practical example of image localization with a convolutional neural network in a self-driving car.	41
Figure 3.8. The architecture of the most popular convolutional neural network for medical image segmentation, U-Net.	46
Figure 3.9. The architecture of the 3D version of U-Net for 3D image segmentation.	47
Figure 3.10. The architecture of V-Net.....	48
Figure 3.11. The overall concept of neural style transfer.	49
Figure 3.12. A simplified representation of the neural style transfer network.	50
Figure 3.13. Photo-realistic neural style transfer the content and style images combined to produce a realistic output.....	51
Figure 3.14. The structure of universal style transfer network consisting of a VGG-Net (and its inverse) and a whitening and colour transform operation.....	52
Figure 4.1. Detection of atrial fibrillation from single-lead electrocardiograms through a mobile device using deep learning.....	58
Figure 4.2. Method for generating realistic electrocardiograms to increase the data available for deep learning synthetically.....	61
Figure 4.3. Generation of realistic synthetic electrocardiograms data for sinus rhythm and atrial fibrillation.	63
Figure 4.4. Illustration of the deep learning algorithm used for classifying electrocardiograms	64
Figure 4.5. Validation of the synthetic electrocardiograms.	67
Figure 4.6. Accuracy evaluation of the deep learning algorithm for electrocardiogram classification for sinus rhythm and atrial fibrillation.	69
Figure 4.7. Visualization of electrocardiograms incorrectly classified by the deep learning algorithm.	70
Figure 5.1. Illustration and description of the manual segmentation procedure, defining the workflow and anatomical extents of the left atrial geometry.	77
Figure 5.2. The architecture of the proposed dual fully convolutional network for left atrial segmentation.	79
Figure 6.1. The overall workflow of medical images for clinical usage and the 2018 Atrial Segmentation Challenge.	87

Figure 6.2. Variation in the quality of the late gadolinium-enhanced magnetic imaging dataset used for this study.....	89
Figure 6.3. Comparative summaries of the 17 algorithms included in this benchmarking study, a well representative subset of the 27 teams that participated in the 2018 Atrial Segmentation Challenge.	93
Figure 6.4. Detailed evaluations of the performance of the 15 convolutional neural network (CNN) pipelines submitted to the challenge grouped into three general categories depending on how the CNNs were applied to segment the left atrium from the late gadolinium-enhanced magnetic resonance imaging.	95
Figure 6.5. Post-challenge analysis of the winning method by Xia et al. in the 2018 Atrial Segmentation Challenge demonstrates the optimality of their approach.....	97
Figure 6.6. Analysis of the key factors of the first convolutional neural network (CNN) in the double CNN methods in enhancing left atrium segmentation performance.....	98
Figure 6.7. Summary of the rankings for all participants under different technical metrics, sorted in descending Dice scores.	101
Figure 6.8. Error analysis of predictions generated from the winning method by Xia et al. in the Atrial Segmentation Challenge.	102
Figure 7.1. Illustration of the Laplace solution on a sample of left and right atrial tissue. .	110
Figure 7.2. Demonstration of thresholding for fibrosis extraction from a 2D slice of a late gadolinium-enhanced magnetic resonance imaging scan on the left atrium.	112
Figure 7.3. Illustration of the AtriaNet pipeline for automatic extraction of pathologic biomarkers from cardiac magnetic resonance imaging.	115
Figure 7.4. Segmentation performance of AtriaNet on 2D cine-magnetic resonance imaging (MRI) and 3D late gadolinium-enhanced MRI.....	121
Figure 7.5. The accuracy of AtriaNet for atrial measurement calculation from 2D cine-magnetic resonance imaging (MRI) and 3D late gadolinium-enhanced MRI, and clinical validation using clinical echocardiography.	124
Figure 7.6. Evaluation and visualization of the wall thickness estimation accuracy of AtriaNet on 3D late gadolinium-enhanced magnetic resonance imaging.	127
Figure 7.7. Fibrosis quantification performance of AtriaNet on 3D late gadolinium-enhanced magnetic resonance imaging and validation low-voltage mapping of the atria.....	129
Figure 8.1. Illustration of the projection of the registered points onto the left atrial surface during data simulation.....	142

Figure 8.2. Example of the impact of the alpha parameter in the concave hull algorithm when generating a mesh from individual points in the point cloud.....	143
Figure 8.3. Illustration of the iterative concave hull algorithm proposed in this study to the discretization of the point clouds into an image volume containing the partial left atrial shell used as the input for the convolutional neural network. ..	144
Figure 8.4. The architecture of the proposed 3D convolutional neural network for predicting the left atrial geometry from a point cloud obtained during clinical mapping.	146
Figure 8.5. 3D visualizations of the left atrial reconstructions of five samples in the <i>paired training</i> dataset.....	150
Figure 8.6. 3D Visualizations of the left atrial reconstruction for one sample each from <i>test #1</i> and <i>test #2</i> clinical datasets.....	152
Figure 8.7. Correlation of the point cloud coverage and Dice score in the test datasets.	154
Figure 8.8. Correlation of the point cloud coverage with the left atrium volume error in the testing datasets. ..	155

List of Tables

Table 5.1. Summary of the 3D late gadolinium-enhanced MRIs dataset from the University of Utah, split by pre- and post-catheter ablation.....	76
Table 5.2. Evaluation metrics for left atrial wall and cavity segmentation for the 22 test 3D late gadolinium-enhanced MRIs	80
Table 5.3. The left atrial diameter and volume measurements and errors for the 22 test 3D late gadolinium-enhanced MRIs	80
Table 6.1. Summary of methods submitted to the 2018 Atrial Segmentation Challenge.....	90
Table 8.1. Summary of the data used in this study.	137
Table 8.2. Configuration of the proposed CNN.....	147
Table 8.3. Technical evaluation for left atrium reconstruction from point clouds in the 300 generated, 4 clinical MRI, and 2 clinical CT data.	149
Table 8.4. Biological evaluation for left atrium reconstruction from point clouds in the 300 generated, 4 clinical MRI, and 2 clinical CT data.	150
Table 8.5. The point cloud coverage over the left atrium for the generated, clinical MRI, and clinical CT datasets.....	153

List of Equations

Equation 3.1. Rectified linear unit activation function.....	32
Equation 3.2. Parametric rectified linear unit activation function.....	32
Equation 3.3. Leaky rectified linear unit activation function	32
Equation 3.4. Softmax activation function for classification	32
Equation 3.5. Highest probability class from classification	32
Equation 3.6. Backpropagation during neural network optimization.....	34
Equation 3.7. Cross-entropy loss function.....	34
Equation 3.8. Convolutional operation in convolutional neural networks	34
Equation 3.9. Max pooling operation in convolutional neural networks	35
Equation 3.10. Dropout operation for neural network regularization	36
Equation 3.11. Batch normalization operation	37
Equation 3.12. Recurrent operation of the $T + 1$ layer in recurrent neural networks.....	37
Equation 3.13. Recurrent operation of the $N + 1$ layer in recurrent neural networks	37
Equation 3.14. Mean squared error loss function for localization.....	42
Equation 3.15. Unpooling operation in fully convolutional networks	44
Equation 3.16. Pixel-wise highest probability classes from segmentation	45
Equation 3.17. Content loss function in neural style transfer.....	49
Equation 3.18. Style loss function in neural style transfer	50
Equation 3.19. Style loss function across multiple layers	50
Equation 3.20. Mean square error evaluation metric.....	53
Equation 3.21. Dice score evaluation metric	53
Equation 3.22. Intersection over union evaluation metric.....	53
Equation 3.23. Sensitivity evaluation metric.....	54
Equation 3.24. Specificity evaluation metric.....	54
Equation 3.25. Hausdorff distance evaluation metric.....	54
Equation 3.26. Surface to surface distance evaluation metric.....	54
Equation 3.27. Atrial diameter calculation.....	55
Equation 3.28. Atrial diameter biological evaluation metric.....	55
Equation 3.29. Atrial volume calculation	55
Equation 3.30. Atrial volume biological evaluation metric.....	55
Equation 4.1. P-wave equation in the ODE formulation for ECG simulation	60
Equation 4.2. QRS-wave equation in the ODE formulation for ECG simulation.....	60

Equation 4.3. T-wave equation in the ODE formulation for ECG simulation	60
Equation 4.4. Content loss for ECG generation with neural style transfer	62
Equation 4.5. Style loss for ECG generation with neural style transfer.....	62
Equation 6.1. Signal to noise ratio for image quality measurement	88
Equation 6.2. Contrast ratio for image quality measurement.....	88
Equation 6.3. Heterogeneity for image quality measurement	88
Equation 7.1. Epicardial part of the PDE formulation for wall thickness estimation	111
Equation 7.2. Endocardial part of the PDE formulation for wall thickness estimation	111
Equation 7.3. Thresholding for fibrosis extraction from atrial imaging.....	113
Equation 7.4. Multi-class dice loss function for segmentation.....	117
Equation 7.5. Class-wise dice score calculation for the dice loss	117
Equation 7.6. Masked mean squared error loss function for wall thickness estimation.....	117
Equation 7.7. Pixel-wise wall thickness evaluation metric	118
Equation 7.8. Kolmogorov-Smirnov evaluation metric	118
Equation 8.1. Point cloud centre of mass translation vector	139
Equation 8.2. Cross product to compute the normal vector for rotation	140
Equation 8.3. Polar coordinate radial calculation.....	140
Equation 8.4. Polar coordinate angle calculation	140
Equation 8.5. Point cloud rotational vector	140
Equation 8.6. Rotated coordinates in the y-axis of a point cloud	141
Equation 8.7. Rotated coordinates in the z-axis of a point cloud	141
Equation 8.8. Point cloud scaling vector	141
Equation 8.9. Projected coordinates of a point cloud given an atrial surface.....	142
Equation 8.10. Array of points along each direction of projection	148

List of Abbreviations

AF	Atrial Fibrillation
ANN	Artificial Neural Network
AWT	Atrial Wall Thickness
CNN	Convolutional Neural Network
CRN	Convolutional Recurrent Network
CT	Computed Tomography
CVD	Cardiovascular Diseases
ECG	Electrocardiogram
FCN	Fully Convolutional Network
GPU	Graphics Processing Unit
IoU	Intersection over Union
LA	Left Atrium
LGE-MRI	Late Gadolinium-Enhanced Magnetic Resonance Imaging
LV	Left Ventricle
MICCAI	Medical Image Computing and Computer Assisted Intervention
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
NST	Neural Style Transfer
PReLU	Parametric Rectified Linear Unit
PV	Pulmonary Vein
PVI	Pulmonary Vein Isolation
RA	Right Atrium
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
ROI	Region of Interest
RV	Right Ventricle
SR	Sinus Rhythm
STACOM	Statistical Atlases and Computational Modelling of the Heart
STSD	Surface to Surface Distance
VGG	Visual Geometry Group

Statement of Contribution

All text in this thesis were written by me including the abstract, introduction (Chapter 1), literature review (Chapters 2-3), methods (Chapters 4-8), and conclusions (Chapter 9). The list of tables and bibliography were also generated by me. All methods presented in this thesis were developed by me including the methodology design, prototyping, tuning, final deployment, and post-analysis of the algorithms. All results, discussions, and conclusions were obtained by me including any summaries, visualisations, and presentations. The figures, equations, and tables shown in this thesis were also generated by me and contained references to the original source when incorporating existing research from published materials.

The data used in this thesis were from external sources, cited where relevant, and obtained with permission from their original sources. These have been mentioned in the acknowledgement and the relevant individuals in charge of approving the release of the data to me were co-authors in the publications presented in the thesis. All other data that did not have any external reference were provided by my co-supervisor A/Prof. Martin Stiles from Waikato hospital.

The manual preparation of the data was performed by me and a few research assistants/interns in our research group for the work presented in a few chapters of this thesis. These have been mentioned in the acknowledgements. In cases where several individuals contributed to the data preparation, my role was the main organiser and leader of the project, performed quality control, and oversaw the entire operation. The individuals who aided the preparation of the datasets were also identified as co-authors in the relevant studies. All other data preparation were performed by me. In all cases, the original sources of the datasets were referenced throughout this thesis.

Chapter 1

Introduction

1.1 Background

Atrial fibrillation (AF) is the most common cardiac arrhythmia [1] affecting millions of people worldwide, and is associated with substantial morbidity and mortality [2]. AF contributes to one in five instances of stroke in people aged over 60 years [2] and increases the risk of heart failure three-fold [3]. The overall prevalence of AF is around 2% in developed countries and is projected to more than double in the following decades due to the aging population and increasing incidence of concomitant diseases [4].

AF is characterised by chaotic electrical activity in the atria, the upper chambers of the heart [5]. This causes the atria to contract uncoordinatedly, or “fibrillate”, leading to a rapid and irregular heartbeat. AF compromises the overall blood supply capacity of the heart to the rest of the body [6], and over time, greatly impacts the quality of life by causing physical distress, shortness of breath, and impaired exercise performance [7].

AF is a progressive disease. A majority of patients initially develop paroxysmal AF which lasts for less than seven days and terminates spontaneously [8]. If left untreated, the disease can progress to persistent or longstanding-persistent AF [9]. Persistent AF involves episodes lasting between 7 days and one year, while longstanding persistent AF lasts for over a year and is considered a chronic condition even with medical intervention [10]. Early detection of AF and early intervention for AF is important to effectively treat and minimise its impact. AF is problematic as it can manifest as a silent and asymptomatic condition, potentially going undetected for years. AF is often diagnosed incidentally during routine medical tests [11] and a stroke may be the first manifestation of AF [12]. Studies have further shown many patients progress from paroxysmal to persistent AF within one year [9, 13], making treatments increasingly ineffective as the severity of the disease rapidly progresses.

Current AF treatments consist of four main approaches: anticoagulation for stroke prevention, pharmacotherapy for rate or rhythm control, electrical cardioversion, and catheter ablation. Although pharmaceutical and cardioversion-based strategies are often the first

choice of treatment, the risk of recurrence remains high [14]. The recent consensus is that catheter ablation is the most effective for treating paroxysmal AF in drug-refractory patients and improves patient outcomes more effectively than medication [15, 16]. The ablation procedure involves “burning” or “freezing” localised tissue in the atria to eliminate AF driver regions which sustain AF [17]. However, the effectiveness of ablation remains poor for persistent and longstanding-persistent AF, with studies showing only a 30% long-term success rate [18, 19]. This is primarily due to the current lack of a basic understanding of the underlying anatomical structures which directly sustain AF [20, 21]. The complex structural alterations in the atria induced by persistent AF further increase the difficulty to accurately identify and target AF driver regions during ablation [22].

1.2 Motivations

Current clinical approaches for tackling AF remain suboptimal, particularly in terms of early detection and effective treatment. Specifically, there are three key gaps which need to be addressed to facilitate AF detection and treatment, and to minimise the risk of AF in order to have better long-term outcomes and increased quality of life for AF patients.

The first research gap is the lack of effective devices and approaches for AF detection. Electrocardiograms (ECG) are commonly used for the non-invasive diagnosis of AF in patients. Conventional assessment of AF utilises 12-lead ECGs which can only be performed in clinics and is inconvenient for patients. The assessment of ECGs then requires physicians to manually identify abnormal heartbeat patterns which is time consuming and subjective depending on levels of experience and expertise [23]. More importantly, the episodic nature of the disease may result in diagnostic delays since ECGs can appear normal between episodes. Recent innovations in portable ECG recording devices [24] could potentially facilitate population screening of AF. This includes the Apple Watch [25] (Apple Inc, California, United States) and KardiaBand/Mobile [26] (AliveCor Inc, California, United States) which aim to detect AF from single-lead ECG signals from the hand-held or wearable ECG sensors. However, the capability and reliability of current commercially available devices remain unclear with several validation studies reporting poor detection rates [25, 26]. As a result, there is an ongoing interest in developing a robust and validated approach for AF detection from portable ECG monitoring devices.

The second research gap involves the current lack of understanding of the atrial structures sustaining AF which prevents accurate identification of ablation targets for patients with persistent AF [20, 21]. Magnetic resonance imaging (MRI) is an effective and non-invasive method of assessing the cardiac structure. Late gadolinium-enhanced (LGE)-MRI, a type of contrast-enhanced MRI, is used to visualise atrial fibrosis [20, 27], the most important biomarker for AF perpetuation [28, 29]. Clinical trials have demonstrated that the degree of fibrosis in the left atrium (LA) is independently associated with ablation success rates and can be used for patient stratification [30]. LA chamber measurements and tissue thickness are also important indicators to aid the diagnosis and treatment of AF [31-33]. More importantly, there are currently no established studies for the analysis of the right atrium (RA) despite its widely accepted role in sustaining persistent AF [34]. In addition, the extraction of the atria and its patient-specific features from LGE-MRIs is extremely difficult due to the complexity of the atrial anatomy and limited imaging resolution. Most existing studies still rely on labour-intensive manual analysis [20, 27-29, 35], and attempts at using automated approaches have achieved limited efficacy [36-39]. Therefore, there is an urgent need for a robust approach to automate the analysis of the atria in medical imaging.

The third research gap relates to the ineffectiveness of current clinical mapping tools used to guide the ablation procedure, compromising the ability of the clinician to accurately target driver regions. Traditionally, catheter ablation has been performed with the aid of fluoroscopy imaging to guide the movement of the catheter inside the atrial chambers [40]. Recent advancements in electro-anatomical mapping (EnsiteNavX, St Jude Medical, Minnesota, United States and CARTO 3, Biosense Webster, California, United States) has allowed the 3D reconstruction and visualisation of the atrial structure from sparsely mapped coordinates recorded along the interior surface of the atria [41]. The 3D virtual models of the atria generated by the systems are used to aid clinicians during catheter manoeuvres. However, despite their wide usage in clinics globally, the accuracy of the atria geometry reconstructed using these proprietary systems is lower than desired [42]. In addition, the systems are also very time-consuming as they require hundreds of data points to be manually sampled for atrial reconstruction. Thus, there is an urgent need to develop an approach to reconstruct the atria more accurately and efficiently to guide the targeted ablation procedure.

1.3 Research Objectives

This thesis aims to develop deep learning-based algorithms for the following objectives to improve the current state of AF detection and treatment (**Figure 1.1**):

- **Objective 1:** Develop a robust AF classification approach for ECGs recorded from a portable sensor. The development is performed using the largest open-source single-lead ECG dataset recorded from the KardiaBand.
- **Objective 2:** Develop a fully automatic method of atrial structural analysis from LGE-MRIs to improve the visualisation and assessment of the patient-specific atrial anatomy for enhancing prognosis and patient stratification. The method is developed and validated using multi-clinical centre imaging containing the world's largest LGE-MRI datasets.
- **Objective 3:** Develop an efficient method to facilitate 3D atrial geometry reconstruction to guide the targeted ablation during the ablation procedure. The method is developed and validated on point clouds acquired from AF patients with anatomical mapping in conjunction with LGE-MRI and computed tomography (CT) data.

Deep learning is a category of modern machine learning algorithms and has been widely applied in the medical field in recent years [43-45]. The effectiveness of deep learning algorithms, particularly neural networks, lies in their ability to automate the feature generation and selection process which were the main bottlenecks of traditional learning-based algorithms [44]. This greatly increases the algorithm's learning efficiency with a dramatic reduction of human intervention as the entire process is fully data-driven, allowing for ease of adaptability to a wider range of tasks [43]. Because of this, deep learning has been the dominant driving force of artificial intelligence research in the past decade [46], surpassing human-level performance in many challenging tasks such as facial recognition [47], natural language processing [48], self-driving cars [49], and medical diagnosis [50, 51]. Furthermore, deep learning has been widely used for analysing biomedical datasets such as brain MRIs/CT [52], lung x-rays [53], liver MRIs [54], bone x-ray/CT [55], and most importantly cardiac MRI/CT/echocardiography [56-58]. Thus, deep learning provides the ideal foundation for tackling the challenging tasks currently present in the field.

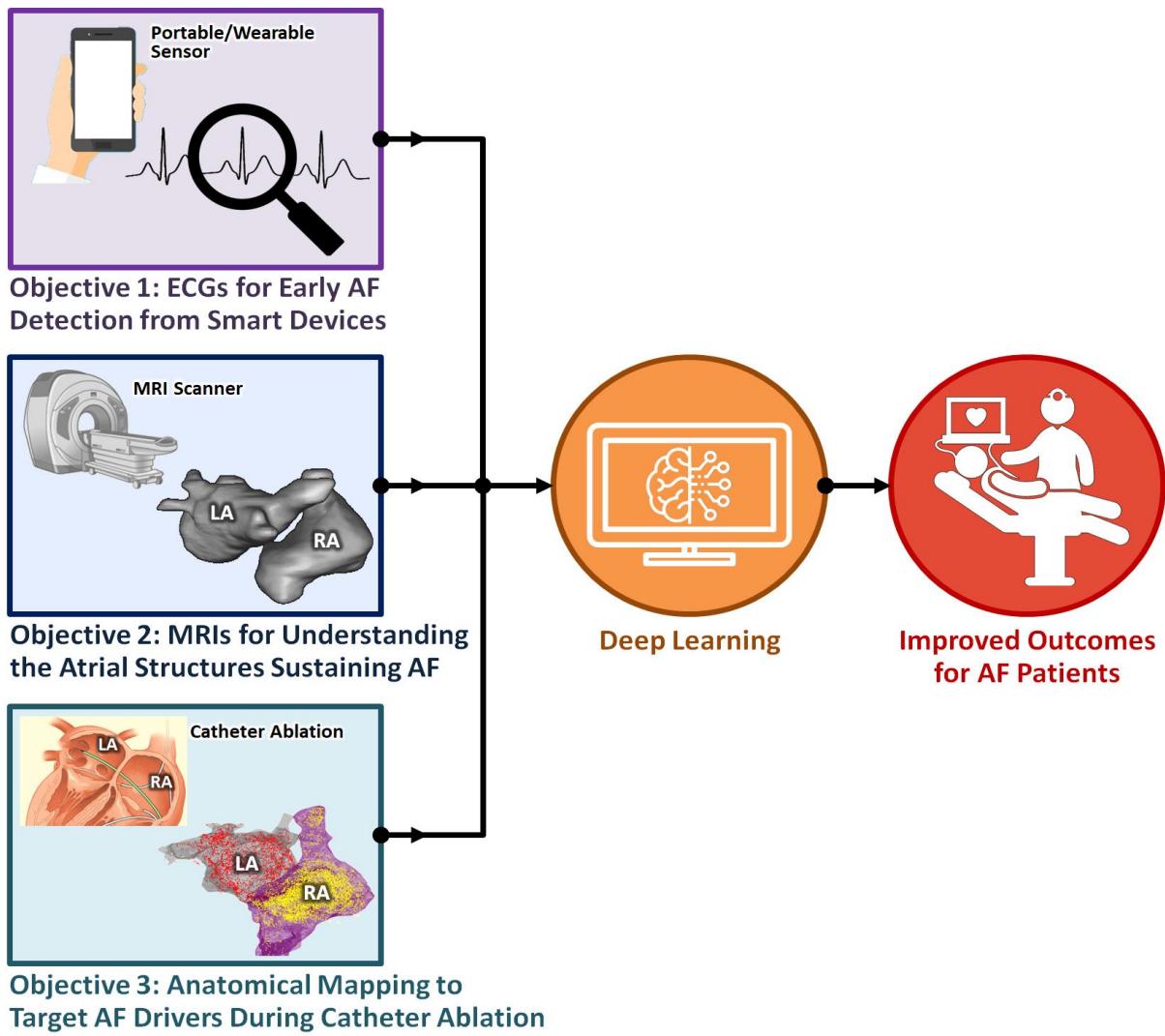


Figure 1.1. Schematic summary of the three main objectives investigated in this thesis. AF, atrial fibrillation; ECG, electrocardiogram; LA, left atrium; MRI, magnetic resonance imaging; RA, right atrium.

1.4 Thesis Outline

The research conducted in this thesis presents significant contributions to improving the detection, diagnosis, and treatment of AF by developing novel deep learning algorithms which are integrable into the clinical pipeline. Ultimately, the results derived from the studies performed will ideally lead to better patient outcomes for those suffering from AF. An illustration of the organizational flow of this thesis is shown in **Figure 1.2**.

Chapters 2 and 3 consist of the background and literature review for the thesis. Chapter 2 describes the cardiac and atrial anatomy, cardiac function, the cardiac activity during AF, and ECG to measure the electrical activity during normal and AF rhythms. The

chapter also details atrial remodelling during AF, with a particular focus on the mechanisms for atrial fibrosis formation during remodelling. The use of LGE-MRIs for visualizing atrial fibrosis is then described to outline its current clinical usage. The chapter lastly details the current AF treatments, catheter ablation, and mapping techniques used to guide ablation. Chapter 3 provides a comprehensive literature review of the current state of the field of deep learning, with particular focus on the application of deep learning for the tasks of classification, localization, and image segmentation. The chapter then focuses on the state-of-the-art methods of medical image segmentation to reconstruct anatomical structures. The chapter lastly describes, in detail, the widely recognized accuracy metrics for evaluating deep learning algorithms which are used extensively throughout this thesis.

To address Objective 1 of this thesis, Chapter 4 presents the development of an automated deep learning algorithm capable of diagnosing AF using ECG recordings from wearable sensors and mobile applications. The chapter further focuses on addressing the current issue of the lack of AF samples in ECG datasets by developing a neural style transfer (NST) generator to model and synthesise additional ECG samples for training.

Objective 2 of this thesis is addressed in Chapters 5 through 7. Chapter 5 describes a contemporary study of applying deep learning for direct and automatic LA segmentation from LGE-MRIs. Chapter 6 conducts a benchmarking study of the algorithms submitted to the 2018 Atrial Segmentation Challenge. The aim of the chapter is to derive the most optimal deep learning architecture for automatic LA segmentation. Chapter 7 further extends the findings from the benchmarking study to develop a deep learning algorithm capable of fully automated LA and RA analysis from both cine-MRIs and LGE-MRIs, providing the automatic derivation of the important pathological biomarkers used in the clinical treatment of AF. These consists of an accurate anatomical structure, degree of atrial dilatation, atrial fibrosis, and atrial wall thickness (AWT) distribution.

To address Objective 3 of this thesis, Chapter 8 presents the development of a deep learning algorithm to accurately reconstruct the 3D LA using sparse point clouds anatomically mapped during catheter ablation. The chapter proposes a novel method of discretizing the point clouds into dense image volumes which will then be effectively processed with deep learning. The deep learning algorithm is trained on synthetic data clouds augmented with LGE-MRI segmentations and validated on clinical anatomical maps from multiple centres.

Chapter 9, finally, provides a summary of the thesis, potential directions for future research, and concluding remarks on the thesis.

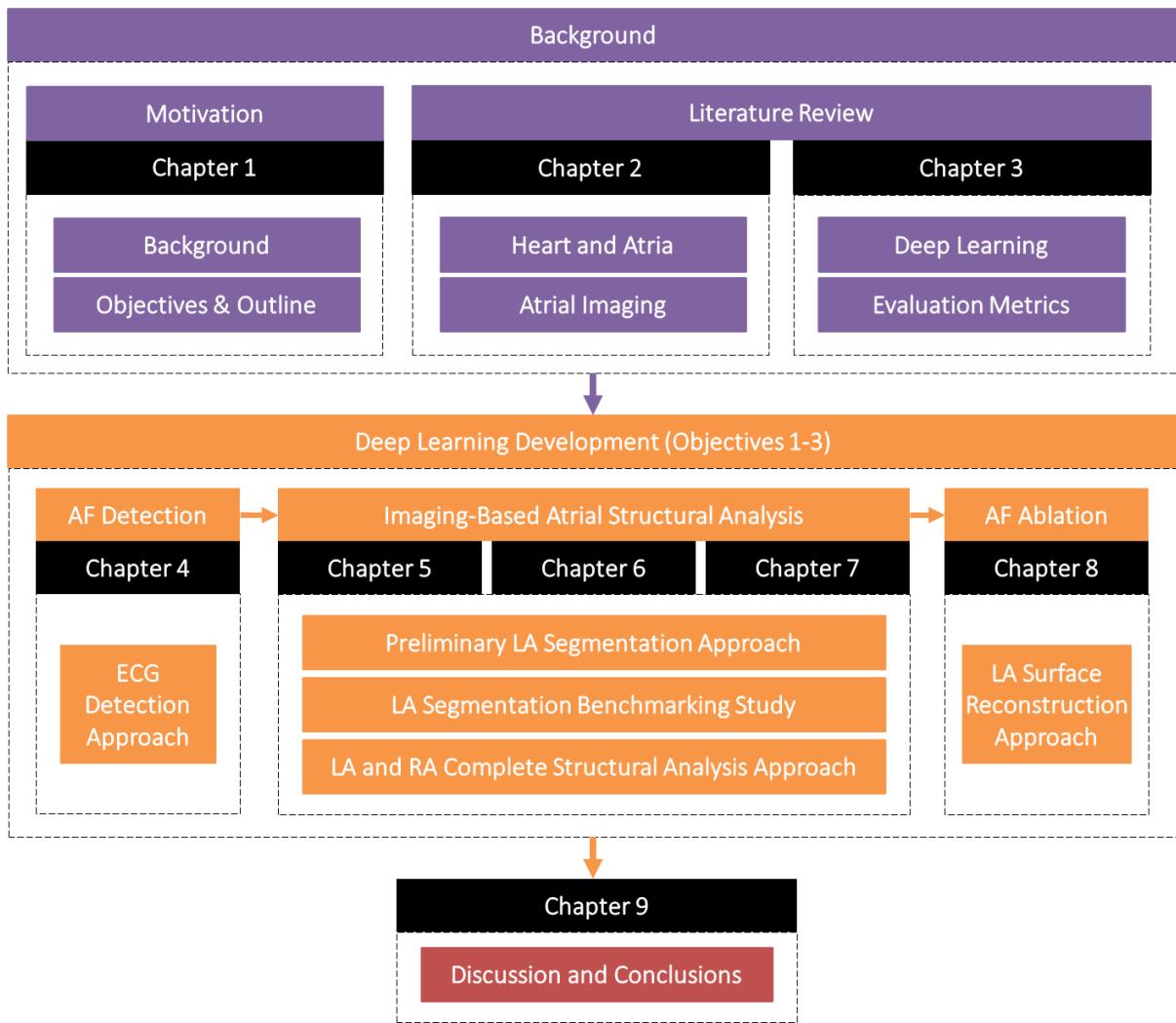


Figure 1.2. The organization overview of this thesis. ECG, electrocardiogram; LA, left atrium; RA, right atrium.

1.5 Publications

The main work presented in this thesis has been published, or is currently under review, in international conference proceedings or journal publications. The list of first-author publications and proceedings is as follows:

1. **Xiong, Z.**, Fedorov, V.V., Fu, X., Cheng, E., Macleod, R. and Zhao, J., 2019. Fully Automatic Left Atrium Segmentation from Late Gadolinium Enhanced Magnetic Resonance Imaging using a Dual Fully Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, 38 (2), pp.515-524.
2. **Xiong, Z.**, Nalar, A., Stiles, M.K., Fedorov, V.V. and Zhao, J., 2019. Fully Automatic 3D Whole Atria Segmentation and Reconstruction from Gadolinium-Enhanced MRIs

Using Convolutional Neural Networks. *The Annual Scientific Meetings of Heart Rhythm Society*.

3. **Xiong, Z.**, Nalar, A., Jamart, K., Stiles, M.K., Fedorov, V.V. and Zhao, J., 2019. Fully Automatic 3D Bi-Atria Segmentation from Late Gadolinium-Enhanced MRIs Using Double Convolutional Neural Networks. *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 63-71.
4. **Xiong, Z.**, Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., Heng, P.A., Ni, D., Li, C., Tong, Q., Si, W., Puybareau, E., Khoudli, Y., Géraud, T., Chen, C., Bai, W., Rueckert, D., Xu, L., Zhuang, X., Luo, X., Jia, S., Sermesant, M., Liu, Y., Wang, K., Borra, D., Masci, A., Corsi, C., de Vente, C., Veta, M., Karim, R., Preetha, C., Engelhardt, S., Qiao, M., Wang, Y., Tao, Q., Nuñez-Garcia, M., Camara, O., Savioli, N., Lamata, P., Zhao, J., 2021. A Global Benchmark of Algorithms for Segmenting the Left Atrium from Late Gadolinium-Enhanced Cardiac Magnetic Resonance Imaging. *Medical Image Analysis*, 67, 101832.
5. **Xiong, Z.**, Nalar, A., Kennelly, J., Petersen, S., Stiles, M.K. and Zhao, J., 2022. Automatic Extraction of Pathologic Biomarkers from Cardiac Magnetic Resonance Imaging Using Convolutional Neural Networks. (Under Review)
6. **Xiong, Z.**, Stiles, M.K., Yao, Y., Shi, R., Nalar, A., Hawson, J., Lee, G. and Zhao, J., 2022. Automatic 3D Surface Reconstruction of the Left Atrium from Clinically Mapped Point Clouds Using Convolutional Neural Networks. (Under Review)
7. **Xiong, Z.**, Stiles, M.K., Gillis, A.M. and Zhao, J., 2022. Enhancing the Detection of Atrial Fibrillation from Wearable Sensors with Neural Style Transfer and convolutional Recurrent Networks. (Under Review)

Contributions have also been made during the period of this doctoral study as the second author in the following publications:

1. Wang, Y., **Xiong, Z.**, Nalar, A., Hansen, B.J., Kharche, S., Seemann, G., Loewe, A., Fedorov, V.V. and Zhao, J., 2019. A Robust Computational Framework for Estimating 3D Bi-Atrial Chamber Wall Thickness. *Computers in Biology and Medicine*, 114, p.103444.
2. Jamart, K., **Xiong, Z.**, Talou, G.M., Stiles, M.K. and Zhao, J., 2019. Two-Stage 2D CNN for Automatic Atrial Segmentation from LGE-MRIs. *International Workshop on Statistical Atlases and Computational Models of the Heart*, pp. 63-71.

3. Jamart, K., **Xiong, Z.**, Maso Talou, G.D., Stiles, M.K. and Zhao, J., 2020. Mini Review: Deep Learning for Atrial Segmentation from Late Gadolinium-Enhanced MRIs. *Frontiers in Cardiovascular Medicine*, 7, p.86.

The work was presented at the following domestic and international conferences:

1. Poster presentation in the 2019 *Statistical Atlases and Computational Modelling of the Heart* workshop held in the *Medical Image Computing and Computer Assisted Intervention* international conference, Shenzhen, China.
2. Poster presentation in the 2019 *Annual Heart Rhythm Society Scientific Sessions*, San Francisco, California, United States.
3. Poster presentation in the 2019 *MedTech CoRE Annual Conference*, Auckland, New Zealand.
4. Poster presentation in the 2019 *Auckland Bioengineering Institute Annual Research Forum*, Auckland, New Zealand.

Chapter 2

Background

2.1 The Heart

2.1.1 The Structure and Function of the Heart

The heart consists of four chambers (**Figure 2.1**). The right heart contains the RA and the right ventricle (RV), and the left heart contains the LA and left ventricle (LV). The left and right sides of the heart are separated by a continuous partition consisting of interatrial septa which separate the LA and RA, and interventricular septa which separates the LV and RV. The atria are also separated from the ventricles by the atrioventricular septa. Blood flow then occurs from the atria to the ventricles through the two atrioventricular orifices which are openings in the atrioventricular septa. The openings are controlled by two atrioventricular valves which periodically open and close during each heartbeat. The valve between the RA and RV is known as the tricuspid valve, and the valve between the LA and LV is known as the mitral valve. The LA and RV are both connected to the lungs via the pulmonary veins (PV) and pulmonary arteries, respectively. The LV and RA both connect to the rest of the body via the aorta and superior/inferior vena cava, respectively. The pulmonary artery and aorta also contain valves to prevent the back flow of blood into the RV and LV.

The heart functions to efficiently pump blood around the body through a series of alternating contractions and relaxations of the atria and ventricles [10]. This process is known as the cardiac cycle and occurs in every heartbeat. The cardiac cycle begins with atrial diastole. This process involves the relaxation of the atrial chambers to allow deoxygenated blood to enter the RA through the vena cavae and oxygenated blood to enter the LA through the PVs. The atria then undergo systole which is the contraction of the atrial chambers to push blood through the atrioventricular valves and into the ventricles. During the diastolic and systolic function of the atria, the ventricles are in a state of diastole to allow the pooling of blood from the atria. The ventricles then undergo systole to eject blood to the rest of the body from the LV via the aorta and to the lungs from the RV via the pulmonary arteries.

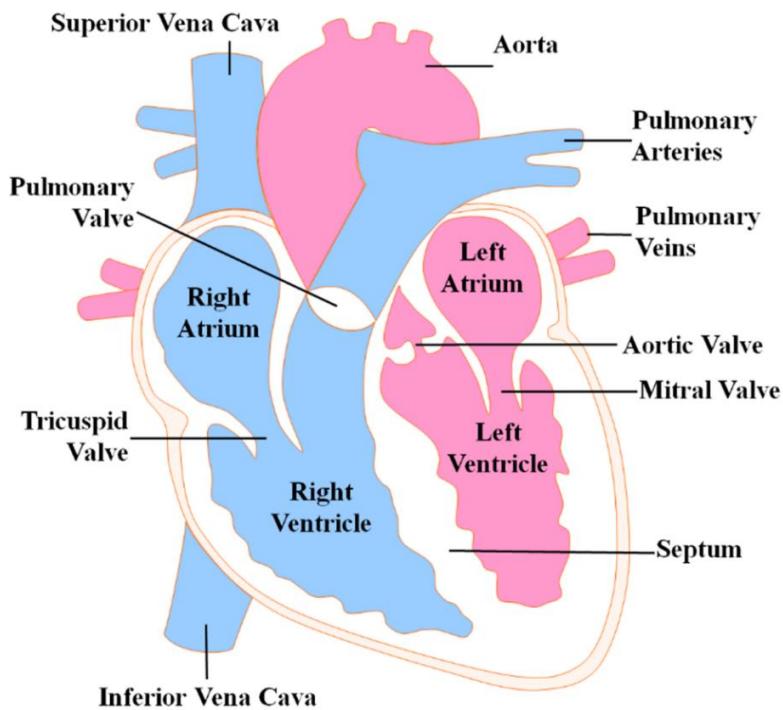


Figure 2.1. Illustration of the detailed anatomy of the heart showing the four chambers, valves, major arteries and veins.

2.1.2 The Structure of the Atria

The atria consist of two chambers (**Figure 2.2**), the LA and the RA on the left and right upper sides of the heart respectively and above the two ventricles. While the two atrial chambers differ significantly in shape and have their own respective identifiable features, they also share the same basic structures including a venous component, an appendage, and a vestibule that leads to the atrioventricular valves. The entrances of the atrial chambers are delineated by the veno-atrial junction, with the two entrances of the vena cavae defining the RA and the PV defining the LA. The atrial chambers then end at an adipose tissue plane that delineates the atrioventricular junction.

The RA contains a large appendage, known as the right atrial appendage, which is triangular in shape projecting anteriorly from the RA chamber. The RA also contains the most visible muscular bundle on the RA endocardium, known as the crista terminalis, which separates the smooth wall of the venous portion and the rough wall of the appendage. The crista terminalis extends from the entrance of the inferior vena cava up to the entrance of the superior vena cava. The sinoatrial node, a group of specialized cells responsible for firing

electrical impulses to stimulate cardiac contractions, is also located close to the crista terminalis and the superior vena cava. The coronary sinus can be found on the RA close to the septum, and is a collection of veins joined together to form a large vessel that collects blood from the heart muscle and delivers less-oxygenated blood to the RA. The atrioventricular node is situated anteriorly to the coronary sinus, and is responsible for electrically connecting the atria and ventricles after firing impulses from the sinoatrial node. Finally, the tricuspid valve separating the RA and RV is composed of three leaflets of fibrous tissue.

The LA also contains an appendage which is distinctly smaller than that in the RA. While the shape of the LA appendage varies depending on patients, the general shape tends to be tubular with one or several bends. The LA endocardial walls are relatively smooth, and contain a network of fine muscular ridges which produce considerable variations in the LA wall thickness. The venous component of the LA consists of four PVs which enter into the chamber posteriorly, two of which lead from the right lung and two of which lead from the left lung. Although four is the most common number of PVs, anatomical variations can result in only three PVs being visible, or two or three PVs leading from the LA which then branch out into four. Finally, the mitral valve separating the LA and LV is composed of two leaflets of fibrous tissue.

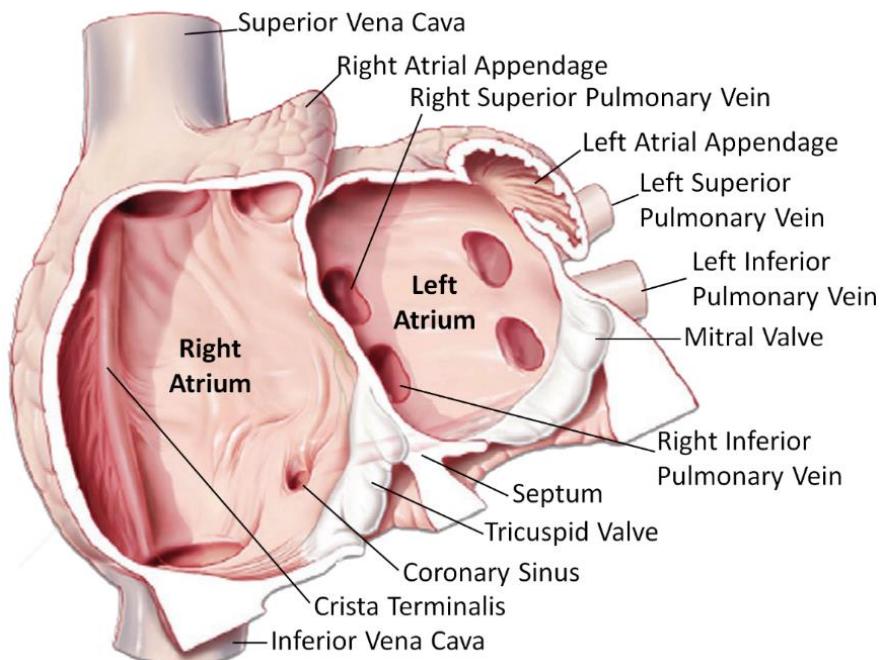
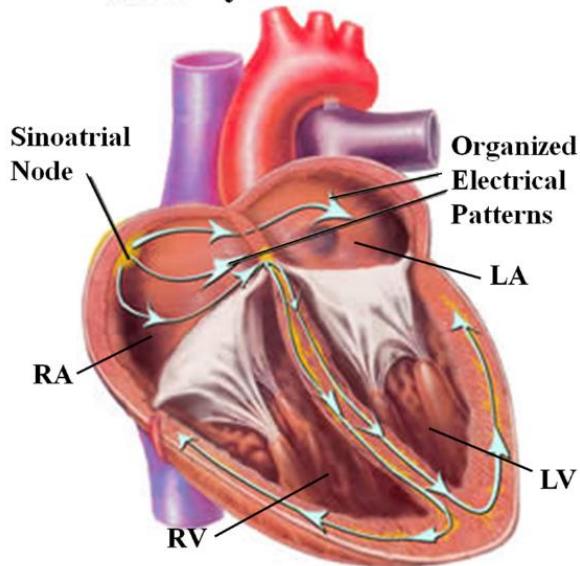


Figure 2.2. The anatomical structure and components of the left and right atrial chambers, with the valves, major arteries and veins adapted from Lee et al. [59].

2.1.3 Atrial Fibrillation

A. Sinus Rhythm



B. Atrial Fibrillation

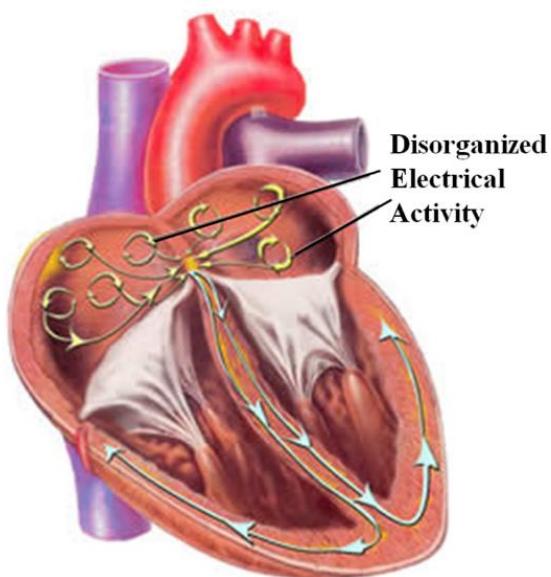


Figure 2.3. Visualization of the electrical propagation patterns present in the heart during **A)** sinus rhythm and **B)** atrial fibrillation (AF) [61]. The electrical waves are initialized from the sinoatrial node and travel throughout the atrial chambers. During AF, these waves are disturbed and become disorganized.

In a normal healthy heart, the cardiac cycle is regulated by an electrical impulse generated by the sinoatrial node. The electrical activations propagate throughout the atria to cause coordinated contractions resulting in normal sinus rhythm (SR) heartbeats [60]. The electrical pulse then travels to the ventricle through the atrioventricular node. To allow sufficient time for the atria to completely empty during each cycle, the atrioventricular node slows down the electrical impulses before the ventricular contraction. When working in harmony, these processes enable the heart to function efficiently.

During AF, the regular electrical impulses from the sinoatrial node are disrupted by foreign electrical patterns which results in chaotic and disorganized electrical activities that lead AF to quiver [5] (**Figure 2.3**). This causes the atria to contract uncoordinatedly, or “fibrillate”, such that they can no longer effectively eject blood into the ventricles which ultimately compromises the blood supply to the rest of the body. While normal resting heartrates are between 60 to 100 beats per minute, patients with AF may experience resting heart rates of 100 to 200 beats per minute.

AF can generally be classified into several categories. Paroxysmal AF is defined as

having episodes lasting less than 7 days. Persistent AF involves episodes lasting anywhere from 7 days to one year, while longstanding-persistent AF involves continuous AF lasting for over a year. Finally, a patient is categorized to have permanent AF if the physician has decided not to pursue the restoration of SR. However, permanent AF is subjective amongst different physicians as the exact differences between longstanding-persistent and permanent AF remains difficult to define.

2.1.4 Electrocardiogram

The synchronous contraction in the heart is regulated by electrical activations which can be visualized using an ECG [62], a body-surface recording which reflects the electrical activity in the heart. Each complete cycle of an ECG represents one heartbeat (**Figure 2.4**). During SR, the ECG cycle consists of a P-wave, QRS-complex, and T-wave. Each cycle begins with the P-wave representing the depolarization of the atria. It is followed by the QRS-complex which represents the depolarization of the ventricles and appears as three closely related waves (Q, R, and S wave). Finally, the T-wave represents the repolarization of the ventricles. The time-gap between the P and R waves is known as the PR interval, and represents the time taken for the electrical signals to travel between the atrial and ventricular chambers. Similarly, the time-gap between the S and T waves, or ST interval, is the time interval between ventricular depolarization and repolarization. The R-peak-to-R-peak (RR) interval is the time gap between the peak of the R wave in one ECG cycle and the peak of the R wave in the following ECG cycle. In basic terms, the RR interval represents the time between two heartbeats and can be used to calculate the heart rate. In ECG signals of SR recordings, the consecutive RR intervals are approximately the same length as the result of the regular beating of the heart.

AF can be diagnosed from an ECG recording by identifying a number of features. A distinct feature is the absence of the P-wave on the ECG which is caused by the rapid and irregular atrial activity (**Figure 2.5**). Instead, low-amplitude and irregular “fibrillatory” waves are seen between the QRS complexes to reflect the disordered atrial depolarization. Irregular and relatively shorter RR intervals can also be seen in ECGs containing episodes of AF. This is due to the ventricular response during the cardiac cycle being affected by the irregular firing activity in the atria which is then variably conducted through the atrioventricular node.

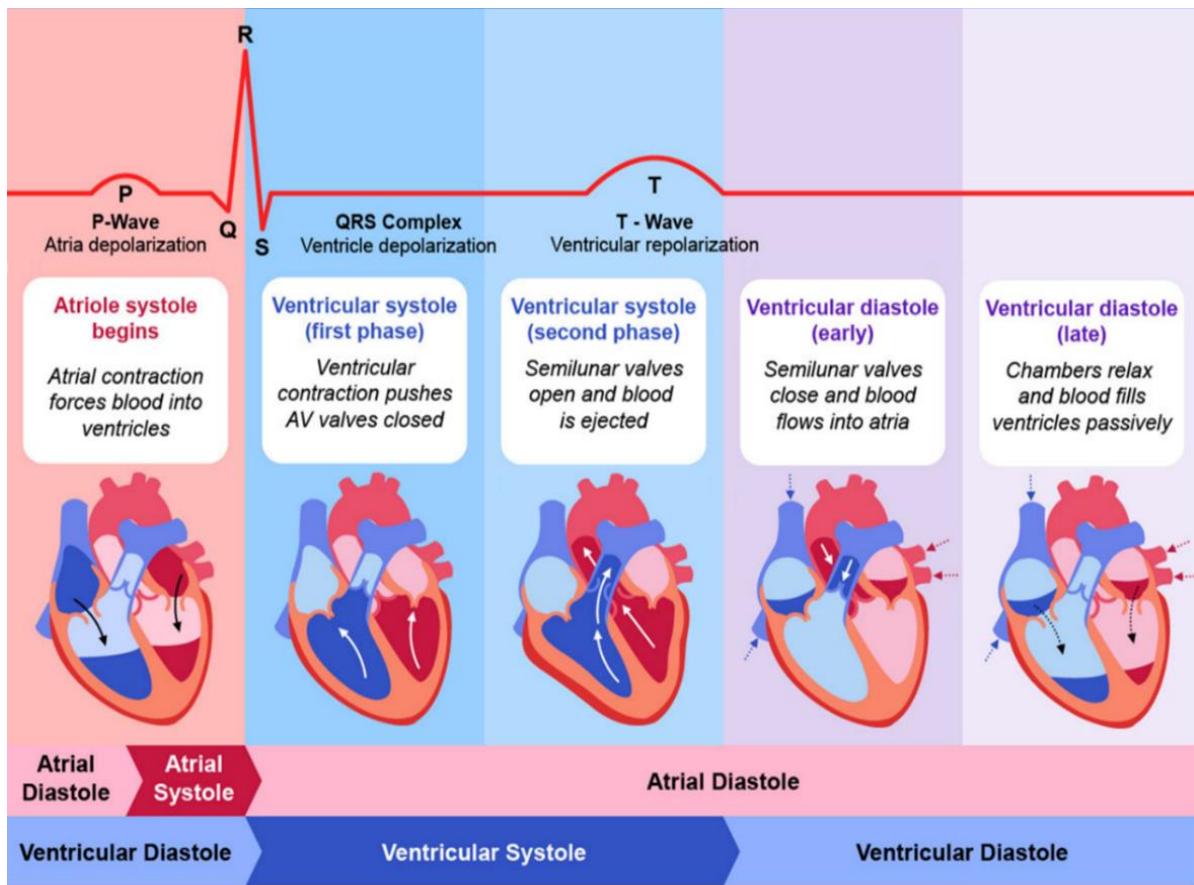


Figure 2.4. Illustration of a single cycle of an electrocardiogram (ECG) along with the corresponding cardiac functions throughout the cardiac cycle [63]. The heart undergoes a series of synchronized contractions, or systole, and relaxations, or diastole, to efficiently pump blood around the body. These muscular activities are regulated by electrical signals, which can be seen from an ECG, to coordinate the atrial and ventricular functions.

A. Electrocardiogram of Sinus Rhythm



B. Electrocardiogram of Atrial Fibrillation

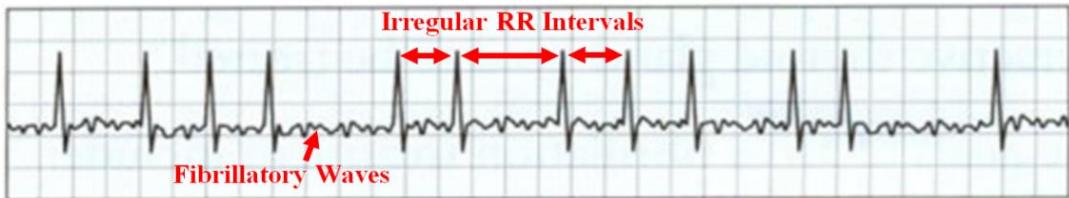


Figure 2.5. Illustration and comparison of typical electrocardiogram (ECG) recordings of A)

sinus rhythm, and **B)** atrial fibrillation (AF) [64]. AF can be characterised from the irregular R-peak-to-R-peak (RR) intervals and the presence of low-amplitude fibrillatory waves.

2.2 Structural Remodelling of the Atria

2.2.1 Atrial Remodelling

The progressive nature of AF is generally concordant with atrial functional and structural changes, known as atrial electrical and structural remodelling [65]. The pathophysiological mechanisms responsible for atrial remodelling can result from cardiac diseases conditions which promote the development of AF. As AF develops, it then causes AF-promoting abnormalities which further enhance the vulnerability of the heart to AF induction and maintenance. The electrical and structural alterations as a result of remodelling cause conduction disturbances which lead to an increased propensity of AF triggers such as re-entry-prone substrates [66]. This auto-reinforcing and self-perpetuating property of AF is often referred to by the term “*AF begets AF*” [67].

Electrical remodelling involves a persistent change in the electrophysiological properties of the atria in response to a change in the sequence of electrical activation. The atrial electrophysiological properties are governed by ion channels, pumps and exchanges in the cardiac cell membrane, any of which can be altered by remodelling. The main components of electrical remodelling include changes in the ionic currents and abnormal distribution of the gap junction connexin hemi-channels that connect cardiomyocytes electrically. In cardiomyocytes, the sodium (Na^+), potassium (K^+), and calcium (Ca^{2+}) channels regulate the inflow and outflow of $\text{Na}^+/\text{K}^+/\text{Ca}^{2+}$ ions to generate action potentials which cause cell contraction. Gap junction ion channels, such as the connexin 40, found mainly in the atria and conduction system, mediate the electrical coupling of action potentials between neighbouring cardiomyocytes [68]. The changes in the ionic currents mainly involve the down-regulation of L-type Ca^{2+} currents [69], up-regulation of rectifier background K^+ currents and constitutive acetylcholine-regulated K^+ currents [70], and reduction in Na^+ current [71]. The down-regulation of the L-type Ca^{2+} current is caused by an increase in intracellular Ca^{2+} due to persistent tachycardia and results in a decrease in action potential duration. The repolarizing abnormalities due to the changes in the K^+ currents contribute to the shortening of atrial refractoriness. The reduced Na^+ current also causes a reduction in the

conduction velocity and shortening of the re-entry wavelength. Alterations in the gap function such as down-regulation and increasingly heterogeneous distribution may further contribute to AF-induced remodelling [72].

Structural remodelling of the atria refers to the adaptive or maladaptive changes in the cardiac cellular architecture in response to external stressors. Such stressors may include age and genetics, as well as concomitant diseases such as heart failure, hypertension, diabetes, and obesity. Remodelling of the atrial structure occurs at both organ and tissue-levels and are time and aetiology dependent. Atrial enlargement is the main characteristic of organ-level remodelling [73], where studies have shown the atrial dimension to be a key determinant of the persistence of AF-maintaining re-entry under certain functional conditions [74]. Significant changes at the tissue-level include cellular hypertrophy, myolysis, dedifferentiation, fibrosis, apoptosis, mitochondria and sarcoplasmic reticulum disruption. In particular, fibrosis is widely regarded as the hallmark of AF [75] as it appears to be a common endpoint of a wide range of AF-promoting conditions, and have been shown to potentially predict recurrences [28]. Fibrosis promotes AF by interrupting fibre bundle continuity, reducing intercellular connections, and causing local conduction disturbances [76]. The formation of fibrotic tissue also contributes to significant functional changes due to its lack of contractive capacity and expansion. The increased stiffness of the atrial wall, as a result of fibrosis development, reduces myocardial compliance and impairs cardiac contraction [77]. Studies have shown fibrotic regions experience a reduction in the oxygen supply which may cause cellular death [78]. As the formation of fibrosis is initially a reparatory mechanism in response to tissue injury or cardiomyocyte death, the death of more cardiomyocytes in fibrotic regions may further perpetuate fibrosis formation [79].

2.2.2 Atrial Fibrosis

The formation of fibrosis is characterized by an overproduction of the extracellular matrix which causes the diseased expansion of connective tissue [47]. The two identified fibrosis types, reparative fibrosis which replaces dead cardiomyocytes and reactive interstitial fibrosis which separates muscle bundles, both interfere with the electrophysiology of the atria [80]. While the mechanisms regulating extracellular matrix remodelling still remain incompletely understood, several secreted factors are known to be pro-fibrotic [81]. These include angiotensin II, transforming growth factor- β_1 , platelet-derived growth factor, and

connective tissue growth factor which have both individual and combined pro-fibrotic effects [66].

Angiotensin II acts by binding angiotensin type I and type II receptors. The type I receptor brings about the pro-fibrotic effects of angiotensin II by stimulating connective tissue growth (fibroblast proliferation), cardiomyocyte hypertrophy, and programmed cell death (apoptosis) [75]. Angiotensin type I activation also stimulates phospholipase C. Phospholipase C is an enzyme which breaks down PIP₂, a signalling phospholipid protein, into diacylglycerol and IP₃, an inositol phosphate signalling molecule. Diacylglycerol activates protein kinase C which is an enzyme responsible for regulating cardiac contractility, pathophysiological growth responses, and pump function [82]. IP₃ causes intracellular Ca²⁺ release. Diacylglycerol combined with IP₃ both induce remodelling [83]. On the other hand, Angiotensin type II activation inhibits mitogen-activated protein kinases and produces anti-proliferative effects that oppose angiotensin type I changes [84]. The balance, or off-balance, of the two counter-regulatory receptors, may therefore have important implications. Cardiomyocytes and fibroblasts both secrete transforming growth factor- β_1 which acts as the primary mediator of angiotensin II. This affects both angiotensin II-producing cells (autocrine) and neighbouring cells (paracrine). Angiotensin II also reciprocally induces transforming growth factor- β_1 synthesis, enhancing pro-fibrotic factors and creating positive feedback [85]. These factors stimulate fibroblast activation and collagen deposition [86], and result in selective atrial fibrosis, atrial conduction heterogeneity, and AF promotion [87].

Platelet-derived growth factors stimulate fibroblast proliferation and differentiation. Fibroblast differentiation occurs when fibroblasts undergo phenotype switching to a more proliferative state, known as myofibroblasts [88]. Platelet-derived growth factors also undergo auto-phosphorylation, a process in which its receptors are self-activated through the activation of the tyrosine kinase enzyme, to initiate signalling via mitogen-activated protein kinases and phospholipase C pathways similar to transforming growth factor- β_1 and angiotensin II. Interestingly, platelet-derived growth factors also appear to be the underlying factor for the fibroblast hyper-responsiveness selectively in the atria, causing the atria to be more susceptible to fibrotic remodelling than the ventricles [89]. Connective tissue growth factor contributes similarly to remodelling as it is a major downstream effector of, and coordinates with, transforming growth factor- β_1 in regions with active fibrosis promotion [90]. Connective tissue growth factor directly activates fibroblasts, and is also upregulated by both angiotensin II and transforming growth factor- β_1 .

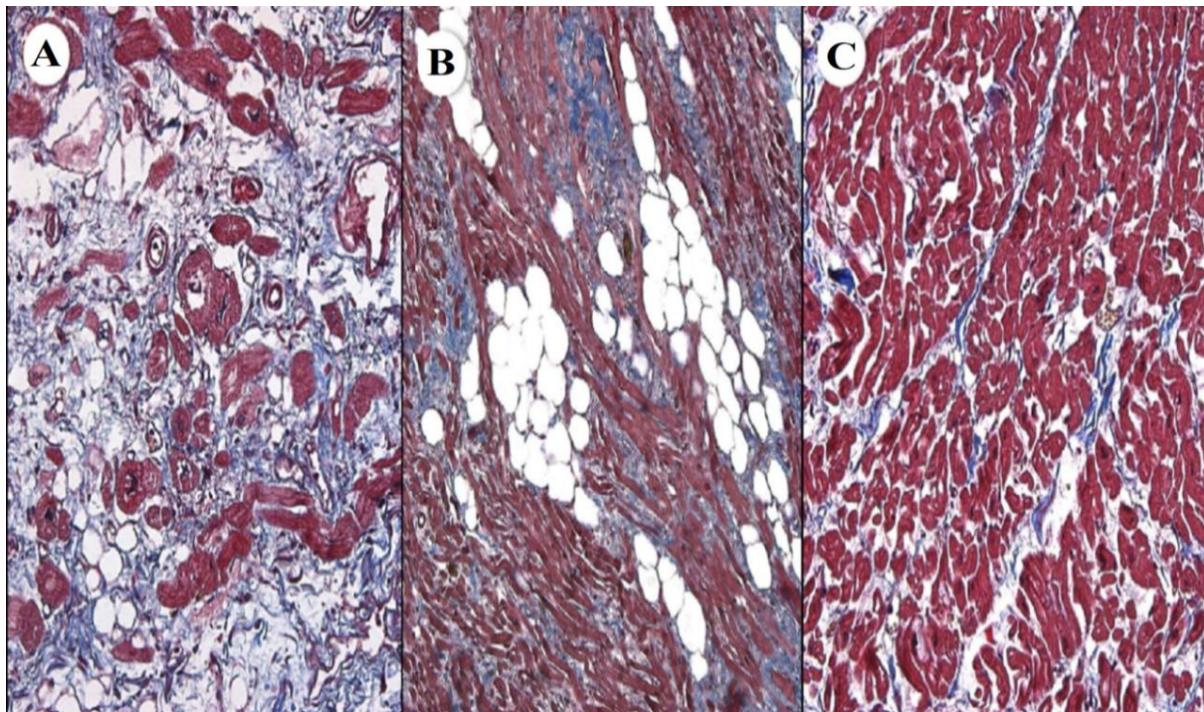


Figure 2.6. Light microscopy imaging with Masson's trichrome stain of the crista terminalis in the right atrium for observing the degree of fibrosis (red) in patients with varying severities of atrial fibrillation (AF) adapted from Platonov et al. [91]. **A)** 51% of fibrosis extent in a patient with permanent AF. **B)** 14% of fibrosis extent in a patient with paroxysmal AF. **C)** 5% of fibrosis extent in a patient without AF.

Visualization and assessment of fibrosis development are important for assessing the progression of atrial remodelling. Histological microscopy is a gold-standard laboratory imaging technique for fibrosis identification from dissected tissue samples (**Figure 2.6**), and have extensively been utilized in experiments involving both animals [92] and human tissue [93]. The protocol for microscopy imaging involves extensive tissue sample preparation before imaging can be performed. This includes staining to enhance cellular features, fixation to preserve the tissue structure, dehydration to remove water, embedding to allow for easier extraction of cellular structures, sectioning for the purpose of mounting on a microscopy slide, and antigen retrieval to further enhance staining [94]. Further steps are then performed to tune the imaging set up for the optimal image quality. However, the time-consuming and destructive nature of the preparation of histological imaging for patients with AF is an impractical method for medical assessment in a clinical setting. Furthermore, the locally extracted atrial tissue may not be representative of the distribution of fibrosis across the entire atria on a macroscopic scale. Due to these issues, LGE-MRI are used instead as it enables the

non-invasive visualization of the extent and distribution of fibrosis on an organ level. This allows for a more direct investigation of the patient-specific changes caused by AF.

2.2.3 Late Gadolinium-Enhanced Magnetic Resonance Imaging

MRI has become a popular imaging modality due to its relatively high image quality, sharp soft tissue contrast, and non-invasive nature in the absence of ionizing radiation. While regular MRI enables the visualization of various atrial structures, additional diagnostic information can be acquired through the use of contrast agents during the imaging procedure. The most commonly used compounds for contrast enhancement are gadolinium-based contrast agencies [95].

In recent years, gadolinium-based contrast agents have been used in approximately a third of all MRI scans to improve the visibility of disease-associated structures, such as fibrosis/scarring, inflammation, tumours, and cancerous tissue, as they often contain increased vascularity compared to surrounding healthy tissue, making them appear brighter after gadolinium enhancement [96]. Contrast-enhanced MRIs also result in images with higher anatomical clarity, allowing for an increase in the image quality and degree of detail visible during clinical assessment [97]. Delayed gadolinium enhancement, or late gadolinium enhancement, is a technique where MRI acquisition is performed after a delay period following intravenous injection of gadolinium. Tissue enhancements in the final LGE-MRI, therefore, indicate cell death as the gadolinium agent is retained within the abnormal tissue [98].

LGE-MRI has been widely established for its use in myocardial tissue characterization such as the detection of myocardial fibrosis for diagnosing myocardial infarction. There has also been a growing interest in the use of LGE-MRI for atrial fibrosis assessment to guide and improve the success rate of ablation [28]. However, the use of conventional imaging for the atria severely is limited by its low spatial resolution (1.6 – 1.8 mm) which is insufficient for capturing the thin-walled chambers such as the LA with an average of 2 – 4 mm [99]. The development of high-resolution LGE-MRI overcomes this limitation by utilizing a free-breathing technique which significantly increases the spatial resolution (**Figure 2.7**). In traditional LGE-MRIs, the spatial resolution is limited by the breath-hold durations of the patient as breath-holding is required to reduce image artefacts [100]. Free-breathing LGE-MRI eliminates this issue with respiratory-gated acquisition to

track and correct the breathing motion [101]. Another advantage of high-resolution LGE-MRI is the whole-heart coverage that offers the ability to reconstruct any desired orientation, while conventional LGE-MRI is limited to the prescribed single axis [102].

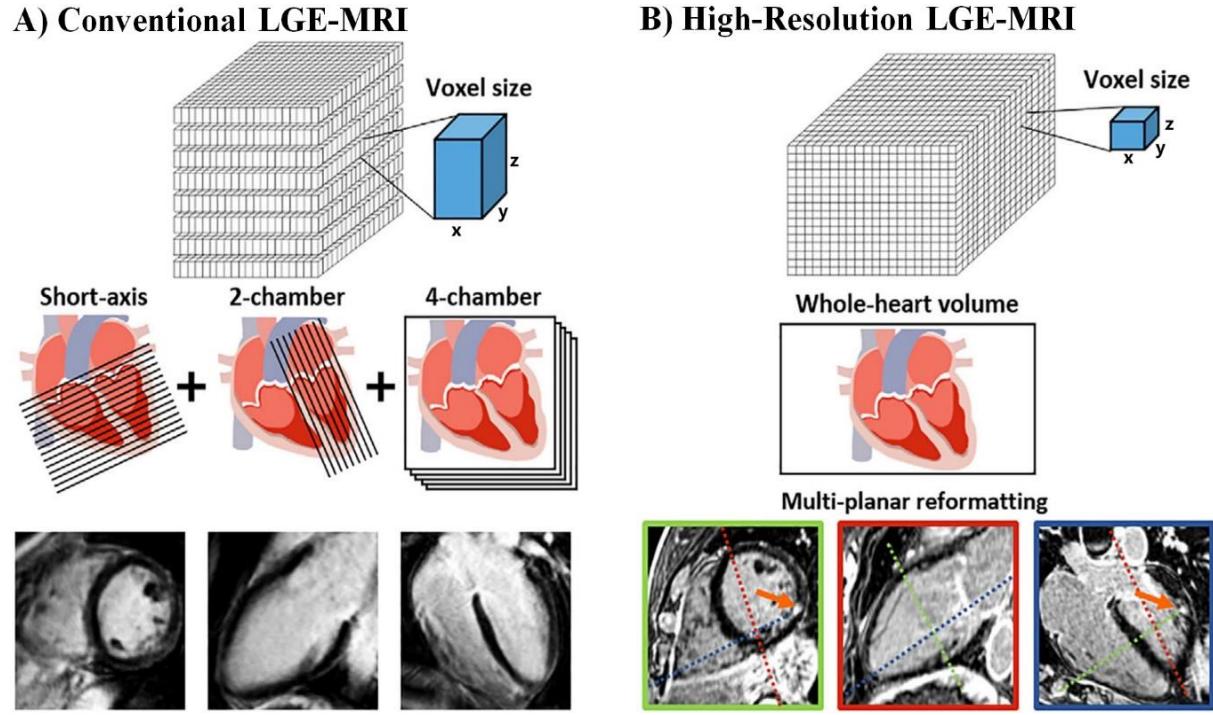


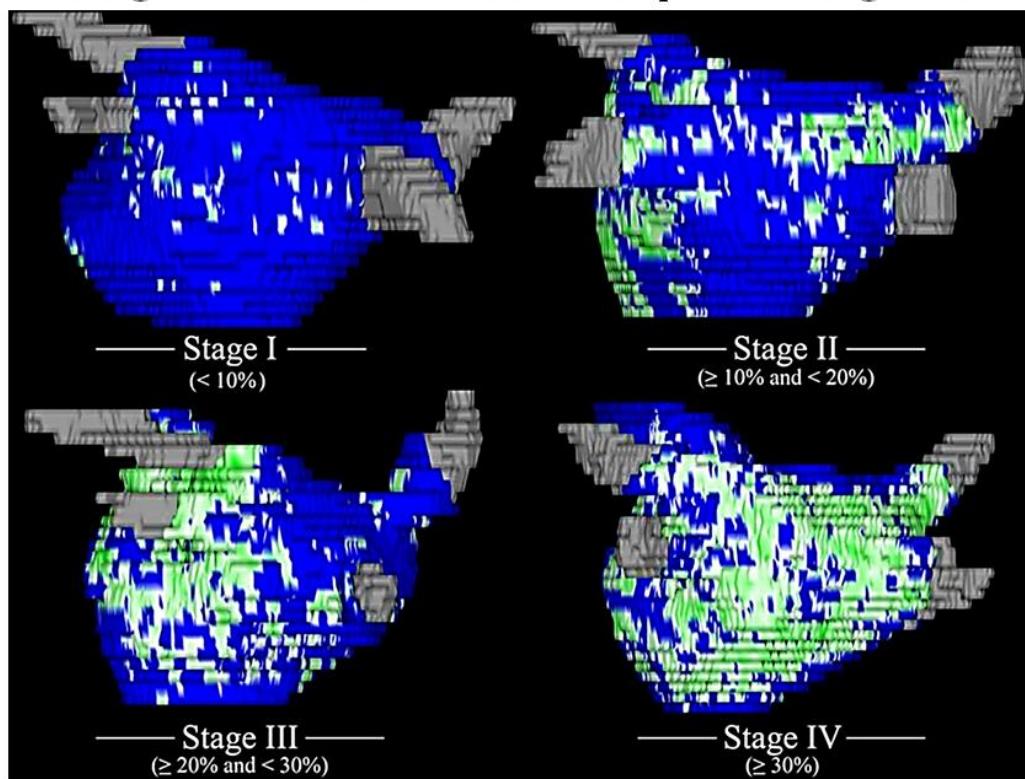
Figure 2.7. Illustration of the difference in imaging resolution and isotropy between conventional and high-resolution late gadolinium-enhanced magnetic resonance imaging (LGE-MRI) adapted from Toupin et al. [103]. Conventional LGE-MRI is acquired at short-axis, two-chamber, and four-chamber views with a moderate spatial resolution and decreased resolution in the z -axis. High-resolution LGE-MRI is acquired for the entire image volume of the heart, allowing the multi-planar slicing of the 3D volume along any arbitrary axis.

High-resolution LGE-MRIs have gained significant momentum in the field of AF, as atrial LGE-MRI is being performed predominantly using this technique. The multicentre, prospective, observational cohort study, Delayed Enhancement MRI-Guided Ablation vs Conventional Catheter Ablation of AF (DECAAF), led by Marrouche et al. from the University of Utah, United States, have demonstrated that the amount of LA fibrosis using LGE-MRI is independently associated with catheter ablation success rate in patients with AF and can be used for patient stratification [30]. Based on this important finding, the group introduced a staging approach to quantify the extent of LA fibrosis as a percentage of the

atrial wall for patient selection of AF trigger ablation: stage I (<10%), II (10%-20%), III (20%-30%), IV ($\geq 30\%$) [28, 29] (**Figure 2.8**). Results after a 1-year follow up showed patients at early stages (I and II) appear to have less structural remodelling and more favourable long-term ablation success rates. On the other hand, patients at later stages (IV) were considered for more conservative treatment methods due to the high risk of AF recurrence. A more recent clinical trial (DECAAF II) involving over 45 clinical centres around the world evaluates conventional catheter ablation vs. catheter ablation guided by LGE-MRI-extracted fibrosis. The study aims to explore the use of atrial fibrosis to improve the definition of catheter ablation target sites [104].

The LA diameter and volume computed from 3D LGE-MRIs have also been shown to provide reliable information for clinical diagnosis and treatment stratification [31-33]. Studies have shown an increased susceptibility of AF in patients whose LA short-axis diameter was greater than 40-45 mm compared to patients with normal measurements of less than 40 mm [105, 106]. It has been observed that an LA diameter greater than 45 mm will likely result in AF recurrence after cardioversion. These suggest that the degree of LA dilatation induced by AF is an important factor when designing long-term therapies and treatment strategies [107]. Furthermore, studies have shown morphological changes in the thickness of atrial wall tissue caused by AF, where a thinner wall contributed to the progression of AF from paroxysmal to persistent [108]. The thickness distribution of the atrial wall has been shown to have an important role in AF arrhythmogenesis as it influences the electrical activation patterns during episodes of AF [27, 109-113].

A) Percentages of Left Atrial Fibrosis Grouped into Stages I-IV



B) Proportion of AF Recurrence Grouped by Stage of Fibrosis

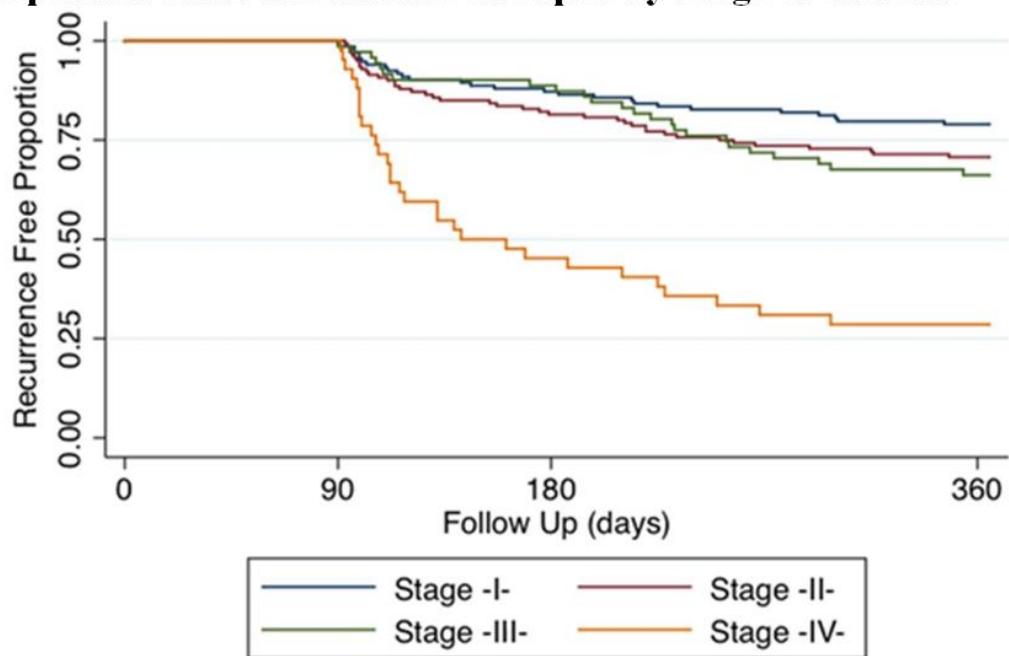


Figure 2.8. The percentage of left atrial (LA) fibrosis calculated from late gadolinium-enhanced magnetic resonance imaging (LGE-MRI) of atrial fibrillation (AF) patients prior to catheter ablation is an independent predictor of the success of ablation, adapted from McGann et al. [114]. **A)** Patients were initially grouped into stages I-IV depending on the extent of LA fibrosis (green). The non-fibrotic, or healthy, tissue is shown in blue. A higher

fibrosis percentage corresponds to a more severe AF. **B)** Kaplan-Meier curve showing the proportion of patients that were free of AF, separated by the four stages over a one-year post-ablation follow-up period. Patients with stage IV fibrosis (yellow line) experienced significantly higher rates of AF recurrence.

2.3 Treatments for Atrial Fibrillation

2.3.1 Overview

The selection of the appropriate treatment for AF patients depends on a number of factors including the type of AF, age, presence of co-morbidities, and severity of symptoms. Current strategies mainly involve strategies for regaining normal heart rhythm (rhythm control), controlling the heart rate (rate control), and prevention of blood clotting to reduce the risk of stroke.

The aim of rhythm control strategies is to revert the heart rhythm back to SR and maintain SR in patients. Rhythm control treatments mainly involve electrical cardioversion, antiarrhythmic medication, and catheter ablation. Electrical cardioversion is a procedure which uses a defibrillator to apply a direct electrical shock to the heart [115]. Cardioversion restores SR immediately and is the preferred method under urgent circumstances if patients are suffering from acute symptoms. Antiarrhythmic medication restores SR over a longer timeframe, and slows the electrical impulses in the heart by targeting the ion channels to reduce conduction. On the other hand, catheter ablation is a minimally invasive approach and involves the “burning” or “cooling” of localized atrial tissue prone to arrhythmia. Through this procedure, tissues which trigger or sustain AF are eliminated in the patient altogether resulting in the termination of AF [17].

After the initial successful restoration of SR, patients often continue taking antiarrhythmic medication to maintain SR and minimize AF recurrence. However, if rhythm restoration is unsuccessful, cannot be maintained for long periods, or undesirable due to patient-specific needs, the rate control strategy is implemented instead. Rate control aims to control and reduce the heart rate to a healthy resting rate with medication such as beta blockers, Ca^{2+} channel blockers, and digoxin. In certain cases where medication is ineffective or unsuitable, a pacemaker may be implanted in the chest to regulate and reduce the heartbeat. Furthermore, preventative treatment due to the increased stroke risk in AF patients involves

anticoagulants, commonly known as blood thinners, to reduce the chance of blood clot formation. Selection of the appropriate antithrombotic treatment also depends on the bleeding risk and the risk factors for stroke present in a patient's medical history.

2.3.2 Catheter Ablation

In recent years, catheter ablation has become one of the most promising methods of AF treatment. Catheter ablation is performed by making an incision, most commonly in the groin area, to insert catheters through the femoral vein and up to the heart and into the atrial chambers. During the procedure, X-ray fluoroscopy is used to guide the insertion and positioning of the catheters. Contrast agents are often used to enhance the clarity of the blood vessels in X-ray images. The technique used for ablation varies between patients, but involves either the use of radiofrequency energy ablation (heat) or cryoablation (extreme cold) to eliminate susceptible tissue and terminate the electrical impulses sustaining the arrhythmia (**Figure 2.9**). Catheter ablation is known to be effective in restoring and maintaining SR in patients with paroxysmal AF, and is generally indicated as second-line therapy after failure or patient intolerance to antiarrhythmic medication. Selection for the ablation procedure is also dependent on the severity of AF, risk of complications such as sudden cardiac death, and patient demographic and history.

The specific locations of the atrial regions to be targeted during catheter ablation have been widely investigated. In 1998, a study done by Haïssaguerre et al. revealed that ectopic beats from the PV are key triggers of AF [116]. The study further demonstrated that electrical isolation of the pro-arrhythmic PV trigger sites via ablation effectively suppressed AF. This approach, known as pulmonary vein isolation (PVI), has since become the recommended procedure during ablation. However, despite the promising effects of PVI in suppressing AF compared to pharmacological strategies, successful prevention of AF recurrence remains suboptimal. Clinical trials across multiple centres have reported a 40-70% AF termination rate one year after ablation for paroxysmal AF patients [117, 118], while the success rates for patients with persistent or longstanding-persistent AF are much lower [119].

Effective ablation targets AF drivers (triggers or substrates that sustain AF). However, the identification of driver locations is a challenging problem [120, 121]. As AF progresses from paroxysmal AF to permanent AF, AF drivers spread throughout both atria beyond the PVs [22].¹² Recent studies found that termination sites for extra-PV ablation increased from

<31% (paroxysmal AF patients), to 55% (persistent AF), to 75% (permanent AF). Importantly, these studies found that >62% of patients with persistent/permanent AF have AF drivers in the RA [122, 123]. In addition, atrial remodelling was found in both atrial chambers for patients with persistent/permanent AF [124]. Therefore, improved patient-specific visualization and reconstruction of the atria geometry may potentially guide the targeting of driver regions and improve the elimination of AF substrates [20, 21].

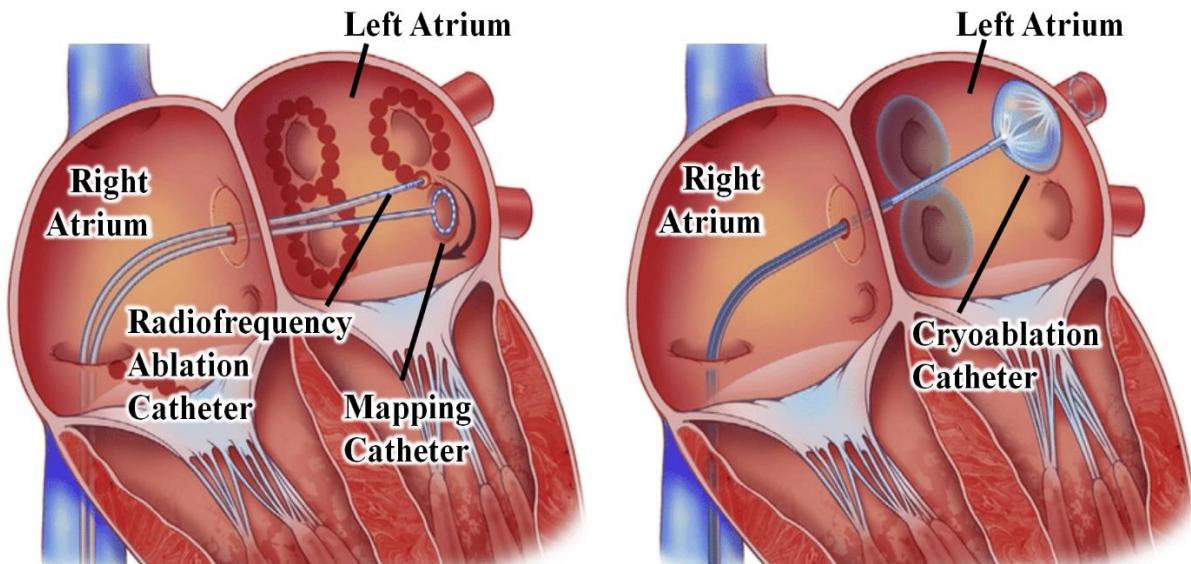


Figure 2.9. Illustration of the pulmonary vein isolation catheter ablation procedure in the left atrial chamber using radiofrequency ablation (left) and cryoablation (right) adapted from Calkins et al. [125].

2.3.3 Electro-Anatomical Mapping

Radiofrequency catheter ablation has been successful in treating many arrhythmias, such as typical atrial flutter or atrioventricular tachycardia, which contain predictable anatomical locations or characteristic electrical activation patterns. However, the treatment of more complex arrhythmias such as AF is more challenging due to the incomplete understanding of the mechanisms of AF and the limitations of conventional mapping techniques to localize arrhythmogenic substrates. The current electro-anatomical mapping systems have improved the resolution, three-dimensional spatial localization, and acquisition of cardiac activation maps to gain new insights into the arrhythmia mechanisms. These systems enable the real-time reconstruction and representation of the 3D cardiac chamber geometry and intra-cardiac electrograms by using novel techniques to accurately localize and

track the mapping catheters.

The most widely used electro-anatomical mapping systems for AF ablation in clinics are the EnsiteNavX (St Jude Medical, Minnesota, United States) and CARTO 3 (Biosense Webster, California, United States) systems [126]. The fundamental principles of the EnsiteNavX system involve the use of electrical impedance to automatically locate the catheters to enable non-fluoroscopic navigation. Three pairs of skin patches (six electrodes) are placed on a patient to create electrical fields along three orthogonal axes. The patch pairs send low-power currents through the patient's chest at slightly different frequencies to form a transthoracic electrical field with the heart in the centre. This forms a voltage gradient across the three axes due to different volume tissue subtended between each surface-electrode pair. An electrical impedance field is then derived by dividing the voltage field with the different currents along each axis to create unique impedance levels that corresponds to specific 3D anatomical locations within the thorax. During the ablation procedure, the catheter electrodes sense the corresponding levels of impedance as it is manoeuvred inside the atrial chambers. The impedance measure, timed with the delivery of the alternating currents, can then be used to identify the exact coordinates of the catheter nodes inside the chamber.

The CARTO 3 system utilizes magnetic technology and electrical current-based methods to produce accurate reconstruction and visualization of the atrial chamber. The system consists of three magnetic field emitters that generate low-intensity magnetic fields and a location sensor in the catheter which measures the strength of the field and its distance from each coil. The location of the catheter is then determined through a triangulation algorithm to calculate the coordinate of the intersection of the catheter with the three fields. In addition to the magnetic field, an electrical field similar to that in the EnsiteNavX system is used to improve the accuracy of catheter localization.

An important feature of the EnsiteNavX and CARTO 3 mapping systems is the proprietary algorithm for automatic atrial chamber reconstruction (**Figure 2.10**) [127]. The spatial points acquired during the manoeuvring of the catheters create a point cloud that forms a sparse representation of the atrial geometry. The reconstruction algorithm generates an initial mesh from the point cloud using Delaunay triangulation or the Ball-Pivoting Algorithm. A smooth surface is generated by connecting line segments and vertices in the mesh with interpolation or extrapolation. Topological validation is performed to ensure the surface generated is entirely closed without openings, apart from naturally occurring anatomical openings such as blood vessels. The presence of non-natural openings implies an incomplete mapping of the region, in which case artificial points are generated to close the

surface. The algorithm then performs a series of filters to check the anatomical plausibility of the generated anatomy, such as the undetected holes in the mesh, the presence of disconnected surfaces, and curvature of each surface. Anatomical landmarks such as the appendage or PVs in the left atrium are identified using skeleton extraction and cylindrical detection algorithms due to their characteristic mathematical geometry. The algorithm also consistently checks if the mesh quality is below a certain threshold and iteratively refines the mesh to improve the anatomical clarity of the reconstruction. Finally, the completed reconstruction is graphically displayed on an interactive interface.

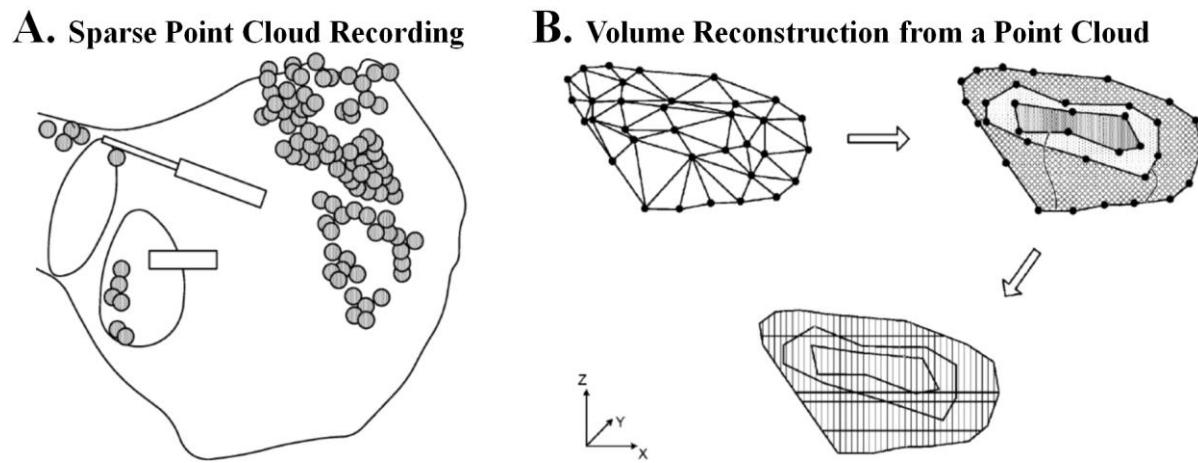


Figure 2.10. Illustration of the cardiac reconstruction algorithm from Biosense Webster (California, United States) adapted from the patent by Merschon et al. [127]. **A)** An example of the sparse point cloud data recording obtained in cardiac chamber. **B)** Diagrams demonstrating the stages of volume reconstruction from a point cloud.

Electro-anatomical mapping of the atrial chambers using EnsiteNavX or CARTO 3 allows for the acquisition of voltage amplitudes at each of the points recorded by the catheter. Voltage mapping is a clinical technique for recording voltages signals to provide a static representation of the electrical activations, and is used as an alternate method of visualizing atrial fibrosis in addition to LGE-MRIs (**Figure 2.11**). A consequence of fibrotic structural remodelling is the low conductivity of the fibrotic tissue as a result of excessive collagen production, forming low-voltage zones during mapping [128]. In clinics, a voltage value of fewer than 0.5 millivolts is commonly used as a threshold to distinguish low-voltage regions from healthy tissue [129]. During mapping-guided ablation, the voltage map is superimposed on the anatomical reconstruction to provide additional information to guide ablation, and have been shown to improve the arrhythmia-free survival rate of AF [130].

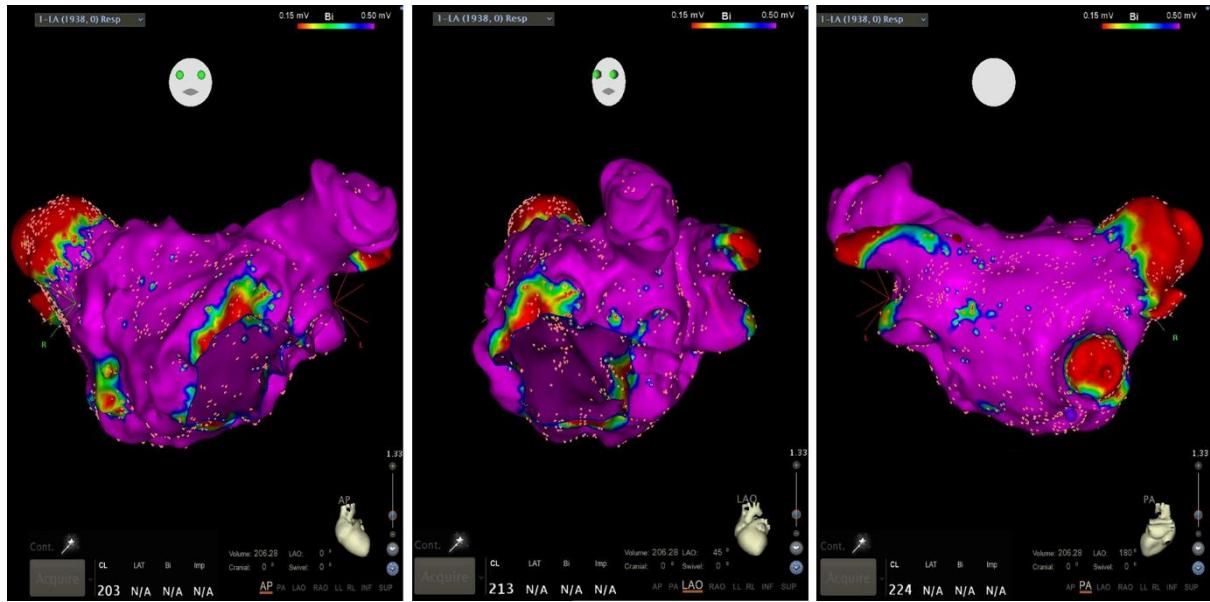


Figure 2.11. Visualization of left atrial reconstruction and voltage map recorded with electro-anatomical mapping using CARTO 3 (Biosense Webster, California, United States) during catheter ablation of atrial fibrillation at Waikato Hospital, Hamilton, New Zealand [131]. The map is shown in three different angles with the orientations of the three views are depicted by the face at the top of each view. The high conduction areas are shown with the colour purple and correspond to healthy tissue. The low conduction areas, or low voltage areas, are shown in red and corresponds to fibrosis.

Chapter 3

Literature Review on Deep Learning

3.1 Background

With the advancements in imaging and anatomical mapping, the increasing availability of larger and information-rich datasets calls for intelligent algorithms capable of streamlining clinical processes to aid clinicians in guiding diagnosis and treatment. Artificial intelligence is the concept of simulating human intelligence to perform difficult tasks at a human level. Machine learning is a subset of artificial intelligence which utilizes learning-based algorithms to independently develop rules for understanding complex datasets. Research into the field of machine learning has gained global momentum in the past decade, driving the development of many intelligent algorithms.

Supervised machine learning is the most successful sub-category of machine learning and involves algorithms capable of learning from data with given ground truths to make predictions. In traditional machine learning algorithms such as support vector machines [132], random forests [133] and K-nearest neighbour [134], a set of meaningful features is manually generated from the raw data by applying certain transformations. These features are then fed into a non-linear classifier, whose parameters are optimized to identify key characteristics of data in different categories. This process requires domain expertise, as a rigorous feature generation and selection procedure is required to find the optimal feature set for learning [135]. Although such methods have been fairly effective, the manual labour required is a major bottleneck for rapid prototyping and driving algorithm performance.

Deep learning is a category of modern machine learning algorithms and mainly involves artificial neural networks (ANNs) which are at the foundation for every deep learning algorithm. The effectiveness of ANNs lies in their ability to automate the feature generation and selection process [44]. This greatly increases the algorithm's learning efficiency as the entire process is fully data-driven, allowing for ease of adaptability to a wider range of tasks [43]. Because of this, ANNs are the dominant driving force of artificial intelligence research in the past decade [46], surpassing human-level performance in many

challenging tasks such as facial recognition [47], natural language processing [48], self-driving cars [49], and medical diagnosis [50, 51]. Furthermore, ANNs are widely used for analysing biomedical datasets in the fields of neurology [52], pulmonology [53], and most importantly, cardiology [56-58].

The following sections introduce the basic principles in ANNs, from simple feed-forward networks to convolutional neural networks (CNN) capable of analysing more complex data structures. We also outline recurrent neural networks (RNN) which are used for time-series data such as bioelectrical signals and often combined with CNNs to form convolutional recurrent networks (CRN). The application of CNNs to classification and object localization is described next, with details on architectural improvements to achieve higher performances. The concept of using CNNs for the important task of image segmentation for delineating important biological structures is then outlined, with an overview of the current popular medical image segmentation CNN architectures such as the fully convolutional network (FCN). We also provide an overview of NST developed in recent years for data augmentation and generation. Lastly, we define the equations of the most common metrics used for evaluating the accuracy of deep learning algorithms which were used throughout this thesis.

3.2 Basic Principles of Neural Networks

3.2.1 Simple Feedforward Neural Networks

This sub-section introduces the basic operations in simple feed-forward ANNs, and how they can be trained with data (**Figure 3.1**). ANNs are comprised of many layers made up of weights. Weights are parameters that can be trained and are set up in the form of matrix operations. An ANN starts with an input layer with fixed dimensions and ends with an output layer that denotes the number of classes to predict. By passing a set of data through various layers of an ANN network, the data is mapped from one shape in space based on its raw input to another throughout the layers and eventually, data from different classes become linearly separable at the last layer (**Figure 3.1B**).

A simple ANN is made up of a series of fully connected layers, each containing a pre-defined number of neurons. At each neuron, an activation (nonlinear) function f is applied to restrict the numerical values to avoid divergence. There are a few common activation

functions including rectified linear unit (ReLU) and softmax. The ReLU function is given by

$$f_{ReLU}(l) = \max(l, 0) \quad (\text{Equation 3.1})$$

given an inputted layer, l . It is used at intermediate layers and has been shown to speed up feature space separation [136]. A similar activation function named the parametric rectified linear unit (PReLU) is also widely used, and is often swapped with ReLU during parameter tuning to improve the overall performance. This function is designed as the zero values produced by ReLU given negative inputs often present difficulties during optimization. It is an adapted ReLU function and is defined as

$$f_{PReLU}(l) = \begin{cases} ax, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (\text{Equation 3.2})$$

where a is a parameter learned by the algorithm, with a value less than 0.01. In the event that the parameter a becomes 0.01, the activation becomes the leaky ReLU defined similarly as

$$f_{LeakyReLU}(l) = \begin{cases} 0.01x, & x < 0 \\ x, & x \geq 0 \end{cases} \quad (\text{Equation 3.3})$$

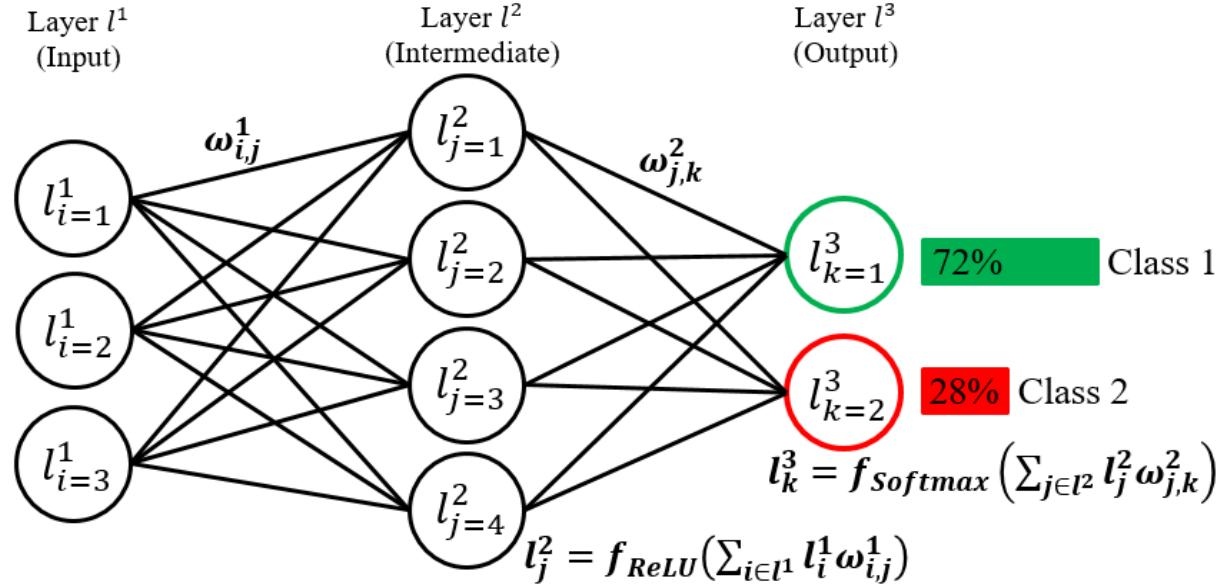
which also serves to address the zeroing problem of the ReLU activation. The softmax function

$$f_{Softmax}(l_j) = \frac{e^{l_j}}{\sum_{i=1}^K e^{l_i}}, \quad \forall j \in [1, K] \quad (\text{Equation 3.4})$$

is used at the last layer of the ANN for the classification of K classes by normalizing the outputs to a probability value within [0,1]. The predicted class in the last layer l^N is hence defined by

$$K_p = argmax_{i \in l^N}(l_i^N). \quad (\text{Equation 3.5})$$

A. Simple Artificial Neural Network



B. Visualization of the Feature Space Transformation

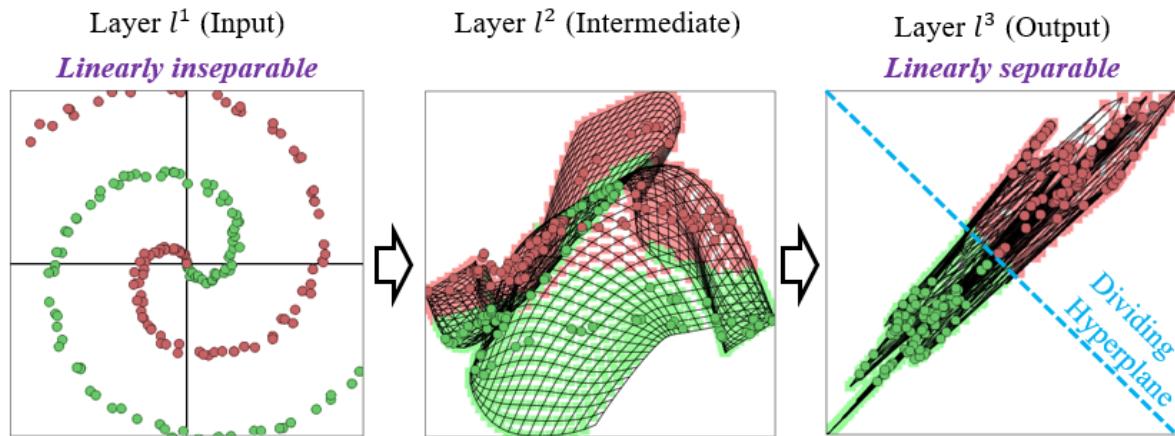


Figure 3.1. The basic concept of an artificial neural network (ANN). **A)** A 3-layer ANN mapping an input sized 3×1 to an intermediate layer sized 4×1 , then to an output sized 2×1 which denotes a binary classification. **B)** A spiral dataset is used to demonstrate how the feature space changes throughout the ANN for the two classes (green and red). The transformed output data is shown to be linearly separable.

The values in the weight matrix are optimized with stochastic gradient descent [137] to improve feature space separation. This is made possible by providing the network with training input data and their corresponding ground truths. During training, the dataset is fed through the network in mini-batches to increase efficiency. A learning rate for stochastic gradient descent is chosen to control the magnitude for updating the parameters in each iteration. The updating of the weight parameters is performed with backpropagation [138] such that the weights, ω , at layer i is updated by

$$\omega^i = \left(1 - \frac{\nu}{N}\right) \omega^i - \frac{\nu}{N_{batch}} \frac{\partial L}{\partial \omega^i} \quad (\text{Equation 3.6})$$

for a learning rate ν , with a total sample N , a mini-batch size N_{batch} , and a loss function L . At the very beginning of the process, the weights are initialized with random values sampled from a normal distribution [139]. The most common optimization loss function is the cross entropy

$$L(f_c(x), y) = -\log(P(y = c|x)) \quad (\text{Equation 3.7})$$

which is the log likelihood of obtaining a correct prediction c given the data x and the ground truth y .

3.2.2 Convolutional Neural Networks

The CNN is a type of ANN which excels in processing 2D data such as images. A convolutional layer differs from a fully connected layer such that adjacent layers are only partially connected. This reduces the computational cost by having fewer weights whilst also allowing for 2D data to be directly fed into the network. The convolution operation is performed by sliding a small filter across the entire input image, and calculating the dot product between the input and the filter at each step. This is written as

$$l_{i,j}^{n+1} = f\left(\sum_{a=1}^K \sum_{b=1}^K \kappa_{ab} l_{i+a,j+b}^n\right), \forall i, j \in l^n \quad (\text{Equation 3.8})$$

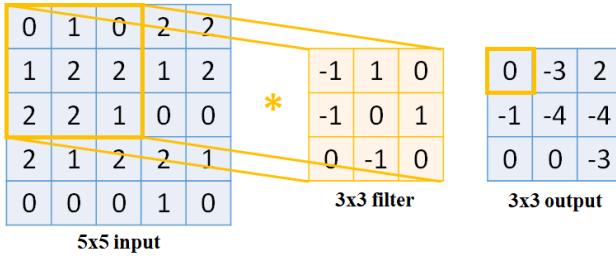
for a $K \times K$ filter, κ , at each (i, j) over the entire input of layer l^n . The values in the filter are the weight parameters to be trained. For a filter size K , $(K - 1)$ pixels are lost in each dimension after each convolution due to the filters being restricted within the boundaries of the image. The entire convolution operation can be repeated F times per layer with different filters to create F feature maps to increase the degrees of freedom for optimization. This transforms an $M \times N$ matrix into an $M \times N \times F$ matrix.

Pooling is a common operation used in CNNs. It increases the efficiency of training by compressing the matrices at certain layers to decrease the total number of parameters in a network. Max-pooling is the most commonly used type of pooling where a layer is down-

sampled by a factor of K through taking the maximum in each sub-grid $K \times K$, i.e.,

$$l_{i,j}^{n+1} = \max(l_{i+k,j+k}^n), \forall k \in [0, K-1] \quad (\text{Equation 3.9})$$

A. Convolution Operation



B. Max Pooling Operation

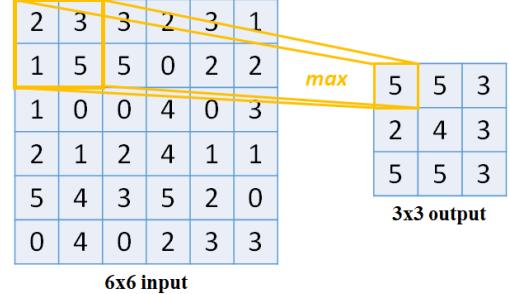


Figure 3.2. Illustration of the basic operations in a convolutional neural network. **A)** A simple convolution operation on a 5×5 image with a 3×3 filter. **B)** A simple max pool operation on a 6×6 image.

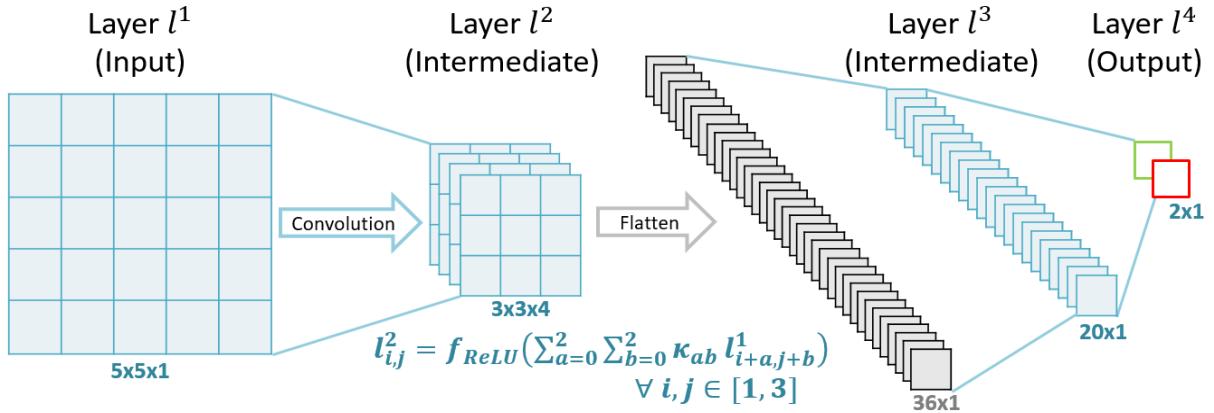


Figure 3.3. A simple convolutional neural network with a single convolutional layer (l^2) and two fully connected layers (l^3 and l^4). l^1 is convolved with a 3×3 filter and the number of feature maps increases from one to four in l^2 . It is then flattened into a 1D vector so that the operations from the fully connected layer can be applied for model prediction as described in the last **Figure 3.1**.

Overfitting is a potential issue in CNNs due to the large number of parameters needed to be trained. Dropout [140] is a technique which may decrease overfitting by randomly setting nodes to zero. This forces the network to find unbiased features in the dataset that is representative of the data in general, instead of specific characteristics that appear only in the training set. A dropout rate, ϕ , is used to update any given layer such that the layer l^n becomes

$$l_{i,j}^n = l_{i,j}^n \phi_{i,j}^n, \forall i, j \in l^n \quad (\text{Equation 3.10})$$

where ϕ equals zero with a pre-determined probability.

3.2.3 Residual Convolutional Neural Networks

The optimization of CNNs is a gradient-based method and is enhanced via the use of residual connections. Residual connections provide an alternative pathway for the gradients to be propagated during backpropagation by using skip connections (**Figure 3.4**) [49]. The motivation of residual blocks stems from an optimization problem which occurs in traditional CNNs that do not contain skip connections. In traditional CNNs, it is observed that by utilizing a larger number of layers, which is previously thought to increase the learning capacity by increasing the degrees of freedom, results in a decrease in performance during training compared to CNNs with fewer layers. The vanishing gradient problem is a possible explanation to this phenomenon, where, in larger CNNs, the gradients of the layers that are deeper in the network progressively approach zero after many iterations due to the repeated derivative calculations in backpropagation, resulting in non-optimization [141]. Skip connections can increase the efficiency of the propagation of gradients during weight parameter updates in gradient descent by combining the original gradient with the gradient after a convolution.

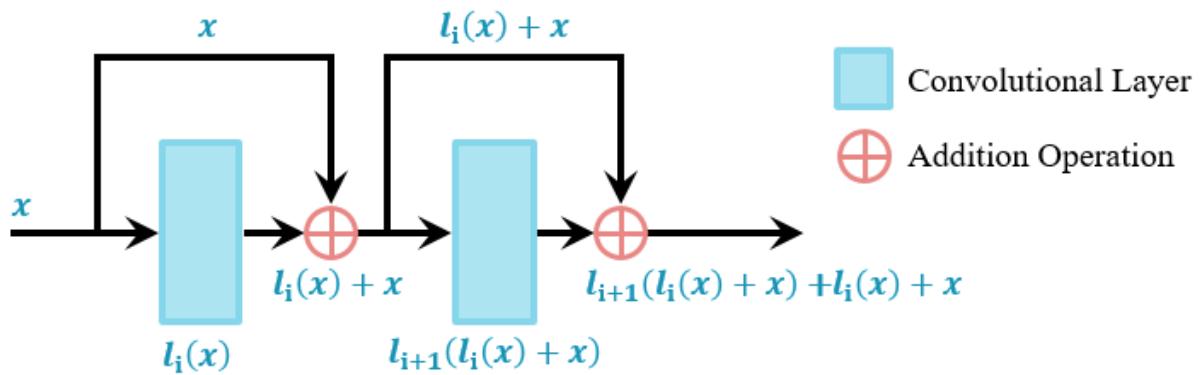


Figure 3.4. Illustration of the basic operations in residual blocks which are used in convolutional neural networks. The skip connections provide an alternative pathway for information to be propagated without introducing additional parameters. The original input data is merged with its respective transformed version via the use of an element-wise sum.

Within a residual block, a typical network includes a batch normalization layer, followed by a ReLU layer, and then a convolutional layer [49, 142]. Batch normalization is a layer which constantly normalizes each mini-batch throughout the entire network, reducing the internal covariant shift caused by progressive transforms [143]. Batch normalization uses the mean of the mini-batch, in a layer, x_{mean} , and their standard deviation, x_{std} , such that

$$Batch\ Norm(l_i) = \gamma \left(\frac{x_i - x_{mean}}{x_{std}} \right) + \beta \quad (\text{Equation 3.11})$$

where x_i represents entries in the current mini-batch and γ and β are trained parameters. By applying ReLU before a convolution layer, as opposed to applying it after, as in the traditional fashion, the input values are further normalized to accelerate optimization [49].

3.2.4 Recurrent Neural Networks

RNN is widely used in the fields of speech recognition [144], language translation [145] and signal processing [146] as the cycling of information within each node allows sequential data to be processed efficiently (**Figure 3.5**). In RNNs, the outputs of each neuron within the intermediate layers are cycled into its original input. Hence, RNNs differ from CNNs as they are able to process data of variable input size. When many recurrent neurons exist in a recurrent layer, the sequential data is processed in parallel through different weights, allowing RNNs to generate multiple representations and create effective feature space separation.

In each RNN layer, l^n , for every node, information at different time steps, t , of an input is stored. The value of the layer in the next time step depends on the current time step of the layer, and the current time step of the previous layer such that

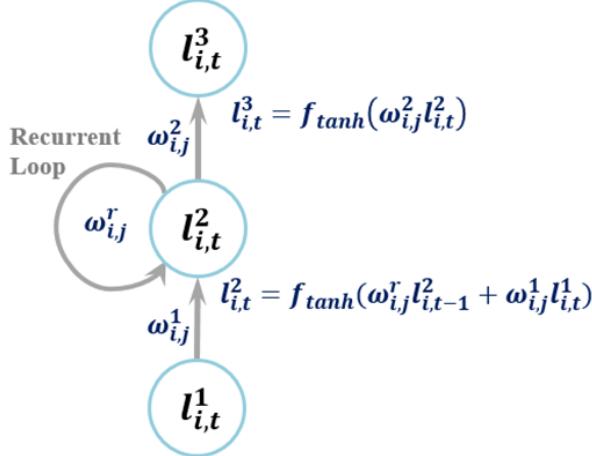
$$l_{i,t+1}^n = f_{tanh}(\omega_{i,j}^r l_{i,t}^n + \omega_{i,j}^{n-1} l_{i,t}^{n-1}) \quad (\text{Equation 3.12})$$

where ω^{n-1} is the weight of the previous layer and ω^r is the weight of the recurrent node. The value of the next layer in the current time step depends on the current layer by

$$l_{i,t}^{n+1} = f_{tanh}(\omega_{i,j}^n l_{i,t}^n) \quad (\text{Equation 3.13})$$

The above operations are repeated to propagate the nodal values through several successive recurrent nodes.

A. Recurrent Node



B. Recurrent Node Unfolded

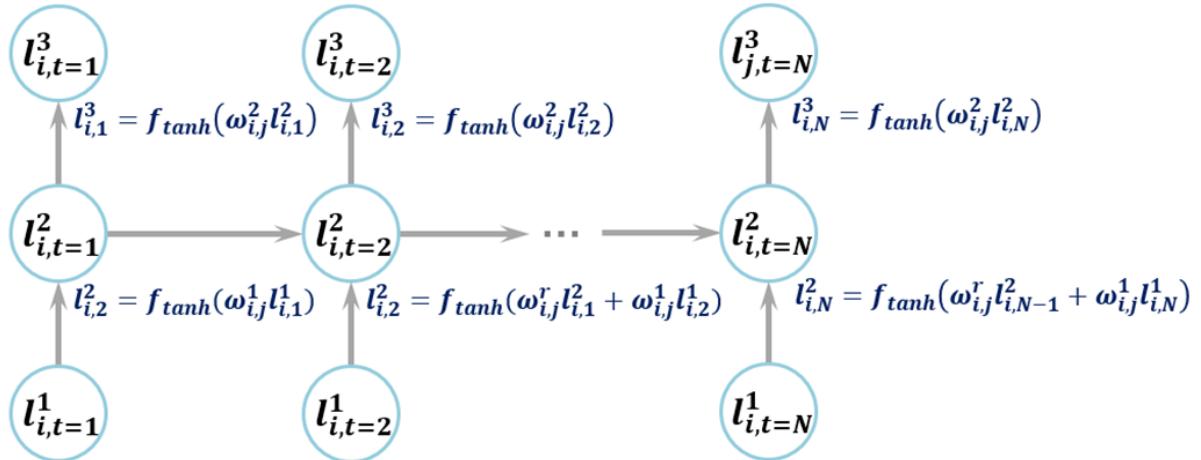


Figure 3.5. Demonstration of the operations of a layer in a recurrent neural network with a single node. **A)** A compressed representation of the recurrent node. **B)** An unfolded version illustrating the cycling of weights inside a recurrent node at different time steps. Note that the weights within each layer are shared but applied to different time steps. The first node $l^2_{i,t=1}$ simply inputs the information from the previous layer as there is no previous time step.

Complex data types such as videos, time-series MRI, and bio-signals require both CNNs for feature extraction and RNNs for sequential learning. CNNs are combined with RNNs to create CRNs. This is done by feeding the output of a CNN as the input to the RNN such that the RNN collates information processed by the CNN over several time-frames into

a single prediction. The CNN-RNN configuration significantly improves upon the feature extraction capabilities of single RNNs as CNNs are able to capture spatial information which can then be used to enhance the RNN’s sequential feature extraction capabilities. By learning directly on more abstract features, the RNN is able to capture additional details which wouldn’t have been readily available directly from the raw time-series datasets.

The CRN architecture is flexible and allows the use of 1D, 2D, or 3D CNNs depending on the task as long as the output of the CNN for every time step, t_i , is a single value with n feature maps. The subsequent RNN input then takes in an input with t time steps with an input number of channels, or features maps, of n . In the case where the input data is of varying time length, the CNN simply iterates through each time step and feeds the time-wise single predictions into a dynamic RNN which handles data of varying length, making CRNs extremely adaptable to different data sizes. Furthermore, bi-directional RNNs are often used to take into account the sequential nature of the data from both front to end and end to front as a further enhancement to the feature learning stage during training [147].

3.3 Convolutional Neural Networks for Classification

Classification is the task of identifying and separating data into different categories automatically given a pre-determined number of classes in the dataset provided. While the concept seems rudimentary, image and signal classification in particular have traditionally been an extremely difficult task for rule-based algorithms, with their performances far below human levels. The configuration for a CNN to perform classification is the simplest, and has already been described in the previous **Chapter 3.2.2**. While the sample provided is the most basic CNN architecture, more complex architectures are required to achieve high performances. The ImageNet Large Scale Visual Recognition Challenge provides an accurate representation of the current state-of-the-art algorithms for image classification, and is used as a benchmark for intelligent algorithms aiming to surpass human-level performance [148]. The ImageNet database contains over 1 million images annotated to 1000 different classes, and annual challenges are held for top research groups such as Microsoft Research (Microsoft Corporation, Washington, United States), Facebook Research (Facebook Inc, California, United States), and Google Brain (Alphabet Inc, California, United States) to compete for increasingly superior performances.

2012 was the first year in which a CNN achieved first place in the ImageNet

challenge [149], with an error rate of 16.4%, improving from the prior year's top performance by around 10%, demonstrating the high potential of CNNs [150]. This network, named "AlexNet" consists of four convolutional layers and three fully connected layers with a maximum depth of 384 feature maps. This CNN configuration was further improved upon in 2014 in Visual Geometry Group (VGG)-Net which contains 16 convolutional layers and three fully connected layers with a maximum depth of 512 feature maps [151]. The VGG-Net study proposed the use of convolutions of increasing depth, starting from 64 and doubling every two to four layers until 512, which significantly improves feature retention and learning deeper into the network where the network parameters became increasingly smaller numerically.

While error rates continues to decrease each year since 2012, 2015 was the most notable year as the winning residual CNN architecture surpassed the estimated human error rate of approximately 5% for the first time in history [152]. The proposed CNN consisted of 152 layers with residual connections every two to three layers. **Figure 3.6** shows a condensed version of the CNN with 34 layers for simplicity, and its counterpart which does not contain any residual connections. The residual network architecture is very similar to that of VGG-Net with convolutions of increasing depth which doubled in depth every few layers. However, the implementation of residual connections allows networks with a much larger number of layers to be trained. The original VGG-Net contains a maximum of 19 layers, as the addition of convolutional layers is observed to cause a decrease in the overall performance. By employing residual connections, the residual network could be trained with over 100 layers, significantly increasing the feature learning capacity of the CNN for larger and more complex datasets. Residual networks have since been a cornerstone in the field of deep learning, as new architectures readily apply this technique to boost performance accuracies.

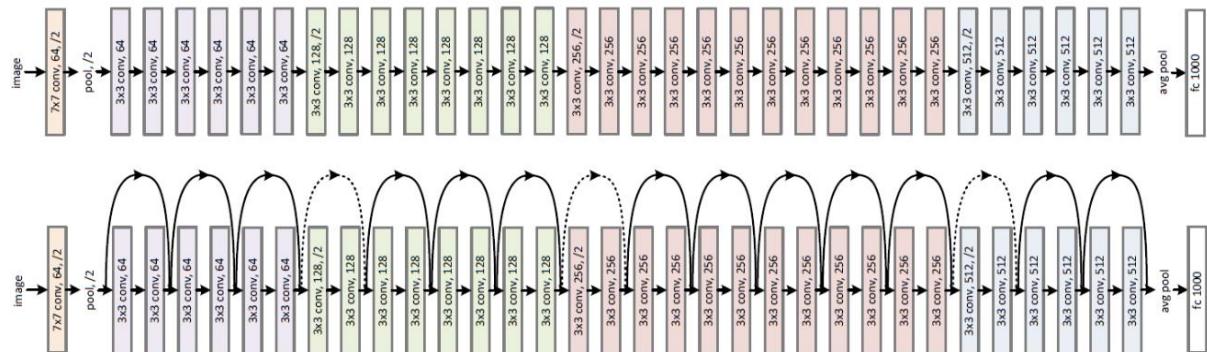


Figure 3.6. Illustration of the residual convolutional neural network (CNN) architecture which won the 2015 ImageNet Large Scale Visual Recognition Challenge, adapted from He

et al. [152]. The normal CNN configuration is shown at the top, while the modified residual CNN is shown at the bottom with the additional residual, or skip, connections.

The ease and light-weight nature of residual networks also allows further investigation into improving its architecture by combining it with other successful CNN architectures. In 2016, the residual CNN was combined with the Inception CNN, the winner of ImageNet 2014, to form “ResNeXt” [153]. In 2017, the “squeeze and excitation” module was invented and added to every residual connection in the residual CNN to improve its performance [154]. The basis of the squeeze and excitation network is the ability to adjust the weightings of each feature map in each convolutional layer. While features maps of each layer have traditionally been treated with an equal weighting during optimization, this new technique proposes to adaptively adjust the weightings with trainable scaling parameters to add additional degrees of freedom during optimization. The result of adding this module is an increase in the residual network performance with a significantly smaller number of layers. This implied the squeeze and excitation module reduced redundancies of each convolutional layer, improving the overall efficiency of the learnable parameters, and reducing computational costs.

3.4 Convolutional Neural Networks for Localization

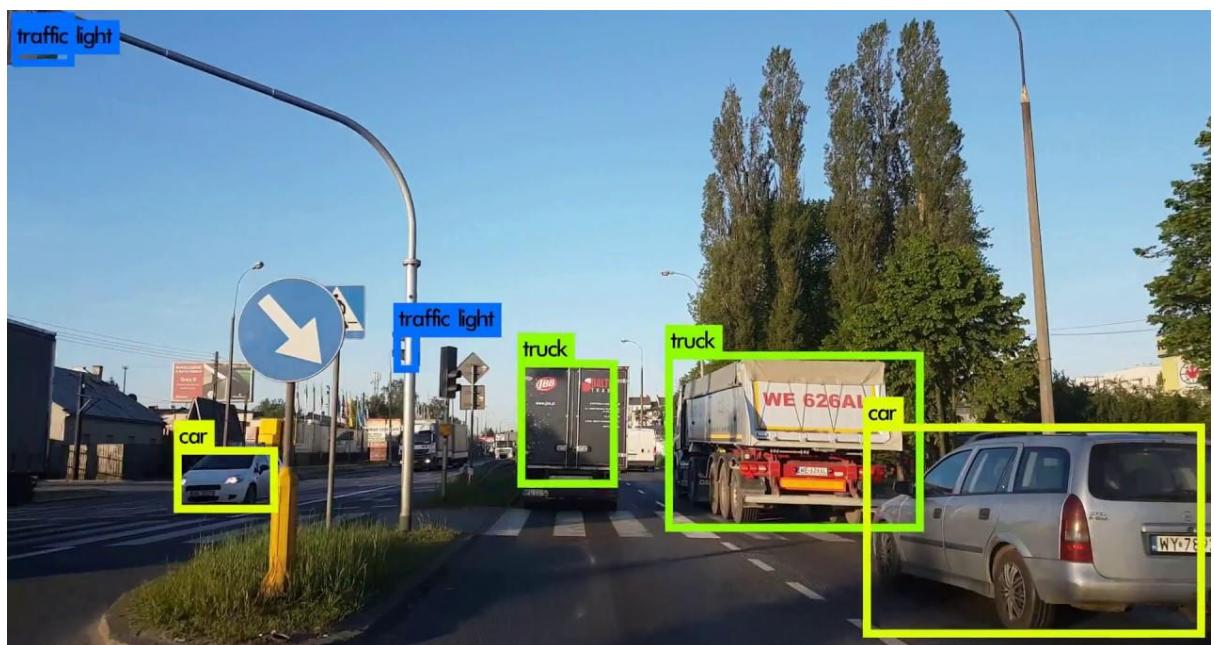


Figure 3.7. A practical example of image localization with a convolutional neural network in

a self-driving car to detect the coordinates of specific objects of interest given an image frame.

Localization is the task of finding the exact coordinates of an object, or several objects or regions of interest (ROI) within an image (**Figure 3.7**). The result of a localization algorithm is often either the centroid of the localized object or a bounding box which fully contains the object. CNNs for localization perform an important task in identifying details beyond just the overall image class, such as the identification of several different classes within the same image to provide more information to the user. A practical application and example of localization is in self-driving cars [49], which require the real-time identification of several different objects of interest such as the road, streetlights, street signs, other cars, people, and obstacles.

The complexity of a localization CNN increases depending on the number of objects of interest required for localization [155]. Instead of classification, localization is seen as a regression problem, as the predictions consist of continuous Cartesian coordinates instead of discrete classes. Implementation-wise, this modification is done by changing the activation function from softmax to ReLU to allow continuous variables as outputs. ReLU also naturally normalizes the output by restricting values to only be positive. Other activations such as sigmoid may also be used depending on the specific output required. The optimization loss function is then changed to a mean squared error (MSE) loss which measures the accuracy of continuous values. For a CNN performing localization on a single object in a 2D image, only the output layer needs to be changed. In this example, the output contains two output nodes for the x and y coordinates representing the object centroid. The overall loss function is then simply defined as

$$f_{MSE}(x, y) = \frac{((x - x')^2 + (y - y')^2)}{2} \quad (\text{Equation 3.14})$$

where the accuracy of the predicted x and y coordinates is compared against the ground truth, x' and y' . If the localization task is performed on a 3D image, the output would contain three nodes for each of the x , y , and z coordinates. Similarly, the detection of N objects would require $2N$ output nodes for 2D images, and $3N$ output nodes for 3D images to represent the centroids of each object of interest.

The detection of multiple objects in a single image introduces significant complexities over single object detection tasks. In recent years, the development of CNNs for object

detection has primarily been focused on either the regional proposal CNNs or the single-shot CNNs such as the single shot detector (SSD) or you only look once (YOLO). Regional proposal CNNs were first proposed to address the issue of object detection problems having an undetermined number of objects in each image, making traditional CNNs with fix-sized outputs infeasible for the task. The initial regional proposal network, named R-CNN [156], utilizes a selective search algorithm which extracted approximately 2,000 potential ROIs for further processing. The selective search involves manual feature engineering and fixed rule-based algorithms. A CNN, such as the VGG-Net is then used to predict each of the proposed ROI to detect the exact coordinates of the bounding box. While this was the first method proposed for multi-object detection, the training process is extremely time-consuming due to a large number of ROIs extracted per image. The next iteration of R-CNN was the fast R-CNN [157], in which a CNN is first used on the entire image to perform feature extraction, and the subsequent feature map is fed through the selective search algorithm. By utilizing the CNN first, the feature for every possible ROI is already computed, and the proposed ROIs are easily optimized utilizing the pre-computed features to produce a bounding box. In essence, instead of having the CNN be executed 2,000 times, it is only executed once per image which exponentially increases the speed. The final iteration was the faster R-CNN [158], in which the selective search algorithm is replaced with a trainable network to predict potential ROIs for objective localization. This further decreases the computational cost as the network was trained to only propose high-confidence ROIs instead of a large number of ROIs.

The invention of the R-CNN group of algorithms significantly improves upon the existing methods for multi-object detection. However, the increased complexity as a result of the improvements made to decrease computation results in the increased difficulty of training and parameter tuning. This has led to the development of single-shot detectors such as the SSD or YOLO which aims to directly detect the objects from a raw image without the need for ROI proposals [159]. For such networks, the input image is first divided into smaller grids, with each grid containing an initial set of bounding boxes which are pre-defined. The networks are then trained by adjusting the coordinates of all bounding boxes from each grid to match with the coordinate of the ground truth. Bounding boxes with a similarity below a certain threshold are removed, while those with a higher likelihood of containing an object are selected to localize the objects within the image. By performing object detection in this manner in a single step, the algorithm vastly outperforms the R-CNN approaches in terms of computational speed and memory efficiency. However, this also leads to weaknesses in detecting small objects, as the pre-defined set of bounding boxes may not be universally

applicable to every case presented.

3.5 Fully Convolutional Networks for Image Segmentation and Reconstruction

Segmentation is the classification of each individual pixel in an image. An FCN is a type of CNN that contains only convolutional layers and no fully connected layers. The last layer of an FCN is also a convolutional layer that is able to produce predictions, which are not restricted to a 1D vector shape, such as in traditional CNNs. This allows for the pixel-to-pixel segmentation prediction of the 2D structure representatives of an image (or 3D structures in the case of 3D images). The unpooling operation in an FCN aids this process by up-sampling intermediate layers to preserve the original input dimension. This can be achieved by simply making copies of each node K^2 times to obtain a K factor up-sampling such that

$$l_{i+k,j+k}^{n+1} = l_{i,j}^n, \forall k \in [0, K - 1] \quad (\text{Equation 3.15})$$

The combination of convolution, pooling and unpooling has been used in many segmentation FCNs. Long et al. used a limited number of convolutional and pooling layers which down-sampled the image, and the final output was up-sampled with deconvolutions to match the input dimensions for prediction [160]. In their study, deconvolution is a reverse process of convolution and achieves a similar purpose as unpooling. Ronneberger et al. utilized a series of convolutional and pooling layers to down-sample the input image, and a series of convolutional and unpooling layers to up-sample the image back to its original size [161]. Noh et al. employed a similar architecture in their study except the images were up-sampled with deconvolutions instead [162]. The advantage of using FCNs for segmentation is in their ability to perform predictions for entire image patches. While making pixel-wise predictions is possible using CNNs, by making patch-wise predictions with FCNs, the network is able to learn the relationship between pixels within each patch. This results in increased performance.

The FCN is set up in a way such that the last layer is convolved from the previous layer to a shape of size $M \times M \times K$ to predict K classes for each pixel in an $M \times M$ image patch. Each value is then passed through the softmax function. For each pixel, a prediction is

made by taking the index along the depth of the feature map with the highest probability such that the output at layer l^N is

$$K_{p_{i,j}} = \operatorname{argmax}_{k \in K} (l_{i,j,k}^N), \forall i, j \in l^N \quad (\text{Equation 3.16})$$

FCNs are applied extensively for various medical image segmentation tasks, and have quickly become the most popular choice due to its superior accuracies compared to traditional thresholding, statistical shape modelling, and atlas registration approaches. While many studies have adapted FCNs used for semantic segmentation [162, 163] on to segmenting biological structures [164, 165], the most popular CNN architecture for segmenting medical images is U-Net (**Figure 3.8**) [161]. Since its proposal in 2015, many variants of U-Net has been developed, including 3D U-Net (**Figure 3.9**) [166], and V-Net (**Figure 3.10**) [167]. These FCNs have therefore been a cornerstone for setting state-of-the-art performances for various medical image segmentation tasks.

The original 2D U-Net consists of a down-sampling segment and an up-sampling segment. The down-sampling part of the network contains 2D convolutions with the number of feature maps doubling from 64 to 1024 every two to three layers. Max pooling is used to decrease the size of the image between the convolutional layers to decrease the memory consumption of the network. The up-sampling part of the network contains 2D convolutions with the number of feature maps halving from 1024 to 64 every two to three layers. Up-convolutions, or deconvolutional layers, are used to increase the size of the image back to a sufficient size for segmentation. While these operations are standard in FCNs, the uniqueness of U-Net lies in the feature-forwarding operations, a type of residual connection. These connections are applied at parts of the network where the image contains different scales to concatenate the unprocessed features in the down-sampling part with the reconstructed features in the up-sampling part of the network. Thus, spatial information directly from the original input image is better preserved, leading to sharper segmentation and improved convergence during training.

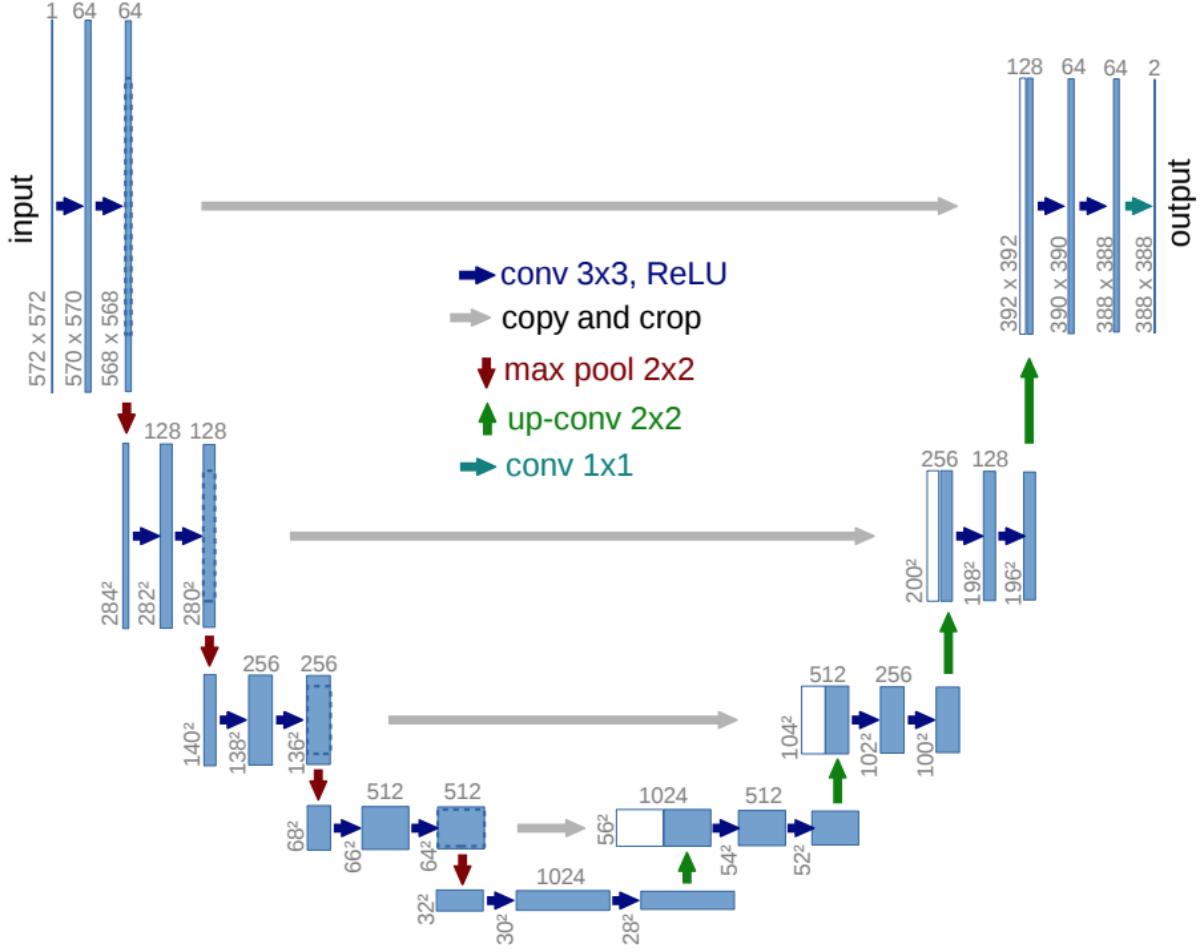


Figure 3.8. The architecture of the most popular convolutional neural network (CNN) for medical image segmentation, U-Net, adapted from Ronneberger et al. [161]. Vertical numbers represent the size of the image as it passes through the CNN, and horizontal numbers represent the number of feature maps present in each convolutional layer. BN, batch normalization; conv, convolutional layer; ReLU, rectified linear unit.

The 3D U-Net architecture is similar to 2D U-Net, except there are a smaller number of convolutional layers and feature maps to reduce the computational burden associated with 3D convolution operations. 3D convolutions have the exact same principle as 2D convolutions, but with 3D filters which slide over a 3D space, resulting in a cubic computational complexity. In 3D U-Net, the number of feature maps doubles from 32 to 256 in the down-sampling part, and halves from 256 to 64 in the up-sampling part. All other components remain the same as in 2D U-Net. Since its proposal in 2016 by the same group that published the original U-Net paper, 3D U-Net has been a popular choice for the segmentation of 3D structures from volumetric medical images such as MRIs and CTs. Although it is able to perform segmentation of 3D images directly, the increased number of

parameters results in a network that is more difficult to train and requires larger amounts of 3D data to effectively optimize.

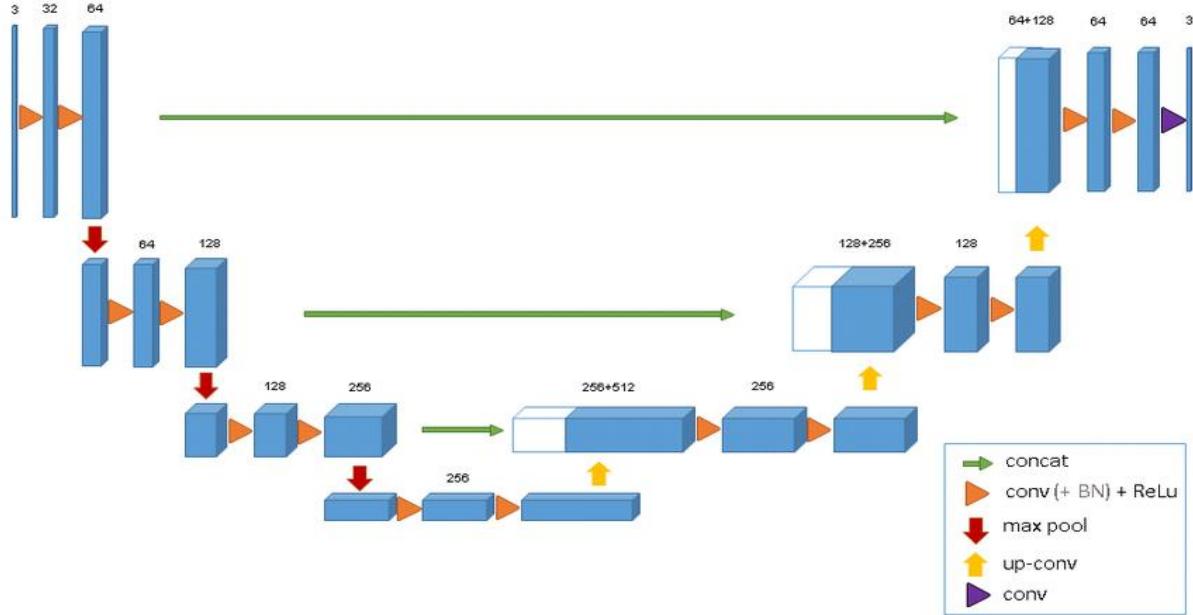


Figure 3.9. The architecture of the 3D version of U-Net for 3D image segmentation, adapted from Cicek et al. [166]. The horizontal numbers represent the number of feature maps present in each convolutional layer. BN, batch normalization; concat, concatenation layer; conv, convolutional layer; ReLU, rectified linear unit.

The V-Net architecture also proposed in 2016 by Milletari et al. [167] aims to enhance 3D FCNs for better convergence on 3D image segmentation. The network uses convolutions with a smaller number of feature maps at the beginning ($N=16$) to increase the total number of layers available for feature extraction without drastically increasing the computational intensity. The network also includes more convolutional layers in the coarse parts at the bottom of the network to further improve feature extraction. To improve convergence, along with the feature-forwarding connections used in the original U-Net architecture, V-Net employs residual connections in convolutional layers of the same spatial resolution. To ensure information is not lost during down-sampling or up-sampling, max pooling layers are replaced with convolutions with a stride of two to achieve the same effect while also extracting useful features. Finally, instead of the classic $3 \times 3 \times 3$ filters used in most 3D CNNs, V-Net uses $5 \times 5 \times 5$ filters to improve the receptive field to better capture more complex structures in a 3D space, leading to smoother segmentations. With these changes implemented, V-Net has been shown to achieve superior segmentation accuracies

compared to 3D U-Net methods with faster convergence times [168].

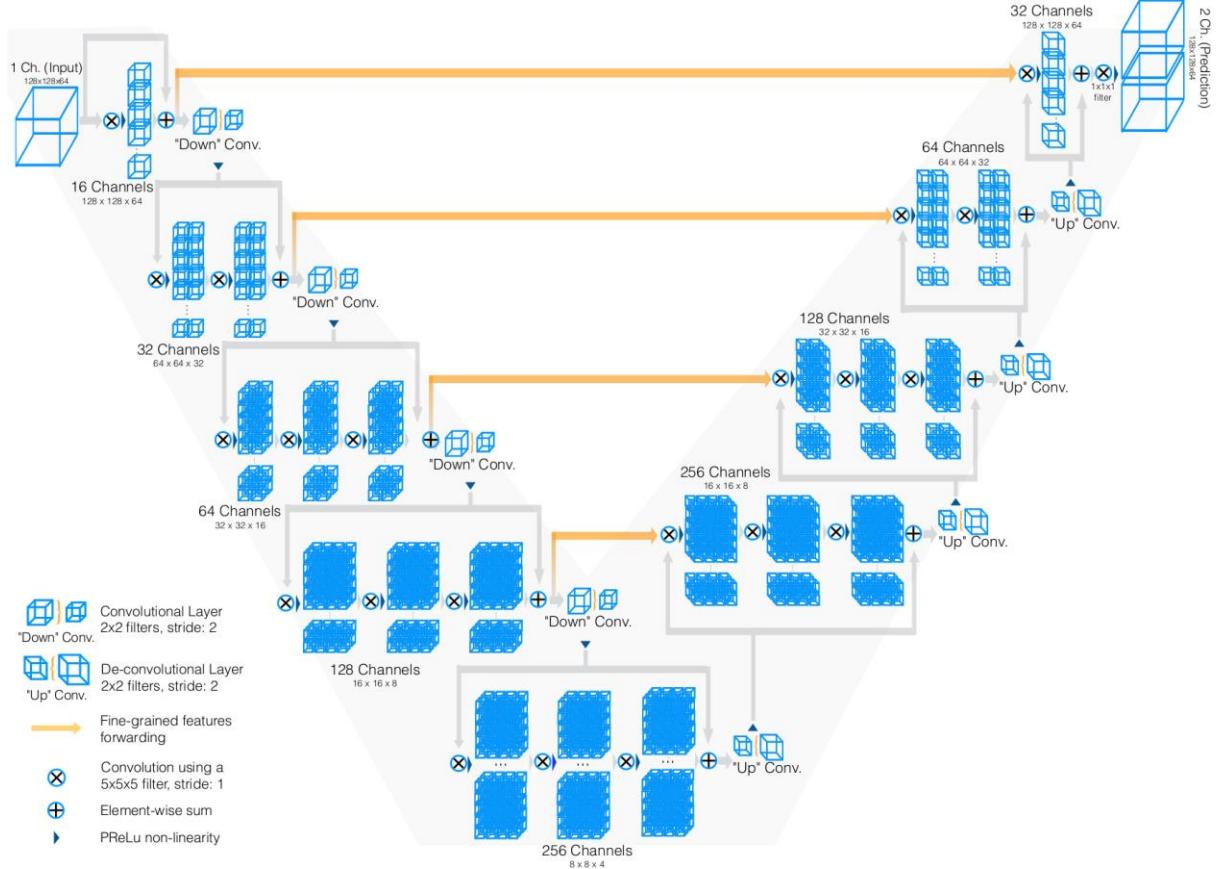


Figure 3.10. The architecture of V-Net proposed by Milletari et al. [167] containing the characteristic “V” shape as shown in the original paper.

3.6 Neural Style Transfer Networks for Data Augmentation

NST networks are a class of neural networks published in 2016 [169] used specifically for the synthesis of a new image by combining features from two images, style features from one image and content features from the other (**Figure 3.11**). NST are primarily used for applications towards texture synthesis [170], artwork generation [171], and photo-realistic scene and lighting transfer [172]. The initial concept of NST is that the networks used do not have to be trained to generate images, and a network pre-trained on a sufficiently large dataset can be used to perform NST between any pair of images. This allows efficient image synthesis as no time and computationally consuming training process is required when applying NST to different datasets. Furthermore, methods have been developed to the point where the current state-of-the-art approaches are able to perform NST effectively between

images from any arbitrary dataset [173-175]. A recent 2019 study has proposed the use of NST for improving 3D cardiac segmentation from MRIs by correcting inconsistent data [176], and NST for converting MRIs to LGE-MRIs in order to increase the number of training samples for cardiac segmentation from LGE-MRIs [177].



Figure 3.11. The overall concept of neural style transfer (NST). Given **A**) a content image providing spatial context, and **B**) a style image providing style features, an NST network is able to generate **C**) an image containing the context of the first image and the style of the second image, creating a completely new image.

The original NST algorithm proposed by Gatys et al. in 2016 [169] uses a pre-trained VGG-Net [151] for feature extraction to learn content and style features from content and style images. **Figure 3.12** shows a simplified version of the overall architecture. The content features are simply the convolutional activations in any selected layer of VGG-Net and the style features are the inner product of the vectorised convolutional activations in any selected layer of VGG-Net. For each pair of content and style images, the generation process begins by initializing the output image with random values. The content features from the content image and the output are compared, and the style features from the style image and the output are compared to compute their differences. The values within the output image are then optimized such that both its content and style features converge towards those of the content and style images. This can be achieved by minimizing a loss function consisting of a weighted sum between the content loss and style loss. In NST, the content loss, L_c , is defined by:

$$L_c(x_c, x_g, l) = \frac{1}{2} \sum_{i,j} (F_{i,j}^l - P_{i,j}^l)^2 \quad (\text{Equation 3.17})$$

given the content image, x_c , generated image, x_g , at layer l of VGG-Net. In this equation, F and P represent the features extracted by VGG-Net for the content and generated images,

respectively. The style loss, L_s , is defined by:

$$L_s(x_s, x_g, l) = \frac{1}{4N_l^2(H_lW_l)^2} \sum_{i,j} (G_{i,j}^l - A_{i,j}^l)^2 \quad (\text{Equation 3.18})$$

given the style image, x_s , generated image, x_g , at layer l of VGG-Net with N feature maps sized $H \times W$. G and A represent the inner product of the vectorised features extracted by VGG-Net for the style and generated images, respectively. However, in the original NST paper, while the content loss was computed at only one layer of VGG-Net, the style loss was computed across multiple layers and summed together such that

$$L_s(x_s, x_g) = \sum_{l=0}^L \omega_l L_s(x_s, x_g, l) \quad (\text{Equation 3.19})$$

for L layers of VGG-Net weighted by ω_l at each layer, with the value of ω being a hyper-parameter tuned during experimentation. The overall loss is thus a weighted sum of L_c and L_s where the weighting factor is also a hyper-parameter.

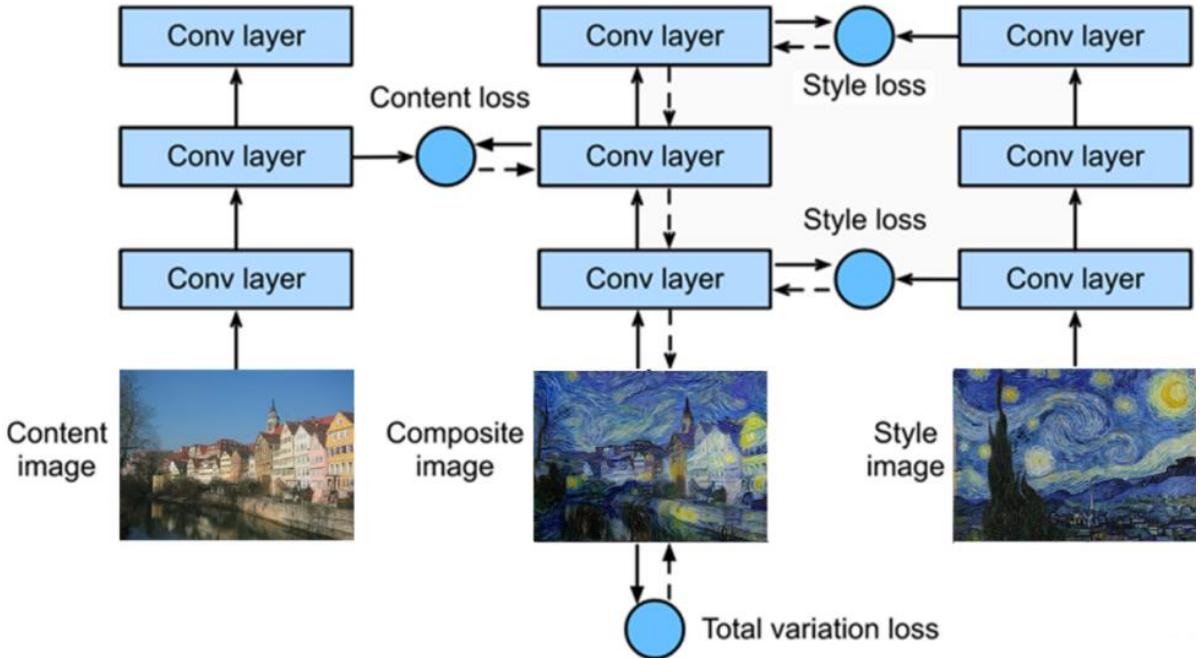


Figure 3.12. A simplified representation of the neural style transfer (NST) network proposed by Gatys et al. [169]. The convolutional neural network (CNN) extracts content features from the content and output (composite) images, and style features from the style and output images. The content loss at a selected layer and the style loss across multiple layers are then minimized such that the output image contains features from both the content and style

image.

The transfer of styles obtained from artworks as shown above to content images containing real-world scenes can be easily achieved with the original NST network. However, the output image has a tenancy to be overly “artistic” in the sense that the realism of the original content image is replaced by artwork-like representations of the original context. Luan et al. therefore proposed a method for photo-realistic NST such that the realism is preserved in the output image, allowing for more effective real-life style transfer applications [172]. The method utilizes the exact same setup as the original NST algorithm, but instead modified the loss function to include an enhanced style loss and a regularization loss term. The enhanced style loss function is formulated by computing the inner product of the features extracted from VGG-Net concatenated with semantic segmentation masks from the content image to preserve sharp context structures. The regularization loss term also serves a similar purpose in preserving the structure of the content image by creating an affine transformation function that maps local colour intensity values onto the output image while penalizing outputs with colour intensities that are widely different from that of the input. Through this formulation, photo-realistic features can be preserved effectively, allowing the method to achieve realistic generated outputs (**Figure 3.13**).



Figure 3.13. Photo-realistic neural style transfer the content (left) and style (middle) images combined to produce a realistic output (right) adapted from Luan et al. [172].

In 2017, researchers from Adobe (Adobe Inc. California, United States) and Nvidia (Nvidia Corporation, California, United States) developed an improved method of performing style transfer [173]. The algorithm named “universal style transfer” aims to be able to perform NST on arbitrary visual styles and content images without the need for any training or optimization of the output image (**Figure 3.14**). This approach attempts to solve the difficult challenge associated with designing a single algorithm capable of generalization across images of any content and style while being computationally fast and memory

efficient. The method also involves the use of a pre-trained VGG-Net. To achieve style transfer, content and style images are first fed through the VGG-Net with its fully connected layers removed. The features extracted for the two images are then merged with the proposed whitening and colouring transform function which directly matches the feature covariance of a content image given a style image. The merged and transformed features are then feed through an inverse of VGG-Net to directly reconstruct an output of the same size. To further improve the generated output, the authors propose to use whitening and colouring transform after every convolutional layer in VGG-Net to merge the content and style features through multiple layers, allowing the algorithm to better capture and match the two feature sets while using minimal additional computation. Through extensive experimentation, the study demonstrated the images generated from the proposed method achieve significantly higher visual scores when validated by users to assess the realism of the output. Furthermore, strong evidence for the superiority of the approach compared to other methods is outlined by being data-independent, more computationally efficient, and learning-free.

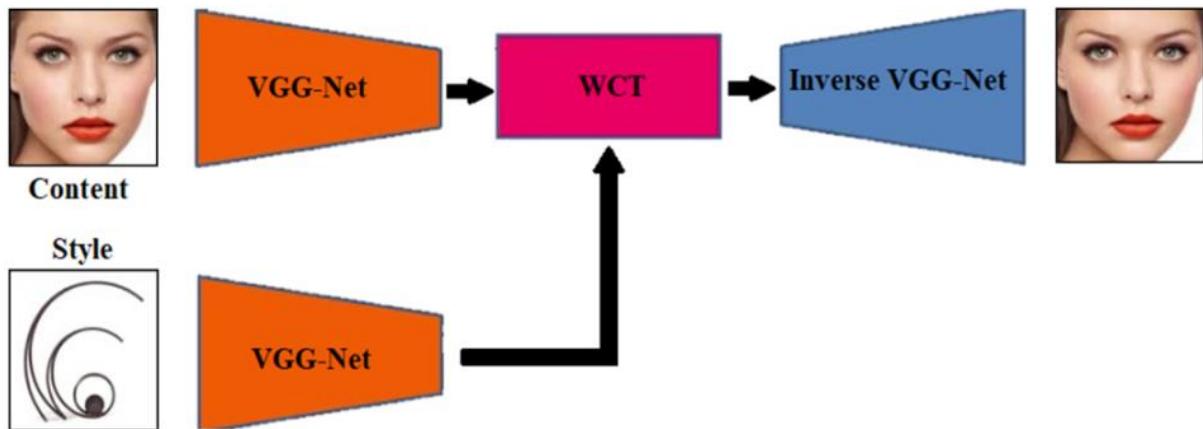


Figure 3.14. The structure of universal style transfer network consisting of a VGG-Net (and its inverse) and a whitening and colour transform (WCT) operation adapted from Li et al. [173].

3.7 Evaluation of Deep Learning Models

Deep learning models are validated through the comparison of their predicted values against a set of pre-defined ground truth labels. Several evaluation metrics for measuring the effectiveness of a newly proposed model are available and standardized across the deep learning community for consistency and comparability with past results. This section outlines

the most widely accepted evaluation metrics used in the field of deep learning, which are also implemented to evaluate the efficacy of the methods proposed in this thesis. This section also includes several metrics adapted and formulated specifically for the task of atrial analysis which is the focus of this thesis.

3.7.1 Technical Measures

Technical metrics measure the absolute accuracy of the predictions from methods such as CNNs for classification, localization, and segmentation. These include the MSE, F_1 score or the Dice score as known for segmentation, intersection over union (IoU) or Jaccard index, sensitivity, specificity, Hausdorff distance, and surface to surface distance (STSD). The MSE is a common measure of the accuracy of location by comparing two sets of coordinates. It is given by

$$MSE = \frac{(x - x')^2 + (y - y')^2}{2} \quad (\text{Equation 3.20})$$

for the ground truth coordinates (x, y) and the predicted coordinates (x', y') of the centre of mass. The F_1 score or the Dice score is the most commonly used metric for evaluating image segmentation accuracy, and is given by

$$DICE(A, B) = \frac{2|A \cap B|}{|A| + |B|} \quad (\text{Equation 3.21})$$

for a prediction mask, A , and a ground truth, B . It is noted the F_1 score is the common name used for evaluating classification accuracy and is formulated in the same manner as the Dice score. The IoU measures the similarity between a prediction and a ground truth with more weighting towards the mistakes in the predictions, and is defined as

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|}. \quad (\text{Equation 3.22})$$

The sensitivity and specificity are used to reflect the success of each algorithm for segmenting the foreground, or positive pixels, and the background, or negative pixels, respectively

$$Sensitivity = \frac{TP}{TP + FN} \quad (\text{Equation 3.23})$$

$$Specificity = \frac{TN}{TN + FP} \quad (\text{Equation 3.24})$$

where TP stands for true positive, TN stands for true negative, FP stands for false positive, and FN stands for false negative. These two metrics are used to evaluate both classification and segmentation accuracies. The Hausdorff distance measures the maximum local distance between the surfaces of the prediction and the ground truth. This metric evaluates geometrical characteristics, unlike the Dice or IoU which purely evaluates pixel-by-pixel comparisons. The 3D version of the Hausdorff distance is used in this thesis to measure the largest error distance of the 3D segmentation defined for a prediction, A , and ground truth, B , as

$$HD(A, B) = \max_{b \in B} \left\{ \min_{a \in A} \left\{ \sqrt{a^2 - b^2} \right\} \right\} \quad (\text{Equation 3.25})$$

where a and b are all pixel locations within A and B . The STSD measures the average distance error between the surfaces of the predicted segmentation, A , and the ground truth, B , and is given by

$$STSD(A, B) = \frac{1}{n_A + n_B} \left(\sum_{p=1}^{n_A} \sqrt{p^2 - B^2} + \sum_{p'=1}^{n_B} \sqrt{p'^2 - A^2} \right) \quad (\text{Equation 3.26})$$

where n_A is the number of pixels in A , n_B is the number of pixels in B , and p and p' describe all points in A and B .

3.7.2 Biological Measures for Atrial Anatomical Accuracy

In addition to the commonly used technical measures, biological measures are also used to validate the clinical usability of the predictions made. These metrics are computed individually for each of the LA and RA. It includes the atrial anterior-posterior diameter error and atrial volume error between the predictions and ground truths which are the two most commonly used clinical biomarkers during the treatment of AF. The atrial diameter, \mathcal{O}_{atria} , is computed by measuring the maximum distance from the anterior side to the posterior side of the atrial chamber. Mathematically for a given 2D MRI slice, this is

written as

$$\varnothing_{atria}(M) = \max_{i \in I} (\sum_{j=1}^J M_{ij}) \quad (\text{Equation 3.27})$$

for a 2D mask, M , with dimensions $I \times J$, where J is the anterior-posterior axis of the atrial chamber. The atrial diameter measured in 3D is simply the maximum value obtained from computing the maximum anterior to posterior distance in each slice of the 3D mask. The diameter error, $Error_\varnothing$, measured in mm, is then given by

$$Error_\varnothing(A, B) = |\varnothing_{atria}(A) - \varnothing_{atria}(B)|r_x \quad (\text{Equation 3.28})$$

for a prediction masks, A , and a ground truth, B , with a spatial resolution of r_x mm in the anterior-posterior axis. The atrial volume, V_{atria} , is calculated by summing all pixels defined in the atrial chamber cavity segmentations. Mathematically, this is written as

$$V_{atria}(M) = \sum_{i=1}^X \sum_{j=1}^Y \sum_{k=1}^Z M_{ijk} \quad (\text{Equation 3.29})$$

for a 3D mask, M , with dimensions $X \times Y \times Z$. The volume error, $Error_V$, measured in cm^3 , is then given by

$$Error_V(A, B) = |V_{atria}(A) - V_{atria}(B)|r_x r_y r_z \times 10^{-3} \quad (\text{Equation 3.30})$$

for a prediction masks, A , and a ground truth, B , with spatial resolution of $r_x \times r_y \times r_z$ mm³. These measures are converted into percentage accuracies by dividing by the ground-truth value for ease of understanding and comparison during analysis.

Chapter 4

Enhancing the Detection of Atrial Fibrillation with Neural Style Transfer and Convolutional Recurrent Networks

Early detection and intervention of AF has shown to be vital for effective treatment of the disease. ECG provides an effective, non-invasive approach for clinical diagnosis and monitoring treatment in patients with AF. Portable ECG recording devices including Apple Watch and Kardia devices have been developed for AF detection. However, the efficacy of these smart devices has not been fully validated as clinical trials have varying results for detecting AF signals from ECGs and the exact algorithms used are not openly available.

In this chapter, we developed a CRN for automatic AF detection using the largest publicly available single-lead ECG dataset acquired through the Kardia mobile device. We proposed a novel NST network for data augmentation to address the current issue with the limited ECG data available. By developing and evaluating our approach on an open-source ECG dataset, we demonstrated that our framework was both robust and verifiable, and potentially can be used to improve portable devices for AF classification. This chapter is presented in a similar form as the potential publication version, cited as:

Xiong, Z., Stiles, M.K., Gillis, A.M. and Zhao, J., 2022. Enhancing the Detection of Atrial Fibrillation from Wearable Sensors with Neural Style Transfer and convolutional Recurrent Networks. (Under Review)

4.1 Introduction

ECG is a widely used, non-invasive technique for clinical diagnosis in patients with cardiac arrhythmia, including AF. However, such diagnoses require specially trained health professionals to manually read and identify irregular ECGs, which is often a time consuming

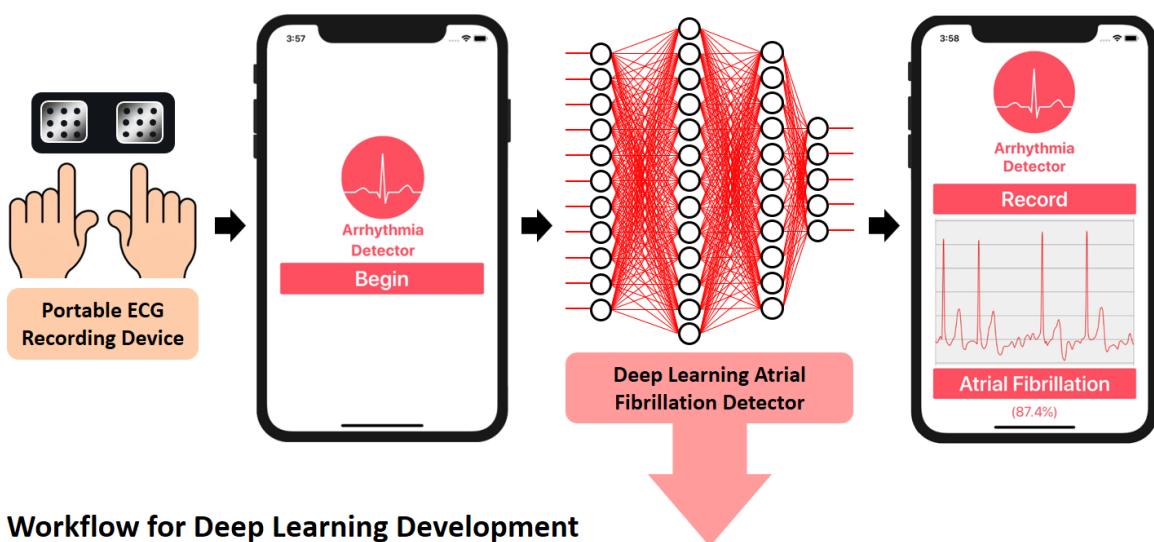
and a rather subjective process in some instances [23]. As a result, there is high interest in developing an automatic approach for AF detection from ECGs.

Currently, two main approaches exist for AF detection from ECGs: atrial activity-based analyses detect AF by identifying the absence of P-waves and the presence of f-waves [178-183], whereas ventricular activity-based analyses look for irregularity in the RR intervals of QRS complexes [184-192]. However, atrial activity-based analyses are often error-prone when performed on noise-contaminated ECGs. This is due to the small signal-to-noise ratio between the low amplitude f-waves and the noise [193]. On the other hand, ventricular activity-based analyses alleviate this problem due to the well-defined, large-amplitude QRS complexes in the ECGs [194]. However, ventricular activity-based analyses require relatively long ECG recordings (>30s) of AF episodes for reliable detection [195]. Recent studies have combined both atrial and ventricular activity-based analyses for more accurate AF detection [196-198], and have obtained promising results on diverse datasets. Traditional machine learning algorithms such as support vector machines [199] have also been used to enhance performance [198, 200-202]. These methods involve the extraction of features from ECGs, such as ventricular and atrial-specific activities, to learn which features are specific to AF and non-AF. This generally requires domain expertise in the field of ECG analysis, as a rigorous feature generation and selection procedure is required to find the optimal feature combination for learning, which can be extremely time-consuming.

In recent years, many portable mobile devices have been developed capable of automatically monitoring the health state of patients, notably Apple Watch [25] (Apple Inc, California), KardiaBand/Mobile [26] (AliveCor Inc, California), Huawei Watch [203] (Huawei Technologies Co. Ltd, Shenzhen), and Cardiio Rhythm [204] (Cardiio Inc, Massachusetts) by utilizing ECG signals recorded from handheld or wearable ECG monitors or through photo-plethysmography. Although many studies have been conducted, the accuracy of current commercially available portable devices for AF detection remains unclear. Bumgarner et al. compared the accuracy of the KardiaBand for AF detection from self-recorded ECGs with physician-interpreted 12-lead ECG diagnoses in 100 patients [26] and concluded the KardiaBand was unable to interpret a third of the patient's recordings while the physicians were able to accurately diagnose AF in 99% of patients. Moreover, a recently published study for the most widely used Apple Watch device on >400,000 patients showed AF was only present in 34% of the patients detected by the device when validated by a physician using a subsequent conventional ECG patch, showing the high propensity for false alerts [25].

The development and validation of automated methods for ECG classification in these smart devices remain difficult due to the lack of a large and well labelled ECG dataset. The issues surrounding the lack of data may potentially be mitigated with well-established data augmentation. However, existing augmentation techniques commonly applied to images such as rotation, flipping, or resizing are ineffective as ECGs are much more rigid in terms of its data structure and will likely result in a decreased physiological accuracy. A method of artificially generating sufficient quasi-physiological ECG samples is therefore ideal. Additionally, artificial intelligence algorithms, such as CRN, have been applied to many challenging tasks in the field of medicine [161, 205] for signal processing.

A. Atrial Fibrillation Detection Through a Deep Learning Powered Mobile Device



B. Workflow for Deep Learning Development

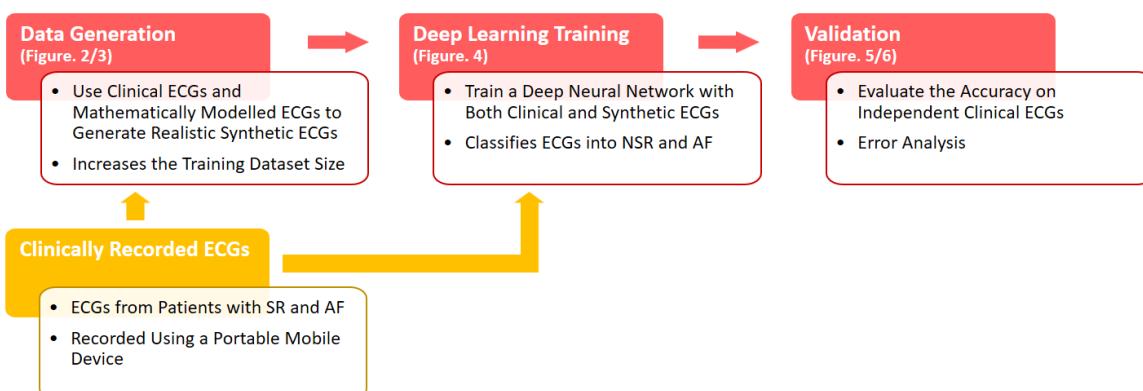


Figure 4.1. Detection of atrial fibrillation (AF) from single-lead electrocardiograms (ECG) through a mobile device using deep learning. **A)** The outline of the proposed pipeline containing portable recording electrodes and a mobile application embedded with a deep learning algorithm. **B)** Workflow for the development of a superior deep learning algorithm trained on both clinical and synthetic ECGs. SR, sinus rhythm.

In this chapter, we have designed a robust CRN using currently the largest open-source single-lead ECG dataset donated by AliveCor [206] for the automatic detection of AF (**Figure 4.1**). We have incorporated multiple CRN layers to improve feature learning from raw ECG signals and an architecture that is capable of processing ECG recordings of variable length. To overcome the issue of the limited data samples, we have proposed a novel generator capable of synthesizing a large number of realistic ECG recordings from current samples of clinical ECGs. This leads to a multi-fold increase in the number of training data available and significantly improves AF detection accuracy of the CRN.

4.2 Methods

4.2.1 Clinical ECG Dataset

The dataset used in this study was from the 2017 PhysioNet/Computing in Cardiology Challenge, donated by AliveCor, which contained 5,834 single-lead ECG recordings recorded from patients with a self-diagnosis device [206]. The device contained two electrodes which were held in each hand to create a lead I equivalent ECG. The signals were recorded for an average length of 30 seconds, with the shortest recording being 9 seconds and the longest recording being 60 seconds in length. The frequency of the signal was 300 Hz with a dynamic range of ± 5 mV. Expert annotations were provided by AliveCor and validated by the 2017 PhysioNet challenge organizers [206] and contained the labels of SR and AF.

Overall, the dataset contained 5,076 SR recordings and 758 AF recordings. The dataset was split into 50% training and 50% testing such that the developed methods could be validated on an independent dataset. A smaller train-test ratio was chosen due to the small number and high variability of AF ECGs present in the entire dataset. This allowed the test set to contain sufficient signals from a broader range of statistical distributions for validation.

4.2.2 Synthetic ECG Generation

A novel method for synthesizing ECGs was developed to provide a deep learning algorithm with more training samples for AF detection. The ECG generator consisted of three

components: 1) clinical ECG recordings as described above, 2) a mathematical model consisting of a system of ordinary differential equations [207], and 3) a convolutional NST network [169] (**Figure 4.2**).

The mathematical model was first used to model idealized ECG waves with a specific heart rate, or pacing of the R-peak-to-R-peak intervals, via parameters for the mean and standard deviation of the heart rate (**Figure 4.2B**). The algorithm used for the modelling of ECG beats consisted of a system of ordinary differential equations (ODEs). The variables controlling the waveform of each ECG beat cycle were initialized and controlled by the amplitude, a_i , the angular spread, b_i , and the locations of the Gaussian functions, θ_i . These three variables were arrays consisting of seven values each, corresponding to the seven components of a simplified ECG: P -, P +, Q , R , S , T -, T +. The minus represented the first half of a wave while the plus represented the second half of a wave. By adjusting the values in the 3 arrays, varying morphologies of ECG signals could be reproduced consistently with high precisions.

The simulation was performed with 3 ODEs, each of which determined the morphology of the P-wave, QRS complex, and T-wave, respectively. The final ECG sample was then the sum of the final solutions of the three ODEs across a pre-defined space determining the length of the ECG signal. The equations were as follows for the three variables with respect to the number of iterations, k , and max iteration, K , in a discrete form

$$P_{k+1} = P_k - \sum_{i \in [P^-, P^+]} \frac{a_{ik}\omega}{b_{ik}^2} \Delta\theta_{ik} \exp\left(-\frac{1}{2} \left(\frac{\Delta\theta_{ik}}{b_{ik}^2}\right)^2\right) \quad (\text{Equation 4.1})$$

$$QRS_{k+1} = QRS_k - \sum_{i \in [Q, R, S]} \frac{a_{ik}\omega}{b_{ik}^2} \Delta\theta_{ik} \exp\left(-\frac{1}{2} \left(\frac{\Delta\theta_{ik}}{b_{ik}^2}\right)^2\right) \quad (\text{Equation 4.2})$$

$$T_{k+1} = T_k - \sum_{i \in [T^-, T^+]} \frac{a_{ik}\omega}{b_{ik}^2} \Delta\theta_{ik} \exp\left(-\frac{1}{2} \left(\frac{\Delta\theta_{ik}}{b_{ik}^2}\right)^2\right) \quad (\text{Equation 4.3})$$

where a_i , b_i and θ_i each contained the components [P -, P +, Q , R , S , T -, T +] and $P_0 = 0$, $QRS_0 = 15$, $T_0 = 0$. ω was defined as $2\pi f$ for a pre-defined heart rate frequency and $\Delta\theta_{ik}$ is the difference between the current θ_{ik} and the phase of the signal modulus 2π . The final output is thus $P_K + QRS_K + T_K$.

The convolutional NST network used to generate synthetic ECGs utilized a simple 1-layer 1D network with 128 feature maps to process features from the real clinical data and the

mathematically modelled data to produce an output containing features from both data samples (**Figure 4.2C**). The aim of the network was to combine the heart rate features from the simulated data, referred to as the content, with the patient-specific P-QRS-T features in the clinical data referred to as the style (**Figure 4.2D**).

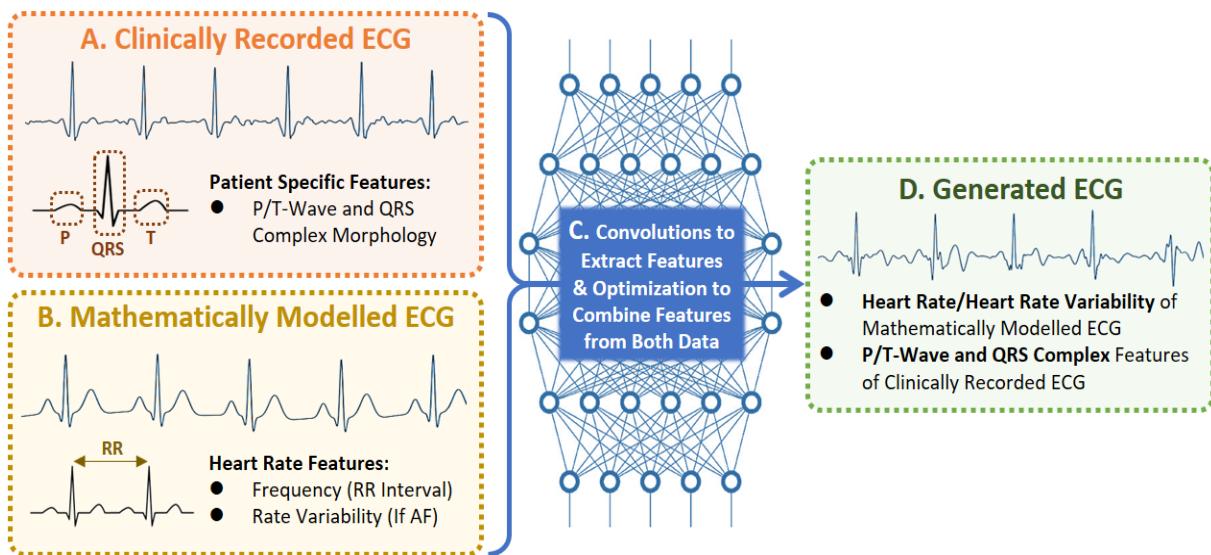


Figure 4.2. Method for generating physiologically-accurate synthetic electrocardiograms (ECG) to increase the data available for deep learning synthetically. **A)** Clinically recorded ECGs containing patient-specific P-QRS-T features. **B)** Heart rate features from mathematically modelled ECGs. **C)** The combination of patient-specific P-QRS-T features of clinical ECGs and heart rate features of mathematically modelled ECGs with a convolutional feature extractor and optimizer. **D)** The generated output consisting of a realistic synthetic ECG. AF, atrial fibrillation; RR, R-peak-to-R-peak.

To produce realistic ECG signals effectively, the features were initially extracted and compared from both the content and style. The formulation was similar to that described in **Chapter 3.6**. The content features were the convolutional activations outputted from the 1D convolutional layer and the style features were the inner product of the vectorised convolutional activations outputted. For each pair of content and style ECGs, the generation process began by initializing the output signal with random values. The content features from the content signal and the output were compared, and the style features from the style signal and the output were compared to compute their differences. The values within the output signal were then optimized such that both its content and style features converge towards those of the content and style signal [208], producing a realistic simulated ECG signal as a

result. This was achieved by minimizing a loss function for each of the content and style features. In our network, the content loss, L_c , was defined by

$$L_c(x_c, x_g, l) = \frac{1}{2} \sum_{i,j} (F_{i,j}^l - P_{i,j}^l)^2 \quad (\textbf{Equation 4.4})$$

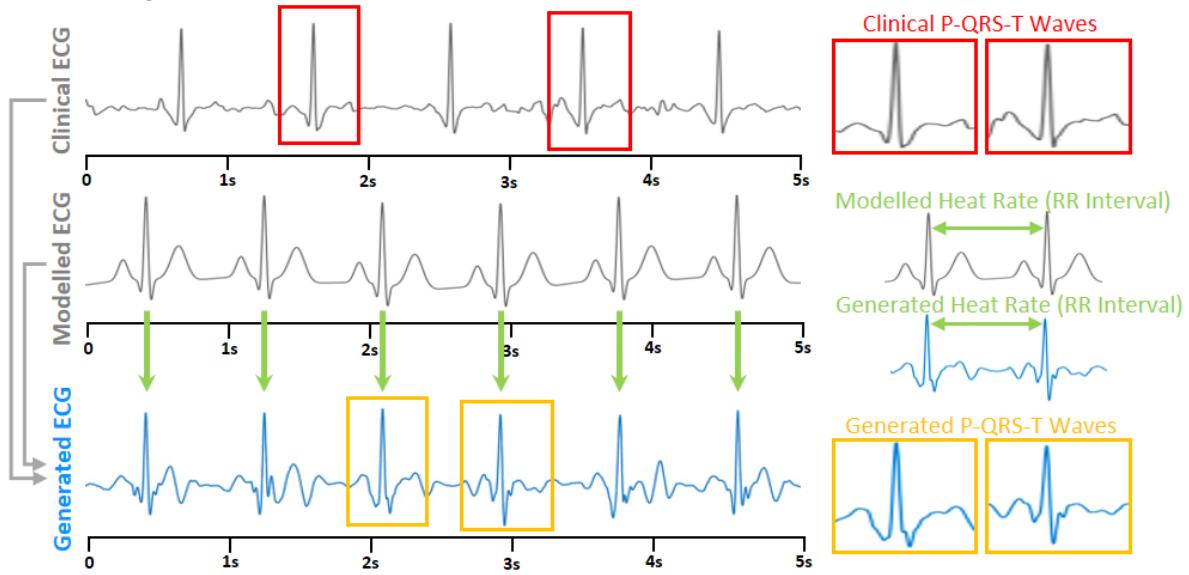
given the content signal, x_c , generated signal, x_g , at layer $l = 1$ of our simple network for indexes i and j . In this equation, F and P represent the features extracted for the content and generated signals, respectively. The style loss, L_s , was defined by

$$L_s(x_s, x_g, l) = \sum_{i,j} (G_{i,j}^l - A_{i,j}^l)^2 \quad (\textbf{Equation 4.5})$$

given the style signal, x_s , generated signal, x_g , at layer $l = 1$ of our simple network. G and A represent the inner product of the vectorised features extracted for the style and generated signals, respectively. The overall loss was finally a sum of L_c and L_s . Through this method, the ECG dataset used for deep learning could be increased multi-fold by using different pairs of clinical ECGs and mathematically modelled ECGs with different heart rate settings.

The generation of SR and AF ECGs were conducted separately to capture the different P-QRS-T features present in different pathologies. Generation of SR signals was conducted with pairs of clinical SR ECGs and mathematically modelled ECGs with heart rates ranging from 50 beats per minute (bpm) to 120 bpm without any heart rate variability (**Figure 4.3A**). Generation of AF ECGs was performed with pairs of clinical AF ECGs and mathematically modelled ECGs with heart rates ranging from 60 bpm to 150 bpm and from a 0 to 10 standard deviation to allow for varying heartbeat rhythms (**Figure 4.3B**). To minimize potential bias, all heart rates and heart rate variability values were sampled randomly based on the statistical distribution of the clinical ECG dataset used as the input to the generator.

A. Sinus Rhythm ECG Generation



B. Atrial Fibrillation ECG Generation

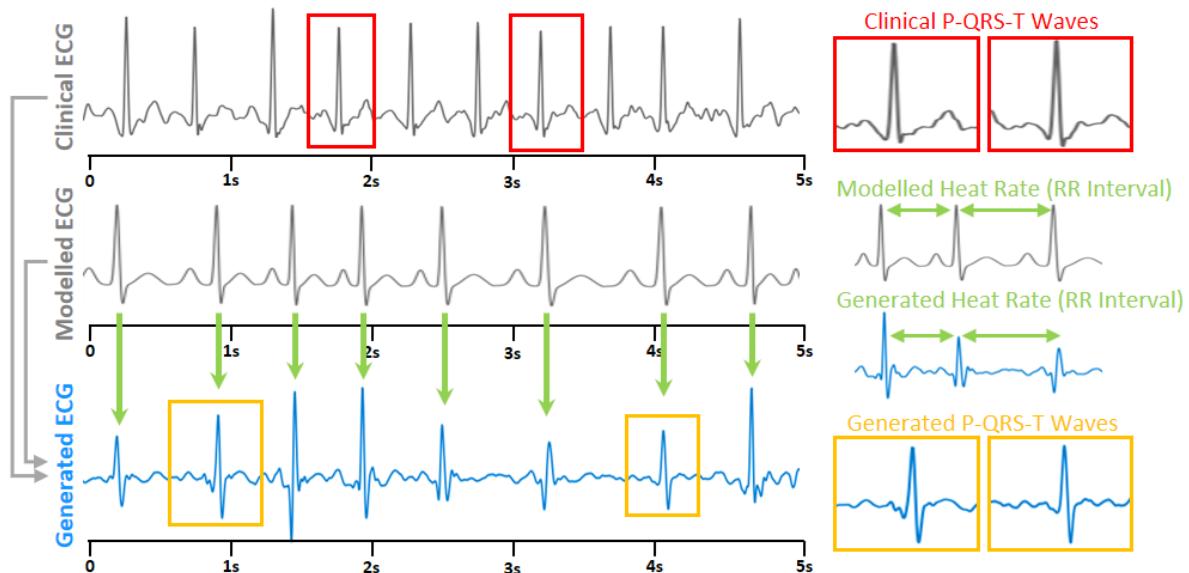


Figure 4.3. Generation of realistic synthetic electrocardiograms (ECG) data for **A**) sinus rhythm (SR) and **B**) atrial fibrillation (AF). The generated ECG (orange) displays the P-QRS-T waveform features of the clinical ECG (red) and the heart rate pattern of the mathematically modelled ECG (green).

4.2.3 Atrial Fibrillation Detection Algorithm

A CRN was used for the detection of AF from raw 1D ECG waveforms [209, 210] (**Figure 4.4**). Since the clinical ECGs were of different lengths, an overlapping sliding window approach was used in which a 5-second window was moved along the entire length

of each ECG in 1-second steps to obtain localized information regarding the detailed P-QRS-T-wave morphologies. The features from each individual window were then collated with a recurrent block to analyse the temporal behaviour of the ECG along the entire time-frame.

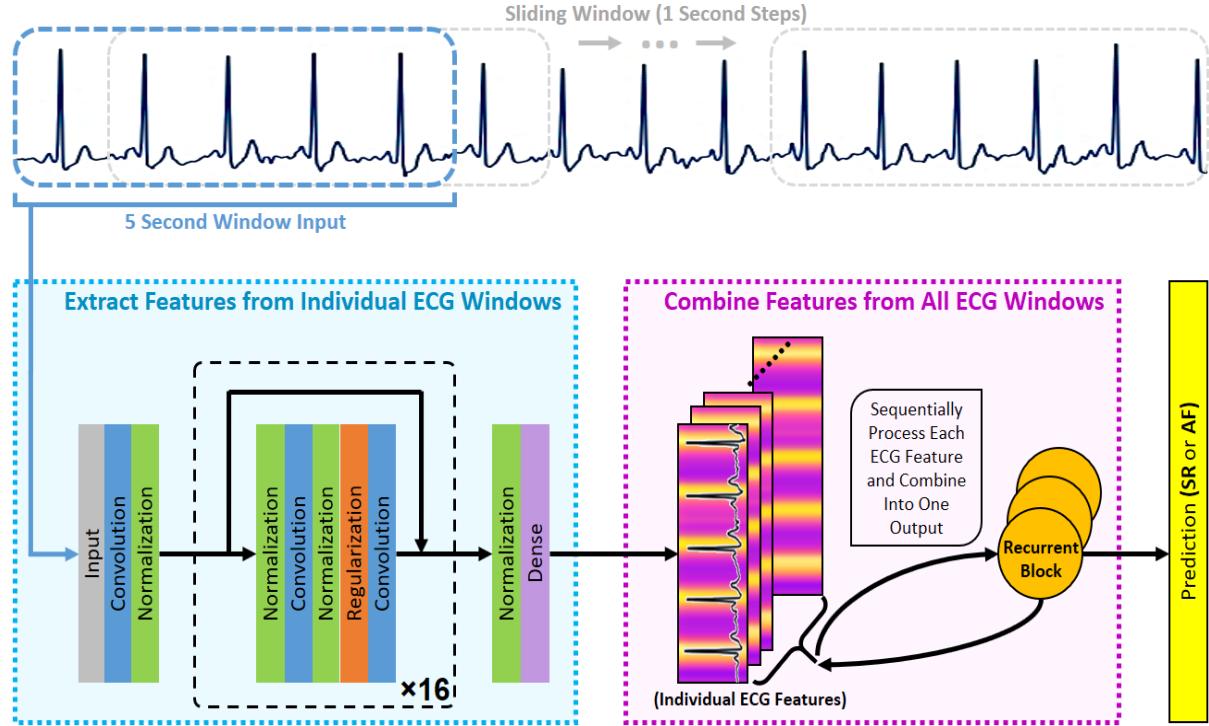


Figure 4.4. Illustration of the deep learning algorithm used for classifying electrocardiograms (ECG). The ECG signal is processed with sliding windows of 5 seconds and then passed through a deep neural network consisting of convolutional layers. Individual features extracted from windows of 5 seconds are then sequentially combined with a recurrent block. The output from the recurrent layers is finally mapped onto a prediction of either normal sinus rhythm (SR) or atrial fibrillation (AF).

The architecture of the convolutional block consisted of 16 residual blocks [152]. In each residual block, the input was first normalized with batch-normalization [211] and a ReLU [212], followed by a convolutional layer for feature extraction. The same normalization was then performed again followed by dropout for regularization [213], and finally another convolutional layer. The convolutional layers were one dimensional, contained a filter size of 16, and had 16 feature maps which doubled for every fourth convolution. Within each residual block, an additional side-connection without any operations was used to increase the efficiency optimization during training by allowing information to flow freely throughout the network in order to counteract the large number of

convolutional layers. After the residual blocks, the output was then compressed into a dense feature map with a fully connected layer. This entire process was repeated for each 5-second ECG window. To collate the individual feature maps of all ECG windows along with the entire signal, 3 long short-term memory recurrent layers each with 512 feature maps [214] were used to sequentially process features maps at each time step. Lastly, the collated features were fed into a softmax activation to obtain the final output which was a prediction for either SR or AF. The network was trained with the Adam optimizer with a learning rate of 0.001 [208] using the cross-entropy loss function. The framework was implemented with Tensorflow [215] on a Titan-V graphics processing unit (GPU) (Nvidia Corporation, California), with each prediction taking approximately 0.1 seconds.

The proposed CRN contained several hyper-parameters which were carefully selected after extensive experimentation. Each hyper-parameter was tuned by keeping all other parameters of the network constant and evaluating the effects of incrementally adjusting the value with 5-fold cross-validation on the training set without augmentation. For each hyper-parameter, the value which resulted in the best cross-validation F_1 score was selected for the final model. Initially, the CRN was tested with 8, 16 or 32 residual blocks, each with two convolutional layers, followed by the use of 2, 3 or 4 recurrent layers. It was observed that as the number of residual blocks was increased from 8 to 16, there was a gradual increase in improvement. However, any subsequent increase in the number of residual blocks from 16 to 32 showed a clear trend in decreasing performance, potentially due to substantial overfitting as a result of the excess number of convolutional layers in the network. Furthermore, the same experiment testing the number of recurrent layers to use also produced a similar trend, where the performance improved from 1 to 3 recurrent layers, but decreased as the numbers of layers were increased from 3 to 6. In the end, 16 residual blocks, or 33 convolutional layers, and 3 recurrent layers were found to contain sufficient parameters to learn the complex features in the ECG signals and perform accurate ECG classification.

The number of parameters in each layer was further tuned to investigate the optimal number of parameters to use in the network by adjusting the filter size of the convolutional layers and the number of feature maps in the recurrent layers. The filter size in each convolutional layer was tuned by testing the entire network with 3×1 , 5×1 , 10×1 and 15×1 sized filter. Filters sized larger than this value was not tested due to the high computational costs which slowed down training and prediction run-time. Experiments testing the filter sizes showed that for 1D input, larger filters were more effective due to the increased receptive field being able to more effectively account for the variations of the signal over

time. Adjusting the number of nodes of the recurrent layer did not result in any significant change as the majority of the processing was done by the residual blocks. Nevertheless, 512 proved to deliver a decent balance between learning capacity and accuracy.

The CRN was validated on the independent testing set to determine the accuracies for the classification of SR and AF with the F_1 score defined in **Chapter 3.6.1**. Experiments were performed to determine the optimal number of synthetic ECGs to generate for training. Feature extraction and analysis were then performed to validate the realism of the synthesized data. Subsequent error analysis was conducted using a confusion matrix, and the incorrectly classified signals in the testing set were extracted for further analysis.

4.3 Results

4.3.1 Synthetic ECG Data

The synthetic ECGs showed that the generator framework was successful in reproducing important features present in real ECG recordings. The generated SR ECGs maintained well-defined P-waves with the RR intervals being consistent throughout the signal. The generated AF ECGs, on the other hand, contained F-waves or a lack of P-waves, and varying R-peak-to-R-peak distances resembling real cases of ECGs obtained from patients with AF. The size of the ECG dataset available for training was increased five-fold by generating a set of synthetic ECGs 400% the size of the training set. This resulted in a total of 11,668 new ECGs available for training including 10,136 SR samples and 1,532 AF samples.

4.3.2 Validation of Synthetic ECGs

To compare the synthesized ECGs against the clinically recorded ECGs, we performed feature extraction on the signals. We extracted the individual P-QRS-T cycles for each heartbeat and overlaid the waveforms for both sets of signals (**Figure 4.5A**). The clinically recorded ECGs had a smaller variation between beats as the P-QRS-T cycles coincided with each other fairly closely without much variation in the amplitude. However, there was a much larger variation for the synthesized ECGs as seen by the larger discrepancies in the amplitude as well as the temporal positioning of the P and T waves

between different beats. We discovered that synthesized ECGs contained significantly higher standard deviations for SR and AF (**Figure 4.5B**). The MSE was then utilized to measure the average difference between synthetic ECG and its original clinical ECG (**Figure 4.5C**). As expected, although there was a difference in signal amplitudes, MSE decreased gradually as more clinical ECGs were used for the data generation process. Our frequency analysis using Fourier transform showed that the synthetic ECGs contained a similar frequency distribution compared to the clinically recorded ECGs, demonstrating that the generator framework was able to effectively capture frequency features in the dataset (**Figure 4.5D**).

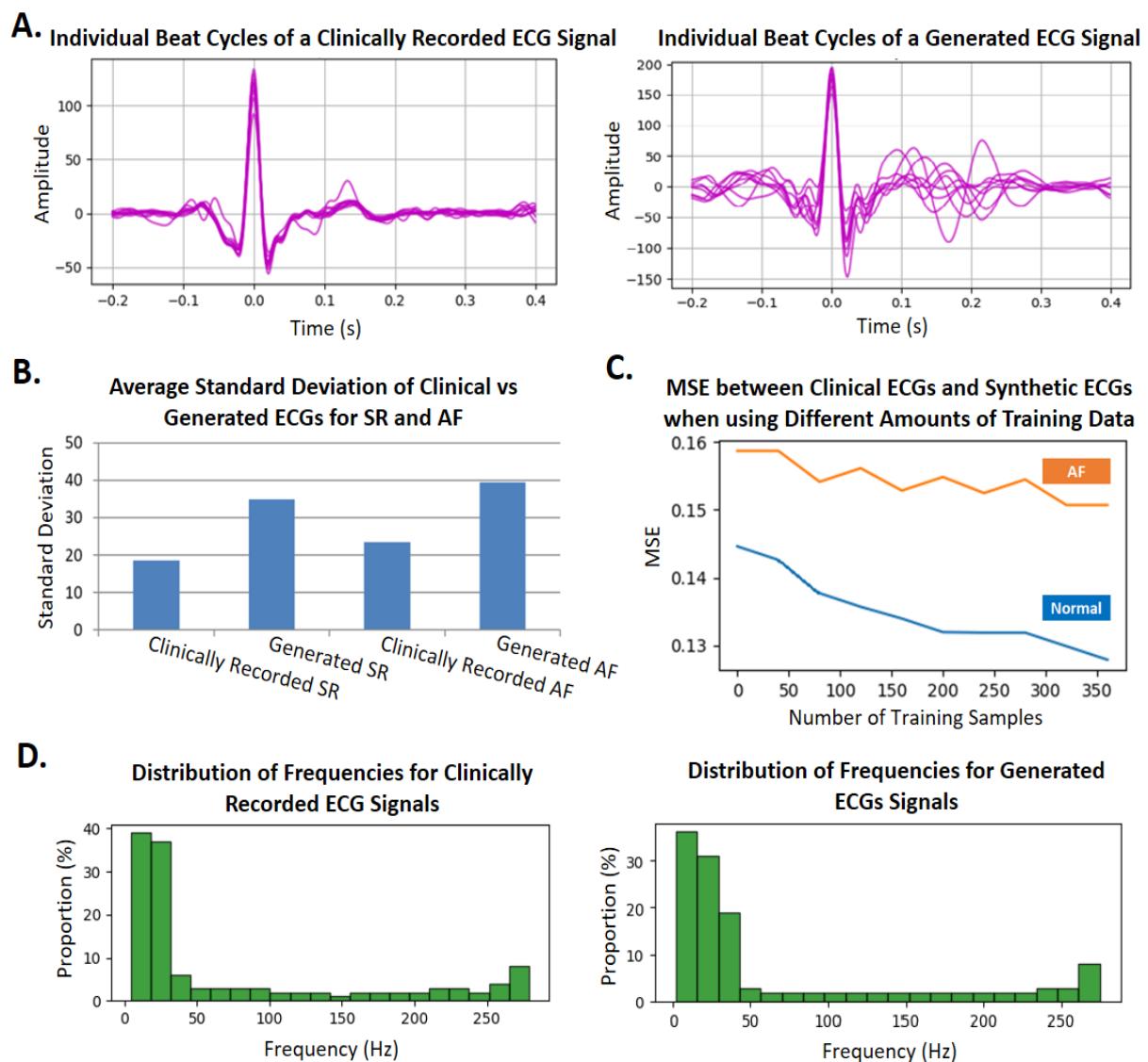


Figure 4.5. Validation of the synthetic electrocardiograms (ECG). **A)** An overlay of individual ECG P-QRS-T wave cycles extracted from each beat in clinically recorded and synthetic ECGs. **B)** The average standard deviation of clinically recorded and synthetic ECGs for sinus rhythm (SR) and atrial fibrillation (AF) signals. **C)** The mean squared error (MSE)

between the clinically recorded and synthetic ECGs plotted against increasing the number of clinically recorded ECGs used. **D)** Comparisons of the averaged frequency distributions obtained from the clinically recorded and generated ECG signals.

The generalizability of the data generated was validated further through several experiments to demonstrate the accuracy and effectiveness of the ECG generator in capturing and reproducing features in the original clinical ECG dataset. Our augmentation strategy resulted in superior performance regardless of the architecture or method used. Comparisons of the accuracy of our CRN with different numbers of convolutional layers ranging from 9 to 33 also showed that the augmentation approach improved the classification accuracy regardless of the configuration used. Finally, we trained our CRN with only synthetic data and the results were an accuracy of 93.5% compared to 94.8% using a CRN trained on only real clinical data. This indicated that although there was a decrease in accuracy, overall, the generated data contained sufficient features to train the network effectively even when no real ECG samples were present.

4.3.3 Atrial Fibrillation Detection Results

The proposed CRN achieved an overall F_1 score of 96.4% on the independent testing dataset, with 99.1% for SR and 93.8% for AF (**Figure 4.6**). Training experiments showed that this was obtained when the size of the synthetic ECG sample used was 300% that of the real clinical ECG dataset, creating a 1.6% improvement in the F_1 score and 3% improvement in AF detection compared without using any synthetic data for training. Results also showed that there was a steady improvement as the amount of synthetic ECG data used for training was incrementally increased from 0% to 300% of the clinical dataset and the subsequent increase from 300% to 400% resulted in a declining accuracy.

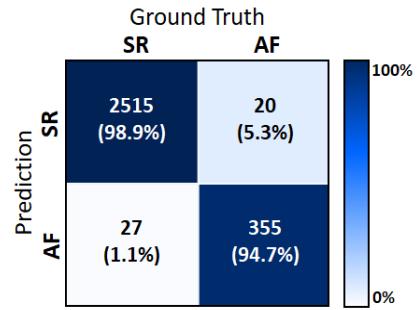
Error analysis via a confusion matrix revealed the greatest proportion of error was due to a reducing in specificity which caused a mislabelling of SR ECGs as AF (**Figure 4.7**). To analyse the discrepancies between the CRN predictions and expert ground truth annotations, ECGs incorrectly classified by the CRN were visualized. **Figure 4.7A** contains two samples of SR misclassified as AF, containing substantial noise blurring the P-QRS-T waveforms. The presence of an inconsistent R-peak causing an inconsistent heart rate in SR also contributed to the difficulty in classifying the ECG. **Figure 4.7B** contains a sample of AF

predicted as SR. The signal contained a fairly consistent heart rate throughout the entire recording apart from small segments which had minor varying RR intervals, potentially being difficult to detect by the CRN.

A. ECG Detection Accuracy Experimental Results

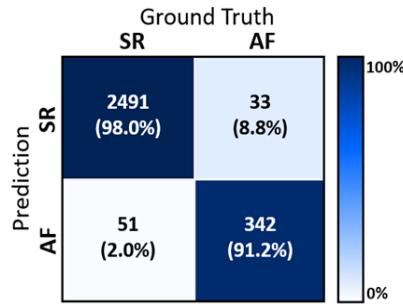
Experiment (Amount of Data Generated)	NSR Accuracy (%)	AF Accuracy (%)	Overall (F ₁ Score)
Baseline (No Generated Data)	98.71	90.86	94.8
50% More Data	98.79	91.34	95.1
100% More Data	98.94	92.76	95.9
150% More Data	98.98	93.09	96.0
200% More Data	98.99	93.15	96.1
250% More Data	98.98	93.32	96.2
300% More Data	99.07	93.79	96.4
350% More Data	98.92	92.73	95.8
400% More Data	98.79	92.13	95.5

B. Confusion Matrix for the Best Model



C.

Confusion Matrix of Prediction Made by the DNN when Trained on Real Data



Confusion Matrix of Prediction Made by the DNN when Trained on Synthetic Data

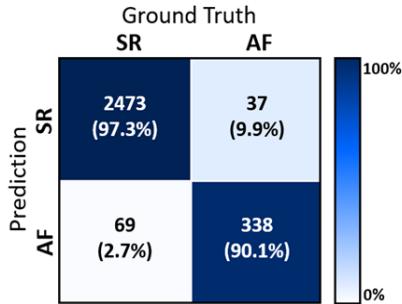


Figure 4.6. Accuracy evaluation of the deep learning algorithm for electrocardiogram (ECG) classification for sinus rhythm (SR) and atrial fibrillation (AF). **A)** Experimental results for the accuracy obtained when different amounts of generated ECGs are used for training compared to the currently available clinical ECG training set. An increase of 300% was shown to be the optimal number of generated ECGs. **B)** Confusion matrix of the best performing algorithm for error analysis and diagnostics. Columns represent the expert annotations and rows represent the prediction made by the deep learning algorithm. The colour map is normalized between 0% and 100%. **C)** Confusion matrix when evaluating the testing set predictions of our proposed deep neural network (DNN) trained on only clinically recorded ECGs and trained on only synthetic ECGs.

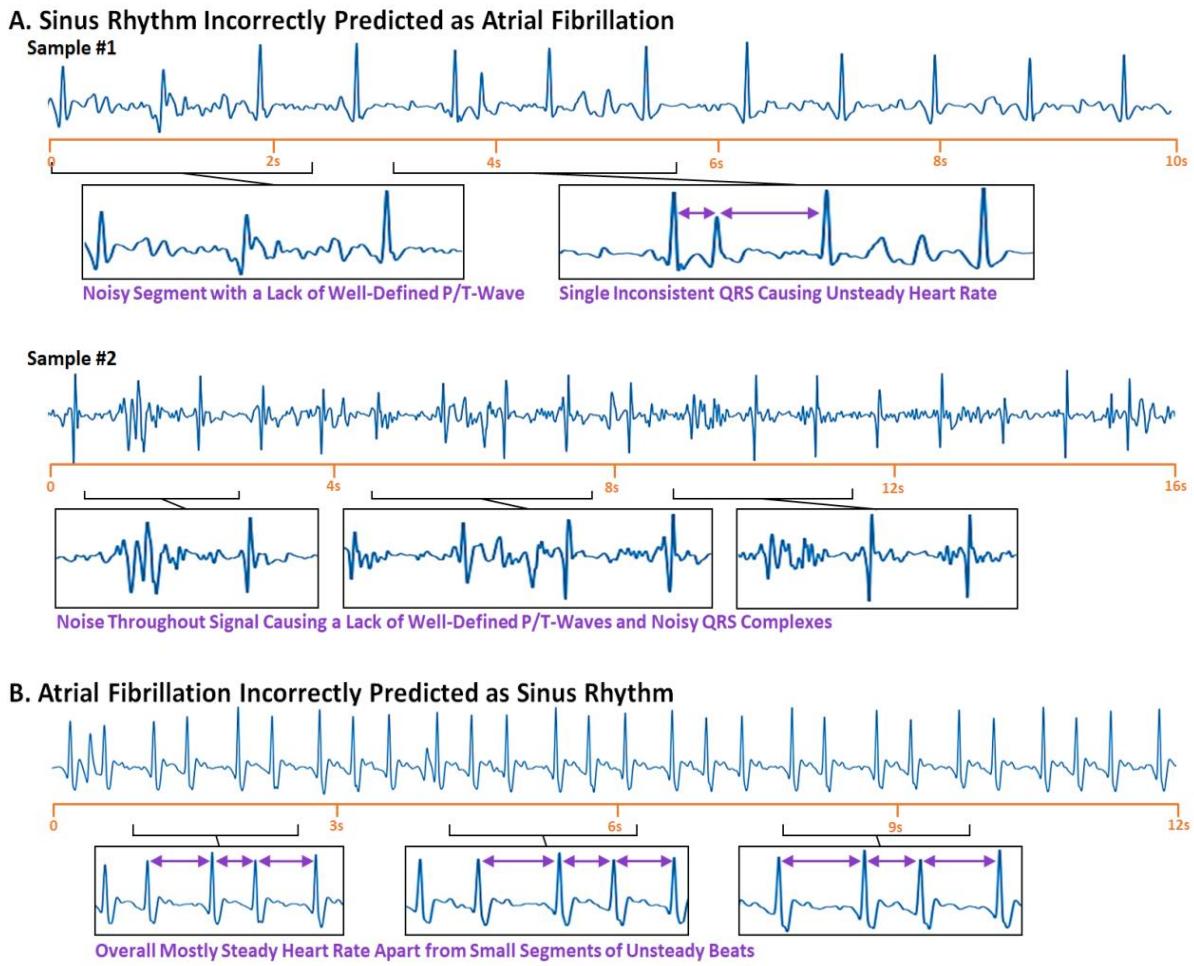


Figure 4.7. Visualization of electrocardiograms (ECG) incorrectly classified by the deep learning algorithm. **A)** Samples of normal sinus rhythm (SR) ECGs mistakenly classified as atrial fibrillation (AF). The enlarged regions with a high degree of noise and irregular R-peak-to-R-peak intervals may contribute to an erroneous prediction. **B)** A sample of AF ECG mistakenly classified as SR. The overall regularity of the heart rate creates difficulties in detecting the small segments of irregular R-peak-to-R-peak intervals (enlarged regions).

4.4 Discussion

Analysis of ECGs for the detection of AF is a challenging task due to the high variability, complexity, and various noise present in ECGs. Our study presents a CRN capable of classifying ECG signals into SR and AF with an accuracy of 96.4%. Furthermore, we have proposed a method of generating physiological accurate synthetic ECGs to aid CRN training, and showed the subsequent improvement in the accuracy of AF detection by 3%. More importantly, our algorithm successfully identified 94.7% of ECGs with AF occurrences

while minimizing false alerts to approximately 1%. By embedding our approach into a mobile application, we believe it may potentially be a more efficient method for monitoring patients with AF, enabling earlier detection and prevention of the serious complications of AF.

4.4.1 Methods of AF Detection

The lack of portability and accessibility of the most widely used 12-lead recording device has led to the development of more efficient, self-monitoring mobile devices [216, 217]. In recent years, the most popular mobile technologies for AF detection are based on either photo-plethysmography or ECG approaches [25, 203, 218]. However, despite the advantages of automated patient monitoring, these devices suffer from a lack of patient samples, particularly AF patients in the trial cohorts. A recent study utilizing the Apple Watch for ECG classification confirmed only around 150 cases of AF from a study cohort of over 400,000 patients [25]. The low prevalence of AF in these cohorts, therefore, is a major limiting factor when considering the development of intelligent algorithms such as CRNs for more accurate AF detection.

Our study aimed to solve this issue with a novel method of ECG generation to provide our CRN with a large number of samples for effective training. By applying our approach, we artificially increased our current sample of 383 AF training cases to over 1,500 which is significantly more than the aforementioned studies.

4.4.2 Deep Learning for ECG Classification

The CRN proposed in this study achieved accuracies for AF detection exceeding most current large-scale studies involving commercially developed ECG and photo-plethysmography smart devices. While direct comparisons are difficult due to varying study designs and methods, we show that our approach achieved a 93.8% accuracy. This was an improvement compared to the 91.7% achieved by Valiaho et al. [218] and 91.7% achieved by Guo et al. using the Huawei Watch [203]. Our results were also competitive with that of Bumgarner et al., who reported a 93% accuracy using the ECGs recorded from the Kardia Band [26].

The ability of the proposed CRN to directly learn from raw ECG signals was another

advantage. While it has been shown that the structural changes caused by AF are likely to lead to changes in the P-QRS-T formation [116], clinical studies on AF patients with regional non-sinus contractions of the left atrial appendage occasionally display apparent SR on ECG recording [219, 220]. This occurrence may be explained by the fact that regional non-sinus electrical activities are actually reflected through waveform features smaller than readily observable P-waves. By training a CRN on sufficient clinical ECG samples, subtle features not typically considered by trained professionals may be implicitly learned and result in a more robust framework.

4.4.3 Limitations and Future Work

A limitation of this study was the small sample size of the clinical ECG dataset. While the sample of approximately 6,000 ECG recordings is currently, to our knowledge, one of the largest publicly accessible datasets available, it is still incredibly small compared to today's standards of deep learning [152]. Larger datasets would also be beneficial for performing more extensive validation by partitioning the dataset into additional testing sets which were independent from the data used for parameter selection during model development. Although our study aimed to tackle this problem through synthetic data generation, it is only possible for the generated data to contain features already present in the clinical ECGs. While results showed an improvement in the AF detection accuracy, the availability of more data samples may result in higher performances. This would be particularly beneficial for evaluating the generalizability and optimality of the proposed network and the amount of synthetic data generated. Furthermore, since the generator exploits features in existing clinical ECGs, the generated data may contain flaws that are inherently present, such as the presence of excessive noise or signals which lack sufficient features to be classified accurately.

In the future, we hope to develop more robust methods of AF detection from ECGs by utilizing more advanced CRNs, and potentially expand our approach to detect other types of rhythm disturbances. By validating our approach on a larger clinical dataset, we may further improve the performance of our algorithms as well as explore opportunities and applications for data synthesis.

4.5 Conclusions

In this chapter, we have developed and validated a CRN on an open dataset for the detection of AF from single-lead ECGs. The CRN achieved an accuracy of 96.4% when validated on clinical ECG recordings and our novel method for synthetic ECG generation resulted in a 3% improvement in the accuracy of AF detection. For the clinical prospect, by embedding our approach into a mobile device, the proposed framework could aid AF diagnosis and be used for patient self-monitoring to improve the early detection of AF. Potentially, improved AF detection could lead to effective treatment and monitoring of patients suffering from AF worldwide.

Chapter 5

Automatic Left Atrium Segmentation from LGE-MRI with Convolutional Neural Networks

Current treatments for AF remain suboptimal due to a lack of understanding of the underlying atrial structures that directly sustain AF, preventing the accurate identification of ablation targets for patients with persistent and longstanding-persistent AF. Existing approaches for analysing atrial structures, especially from LGE-MRIs, rely heavily on manual methods which are extremely labour-intensive and prone to errors. As a result, a robust and automated method for analysing atrial structures in 3D is of high interest.

This chapter presents a preliminary study on designing a CNN for fully automatic LA segmentation as the first step in achieving the challenging task of automatic atrial analysis from medical imaging. We detailed the current state of the field of LA segmentation, and brought to attention the current shortcomings which led to the need of more intelligent algorithms. We defined the LGE-MRI dataset from the University of Utah which was the main imaging dataset used in this thesis. Initial segmentation results were also evaluated to obtain a solid baseline as this was the first attempt of using deep learning for the task.

5.1 Introduction

Current treatments for AF remain suboptimal due to a lack of understanding of the underlying atrial structures that directly sustain AF [20, 21]. As a result, direct study of the atrial structure and its changes in patients with AF is vital to the understanding and treatment of AF. LA segmentation is a crucial task for aiding medical management for AF patients based on structural analysis of the segmented 3D geometry, and has led to many prior studies investigating algorithms for fully automatic LA segmentation. A 2013 study published by Tobon-Gomez et al. compared the performance of nine different algorithms for LA

segmentation from non-gadolinium enhanced MRIs/CT and showed that methodologies combining statistical models with regional growing approaches were the most effective when combined with the well-researched multi-atlas registration method [221]. Similar techniques have also been proposed and further improved upon for segmenting the LA from LGE-MRIs in studies by Veni et al. [36], Zhu et al. [222] and Tao et al. [37]. These studies utilized shape models enhanced with Bayesian optimization and variation regional growth, however, should only be considered as semi-automatic approaches as they require shape priors in the form of rough manual segmentation for initialization.

Despite these recent efforts, most of the existing structural analysis studies, especially those that utilize clinical LGE-MRIs, are still based on labour-intensive and error-prone manual segmentation approaches [28, 29, 35]. This is mainly due to the limitations of current automated methodologies requiring supporting information which are often not available, such as shape priors for initialization [222, 223] or additional magnetic resonance angiography sequences to aid the segmentation process [37]. The nine algorithms mentioned in the study [221] were effective in segmenting the LA from non-gadolinium enhanced MRIs. However, they are also difficult to apply to the LGE-MRIs directly, as the (normal) atrial anatomy is more attenuated by the contrast agent and they are based on conventional approaches. There is therefore an urgent need for an intelligent algorithm that can perform fully automatic atrial segmentation directly from LGE-MRIs, particularly for LA, for accurate reconstruction and measurement of the atrial geometry for clinical usage.

In this chapter, we propose and evaluate a novel CNN for fully automatic LA segmentation. Our method is developed and validated on the largest 3D LGE-MRI dataset from University of Utah, and was a very important step towards patient-specific diagnostics and treatment.

5.2 Methods

5.2.1 Dataset and Labels

The data used in this study consisted of 154 LGE-MRIs from the University of Utah, Utah, United States [28]. The dataset contained 60 patients with AF prior to and 3 to 27 months post clinical ablation (**Table 5.1**). The imaging resolution was $0.625 \text{ mm} \times 0.625 \text{ mm} \times 1.25 \text{ mm}$. Each 3D LGE-MRI scan contained 44 slices along the Z-axis, each with an XY

spatial size of 640×640 pixels or 576×576 pixels. The image acquisition protocol can be found in the original study from the Utah group.

Table 5.1. Summary of the 3D late gadolinium-enhanced (LGE)-MRI dataset from the University of Utah, split by pre- and post-catheter ablation.

	Number of Patients	Pre-Ablation Scans	Post-Ablation Scans	Total
Utah LGE-MRI Dataset	60	60	94	154

To study the LA anatomical structure, the team at Utah manually delineated the LA geometry of each patient. Each 3D LGE-MRI patient data (both pre and post ablation) included two manually segmented masks of the LA cavity and wall by experts [29]. The LA was manually segmented in consensus with three trained observers for each LGE-MRI scan to obtain one segmentation per scan. The segmentation protocol was as follows, and was done in a 2D slice-by-slice manner across all LGE-MRI slices (**Figure 5.1**). Firstly, the endocardium was defined by manually tracing the LA blood pool including the PVs in each slice of the LGE-MRI. The four PVs were included for the LA, and was limited to the PV antrum region and extended to the point where the PVs stopped narrowing. On average, the PV antra were limited to less than 10 mm extending out from the endocardial surface, or approximately three times the thickness of the LA wall. The mitral valve connecting the LA and LV was defined by a 3D plane to create a smooth linear surface. The LA endocardium was then morphologically dilated and manually adjusted according to the LA geometry to obtain the boundary of the epicardium. Finally, the dilated PV and mitral valve were manually removed from the LA epicardium.

In this chapter, the 154 3D LGE-MRI dataset was randomly split into training ($N = 110$), validation ($N = 22$) and testing ($N = 22$) sets. Since each patient had multiple 3D LGE-MRIs for pre and post ablation, the data was split so that all scans from each unique patient were only in one of the training, validation or testing sets. Evaluation could therefore be performed to compare whether pre or post ablation impacts the segmentation accuracy. All individual raw 3D LGE-MRI data were normalized by using its mean and standard deviations of colour intensity.

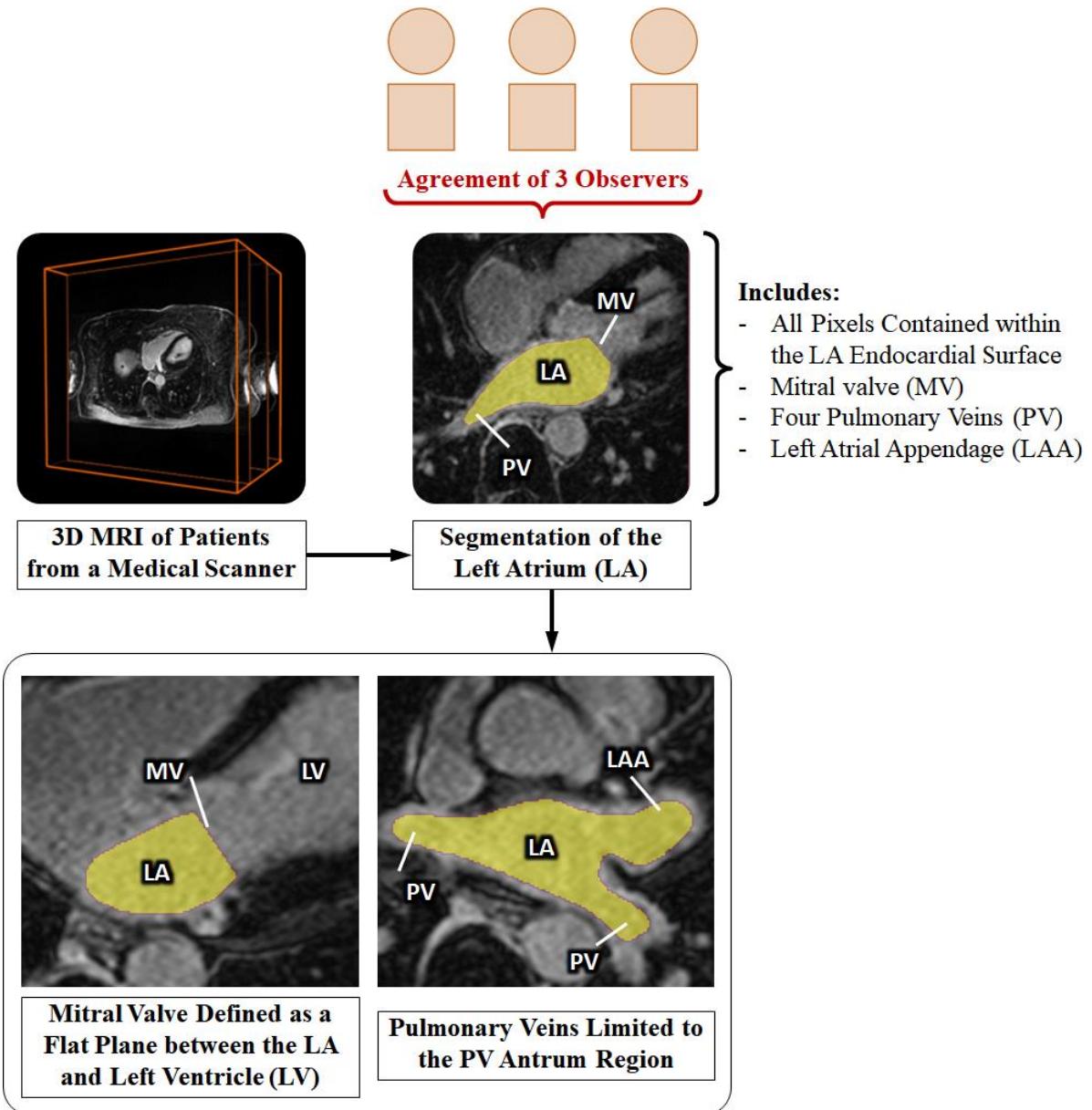


Figure 5.1. Illustration and description of the manual segmentation procedure, defining the workflow and anatomical extents of the left atrial (LA) chamber adapted from Xiong et al. 2021 [224]. LAA, left atrial appendage; LV, left ventricle. MV, mitral valve; PV, pulmonary veins.

5.2.2 Fully Convolutional Network for Segmentation

As a first attempt to tackle the challenging task, we propose and evaluate a novel FCN for fully automatic LA segmentation from LGE-MRIs. The FCN contains two parallel pathways to process both local and global information simultaneously and alleviate the massive class imbalance in the atrial images. The full architecture can be seen in **Figure 5.2**.

Our network performs 2D patch-wise classification at every 15×15 region across each slice of a 3D LGE-MRI. The inputs into the network are a local 41×41 patch, and a global 279×279 patch, both centred on the 15×15 prediction region. The global pathway captures information about the position and gross structure of LA. To reduce the number of parameters used in the resultant network, the large global patch is pooled to a smaller size. The local pathway captures the exact geometry of LA for every pixel within its small neighbourhood. The two patches are separately convolved through 13 convolutional layers (conv 1 to conv 13) and are merged to combine the information by performing an element-wise sum. To maintain dimensional consistency, the last layer in the global path is unpooled to match the size of the local pathway with nearest neighbour interpolation. The network is then convolved two more times (conv 14/conv 15) to further process the combined information where the output is produced in the final layer with two feature maps (conv 16). The ReLU function is used for all layers apart from the last layer which uses a softmax activation function. Dropout is applied to convolutional layers 14 and 15 to reduce overfitting with a dropout rate of 50%. During testing, the network scans through each 15×15 region for every slice of a 3D LGE-MRI without overlapping, and feeds the corresponding 41×41 and 279×279 patches centred around it as input.

The adaptive moment estimation, or Adam, optimizer [225], a type of gradient descent algorithm, was used for optimization. The learning rate was kept constant at 0.0001 without adjustment and the exponential decay rates of the 1st and 2nd moment estimates were set to 0.9 and 0.999, respectively. During training, the accuracy was evaluated on the validation dataset after each iteration of all the training data through the network. This was repeated until the validation accuracy stopped increasing, and the best performing model was selected for evaluation on the test set. The network was developed in Tensorflow [226], an open-source deep-learning library for Python, and was trained on an NVIDIA Titan X-Pascal GPU with 3840 CUDA cores and 12GB RAM. The training phase took approximately four hours and predictions on each 3D LGE-MRI took around one minute to compute.

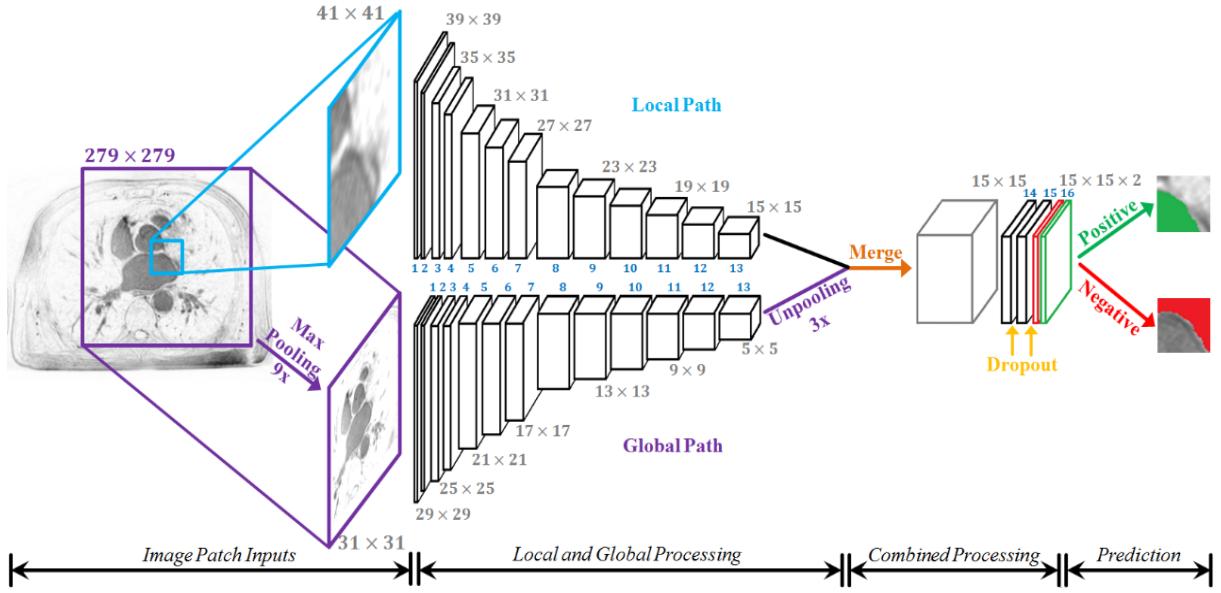


Figure 5.2. The architecture of the proposed dual fully convolutional network for left atrial segmentation. The size of the image at every second layer is shown. The parallel (global and local) pathways process each MRI slice at different resolutions, which are combined at the end of the network. The final output has two feature maps, denoting the probability of a positive or negative pixel classification for each 15×15 patch, respectively.

5.2.3 Evaluation

Evaluation against the ground truths was done to evaluate the accuracy of the FCN for each individual 3D LGE-MRI data. An evaluation was performed for the LA cavity, and the LA wall was evaluated with the inclusion of the cavity. The Dice score, sensitivity, specificity, and Hausdorff distance, defined in **Chapter 3.6.1**, were used for technical evaluation. The LA diameter and volume errors, defined in **Chapter 3.6.2**, were used for biological evaluation.

5.3 Results

Table 5.2 shows the final evaluation metrics for both of the segmentation tasks grouped by pre and post ablation LGE-MRIs. The Dice segmentation accuracy was approximately equal for both the wall and cavity (91.0% vs 91.2%), however, the cavity segmentation had a slightly higher sensitivity of 90.8%, which implies that a slightly greater proportion of the positive pixels were successfully detected compared to that of the wall

segmentation. The 99.9% specificity suggests that our network was extremely effective in detecting background pixels, and the >90% sensitivity shows our network was also very effective in detecting the positive pixels for accurate segmentation.

Table 5.3 shows the predicted and ground truth measurements for the LA diameter and volume, as well as their absolute and relative errors. The predicted masks were accurate within 1.59 mm and 4.01 cm³ of the ground truths on average for the diameter and volume measurements respectively. Overall, the segmentations for the pre-ablation patients were more accurate as seen from the higher Dice scores and lower relative errors of the estimated LA dimensions.

Table 5.2. Evaluation metrics for left atrial (LA) wall and cavity segmentation for the 22 test 3D late gadolinium-enhanced (LGE)-MRIs.

		DICE	Sensitivity	Specificity	Hausdorff Distance
LA Wall(+Cavity)	Pre	91.3 ± 0.8 %	90.0 ± 4.5 %	99.9 ± 0.1 %	10.64 ± 2.08 mm
	Post	90.7 ± 1.9 %	90.2 ± 4.5 %	99.9 ± 0.1 %	10.76 ± 2.43 mm
	Overall	91.0 ± 1.4 %	90.1 ± 4.4 %	99.9 ± 0.1 %	10.70 ± 2.21 mm
LA Cavity	Pre	91.5 ± 0.8 %	90.1 ± 3.5 %	99.9 ± 0.1 %	10.43 ± 2.29 mm
	Post	90.9 ± 1.9 %	91.6 ± 3.8 %	99.9 ± 0.1 %	10.07 ± 2.27 mm
	Overall	91.2 ± 1.4 %	90.8 ± 3.6 %	99.9 ± 0.1 %	10.25 ± 2.24 mm

Table 5.3. The left atrial (LA) diameter and volume measurements and errors for the 22 test 3D late gadolinium-enhanced (LGE)-MRIs.

		Ground Truth Measurement	Predicted Measurement	Absolute Error	Relative Error
LA Diameter (mm)	Pre	40.85 ± 7.07	40.17 ± 7.59	1.36 ± 1.18	3.3 %
	Post	39.43 ± 7.22	39.78 ± 7.48	1.82 ± 2.23	4.6 %
	Overall	40.14 ± 7.01	39.97 ± 7.36	1.59 ± 1.75	3.9 %
LA Volume (cm ³)	Pre	56.98 ± 19.73	58.50 ± 21.15	3.35 ± 2.91	5.9 %
	Post	50.56 ± 17.99	54.18 ± 20.71	4.67 ± 4.09	9.2 %
	Overall	53.77 ± 18.72	56.34 ± 20.55	4.01 ± 3.50	7.5 %

5.4 Discussion

5.4.1 Left Atrium Segmentation from LGE-MRIs

Direct segmentation of atrial chambers from 3D LGE-MRIs is a challenging task.

This is due to a lack of distinguishing features on the LGE-MRIs that enable the separation of atrial tissue from non-atrial regions, in addition to the poor image quality of LGE-MRIs due to motion artefacts, noise interference and significant variations in image intensities.

Currently, our study is among the few that have attempted at direct automatic segmentation of the LA from LGE-MRIs [36, 227, 228]. Out of all existing attempts, most of the LGE-MRI studies, either in-vivo or ex-vivo, have relied heavily on manual segmentation [28, 30, 229, 230]. There are few studies that have attempted to automate the direct segmentation process. The study by Veni et al. proposed was a shape-driven approach in which the posterior probability of the LA surface was expressed via Bayesian equations [36]. Zhu et al. also proposed a shape model for segmenting the LA and was enhanced by variational regional growth [222]. However, the methodologies of both of these studies required a shape prior for initialization which was generated either manually, or based on strong assumptions from observing the data to segment. This is undesirable as it would be more time consuming than a fully automated approach and is also more susceptible to large variations in the data. Tao et al., on the other hand, proposed an approach which used global multi-atlas segmentation followed by a local refinement algorithm [37]. In contrast to the three previous studies, our methods do not require any manual initialization or the use of any additional information apart from LGE-MRIs. Furthermore, benchmarking results of our network showed significant improvements in the accuracy of the LA reconstruction, as well as significant computational advantages.

In contrast to direct segmentation, indirect segmentation of LA from LGE-MRIs is a more popular approach that can be achieved by utilizing paired 3D MRA with the LGE-MRIs from the same patient. The 3D MRA images consist of image features that are more distinguishable than those in the LGE-MRIs, which enable the distinction between atrial tissue and background. As a result, the 3D MRA images are relatively easy to segment, for which comparably more studies have been conducted to develop automatic segmentation approaches for [38, 39, 221, 222]. Although some alternative conventional approaches outlined in a segmentation benchmarking study [39, 221] and the multi-view CNN approach conducted by Mortazi et al. [38] have achieved good LA segmentation performances, a mapping/fusion approach was needed in addition to their indirect segmentation approach to map the segmented masks of MRA images to the LGE-MRIs. This could introduce additional errors which limit the accuracy of such approaches [39]. In addition, the use of MRA images in indirect segmentation of LGE-MRI means that an extra 3D MRA scan is needed for each patient, which is time consuming and rather costly. In comparison, direct segmentation of

LGE-MRIs using our proposed CNN is much more straightforward, efficient and accurate.

5.4.2 Multi-Scale Processing in Convolutional Neural Networks

Multi-scale processing, such as the one used in this chapter, has been explored in previous studies for enhancing the performance of neural networks in other research fields [231-233]. The CNNs proposed by Dou et al. for performing lung nodule classification [234] and by Kamnitsas et al. for brain lesion segmentation [52] both contained multiple pathways for different input resolutions. Despite the similarities of their designs and our proposed approach, our FCN contains a significantly greater number of layers and feature maps in each layer, allowing greater feature learning for segmenting more complex geometries such as the LA wall and cavity. Furthermore, our experimental results showed that our FCN was able to perform high quality slice-by-slice 2D segmentation to achieve accurate 3D reconstruction of the atrial geometry, meaning it is effective for both 2D and 3D tasks.

5.4.3 Limitations and Future Work

The accuracy of our network could possibly be further improved by applying shape constraints which would be imposed on either the intermediate layers or the output to control the generated 3D geometry. This would especially improve the segmentation at the mitral valve which connects the LA with the LV as this region is arbitrarily cut by a straight line in the ground truth masks. Our network attempts to segment the mitral valve region with a smooth rounded shape, which leads to a poor performance value which when evaluated. This issue could potentially be alleviated by manually re-labelling the ground truths masks to improve the definition of the mitral valve, which in turn, will improve the quality of the samples provided during training. The accuracy of the LA 3D reconstruction could also be improved by considering the 3D geometry and continuity between slices. A simple method to achieve this is to incorporate multiple slices as additional channels at the input of our network. However, our preliminary experiments showed that our FCN with three channelled or five channelled inputs performed worse and had substantially greater computational and memory costs. Further attempts at considering the continuity of the LA geometry in 3D warrant future investigation.

5.5 Conclusions

In this chapter, we have developed and evaluated a dual FCN for robust automatic LA segmentation from LGE-MRIs. Our algorithm enables the reconstruction of LA in 3D with a Dice accuracy of 91% as well as accurate estimates of key clinical measurements. The evaluation of our pipeline demonstrates that it is superior to previously proposed methods, setting a new benchmark for future studies. Our study may lead to the development of a more accurate and efficient atrial reconstruction and analysis approach, which can potentially be used for much improved clinical diagnosis, patient stratification and clinical guidance during ablation treatment for patients with AF.

Chapter 6

A Global Benchmark of Algorithms for Segmenting the Left Atrium from LGE-MRI

Following the preliminary development of a deep learning algorithm for LA segmentation from LGE-MRIs in **Chapter 5**, we aimed to improve the pipeline to maximize its accuracy for clinical use.

To achieve this, we organized the 2018 LA Segmentation Challenge using the world's largest atrial LGE-MRI dataset (described in **Chapter 5.2.1**) and ultimately attracted the participation of 27 international teams. Extensive analysis of the submitted algorithms using technical and biological metrics was performed by undergoing subgroup analysis and conducting hyper-parameter analysis, offering an overall picture of the major design choices of CNNs and practical considerations for achieving state-of-the-art LA segmentation. This large-scale benchmarking study made a significant step towards much-improved segmentation methods for atrial LGE-MRIs, and would serve as an important benchmark for evaluating and comparing the future works in the field. Furthermore, the findings from this study could potentially be extended to other imaging datasets and modalities, having an impact on the wider medical imaging community. This chapter is presented as the original published version, cited as:

Xiong, Z., Xia, Q., Hu, Z., Huang, N., Bian, C., Zheng, Y., Vesal, S., Ravikumar, N., Maier, A., Yang, X., Heng, P.A., Ni, D., Li, C., Tong, Q., Si, W., Puybareau, E., Khoudli, Y., Géraud, T., Chen, C., Bai, W., Rueckert, D., Xu, L., Zhuang, X., Luo, X., Jia, S., Sermesant, M., Liu, Y., Wang, K., Borra, D., Masci, A., Corsi, C., de Vente, C., Veta, M., Karim, R., Preetha, C., Engelhardt, S., Qiao, M., Wang, Y., Tao, Q., Nuñez-Garcia, M., Camara, O., Savioli, N., Lamata, P., Zhao, J., 2021. A Global Benchmark of Algorithms

for Segmenting the Left Atrium from Late Gadolinium-Enhanced Cardiac Magnetic Resonance Imaging. *Medical Image Analysis*, 67, 101832.

6.1 Introduction

Direct segmentation and analysis of the LA from LGE-MRIs remain challenging due to the varying intensities caused by the increased enhancement of fibrotic tissue and imaging artefacts, as well as the varying quality of imaging. In recent years, many benchmarking studies have been conducted to investigate methods for automatic 3D segmentation of the LA and other cardiac structures from medical images, mostly for non-contrast imaging. A 2013 benchmarking study held in conjunction with Medical Image Computing and Computer Assisted Intervention (MICCAI) examined methods for automatically segmenting the LA from non-contrast MRIs and CTs ($N = 30$ each) [221]. A more recent 2017 benchmarking study for segmentation of LV and RV from non-enhanced MRIs showed that CNNs significantly outperformed traditional methods [235]. By analysing algorithms from 10 participants, the study revealed that the popular U-Net CNN architecture [161] which was specifically designed for medical image segmentation was particularly effective. The top U-Net implementation achieved a Dice score of 93.1% for the LV and 89.9% for the RV, with Hausdorff distances of 6.9 mm and 12.2 mm, respectively.

Contrary to non-enhanced images, contrast-enhanced MRIs/CTs have received significantly less attention in terms of researching methods of segmentation to aid clinical procedures. A 2016 study was conducted to investigate methods of LA wall segmentation from LGE-MRIs and contrast-enhanced CTs ($N = 10$ each) by analysing the submitted algorithms from three groups [109]. However, due to the limited number of submitted algorithms, the study was unable to draw any definitive conclusions in terms of algorithm development, particularly with the three methods achieving Dice scores of approximately 55%. Other studies on LGE-MRI segmentation also have limited efficacy as the methods proposed required manually initialized shape priors [222, 236] which achieved average STSD of 3.2 mm and 2.8 mm, respectively. Since clinical settings require error distances to be minimized below 1-2 mm due to the thin atrial wall (2-3 mm) [27], the methods mentioned were still lacking in terms of precision. Tao et al. [37] proposed a method to reduce the distance error within 1.5 mm by incorporating paired magnetic resonance angiography into the pipeline. Zhu et al. was the only study that reported a Dice score of 79%.

While LGE-MRI segmentation still heavily relies on traditional methods, recent advancements in CNNs have been made for image segmentation in general. VGG-Net [151] has been widely used for developing FCN [160, 162] for semantic segmentation due to its simplicity, and adaptations of superior architectures, such as ResNet and Inception [237], are currently the state-of-the-art in the field [238]. Recent studies have investigated the use of CNN architectures for LA segmentation built upon the basic U-Net structure. Multi-scale convolutions and attention units have been incorporated into U-Net to form content-aware networks which enhance the learning of discriminative features and abnormalities from the complex LA and PV structure [239, 240]. The incorporation of CNNs into traditional multi-atlas and graph-cut pipelines have been conducted to utilize the feature learning capabilities in deep learning and maintain the anatomical accuracy of the segmentation [241]. However, these studies were evaluated on smaller datasets with different numbers of test samples, making direct comparisons difficult.

It remains unclear if it is possible to create a unified deep learning architecture capable of achieving optimal performance for segmentations across a wide spectrum of applications. While U-Net is currently the most widely used CNN architecture for medical image segmentation, and VGG-Net and ResNet have been adapted as backbones and extended for segmentation in more recent developments, CNNs still have to be individually tuned for each specific application [242, 243]. This is reflected in the current literature which often contains wildly differing implementations of U-Net and other architectures for segmentation tasks in different disciplines, making it difficult to pinpoint design characteristics that can be applied universally. Methods for hyper-parameter optimization, particularly in U-Net, for general problems and/or specific tasks are still an ongoing topic of discussion, and could potentially lead to a more robust framework for segmentation. The lack of accessible, large-scale datasets and the varying quality of the image data (with labels) also hinders the development of optimized methods for LA segmentation from LGE-MRIs. By providing the largest 3D atrial LGE-MRI dataset along with top quality expert labelled LA cavities ($N = 154$), thanks to the great efforts of the University of Utah over the past two decades, we have gathered the community and organized the 2018 LA Segmentation Challenge (**Figure 6.1**) in conjunction with the MICCAI conference and the Statistical Atlases and Computational Modelling of the Heart (STACOM) workshop in Granada, Spain [244]. Throughout the course of the challenge, over 200 research groups/individuals worldwide accessed our challenge dataset. In total, 27 teams participated in the final

evaluation phase of the challenge and their final rankings are shown on the challenge website (<http://atriaseg2018.cardiacatlas.org/>).

To analyse the wide spectrum of conventional and deep learning methods submitted to the challenge, we first sub-grouped the methods by their main architecture designs and then identified the subgroup with the best design features which contributed to its superior performance. We then conducted extensive hyper-parameter tuning experiments on the top-performing method to identify the exact parameter choices leading to achieving state-of-the-art accuracies. By performing the benchmarking study in this manner, we offered both an overall picture on the major design choices necessary, as well as detailed practical considerations.

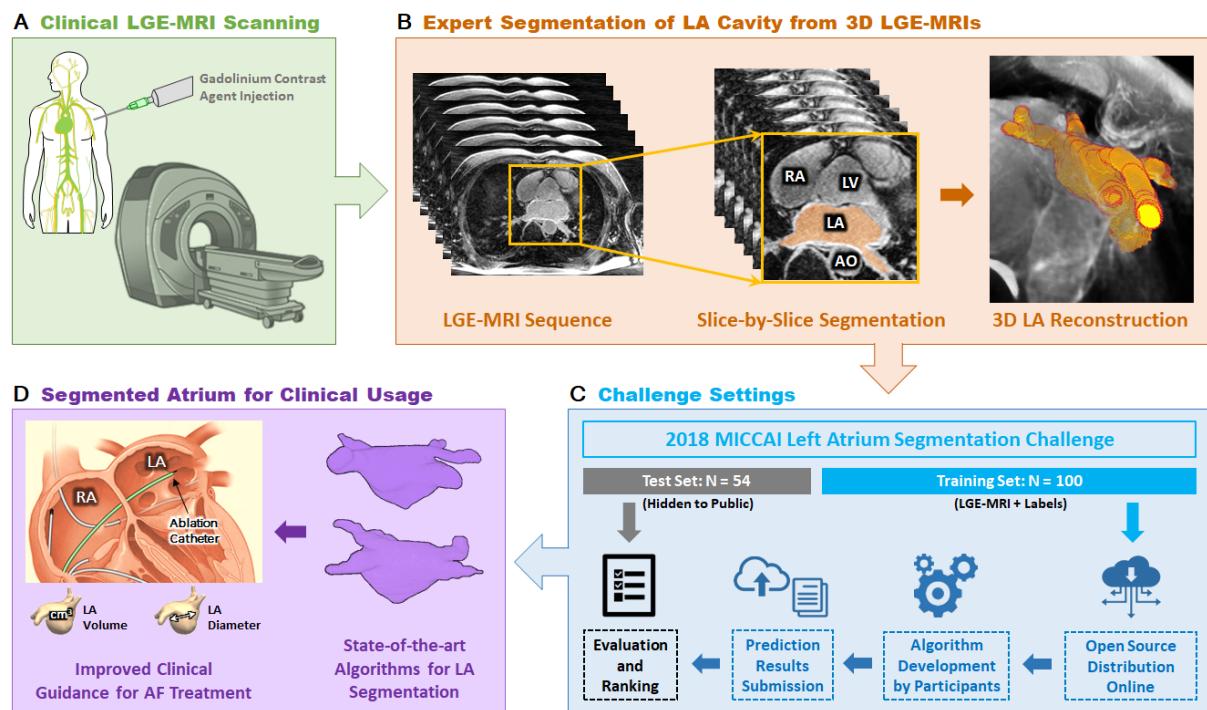


Figure 6.1. The overall workflow of medical images for clinical usage and the 2018 Atrial Segmentation Challenge. **A)** Clinical MRI scanners were used to acquire late gadolinium-enhanced magnetic resonance imaging (LGE-MRIs) from patients with atrial fibrillation (AF). **B)** The LGE-MRIs were manually segmented in a slice-by-slice manner by experts to obtain labels of the LA cavity. The 3D LA geometry can be obtained by stacking the 2D segmentation together. **C)** In the 2018 LA Segmentation Challenge, 154 3D LGE-MRI data (each with a spatial dimension of either $576 \times 576 \times 88$ or $640 \times 640 \times 88$) were split into 100 training and 54 testing sets. The training data and labels were made public to all potential participants of the challenge, and the testing data was used at the end of the challenge for

evaluation. A total of 27 teams participated and were ranked based on the Dice scores. **D**) Accurate reconstruction of the LA anatomical structure provides crucial information for patient stratification and for guiding clinical treatment for patients with AF. AO, aorta; LV, left ventricle; RA, right atrium.

6.2 Methods

6.2.1 Dataset and Labels

The same 154 3D LGE-MRI dataset used in the previous **Chapter 5.2.1** used in this chapter. To provide more thorough analysis to complement the analysis of the methods submitted, we computed quality measures for each LGE-MRI scan. The LGE-MRIs contained a range of imaging qualities, with the differences mostly being attributed to varying patient characteristics and magnetic fields. In order to explore the quality of the dataset, the signal to noise ratio, contrast ratio, and heterogeneity between the foreground containing the LA and the background was assessed (**Figure 6.2**). The signal to noise ratio, contrast ratio, and heterogeneity were defined as:

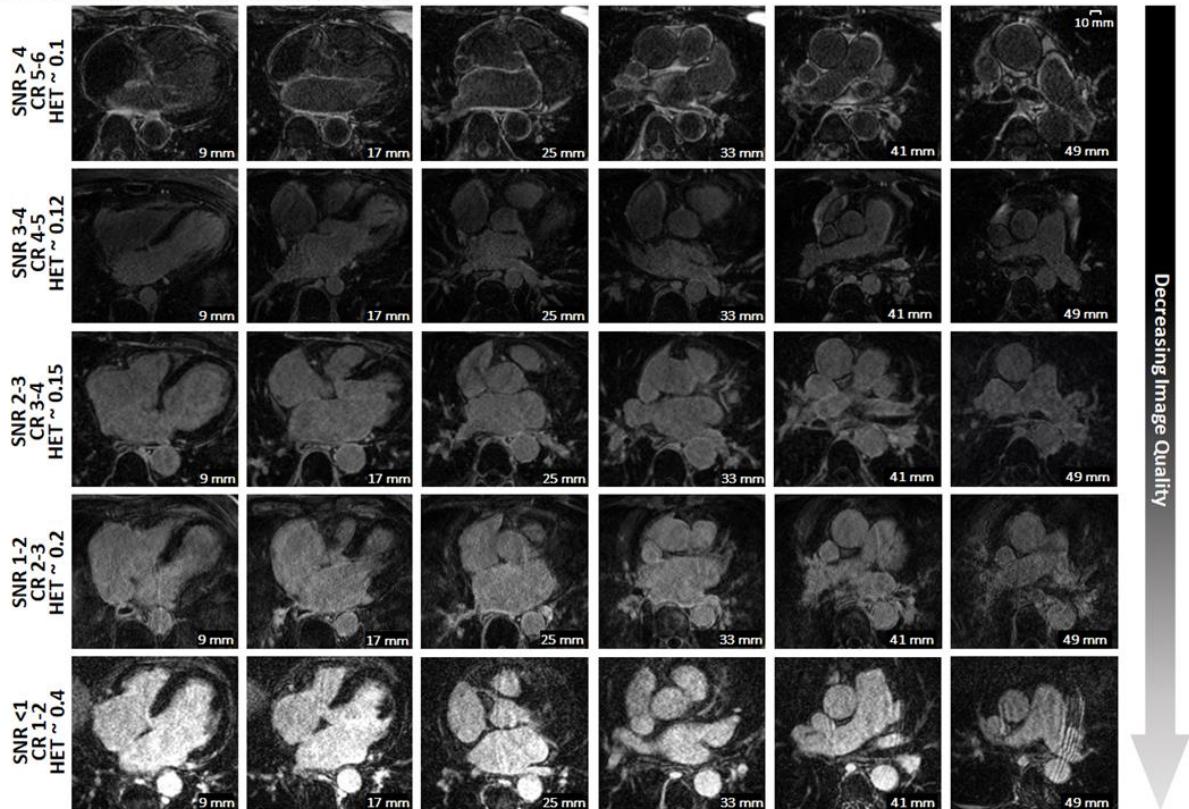
$$\text{Signal to Noise Ratio} = \frac{|\mu_{LA} - \mu_{background}|}{\sigma_{background}} \quad (\text{Equation 6.1})$$

$$\text{Contrast Ratio} = \frac{\mu_{LA}}{\mu_{background}} \quad (\text{Equation 6.2})$$

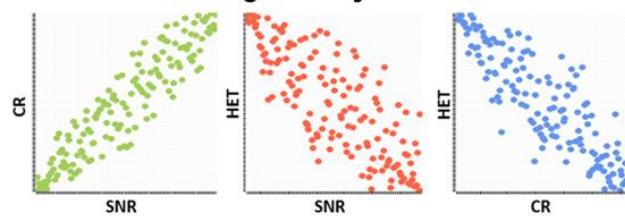
$$\text{Heterogeneity} = \frac{\sigma_{LA}}{|\mu_{LA} - \mu_{background}|} \quad (\text{Equation 6.3})$$

where μ is the mean of either the LA pixels or background pixels of the LGE-MRI and σ is the standard deviation of either the LA pixels or background pixels of the LGE-MRI [245]. The three metrics used were in agreement as the signal to noise ratio had a strong positive correlation with contrast ratio and heterogeneity and while contrast ratio and heterogeneity had a strong negative correlation. Distributions of the quality measurements on all data showed that less than 15% of the data was of high quality (signal to noise ratio > 3), 70% of the data was of medium quality (signal to noise ratio = 1 to 3), and over 15% of the data was of low quality (signal to noise ratio < 1).

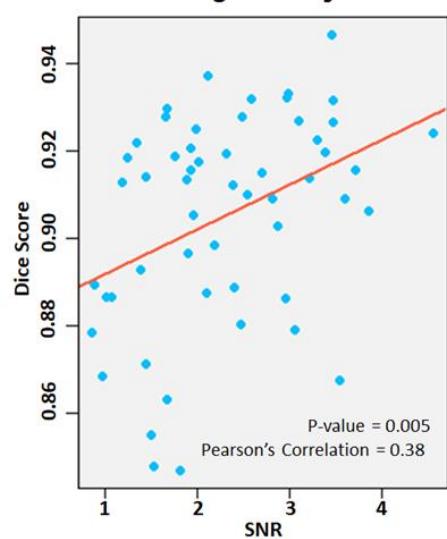
A LGE-MRIs of Different Qualities



B Correlation of Image Quality Metrics



D Correlation of Image Quality vs Dice Score



C Distribution of LGE-MRI Quality in the Dataset

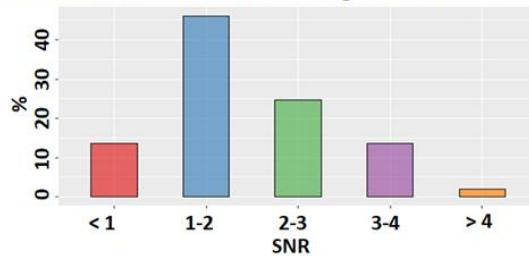


Figure 6.2. Variation in the quality of the late gadolinium-enhanced magnetic imaging (LGE-MRI) dataset used for this study. **A)** Each row represents different LGE-MRIs with various quality measured using signal to noise ratio, contrast ratio, and heterogeneity. The data at the top row has the highest quality and the data at the bottom row has the lowest quality. Each column represents 2D LGE-MRI at the same location within a 3D LGE-MRI and is shown with the distance measurements in millimetres. **B)** Signal to noise ratio and contrast ratio are positively correlated and are both negatively correlated with HET. **C)** Distribution of the

signal to noise ratio for all the 154 3D LGE-MRI datasets used in the 2018 Atrial Segmentation Challenge. **D)** Correlation of the average performance in relation to the LGE-MRI image quality, demonstrating that higher qualities (higher signal to noise ratio) result in higher Dice scores.

6.2.2 2018 Atrial Segmentation Challenge Setup

Table 6.1. Summary of methods submitted to the 2018 Atrial Segmentation Challenge.

#	Dice (%)	Author	Pre-Processing	2D/3D	CNNs	Methodology	Post-Processing	Framework
1	93.2 ± 2.2	Xia et al.	Down sampling, CLAHE	3D	2	U-Net with additional residual connections to locate ROI, same network to segment ROI	None	PyTorch
2	93.1 ± 2.2	*Huang	Down sampling	3D	2	U-Net with additional residual connections, dense connections and dilated convolutions to locate ROI, same network to segment ROI	None	Tensorflow
3	92.6 ± 2.2	Bian et al.	Cropping	2D	1	Dilated ResNet with spatial pyramid pooling to segment images	None	PyTorch
4	92.5 ± 2.7	Vesal et al.	Cropping, CLAHE	3D	1	U-Net with dilated convolutions to segment cropped region	None	Keras
5	92.5 ± 2.3	Yang et al.	Down sampling, cropping	3D	2	Faster-RCNN to locate ROI, U-Net with dense deep supervision to segment the ROI	None	Tensorflow
6	92.3 ± 2.9	Li et al.	None	3D	2	U-Net to locate ROI, U-Net with attention units and hierarchical aggregation units with dilated convolutions to segment the ROI	None	Keras
7	92.3 ± 2.3	Puybareau et al.	Normalization, cropping	2D	1	Fully convolutional network with pre-trained VGG-Net weights and intermediate output maps to segment image	Keep largest component, smoothing	Keras
8	92.1 ± 2.6	Chen et al.	Intensity normalization	2D	1	Multi-task U-Net with an additional classification branch at the center containing spatial pyramid pooling to classify if data is pre or post ablation, training is done on coarse images first then full-sized images to improve feature learning	Dilation and erosion, keep largest component	PyTorch
9	91.5 ± 2.6	*Xu et al.	Resize image to multiple scales, cropping	2D	1	Ensemble of different variants of U-Net to segment images of different scales and average the results	None	Tensorflow
10	90.7 ± 3.1	Jia et al.	Normalization, resizing	3D	2	U-Net to locate ROI and generate distance maps, ensemble of U-Nets to segment ROI and distance maps	None	Keras
11	90.3 ± 3.2	Liu et al.	Cropping	2D	1	U-Net to segment cropped region	None	Keras
12	89.8 ± 3.4	Borra et al.	Cropping based on thresholding	3D	1	U-Net to segment cropped region	Keep largest component	Keras
13	89.7 ± 3.5	De Vente et al.	Cropping	2D	1	U-Net to segment patches which were then stitched together to reconstruct the original image	Keep largest component	Keras
14	88.7 ± 3.1	Preetha et al.	Cropping	2D	1	U-Net with deep supervision to segment images	None	Tensorflow
15	86.1 ± 3.6	Qiao et al.	None	-	-	Convert image to probability map, atlas selection, multi-atlas registration, level-set refinement.	None	Non-Deep Learning
16	85.9 ± 6.1	Nuñez-Garcia et al.	None	-	-	Multi-atlas segmentation, shape modeling, clustering to rank similarity of different atria shapes	None	Non-Deep Learning
17	85.1 ± 5.1	Savioli et al.	Cropping, CLAHE, de-noise filters	3D	1	Fully convolutional network to segment entire image volume	None	Torch

CLAHE, contrast limited adaptive histogram equalization; CNN, convolutional neural network; ROI, region of interest.

The 3D LGE-MRI dataset was randomly split into training ($N = 100$) and testing ($N = 54$) sets, with the entire training set published at the start of the challenge for participants to develop their algorithms. Images from the same patient were only in either the training or testing sets. The testing set was released without the labels during the evaluation period near the end of the challenge, and participants were ranked based on the accuracy of the testing set predictions submitted to the organizers. The Dice score was used as the only evaluation metric in the challenge for simplicity. However, subsequent analyses with STSD and the Euclidean distance error of the LA diameter and volume measurements were conducted after the challenge.

The 17 out of 27 teams who provided their methodologies and performances either in full STACOM papers [168, 246-259] or online only [260, 261] were included in this benchmarking study. Summaries of their methodologies are shown in **Table 6.1** sorted by the final challenge rankings.

6.2.3 Algorithm Evaluation

A range of metrics was used for the benchmarking study including the Dice score, IoU, sensitivity, specificity, Hausdorff distance, and STSD. These metrics were defined in **Chapter 3.6.1**. We also measured the atrial diameter and volume errors which were defined in **Chapter 3.6.2**. Statistical significance was assessed using the two-tailed unpaired t-test to compare the performances of participant sub-groups as well as individual algorithms during analyses. Statistical comparisons between multiple groups were conducted using the Tukey honest significant differences to perform multiple-pairwise comparisons.

6.3 Results

6.3.1 Performance of Submitted Algorithms

We first examined the overall performance of the different methodology categories from the 17 submitted algorithms. CNN was the most popular choice as it was used by 15

teams, and on average, substantially outperformed the other two teams which used traditional atlas-based segmentation methods (92.5% vs 85.1% Dice score, $p < 0.05$, **Figure 6.3A**). Out of the 15 teams using CNNs, 12 teams proposed CNNs designs based on the popular U-Net architecture [161] whilst the other three 3 implemented non-U-Net designs. We observed that the teams using U-Net based CNNs had superior performances (92.4% vs 89.3% Dice score, $p < 0.05$) including Xia et al. [168] and Huang [260] who were ranked 1st and 2nd in the challenge (**Figure 6.3B**). The majority of teams using the U-Net architecture implemented further enhancements to the original architecture in an attempt to improve the segmentation performance. This involved the use of additional residual connections [168], replacing all layers with dilated convolutions [247], improved methods of training such as the use of custom loss functions, deep supervision [248], multi-task learning [251], and attention mechanisms throughout the network [249]. The three teams which did not use U-Net as a baseline approach implemented enhanced versions of existing architectures such as ResNet [237, 246], VGG-Net [151, 250], and Fully-CNNs [160, 259, 262] which have been widely used on the ImageNet database [263]. The Adam optimizer with a learning rate of 0.001 or 0.0001 was the most popular optimization configuration ($N = 12$), while the remaining teams used the stochastic gradient descent with a learning rate of 0.001, 0.0001 or 0.0005 ($N = 3$). The Adam optimizer with a learning rate of 0.001 was used by the top five teams and contributed to a significantly higher Dice score on average compared to teams which used a learning rate of 0.0001 and teams that used the stochastic gradient descent optimizer ($p < 0.05$). There were no significant differences between the batch sizes used ($p = 0.5$). Investigation into the number of parameters in each proposed algorithm also showed no statistically significant correlation with the final challenge performance ($p = 0.21$).

We then evaluated 2D versus 3D approaches based on the 15 submitted CNN algorithms. Since the challenge data was 3D, 8 out of the 15 teams proposed 3D CNNs which simply performed direct 3D segmentation on each set of 3D LGE-MRIs. On the other hand, 2D CNNs used by seven teams segmented each image slice of the image volume along the z-axis and stacked the individual segmentations together to obtain the final 3D results. Summary statistics show that there is no significant difference between the 2D and 3D CNN approaches when the CNN architecture and setup are not considered (92.1% vs 92.5% Dice score, $p = 0.82$, **Figure 6.3C**), even though the 3D CNNs were used by 4 of the top 5 teams. We also observed five teams utilizing a double, sequentially used CNNs in their pipeline to improve the segmentation performance compared to methods that only contain a single CNN ($N = 10$). Through this enhancement, the former approach achieved a significantly better

average Dice score of ~92.8% compared to single CNN methods which obtained a Dice score of ~90.3% on average ($p < 0.05$) (Figure 6.3D).

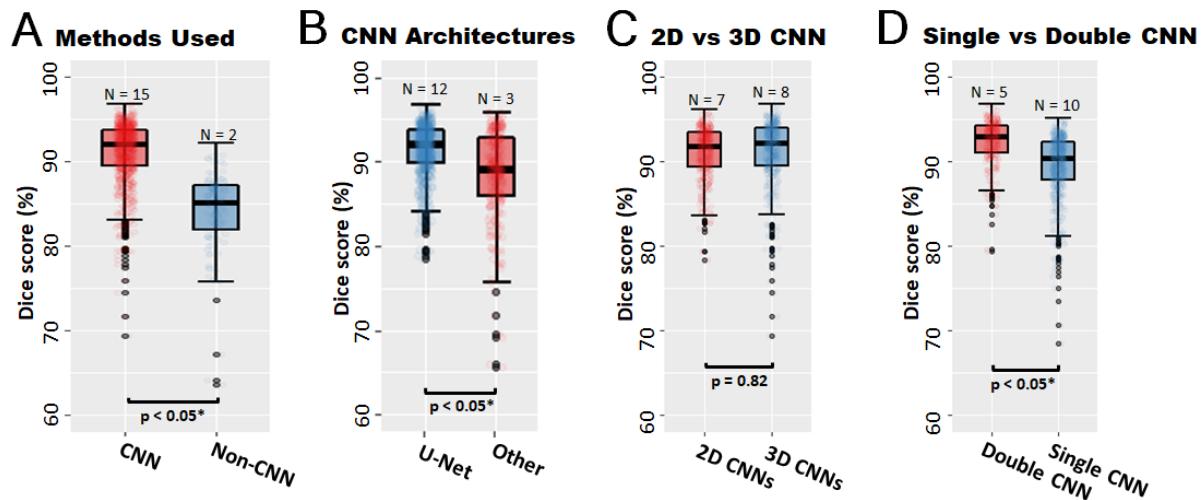


Figure 6.3. Comparative summaries of the 17 algorithms included in this benchmarking study, a well representative subset of the 27 teams that participated in the 2018 Atrial Segmentation Challenge. **A)** The 15 methods utilizing convolutional neural networks (CNNs) outperformed the two traditional multi-atlas-based methods with statistical significance. **B)** The 12 teams using U-Net architectures in their CNN pipelines outperformed the 3 teams using ResNet, VGG-Net, and Fully-CNNs with statistical significance. **C)** Approximately an equal number of teams used methods consisting of 2D CNNs compared to methods consisting of 3D CNNs. However, there was no statistically significant difference in segmentation accuracy between the two groups. **D)** Participants using double, sequentially used CNNs achieved significantly higher accuracy than those using a single CNN for segmentation.

Of the methods discussed in our study, the *single CNN* method focused solely on using one CNN for direct segmentation of the LA from either the original dataset or ROIs cropped at a consistent location across all input images (Figure 6.4A). Alternatively, the double, sequentially used CNNs, termed as *double CNN* throughout this study, involved the first CNN automatically detecting the ROI from LGE-MRIs and the second CNN performing regional segmentation of the LA from the ROIs (Figure 6.4B-C). To further examine the 15 proposed CNN methodologies, we have regrouped the approaches into three categories: double 3D CNNs ($N = 5$), single 2D CNNs ($N = 7$), and single 3D CNNs ($N = 3$). It is noted that no team proposed a double 2D CNN method. The average accuracies of the three

categories of CNN methods were evaluated with both technical and biological performance metrics: Double 3D CNN methods significantly outperformed single 2D and 3D CNN methods in terms of the Dice score (92.8% vs 91.1% and 89.9%, $p < 0.05$), the STSD distance (0.75 mm vs 0.85 mm and 1.1 mm, $p < 0.05$), LA diameter error (2.7% vs 3.2% and 4.3%, $p < 0.05$), and LA volume error (4.5% vs 4.9% and 6.2%, $p < 0.05$) (**Figure 6.4D-G**). The success of the double CNN methodology can also be seen in the challenge rankings as this workflow was utilized by the top 2 teams as well as 4 of the top 6 teams which achieved over 92% Dice score.

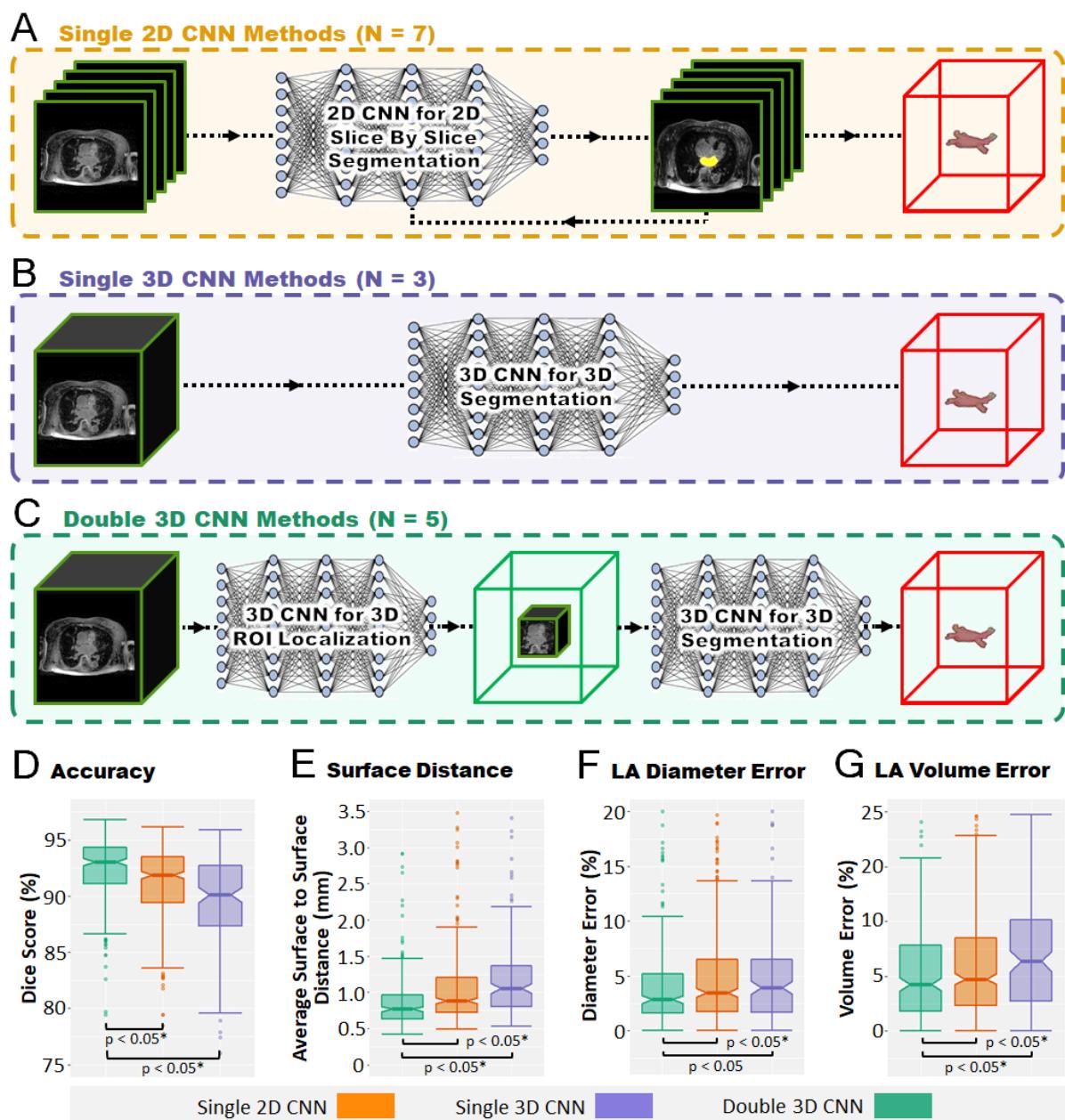


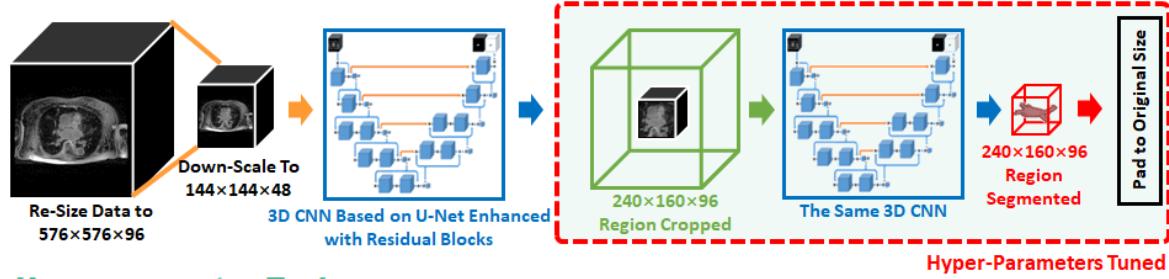
Figure 6.4. Detailed evaluations of the performance of the 15 convolutional neural network (CNN) pipelines submitted to the challenge grouped into three general categories depending on how the CNNs were applied to segment the left atrium (LA) from the late gadolinium-enhanced magnetic resonance imaging (LGE-MRI). **A)** Single 2D CNN methods consisted of one 2D CNN which performed slice-by-slice segmentation of each LGE-MRI. The 3D LA was then reconstructed by stacking slice-by-slice segmentation. **B)** Single 3D CNN methods consisted of a single 3D CNN to segment the LA from the LGE-MRI volume directly. **C)** Double 3D CNN methods consisted of one CNN to detect a 3D region of interest (ROI) from the LGE-MRIs and a second CNN to segment LA from the 3D ROI. Dice score (**D**), the surface to surface distance (STSD) (**E**), LA diameter error (**F**) and LA volume error (**G**) comparisons showed the superiority of 3D double CNN methods among the three different categories of CNN workflows.

6.3.2 Top Performing CNN Methods

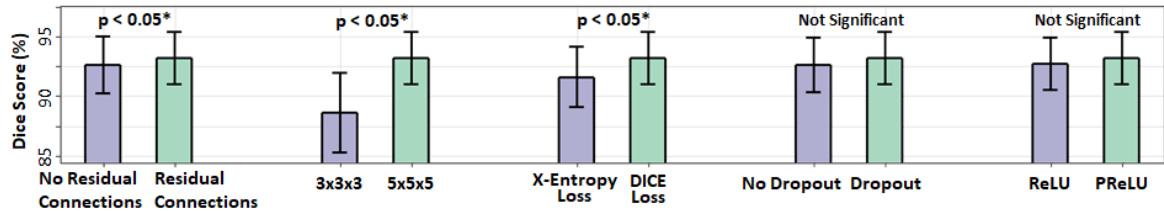
Methods involving double 3D CNNs were shown to have the best overall performance. Interestingly, all five of the double 3D CNN methods adopted U-Net as the baseline architecture, though significant improvements were added to the baseline approach by these participants. These included the addition of residual connections into the U-Net by Xia et al. [168], Huang [260], and Li et al. [249]. Specifically, residual connections were added to each block of two to three sequential convolutional layers along the entire length of the networks to improve gradient flow during backpropagation when training the CNNs. The type of residual connections varied from a simple connection without any additional operations to more advanced connections containing convolutional and pooling layers. Dense connections were also seen in the method proposed by Huang [260] along with dilated convolutions to improve the receptive field of the CNN. Yang et al. [248] did not alter the U-Net architecture but instead elected for an improved dense supervision training scheme and a customized loss function. The proposed loss function was an ensemble of the Dice score, pixel thresholding to improve sensitivity, and an overlap metric for improving segmentations at boundary locations. Apart from the double 3D CNN methods, Vesal et al. [247] also performed highly with single 3D CNN methods. The effectiveness of their CNNs could be potentially attributed to the use of dilated convolutions, allowing them to outperform all other single CNN methods.

To gain further insights into the best performing CNN pipelines, control experiments were performed in this benchmarking study on the winning approach by Xia et al. [168] to examine the factors contributing to their superior performance (**Figure 6.5**). It is noted that since most top-ranking algorithms utilized similar U-Net based designs, experimental observations derived from the approach by Xia et al. [168] would also apply to other methods with similar algorithm setups. For the purposes of analysis, the top algorithm was selected based on its simplicity and light-weight nature. A summary of the participants' double 3D CNN pipeline is shown in **Figure 6.5A** and additional experiments were conducted to evaluate the efficacy of the second CNN network for performing segmentation. Consistent with the benchmarking experiments, the algorithm was also trained using the 100 training data and tested using the 54 test data. Hyper-parameter tuning experiments in **Figure 6.5B** revealed that the extra residual connections added to the U-Net architecture increased the Dice score by 0.7% ($p < 0.05$). The increased receptive field using $5 \times 5 \times 5$ convolutional kernels significantly improved the Dice score by 4% ($p < 0.05$) compared to the widely used $3 \times 3 \times 3$ kernels. The use of the Dice loss improved the accuracy by 2.1% ($p < 0.05$) over the traditional cross-entropy loss which does not account for the major class imbalance present in the dataset. Although not statistically significant, dropout and PReLU further improved performance by approximately 0.5% ($p = 0.1$ and $p = 0.18$). Multi-variate statistical tests using all possible hyper-parameter combinations were also performed, and the experiments yielded consistent results as in **Figure 6.5B**. Colour-intensity normalization or contrast limited adaptive histogram equalization, used by several teams improved the Dice score by 0.7% ($p < 0.05$) (**Figure 6.5C**). Standard data augmentation techniques such as random rotation, elastic deformations, perspective scaling, and random flipping improved the performance by over 2% ($p < 0.05$) (**Figure 6.5D**), while other schemes such as blurring, affine transformations, and sheering did not result in any significant improvements ($p = 0.5$). We also observed that online augmentation, in which the images were augmented on-the-fly to produce differently augmented training sets during every epoch, was approximately similar to offline augmentation, in which an augmented set generated before training was used for all epochs ($p = 0.41$) (**Figure 6.5E**).

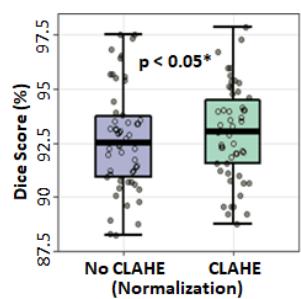
A Top Performing CNN Pipeline (93.2% Dice Score)



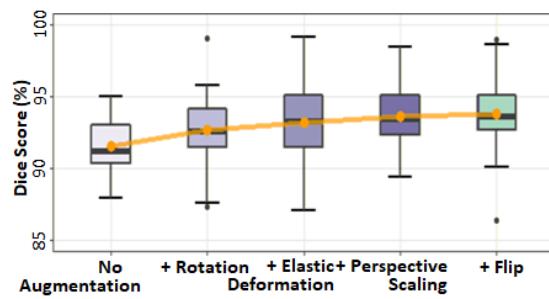
B Hyper-parameter Tuning



C Contrast Norm.



D Data Augmentation Schemes



E Offline vs Online Aug

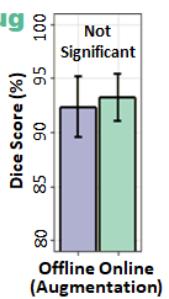
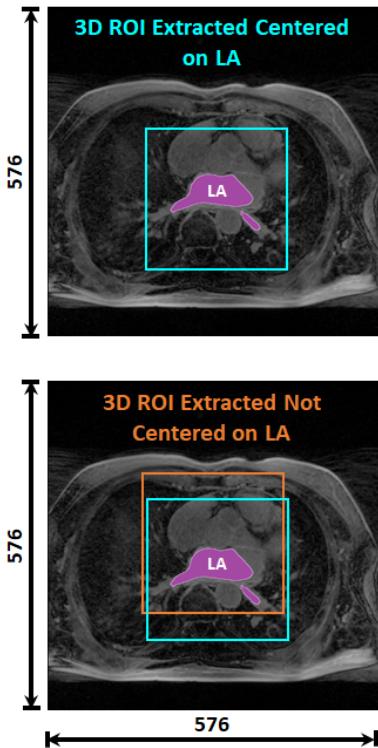


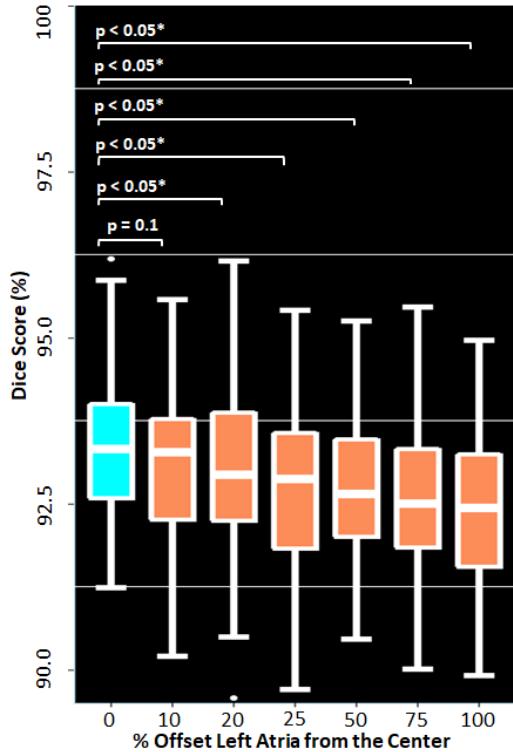
Figure 6.5. Post-challenge analysis of the winning method by Xia et al. [168] in the 2018 Atrial Segmentation Challenge demonstrates the optimality of their approach. **A)** Schematic summary of the 3D double convolutional neural network (CNN) approach. Both CNNs consisted of a U-Net architecture enhanced with batch-normalization in each layer and residual connections along the length of the network. The first CNN detected the centroid of the ROI from a down-sampled version of the initial late gadolinium-enhanced magnetic resonance imaging (LGE-MRI). A $240 \times 160 \times 96$ region centered in the LA cavity was the output and was then processed by the second CNN to segment LA in 3D. The output was padded to obtain the original resolution of the input LGE-MRI. **B)** Hyper-parameter tuning of the U-Net architecture in the second CNN (red box) showed that all parameters used by the winning team were optimal through post-challenge analysis. **C)** The effect of contrast limited adaptive histogram equalization (CLAHE) for normalizing each LGE-MRI during pre-processing on the performance of the CNN. **D)** Data augmentation led to an incremental improvement as more augmentation methods were increasingly added, showing an increasing trend in accuracy. **E)** Comparisons of offline and online data augmentation schemes showing that there were no significant differences between the two schemes.

6.3.3 Key Factors Influencing the Performance of Double CNN Approaches

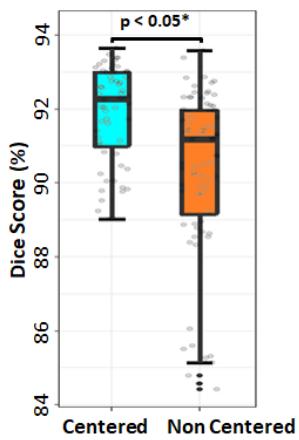
A LA-Centered vs Offset



B Effect of Offsetting the LA



C Effect of Centering the LA on the ROIs



D Effect of CNN Input Size on Accuracy

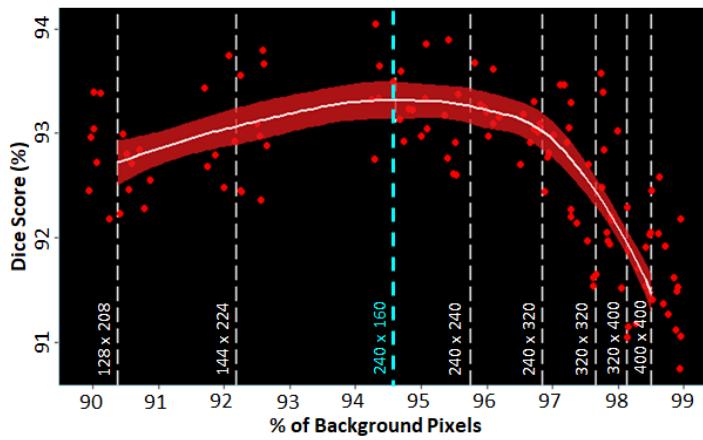


Figure 6.6. Analysis of the key factors of the first convolutional neural network (CNN) in the double CNN methods in enhancing left atrium (LA) segmentation performance. **A)** Illustration of the region of interest (ROI) extracted from the double CNN methods centred on the LA (Top) and ROI extracted not centred on the LA (Bottom). The ROI not centred on the LA can be offset from the centre of the LA by as much as 100 pixels in any direction (orange box). **B)** Displacing the LA by a distance in the ROI from the centre of the patch

reduced segmentation accuracy. A 100% offset implies the LA is pressed against the side of the patch without any loss of LA pixels. **C)** Performance summary comparing the methods proposed in the challenge which used double CNN pipelines to achieve an LA centred patch and the methods which had non-centred patches due to single CNN pipelines. **D)** To achieve the best accuracy, the size of the ROI should be sufficiently small. Effect of the input patch size of a CNN on the segmentation accuracy showing the peak accuracy at a patch size of 240×160 as proposed by the winners of the challenge. The percentage of background pixels is computed along the tested patch sizes to convey the degree of class imbalance in the CNN inputs where a higher percentage represents a more severe class imbalance.

We performed further analyses to target the key factors contributing to the success of double CNN methods compared to single CNNs. Since the optimality of the second CNN was shown through experimentation in the previous subsection, we focused on analysing the key factors of the first CNN for influencing the final segmentation accuracy (**Figure 6.6**). Firstly, it was observed that the ROI yielded as a result of the first CNN in a double CNN method was consistently centred on the LA (**Figure 6.6A**). Without the ROI detection procedure undergone by the first CNN, the centre of the LA would potentially have a ~ 100 pixel or $\sim 17\%$, shift in position from that of the original input images. Our experiments show a decreasing trend in overall CNN performance as the LA (ROI) was purposely shifted away from the centre of the image patch as an input for the second CNN, suggesting that centring the LA is extremely important to obtain superior accuracies (**Figure 6.6B-C**). Secondly, we observed that the smaller the image patch of ROI as the output of the first CNN, the higher final segmentation accuracy. This relationship was true for all ROI sizes that were greater than 240×160 (dashed cyan line in **Figure 6.6D**). The decreased input size of ROIs generated from the first CNN of the double CNN methods reduced the class imbalance as there were significantly fewer background pixels present in the input images, resulting in better performances as seen in our experiments using input sizes with X/Y dimensions of 240×160 to 400×400 . However, our experimentation also showed that the Dice accuracy of the CNN decreased when the size of the ROI was less than 240×160 even though the LA was fully contained within the ROI. We postulate that the observed decrease in performance could be attributed to the boundary of the LA being too close to the edge of the ROI inputted into the CNN. U-Net is known to perform poorly when segmenting boundary regions [161].

6.4 Discussion

6.4.1 Characteristics of Top-Performing CNNs for Segmentation

The key findings of this study were multi-fold. First of all, our study found that CNN approaches, notably those based on the U-Net architecture achieved better performance compared to traditional atlas-based methods and other CNN architectures. Additional residual connections and improved methods of optimization, such as custom loss functions, resulted in higher performance. The use of larger 5×5 also contributed to improved segmentations over the most common 3×3 kernels, potentially due to the increased receptive field causing the CNN to capture an increased amount of structural information. In addition, our post-challenge analysis also showed that data augmentation methods and colour-intensity or contrast normalization enhanced feature learning improved accuracy. More interestingly, 2D and 3D CNN methods had comparable accuracies. In particular, we discovered that double sequentially used CNN architectures achieved superior segmentation results than a single CNN approach. The effectiveness of the double CNN methods relies on two key elements. Firstly, the first CNN automatically detects the location of the ROI resulting in the image patch extracted to be centred on the ROI and sufficiently small to include minimal background pixels. Secondly, the second CNN effectively performs detailed regional segmentation of the ROI from the extracted image patch.

6.4.2 Comparison of Evaluation Metrics

3D double CNNs achieved the highest performance on average, regardless of the evaluation metric used (**Figure 6.7**). Our results show significant improvements over prior studies for LA segmentation from LGE-MRIs in terms of the STSD distance by approaching an average error of ~ 0.7 mm (1 pixel). The Dice scores achieved in our study are also significantly higher than prior studies for LGE-MRIs, and are comparable to the accuracies reported by the studies on non-contrast MRI for segmenting the LA and ventricles. The high level of agreement between the main technical measures of Dice score and STSD and the biological measures of the LA diameter and volume error showed that the top-performing algorithms were capable of producing anatomically accurate segmentation, which is highly important in clinical applications. The most widely used segmentation metrics, the Dice score,

Jaccard Index/IoU, and the STSD, produced fairly consistent rankings amongst the top 5 contestants. However, the Hausdorff distances, sensitivity, and specificity produced significantly different rankings in comparison. A potential explanation of this discrepancy is that the Hausdorff distances measured the most extreme errors, while sensitivity and specificity do not consider both positive and negative pixels simultaneously, leading to biased measurements. This also likely explains why these metrics are not commonly used for evaluating segmentation accuracies. While the conclusions drawn from the study based on the experimental results were validated with multiple technical and biological metrics, these summary results still suggest the ongoing need for an improved global definition of the exact meaning of high segmentation accuracy.

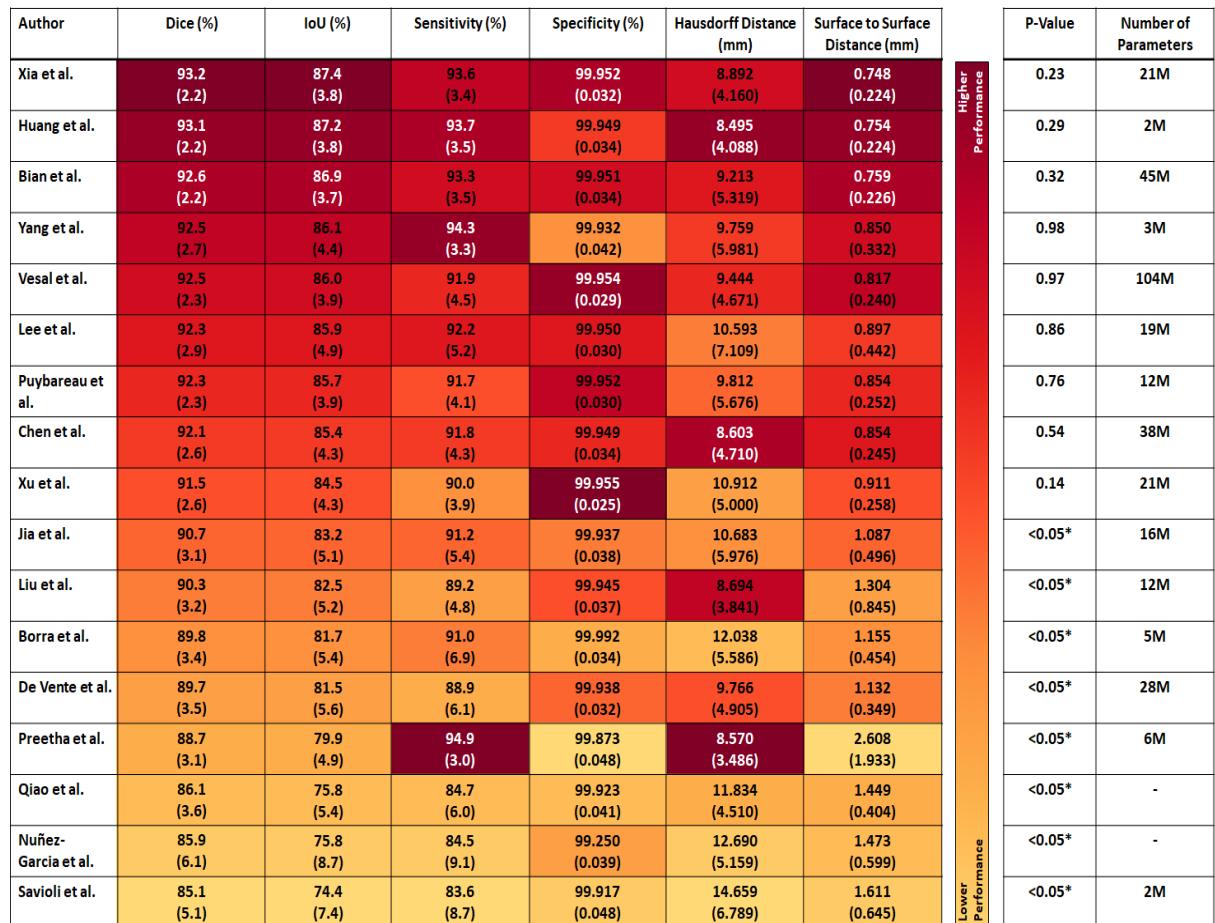


Figure 6.7. Summary of the rankings for all participants under different technical metrics, sorted in descending Dice scores. Metrics included the Dice score, Intersection over Union (IoU/Jaccard Index), sensitivity, specificity, Hausdorff distance (HD) and surface to surface distance (STSD) performance measures. The colour intensities reflect the rankings of the teams such that darker colours represent teams with higher performances. P-values for the statistical significance of each team computed with the Tukey Honest Significant Differences

are shown in the second to last column, with statistically significant values marked with an asterisk (*). The number of parameters for each model in millions (M) is shown in the last column.

6.4.3 Error Analysis and Future Work

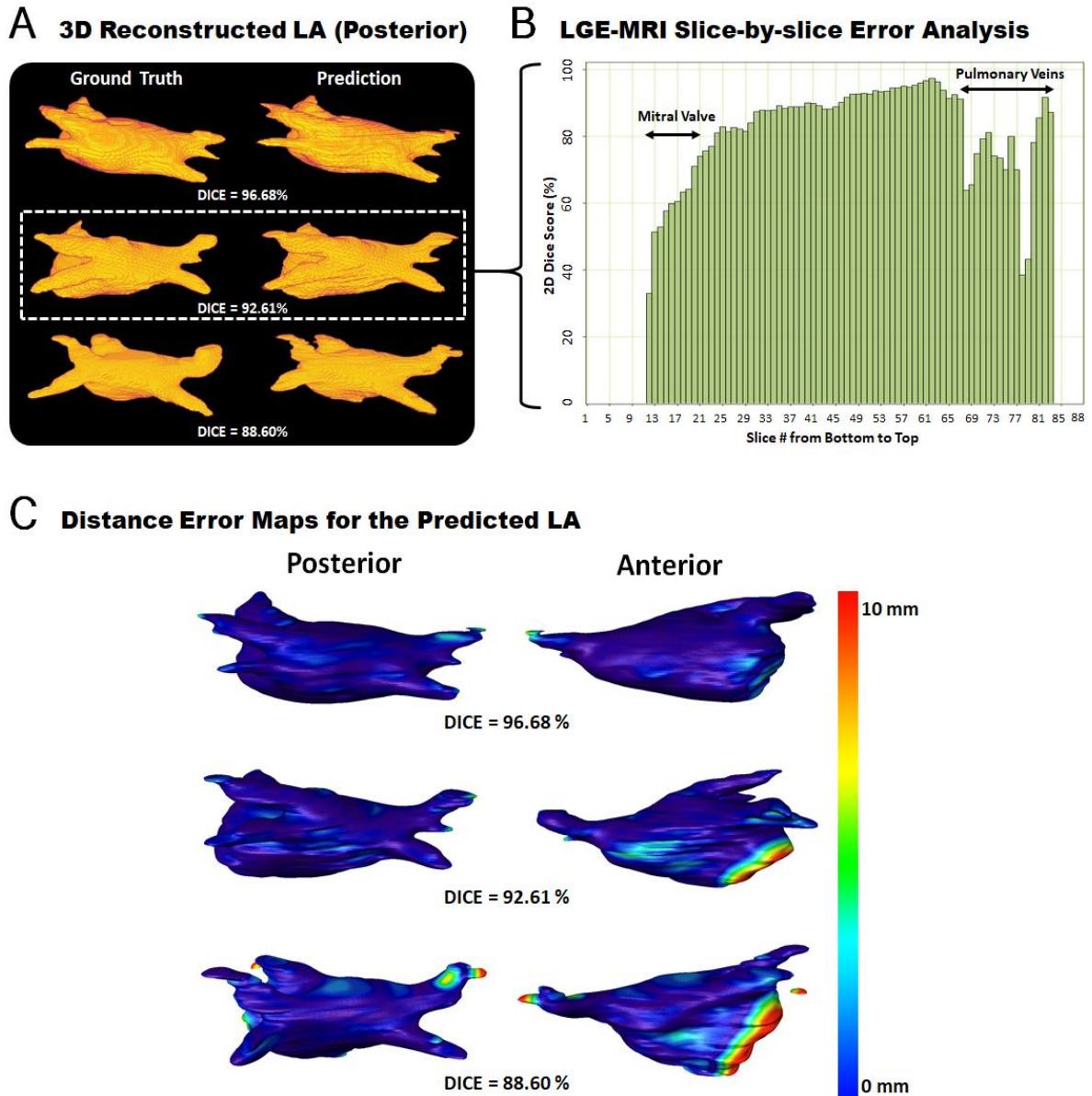


Figure 6.8. Error analysis of predictions generated from the winning method by [168] in the Atrial Segmentation Challenge. **A)** Visualization of predicted 3D left atrium (LA) reconstructions for three hearts with a high, medium, and low Dice scores. **B)** Slice-by-slice evaluation of the predicted LA reconstruction with a 3D Dice score of 92.61% in segmentation accuracy across each of the 88 slices in a 3D late gadolinium-enhanced

magnetic resonance imaging (LGE-MRI). The most erroneous slices were at the bottom and top of the LA corresponding to the mitral valve and pulmonary veins (PV), respectively. Slices 1 to 11 and 83 to 88 did not contain LA; hence, did not have any accuracy values. Note that the 2D Dice score used does not average out to the overall 3D Dice score due to the different dimensionality of the accuracy assessment. C) Surface to surface distance (STSD) error maps of the segmented LA compared with the ground truth for the three hearts at the posterior (left) and anterior (right) views. The distance errors were normalized between 0 and 10 millimetres.

Error analysis of the top methods showed that although the 3D visualizations of the predicted LAs were similar to the ground truths, there are still potential improvements to be made. Slice-by-slice analysis of the predictions from the top team showed that the method performed poorly when segmenting the regions containing the PVs located at the superior slices of the 3D LGE-MRIs and the mitral valve at the bottom connecting the LA with the LV (**Figure 6.8**). The errors at the mitral valve were attributed to the fact that there are no clear landmarks to separate the two chambers. This leads the experts to label this region with a flat plane which potentially contains large variability, making it difficult to be reproduced by the CNNs. On the other hand, the errors at the PVs could be explained by the fact that these structures are often very small in size and vary greatly in shape among patients, making them difficult to detect. The inherently varying and relatively poor LGE-MRI qualities were also a factor which impacted the segmentation performance due to the low contrast of the images. Our analyses showed that the average accuracy for each LGE-MRI was directly correlated to the quality of the particular LGE-MRI ratio for all approaches.

In the future, these issues could potentially be mitigated with an increased number of LGE-MRIs. Due to limitations with the current evaluation metrics, an interesting area to explore in future studies would be to create and design metrics computed through cardiologist assessments. To improve the optimization of the CNN for specific tasks, the influence of the number of layers and kernels may be further investigated and tuned to increase performance. The concept of 2D double CNNs would also be another interesting direction for future research as it was not proposed by any of the teams, and may potentially be utilized to improve segmentation accuracies on more difficult tasks such as bi-atrial chamber segmentation.

6.5 Conclusions

This chapter describes the 2018 LA segmentation challenge which provided 154 3D LGE-MRIs and analyses of the submitted algorithms. Our study found U-Net CNNs achieved the best performance, especially with the use of additional residual connections and advanced methods of optimization such as custom loss functions. Analysis also showed that 2D and 3D CNN methods had comparable accuracies. More interestingly, double sequentially used CNNs achieved superior segmentation results compared to single CNNs. Our extensive hyper-parameter tuning experiments also provide insights into the process of obtaining state-of-the-art accuracies. This large-scale benchmarking study makes a significant step towards much-improved segmentation methods for atrial LGE-MRIs, and will serve as an important benchmark for evaluating and comparing the future works in the field. We believe that findings from this study can potentially be extended to other imaging datasets and modalities, having an impact on the wider medical imaging community.

Chapter 7

Automatic Extraction of Pathologic Biomarkers in the Bi-Atrial Chambers from Cine-MRI and LGE-MRI with Convolutional Neural Networks

After deriving an optimized algorithm for LA segmentation in **Chapter 6**, we extended the findings to finally create a unified method for the comprehensive analysis of the human atria from medical imaging. The aim was to develop a method capable of atrial segmentation, fibrosis quantification, and wall thickness estimation to automate the calculation of key biomarkers used for clinical guidance and ablation target identification in AF patients. We also aimed to adapt the prior methodology to analyse both the LA and RA simultaneously as persistent and permanent AF is driven by complex substrates distributed across both atrial chambers [264].

This chapter presents a fully automated CNN-based pipeline, named *AtriaNet*, able to compute the anatomical structure, chamber diameter and volume measurements, AWT variation, and fibrosis distribution of the bi-atrial chambers directly from cardiac MRIs. *AtriaNet* was validated on the world's largest cardiac cine-MRI and LGE-MRI datasets, surpassing prior state-of-the-art approaches. Further testing on an independent clinical dataset revealed the *AtriaNet* predictions were accurate when evaluated against clinical methods, and had significantly greater efficiency. As *AtriaNet* was the first approach of its kind capable of computing multiple phenotypes for any cardiac imaging modality, our robust pipeline may be transferrable to other challenging medical segmentation tasks. The methods of this chapter were initially conceptualized in the publication:

Xiong, Z., Nalar, A., Jamart, K., Stiles, M.K., Fedorov, V.V. and Zhao, J., 2019. Fully Automatic 3D Bi-Atria Segmentation from Late Gadolinium-Enhanced MRIs Using Double Convolutional Neural Networks. *International*

Workshop on Statistical Atlases and Computational Models of the Heart, pp. 63-71.

The methods were then extended and refined in this chapter, which is presented in a similar form as the potential publication version, cited as:

Xiong, Z., Nalar, A., Kennelly, J., Petersen, S., Stiles, M.K. and Zhao, J., 2022. Automatic Extraction of Pathologic Biomarkers from Cardiac Magnetic Resonance Imaging Using Convolutional Neural Networks. (Under Review)

7.1 Introduction

Quantitative phenotypes derived from MRI, such as the chamber diameter/volume [106], AWT/tissue mass [265], cardiac function [266] and fibrosis [114], are important biomarkers for the determination of pathological states in cardiovascular diseases (CVD) and are crucial for decision making in clinics. For example, the left ventricular ejection fraction is an important clinical biomarker for the early detection of heart failure and cardiomyopathy [267]. In addition, the LA volume is a key indicator for the severity of cardiac arrhythmias [106], such as AF. In spite of this, the extraction of patient-specific phenotypes from MRIs remains a complex task requiring comprehensive expertise and intensive manual labour [28, 29, 35]. This results in important analyses often being impractical and inaccessible during patient assessments. This has been the major limiting factor for both the treatment of CVDs and research into large-scale MRI cohorts to investigate imaging phenotypes at a population level robustly.

Many studies have been conducted in recent years investigating the optimal deep learning algorithms for analysing various cardiac structures. In addition to the studies previously mentioned in **Chapters 5** and **6**, a 2017 whole heart (LA, RA, LV, RV) segmentation study containing 12 teams also showed the superiority of CNNs on 60 3D MRI and CT data, which were shown to achieve more stable and higher scores across all chambers of the heart [268]. In 2018, a study for LA segmentation from 154 3D LGE-MRIs showed the best teams designed U-Net CNNs with additional residual connections and custom optimization loss functions to surpass the vanilla U-Net architecture [224]. Similar findings were obtained by a 2019 benchmarking study for LV and RV segmentation from 45 LGE-

MRIs, where the highest-scoring team utilized an enhanced U-Net approach [269].

Despite these global efforts, current studies have mostly focused on performing cardiac segmentation on small in-house datasets of 100 to 200 independent samples. Recent studies have also only investigated the segmentation of the heart specifically to derive chamber volumes to measure cardiac function [205]. No studies have developed a robust pipeline for extracting all key cardiac imaging phenotypes such as cardiac anatomy, chamber diameter, volume, AWT, and fibrosis. More importantly, no studies have demonstrated the robustness and generality of their approach on multiple independent datasets, which is essential for utilizing it in clinical settings. The UK Biobank is a prospective cohort study consisting of 100,000 participants designed specifically for studying the risk factors of common diseases including CVDs using MRIs [270]. Currently, cine-MRIs from 5,000 individual patients have been manually annotated to derive the phenotypes of the cardiac structure including the ventricle, atria and aorta, leading to the largest cardiac cine-MRI dataset in the world. The University of Utah has been leading a global initiative to phenotype cardiac fibrosis by utilizing LGE-MRIs of the heart [114], resulting in the development of the world's large cardiac LGE-MRI dataset (**Chapter 5.2.1**). This dataset in particular is of exceptionally high quality due to the careful annotations and high-resolution imaging used, allowing a clear representation of the entire heart and its key structures including diseased tissue. While the LA has been heavily investigated, no established studies exist for phenotyping the RA. Our team has manually annotated the RA in the LGE-MRIs dataset [271]. The availability of these two large and high-quality datasets presents an ideal setting for extensive development and validation of more robust CNNs.

In this study, as a first attempt to address the essential issues in the field, we have developed a fully automatic pipeline, with the alias *AtriaNet*, to analyse both atrial chambers (LA and RA) of the heart. More importantly, the computational pipeline is the first stand-alone pipeline capable of providing cardiac quantitative phenotyping including the anatomical structure, diameter, volume, AWT variation, and fibrosis distribution for any cardiac imaging. *AtriaNet* was developed and validated on the world's largest cine-MRI and LGE-MRIs datasets, UK Biobank and the University of Utah, and has the potential to be extended to other cardiac structures in a wider context. More importantly, we further validated the proposed method for its clinical efficacy in guiding the treatment of AF, the most common cardiac arrhythmia. We were the first to automatically compute the phenotype of both atrial chambers from LGE-MRIs as AF is a bi-atrial disease and prior studies are focused on the LA only. Our pipeline was tested on an independent clinical MRI dataset at

Waikato Hospital to demonstrate its applicability on cross-centre imaging [131]. We also validated the accuracy of AtriaNet in estimating clinical biomarkers by comparing the predictions to patient biomarkers recorded during clinical practices alongside the MRI acquisition.

When compared with other approaches in the field, AtriaNet demonstrates multiple advantages and improvements. The designed CNN architecture is capable of estimating multiple cardiac phenotypes using the same network architecture and can be applied to both 2D and 3D imaging. Our method is also the first method for bi-atrial chamber analysis. It achieves high accuracy and can obtain similar accuracy levels even with lower sample size datasets. In terms of computation time, it is orders of magnitude faster than current clinical approaches. Lastly, we demonstrate the performance consistency of AtriaNet when applied to cross-centre imaging, and its effective transferability for computing accurate quantitative phenotypes from independently acquired clinical datasets. Overall, this study is a very important step towards improved and more efficient patient-specific diagnosis and treatment, and the developed methods may potentially be transferable to other imaging tasks, and lead to advancements in cardiac imaging phenotyping for the treatment of other CVDs.

7.2 Methods

7.2.1 Multi-Centre Data for Validation

Clinical datasets from three centres (UK Biobank [270], University of Utah [114], and Waikato Hospital [131]) were used in this study. The UK Biobank dataset contained 4,860 2D long-axis cine-MRIs with two-chamber and four-chamber perspectives and a spatial resolution of $1.8 \times 1.8 \text{ mm}^2$. Each sample was manually annotated with the LA and RA cavities at end-diastole and end-systole. The acquisition of the UK Biobank data has been extensively described in the original publication [272]. The Utah dataset contained 154 3D LGE-MRIs as described in **Chapter 5.2.1**, manually annotated with cavities and walls of the LA and RA. The Waikato dataset contained 11 3D LGE-MRIs, with the same labels and spatial resolution as Utah's.

The Waikato dataset was acquired as a part of our ongoing clinical trial at Midland MRI, Waikato Hospital and Braemar Hospital, Hamilton, New Zealand. Imaging was conducted on patients with AF using LGE-MRIs, led by A/Prof. Martin Stiles, the co-

supervisor of this thesis [131]. 11 patients with AF scheduled for PVI ablation surgery were enrolled in this observational study between June 2017 and February 2020. Each patient underwent an LGE-MRI procedure prior to PVI, with two having follow-up 6 months post-ablation LGE-MRIs. PVI was performed with either radio-frequency ablation or cryo-ablation depending on the most optimal treatment method as decided by the physician. The Waikato clinical study was approved by New Zealand Health and Disability Ethics Committees (Ref: 16/STH/130). An advantage of the Waikato LGE-MRI dataset was the availability of clinical patient information. Unlike the UK Biobank and the Utah datasets which require full de-identification of patients for public release, the Waikato dataset contained patient demographics and characteristics, full echocardiography reports and scans, data recorded during AF ablation such as ablation lines and voltage maps, and follow ups to determine AF recurrence.

In terms of the clinical validation of this study, 2D manually annotated echocardiograms were used to validate the atrial diameter and volume predictions from AtriaNet. Numerical solutions of the AWT were used to validate the wall thickness estimations from AtriaNet. 3D low-voltage maps obtained via the CARTO system (CARTO, Biosense Webster, United States) was used to validate the fibrosis predictions from AtriaNet.

The three datasets were individually split into training, validation, and testing sets. The UK Biobank cine-MRIs were split into 82% training, 6% validation, and 12% testing. The Utah LGE -MRIs were split into 60% training, 10% validation, and 30% testing, and lastly, the Waikato LGE -MRIs were only split into 36% training and 64% testing.

7.2.1.1 Ground Truth Atrial Segmentations for Cine-MRIs and LGE-MRIs

Manual image segmentation on the UK-Biobank cine-MRI dataset was performed by a team of eight observers overlooked by three principal investigators to standardize the segmentation procedure [270]. For each cine-MRI, two of the 50-time frames were manually annotated to capture the extremes in the atrial volume at atrial end-diastolic and end-systolic stages of the cardiac cycle. Labelling was performed for both the two-chamber and four-chamber views. The cavity for both atrial chambers was defined by manually tracing the LA and RA blood pool. PVs were not included in the LA annotations. The mitral valve and tricuspid valve were delineated by a planar cut to separate the atrial and ventricular chambers.

The annotation protocol for the LA in the Utah LGE-MRI dataset was described in **Chapter 5.2.1**. The RA segmentations were manually performed by our team based on the protocols used for LA segmentation to achieve consistency across both atrial chambers.

Firstly, the RA endocardium was defined by manually tracing the RA blood pool in each slice of the LGE-MRI. The tricuspid valve connecting the RA and RV was defined by a 3D plane to create a smooth linear surface. The RA endocardium was then morphologically dilated and manually adjusted according to the visibly enhanced RA wall to obtain the boundary of the epicardium. Next, the septum, the region of tissue connecting the RA and LA, was manually traced such that the epicardial surfaces of the LA and RA joined together. Finally, the dilated tricuspid valve was manually removed from the RA epicardium.

The Waikato LGE-MRI dataset was manually annotated by our team consistent with the Utah protocol for both atrial chambers.

7.2.1.2 Ground Truth Atrial Wall Thickness for LGE-MRIs

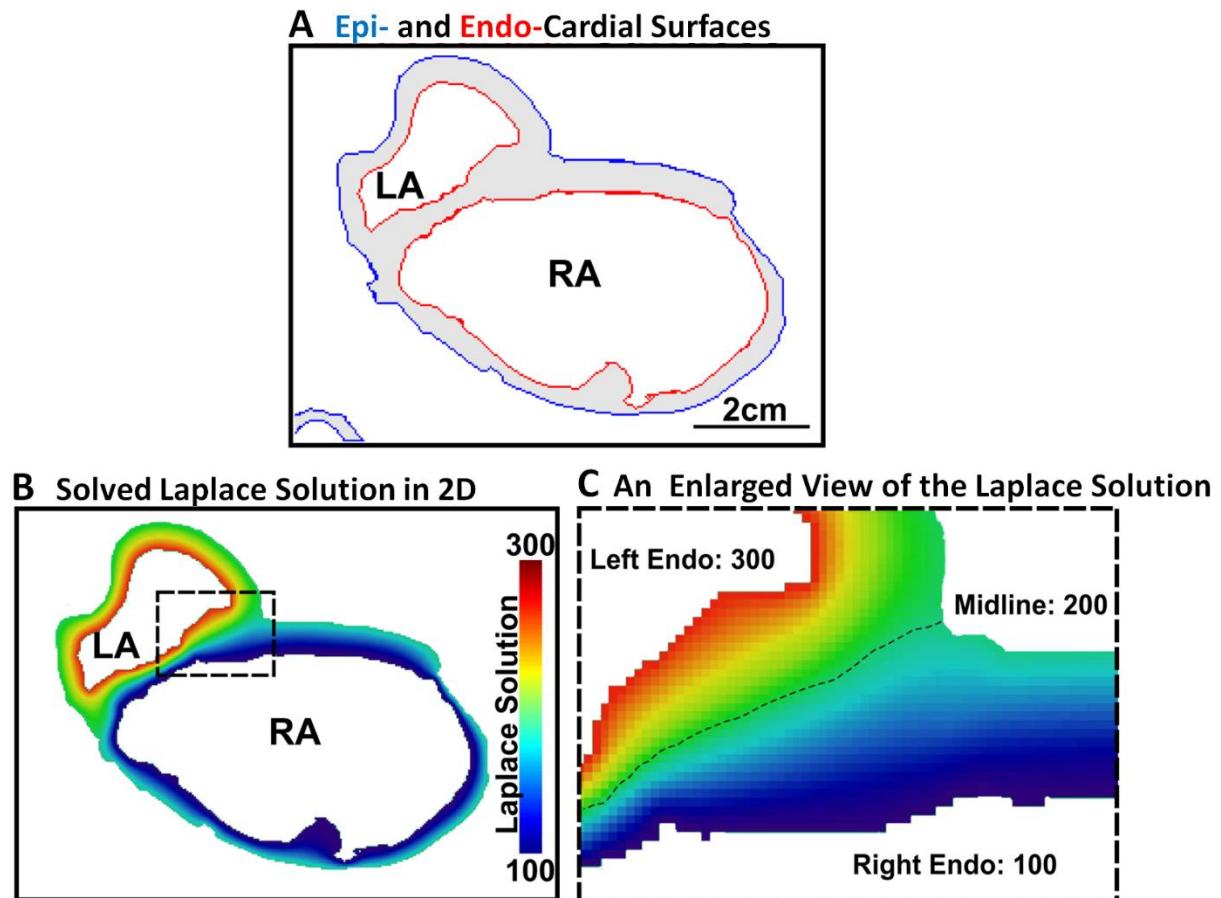


Figure 7.1. Illustration of the Laplace solution on a sample of left atrial (LA) and right atrial (RA) tissue adapted from Wang et al. [273]. **A)** The epicardial (blue) and endocardial (red) surfaces of the atria were highlighted in 2D. **B)** Numerically obtained Laplace solution with fixed values at epicardial (200) and endocardial (300 for RA and 100 for LA) surfaces. The dashed rectangular box region is enlarged and displayed in **C**). The dashed line separated the RA and LA where the Laplace solution equalled 200 (the mid-value).

The calculation of the AWT was based on a method, defined as a coupled partial differential equation (PDE) approach, previously proposed by Wang et al. co-authored by the author of this thesis. This sub-section provides an overview of the main concepts of the method utilized. The full set of equations can be found in the original publication [273] which was motivated by the work of Bishop et al. [265].

The algorithm for computing the AWT produced a 3D map with a value for the thickness of the wall at every pixel inside the manually segmented LA and RA wall masks. The endocardial surface and epicardial surface were firstly manually defined. The solution to the Laplace equation, ϕ , with the endocardial and epicardial surfaces as the boundaries was obtained and then used to formulate two first-order PDEs representing a function of the distance of each pixel from the endocardial and epicardial surfaces (**Figure. 7.1**). The trajectory function of the distance from the epicardial surface, D_{epi} , was formulated as PDE and defined as

$$D_{epi(i,j,k)} = \frac{1 + \frac{\partial u}{\partial x} D_{epi(i \mp 1, j, k)} + \frac{\partial u}{\partial y} D_{epi(i, j \mp 1, k)} + \frac{\partial u}{\partial z} D_{epi(i, j, k \mp 1)}}{\frac{\partial \phi}{\partial x} + \frac{\partial \phi}{\partial y} + \frac{\partial \phi}{\partial z}}$$

where $\mp 1 = \begin{cases} i - 1, & \frac{\partial \phi}{\partial x} > 0 \\ i + 1, & \frac{\partial \phi}{\partial x} < 0 \end{cases}$, $j \mp 1 = \begin{cases} j - 1, & \frac{\partial \phi}{\partial y} > 0 \\ j + 1, & \frac{\partial \phi}{\partial y} < 0 \end{cases}$, and $k \mp 1 = \begin{cases} k - 1, & \frac{\partial \phi}{\partial z} > 0 \\ k + 1, & \frac{\partial \phi}{\partial z} < 0 \end{cases}$

(Equation 7.1)

Similarly, the trajectory function of the distance from the endocardial surface, D_{endo} , was formulated as

$$D_{endo(i,j,k)} = \frac{1 + \frac{\partial u}{\partial x} D_{endo}(i \pm 1, j, k) + \frac{\partial u}{\partial y} D_{endo}(i, j \pm 1, k) + \frac{\partial u}{\partial z} D_{endo}(i, j, k \pm 1)}{\frac{\partial \phi}{\partial x} + \frac{\partial \phi}{\partial y} + \frac{\partial \phi}{\partial z}}$$

where $i \pm 1 = \begin{cases} i + 1, & \frac{\partial \phi}{\partial x} > 0 \\ i - 1, & \frac{\partial \phi}{\partial x} < 0 \end{cases}$, $j \pm 1 = \begin{cases} j + 1, & \frac{\partial \phi}{\partial y} > 0 \\ j - 1, & \frac{\partial \phi}{\partial y} < 0 \end{cases}$, and $k \pm 1 = \begin{cases} k + 1, & \frac{\partial \phi}{\partial z} > 0 \\ k - 1, & \frac{\partial \phi}{\partial z} < 0 \end{cases}$

(Equation 7.2)

and then solved. Lastly, the solutions of the two distance functions were summed to produce a 3D map of the AWT distribution across both atrial chambers.

7.2.1.3 Ground Truth Atrial Fibrosis for LGE-MRIs

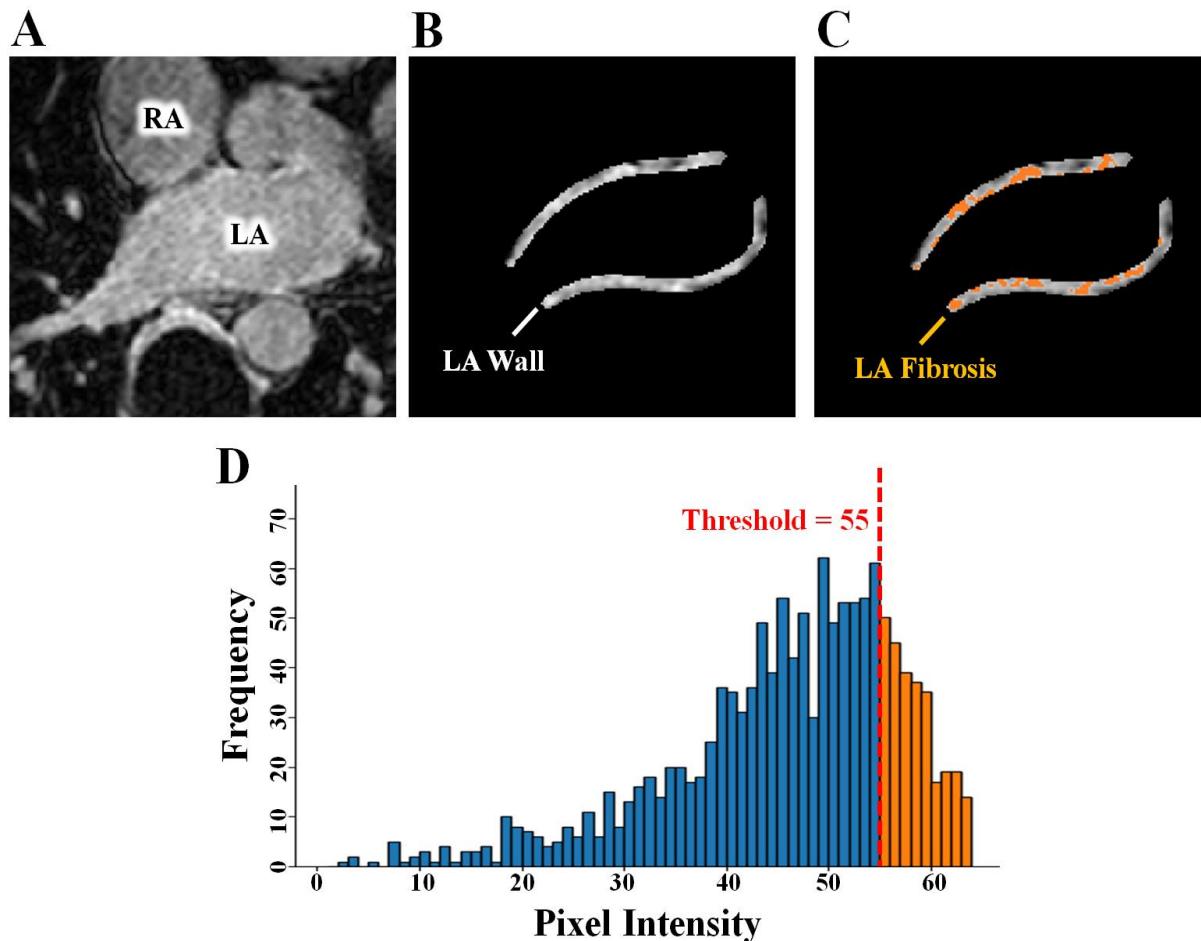


Figure 7.2. Demonstration of thresholding for fibrosis extraction from a 2D slice of a late gadolinium-enhanced magnetic resonance imaging (LGE-MRI) scan on the left atrium (LA). **A)** The MRI image is first acquired. **B)** The LA wall pixels are then extracted through segmentation. **C)** An adaptive thresholding method proposed by the University of Utah is used to obtain the fibrotic pixels within the wall (orange). **D)** Visualization of the pixel intensity distribution of the LA wall with the fibrotic intensity defined as being above 55 in this example. RA, right atrium.

The quantification of the atrial fibrosis was based on methods previously proposed by the University of Utah [28, 29, 114], defined as regions containing pixels with enhanced intensities within the atrial wall tissue, the area between the atrial epicardial and endocardial surfaces. The method consists of setting the threshold at a pre-defined number of standard deviations above the mean pixel intensity. It is also one of the most widespread techniques used in retrospective clinical studies, and considered a gold standard of fibrosis threshold setting. The intensity threshold for determining the fibrotic pixels was computed on a slice-

by-slice manner for the 3D images by estimating the mean and standard deviation of the normal wall tissues. In this context, normal wall tissues were defined as pixels containing intensity values between the 2nd percentile to the 40th percentile of the wall. The fibrosis threshold was then calculated as a pre-determined number of standard deviations above the mean of the normal wall tissue for each MRI slice. Mathematically, this threshold calculation was written as

$$X_{S\text{Threshold}} = \text{mean}(Wall_{norm}) + \text{std}(Wall_{norm}) * n \quad (\text{Equation 7.3})$$

where $Wall_{norm}$ was the normal wall tissue, std was the standard deviation, and n was the scaling factor to control the number of standard deviations above the mean. This method of thresholding allowed reproducibility as the calculations were based on consistent conditions but also allowed flexibility as the number of standard deviations could be tuned for each patient. Common values used in previous studies were 1 to 3 standard deviations above the mean.

7.2.2 AtriaNet for Automatic Atrial Phenotyping

The AtriaNet pipeline consisted of three stages: pre-processing, deep learning-based processing, and post-processing (**Figure 7.3A**). Minor changes in the deep learning stage of AtriaNet were performed to produce the three different outputs visualized in **Figure 7.3A** and estimate the various quantitative phenotypes.

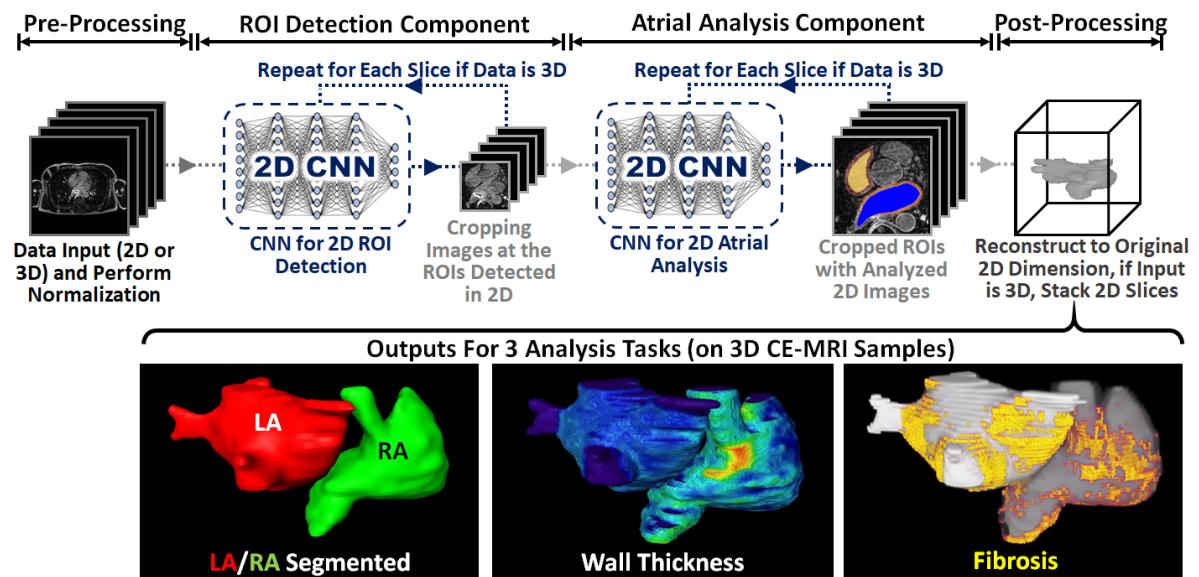
The pre-processing stage was used to normalize the image intensities and slice image volume into 2D images for 3D data. The deep learning stage consisted of two CNNs, both using the same architecture (**Figure 7.3B**). The first CNN localized and cropped a focused sub-region containing the atria to alleviate class imbalance, reduce task complexity and computational cost. The second CNN then performed targeted prediction on the sub-region to produce the desired output based on the given input labels. Extensive hyper-parameter tuning on the number of layers, layer depth, and kernel size of the CNN architecture was performed to optimize the performance. The post-processing stage involved padding the cropped outputs to the original input dimensions, and in the cases where the input was 3D, the individual slices were stacked and reconstructed to the original 3D dimension.

The deep learning stage consisting of the two CNNs used in a sequential manner are

shown in **Figure 7.3A**, using the LGE-MRIs to demonstrate the data flow. The first CNN performed coarse segmentation on a down-sampled version of the input image to decrease computational intensity. The LGE-MRIs were down-sampled four times to $144 \times 144 \times 44$ or $144 \times 144 \times 88$, and the cine-MRI images were down-sampled two times to 104×78 and 78×104 . The centre of mass of the atria was calculated from the coarse segmentation in each image. A patch was then cropped centred on the centre of mass, leaving out the majority of background pixels which significantly improved class imbalance for the next stage of segmentation. The size of the patch was 272×272 for the LGE-MRIs, 80×96 for the four-chamber view cine-MRI, and 64×80 for the two-chamber view cine-MRI. These dimensions were selected based on the maximum size of the atrial chambers in the entire dataset, as well as ensuring it was divisible by 2^4 to satisfy the dimensionalities of the CNN used. The second CNN then performed regional segmentation on the ROIs cropped from the initial images. Finally, the segmentations were zero-padded to their original dimensions to obtain the final segmentation.

The same CNN was used for both stages of the two-stage pipeline and consisted of a modified U-Net architecture with additional residual connections and batch-normalization to improve the convergence. The first half of the CNN was an encoder to learn dense features from the input through several convolutional layers of increasing depth. The convolutional layers contained 5×5 kernels and a stride of 1, and the number of feature maps increased from 16 to 256. At every 1 to 3 convolutional layers, residual connections were added to improve feature learning and 2×2 convolutions with a stride of 2 were used to progressively down sample the image by a factor of 2. The second half of the CNN was a decoder to reconstruct the image back to the original resolution for segmentation through several 5×5 convolutional layers of decreasing depth. The number of feature maps of the convolutions in this part of the network decreased from 128 to 32. The images were progressively up sampled by a factor of 2 with 2×2 deconvolutional, or transpose convolutional, layers with stride of 2. Residual connections were also added at every 1-3 convolutional layers. In order to directly preserve high-resolution features from the input, feature forwarding residual connections were also used to concatenate the outputs of the convolutional layers in the encoder part to those in the decoder path at four different points along the CNN. Overall, apart from the final output layer, batch normalization and PReLU were used after every convolutional layer along the entire CNN for normalization, and 50% dropout was used at every layer for regularization to decrease overfitting. The final output layer of the CNN contained a 1×1 convolution with a stride of 1.

a Deep Learning Pipeline for Automatic Atrial Analysis



b The Same 2D CNN Architecture Used in the Deep Learning Pipeline Above

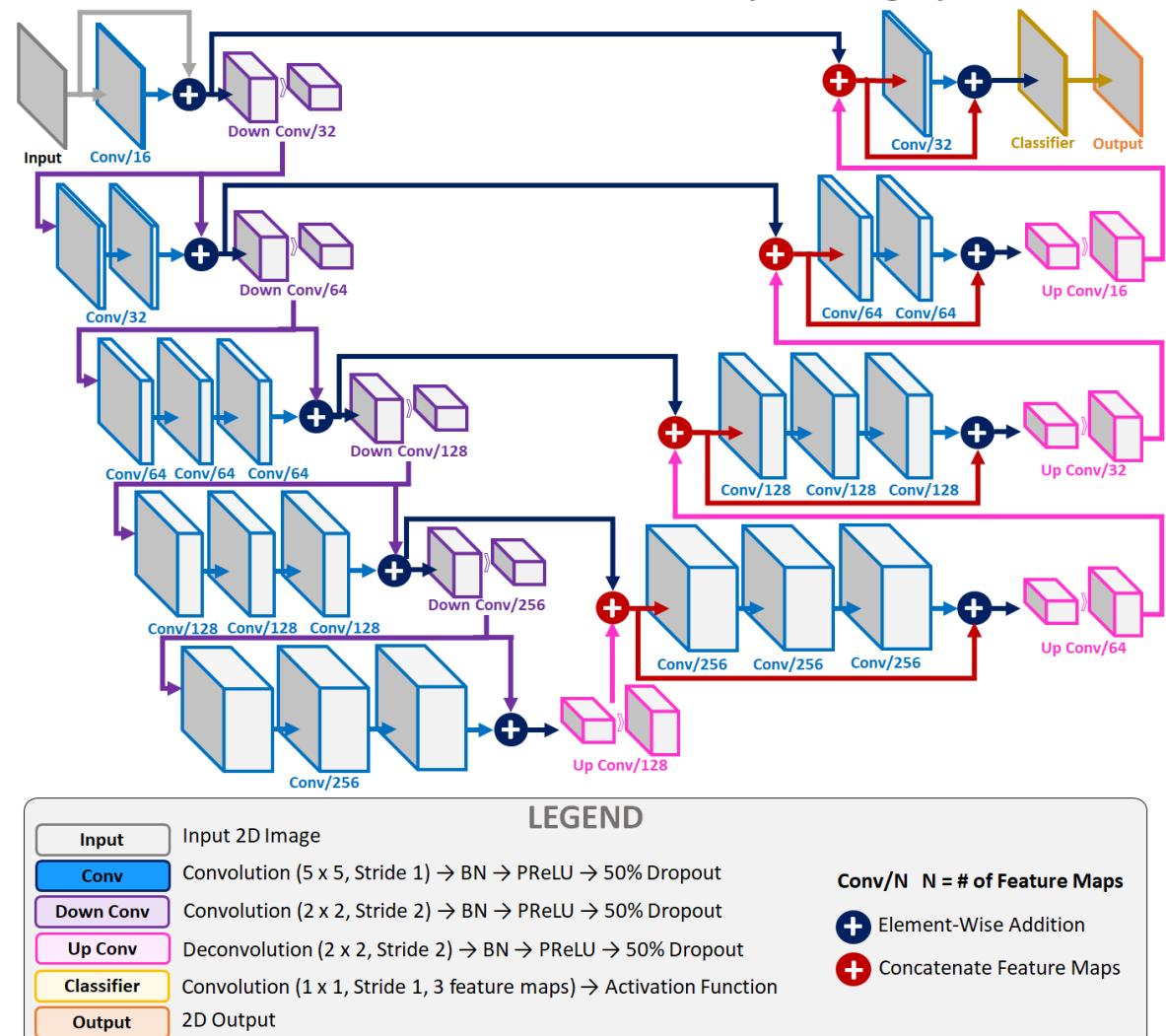


Figure 7.3. Illustration of the AtriaNet pipeline for automatic extraction of pathologic

biomarkers from cardiac magnetic resonance imaging (MRI). **A)** The two-stage convolutional network (CNN) configuration of AtriaNet for fully automatic left atrial (LA) and right atrial (RA) segmentation, AWT estimation, and fibrosis quantification. The first stage detects the region of interest (ROI) from the MRI containing the atrial chamber. The second stage then performs the analysis on the subsequent ROI to produce the final output. **B)** The architecture of the CNN used both stages of AtriaNet. BN, batch normalization; Conv, convolution; PReLU, parametric rectified linear unit.

The Adam optimizer [225] was used to minimize the loss function during training with a constant learning rate of 0.0001 and the exponential decay rates of the 1st and 2nd moment estimates were set to 0.9 and 0.999, respectively. To reduce the computational burden of the large images needed to be processed, all data were stored in the hierarchical data format after pre-processing. Before each epoch, online data augmentation was used to randomly augment each data in the training set with a probability of 50%. The augmentation strategies included random scaling, rotations, flipping, and elastic deformations, providing the CNN with a more diverse set of training images without increasing the memory costs associated with increasing data samples. The CNN was trained with a maximum limit of 1000 epochs, with a criterion to stop training if the accuracy on the validation set did not improve after 50 epochs. A batch size of 8 was used for the LGE-MRI dataset, and 32 for the two cine-MRI datasets. The training set was also shuffled for each epoch to increase randomness. After every epoch, the performance of the CNN was evaluated on the validation set with the Dice score. The parameter set of the CNN which achieved the highest validation accuracy was saved and used on the testing set. The CNN was developed in Tensorflow, an open-source Python deep learning library, and TF-Learn, a high-level Python API for Tensorflow. The training was performed on an Nvidia Titan V GPU with 5120 CUDA cores and 12 GB RAM. The training phase took approximately five hours. Predictions took approximately 5 seconds for each 3D LGE-MRI and 0.1 second for each 2D cine-MRI image.

7.2.2.1 AtriaNet Configuration for Segmentation

In order to train AtriaNet for segmentation on cine-MRIs and LGE-MRIs, the output activation of the second CNN was set to softmax, and the CNN was trained with a dice loss function for categorical classification on a pixel-by-pixel basis. The number of feature maps in the final convolutional layer was 4 for segmenting the LGE-MRI dataset (background, LA cavity, RA cavity, bi-atrial wall), 3 for segmenting the four-chamber view cine-MRI dataset

(background, LA cavity, RA cavity), and 2 for segmenting the two-chamber view cine-MRI dataset (background, LA cavity).

To alleviate class imbalance, a multi-class Dice loss function was used during training to assign higher priorities to the pixels containing the atria during prediction. The Dice loss also increased the speed of convergence, by significantly reducing computational costs. The formulation of the loss function varied when training on datasets with differing number of classes. The general formulation for the loss function, L_{Dice} , was given by

$$L_{dice} = 1 - \frac{1}{N_c} \sum_{i=1}^{N_c} F_{dice}(p_i, g_i) \quad (\textbf{Equation 7.4})$$

where N_c was the number of classes in the label. The function $F_{dice}(p_i, g_i)$ represented the individual dice scores for each class from $i = 1$ to N_c , where p_i and g_i represented the predicted and ground truth 2D binary masks for class i . F_{dice} was defined as

$$F_{dice}(p_i, g_i) = \frac{2 \sum_x \sum_y p_i g_i + 1}{\sum_x \sum_y p_i^2 + \sum_x \sum_y g_i^2 + 1} \quad (\textbf{Equation 7.5})$$

where p and g were of dimensions of x and y . N_c had a value of 3 for the LGE-MRIs, 2 for the four-chamber view cine-MRIs, and 1 for the two-chamber view cine-MRIs.

7.2.2.2 AtriaNet Configuration for Wall Thickness Estimation

The second CNN in AtriaNet was configured to have the output activation as ReLU, and was trained with a MSE loss function due to the task being formulated as a linear regression problem. This was due to the AWT being a continuous variable, such that using ReLU restricts the predicted values to be continuous and positive. The MSE was thus used to accommodate the presence of continuous variables.

During the loss calculation, the background pixels were masked to neglect the effect of the high proportion of zero-pixels which would increase the difficulty for convergence. This alleviated the zero-bias due to the large class imbalance. The loss function, L_{MSE} , was formulated as

$$L_{MSE}(p_{wall}, g_{wall}) = \frac{1}{n} \sum_i^n (g_{wall} - p_{wall})^2 \quad (\textbf{Equation 7.6})$$

where p_{wall} and g_{wall} represents the AWT values in the predicted and ground truth 2D AWT maps. All other parameters in the CNN remained the same as described in the previous section.

7.2.2.3 AtriaNet Configuration for Fibrosis Quantification

The configuration of AtriaNet for fibrosis quantification is similar as in segmentation, except a different number and category of classes were used. The output of the second CNN was modified from its segmentation version to accommodate four output labels (background, non-fibrotic bi-atrial wall, LA fibrosis, and RA fibrosis). The same dice loss function as in **Equation 5.4** was used with the new labels, along with a softmax output activation. All other parameters in the CNN remained the same as the configuration for segmentation.

7.2.3 Evaluation

Technical evaluation metrics included the Dice score, IoU, sensitivity, specificity, Hausdorff distance, and STSD. These metrics were defined in **Chapter 3.6.1**. The technical measure used for AWT estimation was the average pixel-wise error which quantified the error of each predicted AWT value with

$$Error_{pixelwise} = \frac{1}{m} \sum_i^m |x_i|_{atria} - |x'_i|_{atria} \quad (\textbf{Equation 7.7})$$

to show the absolute error of the CNN between the m predicted, $x'|_{atria}$, and ground truth, $x|_{atria}$, AWT values within the atrial wall pixels.

The technical measures used for fibrosis quantification was the Kolmogorov-Smirnov (KS) score which measured the correlation between the ground truth and predicted spatial distributions of the fibrosis in the LA and RA walls. The KS score was defined as

$$KS = 1 - Max|F_A(x) - F_B(x)| \quad (\textbf{Equation 7.8})$$

where $F_A(x)$ and $F_B(x)$ were the cumulative frequency distributions of the prediction, A , and ground truth, B , LA fibrosis coordinates in 3D. The KS score produced values between 0 to 1, where 1 represented the prediction and ground truth containing identical distributions.

We also evaluated biological metrics including the atrial diameter and volume errors

which is defined in **Chapter 3.6.2**. The biological metrics used for AWT estimation was the relative difference between the mean LA and RA wall thicknesses of the predicted and ground truth maps. Lastly, the fibrosis error was measured as the difference between the fibrotic percentage in the LA and RA wall pixels of the predictions and ground truths.

7.3 Results

7.3.1 Accurate Atrial Chamber Segmentation

The segmentation accuracy of AtriaNet was evaluated on the three datasets described (**Figure 7.4**). We first quantified the overlap between the human expert annotations and predicted atrial segmentation using the Dice score. We then assessed the anatomical accuracy of the predictions by analysing STSD errors from the ground truth. All results presented were obtained by evaluating AtriaNet on the testing set during development and parameter tuning.

The Dice score of AtriaNet on the 2D UK Biobank cine-MRIs was 95.6% and 96.1% for the LA and RA cavities respectively (**Figure 7.4A**). Expectedly, the Dice scores were 1% to 2% higher for the end-diastole images due to the increased atrial areas in the scans. The low standard deviation of 2% suggested extremely consistent performance for both chambers. The results surpassed prior studies that used the same ratios of training, testing and validation with their datasets [274]. Furthermore, visual inspection showed excellent agreements between the AtriaNet predictions and ground truth, as reinforced by the low mean STSD of 1 mm, or approximately half a pixel (**Figure 7.4C**). In particular, challenging regions such as the atrioventricular valve and PV antrum which frequently lack clear anatomical boundaries in MRIs were segmented accurately by AtriaNet.

Segmentations of the atrial chamber in 3D LGE-MRIs from Utah and Waikato were more challenging due to the increased anatomical detail present in the high contrast imaging. Despite this increased difficulty, AtriaNet achieved Dice scores of 91.9% and 91.2% for the LA and RA cavities with interquartile ranges between 90% and 93%, showing high precision 3D segmentation across the test datasets (**Figure 7.4B**). Visualizations in 3D further showed the majority of the predicted atrial anatomy was less than 1 pixel (approximately 0.4 mm), from the ground truth (**Figure 7.4D**). The PV, which were not labelled in the cine-MRI due to low resolution, contained the highest distance errors in the LGE-MRIs. However, the errors were mainly attributed to the predicted veins extended out from the LA chamber at

different lengths than the ground truth, a metric which was difficult to define even during the expert annotation. Further 2D comparative visualization revealed other important structures such as the LA appendage and atrioventricular valves were also segmented with high precision (**Figure 7.4E**).

The atrial cavities were segmented to enable the derivation of the diameter and volume measurements. However, analysis of atrial fibrosis requires the segmentation of the atrial wall tissue. To extract LA and RA walls from the LGE-MRIs, AtriaNet first segmented the epicardial surface to delineate the enclosed tissue mass consisting of both the atrial cavity and wall. The wall was therefore the non-overlap region between the epicardial surface and cavity segmentations. AtriaNet segmented the enclosed atrial cavity and wall with a particularly impressive Dice score of 93%. These observed performances were consistent across both LGE-MRI datasets.

Overall, the consistent performance across the three multi-centre datasets demonstrated the generalizability of AtriaNet. In particular, accurate predictions were maintained regardless of using the larger UK Biobank dataset or the smaller Waikato dataset.

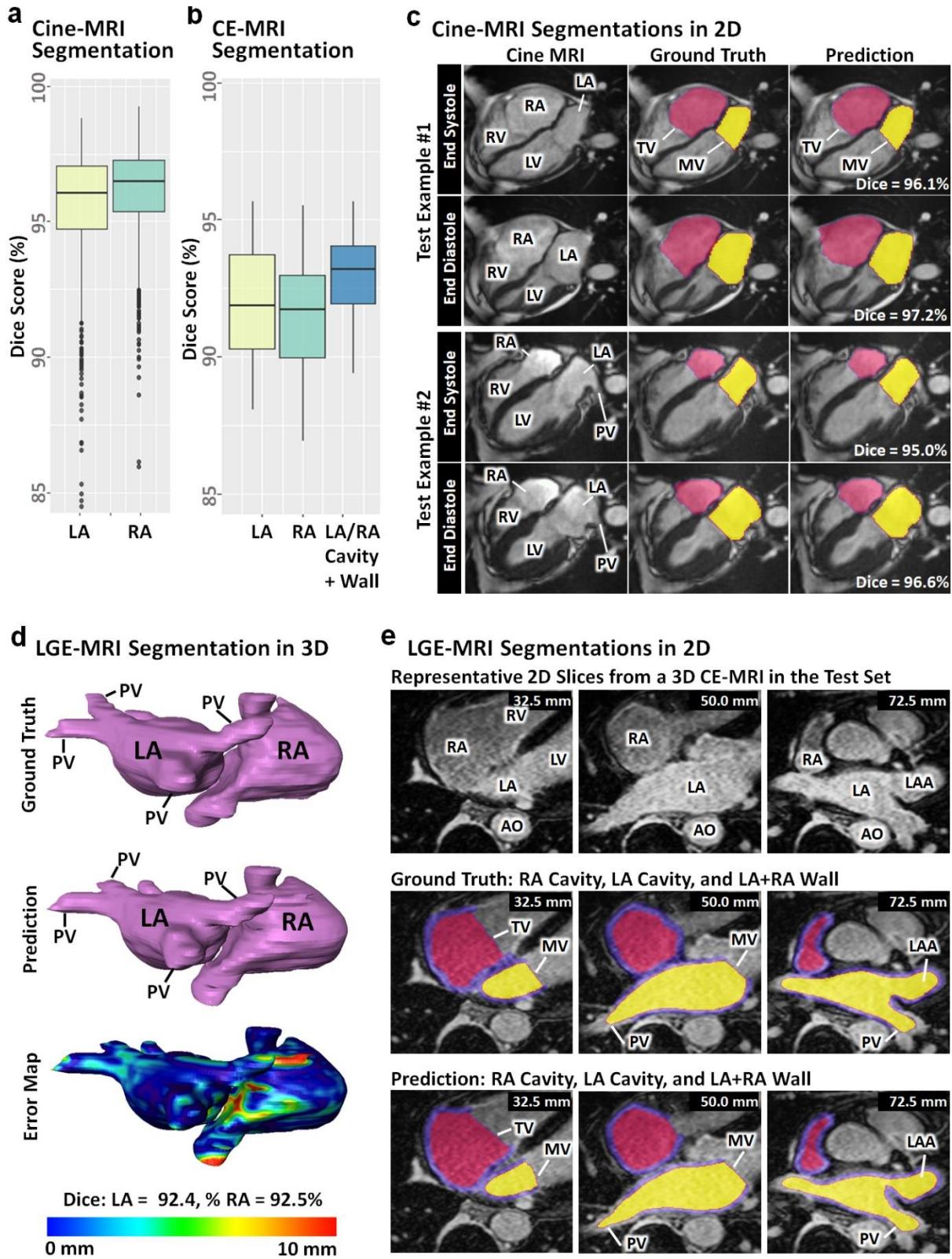


Figure 7.4. Segmentation performance of AtriaNet on 2D cine-magnetic resonance imaging (MRI) and 3D late gadolinium-enhanced (LGE)-MRI. Quantitative results from the left and right atrial (LA/RA) segmentation of **A**) cine-MRIs and **B**) LGE-MRIs measured with the Dice score. **C)** Qualitative results from the segmentation of the two representative cine-MRIs

samples. Each row represents the same cine-MRI sample and each column represents the raw image, ground truth, and AtriaNet prediction for comparison. **D)** Qualitative results from segmenting a representative LGE-MRI in 3D, comparing the ground truth, prediction, and surface to surface distance error map. **E)** Qualitative results for 2D LGE-MRI segmentation for the same patient in **(D)** at representative slices. Each column represents the same slice sample and each row represents the raw LGE-MRIs, ground truth, and prediction. The spatial position of each LGE-MRI slice is labelled for each sample. AO, aorta; LV, left ventricle; LAA, left atrial appendage; MV, mitral valve; PV, pulmonary vein; RV, right ventricle; TV, tricuspid valve.

7.3.2 AtriaNet Calculates the Chamber Volume and Diameter with Clinical Level Precision

AtriaNet is able to efficiently compute accurate cardiac measurements directly from its automatic segmentations. We compared the accuracy between the AtriaNet atrial diameter and volume predictions against those calculated from the expert segmentations in all three datasets. In the cine-MRI dataset, the diameter and volume errors were consistent for each atrial chamber, yielding accuracies of 97% for the RA and 96% for the LA (**Figure 7.5A**). The errors in the LGE-MRI datasets were marginally lower for the RA at 93% for both measures, and higher for the LA which had a volume accuracy of 94% and a diameter accuracy of 96% (**Figure 7.5B**). Similar to the segmentation evaluation, the accuracies on the LGE-MRIs were lower on average, although this was less noticeable, especially for the LA. Surprisingly, AtriaNet achieved higher accuracies for estimating the diameter and volume of the predicted atrial segmentation compared to the accuracy in predicting the actual segmentation themselves. This implied a lower propensity for errors when AtriaNet is used to compute important metrics for phenotyping.

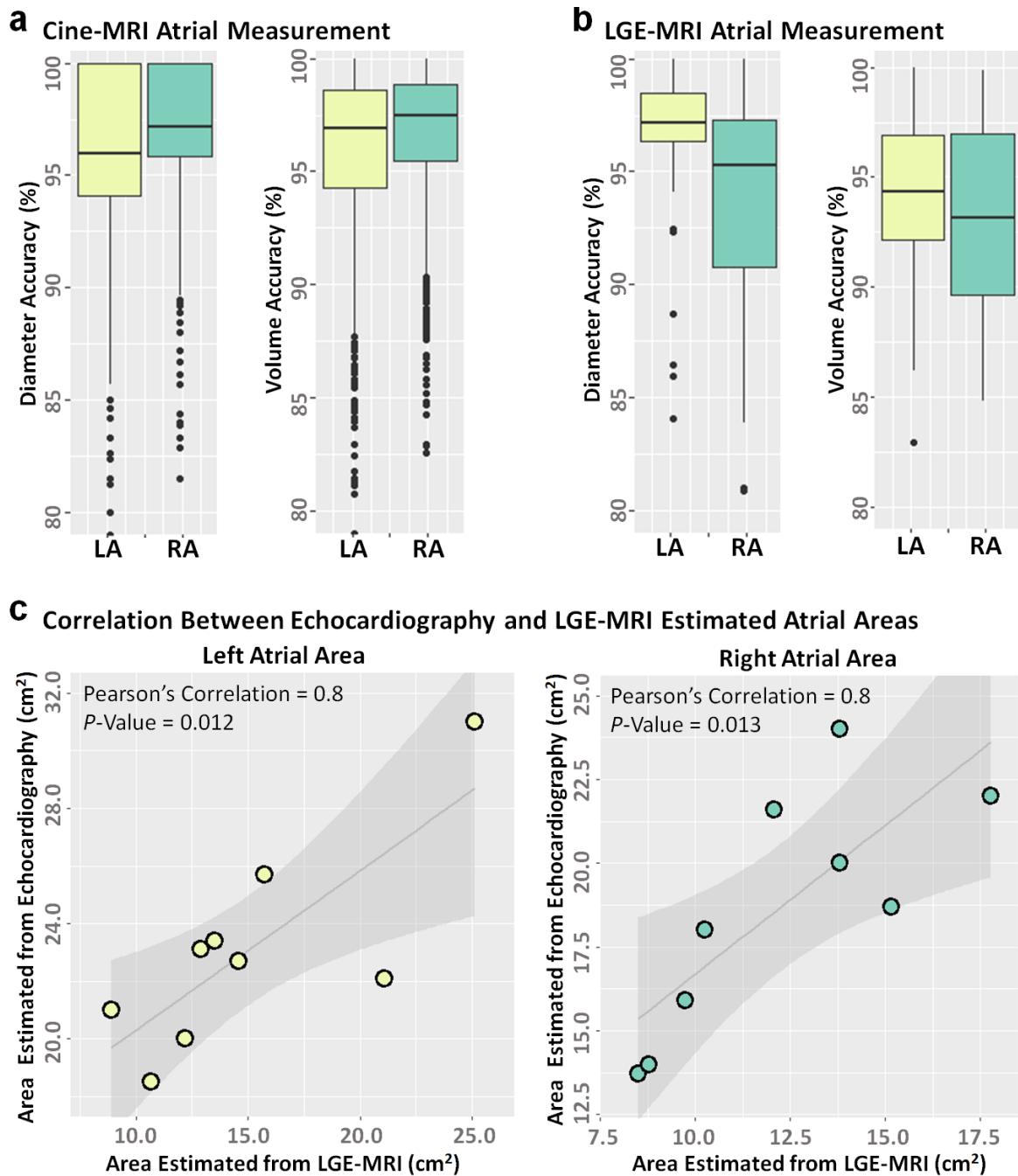


Figure 7.5. The accuracy of AtriaNet for atrial measurement calculation from 2D cine-magnetic resonance imaging (MRI) and 3D late gadolinium-enhanced (LGE)-MRI, and clinical validation using clinical echocardiography. Quantitative results for the percentage accuracy of estimating the left and right atrial (LA/RA) diameter and volume from **A**) cine-MRIs and **B**) LGE-MRIs. **C**) Correlation of the LA and RA areas calculated from clinical assessment of echocardiography scans and computed automatically by AtriaNet from the LGE-MRIs. The linear regression line is shown with a standard error shadow. The correlation coefficients and statistical significance level is also shown.

To further validate the clinical utility of our pipeline, we compared AtriaNet chamber measurement predictions with clinical measures reported by radiologists in 2D echocardiography of the same patients at Waikato (**Figure 7.5C**). Specifically, we compared the correlation between the estimated areas in the AtriaNet 3D segmentation predictions against the areas measured from the echocardiography scans. The areas were estimated from the AtriaNet segmentations by reversing the commonly used bi-plane equation given the volume and diameter of the atrial chambers. **Figure 7.5C** shows the relationship between the LGE-MRI derived areas and the echocardiography-derived areas. The measurements for the both the LA and RA had statistically significant and strong Pearson's correlations of 0.8 ($p = 0.01$). We also compared the correlations directly between the predicted volume/diameter with the echocardiography areas. Both the volume and diameter were correlated with the area for the LA (correlation > 0.7 , $p < 0.05$). The RA volume also had a high correlation (correlation = 0.8, $p = 0.01$), while the diameter was weakly correlated (correlation = 0.5, $p = 0.19$). The strong correlations demonstrated the AtriaNet predictions closely matched the values measured clinically, with the slight variations being attributed to the different imaging modalities used.

Overall, the results demonstrated AtriaNet was able to accurately phenotype the atrial volume and diameter. AtriaNet provides a significant improvement in fully automating atrial chamber measurement calculations. Furthermore, as current clinical methods involve estimating the atrial volume with the bi-plane method on 2D echocardiography scans, AtriaNet provides a more robust alternative by analysing 3D imaging which contains more complete anatomical information.

7.3.3 AtriaNet Estimates Accurate 3D Wall Thicknesses

We extended AtriaNet to compute 3D AWT estimations from the LGE-MRI segmentations. The AWT variation, particularly in the RA, has been demonstrated to have an important role in arrhythmogenesis as it influences the electrical propagation patterns during episodes of arrhythmia [27]. AWT also plays an important role in aiding the selection of regions to target during ablation surgery for treating arrhythmia [111, 265]. The LGE-MRI dataset from Waikato was manually annotated to account for thickness variation in the atrial wall tissue. The dataset is the first of its kind to contain segmentations of both atrial chambers that accounted for the AWT variations in LGE-MRIs. Ground truth maps of the 3D AWT

were obtained by solving computationally expensive differential equations. The mean pixel-wise error was used to evaluate the predicted 3D AWT map.

AtriaNet obtained mean pixel-wise errors of 0.4 mm and 0.6 mm for the LA and RA, respectively, which corresponds to the predictions being within one pixel from the ground truth on average. The mean AWT value was also estimated with a high accuracy of 97% for both atrial chambers (**Figure 7.6A**). AtriaNet successfully reproduced the AWT distribution in all hearts. Visualizations of the 3D predictions showed that key anatomical structures, such as the crista terminalis and the pectinate muscles, were accurately captured and also demonstrated the overall trend of the LA being thinner on average relative to the RA, with the LA having less fluctuation in AWT throughout the chamber (**Figure 7.6C**). The accuracy of the AWT estimation can further be seen in enlarged 2D visualizations (**Figure 7.6D**). Analysis of the AWT distribution graphs showed the distribution of the predicted values was slightly narrower than the ground truth. This revealed the pipeline had difficulty capturing extreme thickness values on both tails of the distribution, although this had a minimal impact on the overall accuracy because the amount of tissue at these extremities was low.

The motivation behind designing AtriaNet to predict the AWT was the extremely labour-intensive pre-processing required and the high computational costs associated with traditional mathematical methods. The generation of the ground truths required the manual delineation of the atrial endocardial and epicardial surfaces for setting up the boundary conditions for the Laplace equations. This process alone can take weeks even by a skilled expert. This, combined with the intensive computation time required to solve the equations, resulted in the entire AWT ground truth taking up to two weeks to produce one sample. On the other hand, the complete AWT map estimation for one 3D LGE-MRI dataset took AtriaNet approximately 10 seconds on average, presenting an over 10000-fold increase in processing speed (**Figure 7.6B**). The high accuracy of our approach and the substantial reduction in processing time can potentially accelerate patient-specific structural analysis in clinics and lead to more efficient and unbiased diagnoses.

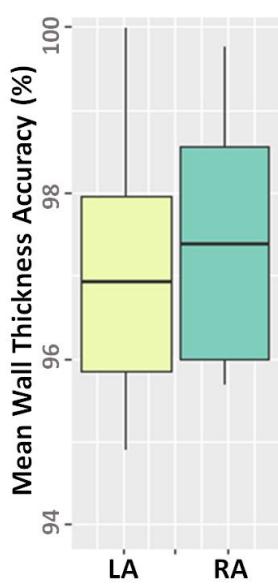
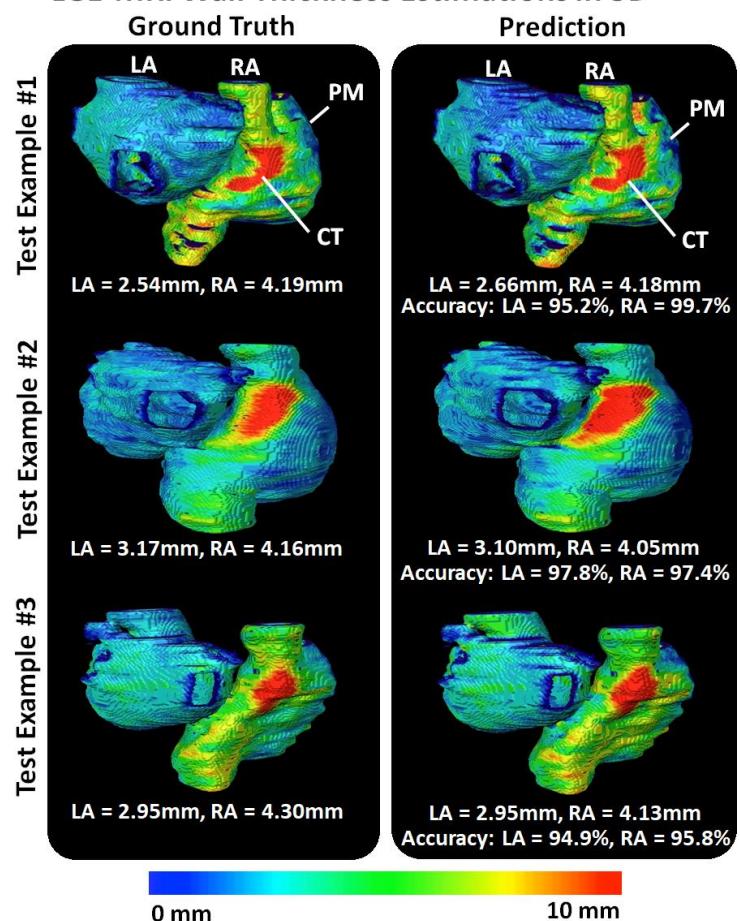
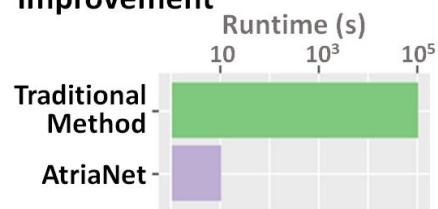
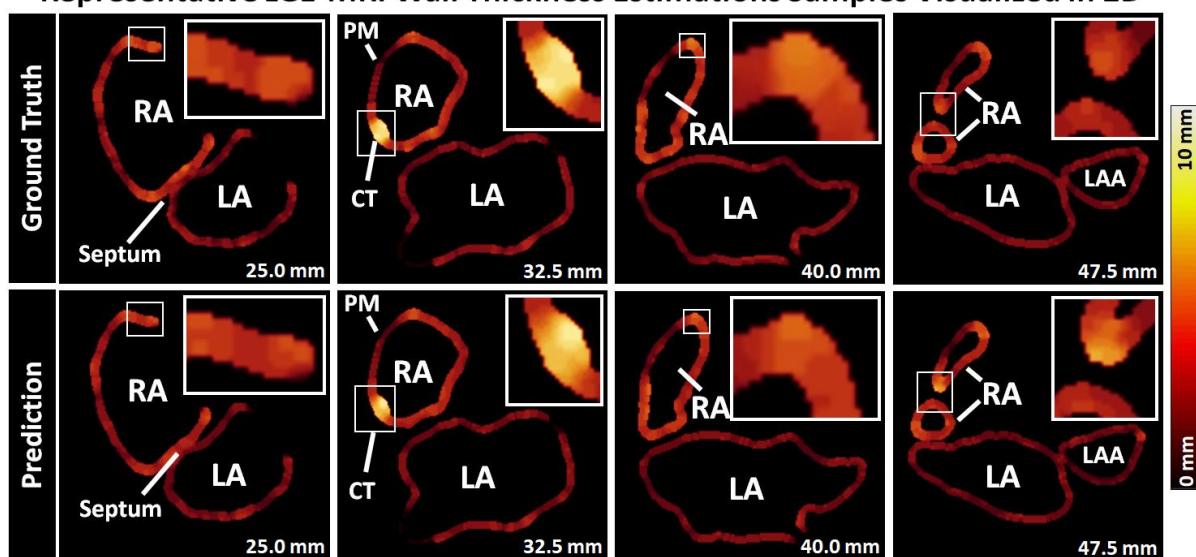
a LGE-MRI Wall Thickness**c LGE-MRI Wall Thickness Estimations in 3D****b Computational Speed Improvement****d Representative LGE-MRI Wall Thickness Estimation Samples Visualized in 2D**

Figure 7.6. Evaluation and visualization of the wall thickness (AWT) estimation accuracy of AtriaNet on 3D late gadolinium-enhanced magnetic resonance imaging (LGE-MRI). **A)** Quantitative results for estimating the left and right atrial (LA/RA) AWT from the LGE-MRIs measured in the mean AWT accuracy. **B)** Comparison of the speed of the traditional method versus our proposed AtriaNet for AWT estimation. The x-axis is displayed in

logarithmic scale. **C)** Qualitative results for comparing the ground truth and predicted AWT distribution maps in 3D for three representative samples, where each row represents the same LGE-MRI sample. Red regions represent thicker walls while blue regions represent thinner walls. The mean AWT of each chamber is also displayed along with the accuracy of AWT estimation. **D)** Qualitative results for the AWT estimations in 2D for the first sample in **(D)**. The ground truths and predictions are compared in the two rows at different slices shown in each column. Thicker regions are shown in the light red/yellow while thinner regions are shown in dark red. Selected regions with high thickness variations are enlarged in the white boxes to improve visualization. CT, crista terminalis; LAA, left atrial appendage; PM, pectinate muscles.

7.3.4 AtriaNet Fibrosis Quantification and Validation with Clinical Electro-Anatomical Mapping

We re-trained AtriaNet to segment the fibrotic and non-fibrotic pixels in the atrial walls of the LGE-MRIs from Utah and Waikato. Ground truth atrial fibrosis was obtained using a semi-automatic, adaptive thresholding method to isolate high-intensity pixels, deemed as fibrotic tissue. The accuracy of AtriaNet was evaluated by comparing the predicted fibrosis distribution with the ground truth distribution using Kolmogorov-Smirnov correlations [275]. Fibrosis segmentations enabled the calculation of the percentage of fibrotic tissue in the wall, an important biomarker which has been shown to predict the success of clinical treatment outcomes. Therefore, we also compared the error in fibrosis percentage calculated from the predictions and ground truths.

AtriaNet obtained correlations of 0.90 and 0.92 ($p < 0.05$) for the predicted LA and RA fibrosis segmentations respectively in the Utah dataset (**Figure 7.7A**), and a 3.0% and 3.7% error for the predicted fibrosis percentage in the LA and RA respectively. The performance on the Waikato dataset was slightly lower with correlations of 0.88 and 0.91 ($p < 0.05$) for the LA and RA fibrosis segmentations, while the fibrosis percentages were estimated with an error of 3.9% for the LA, and a highly accurate 1.9% error for the RA. The average fibrosis percentages were fairly consistent in both datasets, with Utah having 18% for both chambers and Waikato having 20.4% for the LA and 19.5% for the RA. Expectedly, the LGE-MRIs scans acquired in patients who underwent ablation treatment had on average 3% more fibrosis due to the scarring incurred as a result of the surgery. Visualizations of the AtriaNet

fibrosis segmentations also showed excellent matches with the ground truth, and the spatial distribution of fibrotic pixels were accurately reproduced (**Figure 7.7B**). Notably, the regions with a higher fibrosis occurrence were effectively captured, such as the PV antrum and mitral valve in the LA, and the superior vena cava in the RA. The only cases with higher errors were ones with a low fibrosis percentage due to the smaller number of fibrotic pixels being more difficult to detect. However, in these instances, the shape of fibrosis distribution was still accurately predicted.

The accuracy of manually assessing the LA fibrosis from in-vivo imaging has often been scrutinized in previous literature [276]. The varying methods used to define fibrosis results in widely subjective ground truth labels [277], compromising further attempts of developing automated methods to predict fibrosis [278]. We attempted to overcome this issue by comparing our AtriaNet predicted fibrosis maps against low-voltage maps recorded invasively during 3D electro-anatomical mapping of patients at Waikato (**Figure 7.7C**). The low electrical conduction of fibrotic tissue leads to low voltage values being recorded, which could be used to visually assess the accuracy of the fibrosis map generated from the LGE-MRIs. Most studies demonstrated that high intensity regions corresponded fairly closely with lower voltages with the correct thresholds on the LGE-MRIs [129, 279, 280]. **Figure 7.7C** shows three patient data containing the paired LGE-MRI and voltage maps. While direct quantitative comparisons were infeasible due to the differing methods used to anatomically reconstruct the LA [281], the overall pattern of the predicted fibrosis closely matched the low voltage regions (in red) of the voltage maps. This was particularly the case around the PV as well as the mitral valve which contains more fibrotic tissue. While the match between the pairs were not exact due to errors associated with the different LA reconstruction techniques, this experiment successfully demonstrated the ability of the AtriaNet pipeline in quantifying fibrosis in both an accurate and realistic manner.

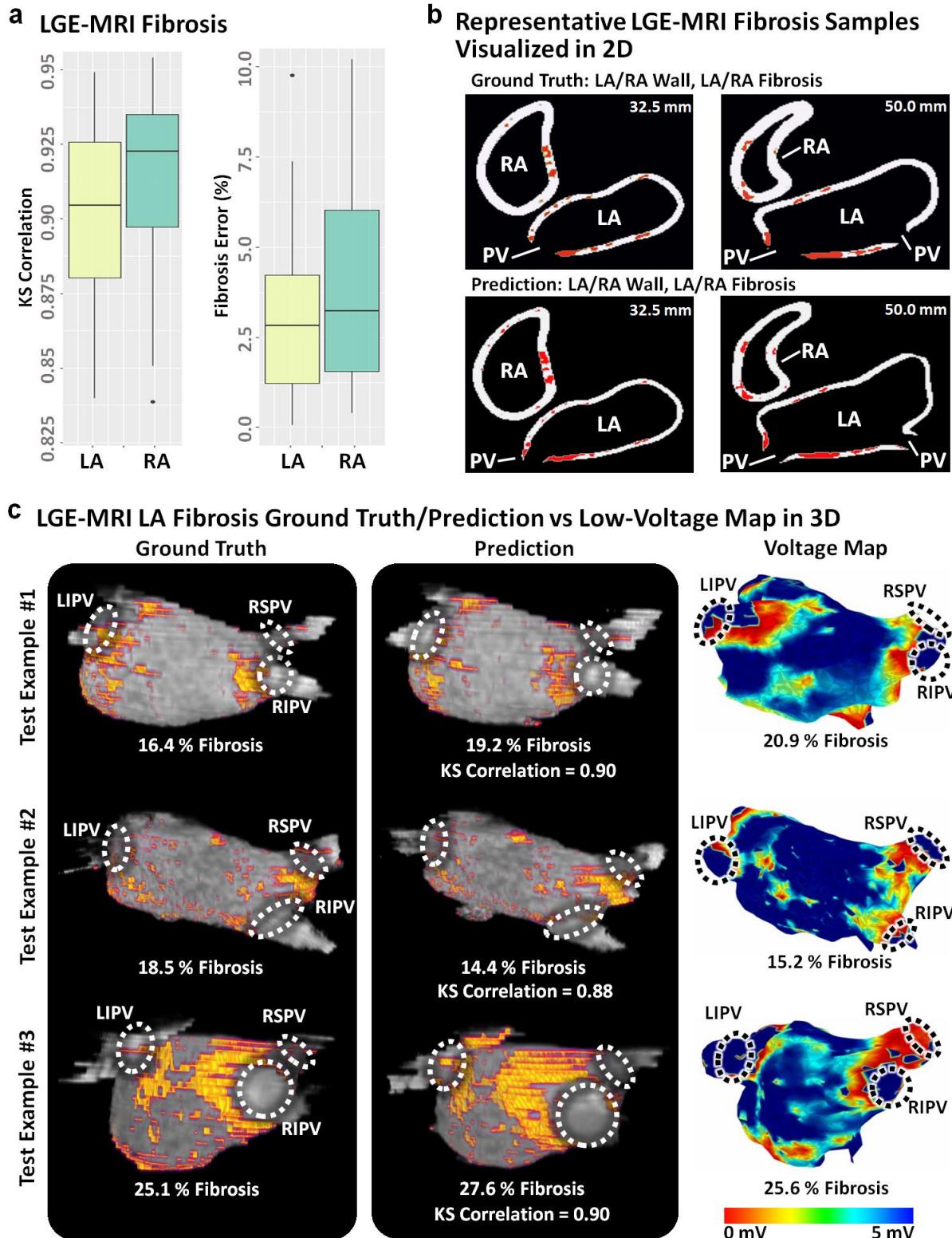


Figure 7.7. Fibrosis quantification performance of AtriaNet on 3D late gadolinium-enhanced magnetic resonance imaging (LGE-MRI) and validation low-voltage mapping of the atria. **A**) Quantitative results for the left and right atrial (LA/RA) fibrosis quantification from LGE-MRI measured in Kolmogorov-Smirnov (KS) correlations and percentage of fibrosis error. **B)** Qualitative results in 2D comparing the ground truth and predicted fibrosis for the one

representative LGE-MRI at two slices. The fibrosis is shown in red. C) Comparison of the 3D LA ground truth and AtriaNet predicted fibrosis maps from LGE-MRIs and the voltage maps showing the fibrotic distribution acquired using electro-anatomical mapping for three patients. Anatomical landmarks are outlined to reference the varying geometries acquired with the different reconstruction methods. Regions of low voltage (red) represent fibrotic tissue in the voltage maps. The correlation between the ground truth and predicted fibrosis are shown, along with the fibrotic percentage of each sample. LIPV, left inferior pulmonary vein; RIPV, right inferior pulmonary vein; RSPV, right superior pulmonary vein.

7.4 Discussion

7.4.1 Automatic Extraction of Bi-Atrial Biomarkers from MRIs

The study presents AtriaNet, the first fully automated pipeline for extracting the atrial anatomy, chamber measurements, 3D AWT, and distribution of fibrosis. In AtriaNet, we designed a simple, effective CNN-based pipeline containing two sequentially arranged CNNs for detecting and subsequent phenotyping of the atria. We enhanced the CNNs from its baseline U-Net architecture with residual blocks at every receptive level and pre-activation normalization at every layer. The proposed method was, to the best of our knowledge, validated on the world's largest cardiac MRI datasets consisting of the UK Biobank 2D cine-MRIs and the University of Utah 3D LGE-MRIs. We further tested the efficacy of AtriaNet on our locally acquired clinical LGE -MRI dataset and compared it with clinical measures. Overall, AtriaNet achieved excellent LA and RA cavity segmentations on the cine-MRIs, and LA and RA cavity and wall segmentations on the LGE-MRIs. Notably, AtriaNet was the first fully automatic approach for RA phenotyping from LGE-MRIs. The atrial measurements subsequently calculated from the segmentations yielded accurate atrial diameter and volume estimates, closely correlating with clinically recorded measures in echocardiography reports. We then demonstrated AtriaNet's ability to accurately compute the AWT distribution on the LGE-MRIs, with a significant improvement in the efficiency of execution and computational speed over traditional methods. Lastly, AtriaNet was capable of predicting the bi-atrial fibrosis distribution on the LGE-MRIs with a high degree of precision, and closely matched low-voltage maps of the atrial fibrosis recorded during non-invasive electro-anatomical mapping.

7.4.2 Importance of Accurate Segmentation of the Atrial Wall

Segmentation of medical imaging is an important step to delineate the anatomical structures of interest for measuring clinical diagnostic biomarkers. In recent years, extensive research has been conducted for the automatic segmentation of the bi-ventricular chambers cavities [235, 282], which includes the LV wall. Similar to our study, state-of-the-art bi-ventricular segmentation approaches utilize enhanced variants of U-Net [205, 249], with the use of the two-CNN configuration becoming popular in recent years [282, 283]. In contrast, atrial segmentation has been investigated in significantly fewer studies, mostly focusing on the LA. Prevailing methods involve extensive hyper-parameter tuning of U-Net based architectures to account for the increased difficulty of segmenting the more morphologically complex LA cavity [224]. In this study, the segmentation capability of our proposed approach is advantageous as AtriaNet is one of few methods able to directly segment the walls of the atrial chambers. Prior studies have mainly focused on cardiac chamber cavity segmentation which is sufficient for calculating chamber measurements [221, 284]. However, the fibrosis distribution and AWT variation which are the key factors of dictating patient stratification and treatment planning [75, 285] can only be extracted from the delineated wall tissue. The atrial walls are significantly more difficult to capture as the walls are approximately two and four times thinner for the RA and LA respectively [27] compared to the LV. Currently available methods initially estimate the atrial wall geometry by extrapolating the cavity, leading to inaccurate extraction of the fibrosis [240] and AWT [286]. Studies which perform accurate fibrosis quantification are still limited to manual wall segmentations [287], while automated AWT estimation has only been attempted for the thick LV wall [288]. In comparison, our proposed AtriaNet directly produced accurate bi-atrial wall segmentations with our simple architecture designed specifically to detect the pixels along the thin atrial wall, which occupies less than 0.01% of the LGE-MRI. Further, our methods has also been extensively validated on the world's largest cardiac LGE-MRI dataset, being more robust than similar studies with small datasets[109]. The precise atrial wall segmentations subsequently led to accurately predicted AWT and fibrosis biomarkers, as seen from both the quantitative and clinical validation results.

7.4.3 Multi-Task CNN Pipeline

The simple design of AtriaNet allowed the pipeline to perform multiple difficult tasks without significant modification. In the AtriaNet architecture, the activation function of the last layer and the loss function of the second CNN were the only parameters altered when executing the three different phenotyping tasks. Since these were fixed parameters, the trained CNN weights were easily shared across different tasks and datasets when data availability was limited. The benefits were more pronounced when developing the model for AWT estimation on the Waikato LGE-MRI dataset, as the CNNs were initialized with pre-trained weights from segmenting the larger Utah dataset to accelerate convergence. The usability of our approach allowed our pipeline to be more efficient by reusing the same model. This also demonstrated the robustness of the proposed configuration of two sequential CNNs for image analysis in general, as our approach was shown to be independent of any specific input, label, or imaging dimension. Computational costs were minimized through the exclusive use of 2D CNNs for both 2D and 3D tasks, as 3D convolutions result in an exponentially greater number of parameters which increases the difficulty of training and convergence [289]. The ease of convergence of AtriaNet was important as the atrial prediction tasks contained data with a high class-imbalance of over 99.9%, increasing the likelihood of locally optimal solutions containing entirely zero-output predictions [290]. These issues were further mitigated with the combination of weighted loss functions to target positive pixels, pre-trained weights to initialize each training session, and convolutions with a relatively lower depth a minimal number of trainable parameters. The lightweight design also allows AtriaNet to be executed on most standard computer hardware without compromising runtimes.

7.4.4 Limitations and Future Work

Future improvements to our study would involve increasing the data quality for validation, especially for the clinical validation of the atrial phenotypes. The clinical data available would also ideally be specifically acquired during the trials for evaluation against the predictions from LGE-MRIs. For example, in future studies, the low-voltage maps recorded would be projected onto the LGE-MRI segmented anatomy during the clinical recording phase to allow for direct comparisons of the fibrosis maps [291]. Further studies

should also be set up to develop more samples of AWT variation. The development of semi-automatic, but accurate methods of AWT delineation would also be beneficial as current manual processes are too time consuming to be applied on a large scale. We also aim to acquire clinical information on our MRI datasets to be able to associate the predicted phenotypes retrospectively with patient outcomes and perform survival analysis. Lastly, the generalizability of AtriaNet would be further validated on more diverse imaging modalities such as CT scans or imaging on other cardiac structures.

7.5 Conclusions

In this chapter, we have proposed a robust pipeline for automated cardiac phenotyping from bi-atrial chambers. In addition, our approach was extensively validated on diverse and cross-centre imaging datasets. AtriaNet was shown to produce high quality segmentations, AWT estimations, and fibrosis quantifications which are crucial prerequisites during clinical assessment and treatment. The increased efficiency and accurate phenotyping may lead to the development of improved clinical diagnosis, patient stratification, and clinical guidance of patients with CVDs. The insights gained from our study may also impact the wider imaging community for transferring our simple yet effective framework to other image phenotyping tasks.

Chapter 8

Automatic Atrial Surface Reconstruction from Clinically Mapped Point Clouds using Convolutional Neural Networks

Current clinical mapping tools used to reconstruct the atria during catheter ablation of AF are labour-intense and time-consuming. Research into atrial reconstruction from anatomical maps consisting of point clouds remains limited, as the currently used algorithms are proprietary and difficult to validate. To improve the accuracy of LA visualization, many clinicians rely on additional imaging methodologies. But this leads to further manual processing and registration of the 3D MRI/CT during the ablation procedure. Therefore, there is an urgent need to develop an approach to reconstruct the atria more accurately and efficiently to guide the targeted ablation procedure.

In this chapter, we proposed a novel CNN method for the automatic 3D surface reconstruction of the LA accurately using a smaller sampled data cloud. To train the CNN, we developed a method of synthesizing a large sample of realistic anatomical maps and corresponding LA geometries using existing mapping data acquired via widely used clinical mapping systems and LGE-MRIs. We further tested the efficacy of our approach on two independent cross-modality clinical datasets and achieved excellent quantitative accuracies when evaluated against reconstructions using traditional methods. Overall, our study may provide a more efficient, cost-effective 3D LA reconstruction approach during catheter ablation, and potentially lead to more targeted treatment of AF. This chapter is presented in a similar form to the potential publication version, with additional details on the methodologies. The publication is cited as:

Xiong, Z., Stiles, M.K., Yao, Y., Shi, R., Nalar, A., Hawson, J., Lee, G. and Zhao, J., 2022. Automatic 3D Surface Reconstruction of the Left Atrium from Clinically Mapped Point Clouds Using Convolutional Neural Networks. Under Review.

8.1 Introduction

Point clouds are a widely used method of storing information acquired in the ever-growing world of data [292, 293]. Current advancements in 3D acquisition technology in the form of sensors, scanners, and imaging capture high-quality data to allow for more refined research of their components and properties [294]. In particular, the acquisition of 3D data in the medical field is an increasingly important area of study in terms of visualizing organ structures, recording real-time anatomical information during surgery and physiological mapping [295, 296]. Compared to 3D imaging, point clouds are significantly more memory-efficient by storing information in a compact and vectorised form. This data format also enables efficient manipulation using simple mathematical operations with low computational costs.

In recent years, medical recording technology, particularly devices in cardiology, has integrated point clouds into the systems for various applications. Catheter ablation is one of the most common clinical procedures for treating complex CVD such as arrhythmia. During the procedure, an estimated geometry of the cardiac chamber is initially constructed using point-by-point catheter recordings on the endocardial surface [297]. The geometry formed from the point cloud is then used to guide and target specific regions containing diseased heart tissue for ablation [20]. Therefore, accurate reconstruction of cardiac chambers from point clouds is vitally important for the effectiveness of the procedure. This is especially the case for atrial chamber reconstruction during catheter ablation of AF, the most common cardiac arrhythmia [205, 224].

Current methods of point cloud to atrial chamber reconstruction, particularly LA, are heavily commercialized and not openly accessible. The two most widely used commercial electro-anatomical mapping systems are the EnsiteNavX (St Jude Medical, Minnesota, United States) and CARTO 3 (Biosense Webster, California, United States). To produce accurate LA models, clinicians further merge the point cloud with anatomical LA segmentations obtained from MRI or CT in advance of the procedure. There is limited research aiming to improve the efficiency and accuracy of LA reconstruction algorithms. The only notable study is Baram et al. who proposed an auto-encoder to perform LA reconstruction from simulated catheter points and LA geometries [298]. The methods proposed were not tested directly on real data and lacked rigorous validation. Therefore, there is a need for a more accurate and robust algorithm capable of fully automatic LA

reconstruction directly from point clouds.

CNNs are currently the main driver of modern computational methods for structured data [299]. The major differences when implementing CNNs for point clouds as opposed to traditional pixels or voxels are the variable lengths and unordered structure of point cloud vectors. This has led many studies to design specialized approaches that adapt CNNs to their respective task, as they have already been proven to be extremely robust in imaging analysis [161, 167]. As the point cloud data is required to be standardized into a consistent shape for the CNN, approaches mainly focus on normalizing the data with pre-processing. Projection-based methods involve mapping 3D point clouds onto 2D surfaces at different angles [300], or onto standardized spherical representations [301], which can be then analysed directly. These studies have focused on selecting the best projection approach, such as using CNNs to analyse multiple projections of the same set of points and aggregate the results to obtain a more robust prediction [302]. Some studies also use CNNs to perform predictions on projections of local points due to the more consistent geometry in a regional area, followed by aggregation of the local outputs into a global prediction [303]. Spherical projections have been more commonly used as more information can be retained in a single 2D representation, although this results in a loss of local details [304]. A more straightforward method is discretization, in which the 3D point clouds are converted into volumetric images which can be directly analysed by CNN [305]. Studies have investigated ways to optimize methods of discretization due to the computationally expensive nature of this type of volume-based analysis. Some approaches have partitioned point clouds into a lattice of voxels, in which each voxel is processed differently depending on the number of points present [306]. To improve accuracy, studies have used adaptive voxel sizes to target regions of high point density and ignore low-density regions [307]. This increases the resolution of the discretized representation of the point set in the regions containing interest without increasing the computational burden.

The recent advancements in CNNs for point cloud analysis have provided a solid baseline for developing a LA point cloud analysis approach. Despite these studies, there still lacks research progress for converting sparse point clouds to volumetric geometries, especially in the medical field. Potential solutions for this complex task could involve state-of-the-art CNNs for 3D medical image segmentation, which specialize in the image reconstruction of extremely fine structures [161, 167]. The popular 3D U-Net architecture [161] has been implemented for a wide range of tasks including heart segmentation [268], and its enhanced version, V-Net [167], achieves further performance improvements. A recent

global benchmarking study has also experimentally deduced the most optimal U-Net CNN configuration for LA segmentation from 3D MRIs [224], surpassing traditional and other CNN methods. Thus, we believe an approach which leverages both leading point cloud analytical techniques and medical imaging CNNs is the best strategy for tackling the task in this study.

In this study, we proposed the first deep learning pipeline for fully automatic surface reconstruction of the LA from point cloud data. Our method achieved anatomically accurate LA predictions directly from point clouds without the need for additional imaging. We tested the method on independent clinical datasets acquired using the two most widely used commercial mapping systems. Our study may potentially be used to improve current mapping systems for guiding ablation procedures to treat AF.

8.2 Methods

8.2.1 Data and Pre-Processing

A summary of the three datasets (*paired training data, test #1, test #2*) used in this study is shown in **Table 8.1**. The CNN was initially trained on a generated dataset (paired training data) and tested on two clinical datasets (*test #1* and *test #2*). The generated dataset was simulated to provide sufficient samples to train the CNN, as clinical data is time-consuming and expensive to acquire. The two clinical datasets both contained LA surface geometries segmented from MRIs or CTs and point clouds acquired with the most widely used commercial mapping systems merged into the same coordinates as the imaging. This provided matching pairs of input point clouds and output LA for testing the CNN. The following sub-sections describe the generation and acquisition of the three datasets in detail.

Table 8.1. Summary of the data used in this study.

Dataset	Training	Validation	Testing
Paired training data	1000	240	300
Test #1: clinical MRI + point cloud	-	-	4
Test #2: clinical CT + point cloud	-	-	2

CT, computed tomography; MRI, magnetic resonance imaging.

8.2.1.1 Paired Training Data

The paired training dataset was generated by merging two separate datasets: 154 LA surface geometries manually segmented from LGE-MRIs [114] and 10 sets of point clouds of the LA recorded with clinical mapping [308]. The point clouds were transformed to fit the same spatial coordinates as the LA segmentations, forming matching pairs of point cloud and LA geometries available for the CNN. Overall, 1540 data samples were generated by exhausting all pairing combinations of the two datasets.

The 154 3D LGE-MRIs from University of Utah were described previously in **Chapter 5.2.1**. In addition, the 3D coordinates of the four PVs in each LA were also recorded for landmark registration with a semi-automatic approach.

The 10 point cloud data were created with clinical mapping during catheter ablation to treat patients with AF in Beijing, China [308]. Similar to the LGE-MRIs, the coordinates of the four PVs were annotated in the maps. The average number of coordinates recorded for the point clouds were 3703 ± 1043 .

The two datasets were merged by transforming the point cloud data using three stages: registration, projection, and discretization. As the coordinates of the PVs were labelled in both datasets, they were first used to register the point cloud through a series of translational, rotational and scaling matrix operations, obtaining the closest possible match of the landmarks. The registered point cloud was then spherically projected onto the surface of the 3D LA geometry using its centre-of-mass as a reference point. Finally, the projected point cloud was discretized using the alpha-concave hull algorithm [309] to generate a dense mesh of the point cloud. The vertices of the mesh were then discretized into integers to produce a 3D image representing a partial shell of the LA.

The paired training dataset was split into training ($N = 1000$), validation ($N = 240$), and testing ($N = 300$). The input data and labels were the point clouds and the LA segmented from the LGE-MRI dataset, respectively. The data was split such that an LA geometry from a given LGE-MRI was only present in one of the three datasets to avoid repeating labels.

8.2.1.2 Test #1: Clinically Paired MRI and Point Cloud Data

The 4 LGE-MRIs from Waikato Hospital were previously described in **Chapter 7.2.1**. The 4 patients selected underwent catheter ablation with the CARTO 3 mapping system [131]. The average number of points recorded for the patients was 2230 ± 790 . Prior to the ablation procedure, the corresponding MRI scans were manually annotated by a team of experts to define the LA geometries. During clinical assessment, the LA was merged with the point

clouds recorded during ablation mapping to spatially match the two data. For pre-processing, the point clouds were discretized to create a 3D input LA shell for the CNN. The corresponding LA geometries from the MRIs were used as the ground truth for evaluation.

8.2.1.3 ***Test #2: Clinically Paired CT and Point Cloud Data***

CTs were obtained from 2 patients at The Royal Hospital Melbourne, Australia, undergoing catheter ablation [310] with the EnsiteNavX mapping system [311]. The average number of mapped points was 2818 ± 206 . Similar to the *test #1* dataset above, the LA were manually segmented from the CTs and merged in the clinic with the point clouds. The point clouds were then discretized to create a 3D input LA shell for the CNN using the method described in **Chapter 8.2.2.3**, and the respective LA geometries from the CTs were used as the ground truth for evaluation.

8.2.2 **Method of Generating the Paired Training Data**

8.2.2.1 **Landmark Registration from Point Cloud to LGE-MRI**

To create a realistic dataset containing paired LA geometry and matching point cloud data, the clinically recorded point clouds from Beijing were first registered onto the Utah LA segmentations. This involved transforming the 3D coordinates of the point clouds such that the location, orientation, and scale were the same as the Utah LGE-MRI LA data. The anatomical landmarks of the two datasets were matched to ensure anatomical accuracy during the registration process.

The *X* and *Y* axes of the point clouds were first flipped to match the geometry of the LA segmentations which was stored in an image format. The point cloud was then translated such that its centre-of-mass aligned with the centre-of-mass of the LA segmentations. The new coordinates, x',y',z' , was calculated by

$$PT_{x',y',z'} = PT_{x,y,z} + COM_{Utah} - COM_{PT} \quad (\text{Equation 8.1})$$

where *COM* was the centre-of-mass of the LA segmentations and the point cloud data. The landmarks of the two data were matched through orientation of the point cloud about its centre-of-mass. This included aligning the PVs using the labelled coordinates available in both data. Since the relative positioning of the PVs were different within the two data due to

patient-specific variations in their LA morphology, the task was simplified to plane-matching. A 2D plane was computed for each data initially using three PVs, the RSPV, LSPV, and LIPV. The normal vector of the plane was therefore the cross-product of two vectors, v , connecting the LIPV and RSPV, and w , connecting the LSPV and RSPV. This was computed as

$$\tilde{n}_{x,y,z} = [v_y w_z - v_z w_y, v_z w_x - v_x w_z, v_x w_y - v_y w_x] \quad (\text{Equation 8.2})$$

for the normal vector, \tilde{n} , in 3D. After the plane was defined, it was then tilted along the vector between the RSPV and LIPV to minimize the sum of the orthogonal Euclidean distance of both the LSPV and RIPV, forming a plane-of-best-fit for all four PV coordinates. Once the two planes were computed for the point cloud and LA cavity, the task was therefore to orientate the normal vector of the point cloud plane to align with the normal vector of the LA cavity plane.

The orientation was done about the centre-of-mass of the four PV coordinates, and the rotational matrix computed from the alignment of the normal vectors was used to subsequently orient the coordinates of the point cloud. The orientation was done in three steps for each axis individually, such that the orientation was performed in 2D along the three different axes. The two normal vectors were first converted into polar coordinates while fixing the x coordinates such that it can be rotated along the x axis. The polar coordinates were computed about the centre-of-mass of the PVs for both data, and was defined by

$$r_x = \sqrt{\tilde{n}_y^2 + \tilde{n}_z^2} \quad (\text{Equation 8.3})$$

$$\theta_x = \tan^{-1} \left(\frac{\tilde{n}_y}{\tilde{n}_x} \right) \quad (\text{Equation 8.4})$$

when fixed in the x axis to obtain the distance, r_x , and angle, θ_x . The rotational vector, R_x , required was then

$$R_x = \theta_x^{PV Utah} - \theta_x^{PV PT} \quad (\text{Equation 8.5})$$

between the polar angular coordinates of the point cloud PVs, $\theta_x^{PV PT}$, and Utah LA PVs, $\theta_x^{PV Utah}$. The entire point cloud data was then converted into polar coordinates using the same **Equation 8.3-8.4**, and R_x was added to the angular component. The points were lastly

converted back into Cartesian coordinates to obtain the rotated point cloud about the x axis. The rotated coordinates were computed as

$$PT_y = r_x^{PT} \cos(R_x + \theta_x^{PT}) + COM^{PT}_y \quad (\text{Equation 8.6})$$

$$PT_z = r_x^{PT} \sin(R_x + \theta_x^{PT}) + COM^{PT}_z \quad (\text{Equation 8.7})$$

to obtain updated y and z coordinates, PT_y and PT_z , about the centre-of-mass, COM^{PT} , of the point cloud. The above process was then repeated to rotate the point cloud about the y and z axes.

The last step of the point cloud registration was to scale the coordinates to the same magnitude as that of the Utah LA cavity. This ensured the registered point cloud achieved maximum coverage of the LA. The scaling was performed by

$$PT_a = (PT_a - COM^{PT}_a) \frac{\max\{LA^{Utah}_a\} - \min\{LA^{Utah}_a\}}{\max\{PT_a\} - \min\{PT_a\}} + COM^{PT}_a, \forall a \in \{x, y, z\} \quad (\text{Equation 8.8})$$

where the point cloud, PT , is scaled about the centre of mass of the point cloud, COM , and LA_{Utah} is the LA, for every axis, a .

8.2.2.2 Projection from Point Cloud to LGE-MRI

A second step was introduced to project the point clouds directly onto the surface of the LA segmentations, simulating realistic mapping catheter recordings obtained on the surface of the LA (**Figure 8.1**). The projection was performed point-by-point by defining a vector, v , through the centre of mass of the LA cavity and one point in the point cloud, and finding the intersection of v with the outer LA surface. Here, the LA surface was obtained by subtracting the original LA segmentation mask with a 1-pixel erosion of the mask to obtain a shell of the inner and outer surfaces of the LA. The inner LA surface was then vectorised by defining a vector containing the pixel-locations of all points in the LA interior. Since the inner LA surface was sparsely defined with 1-pixel gaps and was curved, the surface-projection was performed by minimizing the distance of a point along v with any point on the inner LA surface. Computationally, this was performed by

$$PT_i^{proj} = \min_{v_a} \{v_a - LA_i^{surface}\} \quad \forall i \in LA^{surface}$$

where $v = COM^{Utah} - PT_i$ **(Equation 8.9)**

where v_a is every point along the v within a pre-defined range of a distance of ± 250 in increments of 1 pixel, i was the index of a point in the point cloud data and the LA surface, $LA_{surface}$ was all points in the vectorised LA surface, and PT_{proj} was the projected point. A range of 250 was selected as this was the maximum dimension of any LA in the Utah dataset.

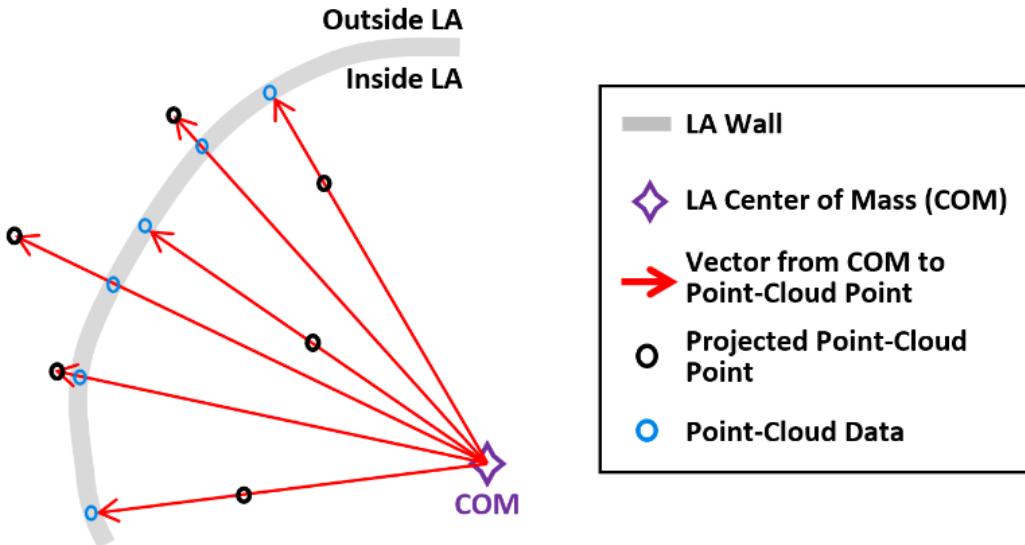


Figure 8.1. Illustration of the projection of the registered points onto the left atrial (LA) surface during data simulation. The point cloud is projected directly onto the LA wall to simulate the points recorded invasively using a catheter during electro-anatomical mapping.

8.2.2.3 Creating Partial Left Atrial Shell from Point Cloud

The vector of the projected point cloud was converted to a raster, or image, format to provide for a fixed input into the CNN. This also simplified the normalization pre-processing stage, as vectors of different lengths and magnitudes would be difficult to standardize. However, discretization of vectors requires densely packed coordinates to form a smooth volume in space. Hence, a novel method of applying the concave hull algorithm was utilized to significantly increase the number of points in the projected point clouds while maintaining its original geometric shape.

An implementation of the alpha-concave hull algorithm was utilized in this study [309], involving the tuneable parameter alpha which was used to adjust the concavity of the output. A high alpha value would increase the convexity of the output, and a low alpha value

may form a geometry which excludes many data points to preserve concavity (**Figure 8.2**). The alpha value was manually tuned and defined as 5 in this study, as this produced an output which included most points in the original dataset while still maintaining the natural curvature of the LA.

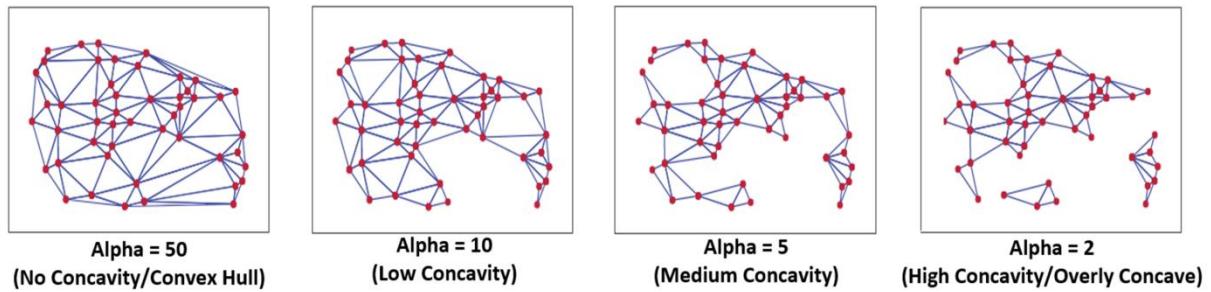


Figure 8.2. Example of the impact of the alpha parameter in the concave hull algorithm when generating a mesh from individual points in the point cloud. High alpha values result in overly convex solutions which lack the detailed curvature present in the original set while low alpha values produce a disconnected and incomplete mesh.

The output of the concave hull algorithm included the points of the concave hull as well as the edges connecting the points. To significantly increase the number of points in the point cloud, the concave hull algorithm was applied three times (**Figure 8.3**). After each iteration, the points along all edges generated by the algorithm were also extracted and added to the vector set defining the point cloud. Thus, inputting this updated point cloud vector into subsequent iterations of the concave hull algorithm would lead the number of points to increase exponentially. Three iterations of the concave hull transformed a point cloud vector of ~4000 samples to over 250,000 samples. The output of the third concave hull iteration was a dense mesh, which was then converted into an image volume by simply rounding each coordinate to the nearest integer and setting the pixel location of the coordinate to 1. This therefore formed a 3D image volume of a partial LA shell from the point cloud.

Due to the high density of the point cloud mesh generated, the image volume was further processed to provide more features to input into the CNN. Since rounding the coordinates of the point cloud mesh introduced many repeated pixel locations the frequency of occurrence at each pixel location was calculated, creating a heat map over the entire image volume. This produced a 3D image volume with a density gradient, providing the CNN with richer features over a binary input.

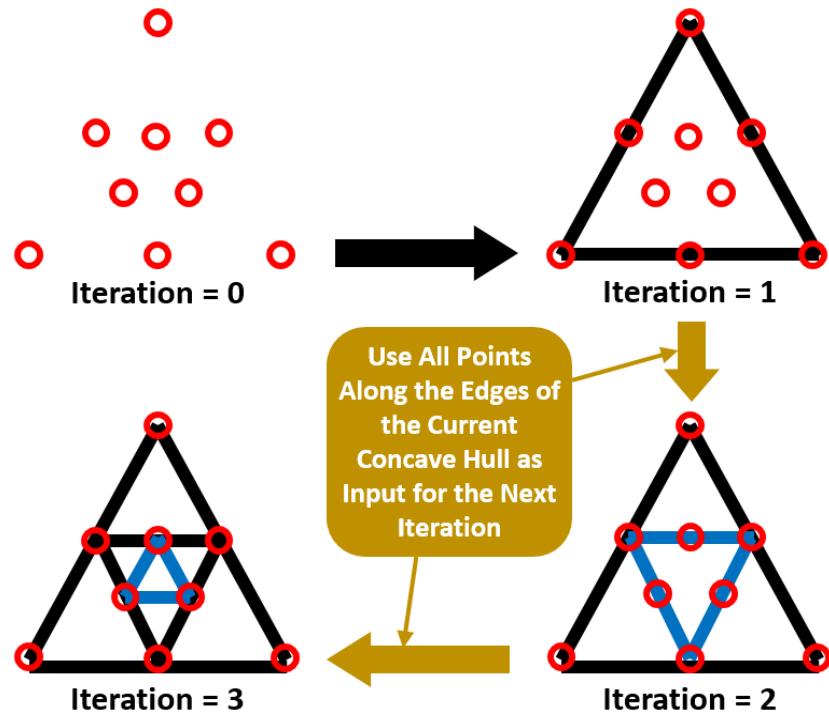


Figure 8.3. Illustration of the iterative concave hull algorithm proposed in this study to the discretization of the point clouds into an image volume containing the partial left atrial (LA) shell used as the input for the convolutional neural network (CNN). Given a set of points in space (iteration = 0), a simple mesh is first generated (iteration = 1), and subsequent iterations utilize all points along the edges of the previous mesh as vertices for the new concave hull (iteration = 2), resulting in an increasingly dense mesh (iteration = 3).

8.2.3 CNN for Left Atrial Reconstruction from Point Clouds

A CNN was developed to predict the 3D surface LA geometry given the point-cloud recording of the LA during clinical mapping. The architecture is shown in **Figure 8.4**, and the full summaries of parameters are shown in **Table 8.2**. The point cloud was first pre-processed into a fixed input volume. All inputs were then cropped to a standard size of $128 \times 208 \times 88$ pixels, removing background pixels to alleviate class imbalance. The CNN architecture consisted of a modified 3D U-Net architecture with additional residual connections to improve the convergence. We used an FCN to decrease computational costs and ensure the CNN operates independent of input size. The CNN was relatively light-weight as the maximum number of convolutional kernels per layer was 128. This further ensured faster training and convergence, as well as being significantly less memory intensive.

The first half of the CNN was an encoder to learn dense features from the input

through several convolutional layers of increasing depth. The convolutional layers contained $5 \times 5 \times 5$ kernels and a stride of 1 for an increased receptive field over traditional $3 \times 3 \times 3$ kernels, and the number of feature maps increased from 8 to 128. At every 1 to 3 convolutional layers, residual connections were added to improve feature learning and $2 \times 2 \times 2$ convolutions with a stride of 2 were used to progressively down-sample the input by a factor of 2. The additional residual connections did not contribute to an increase in parameters but greatly increased information flow throughout the network, allowing important features to be retained as the input is down-sampled. The use of convolutions to down-sample the input as opposed to traditional pooling also implicitly enabled the CNN to learn the important features while removing unimportant information during compression.

The second half of the CNN was a decoder used to reconstruct the input back to the original resolution through several $5 \times 5 \times 5$ convolutional layers of decreasing depth. This was done to facilitate subsequent segmentation. The number of feature maps of the convolutions in this part of the network decreased from 64 to 16. The input was progressively up-sampled by a factor of 2 with $2 \times 2 \times 2$ deconvolutional, or transpose convolutional, layers with stride of 2. Residual connections were added at every 1-3 convolutional layers. In order to directly preserve high-resolution features from the input, feature forwarding connections were also used to concatenate the outputs of the convolutional layers in the encoder path to those in the decoder path at four different points along the CNN. This allowed the CNN to learn from both raw high-level features as well as condensed low-level features. This also greatly improved the consistency of reconstruction by essentially guiding the output to be representative of the input information. Overall, apart from the final output layer, batch normalization and PReLU were used after every convolutional layer along with the entire CNN for normalization, and 50% dropout was used at every layer for regularization to decrease overfitting. The final output layer of the CNN contained a $1 \times 1 \times 1$ convolution with a stride of 1 and a softmax activation function to predict for zeros (background) and ones (LA pixel).

The hyper-parameters in the CNN were selected through controlled experimentation to determine the optimal configuration for the task. The number of convolutional kernels was tuned using 4, 8, and 16 kernels for the first layer, with the remaining layers doubling as described. Experiments showed that using 4 kernels did not provide the network with sufficient depth to predict the LA accurately while 16 kernels were too computationally intensive with minimal improvement over 8 kernels. The number of steps in the encoder and decoder paths was also adjusted to find the degree of compression needed. Similar with the

number of kernels, CNNs without sufficient down-sampling steps were too shallow for the task, while the number of down-sampling steps above the optimal 4 steps did not contribute to an increase in accuracy. We implemented a CNN with $3 \times 3 \times 3$ kernels and compared the results with the $5 \times 5 \times 5$ kernels. Surprisingly, the network had difficulty converging when using size 3 kernels, potentially due to the lack of receptive field which could not effectively process the sparse inputs provided. We found PReLU activations worked more harmoniously with the network architecture compared with ReLU and leaky ReLU as it produced the best performances. The percentage of the dropout was also tuned with dropout rates of 25%, 50%, and 75%. While the performance did not significantly vary, a drop out of 50% provided sufficient regularization without reducing the training time as when applying 75% dropout.

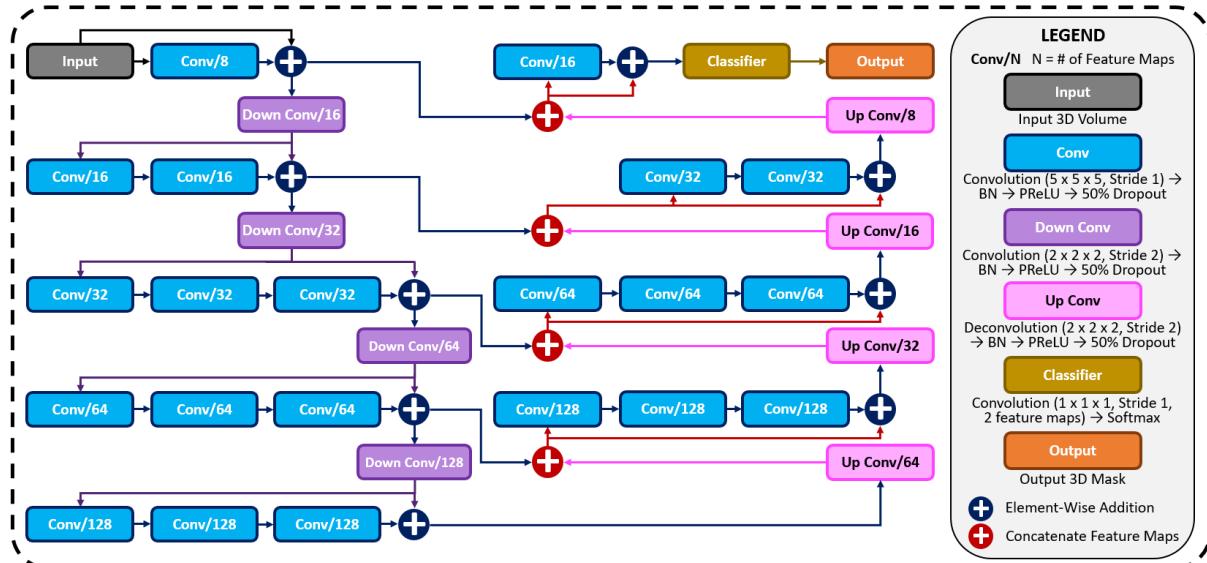


Figure 8.4. The architecture of the proposed 3D convolutional neural network (CNN) for predicting the left atrial (LA) geometry from a point cloud obtained during clinical mapping. The number of kernels in each convolutional layer is shown, along with the type of convolution. The flow of the gradients between layers is also shown, with different operations for merging two layers. The legend shows the exact operations of each layer labelled with different colours. All parameters can be found in **Table 8.2**. BN, batch normalization; conv, convolution; PReLU, parametric rectified linear unit.

Table 8.2. Configuration of the proposed CNN.

Encoder Path Layers	Kernel Size	Stride	Feature Maps	Number of Parameters
Input 3D	-	-	1	-
Conv/8	$5 \times 5 \times 5$	1	8	$5 \times 5 \times 5 \times 1 \times 8$
Down Conv/16	$2 \times 2 \times 2$	2	16	$2 \times 2 \times 2 \times 8 \times 16$
Conv/16	$5 \times 5 \times 5$	1	16	$5 \times 5 \times 5 \times 16 \times 16$
Conv/16	$5 \times 5 \times 5$	1	16	$5 \times 5 \times 5 \times 16 \times 16$
Down Conv/32	$2 \times 2 \times 2$	2	32	$2 \times 2 \times 2 \times 16 \times 32$
Conv/32	$5 \times 5 \times 5$	1	32	$5 \times 5 \times 5 \times 32 \times 32$
Conv/32	$5 \times 5 \times 5$	1	32	$5 \times 5 \times 5 \times 32 \times 32$
Conv/32	$5 \times 5 \times 5$	1	32	$5 \times 5 \times 5 \times 32 \times 32$
Down Conv/64	$2 \times 2 \times 2$	2	64	$2 \times 2 \times 2 \times 32 \times 64$
Conv/64	$5 \times 5 \times 5$	1	64	$5 \times 5 \times 5 \times 64 \times 64$
Conv/64	$5 \times 5 \times 5$	1	64	$5 \times 5 \times 5 \times 64 \times 64$
Conv/64	$5 \times 5 \times 5$	1	64	$5 \times 5 \times 5 \times 64 \times 64$
Down Conv/128	$2 \times 2 \times 2$	2	128	$2 \times 2 \times 2 \times 64 \times 128$
Conv/128	$5 \times 5 \times 5$	1	128	$5 \times 5 \times 5 \times 128 \times 128$
Conv/128	$5 \times 5 \times 5$	1	128	$5 \times 5 \times 5 \times 128 \times 128$
Conv/128	$5 \times 5 \times 5$	1	128	$5 \times 5 \times 5 \times 128 \times 128$

Decoder Path Layers	Kernel Size	Stride	Feature Maps	Number of Parameters
Up Conv/64	$2 \times 2 \times 2$	2	64	$2 \times 2 \times 2 \times 128 \times 64$
Conv/128	$5 \times 5 \times 5$	1	128	$5 \times 5 \times 5 \times 128 \times 128$
Conv/128	$5 \times 5 \times 5$	1	128	$5 \times 5 \times 5 \times 128 \times 128$
Conv/128	$5 \times 5 \times 5$	1	128	$5 \times 5 \times 5 \times 128 \times 128$
Up Conv/32	$2 \times 2 \times 2$	2	32	$2 \times 2 \times 2 \times 128 \times 32$
Conv/64	$5 \times 5 \times 5$	1	64	$5 \times 5 \times 5 \times 64 \times 64$
Conv/64	$5 \times 5 \times 5$	1	64	$5 \times 5 \times 5 \times 64 \times 64$
Conv/64	$5 \times 5 \times 5$	1	64	$5 \times 5 \times 5 \times 64 \times 64$
Up Conv/16	$2 \times 2 \times 2$	2	16	$2 \times 2 \times 2 \times 64 \times 16$
Conv/32	$5 \times 5 \times 5$	1	32	$5 \times 5 \times 5 \times 32 \times 32$
Conv/32	$5 \times 5 \times 5$	1	32	$5 \times 5 \times 5 \times 32 \times 32$
Up Conv/8	$2 \times 2 \times 2$	2	8	$2 \times 2 \times 2 \times 32 \times 8$
Conv/16	$5 \times 5 \times 5$	1	16	$5 \times 5 \times 5 \times 16 \times 16$
Classifier	$1 \times 1 \times 1$	1	2	$1 \times 1 \times 1 \times 16 \times 2$
Output 3D	-	-	2	-

Total Parameters	~32.5 Million
-------------------------	----------------------

Conv, convolution.

The Dice loss was used during training, previously defined in **Chapter 7.2.2.1**. The

Adam optimizer [208] was used to minimize the loss function during training with a constant learning rate of 0.0001 and the exponential decay rates of the 1st and 2nd moment estimates were set to 0.9 and 0.999, respectively. To reduce the computational burden of the large images that needed to be processed, all data was stored in the hierarchical data format after pre-processing. The CNN was trained with a maximum limit of 1000 epochs, with a criterion to stop training if the accuracy on the validation set did not improve after 50 epochs. A batch size of 1 was used due to the high memory costs associated with 3D volumes. The training set was also shuffled for each epoch to increase randomness. After every epoch, the performance of the CNN was evaluated on the validation set with the Dice score. The parameter set of the CNN which achieved the highest validation accuracy was saved and used on the testing set. The training step was performed on an Nvidia Titan V GPU with 5120 CUDA cores and 12 GB RAM. The training phase took approximately ten hours. Predictions took approximately 10 seconds for each partial shell input.

8.2.4 Evaluation

The Dice score, STSD, sensitivity, specificity, LA diameter, and volume errors were used for evaluation. These metrics were defined in **Chapter 3.6.1** and **Chapter 3.6.2**. We also evaluated the coverage of the point cloud in the LA to measure its impact on the evaluation metrics. This was computed by

$$Coverage = \frac{\sum_i^n PT_i}{\sum_j^m LA_{surface}} \quad (\text{Equation 8.10})$$

given the point cloud, PT , with a length of n , and the outer surface of the LA, $LA_{surface}$, with m pixels, and $n < m$ for all points.

8.3 Results

8.3.1 Accuracy for Predicting the Left Atrium from Point Clouds

Table 8.3 and **Table 8.4** show the complete evaluation results for the 3D LA reconstruction from point clouds in generated paired training dataset, and clinical test #1 and

test #2 datasets. Overall, the proposed CNN achieved excellent accuracies for LA prediction, with Dice scores of 93.2% for the paired training set, 92.4% for the *test #1* set, and 93.4% for the *test #2* set. These high accuracies showed that the CNN was able to successfully reconstruct the LA from the sparse inputs provided. The relatively low standard deviation of 2.3% on the 300 testing samples in the paired training set showed that the predictions were also very consistent. This was particularly seen in the two test sets with standard deviations of below 1% for the Dice score. The CNN achieved an STSD of 1.1 pixels on the paired training set, and a more impressive 0.8 and 0.7 pixels on the *test #1* and *test #2* sets, showing the predicted LA was on average within 1 pixel of the ground truth. The high sensitivity of above 90% and the specificities of 99% showed that the CNN was able to distinguish between the positive and negative pixels with high certainties. Surprisingly, the approximately 4% higher sensitivity on the two clinical test sets indicated the CNN was able to capture the LA pixels much more effectively than in the paired training set.

Table 8.3. Technical evaluation for left atrium (LA) reconstruction from point clouds in the 300 generated (*Paired Training*), 4 clinical MRI (*Test #1*), and 2 clinical CT (*Test #2*) data.

Dataset	Dice	STSD	Sensitivity	Specificity
Paired training data	$93.2 \pm 2.3\%$	1.16 ± 0.48 px	$90.6 \pm 3.7\%$	$99.7 \pm 0.1\%$
Test #1	$92.4 \pm 0.8\%$	0.76 ± 0.05 px	$94.9 \pm 0.6\%$	$99.2 \pm 0.2\%$
Test #2	$93.4 \pm 0.6\%$	0.66 ± 0.05 px	$95.0 \pm 0.3\%$	$99.2 \pm 0.1\%$

CT, computed tomography; MRI, magnetic resonance imaging.

The predicted LA were also biologically accurate on average, obtaining low diameter and volume errors of 4.4% and 5.9%, respectively (**Table 8.4**). The higher sensitivities in the two test sets also resulted in lower diameter and volume errors with 2.6% and 3.0% errors for the diameter, and 5.2% and 3.3% for the volume in the *test #1* and *test #2* sets, respectively. We also compared the biological measurements between the ground truth and predicted LA to determine the error source. We found that the mean predicted diameter of 39.9 mm and volume of 49.0 cm^3 were lower when compared to the 41.5 mm and 52.4 cm^3 ground truth measurements. This revealed the CNN had a tendency to slightly underestimate the LA when analysing point clouds.

Table 8.4. Biological evaluation for left atrium (LA) reconstruction from point clouds in the 300 generated (*Paired Training*), 4 clinical MRI (*Test #1*), and 2 clinical CT (*Test #2*) data.

Dataset	Diameter Error	Volume Error
Paired training data	$4.4 \pm 5.2\%$	$5.9 \pm 4.1\%$
Test #1	$2.6 \pm 1.2\%$	$5.2 \pm 1.0\%$
Test #2	$3.0 \pm 1.0\%$	$3.3 \pm 1.9\%$

CT, computed tomography; MRI, magnetic resonance imaging.

8.3.2 Visualization and Error Analysis

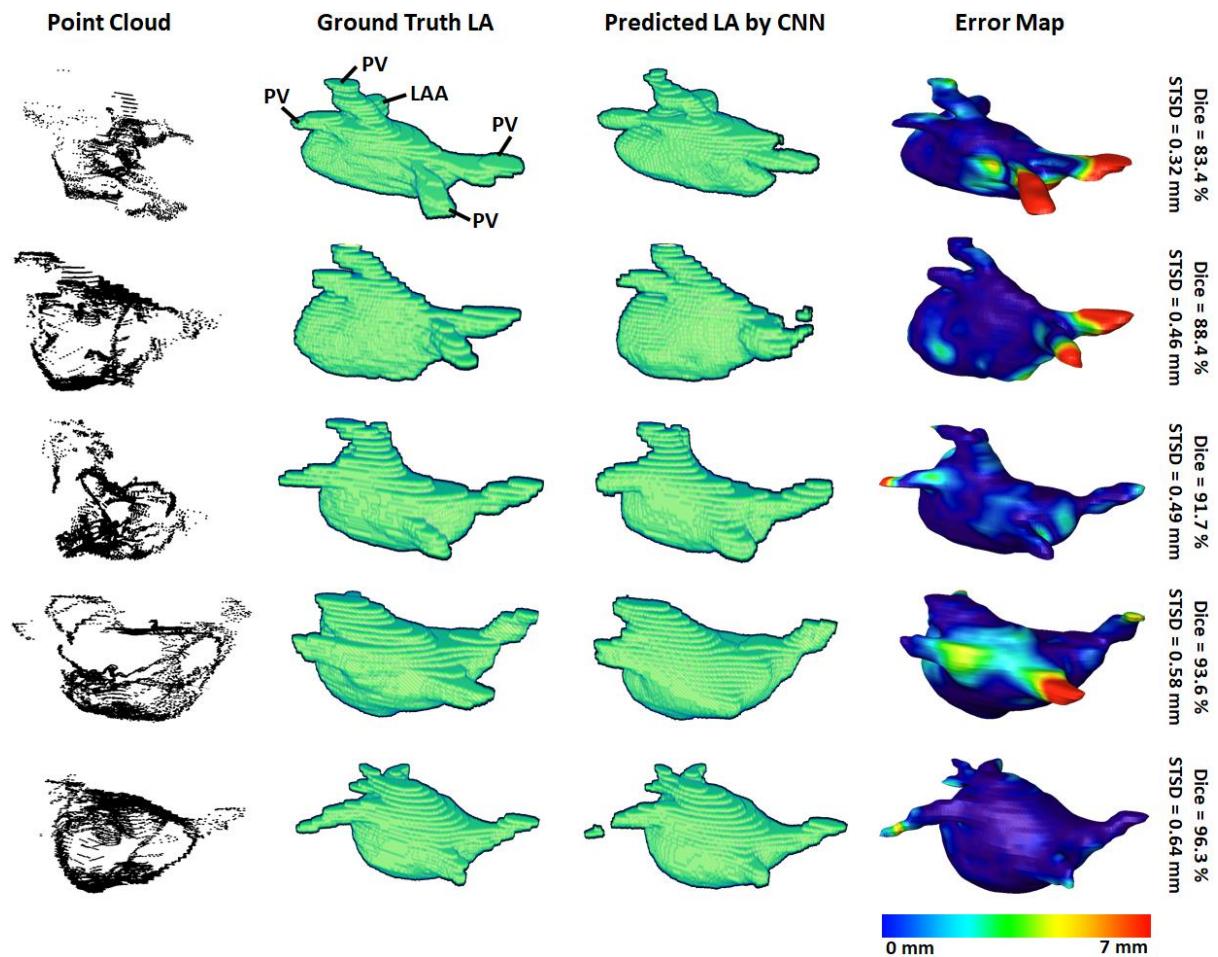


Figure 8.5. 3D visualizations of the left atrial (LA) reconstructions of five samples in the *paired training* dataset. The reconstructions with the highest Dice scores are in the bottom row and the top row contains the reconstructions with the lowest Dice scores. The point-clouds inputs are shown in the first column. The ground truths obtained by manually segmenting the LGE-MRIs are shown in the second column. The reconstructions predicted by the convolutional neural network (CNN) are shown in the third column. The surface-to-

surface distance (STSD) error maps between the ground truths and the predictions are shown in the fourth column, with the colours being normalized between 0 mm and 7 mm for the five samples. LAA, left atrial appendage; PV, pulmonary vein.

3D visualizations of the ground truth and predictions produced by the CNN were produced for further error analysis. **Figure 8.5** shows five samples of predictions made by the CNN in order of increasing accuracy, representing the range of accuracies obtained in the paired training set. The input point cloud was also shown with the corresponding ground truth LA geometry. From the samples shown, it was observed that the degree of coverage depicted by the input data had a significant impact on the accuracy of prediction. This was clearly visible in the first row where the input point cloud had low coverage. The CNN was therefore forced to generate many anatomical features without guidance, based only on the shape of the existing input. The fifth row showed an input containing extremely good coverages, naturally making the prediction much more accurate. However, rows one to four also revealed the power of the CNN for data generation, as the outputs, regardless of Dice score, were all anatomically similar to the ground truths. This also showed that the CNN would be effective on clinically recorded point clouds which do not fully cover the entire LA surface. Expectedly, the most erroneous regions were the PVs when a distance-error map was computed between the predictions and ground truths. This was due to the PVs having a thin and inconsistent shape compared to the rest of the LA, creating difficulties for the CNN to consistently define.

To demonstrate our method is adaptable and feasible on the two real clinical datasets (*test #1* and *test #2*). We displayed the prediction and ground truth of one sample from each dataset in **Figure 8.6**. In general, it can be seen the point cloud in these datasets covered a significantly larger proportion of the LA compared to the paired training dataset. This led to the CNN performing better given the more complete LA shells which were generated from the point clouds. Furthermore, the adaptability of our CNN can be seen in the results for the *test #2* data. The LA was acquired from CTs, as opposed to MRIs which were used in both the paired training and *test #1* datasets, leading to a significantly different geometry. Nevertheless, our CNN effectively predicted the CT geometry although it was only trained on MRIs geometries, showing our approach was independent of the mapping system and image modality.

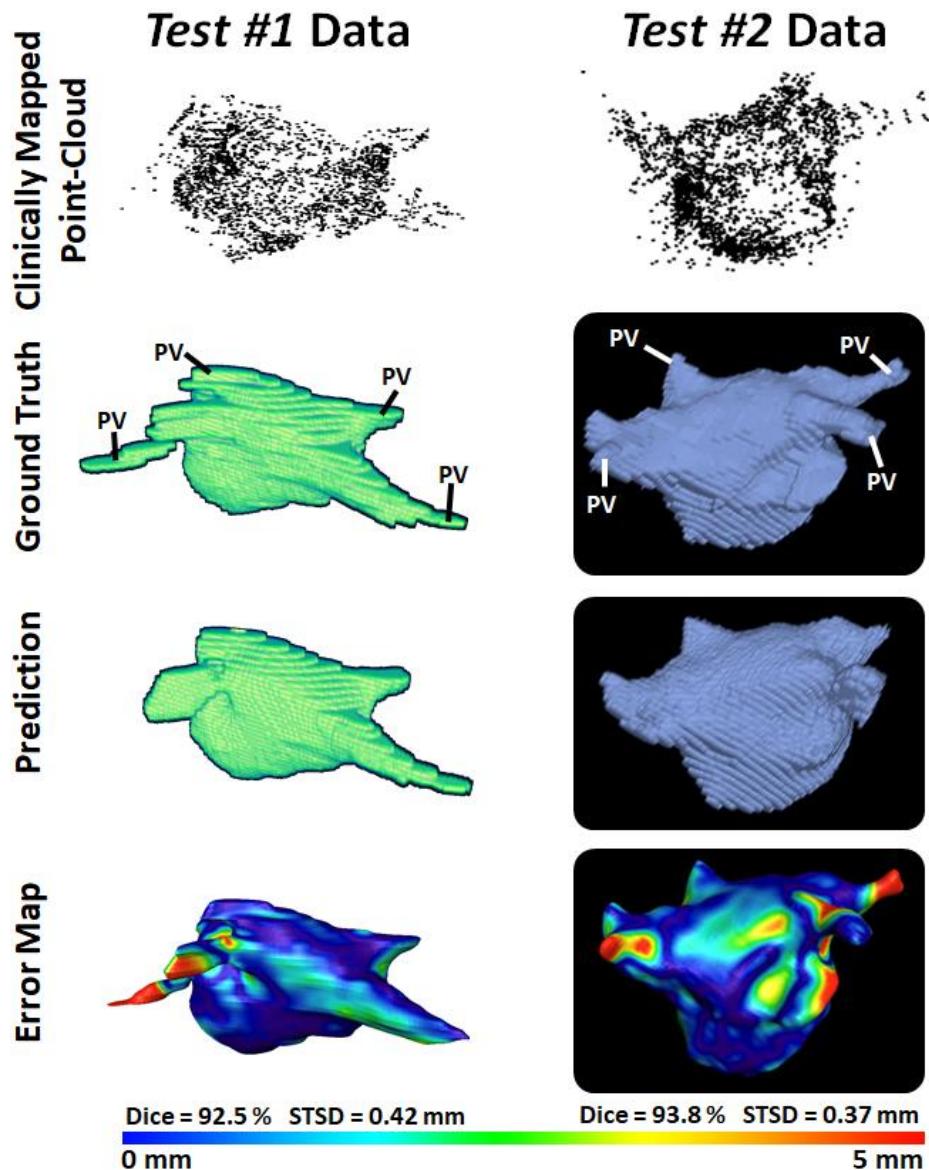


Figure 8.6. 3D Visualizations of the left atrial (LA) reconstruction for one sample each from *test #1* (left column) and *test #2* (right column) clinical datasets. The point cloud recorded with the commercial mapping systems are shown in the first row, along with the LA geometry obtained from segmenting magnetic resonance imaging (MRI) and computed tomography (CT) in the second row. The predicted LA are shown in the third row, and the surface-to-surface distance (STSD) error maps between the ground truths and the predictions are shown in the fourth row. The individual Dice and STSD scores are shown for each sample. PV, pulmonary vein.

8.3.3 Impact of Point Cloud Coverage on Accuracy

We analysed the impact of the coverage of the point cloud over the target output LA

on the evaluation scores obtained in our results (**Table 8.5**). The average coverage across the paired training dataset was 30%, while the *test #1* and *test #2* sets had coverages of 40% and 44%, respectively. The standard deviation of the coverage on the paired training set was 5.4% and contained a range of 19% to 40%. This indicated there was a wide range of point cloud coverages for the CNN during training, allowing it to be applicable to a range of distributions during prediction. Interestingly, the mean coverages of the two test sets were above and outside the range of the paired training set, showing the point clouds acquired in the clinical sets were of higher quality. This was also visible in **Figure 8.5** versus **Figure 8.6** which showed a smoother point cloud distribution for the *test #1* and *test #2* data.

Table 8.5. The point cloud coverage over the left atrium (LA) for the generated (*Paired Training*), clinical MRI (*Test #1*), and clinical CT (*Test #2*) datasets.

Dataset	Coverage
<i>Paired training data</i>	$30.3 \pm 5.4\%$
<i>Test #1</i>	$39.5 \pm 1.5\%$
<i>Test #2</i>	$44.4 \pm 2.6\%$

CT, computed tomography; MRI, magnetic resonance imaging.

We then computed the Pearson's correlation between the point cloud coverage in all data and the accuracies obtained by our CNN (**Figure 8.7-8.8**). Overall, the coverage was significantly and strongly correlated to both the Dice score and sensitivity, with correlations of 0.7. This was a potential explanation for the increased sensitivity on the two clinical sets, as the higher coverage allowed the CNN to predict the entire LA geometry with greater precision. The coverage was also moderately correlated to the STSD with a value of 0.6 and statistical significance. Expectedly, the errors for the diameter and volume were both negatively correlated with the coverage, as higher coverages resulted in better predictions of the biological measurements, and thus lower errors. While the diameter error had a low correlation of -0.1, the volume error had a moderate negative correlation of -0.6. This was due to the diameter only being measured in one dimension, and thus being impacted less by the overall LA reconstruction accuracy, while the volume was influenced by all three dimensions.

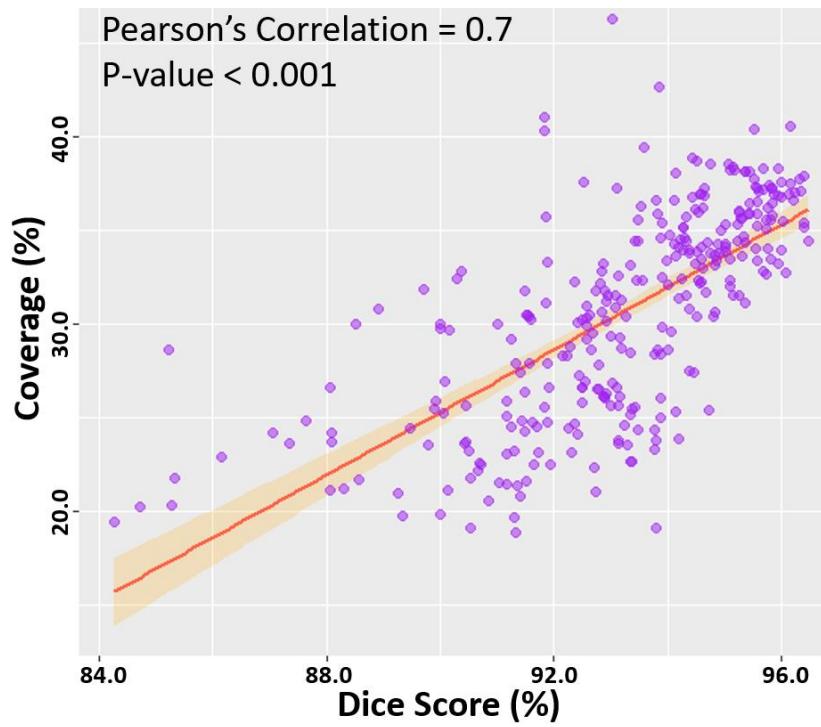


Figure 8.7. Correlation of the point cloud coverage with the Dice score in the testing datasets. The line of best fit is shown, along with the Pearson's correlation value and p-value. The band shows the standard deviation of the points along the line.

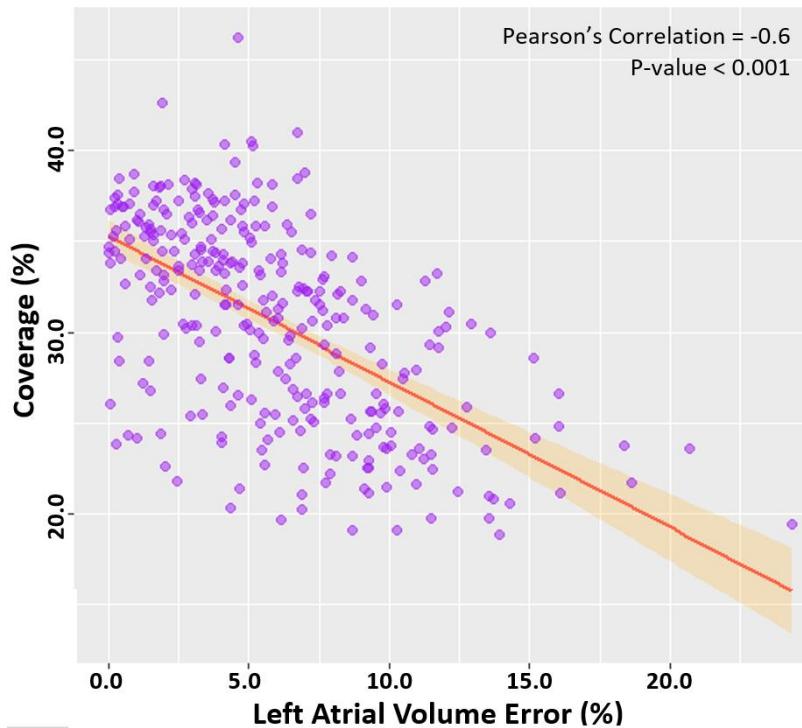


Figure 8.8. Correlation of the point cloud coverage with the left atrium (LA) volume error in the testing datasets. The line of best fit is shown, along with the Pearson's correlation value and p-value. The band shows the standard deviation of the points along the line.

8.4 Discussion

8.4.1 3D Reconstruction of Anatomical Structures Directly from Point Clouds

Direct surface reconstruction of anatomical structures, such as the LA, from point clouds is a challenging task. Prevailing methods of analysis primarily focus on the application of CNNs for the classification and segmentation of point cloud representations of 3D objects or scenery. Well-established research into the reconstruction of 3D surface geometries directly from sparse inputs such as point cloud is therefore limited. Furthermore, the current commercial software used to perform clinical mapping and the subsequent LA reconstruction from the point clouds recorded is inefficient by requiring additional imaging prior to the procedure. The efficacy of the proprietary software also remains difficult to validate, and open research in the area is lacking.

To address the current issues, our study is one of the first to propose a fully automated method for the reconstruction of the LA geometry directly from point clouds. Our study is also one of few to develop a CNN for the surface reconstruction of 3D geometries given a set of partially complete information such as the sparse point clouds data described. Overall, the proposed CNN produced LA predictions with high-performance accuracies across multiple metrics for both technical and biological evaluation. The CNN obtained Dice scores surpassing a prior study which investigated a similar task [298]. The low surface-to-surface distance, LA diameter, and LA volume errors showed our approach produced anatomically accurate predictions, which is a highly important feature for clinical applications. The clinical applicability of our approach was further demonstrated on the two clinical point cloud datasets acquired with the most commonly used CARTO and EnSiteNavX mapping systems. Experimental results showed the CNN achieved similarly accurate and consistent predictions when compared to LA geometries segmented from the MRIs and CTs in the clinical datasets.

8.4.2 Pre-Processing and Network Tuning

An important component of CNN pipelines for point cloud analysis involves the pre-processing of the point clouds data into fixed-sized inputs. Similar to prior studies, we retained the original dimensionality and important spatial information of the inputs by

directly discretizing the 3D point cloud into an image volume [305]. However, the pre-processing step in our study was significantly enhanced by the proposed iterative concave-hull algorithm, which exponentially increased the number of data points with low computational costs. The increased number of points resulted in smooth image volumes after discretization. This was an improvement on past methods which attempted to directly discretize low-density point clouds to produce sparse images which were difficult and computationally expensive to analyse by the CNN. As the pre-processed volumes contained a high density of information, this also benefited the performance of the CNN by providing concentrated data with a relatively low memory cost, leading to more precise predictions with greater efficiency. A further step for ensuring effective feature learning on the pre-processed point clouds involved the utilization of larger convolutional kernels to increase the receptive field of the CNN. The CNN was also enhanced with the use of feature forwarding connections, allowing it to retain and combine features at multiple receptive levels, maximizing the information extracted from the relatively sparse input information provided. Due to the high class-imbalance of the point clouds which often induces CNNs to produce completely empty predictions, we implemented a Dice loss to prioritize non-background pixels. Residual blocks and batch normalization were also included to increase the ease of convergence.

8.4.3 Limitations and Future Work

Our study contains several limitations, which can potentially be addressed in future studies. Experiments on our CNN showed that although it performed well experimentally, its accuracy was directly dependent on the coverage of the point cloud. While most clinical point cloud recordings nowadays maintain good coverage over the entire LA chamber as seen in the samples in this study, future methods should specifically be aimed to address low coverage maps. Such methods could involve statistical shape models which artificially enhance the coverage by using aggregated anatomical features from past data to estimate the location of potential landmarks. Future research should also investigate changes to the CNN architecture to improve its accuracy in general, such as introducing adversarial pathways or auxiliary outputs which are commonly used for image reconstruction. The loss function could be improved by introducing anatomical constraints to ensure the outputs contain all key anatomical landmarks which would be very beneficial in clinical applications. Finally, future

studies should ideally utilize larger samples of clinical data through more extensive collaborations with international clinical centres to further validate the robustness of the method.

8.5 Conclusions

In this chapter, we have developed and evaluated a 3D CNN for robust automatic LA reconstruction from point clouds recorded with clinical mapping during ablation. Our algorithm enables the reconstruction of the LA in 3D with a Dice accuracy of 93%, STSD of approximately 1 pixel, and accurate estimations of clinical measures. The method was further tested on two independent cross-modality clinical datasets, and produced similarly impressive evaluation results. Our study may lead to the development of a more accurate and efficient real-time LA reconstruction approach using a small set of point clouds, which can potentially be used to improve clinical guidance during ablation procedures for the treatment of AF.

Chapter 9

Conclusions

This thesis aimed to facilitate the early detection and effective treatment of AF to minimise the impact of AF and increase the quality of life for patients suffering from AF. The three main goals of the thesis were met through the development of deep learning algorithms to improve AF detection, prognosis, and treatment. Firstly, a robust method of AF detection was developed using ECGs recorded from portable sensors on the largest open-source single-lead ECG dataset. Secondly, a fully automatic method of atrial structural analysis was developed on the world's largest MRI datasets to improve visualisation and assessment of the patient-specific atrial anatomy during diagnosis and stratification. Finally, an efficient method of facilitating the reconstruction of the 3D atrial surface geometry was developed to guide the targeted ablation for AF treatment.

This chapter summarises the key contributions of this thesis and improvements over current research in the field. The potential directions for future research are then discussed, and the thesis ends with a final concluding remark.

9.1 Summary and Contributions

The three main areas of focus in this thesis revolved around the analysis of ECGs, MRIs, and clinical mapping using novel deep learning methodologies to tackle the current clinical shortcomings and challenges of AF.

The development of a CRN for the automatic classification of ECGs recorded with wearable sensors to improve the early detection of AF was presented in **Chapter 4**. The CRN was able to accurately classify ECGs into either SR or AF signals, effectively detecting AF with higher accuracies than studies conducted using commercial devices, and achieving a rate of false AF detection significantly lower than current methods. The CRN was designed to process ECGs of varying input lengths by combining the feature learning capabilities of CNNs and the recursive capacity of RNN, drastically increasing its versatility and transferability to different signals. This was a major improvement over previous methods

which were only able to analyse fixed-sized ECG segments and could not aggregate the individual segment-wise predictions into a single prediction for the entire signal. The added advantage of directly processing the entire length of each ECG signal in a single pass was the result of more effective feature learning as well as an increased ability to detect short AF episodes within a long recording. For the first time, an ECG generator was also proposed to simulate realistic instances of SR and AF ECGs in a controlled manner through the combination of mathematical modelling and an NST network. This addressed the issue with the current lack of AF signals due to its low prevalence which resulted in severe class imbalances and caused previous neural networks to be statistically biased towards SR. The realism of the ECGs synthesis from the novel generator was extensively validated through frequency and wave decomposition analysis. The CRN, trained using the addition of the synthetic data, resulted in a significantly increased accuracy for AF classification.

The development of an automatic CNN framework, AtriaNet, for atrial structural analysis directly from medical imaging to improve the identification of the potential ablation target sites to terminate AF was presented in **Chapters 5, 6, and 7**. AtriaNet was the world's first method capable of LA and RA segmentation, fibrosis quantification, and AWT estimation to enable the automatic computation of multiple crucial biomarkers. The biomarkers consisted of the bi-atrial chamber diameter, atrial volume, anatomic structure, 3D distribution and percentage of fibrosis, and 3D AWT mean and distribution. The extensive multi-stage development process for AtriaNet resulted in a simple, yet effective, unified framework containing two sequentially arranged CNNs for detecting and subsequent phenotyping of the atria. The multi-application capability of AtriaNet was a novel development which enabled an increased training efficiency across the three tasks performed, improving upon prior methods which only performed a single task. AtriaNet was developed on the world's largest cardiac LGE-MRI and cine-MRI datasets, achieving the most robust validation of any method as prior studies only focused on single in-house datasets. The efficacy of AtriaNet was further strengthened by testing it on an independent clinical LGE-MRI dataset to validate its accuracy in a clinical setting. AtriaNet achieved state-of-the-art accuracies on the LA, was the first automated approach for RA analysis, and achieved similarly impressive accuracies for RA analysis. In terms of the clinical accuracy of AtriaNet, the atrial measurements predicted closely correlated with manual measures from echocardiography scans and the 3D fibrosis distribution maps predicted matched low-voltage maps acquired from electro-anatomical mapping during ablation. AtriaNet also significantly improved upon the efficiency of execution and computational speed of traditional methods

for AWT calculation. Overall, the novel development of AtriaNet presented a single framework capable of performing a range of analytical tasks, traditionally conducted manually through multiple different clinical methodologies, with a high degree of confidence and efficiency using a single stand-alone imaging sample for each patient.

The development of a CNN for automatic LA surface reconstruction from clinical mapping to more efficiently guide the targeted ablation during the catheter ablation procedure for AF elimination was presented in **Chapter 8**. A novel method of data generation was developed through the merging of clinical mapping and LA geometries extracted from LGE-MRIs to produce a large and realistic dataset of matching point cloud and LA volumes for training the CNN. The CNN then achieved excellent reconstruction accuracies on the synthetic dataset, and was further validated on clinical LGE-MRI and CT datasets, showing much improved results compared to results from the current commercial software. Importantly, the method was the first to reconstruct the LA directly from the point cloud, as current methods rely on additional mapping and/or imaging, and manual registration to achieve accurate results during ablation. The method was also shown to be capable of predicting anatomical structures of the LA in regions with poor or no mapping coverage. This allowed the generation of the LA surface with a smaller number of points as opposed to previous mapping methods which require high coverage, significantly increasing the efficiency of the mapping procedure. Furthermore, the methodologies developed were an important contribution to the field of organ reconstruction from point cloud data in general, as there was a significant lack of well-established research in this domain for any anatomical structure.

9.2 Limitations and Future Work

9.2.1 Data for Training Deep Learning Models

The most important area for improvement involves the quality, diversity, and quantity of the datasets used to develop deep learning algorithms. While the datasets used throughout this thesis were the largest open-source datasets in each of the respective fields, larger datasets from multiple international sources would enable further improvement in the predictive capabilities of neural networks. Since its popularisation in 2012 from the AlexNet paper, the development of deep learning has primarily been driven by the rapidly increasing

amounts of data. Datasets such as ImageNet, containing over 1 million individual samples from 1,000 different labelled classes, have been the first choice of the dataset used to develop and improve newer neural networks. In comparison, the biomedical field contains significantly less data samples, bottlenecking the accuracy and capabilities of neural networks for performing increasingly difficult tasks. Although recent large-scale studies such as the UK Biobank have collected hundreds of thousands of patient samples, most of the data remained unlabelled due to the extreme time-consuming nature and expertise required for manual annotation. Future efforts should be put towards developing more efficient methods of manually annotating complex datasets such as medical imaging with semi-automatic or automatic methods to drastically accelerate the amount of labelled data available for training neural networks. Neural networks with the capacity to learn from unlabelled or partially labelled datasets, i.e., unsupervised or semi-supervised learning, should also be investigated in the future to leverage the massive amounts of unlabelled data currently available. Another interesting area for investigation would be the use of generative adversarial networks to automatically synthesise high quality labelled datasets without the need for acquisition or manual labelling. Such networks have previously been successful for image generation. However, their application to the field of medical imaging has been relatively unexplored at a larger scale. Developments in these areas would enable the rapid development of deep learning across many biomedical fields by accessing the information contained in more data and at a faster pace.

Future studies would also involve increasing the collaboration between different clinical centres to enable the collection of cross-modality and cross-centre datasets. This would particularly be valuable for the case for atrial LGE-MRIs datasets which currently remain small in comparison to other imaging modalities and imaging of other organ structures. Such efforts are currently underway at Waikato Hospital with the ongoing AF trial, as the LGE-MRIs used in this thesis were acquired in the phase of the trial already conducted. Ideally, the LGE-MRI datasets in the future will allow for the investigation of other contributing factors such as gender or ethnicity differences across several clinical centres, creating the opportunity to design neural networks with equity in mind.

9.2.2 Neural Network Architectures

Another potential area for future work involves improving the design of neural

network architectures. Neural network architectures of higher versatility and accuracy are consistently being released from world-leading research teams. With the steadily increasing amounts of biomedical data, more cutting-edge algorithms could be explored as current methods still revolve around less recent developments such as U-Net. As an example, the AtriaNet CNN pipeline proposed can be potentially improved with the integration of Mask-R-CNN which specifically targets small structures within large frames of view. Such an algorithm would potentially be capable of detecting subtle atrial structures in finer detail. The design of Mask-R-CNN would also allow the capacity for predicting complementary biomarker estimations directly in tandem with segmentation. This can thus create a fully unified framework which can be optimised specifically depending on the diagnostic measure required. The bounding-box approach associated with instance segmentation approaches such as Mask-R-CNN may also improve atrial measurement estimation. The method may also be more easily transferable to cross-modality, cross-centre, multi-resolution, and multi-view imaging datasets given sufficient training samples. Similar improvements to the current CNN architectures could involve replacing the residual blocks with the squeeze and excitation blocks which have been shown to surpass residual networks on the ImageNet dataset. Such enhancements may potentially improve the feature extraction and learning capabilities of CNNs which is extremely important as diagnostic information contained within data such as MRIs or ECGs is often impossible to detect through manual assessment.

9.2.3 Future Work in ECG and Point Cloud Analysis

As image analysis has been the primary driver of artificial intelligence in the past decade due to its diverse applications across almost every sector, tasks such as single-channel signal or point cloud analysis explored in this thesis have received significantly less attention. In this thesis, the development of CNNs for MRI analysis was heavily supported by the existing research into CNNs for general images. However, research into sequential data such as ECGs or spatial data such as point clouds are significantly less in comparison, leading to the approaches developed in this thesis for such datasets being at the front-line in the respective areas. The analysis of ECGs is an extremely difficult task as the events of interest, such as AF, are often rare compared to the remainder of the signal which mostly consists of normal electrical activity. Long ECG recordings with varying lengths further complicate this matter as the extensive research already conducted for CNNs would not be transferrable into

this domain. Similarly, point clouds also consist of dynamic input lengths depending on the mapping acquired, resulting in CNNs being less applicable to this task. Future studies would ideally bring more attention to these non-conventional data types through organising international challenges, similar to the LA segmentation challenge we organised to improve our knowledge of medical image analysis. This would bring together research communities from across the world to collectively tackle and solve these issues as we have already demonstrated for atrial segmentation. An example of a potential improvement would be to utilise CNNs to specifically handle dynamic input ranges to lessen the computational burden associated with pre-processing the point clouds for LA reconstruction. The more direct method of training on the point cloud data will eliminate the need for the current mesh generation algorithm. It would also allow the network to operate end-to-end which is more desirable when scaling to other datasets. A potential solution would consist of leveraging RNNs which allow for inputs of varying sizes to act as the initial data input-processor which then feeds into a CNN for further feature extraction.

9.3 Concluding Remarks

This thesis presented the design and development of deep learning algorithms to aid the treatment of AF which currently places a severe burden on the global health system. The algorithms were developed on ECG, MRIs, and clinical mapping data to improve the current clinical workflow of AF along the stages of detection, prognosis and treatment. The algorithms proposed were extensively validated on the world's largest biomedical datasets in both simulated and clinical settings. The novel methods described in this thesis also provided a solid benchmark for future research into the application of deep learning for AF and other CVDs. This body of work undoubtedly improved on the current state of AF research, potentially leading to the development of more optimised patient-specific strategies and ultimately improve AF patient outcomes in the long term.

Appendix Text

This section presents additional text information regarding the 2018 Atrial Segmentation Challenge as described in **Chapter 6**.

Challenge Details

The 2018 Atrial Segmentation Challenge was launched on 4th April 2018, with the training data released to the public on the same day containing all 100 3D LGE-MRI scans and their corresponding expert labels. The 54 3D test dataset was released two weeks prior to the end of the challenge on 14th September 2018. Each team had a maximum of 5 submission attempts and measures were taken to ensure participants did not exceed this limit. Participant ranking was performed during the last two weeks of the challenge with a leader board which was updated every 8 hours and the participants being ranked only by the Dice score. Thus, the winner of the challenge was the team with the highest overall Dice score for their predictions across all testing data at the end of the challenge.

The challenge was held in conjunction with the 21st International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) on 16th September 2018 in Granada, Spain. Throughout the course of the challenge, 200+ groups from 11 countries accessed the dataset, and 27 teams submitted meaningful results for the predictions on the test set, 17 of which had outlined their methods in papers submitted to the MICCAI workshop or online documents.

Details of Non-Conference Publication Methods Submitted

The full implementation details of 15 teams as listed in **Appendix Table 1** can be found in the 2018 Statistical Atlases and Computational Modelling of the Heart (STACOM) proceedings and was referenced in order of the rankings obtained by each team in the main text. The two closed-source submissions from two teams, Huang [260] (rank #2) and Xu et al. (Xu et al. 2018) (rank #9), are outlined below.

Huang proposed a method consisting of a double 3D CNN approach (**Appendix Figure 3**). In the first stage of prediction, the first CNN was used as a detector, where the

input was the down-sampled volume (with a ratio of $2 \times 2 \times 2$) and the output was a segmentation of the same size. The centre of this mask was extracted by calculating the centre of gravity and multiplied by two in each dimension to get the centre of the LA coordinates in the original resolution. A volume of $320 \times 240 \times 80$ was then extracted around this centre. In the second stage of prediction, the second CNN segmented the input which was a cropped volume of $320 \times 240 \times 80$ and the output was a binary mask of the same size. Overall, the backbone of the segmentation network was a multi-grid network based on a U-Net architecture (Ke et al. 2017). The multi-grid structure was similar to the atrous spatial pyramid pooling but was composed of different strides and dilations such that a multi-scale version of the network was propagated along each convolution. Also, the dense connections were made between the six consecutive convolutions in the coarsest scale of the U-Net architecture, with a growth rate of eight. Convolutions were added to the residual connection of U-Net and global average pooling was also used as an attention mechanism on different channels [109]. The network had channels of 2, 4, 8, 16, 24, 32, 40, 48, 12, and 6 in each block, with each block containing two convolutions with a residual network connection. The cross-entropy minus the logarithm of the Dice score was used as the loss function. Optimization was done using ADAM with an initial learning rate of 0.001 which dropped to 0.0005 after 100 epochs, 0.0002 after 200 epochs, and 0.0001 after 300 epochs. The 100 training samples were split for five-fold cross-validation. During this training stage, the training Dice scores were 0.926, 0.928, 0.923, 0.924 and 0.924 in the five folds respectively, with a mean of 0.925.

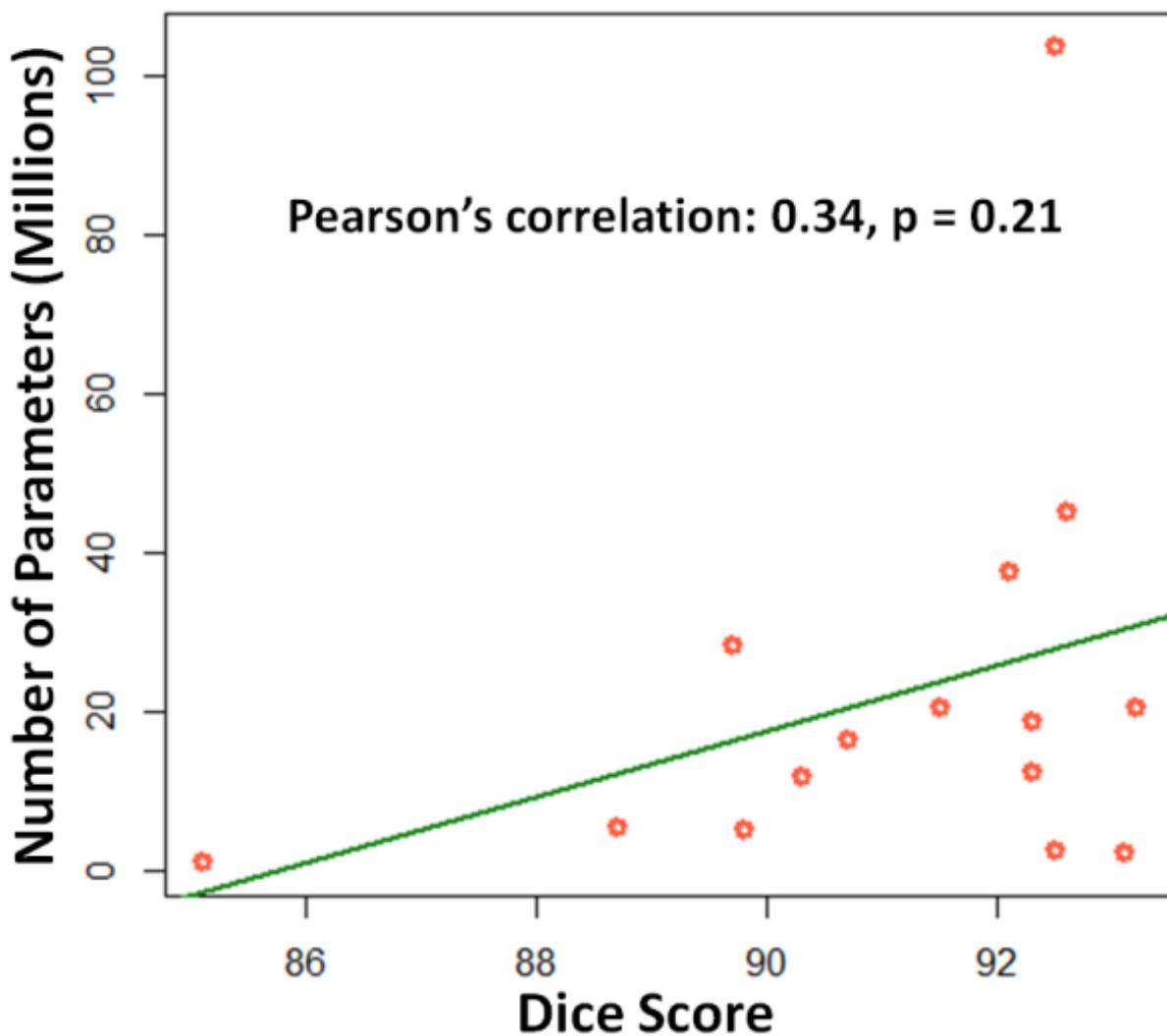
Xu proposed a multi-view LA segmentation framework that ensembles several CNN architectures for segmentation (**Appendix Figure 6**). The proposed processing pipeline consisted of multi-scale images as input data for a set of CNNs. The segmentation results of the multiple CNNs were then ensemble and fused to generate a final output. In the implementation, three scales were used with resolutions from $1 \times 1 \times 1$ mm to $3 \times 3 \times 3$ mm. Since there were only 100 samples, augmentation to the training set was performed on 80 samples which were randomly selected for training. The remaining 20 samples were used as the validation set. All of the CNNs were variants based on the U-Net but with a cost function consisting of both the original cross-entropy and the Dice score. The variants included the 2D U-Net to learn the 2D representation of 3D images from different views, the extended 3D U-Net [166], and the anatomical constrained U-Net [312] which was comprised of 3 cost terms, i.e. the cross-entropy, the Dice, and a term for the anatomical constraint. Finally, an ensemble and fusion strategy was proposed to combine all the multiple results from the CNNs.

Data and code availability

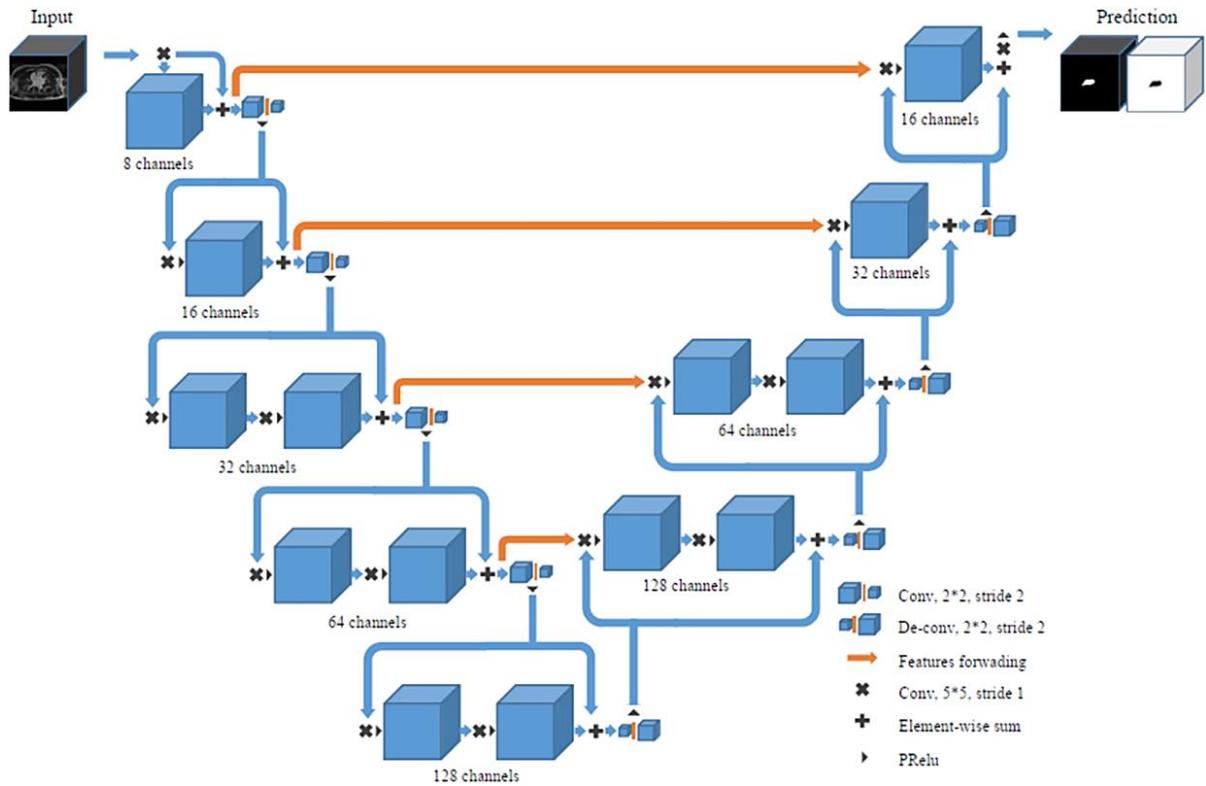
The full dataset containing 154 LGE-MRIs and labels used in this challenge are available on the 2018 Atrial Segmentation Challenge Website: <http://atriaseg2018.cardiacatlas.org/>. The final predictions for the 54 testing data from all teams are also available on the website. The codes used for performing all analyses in this study are also available for download.

Appendix Figures

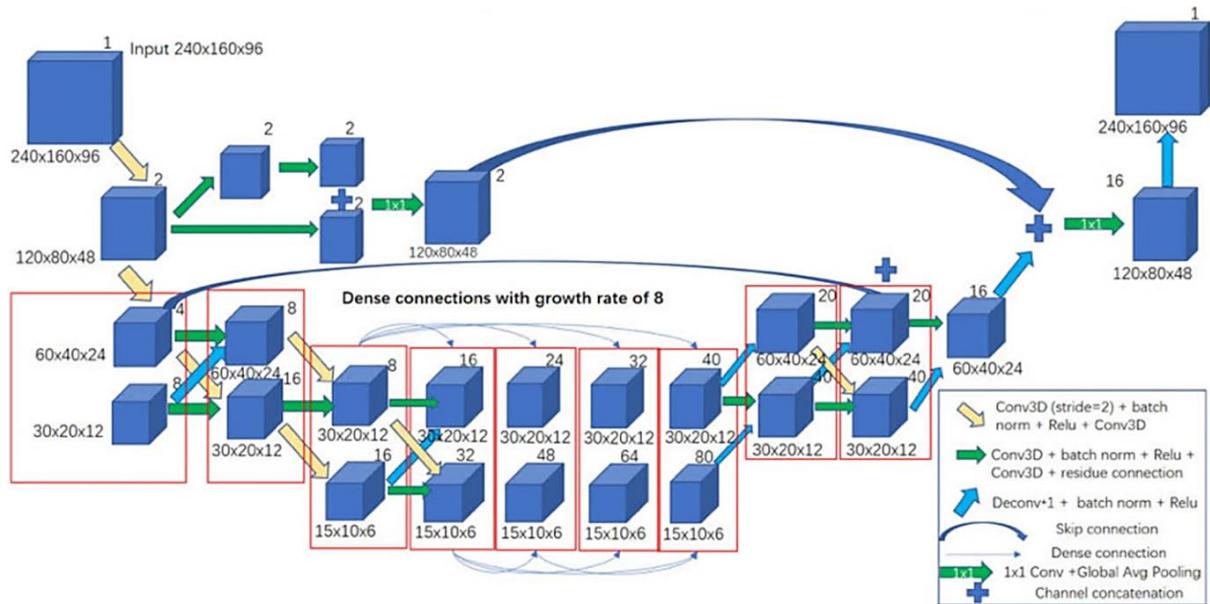
This section presents additional figures regarding the 2018 Atrial Segmentation Challenge as described in **Chapter 6**.



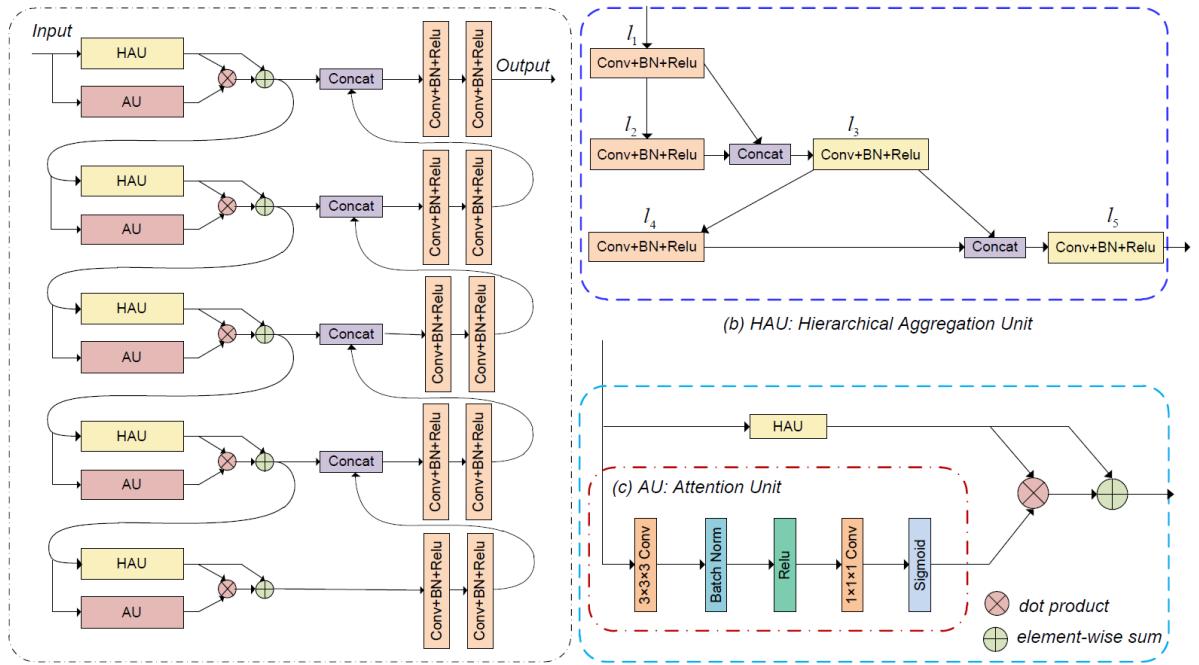
Appendix Figure 1. Comparison of the Dice score performance with the total number of network parameters used in each algorithm. The line of best fit is shown in green.



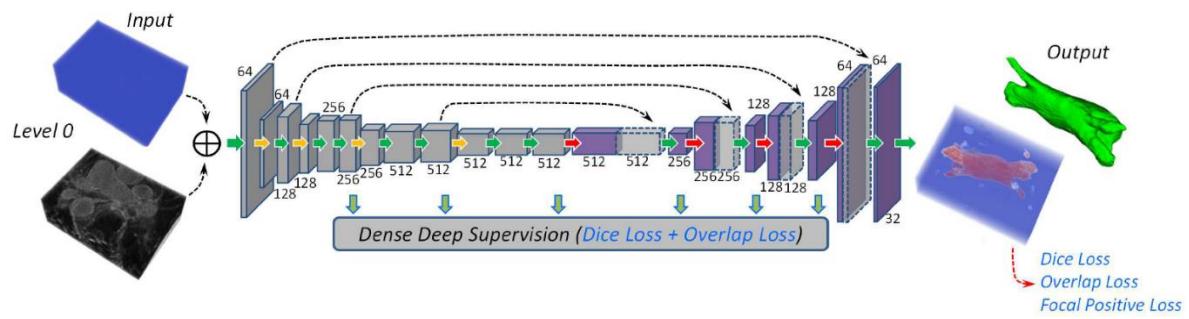
Appendix Figure 2. Convolutional neural network (CNN) architecture proposed by Xia et al. [168] who achieved 1st place in the 2018 Atrial Segmentation Challenge. Conv, convolution; Deconv, deconvolution; PReLU, parametric rectified linear unit.



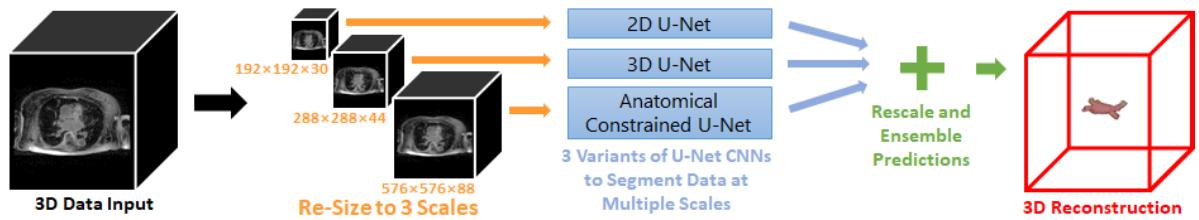
Appendix Figure 3. Convolutional neural network (CNN) architecture proposed by Huang [260] who achieved 2nd place in the 2018 Atrial Segmentation Challenge. Conv, convolution; Deconv, deconvolution; ReLU, rectified linear unit.



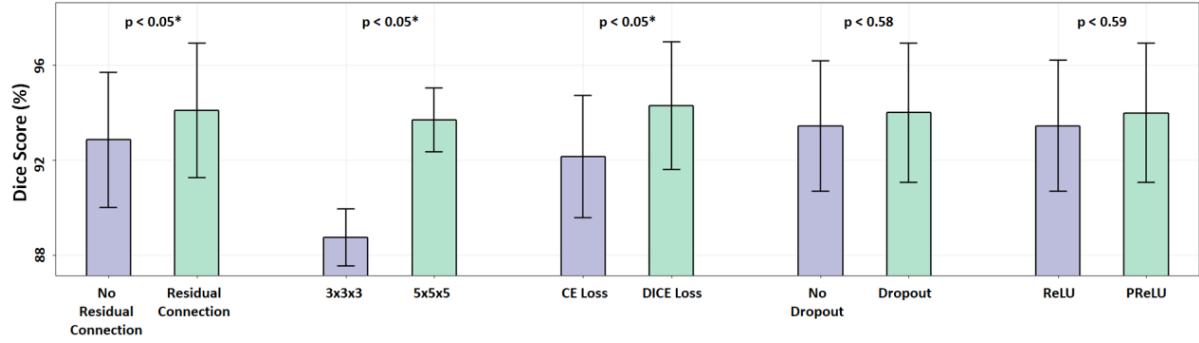
Appendix Figure 4. Convolutional neural network (CNN) architecture proposed by Li et al. [249] who achieved 6th place in the 2018 Atrial Segmentation Challenge. Conv, convolution; BN, batch normalization; ReLU, rectified linear unit; Concat, feature map concatenation.



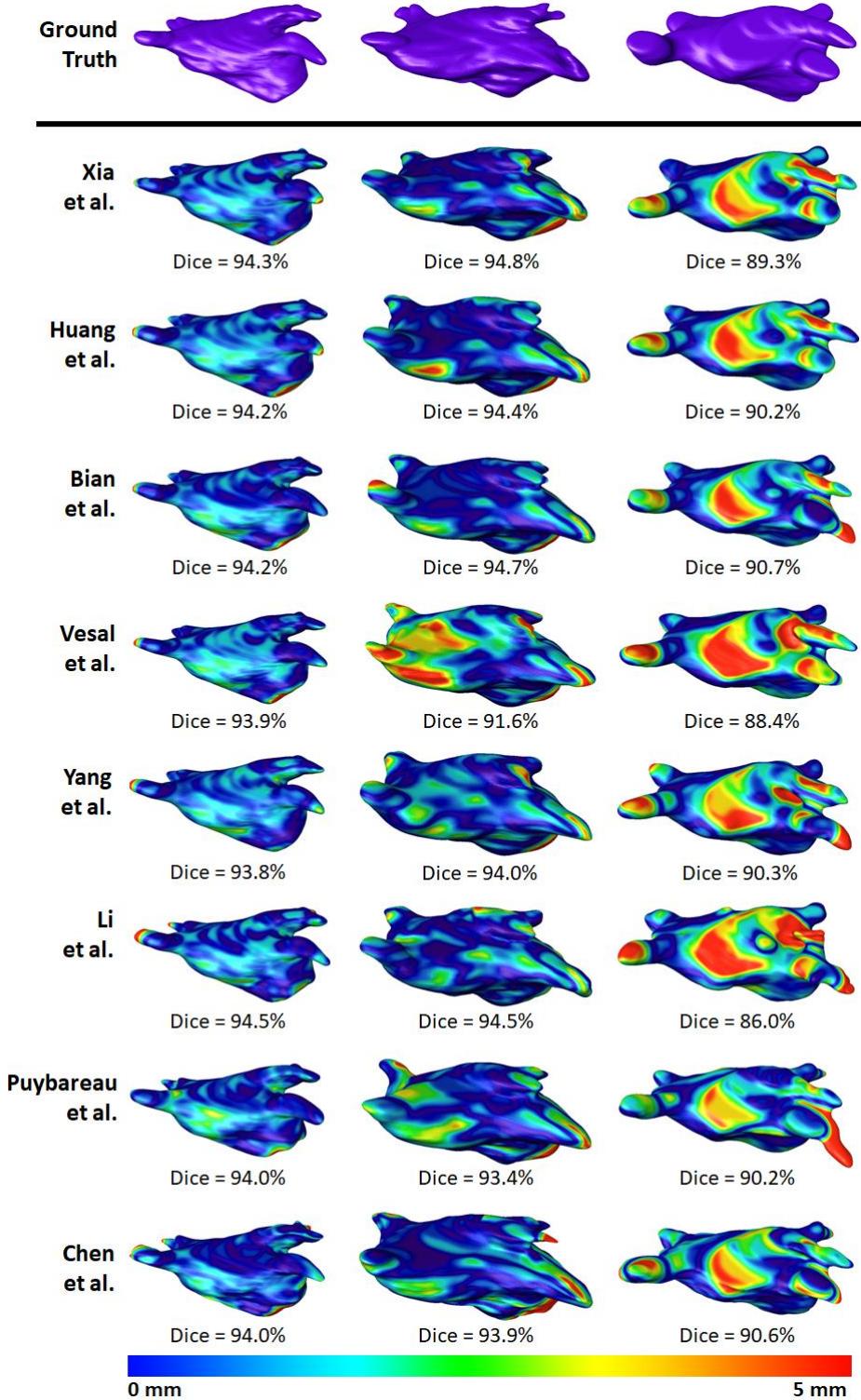
Appendix Figure 5. Convolutional neural network (CNN) architecture proposed by Yang et al. [248] who achieved 5th place in the 2018 Atrial Segmentation Challenge.



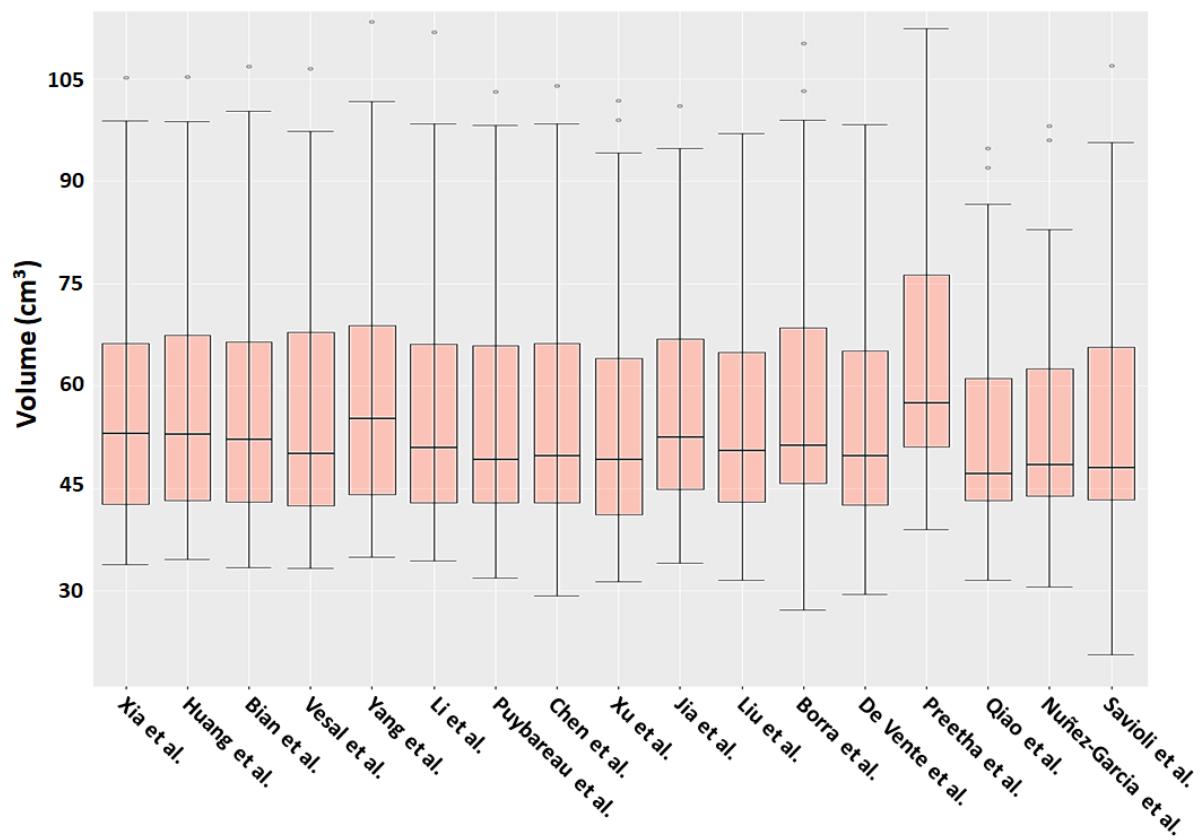
Appendix Figure 6. Proposed convolutional neural (CNN) network pipeline by Xu et al. [261] who achieved 9th place in the 2018 Atrial Segmentation Challenge.



Appendix Figure 7. Hyper-parameter tuning of the U-Net architecture in the second CNN by Xia et al. [168] The comparative data used was obtained from the multi-variate hyper-parameter tuning of all parameter combinations of the five parameters presented.



Appendix Figure 8. Ground truth and predicted left atrium (LA) geometries produced by the top 8 teams in the challenge for three randomly selected late gadolinium-enhanced magnetic resonance imaging (LGE-MRI). The rows represent different participants, and the columns represent LA from different LGE-MRIs. The distance between the predictions and the ground truths for each LGE-MRI is shown with a colour map ranging between 0 mm to 5 mm.



Appendix Figure 9. The distribution of the left atrium (LA) volume predicted by each team in the challenge for the 54 test LGE-MRIs.

Bibliography

- [1] S. Nattel, "New ideas about atrial fibrillation 50 years on," *Nature*, vol. 415, pp. 219-226, 2002.
- [2] S. M. Narayan, M. Rodrigo, C. A. Kowalewski, F. Shenasa, G. L. Meckler, M. N. Vishwanathan, *et al.*, "Ablation of focal impulses and rotational sources: What can be learned from differing procedural outcomes?," *Current Cardiovascular Risk Reports*, vol. 11, p. 27, 2017.
- [3] T. J. Wang, M. G. Larson, D. Levy, R. S. Vasan, E. P. Leip, P. A. Wolf, *et al.*, "Temporal relations of atrial fibrillation and congestive heart failure and their joint influence on mortality: the Framingham Heart Study," *Circulation*, vol. 107, pp. 2920-2925, 2003.
- [4] C. A. Morillo, A. Banerjee, P. Perel, D. Wood, and X. Jouven, "Atrial fibrillation: the current epidemic," *Journal of Geriatric Cardiology: JGC*, vol. 14, p. 195, 2017.
- [5] R. Fogel, "The pathophysiology of atrial fibrillation and implications for therapy," *Internet: www.medscape.com/viewarticle/491110 (accessed 17 December 2004)*.
- [6] S. Mittal, "Differentiating paroxysmal from persistent atrial fibrillation: long-term electrocardiographic monitoring is mightier than the clinician," ed: *Journal of the American College of Cardiology*, 2014.
- [7] P. Dorian, W. Jung, D. Newman, M. Paquette, K. Wood, G. M. Ayers, *et al.*, "The impairment of health-related quality of life in patients with intermittent atrial fibrillation: implications for the assessment of investigational therapy," *Journal of the American College of Cardiology*, vol. 36, pp. 1303-1309, 2000.
- [8] R. L. Page, W. E. Wilkinson, W. K. Clair, E. A. McCarthy, and E. Pritchett, "Asymptomatic arrhythmias in patients with symptomatic paroxysmal atrial fibrillation and paroxysmal supraventricular tachycardia," *Circulation*, vol. 89, pp. 224-227, 1994.
- [9] R. Nieuwlaat, M. H. Prins, J.-Y. Le Heuzey, P. E. Vardas, E. Aliot, M. Santini, *et al.*, "Prognosis, disease progression, and treatment of atrial fibrillation patients during 1 year: follow-up of the Euro Heart Survey on atrial fibrillation," *European Heart Journal*, vol. 29, pp. 1181-1189, 2008.
- [10] B. A. Steinberg, A. S. Hellkamp, Y. Lokhnygina, M. R. Patel, G. Breithardt, G. J. Hankey, *et al.*, "Higher risk of death and stroke in patients with persistent vs.

paroxysmal atrial fibrillation: results from the ROCKET-AF Trial," *European Heart Journal*, vol. 36, pp. 288-296, 2015.

- [11] A. Barbarossa, F. Guerra, and A. Capucci, "Silent atrial fibrillation: a critical review," *Journal of Atrial Fibrillation*, vol. 7, 2014.
- [12] L. Friberg, M. Rosenqvist, A. Lindgren, A. Terént, B. Norrving, and K. Asplund, "High prevalence of atrial fibrillation among patients with ischemic stroke," *Stroke*, vol. 45, pp. 2599-2605, 2014.
- [13] C. R. Kerr, K. H. Humphries, M. Talajic, G. J. Klein, S. J. Connolly, M. Green, *et al.*, "Progression to chronic atrial fibrillation after the initial diagnosis of paroxysmal atrial fibrillation: results from the Canadian Registry of Atrial Fibrillation," *American Heart Journal*, vol. 149, pp. 489-496, 2005.
- [14] T. Fetsch, P. Bauer, R. Engberding, H. P. Koch, J. Lukl, T. Meinertz, *et al.*, "Prevention of atrial fibrillation after cardioversion: results of the PAFAC trial," *European Heart Journal*, vol. 25, pp. 1385-1394, 2004.
- [15] D. B. Mark, K. J. Anstrom, S. Sheng, J. P. Piccini, K. N. Baloch, K. H. Monahan, *et al.*, "Effect of catheter ablation vs medical therapy on quality of life among patients with atrial fibrillation: the CABANA randomized clinical trial," *Jama*, vol. 321, pp. 1275-1285, 2019.
- [16] A. Goette, A. Auricchio, G. Boriani, F. Braunschweig, J. B. Terradellas, H. Burri, *et al.*, "EHRA White Paper: knowledge gaps in arrhythmia management—status 2019," *EP Europace*, vol. 21, pp. 993-994, 2019.
- [17] C. T. January, L. S. Wann, J. S. Alpert, H. Calkins, J. E. Cigarroa, J. B. Conti, *et al.*, "2014 AHA/ACC/HRS guideline for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the Heart Rhythm Society," *Journal of the American College of Cardiology*, vol. 64, pp. e1-e76, 2014.
- [18] E. Buch, M. Share, R. Tung, P. Benharash, P. Sharma, J. Koneru, *et al.*, "Long-term clinical outcomes of focal impulse and rotor modulation for treatment of atrial fibrillation: A multicenter experience," *Heart Rhythm*, vol. 13, pp. 636-641, 2016.
- [19] C. Gianni, S. Mohanty, L. Di Biase, T. Metz, C. Trivedi, Y. Gökoğlan, *et al.*, "Acute and early outcomes of focal impulse and rotor modulation (FIRM)-guided rotors-only ablation in patients with nonparoxysmal atrial fibrillation," *Heart Rhythm*, vol. 13, pp. 830-835, 2016.

- [20] B. J. Hansen, J. Zhao, T. A. Csepe, B. T. Moore, N. Li, L. A. Jayne, *et al.*, "Atrial fibrillation driven by micro-anatomic intramural re-entry revealed by simultaneous sub-epicardial and sub-endocardial optical mapping in explanted human hearts," *European Heart Journal*, vol. 36, pp. 2390-2401, 2015.
- [21] J. Zhao, T. D. Butters, H. Zhang, A. J. Pullan, I. J. LeGrice, G. B. Sands, *et al.*, "An image-based model of atrial muscular architecture: effects of structural anisotropy on electrical activation," *Circulation: Arrhythmia and Electrophysiology*, p. CIRCEP. 111.967950, 2012.
- [22] A. G. Brooks, M. K. Stiles, J. Laborderie, D. H. Lau, P. Kuklik, N. J. Shipp, *et al.*, "Outcomes of long-standing persistent atrial fibrillation ablation: a systematic review," *Heart Rhythm*, vol. 7, pp. 835-846, 2010.
- [23] S. Osowski, L. T. Hoai, and T. Markiewicz, "Support vector machine-based expert system for reliable heartbeat recognition," *IEEE Transactions on Biomedical Engineering*, vol. 51, pp. 582-589, 2004.
- [24] N. Lowres, L. Neubeck, G. Salkeld, I. Krass, A. J. McLachlan, J. Redfern, *et al.*, "Feasibility and cost-effectiveness of stroke prevention through community screening for atrial fibrillation using iPhone ECG in pharmacies," *Thrombosis and Haemostasis*, vol. 111, pp. 1167-1176, 2014.
- [25] M. V. Perez, K. W. Mahaffey, H. Hedlin, J. S. Rumsfeld, A. Garcia, T. Ferris, *et al.*, "Large-scale assessment of a smartwatch to identify atrial fibrillation," *New England Journal of Medicine*, vol. 381, pp. 1909-1917, 2019.
- [26] J. M. Bumgarner, C. T. Lambert, A. A. Hussein, D. J. Cantillon, B. Baranowski, K. Wolski, *et al.*, "Smartwatch algorithm for automated detection of atrial fibrillation," *Journal of the American College of Cardiology*, vol. 71, pp. 2381-2388, 2018.
- [27] J. Zhao, B. J. Hansen, Y. Wang, T. A. Csepe, L. V. Sul, A. Tang, *et al.*, "Three-dimensional integrated functional, structural, and computational mapping to define the structural "fingerprints" of heart-specific atrial fibrillation drivers in human heart ex vivo," *Journal of the American Heart Association*, vol. 6, p. e005922, 2017.
- [28] R. S. Oakes, T. J. Badger, E. G. Kholmovski, N. Akoum, N. S. Burgon, E. N. Fish, *et al.*, "Detection and quantification of left atrial structural remodeling with delayed-enhancement magnetic resonance imaging in patients with atrial fibrillation," *Circulation*, vol. 119, pp. 1758-1767, 2009.

- [29] K. Higuchi, J. Cates, G. Gardner, A. Morris, N. S. Burgon, N. Akoum, *et al.*, "The spatial distribution of late gadolinium enhancement of left atrial MRI in patients with atrial fibrillation," *JACC: Clinical Electrophysiology*, 2017.
- [30] P. Gal and N. F. Marrouche, "Magnetic resonance imaging of atrial fibrosis: Redefining atrial fibrillation to a syndrome," *European Heart Journal*, p. ehv514, 2015; In Press.
- [31] A. Njoku, M. Kannabhiran, R. Arora, P. Reddy, R. Gopinathannair, D. Lakkireddy, *et al.*, "Left atrial volume predicts atrial fibrillation recurrence after radiofrequency ablation: a meta-analysis," *EP Europace*, p. eux013, 2017.
- [32] J. Zhuang, Y. Wang, K. Tang, X. Li, W. Peng, C. Liang, *et al.*, "Association between left atrial size and atrial fibrillation recurrence after single circumferential pulmonary vein isolation: a systematic review and meta-analysis of observational studies," *Europace*, vol. 14, pp. 638-645, 2011.
- [33] A. Njoku, M. Kannabhiran, R. Arora, P. Reddy, R. Gopinathannair, D. Lakkireddy, *et al.*, "Left atrial volume predicts atrial fibrillation recurrence after radiofrequency ablation: a meta-analysis," *Ep Europace*, vol. 20, pp. 33-42, 2017.
- [34] J. L. Cox, T. E. Canavan, R. B. Schuessler, M. E. Cain, B. D. Lindsay, C. Stone, *et al.*, "The surgical treatment of atrial fibrillation: II. Intraoperative electrophysiologic mapping and description of the electrophysiologic basis of atrial flutter and atrial fibrillation," *The Journal of Thoracic and Cardiovascular Surgery*, vol. 101, pp. 406-426, 1991.
- [35] T. A. Csepe, J. Zhao, L. V. Sul, Y. Wang, B. J. Hansen, N. Li, *et al.*, "Novel application of 3D contrast-enhanced CMR to define fibrotic structure of the human sinoatrial node in vivo," *European Heart Journal-Cardiovascular Imaging*, p. jew304, 2017.
- [36] G. Veni, S. Y. Elhabian, and R. T. Whitaker, "ShapeCut: Bayesian Surface Estimation Using Shape-Driven Graph," *Medical Image Analysis*, 2017.
- [37] Q. Tao, E. G. Ipek, R. Shahzad, F. F. Berendsen, S. Nazarian, and R. J. van der Geest, "Fully automatic segmentation of left atrium and pulmonary veins in late gadolinium-enhanced MRI: Towards objective atrial scar assessment," *Journal of Magnetic Resonance Imaging*, vol. 44, pp. 346-354, 2016.
- [38] A. Mortazi, R. Karim, R. Kawal, J. Burt, and U. Bagci, "CardiacNET: Segmentation of left atrium and proximal pulmonary veins from MRI using multi-view CNN," *arXiv preprint arXiv:1705.06333*, 2017.

- [39] G. Yang, X. Zhuang, H. Khan, S. Haldar, E. Nyktari, L. Li, *et al.*, "Fully automatic segmentation and objective assessment of atrial scars for longstanding persistent atrial fibrillation patients using late gadolinium-enhanced MRI," *arXiv preprint arXiv:1705.09529*, 2017.
- [40] V. Y. Reddy, G. Morales, H. Ahmed, P. Neuzil, S. Dukkipati, S. Kim, *et al.*, "Catheter ablation of atrial fibrillation without the use of fluoroscopy," *Heart Rhythm*, vol. 7, pp. 1644-1653, 2010.
- [41] H. L. Estner, I. Deisenhofer, A. Luik, G. Ndrepepa, C. von Bary, B. Zrenner, *et al.*, "Electrical isolation of pulmonary veins in patients with atrial fibrillation: reduction of fluoroscopy exposure and procedure duration by the use of a non-fluoroscopic navigation system (NavX®)," *Europace*, vol. 8, pp. 583-587, 2006.
- [42] E. K. Heist, J. Chevalier, G. Holmvang, J. P. Singh, P. T. Ellinor, D. J. Milan, *et al.*, "Factors affecting error in integration of electroanatomic mapping with CT and MR imaging during catheter ablation of atrial fibrillation," *Journal of Interventional Cardiac Electrophysiology*, vol. 17, pp. 21-27, 2006.
- [43] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85-117, 2015.
- [44] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.
- [45] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep learning* vol. 1: MIT press Cambridge, 2016.
- [46] 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.
- [47] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Advances in Neural Information Processing Systems*, 2014, pp. 1988-1996.
- [48] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of Machine Learning Research*, vol. 12, pp. 2493-2537, 2011.
- [49] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, *et al.*, "End to end learning for self-driving cars," *arXiv preprint arXiv:1604.07316*, 2016.
- [50] Z. Xiong, V. V. Fedorov, X. Fu, E. Cheng, R. Macleod, and J. Zhao, "Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging

using a dual fully convolutional neural network," *IEEE Transactions on Medical Imaging*, 2018.

- [51] Z. Xiong, M. P. Nash, E. Cheng, V. V. Fedorov, M. K. Stiles, and J. Zhao, "ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network," *Physiological Measurement*, 2018.
- [52] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Medical Image Analysis*, vol. 36, pp. 61-78, 2017.
- [53] K.-L. Hua, C.-H. Hsu, S. C. Hidayati, W.-H. Cheng, and Y.-J. Chen, "Computer-aided classification of lung nodules on computed tomography images via deep learning technique," *OncoTargets and Therapy*, vol. 8, 2015.
- [54] Y. Xu, Z. Dai, F. Chen, S. Gao, J. Pei, and L. Lai, "Deep learning for drug-induced liver injury," *Journal of Chemical Information and Modeling*, vol. 55, pp. 2085-2093, 2015.
- [55] H. Lee, S. Tajmir, J. Lee, M. Zissen, B. A. Yeshivas, T. K. Alkasab, *et al.*, "Fully automated deep learning system for bone age assessment," *Journal of Digital Imaging*, vol. 30, pp. 427-441, 2017.
- [56] 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.
- [57] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Işgum, "Automatic coronary calcium scoring in cardiac CT angiography using convolutional neural networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 589-596.
- [58] A. Madani, R. Arnaout, M. Mofrad, and R. Arnaout, "Fast and accurate view classification of echocardiograms using deep learning," *NPJ Digital Medicine*, vol. 1, p. 6, 2018.
- [59] G. Lee, P. Sanders, and J. M. Kalman, "Catheter ablation of atrial arrhythmias: state of the art," *The Lancet*, vol. 380, pp. 1509-1519, 2012.
- [60] D. Glomset and K. Cross, "Morphologic study of the cardiac conduction system: VI. the intrinsic nervous system of the heart," *AMA Archives of Internal Medicine*, vol. 89, pp. 923-930, 1952.
- [61] H. C. Associates, "What is Atrial Fibrillation?," 2021.

- [62] H. Blackburn, A. Keys, E. Simonson, P. Rautaharju, and S. Punzar, "The electrocardiogram in population studies: a classification system," *Circulation*, vol. 21, pp. 1160-1175, 1960.
- [63] D. Bernal, "Phases of the Cardiac Cycle," 2019.
- [64] C. H. S. Scotland, "Diagnosis and Treatment of Atrial Fibrillation," 2021.
- [65] S. Nattel and M. Harada, "Atrial remodeling and atrial fibrillation: recent advances and translational perspectives," *Journal of the American College of Cardiology*, vol. 63, pp. 2335-2345, 2014.
- [66] S. Nattel, B. Burstein, and D. Dobrev, "Atrial remodeling and atrial fibrillation: mechanisms and implications," *Circulation: Arrhythmia and Electrophysiology*, vol. 1, pp. 62-73, 2008.
- [67] M. C. Wijffels, C. J. Kirchhof, R. Dorland, and M. A. Allessie, "Atrial fibrillation begets atrial fibrillation: a study in awake chronically instrumented goats," *Circulation*, vol. 92, pp. 1954-1968, 1995.
- [68] S.-M. Chaloupi, P. Loh, R. N. Hauer, J. M. De Bakker, and H. V. van Rijen, "The role of connexin40 in atrial fibrillation," *Cardiovascular Research*, vol. 84, pp. 15-23, 2009.
- [69] S. Nattel and D. Dobrev, "The multidimensional role of calcium in atrial fibrillation pathophysiology: mechanistic insights and therapeutic opportunities," *European Heart Journal*, vol. 33, pp. 1870-1877, 2012.
- [70] S. V. Pandit, O. Berenfeld, J. M. Anumonwo, R. M. Zaritski, J. Kneller, S. Nattel, *et al.*, "Ionic determinants of functional reentry in a 2-D model of human atrial cells during simulated chronic atrial fibrillation," *Biophysical Journal*, vol. 88, pp. 3806-3821, 2005.
- [71] P. T. Ellinor, K. L. Lunetta, N. L. Glazer, A. Pfeufer, A. Alonso, M. K. Chung, *et al.*, "Common variants in KCNN3 are associated with lone atrial fibrillation," *Nature Genetics*, vol. 42, pp. 240-244, 2010.
- [72] T. Igarashi, J. E. Finet, A. Takeuchi, Y. Fujino, M. Strom, I. D. Greener, *et al.*, "Connexin gene transfer preserves conduction velocity and prevents atrial fibrillation," *Circulation*, vol. 125, pp. 216-225, 2012.
- [73] J. Ausma, M. Wijffels, F. Thoné, L. Wouters, M. Allessie, and M. Borgers, "Structural changes of atrial myocardium due to sustained atrial fibrillation in the goat," *Circulation*, vol. 96, pp. 3157-3163, 1997.

- [74] R. Zou, J. Kneller, L. J. Leon, and S. Nattel, "Substrate size as a determinant of fibrillatory activity maintenance in a mathematical model of canine atrium," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 289, pp. H1002-H1012, 2005.
- [75] B. Burstein and S. Nattel, "Atrial fibrosis: mechanisms and clinical relevance in atrial fibrillation," *Journal of the American College of Cardiology*, vol. 51, pp. 802-809, 2008.
- [76] H. M. van der Velden and H. J. Jongsma, "Cardiac gap junctions and connexins: their role in atrial fibrillation and potential as therapeutic targets," *Cardiovascular Research*, vol. 54, pp. 270-279, 2002.
- [77] T. A. Baudino, W. Carver, W. Giles, and T. K. Borg, "Cardiac fibroblasts: friend or foe?," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 291, pp. H1015-H1026, 2006.
- [78] H. N. Sabbah, V. G. Sharov, M. Lesch, and S. Goldstein, "Progression of heart failure: a role for interstitial fibrosis," in *Cellular Interactions in Cardiac Pathophysiology*, ed: Springer, 1995, pp. 29-34.
- [79] P. Camelliti, T. K. Borg, and P. Kohl, "Structural and functional characterisation of cardiac fibroblasts," *Cardiovascular Research*, vol. 65, pp. 40-51, 2005.
- [80] K. Weber and C. Brilla, "Factors associated with reactive and reparative fibrosis of the myocardium," *Cellular and Molecular Alterations in the Failing Human Heart*, pp. 291-301, 1992.
- [81] C. Kupfahl, D. Pink, K. Friedrich, H. R. Zurbrügg, M. Neuss, C. Warnecke, *et al.*, "Angiotensin II directly increases transforming growth factor β 1 and osteopontin and indirectly affects collagen mRNA expression in the human heart," *Cardiovascular Research*, vol. 46, pp. 463-475, 2000.
- [82] A. C. Newton, C. E. Antal, and S. F. Steinberg, "Protein kinase C mechanisms that contribute to cardiac remodelling," *Clinical Science*, vol. 130, pp. 1499-1510, 2016.
- [83] M. I. Garcia and D. Boehning, "Cardiac inositol 1, 4, 5-trisphosphate receptors," *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, vol. 1864, pp. 907-914, 2017.
- [84] L. Hunyady and K. J. Catt, "Pleiotropic AT1 receptor signaling pathways mediating physiological and pathogenic actions of angiotensin II," *Molecular Endocrinology*, vol. 20, pp. 953-970, 2006.

- [85] S. Rosenkranz, "TGF- β 1 and angiotensin networking in cardiac remodeling," *Cardiovascular Research*, vol. 63, pp. 423-432, 2004.
- [86] L. Attisano and J. L. Wrana, "Signal transduction by the TGF- β superfamily," *Science*, vol. 296, pp. 1646-1647, 2002.
- [87] S. Verheule, T. Sato, T. Everett IV, S. K. Engle, D. Otten, M. Rubart-Von Der Lohe, *et al.*, "Increased vulnerability to atrial fibrillation in transgenic mice with selective atrial fibrosis caused by overexpression of TGF- β 1," *Circulation Research*, vol. 94, pp. 1458-1465, 2004.
- [88] I. A. Darby and T. D. Hewitson, "Fibroblast differentiation in wound healing and fibrosis," *International review of cytology*, vol. 257, pp. 143-179, 2007.
- [89] B. Burstein, E. Libby, A. Calderone, and S. Nattel, "Atrial fibroblasts are different from ventricular: a potential contributor to atrial-ventricular remodeling differences," *Circulation*, vol. 117, pp. 1630-1641, 2008.
- [90] S. M. Chuva De Sousa Lopes, A. Feijen, J. Korving, O. Korchynskyi, J. Larsson, S. Karlsson, *et al.*, "Connective tissue growth factor expression and Smad signaling during mouse heart development and myocardial infarction," *Developmental Dynamics: an Official Publication of the American Association of Anatomists*, vol. 231, pp. 542-550, 2004.
- [91] P. G. Platonov, L. B. Mitrofanova, V. Orshanskaya, and S. Y. Ho, "Structural abnormalities in atrial walls are associated with presence and persistency of atrial fibrillation but not with age," *Journal of the American College of Cardiology*, vol. 58, pp. 2225-2232, 2011.
- [92] H. Fu, G. Li, C. Liu, J. Li, X. Wang, L. Cheng, *et al.*, "Probucol prevents atrial remodeling by inhibiting oxidative stress and TNF- α /NF- κ B/TGF- β signal transduction pathway in alloxan-induced diabetic rabbits," *Journal of Cardiovascular Electrophysiology*, vol. 26, pp. 211-222, 2015.
- [93] E. M. van Oosten, A. H. Boag, K. Cunningham, J. Veinot, A. Hamilton, D. Petsikas, *et al.*, "The histology of human right atrial tissue in patients with high-risk obstructive sleep apnea and underlying cardiovascular disease: a pilot study," *IJC Heart & Vasculature*, vol. 6, pp. 71-75, 2015.
- [94] H. A. Alturkistani, F. M. Tashkandi, and Z. M. Mohammedsaleh, "Histological stains: a literature review and case study," *Global Journal of Health Science*, vol. 8, p. 72, 2016.

- [95] M. R. Prince, "Gadolinium-enhanced MR aortography," *Radiology*, vol. 191, pp. 155-164, 1994.
- [96] S. Morcos, "Extracellular gadolinium contrast agents: differences in stability," *European Journal of Radiology*, vol. 66, pp. 175-179, 2008.
- [97] D. A. Finelli, G. C. Hurst, R. P. Gullapali, and E. M. Bellon, "Improved contrast of enhancing brain lesions on postgadolinium, T1-weighted spin-echo images with use of magnetization transfer," *Radiology*, vol. 190, pp. 553-559, 1994.
- [98] W. P. Bandettini, P. Kellman, C. Mancini, O. J. Booker, S. Vasu, S. W. Leung, *et al.*, "MultiContrast Delayed Enhancement (MCODE) improves detection of subendocardial myocardial infarction by late gadolinium enhancement cardiovascular magnetic resonance: a clinical validation study," *Journal of Cardiovascular Magnetic Resonance*, vol. 14, pp. 1-10, 2012.
- [99] D. C. Peters, J. V. Wylie, T. H. Hauser, K. V. Kissinger, R. M. Botnar, V. Essebag, *et al.*, "Detection of pulmonary vein and left atrial scar after catheter ablation with three-dimensional navigator-gated delayed enhancement MR imaging: initial experience," *Radiology*, vol. 243, pp. 690-695, 2007.
- [100] M. Mannil, J. von Spiczak, R. Manka, and H. Alkadhi, "Texture analysis and machine learning for detecting myocardial infarction in noncontrast low-dose computed tomography: unveiling the invisible," *Investigative Radiology*, vol. 53, pp. 338-343, 2018.
- [101] T. D. Nguyen, P. Spincemaille, J. W. Weinsaft, B. Y. Ho, M. D. Cham, M. R. Prince, *et al.*, "A fast navigator-gated 3D sequence for delayed enhancement MRI of the myocardium: comparison with breathhold 2D imaging," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 27, pp. 802-808, 2008.
- [102] Y. Amano, Y. Matsumura, and S. Kumita, "Free-breathing high-spatial-resolution delayed contrast-enhanced three-dimensional viability MR imaging of the myocardium at 3.0 T: A feasibility study," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 28, pp. 1361-1367, 2008.
- [103] S. Toupin, T. Pezel, A. Bustin, and H. Cochet, "Whole-Heart High-Resolution Late Gadolinium Enhancement: Techniques and Clinical Applications," *Journal of Magnetic Resonance Imaging: JMRI*.

- [104] N. F. Marrouche, T. Greene, J. M. Dean, E. G. Kholmovski, L. M. d. Boer, M. Mansour, *et al.*, "Efficacy of LGE-MRI-guided fibrosis ablation versus conventional catheter ablation of atrial fibrillation: The DECAAF II trial: Study design," *Journal of Cardiovascular Electrophysiology*, vol. 32, pp. 916-924, 2021.
- [105] W. L. Henry, J. Morganroth, A. S. Pearlman, C. E. Clark, D. R. Redwood, S. B. Itscoitz, *et al.*, "Relation between echocardiographically determined left atrial size and atrial fibrillation," *Circulation*, vol. 53, pp. 273-279, 1976.
- [106] T. S. Tsang, M. E. Barnes, K. R. Bailey, C. L. Leibson, S. C. Montgomery, Y. Takemoto, *et al.*, "Left atrial volume: important risk marker of incident atrial fibrillation in 1655 older men and women," in *Mayo Clinic Proceedings*, 2001, pp. 467-475.
- [107] S. M. Vaziri, M. G. Larson, E. J. Benjamin, and D. Levy, "Echocardiographic predictors of nonrheumatic atrial fibrillation. The Framingham Heart Study," *Circulation*, vol. 89, pp. 724-730, 1994.
- [108] K. Nakamura, N. Funabashi, M. Uehara, M. Ueda, T. Murayama, H. Takaoka, *et al.*, "Left atrial wall thickness in paroxysmal atrial fibrillation by multislice-CT is initial marker of structural remodeling and predictor of transition from paroxysmal to chronic form," *International Journal of Cardiology*, vol. 148, pp. 139-147, 2011.
- [109] R. Karim, L.-E. Blake, J. Inoue, Q. Tao, S. Jia, R. J. Housden, *et al.*, "Algorithms for left atrial wall segmentation and thickness—Evaluation on an open-source CT and MRI image database," *Medical Image Analysis*, vol. 50, pp. 36-53, 2018.
- [110] M. Bishop, R. Rajani, G. Plank, N. Gaddum, G. Carr-White, M. Wright, *et al.*, "Three-dimensional atrial wall thickness maps to inform catheter ablation procedures for atrial fibrillation," *Europace*, vol. 18, pp. 376-383, 2015.
- [111] J.-S. Song, J. Wi, H.-J. Lee, M. Hwang, B. Lim, T.-H. Kim, *et al.*, "Role of atrial wall thickness in wave-dynamics of atrial fibrillation," *PloS one*, vol. 12, p. e0182174, 2017.
- [112] A. Roy, M. Varela, and O. Aslanidi, "Image-Based Computational Evaluation of the Effects of Atrial Wall Thickness and Fibrosis on Re-entrant Drivers for Atrial Fibrillation," *Frontiers in Physiology*, vol. 9, 2018.
- [113] R. Andlauer, G. Seemann, L. Baron, O. Dössel, P. Kohl, P. Platonov, *et al.*, "Influence of left atrial size on P-wave morphology: differential effects of dilation and hypertrophy," *EP Europace*, vol. 20, pp. iii36-iii44, 2018.

- [114] C. McGann, N. Akoum, A. Patel, E. Kholmovski, P. Revelo, K. Damal, *et al.*, "Atrial fibrillation ablation outcome is predicted by left atrial remodeling on MRI," *Circulation: Arrhythmia and Electrophysiology*, vol. 7, pp. 23-30, 2014.
- [115] R. A. DeSilva, T. B. Graboys, P. J. Podrid, and B. Lown, "Cardioversion and defibrillation," *American Heart Journal*, vol. 100, pp. 881-895, 1980.
- [116] M. Haissaguerre, P. Jaïs, D. C. Shah, A. Takahashi, M. Hocini, G. Quiniou, *et al.*, "Spontaneous initiation of atrial fibrillation by ectopic beats originating in the pulmonary veins," *New England Journal of Medicine*, vol. 339, pp. 659-666, 1998.
- [117] C. A. Morillo, A. Verma, S. J. Connolly, K. H. Kuck, G. M. Nair, J. Champagne, *et al.*, "Radiofrequency ablation vs antiarrhythmic drugs as first-line treatment of paroxysmal atrial fibrillation (RAAFT-2): a randomized trial," *Jama*, vol. 311, pp. 692-700, 2014.
- [118] J. Cosedis Nielsen, A. Johannessen, P. Raatikainen, G. Hindricks, H. Walfridsson, O. Kongstad, *et al.*, "Radiofrequency ablation as initial therapy in paroxysmal atrial fibrillation," *New England Journal of Medicine*, vol. 367, pp. 1587-1595, 2012.
- [119] R. Parkash, A. S. Tang, J. L. Sapp, and G. Wells, "Approach to the catheter ablation technique of paroxysmal and persistent atrial fibrillation: A meta-analysis of the randomized controlled trials," *Journal of Cardiovascular Electrophysiology*, vol. 22, pp. 729-738, 2011.
- [120] B. J. Hansen, J. Zhao, N. Li, A. Zolotarev, S. Zakharkin, Y. Wang, *et al.*, "Human atrial fibrillation drivers resolved with integrated functional and structural imaging to benefit clinical mapping," *JACC: Clinical Electrophysiology*, vol. 4, pp. 1501-1515, 2018.
- [121] M. K. Stiles, P. Sanders, and D. H. Lau, "Targeting the Substrate in Ablation of Persistent Atrial Fibrillation: Recent Lessons and Future Directions," *Frontiers in Physiology*, vol. 9, 2018.
- [122] H. S. Lim, M. Hocini, R. Dubois, A. Denis, N. Derval, S. Zellerhoff, *et al.*, "Complexity and distribution of drivers in relation to duration of persistent atrial fibrillation," *Journal of the American College of Cardiology*, vol. 69, pp. 1257-1269, 2017.
- [123] J. M. Miller, V. Kalra, M. K. Das, R. Jain, J. B. Garlie, J. A. Brewster, *et al.*, "Clinical benefit of ablating localized sources for human atrial fibrillation: the Indiana University FIRM Registry," *Journal of the American College of Cardiology*, vol. 69, pp. 1247-1256, 2017.

- [124] S. Prabhu, A. Voskoboinik, A. J. McLellan, K. Y. Peck, B. Pathik, C. J. Nalliah, *et al.*, "A comparison of the electrophysiologic and electroanatomic characteristics between the right and left atrium in persistent atrial fibrillation: is the right atrium a window into the left?," *Journal of Cardiovascular Electrophysiology*, vol. 28, pp. 1109-1116, 2017.
- [125] H. Calkins, G. Hindricks, R. Cappato, Y.-H. Kim, E. B. Saad, L. Aguinaga, *et al.*, "2017 HRS/EHRA/ECAS/APHRS/SOLAECE expert consensus statement on catheter and surgical ablation of atrial fibrillation: executive summary," 2017.
- [126] Z. Issa, J. M. Miller, and D. P. Zipes, *Clinical Arrhythmology and Electrophysiology: A Companion to Braunwald's Heart Disease E-Book: Expert Consult: Online and Print*: Elsevier Health Sciences, 2012.
- [127] A. Merschon and F. Massarwa, "Dynamic feature rich anatomical reconstruction from a point cloud," ed: Google Patents, 2016.
- [128] A. Verma, O. M. Wazni, N. F. Marrouche, D. O. Martin, F. Kilicaslan, S. Minor, *et al.*, "Pre-existent left atrial scarring in patients undergoing pulmonary vein antrum isolation: an independent predictor of procedural failure," *Journal of the American College of Cardiology*, vol. 45, pp. 285-292, 2005.
- [129] I. M. Khurram, R. Beinart, V. Zipunnikov, J. Dewire, H. Yarmohammadi, T. Sasaki, *et al.*, "Magnetic resonance image intensity ratio, a normalized measure to enable interpatient comparability of left atrial fibrosis," *Heart Rhythm*, vol. 11, pp. 85-92, 2014.
- [130] S. Rolf, S. Kircher, A. Arya, C. Eitel, P. Sommer, S. Richter, *et al.*, "Tailored atrial substrate modification based on low-voltage areas in catheter ablation of atrial fibrillation," *Circulation: Arrhythmia and Electrophysiology*, vol. 7, pp. 825-833, 2014.
- [131] F. S. Foo, K. K. Poppe, M. Lee, G. C. Clare, M. K. Stiles, K. L. Looi, *et al.*, "Regional variation in cardiac implantable electronic device implants trends in New Zealand over the past decade (ANZACS-QI 54)," *Internal Medicine Journal*, 2020.
- [132] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Processing Letters*, vol. 9, pp. 293-300, 1999.
- [133] T. K. Ho, "Random decision forests," in *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, 1995, pp. 278-282.
- [134] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, pp. 21-27, 1967.

- [135] F. Sebastiani, "Machine learning in automated text categorization," *ACM Computing Surveys (CSUR)*, vol. 34, pp. 1-47, 2002.
- [136] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010, pp. 807-814.
- [137] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, pp. 400-407, 1951.
- [138] D. E. Rumelhart, "Learning representations by back-propagation errors," *Parallel Distributed Processing*, pp. 318-362, 1986.
- [139] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249-256.
- [140] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: a simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929-1958, 2014.
- [141] Y. Bengio, P. Frasconi, and P. Simard, "The problem of learning long-term dependencies in recurrent networks," in *Neural Networks, 1993., IEEE International Conference on*, 1993, pp. 1183-1188.
- [142] P. Rajpurkar, A. Y. Hannun, M. Haghpanahi, C. Bourn, and A. Y. Ng, "Cardiologist-level arrhythmia detection with convolutional neural networks," *arXiv preprint arXiv:1707.01836*, 2017.
- [143] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448-456.
- [144] A. Graves, A.-r. Mohamed, and G. Hinton, "Speech recognition with deep recurrent neural networks," in *Acoustics, Speech and Signal Processing (icassp), 2013 IEEE International Conference on*, 2013, pp. 6645-6649.
- [145] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, et al., "Deep speech 2: End-to-end speech recognition in english and mandarin," in *International Conference on Machine Learning*, 2016, pp. 173-182.
- [146] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt, "Sequential deep learning for human action recognition," in *International Workshop on Human Behavior Understanding*, 2011, pp. 29-39.

- [147] Z. Huang, W. Xu, and K. Yu, "Bidirectional LSTM-CRF models for sequence tagging," *arXiv preprint arXiv:1508.01991*, 2015.
- [148] O. Russakovsky, J. Deng, J. Krause, A. Berg, and L. Fei-Fei, "The ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012)," ed, 2012.
- [149] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1105, 2012.
- [150] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, *et al.*, "Large-scale image classification: fast feature extraction and svm training," in *CVPR 2011*, 2011, pp. 1689-1696.
- [151] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [152] 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.
- [153] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1492-1500.
- [154] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7132-7141.
- [155] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, pp. 3212-3232, 2019.
- [156] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580-587.
- [157] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440-1448.
- [158] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in Neural Information Processing Systems*, vol. 28, pp. 91-99, 2015.
- [159] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779-788.

- [160] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431-3440.
- [161] 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*, 2015, pp. 234-241.
- [162] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proceedings of the IEEE International Conference on Computer vision*, 2015, pp. 1520-1528.
- [163] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.
- [164] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, pp. 1240-1251, 2016.
- [165] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, *et al.*, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214-224, 2015.
- [166] 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*, 2016, pp. 424-432.
- [167] 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)*, 2016, pp. 565-571.
- [168] Q. Xia, Y. Yao, Z. Hu, and A. Hao, "Automatic 3D atrial segmentation from GE-MRIs using volumetric fully convolutional networks," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 211-220.
- [169] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2414-2423.
- [170] D. Ulyanov, V. Lebedev, A. Vedaldi, and V. S. Lempitsky, "Texture networks: Feed-forward synthesis of textures and stylized images," in *ICML*, 2016, p. 4.

- [171] A. Selim, M. Elgharib, and L. Doyle, "Painting style transfer for head portraits using convolutional neural networks," *ACM Transactions on Graphics (ToG)*, vol. 35, p. 129, 2016.
- [172] F. Luan, S. Paris, E. Shechtman, and K. Bala, "Deep photo style transfer," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4990-4998.
- [173] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang, "Universal style transfer via feature transforms," in *Advances in Neural Information Processing Systems*, 2017, pp. 386-396.
- [174] N. Mor, L. Wolf, A. Polyak, and Y. Taigman, "A universal music translation network," *arXiv preprint arXiv:1805.07848*, 2018.
- [175] M.-C. Yeh and S. Tang, "Improved style transfer by respecting inter-layer correlations," *arXiv preprint arXiv:1801.01933*, 2018.
- [176] G. R. Shaleka Agrawal, Kevin Jamart, James Kennelly, Jesse Ashton, Gregory Sands, Manuel Zarzoso, Jichao Zhao, "Structural basis of atrial arrhythmogenesis in metabolic syndrome," *Computing in Cardiology (CinC)*, 2019.
- [177] C. Chen, C. Ouyang, G. Tarroni, J. Schlemper, H. Qiu, W. Bai, *et al.*, "Unsupervised multi-modal style transfer for cardiac MR segmentation," *arXiv preprint arXiv:1908.07344*, 2019.
- [178] S. Ladavich and B. Ghoraani, "Rate-independent detection of atrial fibrillation by statistical modeling of atrial activity," *Biomedical Signal Processing and Control*, vol. 18, pp. 274-281, 2015.
- [179] H. Pürerfellner, E. Pokushalov, S. Sarkar, J. Koehler, R. Zhou, L. Urban, *et al.*, "P-wave evidence as a method for improving algorithm to detect atrial fibrillation in insertable cardiac monitors," *Heart Rhythm*, vol. 11, pp. 1575-1583, 2014.
- [180] X. Du, N. Rao, M. Qian, D. Liu, J. Li, W. Feng, *et al.*, "A novel method for real-time atrial fibrillation detection in electrocardiograms using multiple parameters," *Annals of Noninvasive Electrocardiology*, vol. 19, pp. 217-225, 2014.
- [181] R. Alcaraz, C. Vayá, R. Cervigón, C. Sánchez, and J. Rieta, "Wavelet sample entropy: A new approach to predict termination of atrial fibrillation," in *Computers in Cardiology, 2006*, 2006, pp. 597-600.
- [182] J. Ródenas, M. García, R. Alcaraz, and J. J. Rieta, "Wavelet entropy automatically detects episodes of atrial fibrillation from single-lead electrocardiograms," *Entropy*, vol. 17, pp. 6179-6199, 2015.

- [183] M. García, J. Ródenas, R. Alcaraz, and J. J. Rieta, "Application of the relative wavelet energy to heart rate independent detection of atrial fibrillation," *Computer Methods and Programs in Biomedicine*, vol. 131, pp. 157-168, 2016.
- [184] M. Carrara, L. Carozzi, T. J. Moss, M. de Pasquale, S. Cerutti, M. Ferrario, *et al.*, "Heart rate dynamics distinguish among atrial fibrillation, normal sinus rhythm and sinus rhythm with frequent ectopy," *Physiological Measurement*, vol. 36, p. 1873, 2015.
- [185] J. Park, S. Lee, and M. Jeon, "Atrial fibrillation detection by heart rate variability in Poincare plot," *Biomedical Engineering Online*, vol. 8, p. 38, 2009.
- [186] S. Sarkar, D. Ritscher, and R. Mehra, "A detector for a chronic implantable atrial tachyarrhythmia monitor," *IEEE Transactions on Biomedical Engineering*, vol. 55, pp. 1219-1224, 2008.
- [187] K. Tateno and L. Glass, "Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and Δ RR intervals," *Medical and Biological Engineering and Computing*, vol. 39, pp. 664-671, 2001.
- [188] D. T. Linker, "Accurate, automated detection of atrial fibrillation in ambulatory recordings," *Cardiovascular Engineering and Technology*, vol. 7, pp. 182-189, 2016.
- [189] C. Huang, S. Ye, H. Chen, D. Li, F. He, and Y. Tu, "A novel method for detection of the transition between atrial fibrillation and sinus rhythm," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 1113-1119, 2011.
- [190] R. Alcaraz, D. Abásolo, R. Hornero, and J. J. Rieta, "Optimal parameters study for sample entropy-based atrial fibrillation organization analysis," *Computer Methods and Programs in Biomedicine*, vol. 99, pp. 124-132, 2010.
- [191] D. DeMazumder, D. E. Lake, A. Cheng, T. J. Moss, E. Guallar, R. G. Weiss, *et al.*, "Dynamic analysis of cardiac rhythms for discriminating atrial fibrillation from lethal ventricular arrhythmias," *Circulation: Arrhythmia and Electrophysiology*, p. CIRCEP. 112.000034, 2013.
- [192] D. E. Lake and J. R. Moorman, "Accurate estimation of entropy in very short physiological time series: the problem of atrial fibrillation detection in implanted ventricular devices," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 300, pp. H319-H325, 2010.
- [193] R. COLLOCA, "Implementation and testing of atrial fibrillation detectors for a mobile phone application," 2013.

- [194] J. Zhao, Y. Yao, R. Shi, W. Huang, B. H. Smaill, and N. A. Lever, "Progressive modification of rotors in persistent atrial fibrillation by stepwise linear ablation," *HeartRhythm Case Reports*, vol. 1, pp. 22-26, 2015.
- [195] A. Petrénas, V. Marozas, and L. Sörnmo, "Low-complexity detection of atrial fibrillation in continuous long-term monitoring," *Computers in Biology and Medicine*, vol. 65, pp. 184-191, 2015.
- [196] S. Babaeizadeh, R. E. Gregg, E. D. Helfenbein, J. M. Lindauer, and S. H. Zhou, "Improvements in atrial fibrillation detection for real-time monitoring," *Journal of Electrocardiology*, vol. 42, pp. 522-526, 2009.
- [197] J. Oster and G. D. Clifford, "Impact of the presence of noise on RR interval-based atrial fibrillation detection," *Journal of Electrocardiology*, vol. 48, pp. 947-951, 2015.
- [198] R. Colloca, A. E. Johnson, L. Mainardi, and G. D. Clifford, "A support vector machine approach for reliable detection of atrial fibrillation events," in *Computing in Cardiology Conference (CinC), 2013*, 2013, pp. 1047-1050.
- [199] C. Cortes and V. Vapnik, "Support vector machine," *Machine Learning*, vol. 20, pp. 273-297, 1995.
- [200] R. Couceiro, J. Henriques, R. Paiva, M. Antunes, and P. Carvalho, "Physiologically motivated detection of Atrial Fibrillation," in *Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE*, 2017, pp. 1278-1281.
- [201] Q. Li, C. Rajagopalan, and G. D. Clifford, "Ventricular fibrillation and tachycardia classification using a machine learning approach," *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 1607-1613, 2014.
- [202] Q. Li, C. Rajagopalan, and G. D. Clifford, "A machine learning approach to multi-level ECG signal quality classification," *Computer Methods and Programs in Biomedicine*, vol. 117, pp. 435-447, 2014.
- [203] Y. Guo, H. Wang, H. Zhang, T. Liu, Z. Liang, Y. Xia, *et al.*, "Mobile photoplethysmographic technology to detect atrial fibrillation," *Journal of the American College of Cardiology*, vol. 74, pp. 2365-2375, 2019.
- [204] B. P. Yan, W. H. Lai, C. K. Chan, S. C. H. Chan, L. H. Chan, K. M. Lam, *et al.*, "Contact-free screening of atrial fibrillation by a smartphone using facial pulsatile photoplethysmographic signals," *Journal of the American Heart Association*, vol. 7, p. e008585, 2018.

- [205] Z. Xiong, V. V. Fedorov, X. Fu, E. Cheng, R. Macleod, and J. Zhao, "Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 38, pp. 515-524, 2018.
- [206] G. D. Clifford, C. Liu, B. Moody, H. L. Li-wei, I. Silva, Q. Li, *et al.*, "AF Classification from a short single lead ECG recording: the PhysioNet/Computing in Cardiology Challenge 2017," in *2017 Computing in Cardiology (CinC)*, 2017, pp. 1-4.
- [207] O. Sayadi, M. B. Shamsollahi, and G. D. Clifford, "Robust detection of premature ventricular contractions using a wave-based Bayesian framework," *IEEE Transactions on Biomedical Engineering*, vol. 57, pp. 353-362, 2009.
- [208] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [209] Z. Xiong, M. P. Nash, E. Cheng, V. V. Fedorov, M. K. Stiles, and J. Zhao, "ECG signal classification for the detection of cardiac arrhythmias using a convolutional recurrent neural network," *Physiological Measurement*, vol. 39, p. 094006, 2018.
- [210] Z. Xiong, M. K. Stiles, and J. Zhao, "Robust ECG signal classification for detection of atrial fibrillation using a novel neural network," in *2017 Computing in Cardiology (CinC)*, 2017, pp. 1-4.
- [211] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.
- [212] G. E. Dahl, T. N. Sainath, and G. E. Hinton, "Improving deep neural networks for LVCSR using rectified linear units and dropout," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 8609-8613.
- [213] 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, pp. 1929-1958, 2014.
- [214] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," in *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.
- [215] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, *et al.*, "Tensorflow: A system for large-scale machine learning," in *12th Symposium on Operating Systems Design and Implementation (16)*, 2016, pp. 265-283.

- [216] M. V. McConnell, M. P. Turakhia, R. A. Harrington, A. C. King, and E. A. Ashley, "Mobile health advances in physical activity, fitness, and atrial fibrillation: moving hearts," *Journal of the American College of Cardiology*, vol. 71, pp. 2691-2701, 2018.
- [217] A. N. Koshy, J. K. Sajeev, N. Nerlekar, A. J. Brown, K. Rajakariar, M. Zureik, *et al.*, "Smart watches for heart rate assessment in atrial arrhythmias," *International Journal of Cardiology*, vol. 266, pp. 124-127, 2018.
- [218] E.-S. Väliaho, P. Kuoppa, J. Lipponen, T. Martikainen, H. Jäntti, T. Rissanen, *et al.*, "Wrist band photoplethysmography in detection of individual pulses in atrial fibrillation and algorithm-based detection of atrial fibrillation," *EP Europace*, 2019.
- [219] K. Harjai, S. Mobarek, F. Abi-Samra, Y. Gilliland, N. Davison, K. Drake, *et al.*, "Mechanical dysfunction of the left atrium and the left atrial appendage following cardioversion of atrial fibrillation and its relation to total electrical energy used for cardioversion," *The American Journal of Cardiology*, vol. 81, pp. 1125-1129, 1998.
- [220] Z. I. Attia, P. A. Noseworthy, F. Lopez-Jimenez, S. J. Asirvatham, A. J. Deshmukh, B. J. Gersh, *et al.*, "An artificial intelligence-enabled ECG algorithm for the identification of patients with atrial fibrillation during sinus rhythm: a retrospective analysis of outcome prediction," *The Lancet*, vol. 394, pp. 861-867, 2019.
- [221] C. Tobon-Gomez, A. J. Geers, J. Peters, J. Weese, K. Pinto, R. Karim, *et al.*, "Benchmark for algorithms segmenting the left atrium from 3D CT and MRI datasets," *IEEE Transactions on Medical Imaging*, vol. 34, pp. 1460-1473, 2015.
- [222] L. Zhu, Y. Gao, A. Yezzi, and A. Tannenbaum, "Automatic segmentation of the left atrium from MR images via variational region growing with a moments-based shape prior," *IEEE Transactions on Image Processing*, vol. 22, pp. 5111-5122, 2013.
- [223] G. Veni, Z. Fu, S. P. Awate, and R. T. Whitaker, "Bayesian segmentation of atrium wall using globally-optimal graph cuts on 3D meshes," in *International Conference on Information Processing in Medical Imaging*, 2013, pp. 656-667.
- [224] Z. Xiong, Q. Xia, Z. Hu, N. Huang, C. Bian, Y. Zheng, *et al.*, "A global benchmark of algorithms for segmenting the left atrium from late gadolinium-enhanced cardiac magnetic resonance imaging," *Medical Image Analysis*, vol. 67, p. 101832, 2021.
- [225] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
- [226] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, *et al.*, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," *arXiv preprint arXiv:1603.04467*, 2016.

- [227] D. Ravanelli, E. C. dal Piaz, M. Centonze, G. Casagranda, M. Marini, M. Del Greco, *et al.*, "A novel skeleton based quantification and 3-D volumetric visualization of left atrium fibrosis using late gadolinium enhancement magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 33, pp. 566-576, 2014.
- [228] R. Karim, R. J. Housden, M. Balasubramaniam, Z. Chen, D. Perry, A. Uddin, *et al.*, "Evaluation of current algorithms for segmentation of scar tissue from late gadolinium enhancement cardiovascular magnetic resonance of the left atrium: an open-access grand challenge," *Journal of Cardiovascular Magnetic Resonance*, vol. 15, p. 105, 2013.
- [229] K. Nishida and S. Nattel, "Atrial fibrillation compendium historical context and detailed translational perspective on an important clinical problem," *Circulation Research*, vol. 114, pp. 1447-1452, 2014.
- [230] M. S. Dzeshka, G. Y. Lip, V. Snezhitskiy, and E. Shantsila, "Cardiac fibrosis in patients with atrial fibrillation: Mechanisms and clinical implications," *Journal of the American College of Cardiology*, vol. 66, pp. 943-959, 2015.
- [231] V. Mnih, N. Heess, and A. Graves, "Recurrent models of visual attention," in *Advances in Neural Information Processing Systems*, 2014, pp. 2204-2212.
- [232] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, *et al.*, "Ssd: Single shot multibox detector," in *European Conference on Computer Vision*, 2016, pp. 21-37.
- [233] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Advances in Neural Information Processing Systems*, 2014, pp. 568-576.
- [234] Q. Dou, H. Chen, L. Yu, J. Qin, and P.-A. Heng, "Multilevel contextual 3-d cnns for false positive reduction in pulmonary nodule detection," *IEEE Transactions on Biomedical Engineering*, vol. 64, pp. 1558-1567, 2017.
- [235] O. Bernard, A. Lalande, C. Zotti, F. Cervenansky, X. Yang, P.-A. Heng, *et al.*, "Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: Is the problem solved?," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 2514-2525, 2018.
- [236] G. Veni, S. Y. Elhabian, and R. T. Whitaker, "ShapeCut: Bayesian surface estimation using shape-driven graph," *Medical Image Analysis*, vol. 40, pp. 11-29, 2017.

- [237] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.
- [238] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 801-818.
- [239] X. Du, S. Yin, R. Tang, Y. Liu, Y. Song, Y. Zhang, *et al.*, "Segmentation and visualization of left atrium through a unified deep learning framework," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, pp. 589-600, 2020.
- [240] G. Yang, J. Chen, Z. Gao, S. Li, H. Ni, E. Angelini, *et al.*, "Simultaneous left atrium anatomy and scar segmentations via deep learning in multiview information with attention," *Future Generation Computer Systems*, vol. 107, pp. 215-228, 2020.
- [241] L. Li, F. Wu, G. Yang, L. Xu, T. Wong, R. Mohiaddin, *et al.*, "Atrial scar quantification via multi-scale CNN in the graph-cuts framework," *Medical image analysis*, vol. 60, p. 101595, 2020.
- [242] F. Knoll, A. Maier, and D. Rueckert, *Machine Learning for Medical Image Reconstruction: First International Workshop, MLMIR 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings* vol. 11074: Springer, 2018.
- [243] D. Zhu, *Multimodal Brain Image Analysis and Mathematical Foundations of Computational Anatomy: 4th International Workshop, MBIA 2019, and 7th International Workshop, MFCA 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 17, 2019, Proceedings*: Springer Nature, 2019.
- [244] M. Pop, *Statistical Atlases and Computational Models of the Heart. Atrial Segmentation and LV Quantification Challenges*: Springer, 2019.
- [245] V. Ulman, M. Maška, K. E. Magnusson, O. Ronneberger, C. Haubold, N. Harder, *et al.*, "An objective comparison of cell-tracking algorithms," *Nature Methods*, vol. 14, pp. 1141-1152, 2017.
- [246] C. Bian, X. Yang, J. Ma, S. Zheng, Y.-A. Liu, R. Nezafat, *et al.*, "Pyramid network with online hard example mining for accurate left atrium segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 237-245.

- [247] S. Vesal, N. Ravikumar, and A. Maier, "Dilated convolutions in neural networks for left atrial segmentation in 3D gadolinium enhanced-MRI," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 319-328.
- [248] X. Yang, N. Wang, Y. Wang, X. Wang, R. Nezafat, D. Ni, *et al.*, "Combating uncertainty with novel losses for automatic left atrium segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 246-254.
- [249] C. Li, Q. Tong, X. Liao, W. Si, Y. Sun, Q. Wang, *et al.*, "Attention based hierarchical aggregation network for 3D left atrial segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 255-264.
- [250] É. Puybareau, Z. Zhao, Y. Khoudli, E. Carlinet, Y. Xu, J. Lacotte, *et al.*, "Left atrial segmentation in a few seconds using fully convolutional network and transfer learning," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 339-347.
- [251] C. Chen, W. Bai, and D. Rueckert, "Multi-task learning for left atrial segmentation on GE-MRI," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 292-301.
- [252] S. Jia, A. Despinasse, Z. Wang, H. Delingette, X. Pennec, P. Jaïs, *et al.*, "Automatically segmenting the left atrium from cardiac images using successive 3D U-Nets and a contour loss," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 221-229.
- [253] Y. Liu, Y. Dai, C. Yan, and K. Wang, "Deep learning based method for left atrial segmentation in GE-MRI," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 311-318.
- [254] D. Borra, A. Masci, L. Esposito, A. Andalò, C. Fabbri, and C. Corsi, "A semantic-wise convolutional neural network approach for 3-D left atrium segmentation from late gadolinium enhanced magnetic resonance imaging," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 329-338.
- [255] C. de Vente, M. Veta, O. Razeghi, S. Niederer, J. Pluim, K. Rhode, *et al.*, "Convolutional neural networks for segmentation of the left atrium from gadolinium-enhancement MRI images," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 348-356.
- [256] C. J. Preetha, S. Haridasan, V. Abdi, and S. Engelhardt, "Segmentation of the left atrium from 3d gadolinium-enhanced MR images with convolutional neural

- networks," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 265-272.
- [257] M. Qiao, Y. Wang, R. J. van der Geest, and Q. Tao, "Fully automated left atrium cavity segmentation from 3D GE-MRI by multi-atlas selection and registration," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 230-236.
- [258] M. Nuñez-Garcia, X. Zhuang, G. Sanroma, L. Li, L. Xu, C. Butakoff, *et al.*, "Left atrial segmentation combining multi-atlas whole heart labeling and shape-based atlas selection," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 302-310.
- [259] N. Savioli, G. Montana, and P. Lamata, "V-FCNN: Volumetric fully convolution neural network for automatic atrial segmentation," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2018, pp. 273-281.
- [260] N. Huang, "https://www.dropbox.com/s/yvbj4352dax0q26/description_Ning_Huang.pdf?dl=0," 2018.
- [261] L. Xu, "https://www.dropbox.com/s/e6n2gmjz88ugwi/description_Lingchao%C2%A0Xu.pdf?dl=0," 2018.
- [262] Z. Xiong, V. V. Fedorov, X. Fu, E. Cheng, R. Macleod, and J. Zhao, "Fully automatic left atrium segmentation from late gadolinium enhanced magnetic resonance imaging using a dual fully convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 38, pp. 515-524, 2019.
- [263] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248-255.
- [264] B. J. Hansen, J. Zhao, and V. V. Fedorov, "Fibrosis and atrial fibrillation: computerized and optical mapping: a view into the human atria at submillimeter resolution," *JACC: Clinical Electrophysiology*, vol. 3, pp. 531-546, 2017.
- [265] M. Bishop, R. Rajani, G. Plank, N. Gaddum, G. Carr-White, M. Wright, *et al.*, "Three-dimensional atrial wall thickness maps to inform catheter ablation procedures for atrial fibrillation," *Europace*, vol. 18, pp. 376-383, 2016.
- [266] Y. Matsuda, Y. Toma, H. Ogawa, M. Matsuzaki, K. Katayama, T. Fujii, *et al.*, "Importance of left atrial function in patients with myocardial infarction," *Circulation*, vol. 67, pp. 566-571, 1983.

- [267] M. T. Maeder and D. M. Kaye, "Heart failure with normal left ventricular ejection fraction," *Journal of the American College of Cardiology*, vol. 53, pp. 905-918, 2009.
- [268] X. Zhuang, L. Li, C. Payer, D. Štern, M. Urschler, M. P. Heinrich, *et al.*, "Evaluation of algorithms for multi-modality whole heart segmentation: an open-access grand challenge," *Medical Image Analysis*, vol. 58, p. 101537, 2019.
- [269] G. Yang, T. Hua, and W. Xue, "Left ventricle full quantification challenge," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, 2019.
- [270] S. E. Petersen, N. Aung, M. M. Sanghvi, F. Zemrak, K. Fung, J. M. Paiva, *et al.*, "Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (CMR) in Caucasians from the UK Biobank population cohort," *Journal of Cardiovascular Magnetic Resonance*, vol. 19, pp. 1-19, 2017.
- [271] Z. Xiong, A. Nalar, K. Jamart, M. K. Stiles, V. V. Fedorov, and J. Zhao, "Fully automatic 3D bi-atria segmentation from late gadolinium-enhanced MRIs using double convolutional neural networks," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2019, pp. 63-71.
- [272] S. E. Petersen, P. M. Matthews, J. M. Francis, M. D. Robson, F. Zemrak, R. Boubertakh, *et al.*, "UK Biobank's cardiovascular magnetic resonance protocol," *Journal of Cardiovascular Magnetic Resonance*, vol. 18, pp. 1-7, 2015.
- [273] Y. Wang, Z. Xiong, A. Nalar, B. J. Hansen, S. Kharche, G. Seemann, *et al.*, "A robust computational framework for estimating 3D Bi-Atrial chamber wall thickness," *Computers in Biology and Medicine*, vol. 114, p. 103444, 2019.
- [274] W. Bai, M. Sinclair, G. Tarroni, O. Oktay, M. Rajchl, G. Vaillant, *et al.*, "Automated cardiovascular magnetic resonance image analysis with fully convolutional networks," *Journal of Cardiovascular Magnetic Resonance*, vol. 20, p. 65, 2018.
- [275] M. S. Weiss, "Modification of the Kolmogorov-Smirnov statistic for use with correlated data," *Journal of the American Statistical Association*, vol. 73, pp. 872-875, 1978.
- [276] E. M. Benito, A. Carlosena-Remirez, E. Guasch, S. Prat-González, R. J. Perea, R. Figueras, *et al.*, "Left atrial fibrosis quantification by late gadolinium-enhanced magnetic resonance: a new method to standardize the thresholds for reproducibility," *EJ Europace*, vol. 19, pp. 1272-1279, 2017.

- [277] G. Pontecorboli, R. M. Figueras i Ventura, A. Carlosena, E. Benito, S. Prat-Gonzales, L. Padeletti, *et al.*, "Use of delayed-enhancement magnetic resonance imaging for fibrosis detection in the atria: a review," *EP Europace*, vol. 19, pp. 180-189, 2017.
- [278] M. Sramko, P. Peichl, D. Wichterle, J. Tintera, J. Weichert, R. Maxian, *et al.*, "Clinical value of assessment of left atrial late gadolinium enhancement in patients undergoing ablation of atrial fibrillation," *International Journal of Cardiology*, vol. 179, pp. 351-357, 2015.
- [279] T. J. Badger, M. Daccarett, N. W. Akoum, Y. A. Adjei-Poku, N. S. Burgon, T. S. Haslam, *et al.*, "Evaluation of left atrial lesions after initial and repeat atrial fibrillation ablation: lessons learned from delayed-enhancement MRI in repeat ablation procedures," *Circulation: Arrhythmia and Electrophysiology*, vol. 3, pp. 249-259, 2010.
- [280] D. D. Spragg, I. Khurram, S. L. Zimmerman, H. Yarmohammadi, B. Barcelon, M. Needleman, *et al.*, "Initial experience with magnetic resonance imaging of atrial scar and co-registration with electroanatomic voltage mapping during atrial fibrillation: success and limitations," *Heart Rhythm*, vol. 9, pp. 2003-2009, 2012.
- [281] J. L. Harrison, C. Sohns, N. W. Linton, R. Karim, S. E. Williams, K. S. Rhode, *et al.*, "Repeat left atrial catheter ablation: cardiac magnetic resonance prediction of endocardial voltage and gaps in ablation lesion sets," *Circulation: Arrhythmia and Electrophysiology*, vol. 8, pp. 270-278, 2015.
- [282] X. Zhuang, J. Xu, X. Luo, C. Chen, C. Ouyang, D. Rueckert, *et al.*, "Cardiac segmentation on late gadolinium enhancement MRI: a benchmark study from multi-sequence cardiac MR segmentation challenge," *arXiv preprint arXiv:2006.12434*, 2020.
- [283] 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*, 2017, pp. 120-129.
- [284] H. C. Dittrich, L. A. Pearce, R. W. Asinger, R. McBride, R. Webel, M. Zabalgoitia, *et al.*, "Left atrial diameter in nonvalvular atrial fibrillation: an echocardiographic study," *American Heart Journal*, vol. 137, pp. 494-499, 1999.
- [285] M. Yamazaki, S. Mironov, C. Taravant, J. Brec, L. M. Vaquero, K. Bandaru, *et al.*, "Heterogeneous atrial wall thickness and stretch promote scroll waves anchoring during atrial fibrillation," *Cardiovascular Research*, vol. 94, pp. 48-57, 2012.

- [286] O.-S. Kwon, J. Lee, S. Lim, J.-W. Park, H.-J. Han, S.-H. Yang, *et al.*, "Accuracy and clinical feasibility of 3D-myocardial thickness map measured by cardiac computed tomogram," *International Journal of Arrhythmia*, vol. 21, pp. 1-11, 2020.
- [287] R. Kamali, J. Schroeder, E. DiBella, B. Steinberg, F. Han, D. J. Dosdall, *et al.*, "Reproducibility of clinical late gadolinium enhancement magnetic resonance imaging in detecting left atrial scar after atrial fibrillation ablation," *Journal of Cardiovascular Electrophysiology*, vol. 31, pp. 2824-2832, 2020.
- [288] Q. Huang, E. Z. Chen, H. Yu, Y. Guo, T. Chen, D. Metaxas, *et al.*, "Measure Anatomical Thickness from Cardiac MRI with Deep Neural Networks," in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2020, pp. 44-55.
- [289] I. J. Goodfellow, O. Vinyals, and A. M. Saxe, "Qualitatively characterizing neural network optimization problems," *arXiv preprint arXiv:1412.6544*, 2014.
- [290] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *Journal of Big Data*, vol. 6, pp. 1-54, 2019.
- [291] K. Kettering, G. F. Greil, M. Fenchel, U. Kramer, H.-J. Weig, M. Busch, *et al.*, "Catheter ablation of atrial fibrillation using the Navx-/Ensite-system and a CT-/MRI-guided approach," *Clinical research in cardiology*, vol. 98, pp. 285-296, 2009.
- [292] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun, "Deep learning for 3D point clouds: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [293] R. B. Rusu, Z. C. Marton, N. Blodow, M. Dolha, and M. Beetz, "Towards 3D point cloud based object maps for household environments," *Robotics and Autonomous Systems*, vol. 56, pp. 927-941, 2008.
- [294] F. Pomerleau, F. Colas, and R. Siegwart, "A review of point cloud registration algorithms for mobile robotics," *Foundations and Trends in Robotics*, vol. 4, pp. 1--104, 2015.
- [295] Y. H. Kim, S. A. Chen, S. Ernst, C. E. Guzman, S. Han, Z. Kalarus, *et al.*, "2019 APHRS expert consensus statement on three-dimensional mapping systems for tachycardia developed in collaboration with HRS, EHRA, and LAHRS," *Journal of Arrhythmia*, vol. 36, p. 215, 2020.
- [296] L. M. Ptaszek, B. Moon, G. Rozen, S. Mahapatra, and M. Mansour, "Novel automated point collection software facilitates rapid, high-density electroanatomical

- mapping with multiple catheter types," *Journal of Cardiovascular Electrophysiology*, vol. 29, pp. 186-195, 2018.
- [297] S. Rolf, G. Hindricks, P. Sommer, S. Richter, A. Arya, A. Bollmann, *et al.*, "Electroanatomical mapping of atrial fibrillation: Review of the current techniques and advances," *Journal of Atrial Fibrillation*, vol. 7, 2014.
- [298] A. Baram, M. Safran, A. Ben-Cohen, and H. Greenspan, "Left atria reconstruction from a series of sparse catheter paths using neural networks," in *International Workshop on Machine Learning for Medical Image Reconstruction*, 2018, pp. 138-146.
- [299] J. Zhang, X. Zhao, Z. Chen, and Z. Lu, "A review of deep learning-based semantic segmentation for point cloud," *IEEE Access*, vol. 7, pp. 179118-179133, 2019.
- [300] T. Yu, J. Meng, and J. Yuan, "Multi-view harmonized bilinear network for 3d object recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 186-194.
- [301] F. J. Lawin, M. Danelljan, P. Tosteberg, G. Bhat, F. S. Khan, and M. Felsberg, "Deep projective 3D semantic segmentation," in *International Conference on Computer Analysis of Images and Patterns*, 2017, pp. 95-107.
- [302] N. Audebert, B. Le Saux, and S. Lefèvre, "Semantic segmentation of earth observation data using multimodal and multi-scale deep networks," in *Asian Conference on Computer Vision*, 2016, pp. 180-196.
- [303] M. Tatarchenko, J. Park, V. Koltun, and Q.-Y. Zhou, "Tangent convolutions for dense prediction in 3d," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3887-3896.
- [304] A. Milioto, I. Vizzo, J. Behley, and C. Stachniss, "Rangenet++: Fast and accurate lidar semantic segmentation," in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 4213-4220.
- [305] L. Tchapmi, C. Choy, I. Armeni, J. Gwak, and S. Savarese, "Segcloud: Semantic segmentation of 3d point clouds," in *2017 International Conference on 3D Vision (3DV)*, 2017, pp. 537-547.
- [306] H.-Y. Meng, L. Gao, Y.-K. Lai, and D. Manocha, "Vv-net: Voxel vae net with group convolutions for point cloud segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 8500-8508.

- [307] B. Graham, M. Engelcke, and L. Van Der Maaten, "3D semantic segmentation with submanifold sparse convolutional networks," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018, pp. 9224-9232.
- [308] Y. Yao, L. Zheng, S. Zhang, D. S. He, K. Zhang, M. Tang, *et al.*, "Stepwise linear approach to catheter ablation of atrial fibrillation," *Heart Rhythm*, vol. 4, pp. 1497-1504, 2007.
- [309] H. Edelsbrunner and E. P. Mücke, "Three-dimensional alpha shapes," *ACM Transactions on Graphics (TOG)*, vol. 13, pp. 43-72, 1994.
- [310] B. Pathik, J. M. Kalman, T. Walters, P. Kuklik, J. Zhao, A. Madry, *et al.*, "Transient rotor activity during prolonged 3-dimensional phase mapping in human persistent atrial fibrillation," *JACC: Clinical Electrophysiology*, vol. 4, pp. 72-83, 2018.
- [311] S. Prabhu, B. T. Costello, A. J. Taylor, S. J. Gutman, A. Voskoboinik, A. J. McLellan, *et al.*, "Regression of diffuse ventricular fibrosis following restoration of sinus rhythm with catheter ablation in patients with atrial fibrillation and systolic dysfunction: a substudy of the CAMERA MRI trial," *JACC: Clinical Electrophysiology*, vol. 4, pp. 999-1007, 2018.
- [312] O. Oktay, E. Ferrante, K. Kamnitsas, M. Heinrich, W. Bai, J. Caballero, *et al.*, "Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, pp. 384-395, 2017.