Robust deep learning methods for accountable contrast-agent-free CMR imaging in clinical applications

Ricardo Alonso Gonzales Vera

Balliol College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy in Medical Sciences

Trinity 2024

To my exceptional, supportive, and loving family.

Acknowledgements

Years ago, I embarked on a journey driven by a desire to make a meaningful impact on others. This vision fueled an unrelenting passion for personal growth, pushing me forward on a path of discovery and self-improvement. As time passed, that vision began to take on clearer definition by combining my passion for technology, rooted in electrical engineering, with an altruistic sense aimed at improving the well-being of people, under a medical nuance. This thesis is the culmination of years of effort, and it would not have been possible without the guidance, encouragement, and unwavering support of many invaluable people along the way. Studying at Oxford, surrounded by dedicated and brilliant minds, has been an incredible privilege. I feel fortunate to have pursued my DPhil at the University of Oxford Centre for Clinical Magnetic Resonance Research (OCMR).

I would like to extend my deepest gratitude to my Oxford supervisors, whose guidance and support have been essential throughout this journey. **Prof. Stefan Piechnik**, thank you for taking a chance on me and welcoming me into your group. You taught me how to efficiently pursue impactful research, granting me the flexibility to explore multiple goals while encouraging me to strive for excellence. Your commitment to continuous improvement has been a true inspiration. **Prof. Vanessa Ferreira**, you showed me what it means to conduct science at Oxford. Your emphasis on rigorous research, meticulous planning, and exceptional communication has set a standard I continually aspire to meet. Your expertise in cardiology was invaluable in navigating this interdisciplinary field. **Prof. Qiang Zhang**, I am grateful for your constant support and strategic advice. Your insights on maximising the potential of my DPhil journey, along with our in-depth discussions about artificial intelligence, have profoundly influenced my academic development.

Throughout my DPhil journey, I have been fortunate to collaborate with and learn from many brilliant colleagues at OCMR, and I extend my heartfelt thanks for your insightful discussions and shared pursuit of scientific discovery. In particular,

I would like to acknowledge the members of the **Image Analysis Group**, both past and present: Dr. Evan Hann, Dr. Iulia Popescu, Dr. Matthew Burrage, Dr. Mayooran Shanmuganathan, Mrs. Elena Lukaschuk, Dr. İbrahim Altun, Dr. Konrad Werys, Dr. Katharine Thomas, and Mr. Daniel Atkinson. I am also grateful for the opportunity to mentor and work alongside excellent **visiting students**, including Ms. Leya Barrientos, Mr. Daniel Ibáñez, Mr. Daniel Menacho, Ms. Karina Vasquez, Ms. Ana Lucia Manrique, and Mr. Carlos Suárez.

Developing artificial intelligence models requires access to large datasets, and I am grateful to have collaborated with many researchers who provided the data supporting my projects throughout my DPhil. I would like to thank Prof. Stefan Neubauer, Prof. Betty Raman, Prof. Masliza Mahmod, Prof. Eylem Levelt, Dr. Rina Ariga, and Dr. Rohan Wijesurendra from **OCMR**; Prof. Wei-Ming Huang and Prof. Chun-Ho Yun from the **Mackay Memorial Hospital**; Prof. William Weintraub, Prof. Raymond Kwong, and Prof. Christopher Kramer from the **Hypertrophic Cardiomyopathy Registry**; Prof. Keith Channon from the **Oxford Acute Myocardial Infarction study**; Prof. Jean-Luc Balligand from the **Beta3-LVH clinical trial**; and once again my supervisors for facilitating access to these datasets, including data from the **UK Biobank**. I am thankful to Prof. Bartłomiej Papież, Dr. Felicia Seemann, Dr. Jérôme Lamy, and Dr. Per Arvidsson for their support in the development of methods and analyses crucial to my work. I also thank Mr. Yung Lee, Mr. Kit Yiu, Ms. Irem Çakir, Mr. Raghavskandhan Ramachandran, Mr. Gokul Parameswaran, and Mr. Eren Akmehmet, whose efforts in data labelling contributed enormously to my projects. Finally, I am endlessly grateful for the financial support I received from the Clarendon Scholarship, Radcliffe Department of Medicine Scholars Programme, and Balliol College for fully covering my DPhil training, and to many other institutions for granting me generous travel awards.

While my experience at Oxford has been shaped by many exceptional individuals, my journey began much earlier, starting with my alma mater, UTEC, and including Harvard University, Colorado School of Mines, IBM, the Research Experience for Peruvian Undergraduates Program, Yale University, the University of Alberta, and Lund University. Along the way, I was fortunate to have mentors who greatly

influenced my path. I would especially like to thank **Prof. Victor Murray**, my supervisor at UTEC, who inspired me to pursue an academic career from the start and gave me a clear picture of life in graduate school and the fascinating field of medical imaging. **Prof. Einar Heiberg**, my supervisor at Lund, not only demonstrated outstanding academic performance but also showed me how research in this field can directly impact clinical practice. **Prof. Dana Peters**, my supervisor at Yale, introduced me to the world of cardiovascular magnetic resonance and nurtured my growth in the field, preparing me for world-class research. I am deeply grateful for her support from the very beginning and for welcoming me into her group multiple times, an opportunity for which I am eternally thankful.

This journey would be incomplete without the unwavering support of my friends and family. Starting new chapters in life across the globe has brought invaluable opportunities to grow professionally, but it would have been far more difficult to navigate the personal challenges without the incredible people I am fortunate to call **friends**. Thank you for the unforgettable memories, for the dreams we turned into goals, for the parties we enjoyed, for the places we explored, and for more moments we will create. Amid all the numerous changes, the one constant I cherish most is my **family**. From the countless members of my close and extended family, whom I am proud to call relatives, I am especially grateful to my parents, Rolando and Lida, who have done everything in their power to raise me, setting no limits on their love and support. To my parents-in-law, Serhii and Svitlana, thank you for making Oxford truly feel like home. Above all, my deepest and heartful gratitude goes to my wife, Daria, for her boundless and unwavering love, for celebrating each milestone, lifting me in my lowest moments, and embracing every wild idea along the way.

Declaration

This thesis is entirely my own work, except where indicated otherwise. It has not been submitted, either wholly or substantially, for any other degree at this university or any other institution. The research presented in this work was conducted under the supervision of Prof. Stefan Piechnik, Prof. Vanessa Ferreira, and Prof. Qiang Zhang. All imaging data were retrospectively sourced from the UK Biobank (Chapter 2), the Oxford Centre for Clinical Magnetic Resonance Research (Chapters 3, 4, and 5), Mackay Memorial Hospital (Taiwan) (Chapters 3 and 5), the Hypertrophic Cardiomyopathy Registry (Chapters 4 and 5), and the Oxford Acute Myocardial Infarction Study (Chapter 4). Manual labels included in this thesis were obtained from previous research studies, as referenced in their respective chapters, and were specifically annotated for these projects by Dr. Iulia Popescu, Dr. Matthew Burrage, and Dr. Mayooran Shanmuganathan, who provided qualitative assessments in Chapter 2, and by Dr. Matthew Burrage and Dr. Ibrahim Altun, who provided manual contours in Chapters 3, 4, and 5.

Abstract

Cardiovascular magnetic resonance (CMR) T1 mapping is a crucial non-invasive imaging technique for quantifying myocardial tissue characteristics, providing valuable insights into a range of ischaemic and non-ischaemic cardiac conditions, including myocardial infarction, inflammation, infiltration, and fibrosis. Despite its clinical utility, the application of T1 mapping in routine practice faces several technical challenges, especially: motion artefacts that compromise image quality, limited visualisation tools for ease of interpretation, the need for generalisable segmentation of myocardial tissues, and the absence of reliable automated methods for annotating anatomical landmarks. This thesis addresses these challenges by proposing and validating novel deep learning approaches to enhance the clinical applicability of CMR T1 mapping.

The first goal of this thesis is to provide motion-corrected T1 maps by developing a deep learning-based method for correcting motion artefacts. A convolutional neural network architecture is employed to correct motion artefacts, to improve the quality of the T1 maps to be used for diagnosis. The proposed method is rigorously evaluated against traditional motion correction techniques, demonstrating superior performance in both simulated and real-world datasets.

The second goal is to provide an advanced visualisation technique for assessing T1 maps, including during pharmacologic stress through stress T1 reactivity maps. Leveraging vision transformers, this thesis presents a pathway for generating pixel-wise visualisations of stress-induced changes in left ventricular (LV) myocardial tissue properties, namely the delta T1 (dT1) map. This approach enhances the clinician's ability to interpret the effects of stress on the LV myocardium, providing a better visual display of the spatial extent of myocardial tissue changes during stress conditions.

The third goal is to provide a solution for automated segmentation of myocardial tissue in T1 mapping-based virtual native enhancement images. This is based on a quality control-driven deep ensemble method that combines the strengths of multiple deep learning models, to achieve robust and accurate segmentation, with an integrated quality control mechanism that flags potentially erroneous segmentations. This approach reduced inter-observer variability and improved the consistency of myocardial segmentation.

The fourth goal is to assist the automated segmentation of CMR short-axis slices into the American Heart Association 16-segment model for clinical applications across various CMR modalities. A generalist deep learning model was developed for the automated annotation of anatomical landmarks, specifically the anterior right ventricular insertion point and the LV centre point, using a dual-stage residual neural network framework, which accurately tracks these key landmarks. This work demonstrates many benefits for generalist training, validated against specialist models across diverse CMR datasets.

In conclusion, this thesis presents a comprehensive set of deep learning solutions aimed to enhance the clinical utility of CMR T1 mapping. By addressing the key challenges of motion correction, advanced visualisation, automated segmentation, and landmark annotation, the methods developed in this work pave the way towards more reliable, efficient, and standardised CMR imaging, which may ultimately translate into better clinical decision-making and patient outcomes.

List of Publications

Publications contributing to DPhil Thesis

1. **Gonzales RA**, Ibáñez DH, Hann E, Popescu IA, Burrage MK, Lee YP, Altun İ, Weintraub WS, Kwong RY, Kramer CM, Neubauer S, Ferreira VM, Zhang Q, Piechnik SK. Quality control-driven deep ensemble for accountable automated segmentation of cardiac magnetic resonance LGE and VNE images. *Frontiers in Cardiovascular Medicine*. 2023;10. doi:10.3389/fcvm.2023.1213290.

2. **Gonzales RA**, Zhang Q, Papież BW, Werys K, Lukaschuk E, Popescu IA, Burrage MK, Shanmuganathan M, Ferreira VM, Piechnik SK. MOCOnet: robust motion correction of cardiovascular magnetic resonance T1 mapping using convolutional neural networks. *Frontiers in Cardiovascular Medicine*. 2021;8. doi:10.3389/fcvm.2021.768245.

Abstracts contributing to DPhil Thesis

1. **Gonzales RA**, Burrage MK, Menacho D, Altun İ, Raman B, Ariga R, Wijesurendra RS, Mahmod M, Levelt E, Huang W-M, Yun C-H, Ferreira VM, Zhang Q, Piechnik SK. dT1 maps: a novel approach for visualising myocardial stress without gadolinium-based contrast agents. In: *SCMR 28th Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2025. doi:10.1016/j.jocmr.2024.101483.

2. **Gonzales RA**, Manrique AL, Burrage MK, Thomas KE, Altun İ, Huang W-H, Yun C-H, Zhang Q, Ferreira VM, Piechnik SK. Deep learning for automated insertion point annotation of CMR T1 maps. In: *2024 IEEE 21st International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2024. ORA

3. **Gonzales RA**, Ibáñez DH, Hann E, Popescu IA, Burrage MK, Lee YP, Altun İ, Weintraub WS, Kwong RY, Kramer CM, Neubauer S, Ferreira VM, Zhang Q, Piechnik SK. Quality control-driven framework for reliable automated segmentation of cardiac magnetic resonance LGE and VNE images. In: *International Workshop on Clinically-oriented and Responsible AI for Medical Data Analysis*. Springer International Publishing; 2023. ORA

4. **Gonzales RA**, Zhang Q, Burrage MK, Altun İ, Heiberg E, Peters DC, Neubauer S, Ferreira VM, Piechnik SK. Deep learning for automated insertion point annotation of CMR late gadolinium enhancement and virtual native enhancement images. In: *SCMR 26th Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2023. ORA

5. **Gonzales RA**, Zhang Q, Hann E, Ferreira VM, Piechnik SK. Quality control-driven artificial intelligence for reliable automatic segmentation of LGE images in clinical practice. In: *SCMR Virtual Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2022. ORA

6. **Gonzales RA**, Zhang Q, Papież BW, Werys K, Lukaschuk E, Popescu IA, Burrage MK, Shanmuganathan M, Ferreira VM, Piechnik SK. Fast and robust motion correction of cardiovascular magnetic resonance T1-mapping using data-driven convolutional neural networks for generalisability. In: *SCMR Virtual Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2022. ORA

Other publications during DPhil

1. Lamy J, **Gonzales RA**, Xiang J, Seemann F, Huber S, Steele J, Wieben O, Heiberg E, Peters DC. Tricuspid valve flow measurement using a deep learning framework for automated valve-tracking 2D phase contrast. *Magnetic Resonance in Medicine*. 2024. doi:10.1002/mrm.30163.

2. Zhang Q, Burrage MK, Shanmuganathan M, **Gonzales RA**, Lukaschuk E, Thomas KE, Mills R, Leal Pelado J, Nikolaidou C, Popescu IA, Lee YP, Zhang X, Dharmakumar R, Myerson SG, Rider O, Channon KM, Neubauer S, Piechnik SK, Ferreira VM. Artificial intelligence for contrast-free MRI: scar assessment in myocardial infarction using deep learning–based Virtual Native Enhancement. *Circulation*. 2022;146(20):1492-503. doi:10.1161/circulationaha.122.060137.

3. **Gonzales RA**, Lamy J, Seemann F, Heiberg E, Peters DC. Automated measurements of mitral and tricuspid annular dimensions in cardiovascular magnetic resonance. In: *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2022. p. 1-4. doi:10.1109/isbi52829.2022.9761595.

4. **Gonzales RA**, Seemann F, Lamy J, Mojibian H, Atar D, Erlinge D, Steding-Ehrenborg K, Arheden H, Hu C, Onofrey JA, Peters DC, Heiberg E. MVnet: automated time-resolved tracking of the mitral valve plane in CMR long-axis cine images with residual neural networks: a multi-center, multi-vendor study. *Journal of Cardiovascular Magnetic Resonance*. 2021;23(1):137. doi:10.1186/s12968-021-00824-2.

5. **Gonzales RA**, Lamy J, Seemann F, Heiberg E, Onofrey JA, Peters DC. TVnet: automated time-resolved tracking of the tricuspid valve plane in MRI long-axis cine images with a dual-stage deep learning pipeline. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Springer International Publishing; 2021. p. 567-76. doi:10.1007/978-3-030-87231-1_55.

6. Hann E, **Gonzales RA**, Popescu IA, Zhang Q, Ferreira VM, Piechnik SK. Ensemble of deep convolutional neural networks with Monte Carlo dropout sampling for automated image segmentation quality control and robust deep learning using small datasets. In: *Medical Image Understanding and Analysis*. Springer International Publishing; 2021. p. 280-93. doi:10.1007/978-3-030-80432-9_22.

7. Hann E, Popescu IA, Zhang Q, **Gonzales RA**, Barutçu A, Neubauer S, Ferreira VM, Piechnik SK. Deep neural network ensemble for on-the-fly quality control-driven segmentation of cardiac MRI T1 mapping. *Medical Image Analysis*. 2021;71:102029. doi:10.1016/j.media.2021.102029.

8. **Gonzales RA**, Seemann F, Lamy J, Arvidsson PM, Heiberg E, Murray V, Peters DC. Automated left atrial time-resolved segmentation in MRI long-axis cine images using active contours. *BMC Medical Imaging*. 2021;21(1):101. doi:10.1186/s12880-021-00630-3.

Other abstracts during DPhil

1. Barrientos L, **Gonzales RA**, Lamy J, Seemann F, Mojibian H, Heerdt P, Singh I, Peters DC. Deep learning-based measurement of isovolumic relaxation time from cardiovascular magnetic resonance long-axis cines: validation with pressure-derived IVRT. In: *SCMR 28th Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2025. doi:10.1016/j.jocmr.2024.101286.

2. Atkinson D, Popescu I, Thomas KE, Yun CH, Werys K, Burrage MK, Greenwood JP, Raman B, **Gonzales RA**, Kim YJ, Sanchez Panchuelo RM, Davies NP, Chow K, Neubauer S, Piechnik SK, Ferreira VM. Effects of software on myocardial T1 measurements: a 3 T multicentre single vendor phantom study. In: *SCMR 28th Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2025. doi:10.1016/j.jocmr.2024.101653.

3. **Gonzales RA**, Arvidsson PM, Lukaschuk E, Balligand J-L, Heiberg E, Peters DC, Zhang Q, Ferreira VM, Piechnik SK. Automated CMR index of left ventricular diastolic function (e'): a validation study against echocardiography in the large-scale Beta3-LVH trial. In: *CMR 2024 – The Global CMR Conference*. Society for Cardiovascular Magnetic Resonance; 2024. doi:10.1016/j.jocmr.2024.100955.

4. Shanmuganathan M, **Gonzales RA**, Burrage MK, Arvidsson PM, Banerjee A, Çakir I, Seemann F, Heiberg E, Peters DC, Zhang Q, Channon KM, Piechnik SK, Ferreira VM. Automated CMR index of left ventricular diastolic function post-acute myocardial infarction provides independent and incremental prediction of long-term prognosis when added to conventional indices. In: *CMR 2024 – The Global CMR Conference*. Society for Cardiovascular Magnetic Resonance; 2024. doi:10.1016/j.jocmr.2024.100268.

5. **Gonzales RA**, Lamy J, Thomas KE, Yiu K, Zhang Q, Shanmuganathan M, Heiberg E, Ferreira VM, Piechnik SK, Peters DC. TVnet: a deep-learning approach for enhanced right ventricular function analysis through tricuspid valve motion tracking. In: *International Workshop on Applications of Medical AI*. Springer International Publishing; 2023. ORA

6. Lamy J, Xiang J, Seemann F, **Gonzales RA**, Huber S, Steele J, Heiberg E, Peters DC. 2.5D Flow MRI: 2D phase-contrast of the tricuspid valvular flow with automated valve-tracking. In: *ISMRM & ISMRT Annual Meeting & Exhibition*. International Society for Magnetic Resonance in Medicine; 2023. doi:10.58530/2023/0930.

7. Lamy J, Seemann F, **Gonzales RA**, Xiang J, Heiberg E, Peters DC. 2.5D flow MRI of tricuspid valvular flow: an accurate automated valve-following phase-contrast approach. In: *SMRA 34th Annual International Conference*. Society for Magnetic Resonance Angiography; 2022. ORA

8. **Gonzales RA**, Lamy J, Thomas KE, Zhang Q, Shanmuganathan M, Heiberg E, Ferreira VM, Piechnik SK, Peters DC. TVnet: automated global analysis of tricuspid valve plane motion in CMR long-axis cines with residual neural networks for assessment of right ventricular function. In: *Artificial Intelligence in Cardiovascular Magnetic Resonance Imaging - A Joint Summit of the EACVI and SCMR*. European Heart Journal - Cardiovascular Imaging; 2022. doi: 10.1093/ehjci/jeac141.019.

9. Lamy J, Seemann F, **Gonzales RA**, Heiberg E, Peters DC. Towards 2D phase-contrast of tricuspid valve flow. In: *Joint Annual Meeting ISMRM-ESMRMB & ISMRT 31st Annual Meeting 2022*. International Society for Magnetic Resonance in Medicine; 2022. doi: 10.58530/2022/0798.

10. Zhang Q, Burrage MK, Shanmuganathan M, **Gonzales RA**, Nikolaidou C, Popescu IA, Lukaschuk E, Neubauer S, Ferreira VM, Piechnik SK. Development of deep learning virtual native enhancement for gadolinium-free myocardial infarction and viability assessment. In: *SCMR Virtual Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2022. ORA

11. Peters DC, Lamy J, Seemann F, Heiberg E, **Gonzales RA**. Isovolumic relaxation time and e' metrics evaluated by deep-learning analysis of long-axis cine: correlations to atrial pressure and fibrosis. In: *2021 ISMRM & SMRT Annual Meeting & Exhibition*. International Society for Magnetic Resonance in Medicine; 2021. ORA

12. **Gonzales RA**, Lamy J, Seemann F, Heiberg E, Onofrey J, Peters DC. Automated tracking of the tricuspid valve plane in long-axis cine images with a 2-step deep learning pipeline. In: *SCMR Virtual Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2021. ORA

13. **Gonzales RA**, Lamy J, Peters DC. Automated tracking of the tricuspid valve plane in long-axis cine images with a 2-step deep learning pipeline. In: *SMRA 32nd Annual International Conference*. Society for Magnetic Resonance Angiography; 2020. ORA

14. **Gonzales RA**, Onofrey J, Lamy J, Seemann F, Heiberg E, Peters DC. Time-resolved tracking of the atrioventricular plane displacement in long-axis cine images with residual neural networks. In: *2020 ISMRM & SMRT Virtual Conference & Exhibition*. International Society for Magnetic Resonance in Medicine; 2020. ORA

Recognitions and Awards during DPhil

1. **Travel Award**. Society for Cardiovascular Magnetic Resonance, 2025.
2. **Travel Award**. Balliol College and The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 2024.
3. **Editor's Recognition Award**. Radiology: Artificial Intelligence, 2024.
4. **Honorable Mention as Outstanding Reviewer**. Medical Image Computing and Computer Assisted Intervention Society, 2023.
5. **Travel Award**. British Heart Foundation – Centre of Research Excellence, 2023.
6. **Fellow (FSCMR)**. Society for Cardiovascular Magnetic Resonance, 2023.
7. **Travel Award**. Deep Learning for Medical Imaging School, 2023.
8. **RENACYT Researcher – Level V**. National Council for Science, Technology, and Innovation - Peru, 2022.
9. **Travel Award**. Oxford Centre for Personalised Medicine, 2022.
10. **Travel Award**. Medical Image Computing and Computer Assisted Intervention Society, 2021.
11. **Clarendon Scholar**. University of Oxford, 2020-2024.

Abbreviations

AHA	American Heart Association
AI	artificial intelligence
CAD	coronary artery disease
CMR	cardiovascular magnetic resonance
CNN	convolutional neural network
CVD	cardiovascular disease
DSC	Dice similarity coefficient
DVF	displacement vector field
GAN	generative adversarial network
GBCA	gadolinium-based contrast agent
HCM	hypertrophic cardiomyopathy
HCMR	Hypertrophic Cardiomyopathy Registry
ICC	intra-class correlation
IQR	interquartile range
IRW	inversion recovery-weighted
LGE	late gadolinium enhancement
LV	left ventricle
MI	myocardial infarction
MMH	Mackay Memorial Hospital (Taiwan)
MOCO	motion correction
OCMR	Oxford Centre for Clinical Magnetic Resonance Research
OxAMI	Oxford Acute Myocardial Infarction
PSIR	phase-sensitive inversion recovery
QC	quality control
QCD	quality control-driven
ROI	region of interest
RV	right ventricle
SD	standard deviation
ShMOLLI	Shortened Modified Look-Locker Inversion Recovery
VNE	virtual native enhancement

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Cardiovascular diseases

Cardiovascular diseases (CVDs) are the leading cause of death globally [1], posing a major health challenge in the United Kingdom [2], where they account for approximately a fourth of all deaths [3]. This translates to around 170,000 deaths each year in the UK alone, with a substantial financial burden on the National Health Service, estimated to exceed £10 billion annually [3]. CVDs encompass a range of conditions, including coronary artery disease (CAD), myocardial infarction (MI), and non-ischaemic diseases, such as myocarditis and cardiomyopathies. CAD, the most common type of CVD, involves the narrowing or blockage of coronary arteries due to atherosclerotic plaque buildup, leading to reduced blood flow and oxygen supply to the heart muscle [4]. MI, commonly known as a heart attack, occurs when a coronary artery blockage deprives part of the heart muscle of oxygen, causing tissue death and significant morbidity and mortality [5]. An example of non-ischaemic heart disease is hypertrophic cardiomyopathy (HCM), the most common genetic heart disease. It is characterised by abnormal thickening of the heart muscle, particularly the left ventricle (LV), which can obstruct LV forward flow and complicate the heart's ability to pump blood effectively, with a prevalence of about 1 in 500 individuals [6].

Advancements in medical research, including new therapies, minimally invasive surgeries, and advanced diagnostic tools, have greatly improved the treatment and management of CVDs, leading to better patient outcomes and lower mortality rates. Innovations in radiological imaging technologies, such as cardiovascular magnetic resonance (CMR) and computed tomography, have improved diagnostic accuracy and monitoring of CVDs. CMR is the imaging gold standard for cardiac structure and function, and offers reproducible non-invasive myocardial tissue characterisation, making it invaluable for assessing heart disease in both clinical and research settings [7]. Despite these advancements, challenges remain in the prevention, diagnosis, and treatment of heart disease. Ongoing research aims to further elucidate the molecular mechanisms of CVDs, identify novel biomarkers for early detection, and develop targeted therapies. The integration of artificial intelligence (AI) and machine learning in cardiovascular medicine holds promise for enhancing diagnostic accuracy, personalising treatment strategies, and optimising healthcare delivery [8].

1.2 Cardiovascular magnetic resonance

CMR imaging is the gold standard for the non-invasive assessment of cardiac function and morphology [7]. Unlike computed tomography, positron emission tomography and single-photon emission computed tomography imaging, CMR does not use ionising radiation. Compared to echo imaging, CMR offers superior spatial resolution and reproducibility in assessing cardiac structure, function and myocardial tissue characterisation, making it an attractive choice of cardiac imaging modality in many clinical and research settings [9, 10, 11].

The physics behind CMR imaging lies in the concept of nuclear spin, a quantum mechanical property of a given nucleus that determines how it interacts with magnetic fields. Nuclei like hydrogen possess a magnetic moment that, when exposed to an external magnetic field, tend to align in a way that is energetically more favourable, leading to a net magnetisation. This magnetisation can be manipulated using radio frequency pulses that alter its orientation, leading to the generation of a detectable electro-magnetic signals [12]. Among the major magnetic resonance parameters, the rate at which this signal from the individual nuclei desynchronises (T2, or spin-spin, relaxation time) and the rate at which the gross original magnetisation is restored (T1, or spin-lattice, relaxation time) vary depending on the tissue type, which is essential for differentiating between various cardiac conditions [13]. By adjusting the parameters of the pulse sequences, detailed images of the heart are produced, allowing for precise diagnoses and treatment monitoring [14].

Several specialised modalities within CMR are particularly valuable for assessing cardiac conditions [15]. Cine imaging, a dynamic modality that captures moving images of the heart, is crucial for evaluating cardiac motion during the heart-beat cycle, enabling precise measurements of changes in tissue volumes and myocardial strain [16]. Late gadolinium enhancement (LGE) imaging highlights areas of scar tissue or fibrosis by using gadolinium-based contrast agents (GBCAs) to highlight the areas with damaged myocardial tissue, which is essential for diagnosing myocardial infarction and other structural changes [17, 18]. T1 mapping, which quantifies the longitudinal (T1) relaxation time of myocardial tissue, offers insight into the myocardial tissue composition, aiding in the detection of focal as well as diffuse fibrosis and

other pathologies that LGE is unable to detect [19]. Lastly, perfusion imaging, also typically employing GBCA, assesses the blood flow to the myocardium during stress and rest conditions, crucial for identifying ischaemic regions and guiding interventions in CAD [20]. Each of these commonly-used CMR modalities plays a vital role in diagnosing different aspects of cardiovascular health, contributing to a comprehensive evaluation of heart disease.

1.3 T1 mapping

T1 mapping is recognised as a clinically useful contrast-free, directly quantitative myocardial tissue characterisation technique. Native T1 maps visualise the T1 relaxation time of myocardial tissues on a pixel-by pixel basis. This metric has characteristics of an excellent biomarker, having a narrow normal range and strong sensitivity to changes in tissue composition, especially water content [21]. T1 mapping has proven particularly useful for detecting myocardial oedema and inflammation, focal and diffuse fibrosis, infiltration, and other pathologies [21]. Clinicians can use T1 mapping to identify subtle variations in tissue properties that may indicate early, focal or diffuse disease, which may not be detectable by conventional imaging methods. Native T1 mapping is recognised for its non-invasive ability to assess myocardial conditions [22] such as cardiomyopathies and myocarditis without necessity for GBCA administration [23]. Another advantage of T1 mapping lies in applications such as stress T1 mapping to detect changes in myocardial blood flow and volume during stress compared to rest [24], and post-contrast T1 mapping to derive extracellular volume maps for improved detection of diffuse fibrosis [21]. Recent developments integrate native T1 data with native cine imaging to produce LGE-like visualisations of myocardial scar [25].

Stress T1 mapping is an advanced application of CMR that measures myocardial T1 relaxation times under either pharmacological stress, typically using agents that induce vasodilation [24], or during exercise [26]. This technique can detect changes in myocardial blood flow and volume that occur with increased cardiac demand, simulating exercise conditions [27]. Blunted responses in T1 reactivity during stress

indicate abnormalities in myocardial perfusion reserve [28, 29], offering a contrast-free method to assess ischaemic heart disease [26, 30, 31, 32, 33, 34], microvascular dysfunction [35], and depletion of coronary vasodilatory reserve [28].

Post-contrast T1 mapping involves the measurement of T1 times following the administration of GBCA [36]. These T1 maps can then be used to produce extracellular volume maps by comparing T1 relaxation times of myocardial tissue and blood before and after GBCA administration, adjusted for haematocrit levels. Extracellular volume mapping provides a way to detect the presence of diffuse myocardial fibrosis (after excluding confounders such as myocardial oedema and amyloidosis), which has prognostic implications especially in heart failure, but at a cost of GBCA and additional imaging time [21].

More recently, it was discovered that merging T1 maps with cine imaging can create LGE-like images as a powerful, non-invasive approach to myocardial tissue characterisation without the use of GBCA [25]. This integrated approach, called virtual native enhancement (VNE), allows for the visualisation of myocardial scars, fibrosis, and other pathologies with an accuracy comparable to conventional LGE, but without the associated risks of GBCA and a significantly shorter imaging protocol [37]. The resulting images offer significantly improved image quality compared to conventional LGE, with high spatial resolution and tissue contrast, improving the detection of subtle myocardial changes and aiding in accurate disease staging. Ultimately, VNE may pave way towards safer, faster, and more cost-effective CMR imaging, making it particularly advantageous for patients with contraindications to GBCA and facilitating more frequent monitoring of disease progression or response to therapy, without exposure to radiation or contrast agents [38].

1.4 CMR image analysis

Building on the advanced capabilities of CMR imaging, the transformation of raw imaging data into actionable clinical insights is essential. Initially, the analysis of CMR images relied heavily on manual image analysis by human experts, a process that, while valuable, is time-consuming and subject to substantial variability and

inter-observer bias [39]. As the volume and complexity of imaging data have grown, there has been a clear need to develop automated methods to enhance speed and efficiency, as well as reproducibility and accuracy in CMR image analysis. This evolution has given rise to the field of CMR image analysis, which leverages sophisticated image processing techniques to systematically analyse and evaluate CMR data [40]. The major tasks in CMR image analysis can be divided into image classification, regression, detection, segmentation, registration, and generation [41].

Classification and regression are fundamental tasks in image analysis. Classification categorises input data into predefined classes based on their features, usually implemented using models that assign probabilities to each class, selecting the most probable as the output. Regression, in contrast, predicts continuous outcomes based on input variables by modelling the relationship between a dependent variable and one or more independent variables. For instance, classification models developed for short-axis cine images can estimate the absence or presence of a myocardial scar [42] and distinguish hypertrophic cardiomyopathy from Fabry cardiomyopathy [43], directly from the imaging data. With the same type of input data, regression models can directly estimate the ejection fraction of the LV [44] and right ventricle (RV) [45].

The remaining tasks are applications that build upon the principles of classification and regression. Object detection combines classification and localisation to identify and delineate objects within images, typically in a bounding box or in Cartesian coordinates—for example, tracking the mitral [46] and tricuspid valves [47] in long-axis cines to measure valvular motion. Segmentation divides an image into predesignated regions of interest (ROIs) by classifying each pixel based on its characteristics, simplifying the representation—for instance, segmenting the chambers [48] or epicardial adipose tissue [49] to measure volumes. Registration involves aligning or overlaying two or more images to achieve exact tissue correspondence between them, such as pixel-wise alignment of the LV myocardium in cine images to measure deformation or strain [50]. Image generation modifies images from one domain to another to enhance specific features or adapt images to standardised conditions, like synthesising missing or corrupted slices based on the remaining data to improve analysis accuracy [51].

Historically, medical image analysis relied heavily on classical image processing methods [52], within the realm of the broad AI field. The initial approaches were primarily rule-based and used basic signal processing techniques and geometric models to enhance, detect, and classify anatomical structures within medical images. Thresholding [53], edge detection algorithms [54], and region-growing techniques [55] were commonly applied to delineate organ boundaries and identify lesions. Registration of images across different modalities or temporal stages involved rigid or affine transformations based on manually selected landmarks or iteratively matched regions [56]. Although effective to some extent, these methods were limited by their dependency on human input for parameter selection and their inability to adapt to the high variability in medical images across different patients [57].

The incorporation of machine learning into cardiac imaging marked a significant advancement over traditional image processing techniques [58]. Initially, machine learning in cardiac imaging employed shallow learning algorithms such as decision trees [59], linear classifiers [60], and support vector machines [61], which required careful engineering and selection of image features by domain experts. These models were trained on handcrafted features extracted from the images, such as texture patterns, edge statistics, and geometric properties, which were used to classify regions of interest or predict disease states. These methods improved the reproducibility of quantitative analyses and helped standardise measurements across studies and clinical sites [62]. However, the performance of these machine learning approaches was closely tied to the quality and relevance of the manually designed features, which often limited their effectiveness in complex imaging scenarios [63].

The transition from traditional machine learning to deep learning in CMR represented a paradigm shift driven by the need for more robust and generalisable AI-based solutions for image analysis [40]. Deep learning models, particularly convolutional neural networks (CNNs), emerged as powerful tools capable of learning directly spatial features from raw image data, thus bypassing the need for manual feature extraction [64]. These networks automatically learn hierarchical features that are more adaptive and discriminative for various tasks, such as segmenting complex anatomical structures [65] and reconstructing images from undersampled data [66]. The appli-

cation of deep learning in CMR has enabled more accurate and detailed analyses of cardiac function and structure [67], enhancing the ability to detect diseases at earlier stages and to assess myocardial tissue characteristics more precisely, with a higher emphasis on cine imaging [68]. Deep learning has facilitated the generation of synthetic contrast-enhanced images from native scans [25, 37], providing a significant innovative step in the technological advancement of CMR. More recently, vision transformers [69], employing an architecture distinct from traditional CNNs, have shown promising results that could potentially surpass the performance of CNNs, setting a new benchmark in the development of AI for medical imaging [70]. This shift towards transformers is driven by their ability to capture long-range dependencies within the data, which enhances their capability to capture complex patterns and relationships that are critical for accurate and detailed analysis of cardiac images.

Training deep learning models for CMR involves several stages, starting with the compilation of a large and diverse dataset of annotated images [40]. These datasets must encompass a wide range of pathologies, anatomical variations, and imaging artefacts, to ensure that the models are robust and generalisable [71]. Once assembled, the data are pre-processed to normalise intensity ranges, align images, and augment the dataset through techniques such as rotation, scaling, and elastic deformations to increase the model's exposure to various imaging conditions. The neural networks are then trained using backpropagation and a suitable optimisation algorithm to minimise a loss function that quantifies the difference between the predicted outputs and the ground truth. During training, model parameters are adjusted to reduce prediction errors. Techniques such as dropout, batch normalisation, and early stopping, are employed to prevent overfitting. Finally, the trained model is evaluated on independent test datasets to assess its performance, and further fine-tuning is conducted if necessary to optimise its accuracy and efficiency for clinical deployment [72].

1.5 Goals

This thesis aims to develop novel data-driven approaches to support robust automated CMR T1 mapping processing, addressing key challenges in advanced visualisation and image post-processing (Figure 1.1). The work utilises AI capabilities by training deep learning models to perform on par with expert human image analysts. These developments may enable the widespread clinical use of CMR T1 mapping, increasing the consistency and reliability of diagnoses while improving patient throughput. This would ultimately reduce costs, scan times, and the time burden of routine image analysis in clinical reporting.

Specifically, the following goals are explored (Figure 1.1 1 2 3 4):

1. Devise a CNN-based model designed to correct motion artefacts in T1 maps (Figure 1.1 1). Motion artefacts and misalignment during scans can lead to errors in the quantitative assessment of myocardial tissues, potentially resulting in misdiagnoses or inaccuracies in disease characterisation. This project, described in Chapter 2, co-registers the T1-weighted images within a T1-mapping sequence to improve the precision of the reconstructed T1 maps, mitigating the impact of cardiac motion.

2. Design a vision transformer-based model for automated, pixel-wise visualisation of myocardial blood volume changes on T1 maps during stress conditions compared to resting conditions (Figure 1.1 2). Currently, the lack of automated tools for pixel-wise registration and visualisation complicates the analysis of stress and rest T1 maps, making it time-consuming and prone to variability. This project, described in Chapter 3, registers a stress T1 map to its paired rest T1 map enabling a pixel-wise percentage difference of stress T1 reactivity, or delta T1 (dT1) map, without the need for manual measurements, to enhance diagnostic accuracy and facilitate the work towards broader adoption of stress T1 mapping.

3. Develop a quality control-driven deep ensemble method to segment the myocardium in T1 map-derived VNE images, also applicable to LGE images (Figure 1.1 3). The absence of automated segmentation tools for VNE images necessitates manual analysis, reducing efficiency and introducing variability. This project, described in Chapter 4, exploits the differences between a group of CNN models to derive an optimal myocardial segmentation with a confidence metric, to enhance the precision, reliability, and efficiency of CMR image analysis.

4. Automate the detection of key anatomical landmarks necessary for detailed segmental analysis of the myocardium across all T1 map-based data (native, stress, post-contrast, and VNE), and also applicable to LGE images (Figure 1.1 4). Standardised segmentation, according to the American Heart Association (AHA) 16-segment model, requires precise annotation of landmarks such as the anterior RV insertion point and the LV centre point. The lack of automated methods for landmark annotation leads to the need for manual analysis, introducing variability and reducing efficiency. This project, described in Chapter 5, leverages multi-modality information from different sequences to ensure robust performance, to facilitate standardised segmentation and accurate myocardial quantification across all modalities.

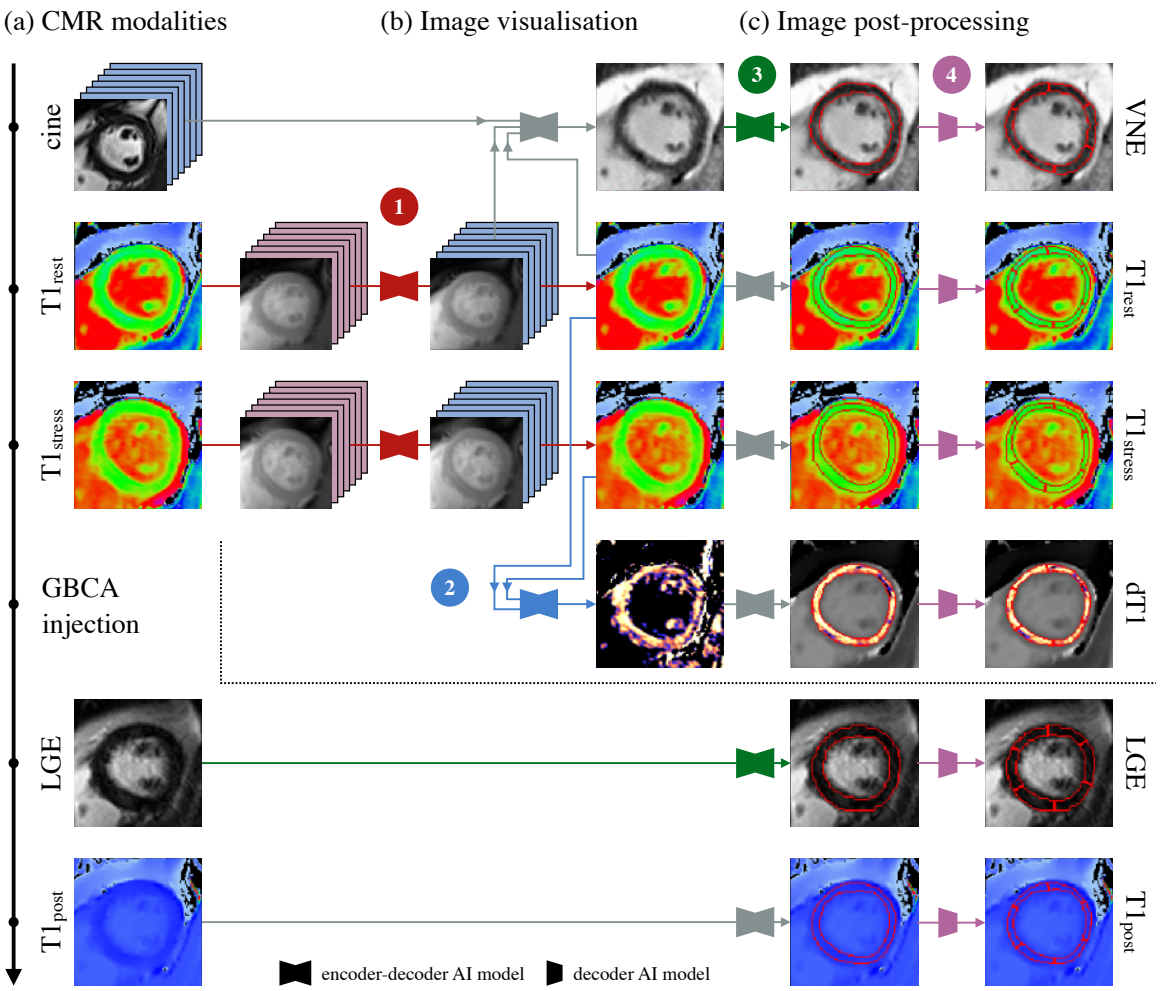


Figure 1.1: Illustration of the proposed deep learning framework for enhanced CMR T1 mapping analysis. (a) CMR modalities used: cine sequences, rest and stress T1 maps, LGE images, and post-contrast T1 maps. (b) Image visualisation steps: 1 motion correction of T1 mapping images; 2 generation of stress T1 reactivity maps (dT1 maps). (c) Image post-processing steps: 3 automated myocardial segmentation using a quality control-driven deep ensemble method; 4 automated annotation of key anatomical landmarks for standardised segmentation and accurate myocardial quantification across all modalities. Elements in grey indicate existing methods developed by our research group, outside the scope of this thesis.

Chapter 2

Image quality: motion correction of T1 mapping with convolutional neural networks

Chapter 1 provided an overview of the challenges and advancements in CMR T1 mapping, establishing the context for the need to improve image quality and automated analysis. Building on this foundation, this Chapter focuses on the correction of motion artefacts in T1 maps using CNNs. Motion artefacts in pixel-wise T1 maps significantly impair the accuracy of myocardial tissue characterisation, which is critical for effective diagnosis and treatment planning. A novel deep learning-based approach is introduced to address these artefacts, aiming to enhance the reliability of T1 maps. The application of CNNs allows for the correction of motion distortions, ensuring that the resultant images maintain high fidelity to the original anatomical structures. This Chapter's advancements in image quality serve as a crucial prerequisite for the subsequent Chapters, where the improved data will be used for more precise visualisation, segmentation, and landmark annotation discussed in Chapters 3, 4, and 5, respectively. By resolving the motion artefacts, the groundwork is laid for the automated and accurate post-processing techniques discussed later.

The work in this Chapter has been published [73] and presented [74] in:

73. **Gonzales RA**, Zhang Q, Papież BW, Werys K, Lukaschuk E, Popescu IA, Burrage MK, Shanmuganathan M, Ferreira VM, Piechnik SK. MOCOnet: robust motion correction of cardiovascular magnetic resonance T1 mapping using convolutional neural networks. *Frontiers in Cardiovascular Medicine*. 2021;8. doi:10.3389/fcvm.2021.768245.

74. **Gonzales RA**, Zhang Q, Papież BW, Werys K, Lukaschuk E, Popescu IA, Burrage MK, Shanmuganathan M, Ferreira VM, Piechnik SK. Fast and robust motion correction of cardiovascular magnetic resonance T1-mapping using data-driven convolutional neural networks for generalisability. In: *SCMR Virtual Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2022. ORA

2.1 Background

T1 mapping involves fitting exponential recovery curves to individual pixels in multiple T1-weighted images. Developed from the original Look-Locker spectroscopic method [75], modern techniques use electrocardiographic gating across several heartbeats to acquire intermittent images [21]. The shortened modified Look-Locker inversion recovery (ShMOLLI) method, which requires only nine heartbeats, enhances precision and reproducibility [76]. However, despite reducing cardiac motion impact by acquiring images at the same cardiac phase, respiratory motion remains a significant challenge [77], leading to potential errors in T1 estimation and misdiagnoses if not properly corrected [76].

Retrospective motion correction (MOCO) can substantially improve the reliability and clinical utility of T1 mapping by aligning T1-weighted images before reconstruction [78]. This alignment process addresses variations in image contrast and signal nulling that occur at different inversion times. Model-driven registration methods for MOCO [79, 80, 81, 82] have shown promise in overcoming these challenges. However, careful visual inspection is still necessary to identify any uncorrected residual motion or distortions from failed corrections [83]. While visual assessment remains the clinical standard for CMR interpretation [84], it is labour-intensive, prone to error from inconsistency and operator fatigue, and slows down clinical workflows when processing large volumes of images.

The emergence of deep learning, particularly CNNs, has revolutionised image processing by replacing predefined, hand-crafted rules with automated learning from large datasets. These approaches have been rapidly adopted in CMR, providing fast, consistent, and accurate image segmentation and analysis pipelines, significantly reducing the manual labour required from physicians [40]. Deep learning also shows potential for enhancing clinical image registration [85], offering improvements in accuracy, efficiency, and quality control, with potential to address the unmet needs of MOCO in T1 maps.

This Chapter introduces MOCOnet, a novel deep learning approach for myocardial motion correction in CMR T1 mapping, using data from the UK Biobank [86]. MOCOnet employs an encoder-decoder architecture with warping layers to facilitate

learning deformations in a coarse-to-fine manner. By processing a set of T1-weighted images, MOCOnet can efficiently predict the necessary spatial deformations to correct motion artefacts. The model's performance was evaluated against a well-validated multi-modal image registration method, with multiple blinded expert observers assessing the effectiveness of the motion correction.

2.2 Methods

2.2.1 Cardiac T1 mapping and motion artefact

The ShMOLLI technique for cardiac T1 mapping involves calculating T1 values by fitting exponential recovery curves to 7 inversion recovery-weighted (IRW) images (Figure 2.1a), captured at various inversion times during a short nine-heartbeat breathhold [76]. The reconstructed T1 map (Figure 2.1b) allows for pixel-level quantification of T1 values. Additionally, an associated R2 map (Figure 2.1c), representing the coefficient of explained variance, monitors the quality of the curve fitting relative to a mono-exponential T1 relaxation recovery model. Ideally, this results in a uniformly white appearance in the R2 map for relevant regions, but motion in the IRW images (Figure 2.1d, arrowed) can decrease T1 map interpretability (Figure 2.1e, arrowed), as indicated by dark bands in the R2 map (Figure 2.1f, arrowed). Besides motion artefacts, the R2 map also detects issues like off-resonance effects, fat inclusion, and mistriggering [87, 88].

2.2.2 Non-rigid registration approach

A T1 map affected by motion artefacts consists of 7 unaligned IRW images. Consequently, to correct these artefacts, the IRW images need to be aligned. The artefact can be modelled as a deformation of the aligned IRW images using a displacement vector field (DVF). Thus, the non-rigid registration problem is addressed by estimating the inverse DVF for a set of unaligned IRW images.

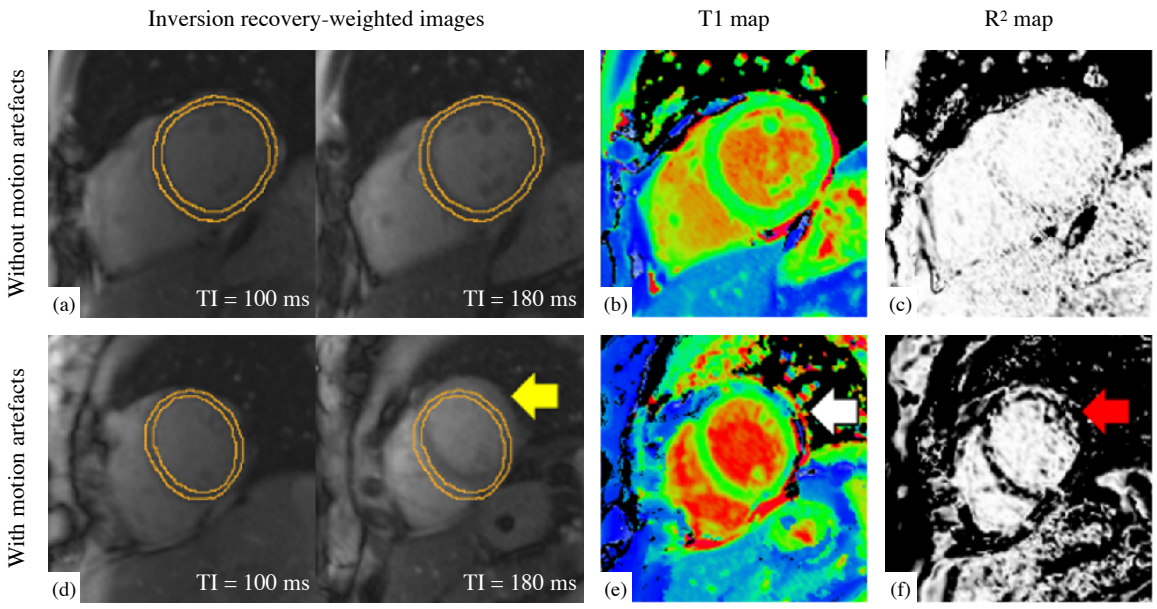


Figure 2.1: Examples of T1 maps with good quality (top row) contrasted with those affected by motion artefacts (bottom row). (a)(d) Two examples out of seven of inversion-recovery weighted (IRW) images required for T1 map reconstruction are shown, time-stamped with their corresponding inversion times (TI) and overlaid by identical manual myocardial contours for identifying motion. (b)(e) ShMOLLI T1 maps. (c)(f) R^2 quality control maps. A good quality T1 map is indicated by (a) myocardium in same position and (c) 'all white' in the left ventricular myocardium indicating high T1 fitting confidence. A T1 map with motion artefact is evident by misalignment in IRW images (yellow arrow), suspicious features in T1 map (white arrow) and dark bands in R^2 map in the left ventricular myocardium as evidence of poor T1 fitting (red arrow). *Reproduced from Gonzales et al (2021) Frontiers in Cardiovascular Medicine 8* [73], *published under CC-BY 4.0.*

2.2.3 Multi-scale registration neural network

The proposed model corrects a T1 map by estimating the inverse DVF for each of the 7 IRW images, enabling non-rigid registration before T1 map reconstruction. The multi-scale registration CNN (Figure 2.2) uses an encoder-decoder U-Net-like architecture [89] with warping layers [90] between the contracting and expansive paths at each scale. Feature maps are downsampled using a series of 3×3 convolutional layers followed by batch normalisation, a leaky rectified linear unit, and a max pooling layer, and upsampled with a transposed convolutional layer. Warping layers accelerate training by imposing a multi-scale loss function and improve registration accuracy by addressing motion starting from coarse levels, refining it at higher resolutions.

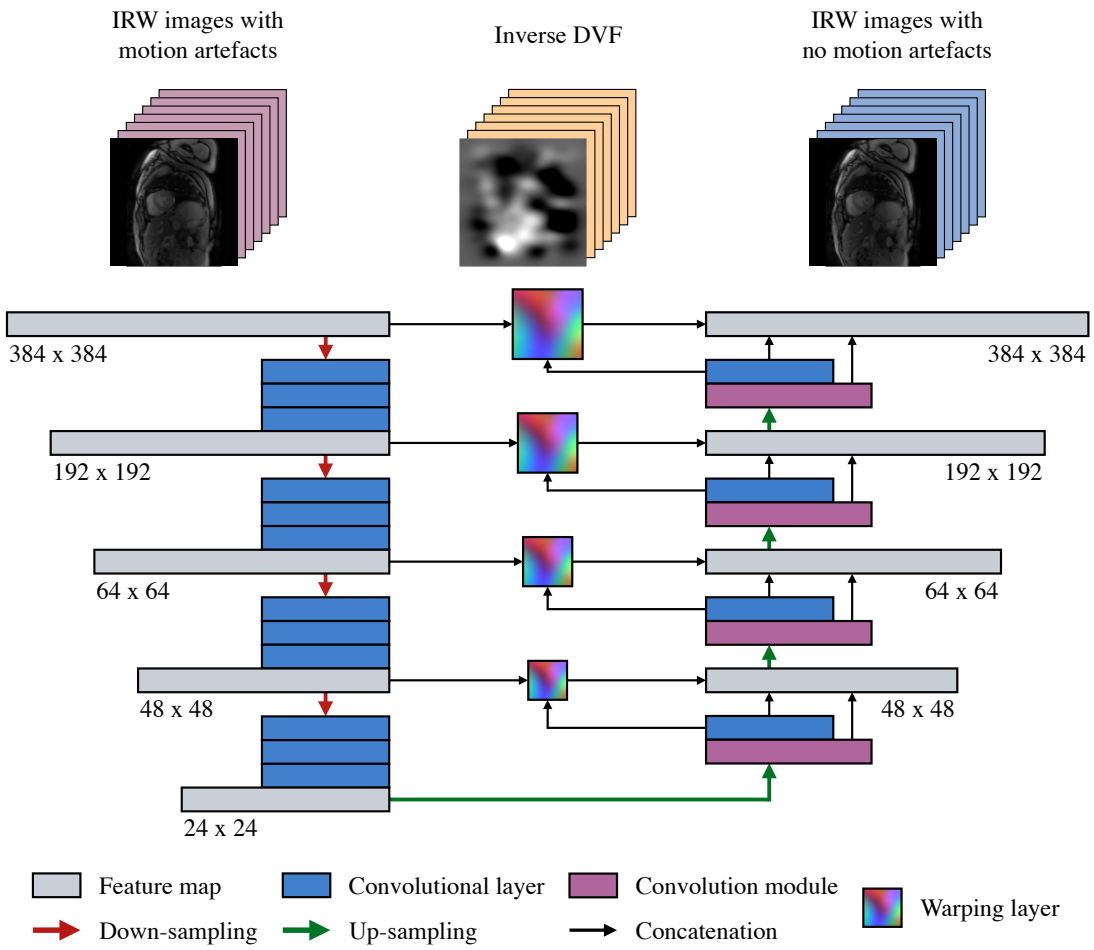


Figure 2.2: Structure of the proposed motion correction convolutional neural network (MOCOnet). A stack of 7 inversion recovery-weighted (IRW) images is input into the encoder-decoder structure on a per-channel basis. The warping layers estimate the optical flow from all the channels in a coarse-to-fine manner at each scale. The last warping layer generates the inverse distance vector field (DVF), i.e., the deformation required to correct the motion artefacts, in a groupwise manner. *Reproduced from Gonzales et al (2021) Frontiers in Cardiovascular Medicine 8* [73], *published under CC-BY 4.0.*

The IRW images undergo a sequence of convolution and downsampling operations to generate multi-scale features for each channel. These features, ranging from low to high resolution, are then used in convolution modules to produce DVFs. Each module takes input from the previous step, the DVF from the prior scale, and the warped features from the downsampling stage. Warping at each of the four scales allows residual motion information to be corrected and refined progressively. Consequently, the neural network generates DVFs in a coarse-to-fine manner, adding detail at higher resolutions with a loss function at each scale to guide learning.

2.2.4 Imaging data and inclusion criteria

The imaging dataset comprised over 5,000 CMR native T1 maps from the UK Biobank Imaging Component [86], acquired in the mid-ventricular short-axis view using the ShMOLLI T1 mapping sequence [76]. For quality control, a trained human operator with over a decade of experience in CMR image analysis assessed artefacts in the LV myocardium across the 7 IRW images for each T1 map. A total of 1,536 T1 maps were classified as high quality with no artefacts. Data with mild to severe motion or other artefacts were excluded from the training dataset. This strict quality control ensured the neural network learned to align images accurately without distractions from residual motion artefacts in the training data.

2.2.5 Training procedure

The quality-controlled images were used to create a training dataset, reserving 10% for validation. Artificial DVFs were generated as described previously [91] and applied to the IRW images without motion artefacts to simulate random non-rigid motion without segmentations [91]. Specifically, 7 DVFs with random parameters preserving anatomical topology were generated. The mean displacement value at each pixel was removed from all 7 DVFs to focus on relative displacement between images. These DVFs were applied to the IRW images, creating deformed IRW images. The model was trained to predict 7 inverse DVFs from the 7 deformed IRW images, with synthetic inverse DVFs serving as ground truth (Figure 2.3a).

2.2.6 Testing procedure

Once trained, MOCOnet processes a set of 7 IRW images, with or without motion artefacts, to estimate the necessary deformations for correcting motion (Figure 2.3b), without needing ground truth. The T1 map is then reconstructed offline using the corrected images with an open-source library for CMR parametric mapping [92].

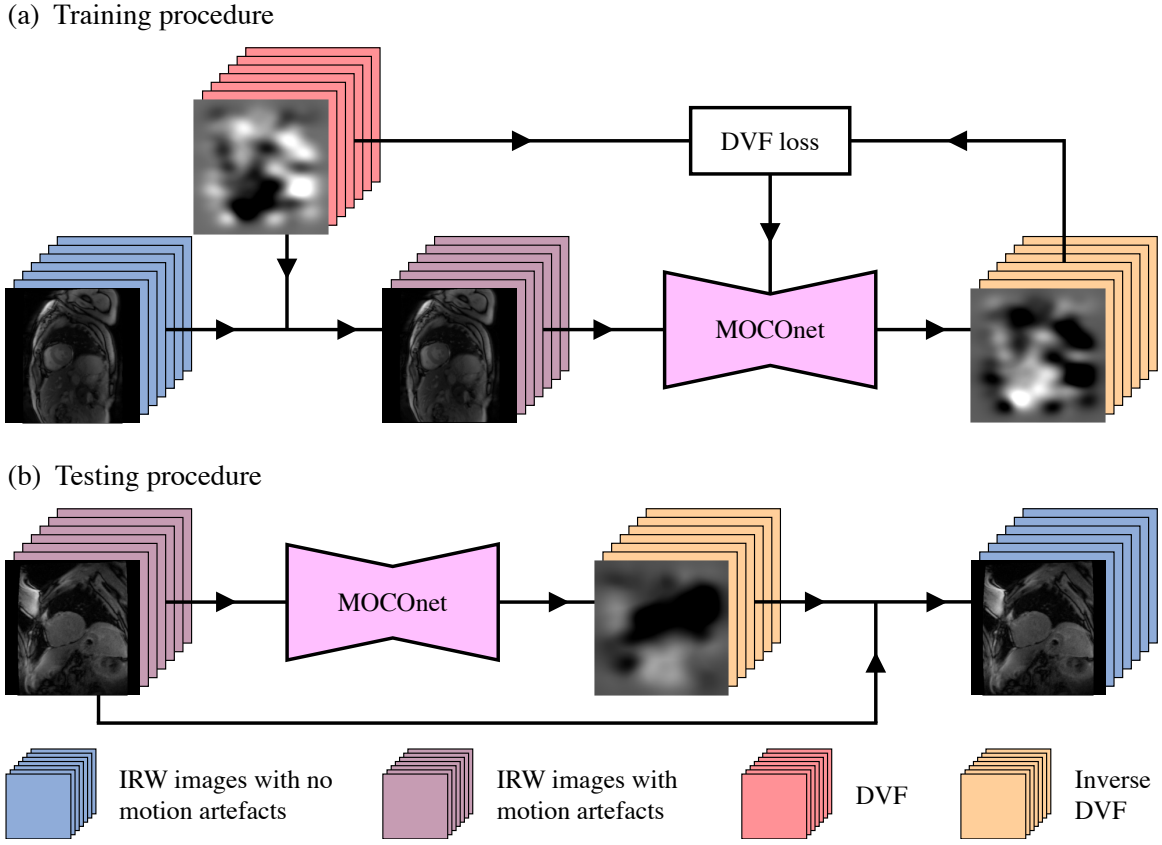


Figure 2.3: Development workflow of the proposed motion correction convolutional neural network (MOCOnet) for myocardial ShMOLLI T1 mapping. (a) MOCOnet was trained on 1,536 sets of 7 inversion recovery-weighted (IRW) images with no motion artefacts which were synthetically deformed with displacement vector fields (DVFs), to predict the inverse DVF required to correct the motion. (b) MOCOnet was tested on 200 T1 maps with a varied degree of motion artefacts. Each stack denotes a set of 7 images; each junction denotes the DVFs application to the IRW images; the box with DVF loss represents the weight adjustment during training. *Reproduced from Gonzales et al (2021) Frontiers in Cardiovascular Medicine 8* [73], *published under CC-BY 4.0.*

2.2.7 Implementation specification

All images were zero-padded to a uniform size of 384 × 384 pixels and intensities were pre-processed with quantile normalisation to ensure generalisability [93]. The multi-scale loss was computed as the average mean square errors of the predicted DVFs at each scale and resolution. The neural network was optimised using the Adam method [94] with an initial learning rate of 0.001, adjusted via a scheduler during training, and a mini-batch size of 4. Training continued until the validation loss did not decrease for 50 epochs, totalling approximately 48 hours until convergence with

low bias and variance. The network was trained using a NVIDIA TITAN XP GPU and implemented in TensorFlow [95]. After training, motion correction for each set of 7 IRW images took under one second on a GPU or modern CPU. Additional details of the architecture are described in Appendix A.1.

2.2.8 Validation

The proposed method was compared to a validated multi-modal image registration algorithm [96], used as the baseline method. This algorithm mitigates artificial motion discontinuities by combining a bilateral filter, a deformation field-based filter, and a diffusion regularisation algorithm. It serves as an effective registration approach without needing prior image segmentation. The baseline method, implemented in C, used the first IRW image as a reference for subsequent pairwise registrations, taking around 30 seconds per T1 map on a modern CPU.

A multi-observer experiment was devised to assess the effectiveness and robustness of motion correction and potential noise introduction in cases originally without motion. A test set of 200 T1 maps with varying degrees of motion artefacts was selected from the UK Biobank based on existing quality scores. Specifically, 50 samples had severe motion artefacts affecting all myocardial segments, 100 had moderate motion affecting individual segments, and 50 had mild to no motion.

Motion extent in the test set was assessed on a five-point scale: 'no motion', 'mild motion', 'moderate motion', 'severe motion', and 'very severe motion', with a numerical scale from 0 to 100% for practicality and intuitive use by human operators. Both the baseline and proposed methods were applied to all samples, creating a total of 400 motion-corrected samples. From the mixed 600 samples, 120 (20%) were randomly duplicated to evaluate intra-observer variability. 3 trained human observers scored the 720 samples for motion extent, blinded to original artefact scores and correction methods. To reduce score variance (X_i), the weighted average score ($\overline{X}$) of the three observers was calculated as $\overline{X} = \sum W_i X_i / \sum W_i$, with weights ($W_i$) derived from the inverse of intra-observer variance (σ_i) [97, 98] based on duplicated cases, i.e., $W_i = 1/\sigma_i^2$. The expected standard error of the weighted average scores was $SE(\overline{X}) = \sqrt{\sum W_i}^{-1}$. Quality scores were reported as mean ± standard deviation.

The non-parametric Wilcoxon signed-rank test assessed statistical differences between data with and without motion correction by the baseline and proposed methods. Given the modest number of repeated comparisons, the significance threshold was set at standard $p < 0.05$ [99]. Statistical analyses were conducted using Python.

2.3 Results

The human observer validation results for the 200 cases from the UK Biobank material are summarised in Table 2.1. The intra-observer variability for the three observers on the 20% duplicated cases were 10.6, 17.3, and 21.9, respectively. The standard error of the final weighted-average scores, used to compare the motion correction methods, was 8.3 on a scale from 0 to 100%. Both the baseline method and MOCOnet significantly reduced motion artefacts, with average motion scores decreasing from 37.1 ± 21.5 to 15.8 ± 15.6 using the baseline method, and to 13.3 ± 10.5 using MOCOnet (both $p < 0.001$). MOCOnet demonstrated a significantly greater reduction in motion artefacts compared to the baseline method in the subgroups with severe motion ($n = 50$, $p = 0.006$) and moderate motion ($n = 100$, $p = 0.04$). For the subgroup with mild to no motion ($n = 50$), both methods significantly reduced motion artefacts further (both $p < 0.001$), and neither introduced additional noise, nor showed significant differences from each other ($p = 0.2$). Overall, MOCOnet provided a higher and more consistent suppression of motion artefacts than the baseline method, as indicated by its lower maximum score and variability ($n = 200$, $p = 0.007$). The boxplot of motion scores (Figure 2.4) illustrates these differences, showing that MOCOnet achieved a narrower range of perceived motion estimates and better robustness to outliers.

MOCOnet learnt from synthetic random motion to predict the necessary DVFs for correcting motion in IRW images, producing motion-corrected T1 maps in real-world data. Figure 2.5 demonstrates the method's robustness, with one training sample mistakenly classified as having no motion artefacts, as indicated by overlaid myocardium and stomach contours. However, this did not lead to overfitting or affect final outcomes, as the data-driven process facilitated learning general principles, allowing MOCOnet to correct the error in such sample rather than replicating it.

Table 2.1: Human observer assessment of motion extent (%) on 200 T1 maps before motion correction, and after the baseline and proposed method for motion correction

	All data (n=200)	**Group 1 Severe motion** (n=50)	**Group 2 Moderate motion** (n=100)	**Group 3 Mild to no motion** (n=50)
Before MOCO	37.1 ± 21.5 (99.3)	55.8 ± 18.7 (99.3)	35.5 ± 18.9 (80.5)	21.7 ± 13.8 (62.1)
Baseline method	15.8 ± 15.6 (93.4)	25.8 ± 19.8 (93.4)	14.7 ± 13.9 (65.7)	**8.1** ± 6.5 (34.2)
MOCOnet	**13.3 ± 10.5 (86.9)**	**18.6 ± 14.3 (86.9)**	**12.7± 9.2 (46.4)**	9.4 ± **6.4 (19.8)**

The quality scores are inverse variance-weighted scores of three human observers and reported in mean ± SD (maximum value). The best results are highlighted in bold. *Reproduced from Gonzales et al (2021) Frontiers in Cardiovascular Medicine 8* [73], *published under CC-BY 4.0.*

2.4 Discussion

In this study, MOCOnet, an innovative end-to-end neural network, was designed to correct motion in CMR T1 maps, leveraging a large-scale dataset and validated by expert human analysts. MOCOnet can automatically predict the necessary deformations to correct actual motion artefact cases. The method offers a rapid processing speed of under 1 second per T1 map, without needing modifications to image acquisition protocols, external hardware, or user input, making it suitable for immediate clinical application.

While the approach of estimating the necessary DVFs to align images was tested on myocardial ShMOLLI T1 maps, the underlying problem formulation and solution are broadly applicable. The deformation estimation relies on a data-driven approach [71] that treats the images 'as is' without depending heavily on variations in contrast, specific inversion recovery times, or prior user input. This methodology can be extended to other T1 mapping techniques requiring alignment of multiple T1-weighted images for accurate exponential recovery curve fitting [24], as well as to parametric mapping in other organs such as the brain [100] and liver [101], and other imaging modalities with different image contrasts [102].

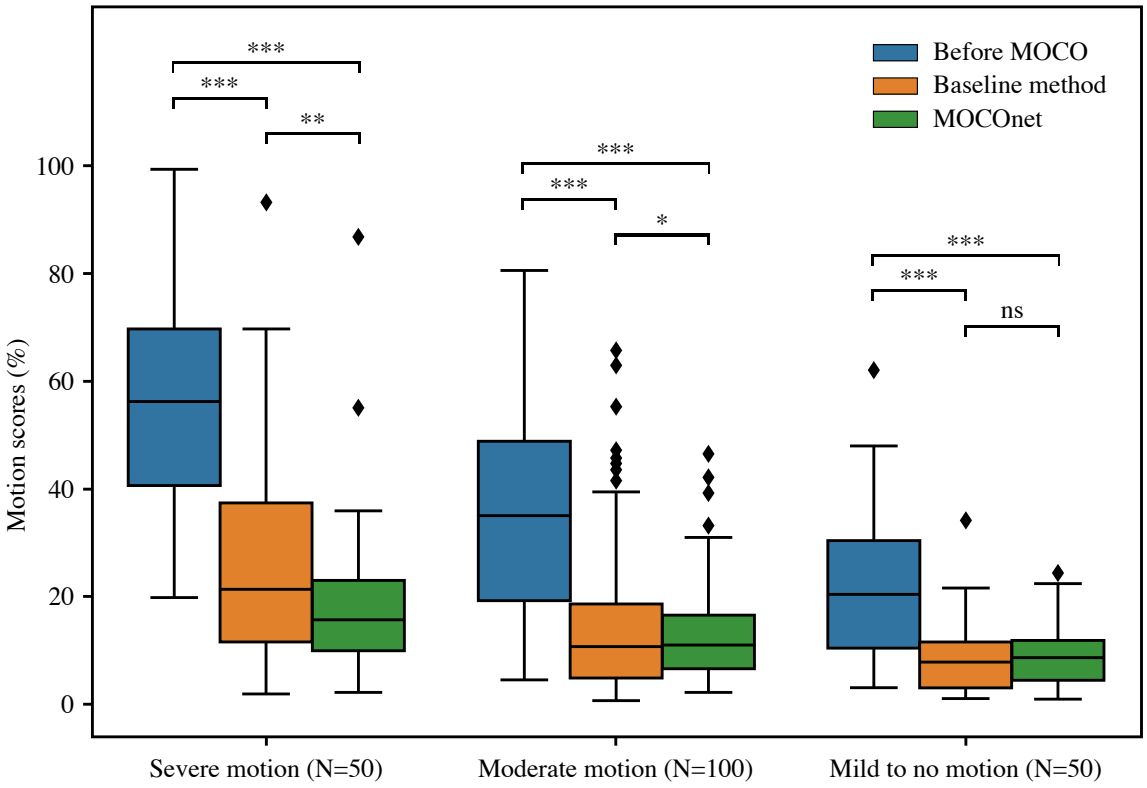
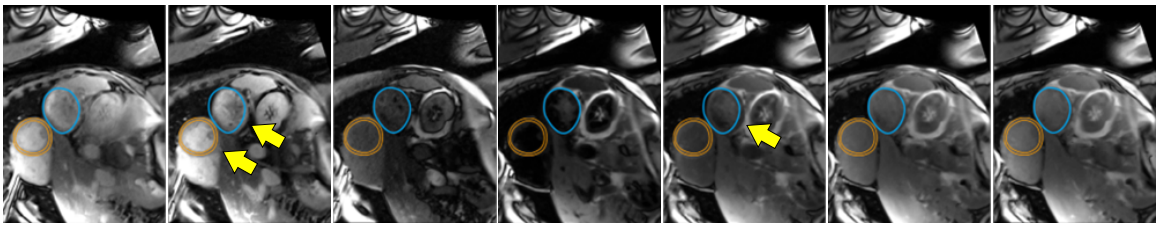


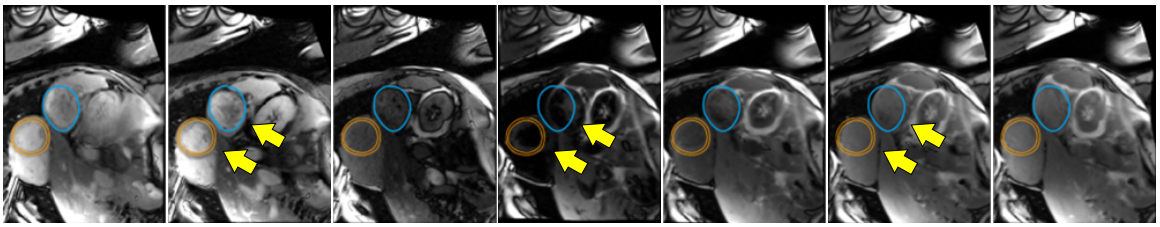
Figure 2.4: Motion correction (MOCO) performance of the baseline and the proposed deep learning-based motion correction (MOCOnet) methods. Box and whisker plot of motion scores in non-parametric terms of three data groups, before (blue) and after motion correction by the baseline (orange) and proposed MOCOnet (green) methods. Reported values are inverse variance-weighted scores of three human observers. MOCOnet achieved the best results and significantly reduced the motion artefacts. * p=0.04; ** p<0.01; *** p<0.001; ns = not significant. *Reproduced from Gonzales et al (2021) Frontiers in Cardiovascular Medicine 8* [73], *published under CC-BY 4.0.*

The clinical implications of this method are promising. A substantial portion of the UK Biobank T1 mapping data analysed showed mild to severe motion, affecting the diagnostic value of T1 mapping. Although advancements in automated motion artefact detection could help with quality monitoring [103], rescheduling scans to obtain motion-free T1 maps would prolong scan times and reduce patient throughput. MOCOnet offers a compelling solution by retrospectively correcting motion in most of the acquired data, enhancing T1 map quality, potentially salvaging data compromised by motion, reducing the need for rescan, and improving diagnostic accuracy. MOCOnet also shows potential for stress T1 mapping applications [24, 30, 34, 104], which may be more prone to motion artefacts. As deep learning continues to evolve, further

(a) Defective ground truth



(b) Defective ground truth with synthetic deformation used for training



(c) Ground truth corrected by MOCOnet

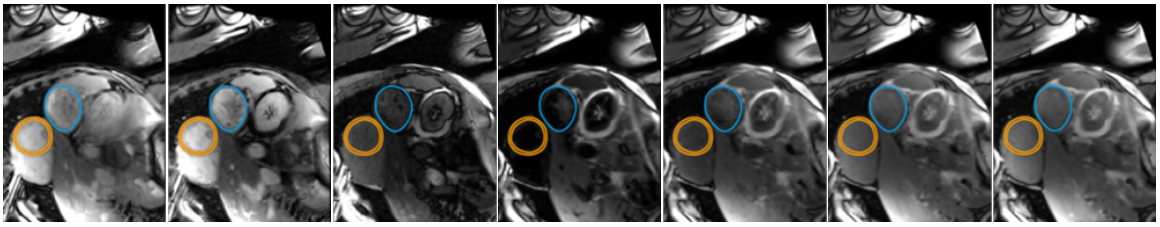


Figure 2.5: Robustness of the proposed motion correction convolutional neural network (MOCOnet) for myocardial ShMOLLI T1 mapping from a noisy training sample. (a) Training sample falsely considered free of motion (1 in 1536) as manually depicted with unaligned myocardium (orange) and stomach (blue) with yellow arrows throughout the inversion recovery-weighted images. (b) Applied deformation to the training sample used for training. (c) Sample corrected by MOCOnet after training demonstrating the successful learning of the general rule without replicating the data. *Reproduced from Gonzales et al (2021) Frontiers in Cardiovascular Medicine 8* [73], *published under CC-BY 4.0.*

studies could explore integrating a wider range of learning-based registration methods [85, 105] into a quality-control-driven pipeline [106, 107] to verify registration accuracy on-the-fly, including the R2 maps. With additional development, MOCOnet, combined with T1 protocol quality assurance [108, 109] and automated myocardial segmentation [106], could establish a robust framework for clinical T1 mapping.

The architectural design and methodological choices for the motion correction framework are informed by established techniques validated through prior studies, particularly the original ablation study of PWC-Net [90]. Each component of the model, including the hierarchical feature extraction in the encoder, the warping lay-

ers, and the multi-scale supervision loss, was integrated based on their demonstrated effectiveness in improving performance. The encoder-decoder structure with skip connections was selected to enable efficient integration of low- and high-resolution features, as studies have shown that the absence of skip connections can reduce spatial detail retention and impair the alignment of input images [89, 110]. Warping layers were incorporated to explicitly handle spatial transformations between feature maps, ensuring progressive alignment across scales, which is critical for complex motion correction tasks; ablation experiments from PWC-Net and similar works [111] have highlighted significant performance degradation when this component is removed. Additionally, the multi-scale loss function, which supervises predictions at each resolution, was designed to encourage coarse-to-fine refinement, a strategy proven to enhance accuracy, particularly in cases involving varying motion magnitudes [112]. The inclusion of smoothness regularisation further ensures physiologically plausible deformations, addressing abrupt and unrealistic predictions, which has been extensively validated in the context of image registration [113]. While the model evaluation in this chapter relies on qualitative scoring by observers, the design choices align with configurations shown to perform robustly in the literature.

Despite MOCOnet's strong performance in correcting motion artefacts, as evidenced by significant improvements in motion scores, human observer experiments indicated that it might not fully correct images with severe motion. This limitation is partly due to the complexity of correcting severe motion and the challenge of through-plane motion, which can cause T1 values to be derived from signals at different tissue locations. Therefore, maintaining breath-hold is essential for acquiring high-quality T1 maps. Future work will involve validating MOCOnet across multi-vendor, multi-centre datasets, extending its application to other anatomical regions, and implementing the solution directly on scanners for robust real-time motion artefact correction, ensuring high-quality, reliable images for immediate clinical use.

2.5 Conclusion

MOCOnet proves to be an effective and reliable convolutional neural network for correcting myocardial motion artefacts. This technique is well-suited for post-processing T1 mapping, enabling the restoration of T1 values in images compromised by motion artefacts. Furthermore, this non-rigid registration solution can be adapted to other mapping methods, facilitating the production of high-quality, reliable images ready for immediate clinical interpretation. By enhancing parametric mapping methods, MOCOnet has the potential to improve the reliability of quantitative CMR medical imaging. The substantial improvements in accuracy and reliability of myocardial T1 maps provide a stable foundation for the detailed evaluations that follow in subsequent Chapters. Enhanced T1 maps ensure higher diagnostic confidence, crucial for the effective application of further inter-sequence registration, segmentation and landmark annotation techniques. This foundational improvement in image quality sets the stage for robust and comprehensive cardiac assessments, integrating seamlessly into the broader analysis framework discussed throughout the thesis.

Chapter 3

Advanced visualisation: derivation of stress T1 reactivity maps using vision transformers

Chapter 2 addressed the issue of motion artefacts in T1 maps, enhancing image quality by correcting intra-acquisition motion. Building on this, the current Chapter tackles the challenge of inter-acquisition motion correction between rest and stress T1 maps, a process that could benefit from the intra-acquisition corrections discussed previously. In addition to the technical advancements, this Chapter introduces the novel concept of stress T1 reactivity maps, or delta T1 (dT1) maps, which offer a new dimension in myocardial tissue evaluation by capturing dynamic changes between rest and stress conditions. These maps may provide critical insights into myocardial health, enabling more precise assessment of stress-induced variations in T1 values [114]. The Chapter details the development and implementation of a robust image registration process that aligns rest and stress T1 maps for accurate dT1 computation. The ability to conduct subsequent analyses on these images, including standardised segmentation, is facilitated by the methods discussed in Chapter 5, which ensures that the enhanced data quality and alignment achieved here can be effectively used in downstream tasks.

The work in this Chapter has been presented [115] in:

115. **Gonzales RA**, Burrage MK, Menacho D, Altun İ, Raman B, Ariga R, Wijesurendra RS, Mahmod M, Levelt E, Huang W-M, Yun C-H, Ferreira VM, Zhang Q, Piechnik SK. dT1 maps: a novel approach for visualising myocardial stress without gadolinium-based contrast agents. In: *SCMR 28th Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2025. doi:10.1016/j.jocmr.2024.101483.

3.1 Background

CMR stress perfusion imaging is a crucial technique for evaluating myocardial ischemia [20, 116, 117, 118] and obstructive CAD. Recent developments in automated pixel-wise myocardial perfusion maps [119] and the quantification of myocardial blood flow have enhanced clinical capabilities and understanding of various cardiovascular diseases [120, 121, 122, 123, 124, 125, 126, 127]. Despite these advancements, the use of GBCA is still required, which adds to the cost and duration of scans, poses potential risks for certain patients, and may accumulate in body tissues [128].

Stress T1 mapping [24] offers a further advancement in CMR technology, capable of assessing microvascular dysfunction [35, 114], obstructive CAD [30, 32, 33, 34, 104], and reduced coronary vasodilatory reserve in valvular heart disease [28], without the use of GBCAs. This method evaluates changes in myocardial blood volume properties during vasodilatory stress, potentially providing a more comprehensive indicator of ischaemia than traditional GBCA-based myocardial blood flow assessments [27]. The measure of stress T1 reactivity involves calculating the percentage change between rest and stress T1 values on a per-slice and per-segment basis [24]. However, deriving these values currently requires multiple manual measurements using ROIs and offline post-processing, making it a labour-intensive, subjective, and time-consuming process. Additionally, current stress T1 results lack a visual component, unlike standard pixel-wise CMR perfusion maps, which complicates visual diagnosis. Consequently, there is a pressing need for a pixel-wise map display for stress T1 reactivity, or a 'stress T1 perfusion map,' that can assess myocardial perfusion and blood flow responses without using GBCAs. Such a tool would facilitate the immediate visual diagnosis of myocardial changes, delineating their spatial extent and severity within anatomical borders, thereby streamlining routine clinical workflows.

This Chapter introduces an innovative pixel-wise dT1 mapping concept and method to visualise and quantify stress T1 reactivity. The proposed model uses a deep-learning transformer-based approach to register rest and stress T1 maps and quantify the percentage changes in myocardial T1 values. The model's accuracy and practicality are validated across standard clinical magnetic field strengths, multiple centres, a broad spectrum of pharmacological agents, and a diverse range of cardiovascular

conditions, encompassing both sexes. This research marks a pioneering step in using transformer-based models for CMR image registration, aiming to deliver a more precise, efficient, and GBCA-free solution for assessing myocardial ischaemia.

3.2 Methods

3.2.1 Imaging data

The study retrospectively included subjects who underwent stress CMR scans at the University of Oxford Centre for Clinical Magnetic Resonance Research (OCMR), John Radcliffe Hospital (Oxford, UK) [28, 34, 35, 129, 130, 131], and Mackay Memorial Hospital (MMH; Taipei, Taiwan). Informed written consent was obtained from all participants, following local ethical approvals at each centre. The CMR scans were performed using Siemens magnetic resonance scanners: Avanto, Avanto Fit, and Tim Trio (1.5T and 3T) at OCMR, and the Aera scanner (1.5T) at MMH. The pharmacological agents used were adenosine, regadenoson, and dobutamine at OCMR, and dipyridamole at MMH. Rest and stress T1 maps were acquired using the ShMOLLI sequence prototype [76, 108], ranging from 1 to 3 short-axis slices (basal, mid-ventricular, and apical). The stress T1 maps were primarily obtained during vasodilator stress, following standard clinical protocols for stress perfusion [15], with some research protocols using increasing doses of the pharmacologic stress agents to account for variations in stress response.

3.2.2 Data preparation

Each rest and stress T1 map was automatically matched by slice position. Data partitioning was stratified and randomised, ensuring balanced representation across the study, magnetic field strength, pharmacological agents, biological sex, and slice positions. Manual inspection excluded pairs with imaging artefacts and out-of-plane motion. For training, all possible combinations of matched rest and stress T1 maps, including within-group permutations, were used to enhance T1 change diversity. For testing, only standard pairs (rest with stress) collected at identical trigger times were

included. These pairs were manually segmented and labelled by two experienced observers using specialised in-house software, blinded to the study's results.

3.2.3 Deep learning model design

The proposed method (Figure 3.1) involved non-rigid registration of a stress T1 map (moving image) to its corresponding rest T1 map (fixed image). This was achieved using the VoxelMorph registration framework [132], which was modified to incorporate a Swin transformer architecture, Swin U-Net [133], replacing the traditional U-Net backbone [89]. This architectural upgrade improved the model's capacity to capture long-distance dependencies within the images, facilitating the registration of a broader range of displacements. The input images were divided into non-overlapping patches, treated as embedded tokens. These tokens were processed by a hierarchical encoder with Swin transformer blocks for effective context feature extraction, allowing for both local and global semantic feature learning. The decoder, mirroring the encoder, included an attention mechanism during the up-sampling process, ensuring accurate alignment of critical anatomical features between the stress and rest T1 maps for precise non-rigid registration. The dT1 map, representing the pixel-wise percentage change, was then calculated as: $\mathrm{dT1} = (\mathrm{T1}_{\mathrm{stress}} - \mathrm{T1}_{\mathrm{rest}})/\mathrm{T1}_{\mathrm{rest}} \times 100\%$.

3.2.4 Deep learning model training

During training, the model learnt to align a given stress T1 map with its corresponding rest T1 map by minimising the image-based loss between them, regularised by the smoothness loss [132], in an unsupervised manner. Each T1 map's intensity was scaled from 0 to 1, centred on the myocardium, and cropped to 192 x 192 pixels. Additional data augmentation techniques included flipping, rotation, and mild synthetic deformation [91]. The neural network was optimised using the Adam method [94] for 200 epochs, implemented in TensorFlow [95] with an NVIDIA GeForce RTX 3090 GPU, taking approximately 59 hours. Additional technical details are described in Appendix A.2.

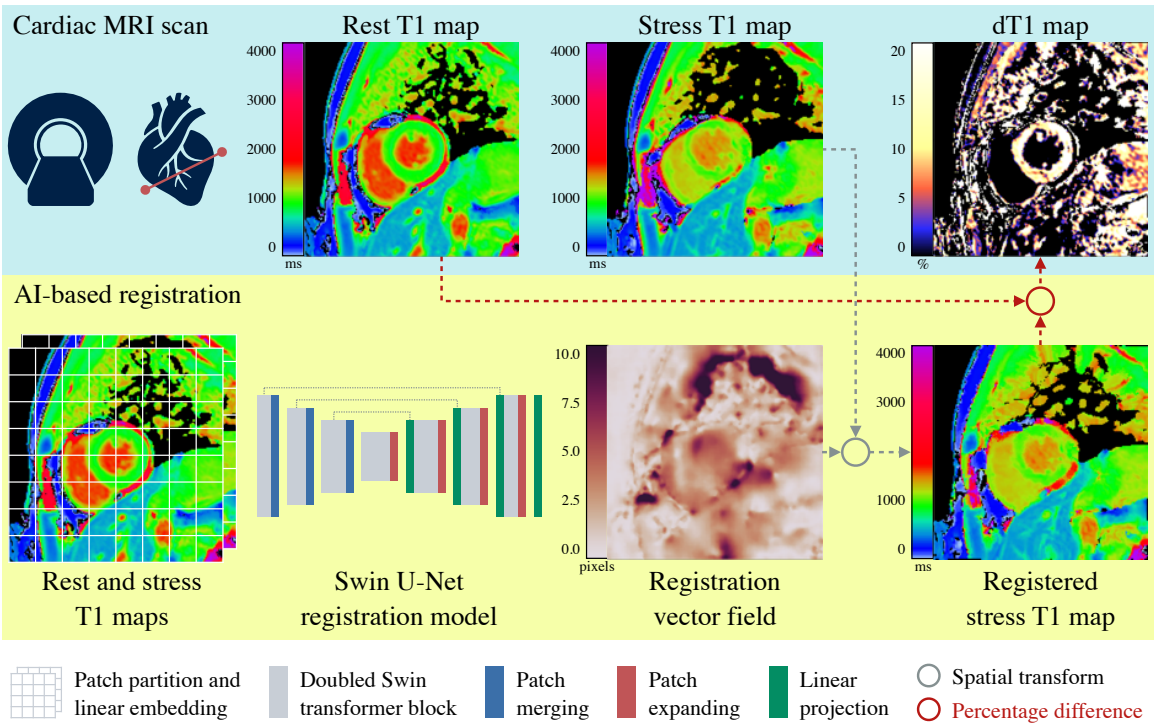


Figure 3.1: Deep learning approach for deriving delta T1 (dT1) maps using a transformers-based model. The process begins by stacking the rest and stress T1 maps, followed by patch partitioning and linear embedding. This prepares the maps for the Swin U-Net registration model, which computes a registration vector field. The field is then used to align the stress T1 map with the rest T1 map. After alignment, a pixel-wise comparison between the rest and the registered stress T1 maps calculates the percentage difference, resulting in the dT1 map, which indicates T1 reactivity due to stress.

3.2.5 Performance evaluation and statistical analysis

The model's registration capability was evaluated by comparing the derived myocardial stress T1 reactivity from the dT1 map with manually derived stress T1 reactivity values from the separate rest and stress T1 maps. This comparison involved overlaying manual contours from the rest T1 maps onto the model-generated dT1 maps and calculating the mean dT1 values within these contours. The correlation between the manually derived and model-generated dT1 values was assessed using the intraclass correlation (ICC) coefficient and Bland-Altman analysis. The evaluation also considered reversed dataset order (to simulate negative stress responses [134]), different AHA segments, pharmacological agents, magnetic field strengths, and biological sexes. A mixed-model ANOVA was performed to assess the effects of these categories

on discrepancies in myocardial measurements. Patient ID was treated as a random effect to account for intra-subject variability, while the aforementioned factors were treated as fixed effects. Statistical analyses were conducted using Python, with significance set at $p < 0.05$. For benchmarking, a U-Net-based VoxelMorph model [132], trained similarly, was compared to the proposed model using myocardial alignment, measured with the Dice similarity coefficient (DSC) [135], and paired t-test analyses across each condition, evaluating the benefits of the transformer-based architecture.

3.2.6 Evaluation of data-centric training strategies

To investigate the impact of training configurations on the performance of the Swin U-Net model, an evaluation of data-centric strategies was conducted. The study examined three data augmentation approaches: (i) no augmentation, (ii) rotation and flipping, and (iii) rotation, flipping, and synthetic deformation vector fields (DVFs), designed to simulate non-rigid deformations. These augmentations were applied in conjunction with two pairing mechanisms: regular pairing, where rest and stress T1 maps were presented in their standard order, and permuted pairing, which introduced within-group combinations of rest and stress maps to increase variability in T1 changes. The pairing strategies ensured diverse representations of myocardial stress responses, while the augmentations aimed to improve the model's robustness to variations in myocardial anatomy and imaging conditions. Training configurations were tested by varying augmentation and pairing methods, and the model's outputs were evaluated in terms of slice-level and segmental agreement with manual reference.

3.3 Results

3.3.1 Imaging data

The study included 5,001 native T1 maps, forming 4,491 regular pairs, collected from 563 subjects (average age 55±16 years, 29% female) prior to quality control. Specifically, data from 549 subjects at OCMR provided 3,798 pairs, while 14 subjects at MMH contributed 693 pairs. Post-quality control (Figure 3.2), 500 subjects (average

age 54±16 years, 29% female) were included in the training dataset, and 61 subjects (average age 62±16 years, 31% female) in the test dataset. During training, 41 T1 maps (1%) were excluded due to severe imaging artefacts, and 525 regular pairs (16%) were discarded due to significant out-of-plane displacement. This process resulted in 22,018 permuted pairs after filtering. For testing, 91 T1 maps (12%) were initially excluded for artefacts, and 20 regular pairs (3%) were removed due to out-of-plane motion, leaving 701 regular pairs. The stratified partitioning ensured a balanced representation across magnetic field strengths, pharmacological agents, biological sex, and slice positions, between training and testing sets, as detailed in Table 3.1. The dataset comprised healthy volunteers (n=103) and patients (n=458) with various conditions, including coronary artery disease (n=146), hypertrophic cardiomyopathy (n=128), atrial fibrillation (n=93), aortic stenosis (n=37), type 2 diabetes (n=26), hypertension (n=21), and breast cancer (n=7).

3.3.2 Stress T1 reactivity derivation

The transformer-based model demonstrated high accuracy in replicating manual stress responses directly from co-registered rest-stress T1 maps (Figure 3.3), with processing times of less than 1.5 seconds per sample on a standard computer. At the pair level, the agreement between model-generated myocardial dT1 values (3.63 ± 3.27%) and manually derived values (3.69 ± 3.23%) was strong, with a low mean difference of -0.06 ± 0.74% and an excellent ICC of 0.97. In a simulated reversed response scenario—registering a rest T1 map to its stress T1 map—the mean difference remained low (0.60 ± 0.88%), with a high ICC of 0.93. At a segmental level, where small changes in the segmental mask can affect stress response, the bias was still low (0.06%) but had a wider spread (1.72%), with a high ICC of 0.91. The stratified analysis showed consistently high performance across different magnetic field strengths, pharmacological agents, biological sexes, and slice positions (Table 3.2). No statistically significant differences were found for any fixed effects ($p > 0.1$), indicating near-perfect agreement across all subdivisions (ICC > 0.97). The lowest agreement levels were observed with regadenoson-induced stress responses (ICC = 0.90) and in apical slices (ICC = 0.94).

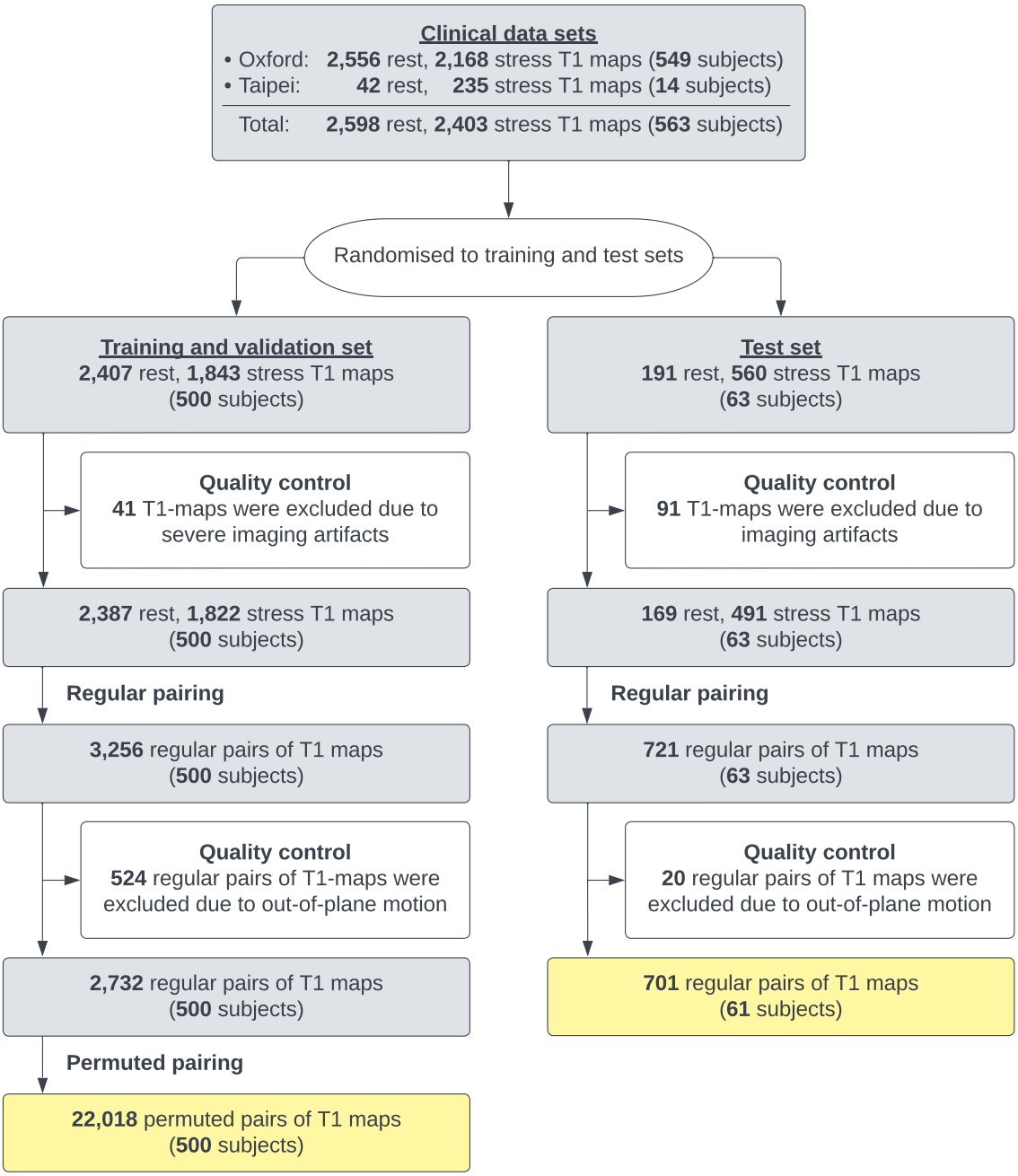


Figure 3.2: Data flow for the material selection for developing and testing the dT1 map registration model. It includes the number of rest and stress T1 maps, the number of subjects from both cohorts, and the subsequent steps taken to form the training and validation set, as well as the test set. Quality control steps involved the exclusion of T1 maps with imaging artifacts and regular pairs with out-of-plane motion. The final numbers of T1 maps and regular pairs included in the training/validation and test datasets are shown, along with the number of permuted pairs created for the study.

Table 3.1: Composition of the included clinical data

Parameter	All data	Training	Test	Ratio
All data	561 (4,840)	500 (4,209)	61 (631)	11% (13%)
Magnetic field strength				
1.5 T	286 (3,380)	242 (2,819)	44 (561)	15% (17%)
3 T	275 (1,460)	258 (1,390)	17 (70)	6% (5%)
Stress agent				
Rest (no stress agent)	561 (2,538)	500 (2,387)	61 (151)	11% (6%)
Adenosine	474 (1,387)	433 (1,273)	41 (114)	9% (8%)
Regadenoson	63 (159)	55 (145)	8 (14)	13% (9%)
Dobutamine	10 (544)	5 (303)	5 (241)	50% (44%)
Dipyridamole	14 (212)	7 (101)	7 (111)	50% (52%)
Biological sex				
Female	163 (1,445)	144 (1,166)	19 (279)	12% (19%)
Male	398 (3,395)	356 (3,043)	42 (352)	11% (10%)
Slice position				
Basal	517 (1,406)	481 (1,305)	36 (101)	7% (7%)
Mid-ventricular	541 (2,408)	498 (1,946)	43 (462)	8% (19%)
Apical	493 (1,026)	463 (958)	30 (68)	6% (7%)

Count represents the number of patients (number of T1 maps).

3.3.3 Case studies

Figure 3.4 showcases two contrasting examples from the test set. The first example (Figure 3.4a) features a healthy volunteer who showed a uniform stress response across all myocardial segments. The second example (Figure 3.4b) involves a patient with coronary artery disease, characterised by a lack of stress response in a specific territory, indicating localised impairment due to the disease. The illustrations include corresponding rest and registered stress T1 maps, alongside derived dT1 maps. Additionally, dT1 maps are overlaid on grayscale rest T1 maps to highlight specific areas of myocardial stress response. The accompanying bull's eye plots provide a detailed representation of myocardial stress reactivity distribution, emphasising the difference in myocardial perfusion between the healthy volunteer and the patient with CAD.

3.3.4 Comparison with baseline model

The proposed model outperformed the baseline model. Despite the baseline U-Net model achieving high performance at the pair level (ICC=0.96), it performed worse in reversed response scenarios and had lower agreement at the segmental level ($p < 0.0001$), with an ICC of 0.88 and 0.86, respectively. Regarding myocardial alignment, the T1 map pairs were initially misaligned with a mean DSC of 0.56 ± 0.22. After registration, alignment improved to a DSC of 0.85 ± 0.06 with the baseline model and further improved ($p < 0.0001$) to a DSC of 0.86 ± 0.04 with the proposed model.

Table 3.2: Performance of myocardial dT1 mapping in the test data set across various conditions, with the proposed model

Results	**# patients**	**# dT1 maps**	**Error** (mean ± SD)	**ICC**
Overall				
Slice	61	701	-0.06 ± 0.74	0.97
Slice – reversed	61	701	0.60 ± 0.88	0.93
Segments	61	4,110*	0.06 ± 1.72	0.91
Magnetic field strength				
1.5 T	44	651	-0.08 ± 0.75	0.97
3 T	17	50	0.20 ± 0.57	0.99
Stress agent				
Adenosine	41	146	0.08 ± 0.68	0.97
Regadenoson	8	16	0.33 ± 1.21	0.90
Dobutamine	5	241	-0.29 ± 0.85	0.97
Dipyridamole	7	298	0.05 ± 0.58	0.98
Biological sex				
Female	19	375	-0.03 ± 0.83	0.97
Male	42	326	-0.09 ± 0.63	0.98
Slice position				
Basal	36	70	0.05 ± 0.41	0.99
Mid-ventricular	43	593	-0.09 ± 0.73	0.97
Apical	30	38	0.19 ± 1.15	0.94

* Count represents the total amount of segments from the test set of 701 dT1 maps.

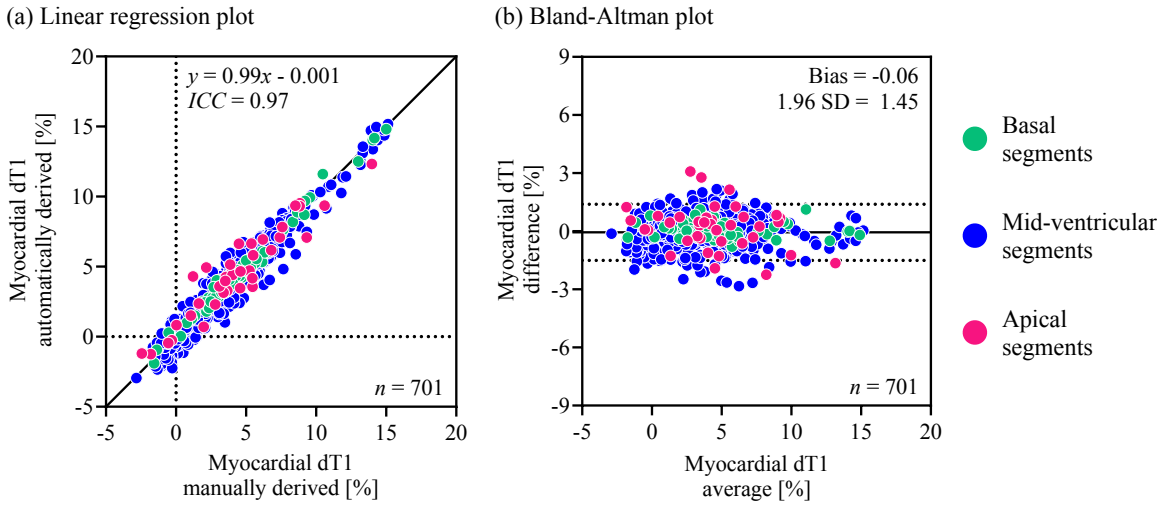


Figure 3.3: Evaluation of the proposed transformer-based model for myocardial dT1 mapping. (a) Linear regression plot showing the relationship between myocardial dT1 values derived automatically by the model versus those obtained manually, with each dot representing a regular pair. The regression line, intra-class correlation coefficient (ICC), and equation are indicated. (b) Bland-Altman plot illustrating the agreement confidence range between the model-generated and manually derived myocardial dT1 values.

3.3.5 Effects of data-centric training strategies

The results demonstrated that both data augmentation and pairing mechanisms played a significant role in enhancing the accuracy of the model's dT1 maps (Table 3.3). When evaluating regular pairing configurations (n=2,732 pairs), segmental analysis, which is more sensitive to variability, showed an ICC of 0.87 without data augmentation. Introducing rotation and flipping slightly reduced this to 0.86, while adding DVFs resulted in a notable improvement, increasing ICC to 0.89 for segmental analysis and maintaining high slice-level accuracy at 0.96. Permuted pairing (n=22,018 pairs) demonstrated considerable improvements even without data augmentation, achieving ICC values of 0.97 and 0.91 for slice-level and segmental analysis, respectively. Incorporating data augmentation into permuted pairing provided marginal gains, with reduced error spread (lower standard deviations) while maintaining the same high ICC values. This indicates that permuted pairing inherently provides the diversity required for robust model performance, with data augmentation offering incremental refinements rather than substantial changes in ICC.

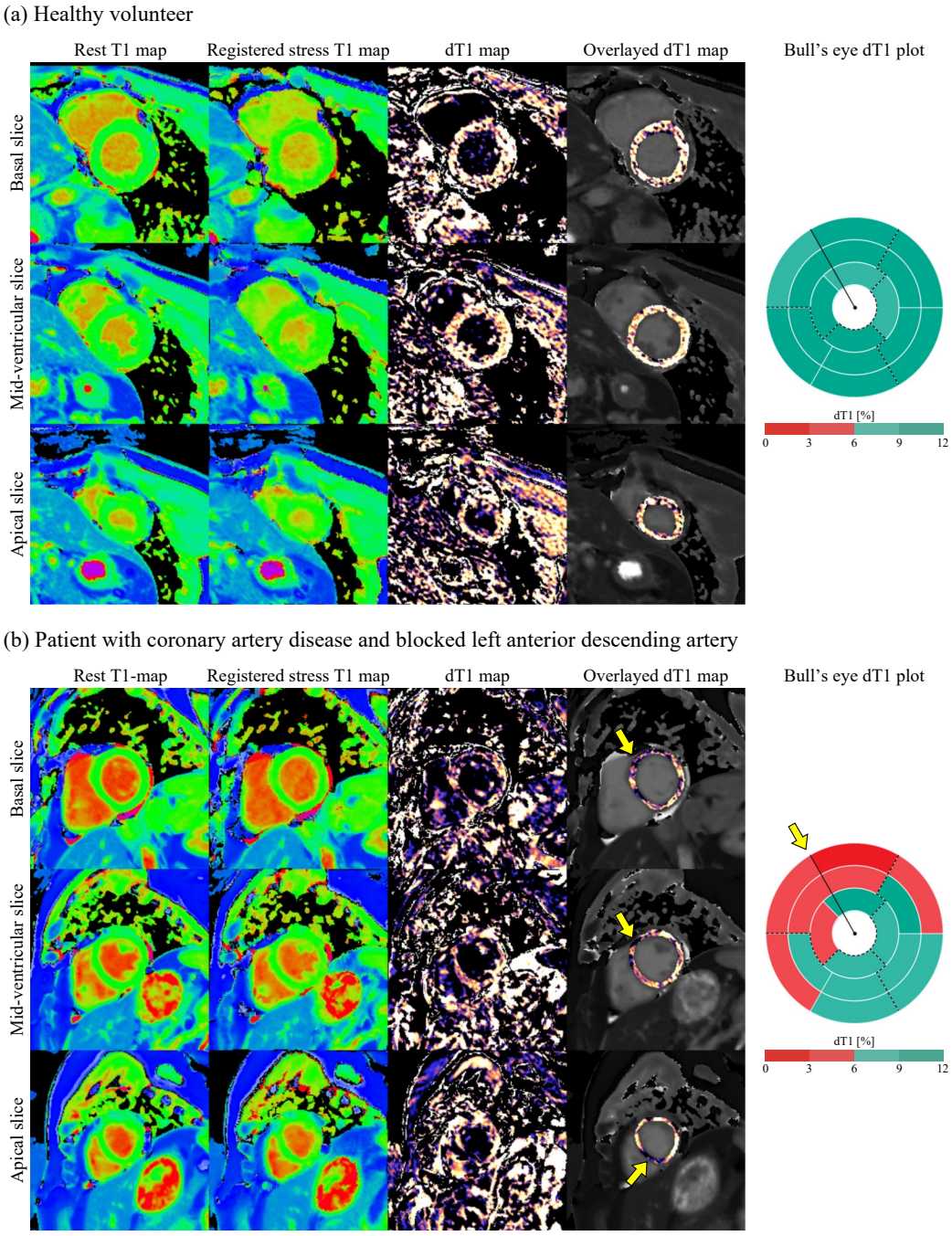


Figure 3.4: Case studies of myocardial dT1 mapping. (a) A healthy volunteer; and (b) a patient with coronary artery disease and obstructed left anterior descending artery. From left to right, the columns display: rest T1 maps, stress T1 maps registered via our model, dT1 maps, and grayscale rest T1 maps overlaid with myocardium dT1 maps to highlight areas of stress reactivity. The last column features simplified bull's eye plots of dT1 values, visually illustrating the myocardial stress response distribution across basal, mid-ventricular, and apical slices.

Table 3.3: Effects of data augmentation and pairing mechanisms on model performance for dT1 mapping

Training data		**Slice**		**Segments**	
Data augmentation	Pairing	Error	ICC	Error	ICC
none	regular	0.23 ± 0.91	0.96	0.37 ± 2.15	0.87
rot + flip	regular	0.09 ± 1.05	0.95	0.24 ± 2.31	0.86
rot + flip + DVF	regular	-0.28 ± 0.85	0.96	-0.13 ± 1.99	0.89
none	permuted	0.01 ± 0.75	0.97	0.13 ± 1.75	0.91
rot + flip	permuted	-0.11 ± 0.75	0.97	0.01 ± 1.78	0.91
rot + flip + DVF	permuted	-0.06 ± 0.74	0.97	0.06 ± 1.72	0.91

Comparison of the Swin U-Net model's performance under different training configurations, evaluating the effects of data augmentation techniques (none, rotation and flipping, rotation, flipping, and deformation vector fields [DVF]) and pairing mechanisms (regular and permuted). Slice-level and segmental analyses are presented in terms of error (mean ± standard deviation) and intra-class correlation coefficient (ICC). Regular pairing included 2,732 pairs for training, while permuted pairing utilized 22,018 pairs.

3.4 Discussion

This work presents an automated approach for generating pixel-wise stress T1 reactivity maps in CMR imaging, employing a transformer-based model. The method was validated retrospectively using imaging data from subjects who underwent vasodilator stress CMR scans, comparing the model's outputs against manually derived stress T1 reactivity measurements. The approach uses the VoxelMorph registration framework enhanced with a Swin transformer architecture to improve the accuracy and robustness of non-rigid registration between stress and rest T1 maps. The model showed a high level of agreement with manual measurements (ICC = 0.97) across various conditions, including different pharmacological stress agents and patient demographics. This innovation offers a novel tool for detailed visualisation and quantification of myocardial stress responses, eliminating the need for GBCAs.

Automating the derivation of pixel-wise stress T1 reactivity maps offers notable clinical advantages over traditional perfusion imaging methods. This technique eliminates the need for GBCAs, reducing the risk of nephrogenic systemic fibrosis in

patients with advanced renal failure [136] and avoiding GBCA accumulation in body tissues [128]. By enabling precise quantification and visualisation of myocardial stress responses, this method can facilitate early detection of ischaemic conditions and assessment of myocardial viability with high spatial resolution [34]. Moreover, the rapid, automated nature of this approach (<1.5s per sample) standardises the analysis process, potentially reducing variability and increasing throughput in clinical settings.

This mapping technique supports structured analysis of stress T1 reactivity, allowing exploration of age- and sex-specific normal ranges [137] and the effects of variables such as pharmacological agents, response timing, and organ-specific differences. These insights can refine dT1 maps, providing a more nuanced visualisation of these variations. Improved evidence on appropriate thresholds for the bull's eye plot, proposed here at around 3, 6, and 9%, guided by existing literature [24], will enhance diagnostic accuracy.

This research marks the first application of transformer-based models for CMR image registration, tailored to rest and stress T1 maps, and applicable to other imaging sequences. Transformers' ability to capture long-term spatial dependencies and displacements offers a significant advantage over conventional convolutional neural networks in non-rigid image registration. By integrating a Swin transformer architecture [133] into the VoxelMorph framework [132], the model achieved significant improvements in mapping accuracy and detail orientation. Additionally, the evaluation of data-centric strategies, including diverse pairing mechanisms and advanced augmentation techniques, revealed their critical role in optimizing model performance. Permuted pairing, in particular, significantly enhanced the robustness of the model by introducing a broader range of myocardial T1 changes during training, while synthetic deformation vector fields further improved the model's ability to handle complex anatomical mismatches. These findings demonstrate that combining advanced architectural designs with comprehensive training configurations is essential for achieving the high accuracy and consistency required in clinical applications. This methodological innovation aligns with the broader trend in medical image analysis, increasingly favouring transformers over convolutional neural networks for their superior performance in handling complex spatial relationships [70].

The dataset used in this study includes 4,840 native T1 maps from 561 subjects across two centres, employing four distinct pharmacological agents (adenosine, regadenoson, dobutamine, and dipyridamole) and capturing different stress timings and magnetic field strengths. This diversity allowed the model to learn from a wide range of conditions, ensuring no significant differences arose from this variety. The images, acquired using the ShMOLLI sequence [76], known for reduced intra-individual variability [104], are crucial in stress T1 mapping to detect minute changes in myocardial blood volume and perfusion under stress conditions. This comprehensive dataset supports the development of a model with potentially more generalisable performance across various clinical scenarios, though further testing is needed to confirm this.

The study has several limitations. Firstly, the analysis was limited to short-axis images, though the training framework and model could potentially apply to a broader range of image orientations. Secondly, like any motion correction method, this model addresses in-plane motion and may be sensitive to through-plane spin history and three-dimensional misregistration of anatomically divergent tissue samples. Care during data acquisition is advised, and further research is needed to address these issues. Thirdly, the registration process applies to reconstructed T1 maps rather than T1-weighted images, suggesting a potential extension to address both within- and between-dataset motion. Future work includes on-the-scanner implementation and direct validation against pathological findings, initially relying on ROI-based processing.

3.5 Conclusion

The transformer-based model introduced in this study has effectively automated the creation of pixel-wise stress T1 reactivity maps (dT1 maps), showcasing high accuracy and reliability across a broad spectrum of conditions and patient demographics. This innovative approach enables detailed visualisation and precise measurement of myocardial stress responses, closely resembling traditional perfusion mapping but without the necessity for GBCAs. By integrating these capabilities into a pipeline that could potentially include the motion-corrected T1 maps from Chapter 2, and

posterior automated analysis outlined in Chapter 5, the precision and practicality of inter-sequence registration could be enhanced further. Such advancements hold promise for refining diagnostic processes and treatment planning for myocardial ischaemia and other cardiac disorders, providing clinicians with safer, more accessible, and efficient tools for cardiac evaluation. This progression toward integrated diagnostic solutions underscores the thesis's overarching goal of enhancing the coherence and effectiveness of CMR T1 mapping analysis through advanced post-processing technologies.

Chapter 4

Automated post-processing I: accountable myocardial segmentation via quality control-driven deep ensemble

Chapters 2 and 3 addressed key challenges in CMR imaging by improving T1 mapping, first through motion correction and then by introducing stress T1 reactivity maps. Building on these advancements, this Chapter shifts focus to the segmentation of virtual native enhancement (VNE) images, derived from T1 maps and cine sequences. While previous work by our research group has automated the segmentation of T1 maps [106], VNE images have a different LGE-like appearance and contrast ranges. These VNE images also require a separate myocardial contour for accurate analysis. The techniques discussed here may leverage the motion-corrected T1 maps from Chapter 2 to enhance the quality of VNE image derivation by avoiding artefacts. Furthermore, this method is directly applicable to the segmentation of LGE images, ensuring consistency across different CMR modalities. A quality control-driven deep ensemble is introduced to guarantee reliable segmentations, addressing critical needs in clinical practice. The imaging modalities discussed here could ultimately benefit from the annotation techniques covered in Chapter 5, building towards a cohesive approach to CMR analysis.

The work in this Chapter has been published [138] and presented [139, 140] in:

138. **Gonzales RA**, Ibáñez DH, Hann E, Popescu IA, Burrage MK, Lee YP, Altun İ, Weintraub WS, Kwong RY, Kramer CM, Neubauer S, Ferreira VM, Zhang Q, Piechnik SK. Quality control-driven deep ensemble for accountable automated segmentation of cardiac magnetic resonance LGE and VNE images. *Frontiers in Cardiovascular Medicine*. 2023;10. doi:10.3389/fcvm.2023.1213290.

139. **Gonzales RA**, Ibáñez DH, Hann E, Popescu IA, Burrage MK, Lee YP, Altun İ, Weintraub WS, Kwong RY, Kramer CM, Neubauer S, Ferreira VM, Zhang Q, Piechnik SK. Quality control-driven framework for reliable automated segmentation of cardiac magnetic resonance LGE and VNE images. In: *International Workshop on Clinically-oriented and Responsible AI for Medical Data Analysis*. Springer International Publishing; 2023. ORA

140. **Gonzales RA**, Zhang Q, Hann E, Ferreira VM, Piechnik SK. Quality control-driven artificial intelligence for reliable automatic segmentation of LGE images in clinical practice. In: *SCMR Virtual Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2022. ORA

4.1 Background

LGE has long been regarded as the gold standard in CMR imaging for non-invasive myocardial tissue characterisation. LGE provides vital information regarding the extent and location of myocardial damage, enabling clinicians to make accurate diagnoses and informed treatment decisions [17]. It is particularly useful in identifying areas of scar tissue or fibrosis [141], which are commonly associated with conditions such as MI [142] and HCM [143]. Quantifying these areas can offer insights into scar burden, which is predictive of adverse clinical outcomes like heart failure and sudden death, and can guide risk-modification strategies, such as the implantation of cardioverter-defibrillator devices [144].

To determine the extent and location of myocardial pathology in LGE images, precise segmentation of the LV myocardium is required. Traditionally, this segmentation has been performed manually by experts, a method that is both time-consuming and subjective. Recent advancements have seen the development of automated segmentation techniques to enhance efficiency and reduce inter-observer variability. These techniques can be broadly classified as either model-driven [145] or data-driven [146]. Model-driven methods employ prior knowledge about the LV myocardium's structure to guide segmentation, whereas data-driven methods involve machine learning algorithms to learn from examples in a training dataset, often yielding superior results [147]. Despite these advancements, two significant challenges remain for clinical application: the need for large amounts of high-quality training data [71], particularly for rare or heterogeneous diseases, and the occurrence of unflagged segmentation errors [148], which can lead to inaccurate scar quantification and potentially impact clinical decision-making. Consequently, there is a pressing need for a validated, automated quality control (QC) mechanism to reliably detect and flag segmentation errors [149].

Various approaches have been proposed to address the challenge of data scarcity in medical applications, including transfer learning, domain adaptation, and data augmentation [150]. Transfer learning and domain adaptation leverage knowledge from pre-existing datasets, while data augmentation creates new data by applying transformations to existing datasets. Among these, data augmentation using Generative Adversarial Networks (GANs) has gained popularity for its ability to generate large

amounts of diverse and realistic data, which is especially useful for limited datasets [151, 152]. However, the application of synthetic data in medical contexts poses a validation challenge, as the generated data may not accurately reflect the true biological and pathological variations seen in real-world scenarios [153]. Thus, clinical validation of synthetic data is essential before its use in medical applications.

Recently, automated methods for detecting inaccuracies in automatic segmentation have gained traction [149]. Post-analysis QC tools have been developed to assess the reliability of segmentation outputs, serving as the final performance indicator for models. These methods typically function as binary classifiers [154, 155], assigning correct/incorrect labels to segmentations, or as regressors [156, 157], which aim to infer well-known validation metrics or uncertainty estimates. While these QC approaches have been successfully applied to CMR T1 mapping [106] and short-axis cines [158], a QC pipeline for LGE segmentation remains absent.

In this Chapter, we introduce a novel approach for segmentation that not only addresses the challenges of limited training data and quality control but also leverages the emerging VNE imaging technique. VNE is gaining traction as it can derive LGE-like images using native signals from T1 mapping and cine sequences, eliminating the need for GBCA [25, 37]. This contrast-agent-free modality significantly expands the training dataset with clinically-validated data, enhancing the utility of CMR. Furthermore, we extend an automated quality control mechanism to flag problematic cases for focused inspection prior to clinical use. Our approach builds upon the quality control-driven (QCD) framework [157], which can predict a confidence metric in the absence of ground truth, thereby ensuring robust and reliable segmentation for clinical applications.

4.2 Methods

4.2.1 Imaging data

The development dataset comprised 4,716 LGE images from 1,363 patients, sourced from: (1) the multi-centre Hypertrophic Cardiomyopathy Registry (HCMR) study [159], which included 3,286 images from 1,129 patients across 24 centres; (2) the clin-

ical service at the Oxford Centre for Clinical Magnetic Resonance Research (OCMR), providing 712 images from 109 patients; and (3) the Oxford Acute Myocardial Infarction (OxAMI) study [160], which contributed 718 images from 125 patients. Institutional review committee and ethics approvals were obtained for these studies. The dataset included 3,286 LGE images from 1,129 patients with hypertrophic cardiomyopathy, and 1,430 LGE images from 234 patients with MI (255 images from 65 patients with chronic MI and 1,175 images from 169 patients with acute MI). CMR scanning was performed on Siemens magnetic resonance scanners (Siemens Healthcare, Germany) with 1.5T (71% of data) and 3T (29% of data) magnetic field strengths. The CMR protocols encompassed cine steady-state free precession imaging, native and post-contrast T1 mapping using the ShMOLLI sequence [76, 108], and LGE imaging conducted approximately 10 minutes post-intravenous administration of 0.1 to 0.2 mmol/kg of gadolinium-based contrast agent, typically using the phase-sensitive inversion recovery (PSIR) sequence [159]. Manual quality control included selecting uncorrupted, paired cines, T1 maps, and LGE images, which were segmented manually by three trained observers in previous studies [25, 37, 161].

4.2.2 Data augmentation using a generative adversarial network

A conditional GAN approach was employed to augment the data by generating VNE images [25, 37] from paired short-axis cine and T1 maps. These VNE images leveraged native components, including native T1 mapping and pre-contrast cine frames throughout the cardiac cycle. These inputs provided image contrast, alterations in myocardial tissue properties, myocardial structure (such as wall thickness), motion data of the cardiac wall, and more distinct myocardial borders. The deep learning generator used these inputs to produce VNE images that closely resembled LGE images in terms of structure and contrast. The clinical benefit of VNE lies in its ability to create 'virtual' LGE images without the need for GBCA, allowing for quicker, more cost-effective, and contrast-free CMR scans.

The VNE generator (Figure 4.1) consisted of parallel convolutional neural network streams that processed cine frames and T1 maps separately. Each stream used a six-

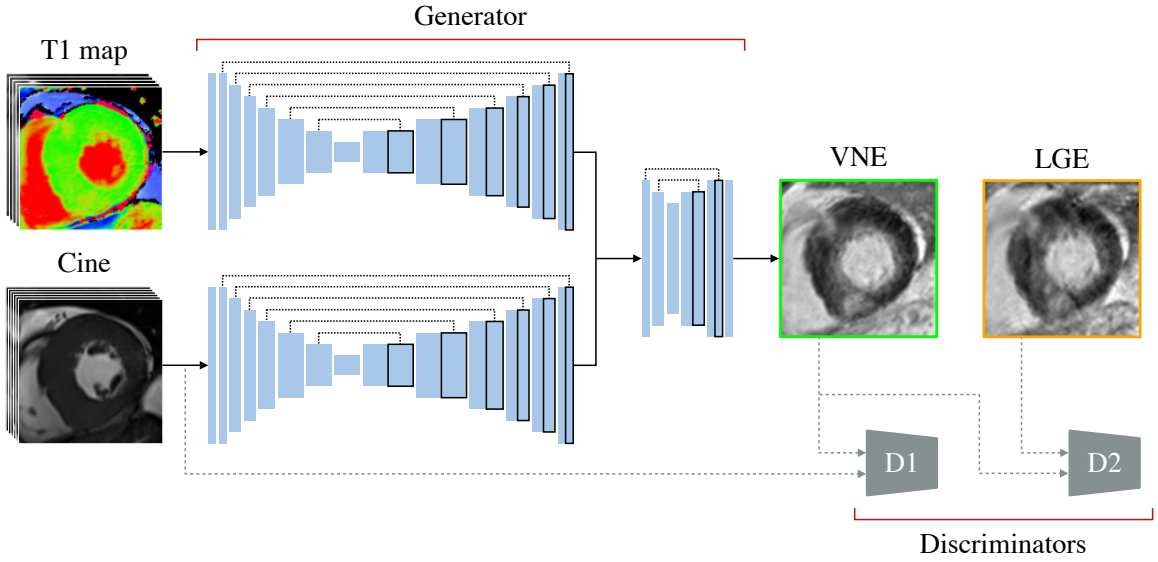


Figure 4.1: Data augmentation framework. A late gadolinium enhancement (LGE) image is augmented by using its paired short-axis cine and T1 map, producing a virtual native enhancement (VNE) image, using a modified conditional generative adversarial network approach. Parallel deep encoder-decoders extract features from native signals, which are fused through a shallow encoder-decoder to derive a VNE image. The discriminators, D1 and D2, during training, are used to enhance the D1 image 'clarity' and the image 'realness' with perceptual similarity, respectively. *Reproduced from Gonzales et al (2023) Frontiers in Cardiovascular Medicine 10* [138], *published under CC-BY 4.0.*

level encoder-decoder U-Net structure [89]. The encoder extracted image features at various scales, employing successive convolutional layers for feature extraction and downsampling at each level, providing a multiscale feature representation. The corresponding decoder fused these multiscale features to generate the final feature maps, using symmetrical upsampling layers and convolutions to sequentially combine the multiscale features. These feature maps from the streams were then concatenated and passed through an additional two-level encoder-decoder block, which integrated information from the different modalities to produce the final VNE image using a late fusion approach. Each encoder-decoder block concluded with a hyperbolic tangent activation function.

In the customised conditional GAN setup [162], the architecture featured two discriminators, D1 and D2, modelled after the VGG16 model [163]. Discriminator D1, designed to verify the 'clarity' of larger images, utilised an expanded architec-

ture with an input layer accommodating the resultant VNE and the input cine stack, ensuring sharper, clearer images. This involved a series of convolutional layers alternating between feature extraction and downsampling, each followed by leaky rectified linear unit activation functions. Discriminator D2, aimed at assessing the 'realness' in single-channel images, had a similar yet more compact structure, processing both the resultant VNE and the paired LGE. The generator's goal was to create VNE images that exhibited high perceptual similarity [164] to LGE images, making them indistinguishable from LGE contrast images. The discriminators' role was to differentiate between VNE and LGE images. Following adversarial training of the neural networks, the trained generator could translate native CMR signals into LGE-like representations.

With the previously trained VNE generator [25, 37], the LGE images in the development data were expanded by producing corresponding VNE images (Figure 4.2) in independent datasets. Data augmentation was successfully achieved for all cases, except for the subset related to acute myocardial infarction, which is awaiting further validation before inclusion. All augmented data were also manually segmented. The VNE images closely matched the position-matched LGE images, despite minor differences in slice position between the paired T1/cine and the final LGE, which were occasionally due to patient movement between image acquisitions (Figure 4.2, cases 5 and 6). This serendipitously introduced greater diversity and realism into the training data, thereby enhancing the model's robustness.

4.2.3 Quality control-driven ensemble framework

A QCD ensemble framework [157] (Figure 4.3) was designed to enhance the accuracy and reliability of segmentation by exploiting the strengths of multiple CNNs. This framework employed various U-Nets [89] with different depths, creating a diverse set of candidate segmentations. These segmentations were combined using statistical rank filters on a pixel-wise basis [165], increasing the pool of segmentation options and improving robustness.

The ensemble framework consisted of six U-Nets [89], ranging in depth from 1 to 6 levels. Each U-Net included an encoder and a decoder. The encoder featured con-

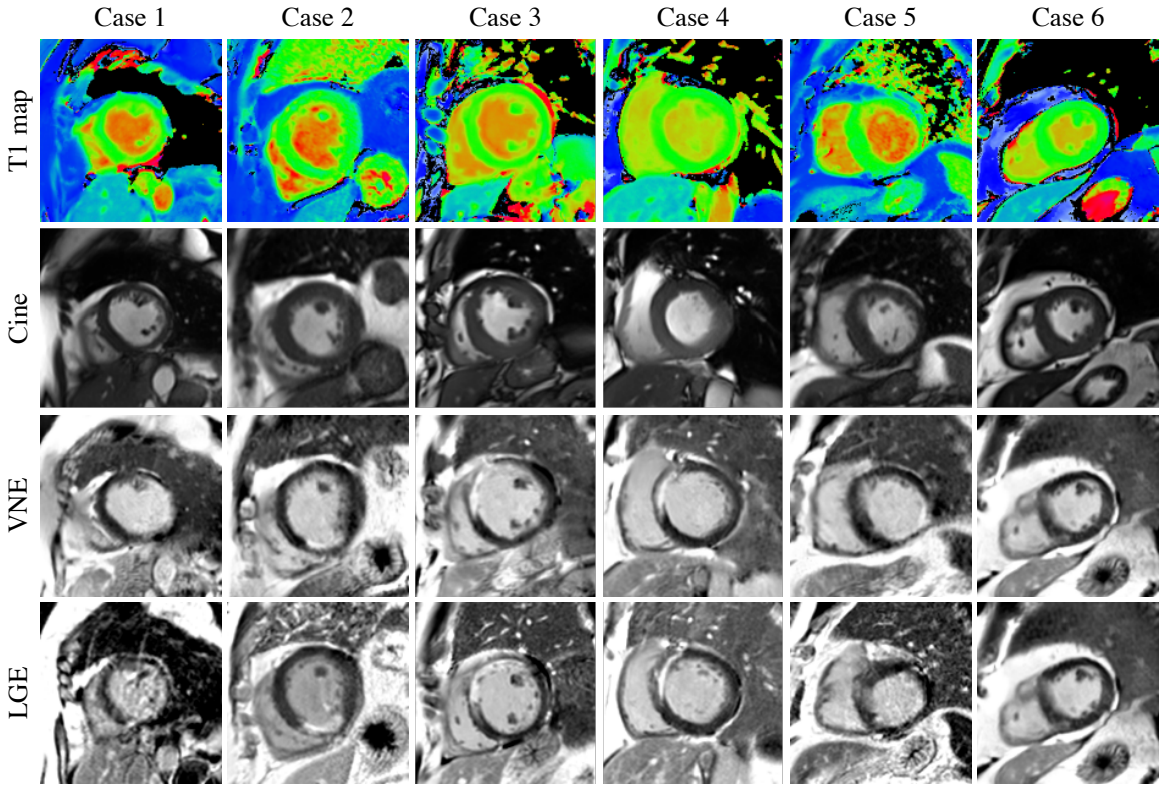


Figure 4.2: Cases with their T1 maps, cines, late gadolinium enhancement (LGE) and virtual native enhancement (VNE) images. The resultant database includes the LGE data and VNE data, as ways of data augmentation and additional validation. *Reproduced from Gonzales et al (2023) Frontiers in Cardiovascular Medicine 10* [138], *published under CC-BY 4.0.*

volutional layers followed by dropout layers [166] for regularisation, with the dropout rate increasing with each layer to prevent overfitting. After convolution and dropout, a max pooling operation was applied. The decoder mirrored the encoder but used transposed convolutional layers for upscaling. It also used skip connections, linking outputs from the decoder with corresponding encoder layers. The final layer underwent additional convolutions and a softmax activation to generate the final segmentation. This process, repeated for each depth, allowed for the creation of diverse candidate segmentations, enhancing the ensemble's performance.

At the core of the framework was an automatic quality scoring mechanism that predicted the DSC for each candidate segmentation by leveraging their differences. It calculated the pairwise agreement, or inter-segmentation DSC matrix, between segmentations, capturing the overlap and differences among candidates. These DSC

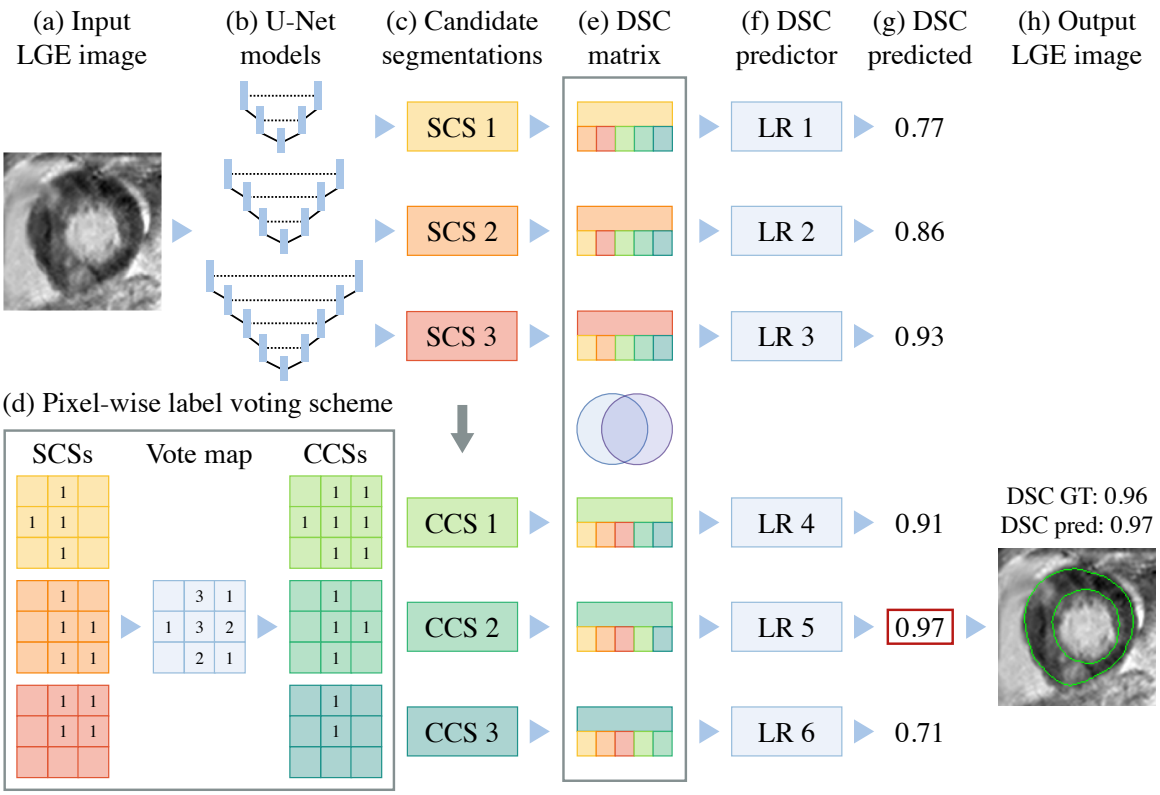


Figure 4.3: Illustration of quality control-driven ensemble framework concepts; depicted here with 3 (out of 6) U-Nets. (A) A late gadolinium enhancement (LGE) image is processed by (B) an ensemble of independent U-Net segmentation models to produce (C) single candidate segmentations (SCSs). (D) The SCSs are then combined via a pixel-wise label voting scheme to derive combined candidate segmentations (CCSs). (E) An association matrix of Dice similarity coefficients (DSC) is generated upon the agreement between SCSs and CCSs. The inter-candidate DSCs are supplied to the (F) linear regressors (LR), and (G) each model outputs the predicted the DSC, in absence of ground truth (GT); finally, (H) the model with the highest predicted DSC and its corresponding automated segmentation output are selected on-the-fly. *Reproduced from Gonzales et al (2023) Frontiers in Cardiovascular Medicine 10* [138], *published under CC-BY 4.0.*

matrices were then input into separate linear regression models for each candidate, with the target being the DSC between the candidate and the ground truth.

For each input image, the framework assigned a predicted DSC for every candidate segmentation, both individually and combined, relative to the ground truth segmentation. The final segmentation was selected by identifying the candidate with the highest predicted DSC, indicating the most accurate and reliable result. This selection process was entirely automated, with no manual intervention required. This approach effectively mimicked a multidisciplinary clinical team, where consistency

among multiple expert opinions served as an indicator for the best approach to managing complex cases. By incorporating this quality control-driven strategy, the ensemble framework aimed to enhance overall segmentation performance and provide confidence metrics, particularly valuable in clinical settings.

4.2.4 Implementation

The dataset was expanded using VNE technology, which used co-located short-axis cines and ShMOLLI T1 maps, resulting in 3,541 VNE images. The development dataset was randomly divided into: (1) 85% for training (4,092 LGE images and 2,917 VNE images from 1,158 patients); (2) 7.5% for validation (309 LGE/VNE images from 102 patients); and (3) 7.5% for testing (309 LGE/VNE images from 103 patients), following recommended guidelines [167]. Image pixel values were normalised from 0 to 1 and zero-padded to 256 $\times$ 256. For the segmentation models, the Adam optimiser [94] was used to minimise the categorical cross-entropy loss, with a learning rate of 0.00005 for 200 epochs; an automated early stopping mechanism was employed to prevent overfitting, using the validation set. For the quality prediction models, a linear regressor was fit for each candidate segmentation based on the inter-agreement between its corresponding candidate segmentation and the others. These regressors were trained on the validation set to avoid autocorrelation with the training set. Specific details for the architectures and training procedures are further explained in Appendix A.3. The models were trained and tested on TensorFlow [95] with an NVIDIA GeForce RTX 3090 GPU, taking approximately 11.5 hours.

4.2.5 Evaluation

The QCD ensemble framework's performance was evaluated for myocardial contours on both LGE and VNE test datasets, and across the main pathologies. Segmentation accuracy was measured using the DSC, comparing the agreement between the selected optimal mask and the ground truth mask. The predicted segmentation accuracy was evaluated in terms of the mean absolute error (MAE) and binary classification accuracy, with a DSC threshold of 0.7 [168]. The MAE measured the difference between the predicted DSC and the observed ground truth DSC from manual segmentation.

Binary classification accuracy determined whether segmentations were classified as good ($\geq$0.7) or poor quality ($<$0.7) to illustrate the practical application of DSC prediction. In the evaluation, false positives occurred when the predicted DSC was $\geq$ 0.7 but the actual accuracy was $<$0.7, and false negatives when the predicted DSC was $<$0.7 despite the actual accuracy being $\geq$ 0.7. The 0.7 threshold ensured a balance between sensitivity and specificity in classifying segmentations. A Wilcoxon signed-rank test was conducted using Python to determine if there was a statistically significant difference between segmentation results on LGE and VNE data, paired where possible, and within pathology groups. A p-value of < 0.05 was considered significant. This analysis helped assess the model's robustness in segmenting both image types and the potential benefits of incorporating VNE data into the training process.

A comparative analysis was conducted to examine the performance of the pipeline's key components and the impact of integrating VNE data. First, the performance of the deepest U-Net (depth of 6 levels) and the QCD segmentation framework was assessed to highlight the advantage of achieving higher segmentation accuracy with a quality prediction capability. Second, experiments were conducted using training data comprising LGE, VNE, and both types combined to thoroughly evaluate the data augmentation capability of GAN-generated VNE data. Each experiment was also tested on LGE data, VNE data, and both combined, to demonstrate the robustness of the proposed method. The segmentation accuracy and quality prediction accuracy were assessed in all experiments to compare differences.

4.3 Results

4.3.1 Segmentation and prediction accuracy

The scatter plots (Figure 4.4) illustrate the correlation between the ground-truth DSC and the predicted DSC for the outputs of the framework and each candidate model in the test set, showing accurate predictions across a range from underperforming to high-performing models. The QCD framework efficiently segmented the LV myocardium in both LGE and VNE images. It demonstrated consistent segmentation

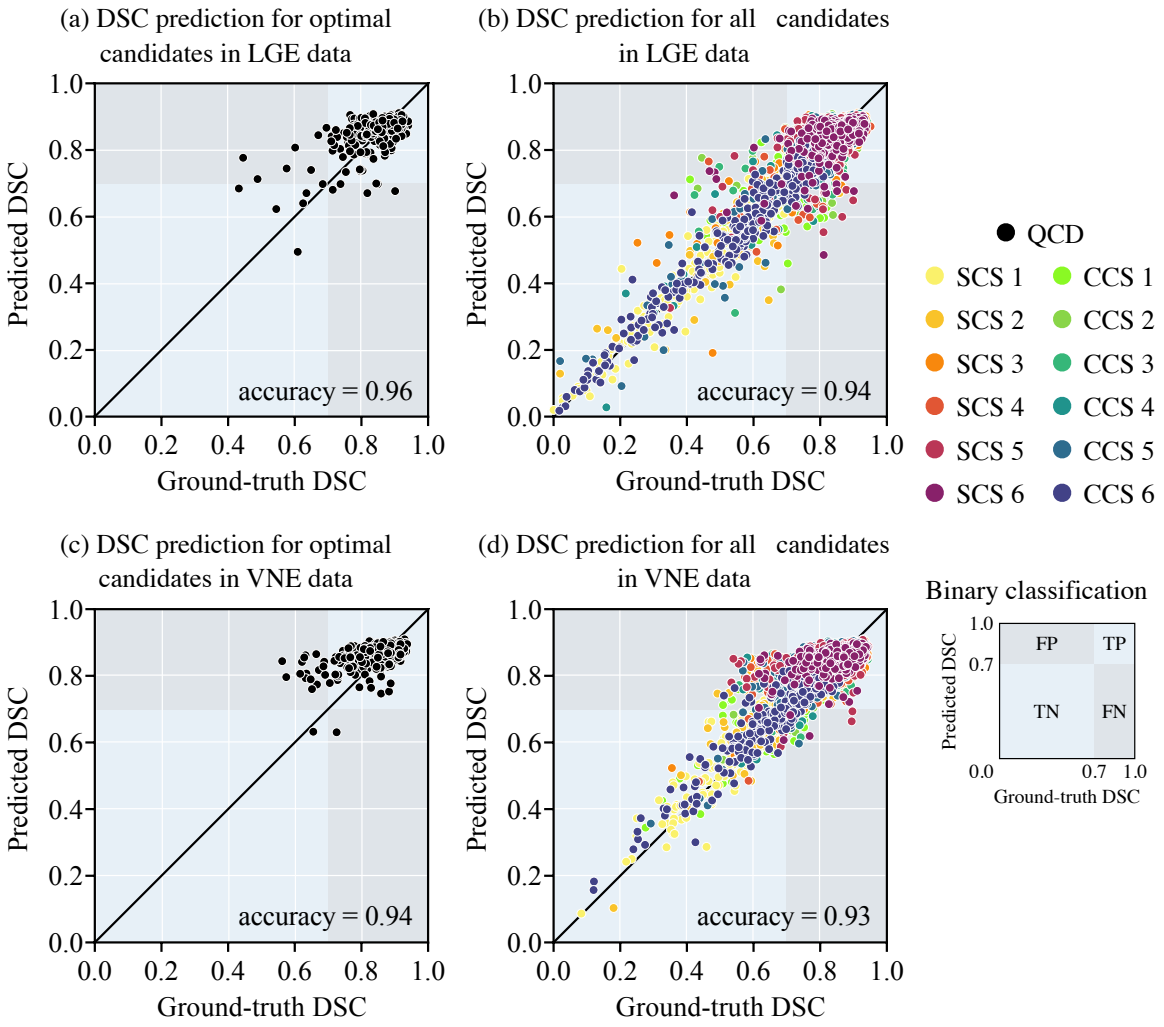


Figure 4.4: Performance evaluation of the quality control-drive segmentation framework for late gadolinium enhancement (LGE) and virtual native enhancement (VNE) images. Scatter plots of the observed ground-truth Dice similarity coefficient (DSC) (x-axis) versus the predicted DSC (y-axis) for myocardial contours in (A,B) LGE and (C,D) VNE images for the optimal candidates (A,C) and for all single (SCS) and combined segmentation (CCS) models (B,D). The shown overall binary classification accuracy is measured as the proportion of true results (true positive (TP) or true negative (TN)—light blue background), in a population of both true and negative results (false positive (FP) or false negative (FN)—grey background), with a binary threshold of DSC ≥ 0.7. *Reproduced from Gonzales et al (2023) Frontiers in Cardiovascular Medicine 10* [138], *published under CC-BY 4.0.*

performance on both the LGE and VNE test datasets, with comparable mean DSC values (LGE: 0.845 $\pm$ 0.075; VNE: 0.845 $\pm$ 0.071; p = not significant). The framework also showed robust performance across key pathologies, including hypertrophic cardiomyopathy (0.845 $\pm$ 0.069) and myocardial infarction (0.844 $\pm$ 0.085; p = not

significant). The MAE for the predicted DSC was low at 0.043 ± 0.043, highlighting the accuracy of the quality control-driven approach in predicting segmentation quality. Furthermore, using a DSC threshold of 0.7, the binary classification accuracy was high at 0.951, underscoring the practical utility of the proposed ensemble framework in clinical environments. Figures 4.5 and 4.6 provide representative test cases of the QCD framework on LGE and VNE images, illustrating true positive, true negative, false positive, and false negative instances.

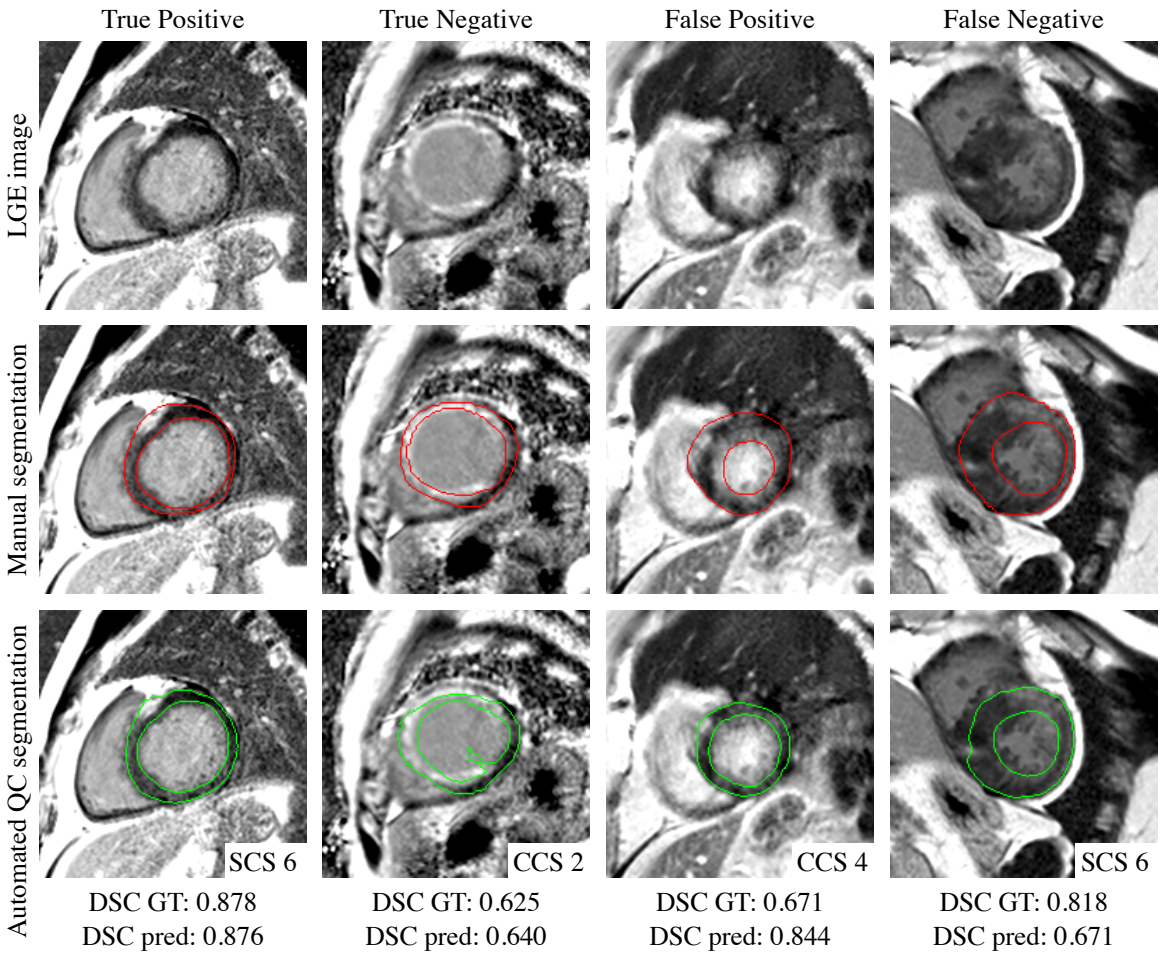


Figure 4.5: Examples of true positive (93.9%), true negative (1.9%), false positive (2.3%) and false negative (1.9%) for predicted quality-controlled (QC) segmentations in late gadolinium enhanced (LGE) images. The left ventricular myocardium is manually segmented in red and automatically segmented in green, from different single (SCS) and combined candidate segmentation (CCS) models. The corresponding observed ground-truth (GT) Dice similarity coefficient (DSC) and predicted DSC are provided at the bottom. *Reproduced from Gonzales et al (2023) Frontiers in Cardiovascular Medicine 10* [138], *published under CC-BY 4.0.*

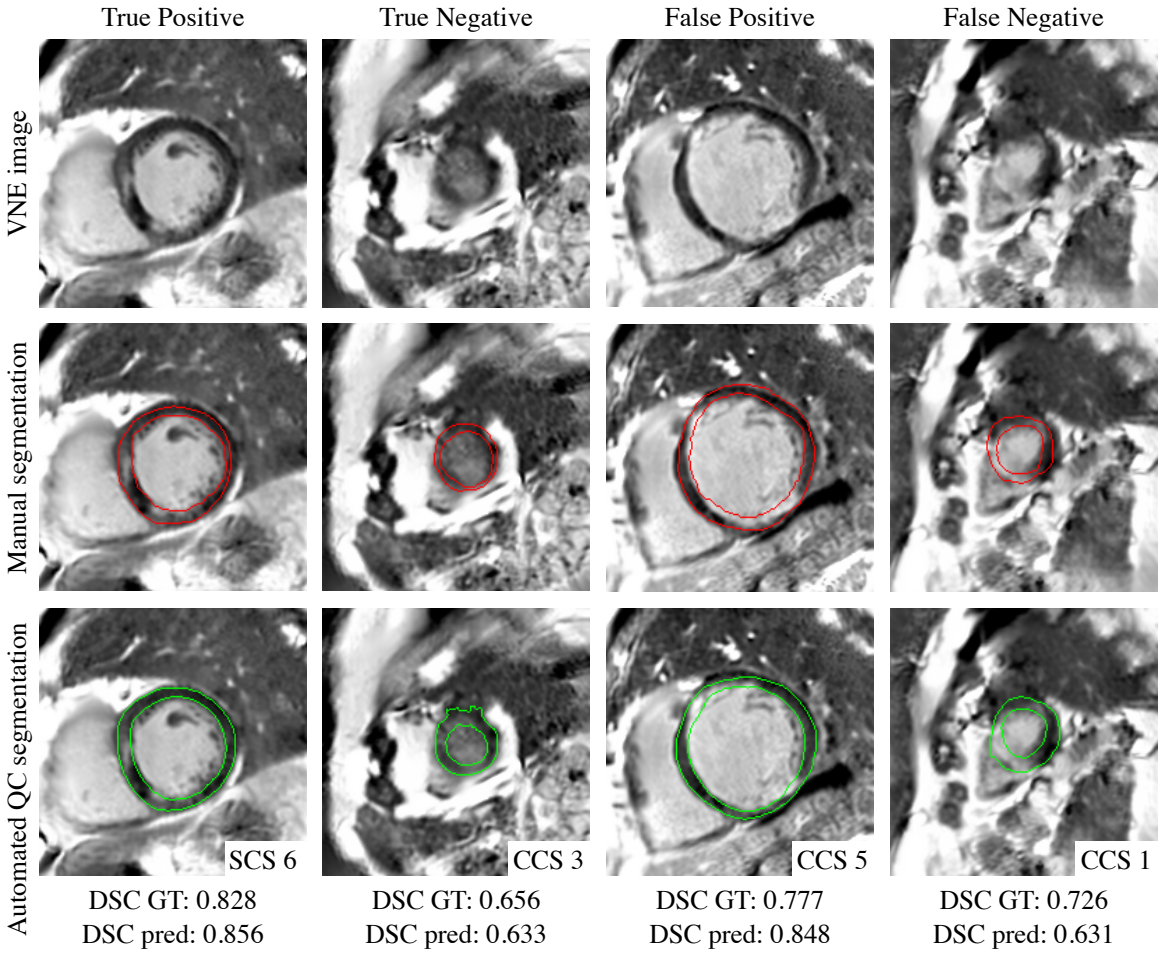


Figure 4.6: Examples of true positive (94.2%), true negative (0.3%), false positive (5.2%) and false negative (0.3%) for predicted quality-controlled (QC) segmentations in virtual native enhancement (VNE) images. The left ventricular myocardium is manually segmented in red and automatically segmented in green, from different single (SCS) and combined candidate segmentation (CCS) models. The corresponding observed ground-truth (GT) Dice similarity coefficient (DSC) and predicted DSC are provided at the bottom. *Reproduced from Gonzales et al (2023) Frontiers in Cardiovascular Medicine 10* [138], *published under CC-BY 4.0.*

4.3.2 Comparative analysis

The comparative analysis (Table 4.1) underscores the contributions of GAN-generated VNE data as a form of data augmentation and the QCD segmentation framework as an automated quality control mechanism. Firstly, the individual performance of the deepest U-Net matched the ensemble performance of the QCD segmentation framework, which also assessed the quality of the resulting segmentation. When evaluating the results across all datasets for both training and testing, the deepest

Table 4.1: Comparative analysis of individual models (U-Net with depth of 6 levels) and quality control-driven (QCD) segmentation framework

Models and traning data		**Test data**								
		LGE (n = 309)			**VNE (n = 309)**			**both (n = 618)**		
Model	Set	DSC	MAE	Acc.	DSC	MAE	Acc.	DSC	MAE	Acc.
U-Net	LGE	0.836	-	-	0.838	-	-	0.837	-	-
U-Net	VNE	0.791	-	-	0.824	-	-	0.807	-	-
U-Net	both	0.844	-	-	**0.846**	-	-	0.845	-	-
QCD	LGE	0.835	**0.042**	**0.971**	0.838	0.046	0.922	0.837	0.044	0.947
QCD	VNE	0.799	0.057	0.922	0.833	**0.041**	**0.958**	0.816	0.049	0.940
QCD	both	**0.845**	0.042	0.958	0.845	0.043	0.945	**0.845**	**0.043**	**0.951**

The models were trained on late gadolinium enhancement (LGE; n = 4,092) and / or virtual native enhancement (VNE; n = 2,917) data and tested on LGE (n = 309) and / or VNE (n = 309) data, evaluated by their segmentation performance with the mean Dice similarity coefficient (DSC) and the quality predictive capacity with the mean absolute error (MAE) and the binary classification accuracy (Acc.). The best results are highlighted in bold. *Reproduced from Gonzales et al (2023) Frontiers in Cardiovascular Medicine 10* [138], *published under CC-BY 4.0.*

U-Net achieved a segmentation accuracy of 0.845 ± 0.070, which was similar to the segmentation accuracy of the QCD framework. Secondly, the deepest U-Net trained solely on LGE data fully generalised to the VNE test set, achieving DSC values of 0.836 ± 0.082 for LGE and 0.838 ± 0.075 for VNE. Likewise, the deepest U-Net trained only on VNE data, with 30% less data, also generalised well to the LGE test set, showing DSC values of 0.791 ± 0.119 for LGE and 0.824 ± 0.084 for VNE. These results, consistent with the findings for the ensemble framework, support the similarity between VNE and LGE images and validate the data augmentation approach as yielding comparable performance to traditional methods.

Thirdly, the inclusion of GAN-generated data consistently improved performance in all experiments for both the individual model and the QCD segmentation framework. For example, the QCD framework trained exclusively on LGE data achieved a mean DSC of 0.835 ± 0.082 for LGE and 0.838 ± 0.080 for VNE test sets. In contrast, the framework trained on both LGE and VNE data showed improved segmentation accuracy, with mean DSC values of 0.845 ± 0.075 for LGE and 0.845 ± 0.071 for

VNE. This highlights the benefits of incorporating VNE data into the training process. The quality prediction capability also improved when training and testing were conducted on both datasets. Overall, there were no significant differences observed between the LGE and VNE test sets across the experiments.

4.4 Discussion

This Chapter demonstrated that the proposed framework, which integrates GAN-generated VNE data and an automated quality control mechanism, significantly enhances the accuracy and reliability of LGE and VNE segmentation under most scenarios. The comparative analysis highlighted the benefits of incorporating VNE data, which generally offers better image quality and consistency [25, 37], into the training process. This, in turn, improved the effectiveness of the ensemble framework in segmentation tasks. The framework proved robust when applied to both LGE and VNE data, establishing a reliable pipeline for automated segmentation in both the established and emerging contrast-agent-free imaging modalities. This development paves the way for faster and more accurate diagnoses of myocardial damage.

One of the persistent challenges in developing robust and reliable deep learning models for medical image segmentation, particularly in LGE segmentation, is data scarcity or limited access. The high costs and ethical considerations involved in acquiring, labelling, and sharing large-scale annotated datasets often lead to an insufficient representation of diverse and rare pathological cases [71]. As a result, models may underperform or fail to generalise well to unseen cases [148], limiting their clinical utility. The generation of VNE images [25, 37] provides a data augmentation technique with more reliable image contrast, significantly improving over previous methods using synthetic LGE images [169, 170], which were not clinically validated and not specifically designed to display LGE lesion signals. Addressing data scarcity by generating VNE images, as demonstrated in this study, can augment or potentially replace existing LGE data, leading to more robust and reliable models that can better manage the complexity of clinical cases, ultimately enhancing patient care.

The study emphasised the integration of an automated quality assurance framework, developed using a traditional encoder-decoder U-Net architecture [89] with varying depths. This quality control-driven strategy delivered reliable quality predictions, which are crucial for clinical decision-making. The regression-based quality prediction approach allows for further exploitation of the diversity among different candidate models, providing a deterministic way to assess the agreement between candidates, which has proven more effective than emerging Monte Carlo-based quality assurance methods [107]. Future work may involve incorporating newer network architectures and advanced pre-processing techniques to further increase diversity, from existing LGE segmentation methods [171] to spatial transformation-based preprocessing [67, 172]. Future research will explore extending the dataset, evaluating different candidate models, and assessing scar burden.

The clinical implications of the proposed QCD ensemble framework are significant, introducing an automated quality control mechanism for the first time in automated LGE segmentation, thereby improving both accuracy and reliability. The quality control-driven strategy facilitates the identification and refinement of suboptimal segmentations, enhancing the system's efficiency and reliability, and supporting broader adoption in clinical workflows. This improvement streamlines the diagnostic process, reduces variability in contouring, and increases clinician confidence in automated segmentation results. It potentially simplifies routine scar burden quantification, informs better treatment decisions, and improves patient outcomes.

However, this work has some limitations. The primary focus has been on the fundamental task of myocardial delineation, without delving into the complexities of scar tissue quantification [173]. In particular, diffuse, less structured, and scattered lesions require lower segmentation thresholds and are subject to numerous methodological choices and biases in ground truth data [174, 175]. While our model was evaluated on a comprehensive international database, allowing for potential generalisation to various conditions requiring LGEs, the findings were mainly concentrated on patients with hypertrophic cardiomyopathy and MI, suggesting a need for future validation across a broader range of pathologies. Lastly, the choice of a DSC threshold of 0.7 for distinguishing between acceptable and unacceptable segmentations, although effec-

tive in our prior work [106, 107, 157], may not encompass all geometric properties of cardiac structures [148]. We anticipate that the proposed methods will generalise well across various thresholds and applications, but further research is needed to explore the available options for routine clinical use beyond the scope of this study.

4.5 Conclusion

This study introduces a novel method for automated LGE and VNE segmentation that addresses the issues of limited training data and the absence of quality control in clinical applications. By utilising GAN-generated VNE images and integrating an automated quality control mechanism, we have demonstrated the potential for enhanced segmentation performance and reliability. This framework can be effortlessly incorporated into clinical workflows, offering an efficient, quality-controlled, and dependable tool for clinicians in the diagnosis and management of patients with myocardial damage. The utilisation of motion-corrected T1 maps from Chapter 2 could further enhance the quality of VNE images derived from T1 maps and cines, potentially increasing the accuracy of this segmentation method. Moreover, the detailed landmark annotation described in Chapter 5 can enable its segmental analysis, thus broadening the clinical utility of this approach.

Chapter 5

Automated post-processing II: landmark annotation for standardised segmentation through residual neural networks

Chapters 2, 3, and 4 have progressively enhanced the clinical applicability of T1 mapping-based modalities, by correcting motion artefacts, introducing a new visualisation concept, and establishing robust automatic segmentation methods for virtual native enhancement (VNE) images. This Chapter focuses on one further necessary stage for a fully-automated T1 mapping image post-processing framework: the annotation of anatomical landmarks, such as the anterior RV insertion point and the LV centre point in ventricular short-axis images, which are critical for standardised AHA segmentation of the LV myocardium [176]. For this purpose, a residual neural network framework is employed to accurately track these key landmarks across various CMR modalities: across all T1 map-based data (native, rest, stress, and VNE), and also applicable to LGE images in both magnitude phase-sensitive inversion recovery (PSIR) formats. The methods developed in this Chapter illustrate the possibility for integrating multiple CMR imaging modalities for a unified approach to comprehensive CMR image analysis.

The work in this Chapter has been presented [177, 178] in:

177. **Gonzales RA**, Manrique AL, Burrage MK, Thomas KE, Altun İ, Huang W-H, Yun C-H, Zhang Q, Ferreira VM, Piechnik SK. Deep learning for automated insertion point annotation of CMR T1 maps. In: *2024 IEEE 21st International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2024. ORA

178. **Gonzales RA**, Zhang Q, Burrage MK, Altun İ, Heiberg E, Peters DC, Neubauer S, Ferreira VM, Piechnik SK. Deep learning for automated insertion point annotation of CMR late gadolinium enhancement and virtual native enhancement images. In: *SCMR 26th Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2023. ORA

5.1 Background

As discussed earlier in this thesis, CMR imaging has become a cornerstone in the assessment and diagnosis of various cardiac conditions, offering detailed insights into myocardial tissue characteristics and function [179]. Among the various CMR techniques, quantitative T1 mapping, LGE, and VNE can provide comprehensive myocardial tissue characterisation. T1 mapping aids in characterising myocardial tissue properties [19], LGE identifies and quantifies focal myocardial scar tissue [17], and VNE offers a contrast-agent-free alternative that mimics LGE appearances using native signals from cines and T1 maps [25]. Accurate segmentation of the myocardium, adhering to the AHA 16-segment model [176], is essential for accurate assessment of the spatial information in myocardial tissue characterisation, such as correspondence to coronary artery territories in ischaemic heart disease, and spatial description in non-ischaemic pathologies. However, this requires precise annotation of anatomical landmarks such as the anterior RV insertion point and the LV centre point in short-axis images, which are critical for segmental quantification.

Despite advancements in automated myocardial segmentation [65], annotating the anterior RV insertion point and the LV centre point remains a challenging and time-consuming task, typically taking approximately 20 seconds per sample [180]. When applied to large datasets, this cumulative time significantly increases the overall workload. Manual annotation is not only time-consuming but also prone to inter- and intra-observer variability [181], leading to inconsistencies in segmental analysis. Previous studies have attempted to address this issue using machine learning [182, 183] and deep learning [180, 184, 185, 186, 187] approaches, demonstrating promising results in specific CMR modalities such as cines [182, 184, 185, 187], LGE images [186] and specialised sequences [180, 183]. However, these models were typically trained and validated on modality-specific datasets, or without including ShMOLLI T1 maps or VNE images, limiting their generalisability and utility in clinical practice, where multiple CMR modalities are often used in combination, or where these methods are used, for a comprehensive cardiac assessment.

In this Chapter, we aim to overcome these limitations by developing a single, robust residual neural framework that can accurately regress the Cartesian coordi-

nates of the anterior RV insertion point and the LV centre point across various CMR modalities (rather than just a single modality), including T1 maps (rest, stress, and post-contrast), LGE (magnitude and PSIR), and VNE. By leveraging a diverse, multimodal dataset for training, we hypothesised that a generalist model can achieve superior performance compared to modality-specific models, providing a unified solution for automated myocardial segmentation. This approach not only enhances the accuracy and consistency of segmental quantification but can also streamline the workflow in clinical and research settings, facilitating the broader adoption of advanced CMR techniques.

5.2 Methods

5.2.1 Imaging data

The imaging dataset (Table 5.1) consisted of 16,160 short-axis CMR images obtained from 1,299 human participants, sourced from the Hypertrophic Cardiomyopathy Registry (HCMR) [159], Oxford Centre for Clinical Magnetic Resonance Research (OCMR) [34], and Mackay Memorial Hospital (MMH), all using Siemens magnetic resonance scanners. The modalities used in this study were ShMOLLI native and post-contrast T1 maps [76], LGE images in both magnitude and PSIR formats, and VNE images [25]. The native T1 maps were acquired at rest (5,159 T1 maps from 1,291 human participants) and stress (1,592 T1 maps from 186 human participants). Each scan included up to three short-axis slices, and the images were manually annotated by four trained observers to identify the LV centre point and the anterior RV insertion point, using MC-ROI (dedicated inhouse software developed by SKP in Interactive Data Language v6.1, Exelis Visual Information Solutions, Boulder, Colorado, USA). The dataset comprised 43 healthy human volunteers and 1,256 human patients with conditions including HCM (n=1,111), CAD (n=94), end-stage renal disease (n=33), breast cancer (n=14), and aortic stenosis (n=4). The median (IQR: interquartile range) interpolated spatial resolution for each modality was 0.94 (0.94-1.04) mm for native T1 maps, 0.9 (0.94-0.94) mm for post-contrast T1 maps, 1.41 (1.37-1.46) mm for LGE magnitude and PSIR images, and 0.94 (0.94-1.04) mm for

Table 5.1: Composition of the data used in this Chapter

Modality	HCMR	OCMR	MMH	Training	Test
$T1_{native}$	1,104 (3,076)	155 (3,163)	33 (512)	1,035 (5,201)	257 (1,550)
$T1_{post}$	-	20 (464)	9 (378)	26 (762)	3 (80)
LGE_{mag}	987 (2,621)	-	-	778 (1,975)	209 (646)
LGE_{PSIR}	999 (2,741)	-	-	790 (2,095)	209 (646)
VNE	1,107 (3,205)	-	-	898 (2,559)	209 (646)
All	1,111 (11,643)	155 (3,627)	33 (890)	1,042 (12,592)	257 (3,568)

Count represents the number of human participants (number of images).

VNE images. The field of view for the VNE data was notably smaller than the other modalities, as the images were generated with a fixed, centre-cropped view of the heart in line with previously published work [25]. The dataset was randomly divided into training and testing sets, with approximately 80% of each modality allocated for training and the remaining 20% for testing, avoiding data leakage. In total, 1,042 subjects (12,592 images) were used for training, and separate 257 subjects (3,568 images) were reserved for testing.

5.2.2 Data preprocessing

The entire field of view was retained for all images (Figure 5.1a). The pixel spacing for each image was interpolated to a uniform 1 mm to standardise spatial resolution. To ensure consistent input dimensions, all images were zero-padded, and centre-cropped if needed, to a fixed size of 384 x 384 pixels. Each image underwent standardisation based on mean and standard deviation of the non-zero pixel intensities to enhance training stability [93]. Additionally, the anterior RV insertion point annotation was automatically refined to align with the intersection between the RV entry line—formed by the line connecting the LV centre point and the anterior RV insertion point—and the LV epicardial contour. For ease of training and to improve gradient flow during the learning process, the coordinates of all annotations were normalised to a 0-1 range, dividing each Cartesian coordinate by the image length [188].

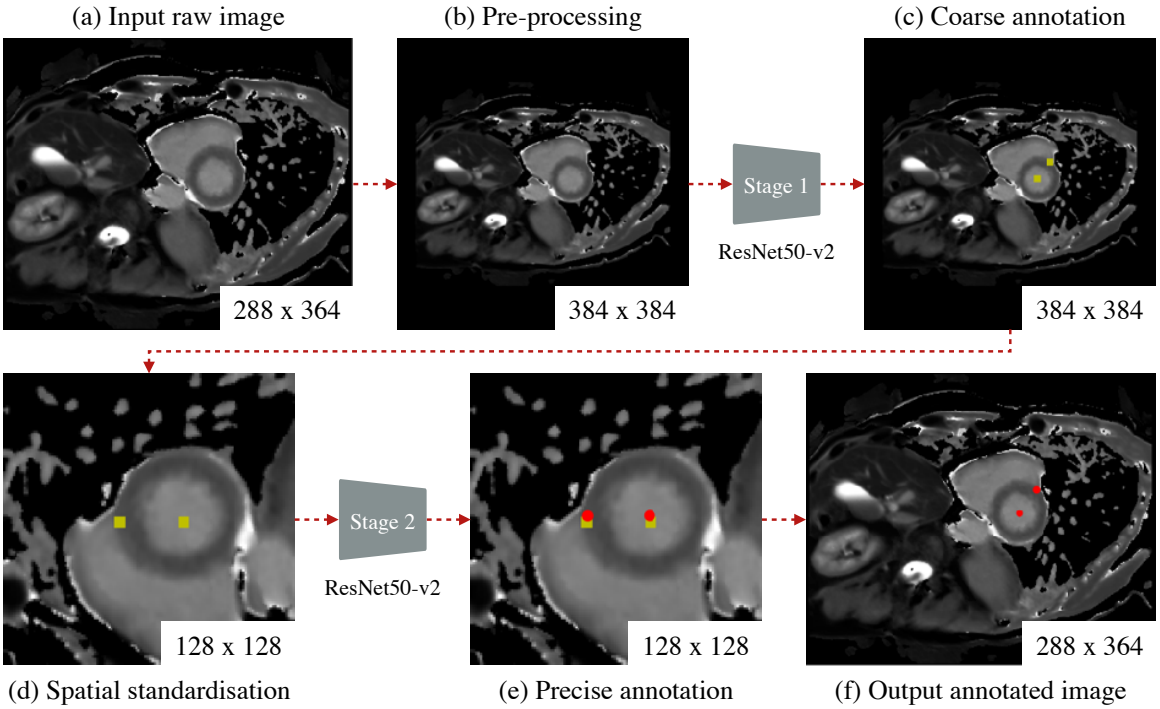


Figure 5.1: Dual-stage tracking framework for automated annotation of the anterior right ventricular insertion point and left ventricular centre point. The process begins with (a) the input of the raw image, followed by (b) pre-processing where the image is zero-padded to standard dimensions (384 x 384 pixels) and the pixel spacing is adjusted to 1 mm. (c) In Stage 1, a coarse annotation (in yellow) is generated. (d) The image is then spatially standardised, rotating and cropping it to 128 x 128 pixels, ensuring the RV entry line is oriented horizontally towards 9 o'clock. (e) In Stage 2, the model refines the annotation (in red), producing a precise output that is then (f) mapped back onto the original image dimensions. Image brightness and contrast were increased by 40% and 20%, respectively, for visualisation purposes.

5.2.3 Deep learning framework

The deep learning framework employs a dual-stage residual neural network approach [47], using ResNet-50 models [189] in both stages for precise landmark annotation (Figure 5.1). In Stage 1, a ResNet-50v2 model [190], pre-trained on ImageNet [191] and fine-tuned on the CMR dataset, is re-trained on pre-processed images (Figure 5.1b) to predict the Cartesian coordinates of the anterior RV insertion point and the LV centre point. This stage provides a coarse annotation that serves as the initial localisation of these landmarks (Figure 5.1c). The output coordinates are then used in Stage 2, where images undergo spatial standardisation: they are rotated so that the RV entry line is flattened with the RV anterior point on the left and the LV centre

point on the right, with the RV positioned in the lower left part of the LV. The image is then cropped around the midpoint between these points to a size of 128 x 128 pixels (Figure 5.1d). The second ResNet-50v2 model, trained on these standardised images, refines the prediction with higher accuracy (Figure 5.1e-f). This two-stage approach has been previously proposed to progressively enhances the model's ability to accurately annotate landmarks by leveraging both the initial coarse localisation and subsequent image standardisation [47].

5.2.4 Model training

Model training was initiated using transfer learning from ImageNet [191], where the pre-trained weights were employed for initialisation, but the layers were not kept frozen during training. To enhance the model's generalisation capability, extensive data augmentation techniques were applied [192]. These augmentations included random rotations, scaling, Gaussian noise addition, Gaussian blur, brightness adjustment, contrast adjustment, simulation of low resolution, gamma augmentation, and, only for the first stage, random horizontal and vertical flipping of the images. The training process employed mean squared error as the loss function and the Adam optimiser [94], with a learning rate of 0.0001 for optimal convergence. Training continued for up to 1,000 epochs, with early stopping implemented when no further improvement was observed between training and validation loss. To monitor the model's performance throughout the training process, 10% of the training data was internally allocated for validation purposes. Further technical specifications are detailed in Appendix A.4. The entire training pipeline was executed within the TensorFlow framework [95], with an NVIDIA GeForce RTX 3090 GPU.

5.2.5 Specialist model evaluation

The specialist models were trained on the specified individual CMR modalities, including native T1 maps, post-contrast T1 maps, LGE magnitude images, LGE PSIR images, and VNE images (Table 5.1). The models were then tested on both the same modality they were trained on and all other modalities to assess statistically their generalisation capabilities. Performance was evaluated using two primary metrics:

Euclidean distance, measured as the distance between the predicted and actual coordinates of the LV centre and anterior RV insertion points, and angular distance, calculated by comparing the angle formed by the predicted RV entry line to the actual RV entry line. The overall generalisation error rate (GER) was expressed as the median (IQR) of the ratios of the specialist model's errors deployed to other modalities (out-of-distribution error) over its error on the same modality (in-distribution error).

5.2.6 Generalist model evaluation

The generalist model was trained on the whole dataset that consisted of all CMR modalities for specialist models mixed together (Table 5.1). It was tested both on individual modalities and the entire dataset to statistically evaluate its ability to generalise across different CMR modalities. The same performance metrics used for the specialist models—Euclidean distance and angular distance—were applied to the generalist model. The GER was expressed as the median (IQR) of the ratios of the generalist model's errors deployed to each modality over its error when individually deployed to all the modalities. The improvement from the specialist models was expressed as the median (IQR) of the ratios of the specialist models' in-distribution errors over the generalist model error deployed to all modalities.

5.2.7 Human observer variability analysis

To provide a benchmark for the model performance we used preexisting data from training of 11 human observers manually annotating a separate set of 42 native T1 maps and up to five sessions each [161]. Annotations from the first session were excluded from the analysis. Inter-observer variability was assessed by comparing the average annotations of each observer against the averages of the others, while intra-observer variability, available for 7 observers, was calculated by comparing the annotations (Cartesian coordinates of both points) made by the same observer across the repeated sessions. The specialist native T1 and generalist 2-stage models, trained with data augmentation, were deployed on the same 42 native T1 maps and compared against the average annotations of the available observers. The variability was evaluated using the same metrics as the main analysis.

5.2.8 Statistical analysis

Model performance was compared across three configurations: (1) a baseline configuration with no data augmentation, (2) a configuration with data augmentation, and (3) a two-stage configuration with data augmentation. These configurations were used for both specialist and generalist models to evaluate improvements in accuracy and generalisation. Non-parametric descriptors were used as the metrics' distribution did not pass the Shapiro-Wilk test. For statistical comparisons involving paired (for model evaluation) and non-paired data (for human variability analysis), the Wilcoxon and Mann-Whitney U Test tests were conducted to assess statistical significance, respectively. A p-value threshold of <0.05 was set for statistical significance.

5.3 Results

5.3.1 Specialist model performance

In the baseline configuration (Table 5.2a), the specialist models showed modest performance on their respective modalities and substantially worsened when applied to other modalities. The overall GER in this configuration was 2.4 (1.7-3.0), i.e., the out-of-distribution errors were typically 2.4 times significantly higher than the in-distribution errors (all $p<0.001$). For instance, the VNE specialist model had a median Euclidean distance error of 7.1 (4.6-10.9) mm on its own data, but this error significantly increased when tested on other modalities (all $p<0.001$), with a GER of 2.9 (2.8-3.0), due to the inherent image variability. This is particularly noticeable for the combinations with VNE data where there were substantial differences in the fields of view.

Performance improved as more advanced configurations were applied. With data augmentation (Table 5.2b), the VNE model's error dropped from 7.1 to 3.4 (2.1-4.8) mm on its own data, though cross-modality performance remained suboptimal, with a GER of 6.4 (5.9-7.1). Overall, compared to the baseline configuration, every specialist model significantly improved 1.9 (1.5-2.1) times (all $p<0.001$), but the overall GER was 3.4 (2.3-6.2). The best performance was observed in the final configuration,

with two stages and data augmentation (Table 5.2c). In this configuration, specialist models showed an improvement on their own data and across modalities: every specialist model significantly improved 2.6 (1.5-2.8) times compared to the baseline configuration (all $p<0.001$), and the overall GER was 2.6 (1.4-3.6).

The same evaluation in angular distance exhibited similar trends but at a lower magnitude (Table 5.3). Every specialist model yielded a better performance within in-distribution data compared to its out-of-distribution data testing. The GERs were 1.8 (1.6-2.6), 2.0 (1.2-4.3), and 2.0 (1.3-4.4) for the first, second and third configurations, respectively. The improvements from the baseline configuration were 1.1 (1.0-1.2) and 1.1 (1.0-1.1) for the second and third configurations, respectively. All these improvements were significant (all $p<0.05$), except for the specialist native and post-contrast T1 models with no significant difference in both configurations (all $p\approx0.2$).

5.3.2 Generalist model performance

The generalist model performed significantly better than almost all specialist models, across all configurations in both Euclidean and angular distance. For the Euclidean distance error assessment, in the first configuration (Table 5.2a), the generalist model significantly improved the specialist models' in-distribution errors 2.19 (1.45-2.28) times (all $p<0.001$). In the advanced configurations, this improvement was also significant but marginal with a rate of 1.04 (1.04-1.34) for the second configuration (Table 5.2b, all $p<0.01$) and 1.03 (1.01-1.04) for the third configuration (Table 5.2c, all $p<0.05$, except VNE with $p=0.11$). In both configurations, the specialist native T1 models were significantly better than the generalist models (both $p<0.001$).

For the angular distance error assessment, almost all improvements were significant. Specifically, these improvements were 1.15 (1.02-1.27) for the first configuration (Table 5.3a, all $p<0.05$), 1.03 (1.03-1.09) for the second configuration (Table 5.3b, all $p<0.05$, except for LGE PSIR and VNE with $p=0.09$ and $p=0.23$); and 1.08 (1.04-1.16) for the third configuration (Table 5.3c, all $p<0.05$, except for LGE PSIR with $p=0.08$). The specialist models that were significantly better ($p<0.05$) than the generalist models were the ones trained on post-contrast T1 maps in the first configuration, and on native T1 maps in the second and third configurations.

Table 5.2: Model performance in tracking the right ventricular entry line across short-axis CMR modalities and configurations, measured in median (IQR) Euclidean distance (mm) against manually annotated landmarks

(a) Performance with 1 stage and without data augmentation

Train \ Test	$T1_{native}$	$T1_{post}$	LGE_{mag}	LGE_{PSIR}	VNE	All	GER
$T1_{native}$	2.6 (1.4-5.2)	6.4 (3.2-9.7)	12.3 (7.3-19.7)	13.7 (8.3-21.4)	9.5 (4.2-16.0)	6.4 (2.6-13.5)	4.1 (3.3-4.8)
$T1_{post}$	16.6 (10.4-24.2)	10.5 (7.2-13.0)	20.6 (13.6-28.1)	21.5 (14.2-30.2)	13.0 (9.1-16.5)	16.6 (10.9-24.3)	1.8 (1.5-2.0)
LGE_{mag}	12.2 (6.4-19.9)	21.2 (14.4-41.4)	4.5 (2.7-7.8)	9.5 (5.4-15.4)	7.5 (4.7-11.4)	8.9 (4.7-15.6)	2.4 (2.0-3.2)
LGE_{PSIR}	12.3 (7.2-19.8)	17.2 (12.1-43.0)	9.5 (5.7-14.5)	7.4 (4.5-11.6)	11.2 (8.2-14.9)	10.7 (6.5-16.5)	1.6 (1.5-1.8)
VNE	18.1 (11.3-26.8)	21.1 (9.5-26.6)	20.6 (11.6-29.0)	22.8 (14.0-31.1)	7.1 (4.6-10.9)	16.6 (9.1-26.3)	2.9 (2.8-3.0)
All	**2.2** (1.2-4.4)	**4.3** (2.4-7.8)	**2.9** (1.8-5.1)	**2.9** (1.7-5.0)	**2.8** (1.8-4.3)	**2.7** (1.5-4.7)	**1.09**† (1.03-1.10)

(b) Performance with 1 stage and data augmentation

Train \ Test	$T1_{native}$	$T1_{post}$	LGE_{mag}	LGE_{PSIR}	VNE	All	GER
$T1_{native}$	**1.9**† (1.0-4.1)	5.5 (4.4-8.4)	11.7 (4.3-23.4)	25.2 (16.9-36.3)	10.0 (2.9-22.1)	5.3 (1.8-18.8)	5.8 (4.8-8.1)
$T1_{post}$	9.9 (5.9-15.2)	5.1 (3.9-6.1)	19.0 (11.7-28.7)	15.2 (9.6-23.5)	14.9 (9.8-25.0)	12.6 (7.4-20.8)	3.0 (2.7-3.2)
LGE_{mag}	6.9 (2.6-17.4)	24.4 (15.0-40.4)	2.9 (1.7-4.8)	8.4 (4.1-14.4)	4.1 (2.2-10.2)	5.3 (2.4-13.5)	2.7 (2.2-4.3)
LGE_{PSIR}	4.0 (2.0-9.4)	16.2 (6.4-25.5)	3.1 (1.8-5.2)	2.8 (1.6-5.0)	2.8 (1.7-5.2)	3.4 (1.8-6.8)	1.3 (1.1-2.5)
VNE	28.1 (13.8-42.4)	18.0 (8.6-26.3)	22.6 (9.4-31.0)	20.6 (9.5-30.3)	3.4 (2.1-5.8)	19.0 (6.6-32.8)	6.4 (5.9-7.1)
All	2.1 (1.2-4.4)	**2.5** (1.5-4.6)	**2.7**† (1.6-4.6)	**2.7**† (1.6-4.6)	**2.4** (1.5-3.9)	**2.4** (1.4-4.4)	**1.04*** (0.98-1.12)

(c) Performance with 2 stages and data augmentation

Train \ Test	$T1_{native}$	$T1_{post}$	LGE_{mag}	LGE_{PSIR}	VNE	All	GER
$T1_{native}$	**1.8*** (1.0-4.0)	5.6 (3.6-7.5)	4.6 (2.3-11.3)	8.1 (3.1-20.9)	4.3 (1.8-15.4)	3.3 (1.5-8.5)	2.8 (2.5-3.4)
$T1_{post}$	4.5 (2.4-8.2)	3.1 (1.1-5.5)	12.2 (6.3-21.7)	9.0 (4.8-16.3)	8.3 (4.9-14.9)	6.8 (3.4-13.1)	2.8 (2.4-3.2)
LGE_{mag}	6.3 (2.2-18.3)	25.7 (7.9-46.3)	2.8 (1.7-4.6)	2.9 (1.6-5.8)	3.2 (1.7-7.9)	3.7 (1.9-10.0)	1.7 (1.1-4.0)
LGE_{PSIR}	3.5 (1.6-9.6)	9.3 (5.5-28.1)	2.8 (1.7-4.9)	2.6 (1.6-4.4)	2.9 (1.7-6.9)	3.1 (1.7-7.0)	1.2 (1.1-1.9)
VNE	14.8 (3.5-44.7)	12.3 (8.2-16.8)	5.3 (2.2-13.4)	4.9 (2.2-13.6)	1.9 (1.1-4.0)	5.9 (2.1-20.3)	4.6 (2.7-6.8)
All	1.9 (1.0-4.4)	**2.3*** (1.2-4.6)	**2.6*** (1.6-4.6)	**2.5*** (1.5-4.7)	**1.9*** (1.1-4.0)	**2.2*** (1.2-4.4)	**1.07**† (0.88-1.16)

Colours represent a gradient of error magnitude from green (≤2 mm) through yellow (11 mm) to red (≥20 mm). Bold values indicate the lowest median error for each test modality (column), or generalisation error rate (GER), per configuration. Values marked with * represent the overall lowest error in each column across all configurations, while † indicates statistical similarity to the * value (Wilcoxon test).

Table 5.3: Model performance in tracking the right ventricular entry line across short-axis CMR modalities and configurations, measured in median (IQR) angular distance (°) against manually annotated landmarks

(a) Performance with 1 stage and without data augmentation

Train \ Test	$T1_{native}$	$T1_{post}$	LGE_{mag}	LGE_{PSIR}	VNE	All	GER
$T1_{native}$	6.1 (2.7-10.6)	14.6 (8.8-18.6)	13.7 (6.6-22.4)	10.3 (4.9-18.1)	11.0 (5.3-17.9)	8.6 (3.8-15.9)	2.0 (1.8-2.3)
$T1_{post}$	15.6 (7.0-28.5)	**7.9** (5.0-15.8)	14.5 (7.0-24.5)	20.0 (8.9-33.1)	14.5 (6.9-25.5)	15.6 (7.1-27.7)	1.9 (1.8-2.1)
LGE_{mag}	20.3 (10.6-32.4)	42.9 (19.8-50.1)	6.7 (3.0-11.8)	6.8 (3.2-12.3)	6.4 (2.6-12.3)	10.8 (4.6-22.3)	2.0 (1.0-3.9)
LGE_{PSIR}	12.7 (6.1-24.6)	50.3 (16.7-59.0)	7.4 (3.4-12.5)	7.2 (3.4-12.6)	7.5 (3.4-13.3)	9.4 (4.3-17.8)	1.4 (1.0-3.1)
VNE	16.4 (8.0-30.9)	39.7 (32.2-46.8)	9.0 (4.2-15.0)	11.0 (4.6-20.1)	6.2 (3.1-11.0)	11.4 (5.1-21.6)	2.2 (1.7-3.6)
All	**5.8**† (2.6-10.4)	13.7 (7.9-18.8)	**5.4*** (2.5-10.6)	**5.1*** (2.6-10.1)	**4.9**† (2.4-8.7)	**5.5*** (2.6-10.4)	**1.0*** (0.9-1.1)

(b) Performance with 1 stage and data augmentation

Train \ Test	$T1_{native}$	$T1_{post}$	LGE_{mag}	LGE_{PSIR}	VNE	All	GER
$T1_{native}$	**5.6*** (2.4-10.0)	6.4 (2.4-12.2)	12.3 (4.8-23.2)	11.3 (4.9-19.6)	23.8 (11.1-41.6)	9.0 (3.8-19.0)	2.1 (1.8-2.7)
$T1_{post}$	11.8 (5.2-21.3)	11.0 (7.3-15.7)	13.1 (6.3-23.3)	12.7 (5.9-21.8)	16.1 (7.8-28.7)	12.7 (6.0-22.9)	1.2 (1.1-1.3)
LGE_{mag}	26.6 (13.9-45.2)	34.4 (14.3-56.6)	6.2 (3.0-10.5)	8.1 (3.4-13.9)	12.0 (5.7-21.6)	13.8 (6.0-29.3)	3.1 (1.8-4.6)
LGE_{PSIR}	13.5 (6.6-23.7)	28.0 (10.5-54.2)	6.0 (3.1-11.5)	5.7 (2.6-10.8)	6.3 (3.0-11.1)	8.6 (3.9-16.8)	1.7 (1.1-3.0)
VNE	40.3 (19.1-64.9)	35.5 (25.1-60.8)	8.8 (4.2-15.5)	10.4 (4.2-19.2)	5.1 (2.3-8.7)	14.7 (5.6-38.5)	4.5 (2.0-7.2)
All	6.2 (2.9-10.9)	**5.6*** (2.8-9.2)	**5.7**† (2.6-9.7)	**5.7**† (2.5-10.1)	**4.9**† (2.2-8.7)	**5.8**† (2.6-10.2)	**1.0**† (1.0-1.0)

(c) Performance with 2 stages and data augmentation

Train \ Test	$T1_{native}$	$T1_{post}$	LGE_{mag}	LGE_{PSIR}	VNE	All	GER
$T1_{native}$	**5.6**† (2.4-10.1)	6.2 (2.4-13.0)	11.4 (5.0-21.5)	11.4 (5.0-19.5)	23.9 (11.9-38.9)	9.0 (3.7-19.0)	2.0 (1.8-2.6)
$T1_{post}$	10.1 (4.8-17.9)	10.2 (7.7-16.1)	14.4 (6.9-25.5)	12.0 (5.6-20.9)	12.6 (5.8-22.4)	11.6 (5.4-20.9)	1.2 (1.1-1.3)
LGE_{mag}	28.6 (16.1-49.3)	44.9 (16.5-77.4)	6.1 (3.0-10.6)	7.7 (3.6-13.8)	14.0 (6.5-22.3)	14.9 (6.3-30.4)	3.5 (2.0-5.4)
LGE_{PSIR}	13.8 (6.8-24.0)	27.4 (10.6-52.5)	6.0 (3.1-11.4)	5.6 (2.6-10.9)	10.7 (4.5-22.5)	9.7 (4.2-19.3)	2.2 (1.7-3.1)
VNE	32.0 (15.7-57.0)	31.6 (21.9-43.4)	10.8 (4.9-18.3)	11.0 (5.0-20.1)	5.9 (2.8-10.4)	14.9 (6.1-34.1)	3.6 (1.9-5.4)
All	6.1 (2.8-11.0)	**5.9**† (3.0-9.2)	**5.7**† (2.6-9.7)	**5.6**† (2.5-10.1)	**4.8*** (2.2-8.7)	**5.7**† (2.6-10.2)	**1.0**† (1.0-1.0)

Colours represent a gradient of error magnitude from green (≤5°) through yellow (7.5°) to red (≥20°). Bold values indicate the lowest median error for each test modality (column), or generalisation error rate (GER), per configuration. Values marked with * represent the overall lowest error in each column across all configurations, while † indicates statistical similarity to the * value (Wilcoxon test).

When escalating to the second and third configurations, the performance was overall improved. In the Euclidean distance error assessment, the best performing models from the third configuration (Table 5.2c, marked with *) were significantly better than their counterparts (all $p<0.001$), except for the specialist native T1 model and the generalist model, tested on LGE magnitude and PSIR data, from the second configuration (Table 5.2b, marked with †) that had no significant difference (all $p>0.05$). For the angular distance error assessment, the best performing models from the third configuration (Table 5.3c, marked with * and †) had no significant difference with their counterparts in the other configurations (all $p>0.1$), except with the specialist post-contrast T1 model that had significantly higher error ($p<0.001$). The GERs, for both assessments, were near 1.0, indicating that the generalist model maintained strong performance across all modalities without performance loss when compared to any single one.

5.3.3 Human observer variability analysis

The observer variability analysis (Table 5.4) revealed distinct differences between intra-observer and inter-observer variability, as well as when these were compared to the native T1 specialist and generalist models' performance on native T1 maps. For Euclidean distance, inter-observer variability was 33% higher than intra-observer variability ($p<0.001$). The variability of both specialist and generalist models were 32% and 34% significantly higher than the human inter-observer variability (both $p<0.001$), respectively. In terms of angular distance, inter-observer variability was 31% significantly higher than intra-observer variability ($p<0.001$). The inter-observer variability was numerically higher, but without significant difference, than the specialist ($p=0.74$) and the generalist ($p=0.35$) model. There were no significant differences between the specialist and generalist models in both Euclidean ($p=0.50$) and angular ($p=0.60$) distances. The IQRs of the generalist model were lower than specialist model's. At batch-processing inference with GPU, each sample took approximately 60 ms for the model with 1 stage and 165 ms for the model with 2 stages.

Table 5.4: Intra- and inter-observer human variability compared against the specialist native T1 and generalist 2-stage models with data augmentation tested on a separate set of 42 native T1 maps

Error metric	**Intra-observer variability** (n=825)	**Inter-observer variability** (n=1,838)	**Specialist $T1_{native}$ model** (n=414)	**Generalist model** (n=414)
Euclidean distance (mm)	0.9 (0.4-2.6)	1.3 (0.6-3.2)	1.7 (1.0-3.1)	1.7 (1.1-3.0)
Angular distance (°)	3.7 (1.7-7.2)	4.9 (2.4-9.3)	4.8 (2.4-9.2)	4.8 (2.1-8.4)

Error reported in median (IQR). Count (n) represents the number of evaluations of available samples and sessions from all the 11 observers.

5.4 Discussion

The results of this study demonstrate that the generalist model performed well (Figure 5.2) and significantly outperformed the specialist models across nearly all configurations and modalities, highlighting the value of a unified approach in automated CMR image analysis. The generalist model regularly achieved lower median errors in both Euclidean and angular distances compared to the specialist models, regardless of the modality on which it was tested. This suggests that training on a larger, diverse, and multi-modal datasets enables the model to generalise more effectively across different CMR modalities, reducing the necessity for individually trained modality-specific models. The addition of data augmentation and the two-stage processing approach usually aided further refinements, particularly in reducing the Euclidean distance error, although these enhancements were secondary to the broader advantage conferred by the generalist model. These findings support the potential for a generalised approach to improve the efficiency and accuracy of automated myocardial segmentation in diverse clinical contexts.

The specialist models, while effective within their respective modalities, showed limitations in generalisation when tested on different modalities. This generalisation gap underscores the overfitting risks that arise when models are trained in narrow domains. This finding also highlights the inherent trade-offs in developing models tailored to specific imaging types, where their specialised focus restricts their appli-

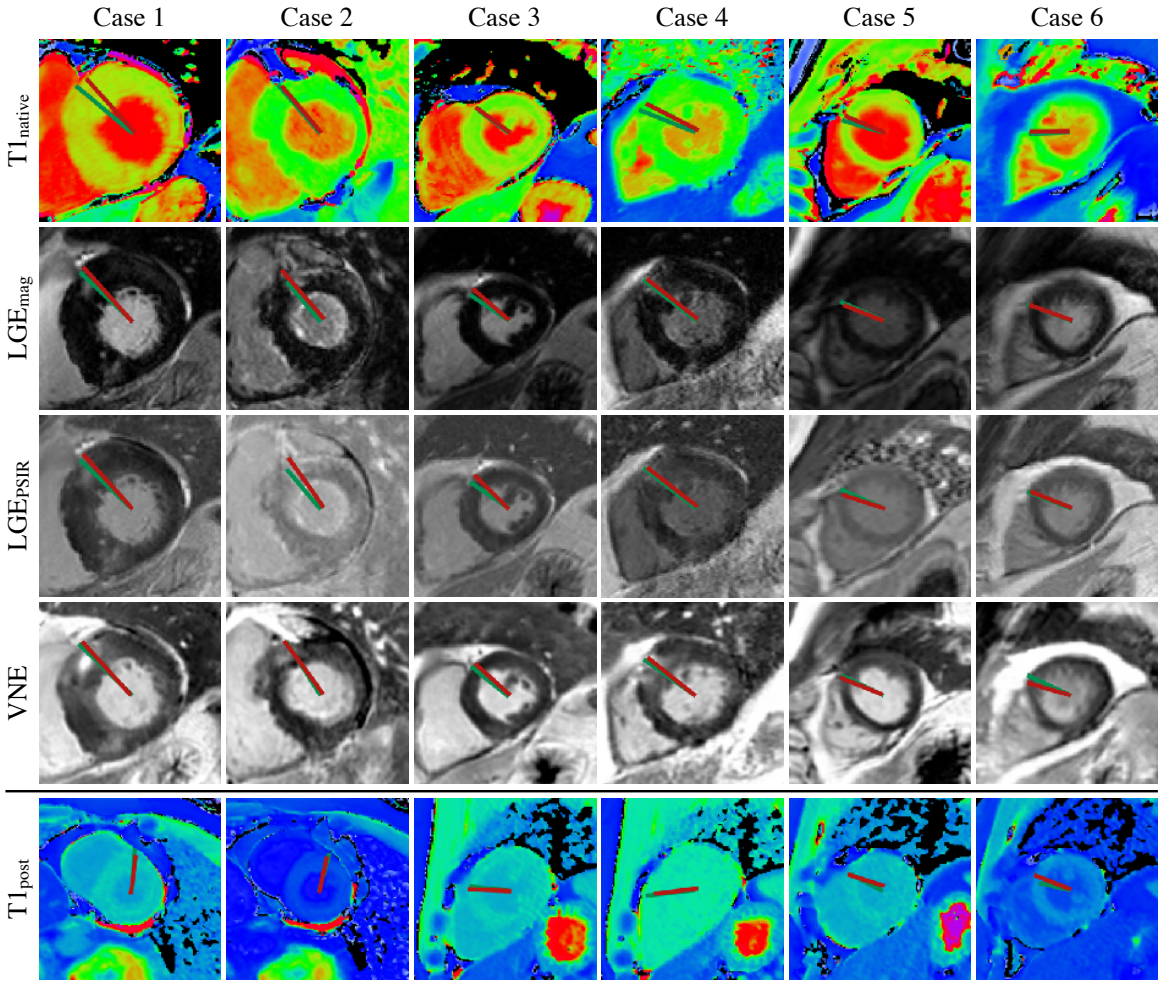


Figure 5.2: Representative examples of manual (green) and predicted (red) right ventricular entry lines in the test imaging data, by the final generalist model. CMR images from 6 different human subjects are displayed with their paired native T1 maps, late gadolinium enhancement (LGE) magnitude and phase-sensitive inversion recovery (PSIR) images, and virtual native enhancement (VNE) images, ranging from basal to apical slices. For the post-contrast T1 maps, given these do not have paired LGE or VNE data, 2 mid-ventricular samples for each of the 3 human subjects from the test set are consecutively displayed. Each sample was automatically centre-cropped to 128 mm x 128 mm for visualisation purposes.

cability in more diverse clinical scenarios. The increase in error (GER) when these models were tested outside their trained modality highlights the risks of overfitting and the potential inefficiencies in requiring multiple models to handle different CMR modalities. The generalist model's superior performance across all tested modalities thus represents a more versatile and scalable solution for comprehensive cardiac assessments, especially in cases where the is little training data [107]. The generalist

approach is better or not much worse than any specialist in all tested scenarios, which potentially simplifies the deployment of automated systems in clinical settings by reducing the computational and operational burden associated with deploying multiple specialist models one at a time to any specific sequence separately.

This potential of generalist models arises from their adaptability, making them particularly suited for handling the diverse and multimodal nature of clinical imaging. Although this study focuses on the single task of regressing coordinates, the generalist model concept is rooted in leveraging diverse data, enabling the model to perform the same task across different modalities. This aligns with the broader trend in current research toward developing foundation models in the medical field [193], which aim to handle multiple tasks across various modalities using a top-down approach. While foundation models theoretically promise superior performance [194], they currently face challenges, particularly in achieving the same level of accuracy as single-task, single-modality models in specific medical applications [195]. The approach in this study represents a bottom-up effort, demonstrating that a less specific, more general model can achieve performance comparable to or even exceeding that of specialist models. This work exemplifies that to build accountable foundation models, the development process should begin with specific tasks and gradually expand to general tasks, ensuring that performance in each task and modality matches or surpasses that of specialist models and ideally reaches expert-level accuracy. For context, the Segment Anything Model [196], a foundation model trained on an extensive dataset of photographs, performs well in tasks that do not demand extremely high accuracy, such as object identification in self-driving cars. However, its performance declines significantly when applied to medical images [197]. Conversely, TotalSegmentator [198], a generalist model tailored for segmentation using CT datasets, shows much better performance in that domain, although it still may not match the accuracy of state-of-the-art specialist models [199]. The development of generalist models is ongoing, and future advancements are likely to bring even more impressive results especially if the training and deployment reliability are as rigorously documented as outlined in this work.

When comparing the performance of the trained models to that of human observers, the model exhibited numerically higher errors in Euclidean distance and no significant difference in angular distance. This suggests that the generalist model performs at a level comparable to trained human observers, at least in one T1 mapping modality. These errors, approximately equivalent to a single pixel's resolution in the CMR images, remain within an acceptable range for clinical application and highly practical for the automated component. The model's ability to produce consistent and reliable annotations across various modalities is a notable achievement given the inherent variability in manual annotations.

5.5 Conclusion

This work underscores the effectiveness of generalist deep learning models in enhancing the precision and efficiency of anatomical landmark annotation within the CMR imaging modalities tested, serving as the capstone in the fully automated T1 mapping image post-processing framework described in this thesis. It was found that augmentation and two stage processing improve the accuracy substantially. By leveraging a dual-stage ResNet50v2 framework and applying extensive data augmentation, the generalist approach demonstrated superior accuracy over nearly all modality-specific models, save only for T1 mapping. The generalist approach showed most prominent benefits in modalities with lower training dataset sizes. Such integration of landmark annotation is pivotal for streamlining standardised myocardial segmentation, following the AHA guidelines. This holistic method not only meets clinical needs by significantly reducing the variability and time involved in manual annotation but also paves the way for future advancements in automated CMR analysis, aiming to standardise and refine diagnostic processes across varied imaging settings directly building upon the enhancements introduced in earlier Chapters, 2, 3, and 4.

Chapter 6

Summary and future work

6.1 Summary

CMR stands as an imaging gold standard to assess cardiac structure and function, additionally offering unparalleled insights into myocardial tissue characteristics, such as inflammation, infiltration and fibrosis. Among the various CMR techniques, T1 mapping has gained prominence due to its ability to provide quantitative assessments of myocardial tissue properties beyond what LGE can offer, enabling precise diagnosis and management of a wide array of cardiac conditions. However, despite its potential, the implementation of T1 mapping in routine clinical practice is hampered by several challenges. These include the susceptibility of T1 maps to motion artefacts, which can severely degrade image quality, the limited tools available for visualising T1-maps for immediate diagnostics (including stress-induced myocardial changes), the need for consistent and accurate segmentation of the LV myocardium across different modalities, and the lack of reliable, automated methods for annotating key anatomical landmarks essential for segmental analysis. In response to these challenges, this thesis sets forth four primary goals targeting selected aspects of CMR T1 mapping to enhance its clinical applicability, which may have generalisability for machine-learning-based automated imaging analysis for other similar cardiac imaging, and beyond.

The first goal (Chapter 2) addressed the pervasive issue of motion artefacts in pixel-wise T1 mapping, which can lead to significant inaccuracies in the quantification of myocardial properties. To mitigate this, the thesis proposed a CNN-based method for motion correction, specifically designed to correct such artefacts in a selected T1 mapping sequence. The CNN architecture was tailored to handle the unique challenges of CMR images, including the variability in image quality and the presence of subtle artefacts that may not be easily identifiable through traditional correction methods. The proposed method was rigorously evaluated and demonstrated superior performance to pre-existing motion correction technique.

The second goal (Chapter 3) introduced an advanced visualisation technique for assessing myocardial stress responses in rest and stress T1-maps, which are relevant in diagnosing conditions such as ischaemia, microvascular dysfunction and other conditions associated with depletion of coronary vasodilatory reserve. The traditional

approach to evaluating myocardial stress involves dual manual segmentations of images acquired at rest and during stress - an approach that can be subjective and prone to variability. To overcome these limitations, vision transformers were used to generate stress T1 reactivity maps, a novel form of visualisation that provides a pixel-wise representation of stress-induced changes in myocardial tissue properties - the delta T1 (dT1) map. This method allows for a more detailed and quantitative assessment of myocardial changes during stress compared to rest, enabling clinicians to identify subtle changes that may be indicative of underlying pathology. The resultant stress T1 reactivity maps were validated through extensive testing, showing a high level of consistency and reliability across different pharmacological stress conditions in different magnetic resonance scanners.

The third goal (Chapter 4) tackled the challenge of automating the segmentation of the LV myocardium in short-axis CMR images, with a particular focus on the T1 mapping-based virtual native enhancement (VNE) images, but also applicable to LGE images. LV segmentation is a critical step in the analysis of CMR images, as it delineates the boundaries of myocardial tissues, allowing for accurate quantification of tissue characteristics. However, manual segmentation is time-consuming and subject to inter-observer variability. To address this, a quality control-driven deep ensemble method was deployed that combines the outputs of multiple CNN models to produce robust and accurate LV myocardium segmentations. The ensemble approach leverages the strengths of different models, while the integrated quality control mechanism automatically identifies and flags potentially erroneous segmentations for review. This method was extensively validated on large, diverse CMR datasets, demonstrating significant improvements in time and segmentation accuracy and consistency compared to traditional methods. By automating the segmentation process, this work paves the way to reducing the burden on clinicians while improving the reproducibility of CMR analyses.

The fourth goal (Chapter 5) focused on the development of a generalist deep learning model for the automated annotation of key anatomical landmarks, specifically the anterior RV insertion point and the LV centre point for LV segmental analysis. These landmarks are crucial for standardised LV segmental analysis according to the AHA-

16 segment model, which are widely used in clinical practice for assigning coronary artery territory and providing spatial information in myocardial abnormalities detected in CMR images. The proposed best solution employs a dual-stage residual neural network framework, with the first stage providing a coarse localisation of the landmarks, and the second stage refining these predictions through spatial standardisation. With additional data augmentation, a single generalist two-stage model was shown to be sufficient to ensure high precision in landmark annotation across a wide selection of CMR modalities, including native and post-contrast T1 maps, LGE, and VNE images. The implementation of this model moves the field closer to standardised myocardial segmentation.

6.2 Future directions

The advancements presented in this thesis have laid a strong foundation in machine-learning-based methods for automated image analysis for CMR T1 mapping, improving their accuracy, reliability, and potential clinical applicability. This opens several avenues for further exploration and development in this dynamically evolving field. The following future directions outline potential enhancements and expansions of the methods introduced herein, aiming to address current limitations and to push the boundaries of what is achievable in the domain of cardiac imaging.

1. The motion correction method developed in Chapter 2 effectively addresses intra-acquisition artefacts in T1 mapping. However, expanding this technique to handle both intra- and inter-acquisition motion at the inversion recovery-weighted (IRW) image level could potentially further enhance its image quality. By adapting the model to correct motion not only within individual T1 maps but also across different acquisitions, a unified framework could be established to suppress motion artefacts more comprehensively. Incorporating the framework from Chapter 3, where a T1 IRW image with the highest contrast serves as a fixed reference, this may enable the alignment of all T1 IRW images, whether from the same or different acquisitions, or perhaps from even different sequences if generalised properly (Chapter 5). This approach could result in motion-free T1

maps, ready for further quantitative analysis, such as assessing stress response using stress T1 maps, calculating the extracellular volume with post-contrast T1 maps and even inter-modality image co-registration.

2. The introduction of the dT1 map concept (Chapter 3) opens new opportunities for visualising changes in T1 maps, beyond using just the numerical segmental T1 values. Future studies should investigate whether the dT1 map itself, as a visual diagnostic tool, offers additional clinical value. This map could be particularly useful for studying the stress response along the subendocardial and subepicardial myocardium, leveraging the pixel-wise spatial resolution of the technique. The improved processing and visualisation move the field towards further studies needed to validate the added utility of dT1 mapping in clinical practice.

3. The quality control-driven (QCD) ensemble framework developed for automated segmentation in Chapter 4 could be further refined by incorporating diverse architectures, such as CNNs, transformers (Chapter 3), or hybrid and inter-modality models (Chapter 5). These enhancements may improve the robustness and accuracy of the LV segmentation process. Additionally, the framework could be adapted to directly segment specific pathological features, such as myocardial scar tissue and microvascular obstruction. The QCD concept may also be extended to federated learning environments, where data privacy concerns limit the sharing of imaging data between centres. In this scenario, individual models could be trained at each corresponding centre, and only the trained segmentation models and their associated metrics (such as Dice similarity coefficients) would be shared. This approach might allow for collaborative model development, without compromising data protection and information governance, while still enabling the exchange of valuable information for training regression models that predict segmentation confidence and exploit the learning from different centres.

4. The RV entry line annotation, introduced in Chapter 5, has several potentially wider applications in CMR imaging. Many deep learning methods for CMR begin by cropping the image to focus on the region of interest, rather than processing the entire field of view. The proposed technique can potentially be used to robustly identify the centre of the LV, facilitating more precise cropping and standardisation. This capability may enhance a variety of computer vision tasks, including segmentation, registration, and image generation. Future research should explore the generalisability of the trained generalist model, applying it to other CMR imaging modalities, such as T2 mapping, cine sequences, and perfusion maps. Further work is needed to explore generalised model applications in pre-processing (e.g., motion correction) and post-processing (segmentation and quantification) trained on vast datasets, including mixed sequences.

5. An important future direction is the integration of these AI methods (Figure 1.1) directly into MRI scanner software. By incorporating the motion correction algorithm from Chapter 2, the stress T1 reactivity mapping technique from Chapter 3, the QCD ensemble segmentation framework from Chapter 4, the automated landmark annotation model from Chapter 5, as well as our group's automated myocardial segmentation of T1 maps [106] and generation of VNE images using T1 maps and cines [25, 37] directly into the scanner's processing pipeline, clinicians could receive immediate image correction and quality-benchmarked analysis results during scanning sessions. This comprehensive integration would streamline the imaging workflow, providing rapid access to high-quality images and quantitative data, potentially enhancing diagnostic efficiency and patient care. Future work should focus on adapting these algorithms for compatibility with scanner hardware and software, and conducting clinical evaluations to assess their impact on routine practice.

6.3 Conclusion

This thesis presents a cohesive framework for enhancing CMR T1 mapping processing through a series of deep learning innovations, each addressing selected limitations in current clinical and research practices. By developing novel methods for motion correction, artefacts that compromise image quality can be mitigated, ensuring that T1 maps used for diagnosis are accurate and reliable. The introduction of stress T1 reactivity maps provides a precise tool for visualising myocardial stress responses, refining diagnostic capabilities. The QCD deep ensemble method for automated segmentation and the generalist deep learning model for landmark annotation contribute towards accountable segmentation accuracy and consistency improvements across various modalities. Together, these advancements pave the way for broader adoption of standardised, automated methods in clinical workflows, potentially enabling more accurate diagnoses and better patient outcomes. Future efforts should focus on integrating these techniques into a unified pathway that can be implemented directly onto clinical scanner systems, streamlining image processing and analysis for efficient diagnostic workflows in routine practice.

Appendix A

Technical implementation details

The appendix provides detailed methodological information supporting the computational advancements presented in this thesis. Each section corresponds to a results chapter (Chapters 2, 3, 4, and 5), focusing on the technical aspects of the developed frameworks to ensure reproducibility. These descriptions offer an in-depth view of the deep learning architectures and training configurations. The goal is to bridge the gap between high-level descriptions in the main text and the technical details required for implementation.

A.1 Goal 1: image quality

This section outlines the technical implementation of the convolutional neural network-based framework developed for motion correction in T1 maps (Chapter 2). The methodology addresses the challenge of motion artefacts that degrade the quality of myocardial tissue characterisation. By co-registering inversion recovery-weighted (IRW) images within T1 mapping sequences, the proposed framework enhances the precision of reconstructed T1 maps.

Neural network architecture

Input and Output Specifications

- **Input:** A stack of 7 IRW images with spatial dimensions of 384×384. The input shape is defined as (384, 384, 7).
- **Output:** The network outputs inverse displacement vector fields (DVFs) for motion correction, with shape (384, 384, 14). The 14 channels represent the x and y displacement components for each of the 7 IRW images.

Core architecture

The network employs a multi-scale encoder-decoder architecture with skip connections and intermediate warping layers for motion refinement. This design ensures coarse-to-fine registration, progressively correcting motion artifacts at multiple scales.

Encoder The encoder path extracts hierarchical features from the input image stack by progressively reducing spatial resolution while increasing feature depth. Each encoder block consists of three convolutional layers (kernel size 3×3, stride 1, same padding), followed by batch normalisation to stabilise training, Leaky ReLU activation with a negative slope of 0.1, and a max pooling layer for downsampling.

- **Block 1:** Initial input shape: (384, 384, 7). 16 convolutional filters. Max pooling with 2×2 kernel and stride 2. Output shape: (192, 192, 16).
- **Block 2:** Encoder input shape: (192, 192, 16). 32 convolutional filters. Max pooling with 3×3 kernel and stride 3. Output shape: (64, 64, 32).
- **Block 3:** Encoder input shape: (64, 64, 32). 64 convolutional filters. Max pooling with 4×4 kernel and stride 4/3. Output shape: (48, 48, 64).
- **Block 4:** Encoder input shape: (48, 48, 64). 128 convolutional filters. Max pooling with 2×2 kernel and stride 2. Output shape: (24, 24, 128).

- **Bottleneck:** Encoder input shape (24, 24, 128). 256 convolutional filters. Output shape: (24, 24, 256).

Warping Layers The warping layers spatially align feature maps using the predicted displacement vector fields (DVFs). These layers integrate encoder and decoder features at each scale and ensure multi-scale, coarse-to-fine motion correction. The DVFs are progressively refined at each scale, beginning with a coarse estimate and adding finer details at higher resolutions.

- **Inputs:**
  - Encoder feature map before max pooling $(H, W, C_{\text{encoder}})$, where H, W, and C represent the height, width, and channels of the respective levels.
  - Decoder feature map from the previous decoder block $(H, W, C_{\text{decoder}})$.
  - Predicted DVFs $(H, W, 14)$, representing x and y displacements for each of the 7 IRW images, refined at the previous decoder step.
- **DVF prediction:**
  - At each scale, the decoder predicts the DVFs through a convolutional layer (kernel size 3×3, stride 1, 14 filters) applied to the decoder feature map.
  - These predicted DVFs are used to warp the decoder feature map via bilinear interpolation.
- **Operations:**
  - The decoder feature map is warped using bilinear interpolation, guided by the DVFs predicted at the same scale.
  - The warped decoder feature map is concatenated with the encoder feature map to integrate multi-scale context.
- **Output:** Warped feature map $(H, W, C_{\text{encoder}} + C_{\text{decoder}})$, passed to the next decoder block.

Decoder The decoder reconstructs the spatial resolution and refines the DVFs. Each decoder block comprises: a transposed convolutional layer for upsampling; skip connections concatenating corresponding warped features; and a convolutional layer (kernel size 3×3) for feature refinement.

- **Block 1:** Decoder input shape: (24, 24, 256). Transposed convolution (kernel size 2×2, stride 2, 128 filters). Refined output shape: (48, 48, 128).

- **Block 2:** Decoder input shape: (48, 48, 128). Transposed convolution (kernel size 4×4, stride 3, 64 filters). Concatenated warped feature map shape: (64, 64, 192). Refined output shape: (64, 64, 64).
- **Block 3:** Input shape (64, 64, 64). Transposed convolution (kernel size 3×3, stride 3, 32 filters). Concatenated warped feature map shape: (192, 192, 96). Refined output shape: (192, 192, 32).
- **Block 4:** Input shape (192, 192, 32). Transposed convolution (kernel size 2×2, stride 2, 16 filters). Concatenated warped feature map shape: (384, 384, 64). Refined output shape: (384, 384, 14).

Training procedure

Loss function

The network is trained using a multi-scale loss function that combines displacement accuracy and spatial smoothness. The loss operates at each resolution level of the decoder, ensuring coarse-to-fine supervision for the predicted displacement vector fields (DVFs). The total loss is formulated as:

$$\mathcal{L}_{\text{total}} = \sum_{i=1}^{N} w_i \cdot \mathcal{L}_{\text{MSE}}^{(i)} + \lambda \cdot \mathcal{L}_{\text{smooth}}$$

where:

- $N = 4$: Number of scales in the decoder.
- w_i: Scale-specific weights assigned to the MSE loss at scale i, with values $w_1 = 0.08$, $w_2 = 0.02$, $w_3 = 0.01$, and $w_4 = 0.005$.
- $\lambda = 0.1$: Regularisation weight for the smoothness term.

Mean Squared Error (MSE) loss At each scale i, the MSE loss measures the squared difference between the predicted DVF $\mathbf{u}_i^{\text{pred}}$ and the ground truth DVF $\mathbf{u}_i^{\text{gt}}$:

$$\mathcal{L}_{\text{MSE}}^{(i)} = \frac{1}{H_i W_i} \sum_{x=1}^{H_i} \sum_{y=1}^{W_i} \left\| \mathbf{u}_i^{\text{pred}}(x, y) - \mathbf{u}_i^{\text{gt}}(x, y) \right\|^2$$

where H_i and W_i are the height and width of the feature map at scale i.

Smoothness regularisation To enforce spatial consistency, a smoothness term penalises large gradients in the DVFs:

$$\mathcal{L}_{\text{smooth}} = \frac{1}{HW} \sum_{x=1}^{H} \sum_{y=1}^{W} \left(\|\nabla_x \mathbf{u}(x,y)\|^2 + \|\nabla_y \mathbf{u}(x,y)\|^2 \right)$$

where ∇_x and ∇_y denote gradients of the DVF along the x and y axes.

Optimisation and training parameters

- **Optimiser:** Adam method with first and second moment decay rates of $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$.
- **Initial learning rate:** 0.001.
- **Learning rate scheduler:** The learning rate was reduced by a factor of $\sqrt{0.1}$ when the validation loss plateaued, with a minimum learning rate of 0.00005.
- **Batch size:** 4.
- **Epochs:** 200.

A.2 Goal 2: advanced visualisation

This section details the computational framework leveraging a vision transformer-based model to derive stress T1 reactivity maps (dT1 maps, Chapter 3). The methodology facilitates pixel-wise visualisations of myocardial blood volume changes during stress compared to rest, addressing variability in manual measurements and improving diagnostic accuracy. The automated approach enhances the feasibility and adoption of stress T1 mapping in clinical workflows.

Neural network architecture

Input and output specifications

- **Input:** Paired rest and stress T1 maps are provided as two-dimensional images of size 192×192 pixels.
- **Output:** The model outputs a displacement vector field (DVF) of size $192 \times 192 \times 2$, representing the pixel-wise transformation needed to align the stress T1 map to the rest T1 map.

Core architecture

The network is a hierarchical encoder-decoder architecture based on Swin transformers, designed for high-resolution image registration.

Encoder

- **Patch partitioning:**
  - Input images of size $192 \times 192 \times 1$ (single channel for grayscale T1 maps) are divided into non-overlapping patches of size 4×4.
  - Each patch is flattened into a vector of length 16 (i.e., 4×4).
- **Patch embedding:**
  - Flattened patches are projected into a high-dimensional space (128 dimensions) using a dense layer.
  - Positional embeddings of size 128 are added to retain spatial context.
- **Swin transformer blocks:** Each block consists of:
  - **Window-based Multi-head Self-Attention (W-MSA):**

    - Windows of size 7×7 patches (or 28×28 pixels) are formed within the feature map.
    - Attention operates independently within each window, splitting the feature map into multiple heads ($4, 8, 8, 8$ heads for each hierarchical level).
    - Attention is scaled by the square root of the embedding dimension per head to normalise values.
  - **Shifted window mechanism:**
    - Alternate layers apply a cyclic shift of $-3, -3$ pixels to each window, ensuring overlap and promoting cross-window communication.
  - **Feedforward network:**
    - Two fully connected layers:
      - The first layer expands the embedding dimension by a factor of 4 (e.g., $128 \rightarrow 512$).
      - The second layer reduces the dimension back to the original size (e.g., $512 \rightarrow 128$).
    - GELU activation is applied after the first dense layer.
    - Dropout (probability $= 0.1$) is applied after each dense layer for regularisation.
  - **Residual connections:**
    - Outputs of W-MSA and feedforward sub-layers are added to their respective inputs, enabling gradient flow across layers.
  - **Layer normalisation:**
    - Applied before W-MSA and feedforward sub-layers.
- **Patch merging:**
  - Input patches are grouped into 2×2 blocks.
  - Each block is concatenated, doubling the embedding dimension while halving the spatial resolution.

Decoder

- **Patch expanding:**
  - Reverses the effects of patch merging by splitting patches into 2×2 sub-patches.

  - Reduces embedding dimensions to match the encoder's spatial resolution.

- **Swin transformer blocks:**
  - Mirrors the encoder with identical configurations for attention heads, feed-forward layers, and residual connections.
  - Window-based attention ensures consistency between encoder and decoder operations.

- **Skip connections:**
  - Outputs from the encoder layers are concatenated with their corresponding decoder layers.
  - Channel dimensions are aligned using dense layers where necessary.

- **Final layer:**
  - A dense convolutional layer with a 3×3 kernel maps the decoder output to the DVF of size $192 \times 192 \times 2$.
  - The DVF represents horizontal and vertical displacements for pixel alignment.

Architectural parameters

- **Attention heads:** Scaled proportionally to feature dimensions: 4, 8, 8, 8.
- **Embedding dimensions:** Start at 128 and double at each downsampling stage.
- **Window size:** Fixed at 7×7 for local self-attention.
- **Depth:** Four hierarchical levels, each with two Swin transformer blocks.

Training procedure

Loss Function

The network is trained using a composite loss function that combines image similarity and smoothness regularisation. The loss ensures alignment between the registered stress T1 map and the fixed rest T1 map while enforcing spatial consistency in the DVFs. The total loss is formulated as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}} + \lambda \cdot \mathcal{L}_{\text{smooth}}$$

where:

- $\mathcal{L}_{\text{MSE}}$: Mean Squared Error (MSE) loss, measuring pixel-wise similarity between the registered stress T1 map and the fixed rest T1 map.
- $\mathcal{L}_{\text{smooth}}$: Smoothness regularisation loss, penalising large gradients in the predicted DVF.
- $\lambda = 0.02$: Weight assigned to the smoothness term, balancing alignment accuracy and spatial regularisation.

Mean Squared Error (MSE) loss The MSE loss measures the squared difference between the pixel intensities of the registered stress T1 map (I^{reg}) and the fixed rest T1 map (I^{fixed}):

$$\mathcal{L}_{\text{MSE}} = \frac{1}{HW} \sum_{x=1}^{H} \sum_{y=1}^{W} \left(I^{\text{reg}}(x, y) - I^{\text{fixed}}(x, y) \right)^2$$

where H and W are the height and width of the input image, respectively.

Smoothness regularisation The smoothness loss penalises large spatial gradients in the DVF to ensure realistic deformations. It is computed as:

$$\mathcal{L}_{\text{smooth}} = \frac{1}{HW} \sum_{x=1}^{H} \sum_{y=1}^{W} \left(\|\nabla_x \mathbf{u}(x, y)\|^2 + \|\nabla_y \mathbf{u}(x, y)\|^2 \right)$$

where:

- $\mathbf{u}(x, y)$: Displacement vector at pixel location (x, y).
- ∇_x and ∇_y: Gradients of the DVF along the x- and y-axes.

Optimisation and training parameters

- **Optimiser:** Adam method with first and second moment decay rates of $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$.
- **Learning rate:** 0.0001.
- **Batch size:** 32.
- **Epochs:** 200.
- **Gradient clipping:** Gradients were clipped to a maximum magnitude of 0.5 to stabilise training and prevent exploding gradients.

A.3 Goal 3: automated post-processing I

This section describes the development of a quality control-driven (QCD) deep ensemble framework for myocardial segmentation in T1 map-derived virtual native enhancement and late gadolinium enhancement images (Chapter 4). The method ensures robust and consistent segmentation by integrating outputs from multiple models with a confidence metric, addressing efficiency and variability challenges in manual analysis. The technical details for the generation of virtual native enhancement images can be found in the Supplemental Material from Zhang et al. Circulation 146 [37].

1. Ensemble of U-Net models

Input and output specifications

- **Input:** Single-channel grayscale images of size $256 \times 256 \times 1$, zero-padded.
- **Output:** Multi-class segmentation masks of size $256 \times 256 \times 3$, where each channel corresponds to myocardial tissue, blood pool, and background.

Core architecture

The U-Net ensemble consists of six models, each based on an encoder-decoder framework with depth d varying from 1 to 6.

Encoder

- Each depth level i includes:
  - Two consecutive convolutional layers (kernel size 3×3, stride 1, $2^{(4+i)}$ filters) with ReLU activation.
  - A dropout layer with rate $0.2 \cdot (1 + 0.25 \cdot (i - 1))$.
  - A max pooling operation (2×2) that halves the spatial dimensions.
- The encoder depth d determines the total number of levels.

Decoder

- Symmetrical to the encoder, featuring:
  - Transposed convolutions (2×2) for upsampling.
  - Concatenation with encoder feature maps for skip connections.

- Two 3×3 convolutional layers with ReLU activation for refinement.
- A dropout layer applied using the same rate as its corresponding encoder level.

Output layer

- A final 1×1 convolution maps the output to three channels.
- Softmax activation generates per-pixel class probabilities.

Training procedure

Loss functions The categorical cross-entropy loss is used for multi-class segmentation, defined as:

$$\mathcal{L}_{CCE} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{C} y_{n,k}\log(p_{n,k})$$

where N is the total number of pixels in the image, C is the number of classes (myocardium, blood pool, background), $y_{n,k}$ is the fround truth label for pixel n and class k, and $p_{n,k}$ is the predicted probability for pixel n and class k.

Optimisation and training parameters

- **Optimiser:** Adam method with first and second moment decay rates of $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$.
- **Learning rate:** 0.001.
- **Batch size:** 16.
- **Epochs:** Maximum of 300, with early stopping triggered by 20 epochs without validation improvement.

2. QCD framework

Input and output specifications

- **Input:** A matrix of Dice Similarity Coefficient (DSC) scores for both single candidate segmentations (SCSs) and combined candidate segmentations (CCSs).
- **Output:** Predicted DSC scores for each candidate mask, indicating the expected quality of segmentation.

Core architecture

The QCD framework employs a set of linear regression models to evaluate and predict the quality of segmentation candidates.

- **Candidate masks:**
  - Generated by the ensemble of U-Nets (SCSs) and via pixel-wise label voting scheme (CCSs).
  - Let there be M U-Net models. For each input image, $2M$ segmentation candidates are created: M SCSs and M CCSs.
- **DSC matrix:**
  - A pairwise DSC matrix $\mathbf{D}$ is constructed, where each entry D_{ij} represents the DSC between candidate i and candidate j.
  - For N observations and $2M$ candidates per observation, $\mathbf{D} \in \mathbb{R}^{N\times(2M\times 2M)}$.
- **Linear regression model:**
  - For each candidate i, a linear regression model predicts its DSC score relative to the ground truth:

    $$\hat{y}_i = \mathbf{w}^\top \mathbf{x}_i + b$$

    where:
    - $\mathbf{x}_i \in \mathbb{R}^{2M}$ is the vector of pairwise DSC scores for candidate i.
    - $\mathbf{w} \in \mathbb{R}^{2M}$ is the weight vector.
    - $b \in \mathbb{R}$ is the bias term.
    - $\hat{y}_i$ is the predicted DSC score for candidate i.
- **Candidate selection:**
  - The trained regression models evaluate all candidates for a given input.
  - The candidate with the highest predicted DSC score is selected as the final segmentation:

    $$\text{Final candidate} = \arg\max_i \hat{y}_i.$$

Training procedure

Loss function Linear regression models are trained using the mean squared error (MSE) loss function:

$$\mathcal{L}_{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

where:

- N: Total number of training candidates.
- y_i: Ground truth DSC score for candidate i, computed by comparing its segmentation mask to the ground truth mask.
- $\hat{y}_i$: Predicted DSC score for candidate i.

Optimisation

- **Solver:** Ordinary least squares (OLS), using the closed-form solution:

  $$\mathbf{w} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$$

  where:

  - $\mathbf{X} \in \mathbb{R}^{N \times 2M}$ is the matrix of input features (DSC values).
  - $\mathbf{y} \in \mathbb{R}^N$ is the vector of ground truth DSC scores.

- Convergence is typically achieved within 100-200 iterations, as each regression model solves a low-dimensional optimisation problem.

A.4 Goal 4: automated post-processing II

This section introduces the technical details of a dual-stage residual neural network designed for automated landmark annotation in cardiovascular magnetic resonance imaging (Chapter 5). The framework standardises segmentation across multiple modalities, including native, stress, and post-contrast T1 mapping, as well as virtual native enhancement and late gadolinium enhancement images. By automating the detection of key anatomical landmarks, the method enhances consistency and reduces variability in segmental myocardial analysis.

Neural network architecture

Input and output specifications

- **Input**:
  - **Stage 1**: $384 \times 384 \times 1$ grayscale images.
  - **Stage 2**: $128 \times 128 \times 1$ grayscale images, cropped and centred on regions of interest.
- **Output**:
  - A vector of four floating-point values representing normalised landmark coordinates in $[0, 1]$, which are then rescaled to the original input dimensions.

Core architecture

The architecture utilises a modified ResNet50V2 backbone pre-trained on ImageNet, with the following components:

Input preprocessing layer

- Single convolutional layer with 3 filters, kernel size 3×3, ReLU activation, and same padding.

Feature extraction backbone The ResNet50V2 backbone comprises convolutional and residual blocks. Each **bottleneck block** includes: a 1×1 convolution for dimensionality reduction or expansion, a 3×3 convolution for spatial feature extraction, a second 1×1 convolution to restore dimensions, and shortcut (identity) connections for efficient gradient flow.

- **Convolutional block 1**:
  - 64 filters, kernel size 7×7, stride 2, batch normalisation, and ReLU activation.
  - 3×3 max pooling, stride 2.
- **Residual block stack 1**: Three residual blocks, each with:
  - Bottleneck: 1×1 convolution with 64 filters, 3×3 convolution with 64 filters, 1×1 convolution with 256 filters.
- **Residual block stack 2**: Four residual blocks, each with:
  - Bottleneck: 1×1 convolution with 128 filters, 3×3 convolution with 128 filters, 1×1 convolution with 512 filters.
- **Residual block stack 3**: Six residual blocks, each with:
  - Bottleneck: 1×1 convolution with 256 filters, 3×3 convolution with 256 filters, 1×1 convolution with 1024 filters.
- **Residual block stack 4**: Three residual blocks, each with:
  - Bottleneck: 1×1 convolution with 512 filters, 3×3 convolution with 512 filters, 1×1 convolution with 2048 filters.

Global average pooling layer

- Reduces the spatial dimensions to a 1×1 feature vector with 2048 channels.

Fully connected layers

- Dense layer with 1000 units, ReLU activation, and dropout rate of 0.35.
- Output layer with 4 units and sigmoid activation.

Training procedure

Loss function

The model minimises the mean squared error (MSE) loss function, which measures the squared differences between the predicted and true landmark coordinates. The landmarks are represented as two pairs of normalised x and y coordinates: (x_1, y_1) and (x_2, y_2).

The MSE loss is calculated as:

$$\text{MSE} = \frac{1}{4} \sum_{i=1}^{2} \left[(x_{i,\text{true}} - x_{i,\text{pred}})^2 + (y_{i,\text{true}} - y_{i,\text{pred}})^2 \right]$$

- $x_{i,\text{true}}$ and $y_{i,\text{true}}$: True coordinates of the i-th landmark, normalised to $[0, 1]$.
- $x_{i,\text{pred}}$ and $y_{i,\text{pred}}$: Predicted coordinates of the i-th landmark, also normalised to $[0, 1]$.
- The factor $\frac{1}{4}$ ensures that the loss accounts for all four coordinate components (two x and two y values).

This formulation penalises discrepancies between the predicted and true coordinates equally, ensuring precise localisation of landmarks.

Optimisation and training parameters

- **Optimiser:** Adam method with first and second moment decay rates of $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$.
- **Learning rate:** 1×10^{-4}.
- **Weight initialisation:** Pre-trained ResNet weights for the backbone and random initialisation for custom layers.
- **Batch size:** 8.
- **Number of epochs:** 1000.

Bibliography

[1] Mensah GA, Fuster V, Murray CJL, Roth GA, Habtegiorgis Abate Y, Abbasian M, et al. Global burden of cardiovascular diseases and risks, 1990-2022. *Journal of the American College of Cardiology*. 2023;82(25):2350-473. doi:10.1016/j.jacc.2023.11.007.

[2] Cheema KM, Dicks E, Pearson J, Samani NJ. Long-term trends in the epidemiology of cardiovascular diseases in the UK: insights from the British Heart Foundation statistical compendium. *Cardiovascular Research*. 2022;118(10):2267-80. doi:10.1093/cvr/cvac053.

[3] British Heart Foundation. Heart & Circulatory Disease Statistics, 2023 Compendium; 2023. BHF.

[4] Grech ED. Pathophysiology and investigation of coronary artery disease. *BMJ*. 2003 May;326(7397):1027. doi:10.1136/bmj.326.7397.1027.

[5] Stephens NR, Restrepo CS, Saboo SS, Baxi AJ. Overview of complications of acute and chronic myocardial infarctions: revisiting pathogenesis and cross-sectional imaging. *Postgraduate Medical Journal*. 2019;95(1126):439-50. doi:10.1136/postgradmedj-2018-136279.

[6] Maron BJ, Gardin JM, Flack JM, Gidding SS, Kurosaki TT, Bild DE. Prevalence of hypertrophic cardiomyopathy in a general population of young adults. *Circulation*. 1995;92(4):785-9. doi:10.1161/01.cir.92.4.785.

[7] Salerno M, Sharif B, Arheden H, Kumar A, Axel L, Li D, et al. Recent advances in cardiovascular magnetic resonance. *Circulation: Cardiovascular Imaging*. 2017;10(6):e003951. doi:10.1161/circimaging.116.003951.

[8] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*. 2019;25(1):44-56. doi:10.1038/s41591-018-0300-7.

[9] Grothues F, Smith GC, Moon JCC, Bellenger NG, Collins P, Klein HU, et al. Comparison of interstudy reproducibility of cardiovascular magnetic resonance with two-dimensional echocardiography in normal subjects and in patients with heart failure or left ventricular hypertrophy. *The American Journal of Cardiology*. 2002;90(1):29-34. doi:10.1016/S0002-9149(02)02381-0.

[10] Walker S, Girardin F, McKenna C, Ball SG, Nixon J, Plein S, et al. Cost-effectiveness of cardiovascular magnetic resonance in the diagnosis of coronary heart disease: an economic evaluation using data from the CE-MARC study. *Heart*. 2013;99(12):873. doi:10.1136/heartjnl-2013-303624.

[11] Arnold JR, McCann GP. Cardiovascular magnetic resonance: applications and practical considerations for the general cardiologist. *Heart*. 2020;106(3):174. doi:10.1136/heartjnl-2019-314856.

[12] Ridgway JP. Cardiovascular magnetic resonance physics for clinicians: part I. *Journal of Cardiovascular Magnetic Resonance*. 2010;12(1):71. doi:10.1186/1532-429X-12-71.

[13] Ferreira VM, Piechnik SK. CMR parametric mapping as a tool for myocardial tissue characterization. *Korean Circulation Journal*. 2020;50(8):658-76. doi:10.4070/kcj.2020.0157.

[14] Thomas KE, Fotaki A, Botnar RM, Ferreira VM. Imaging methods: magnetic resonance imaging. *Circulation: Cardiovascular Imaging*. 2023;16(1):e014068. doi:10.1161/circimaging.122.014068.

[15] Kramer CM, Barkhausen J, Bucciarelli-Ducci C, Flamm SD, Kim RJ, Nagel E. Standardized cardiovascular magnetic resonance imaging (CMR) protocols: 2020 update. *Journal of Cardiovascular Magnetic Resonance*. 2020;22(1):17. doi:10.1186/s12968-020-00607-1.

[16] Bellenger NG, Burgess MI, Ray SG, Lahiri A, Coats AJS, Cleland JGF, et al. Comparison of left ventricular ejection fraction and volumes in heart failure by echocardiography, radionuclide ventriculography and cardiovascular magnetic resonance. Are they interchangeable? *European Heart Journal*. 2000;21(16):1387-96. doi:10.1053/euhj.2000.2011.

[17] Kim RJ, Wu E, Rafael A, Chen EL, Parker MA, Simonetti O, et al. The use of contrast-enhanced magnetic resonance imaging to identify reversible myocardial dysfunction. *New England Journal of Medicine*. 2000;343(20):1445-53. doi:10.1056/nejm200011163432003.

[18] Mahrholdt H, Wagner A, Judd RM, Sechtem U, Kim RJ. Delayed enhancement cardiovascular magnetic resonance assessment of non-ischaemic cardiomyopathies. *European Heart Journal*. 2005;26(15):1461-74. doi:10.1093/eurheartj/ehi258.

[19] Carrabba N, Amico MA, Guaricci AI, Carella MC, Maestrini V, Monosilio S, et al. CMR mapping: the 4th-era revolution in cardiac imaging. *Journal of Clinical Medicine*. 2024;13(2). doi:10.3390/jcm13020337.

[20] Schwitter J, Wacker CM, Wilke N, Al-Saadi N, Sauer E, Huettle K, et al. MR-IMPACT II: Magnetic Resonance Imaging for Myocardial Perfusion Assessment in Coronary artery disease Trial: perfusion-cardiac magnetic resonance vs. single-photon emission computed tomography for the detection of coronary artery disease: a comparative multicentre, multivendor trial. *European Heart Journal*. 2013;34(10):775-81. doi:10.1093/eurheartj/ehs022.

[21] Messroghli DR, Moon JC, Ferreira VM, Grosse-Wortmann L, He T, Kellman P, et al. Clinical recommendations for cardiovascular magnetic resonance mapping of T1, T2, T2* and extracellular volume: A consensus statement by the Society for Cardiovascular Magnetic Resonance (SCMR) endorsed by the European Association for Cardiovascular Imaging (EACVI). *Journal of Cardiovascular Magnetic Resonance*. 2016;19(1):75. doi:10.1186/s12968-017-0389-8.

[22] Arbelo E, Protonotarios A, Gimeno JR, Arbustini E, Barriales-Villa R, Basso C, et al. 2023 ESC Guidelines for the management of cardiomyopathies: developed by the task force on the management of cardiomyopathies of the European Society of Cardiology (ESC). *European Heart Journal*. 2023;44(37):3503-626. doi:10.1093/eurheartj/ehad194.

[23] Ferreira VM, Schulz-Menger J, Holmvang G, Kramer CM, Carbone I, Sechtem U, et al. Cardiovascular magnetic resonance in nonischemic myocardial inflammation: expert recommendations. *Journal of the American College of Cardiology*. 2018;72(24):3158-76. doi:10.1016/j.jacc.2018.09.072.

[24] Piechnik SK, Neubauer S, Ferreira VM. State-of-the-art review: stress T1 mapping—technical considerations, pitfalls and emerging clinical applications. *Magnetic Resonance Materials in Physics, Biology and Medicine*. 2018;31(1):131-41. doi:10.1007/s10334-017-0649-5.

[25] Zhang Q, Burrage MK, Lukaschuk E, Shanmuganathan M, Popescu IA, Nikolaidou C, et al. Toward replacing Late Gadolinium Enhancement with artificial intelligence Virtual Native Enhancement for gadolinium-free cardiovascular magnetic resonance tissue characterization in hypertrophic cardiomyopathy. *Circulation*. 2021;144(8):589-99. doi:10.1161/circulationaha.121.054432.

[26] Nakamori S, Fahmy A, Jang J, El-Rewaidy H, Neisius U, Berg S, et al. Changes in myocardial native T1 and T2 after exercise stress: a noncontrast CMR pilot study. *JACC: Cardiovascular Imaging*. 2020;13(3):667-80. doi:10.1016/j.jcmg.2019.05.019.

[27] McCommis KS, Zhang H, Goldstein TA, Misselwitz B, Abendschein DR, Gropler RJ, et al. Myocardial blood volume is associated with myocardial oxygen consumption. *JACC: Cardiovascular Imaging*. 2009;2(11):1313-20. doi:10.1016/j.jcmg.2009.07.010.

[28] Mahmod M, Piechnik SK, Levelt E, Ferreira VM, Francis JM, Lewis A, et al. Adenosine stress native T1 mapping in severe aortic stenosis: evidence for a role of the intravascular compartment on myocardial T1 values. *Journal of Cardiovascular Magnetic Resonance*. 2014;16(1):92. doi:10.1186/s12968-014-0092-y.

[29] Nickander J, Steffen Johansson R, Lodin K, Wahrby A, Loewenstein D, Bruchfeld J, et al. Stress native T1 and native T2 mapping compared to myocardial perfusion reserve in long-term follow-up of severe Covid-19. *Scientific Reports*. 2023;13(1):4159. doi:10.1038/s41598-023-30989-y.

[30] Liu A, Wijesurendra RS, Francis JM, Robson MD, Neubauer S, Piechnik SK, et al. Adenosine stress and rest T1 mapping can differentiate between ischemic, infarcted, remote, and normal myocardium without the need for gadolinium contrast agents. *JACC: Cardiovascular Imaging*. 2016;9(1):27-36. doi:10.1016/j.jcmg.2015.08.018.

[31] Bohnen S, Prüßner L, Vettorazzi E, Radunski UK, Tahir E, Schneider J, et al. Stress T1-mapping cardiovascular magnetic resonance imaging and inducible myocardial ischemia. *Clinical Research in Cardiology*. 2019;108(8):909-20. doi:10.1007/s00392-019-01421-1.

[32] van Assen M, van Dijk R, Kuijpers D, Vliegenthart R, Oudkerk M. T1 reactivity as an imaging biomarker in myocardial tissue characterization discriminating normal, ischemic and infarcted myocardium. *The International Journal of Cardiovascular Imaging*. 2019;35(7):1319-25. doi:10.1007/s10554-019-01554-4.

[33] Yimcharoen S, Zhang S, Kaolawanich Y, Tanapibunpon P, Krittayaphong R. Clinical assessment of adenosine stress and rest cardiac magnetic resonance T1 mapping for detecting ischemic and infarcted myocardium. *Scientific Reports*. 2020;10(1):14727. doi:10.1038/s41598-020-71722-3.

[34] Burrage MK, Shanmuganathan M, Masi A, Hann E, Zhang Q, Popescu IA, et al. Cardiovascular magnetic resonance stress and rest T1-mapping using regadenoson for detection of ischemic heart disease compared to healthy controls. *International Journal of Cardiology*. 2021;333:239-45. doi:10.1016/j.ijcard.2021.03.010.

[35] Levelt E, Piechnik SK, Liu A, Wijesurendra RS, Mahmod M, Ariga R, et al. Adenosine stress CMR T1-mapping detects early microvascular dysfunction in patients with type 2 diabetes mellitus without obstructive coronary artery disease. *Journal of Cardiovascular Magnetic Resonance*. 2016;19(1):81. doi:10.1186/s12968-017-0397-8.

[36] Iles L, Pfluger H, Phrommintikul A, Cherayath J, Aksit P, Gupta SN, et al. Evaluation of diffuse myocardial fibrosis in heart failure with cardiac magnetic resonance contrast-enhanced T1 mapping. *Journal of the American College of Cardiology*. 2008;52(19):1574-80. doi:10.1016/j.jacc.2008.06.049.

[37] Zhang Q, Burrage MK, Shanmuganathan M, Gonzales RA, Lukaschuk E, Thomas KE, et al. Artificial intelligence for contrast-free MRI: scar assessment in myocardial infarction using deep learning–based Virtual Native Enhancement. *Circulation*. 2022;146(20):1492-503. doi:10.1161/circulationaha.122.060137.

[38] Mallio CA, Radbruch A, Deike-Hofmann K, van der Molen AJ, Dekkers IA, Zaharchuk G, et al. Artificial intelligence to reduce or eliminate the need for gadolinium-based contrast agents in brain and cardiac MRI: a literature review. *Investigative Radiology*. 2023;58(10). doi:10.1097/rli.0000000000000983.

[39] Suinesiaputra A, Bluemke DA, Cowan BR, Friedrich MG, Kramer CM, Kwong R, et al. Quantification of LV function and mass by cardiovascular magnetic resonance: multi-center variability and consensus contours. *Journal of Cardiovascular Magnetic Resonance*. 2015;17(1):63. doi:10.1186/s12968-015-0170-9.

[40] Leiner T, Rueckert D, Suinesiaputra A, Baeßler B, Nezafat R, Išgum I, et al. Machine learning in cardiovascular magnetic resonance: basic concepts and

applications. *Journal of Cardiovascular Magnetic Resonance*. 2019;21(1):61. doi:10.1186/s12968-019-0575-y.

[41] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. *Medical Image Analysis*. 2017;42:60-88. doi:10.1016/j.media.2017.07.005.

[42] Amyar A, Nakamori S, Morales M, Yoon S, Rodriguez J, Kim J, et al. Gadolinium-free cardiac MRI myocardial scar detection by 4D convolution factorization. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*. Springer Nature Switzerland; 2023. p. 639-48. doi:10.1007/978-3-031-43895-0_60.

[43] Chen WW, Kuo L, Lin YX, Yu WC, Tseng CC, Lin YJ, et al. A deep learning approach to classify Fabry cardiomyopathy from hypertrophic cardiomyopathy using cine imaging on cardiac magnetic resonance. *International Journal of Biomedical Imaging*. 2024;2024(1):6114826. doi:10.1155/2024/6114826.

[44] Li T, Wei B, Cong J, Hong Y, Li S. Direct estimation of left ventricular ejection fraction via a cardiac cycle feature learning architecture. *Computers in Biology and Medicine*. 2020;118:103659. doi:10.1016/j.compbiomed.2020.103659.

[45] Inomata S, Yoshimura T, Tang M, Ichikawa S, Sugimori H. Estimation of left and right ventricular ejection fractions from cine-MRI using 3D-CNN. *Sensors*. 2023;23(14). doi:10.3390/s23146580.

[46] Gonzales RA, Seemann F, Lamy J, Mojibian H, Atar D, Erlinge D, et al. MVnet: automated time-resolved tracking of the mitral valve plane in CMR long-axis cine images with residual neural networks: a multi-center, multi-vendor study. *Journal of Cardiovascular Magnetic Resonance*. 2021;23(1):137. doi:10.1186/s12968-021-00824-2.

[47] Gonzales RA, Lamy J, Seemann F, Heiberg E, Onofrey JA, Peters DC. TVnet: automated time-resolved tracking of the tricuspid valve plane in MRI long-axis cine images with a dual-stage deep learning pipeline. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2021*. Springer International Publishing; 2021. p. 567-76. doi:10.1007/978-3-030-87231-1_55.

[48] Bai W, Sinclair M, Tarroni G, Oktay O, Rajchl M, Vaillant G, et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*. 2018;20(1):65. doi:10.1186/s12968-018-0471-x.

[49] Daudé P, Ancel P, Confort Gouny S, Jacquier A, Kober F, Dutour A, et al. Deep-learning segmentation of epicardial adipose tissue using four-chamber cardiac magnetic resonance imaging. *Diagnostics*. 2022;12(1). doi:10.3390/diagnostics12010126.

[50] Lamacie MM, Houbois CP, Greiser A, Jolly MP, Thavendiranathan P, Wintersperger BJ. Quantification of myocardial deformation by deformable registration–based analysis of cine MRI: validation with tagged CMR. *European Radiology*. 2019;29(7):3658-68. doi:10.1007/s00330-019-06019-9.

[51] Xia Y, Ravikumar N, Frangi AF. Learning to complete incomplete hearts for population analysis of cardiac MR images. *Medical Image Analysis*. 2022;77:102354. doi:10.1016/j.media.2022.102354.

[52] Giger ML, Chan HP, Boone J. Anniversary paper: History and status of CAD and quantitative image analysis: The role of medical physics and AAPM. *Medical Physics*. 2008;35(12):5799-820. doi:10.1118/1.3013555.

[53] van der Geest RJ, de Roos A, van der Wall EE, Reiber JHC. Quantitative analysis of cardiovascular MR images. *The International Journal of Cardiac Imaging*. 1997;13(3):247-58. doi:10.1023/a:1005869509149.

[54] Singleton HR, Pohost GM. Automatic cardiac MR image segmentation using edge detection by tissue classification in pixel neighborhoods. *Magnetic Resonance in Medicine*. 1997;37(3):418-24. doi:10.1002/mrm.1910370320.

[55] v Kaushikkar S, Li D, Haacke EM, Dávila-Román VG. Adaptive blood pool segmentation in three-dimensions: Application to MR cardiac evaluation. *Journal of Magnetic Resonance Imaging*. 1996;6(4):690-7. doi:10.1002/jmri.1880060419.

[56] Makela T, Clarysse P, Sipila O, Pauna N, Pham QC, Katila T, et al. A review of cardiac image registration methods. *IEEE Transactions on Medical Imaging*. 2002;21(9):1011-21. doi:10.1109/tmi.2002.804441.

[57] Duncan JS, Ayache N. Medical image analysis: progress over two decades and the challenges ahead. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2000;22(1):85-106. doi:10.1109/34.824822.

[58] Wernick MN, Yang Y, Brankov JG, Yourganov G, Strother SC. Machine learning in medical imaging. *IEEE Signal Processing Magazine*. 2010;27(4):25-38. doi:10.1109/msp.2010.936730.

[59] Gopalakrishnan V, Menon PG, Madan S. cMRI-BED: A novel informatics framework for cardiac MRI biomarker extraction and discovery applied to pediatric cardiomyopathy classification. *BioMedical Engineering OnLine*. 2015;14(2):S7. doi:10.1186/1475-925x-14-S2-S7.

[60] Gotardo PFU, Boyer KL, Saltz J, Raman SV. A new deformable model for boundary tracking in cardiac MRI and its application to the detection of intraventricular Dyssynchrony. In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*. vol. 1; 2006. p. 736-43. doi:10.1109/cvpr.2006.34.

[61] Kotu LP, Engan K, Borhani R, Katsaggelos AK, Ørn S, Woie L, et al. Cardiac magnetic resonance image-based classification of the risk of arrhythmias in post-myocardial infarction patients. *Artificial Intelligence in Medicine*. 2015;64(3):205-15. doi:10.1016/j.artmed.2015.06.001.

[62] Henglin M, Stein G, Hushcha PV, Snoek J, Wiltschko AB, Cheng S. Machine learning approaches in cardiovascular imaging. *Circulation: Cardiovascular Imaging*. 2017;10(10):e005614. doi:10.1161/circimaging.117.005614.

[63] Greenspan H, van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. *IEEE Transactions on Medical Imaging*. 2016;35(5):1153-9. doi:10.1109/tmi.2016.2553401.

[64] Litjens G, Ciompi F, Wolterink JM, de Vos BD, Leiner T, Teuwen J, et al. State-of-the-art deep learning in cardiovascular image analysis. *JACC: Cardiovascular Imaging*. 2019;12(8_Part_1):1549-65. doi:10.1016/j.jcmg.2019.06.009.

[65] Chen C, Qin C, Qiu H, Tarroni G, Duan J, Bai W, et al. Deep learning for cardiac image segmentation: a review. *Frontiers in Cardiovascular Medicine*. 2020;7. doi:10.3389/fcvm.2020.00025.

[66] Morales MA, Manning WJ, Nezafat R. Present and future innovations in AI and cardiac MRI. *Radiology*. 2024;310(1):e231269. doi:10.1148/radiol.231269.

[67] Davies RH, Augusto JB, Bhuva A, Xue H, Treibel TA, Ye Y, et al. Precision measurement of cardiac structure and function in cardiovascular magnetic resonance using machine learning. *Journal of Cardiovascular Magnetic Resonance*. 2022;24(1):16. doi:10.1186/s12968-022-00846-4.

[68] Zhang Q, Fotaki A, Ghadimi S, Wang Y, Doneva M, Wetzl J, et al. Improving the efficiency and accuracy of CMR with AI – review of evidence and proposition of a roadmap to clinical translation. *Journal of Cardiovascular Magnetic Resonance*. 2024:101051. doi:10.1016/j.jocmr.2024.101051.

[69] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In: *ICLR 2021 - 9th International Conference on Learning Representations*; 2021. p. 1-22. doi:10.48550/arXiv.2010.11929.

[70] Li J, Chen J, Tang Y, Wang C, Landman BA, Zhou SK. Transforming medical imaging with Transformers? A comparative review of key properties, current progresses, and future perspectives. *Medical Image Analysis*. 2023;85:102762. doi:10.1016/j.media.2023.102762.

[71] Willemink MJ, Koszek WA, Hardell C, Wu J, Fleischmann D, Harvey H, et al. Preparing medical imaging data for machine learning. *Radiology*. 2020;295(1):4-15. doi:10.1148/radiol.2020192224.

[72] Zhang A, Xing L, Zou J, Wu JC. Shifting machine learning for healthcare from development to deployment and from models to data. *Nature Biomedical Engineering*. 2022;6(12):1330-45. doi:10.1038/s41551-022-00898-y.

[73] Gonzales RA, Zhang Q, Papież BW, Werys K, Lukaschuk E, Popescu IA, et al. MOCOnet: robust motion correction of cardiovascular magnetic resonance T1 mapping using convolutional neural networks. *Frontiers in Cardiovascular Medicine*. 2021;8. doi:10.3389/fcvm.2021.768245.

[74] Gonzales RA, Zhang Q, Papież BW, Werys K, Lukaschuk E, Popescu IA, et al. Fast and robust motion correction of cardiovascular magnetic resonance T1-mapping using data-driven convolutional neural networks for generalisability. In: *SCMR Virtual Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2022. ORA.

[75] Look DC, Locker DR. Time saving in measurement of NMR and EPR relaxation times. *Review of Scientific Instruments*. 1970;41(2):250-1. doi:10.1063/1.1684482.

[76] Piechnik SK, Ferreira VM, Dall'Armellina E, Cochlin LE, Greiser A, Neubauer S, et al. Shortened Modified Look-Locker Inversion recovery (ShMOLLI) for clinical myocardial T1-mapping at 1.5 and 3 T within a 9 heartbeat

breathhold. *Journal of Cardiovascular Magnetic Resonance*. 2010;12(1):69. doi:10.1186/1532-429x-12-69.

[77] Messroghli DR, Plein S, Higgins DM, Walters K, Jones TR, Ridgway JP, et al. Human myocardium: single-breath-hold MR T1 mapping with high spatial resolution—reproducibility study. *Radiology*. 2006;238(3):1004-12. doi:10.1148/radiol.2382041903.

[78] Kellman P, Wilson JR, Xue H, Ugander M, Arai AE. Extracellular volume fraction mapping in the myocardium, part 1: evaluation of an automated method. *Journal of Cardiovascular Magnetic Resonance*. 2012 Jan;14(1):60. doi:10.1186/1532-429x-14-63.

[79] Chefd'hotel C, Hermosillo G, Faugeras O. Flows of diffeomorphisms for multimodal image registration. In: *Proceedings IEEE International Symposium on Biomedical Imaging*; 2002. p. 753-6. doi:10.1109/isbi.2002.1029367.

[80] Xue H, Shah S, Greiser A, Guetter C, Littmann A, Jolly MP, et al. Motion correction for myocardial T1 mapping using image registration with synthetic image estimation. *Magnetic Resonance in Medicine*. 2012;67(6):1644-55. doi:10.1002/mrm.23153.

[81] Zhou R, Huang W, Yang Y, Chen X, Weller DS, Kramer CM, et al. Simple motion correction strategy reduces respiratory-induced motion artifacts for k-t accelerated and compressed-sensing cardiovascular magnetic resonance perfusion imaging. *Journal of Cardiovascular Magnetic Resonance*. 2018;20(1):6. doi:10.1186/s12968-018-0427-1.

[82] Becker KM, Blaszczyk E, Funk S, Nuesslein A, Schulz-Menger J, Schaeffter T, et al. Fast myocardial T1 mapping using cardiac motion correction. *Magnetic Resonance in Medicine*. 2020;83(2):438-51. doi:10.1002/mrm.27935.

[83] Robinson AA, Chow K, Salerno M. Myocardial T1 and ECV measurement: Underlying concepts and technical considerations. *JACC: Cardiovascular Imaging*. 2019;12(11 Part 2):2332-44. doi:10.1016/j.jcmg.2019.06.031.

[84] Schulz-Menger J, Bluemke DA, Bremerich J, Flamm SD, Fogel MA, Friedrich MG, et al. Standardized image interpretation and post-processing in cardiovascular magnetic resonance - 2020 update: Society for Cardiovascular Magnetic Resonance (SCMR): Board of Trustees Task Force on Standardized Post-Processing. *Journal of Cardiovascular Magnetic Resonance*. 2020;22(1):19. doi:10.1186/s12968-020-00610-6.

[85] Fu Y, Lei Y, Wang T, Curran WJ, Liu T, Yang X. Deep learning in medical image registration: a review. *Physics in Medicine & Biology*. 2020;65(20):20TR01. doi:10.1088/1361-6560/ab843e.

[86] Petersen SE, Matthews PM, Bamberg F, Bluemke DA, Francis JM, Friedrich MG, et al. Imaging in population science: cardiovascular magnetic resonance in 100,000 participants of UK Biobank - rationale, challenges and approaches. *Journal of Cardiovascular Magnetic Resonance*. 2013;15(1):46. doi:10.1186/1532-429x-15-46.

[87] Ferreira VM, Piechnik SK, Dall'Armellina E, Karamitsos TD, Francis JM, Choudhury RP, et al. Non-contrast T1-mapping detects acute myocardial edema with high diagnostic accuracy: a comparison to T2-weighted cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance*. 2012;14(1):53. doi:10.1186/1532-429x-14-42.

[88] Piechnik SK, Jerosch-Herold M. Myocardial T1 mapping and extracellular volume quantification: an overview of technical and biological confounders. *The International Journal of Cardiovascular Imaging*. 2018;34(1):3-14. doi:10.1007/s10554-017-1235-7.

[89] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing; 2015. p. 234-41. doi:10.1007/978-3-319-24574-4_28.

[90] Sun D, Yang X, Liu MY, Kautz J. PWC-Net: CNNs for optical flow using pyramid, warping, and cost volume. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2018. p. 8934-43. doi:10.1109/cvpr.2018.00931.

[91] Sokooti H, de Vos B, Berendsen F, Lelieveldt BPF, Išgum I, Staring M. Nonrigid image registration using multi-scale 3D convolutional neural networks. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*. Springer International Publishing; 2017. p. 232-9. doi:10.1007/978-3-319-66182-7_27.

[92] Werys K, Dragonu I, Zhang Q, Popescu I, Hann E, Puchta H, et al. Total Mapping Toolbox (TOMATO): An open source library for cardiac magnetic resonance parametric mapping. *SoftwareX*. 2020;11:100369. doi:10.1016/j.softx.2019.100369.

[93] Onofrey JA, Casetti-Dinescu DI, Lauritzen AD, Sarkar S, Venkataraman R, Fan RE, et al. Generalizable multi-site training and testing of deep neural networks using image normalization. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*; 2019. p. 348-51. doi:10.1109/isbi.2019.8759295.

[94] Kingma DP, Ba JL. Adam: A method for stochastic optimization. In: *3rd International Conference on Learning Representations – ICLR 2015*; 2015. doi:10.48550/arXiv.1412.6980.

[95] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: A system for large-scale machine learning. In: *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*. USENIX Association; 2016. p. 265-83. doi:10.48550/arXiv.1605.08695.

[96] Papież BW, Heinrich MP, Fehrenbach J, Risser L, Schnabel JA. An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration. *Medical Image Analysis*. 2014;18(8):1299-311. doi:10.1016/j.media.2014.05.005.

[97] Cochran WG. The combination of estimates from different experiments. *Biometrics*. 1954;10(1):101-29. doi:10.2307/3001666.

[98] Lee CH, Cook S, Lee JS, Han B. Comparison of two meta-analysis methods: Inverse-variance-weighted average and weighted sum of z-scores. *Genomics Inform*. 2016;14(4):173-80. doi:10.5808/gi.2016.14.4.173.

[99] Armstrong RA. When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*. 2014;34(5):502-8. doi:10.1111/opo.12131.

[100] Bergamino M, Bonzano L, Levrero F, Mancardi GL, Roccatagliata L. A review of technical aspects of T1-weighted dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) in human brain tumors. *Physica Medica*. 2014;30(6):635-43. doi:10.1016/j.ejmp.2014.04.005.

[101] Jaubert O, Arrieta C, Cruz G, Bustin A, Schneider T, Georgiopoulos G, et al. Multi-parametric liver tissue characterization using MR fingerprinting: Simultaneous T1, T2, T2*, and fat fraction mapping. *Magnetic Resonance in Medicine*. 2020;84(5):2625-35. doi:10.1002/mrm.28311.

[102] Zheng JQ, Lim NH, Papież BW. D-net: Siamese based network for arbitrarily oriented volume alignment. In: *Shape in Medical Imaging*. Springer International Publishing; 2020. p. 73-84. doi:10.1007/978-3-030-61056-2_6.

[103] Zhang Q, Hann E, Werys K, Wu C, Popescu I, Lukaschuk E, et al. Deep learning with attention supervision for automated motion artefact detection in quality control of cardiac T1-mapping. *Artificial Intelligence in Medicine*. 2020;110:101955. doi:10.1016/j.artmed.2020.101955.

[104] Burrage MK, Shanmuganathan M, Zhang Q, Hann E, Popescu IA, Soundarajan R, et al. Cardiac stress T1-mapping response and extracellular volume stability of MOLLI-based T1-mapping methods. *Scientific Reports*. 2021;11(1):13568. doi:10.1038/s41598-021-92923-4.

[105] Lara Hernandez KA, Rienmüller T, Baumgartner D, Baumgartner C. Deep learning in spatiotemporal cardiac imaging: A review of methodologies and clinical usability. *Computers in Biology and Medicine*. 2021;130:104200. doi:10.1016/j.compbiomed.2020.104200.

[106] Hann E, Popescu IA, Zhang Q, Gonzales RA, Barutçu A, Neubauer S, et al. Deep neural network ensemble for on-the-fly quality control-driven segmentation of cardiac MRI T1 mapping. *Medical Image Analysis*. 2021;71:102029. doi:10.1016/j.media.2021.102029.

[107] Hann E, Gonzales RA, Popescu IA, Zhang Q, Ferreira VM, Piechnik SK. Ensemble of deep convolutional neural networks with Monte Carlo dropout sampling for automated image segmentation quality control and robust deep learning using small datasets. In: *Medical Image Understanding and Analysis*. Springer International Publishing; 2021. p. 280-93. doi:10.1007/978-3-030-80432-9_22.

[108] Zhang Q, Werys K, Popescu IA, Biasiolli L, Ntusi NAB, Desai M, et al. Quality assurance of quantitative cardiac T1-mapping in multicenter clinical trials – A T1 phantom program from the hypertrophic cardiomyopathy registry (HCMR) study. *International Journal of Cardiology*. 2021;330:251-8. doi:10.1016/j.ijcard.2021.01.026.

[109] Atkinson D, Popescu I, Thomas KE, Yun CH, Werys K, Burrage MK, et al. Effects of software on myocardial T1 measurements: a 3 T multicentre single vendor phantom study. In: *SCMR 28th Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2025. doi:10.1016/j.jocmr.2024.101653.

[110] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017. p. 2261-9. doi:10.1109/cvpr.2017.243.

[111] Ranjan A, Black MJ. Optical flow estimation using a spatial pyramid network. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017. p. 2720-9. doi:10.1109/cvpr.2017.291.

[112] Ilg E, Mayer N, Saikia T, Keuper M, Dosovitskiy A, Brox T. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017. p. 1647-55. doi:10.1109/cvpr.2017.179.

[113] Haskins G, Kruger U, Yan P. Deep learning in medical image registration: a survey. *Machine Vision and Applications*. 2020;31(1):8. doi:10.1007/s00138-020-01060-x.

[114] Kosuge H, Hachiya S, Fujita Y, Hida S, Chikamori T. Potential of non-contrast stress T1 mapping for the assessment of myocardial injury in hypertrophic cardiomyopathy. *Journal of Cardiovascular Magnetic Resonance*. 2023;25(1):53. doi:10.1186/s12968-023-00966-5.

[115] Gonzales RA, Burrage MK, Menacho D, Altun İ, Raman B, Ariga R, et al. dT1 maps: a novel approach for visualising myocardial stress without gadolinium-based contrast agents. In: *SCMR 28th Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2025. ORA.

[116] Nagel E, Greenwood JP, McCann GP, Bettencourt N, Shah AM, Hussain ST, et al. Magnetic resonance perfusion or fractional flow reserve in coronary disease. *New England Journal of Medicine*. 2019;380(25):2418-28. doi:10.1056/nejmoa1716734.

[117] Greenwood JP, Maredia N, Younger JF, Brown JM, Nixon J, Everett CC, et al. Cardiovascular magnetic resonance and single-photon emission computed tomography for diagnosis of coronary heart disease (CE-MARC): a prospective trial. *The Lancet*. 2012;379(9814):453-60. doi:10.1016/s0140-6736(11)61335-4.

[118] Kwong RY, Ge Y, Steel K, Bingham S, Abdullah S, Fujikura K, et al. Cardiac magnetic resonance stress perfusion imaging for evaluation of patients with chest pain. *Journal of the American College of Cardiology*. 2019;74(14):1741-55. doi:10.1016/j.jacc.2019.07.074.

[119] Xue H, Brown LAE, Nielles-Vallespin S, Plein S, Kellman P. Automatic in-line quantitative myocardial perfusion mapping: Processing algorithm and implementation. *Magnetic Resonance in Medicine*. 2020;83(2):712-30. doi:10.1002/mrm.27954.

[120] Kotecha T, Chacko L, Chehab O, O'Reilly N, Martinez-Naharro A, Lazari J, et al. Assessment of multivessel coronary artery disease using cardiovascular magnetic resonance pixelwise quantitative perfusion mapping. *JACC: Cardiovascular Imaging*. 2020;13(12):2546-57. doi:10.1016/j.jcmg.2020.06.041.

[121] Kotecha T, Martinez-Naharro A, Boldrini M, Knight D, Hawkins P, Kalra S, et al. Automated pixel-wise quantitative myocardial perfusion mapping by CMR to detect obstructive coronary artery disease and coronary microvascular dysfunction. *JACC: Cardiovascular Imaging*. 2019;12(10):1958-69. doi:10.1016/j.jcmg.2018.12.022.

[122] Seraphim A, Dowsing B, Rathod KS, Shiwani H, Patel K, Knott KD, et al. Quantitative myocardial perfusion predicts outcomes in patients with prior surgical revascularization. *Journal of the American College of Cardiology*. 2022;79(12):1141-51. doi:10.1016/j.jacc.2021.12.037.

[123] Chacko L, Kotecha T, Ioannou A, Patel N, Martinez-Naharro A, Razvi Y, et al. Myocardial perfusion in cardiac amyloidosis. *European Journal of Heart Failure*. 2024;26(3):598-609. doi:10.1002/ejhf.3137.

[124] Knott KD, Augusto JB, Nordin S, Kozor R, Camaioni C, Xue H, et al. Quantitative myocardial perfusion in Fabry disease. *Circulation: Cardiovascular Imaging*. 2019 Jul;12(7):e008872. doi:10.1161/circimaging.119.008872.

[125] Hughes RK, Camaioni C, Augusto JB, Knott K, Quinn E, Captur G, et al. Myocardial perfusion defects in hypertrophic cardiomyopathy mutation carriers. *Journal of the American Heart Association*. 2021;10(15):e020227. doi:10.1161/jaha.120.020227.

[126] Joy G, Kelly CI, Webber M, Pierce I, Teh I, McGrath L, et al. Microstructural and microvascular phenotype of sarcomere mutation carriers and overt hypertrophic cardiomyopathy. *Circulation*. 2023;148(10):808-18. doi:10.1161/circulationaha.123.063835.

[127] Camaioni C, Knott KD, Augusto JB, Seraphim A, Rosmini S, Ricci F, et al. Inline perfusion mapping provides insights into the disease mechanism in hypertrophic cardiomyopathy. *Heart*. 2020 Jun;106(11):824. doi:10.1136/heartjnl-2019-315848.

[128] Weinreb JC, Rodby RA, Yee J, Wang CL, Fine D, McDonald RJ, et al. Use of intravenous gadolinium-based contrast media in patients with kidney disease: Consensus Statements from the American College of Radiology and the National Kidney Foundation. *Radiology*. 2021;298(1):28-35. doi:10.1148/radiol.2020202903.

[129] Wijesurendra RS, Liu A, Eichhorn C, Ariga R, Levelt E, Clarke WT, et al. Lone atrial fibrillation is associated with impaired left ventricular energetics that persists despite successful catheter ablation. *Circulation*. 2016;134(15):1068-81. doi:10.1161/circulationaha.116.022931.

[130] Liu A, Wijesurendra RS, Ariga R, Mahmod M, Levelt E, Greiser A, et al. Splenic T1-mapping: a novel quantitative method for assessing adenosine stress adequacy for cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance*. 2016;19(1):1. doi:10.1186/s12968-016-0318-2.

[131] Raman B, Ariga R, Spartera M, Sivalokanathan S, Chan K, Dass S, et al. Progression of myocardial fibrosis in hypertrophic cardiomyopathy: mechanisms and clinical implications. *European Heart Journal - Cardiovascular Imaging*. 2019;20(2):157-67. doi:10.1093/ehjci/jey135.

[132] Balakrishnan G, Zhao A, Sabuncu MR, Guttag J, Dalca AV. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*. 2019;38(8):1788-800. doi:10.1109/tmi.2019.2897538.

[133] Cao H, Wang Y, Chen J, Jiang D, Zhang X, Tian Q, et al. Swin-UNet: UNet-like pure transformer for medical image segmentation. In: *Computer Vision – ECCV 2022 Workshops*. Springer Nature Switzerland; 2023. p. 205-18. doi:10.1007/978-3-031-25066-8_9.

[134] Kuijpers D, Prakken NH, Vliegenthart R, van Dijkman PRM, van der Harst P, Oudkerk M. Caffeine intake inverts the effect of adenosine on myocardial perfusion during stress as measured by T1 mapping. *The International Journal of Cardiovascular Imaging*. 2016;32(10):1545-53. doi:10.1007/s10554-016-0949-2.

[135] Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26(3):297-302. doi:10.2307/1932409.

[136] Grobner T. Gadolinium – a specific trigger for the development of nephrogenic fibrosing dermopathy and nephrogenic systemic fibrosis? *Nephrology Dialysis Transplantation*. 2006;21(4):1104-8. doi:10.1093/ndt/gfk062.

[137] Thomas KE, Lukaschuk E, Shanmuganathan M, Kitt JA, Popescu IA, Neubauer S, et al. Misclassification of females and males in cardiovascular magnetic resonance parametric mapping: the importance of sex-specific normal ranges for diagnosis of health vs. disease. *European Heart Journal - Cardiovascular Imaging*. 2024;25(3):339-46. doi:10.1093/ehjci/jead247.

[138] Gonzales RA, Ibáñez DH, Hann E, Popescu IA, Burrage MK, Lee YP, et al. Quality control-driven deep ensemble for accountable automated segmentation of cardiac magnetic resonance LGE and VNE images. *Frontiers in Cardiovascular Medicine*. 2023;10. doi:10.3389/fcvm.2023.1213290.

[139] Gonzales RA, Ibáñez DH, Hann E, Popescu IA, Burrage MK, Lee YP, et al. Quality control-driven framework for reliable automated segmentation of cardiac magnetic resonance LGE and VNE images. In: *International Workshop on Clinically-oriented and Responsible AI for Medical Data Analysis*. Springer International Publishing; 2023. ORA.

[140] Gonzales RA, Zhang Q, Hann E, Ferreira VM, Piechnik SK. Quality control-driven artificial intelligence for reliable automatic segmentation of LGE images in clinical practice. In: *SCMR Virtual Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2022. ORA.

[141] Kelle S, Roes SD, Klein C, Kokocinski T, de Roos A, Fleck E, et al. Prognostic value of myocardial infarct size and contractile reserve using magnetic resonance imaging. *Journal of the American College of Cardiology*. 2009;54(19):1770-7. doi:10.1016/j.jacc.2009.07.027.

[142] Roes SD, Kelle S, Kaandorp TAM, Kokocinski T, Poldermans D, Lamb HJ, et al. Comparison of myocardial infarct size assessed with contrast-enhanced magnetic resonance imaging and left ventricular function and volumes to predict mortality in patients with healed myocardial infarction. *The American Journal of Cardiology*. 2007 Sep;100(6):930-6. doi:10.1016/j.amjcard.2007.04.029.

[143] Weng Z, Yao J, Chan RH, He J, Yang X, Zhou Y, et al. Prognostic value of LGE-CMR in HCM: a meta-analysis. *JACC: Cardiovascular Imaging*. 2016;9(12):1392-402. doi:10.1016/j.jcmg.2016.02.031.

[144] Singh A, Chen W, Patel HN, Alvi N, Kawaji K, Besser SA, et al. Impact of wideband late gadolinium enhancement cardiac magnetic resonance imaging on device-related artifacts in different implantable cardioverter-defibrillator types. *Journal of Magnetic Resonance Imaging*. 2021;54(4):1257-65. doi:10.1002/jmri.27608.

[145] Karim R, Bhagirath P, Claus P, James Housden R, Chen Z, Karimaghaloo Z, et al. Evaluation of state-of-the-art segmentation algorithms for left ventricle infarct from late gadolinium enhancement MR images. *Medical Image Analysis*. 2016;30:95-107. doi:10.1016/j.media.2016.01.004.

[146] Zhuang X, Xu J, Luo X, Chen C, Ouyang C, Rueckert D, et al. Cardiac segmentation on late gadolinium enhancement MRI: a benchmark study from multi-sequence cardiac MR segmentation challenge. *Medical Image Analysis*. 2022;81:102528. doi:10.1016/j.media.2022.102528.

[147] Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D. Image segmentation using deep learning: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2022;44(7):3523-42. doi:10.1109/tpami.2021.3059968.

[148] Bernard O, Lalande A, Zotti C, Cervenansky F, Yang X, Heng PA, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Transactions on Medical Imaging*. 2018;37(11):2514-25. doi:10.1109/tmi.2018.2837502.

[149] Galati F, Ourselin S, Zuluaga MA. From accuracy to reliability and robustness in cardiac magnetic resonance image segmentation: a review. *Applied Sciences*. 2022;12(8). doi:10.3390/app12083936.

[150] Tajbakhsh N, Jeyaseelan L, Li Q, Chiang JN, Wu Z, Ding X. Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. *Medical Image Analysis*. 2020;63:101693. doi:10.1016/j.media.2020.101693.

[151] Garcea F, Serra A, Lamberti F, Morra L. Data augmentation for medical imaging: a systematic literature review. *Computers in Biology and Medicine*. 2023 Jan;152:106391. doi:10.1016/j.compbiomed.2022.106391.

[152] Al Khalil Y, Amirrajab S, Lorenz C, Weese J, Pluim J, Breeuwer M. On the usability of synthetic data for improving the robustness of deep learning-based segmentation of cardiac magnetic resonance images. *Medical Image Analysis*. 2023;84:102688. doi:10.1016/j.media.2022.102688.

[153] Chen RJ, Lu MY, Chen TY, Williamson DFK, Mahmood F. Synthetic data in machine learning for medicine and healthcare. *Nature Biomedical Engineering*. 2021;5(6):493-7. doi:10.1038/s41551-021-00751-8.

[154] Albà X, Lekadir K, Pereañez M, Medrano-Gracia P, Young AA, Frangi AF. Automatic initialization and quality control of large-scale cardiac MRI segmentations. *Medical Image Analysis*. 2018;43:129-41. doi:10.1016/j.media.2017.10.001.

[155] Sander J, de Vos BD, Išgum I. Automatic segmentation with detection of local segmentation failures in cardiac MRI. *Scientific Reports*. 2020;10(1):21769. doi:10.1038/s41598-020-77733-4.

[156] Valindria VV, Lavdas I, Bai W, Kamnitsas K, Aboagye EO, Rockall AG, et al. Reverse classification accuracy: predicting segmentation performance in the absence of ground truth. *IEEE Transactions on Medical Imaging*. 2017;36(8):1597-606. doi:10.1109/tmi.2017.2665165.

[157] Hann E, Biasiolli L, Zhang Q, Popescu IA, Werys K, Lukaschuk E, et al. Quality control-driven image segmentation towards reliable automatic image analysis in large-scale cardiovascular magnetic resonance aortic cine imaging. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Springer International Publishing; 2019. p. 750-8. doi:10.1007/978-3-030-32245-8_83.

[158] Ruijsink B, Puyol-Antón E, Oksuz I, Sinclair M, Bai W, Schnabel JA, et al. Fully automated, quality-controlled cardiac analysis from CMR. *JACC: Cardiovascular Imaging*. 2020;13(3):684-95. doi:10.1016/j.jcmg.2019.05.030.

[159] Kramer CM, Appelbaum E, Desai MY, Desvigne-Nickens P, DiMarco JP, Friedrich MG, et al. Hypertrophic Cardiomyopathy Registry: the rationale and design of an international, observational study of hypertrophic cardiomyopathy. *American Heart Journal*. 2015;170(2):223-30. doi:10.1016/j.ahj.2015.05.013.

[160] Scarsini R, Shanmuganathan M, De Maria GL, Borlotti A, Kotronias RA, Burrage MK, et al. Coronary microvascular dysfunction assessed by pressure wire

and CMR after STEMI predicts long-term outcomes. *JACC: Cardiovascular Imaging*. 2021;14(10):1948-59. doi:10.1016/j.jcmg.2021.02.023.

[161] Carapella V, Puchta H, Lukaschuk E, Marini C, Werys K, Neubauer S, et al. Standardized image post-processing of cardiovascular magnetic resonance T1-mapping reduces variability and improves accuracy and consistency in myocardial tissue characterization. *International Journal of Cardiology*. 2020;298:128-34. doi:10.1016/j.ijcard.2019.08.058.

[162] Isola P, Zhu JY, Zhou T, Efros AA. Image-to-image translation with conditional adversarial networks. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017. p. 5967-76. doi:10.1109/cvpr.2017.632.

[163] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *ICLR 2015 - 3rd International Conference on Learning Representations*; 2015. p. 1-14. doi:10.48550/arXiv.1409.1556.

[164] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution. In: *Computer Vision – ECCV 2016*. Springer International Publishing; 2016. p. 694-711. doi:10.1007/978-3-319-46475-6_43.

[165] Li X, Aldridge B, Fisher R, Rees J. Estimating the ground truth from multiple individual segmentations incorporating prior pattern analysis with application to skin lesion segmentation. In: *2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*; 2011. p. 1438-41. doi:10.1109/isbi.2011.5872670.

[166] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*. 2014;15(1):1929–1958. doi:10.5555/2627435.2670313.

[167] Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*. 2018;286(3):800-9. doi:10.1148/radiol.2017171920.

[168] Robinson R, Valindria VV, Bai W, Oktay O, Kainz B, Suzuki H, et al. Automated quality control in image segmentation: application to the UK Biobank cardiovascular magnetic resonance imaging study. *Journal of Cardiovascular Magnetic Resonance*. 2019;21(1):18. doi:10.1186/s12968-019-0523-x.

[169] Chen C, Ouyang C, Tarroni G, Schlemper J, Qiu H, Bai W, et al. Unsupervised multi-modal style transfer for cardiac MR segmentation. In: *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*. Springer International Publishing; 2020. p. 209-19. doi:10.1007/978-3-030-39074-7_22.

[170] Campello VM, Martín-Isla C, Izquierdo C, Petersen SE, Ballester MAG, Lekadir K. Combining multi-sequence and synthetic images for improved segmentation of late gadolinium enhancement cardiac MRI. In: *Statistical Atlases and Computational Models of the Heart. Multi-Sequence CMR Segmentation, CRT-EPiggy and LV Full Quantification Challenges*. Springer International Publishing; 2020. p. 290-9. doi:10.1007/978-3-030-39074-7_31.

[171] Wu Y, Tang Z, Li B, Firmin D, Yang G. Recent advances in fibrosis and scar segmentation from cardiac MRI: a state-of-the-art review and future perspectives. *Frontiers in Physiology*. 2021;12. doi:10.3389/fphys.2021.709230.

[172] Gonzales RA, Seemann F, Lamy J, Arvidsson PM, Heiberg E, Murray V, et al. Automated left atrial time-resolved segmentation in MRI long-axis cine images using active contours. *BMC Medical Imaging*. 2021;21(1):101. doi:10.1186/s12880-021-00630-3.

[173] Heiberg E, Engblom H, Carlsson M, Erlinge D, Atar D, Aletras AH, et al. Infarct quantification with cardiovascular magnetic resonance using "standard deviation from remote" is unreliable: validation in multi-centre multi-vendor data. *Journal of Cardiovascular Magnetic Resonance*. 2022;24(1):53. doi:10.1186/s12968-022-00888-8.

[174] Mikami Y, Kolman L, Joncas SX, Stirrat J, Scholl D, Rajchl M, et al. Accuracy and reproducibility of semi-automated late gadolinium enhancement quantification techniques in patients with hypertrophic cardiomyopathy. *Journal of Cardiovascular Magnetic Resonance*. 2014;16(1):85. doi:10.1186/s12968-014-0085-x.

[175] Flett AS, Hasleton J, Cook C, Hausenloy D, Quarta G, Ariti C, et al. Evaluation of techniques for the quantification of myocardial scar of differing etiology using cardiac magnetic resonance. *JACC: Cardiovascular Imaging*. 2011;4(2):150-6. doi:10.1016/j.jcmg.2010.11.015.

[176] Cerqueira MD, Weissman NJ, Dilsizian V, Jacobs AK, Kaul S, Laskey WK, et al. Standardized myocardial segmentation and nomenclature

for tomographic imaging of the heart. *Circulation*. 2002;105(4):539-42. doi:10.1161/hc0402.102975.

[177] Gonzales RA, Manrique AL, Burrage MK, Thomas KE, Altun İ, Huang WM, et al. Deep learning for automated insertion point annotation of CMR T1 maps. In: *2024 IEEE 21st International Symposium on Biomedical Imaging (ISBI)*. IEEE; 2024. ORA.

[178] Gonzales RA, Zhang Q, Burrage MK, Altun İ, Heiberg E, Peters DC, et al. Deep learning for automated insertion point annotation of CMR late gadolinium enhancement and virtual native enhancement images. In: *SCMR 26th Annual Scientific Sessions*. Society for Cardiovascular Magnetic Resonance; 2023. ORA.

[179] Rajiah PS, François CJ, Leiner T. Cardiac MRI: state of the art. *Radiology*. 2023;307(3):e223008. doi:10.1148/radiol.223008.

[180] Ghadimi S, Auger DA, Feng X, Sun C, Meyer CH, Bilchick KC, et al. Fully-automated global and segmental strain analysis of DENSE cardiovascular magnetic resonance using deep learning for segmentation and phase unwrapping. *Journal of Cardiovascular Magnetic Resonance*. 2021;23(1):20. doi:10.1186/s12968-021-00712-9.

[181] Brady AP. Error and discrepancy in radiology: inevitable or avoidable? *Insights into Imaging*. 2017;8(1):171-82. doi:10.1007/s13244-016-0534-1.

[182] Sedai S, Roy PK, Garnavi R. Right ventricle landmark detection using multiscale HOG and random forest classifier. In: *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*; 2015. p. 814-8. doi:10.1109/isbi.2015.7163996.

[183] Kim YC, Chung Y, Choe YH. Automatic localization of anatomical landmarks in cardiac MR perfusion using random forests. *Biomedical Signal Processing and Control*. 2017;38:370-8. doi:10.1016/j.bspc.2017.07.001.

[184] van Zon M, Veta M, Li S. Automatic cardiac landmark localization by a recurrent neural network. In: *Progress in Biomedical Optics and Imaging - Proceedings of SPIE*. vol. 10949. SPIE; 2019. doi:10.1117/12.2512048.

[185] Wang X, Zhai S, Niu Y. Left ventricle landmark localization and identification in cardiac MRI by deep metric learning-assisted CNN regression. *Neurocomputing*. 2020;399:153-70. doi:10.1016/j.neucom.2020.02.069.

[186] Xue H, Artico J, Fontana M, Moon JC, Davies RH, Kellman P. Landmark detection in cardiac MRI by using a convolutional neural network. *Radiology: Artificial Intelligence*. 2021;3(5):e200197. doi:10.1148/ryai.2021200197.

[187] Koehler S, Sharan L, Kuhm J, Ghanaat A, Gordejeva J, Simon NK, et al. Comparison of evaluation metrics for landmark detection in CMR images. In: *Bildverarbeitung für die Medizin 2022*. Springer Fachmedien Wiesbaden; 2022. p. 198-203. doi:10.1007/978-3-658-36932-3_43.

[188] Lv J, Shao X, Xing J, Cheng C, Zhou X. A deep regression architecture with two-stage re-initialization for high performance facial landmark detection. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2017. p. 3691-700. doi:10.1109/cvpr.2017.393.

[189] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016. p. 770-8. doi:10.1109/cvpr.2016.90.

[190] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In: *Computer Vision – ECCV 2016*. Springer International Publishing; 2016. p. 630-45. doi:10.1007/978-3-319-46493-0_38.

[191] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*; 2009. p. 248-55. doi:10.1109/cvpr.2009.5206848.

[192] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*. 2021;18(2):203-11. doi:10.1038/s41592-020-01008-z.

[193] Tu T, Azizi S, Driess D, Schaekermann M, Amin M, Chang PC, et al. Towards generalist biomedical AI. *NEJM AI*. 2024;1(3):AIoa2300138. doi:10.1056/aioa2300138.

[194] Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. *Nature*. 2023;616(7956):259-65. doi:10.1038/s41586-023-05881-4.

[195] Zhang S, Metaxas D. On the challenges and perspectives of foundation models for medical image analysis. *Medical Image Analysis*. 2024;91:102996. doi:10.1016/j.media.2023.102996.

[196] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment Anything. In: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*; 2023. p. 3992-4003. doi:10.1109/iccv51070.2023.00371.

[197] Mazurowski MA, Dong H, Gu H, Yang J, Konz N, Zhang Y. Segment Anything Model for medical image analysis: An experimental study. *Medical Image Analysis*. 2023;89:102918. doi:10.1016/j.media.2023.102918.

[198] Wasserthal J, Breit HC, Meyer MT, Pradella M, Hinck D, Sauter AW, et al. TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. *Radiology: Artificial Intelligence*. 2023;5(5):e230024. doi:10.1148/ryai.230024.

[199] Mitchell C, Xing S, Fenster A. 3D U-Net with ROI segmentation of kidneys and masses in CT scans. In: *Kidney and Kidney Tumor Segmentation*. Cham: Springer Nature Switzerland; 2024. p. 93-6. doi:10.1007/978-3-031-54806-2_13.