Machine Learning Approaches to Extract Higher-Order Features from Non-Contrast Computerised Tomography Images Enables Stratification of Diseases

Anirudh Chandrashekar
Keble College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity 2021

सुखर्थिनं कुतो विद्या नास्ति विद्यार्थीनं सुखम्
सुखार्थी वा त्यागविद्यां विद्यार्थी वा त्यागेसुखम

Mahabharata
Book 5, Chapter 40

"The important thing is to not stop questioning.
Curiosity has its own reason for existing."
- Albert Einstein

"It always seems impossible until it's done".
- Nelson Mandela

Acknowledgments

There are number of people without whom this doctorate would have been impossible. I would first like to express my heartfelt appreciation for my clinical advisors at Oxford, Dr. Regent Lee and Prof. Ashok Handa. Throughout the entire four-year process – from my first summer visit back in 2017 all the way to the final weeks of thesis-writing in 2021 – they have been an invaluable source of support and advice. They have always encouraged my independent research, giving me the confidence and trust to explore new ideas. At the same time, I knew that I could always count on their guidance, particularly when I needed it most. This combination of trust, encouragement and counsel have been vital to my development as a clinician scientist.

Similarly, I would like to thank my engineering supervisor, Prof. Vicente Grau. His advice and instructions on tackling clinical problems from an engineering perspective were invaluable – often providing me with the critical piece to the puzzle of my research. I am greatly indebted to his patience and perseverance, especially during the long periods of methods development and validation.

Work in the lab was uniquely enjoyable and intellectually stimulating thanks to both the clinical team and my companions at the IBME. Their willingness to provide inputs and constructive criticism on my work was essential to my progress. In particular, from the clinical team, I would like to thank Pierfrancesco Lapolla, Joel Ward, Elisha Ngetich and Natesh Shivakumar as they helped immensely with data collection and curation. Thank you to Jorge Corral Acero and Abhirup Banerjee from IBME for their invaluable methodological input.

I would also like to extend my thanks to everyone who made Keble College such a welcoming and supportive home away from home. In particular, thank you to Florian Keulers, Jasper Verschuur, Chloe Tubman and David Yang for their friendship in and out of the rowing boat, for our many early-morning and late-night conversations – about matters both serious and trivial – and three years of adventures through Oxford. And I would be remiss to not acknowledge Kunal Agarwal, Annie Welden, Molly Masterson, and Andrew Unsworth for their continuous friendship these past years.

I must also acknowledge the influence of Nicos Labropoulos on me. He has been a mentor, advisor, and role model, playing a key role in my personal and academic development. Since we met in 2013, he has encouraged me to grasp every opportunity with both hands and to grow as a *thinker*. He introduced me to vascular surgery and I will be forever grateful for his support and mentorship.

Last but by no means least, thanks to my amazing sister, Shruthi, and to my parents. No words would suffice to express my gratitude to them. Every day, I realize how lucky I am to have this loving and supportive family. My parents have sacrificed so much to give us the best experiences and opportunities and I do not nearly thank them enough for it. I am also acutely aware of the fact that if my mom had not convinced me to study abroad in 2017, I would not have had this rich and fulfilling chapter of my life. This thesis is dedicated to my sister, parents, and extended family; the people - who have made me the person I am today. I look forward to making you proud.

Abstract

Medical Imaging, which allows for the non-invasive assessment of biological tissues, is a rapidly growing health care service that has evolved from a diagnostic tool to a platform for personalized precision medicine. Computerized Tomography (CT), which is a commonly obtained imaging study worldwide, utilizes X-ray radiation to differentiate tissues based on density differences. More complex imaging studies are obtained based on the clinical question to supplement the CT. Image overutilization, which is the acquisition of medical images that have minimal impact on patient care and increasing global inequities due to the cost/availability of medical imaging services are significant problems that need to be addressed.

One possible solution would be to maximize the amount of clinically relevant information extracted from routine CT images using machine learning approaches. Recent developments in machine and deep learning have provided a powerful set of tools for the automated and complex analysis of medical images. Leveraging these techniques, it is possible to develop algorithms capable of learning from human annotations or paired images. These methods can also be applied to large datasets to minimise human input and to enable the rapid scale up of analyses.

In this thesis, I hypothesize that Non-Contrast CT images contain higher-order information to differentiate tissue anatomy or pathology without the need of intravenous contrast agents or radioactive tracers. I further hypothesized that such higher-order information enables stratification of disease progression without the need of additional imaging studies. I focus on two pathologies, abdominal aortic aneurysms (AAA) and Head and Neck Squamous Cell Carcinoma (HNSCC) as demonstration of feasibility. The methods described in this thesis can be applied to other pathologies and will be poised to disrupt clinical pathways in the future.

Table of Contents

LIST OF FIGURES **10**

LIST OF TABLES **14**

LIST OF ABBREVIATIONS **15**

1. **PREFACE** **21**

1.1 Motivation 22

1.2 Contribution 23

1.3 Publications Derived from thesis 24

1.4 Originality and Individual Role 26

1.5 Authorship 28

2. **INTRODUCTION** **29**

2.1 Increasing role of medical imaging in clinical care 30

2.1.1 Computerized Tomography (CT) Imaging 30

2.1.2 Computerized Tomography Imaging with Intravenous Contrast (CECT) 31

2.1.3 Positron Emission Tomography (PET) 31

2.2 Overutilization of medical imaging is an increasing health care concern 32

2.3 Potential solutions to limit medical image overutilization 33

2.4 Organ Isolation from medical images is required for feature extraction 33

2.5 Integration of deep learning in medical imaging improves clinical utility 34

2.6 Radiomic extraction from medical images allows for personalized medicine 35

2.7 Geometric analysis (3D) provides unique insight into disease progression 36

2.8 Hypothesis and Aims 37

2.8.1 Study Objectives 37

2.8.2 Hypothesis 37

2.8.3 Aims 37

3. **OVERVIEW OF IMPLEMENTED DATASETS** **38**

3.1 Imaging Cohorts from the OxAAA Study (OxAAA-IMG) 39

3.1.1 Defining the OxAAA-IMG-I dataset 41

3.1.2 Defining the OxAAA-IMG-II dataset 41

3.1.3 Defining the OxAAA-IMG-III dataset – Subset of Imaging Data with Demographic Information 42

3.1.4 Defining the OxAAA-IMG-IV cohort - Subset of Patients with Growth Measurements 42
3.2 Imaging Cohort from the MA3RS Study 44
3.3 Imaging cohort of patients with Head and Neck Cancer 45
4. AORTIC SEGMENTATION FROM CT WITH OR WITHOUT IV CONTRAST 46
4.1 Introduction 48
4.2 Methods 50
4.2.1 Curation of CT images from a clinical cohort 50
4.2.2 Manual Segmentation of CT Images (Defining the Ground Truth data) 50
4.2.3 Validation of Data Extraction Methods 51
4.2.4 Assessment of intra- and inter- observer variation of manual segmentation 57
4.2.5 Data augmentation 57
4.2.6 U-Net Architecture 58
4.2.7 Attention Gating to strengthen U-Net Performance 58
4.2.8 Loss Function to evaluate Model Performance 59
4.2.9 Aortic Segmentation Pipeline: Image Pre-Processing 60
4.2.10 Aortic Segmentation Pipeline: Aortic ROI Detection 60
4.2.11 Aortic Segmentation Pipeline: Aortic Segmentation 60
4.2.12 Semi-automatic segmentation of the aortic side branches from CECT images 62
4.2.13 Automatic Classification of Aortic Side Branches from CECT images 63
4.2.14 Assessment of model accuracy using aortic morphological features 67
4.2.15 Evaluation of Aortic Segmentation Pipeline on an independent validation cohort 68
4.3 Results 68
4.3.1 CT Image Characteristics 68
4.3.2 Validation of Data Extraction Methods 69
4.3.3 Intra- and Inter- observer variability assessment 70
4.3.4 Model Selection 74
4.3.5 Aortic Segmentation pipeline: ROI selection accuracy 78
4.3.6 Aortic Segmentation Pipeline: Aortic Segmentation accuracy 78
4.3.7 Aortic Segmentation Pipeline: Side-Branch Segmentation and Classification Accuracy 80
4.3.8 Assessment of aortic morphology from CECT-derived segmentations 81
4.3.9 Assessment of aortic morphology from NCCT-derived segmentations 86
4.3.10 Assessment of Aneurysm morphology on an independent validation cohort (OxAAA-IMG-II Cohort) 88
4.4 Discussion 90

4.5 Conclusion 93
4.5.1 Limitations 94
5. VISUALIZATION OF ANEURYSM MORPHOLOGY WITHOUT IV CONTRAST 96
5.1 Introduction 98
5.2 Methodology 100
5.2.1 Curation of CT images from a clinical cohort 100
5.2.2 Registration (alignment) of contrast-enhanced scans to non-contrast scans 100
5.2.3 Extraction, 2D sub-sampling and spatial normalization of aortic structures from NCCT and CECT images 101
5.2.4 Hounsfield unit sampling between AAA regions (Experiment 1) 101
5.2.5 Regional classification of AAAs using radiomic signature (Experiment 2) 102
5.2.6 Generative Adversarial Networks: Cycle-/Conditional-GAN (Experiment 3) 106
5.2.7 Evaluation of clinically important metrics using simulated pseudo-contrast images (Experiment 4) 108
5.2.8 Evaluation of GANs on the independent validation cohort (Experiment 5) 109
5.2.9 GAN model training and evaluation for the simulation of aortic side branches (Experiment 6) 110
5.3 Results 112
5.3.1 HU intensities differ between regions of a AAA in NCCT images 112
5.3.2 Radiomic signatures can be used to classify AAA regions from a NCCT 113
5.3.3 Generative Models can simulate contrast images using non-contrast images. 117
5.3.4 CT Image quality affects NCCT Transformation Accuracy 118
5.3.5 Cycle-GAN outperforms Con-GAN in simulating contrast CT images based on the measured clinical metrics 122
5.3.6 Similar GAN performance is observed within the validation cohort (OxAAA-IMG-II cohort). 127
5.3.7 Cycle-GAN can simulate extra-aortic/AAA features including aortic side branches. 128
5.4 Discussion 131
5.5 Conclusion 134
5.5.1 Limitations 135
6. PREDICTION OF AAA GROWTH IN HUMANS USING MORPHOLOGICAL FEATURES 137
6.1 Introduction 139
6.2 Methods 141

6.2.1 Patient Cohort 141

6.2.2 Automated Segmentation of CT images 143

6.2.3 Hypothesis-driven Feature Extraction of AAA Volumes 143

6.2.4 Semi-supervised Feature Extraction from Aortic Aneurysms 144

6.2.5 Developing the growth prediction models using both hypothesis-driven and PCA-derived features 150

6.2.6 Evaluation of the growth prediction models on an independent validation cohort (MA3RS Study) 151

6.3 Results 152

6.3.1 Patient demographic features do not influence the geometric features of AAA 152

6.3.2 Hypothesis-driven Geometric Feature Extraction from the OxAAA Growth Prediction Cohort 152

6.3.3 Semi-Supervised Geometric Feature Extraction from the OxAAA Growth Prediction Cohort 154

6.3.4 Prediction of AAA growth as a categorical outcome 156

6.3.5 Prediction of AAA growth rate as a continuous variable 167

6.3.6 Similar AAA Growth predictive accuracy is observed within the independent validation cohort (MA3RS Study) 170

6.4 Discussion 177

6.5 Conclusion 180

6.5.1 Limitations 181

7. VISUALIZATION OF FDG UPTAKE WITHOUT USING RADIOACTIVE TRACER ENABLES PREDICTION OF CLINICAL OUTCOME IN HNSCC ... 182

7.1 Introduction 184

7.2 Methods 185

7.2.1 Patient Population 185

7.2.2 SUV Map Generation from PET Images 186

7.2.3 Patient Contour Segmentation from NCCT/SUV images and Image Registration 187

7.2.4 Tumour Segmentation of the SUV Map for Radiomic Analysis 187

7.2.5 Radiomic Feature Extraction from defined segmentations 191

7.2.6 Radiomic Feature Reduction 193

7.2.7 Random Forest classification of metabolic activity based on radiomic signatures 193

7.2.8 Generative Models: Non-Contrast-CT-to-SUV Image Transformation 193

7.3 Results 196

7.3.1 Patient Population and SUV Map Characteristics 196

7.3.2 NCCT/SUV Registration and Tumour Segmentation 196

7.3.3 Experiment 1A: Radiomic features in NCCT images can differentiate regions of elevated vs negligible FDG uptake 197

7.3.4 Experiment 1B: Radiomics features in NCCT can differentiate high versus low FDG uptake within an individual tumour 199

7.3.5 Experiment 2: Generation of SUV Maps from NCCT 200

7.4 Discussion 204

7.5 Conclusion 207

7.5.1 Limitations 207

8. CONCLUSION AND FUTURE DIRECTIONS 209

8.1 Conclusion 210

8.2 Future Directions 211

8.2.1 Expanding the Data Cohorts 211

8.2.2 Mathematical modelling to simulate AAA shape and morphology 212

8.2.3 Expanding the DL pipelines to focus on other pathologies 213

8.2.4 Investigating the clinical impact and acceptance of DL algorithms 213

9. APPENDIX 214

9.1 Bland-Altman plot analysis to assess measurement accuracy 215

9.2 Radiomic Feature Extraction 216

9.3 Feature Reduction and Selection 217

9.4 Generative Model Training (Non-Contrast-CT to SUV Transformation) 218

9.4.1 CycleGAN Architecture 218

9.4.2 GAN model training 218

9.5 Results 218

9.5.1 Outcome prediction using Simulated SUV maps 218

10. REFERENCES 222

11. PUBLICATIONS 233

List of Figures

Fig. 1: Defining the OxAAA Imaging Cohorts used in this thesis. 40
Fig. 2: Comparison of the OxAAA-IMG-IV and MA3RS study datasets. 44
Fig. 3: Axial slice through an abdominal aortic aneurysm (AAA) from a NCCT and CECT image[10]. 48
Fig. 4: Manual Segmentation of CT Images (CECT/Non-Contrast). 52
Fig. 5: Diameter Profiles in Simulated Aneurysm Volumes. 53
Fig. 6: Varying the curvature of the aortic centreline. 55
Fig. 7: Mathematically-derived AAA shapes using multiple input parameters. 56
Fig. 8: Data Augmentation of Axial Slice with Aortic Aneurysm. 57
Fig. 9: Aortic segmentation pipeline for the simultaneous detection of the aortic lumen, and outer wall structure. 59
Fig. 10: Semi-automatic segmentation and classification of aortic side branches. 62
Fig. 11: Side-Branch Classification using manually constructed decision tree algorithm 1 (DT - 1). 64
Fig. 12: Side-Branch Classification using manually constructed decision tree algorithm 2 (DT - 2). 65
Fig. 13: Side-Branch Classification using manually constructed decision tree algorithm 3 (DT - 2). 66
Fig. 14: Validation of Automatic Surface Area and Volume Measurements from Synthetic AAA shapes. 70
Fig. 15: Clinical assessment of Inter- and Intra- observer-generated segmentations versus the ground truth segmentations from Contrast-enhanced CT. 72
Fig. 16: Clinical assessment of Inter- and Intra- observer-generated segmentations versus the ground truth segmentations from Non-Contrast CT. 73
Fig. 17: Attention U-Net vs 3D U-Net for AAA Segmentation. 74
Fig. 18: 1-,2- and 3-D evaluation of model predictions derived from the Attn U-Net (+Aug) and. 3D U-Net (+Aug). 76
Fig. 19: Attention-based 3D U-Net outputs from two patients with the labelled GT masks (OxAAA-IMG-I dataset). 77
Fig. 20: Training paradigm of the Attention-based U-Nets for Aortic ROI detection from CECT images (A) and Non-Contrast (B). 78
Fig. 21: Training paradigm of the Attention-based U-Nets for Aortic Segmentation from CECT and Non-Contrast images. 79
Fig. 22: DICE accuracy of model predictions from CECT/NCCT images compared against ground truth (GT) segmentations. 79

Fig. 23: Aortic segmentation pipeline (Attn UNets A – C) for a patient within the testing cohort. 80
Fig. 24: Classification of Aortic Side Branches based on branch artery. 81
Fig. 25: Correlation-coefficient and Bland-Altman Plot analysis for aortic morphological measurements from CECTs. 82
Fig. 26: Model predictions from Contrast (A-B) and Non-Contrast CT (C) Images. 83
Fig. 27: Maximum AAA Diameter profiles along planes orthogonal to AAA centerline. 85
Fig. 28: Correlation coefficient and Bland-Altman plot analysis for aortic morphological measurements from NCCTs. 87
Fig. 29: Bland-Altman Plots evaluating the accuracy of the segmentation model in extracting the aneurysmal/AAA shape from the validation cohort (OxAAA-IMG-II). 89
Fig. 30: Axial slice from a Computed Tomography (CT) scan with and without the use of an intravenous iodinated contrast agent. 98
Fig. 31: Axial slices from both the Contrast and Non-contrast CT scans. 102
Fig. 32: Sub-sampling of AAA shape and regional classification. 104
Fig. 33: Pipeline for the regional Classification of AAA using higher-order radiomic features. 105
Fig. 34: Con-GAN (A.) and CycleGAN (B.) architectures used for the transformation of NCCT images to contrast CECT images. 106
Fig. 35: Regional classification of thrombus morphology. 108
Fig. 36: Input data for the GAN architectures. 111
Fig. 37: Axial slices from both the Contrast and Non-contrast CT scans. 112
Fig. 38: Random forest model performance for the regional classification of AAA. 115
Fig. 39: Transformation accuracy within the OxAAA-IMG-I cohort. A. 117
Fig. 40: Increased DICE score variability in NCCT images obtained with low X-Ray tube currents (mA). 118
Fig. 41: Transformation accuracy of the refined cohort. 119
Fig. 42: Pseudo-Contrast images generated using the Con-GAN and Cycle-GAN architectures from 6 different patients alongside their respective NCCT and CECT axial slices. 120
Fig. 43: Segmented AAA shapes generated from the Con-GAN and Cycle-GAN architectures. 121
Fig. 44: One-dimensional diameter (A-B), two-dimensional area (C-D) and three-dimensional volume assessment (E-F) of generated images. 123
Fig. 45: Confusion Matrices comparing OWS regional classifications between generated images and ground truth segmentations. 124

Fig. 46: Points of discrepancy between the Con-GAN and Cycle-GAN Models. 125
Fig. 47: Failure of both generative models. 126
Fig. 48: Influence of AAA Size and Shape on GAN Performance. 129
Fig. 49: Pseudo-Contrast CT images are displayed alongside their respective Contrast and Non-Contrast CT Images for four patients. 130
Fig. 50: Radius of Curvature (RC) (A) and undulation index (UI) (B) measurements obtained from the sample abdominal aortic aneurysm (AAA) shapes[12]. 144
Fig. 51: Reference axis reorientation of AAAs prior to PCA-mediated Feature Extraction. 145
Fig. 52: Characterization of the aneurysmal shape prior to Shape-Based PCA. 146
Fig. 53: Method to generate surface lines from the Aortic/AAA segmentations. 147
Fig. 54: Characterization of the aneurysmal surface prior to Surface-Based PCA. 148
Fig. 55: Origin Point extraction of the aortic side branches from aortic/AAA segmentations. 149
Fig. 56: Linear correlations between extracted geometric featuresfrom AAAs within the OxAAA-IMG-IV dataset. 152
Fig. 57: Six AAAs of similar size displayed alongside their respective convex hull. 153
Fig. 58: Six AAAs of similar size displayed with calculated centrelines. 153
Fig. 59: Cumulative mode contribution for all three sets of PCA. 155
Fig. 60: AuROC for each combination of features to assess the performance of the multinomial regression model in discerning AAA growth phenotype (A. > 5.0 mm and B. < 2.5 mm Growth). 157
Fig. 61: AuROC for models classifying Slow (< 2.5 mm) and Fast (>5.0 mm) growth trained with PCA-derived features. 158
Fig. 62: AuROC for models classifying Slow (< 2.5 mm) and Fast (>5.0 mm) growth trained with Hypothesis-Driven and PCA-derived features. 160
Fig. 63: The relevance of Mode 3 (Shape-based PCA) on the average AAA shape and its impact on aneurysmal growth (Slow vs Not-Slow Growth). 162
Fig. 64: The relevance of Mode 7 (Shape-based PCA) on the average AAA shape and its impact on aneurysmal growth (Slow vs Not-Slow Growth). 163
Fig. 65: The relevance of Mode 4 (Shape-based PCA) on the average AAA shape and its impact on aneurysmal growth (Fast vs Not-Fast Growth). 164
Fig. 66: The influence of Mode 10 (Branch-Point PCA) on average branch locations and its impact on aneurysmal growth (Slow vs Not-Slow Growth). 165
Fig. 67: The influence of Mode 3 (Branch-Point PCA) on average branch locations and its impact on aneurysmal growth (Fast vs Not-Fast Growth). 166

Fig. 68: Outputs of Linear Regression models trained to predict annual AAA growth. 168
Fig. 69: Comparison of the OxAAA-IMG-IV and MA3RS study cohorts (Hypothesis-driven Features) in preparation for AAA growth prediction. 171
Fig. 70: Linear correlations between extracted geometric featuresfrom AAAs within the MA3RS Dataset 172
Fig. 71: Comparison of the OxAAA-IMG-IV and MA3RS study datsets (PCA-derived Features) in preparation for AAA growth prediction. 172
Fig. 72: Receiver operator characteristic (ROC) alongside the area under curve (AUROC) to assess model performance on MA3RS dataset. 175
Fig. 73: Outputs of Linear Regression models trained to predict annual AAA growth (MA3RS Dataset). 176
Fig. 74: Conversion of a FDG-PET image to Standard Uptake Value (SUV) map. 186
Fig. 75: Patient Surface Contour Segmentation from Non Contrast CT and SUV Images and Image Registration. 188
Fig. 76: SUV-Map-based Tumour Segmentation for Radiomic Analysis. 189
Fig. 77: Visualisation of the thyroid (green) and tumour (red) regions from two patients within the dataset. 190
Fig. 78: Regions of different FDG uptake avidity as defined by the PET SUV map. 190
Fig. 79: Workflow for the classification of volumes extracted from NCCT images based on metabolic activity using radiomic signature (Experiments 1A and 1B). . 192
Fig. 80: Pipeline for the Clinical Evaluation of Simulated SUV Maps. 195
Fig. 81: Differentiating regions of FDG uptake within a CT image. 198
Fig. 82: Area under Receiver Operation Curves for 4 random forest models trained with a combination of radiomic features to classify CT regions based on metabolic activity. 199
Fig. 83: Simulated SUV Map (Output of Cycle-GAN) displayed alongside its ground truth (Real SUV Map) and Non-Contrast CT axial slice for six patients. 201
Fig. 84: Technical Assessment of Simulated SUV-Map Accuracy: 202
Fig. 85: Area under ROC curve for logistic regression models trained to predict clinical outcomes 203

List of Tables

Table 1: Image cohort comparison: (1) OxAAA-IMG-I: used for model training, n = 75 and (2) OxAAA-IMG-II used for model validation, n = 200. 41
Table 2: Summary of participant demographics at the pre-surgical assessment (OxAAA-IMG-III dataset) 43
Table 3: U-Nets trained for model selection and the segmentation pipeline with learning parameters. 61
Table 4: Image characteristics within training (n = 50, Folds 2 + 3) and validation (n = 25, Fold 1) cohorts (OxAAA-IMG-I dataset). 69
Table 5: DICE scores and intra-class correlations for Intra-/Inter- operator segmentations using the images within the OxAAA-IMG-I dataset (*p < 0.001). 71
Table 6: Aortic Segmentation Accuracy of the Attn U-Net (+Aug) vs 3D-U-Net (+Aug). 75
Table 7: Clinical assessment of segmented volumes from CECT and NCCT images. 84
Table 8: Evaluating the maximum diameters of the aortic side branches from segmentations against manually-derived measurements. 86
Table 9: AAA Segmentation Accuracy on the OxAAA-IMG-II dataset (n = 200). 88
Table 10: Maximum Diameter of the aortic side branches from the CECT-derived segmentations on the OxAAA-IMG-II dataset 90
Table 11: Sub-volume splits for each of the three folds during random forest model training/ testing 113
Table 12: Most important 1st and 2nd Order radiomic features for regional classification of AAA 116
Table 13: Ground truth CECT images vs. generated CECT images of the validation cohort. 127
Table 14: Diameter measurements of aortic side branches from CECT and Pseudo-contrast CT images. 131
Table 15: Summary of participant demographics (OxAAA-IMG-III) at the pre-surgical assessment and significance of spearman correlation with extracted geometric parameters[12]. .. 142
Table 16: Groups split based on annual AAA growth within the OxAAA Imaging cohort. ... 156
Table 17: Optimized regression model combining both Hypothesis-driven and PCA-derived features to predict slow/fast growth. 160
Table 18: Feature coefficients for the four linear regression trained to predict AAA growth.. 169
Table 19: Groups split based on annual AAA growth within the MA3RS dataset. 173
Table 20: RMSE difference between GT- and GAN- SUV Maps. 201

List of Abbreviations

ΔOWS	Maximum thickness of the Outer Wall Structure (OWS) within the aortic aneurysm. May include the thickness of the intra-luminal thrombus (ILT) - used to synthesize AAA shapes (Section 4.2.4)
ΔWall	Wall thickness of the aortic aneurysm. If ΔWall = ΔOWS, then the AAA does not contain an ILT component – used to synthesis AAA shapes (Section 4.2.4)
1,2, or 3-D	1,2, or 3-Dimensional – Characterizes the spatial dimensions of the calculated metrics used to evaluate the congruence between the predicted and ground truth (GT) AAA shapes (Section 4.2.15, 5.3.5, and 6.3.2-3)
AAA	Abdominal Aortic Aneurysm – Abnormal and progressive ballooning of the infra-renal Aorta. This pathology is the focus for Chapter 4 – 6.
AP, APD	Anteroposterior – Common and standardized direction to measure the size of the aneurysmal sac. Maximum AP diameters (APDs) from serial scans were used to calculate aneurysmal growth (Section 6.2.3)
Attn U-Net	Attention-Gated U-Network – Deep learning network used for segmentation of the aorta/AAA from both non-contrast and contrast-enhanced CT images. Model Performance was compared against that of the generic 3D U-Net (Sections 4.2.7-8)
Aug	Data augmentation, via non-linear transformations, was used to artificially increase the training dataset size for aortic/AAA segmentation. This was used to maximize deep learning performance (Section 4.2.6).
AuROC	Area under the Receiver Operator Characteristic (ROC) curve is a quantitative assessment of classifier performance for a particular task (Section 5.3.2, 6.3.4, 6.3.6, 7.3.3-4, and 7.3.5.2)
Bq	Becquerel – SI derived unit of radioactivity. Used to capture the initial radiation dose secondary to radioactive tracer (FDG) injection (Section 7.2.2)
CAD	Coronary artery disease develops when the major blood vessels that supply the heart become damaged and is a key component of a patient's past medical history (Section 3.1)
CD	Average Euclidean Distance Deviation (between centrelines) – Metric

	used to compare the difference between calculated centrelines derived from predicted and GT AAA shapes. (Section 4.3.7-8).
CECT	Contrast-Enhanced CT. CT imaging following IV contrast injection. Intravenous iodinated contrast increases the apparent density of blood and is usually introduced to enhance visualization of vasculature.
CNN	Convolutional Neural Network. Foundation of deep learning architectures, which consist of multiple layers that transform the input using various pre-defined methods
Con-GAN	Conditional-Generative Adversarial Network
CT	Computerized Tomography, an imaging modality whereby a 3-D volume is constructed from multiple 2-D X-ray images. Voxel intensity within the image are in Hounsfield Units (HU) and The intrinsic contrast between tissues is dependent on tissue density.
CV	Coefficient of Variation, statistical measure of the relative dispersion within a data set (ratio of standard deviation to the mean, Section 4.2.15, 4.3.1)
Cycle-GAN	Cycle-Generative Adversarial Network - Deep learning network used for cross-modality image transformation problems (Chapter 4 and Chapter 6). Performance was compared against that of Conditional GAN (Section 5.3.3)
DICE	Sørensen–Dice score – Metric used to assess the congruence between two segmentation masks (ex. Prediction vs Ground-truth). Also used as the loss function to train DL segmentation models (Defined in Section 4.2.9).
$DICE_C$	DICE score metric to assess the segmentation accuracy of the Combined Aorta (Lumen + OWS). Both Lumen and OWS are represented as one entity (Section 4.3.3-4,9; 5.3.3-4)
$DICE_I$	DICE score metric to assess the segmentation accuracy of only the Inner Lumen (Section 4.3.3-4,9; 5.3.3-4)
DICOM	Digital Imaging and Communications in Medicine – Common format for medical image storage. Images were obtained, de-identified and stored in this format.
DL	Deep Learning is a branch of Machine learning that employs artificial neural networks to complete complex and difficult tasks.
D_{Max}	Maximum aneurysmal diameter within the aortic aneurysm. Includes the thickness of the intra-luminal thrombus (ILT), if present, and the

	aortic wall (ΔWall). Is used to synthesize AAA shapes (Section 4.2.4)
D_{Neck}	Proximal Neck diameter of the aortic aneurysm. Includes the aortic wall thickness (ΔWall) and is used to synthesize AAA shapes (Section 4.2.4)
DT	Decision Tree. Expert-based classification algorithms used to classify aortic side branches by artery. Three decision trees were created and a majority voting system was used to combine outputs (Section 4.2.14, 4.3.6).
eGFR	Estimated Glomerular Filtration Rate is a value that characterizes a patient's kidney function and is a key component of a patient's current medical status (Section 3.1)
FDG	2-[fluorine-18]fluoro-2-deoxy-d-glucose is the radioactive tracer used during Positron Emission Tomography to capture regions of increased metabolic activity.
GAN	Generative Adversarial Network – Family of Deep learning networks used for cross-modality image transformation problems (Chapter 4 and Chapter 6). Cycle- and Conditional- GANs were evaluated.
GT	Ground Truth. All model predictions are compared against the GT to assess model performance. GT for the aortic segmentation (Chapter 4) and image transformation (Chapter 5) were the manual segmentations and the contrast-enhanced CT, respectively.
Ht	Height of the aortic aneurysm - used to synthesize AAA shapes (Section 4.2.4)
H	High Pass Decomposition Filter (Wavelet Transforms) – used in image processing to selectively emphasize High-frequency image details within an image input (Sections 5.3.2, 7.3.3-4)
HNSCC	Head and Neck Squamous Cell Carcinoma is a cancer that often develops from the mucosal epithelium in the oral cavity, pharynx and larynx. Paired PET-CT imaging was obtained in patients with advanced HNSCC and served as the dataset for Chapter 7.
HU	Hounsfield Unit – unit prescribed to each voxel within a CT image, to describe its density or attenuation coefficient.
ICC	Intraclass correlation coefficient is a descriptive statistic used to assess the consistency, or conformity of measurements made by multiple observations or observers (Section 4.3.2)
ILT	Intra-luminal Thrombus is a blood clot that forms within the growing

	aneurysmal sac. Most commonly, it is visualized following IV contrast injection in CT. It is believed to play a role in AAA growth and progression.
IQR	Interquartile range – a descriptive statistic used to represent the spread of data between the 75^{th} and 25^{th} percentile.
I_{Tube}	X-Ray tube current is an image setting that affects the energy and number of electrons released during CT image acquisition (Section 5.2.6.3)
IV	Intravenous – method of contrast and radioactive FDG tracer injection prior to CECT and PET imaging, respectively.
L	Low Pass Decomposition Filter (Wavelet Transforms) - used in image processing to selectively emphasize Low-frequency image details within an image input (Sections 5.3.2, 7.3.3-4)
LDA	Linear Discriminant Analysis is a statistical method to find a linear combination of significant features that separates the input data into two classes/events (Sections 6.3.4.1-2)
LOG	Laplacian of Gaussian using a defined sigma is an edge enhancement filter that emphasizes areas of grey level change within the input image. Low and high sigma emphasizes fine ((grey level change over short distances) and coarse features (grey level change over large distances)
MA3RS study	MRI in AAA to predict Rupture or Surgery Study – An ongoing prospective observational multicentre cohort study of 350 patients with diagnosed AAA to monitor aneurysmal growth/progression (Section 3.2).
MAE	Mean Average Error - Metric used to assess the congruence between two sets of results.
MRI	Magnetic Resonance Imaging, an imaging modality using magnetic fields and radio waves to generate images.
MRMR	Minimum Redundancy, Maximum Relevance algorithm is a Feature Selection algorithm in MATLAB used to capture features that are integral for a particular classification task.
NCCT	Non-Contrast CT - This is identical to a CT; however, it draws attention to the fact that IV contrast was not introduced (vs. CECT).
NFI	Non-Fusiform Index is a 3-D shape index that describes the deviation of the aneurysmal sac from an ideal fusiform shape.

NHS	National Health Service is the umbrella term for the health care system within the UK. All images obtained from the Oxford cohort were obtained from consented patients within the OUH-NHS system.
OUH	Oxford University Hospitals is the local health care network operating within the NHS.
OWS or OW	OWS or Outer Wall Structure is the combination of the intra-luminal thrombus (ILT), if present, and outer wall within the abdominal region of the aorta. The two components are not easily distinguishable due to CT image resolution. OW is the outer wall of the aorta within the thoracic aorta.
OxAAA study	Oxford Abdominal Aortic Aneurysm study is a prospective study designed to longitudinally examine the natural progression of AAA disease.
PACS	Picture archiving and communication system
PAOD	Peripheral arterial occlusive Disease develops when the peripheral blood vessels become damaged and is a key component of a patient's past/current medical history (Section 3.1)
PCA	Principal Component Analysis is used for dimensionality reduction. It is performed by constructing principal components to obtain lower-dimensional data while preserving the data's variation (Section 6.2.4).
PET	Positron Emission Tomography is an imaging modality that relies upon the detection of gamma ray pair emitted from the decay of an injected radioactive tracer.
r	Spearman correlation coefficient is a numerical measure (-1 – 1) that describes the statistical relationship between two variables.
RC	Radius of Curvature is a geometric feature calculated from adjacent points along the aortic centreline as a measure of local curvature. The smaller the RC, the greater the local curvature (Section 6.2.3.2)
RMSE	Root Mean Square Error - Metric used to assess the congruence between two sets of results.
ROC curve	Receiver Operator Characteristic Curve is a representation of classifier performance for a particular task over a sensitivity range. (Section 5.3.2, 6.3.4, 6.3.6, 7.3.3-4, and 7.3.5.2)
ROI	Region-of-Interest describes a sub-region within the larger CT image with the relevant pathology (ex. the segmentation pipeline first identifies the ROI with the Aorta and uses that ROI for high-

	resolution segmentation, Section 4.2.11, and 4.3.4).
SD	Standard Deviation - a descriptive statistic used to represent the spread of data.
SMA	Superior Mesenteric Artery – one of the six major side branches originating from the descending aorta (Section 4.3.6)
SUV	Standard Uptake Value are the pixel values of a PET image that are standardized by patient weight, radiation dose and the interval between injection and imaging (Section 7.2.2).
TCIA	The Cancer Imaging Archive – Online resource with publically available imaging and clinical datasets. This database was used to identify the HNSCC dataset utilised in Chapter 7.
Thor	Thoracic – anatomical region bounded superiorly by the neck and inferiorly by the diaphragm. Contains the ascending and descending aorta along with the aortic arch (Section 4.2.12)
UI	Undulation Index is a measure of the degree of surface irregularity and asymmetry and was used as an indicator to predict aneurysmal growth (Section 6.2.3.3).

1 Preface

Contents

1.1 Motivation 22
1.2 Contribution 23
1.3 Publications Derived from thesis 24
1.4 Originality and Individual Role 26
1.5 Authorship 28

1.1 Motivation

Medical imaging is an important disruptive technology that provides clinicians insight into the intrinsic properties of biological tissues and allows for the non-invasive delineation, characterisation, and assessment of the target tissue[1]. It is a rapidly growing health care service that has evolved from being primarily a diagnostic tool to a platform that is integral in delivering personalized precision medicine[2].

The most common large-device imaging modality both in the UK and worldwide is non-contrast Computerized Tomography (CT)[3,4]. This modality visualizes internal structures by monitoring the attenuation of X-Rays as they pass through tissues of different density, within the patient. Hardware and software advancements have allowed for the rapid acquisition of a larger region-of-interest from multiple orientations while reducing the radiation dosage delivered to the patients[1,5]. Similarly, post-processing and 3-dimensional visualization algorithms can help provide a patient-specific understanding of the diagnosed pathology[6].

In certain cases, additional imaging is obtained when a unique visualization is required and is often the standard of care. When treatment of an artery is being considered and a detailed view of the arterial anatomy is required, a CT study with intravenous iodinated contrast is obtained. On the other hand, when staging a malignancy obtaining a positron emission tomography (PET) study, which captures the metabolic activity within the region-of-interest, is required.

However, these imaging studies are associated with a variety of disadvantages that prevent universal implementation. Contrast agents are injected intravenously and are nephrotoxic in patients with a decreased baseline renal function or concomitant chronic kidney disease[7]. Specific staging criteria and protocols, endorsed by the American College of Radiology, are applied for patients undergoing contrast-enhanced CT imaging, based upon the patient's Glomerular Filtration Rate (eGFR), to minimize the burden on the kidneys. Similarly, both studies increase the radiation dosage to the patient and can cause significant patient discomfort. Additionally, these imaging modalities are also expensive and require specialist training that are associated with alarming global inequities[8].

Additionally, overutilization of medical imaging is an increasing health care concern. Few studies have highlighted that 20-50% of imaging procedures fail to impact patient outcome and therefore cause an "unnecessary" burden to both health care systems and, more importantly, to the patient[3,9]. One method to limit image overutilization is to maximize the amount of clinically relevant information extracted from each imaging study. This would ensure that additional imaging would only be obtained when necessary.

Here, we investigated if non-contrast CTs, which are currently of minimal clinical value in efficiently differentiating soft tissue components, could be used to stratify disease severity and improve clinical management. Furthermore, focusing on non-contrast CTs would allow for the use of previously obtained or 'historic' medical images. This can be a means to accumulate large data cohorts for model optimization. Highlighting the value of non-contrast CTs could help optimize patient surveillance protocols and medical imaging infrastructures within health care systems, reduce image overutilization and the total per-patient radiation exposure.

1.2 Contribution

The contribution of this thesis is a series of algorithms that allow for the extraction of higher-order features from non-contrast CT imaging studies. These features can subsequently be used to monitor and optimize clinical management. In this thesis, I focus on two specific pathologies, Abdominal Aortic Aneurysmal (AAA) disease and Head and Neck Squamous Cell (HNSCC) Carcinoma as paired imaging (non-contrast and contrast-enhanced CTs for AAA and non-contrast CT and PET for HNSCC) is obtained for both diseases. A description of the datasets and the methods used for de-identification and curation are detailed in Chapter 3.

First, and as described in Chapter 4, an automatic region-of-interest detection and 3-D segmentation pipeline was developed to extract the pathological aorta in AAA patients from both non-contrast and contrast-enhanced CT images. While there are previous studies that investigate semi-automatic and automatic segmentation methods for aneurysms from contrast-enhanced CTs, this is the first time that such a segmentation pipeline has been used to extract the pathological aorta/aneurysm from non-contrast CTs. This approach was trained and independently validated to establish its utility as a post-hoc analysis tool.

Image transformation methods were developed, as described in Chapter 5, to investigate the ability to visualize anatomic features in a non-contrast CT that are commonly only visible following contrast injection. This is the first time such an image transformation has been performed and clinically evaluated. In the aortic aneurysm, this would include the blood flow lumen and the pathological thrombus or blood clot, if present. Obtaining insight into the aneurysmal structure and morphology without the need for intravenous contrast has the potential to improve the patient experience while maintaining the current standard of care.

Subsequently, these deep-learning-based tools were used to derive a 3-D volume of the pathological aorta, which serves as the basis to derive higher-order features to predict clinical outcome (ex. aneurysmal growth), as described in Chapter 6. This thesis describes a novel collection of geometric features, which are independent of demographic features, that characterize the pathological aneurysm and are predictive of aneurysmal growth. Successful models, trained and optimized using the internal imaging cohort, were evaluated against an external validation cohort.

Finally, in Chapter 7, I extended the methods beyond abdominal aortic aneurysms. I implemented image transformation methods to investigate if tumour tissues, with elevated tracer uptake, from patients diagnosed with Head and Neck Squamous Cell Carcinoma could be distinguished in a non-contrast CT image. These machine and deep learning methods, similar to that observed in Chapter 5, highlight for the first time the ability to simulate clinically-accurate PET-like outputs from non-contrast CTs inputs. This suggests that the methods optimized in this thesis for the incorporation of non-contrast images into the clinical pathway can be translatable to similar pathologies.

1.3 Publications Derived from thesis

Much of the work contained in this thesis has been published during the duration of my DPhil. Additional investigations and completed future work are currently being prepared for subsequent publication.

The automatic high-resolution aortic segmentation pipeline from both non-contrast and contrast-enhanced CTs was published in the Annals of Surgery (2020)[10] and is subject to international patent filing (WO2021038203A1). The work presented there included the analysis

from only 75 patients within the Oxford Abdominal Aortic Aneurysm (OxAAA) Imaging cohort. A 3-fold cross-validation approach was used with a data split of 50 and 25 between training and validation. The derived 3-D AAA shapes were evaluated using 1st and 2nd order metrics against manually segmented AAA shapes. It did not include the independent validation cohort of 200 patients within the expanded OxAAA imaging cohort.

The non-contrast to contrast-CT image transformation pipeline in patients diagnosed with AAAs was published in the Annals of Surgery (2021)[11] and is subject to international patent filing (WO2021038202A1). This publication evaluated the transformation capacity of multiple deep generative networks on not only the aortic aneurysm but also the aorta's major side branches. The simulated results were evaluated both technically and clinically. Unlike the more complete analysis described in this thesis, the publication did not include the regional classification (aortic lumen vs. thrombus) based on radiomic features. This was completed to obtain some mechanistic insight into the deep learning networks.

Using geometric features derived from aortic aneurysms to predict aneurysmal growth was also published in the Annals of Surgery (2020)[12] and is subject to priority patent filing within the UK (RTA/P299591GB). This publication utilized the retrospectively curated OxAAA imaging cohort to identify certain geometric features that not only characterized the aneurysmal shape but were also predictive of growth as a categorical ("Slow" or "Fast") or continuous outcome. However, it did not incorporate the characterization of the aneurysmal shape using principal component analysis or validation of the growth prediction models on an independent validation cohort that comprise the original contribution described in this thesis.

Finally, the non-contrast to PET image transformation task described in detail Chapter 7, has been submitted for publication and is subject to international patent filing (PCT/GB2021/051141).

1.4 Originality and Individual Role

Generally, all image segmentation, transformation, feature extraction, and machine/deep learning model optimization and validation are my individual, original work. Clinical study design, patient recruitment, clinical image acquisition/curation were performed by the clinical collaborators. Validation of image segmentation methods were also performed with the help of clinical collaborators. Specifically:

Chapter 3: Clinical collaborators designed the overarching OxAAA study and acquired the Non-contrast and contrast-enhanced CT images that created the multiple OxAAA imaging datasets (OxAAA-IMG-I – IV). Pierfrancesco Lapolla (PS) was integral in retrospectively identifying AAA patients that (1) underwent elective repair of AAAs and (2) provided written consent to use their current and previous clinical data (ex. medical images) for research purposes during the consent process for surgery. Following identification, he collected and curated the OxAAA imaging cohort that was used extensively in this thesis.

The MA3RS imaging dataset was curated by the ongoing prospective multicentre observational cohort MAR3RS study organized by the University of Edinburgh[13]. This dataset was kindly provided for model validation in Section 6.3.6.

Finally, the head and neck cancer dataset is a cohort of paired PET/CT Images from four different institutions in Quebec, Canada[14]. The complete dataset was made publicly available via The Cancer Imaging Archive (TCIA) at http://www.cancerimagingarchive.net. This dataset was used extensively in Section 7.

Chapter 4: The methods developed in this chapter were used to automatically segment the diseased aorta (with abdominal aortic aneurysms) and its associated side branches from the curated non-contrast and contrast-enhanced CT images. With collaborators, I used publicly available software to manually segment the CT images. PL and Natesh Shivakumar (NS), both members of the larger OxAAA study, helped with the inter-observer variability assessments of the manually annotated segmentations.

Independently, I designed and implemented a multi-step pipeline using deep learning methods for the automatic region-of-interest detection and segmentation of both the aortic

lumen and the pathological aneurysm. I expanded this method to capture and classify the aortic side branches. I then evaluated the utility of this pipeline by (1) validating the accuracy of automatically-extracted measurements and (2) comparing clinically relevant measurements extracted from the model predictions and against that from the manual segmentations.

Chapter 5: The methods developed in this chapter focused on optimizing the non-contrast to contrast-enhanced CT image transformation task using machine and deep learning methods. PL and NS helped with the initial proof-of-concept experiments by visually inspecting and evaluating the registration accuracy between Non-contrast and contrast-enhanced CT axial slices in Section 5.2.4.

Independently, I optimized and evaluated multiple methods for this image transformation task. I used (1) random forest methods to extract the radiomic signatures between visually indistinct regions within the aortic aneurysm from a non-contrast CT image and (2) deep learning generative algorithms to simulate a contrast-enhanced visualization. I, then, evaluated this simulated pseudo-contrast images for both technical and clinical accuracy.

Chapter 6: Collaborators were primarily involved in collecting and curating the CT imaging cohorts used for both growth prediction model training, optimization and validation. Specifics involving the study datasets are found in detail within Sections 3.1 and 3.2.

Independently, I extracted geometric features that characterize and capture the aneurysmal shape from the derived segmentations; applied regression to generate predictive models from the derived feature sets; evaluated the optimized models on an independent validation cohort. Jorge Corral Acero, a fellow DPhil student working under Dr. Grau, provided significant guidance on regression model optimization.

Chapter 7: This chapter focused on optimizing the non-contrast CT to PET image transformation task using both radiomic and deep learning methods. Clinical collaborators assisted with the manual segmentation of metabolically-active tumor regions. Joel Ward, a clinician within the ENT surgical service, was integral in segmenting the Thyroid tissue. This

segmentation served as a control to the metabolically-active tumor tissue within the radiomic experiments.

I used (1) random forest methods to extract the radiomic signatures between regions of negligible, low and high FDG uptake on a non-contrast CT image with a focus on malignant tissues and (2) deep learning generative algorithms to simulate a PET-like visualization. I, then, evaluated this simulated pseudo-PET images for both technical and clinical accuracy.

1.5 **Authorship**

I am the sole author of this thesis document and produced all included tables and figures. Figures that have been included my publications are referenced within their respective captions. Critique, proofreading, and comments were provided by Dr. Regent Lee, Prof. Vicente Grau, Prof. Ashok Handa, Pierfrancesco Lapolla, Natesh Shivakumar, Raman Uberoi and Joel Ward. All errors and omissions are my own.

2
Introduction

Contents

2.1 Increasing role of medical imaging in clinical care 30
2.1.1 Computerized Tomography (CT) Imaging 30
2.1.2 Computerized Tomography Imaging with Intravenous Contrast (CECT) 31
2.1.3 Positron Emission Tomography (PET) 31
2.2 Overutilization of medical imaging is an increasing health care concern 32
2.3 Potential solutions to limit medical image overutilization 33
2.4 Organ Isolation from medical images is required for feature extraction 33
2.5 Integration of deep learning in medical imaging improves clinical utility 34
2.6 Radiomic extraction from medical images allows for personalized medicine 35
2.7 Geometric analysis (3D) provides unique insight into disease progression 36
2.8 Hypothesis and Aims 37
2.8.1 Study Objectives 37
2.8.2 Hypothesis 37
2.8.3 Aims 37

2.1 Increasing role of medical imaging in clinical care

The intrinsic properties of biological tissues vary spatially and temporally secondary to structural and functional changes, including those caused by disease and disability[1,5]. These properties are often captured by medical imaging and permits the accurate and non-invasive delineation, characterization, and assessment of the targeted diseased tissue. As a result, medical imaging is an important technology that is actively used in clinical practice to aid in decision making. Over the years, this disruptive technology has evolved from being primarily a diagnostic tool to a platform that is vital for personalized precision medicine[15]. Furthermore, medical imaging has allowed millions of persons to avoid invasive and costly procedures by (1) yielding definitive diagnostic information, and thereby by removing the need for surgical repair or (2) enabling minimally-invasive surgical alternatives.

During the past decade, imaging services have grown about twice the rate of other medical technologies/services (laboratory procedures and pharmaceuticals)[16]. This rapid integration of medical imaging into clinical care is linked with recent technological advances involving image acquisition, reconstruction, post-processing and visualization. Optimization of the image acquisition hardware has allowed for the rapid/simultaneous acquisition of a larger region-of-interest (ex. whole body) from multiple planes/orientations (ex. coronal, sagittal planes) while reducing the radiation dosage delivered to the patient[1]. Subsequently, 3D reconstruction algorithms in-built within the imaging module allow for the efficient conversion of acquired image projections (2-Dimensional, 2D) to an accurate 3D representation of the region-of-interest in the form of a 3D DICOM array[6]. Post-processing and visualization can then be used not only to diagnose but also to obtain a patient-specific understanding of the diagnosed pathology. Two of the largest growing imaging modalities, not only in the UK but also worldwide, are Computerized Tomography (CT) and Positron Emission Tomography (PET) imaging.

2.1.1 Computerized Tomography (CT) Imaging

According to the Diagnostic Imaging Dataset Statistical Release published by the NHS, Computerized Axial Tomography (CT) scans accounted for ~15% of all imaging tests reported in England between September 2018 and September 2019, with approximately 0.47 million scans being performed each month[17]. Similarly, CT imaging is the most commonly obtained study in the EU and the US, with ~80-85 million scans being performed each year[3,4].

Mechanistically, a CT operates by using an X-ray generator that rotates around the patient along with an X-ray detector, opposite to the x-ray source. As the X-rays pass through the patient to the detector, they are attenuated differently by various tissues, based on tissue density[1]. The resulting distribution is then processed internally using a form of tomographic reconstruction, which produces the 3-D cross-sectional image set that is then interpreted by clinicians. The pixel values within the reconstructed CT are displayed according to the mean attenuation of the local tissue on the Hounsfield Unit (HU) scale[5].

2.1.2 Computerized Tomography Imaging with Intravenous Contrast (CECT)

In certain cases, intravenous (IV) iodinated contrast is injected to enhance the intrinsic tissue contrast between regions that are difficult to differentiate on the non-contrast CT (NCCT) image. This is the most commonly used contrast agent overall[7]. Injected into the vascular lumen, IV iodinated contrast increases the density and thus the attenuation of blood with which it mixes[7]. On the reconstructed 3D DICOM array, the vasculature appears bright adjacent to darker soft tissues. This is especially important when treatment of an artery is being considered and a detailed view of the arterial anatomy is required.

However, Contrast-enhanced CTs (CECTs) are associated with several disadvantages[18,19]. Contrast media have been associated with hypersensitivity reactions (HSRs), with a recent study suggesting that 16.8% of contrast-related HSRs (incidence: ~1.0%) to be classified as moderate to severe. Similarly, these agents are nephrotoxic in patients with a decreased baseline renal function or concomitant chronic kidney disease[19]. In both instances, modified contrast injection protocols are widely being applied with the aim of minimizing the risk of adverse reactions.

2.1.3 Positron Emission Tomography (PET)

Positron Emission Tomography (PET) is an imaging modality that can be used to visualize abnormal metabolic activity. This is especially important in biological tissues that do not appear pathological based on their morphology[20]. In the UK, approximately 200,000 paired PET-CTs are performed annually with an annual increase of 14-16%. It is a widely adopted clinical tool for the diagnosis, staging and follow-up for a variety of malignancies (pulmonary nodule[21], melanoma[22], head and neck squamous cell carcinoma[14,23], etc). It provides clinicians with a semi-

quantitative representation of the tissue's metabolic activity and can be used to guide further treatment[20,24,25].

Functional PET-imaging is based on the detection of photons released when injected radionuclides are metabolized in tissues[20,25]. This modality takes advantage of the differences between tumour and healthy tissue physiology. The hallmarks of such malignant tissues are rapid proliferation/angiogenesis, increase in size, local invasion, and distant metastasis[26]. At the molecular level, malignant cells have increased glucose utilization due to an upregulation of enzymatic activity. As a result, injection of a glucose-based radionuclide, 2-[fluorine-18]fluoro-2-deoxy-d-glucose (FDG), can be used to identify these abnormal metabolically-active tissues. The rate of uptake of FDG into malignant tissues has been shown to be proportional to its metabolic activity[20]. However, unlike glucose, FDG is not fully metabolized and becomes trapped within active cells. This accumulation of FDG is what is observed in a PET image.

PET images are obtained alongside a NCCT images to enable the localization of areas of increased metabolic activity with their underlying anatomic structures. Co-registering functional (PET) and anatomic (CT) information improves clinical confidence in decision making and is common practice[23,27,28].

2.2 Overutilization of medical imaging is an increasing health care concern

A substantial component of the growth of medical imaging services within clinical practice is due to its overutilization for both diagnosis and image-guided therapy[16,29]. The definition of overutilization, in this instance, is the acquisition of additional medical images that have minimal impact on patient care or patient outcome[29,30]. Some publications have suggested that 20-50% of imaging procedures fail to provide additional information to improve patient welfare and were deemed "unnecessary"[3,9]. Additionally, an increase in medical image acquisition is linked to an increase in average per-person radiation dosage[3,31]. This was observed by a 7.1-factor increase in the per-person radiation dosage between 2006 and 1980[31,32]. There are many causes of overutilization in medical imaging including the underlying health care system (e.g. fee-for-system), lack of comparative effectiveness research between imaging modalities[33,34], poor adherence to image appropriateness criteria[35], referring physicians[36], patient request[37], and duplicate imaging studies.

2.3 Potential solutions to limit medical image overutilization

Methods to limit image overutilization and reduce the radiation dose to the patient are to (1) improve the collaboration and communication between all relevant groups and (2) maximize the amount of clinically-relevant information extracted from each imaging modality. The latter requires comparative effectiveness research to refine appropriateness criteria for each imaging modality and to clearly delineate disease practice guidelines.

One existing solution is to implement post-processing methods to decipher clinically-relevant information from obtained medical images. These methods can be broadly characterized into (1) radiomic and (2) geometric evaluation of the region-of-interest from medical images. Deep learning techniques are commonly used for region-of-interest isolation to automate this process.

Given that NCCTs are the most common type of large imaging modality (vs MRI and PET)[3,4,8,38] and are commonly obtained alongside both contrast-enhanced CT and PET studies, we investigated if they could provide information normally only obtained from the other imaging modalities. Currently, these NCCTs are of limited clinical value and their ability to stratify disease severity and improve clinical management has never before been investigated.

Furthermore, focusing on NCCTs would allow for the use of previously obtained or 'historic' medical images to predict disease progression and clinical outcome. This can be used to accumulate large data cohorts for training/ validation during model optimization. Highlighting the value of NCCTs could help optimize patient surveillance protocols and medical imaging infrastructures within health care systems, reduce image overutilization and total per-patient radiation exposure.

2.4 Organ Isolation from medical images is required for feature extraction

A common pre-requisite for both radiomic and geometric-based approaches is a robust segmentation method to isolate the 3D-region-of-interest from the medical image[39,40]. This is even more essential when focusing on NCCTs. Manual segmentation approaches are labour-intensive and are susceptible to bias/errors that may confound the analysis[40]. Semi-automatic segmentation methods are more common and implement a combination of threshold-based[41,42] or region-growing[43] based approaches. Both methods take advantage of the original pixel relationships within the region-of-interest to try and separate it from surrounding tissues and

structures. The former defines a fixed threshold, above which is the region-of-interest (foreground) and below which is the background. The latter is an adaptive method which expands the segmentation into adjacent regions of similar pixel distributions[43]. These methods can only work for structures that contain a HU distribution that is different from surrounding tissues (ex. to isolate the IV contrast in a contrast-enhanced CT). On the other hand, automatic segmentation methods usually employ machine or deep learning models to rapidly isolate the region-of-interest[40]. The benefit of these models are that (1) they can be trained to isolate volumes irrespective of HU distribution (ex. both non-contrast and contrast-enhanced CT images), and (2) they do not require complex user input. This is an active area of research, and the results are organ- and pathology- dependent. The segmentations derived from each method need to be evaluated and corrected by a trained clinician to ensure accuracy.

2.5 Integration of deep learning in medical imaging improves clinical utility

In medical imaging, the implementation of sophisticated deep learning algorithms has been heralded as the greatest disruptive technology in recent history[44]. The primary driver for this integration was the need to improve clinical decision algorithms and decrease clinician error rates[45]. The most commonly quoted applications of DL in medical imaging are for object segmentation (delineation of organ/lesion boundary), object classification (ex. determine clinical severity of organ/lesion), multimodal image coregistration (non-rigid warping of a target image to ensure anatomic matching to a reference image) and image transformation (converting one image type to another to visualize intrinsic features)[46-49].

DL employs artificial neural networks, which are built using thousands to millions of nodes arranged in a pre-defined set of layers[47,50,51]. As the number of layers increase so too does the complexity and richness of the final model output; having more layers has been shown to increase test time accuracy[47]. The general architecture of the implemented DL networks is dependent on the prescribed task (ex. object segmentation or image transformation). Each node receives information from other nodes and its outputs at each node are weighted. During model training, the DL network minimizes the difference (error) between its final output and the reference (ground truth) by iteratively modifying its internal weights. The trained model is then evaluated against a smaller data set (testing) to assess its accuracy and robustness. Although

many of these DL models indicate performance at or above the level of a trained clinician, they require vast amounts of diverse data and an advanced infrastructure to be robust, viable and applicable within the clinic[44,47].

2.6 Radiomic extraction from medical images allows for personalized medicine

Radiomics is a methodology coined by Lambin et al.[2] that transforms pathological medical images into mineable high-dimensional data that can eventually be used to monitor and predict clinical outcome. It is a quantitative and reproducible-method to calculate pixel-based relationships involving intensity and distribution ("texture") within a pre-defined region-of-interest that can be integrated into clinical decision support systems to improve medical decision making[2,15]. Many radiomic methods have been previously validated to rapidly calculate quantitative features from a variety of medical images such as computerised tomography (CT)[52], magnetic resonance (MR)[53] and/or positron emission tomography (PET)[23,54]. These methods are most commonly used in cancer-based imaging to potentially obtain information complimentary to other pertinent data (ex. clinically-obtained, treatment-related, genomic/genetic information and/or pathological data). It attempts to characterize the region-of-interest (ex. tumour) through image-based features that may provide insight into its phenotype (ex. cancer severity, stage/grade) and the surrounding microenvironment[15]. Many clinical studies have investigated the impact of integrating radiomics-derived features with other clinically-relevant features for clinical-decision making and its potential and impact have been repeatedly confirmed[55,56].

However, many limitations exist that prevent the immediate integration of radiomics into the clinical pipeline[15,46,57]. Radiomics methods require (1) standardized high-quality image acquisition protocols between different scanners/institutions that minimizes the technical differences between the obtained images, (2) large, curated image databases across multiple institutions with a diverse array of disease phenotypes for both model training and validation and (3) powerful computing resources for simultaneous radiomic feature extraction, clinical feature integration and clinical decision-support model training. Although the infrastructure to allow for immediate prospective analysis of radiomic methods is lacking, many institutions

contain large repositories of retrospectively-collected 'historic' imaging data. This can be used to provide insight into whether such methods can increase the personalized delivery of medicine.

2.7 Geometric analysis (3D) provides unique insight into disease progression

Geometric analysis focuses on the 3D shape characterization of the pathological region-of-interest and is independent of its underlying pixel/HU distribution. Similar to radiomic methods, this analysis deconstructs the region-of-interest into features that can then be used to predict clinical outcome (ex. disease progression) and has been employed for a variety of pathological conditions[58-60]. There is well-documented evidence that vascular geometry has a major impact in blood dynamics and in the origin and development of vascular disease (ex. atherosclerotic aneurysms, cerebral aneurysms) secondary to the hemodynamic forces on the vascular wall[58-61]. Therefore, the study of vascular geometry with regards to a particular pathology can provide insight on the hemodynamic triggers for disease pathogenesis and/or progression. Directly investigating these triggers, as opposed to the geometric features, is a computationally intensive ordeal that is not feasible in a larger scale study[62]. In the case of cerebral aneurysms, which is a localized dilatation of the cerebral arterial wall, a collection of geometric features describing its size and shape are predictive of aneurysmal rupture risk[58,61]. These easily-derived features act as surrogates for specific flow parameters and are amenable for inclusion in larger scale studies.

2.8 Hypothesis and Aims

2.8.1 Study Objectives

In this thesis, I investigate the role of Non-contrast CT imaging in the disease progression of two specific pathologies, Abdominal Aortic Aneurysmal (AAA) disease and Head and Neck Squamous Cell (HNSCC) Carcinoma. For these pathologies, current guidelines require obtaining more intensive rounds of imaging including a CT with IV contrast or PET images, which are used in conjunction with the NCCT for clinical decision making.

2.8.2 Hypothesis

The hypothesis of this thesis include:

1. Non-Contrast CT images contain sufficient information to differentiate visually indistinct soft-tissues and to stratify disease progression without the need to obtain additional or more intensive imaging studies. Here, I focus on two pathologies, AAA and HNSCC.
2. Machine learning methods facilitates the extraction of higher order information from NCCT images.

2.8.3 Aims

The aims of the study include:

1. Train a deep-learning-based method for the automatic segmentation of the aneurysmal aorta and its side-branches from non-contrast and contrast-enhanced CT images and assess its clinical accuracy (Chapter 4).

2. Employ a radiomics and deep-learning based approach for the simulation of clinically-appropriate contrast-enhanced CTs from NCCTs (Chapter 5).

3. Develop a model to predict abdominal aortic aneurysm growth and progression using 3D geometric indices extracted from aneurysmal volumes (Chapter 6)

4. Employ a radiomics and deep-learning based approach for the simulation of clinically appropriate Positron Emission Tomography (PET) maps from NCCTs (Chapter 7).

3
Overview of Implemented Datasets

Contents

3.1 Imaging Cohorts from the OxAAA Study (OxAAA-IMG) 39
3.1.1 Defining the OxAAA-IMG-I dataset 41
3.1.2 Defining the OxAAA-IMG-II dataset 41
3.1.3 Defining the OxAAA-IMG-III dataset – Subset of Imaging Data with Demographic Information 42
3.1.4 Defining the OxAAA-IMG-IV cohort - Subset of Patients with Growth Measurements 42
3.2 Imaging Cohort from the MA3RS Study 44
3.3 Imaging cohort of patients with Head and Neck Cancer 45

The primary objective of this thesis was to investigate if NCCT images contain sufficient information to stratify disease progression. Current stratification methods involve obtaining more intensive rounds of imaging (ex. with intravenous contrast (CECT) or with a radioactive tracer (PET)) for clinical decision making. In this thesis, I focus on two specific pathologies, Abdominal Aortic Aneurysms (AAA) and Head and Neck Squamous Cell carcinoma (HNSCC) as paired imaging is obtained for both diseases. This chapter introduces and explores the datasets that will be used to tackle each of the objectives detailed in Section 2.8.

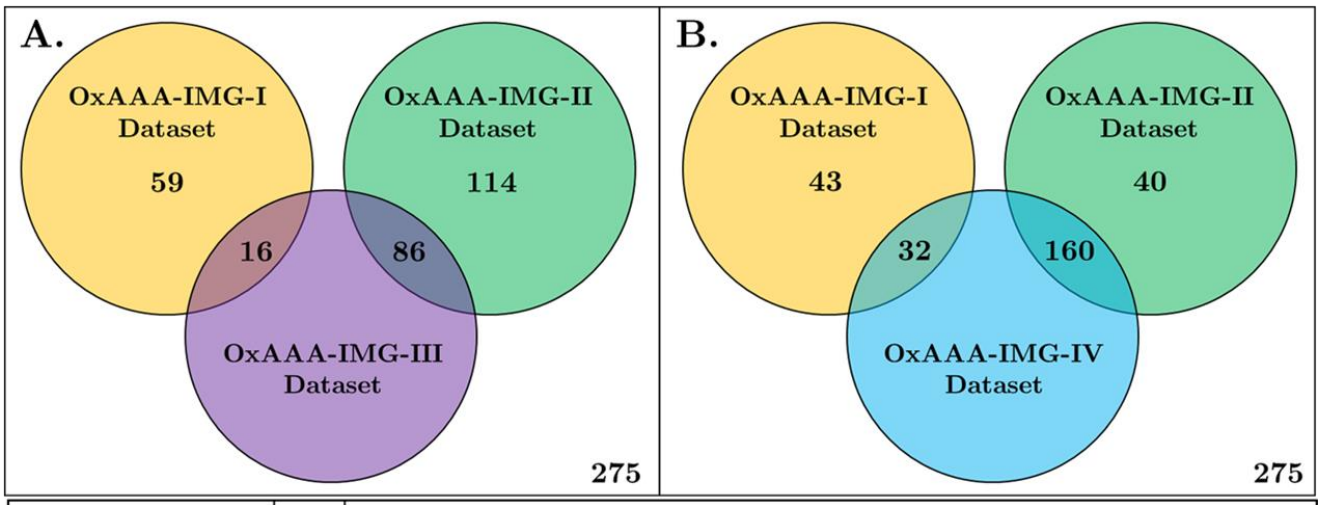
3.1 Imaging Cohorts from the OxAAA Study (OxAAA-IMG)

CT imaging data from patients with diagnosed AAA was acquired from the ongoing Oxford Abdominal Aortic Aneurysm (OxAAA) study. The study received full regulatory and ethics approval from both Oxford University and Oxford University Hospitals (OUH) National Health Services (NHS) Foundation Trust (Ethics Ref 13/SC/0250). Generally, this study is a prospective study designed to longitudinally examine the natural progression of AAA disease. Details regarding the OxAAA study cohort and recruitment process have been published[63]. Each patient gave written consent for the utilization of clinical images collected during the routine clinical management pathway for research analysis.

In addition, we utilised the clinical database (Oxnet Janus), which prospectively registered patients who underwent elective repair of AAAs at the John Radcliffe Hospital, Oxford, UK. During the consent process for surgery, patients gave written consent for the utilisation of their clinical data (including images) for utilisation by research with ethics and regulatory approval. Medical records of all non-emergency infra-renal AAA repairs (open or endovascular repair) from 1st of Feb. 2009 to 30th June 2018 were examined to identify patients who gave consent for utilisation of their clinical data for research.

As part of the routine pre-operative assessment for AAA, a NCCT of the abdomen and a CECT of both the chest and abdomen were performed. CECT images were obtained following contrast injection in helical mode with a pre-defined slice thickness of 1.25 mm. NCCT images included only the abdominal aorta and were obtained with a pre-defined slice thickness of 2.5 mm. Paired images were anonymised within the OUH PACS system before being downloaded

onto the secure study drive. These images formed the OxAAA-IMG cohort (n = 275 patients, **Fig. 1**)



Dataset	N	Dataset Implementation	
OxAAA-IMG-I	75	Sections 4.2.1-4.2.14	Model Training + Optimization of the Automatic AAA Segmentation Pipeline (3-fold Cross-Validation)
		Sections 5.2.1-5.2.7	Model Training + Optimization of the Non-Contrast to Contrast Image Generation Pipeline (3-fold Cross-Validation)
OxAAA-IMG-II	200	Sections 4.2.15	Internal validation of the automatic AAA segmentation pipeline
		Sections 5.2.8	Independent validation of the Non-Contrast to Contrast Image Generation Pipeline
OxAAA-IMG-III	102	Sections 6.3.1	Evaluate the correlations between demographic features and extracted biomarkers
OxAAA-IMG-IV	192	Sections 6.3.4-6.3.5	Model Training + Optimization of the AAA Growth Prediction Algorithm (Iterative 10-Fold Cross-Validation)

Fig. 1: Defining the OxAAA Imaging Cohorts used in this thesis. This cohort comprised of NCCT and/or CECT images obtained from patients (n = 275) diagnosed with Abdominal Aortic Aneurysms (AAA). This cohort was divided into 4 datasets. The OxAAA-IMG-I dataset comprised of imaging data from 75 patients and were used for model training and optimization used in Sections 4 and 5. Subsequently, OxAAA-IMG-II dataset comprised of the remaining 200 patients and were used as an independent dataset for model validation (Chapters 4 and 5). OxAAA-IMG-III dataset was a subset of 102 patients from the larger cohort of 275 patients with prospectively collected demographic data (A.). This was used in Section 6.3.1 to evaluate the relationships between the demographic features and extracted biomarkers. Finally, OxAAA-IMG-IV dataset comprised of imaging data from 192 patients with serial CT scans (B.) that could be used to calculate AAA growth (Section 6.2.1).

3.1.1 Defining the OxAAA-IMG-I dataset

From this OxAAA-IMG cohort of 275 patients, CT imaging data from 75 patients (NCCT and CECT) were randomly selected for model training (**Fig. 1)**. This sub-cohort was used to train and optimize the methods for Aortic/AAA segmentation (Section 4.2.4 - 4.2.11), aortic side-branch segmentation and classification (Section 4.2.12-4.2.13), and Non-Contrast-to-Contrast CT image transformation (Section 5.2.6). A 3-fold cross-validation approach was applied with this cohort to maximize the images used for model evaluation. For each fold, the testing cohort consisted of a unique set of 25 patients and the training cohort consisted of the remaining 50 patients.

3.1.2 Defining the OxAAA-IMG-II dataset

The remaining 200 patients (NCCT and CECT) were introduced as an internal validation cohort (**Fig. 1)** to evaluate the trained and optimized models (Aortic/AAA segmentation – Section 4.2.15, Aortic side-branch segmentation/classification - Section 4.2.15 and Non-Contrast-to-Contrast Image Transformation - Section 5.2.8). **Table *1*** highlights the similarities between the two imaging datasets derived from the DICOM header information or image.

Table 1: Image cohort comparison: (1) OxAAA-IMG-I: used for model training, n = 75 and (2) OxAAA-IMG-II used for model validation, n = 200.

		OxAAA-IMG-I Dataset (n = 75)		OxAAA-IMG-II Dataset (n = 200)		p-value
Contrast (CECT)	25th Percentile HU [95% CI]	-1009	[-1006 -1021]	-1008	[-1007 -1012]	0.48
	Mean HU [95% CI]	-562.6	[-621.1 -515.3]	-579.4	[-617.2 -536.6]	0.32
	75th Percentile [95% CI]	-76.5	[-83.3 -47.8]	-80.4	[-90.8 -57.4]	0.22
	Standard Deviation [95% CI]	488.6	[464.0 497.2]	493.8	[483.1 500.9]	0.10
	Voxel Length [95% CI]	0.82 mm	[0.77 0.84]	0.84 mm	[0.80 0.86]	0.62
	Voxel Height [95% CI]	0.82 mm	[0.77 0.84]	0.84 mm	[0.80 0.86]	0.62
	Voxel Thickness	1.25 mm		1.25 mm		-
	KiloVoltage Peak (kVP)	120		120		-
	Exposure Time [95% CI]	454	[369.2 518.8]	484.2	[389.3 529.0]	0.42
	X-Ray Tube Current (mA)	215.3	[123.3 445.7]	279.1	[140.2 437.4]	0.46
Non-Contrast (NCCT)	25th Percentile HU [95% CI]	-1004	[-987 -1012]	-986.7	[-993.4 -996.7]	0.09
	Mean HU [95% CI]	-558	[-522.4 -601.1]	-551.3	[-586.5 -506.3]	0.63
	75th Percentile [95% CI]	-52.8	[-66.0 -39.6]	-46.6	[-61.9 -31.2]	0.39
	Standard Deviation [95% CI]	476.7	[472.0 488.4]	483.0	[478.4 480.3]	0.37
	Voxel Length [95% CI]	0.82 mm	[0.74 0.84]	0.81 mm	[0.77 0.86]	0.59
	Voxel Height [95% CI]	0.82 mm	[0.74 0.84]	0.81 mm	[0.77 0.86]	0.59
	Voxel Thickness	2.5 mm		2.5 mm		-
	KVP	120		120		-
	Exposure Time [95% CI]	447.3	[425.2 455.8]	452.2	[428.3 465.5]	0.63
	X-Ray Tube Current (mA)	353.2	[218.8 499.4]	357.1	[202.7 515.2]	0.49

3.1.3 Defining the OxAAA-IMG-III dataset – Subset of Imaging Data with Demographic Information

In the OxAAA-IMG-III cohort, 102 participants were prospectively recruited to the OxAAA study at the time of surgery (Male n= 99; Females n= 3). All imaging studies used within this cohort were found within the larger OxAAA-IMG cohort (**Fig. 1A)**. This purpose of this cohort was to assess the correlation between the extracted image biomarkers and patient demographic/clinical characteristics (Section 6.3.1). Demographic information was recorded from each patient, which were matched to the pre-operative CT scan.

For this prospective cohort, the median age at the time of consent was 72 (IQR: 67-79) years old. The majority were ex-smokers (67%) and 24% were current smokers. A history of symptomatic atherosclerotic arterial disease was prevalent in this group (ischaemic heart disease: 32%; peripheral arterial occlusive disease: 16%; cerebral vascular disease: 12%). The majority of participants reported a prior diagnosis of arterial hypertension (73%) and hypercholesterolemia (60%). However, these were well controlled by long term pharmacological therapy [anti-hypertensive(s): 67%, statin: 73%, anti-platelet(s): 46%], as reflected by their controlled mean arterial pressure (102 ± 13 mmHg) and overall normal cholesterol profiles [median = 3.8 mmol/L (IQR 3.2-4.6), lower than 5.2 mmol/L in 82% of participants] at the time of recruitment. Sixteen percent of the participants reported a history of diabetes mellitus, and 27% had chronic kidney disease with eGFR<60. Baseline demographic data from the prospective cohort (n=102) are presented in **Table *2***. Median AAA size within this cohort was 63.0 mm with an interquartile range from 58.0 to 72.5 mm.

3.1.4 Defining the OxAAA-IMG-IV cohort - Subset of Patients with Growth Measurements

This arm consisted of 192 AAA patients from the OxAAA imaging cohort with at least 1 historic CT scan conducted greater than 8 months prior to the pre-operative scan (**Fig. 1B)**. This enabled the calculation of 'prospective' growth data using the baseline scan. The remaining 83 patients did not contain a historic CT scan and could not be used to calculate aneurysmal growth. In this study, the historic CT scans could either be a NCCT or CECT imaging study. Median AAA size and median follow-up time within this cohort was 54.5 mm (interquartile range, IQR of 48.0 to 60.0 mm) and 2.0 years (IQR of 1.0 - 3.7 years), respectively. Similarly,

median aneurysmal growth (ΔMax Diameter /Δ time, yrs) was 3.7 mm/yr with an interquartile range of 2.5 to 5.0 mm/yr.

Table 2: Summary of participant demographics at the pre-surgical assessment (OxAAA-IMG-III dataset)

	All Participants (n = 102)
Male (%)	99(97)
Age at Consent (Median/IQR)	72 (67-79)
Height (±SD)	1.75 ± 0.08
Weight (Median/IQR)	81.9 (74-90.2)
BMI (Median/IQR)	26.8 (24.3-28.7)
MAP (±SD)	102.2 ± 12.8
Current Smoker (%)	24 (24)
Past Smoking Hx (%)	68 (67)
Never Smoked (%)	13 (13)
CAD Hx (%)	33 (32)
Coronary Intervention (%)	26 (25)
PAD History (%)	16 (16)
Cerebral Art. Disease (%)	12 (12)
HTN History (%)	74 (73)
Hypercholesterolemia (%)	61 (60)
Tot. Cholesterol (Median/IQR)	3.8 (3.2-4.6)
HDL (Median/IQR)	1.1 (0.9-1.3)
LDL (Median/IQR)	1.5 (0-2.5)
TG (Median/IQR)	1.2 (0.8-1.6)
Diabetes (%)	16 (16)
HbA1C (Median/IQR)	5.6 (5.4-5.9)
Diabetes - Oral/Insulin (%)	12 (12)
CKD - eGFR < 60 (%)	28 (27)
Creatinine (Median/IQR)	86.5 (73.3-101.3)
Beta-Blockers (%)	32 (31)
ACEI/A2RA (%)	56 (55)
Aspirin (%)	47 (46)
Thienopyridine (%)	9 (9)
Ticragrelor (%)	3 (3)
Anticoagulant (%)	12 (12)
CCBs (%)	43 (42)
Diuretics (%)	22 (22)
Gastro-restraint (%)	31 (30)
Steroids (%)	7 (7)
Statins (%)	74 (73)
AAA Diam (Median/IQR)	63 (58 - 72.5)

3.2 Imaging Cohort from the MA3RS Study

Our growth prediction models were validated using an independent cohort obtained from the ongoing MRI in AAA to predict Rupture or Surgery (MA3RS) study[13]. The MA3RS study is an prospective observational multicentre cohort study of 350 patients with diagnosed AAA in three centres across Scotland[13]. The purpose of this study was to investigate a novel magnetic resonance imaging (MRI) technique for the ability to identify aneurysmal inflammation and expansion. All AAA participants of this study were monitored for aneurysmal expansion over 2 years using either CECT or ultrasound surveillance[13]. Of the 350 patients, 146 patients were excluded as AAA follow-up was performed using only ultrasound measurements and 15 patients were excluded as baseline CT/CECT images were not documented/available. Finally, 41 patients were excluded as AAA intervention was performed prior to the second scan. As a result, CECT images from 148 of the 350 patients were used to validate the trained growth prediction models.

Median AAA size for the MA3RS dataset was 53.4 mm (IQR of 48.4 - 59.1 mm). A secondary CT imaging study was obtained 2 years after the initial study and AAA growth was calculated. Median aneurysmal growth was 2.0 mm/yr with an interquartile range of 1.0 to 3.1 mm/yr. The similarities and differences between the OxAAA-IMG-IV and MA3RS study cohorts are highlighted in **Fig. *2***.

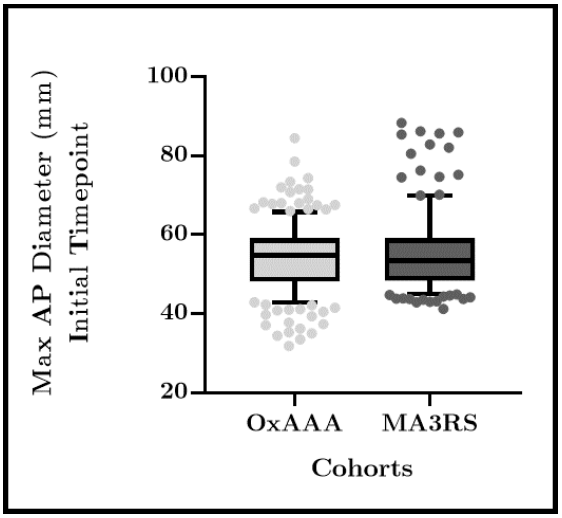


Fig. 2: Comparison of the OxAAA-IMG-IV and MA3RS study datasets. AAA growth rates between the two studies were statistically different ($p < 0.001$).

3.3 Imaging cohort of patients with Head and Neck Cancer

The final chapter of my thesis focuses on the NCCT to PET image transformation task (Chapter 7). The methods optimized here are similar to that used for the NCCT to CECT image transformation task (Chapter 5). Here, we employ machine-learning methods to extract information from NCCT images that are correlated with uptake of the radioactive tracer, FDG in malignant tissues. We extend this concept to simulate clinically appropriate PET maps from NCCT images.

A collection of paired FDG-PET and CT images of 298 patients with diagnosed head & neck squamous cell carcinoma (HNSCC) was prospectively recruited from four different institutions in Quebec, Canada. Vallières et al.[14] utilised this clinical cohort to investigate the impact of radiomic methods for the risk assessment of tumour progression. The complete dataset was made publicly available via The Cancer Imaging Archive (TCIA) at http://www.cancerimagingarchive.net. Details of this study, including patient characteristics and clinical outcomes for each of the patients, are as published and also available through the TCIA repository.

All 298 patients underwent routine treatment management (radiation – 48, 16%; chemo-radiation – 252, 84%). Imaging was obtained within a median of 18 days (range 6 – 66) prior to the start of treatment[14]. At the time of imaging, the median patient weight was 75 Kg (range – 43 -142 Kg) and the median dosage of FDG injected was 1.65 x 10^8 Bq (range – 3.81 x 10^8 – 31.82 x 10^8). Additionally, the median duration between injection and scan time (Δt) was 1.80 x 10^4 s (range – 1.04 x 10^4 – 3.01 x 10^4).

The median follow up time after treatment was 43 months (range – 6 -112 months). Of the 298 patients, 45 patients developed locoregional recurrence, 40 patients developed distant metastasis and 56 patients died. Additional information regarding the patient cohort characteristics can be found within the previously published data documentation[14]. This cohort was used extensively in chapter 7.

4
Aortic Segmentation from CT with or without IV Contrast

Contents

4.1 Introduction 48
4.2 Methods 50
4.2.1 Curation of CT images from a clinical cohort 50
4.2.2 Manual Segmentation of CT Images (Defining the Ground Truth data) 50
4.2.3 Validation of Data Extraction Methods 51
4.2.4 Assessment of intra- and inter- observer variation of manual segmentation 57
4.2.5 Data augmentation 57
4.2.6 U-Net Architecture 58
4.2.7 Attention Gating to strengthen U-Net Performance 58
4.2.8 Loss Function to evaluate Model Performance 59
4.2.9 Aortic Segmentation Pipeline: Image Pre-Processing 60
4.2.10 Aortic Segmentation Pipeline: Aortic ROI Detection 60
4.2.11 Aortic Segmentation Pipeline: Aortic Segmentation 60
4.2.12 Semi-automatic segmentation of the aortic side branches from CECT images 62
4.2.13 Automatic Classification of Aortic Side Branches from CECT images 63
4.2.14 Assessment of model accuracy using aortic morphological features 67
4.2.15 Evaluation of Aortic Segmentation Pipeline on an independent validation cohort 68
4.3 Results 68
4.3.1 CT Image Characteristics 68

4.3.2 Validation of Data Extraction Methods ... *69*
4.3.3 Intra- and Inter- observer variability assessment ... *70*
4.3.4 Model Selection ... *74*
4.3.5 Aortic Segmentation pipeline: ROI selection accuracy ... *78*
4.3.6 Aortic Segmentation Pipeline: Aortic Segmentation accuracy ... *78*
4.3.7 Aortic Segmentation Pipeline: Side-Branch Segmentation and Classification Accuracy ... *80*
4.3.8 Assessment of aortic morphology from CECT-derived segmentations ... *81*
4.3.9 Assessment of aortic morphology from NCCT-derived segmentations ... *86*
4.3.10 Assessment of Aneurysm morphology on an independent validation cohort (OxAAA-IMG-II Cohort) ... *88*
4.4 Discussion ... 90
4.5 Conclusion ... 93
4.5.1 Limitations ... *94*

4.1 Introduction

A computerised tomography (CT) scan uses multiple X-ray measurements to provide a non-invasive visualisation of internal structures. Since the invention of the first commercially-available CT scanner in 1972[64], the use of CT for the diagnosis and disease management is extensively embedded in modern medicine. Visualisation of vasculature on a routine CT is challenging as vessels have similar radio-densities (measured in Hounsfield Unit, HU) to adjacent soft tissues. Injection of intravenous contrast enhances the radio-density within the vessel, enables its visualization and permits rapid segmentation. The produced CT angiogram (CECT) is routinely utilised to for diagnosis. On the other hand, vascular segmentation from NCCT images is a time-intensive and challenging task. Such methods are not readily available to clinicians.

Furthermore, pathological changes, present in the lumen, vessel wall or a combination of both, impede automatic segmentation. In the example of abdominal aortic aneurysms (AAA, abnormal ballooning of the aorta) (**Fig. 3a**, red arrow), a thrombus is adherent to the aneurysmal aortic wall (**Fig. 3b**, red arrow points toward the AAA) in >90% of cases[65]. Existing methods to segment these CECTs are unable to consistently extract the thrombus and the complex thrombus-lumen interface with accuracy. As such, no automated and standardized methods exist to assess aneurysmal diameter (**Fig. 3c**) or thrombus volume. These are vital pieces of clinical information used in the care of AAA patients.

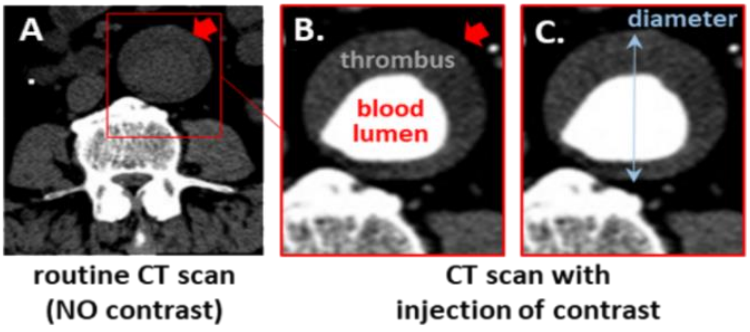


Fig. 3: Axial slice through an abdominal aortic aneurysm (AAA) from a NCCT and CECT image[10].

Prior to the advent of deep learning (DL), vascular segmentation methods incorporated traditional tools including edge detection and/or mathematical models. These methods are

complex, difficult to execute and are poorly generalizable. In the early 2000s, image-based DL methods became more approachable, given significant improvements in hardware. Convolutional neural networks (CNNs), which are the foundation of DL architectures, consist of multiple layers that transform the input using various pre-defined methods (convolution, non-linear activation, pooling, etc). The derived high-level abstractions are then extracted by fully connected layers. Finally, the weights of each neural layer and by extension the model are optimized during training[66,67]. In recent years, many groups have strived to identify improvements to this conventional approach.

One well-known architecture for biomedical image segmentation is the U-Net[68]. This model employs skip connections between layers, which serve to integrate the spatial and contextual information, to assemble a more precise output. Furthermore, these methods, which were initially limited to 2D, have been applied to 3D images to fully utilize spatial information[66,68]. However, due to memory limitations, many 3D U-Net methods utilise down-sampled input images. This input size may not have enough resolution to represent its diverse anatomical variety. This is especially relevant when evaluating structures with variation that can only be captured at higher resolutions[66,69]. Additionally, most methods are not automatic and require complex user input.

In this study, a modified U-Net architecture was implemented to achieve high-throughput, automated segmentation of pathological vessels (AAA) in CT images acquired with or without the use of IV contrast. In CECT images, our method enables simultaneous segmentation of both the outer wall structure (OWS) and lumen to enable characterisation of pathological contents. Here, OWS is the pathological component within AAA and is the combination of the intra-luminal thrombus (ILT), if present, and the aortic outer wall. The model's efficacy was demonstrated by segmenting the thoracic and abdominal aortic regions. Clinical relevance of the trained models was extensively evaluated. Finally, an iterative active contour algorithm was used to expand the CECT-derived segmentations into its side branches. This allowed for evaluation of not only the pathological aortic aneurysm but also its relationship with its major aortic side branches.

4.2 Methods

4.2.1 Curation of CT images from a clinical cohort

Chest and abdominal CT images from 275 patients (OxAAA Imaging Cohort, OxAAA-IMG) were acquired through the ongoing Oxford Abdominal Aortic Aneurysm (OxAAA) study. Full details regarding ethics approval, and image acquisition are detailed in Section 3. As part of the routine pre-operative assessment for AAA, a NCCT of the abdomen and a CECT of both the chest and abdomen were performed. This collection was divided into two cohorts: (1) OxAAA-IMG-I dataset: Seventy-five patients were randomly selected and were used for pipeline training and optimization and (2) OxAAA-IMG-II dataset: paired NCCT and CECT images from the remaining 200 patients served as an internal validation cohort to evaluate the performance of the developed methodology.

4.2.2 Manual Segmentation of CT Images (Defining the Ground Truth data)

In the CECT, both the aortic lumen and OWS were segmented from the aortic root to the iliac bifurcation using the ITK-Snap segmentation software[70] by trained clinicians. Semi-automatic segmentation of the aortic lumen was easily achieved using region-growing by manually delimiting the target intensities between the contrast-enhanced lumen and surrounding tissue. Segmentation of the wall was performed manually by drawing along its boundary using the previously obtained inner lumen as a base. Removing the lumen from the larger segmentation results in a mask highlights the OWS (aortic wall + intra-luminal thrombus (ILT), if present). In the NCCT image, the aorta was manually segmented. These segmentations served as the ground truth (GT) when evaluating model performance.

Clinicians trained to segment NCCT and CECT- images included researchers, radiologists and/or vascular surgeons involved with the OxAAA study. All collaborators have extensive experience in interpreting CT-image data and were subsequently trained to use ITK Snap segmentation software to capture the aortic segmentation. The derived segmentations were extensively validated (Sections 4.2.3, 4.2.4) before model training and optimization.

Axial CECT images depicting the ascending thoracic (yellow arrow), descending thoracic (blue arrow), and abdominal (red arrow) aortic regions are shown in **Fig. 4a-b.** The latter is aneurysmal and contains crescentic layers of thrombus. **Fig. 4c** displays the cross-section of the

abdominal aorta (red arrow) in the NCCT scan. **Fig. 4d-f,** show the CT images with the overlying manual segmentations. 3D volumes derived from the manual 2D segmentations are depicted in **Fig. 4g-h.**

4.2.3 Validation of Data Extraction Methods

All subsequent steps involve the evaluation of either manual- or model- derived segmentations to assess their clinical accuracy. Therefore, it was essential to develop an evaluation pipeline that is both accurate and robust. In order to achieve this, sample aortic aneurysms were simulated using mathematically-derived expressions for its centreline and the diameter profile along the volume. Clinical metrics were calculated from these generated volumes using multiple methods to compare accuracy.

Fig. 5 highlights methods to vary the diameter profile along the developing aortic volume. Diameter profiles for the lumen and combined aneurysm, initially proposed by Finol et al (2002)[71], are dependent on the AAA neck (D_{Neck}) and max diameter (D_{Max}), and height of the aneurysm (Ht). Additional inputs include both thickness of the aortic wall (ΔWall) and maximum thickness of the outer wall structure at D_{Max} (ΔOWS).

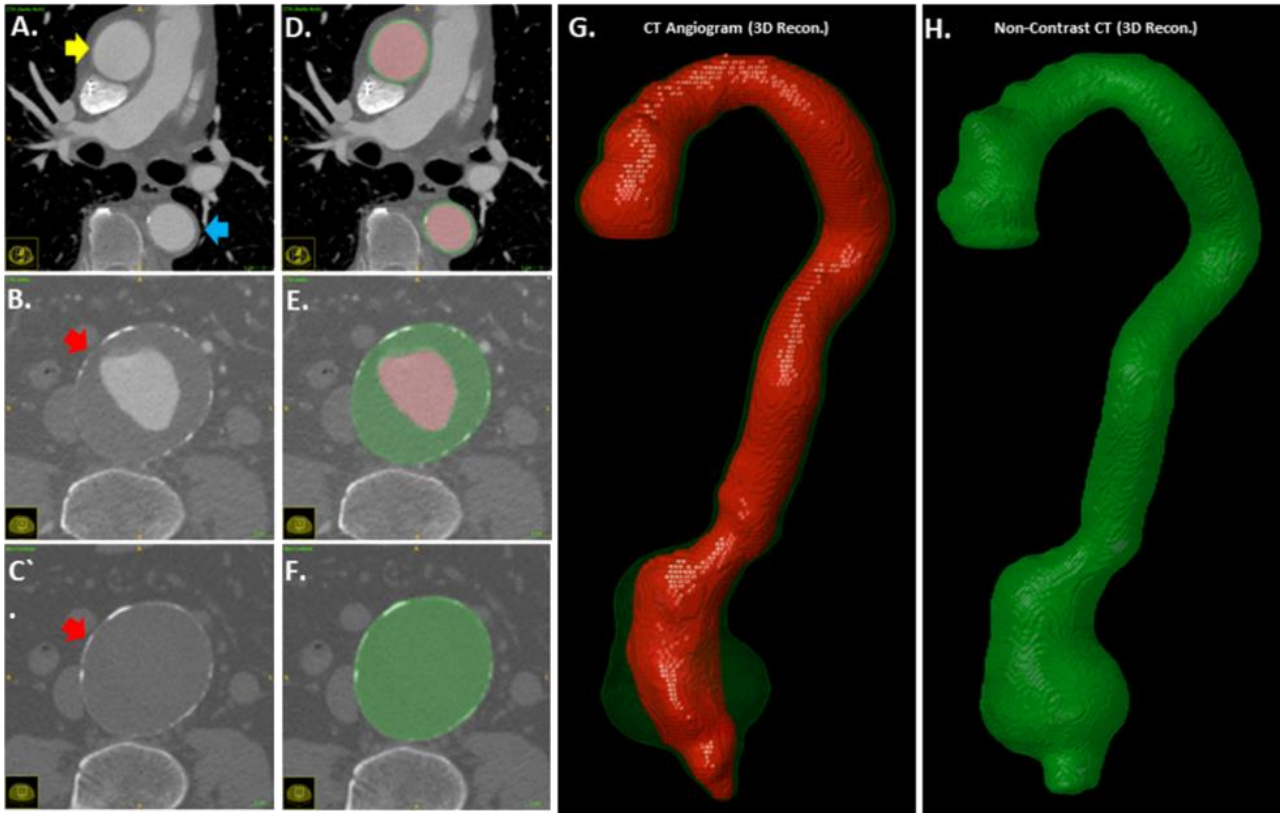


Fig. 4: Manual Segmentation of CT Images (CECT/Non-Contrast). ***A-F.*** Axial Slices obtained from a CECT and NCCT scan with overlaying manually segmented labels. The lumen is illustrated in red and is typically surrounded by the outer wall in green. In the abdominal region, the green label includes the intra-luminal thrombus, if present. **G-H.** 3D-reconstructed volumes representing the aortic lumen (red) and wall structure (green), which contains the intra-luminal thrombus, are generated from the masks[10].

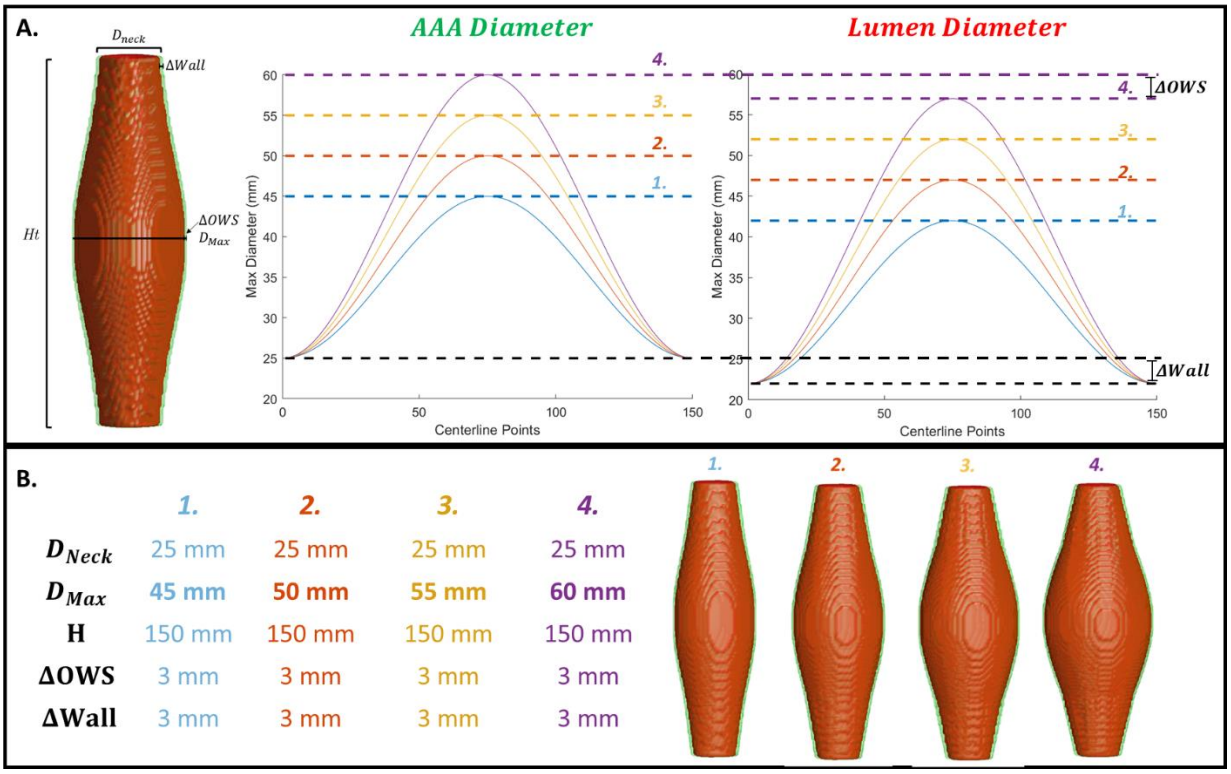


Fig. 5: Diameter Profiles in Simulated Aneurysm Volumes. A. Aneurysm and Luminal diameter profiles were constructed using five input parameters, as indicated on the example aneurysm. Four diameter profiles with increasing D_{Max} are constructed using Equations 1 and 2. ΔOWS is equivalent to ΔWall for all constructed volumes. **B.** Generation of four 3-D AAA volumes was performed by constructing circular planes along the aortic centreline with a diameter, as defined by the diameter profile. Shapes illustrated in red and green indicate the inner lumen and OWS, respectively. The OWS is thicker and is more apparent in simulated AAA shapes in Fig. 7.

The equations for the lumen and AAA diameter profiles are indicated below:

$$\boldsymbol{LumenDiameter(z)} = 2 * \left(\frac{(D_{Max} - \Delta OWS) - (D_{neck} - \Delta Wall)}{4}\right) * \left[1 + \sin\left(\left(\frac{2 * \pi * z}{Ht}\right) - \frac{\pi}{2}\right)\right] + (D_{neck} - \Delta Wall) \qquad Eq.\ 1$$

$$\boldsymbol{AAADiameter(z)} = 2 * \left(\frac{D_{Max} - D_{neck}}{4}\right) * \left[1 + \sin\left(\left(\frac{2 * \pi * z}{Ht}\right) - \frac{\pi}{2}\right)\right] + D_{neck} \qquad Eq.\ 2$$

If ΔOWS is equivalent to the ΔWall, then the OWS of the aneurysm is entirely composed of the aortic outer wall. Four diameter profiles from a combination of parameters (Increasing D_{Max}) are illustrated in **Fig. 5b**. The aneurysmal volumes were generated from the diameter profiles by constructing circular planes with a defined diameter, along the aortic centreline.

The aortic/aneurysmal centreline ($r(z)$) was derived using Eq. 3 and is a function of the amplitude (A) and the height of the aneurysmal sac (Ht). The greater the amplitude, the greater the degree of curvature and the smaller the radius of curvature (inverse of curvature) within the centreline. Radius of curvature was calculated between each set of three points at 1 mm increments along the centreline using Eq 4. In this analysis, the centrelines for the lumen and OWS were identical. Although, this is a rare clinical scenario, this was sufficient as the primary goal for this experiment was to evaluate the data extraction methods used for aortic aneurysms.

$$\boldsymbol{r(z)} = \begin{pmatrix} \boldsymbol{z} \\ \boldsymbol{A\,sin(\frac{\pi * z}{Ht})} \end{pmatrix} \qquad Eq.\ 3$$

$$\boldsymbol{Radius\ of\ Curvature} = \frac{||\boldsymbol{r'(z)}||^{\boldsymbol{3}}}{\left|\left|\boldsymbol{r'(z)\ x\ r''(z)}\right|\right|} \qquad Eq.\ 4$$

Fig. *6*a highlights how the minimum radius of curvature of the aortic centreline varies with increasing amplitude (A). Four volumes with increasing amplitude and consequently decreasing radii of curvature are illustrated in **Fig. *6*b**.

Varying both lumen/AAA diameter profiles and the curvature of the aortic centreline captures a great deal of variability associated with aortic aneurysm morphology as seen in **Fig. *7***. 150 volumes were generated using five D_{Max} (45mm, 50mm, 55mm, 60mm, 65mm), 5 ΔOWS (3 mm, 5 mm, 10mm, 15mm, 20mm) and 6 centrelines with increasing amplitude (0mm, 10mm, 20mm, 30mm, 40mm, 50mm). D_{Neck}, height of the aneurysmal sac and ΔWall was kept consistent at 25 mm, 150 mm and 3 mm.

Maximum diameter in planes orthogonal to the centreline, volume and surface area of each of these aneurysms were derived mathematically and compared against the outputs derived from MATLAB's in-built functions. Correlation coefficient analysis and Bland-Altman plots were constructed to display the overlap between measurements. The utility of the Bland-Altman plot in describing the differences between two sets of measurements is detailed and illustrated on a simulated data set within the appendix (Section 8.1, **Sup Fig. *1***). This is a common method of comparison used in this thesis. Root-Mean-Square error difference was used to assess the similarity between diameter profiles.

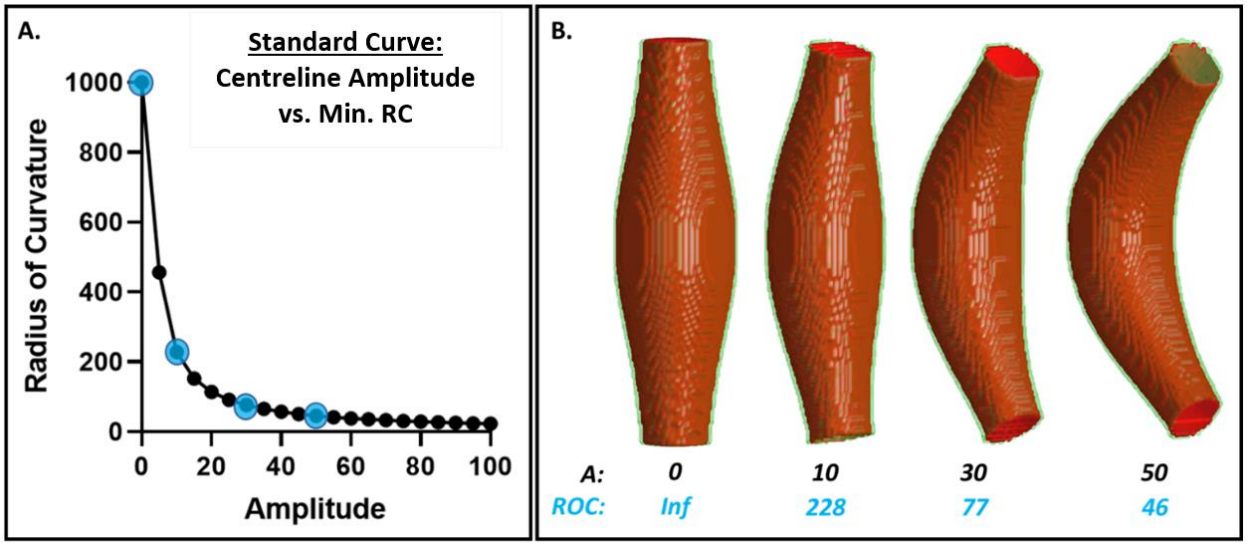


Fig. 6: Varying the curvature of the aortic centreline. **A.** Standard curve illustrating the relationship between Amplitude of the aortic centreline (defined by Eq. 3) and minimum radius of curvature (RC, calculated by Eq. 4). **B.** Generation of 4 Aneurysmal volumes (D_{Max} = 45, D_{Neck} = 25, H = 150 mm, ΔWall = 3 mm, ΔOWS = 3 mm) with centrelines increasing in amplitude (highlighted in blue in Fig. 6A).

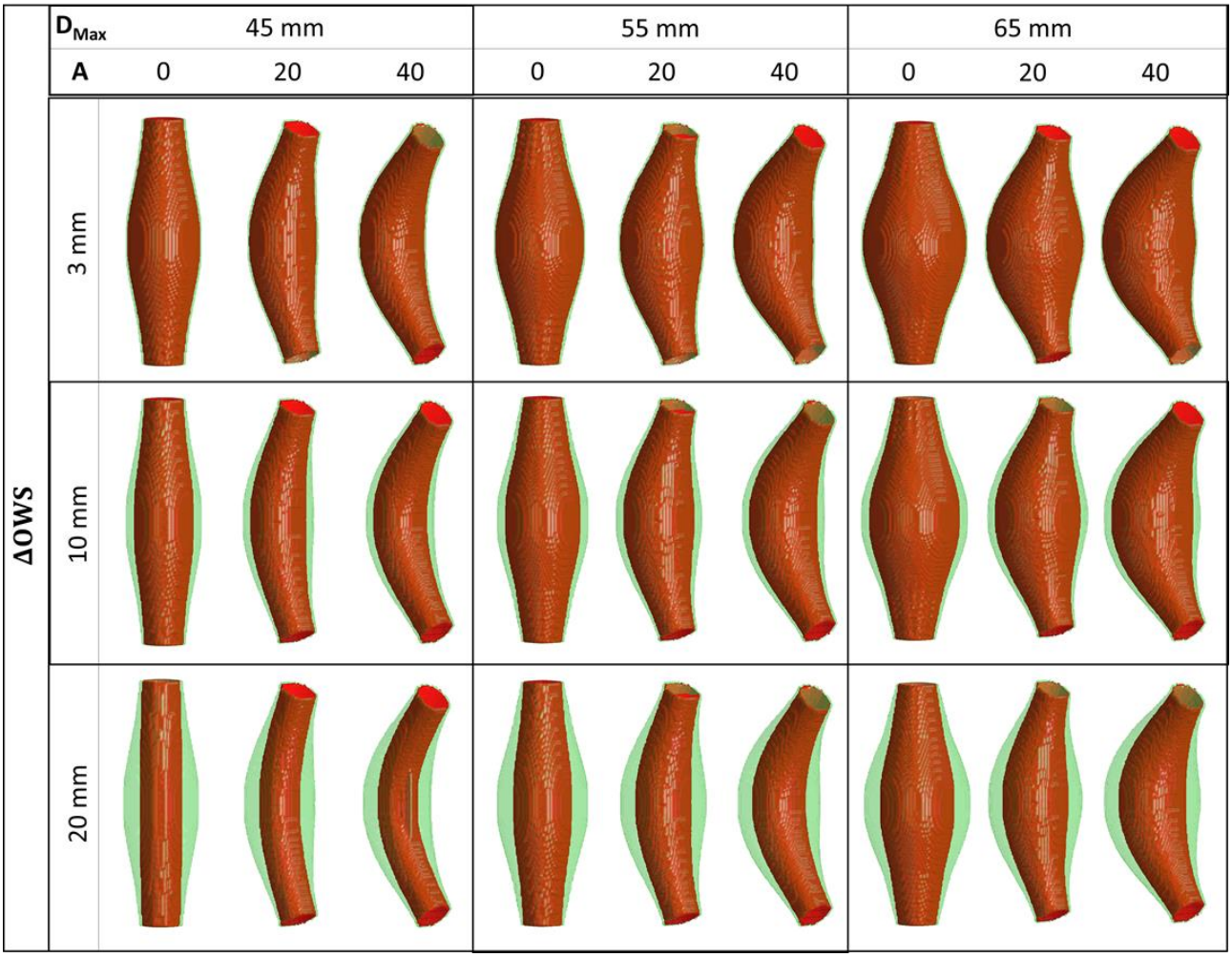


Fig. 7: Mathematically-derived AAA shapes using multiple input parameters. A subset of 27 volumes from the generated 150 shapes are illustrated. This subset was generated using a combination of 3 DMax (45 mm, 55 mm and 65 mm), 3 centreline amplitudes (0 mm, 20 mm, 40 mm) and 3 ΔOWS (3 mm, 10 mm and 20 mm). All other input parameters (Ht, D_{Neck} and ΔWall) were kept consistent. This method provides the ability to generate a diverse range of synthetic AAA volumes.

4.2.4 Assessment of intra- and inter- observer variation of manual segmentation

All patients within OxAAA-IMG-I dataset were used for intra- and inter-observer variability evaluation. This evaluates the validity of the manual segmentations. For the intra-observer assessment, manual segmentation was performed for the second time by AC after a gap of 2 weeks. For the inter-observer assessment, trained clinicians (NS or PL) performed the segmentations independent of the primary observer. The intraclass correlation coefficient (ICC) was calculated for the intra-/inter- observer analysis to assess the consistency of inner lumen and OWS segmentations. Additionally, clinical metrics were calculated from each set of segmentations. Bland-Altman analysis was performed to compare the extent of differences in clinical metrics between the segmentations. A full description regarding the extracted clinical metrics can be found in section 4.2.15.

4.2.5 Data augmentation

To diversify the training set, CT images and their corresponding segmentations were augmented using divergence transformations. These augmentations employ non-linear warping techniques to manipulate the image in predefined locations. Each image within OxAAA-IMG-I dataset was augmented 10:1 to obtain a total of 825 post-augmented scans. **Fig. *8*** illustrates an axial slice augmented 10 times. The impact of these non-linear data augmentations for AAA segmentation was evaluated alongside model selection. During model training, images were further augmented in 3D using random rotation (0-15°), translation and scaling (0.7 - 1.3).

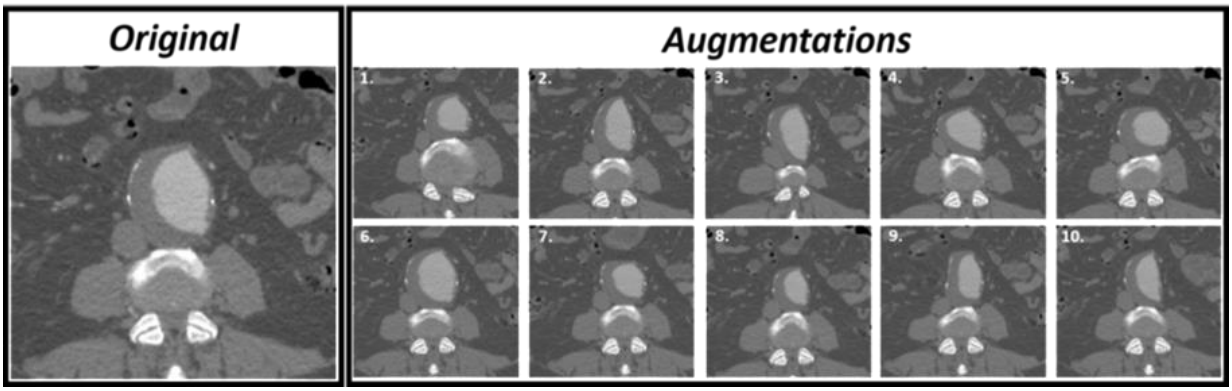


Fig. 8: Data Augmentation of Axial Slice with Aortic Aneurysm. Axial slice (**Original**) is augmented 10:1 using divergence transformations[10].

4.2.6 U-Net Architecture

In this study, a variation of the U-Net was used for the aortic segmentation pipeline (**Fig. 9a**)[68,69]. Its general architecture consists of two components: the contraction and expansion path (**Fig. 9b**). The contraction path (red) extracts information to capture the context of the input at the expense of losing spatial information. This is followed by an expansion path (green), where the size of the image increases to produce a predictive binary mask. The lost image detail is restored using skip connections and is merged via concatenation. This integrates the spatial and contextual information to assemble a more precise prediction of the aortic structure.

4.2.7 Attention Gating to strengthen U-Net Performance

An attention-gated 3D U-Net was evaluated for the segmentation of the aneurysmal aorta. Attention gates utilize information extracted from the coarse scale to filter out irrelevant data exchanged via the skip connections before the concatenation step. The output of each attention gate is the element-wise multiplication of input feature-maps and a learned attention coefficient $[0 - 1]$. Given the goal to simultaneously predict the location of the aortic lumen and wall structure, multi-dimensional attention coefficients were used. These coefficients were determined using additive addition [72], which is more accurate than multiplicative addition [73]. The integration of attention gates for the purpose of pancreatic segmentation has produced superior results when compared to that of prior models[69]. A similar attention mechanism was investigated in this study for aortic segmentation. The performance of this modified U-Net architecture against that of a generic 3D U-Net for segmentation of the aneurysm with and without data augmentation was compared during model selection. Training parameters utilized in this experiment are highlighted in **Table 3.** **Fig. 9b** illustrates the 3D U-Net architecture with attention gates utilized in this study.

4.2.8 Loss Function to evaluate Model Performance

The DICE score was used to quantify model performance at each step. This metric evaluates the similarity between two binary images (A and B) and is defined as follows:

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

Here, this index equals twice the number of elements common to both binary images (2*True Positives) divided by the total number of elements in both images (2* True Positives + False Positives + False Negatives). This similarity quotient ranges between 0 and 1.

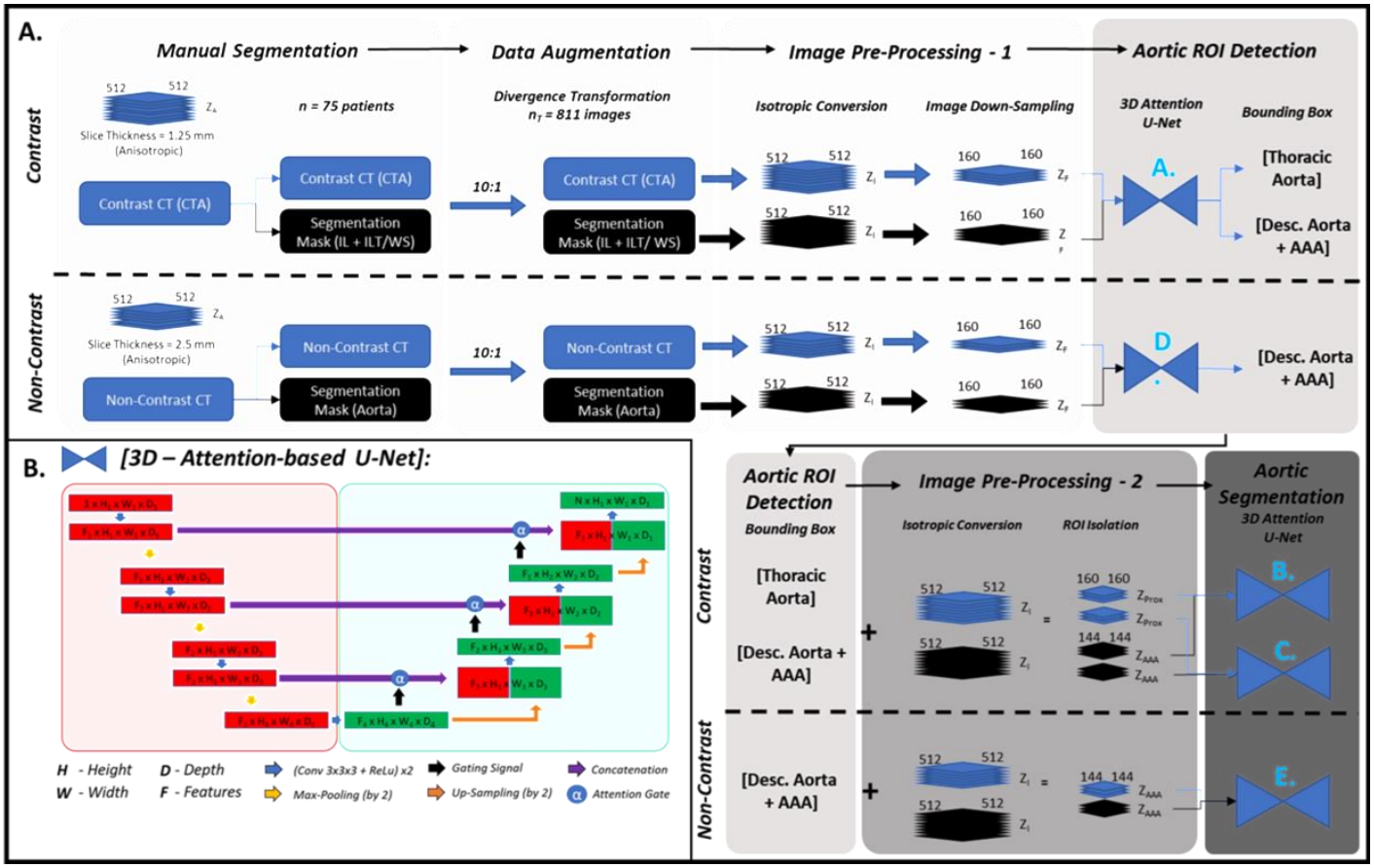


Fig. 9: Aortic segmentation pipeline for the simultaneous detection of the aortic lumen, and outer wall structure. A. Pipeline training required manual segmentation, 2D/3D-data augmentation and pre-processing of both CECT and NCCT images. Aortic ROI detection is coordinated by U-Net A for CECT images and U-Net D for NCCT images. This is followed by aortic segmentation and is coordinated by U-Net B+C for CECT images and U-Net E for NCCT images. **B.** The base architecture for this pipeline is a 3D Attention-based U-Net[10].

4.2.9 Aortic Segmentation Pipeline: Image Pre-Processing

Following data augmentation, all 825 CT images from 75 patients (284,624 CECT axial slices, 145,320 NCCT axial slices, OxAAA-IMG-I dataset) were pre-processed. Pre-processing steps included isotropic voxel conversion and down-sampling by a factor of 3.2 (512 x 512 x $Z_{initial/(i)}$ → 160 x 160 x $Z_{final/(f)}$: $Z_f = Z_i / 3.2$). Here, Z_i and Z_f represents the number of axial slices within the study series before and after pre-processing. The down-sampled images were only used for aortic detection. The higher resolution images were used for aortic segmentation.

4.2.10 Aortic Segmentation Pipeline: Aortic ROI Detection

Attention U-Nets A and D *(Attn U-Net A, D,* refer to **Fig. *9*)** were trained to segment the aorta from these decreased resolution, isotropic CECT and NCCT images, respectively. These architectures were trained using fifty randomly-selected patients (Folds 2+3) and evaluated on the remaining 25 patients (Fold 1 of 3, OxAAA-IMG-I dataset). Aortic bounding boxes were generated from the model predictions. Two bounding boxes were generated from the CECT image (1. Thoracic [Thor.] and 2. Descending/Abdominal Aorta [AAA]) and one was generated from the NCCT image (1. Descending/ Abdominal Aorta [AAA]). The ROIs derived from the bounding boxes served as the input data for aortic segmentation. Z_{Thor} or Z_{AAA} represent the number of axial slices within the thoracic aorta and descending aorta/AAA ROIs, respectively. All subsequent U-Nets (*Attn U-Nets B,C and E*) were trained using the entire OxAAA-IMG-I dataset (75 patients, 825 augmented images). This was done to expose the DL models to the diverse and complex aortic/aneurysmal morphology.

4.2.11 Aortic Segmentation Pipeline: Aortic Segmentation

U-Nets B and C *(Attn U-Net B + C,* refer to **Fig. *9*a)** were trained on the CECT ROIs to simultaneously segment the aortic lumen and OW regions of the thoracic and OWS of abdominal aorta, respectively. U-Net E was trained on the non-contrast ROIs and was tasked to segment the abdominal aorta. For all 3 U-Nets, *3-fold cross-validation* experiments were performed with a data-split of 50:25 patients between training and testing cohorts for each fold (OxAAA-IMG-I dataset). Each fold consisted of 550 post-augmented images from 50 patients for training. The testing cohort consisted of 25 pre-augmented images from 25 patients (testing cohort). There was no overlap between the train/validation and testing cohorts. **Table *3*** delineates all the U-Nets trained and evaluated in this study along with their learning

parameters, which were optimized individually. Model training was performed simultaneously on a workstation with 2 11gb NVIDIA RTX 2080 TI graphics cards.

Table 3: U-Nets trained for model selection and the segmentation pipeline with learning parameters.

Model	Epochs	Learning Rate	Weight Decay	Batch Size	Task	Implementation
Model Selection						
U-Net	1000	$1.0 * 10^{-3}$	$1.0 * 10^{-6}$	2	Multi-Class AAA Segmentation	*Attn U-Net* vs *U-Net* for AAA Segmentation
Attn U-Net	1000	$1.0 * 10^{-3}$	$1.0 * 10^{-6}$	2		
Aortic Segmentation Pipeline						
Attn U-Net A	600	$1.0 * 10^{-3}$	$1.0 * 10^{-6}$	2	Aortic Segmentation from low-resolution isotropic CECT	Aortic ROI Detection (Contrast)
Attn U-Net B	750	$1.0 * 10^{-3}$	$1.0 * 10^{-6}$	2	Multi-Class Aortic Arch Segmentation from high-resolution isotropic CECT	Aortic Segmentation (Contrast)
Attn U-Net C	1000	$1.0 * 10^{-3}$	$1.0 * 10^{-6}$	2	Multi-Class Descending Aorta + AAA Segmentation from high-res. isotropic CECT	Aortic Segmentation (Contrast)
Attn U-Net D	600	$1.0 * 10^{-3}$	$1.0 * 10^{-6}$	2	Aortic Segmentation from low-resolution isotropic NCCT	Aortic ROI Detection (Non-Contrast)
Attn U-Net E	1000	$1.0 * 10^{-3}$	$1.0 * 10^{-6}$	2	Aortic Segmentation from high-resolution isotropic NCCT	Aortic Segmentation (Non-Contrast)

4.2.12 Semi-automatic segmentation of the aortic side branches from CECT images

The output of the aortic segmentation pipeline is a 3D-shape that captures the inner lumen and pathological OWS in patients diagnosed with abdominal aortic aneurysms from the aortic root to the iliac bifurcation. An active contour technique (snakes), which is an iterative region-growing image segmentation algorithm (# of iterations = 50), was implemented in MATLAB to expand the automatically-derived AAA segmentation into its aortic side branches. At this stage, the expanded shape was assessed for quality control and any errors in segmentation were manually adjusted. Branches were then isolated from the main aortic volume and classified based on branch artery using an expert-based classification system. **Fig. *10.*** Illustrates this side-branch segmentation supplement to the main aortic-segmentation pipeline and was developed/optimized using the OxAAA-IMG-I dataset.

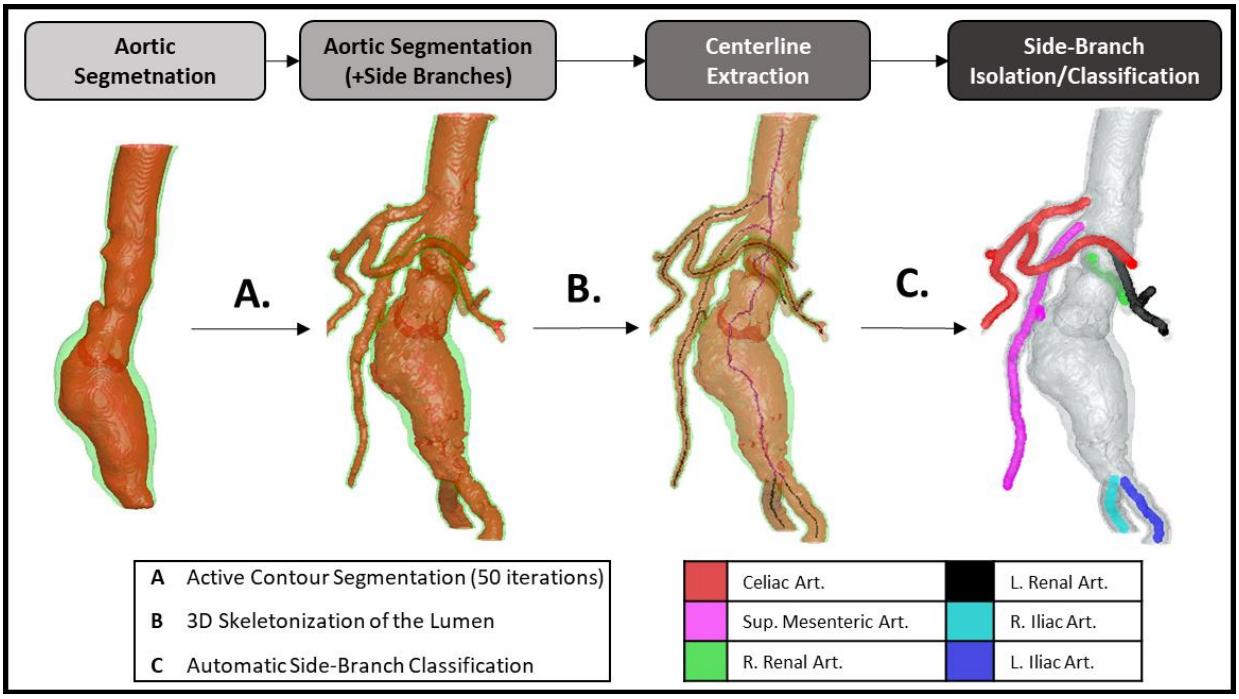


Fig. 10: Semi-automatic segmentation and classification of aortic side branches. This supplement to the aortic segmentation algorithm uses an active-contour segmentation algorithm to capture the origin and direction of the 6 main side-branches that evolve from the descending and abdominal aorta. Branches were automatically classified using the spatial relationship between branch origins (Section 4.2.13).

4.2.13 Automatic Classification of Aortic Side Branches from CECT images

An expert-based classification method was used to classify the isolated branches from the expanded AAA segmentation. This classification method encompassed three sets of decision trees manually generated using expert-based observations. Each decision tree (DT) utilizes an independent classification algorithm to classify the branches first into 3 categories (1. Celiac/SMA, 2. Left/Right Renals, and 3. Left/Right Iliacs). The algorithms and their assumptions for each of the DTs are highlighted in **Fig. *11*, *Fig. 12*,** and ***Fig. 13*.**

Assumptions underlying DT-1 include: 1. all side-branch segmentations have iliac branches, 2. The AAA centroid is between the iliac and other branches and 3. The distance between the Celiac/SMA branches is smaller than the distance between the renal arteries (**Fig. *11***). Assumptions supporting DT-2 include 1. Iliacs are present only if the maximum distance between origin points $>$ 100 mm, 2. First origin point is classified as Celiac/SMA and 3. The angle between the two Celiac/SMA origins and TopCent is $<$ 30° (***Fig. 12***). Finally, the assumptions underlying DT-3 include 1. Iliacs are present only if the maximum distance between origin points $>$ 100 mm, 2. First origin point is classified as Celiac/SMA and 3. The angle between the 2 Celiac/SMA direction vectors are $<$ 45° (***Fig. 13***).

A majority voting system was used to combine the outputs of each DT to obtain the final output. This method was developed/optimized using 50 patients (Folds 2+3 of OxAAA-IMG-I dataset) and evaluated on a validation cohort of 25 patients (Fold 1 of OxAAA-IMG-I dataset). Subsequently, each of the three categories was further divided into their respective arteries using the spatial orientation of the branch origins (1. Celiac Artery is superior to the Superior Mesenteric Artery (SMA). 2. The origin of the left/right renal arteries are closer to the patient's left/right sides, respectively. 3. The origin of the left/right iliac arteries are closer to the patient's left/right sides, respectively.).

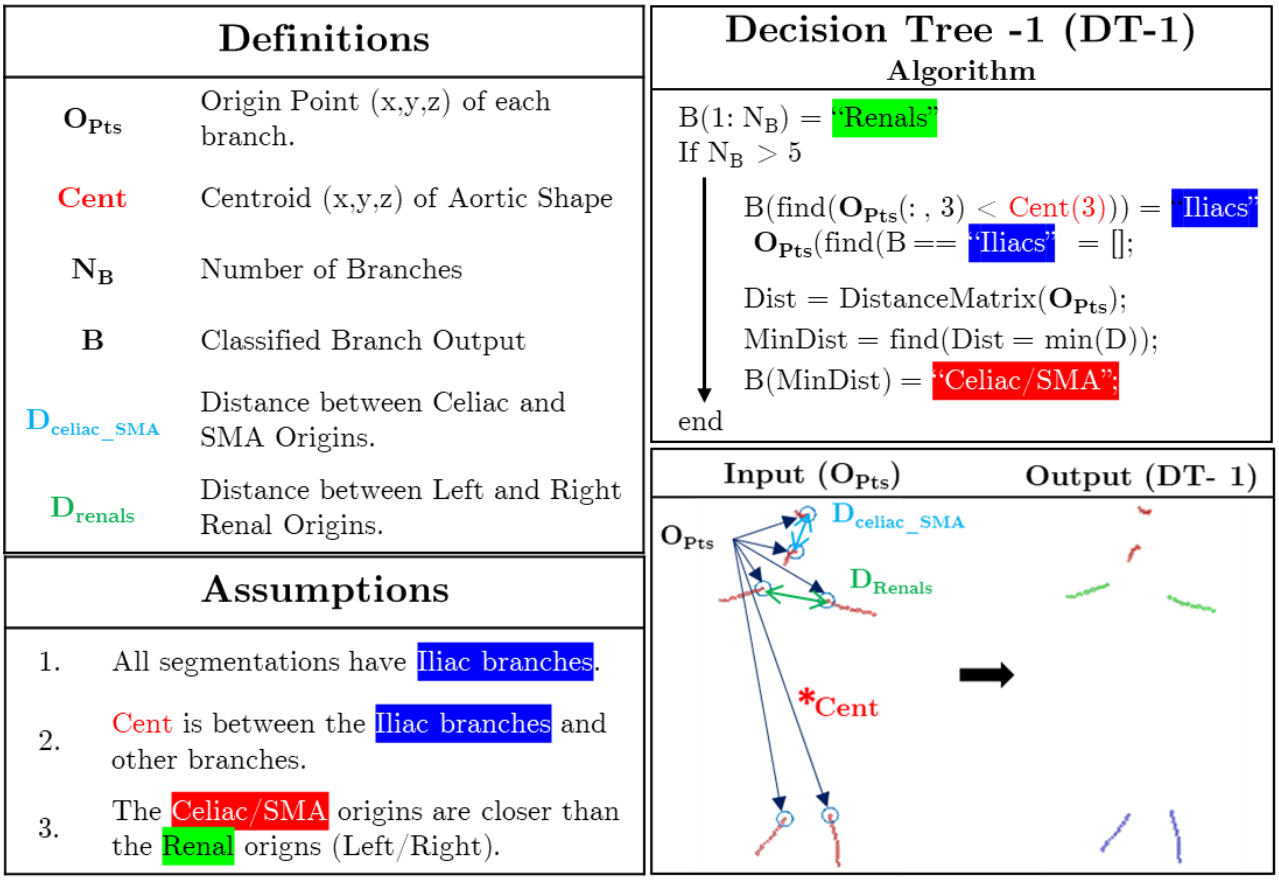


Fig. 11: Side-Branch Classification using manually constructed decision tree algorithm 1 (DT - 1). Decision tree- 1 was constructed to classify the side-branches originating from the abdominal aorta into 3 categories. The algorithm is illustrated on the top right panel. DT-1 compares the distances between origin points. This algorithm assumes that all branches below the aortic centroid are the iliac branches and that the Celiac/SMA origins are closer than that of the Renal arteries. An example patient is illustrated on the bottom right panel. This output is integrated with that from the other DTs (Fig. 12, Fig. 13) and a majority voting method is used to produce the final classifications.

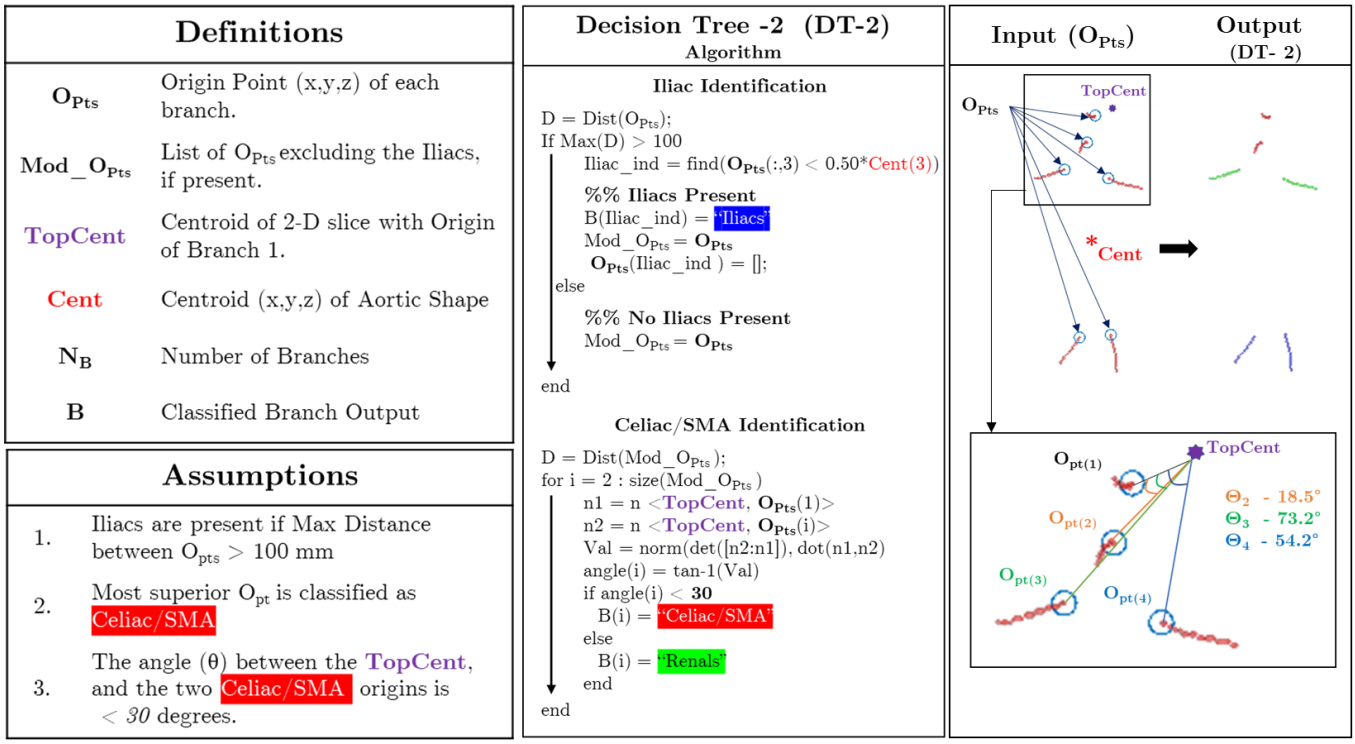


Fig. 12: Side-Branch Classification using manually constructed decision tree algorithm 2 (DT - 2). Decision tree- 2 was constructed to classify the side-branches originating from the abdominal aorta into 3 categories. The algorithm is illustrated on the middle panel. DT-2 utilizes the angle between the origin points and the Top Centroid (TopCent). The angle between the two Celiac/SMA branches and TopCent is < 30 degrees. Additional assumptions for the algorithm are highlighted in the bottom left panel. An example patient with the angle calculations is illustrated on the right panel. This output is integrated with that from the other DTs (Fig. 11, Fig. 13) and a majority voting method is used to produce the final classifications.

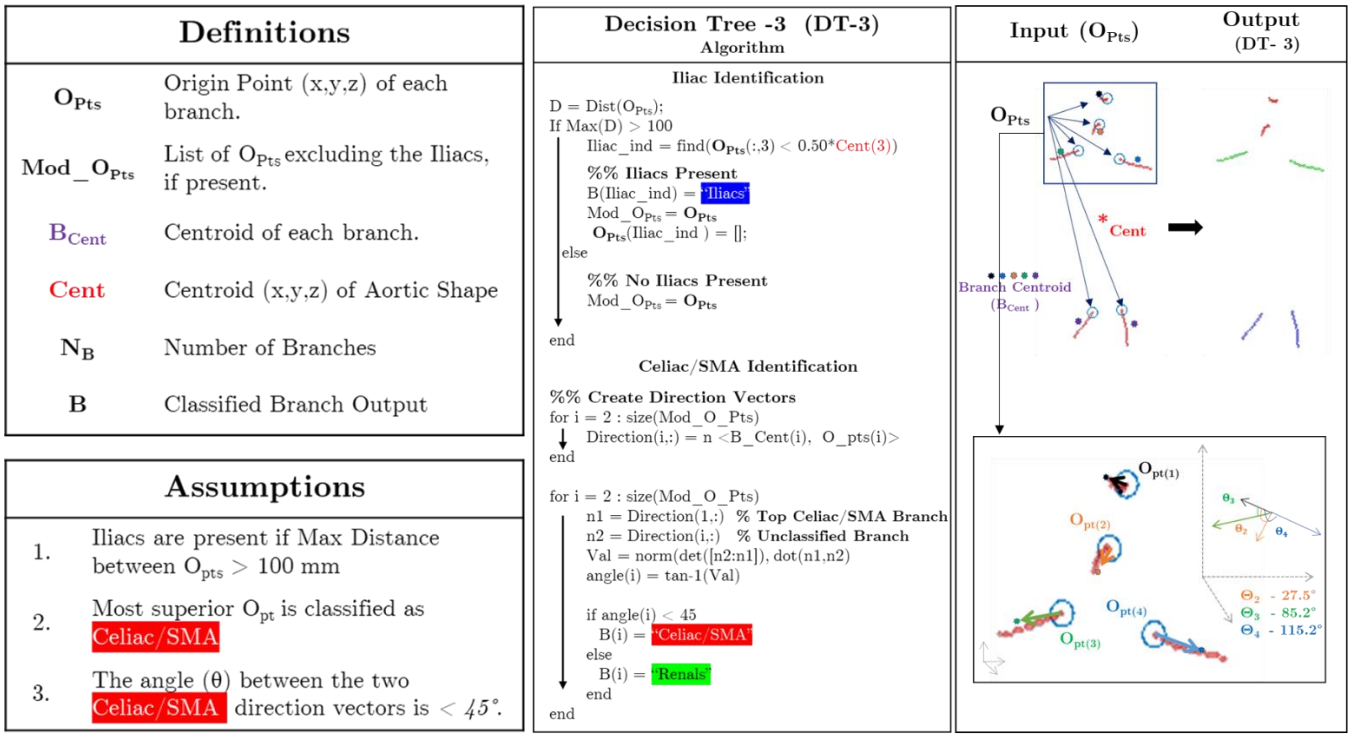


Fig. 13: Side-Branch Classification using manually constructed decision tree algorithm 3 (DT - 2). Decision tree- 3 was constructed to classify the side-branches originating from the abdominal aorta into 3 categories. The algorithm is illustrated on the middle panel. DT-3 utilizes the angles between the direction vectors of each of the branches. The angle between the two Celiac/SMA branches is < 45 degrees. Additional assumptions for the algorithm are highlighted in the bottom left panel. An example patient with the angle calculations is illustrated on the right panel. This output is integrated with that from the other DTs (Fig. 11, Fig. 12) and a majority voting method is used to produce the final classifications.

4.2.14 Assessment of model accuracy using aortic morphological features

An in-house program in MATLAB was developed to automate the measurement extraction from AAA volumes. In addition to validating this method on the synthetic AAA shapes (Section 4.3.2), automatically calculated measurements from GT segmentations were compared against manually extracted measurements from the same CT images. This was accomplished to assess the accuracy and clinical validity of the manually-segmented 3D volumes. From each patient, measurements were obtained both along the axial plane and the plane orthogonal to the aortic centreline. Six measurements were obtained from three slices (1. slice with the max anteroposterior (AP) diameter, 2-3. 1 cm above and below the slice with the max AP diameter). Max antero-posterior and transverse diameters were measured in each of the three slices. Coefficients of variation between the manual delineation and automatic methods are reported.

In addition to the DICE score, 1-, 2- and 3-D measurements of aortic morphology were extracted from the aorta. These were used to evaluate the clinical validity of this high-resolution segmentation pipeline. Maximum AP diameter (1-D) along the axial plane and axial area (2-D) of the aneurysmal region were automatically extracted from each 3-D image. Finally, 3-D measurements included volume assessment of the lumen and OWS from CECT images and of the total aortic volume from NCCT images. All metrics were calculated on model predictions and GT segmentations using an in-house program in MATLAB. Bland Altman plots and correlation coefficient analysis assessed bias and the strength of association between the output of the DL models and the GT. Bias for all measurements was reported along with its 95% confidence interval (95% CI).

Second-order features including lumen and OWS centrelines were calculated using an implementation of the homotopic thinning algorithm[74]. Centreline deviation between model predictions and GT annotations was calculated using 1. Average Euclidean distances and 2. Hausdorff distance. The former calculates the distances between two closest points in the two centrelines. On the other hand, the Hausdorff metric reflects the upper bounds of the former[75]. It is the greatest distance between a point in one centreline and the closest point in the adjacent line. Additionally, maximum diameter in planes perpendicular and orthogonal to the

generated centrelines were extracted and compared between the ground truth measurements and the model predictions. Root-mean-square-error and percentage deviation were used to assess the similarity between the diameters orthogonal to the AAA centreline. This second-order feature assessment ensures the utility of model predictions for complex geometric/morphological analysis.

Finally, to evaluate the accuracy of the expanded aortic segmentation (**Fig. *10***), maximum diameters from each of the aortic side branches was automatically calculated from the expanded annotations and compared against measurements manually-obtained from the CT image.

4.2.15 Evaluation of Aortic Segmentation Pipeline on an independent validation cohort

The optimized aortic segmentation pipeline was internally validated on the OxAAA-IMG-II dataset, which consisted of an independent set of 200 patients with paired NCCT and CECT images (total of 29,468 pairs of 2-D images) of the abdominal aorta. This cohort was previously described in Section 3.1. All images were manually segmented prior to model evaluation. AAA segmentation accuracy was evaluated using both DICE score and the aortic morphological features (Section 4.2.14). From CECT-derived segmentations, model predictions were compared against GT using 1. Max AP Diameter (Axial), 2. Max AP diameter (Orthogonal), 3. Max Area (Axial), 4. Lumen Volume, 5. OWS Volume, and 6. Centreline Deviations (average Euclidean/ Hausdorff distance). Side-branch classification accuracy was investigated and maximum branch diameters were automatically extracted from model predictions and compared against that from the GT. On the other hand, from NCCT-derived segmentations, model predictions were compared against GT using 1. Max AP Diameter (Axial), 2. Max AP diameter (Orthogonal), 3. Max Area (Axial), 4. AAA sac volume and 5. centreline Deviations.

4.3 Results

4.3.1 CT Image Characteristics

Three-fold cross-validation within OxAAA-IMG-I dataset was used during training of the aortic segmentation pipeline. CT image characteristics between the training/validation and testing cohorts were extracted for each fold. Statistical comparison (two-tailed unpaired t-tests)

between the training and testing cohorts for both CECT and NCCT images across all three folds, revealed no significant differences. The CT image characteristics between the training (folds 2+3, 50 patients) and testing (fold 1, 25 patients) are summarised in **Table *4***.

4.3.2 Validation of Data Extraction Methods

A total of 150 volumes were generated using a combination of D_{Max}, ΔOWS and aortic centrelines parameters. D_{Neck}, height of the aneurysmal sac and ΔWall was kept consistent at 25 mm, 150 mm and 3 mm. A subset of these volumes can be visualized in **Fig. *7***. Diameter profiles along the aortic centreline were measured and compared against that used to generate the volumes. Root-Mean-Square-Error difference between the automatically derived and input maximum diameter profiles for the 150 volumes was 0.81 ± 0.04 mm. **Fig. *14*** illustrates the correlation coefficient analysis and Bland-Altman plots for the similarity between the calculated volume and surface area measurements against those derived automatically from the in-built MATLAB algorithm. As seen within the Section 9.1, this is an appropriate method to compare the differences between data sets. Volume and surface area measurements of the lumen and OWS are similar between the two methods (negligible bias and narrow limits of agreement). This supports the validity of the in-built data-extraction method to extract 1-D metrics (Max Diameter along the aortic centreline) and 3-D metrics (Volume, Surface Area).

Table 4: Image characteristics within training (n = 50, Folds 2 + 3) and validation (n = 25, Fold 1) cohorts (OxAAA-IMG-I dataset).

		Training Cohort (n = 50)		Validation Cohort (n = 25)		p-value
Contrast	25th Percentile HU [95% CI]	-1008	[-1003 -1014]	-1006	[-1002 -1010]	0.42
	Mean HU [95% CI]	-587	[-646.1 -527.6]	-568.4	[-610.2 -526.6]	0.48
	75th Percentile [95% CI]	-67	[-45.8 -85.8]	-54.1	[-40.8 -67.4]	0.18
	Standard Deviation [95% CI]	48.1	[475.0 493.2]	490.6	[485.1 495.9]	0.08
	Voxel Length [95% CI]	0.81 mm	[0.76 0.86]	0.83 mm	[0.79 0.87]	0.50
	Voxel Height [95% CI]	0.81 mm	[0.76 0.86]	0.83 mm	[0.79 0.87]	0.50
	Voxel Thickness	1.25 mm		1.25 mm		-
	KiloVoltage Peak (kVP)	120		120		-
	Exposure Time [95% CI]	434	[359.2 508.8]	474.2	[369.3 579.0]	0.50
	X-Ray Tube Current (mA)	265.3	[105.3 425.7]	299.1	[140.2 467.4]	0.32
Non-Contrast	25th Percentile HU [95% CI]	-1009	[-1005 -1013]	-1005	[-1002 -1008]	0.07
	Mean HU [95% CI]	-565	[-524.3 -606.7]	-550.6	[-589.4 -511.9]	0.55
	75th Percentile [95% CI]	-53.4	[-39.9 -66.9]	-46.8	[-62.2 -31.3]	0.49
	Standard Deviation [95% CI]	483.2	[476.0 490.0]	484.8	[479.8 489.9]	0.69
	Voxel Length [95% CI]	0.80 mm	[0.75 0.85]	0.82 mm	[0.78 0.86]	0.58
	Voxel Height [95% CI]	0.80 mm	[0.75 0.85]	0.82 mm	[0.78 0.86]	0.58
	Voxel Thickness	2.5 mm		2.5 mm		-
	KVP	120		120		-
	Exposure Time [95% CI]	457.3	[435.2 467.8]	462.2	[448.3 475.0]	0.63
	X-Ray Tube Current (mA)	355.6	[221.8 489.4]	367.5	[209.7 525.3]	0.49

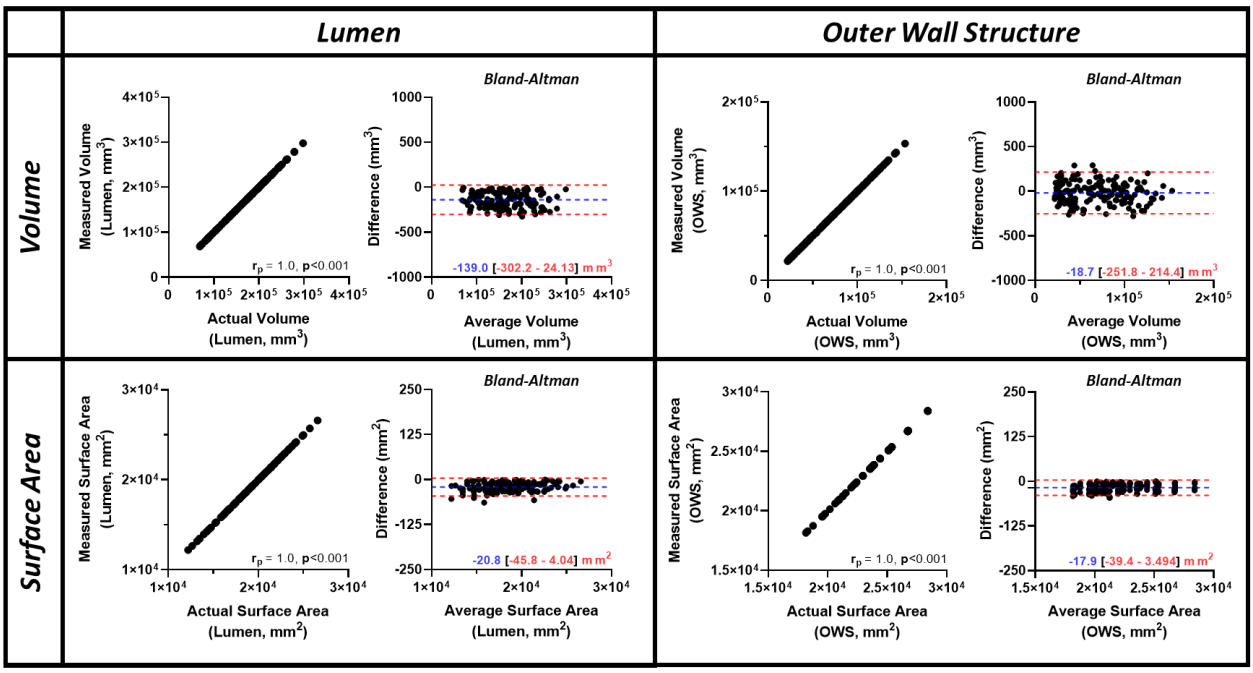


Fig. 14: Validation of Automatic Surface Area and Volume Measurements from Synthetic AAA shapes. Correlation coefficient and Bland-Altman plot analysis between measured metrics (in-built MATLAB functions) versus actual measurements (derived using the mathematical expressions for the aneurysmal surface). The metrics include volume and surface area of the aortic inner lumen and OWS. A very strong correlation was observed between all measurements. Bland-Altman analysis indicated a negligible bias (blue dotted line) with narrow limits of agreement (red dotted line, < ± 0.50 %).

Additional methods validation was performed by comparing measurements derived from the GT segmentation against that derived directly from CT image. Coefficients of variation (%CV) between manual and automatic measurements for the maximum AP and transverse diameters along the axial plane were 0.7 ± 0.05% and 1.1 ± 0.03%, respectively. Additionally, %CV between manual and automatic measurements for the maximum AP and transverse diameters along the plane orthogonal to the aortic centreline were 0.9 ± 0.04% and 1.4 ± 0.1%, respectively. This suggests high concordance between the manual and automatic methods and supports the use of the automatic extraction algorithm for subsequent steps.

4.3.3 *Intra-* and *Inter-* observer variability assessment

There were strong agreements for *inter- and intra-* observer measurements as measured by DICE (±SD) score and intra-class correlation coefficients from CECT and NCCT images (**Table *5***). This supports the accuracy of the manual segmentations used for model training.

Table 5: DICE scores and intra-class correlations for Intra-/Inter- operator segmentations using the images within the OxAAA-IMG-I dataset (*p < 0.001).

DICE ± SD (%) N = 75	**Region**	**Intra-**		**Inter-**	
		DICE ± SD (%)	ICC	DICE ± SD (%)	ICC
Contrast	Inner Lumen	97.8 ± 0.1 %	1.000 *	96.8 ± 0.2 %	0.995 *
	OWS only	94.5 ± 0.4 %	0.981 *	93.1 ± 0.5 %	0.974 *
	Entire Aorta	98.8 ± 0.2 %	0.989 *	95.1 ± 0.5 %	0.981 *
Non-Contrast	Entire Aorta	97.8 ± 0.4%	0.988 *	96.5 ± 0.6 %	0.977 *

In addition to DICE score metric, aneurysm-specific metrics (1st order [1-3D] and 2nd order) were extracted from the intra-/inter- observer derived segmentations. Bland-Altman analysis comparing the GT segmentations with that of the inter-/intra- observer for both the contrast-enhanced (**Fig. *15***) and non-contrast displayed considerable overlap (**Fig. *16*).** The limits of agreement were wider for the comparison between observers versus as opposed to the comparison between intra- observer segmentations. This establishes the threshold for segmentation performance in both non-contrast and contrast-enhanced CTs with regards to the pathological aortic aneurysm.

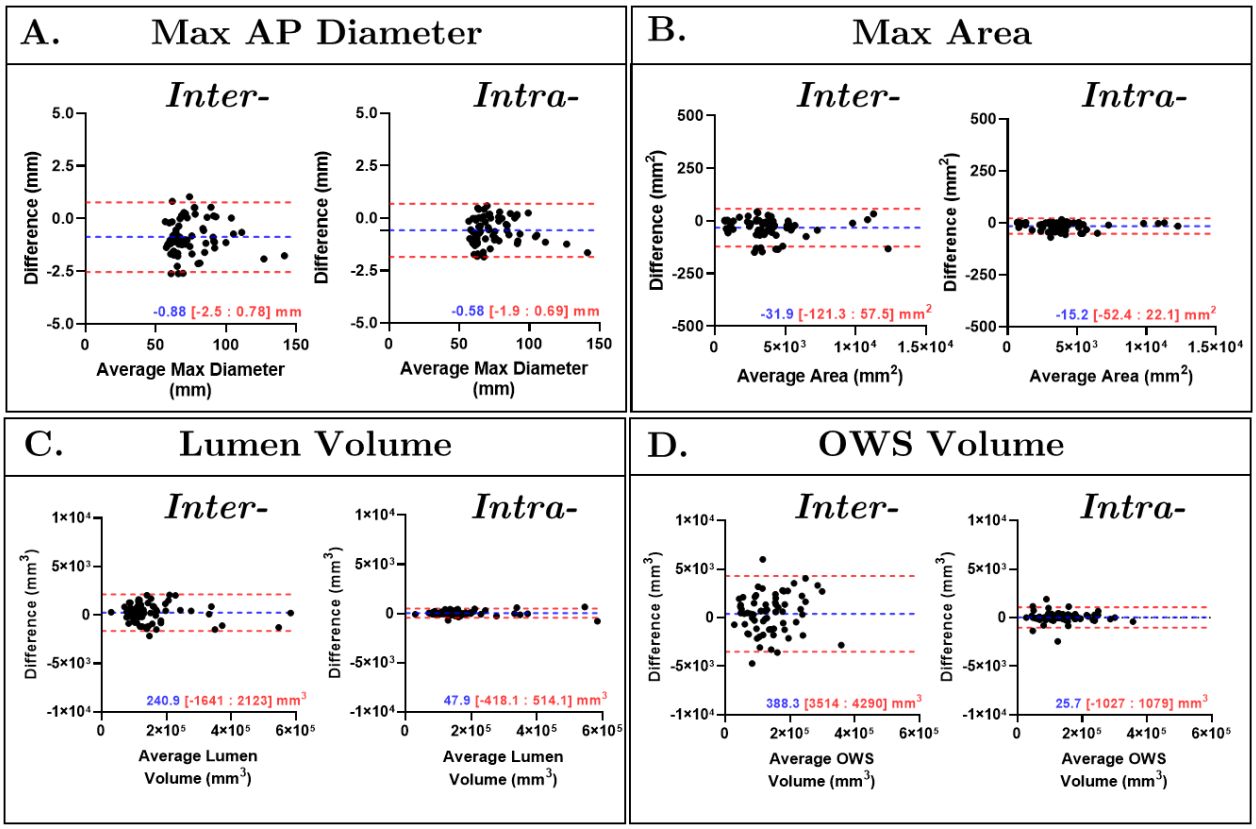


Fig. 15: Clinical assessment of Inter- and Intra- observer-generated segmentations versus the ground truth segmentations from Contrast-enhanced CT. Max AP diameter, max area, and lumen and OWS volume were automatically extracted and compared between the segmentations via Bland Altman Analysis. Bias (blue) and the limits of agreement (red) are illustrated in each panel.

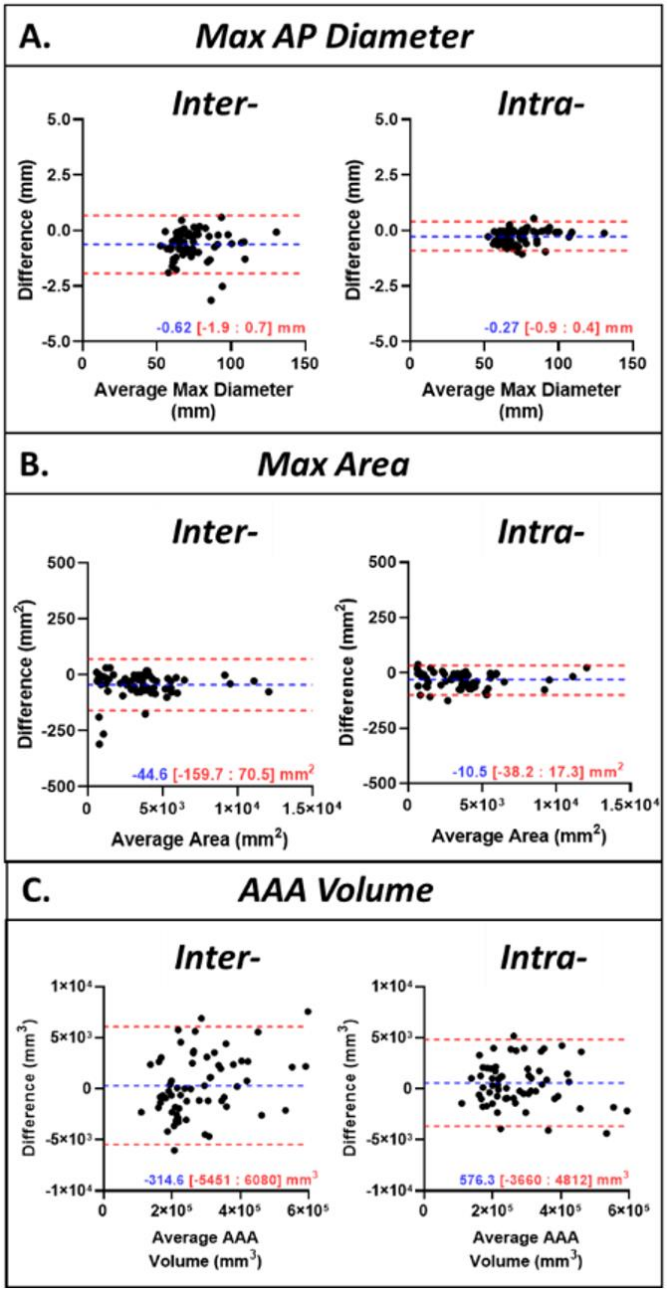


Fig. 16: Clinical assessment of Inter- and Intra- observer-generated segmentations versus the ground truth segmentations from Non-Contrast CT. Max AP diameter, max area, and AAA volume were automatically extracted and compared between the segmentations via Bland Altman Analysis. Bias (blue) and the limits of agreement (red) are illustrated in each panel.

4.3.4 Model Selection

We hypothesized that the 3D U-Net with attention gating (Attn U-Net) would outperform the generic 3D UNet[76] for segmentation of the AAA. Similarly, we hypothesized that the incorporation of non-linear data augmentation (+ Aug) techniques would improve model performance. Fifty patients within OxAAA-IMG-I dataset were used for model training and optimization. The remaining 25 patients were used to test the performance of these networks for aneurysmal segmentation.

Four networks were trained to assess the importance of data augmentation and attention gating for aneurysmal segmentation. **Fig. *17*** illustrates the evolving DICE score metric for the validation group during model training for all four networks. During the training of the Attn U-Net with data augmentation, the overall DICE score plateaus at 95.3% after 1000 epochs (Inner Lumen: 97.4%, OWS: 89.2%). On the other hand, the performance of the control 3D UNet with data augmentation plateaus at approximately 91.8% (Inner Lumen: 96.4%, OWS: 87.2%). The networks trained with the non-linear data-augmentation methods outperform networks without the augmentation.

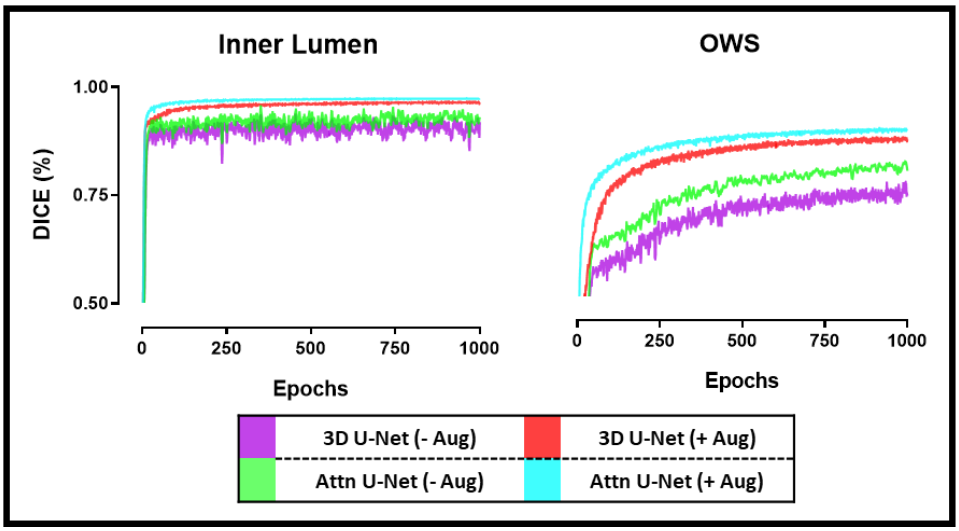


Fig. 17: Attention U-Net vs 3D U-Net for AAA Segmentation. Model performance on validation cohort during training. Four networks (Attention U-Net ± augmentation and generic 3D U-Net ± augmentation) were trained for a total of 1000 epochs. Model outputs were assessed at each iteration and were compared against the GT segmentation using the DICE metric.

Model outputs on the testing sub-cohort (25 patients) were compared against the manually segmented GT images utilizing the DICE score metric. The results of this analysis are found in

Table *6*. Given the apparent importance of data augmentation in model performance, the performance of the Attn U-Net (+ Aug) was evaluated against that of the 3D U-Net (+Aug).

The accuracy of the Attn U-Net (+Aug) in in extracting the OWS of the abdominal region is significantly superior to that of the generic 3D U-Net (+Aug). Additionally, the results show that the output produced by Attn U-Net (+Aug) has a stronger correlation to the GT segmentations for all metrics evaluated than that of the generic 3D U-Net (+Aug) (***Table 6***). Similarly, the Bland-Altman plots for the attention-based U-Net (+Aug) indicate smaller biases with notably smaller bounds (95% confidence interval) when compared against the 3D U-Net (+Aug) outputs (**Fig. *18***). Going forward the Attn U-Net (+Aug) and 3D U-Net (+Aug), will be referred to Attn U-Net and 3D U-Net, respectively.

Table 6: Aortic Segmentation Accuracy of the Attn U-Net (+Aug) vs 3D-U-Net (+Aug).

	Attn UNet vs. Ground Truth	**3D UNet vs. Ground Truth**	
Region	**DICE (± SD)**	**DICE (± SD)**	**p-value**
Inner Lumen	96.8 ± 1.2 %	94.4 ± 1.4 %	0.76
Entire AAA	94.8 ± 0.9 %	89.5 ± 1.1 %	**0.01**
OWS Only	88.2 ± 1.9 %	85.2 ± 1.9 %	$<$ **0.01**
	% Difference (± SD)	**% Difference (± SD)**	
Max AP Diameter	1.1 ± 0.9 %	3.8 ± 2.2 %	$<$ **0.01**
Max Axial Area	2.5 ± 1.6 %	5.0 ± 2.8 %	**0.01**
IL Volume	0.3 ± 0.9 %	1.4 ± 1.8 %	**0.02**
OWS Volume	3.1 ± 2.1%	8.8 ± 8.4 %	**0.01**

Compared to the manual segmentation (ground truth), the difference of AP diameter as measured by the *Attn U-Net* and *3D U-Net* is 0.66 ± 0.56 mm and 2.30 ± 1.36 mm, respectively ($p < 0.01$). The *Attn U-Net* architecture was able to measure the maximum AP diameter to within $<$ 1mm accuracy in 76% (or 19/25) of the cases, as compared to the standard 3D U-Net which can achieve this accuracy margin in 16% (or 4/25) of cases. Similar levels of accuracy were documented when evaluating the inner lumen and OWS volumes. This rationalizes the incorporation of the attention-gating unit into the segmentation pipeline. Example model outputs within the test set are shown in **Fig. *19*** along with their respective gold standards and DICE similarity scores.

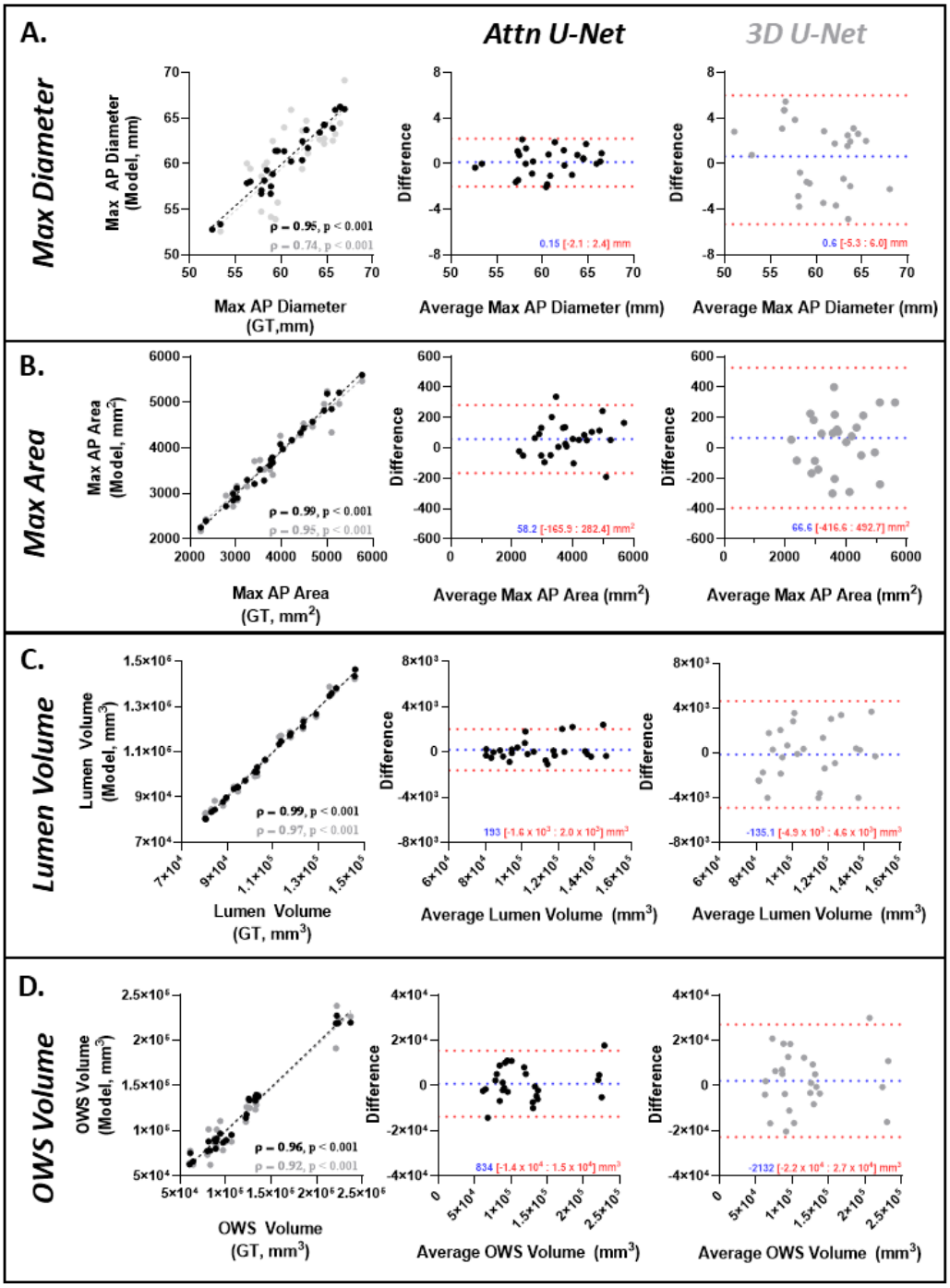


Fig. 18: 1-,2- and 3-D evaluation of model predictions derived from the Attn U-Net (+Aug) and. 3D U-Net (+Aug). Intra-class correlation and Bland-Altman plot analysis measurements obtained from the model predictions and the GT segmentations. Bias [95% CI] within the respective models (Attention U-Net, 3D U-Net) is displayed for each measurement[10].

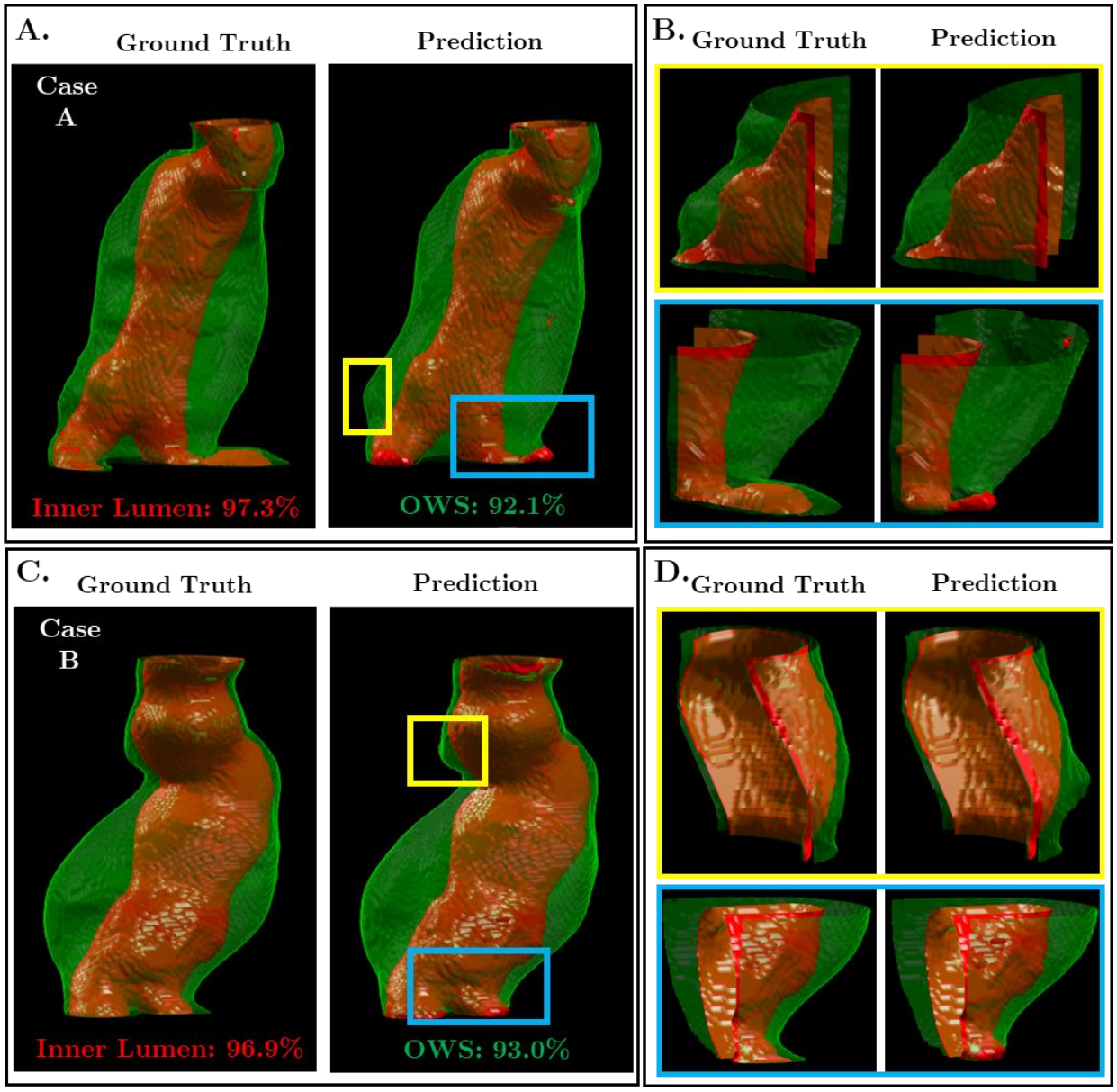


Fig. 19: Attention-based 3D U-Net outputs from two patients with the labelled GT masks (OxAAA-IMG-I dataset). DICE scores for both the inner lumen and OWS (aortic outer wall + ILT) predictions are indicated for each patient **(A,C)**. Various points of discrepancy within the two predictions are highlighted **(B,D)**[10].

4.3.5 Aortic Segmentation pipeline: ROI selection accuracy

The first stage of the segmentation pipeline focused on aortic region-of-interest identification. *Attn-U-Nets A and D* were trained and tested using 50 and 25 patients, respectfully (OxAAA-IMG-I). These networks were tasked to extract the aortic shape from low-resolution isotropic CT images. Both model performances plateaued rapidly after 200 epochs of training. **Fig. *20*** illustrates the evolving DICE score for the validation group during training of these architectures. The segmentation accuracies on the testing cohort for extracting the aortic mask from the CECT and NCCT images were 93.4 ± 1.2% and 88.7 ± 4.2%, respectively. Applying this network allowed for accurate aortic ROI selection on *all images* within the OxAAA-IMG-II dataset.

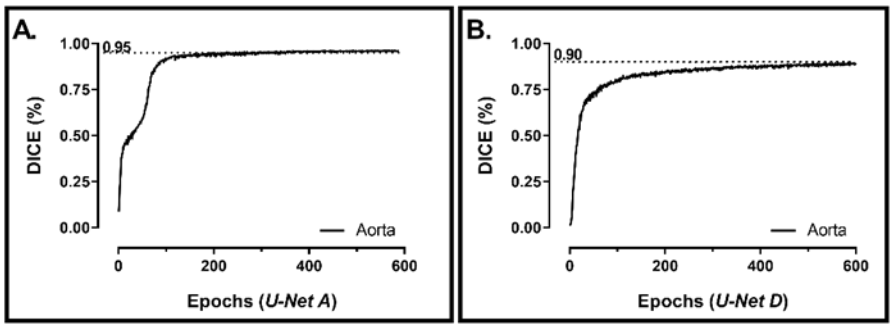


Fig. 20: Training paradigm of the Attention-based U-Nets for Aortic ROI detection from CECT images (A) and Non-Contrast (B). Attn U-Net A was trained for 600 epochs on down-sampled isotropic CT images. A. Attn U-Net D was trained for 600 epochs on down-sampled isotropic non-contrast images[10].

4.3.6 Aortic Segmentation Pipeline: Aortic Segmentation accuracy

Following ROI selection, 3-fold cross-validation was used to train the final segmentation models. Fig. *21a* illustrates the evolving DICE score metric for the validation group during training of these architectures. Consequently, **Fig. *22*a** displays the performance of *Attn U-Nets B and C* on the ability to segment CECT images via the DICE score metric. The inner lumen DICE accuracy is comparable between the thoracic and the abdominal aorta regions. However, OWS DICE accuracy is lower in the thoracic aorta compared to the abdominal aortic region. This is primarily because the thoracic aorta is mostly devoid of ILT and in most cases is a thin circular 'ring' surrounding the lumen. Slight differences in this segmentation result in a relatively larger

proportion of error, as compared to the abdominal region where OWS differences will be proportionally less due to ILT presence.

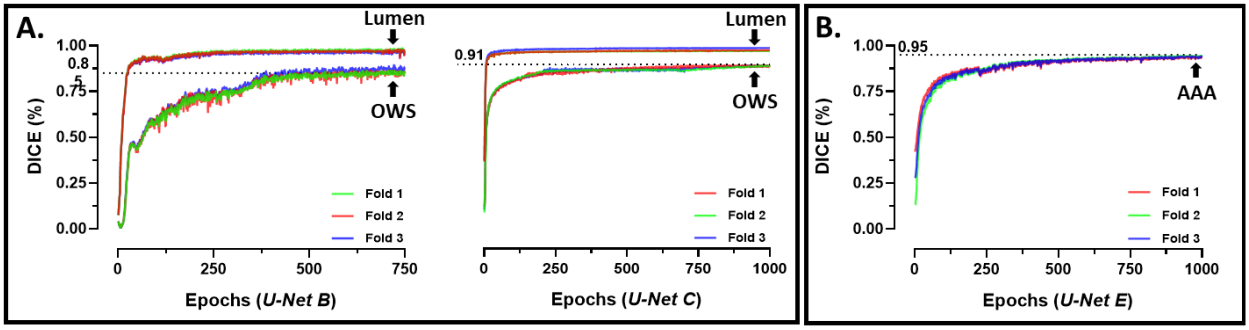
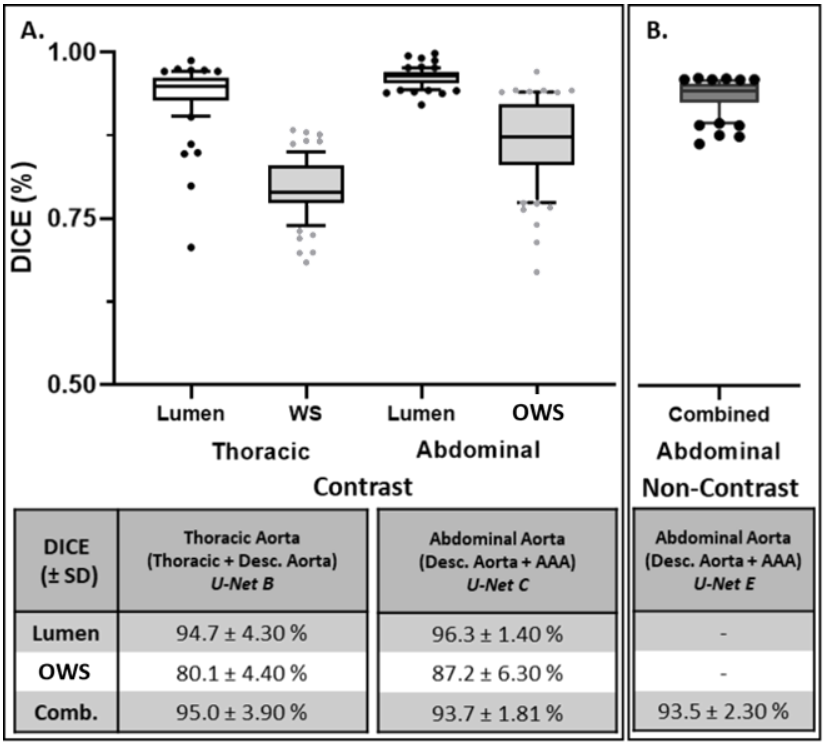


Fig. 21: Training paradigm of the Attention-based U-Nets for Aortic Segmentation from CECT and Non-Contrast images. A. U-Nets B and C were trained on the contrast-enhanced ROIs derived from U-Net A – thoracic Aorta and descending aorta/AAA ROIs to simultaneously segment the lumen and OWS (labelled on each plot) B. U-Net E was trained for 1000 epochs on the non-contrast ROIs derived from U-Net D to segment the aortic aneurysm as a singular label (labelled as AAA). A 3-fold cross-validation approach was used this training step (OxAAA-IMG-I).



DICE (± SD)	Thoracic Aorta (Thoracic + Desc. Aorta) *U-Net B*	Abdominal Aorta (Desc. Aorta + AAA) *U-Net C*	Abdominal Aorta (Desc. Aorta + AAA) *U-Net E*
Lumen	94.7 ± 4.30 %	96.3 ± 1.40 %	-
OWS	80.1 ± 4.40 %	87.2 ± 6.30 %	-
Comb.	95.0 ± 3.90 %	93.7 ± 1.81 %	93.5 ± 2.30 %

Fig. 22: DICE accuracy of model predictions from CECT/NCCT images compared against ground truth (GT) segmentations. A. DICE scores for the CECT-derived segmentations are divided based on the non-overlapping ROIs (Thoracic Aorta [U-Net B] and Abdominal Aorta [U-Net C]). Scores for the lumen, Outer wall structure (OWS, ± ILT) only and the combined aortic region are calculated. **B.** DICE scores for the non-contrast-derived segmentations of the combined aortic region are calculated for the descending aorta/AAA only [U-Net E][10].

For NCCT images, 3-fold cross-validation (OxAAA-IMG-I dataset) was utilized to train *Attn U-Net E* (**Fig. *21*b**). **Fig. *22b*** displays the performance of *Attn U-Net E* to segment NCCT images via the DICE metric. The aortic segmentation pipeline for a patient within the test cohort compared against GT annotations is illustrated in **Fig. *23***.

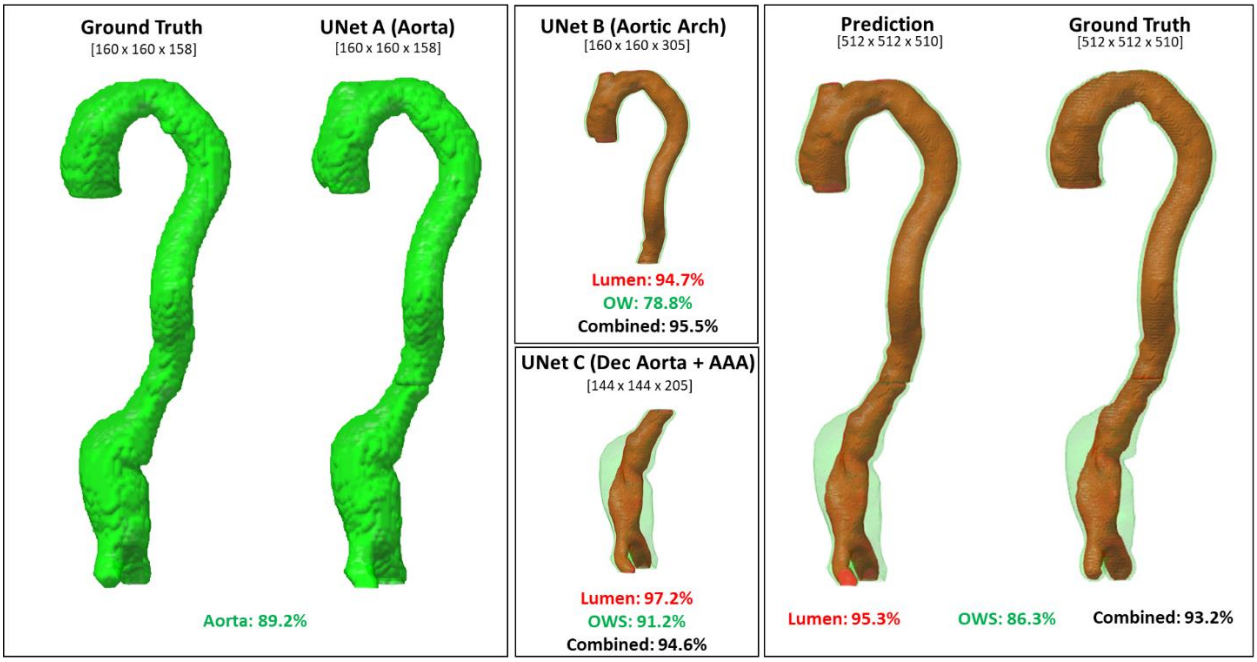


Fig. 23: Aortic segmentation pipeline (Attn UNets A – C) for a patient within the testing cohort. Attn UNet A identified the aortic structure from down-sampled images and was the basis for thoracic and abdominal/AAA aortic ROI detection. Attn UNets B + C identified the lumen and OWS predictions for their indicated region. Region predictions were combined and compared against GT to assess overall accuracy[10].

4.3.7 Aortic Segmentation Pipeline: Side-Branch Segmentation and Classification Accuracy

An iterative region-growing image segmentation algorithm was implemented in MATLAB to expand the AAA segmentation into its side branches. Branch centrelines were isolated and were classified into 3 categories based on branch arteries (1. Celiac/SMA, 2. Renal Arteries, 3. Iliac Arteries) using the developed classification algorithm. These decision trees were developed using the OxAAA-IMG-I dataset (training: 50 patients, validation: 25 patients). **Fig. *24*a** displays the classification accuracy of each of the decision trees on the validation cohort. The classification accuracy combining all three DTs via majority voting was superior to that using each DT alone.

Fig. *24*b-d illustrate three examples within the validation cohort along with the classifications designated by each DT.

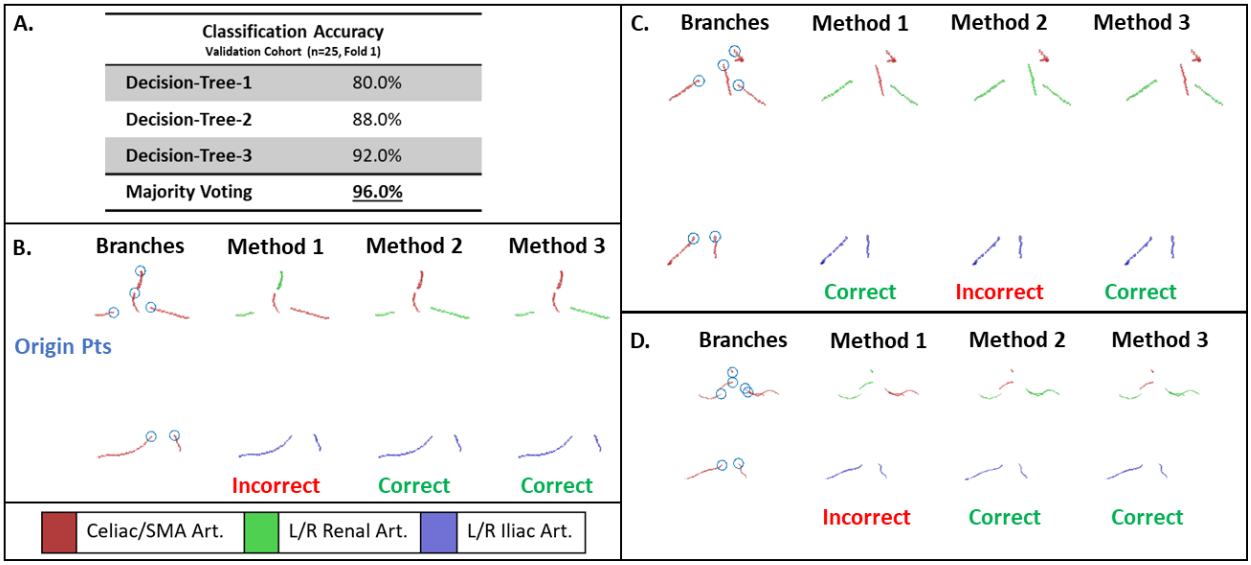


Fig. 24: Classification of Aortic Side Branches based on branch artery. (A.) A multi-decision tree approach was used to classify the branch centrelines in 3 categories. Subsequently, a majority voting method was used to consolidate the predictions. (B-D.) The classification results for three example patients are illustrated.

4.3.8 Assessment of aortic morphology from CECT-derived segmentations

Maximal AP diameter (r_p = 0.99, P < 0.001) and cross-sectional area (r_p = 0.98, P < 0.001) derived from model predictions of CECT images were very strongly correlated with manually derived measurements (**Fig. *25*a-b**). The variability in diameter measurements was < 1.5 mm (1.2 ± 0.80%). Inner lumen (r_P = 0.98, p< 0.001) and OW (r_P = 0.78, p< 0.001) volumes of the thoracic aorta, derived from the output of U-Net B, were strongly correlated with those obtained from the GT annotations (**Fig. *25*c-d**). The variability in the OW volume measurements (12.40 ± 8.10 %) were noticeably greater than for the lumen (3.90 ± 2.64 %) in the thoracic aorta. This is inherently linked with its thin circumferential distribution in the thoracic aorta.

Inner lumen (r_p = 0.99, P < 0.001) and OWS (r_p = 0.97, P < 0.001) volumes from the abdominal aorta, derived from the output of U-Net C, were very strongly correlated with manually-extracted volumes (**Fig. *25*e-f**). In this case, the OWS variability in the abdominal region (5.50 ± 3.01 %) is lower than that in the thoracic aorta (**Table *7***). Model predictions of

the thoracic (*U-Net B*) and the abdominal aorta (*U-Net C*) from four patients in the testing cohort alongside their GT masks are shown in **Fig. *26*a-b**.

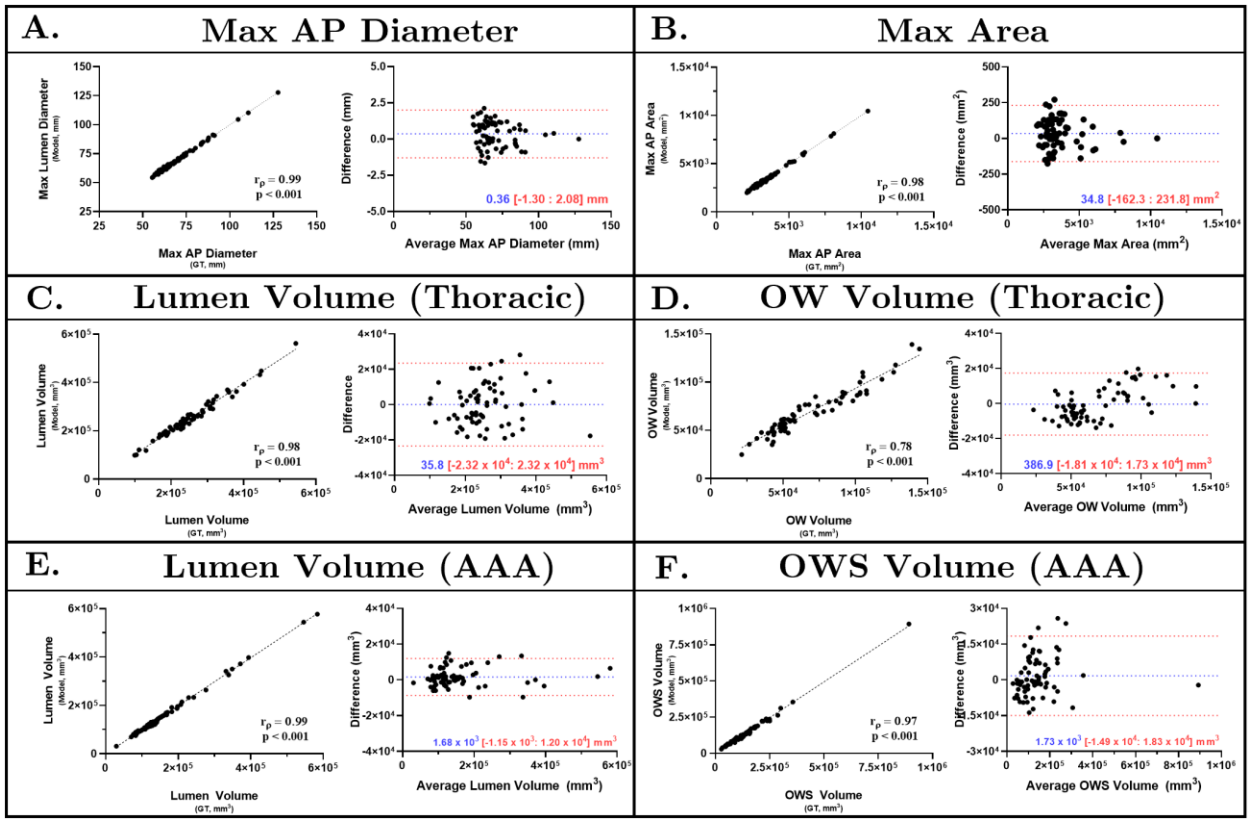


Fig. 25: Correlation-coefficient and Bland-Altman Plot analysis for aortic morphological measurements from CECTs. 1-D (Max AP Diameter of AAA - **A**), 2-D (Max axial area of AAA - **B**), and 3-D (Lumen/OW volume of Thoracic Aorta – **C-D**, and lumen/OWS volumes of the Abdominal Aorta - **E-F**) measurements derived from model predictions were compared against those derived from the GT. This analysis was limited to volumes extracted from CECT images. Spearman correlation coefficients (r_ρ) and p-values are indicated on the graphs.

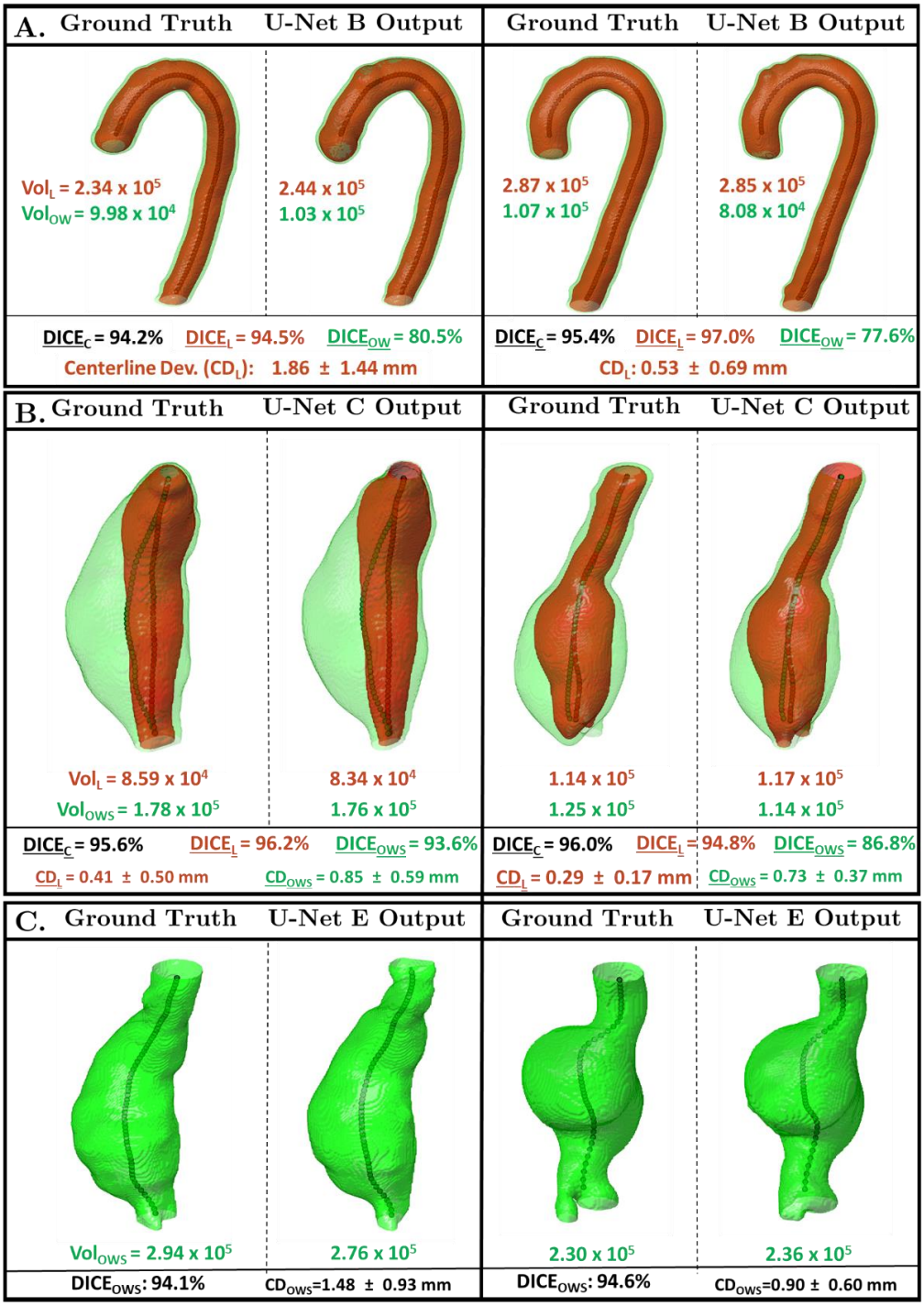


Fig. 26: Model predictions from Contrast (A-B) and Non-Contrast CT (C) Images. Thoracic (**A)** and abdominal (**B)** aortic regions from CECT images and the abdominal region from NCCT images (**C**) are displayed alongside labelled GT masks. Lumen and OW/OWS volumes, when available, are indicated next to each segmentation. DICE scores for the lumen (red), OW/OWS (green) and the combined aortic predictions are indicated for each patient. The difference in

centrelines derived from the lumen masks are indicated as average Euclidean distance deviation (CD_L ± SD)[10].

Furthermore, the similarity in the lumen and OWS centrelines generated from the model predictions and GT annotations is highlighted in **Table *7***. Centreline deviations within the thoracic aorta are greater than those observed within the abdominal aorta. This may be due to the difficulty in delineating the border between the aorta and branching arteries within the thoracic region. Model-derived segmentations of these outlets may affect centreline properties greater than other metrics. However, the average Euclidean distance deviation is less than 2 mm for the thoracic aorta in 89% of cases (67/75) and for the abdominal aorta in 92% of cases (69/75).

Table 7: Clinical assessment of segmented volumes from CECT and NCCT images.

CECT			**% Difference (± SD)**		**% Difference (± SD)**	**NCCT**		**% Difference (± SD)**
Max AP Diameter		***Thoracic Aorta (U-Net B)***	-	***Abdominal Aorta / AAA (U-Net C)***	1.20 ± 0.80 %	**Max AP Diameter**	***Abdominal Aorta / AAA (U-Net E)***	1.67 ± 1.10 %
Max Axial Area			-		2.96 ± 2.45 %	**Max Axial Area**		3.60 ± 3.02 %
Lumen Volume			3.90 ± 2.64 %		2.90 ± 2.60 %	**AAA Volume**		1.67 ± 1.10 %
OW or OWS Volume			12.40 ± 8.10 %		5.50 ± 3.01 %			
			Deviation (Euclidean Distance, ± SD)		**Deviation (Euclidean Distance, ± SD)**			**Deviation (Euclidean Distance, ± SD)**
Lumen	**Centreline Dev.**		1.07 ± 0.67 mm		0.85 ± 0.52 mm	**Centreline Deviation**		1.94 ± 1.00 mm
	Hausdorff Dist.		2.60 ± 2.02 mm		2.54 ± 1.98 mm			
OW or OWS	**Centreline Dev.**		1.64 ± 0.80 mm		1.04 ± 0.57 mm	**Hausdorff Distance**		3.58 ± 2.08 mm
	Hausdorff Dist.		2.84 ± 1.59 mm		2.73 ± 1.73 mm			

Note: OW is the aortic outer wall within the thoracic aorta. OWS is the combination of the aortic outer wall and ILT, when present, within the pathological abdominal aorta.

The generated centrelines allow for the automatic calculation of max AP diameter along planes orthogonal to the aortic centreline. The resulting diameter profiles between ground truth and predictions were compared and showed a RMSE of 1.45 ± 1.65 mm, which is equivalent to a percentage difference of 2.3 ± 1.1 %. **Fig. *27*a** illustrates two examples of AAAs with planes orthogonal to centreline. These planes were used to generate the re-aligned or straightened view of the AAA. Corresponding DICE scores and average Euclidean distance deviations between centrelines are indicated. Maximum diameter profiles along the aneurysmal sac are illustrated for both ground truth and model predictions.

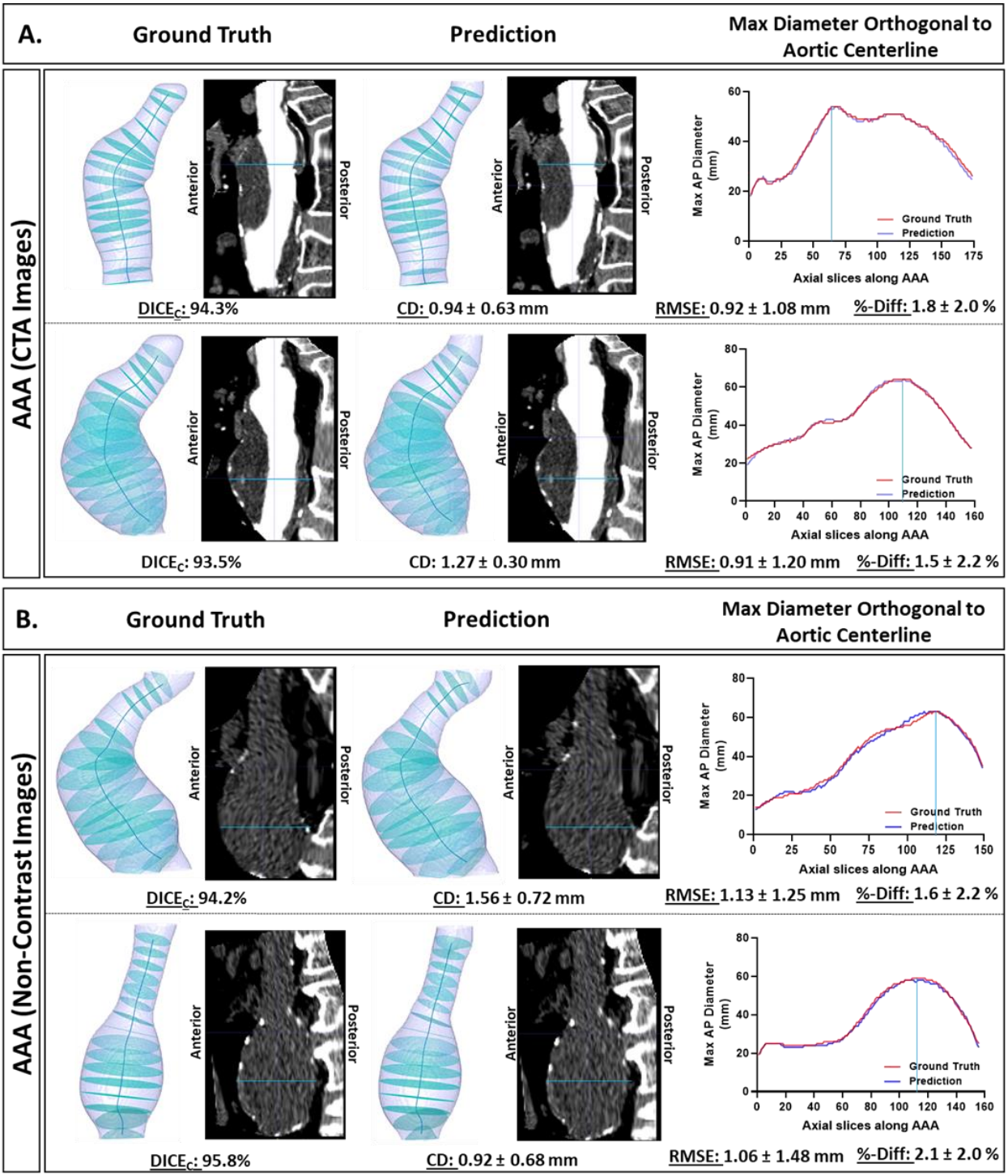


Fig. 27: Maximum AAA Diameter profiles along planes orthogonal to AAA centerline. Profiles were generated from ground truth and model predictions of AAA shapes derived from contrast-enhanced (**A**) and non-contrast (**B**) CT images. The perpendicular planes were used to generate the re-aligned or straightened views of AAA. Corresponding DICE scores, average Euclidean distance deviations between centerlines, RMSE and %-difference of diameter profiles are indicated[10].

For all size aortic side branches, maximum diameters from the derived segmentations were similar to the manually-obtained measurements. This is highlighted in **Table *8*.** These results support the clinical strength of this segmentation pipeline from CECT images to evaluate not only the aortic/AAA shape but also its side branches.

Table 8: Evaluating the maximum diameters of the aortic side branches from segmentations against manually-derived measurements.

Max Branch Diameter	Celiac Artery	Superior Mesenteric Artery	Left Renal Artery	Right Renal Artery	Left Iliac Artery	Right Iliac Artery
Mean (Automatic)	8.4 ± 3.0	8.6 ± 3.2	6.2 ± 3.2	6.7 ± 2.9	15.2 ± 2.8	14.9 ± 2.6
Mean (Manual)	8.5 ± 3.1	8.3 ± 3.0	6.6 ± 3.7	7.0 ± 2.5	15.0 ± 3.1	14.8 ± 2.4
MAE	1.5 ± 0.3 mm	1.2 ± 1.0 mm	1.6 ± 0.9 mm	1.5 ± 1.0 mm	1.4 ± 1.0 mm	1.2 ± 1.1 mm
p	0.94	0.35	0.14	0.19	0.77	0.64

4.3.9 Assessment of aortic morphology from NCCT-derived segmentations

Maximal AP diameter ($r_p = 0.99$, $P < 0.001$), cross-sectional area ($r_p = 0.99$, $P < 0.001$) and volume ($r_p = 0.99$, $P < 0.001$) measurements extracted from the model predictions of NCCT images are very strongly correlated with those derived from the GT segmentations (**Fig. *28*).** The variability in extracting these measurements from NCCT-derived segmentations is like that of CECT-derived annotations (**Table *7***). Model predictions from two patients within the testing cohort are illustrated in **Fig. *26*c.** The resulting diameter profiles between ground truth and predictions were compared and showed a RMSE of 2.11 ± 1.32 mm, which is equivalent to a percentage difference of 2.8 ± 2.1 %. **Fig. *27*b** illustrates two examples of AAAs with planes orthogonal to centreline. This study shows for the first time the ability to segment the aneurysmal aorta from NCCT images at a level comparable to a human observer.

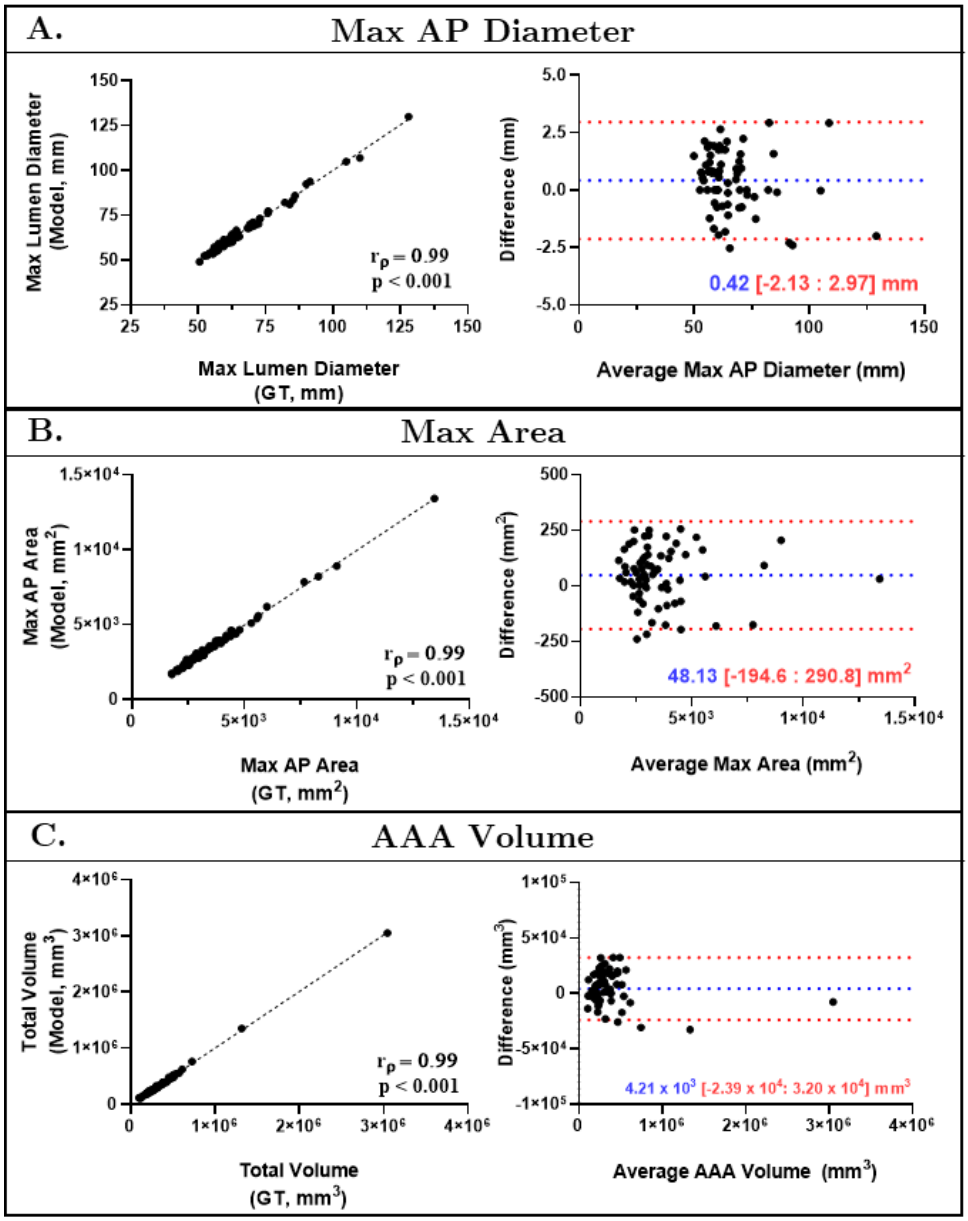


Fig. 28: Correlation coefficient and Bland-Altman plot analysis for aortic morphological measurements from NCCTs. 1-D (Max AP Diameter of AAA - **A**), 2-D (Max axial area of AAA - **B**), and 3-D (Total volume of AAA - **C**) measurements from model predictions were compared against those derived from GT. This analysis was limited to segmentations extracted from NCCT images. Spearman correlation coefficients (r_ρ) and p-values are indicated on the graphs[10].

4.3.10 Assessment of Aneurysm morphology on an independent validation cohort (OxAAA-IMG-II Cohort)

The aortic segmentation pipeline was internally validated on the OxAAA-IMG-II dataset, which consisted of an independent set of 200 patients with paired NCCT and CECT images (total of 29,468 pairs of 2-D images) of the descending and abdominal aorta. AAA segmentation accuracy from both NCCT and CECT images is highlighted in **Table *9*** and was similar to that previously observed. DICE score accuracy of the OWS volume of the aortic aneurysm was slightly higher than that observed with the OxAAA-IMG-I dataset. The % differences between GT and model prediction measurements for both the CECT and NCCT images were similar, supporting model robustness (**Table *9***). Bland Altman analysis for both CECT and NCCT measurements display similar bias and limits of agreement to those obtained with the previously obtained (OxAAA-IMG-I dataset, **Fig. *29***).

Table 9: AAA Segmentation Accuracy on the OxAAA-IMG-II dataset (n = 200).

CECT		DICE (%)	NCCT	DICE (%)
Abdominal Aorta / AAA	Inner Lumen	97.5 ± 0.6%	Abdominal Aorta / AAA	Inner Lumen -
	OWS	92.0 ± 4.2%		OWS -
	Combined	96.3 ± 0.7%		Combined 95.7 ± 0.7%
		% Difference (± SD)		% Difference (± SD)
	Max AP Diameter	1.1 ± 0.8 %		Max AP Diameter 1.5 ± 1.4 %
	Max Axial Area	3.0 ± 3.0 %		Max Axial Area 3.6 ± 4.4 %
	Lumen Volume	1.1 ± 0.9 %		AAA Volume 2.3 ± 1.6 %
	OWS Volume	3.4 ± 3.1 %		
		Deviation (Euclidean Distance, ± SD)		Deviation (Euclidean Distance, ± SD)
	Lumen Centreline Dev.	0.7 ± 0.7 mm		Centreline Deviation 1.8 ± 1.1 mm
	Lumen Hausdorff Dist.	2.2 ± 2.1 mm		
	WS Centreline Dev.	1.1 ± 0.66 mm		Hausdorff Distance 3.7 ± 2.1 mm
	WS Hausdorff Dist.	2.8 ± 1.2 mm		

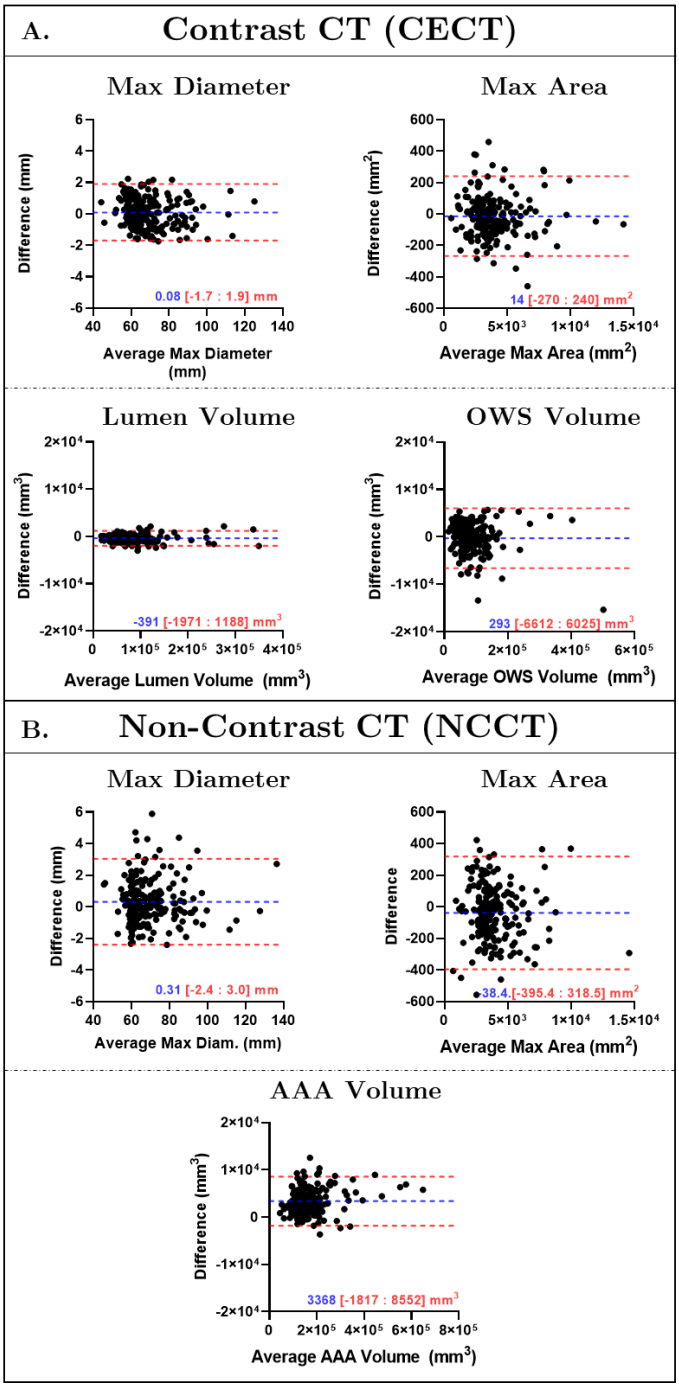


Fig. 29: Bland-Altman Plots evaluating the accuracy of the segmentation model in extracting the aneurysmal/AAA shape from the validation cohort (OxAAA-IMG-II). These images comprise the independent validation cohort and were used to validate model performance. A. From CECT images, maximum axial diameter, maximum axial area, and lumen/OWS volumes were extracted automatically from model predictions and compared against measurements obtained from the GT segmentations. B. From NCCT images, maximum axial diameter, maximum axial area and AAA volume were extracted from model predictions and compared against that obtained from GT segmentations. Bias and Limits of Agreement are indicated in each comparison.

Following semi-automatic segmentation of the aortic side branches, branch classification accuracy was assessed using the expert-based classification algorithm on the CECT-derived aneurysm segmentations. Decision Trees 1 to 3 classified the branches with an accuracy of 83.5%, 93,9% and 96.9%, respectively. Majority voting of the DTs output increased classification accuracy to 98.4%. Maximum branch diameters were extracted from the classified segmentations and were compared against manually-obtained measurements. Mean absolute error was calculated between the diameter measurements and no statistically significant differences were observed for all six aortic side branches (**Table *10***).

Table 10: Maximum Diameter of the aortic side branches from the CECT-derived segmentations on the OxAAA-IMG-II dataset.

Max Branch Diameter	Celiac Artery	Superior Mesenteric Artery	Left Renal Artery	Right Renal Artery	Left Iliac Artery	Right Iliac Artery
Mean (Automatic)	8.2 ± 2.1	8.5 ± 2.3	6.8 ± 2.5	7.0 ± 2.2	14.7 ± 3.0	14.7 ± 2.9
Mean (Manual)	8.4 ± 1.4	8.2 ± 1.3	6.7 ± 1.3	6,9 ± 1.4	14.4 ± 2.3	14.6 ± 2.2
MAE	1.9 ± 0.8 mm	1.4 ± 0.9 mm	1.7 ± 1.0 mm	1.3 ± 0.9 mm	1.6 ± 1.0 mm	1.6 ± 1.0 mm
p	0.55	0.14	0.88	0.28	0.54	

4.4 Discussion

This study describes a fully automatic and high-resolution algorithm able to extract the aortic shape from both CECT and NCCT images at a level superior to that of other currently published methods [77,78]. High accuracy of our segmentation pipeline was supported by the DICE score metric between model predictions and ground truth annotations for both the thoracic and abdominal aorta. However, this metric, which evaluates the similarity between two binary images by evaluating the extent of pixel overlap, has its limitations. In cases of small volumes, minimal changes lead to lower DICE score percentages. This is especially true in the OW region of the thoracic aorta. This region is usually a thin, circumferential ring around the aortic inner lumen. Small variations within this small volume lead to relatively greater variability and lower DICE scores. Concurrently, if there are small but critical errors in a relatively large volume, the DICE score would remain elevated, but the clinical utility of the image would be diminished.

However, this is a common way to compare the performance of segmentation algorithms across methods.

In order to address the limitations of DICE score metric, clinical utility was demonstrated by comparing 1st and 2nd-order measurements, which are important parameters for characterising AAA progression. Extracting max AP diameter measurements enables the calculation of growth during surveillance and determines timing of surgery[79,80]. Assessing the accuracy of diameter extraction is essential to integrate this DL platform with current methods of AAA management. Cross-sectional area (2-D) has been shown to have the lowest variability in assessing aneurysm change and supplements the 1-D diameter measurements [81]. Evolution of 3-D, especially thrombus volume, and 2nd order indices, including centreline are linked to AAA progression, rupture risk and the incidence of adverse cardiovascular events[82,83]. This automatic method of volume extraction can be used to standardize current methods of aneurysmal disease management and sets the foundation for subsequent complex geometric analysis. Furthermore, the proposed pipeline can be extended to other vascular pathologies.

In addition to the AAA volume from CECT images, this pipeline allows for the segmentation and accurate extraction of the aortic side branches. Understanding the relationship of the AAA volume to the location and evolution of the aortic side branches is an important consideration when preparing and planning for AAA intervention[84,85]. Preoperative planning for endovascular AAA repair requires multiple measurements that involve the side branches including 1. the distance between lowest renal artery to aortic bifurcation, 2. diameters of the right/left common iliac arteries, and 3. diameters of the right/left external iliac arteries[84]. Although this pipeline investigates the origin of each of the side branches, its sets the foundation for more thorough investigations. A few studies have investigated the application of 3D-printing for pre-operative planning for complex AAA anatomies, especially with tortuous neck regions, due to the current limitations in AAA visualization[85]. This pipeline would generate segmentations of not only the aorta but also its side branches that would allow for comprehensive visualization and can be used to computational fluid dynamics modelling to understand the intrinsic fluid perturbations within the aneurysmal sac.

Prior to the advent of machine-learning approaches, AAA segmentations were performed using intensity-based semi-automatic algorithms (e.g. level-sets, active shape models and graph cut methods) [77,86-88]. Their primary drawback was the failure to accurately detect the aneurysm's outer boundary as its intensity is like that of adjacent structures. Although these models may provide good results, there are significant limitations that prevent clinical implementation. These methods are semi-automatic and require significant model optimization. Furthermore, these methods require complex user-input (ex. prior lumen segmentations/centrelines) and are highly data-set dependent [77,86]. The latter limits model robustness and generalizability.

Recently, DL methods on CECTs have been proposed to tackle this problem without encountering many of the limitations of their predecessors. Variations on Deep Belief and U-net based networks have been used to segment the infra-renal region of the aorta [78,89]. Unfortunately, many of these networks are limited to 2-D inputs (axial CT slices), which may fail to appropriately capture the aneurysm's 3D geometry. The accuracy and reproducibility of these models is like that of earlier methods as they are trained and validated on small data sets. Lopez-Linares et al. recently proposed a Holistically-Nested Edge Detection (HED) network trained in both 2D and 3D that out-performs currently available methods in both pre- and post-operative AAA segmentation [90]. However, this method is limited to single-class segmentation of the aneurysmal wall and performs poorly with small aneurysms and those with a small thrombus burden.

Current CNN architectures can capture semantic contextual information by generating a coarse feature-map grid through iterative down-sampling of the input. Features on this coarse map represent location and relationship between structures/tissues at the organ level. However, these architectures struggle to capture small target objects with increased shape variability. This is especially important for pathological vascular cases. Integrating attention gates, which is commonly used in natural image analysis and classification tasks, into this architecture has shown promise in focusing on target structures without the need for additional training[69]. These attention gates can suppress predictions in irrelevant background region and can be trained simultaneously with the underlying network using standard back-propagation methods. The strength of this attention-based U-Net has been documented on the segmentation of abdominal

structures[69]; however, its role in aortic segmentation has never before been evaluated. Its superior segmentation performance for aneurysmal segmentation rationalized its incorporation within the full aortic segmentation pipeline.

This is the first time a DL method is used to isolate the aorta/AAA from a NCCT scan. This allows for the extraction of complex morphological information from non-contrast images. Furthermore, the same methodology underpinning this work can be extended to enable automatic segmentation of other structures with or without the use of IV contrast agents.

Although CECTs provide unique insight into aneurysm morphology and the vascular tree, it is not without its disadvantages. Administration of contrast requires needle insertion, which is associated with multiple complications including inadvertent arterial puncture and contrast leak from veins causing skin irritation/damage. Additionally, contrast agents are nephrotoxic in patients with a decreased baseline renal function or concomitant chronic kidney disease[7]. Given that a large sub-cohort of patients with aortic aneurysmal disease may have diagnosed renal disease, this study highlights the necessity to re-evaluate the role of NCCT imaging for the management of aneurysmal disease.

4.5 Conclusion

In this chapter, a novel automated pipeline was developed to enable high resolution segmentation of blood vessels (ex. Aorta and its side branches) using deep learning approaches. This pipeline implements a modified U-Net segmentation network for two sequential stages, (1) aortic region-of-interest detection and (2) segmentation. Aortic region-of-interest detection was performed using low-resolution CT images to capture a bounding box region surrounding the aorta. The aorta/AAA within this sub-region was then segmented from the high-resolution CT using an Attention-gated U-Net. We highlight the superiority of this pipeline and the importance of the attention gating mechanism in capturing the aneurysmal pathology by comparing its performance (1) against a base U-Net and (2) with and without data augmentation methods. Subsequently, we were able to automatically extract clinically relevant metrics from the predicted segmentations of the thoracic and abdominal aortas, which can be used within the clinic to efficiently characterize and monitor pathological progression. Finally, this segmentation pipeline was able to capture the aorta/AAA from a NCCT image. This novel

approach can be used to expand the clinical utility of NCCT images in the management of AAA disease.

4.5.1 Limitations

Few limitations exist with this pipeline that need to be considered before clinical implementation. Firstly, this pipeline has multiple steps that are completed in sequence. This is a common method when tackling complex segmentation tasks and reduces the hardware/computational resource requirement that is common to many DL tasks. However, from an efficiency perspective, it requires the training, optimization and validation of multiple models. Given that the models implemented for ROI detection and segmentation are both modified U-Nets, it is highly likely that these networks learn similar information and can theoretically be integrated. The performance of this integrated network (ex. high-resolution aortic/AAA segmentation from the entire CT image) should be comparable to our current pipeline.

Additionally, in this chapter, we implement non-linear divergence transformations to augment the dataset in a ratio of 10:1, prior to model training. In this instance, augmentation was performed to generate CT images with a diverse array of AAA shapes. This is apparent in **Fig. 8** (Section 4.2.6). Of the generated augmentations, few of the generated aneurysmal shapes are not necessarily clinically/anatomically possible, which brings into the question the validity of the selected augmentation method. However, such an approach was used to address the limitation of manually annotated GT data. Furthermore, in **Fig. 17**, we highlight the importance of the implemented augmentation for ensuring model robustness. Future work to address this limitation would be (1) to investigate more clinically appropriate augmentation methods and (2) increase the number of manually-annotated GT images. Possible augmentation techniques include systematically transforming the boundary between the aortic lumen and OWS regions and generating NCCT/CECT 3-D images using anatomically-correct simulations of aorta/AAA volumes using the methods highlighted in section 4.2.3.

Finally, we highlight the potential of this automatic segmentation pipeline to extract arteries; however, we primarily focus on the clinical validation of the aortic segmentation from the aortic root to the iliac bifurcation, with a specific focus on the aortic aneurysm. A semi-

automatic approach was subsequently developed to extend the aortic/AAA segmentation into its major side-branches; however, this step does require user input intended to correct the generated segmentation. Here, future work is necessary to (1) automate this process by incorporating deep learning methods and (2) implement rigid quality control steps to ensure accuracy of the segmentation and the derived measurements.

5

Visualization of Aneurysm Morphology without IV Contrast

Contents

5.1 Introduction 98

5.2 Methodology 100

5.2.1 Curation of CT images from a clinical cohort 100

5.2.2 Registration (alignment) of contrast-enhanced scans to non-contrast scans 100

5.2.3 Extraction, 2D sub-sampling and spatial normalization of aortic structures from NCCT and CECT images 101

5.2.4 Hounsfield unit sampling between AAA regions (Experiment 1) 101

5.2.5 Regional classification of AAAs using radiomic signature (Experiment 2) 102

5.2.5.1 Sub-sampling of the AAA within the NCCT/CECT image 103

5.2.5.2 Radiomic feature extraction and feature selection 103

5.2.5.3 Random Forest Model Training and Evaluation 105

5.2.6 Generative Adversarial Networks: Cycle-/Conditional-GAN (Experiment 3) 106

5.2.6.1 Model Architectures: Cycle-GAN and Conditional GAN 106

5.2.6.2 GAN models training and evaluation 107

5.2.6.3 CT Image quality and GAN model performance 107

5.2.7 Evaluation of clinically important metrics using simulated pseudo-contrast images (Experiment 4) 108

5.2.8 Evaluation of GANs on the independent validation cohort (Experiment 5) 109

5.2.8.1 Influence of AAA shape on GAN performance 109

5.2.9 GAN model training and evaluation for the simulation of aortic side branches (Experiment 6) 110

5.3 Results 112

5.3.1 HU intensities differ between regions of a AAA in NCCT images 112

5.3.2 Radiomic signatures can be used to classify AAA regions from a NCCT 113

5.3.3 Generative Models can simulate contrast images using non-contrast images. 117

5.3.4 CT Image quality affects NCCT Transformation Accuracy 118

5.3.5 Cycle-GAN outperforms Con-GAN in simulating contrast CT images based on the measured clinical metrics 122

5.3.6 Similar GAN performance is observed within the validation cohort (OxAAA-IMG-II cohort). 127

5.3.7 Cycle-GAN can simulate extra-aortic/AAA features including aortic side branches. 128

5.4 DISCUSSION 131

5.5 CONCLUSION 134

5.5.1 Limitations 135

5.1 Introduction

In a computed tomography (CT) image, density variations across tissues translate to differences in tissue attenuation and subsequent radiodensities (measured in Hounsfield Units, HU)[91]. HU values are generally displayed as a greyscale, with brighter regions corresponding to higher attenuation (e.g. bone and calcification) and conversely for darker regions (e.g. fat and air). Where treatment of an artery is being considered, a detailed view of the arterial anatomy is required. In the example of abdominal aortic aneurysms (AAA, abnormal ballooning of the abdominal aorta), an intra-luminal thrombus adherent to the aortic wall within the enlarging aneurysmal sac is present in 95% of cases [92] (**Fig. *30***). Given the similarities in density, these regions cannot be readily distinguished using a conventional NCCT image.

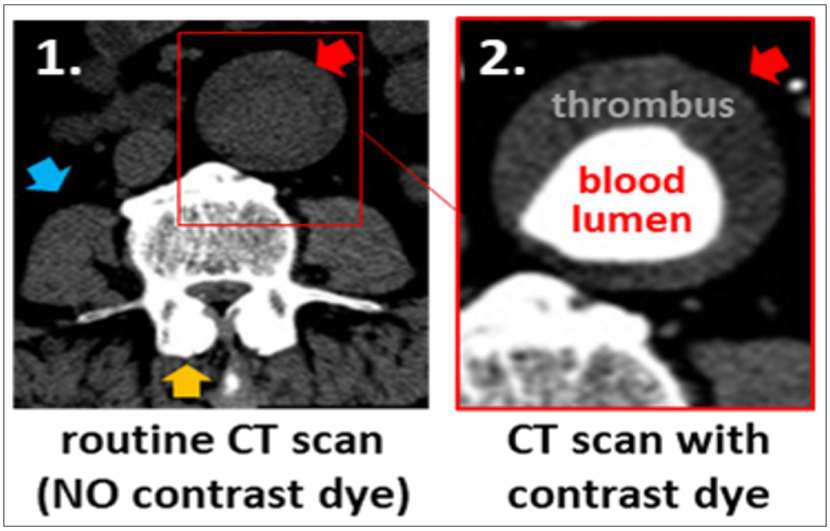


Fig. 30: Axial slice from a Computed Tomography (CT) scan with and without the use of an intravenous iodinated contrast agent. IV contrast enhances visualization of the vasculature and allows for diagnosis of vascular pathology[11].

Clear visualisation of these regions can only be achieved by injecting an intravenous (IV) iodinated contrast agent in a CT angiogram (CECT). IV contrast increases the luminal density and attenuation to distinguish the vascular tree from surrounding soft tissues[18,91]. However, CECTs are associated with several disadvantages [18,19]. CECTs are contraindicated in patients with iodine allergies, as most contrast agents are iodine-based. Additional complications include inadvertent arterial puncture by needle, contrast leak causing skin irritation/damage[19] and

possibly acute kidney injury in susceptible patients[19]. There is a recognised risk of complete kidney failure in these patients, leading to renal dialysis and mortality.

We hypothesise that the raw data acquired from a NCCT be used to differentiate blood and other soft tissues. Blood is predominantly fluid, with red/white blood cells whereas the adjacent outer wall structure (OWS, may contain an intra-luminal thrombus) is predominantly fibrinous and collagenous, with red cells/platelets. These individual components vary in ultrastructure and physical density, which should reflect in different (albeit subtle) HUs on a CT scan (either in individual HU values or in their spatial distribution/"texture"). We further hypothesised that using deep learning (DL) generative methods, these subtle differences can be amplified to enable simulation of contrast-enhanced images without the use of contrast agents.

In this study, we investigate the ability of Generative Adversarial Networks (GANs) for this non-contrast to contrast image transformation task. These networks are a class of DL architectures whereby two neural networks train simultaneously, with one network focused on data generation (generator) and the other focused on data discrimination (discriminator). Designed to mimic how the human brain operates, these neural networks are combinations of simple processing nodes that are heavily interconnected. Each of these nodes performs a mathematical operation, which are adjusted in a "training" process, to capture the underlying relationships in the provided training data for a particular task. In this instance, these networks *compete* against each other to better learn the underlying statistical distribution of the training data. This allows for the generation of new examples from the same distribution[93]. Many variations of the original GAN have been developed, including conditional GANs (Con-GANs[1]) and cycle-GANs. The former learns the transformation between two paired distributions using a pixel-to-pixel approach [94]. On the other hand, the Cycle-GAN is able to learn transformations between two distributions without the need for direct pairings between samples[93]. Here, we applied these DL generative models to generate simulated CECT scans from corresponding NCCT ones. We assessed the performance of these DL generative models using two technical metrics (root-mean-square-error (RMSE), Sørensen–Dice score (DICE)) and

[1] Conditional GANS are usually abbreviated as CGANs in the machine learning literature. Here, Con-GAN is used to avoid confusion with Cycle-GAN.

four clinically important metrics (1. diameter, 2. cross-sectional area, 3. volume of the lumen and OWS, and 4. AAA thrombus morphology).

5.2 Methodology

5.2.1 Curation of CT images from a clinical cohort

CT scans were retrospectively acquired through the ongoing OxAAA study. Details regarding the OxAAA study have been previously published[95]. Each patient gave consent for the use of clinical and imaging data for research analyses. This chapter was inspired by our attempt to utilise historic CT scans acquired during the AAA surveillance period to discover novel signatures of AAA growth[12], as many of these historic CT scans were non-contrast scans.

As part of the routine pre-operative assessment for aortic aneurysmal disease, a NCCT of the abdomen/pelvis and an arterial phase contrast-enhanced CT (CECT) was performed. CECT images were obtained in helical mode with a pre-defined slice thickness of 1.25 mm. On the other hand, NCCT images were obtained with a pre-defined slice thickness of 2.5 mm. Paired contrast and NCCT were anonymised before subsequent analysis. Initial investigations and model training/optimization was performed with the OxAAA-IMG-I dataset of 75 patients with paired NCCT and CECT images (11,243 pairs of images). The OxAAA-IMG-II dataset of 200 patients with paired NCCT and CECT images (29,468 pairs of images) was used to validate the optimized models. Paired images were segmented using our proprietary automated deep learning pipeline[10] (Chapter 4) and aligned to ensure aortic overlap for subsequent analysis.

5.2.2 Registration (alignment) of contrast-enhanced scans to non-contrast scans

To account for voluntary and involuntary movement by the patient between scans, it was necessary to register (align) the contrast-enhanced images obtained to the non-contrast image plane. In order to optimize the registration of the aorta between these two images, the segmented volumes from the CECT and NCCT images were first registered using an non-rigid b-spline registration program in MATLAB[96]. In addition to the registration accuracy calculated within the program, the Sørensen–Dice (DICE) score was calculated to gauge the similarity of the moving/registered image (CECT) with the static image (NCCT). The DICE score metric was defined as in Section 4.2.9. The resultant transformation matrix was applied to the source CECT

image. This method maximized the registration accuracy within the aorta over other anatomical regions.

5.2.3 Extraction, 2D sub-sampling and spatial normalization of aortic structures from NCCT and CECT images

The segmented and registered volumes were used to extract the aortic structure within the entire CT image series of each patient. This was done to focus the subsequent deep learning architectures on the aorta/aneurysm and not on surrounding structures. Additionally, the 3D-isolated aorta from both NCCT and CECT images was divided into 2D axial slices. Subsequently, the aorta in each axial slice (512 x 512) was repositioned to the centre of that slice (at [256,256]). This isolation, 2D-extraction and spatial-normalization was fully automated and was performed to each pair of images to generate the dataset for training and testing.

5.2.4 Hounsfield unit sampling between AAA regions (Experiment 1)

To investigate the regional differences within an aneurysm, 100 paired axial slices within the AAA were selected from the OxAAA-IMG-I dataset. These slices were visually validated for registration accuracy by two blinded reviewers (NS and PL) and were sampled for the underlying Hounsfield unit (HU) distribution at the lumen, intra-luminal thrombus and interface locations (**Fig. 31**). These visually indistinct regions on the NCCT images were identified from their paired CECT images. To account for slight discrepancies in the image registration process, and minimise sampling errors, the thrombus (blue) and lumen (yellow) areas were shrunk by 20% at the adjoining border. The zone between the two regions was demarcated as the interface (red). This delineation is clearly indicated in **Fig. 31b.** The average HU intensities within each region were compared using a One-way-ANOVA. As a negative control, concentric sampling within the blood lumen in each of the NCCT images was performed (**Fig. 31c**).

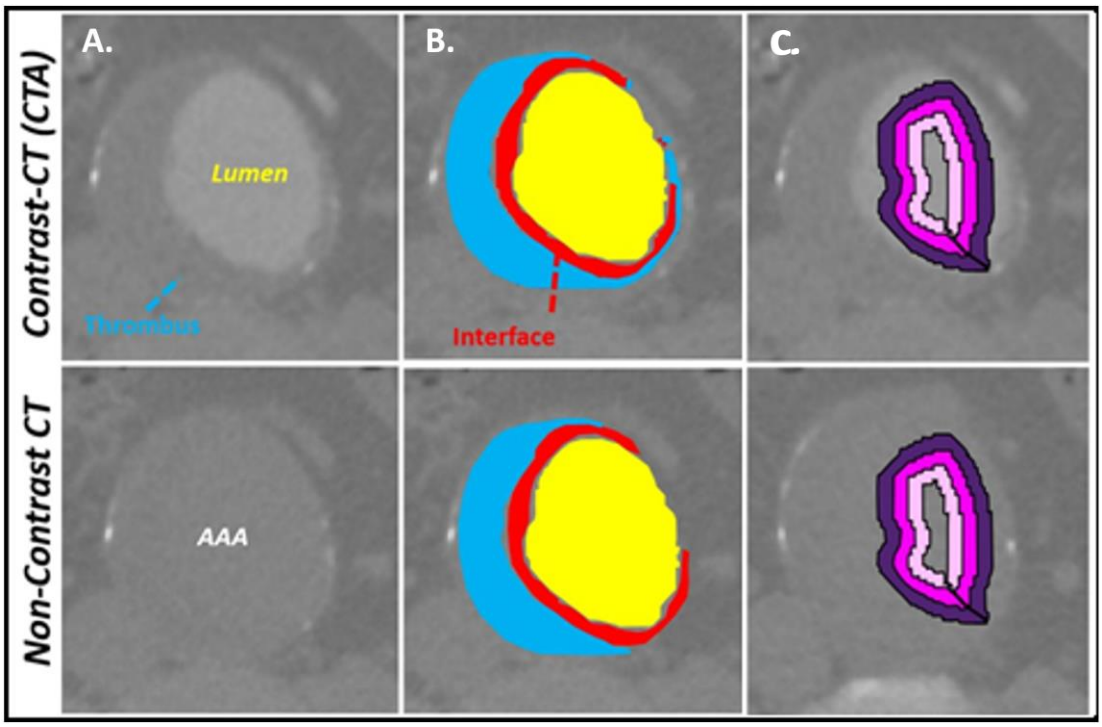


Fig. 31: Axial slices from both the Contrast and Non-contrast CT scans. Demarcated regions display the thrombus (blue), lumen (yellow) and the interface between the regions (red). Hounsfield unit sampling of the lumen, interface and thrombus was performed within each region. Concentric sampling within the lumen as demarcated by the pink, magenta and purple, was used as negative control for this experiment Hounsfield unit sampling within the lumen was hypothesized to have minimal difference compared against that of the interface and thrombus[11].

5.2.5 Regional classification of AAAs using radiomic signature (Experiment 2)

In addition to assessing the average HU intensity between AAA regions, we hypothesized that more complex image-based features are present that are able to distinguish these visually-indistinct regions. As a result, we sought to implement radiomics-based methods to this problem. The field of radiomics employs advanced data-characterisation algorithms to extract higher-order image-based features within a pre-defined region-of-interest and has been used previously to uncover potential disease features that fail to be appreciated visually. This experiment was essential prior to investigating the impact of DL generative networks for the non-contrast to contrast image transformation task (Experiment 3) as it not only provides a baseline level of performance but also may provide insight into the mechanisms of the DL network.

5.2.5.1 Sub-sampling of the AAA within the NCCT/CECT image

In order to investigate the radiomic signatures between AAA regions, sub-volumes with a pre-defined volume were extracted from the entire AAA in fixed increments. Segmentation masks from both the NCCT and CECT images were used to guide sub-volume extraction and all sub-volumes were located entirely within the AAA volume. We hypothesized that sub-volumes of 1 cm^3 would be able to capture the radiomic signatures common to each region and would be small enough to sample the entire AAA. 5 mm increments (stride) were used to minimize the number of sub-volumes extracted from each aneurysm while allowing for sufficient overlap between sub-volumes. Classification of the sub-volume as either lumen or OWS was determined by the CECT-derived segmentation mask. Sub-volumes located partially in both the lumen and OWS (ex. interface) were classified based on the whichever region was mostly represented. The extracted and classified sub-volumes for three patients are illustrated in **Fig. *32***.

5.2.5.2 Radiomic feature extraction and feature selection

Radiomic features were extracted from the NCCT-derived aneurysmal sub-volumes (1 cm^3). Each sub-volume was isotropic with each voxel measuring 1 mm x 1 mm x 1mm. All radiomic features were extracted using Pyradiomics, an open-source python package[55]. For each sub-volume, 18 first-order, 68 second-order and 1118 filter-based features were calculated.

First-order features consist of image-based statistics (ex. minimum, mean, median, maximum, kurtosis, etc), which describe the distribution of voxel intensities within the defined region of interest. On the other hand, second-order features are matrix-based features that capture the radiomic "texture" within a defined region of interest shape/volume. Finally, filter-based features are 1^{st} and 2^{nd}-order features calculated on filtered images. This is a common method used to extract higher-order representations of the data set. The two filtering methods applied to the images included 1. Laplacian of gaussian (LOG) and 2. Wavelet filtering. The former is an edge enhancement filter that seeks to emphasize area of grey level change, where a predefined sigma indicates the degree of coarseness in the filtered image. Here, a high sigma favours coarse textures (grey level variations over a large scale), whereas a low sigma favours finer textures (grey level variations over a smaller scale). In this study, 5 levels of sigma were used (σ = 1,2,3,4,5) with 86 $1^{st}/2^{nd}$ order features extracted for each sigma (425 - 3^{rd} order LOG features). Similarly, wavelet filtering produces 8 decompositions for each input image; these

decompositions involve applying either a high (H) and/or low (L) pass filter in 3-dimensions (HHH, HHL, HLH, HLL, LLL, LLH, LHL, LHH). For each decomposition, 86 1st and 2nd order features are extracted (680 – 3rd Order Wavelet features). Full documentation for each of the extracted features can be found along with the pyradiomics source code[55].

Features were ranked using the minimum redundancy, maximum relevance (MRMR) algorithm in MATLAB. The top 25 features were selected for model training and optimization. This algorithm was selected as it tends to efficiently parse through a large array of features to select a subset with a high correlation to an output class and a low correlation between the features[97,98]. By scaling down the feature vector, it prevents overfitting and maximizes model interpretability, which is extremely valuable for clinical problems.

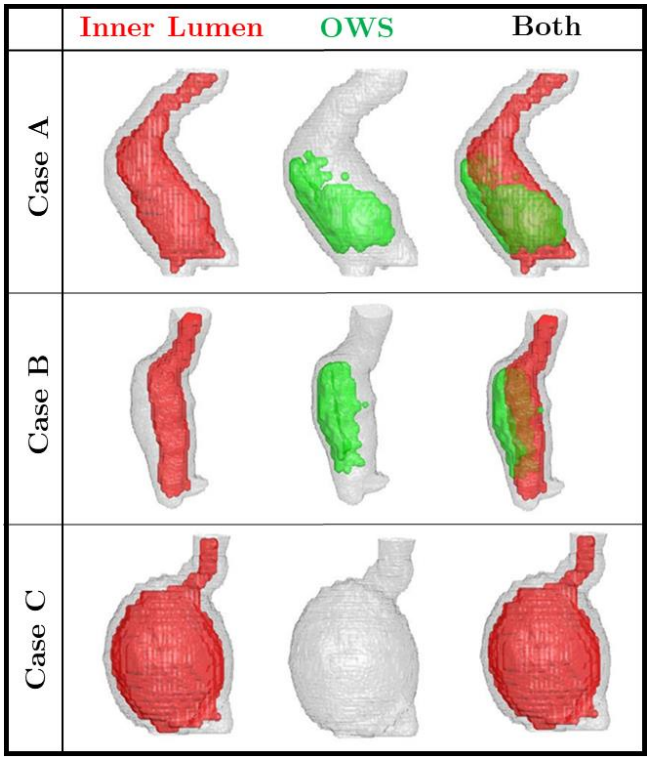


Fig. 32: Sub-sampling of AAA shape and regional classification. Sub-volumes (1 cm x 1cm x 1cm) were extracted from each AAA in 5 mm increments. Sub-volumes were classified into either Lumen (red) or OWS (green) based on CECT-derived segmentations. The OWS contains both the aneurysmal outer wall and intra-luminal thrombus when present. Radiomic features were extracted from each of the sub-volumes and classification methods were used to decipher the signatures specific to each region.

5.2.5.3 Random Forest Model Training and Evaluation

Three random forest models were trained for the classification of sub-volumes from NCCTs (1. Lumen, 2. OWS). Random forests are a family of machine learning classification algorithms that employ many individual DTs that operate as an ensemble for a particular task. During training, the weights for each tree are fine-tuned to create an uncorrelated forest of trees whose prediction by ensemble is more accurate than that of any individual DT. Here, each model was trained/optimized (10-fold cross-validation) using a dataset of 50 patients (OxAAA-IMG-I dataset) . Following optimization, the best of the cross-validation models was tested on a unique set of 25 patients (OxAAA-IMG-I dataset). Area under the receiver operator characteristic (AuROC) curve on the testing cohort was used to assess model performance. Additionally, model predictions for each sub-volume were used to reconstruct a 3D representation, which was then compared against that derived from the GT segmentation. This pipeline for one of the three models is illustrated in **Fig. *33***. This pipeline is identical for the remaining two models.

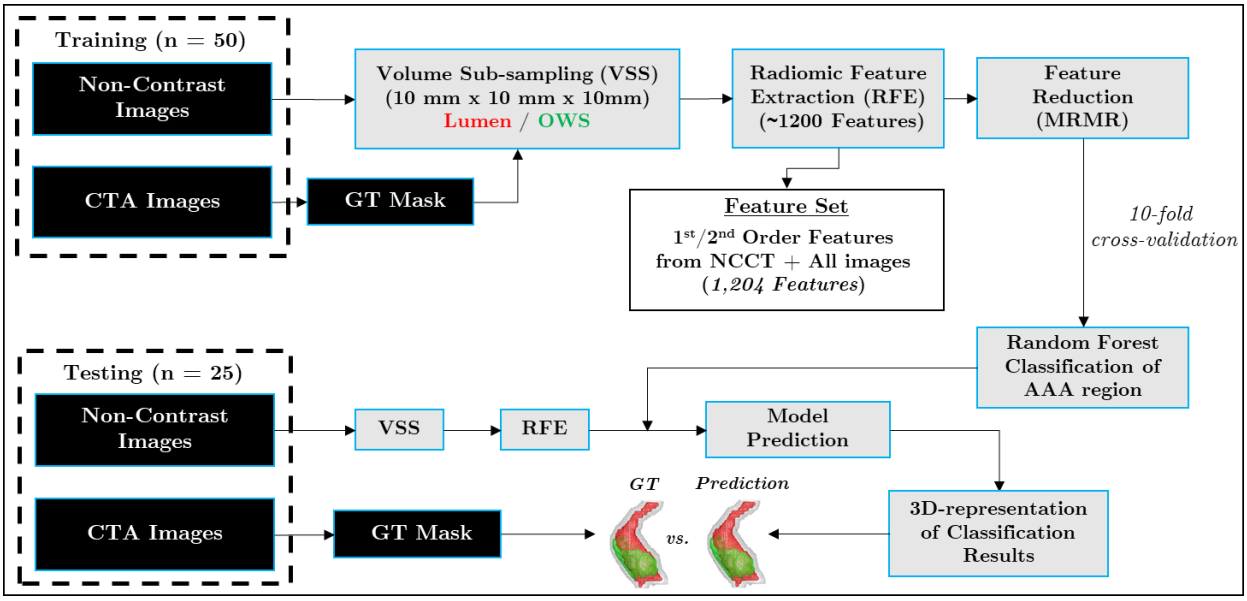


Fig. 33: Pipeline for the regional Classification of AAA using higher-order radiomic features. Three testing folds were created, each with a unique set of 25 patients from the OxAAA-IMG-I dataset. For each fold, the remining 50 patients were used as the training cohort. Multiple 1 cm^3 sub-volumes were extracted in 5 mm increments throughout the entirety of the AAA shape from the NCCT image. 1,205 radiomic features were extracted from each sub-volume. Within each training cohort, 10-fold cross-validation was performed. The best of the cross-validation models was then applied to the pre-defined testing cohort. Area under the ROC curve for the testing cohort was used to assess classification accuracy. Classification results were used to generate a crude/down-sampled model prediction which could be compared against that of the GT using DICE score (%).

5.2.6 Generative Adversarial Networks: Cycle-/Conditional-GAN (Experiment 3)

In this study, the Cycle-GAN and the conditional GAN (Con-GAN) were used for the non-contrast to contrast image transformation task (**Fig. *34***). Model architectures and training details are described in the subsequent sections.

5.2.6.1 Model Architectures: Cycle-GAN and Conditional GAN

The generator and discriminator components in the Cycle-GAN model architecture (**Fig. *34*a**) were explicitly defined as least-squares GAN and a 70 x 70 pixel PatchGAN, respectively[93]. The former incorporates an additional least-squares loss function for the discriminator, which in turn, improves the training of the generative model. On the other hand, the discriminator goes through the image pairs, in 70 x 70 patches, and is trained to classify whether the image under question is "real" or "fake". In addition to the cycle-GAN, a Pix2Pix-Conditional-GAN was trained and evaluated. Unlike the cycle-GAN, conditional GANs require paired non-contrast and contrast images, which need to be registered to ensure a pixel-to-pixel correspondence between them. The generator and discriminator components in the conditional GAN model architecture were identical to those used in the Cycle GAN (**Fig. *34***b).

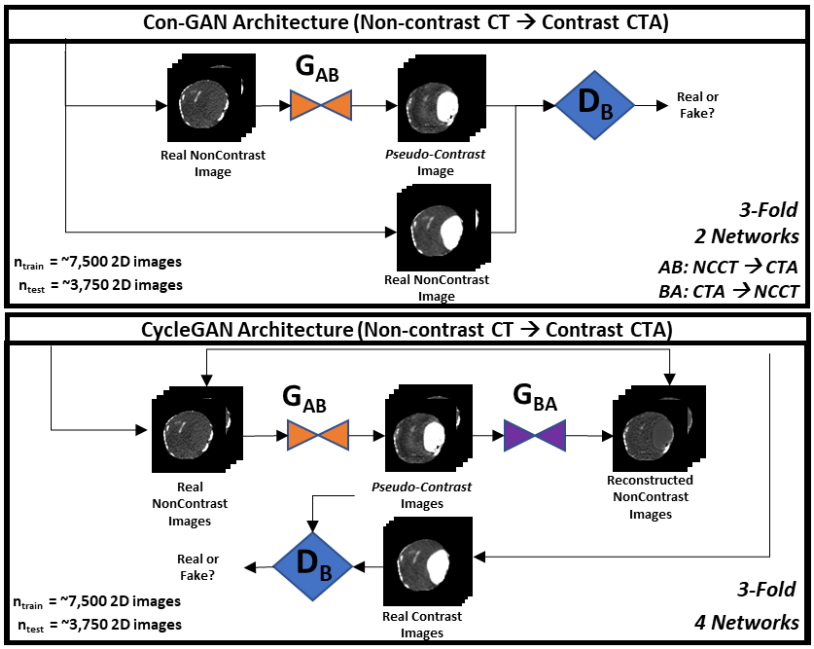


Fig. 34: Con-GAN (A.) and CycleGAN (B.) architectures used for the transformation of NCCT images to contrast CECT images. The former consists of 2 networks (1 generator and 1 discriminator) and the latter consists of 4 networks (2 generators, and 2 discriminators). Each model was trained using a 3-fold cross-validation paradigm[11].

5.2.6.2 GAN models training and evaluation

A 3-fold cross-validation paradigm with a training/test data split of 50:25 patients (~7,500: ~3,750 2D axial slices, OxAAA-IMG-I dataset) was employed for both networks. The Con-GAN and Cycle-GAN models were trained with a learning rate of $2.0 * 10^{-5}$ for 200 epochs on 256 x 256 images centered around the aorta. For the Cycle-GAN architecture, four networks (2 generators + 2 discriminators) were trained simultaneously and various loss functions were evaluated at each iteration to document model training. In addition to the loss metrics inherent to the networks, an identity mapping and a cycle consistency loss functions were included to ensure appropriate style transfer and regularization of the generator to allow for image translation, respectively. On the other hand, two networks (1 generator + 1 discriminator) were trained for the Con-GAN. Model weights were saved every 10 epochs and intermediate model predictions were generated from the NCCT images within the training cohort. The predictions were evaluated against the ground truth CECT images to assess model training.

Model performances were evaluated using two metrics: Root-Mean-Square-Error (RMSE) and Sorensen-DICE score. The former is a commonly used metric in image transformation tasks to assess the pixel-by-pixel difference between the simulated CECT image and the ground truth (i.e. Paired CECT Image). The latter is a quantitative similarity assessment of regions within sets of images[99]. Here, $DICE_I$ was calculated to assess overlap in the inner lumen between the simulated pseudo-contrast and CECT images. Additionally, $DICE_C$ measures aortic shape overlap between the simulated and ground truth images. This is a surrogate measure to evaluate registration accuracy and GAN-induced shape alterations.

5.2.6.3 CT Image quality and GAN model performance

The NCCT images used in this study were acquired using a variety of settings (ex. x-ray tube current, rotation time, and pitch factor) that may impact image quality. The combination of x-ray tube current and rotation time determines the amount of incident x-ray photons, which is strongly correlated with image quality. Given that these DL platforms utilise these images and their pixel relationships as the basis for transformation, we hypothesised that NCCT images obtained at greater x-ray tube currents would result in qualitatively and quantitatively better simulated CECT images. This was evaluated by comparing patients with the best and worst image transformations to define an optimal cut-off criterion of image quality.

5.2.7 Evaluation of clinically important metrics using simulated pseudo-contrast images (Experiment 4)

The following metrics were obtained from the CECT and simulated pseudo-contrast images using an in-house algorithm in MATLAB[100]: 1-D measurements: maximum lumen and aortic wall diameters; 2-D measurements: blood lumen and OWS areas; 3-D measurements: blood lumen and OWS volumes. This in-house algorithm for the characterization of AAA volumes was validated in chapter 4. Bland Altman plots and correlation coefficient analysis were performed for each to assess bias and the strength of association between the output of the GAN models and the ground truths. Bias measurements are reported along with its 95% confidence interval (95% CI). In addition, the spatial morphology of the OWS was extracted using an in-house algorithm in MATLAB. The general algorithm is highlighted in **Fig. *35*** and involves classifying the OWS along axial planes through the AAA shape. Based on our previous work, spatial morphology of the OWS can be categorised into seven categories (**Fig. *35***)[95]. This is an important feature used to plan surgical intervention and can influence post-surgical outcomes[95].

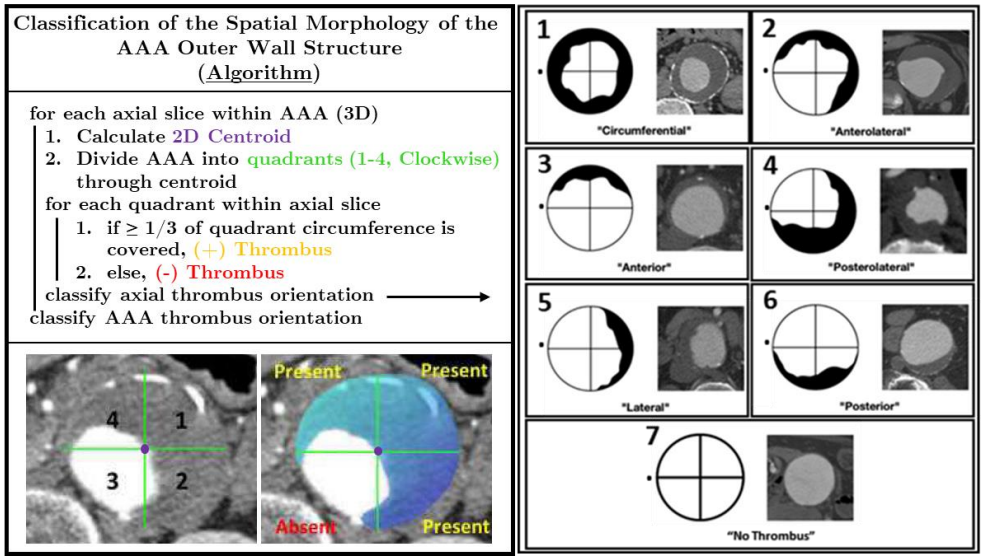


Fig. 35: Regional classification of thrombus morphology. Thrombus presence was classified into seven distinct categories (1-7) based on methods adopted from our previous work[95]. The algorithm requires moving through the AAA shape and classifying the OWS within the axial slice into 7 categories. For each axial slice, perpendicular anteroposterior and transverse lines through the centroid divide it into 4 quadrants, labeled 1-4 in a clockwise fashion. OWS within each quadrant was classified as 'visible' if it occupied ≥ 1/3 of the quadrant circumference. This threshold was chosen as it accounts for minimal coverage area within the quadrant, while being large enough to be easily appraised. AAA thrombus orientation was classified by choosing the most common thrombus classification.

5.2.8 Evaluation of GANs on the independent validation cohort (Experiment 5)

The OxAAA-IMG-II dataset was used as an independent set of 200 patients with paired NCCT and CECT images (total of 29,468 pairs of images) for model validation. The con- and cycle-GAN models from each validation fold (3) were used to generate pseudo-contrast images of the aorta/aneurysm. The outputs of each model were averaged to generate the final pseudo-contrast images. As a result, for each NCCT image, two pseudo-contrast images were generated: one from the averaged Con-GAN outputs and the other from the averaged Cycle-GAN outputs. Following image transformation, CECT images were provided to assess transformation accuracy. This was completed using the same methods and metrics as described above.

5.2.8.1 Influence of AAA shape on GAN performance

Subsequently, the effect of aneurysmal size (as measured by maximum AAA diameter) and aneurysm shape (fusiform vs saccular) on GAN performance was investigated. The shape of an aortic aneurysm is described as either fusiform or saccular. The former suggests that the aneurysm dilates on all sides of the aorta and appears uniform in shape. On the other hand, a saccular-shaped aneurysm bulges/dilates on one side (asymmetric). Difference in AAA shape may lead to altered hemodynamic conditions and wall stresses within the vessel and have been shown to lead to different clinical outcomes[101]. In this study, aneurysmal shape was quantified using the non-fusiform index (NFI), which is a 3-D shape index that describes the deviation of the aneurysmal sac from an ideal fusiform shape. Derived from Martufi et al., the NFI is based on luminal surface area (S) to the volume of the aneurysmal sac (V) and is compared against that from an idealized fusiform model[102].

$$NFI = 1 - \left(\frac{V_{fusiform}^{\frac{2}{3}}}{S_{fusiform}}\right)\left(\frac{S}{V^{2/3}}\right)$$

This idealized fusiform aneurysm, initially proposed by Finol et al. (2002) and used previously in chapter 4 to generate AAA shape for methods validation, utilizes the patient-specific neck diameter (D_{neck}), maximum AAA diameter (D_{max}), and height of the aneurysmal sac (Ht) and is modelled by the following equation:

$$f_{fusiform}(z) = \left(\frac{D_{max} - D_{neck}}{4}\right)\left[1 + \sin\left(\frac{2\pi z}{Ht} - \frac{\pi}{2}\right)\right] + \frac{D_{neck}}{2}$$

In order to investigate the role AAA size and shape on GAN performance, maximum diameter and NFIs of AAAs were correlated against GAN transformation accuracy, as measured by the DICE score accuracy of the inner lumen.

5.2.9 GAN model training and evaluation for the simulation of aortic side branches (Experiment 6)

From Experiment 5, we selected the GAN model with superior performance in simulating aortic/aneurysmal-specific features for subsequent analyses. The training data for this experiment consisted of 2-D axial slices centred around the aorta/AAA with surrounding tissue present (**Aortic Region-of-Interest, ROI; Fig. *36*b**). As a result, this generative network was trained to concurrently simulate both intra- (aortic lumen, thrombus morphology) and extra-aortic/AAA features (main aortic side branch origins: Coeliac artery, superior mesenteric artery, renal arteries, and iliac arteries). Same model training (3-fold cross-validation paradigm, n = 75 patients, OxAAA-IMG-I dataset) and evaluation parameters were utilised for this experiment. Branch capture rate was calculated for each artery to quantify the ability of the GAN to identify/transform the aortic side branches. Maximum aortic side branch diameter was obtained from both the contrast-enhanced and the simulated pseudo-contrast CT images for the six major branches originating from the descending/abdominal aorta (1. Coeliac artery, 2. Superior Mesenteric Artery, 3-4. Left/Right Renal Arteries, and 5/6. Left/Right Iliac Arteries). Mean average error (MAE) between measurements was reported for each branch. Student t-testing was performed to assess for statistical significance. Model performance was validated using the OxAAA-IMG-II dataset.

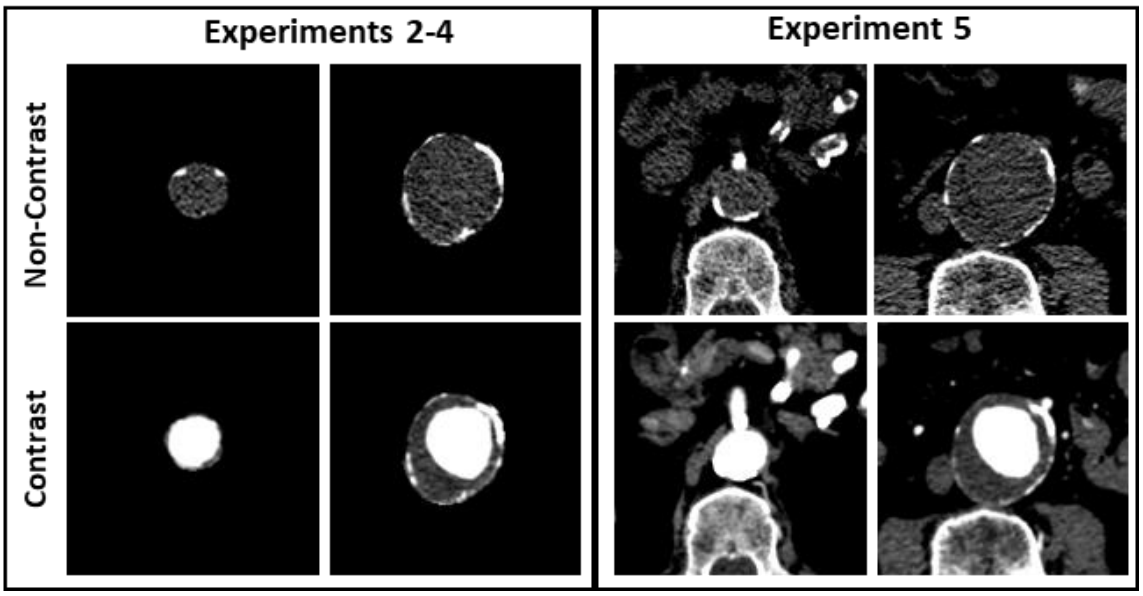


Fig. 36: Input data for the GAN architectures. A. For experiments 2-4 (Cycle-GAN vs Con-GAN), the input data consisted of 240*240 mm images (256 x 256 pixels) centred around the aorta with all surrounding structures removed. This allowed for the GAN to only focus on the AAA transformation. **B.** For experiment 5, the input data consisted of 135*135 mm (144 x 144 pixels) images centred around the aorta. Extra-aortic structures including side-branches were not removed. This allowed the GAN to focus also on the transformation of adjacent structures.

5.3 Results

5.3.1 HU intensities differ between regions of a AAA in NCCT images

There are discernible differences between the HU frequency distributions of each region (**Fig. 37b**). Average HU intensity in the NCCT images differed significantly between all three regions (blood lumen vs thrombus, blood lumen vs interface and interface vs thrombus) for all patients assessed. **Fig. 37b** demonstrates the significant differences in HU intensities between the thrombus and lumen for eight axial slices obtained from one patient. Furthermore, the blood lumen/thrombus interface also differed significantly from the other two regions, indicating a gradual change from lumen to the thrombus. As a comparison, no significant HU differences were observed following concentric sampling within the lumen (**Fig. 37c-d**).

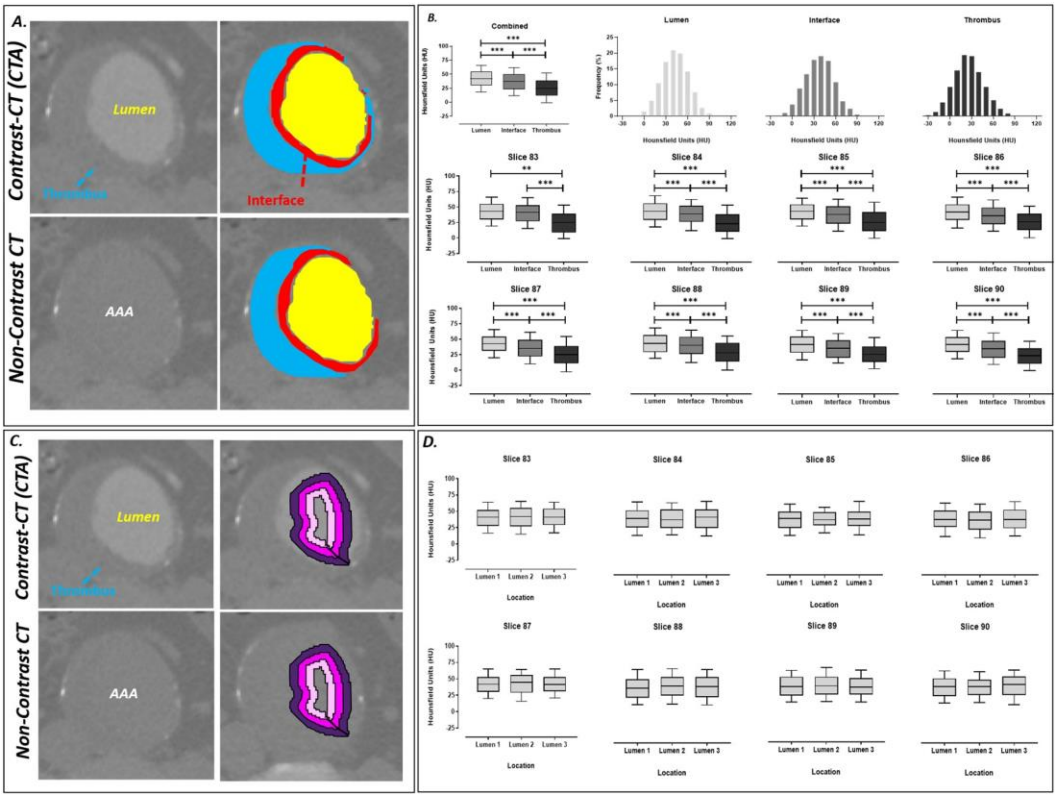


Fig. 37: Axial slices from both the Contrast and Non-contrast CT scans. A. Demarcated regions display the thrombus (blue), lumen (yellow) and the interface between the regions (red). **B.** Hounsfield unit sampling of the lumen, interface and thrombus with histograms displaying the frequency of HUs within each region. Analysis at each axial slice was performed and showed differences in HU intensity between the lumen, interface and thrombus regions. **C.** Concentric sampling within the lumen as demarcated by the pink, magenta and purple, was used as negative control for this experiment **D.** Hounsfield unit sampling of the lumen at multiple locations indicated minimal difference in HU intensity[11].

5.3.2 Radiomic signatures can be used to classify AAA regions from a NCCT

In addition to average HU intensity, we hypothesized that the aortic lumen and OWS, of a AAA from a NCCT image display differences in radiomic signature that can be deciphered and used for classification. Anisotropic NCCT images with AAAs were re-sampled to have isotropic dimensions (voxel dimensions: 1 mm^3). From 75 patients (OxAAA-IMG-I dataset), 86,411 sub-volumes (1 cm^3) were extracted from the AAA volume. Sub-volumes were then classified based on location within the AAA into two categories: 1. Lumen (n = 58,936) and 2. OWS (n = 27,475). Three testing folds were created, each with a unique set of 25 patients. The remaining 50 patients for each fold were used for model training and optimization. The training/validation and testing splits for each model are illustrated in **Table *11***.

Table 11: Sub-volume splits for each of the three folds during random forest model training/ testing (Regional Classification of AAA using Radiomic Signatures, OxAAA-IMG-I dataset).

Fold-1	**Train/Validation Cohort (n = 50)**		**Test Cohort (n = 25)**
	Train	Validation	
Lumen	27,742	11,890	19,304
OWS	13,252	5,679	8,544
Total	40,994	17,569	27,848
Fold-2	**Train/Validation Cohort (n = 50)**		**Test Cohort (n = 25)**
	Train	Validation	
Lumen	27,707	11,875	19,354
OWS	11,836	5,072	10,567
Total	39,543	16,947	29,921
Fold-3	**Train/Validation Cohort (n = 50)**		**Test Cohort (n = 25)**
	Train	Validation	
Lumen	31,053	13,320	14,574
OWS	15,009	6,433	6,033
Total	46,062	19,753	20,607

1st and 2nd-order features from the NCCT-derived sub-volumes were extracted. Feature reduction (MRMR algorithm) was performed on the training cohort within each fold to produce three sets of 25 features. These features were then used to train 3 independent random-forest models using a 10-fold cross validation paradigm. The performance of the best performing model for each fold was then applied to the previously defined testing cohort. Model performances on the testing cohorts are illustrated in **Fig. *38a***. The top five features from each model, in terms of feature importance, are highlighted in **Table *12***. Model predictions for each sub-volume were reconstructed into a 3-D segmentation mask, which was then compared against that of the ground-truth. The average $DICE_I$ metric between the inner lumen segmentation derived from the random-forest model classification and that of the ground truth was 88.0 ± 12.8%. Three examples of 3D reconstructions from each of the three models are illustrated in **Fig. *38b*** along with their respective $DICE_I$ accuracy. The produced reconstructions are at a lower resolution than the CECT-derived segmentations, as each sub-volume is 1 cm^3 with a stride of 5 mm. The visualizations generated through this method are able to differentiate visually-indistinct regions of a NCCT using 1st and 2nd order radiomic features and sets the foundation for DL generative methods.

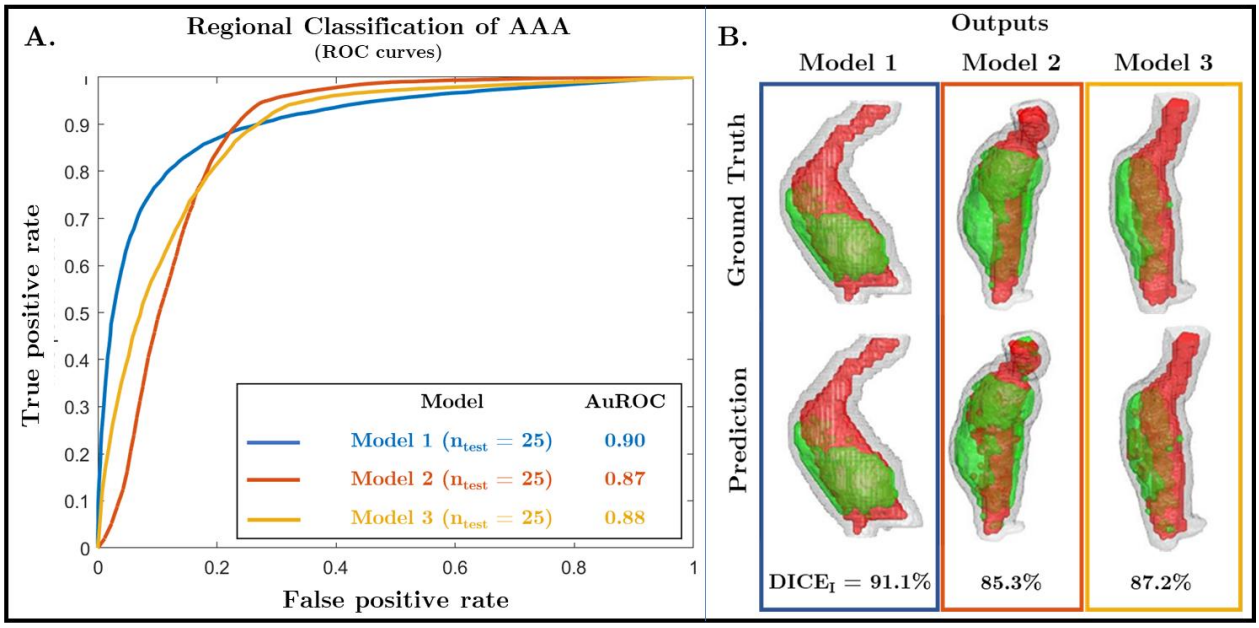


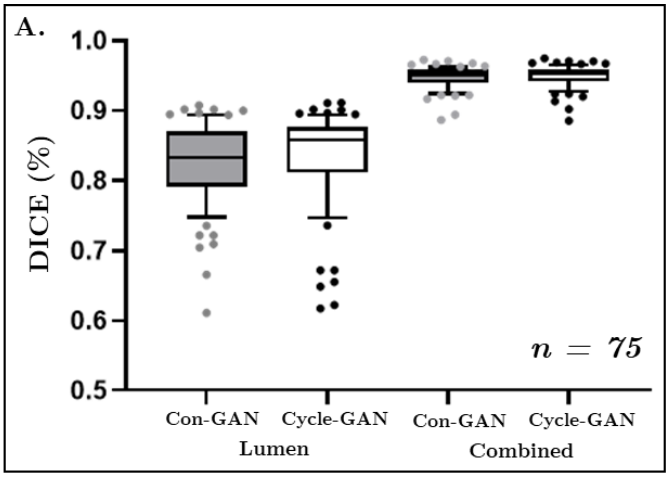
Fig. 38: Random forest model performance for the regional classification of AAA. 1st and 2nd order radiomic features were extracted from the defined sub-volumes. The Minimum redundancy, Maximum relevance (MRMR) algorithm in MATLAB was used to reduce the feature set to 25 features. Three models were trained/optimized (10-Fold Cross-validation) using a training dataset of 50 patients (Table 11). Model performances was evaluated by measuring the area under the ROC (AuROC) curves on the testing datasets (n = 3), which each consisted of 25 unique patients. B. Model predictions were used to construct a down-sampled segmentation mask of the AAA and were compared against that derived from the ground truth using DICE accuracy of the inner lumen (red). Three predictions, one from each model, are visualized alongside their respective ground truth segmentations. The results support the ability to differentiate visually-indistinct regions using 1st and 2nd order radiomic features.

Table 12: Most important 1st and 2nd Order radiomic features for regional classification of AAA (for each trained/optimized model) along with feature importance (range: 0 → 1; 0 – not important, 1 – very important).

	Features	Importance
Model 1	LOG (σ = 3.0 mm) -2nd Order (GLDM) Dependence NonUniformity	0.12
	LOG (σ = 2.0 mm) -1st Order 10th Percentile	0.09
	Wavelet (LLL) – 1st Order 90th Percentile	0.08
	LOG (σ = 5.0 mm) -1st Order Maximum	0.07
	NCCT (Original) – 1st Order Median	0.06
	Features	**Importance**
Model 2	LOG (σ = 3.0 mm) -2nd Order (GLDM) Dependence NonUniformity	0.13
	LOG (σ = 2.0 mm) -1st Order 10th Percentile	0.09
	LOG (σ = 4.0 mm) -2nd Order (GLSZM) SizeZoneNonUniformity	0.08
	NCCT (Original) – 1st Order Median	0.07
	LOG (σ = 5.0 mm) -1st Order Maximum	0.07
	Features	**Importance**
Model 3	NCCT (Original) – 1st Order 90th Percentile	0.14
	LOG (σ = 3.0 mm) -2nd Order (GLDM) Dependence NonUniformity	0.11
	LOG (σ = 2.0 mm) -1st Order 10th Percentile	0.09
	Wavelet (LLL) – 1st Order 90th Percentile	0.05
	LOG (σ = 5.0 mm) -1st Order Maximum	0.05

5.3.3 Generative Models can simulate contrast images using non-contrast images.

During training of the generative algorithms, the RMSE between the simulated pseudo-contrast images and ground truth (CECT) images decreased with each epoch (training cycle) to plateau at 3.8 ± 0.8 and 3.9 ± 0.6 for the con-GAN and cycle-GAN, respectively. Similarly, the $DICE_I$ increases with epoch duration to plateau at 91.8 ± 0.6% and 92.0 ± 0.4% for the con-GAN and cycle-GAN, respectively. **Fig. *39*** indicates the RMSE and DICE scores for the images generated from the testing cohorts across the three folds using both GANs (n = 11,243 images, 75 patients, OxAAA-IMG-I dataset). A per-patient transformation accuracy was derived by grouping the 2D-axial images and their respective DICE scores by the patient. With regards to $DICE_I$, the median performance of the Cycle-GAN is greater than that of the Con-GAN. Of note, there were multiple overlaps between the outliers below the 10th percentile of cycle-GAN and Con-GAN (6/7) networks. This suggests that there may be image properties inherent to this subgroup of NCCTs leading to decreased transformation accuracy.



B. n = 75		Con-GAN	Cycle-GAN
RMSE		4.4 ± 3.7	4.6 ± 4.1
DICE	*Inner*	85.2 ± 11.4 %	85.6 ± 14.5 %
	Comb.	94.9 ± 4.3 %	95.4 ± 4.0 %

Fig. 39: Transformation accuracy within the OxAAA-IMG-I cohort. A. Box Plots of Averaged DICE scores per patient for the lumen ($DICE_I$) and the combined aortic mask ($DICE_C$) segmentations. Segmentations are derived from model (Cycle-GAN/Con-GAN) predictions. Individual data points outside the 10-90th percentile are highlighted. **B**. RMSE/DICE scores of generated pseudo-contrast axial images[11].

5.3.4 CT Image quality affects NCCT Transformation Accuracy

Comparing the image properties of the NCCT scans below the 10th percentile and above the 90th percentile highlighted one key difference with regards to the x-ray tube current (I_{tube}) used during image acquisition (**Fig. *40***). Scans obtained with lower I_{tube} values produced images with poor transformation accuracy. Subsequently, a threshold criterion of above the 15th percentile of I_{tube} was implemented. This equated to a selection criterion of images obtained with $I_{tube} > 80$ mA. Apart from tube current, there were no observable differences between other CT acquisition parameters (Mean ± SD) for the two subsets of cases. Notably, no difference was observed in spiral pitch factor (Excluded: 1.08 ± 0.17, Included: 1.10 ± 0.18, p = 0.42), slice thickness (Excluded: 2.5 mm, Included 2.5 mm, p = 1.0), and total collimation width (Excluded: 35.9 ± 8.2, Included: 34.2 ± 9.3, p = 0.33).

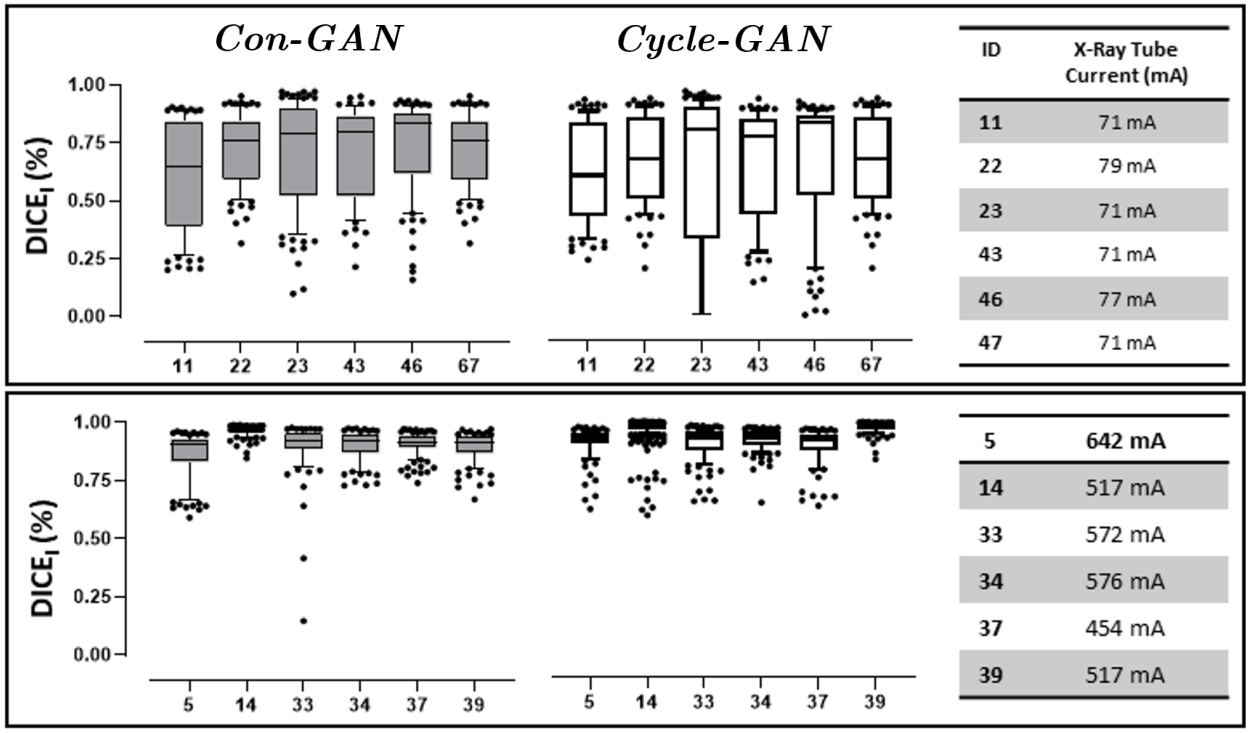
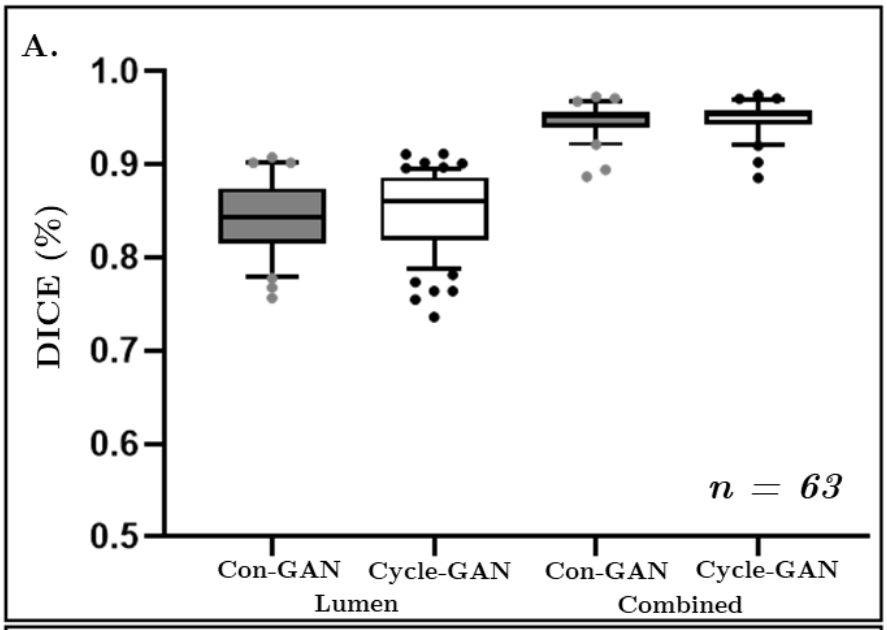


Fig. 40: Increased DICE score variability in NCCT images obtained with low X-Ray tube currents (mA). Patients (n =6) with the lowest reconstruction accuracy **(A**, Tube Current: 73.3 ± 3.67 mA**)** are compared against those (n = 6) with the highest reconstruction accuracy **(B**, Tube Current: 546.3 ± 64.7 mA**).** X-ray tube currents for the NCCT images are shown accordingly[11].

Exclusion of these 12 patients (I_{tube} < 80 mA), resulted in improvements in RMSE, and in $DICE_I$ for both generative models (**Fig. *41***). $DICE_C$ scores remained unaffected by this exclusion. The blood lumen generated from the Cycle-GAN bears a closer resemblance to the ground truth image as compared to that generated by the Con-GAN, as reflected by the superior $DICE_I$ scores. On the other hand, $DICE_C$ scores were identical for both sets of model predictions. This is apparent in **Fig. 42**, which illustrates paired CECT/NCCT axial slices alongside their Pseudo-*Contrast* images from six different patients. The segmented aneurysm shape for each of these patients is illustrated in **Fig. *43.*** Correspondingly, the inner lumen areas and thrombus volumes derived from the cycle-GAN model outputs are better approximations to those derived from the ground truth compared to that of the con-GAN.



B. n = 63		Con-GAN	Cycle-GAN
RMSE		3.1 ± 2.6	3.0 ± 2.4
DICE	*Inner*	85.7 ± 10.4 %	86.1 ± 12.2 %
	Comb.	94.9 ± 4.3 %	95.4 ± 4.0 %

Fig. 41: Transformation accuracy of the refined cohort. A. Box Plots of averaged DICE scores per patient within the refined cohort (I_{tube} > 80 mA, n = 63) for the lumen ($DICE_I$) and the combined aortic mask ($DICE_C$) segmentations. **B.** Overall RMSE and DICE scores of pseudo-contrast images from NCCT images obtained at >80 mA[11].

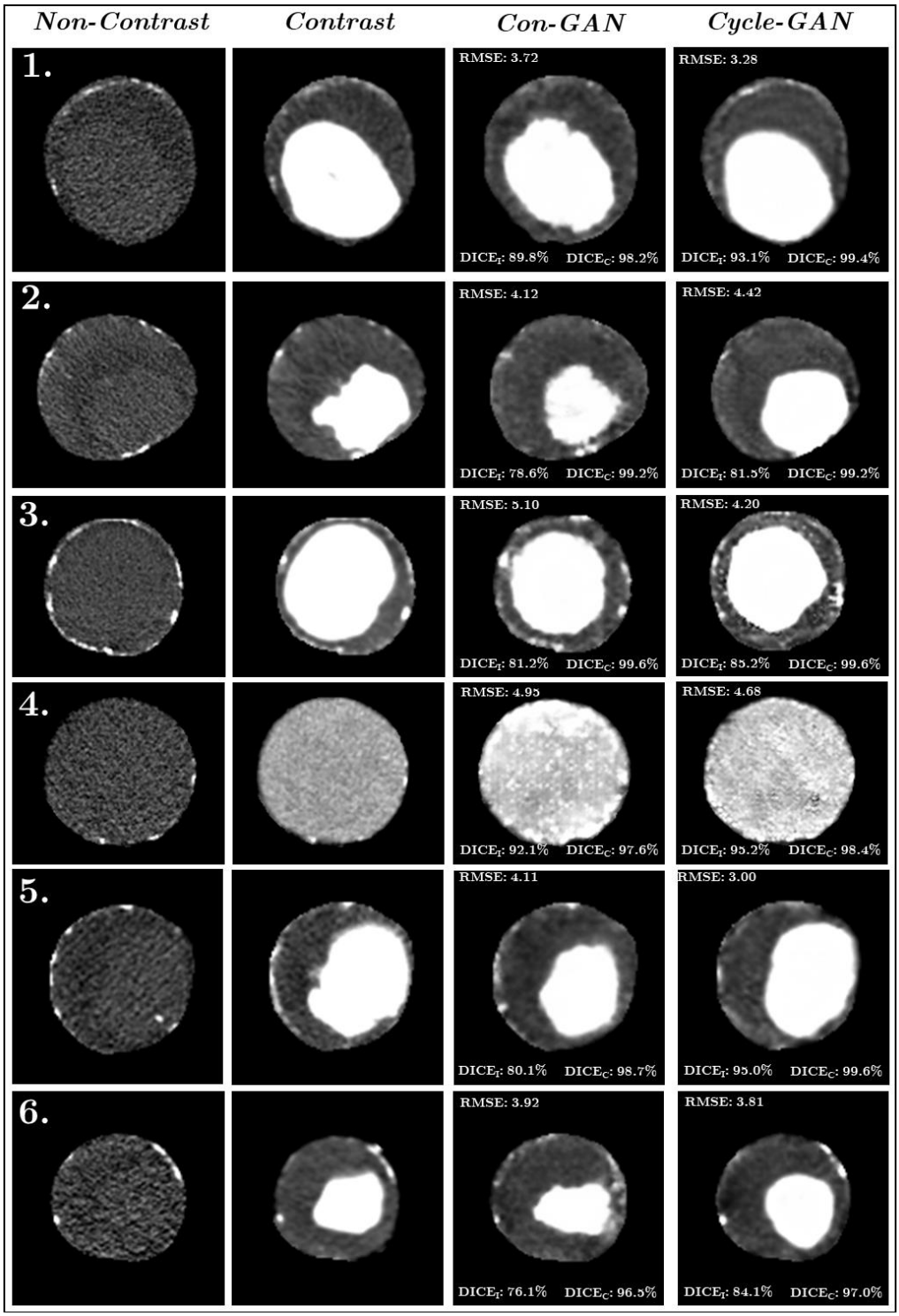


Fig. 42: Pseudo-Contrast images generated using the Con-GAN and Cycle-GAN architectures from 6 different patients alongside their respective NCCT and CECT axial slices. Using the CECT axial images as GT, RMSE and DICE scores for the inner lumen ($DICE_I$)and combined aorta ($DICE_C$) are calculated[11].

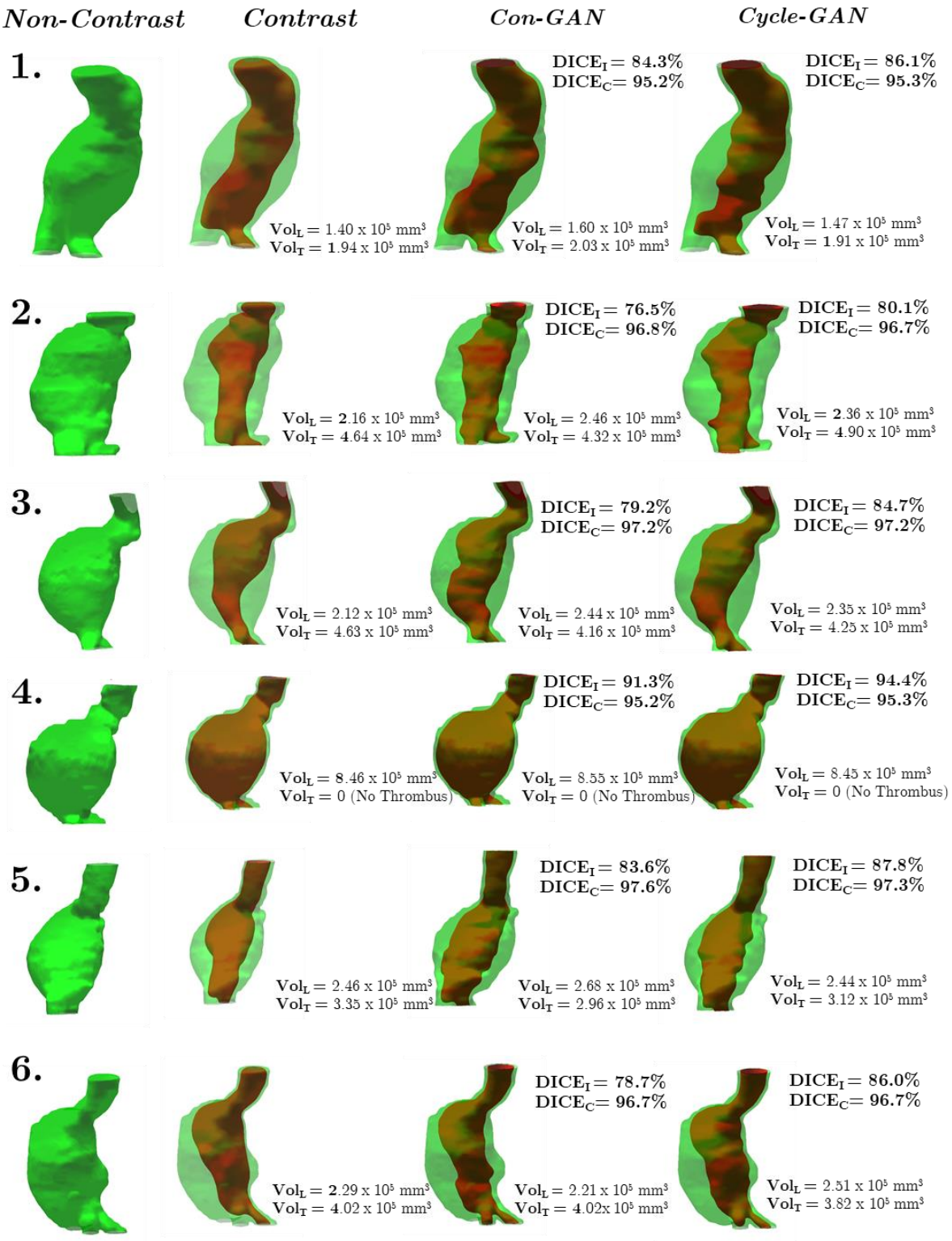


Fig. 43: Segmented AAA shapes generated from the Con-GAN and Cycle-GAN architectures. AAA shapes from 6 patients, the same as seen in Figure 6, are displayed alongside their respective segmentations from NCCT and CECT images. $DICE_I$, $DICE_c$ and lumen/thrombus volumes are calculated and displayed for each aneurysmal region[11].

5.3.5 Cycle-GAN outperforms Con-GAN in simulating contrast CT images based on the measured clinical metrics

Evaluation of aneurysm morphology is useful in defining the biological behaviour of an AAA during the natural history of the disease[79,81,83,95,103]. Measurements derived from the two GAN models' outputs were compared against those obtained from the ground truth (GT).

1-D measurements: The Cycle-GAN model is better at approximating the maximum lumen diameter than the Con-GAN model (**Fig. *44*a**). On the other hand, both models have similar strengths in determining the outer vessel wall diameter (**Fig. *44*b**). Maximum inner lumen and outer vessel wall diameters extracted from the model outputs correlated strongly with GT measurements (Cycle-GAN, ρ = 0.85 and 0.99, p <0.01; Con-GAN, ρ = 0.83 and 0.98, p <0.01).

2-D measurements: The Cycle-GAN model again performed better than the Con-GAN model (**Fig. *44*c-d**). Thrombus area in each axial slice as determined by the Cycle-GAN and Con-GAN models is on average 9.3 ± 11.5% and 9.4 ± 12.2% different from GT measurements.

3-D measurements: The Cycle-GAN better approximates the 3D- lumen and thrombus volume measurements than the Con-GAN (**Fig. *44*e-f**). Lumen volumes derived from both models (Cycle-GAN: ρ = 0.84, p < 0.01; Con-GAN: ρ = 0.82, p < 0.01) and thrombus volumes (ρ = 0.86, p < 0.01) from the Cycle-GAN correlated strongly with the manually derived measurements.

Pseudo-contrast images within the aneurysmal region produced by Cycle-GAN (**Fig. 45a**) had an OWS classification accuracy of 93.5%, which outperforms that produced by the generative images of the Con-GAN model (85.7%, **Fig. 45b**). Examples of the discrepancies between the models are illustrated in **Fig. *46* and Fig. *47*.**

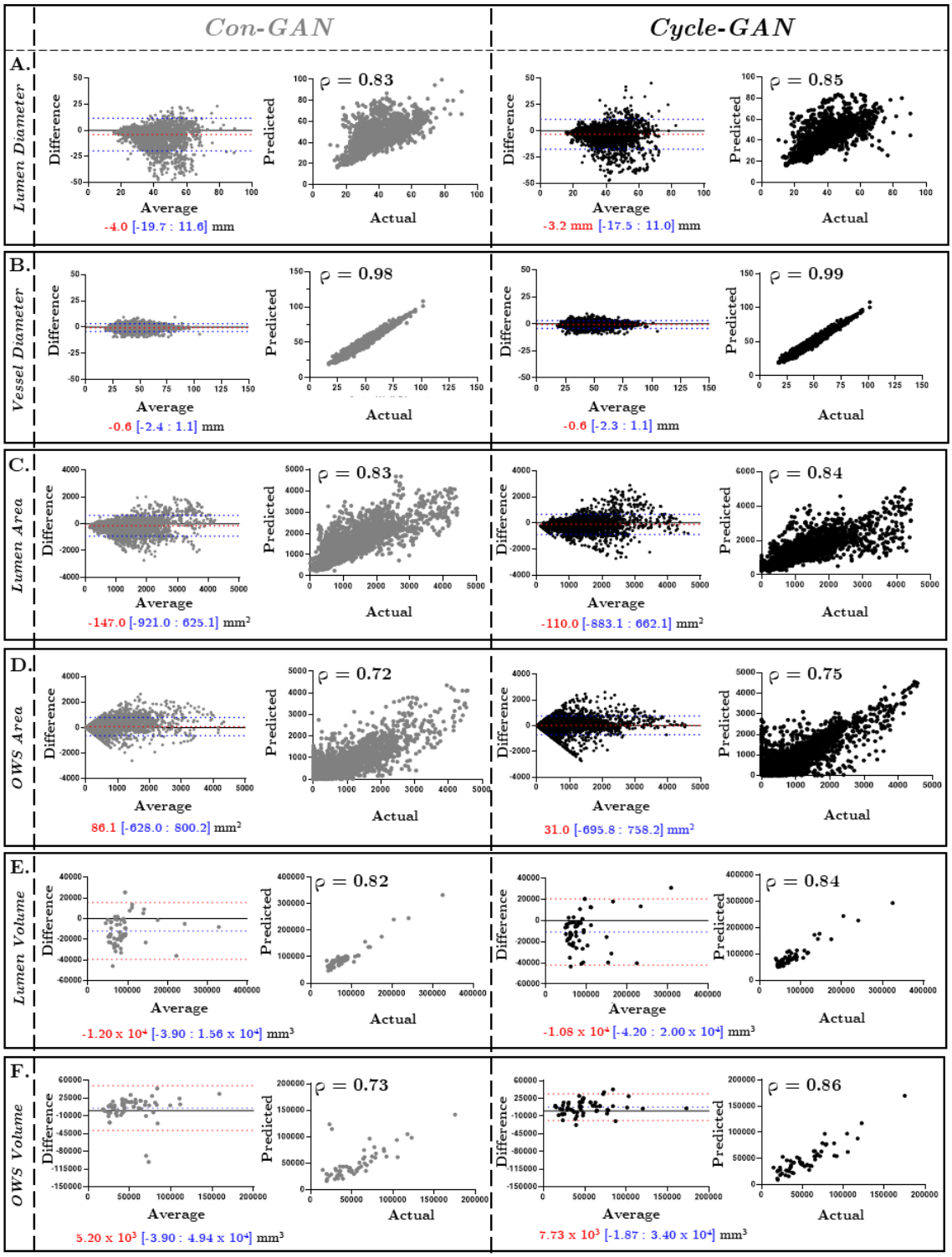
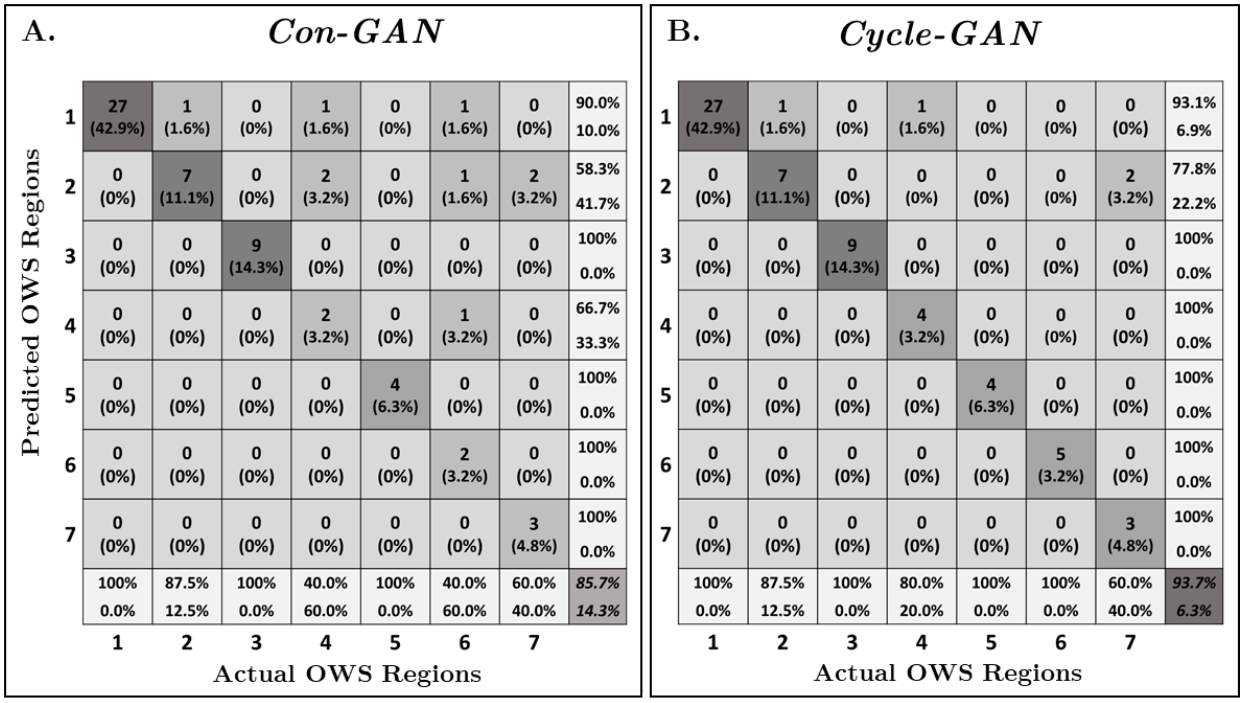


Fig. 44: One-dimensional diameter (A-B), two-dimensional area (C-D) and three-dimensional volume assessment (E-F) of generated images. Bland-Altman plots and correlation-coefficient analysis comparing the measurements of generated images compared against those derived from ground truth segmentations. Measurements derived from Con-GAN and Cycle-GAN outputs are illustrated in grey and black, respectively. Spearman correlation coefficients (ϱ) are indicated on the graphs ($p < 0.01$ for all comparisons)[11].



A. Con-GAN	Actual 1	Actual 2	Actual 3	Actual 4	Actual 5	Actual 6	Actual 7	
Predicted 1	27 (42.9%)	1 (1.6%)	0 (0%)	1 (1.6%)	0 (0%)	1 (1.6%)	0 (0%)	90.0% 10.0%
Predicted 2	0 (0%)	7 (11.1%)	0 (0%)	2 (3.2%)	0 (0%)	1 (1.6%)	2 (3.2%)	58.3% 41.7%
Predicted 3	0 (0%)	0 (0%)	9 (14.3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	100% 0.0%
Predicted 4	0 (0%)	0 (0%)	0 (0%)	2 (3.2%)	0 (0%)	1 (3.2%)	0 (0%)	66.7% 33.3%
Predicted 5	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (6.3%)	0 (0%)	0 (0%)	100% 0.0%
Predicted 6	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (3.2%)	0 (0%)	100% 0.0%
Predicted 7	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (4.8%)	100% 0.0%
	100% 0.0%	87.5% 12.5%	100% 0.0%	40.0% 60.0%	100% 0.0%	40.0% 60.0%	60.0% 40.0%	*85.7%* *14.3%*

B. Cycle-GAN	Actual 1	Actual 2	Actual 3	Actual 4	Actual 5	Actual 6	Actual 7	
Predicted 1	27 (42.9%)	1 (1.6%)	0 (0%)	1 (1.6%)	0 (0%)	0 (0%)	0 (0%)	93.1% 6.9%
Predicted 2	0 (0%)	7 (11.1%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (3.2%)	77.8% 22.2%
Predicted 3	0 (0%)	0 (0%)	9 (14.3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	100% 0.0%
Predicted 4	0 (0%)	0 (0%)	0 (0%)	4 (3.2%)	0 (0%)	0 (0%)	0 (0%)	100% 0.0%
Predicted 5	0 (0%)	0 (0%)	0 (0%)	0 (0%)	4 (6.3%)	0 (0%)	0 (0%)	100% 0.0%
Predicted 6	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	5 (3.2%)	0 (0%)	100% 0.0%
Predicted 7	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	3 (4.8%)	100% 0.0%
	100% 0.0%	87.5% 12.5%	100% 0.0%	80.0% 20.0%	100% 0.0%	100% 0.0%	60.0% 40.0%	*93.7%* *6.3%*

Fig. 45: Confusion Matrices comparing OWS regional classifications between generated images and ground truth segmentations. OWS segmentations derived from Model Outputs (Con-GAN(**A**), and Cycle-GAN(**B**)) were classified into seven categories as per Fig. *35* and evaluated against that determined by the ground truth segmentations[11].

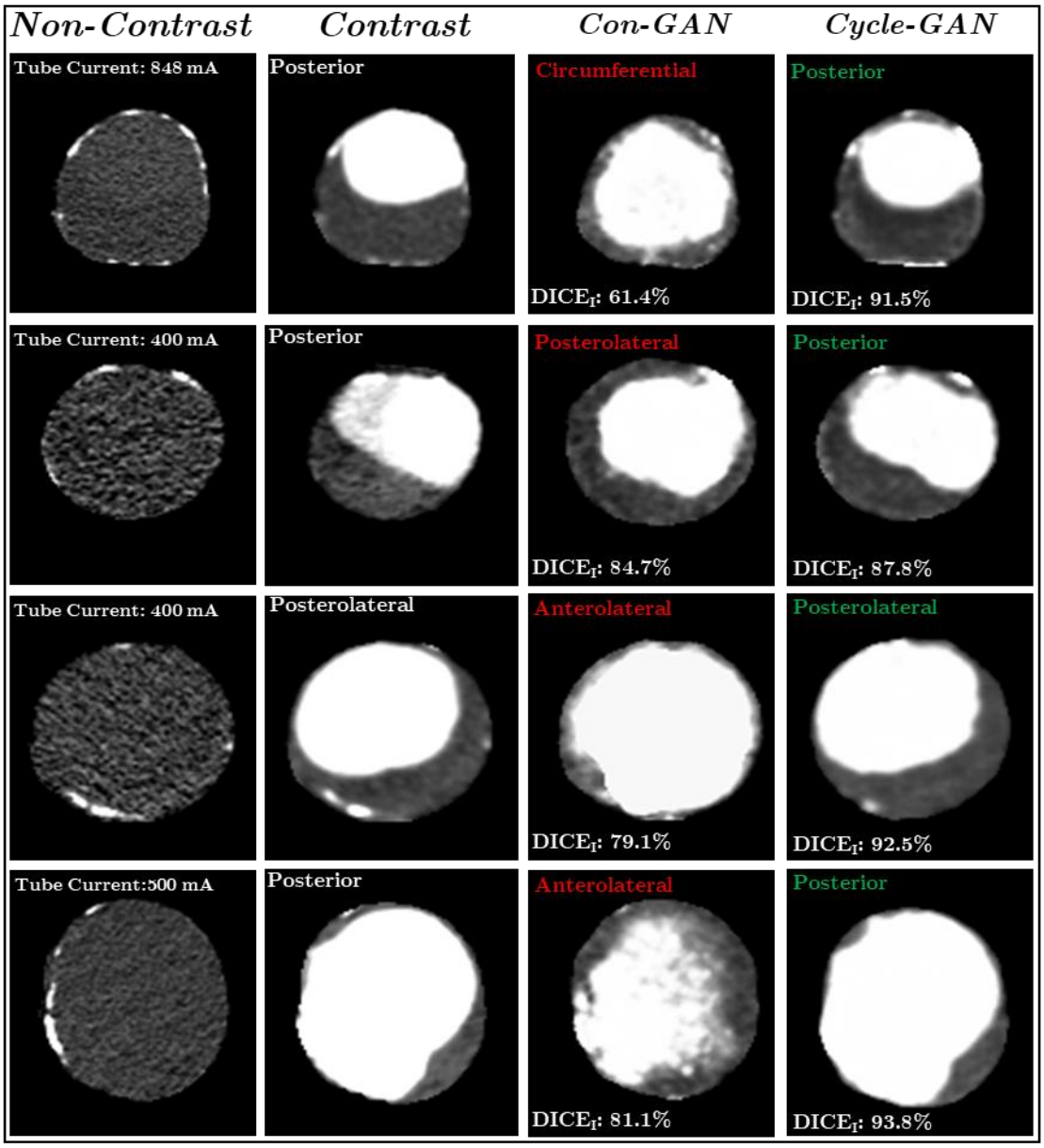


Fig. 46: Points of discrepancy between the Con-GAN and Cycle-GAN Models. Axial images of four patients within the OxAAA-IMG-I dataset are displayed where the Con-GAN fails to properly classify the OWS. However, the Cycle-GAN is able to properly classify the OWS. Corresponding x-ray tube currents and DICE scores are highlighted within each panel[11].

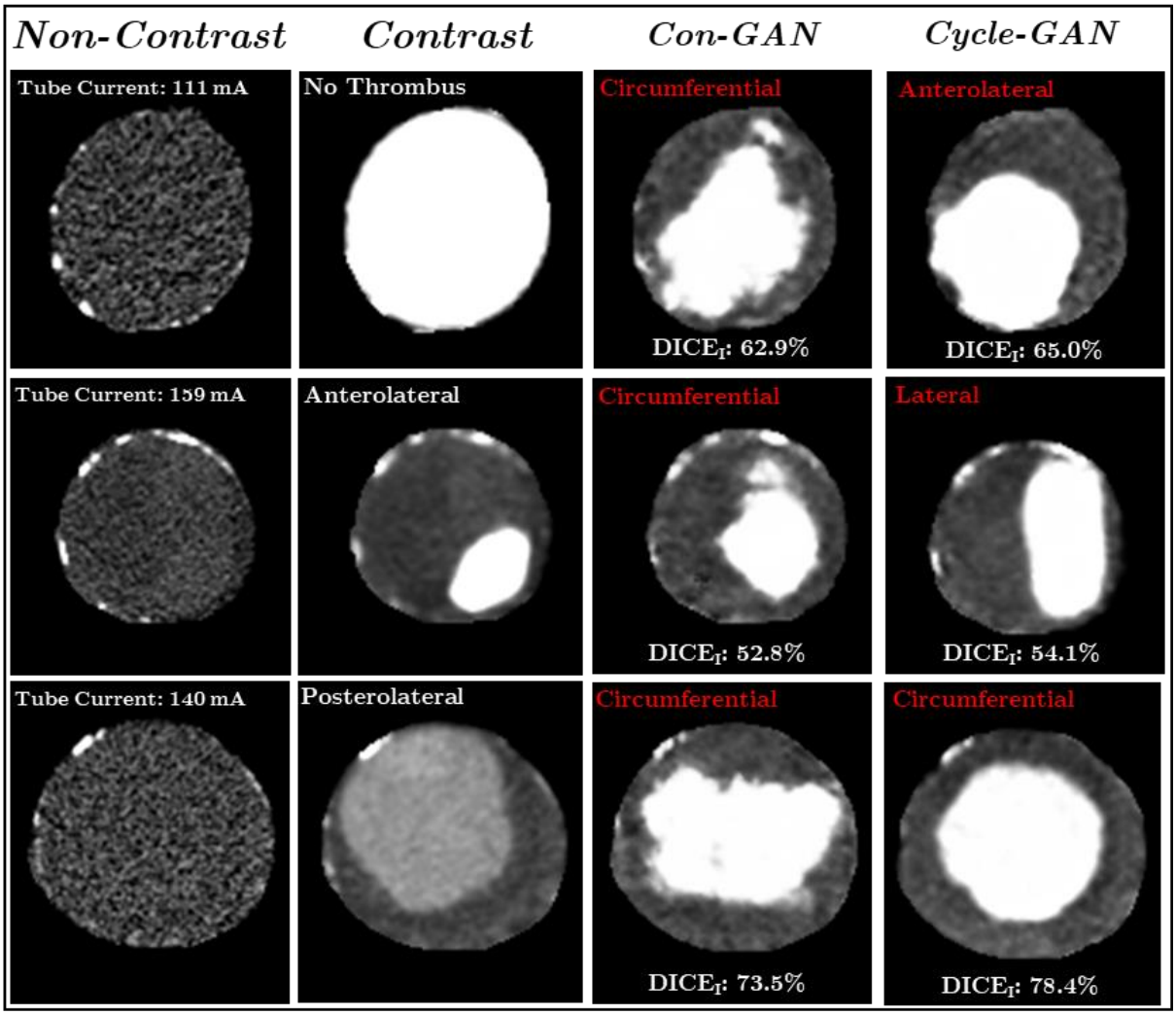


Fig. 47: Failure of both generative models. Axial images of three patients within the OxAAA-IMG-I datasetare displayed where both models fail to properly classify the OWS. Corresponding tube currents and DICE scores are highlighted within each panel[11].

5.3.6 Similar GAN performance is observed within the validation cohort (OxAAA-IMG-II cohort).

Of the 200 independent cases, 35 were subsequently excluded as they were obtained at tube currents <80 mA (n=25) or at unknown tube currents (n=10). RMSE between the generated Pseudo-contrast and CECT images for the 165 patients (Cycle-GAN: 4.2 ± 3.8 and Con-GAN: 4.2 ± 3.7) were similar to that observed in the training set. DICE accuracy of the inner lumen again showed superior performance of the Cycle-GAN (84.1 ± 7.2%) when compared against that of the Con-GAN (83.2 ± 7.7%). Extracted 2- and 3-D measurements from this refined testing cohort further supported the finding that the Cycle-GAN is superior at simulating CECT images of the AAA.

As a post-hoc analysis, we compared the GAN performance between the excluded cohort (n=35) and the rest (n=165). Area and volume measurements of the AAA lumen/thrombus derived from NCCT images obtained at >80 mA were closer to measurements derived from GT annotations than the excluded cohort ($p < 0.001$ for all comparisons, **Table *13a-b*).** Additionally, *Pseudo-contrast* images within the aneurysmal region produced by Cycle-GAN had an OWS classification accuracy of 90.6%, which outperforms the one produced by the generative images of the Con-GAN model (83.8%). This further supports the use of tube currents as a criterion to assess the quality of DICOM data for analysis.

Table 13: Ground truth CECT images vs. generated CECT images of the validation cohort. Of the 200 cases, 35 met the exclusion criteria (tube current < 80 mA) as described in the manuscript[11].

	Ground-Truth vs Generated CTA	A. Conditional-GAN			B. Cycle-GAN			C. Cycle-GAN (expanded Aortic ROI)		
		Excluded (n=35)	Included (n=165)	P value	Excluded (n=35)	Included (n=165)	P value	Excluded (n=35)	Included (n=165)	P value
	DICE (%)	70.0 ± 18.3 %	83.2 ± 7.7 %	<0.001	70.4 ± 20.9 %	84.1 ± 7.2 %	<0.001	78.8 ± 14.9 %	84.6 ± 9.0 %	**0.01**
Root Mean Square Error (RMSE)	Pixel by Pixel comparison	4.5 ± 3.8	4.2 ± 3.7	0.41	4.9 ± 4.2	4.2 ± 3.8	0.32	4.3 ± 2.5	4.1 ± 2.3	0.19
	Max Lumen Diameter (mm)	12.4 ± 8.5	5.2 ± 6.2	<0.001	10.4 ± 7.5	4.6 ± 6.7	<0.001	9.9 ± 10.1	4.6 ± 5.5	**<0.01**
	Max Outer Diameter (mm)	1.8 ± 1.7	1.5 ± 1.6	0.46	1.2 ± 1.6	1.1 ± 1.7	0.56	1.5 ± 1.2	1.2 ± 1.0	0.46
%-Difference	Lumen Area	51.2 ± 24.6 %	25.1 ± 15.2 %	<0.001	41.2 ± 33.9 %	24.1 ± 13.8 %	<0.001	46.2 ± 35.9 %	22.1 ± 15.9 %	**<0.001**
	Lumen Volume	45.6 ± 29.9 %	22.1 ± 18.6 %	<0.001	33.3 ± 26.9 %	19.1 ± 17.8 %	<0.001	34.3 ± 24.6 %	17.1 ± 15.8 %	**<0.001**
	ILT Volume	36.2 ± 22.2 %	26.2 ± 24.0 %	<0.001	26.0 ± 28.2 %	22.4 ± 23.4 %	<0.001	28.0 ± 30.2 %	19.9 ± 16.9 %	**<0.001**

The role of aneurysmal size (maximum AAA diameter) and shape (NFI) was compared against Cycle-GAN performance. **Fig. *48*a-d** illustrates four aneurysms alongside their respective 'ideal' fusiform shape. Aneurysms with a predominantly fusiform shape had lower NFIs compared to those aneurysms with a predominantly saccular shape. This supports the classification of aneurysm shape using NFI. Both AAA size and NFI had no significant impact on transformation accuracy as assessed by DICE score of the inner lumen from the Cycle-GAN (**Fig. *48*e**). This suggests that the trained Cycle-GAN do not have an implicit AAA size or shape preference.

5.3.7 Cycle-GAN can simulate extra-aortic/AAA features including aortic side branches.

Given that the Cycle-GAN generated superior results against that of the Con-GAN, we trained a Cycle-GAN to identify an expanded ROI surrounding the aorta. This concurrently simulates both intra- (aortic lumen, thrombus morphology) and extra-aortic/AAA features (major abdominal aortic side branches). Despite the expanded input data size (i.e. more information to 'learn' from), its ability to extract intra-aortic/aneurysmal features are comparable to the previously trained models (smaller ROI / less information to 'learn' from) (**Table *13*c). Fig. *49*** illustrates generated pseudo-contrast images alongside their respective CECT and NCCT images for four patients within the OxAAA-IMG-II dataset. Branch arteries arising from the aorta are visible in all patients. **Table *14*** highlights capture rate and maximum diameters for each branch, obtained in a blinded fashion, from pseudo-contrast CT and its corresponding CECT within the OxAAA-IMG-II dataset. There is no difference between the diameter for all six major aortic side branches measured using the CECT (GT) or the pseudo-contrast CT (derived from the NCCT images).

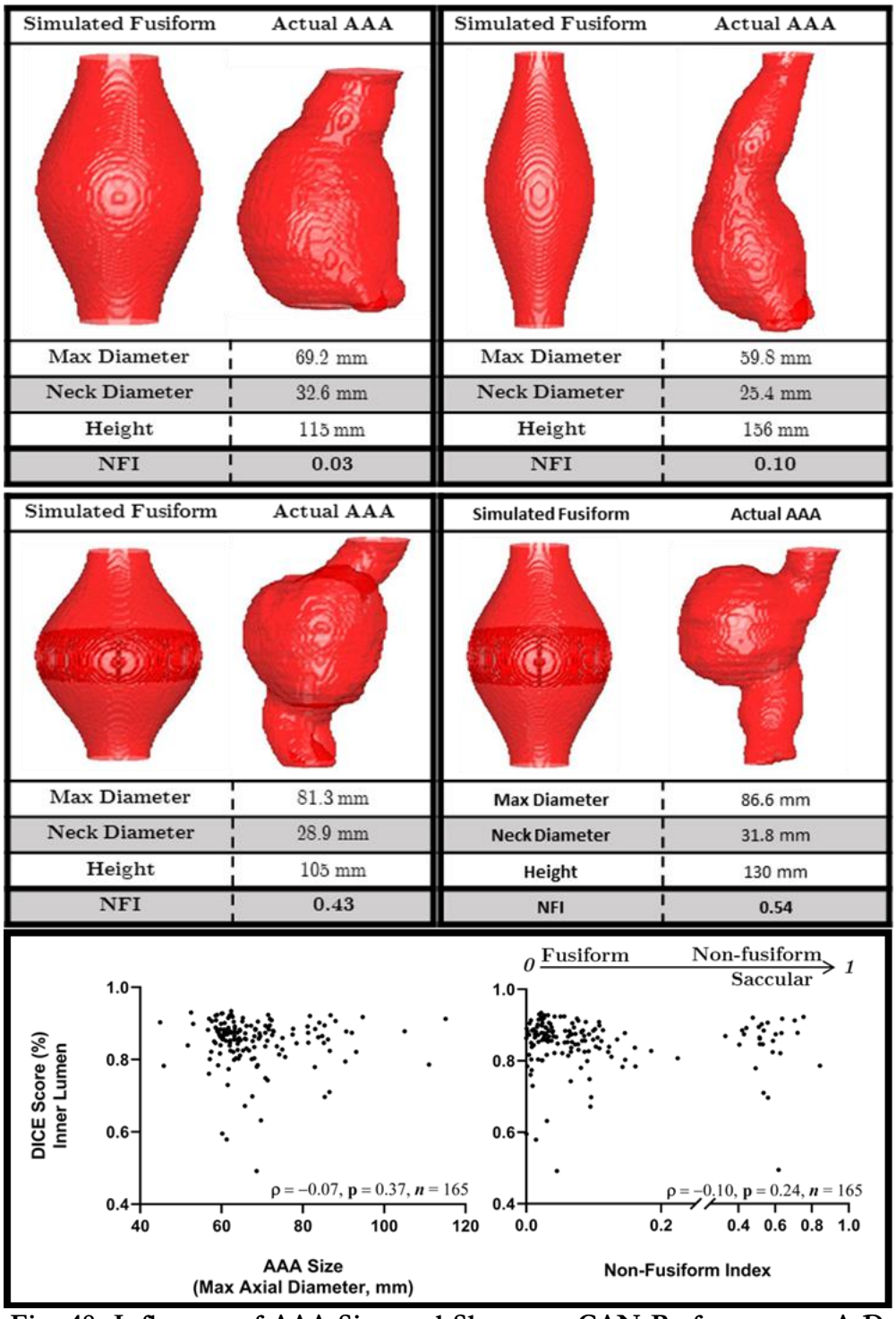


Fig. 48: Influence of AAA Size and Shape on GAN Performance. A-D: Four abdominal aortic aneurysms from the validation cohort alongside their respective idealized fusiform models, generated using patient-specific AAA maximum/neck diameters and height of aneurysmal sac. NFIs, a surrogate measure of AAA shape, are displayed for the 4 patients. **E.** Impact of AAA size (Max Diameter) and shape (NFI) on Cycle-GAN transformation accuracy (DICE score of Inner Lumen) on the OxAAA-IMG-I dataset[11].

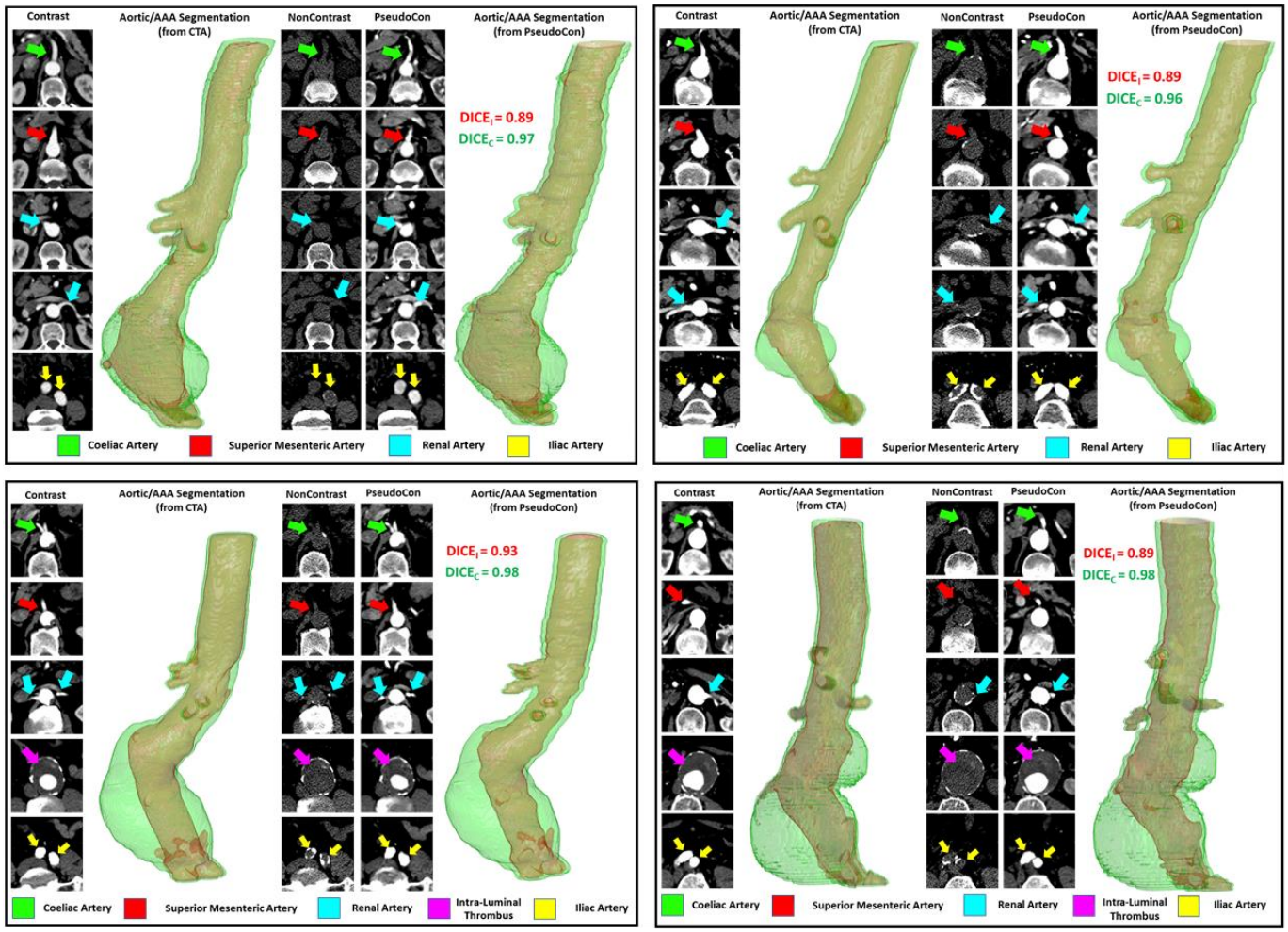


Fig. 49: Pseudo-Contrast CT images are displayed alongside their respective Contrast and Non-Contrast CT Images for four patients. Arrows indicating the branch arteries (Celiac artery, Superior Mesenteric Artery, Renal arteries, and Iliac arteries) and the intraluminal thrombus are highlighted wherever visible. Aortic segmentations including the side branches are generated from both the CECT (gold standard) and the Pseudo-Contrast CT and DICE scores comparing the segmentations are displayed[11].

Table 14: Diameter measurements of aortic side branches from CECT and Pseudo-contrast CT images.

Total (n = 165)	Celiac Artery	Sup. Mesenteric Artery	Left Renal Artery	Right Renal Artery	Left Iliac Artery	Right Iliac Artery
Branch Capture Rate	163/165	164/165	159/165	161/165	163/165	162/165
CTA	8.2 ± 1.3 mm	8.5 ± 1.5 mm	6.8 ± 1.3 mm	6.7 ± 1.3 mm	14.9 ± 2.6 mm	14.8 ± 2.4 mm
Pseudo-Contrast	8.3 ± 1.2 mm	8.4 ± 1.2 mm	6.7 ± 1.1 mm	6.6 ± 1.0 mm	15.0 ± 2.5 mm	15.1 ± 2.5 mm
MAE	0.8 ± 0.5 mm	0.9 ± 0.5 mm	0.7 ± 0.6 mm	0.7 ± 0.5 mm	0.8 ± 0.5 mm	0.9 ± 0.4 mm
P	0.69	0.21	0.36	0.31	0.52	0.56

5.4 Discussion

The primary objective of this study was to investigate whether there are subtle differences between visually indistinguishable regions within the NCCT image. This was required to ensure that the images contained the necessary information for the DL method to generate anatomically-correct visualisations. This was achieved by comparing the HU intensity distributions between regions in the thrombus-filled aneurysm. Visually, axial slices within the AAA appear uniform on the NCCT image and the histograms for each of the regions (lumen, interface and thrombus) display considerable overlap. However, average HU intensity was significantly different between all three regions for all patients assessed and a gradual HU change was observed from the aortic lumen to the intra-luminal thrombus. This highlights that there are differences, albeit subtle, between the regions that can be exploited and enhanced to estimate the CECT image.

In addition to average HU intensity, the differences between these visually in-distinct regions can be captured using $1^{st}/2^{nd}$-order radiomic features. Radiomics employs advanced data-characterisation algorithms to extract underlying pixel relationships and has been used to uncover potential disease features that fail to be appreciated visually. Random forest models incorporating these radiomic features were used to accurately classify sub-volumes within the aneurysmal volume into either the aortic inner lumen and OWS, which may contain the intra-luminal thrombus. A 3-fold method was used to maximize the images used for model evaluation as the testing cohort for each fold consisted of sub-volumes from a unique set of 25 patients. Models were trained/optimized using the remaining 50 patients. The most important features for this classification task were highlighted in Table *12*. These features consist of a combination of 1^{st} and 2^{nd} order features on

either the original or filtered NCCT sub-volume. The filters used during radiomic feature extraction consisted of 1. Laplacian of Gaussian (edge detection filter) and 2. Wavelet modifications. These are commonly used methods to diversify the extracted radiomic features. There was strong overlap between these features. Crude/down-sampled segmentations (resolution – 5mm) from these classification outputs were compared against that from the ground truth classifications using DICE score of the inner lumen. The results suggest that radiomic differences between inner lumen, and OWS in NCCT images can be used to visualize AAA regions and supports the validity of this image transformation task. The generative networks likely utilise this higher-order information to enhance its performance.

The study's secondary objective was to investigate if DL models (Cycle- and Con-GAN) could extract the subtle differences between soft tissue components in NCCT images within the context of AAA disease and generate CECT images. For both GANs, a 3-fold cross-validation approach was employed during training. There was no data leakage between cohorts and the testing cohort for each fold contained a unique set of patients. The 2-D training/testing data was obtained by first isolating the aorta, axial sub-sampling within the volume, and centring the aorta within the extracted 2-D slice. Isolating the aorta by removing surrounding organs and tissues reduces the amount of noise presented to the model. Additionally, centring the aorta in each axial 2D slice reduces the spatial variation of the aorta seen by the generative models. This theoretically should maximise the information learned by the GAN networks for the NCCT to CECT image transformation task. RMSE, a quantitative measure of image difference, and DICE scores for the inner lumen and combined aortic mask were used to optimise the training parameters. DICE accuracy of the inner lumen is an appropriate measure of transformation accuracy as it evaluates the primary goal of the generative models.

It was apparent that CT image quality determines transformation accuracy. Exploring the $DICE_I$ metric within the testing cohort highlighted certain patients with not only decreased transformation accuracies but also greater $DICE_I$ variability within the aortic volume. Here, we hypothesised that poor quality NCCT images have correspondingly poor CECT reconstruction accuracies. Image quality and the extent of distortion ('signal to noise' ratio) of NCCT images is directly related to the number of x-ray photons incident within the target volume[104]. Accordingly, the number of x-ray photons is directly related to the x-ray tube current (I_{tube}),

which is the rate of photon production within the x-ray tube, and CT rotation time. Given that CT rotation time is usually constant, we reasoned that images obtained with decreased I_{tube} have an increased likelihood of generating CECT images with decreased accuracy. Therefore, a stringent selection criterion was imposed to exclude potential outliers with I_{tube} below the 15^{th} percentile (<80mA). Future studies investigating the role of CT acquisition parameters (ex. slice thickness, kVP, tube current) for this transformation task in greater detail is required to determine the limitations of this technique.

An unexpected observation was that cycle-GAN outperformed the con-GAN. In the original context, we hypothesised that the Con-GAN would have superior performance to the Cycle-GAN as it is able to learn pixel-to-pixel transformations between image pairs. (This is not the case for the Cycle-GAN as data is not introduced as paired images.) This superiority was unexpected but can be rationalised to its underlying network architecture and the multiple training losses (ex. cycle-consistency and identity loss terms). This interesting observation sheds some unique philosophical insight into the nature of 'learning' by neural networks and its similarities of human learning for this particular task. For a cycle-GAN, the neural network learns heuristically from all images of the paired ground truth data, without being constrained to specific pixel relationships as that of a Con-GAN. It is possible, however, this observation is specific to the dataset utilised here.

We showed that these generative models enable the visualisation of aortic aneurysm morphology in CT scans obtained without the use of intravenous contrast and that transformation accuracy is independent of AAA size or shape. Extracting diameter measurements is required in AAA management as it guides the frequency of aneurysm surveillance and determines the timing of surgery[79,80]. 2-D cross-sectional area measurements of the aneurysm have been shown to complement the 1-D diameter measurements as diameter measurements can be subject to substantial interobserver variability and at times may fail to represent the 3-D growth of the aneurysm. As a result, cross-sectional area measurements have been shown to have the lowest variability in assessing aneurysm size [81]. Evolution of 3-D indices, especially thrombus volume, is linked to AAA progression, rupture risk and even the incidence of adverse cardiovascular events[82,83]. Assessing thrombus/OWS spatial morphology is important for surgical planning and has been shown to influence postoperative outcomes. We have

previously reported that the spatial morphology of native intra-luminal thrombus correlates with the onset of type 2 endoleak, which is an adverse outcome following endovascular surgical repair of aneurysms[95]. This is a task that is not possible from the original NCCT image and if achieved, reinforces the clinical impact of using generative networks for this image transformation task. For the first time, the results highlight the ability to assess the thrombus regional location from the NCCT with high accuracy.

In addition, we showed that a Cycle-GAN trained on 135*135 mm (144 x 144 pixels) ROIs surrounding the aorta is able to robustly visualise not only aortic/AAA morphology but also extra-aortic structures, including its side branches. The accuracy of the side branch visualisation was assessed by measuring the maximum diameter of each branch from the descending/abdominal aorta. The results support the ability to capture this measurement with a mean average error of ~1.5 mm (1-2 pixels). Using GANs for pseudo-contrast CT visualisation of the aorta/AAA, its intra-luminal thrombus as well as its side branches from a NCCT is a novel technique and presents clinicians with a safer alternative to the routinely obtained contrast-enhanced CECT. Future studies are required to determine its clinical utility, such as using this alternative imaging method to plan for endovascular grafting.

This deep learning approach described here can also be applied to reconstruct other anatomical structures (veins, solid organs, etc.) without the need for contrast administration. Beyond the potential clinical utility, our method can be applied for research analysis. There is a growing body of literature on the role of the intraluminal thrombus and abdominal aortic aneurysm growth[83,105,106]. The ability to characterise the OWS using non-contrast-enhanced CT scans greatly expands the scope for research in this topic, without the need to subject research participants to contrast injection.

5.5 Conclusion

Here, we present and internally validate a deep learning approach to visualise aortic aneurysm morphology without the use of intravenous iodinated contrast agents. This experiment was possible as paired imaging (NCCT and CECT) is commonly obtained from AAA patients during surgical planning. We implemented a non-rigid registration algorithm to register/align the CECT to the reference NCCT image. Initial proof-of-concept experiments on these paired and

registered images highlighted differences in radiomic signature between different regions within the NCCT image (ex. OWS and aortic lumen). These differences were used to simultaneously train/optimize machine-learning (ex. random forest, Section 5.3.2) and deep learning (ex. Generative Adversarial networks, Section 5.3.3) networks for this non-contrast to contrast image transformation task. Multiple DL generative networks were compared (Con-GAN vs Cycle-GAN) using both technical (ex. RMSE between the generated and GT image) and clinical metrics. Both generative networks were able to simulate acceptable contrast-enhanced CT images from a NCCT input, however the Cycle-GAN displayed superior results from both a technical and clinical perspective. The random forest machine learning model highlighted an alternative method to achieve similar results and allowed for model interpretability. Subsequently, we identified the importance of image/scanner parameter settings (ex. tube current) for this image transformation task and set the foundation for more comprehensive experiments. Finally, we showed that these methods are able to transform not only the aorta/AAA but also its major aortic side branches. This platform technique can be applied to different anatomical structures for research and eventually clinical applications.

5.5.1 Limitations

There are few limitations within this study that need to be addressed to maximize the potential of this disruptive technology. Firstly, this study is a retrospective study using de-identified pre-surgical images obtained from AAA patients. These images were obtained between 2009 – 2018 and all clinical information was stripped from the DICOM headers prior to storage. Given the lack of methods standardization with regards to acquisition, all images were obtained using a variety of image/parameter settings. In this study, we show the importance of tube current in image transformation quality. We claim that, like tube current, there are other image/scanner settings that need to be optimized to ensure appropriate and acceptable non-contrast to contrast image transformation. Additionally, the algorithms trained and validated in this chapter incorporate imaging data obtained from a single center. There is poor representation of multiple scanner models/types. Although these scanners are meant to be standardized and the outputs are meant to be comparable, there are intrinsic scanner differences that distinguish the outputs and is an active area of consideration when testing DL generative model robustness or generalizability.

In addition to heterogeneity associated with image/parameter settings, there is also intrinsic patient variability that needs to be taken into consideration. For example, it was apparent that patients with different body-mass-indices (BMI) had different levels of fat content within the abdomen, which may impact the image quality of the aorta within the NCCT image. This is an active area of inquiry that could not be comprehensively tackled due to the retrospective nature of this study but needs to be clarified prior to clinical implementation.

Another limitation of this current technique is the 2-D nature of the DL generative network. In this study, the 3D-CT image was sampled axially, and the axial images were used for model training, optimization and validation. Axial views were selected for this analysis, instead of the coronal or sagittal views, as the aorta/AAA displays minimal slice-to-slice variability (ex. variation of a circle/oval). Axial slices were introduced into the DL network randomly and independently, without any connection to adjacent slices. This significantly increased the dataset size at the expense of losing inter-slice contextual information. Future studies involve introducing multiple views or developing a 3-D generative algorithm for cross-modality medical image synthesis.

6
Prediction of AAA growth in humans using morphological features

Contents

6.1 Introduction 139

6.2 Methods 141

6.2.1 Patient Cohort 141

6.2.2 Automated Segmentation of CT images 143

6.2.3 Hypothesis-driven Feature Extraction of AAA Volumes 143

6.2.3.1 Maximal Anteroposterior Diameter (APD) 143

6.2.3.2 Radius of Curvature (RC) 143

6.2.3.3 Undulation Index 144

6.2.4 Semi-supervised Feature Extraction from Aortic Aneurysms 144

6.2.4.1 Reference Axis Reorientation prior to Feature Extraction (PCA) 145

6.2.4.2 Feature Extraction: Characterization of the aortic/AAA Shape 146

6.2.4.3 Feature Extraction: Characterization of the Aortic/AAA Surface 146

6.2.4.4 Feature Extraction: Characterization of Aortic Side Branch Origins 148

6.2.4.5 Feature Extraction: Principal Component Analysis 150

6.2.5 Developing the growth prediction models using both hypothesis-driven and PCA-derived features 150

6.2.6 Evaluation of the growth prediction models on an independent validation cohort (MA3RS Study) 151

6.3 Results 152

6.3.1 Patient demographic features do not influence the geometric features of AAA 152

6.3.2 Hypothesis-driven Geometric Feature Extraction from the OxAAA Growth Prediction Cohort 152

6.3.3 Semi-Supervised Geometric Feature Extraction from the OxAAA Growth Prediction Cohort 154

6.3.4 Prediction of AAA growth as a categorical outcome 156

6.3.4.1 Hypothesis-Driven Geometric Features predict AAA growth 156

6.3.4.2 PCA-derived Geometric Features improve AAA growth prediction 157

6.3.4.3 Refinement of the individual prediction model for fast/slow growth 159

6.3.4.4 Examination of PCA features provides insight into AAA growth 161

6.3.5 Prediction of AAA growth rate as a continuous variable 167

6.3.6 Similar AAA Growth predictive accuracy is observed within the independent validation cohort (MA3RS Study) 170

6.3.6.1 Aneurysmal growth within the MA3RS study dataset 170

6.3.6.2 Geometric feature extraction within the MA3RS study cohort 170

6.3.6.3 Validation of growth prediction models within the MA3RS dataset 173

6.4 Discussion 177

6.5 Conclusion 180

6.5.1 Limitations 181

6.1 Introduction

Abdominal Aortic Aneurysms (AAA) are an abnormal degenerative condition characterized by pathological dilatations of the abdominal region of the aorta. Clinically, an AAA is defined when the aortic diameter is >50% of the healthy aorta adjacent to the aneurysm. The natural history of an untreated AAA consists of progressive dilatation with eventual rupture and death. The clinical management of AAAs consist of screening/diagnosis, regular surveillance and timely surgical intervention by open surgical repair or endovascular stent grafting[79,107].

Methods for the prediction of AAA growth are considered as a priority for research in the opinions of vascular and endovascular surgeons[108]. Previous investigations have focused on predicting AAA rupture risk. Various biomechanical analysis using geometric features[109,110] (different diameter measurements), aortic tortuosity[111], morphological parameters[101] and presence/extent of intra-luminal thrombus (ILT)[112,113] have been investigated to predict rupture risk potential. Although these studies provide insight into the biomechanical stress profiles of AAAs, they are complex and present large degrees of uncertainty that prevent clinical implementation. Furthermore, with the adoption of aneurysmal screening programs, public awareness, and advanced technology for AAA detection/surveillance most AAA patients are monitored and treated well before the point of rupture[114]. Accurate prediction of AAA growth in patients can allow for the optimization of surveillance intervals and better inform the timing for surgery. AAA size (maximum AAA diameter), which is commonly obtained in the clinical setting, has been shown repeatedly to be a poor independent indicator of aneurysmal growth or rupture potential[115]. Our prior work has highlighted the feasibility of AAA growth prediction using physiological and biochemical measurements obtained from patients[116-118]. These measurements, however, require additional research steps to the routine clinical care pathway.

Computerised tomography (CT) scans are utilized extensively as diagnostic tests in medicine and surgery. Globally around 150 million CT scans are performed each year[119,120]. In the management pathway of AAAs, each patient requires one dedicated CT scan prior to surgery to plan for the operative approach. During the small AAA surveillance period, a proportion of these patients would have also undertaken CT scan(s) for other medical reasons. In these patients, these historic CT scan(s) performed sometime prior to AAA surgery could serve as the

'baseline' scan. The subsequent pre-operative CT scan would serve as a serial scan and enables growth rate calculation during the time period between the two. With relevant regulatory approval, these historic baseline CT scans can be anonymised and utilised to discover novel features to predict future AAA growth. Machine/deep learning methods developed in Chapters 4 and 5 enable the automatic extraction of the AAA shape and set the foundation for this morphological analysis.

As AAAs enlarge, a variety of geometrical changes are observed including altered aortic tortuosity[121] and increased aneurysmal asymmetry[122]. Several of these changes result in a unique non-uniform distribution of wall stress and have been hypothesized to either favour AAA growth deceleration or increase rupture risk[123,124]. However, there is no prior literature on the prediction of AAA growth using complex geometric features extracted from CTs. This has been examined in the context of cerebral aneurysms. Dhar et al., investigated the use of six image-based morphological features to predict intracranial aneurysm rupture[125]. This study documented the importance of vascular geometry in altering blood flow dynamics and promoting intracranial aneurysmal rupture[125-127]. Although there are many similarities between intracranial and aortic aneurysms, these features have not been investigated in the latter. One such parameter involved in predicting cerebral aneurysm rupture was undulation index (UI), which captures the degree of surface concavity and increases with surface irregularity[125,126]. A highly asymmetric and/or tortuous infra-renal region would result in an increased UI parameter. In this regard, regions of increased curvature (radius of curvature, RC) within the aneurysm contributes to non-laminar fluid flow, and non-uniform wall shear stress[128,129]. No prior literature has examined the utility of undulation index (UI) or radius of curvature (RC) for AAA growth prediction.

Previous studies published through the prospective OxAAA study suggest that the intra-luminal thrombus (ILT), which is unique to AAAs and present in >90% of cases, is biologically active and plays a critical role in promoting AAA growth and progression[63]. Hence, we further hypothesized that aneurysms with local regions of increased curvature and an increase degree of surface undulation could indirectly affect its growth through the interaction between the altered hemodynamic forces and the biologically active intra-luminal thrombus. In addition to this

hypothesis-driven approach (UI and RC), we investigated if a semi-supervised method of AAA shape characterization could be used to supplement prediction models.

6.2 Methods

6.2.1 Patient Cohort

The study was conducted as part of the ongoing Oxford Abdominal Aortic Aneurysm (OxAAA) study (Ethics approval Ref: SC/0250/13). Details regarding the OxAAA study cohorts used in this chapter are highlighted in Section 3.1. This study cohort consists of two arms.

The first arm consisted of OxAAA-IMG-III cohort and included participants (n = 102) from the larger OxAAA-IMG cohort that were prospectively recruited to the prospective OxAAA study at the time of surgery. Demographic information from each patient were matched to the pre-operative CT scan. Height and weight of the patient were measured to calculate their body mass index. The history of coronary artery disease (CAD) is defined as symptoms angina +/- previous coronary interventions (angioplasty or bypass). The history of peripheral arterial occlusive disease (PAOD) is defined as symptoms of intermittent claudication +/- previous lower limb arterial intervention (angioplasty or bypass). The history of cerebral arterial disease is defined by transient ischaemic attack or stroke. The history of hypertension, diabetes mellitus, and hypercholesterolaemia were as diagnosed by the primary care/general practitioner. We further measured the individual's blood pressure, blood cholesterol profile and HbA1c level to ascertain the effect of their pharmacological therapy. Their current medications were further recorded (**Table 15**).

The second arm consisted of the OxAAA-IMG-IV cohort, which included participants (n = 192) within the larger OxAAA-IMG cohort with aneurysmal growth measurments. Here, the clinical database (Oxnet Janus), which prospectively registered every patient who underwent elective repair of AAAs at the John Radcliffe Hospital (Oxford, UK) was used. Each of the patients gave written consent for the utilization of clinical images collected during the routine clinical management pathway for research analysis. Radiology records of all non-emergency infra-renal AAA repairs (open surgery or endovascular repair) from 1st of February 2009 to 30th June 2018 were examined. Only those patients with at least 1 historic CT scan conducted greater than 8 months prior to the pre-operative scan were included for analyses.

Table 15: Summary of participant demographics (OxAAA-IMG-III) at the pre-surgical assessment and significance of spearman correlation with extracted geometric parameters[12].

	All Participants (n = 102)	Significance of Spearman Correlation Coefficient		
		Diam.	UI	RC
Male (%)	99(97)	0.35	0.36	0.09
Age at Consent (Median/IQR)	72 (67-79)	0.12	0.83	0.70
Height (±SD)	1.75 ± 0.08	0.07	0.45	0.06
Weight (Median/IQR)	81.9 (74-90.2)	0.17	0.15	0.07
BMI (Median/IQR)	26.8 (24.3-28.7)	0.61	0.35	0.18
MAP (±SD)	102.2 ± 12.8	0.13	0.70	0.20
Current Smoker (%)	24 (24)	0.57	0.99	0.09
Past Smoking Hx (%)	68 (67)	0.97	0.48	0.59
Never Smoked (%)	13 (13)	0.65	0.34	0.44
CAD Hx (%)	33 (32)	0.25	0.15	0.56
Coronary Intervention (%)	26 (25)	0.63	0.41	0.88
PAD History (%)	16 (16)	0.85	0.59	0.22
Cerebral Art. Disease (%)	12 (12)	0.10	0.16	0.34
HTN History (%)	74 (73)	0.53	0.77	1.00
Hypercholesterolemia (%)	61 (60)	0.61	0.67	0.75
Tot. Cholesterol (Median/IQR)	3.8 (3.2-4.6)	0.23	0.76	0.63
HDL (Median/IQR)	1.1 (0.9-1.3)	0.47	0.41	0.06
LDL (Median/IQR)	1.5 (0-2.5)	0.26	0.92	0.37
TG (Median/IQR)	1.2 (0.8-1.6)	0.49	0.65	0.66
Diabetes (%)	16 (16)	0.88	0.26	0.78
HbA1C (Median/IQR)	5.6 (5.4-5.9)	0.19	0.81	0.41
Diabetes - Oral/Insulin (%)	12 (12)	0.19	0.22	0.17
CKD - eGFR < 60 (%)	28 (27)	0.15	1.00	0.96
Creatinine (Median/IQR)	86.5 (73.3-101.3)	0.19	0.07	0.74
Beta-Blockers (%)	32 (31)	0.68	0.30	0.49
ACEI/A2RA (%)	56 (55)	0.27	0.77	0.97
Aspirin (%)	47 (46)	0.83	0.89	1.00
Thienopyridine (%)	9 (9)	0.55	0.17	0.44
Ticragrelor (%)	3 (3)	0.81	0.64	0.72
Anticoagulant (%)	12 (12)	0.93	0.46	0.21
CCBs (%)	43 (42)	0.13	0.64	0.34
Diuretics (%)	22 (22)	0.47	0.38	0.58
Gastro-restraint (%)	31 (30)	0.57	0.68	0.32
Steroids (%)	7 (7)	0.36	0.97	0.15
Statins (%)	74 (73)	0.57	0.76	0.24
AAA Diam (Median/IQR)	63 (58 - 72.5)			
UI (±SD)	0.23 ± 0.08			
Min. RC (Median/IQR)	35.9 (29.7 - 46.9)			

Note: Participant demographics were collected at the pre-surgical assessment and were correlated against the extracted geometric parameters of AAA diameter, undulation index, and radius of curvature. Characteristics that follow a Gaussian/Normal Distribution are indicated with a +. For such variables, mean ± SD are presented, and cohort differences are compared using a student t-test. For variables that don't follow a Gaussian distribution, median and inter-quartile range (IQR) are presented and cohort differences are compared using a Mann-Whitney test.

6.2.2 Automated Segmentation of CT images

AAA segmentation was performed using a proprietary automated deep learning segmentation platform[130]. The deep learning model, detailed in Chapter 4, generated aortic/aneurysm segmentations which were visually assessed against the source DICOM (Digital Imaging and Communications in Medicine standard) images. Where required, further manual adjustments were performed using the open source ITK-Snap software[131].

6.2.3 Hypothesis-driven Feature Extraction of AAA Volumes

The geometric features included in our analyses were derived from the entire AAA and can therefore be derived from either contrast or non-contrast enhanced CT images. All geometric features were extracted from the above derived segmentations using MATLAB. AAA size was measured by calculating the maximum anteroposterior diameter. Annual aneurysmal growth was derived by subtracting the historic/baseline AAA size from pre-operative AAA size and dividing the difference by the time duration (yrs) between scans (Annual Growth = Δ AAA size in mm /(number of days lapsed between scans / 365 days)).

6.2.3.1 Maximal Anteroposterior Diameter (APD)

Maximum APD was automatically extracted from the aneurysmal segmentations using MATLAB. For each axial slice along the aortic volume, the diameter was obtained by measuring the maximum distance between two points on the aneurysmal boundary in the sagittal plane. The maximum APD of the AAA was the maximum value from all axial slices.

6.2.3.2 Radius of Curvature (RC)

Radius of curvature is a centreline-based metric that captures the degree of curvature along the centerline[129]. Here, RC equals the radius of the circular arc that best approximates the curve between a set of adjacent points. The smaller the circular arc, the smaller the RC and the greater the local curvature (**Fig. *50a***). On the other hand, the larger the circular arc, the greater the RC, and the lower the local curvature (ex. a straight line). AAA centrelines were calculated using an implementation of the homotopic thinning algorithm[132] and were subsampled using b-spline interpolation methods based on the number of axial slices (Chapter 4.2.14). RC was calculated for adjacent sets of triplet points and the minimum value was obtained. This described the greatest region of curvature within the aneurysmal volume.

6.2.3.3 Undulation Index

Aneurysmal UI was defined as $UI = 1 - (\frac{V}{V_{CH}})$ [125,126]. Here, V is the volume of the infra-renal abdominal aorta defined as the region between the renal arteries and the iliac bifurcation and V_{CH} is the volume of its convex hull (**Fig. *50b***). In this instance, the convex hull of AAA is the smallest volume that encompasses the entire region and is convex at all points. It effectively resembles a plastic wrap attached to the inlet and stretched over the entire aneurysmal surface. This parameter captures the degree of surface concavity and increases with surface/shape irregularity. Conversely, a shape that is nonconcave (ex. a perfect sphere or cylinder) will have a UI of 0.

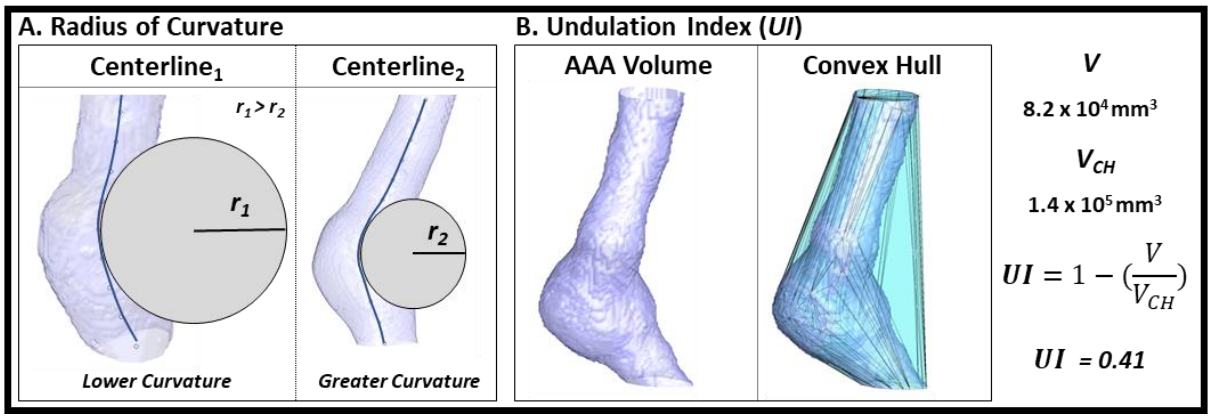


Fig. 50: Radius of Curvature (RC) (A) and undulation index (UI) (B) measurements obtained from the sample abdominal aortic aneurysm (AAA) shapes[12].

6.2.4 Semi-supervised Feature Extraction from Aortic Aneurysms

In addition to this hypothesis-driven approach, we investigated if principal component analysis could be used to characterize the aneurysmal shape and provide additional growth predictive capacity to that observed with the hypothesis-driven geometric features. Principal component analysis (PCA) is a commonly-used method used for dimensionality reduction. It is performed by projecting each data point onto a range of principal components to obtain lower-dimensional data while preserving as much of the data's variation as possible. In this instance, we hypothesized that most of the diversity or variance of the aneurysmal OWS can be captured using PCA and that the reduced principal components, which may describe a certain aspect of the AAA geometry, may be helpful in predicting AAA growth. The advantage of this approach when compared against that detailed in section 6.2.3 is its ability to highlight

features/relationships without any implicit user bias. Additionally, this framework allows for the ability to interpret the relevant features, which is extremely valuable in the clinical setting. For this experiment, an iterated 10-Fold optimization with shuffling (n = 100 iterations) approach was applied using all 192 patients for model training/optimization and validation.

6.2.4.1 Reference Axis Reorientation prior to Feature Extraction (PCA)

Inherently, aortic aneurysms display high levels of variability not only in terms of geometry but also in terms of location and extent within the infra-renal region. The goal of the PCA is to capture this variability in meaningful principal components. However, this is complicated by the presence of additional variability secondary to the movement of patient locations between imaging studies. Minimization of this variability was performed by reorienting all images to the same coordinate axis based on a fixed point. In this instance, the fixed point was selected to be the 2D aortic centroid at the level of the origin of the celiac artery. This was selected as the reference point as it was present in all patients, and consisted of a non-pathological point proximal to the pathological aneurysm. **Fig. *51*a** displays the AAA centrelines before and after the re-orientation. It is visually apparent that following reorientation, any variability in the OWS centreline was primarily due to the aneurysm.

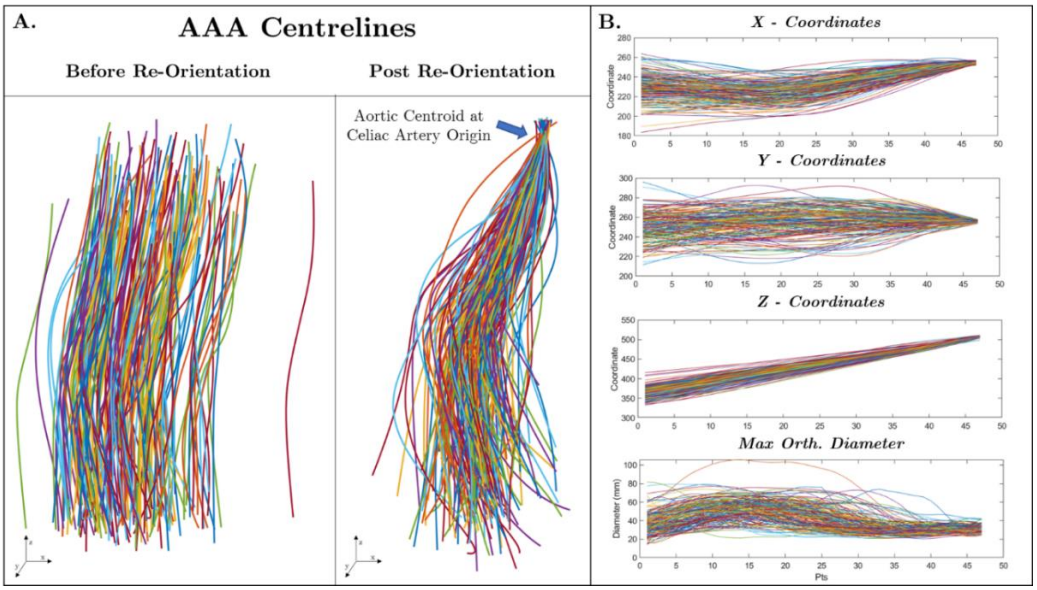


Fig. 51: Reference axis reorientation of AAAs prior to PCA-mediated Feature Extraction. Centrelines are extracted from the AAA shape using a homotopic thinning algorithm (Section 4.2.15). Centrelines were registered to the location of the aortic centroid at the celiac artery origin. 46 equidistant points were sampled along the derived centrelines. The X, Y and Z coordinates along with the maximum aneurysmal diameter perpendicular to the centreline at the given points were extracted and served as the input for PCA.

6.2.4.2 Feature Extraction: Characterization of the aortic/AAA Shape

Following re-orientation, the shape of each aorta/AAA was characterized into a singular vector, which consisted of its centreline coordinates and the maximum diameter along planes orthogonal to the centreline at the pre-defined coordinates. Fifty equidistant points along the centreline were interpolated and orthogonal planes were sampled using methods highlighted in Section 4.2.15 (**Fig. *51*b**). **Fig. *52*** displays 4 aortas with AAAs characterized in this format. Subsequently, all x, y, and z coordinates along with the maximum diameter measurements were normalized by subtracting the mean and dividing by the standard deviation. The first and last two points/planes along the aortic centreline were excluded from all patients as these orthogonal planes were predominantly located outside the AAA volume and would have incorrect max diameters.

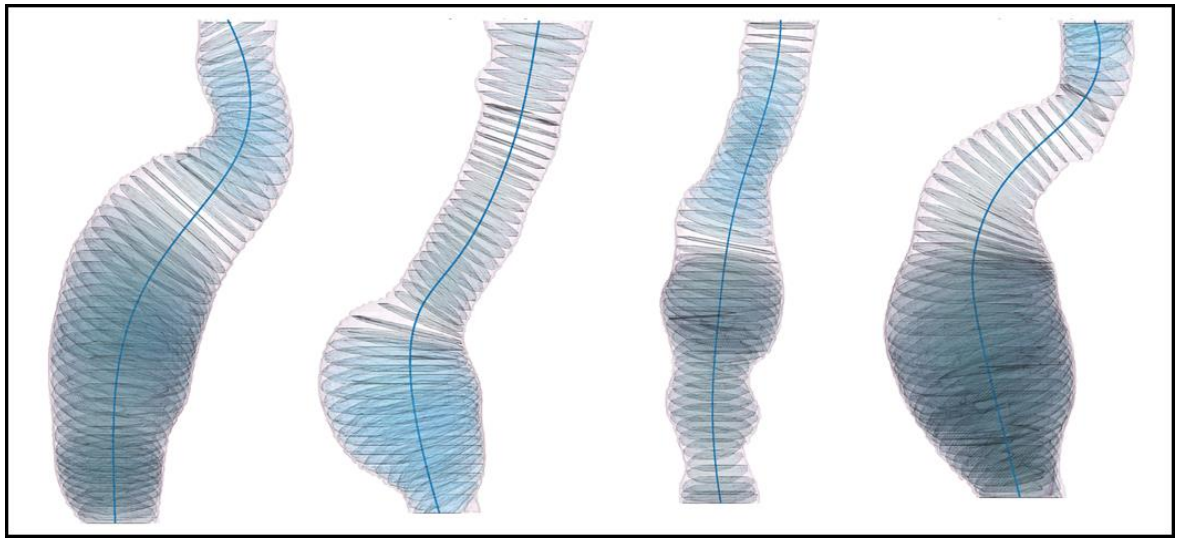


Fig. 52: Characterization of the aneurysmal shape prior to Shape-Based PCA. Equidistant points and orthogonal planes were sampled from the aortic centrelines. Maximum diameter along the perpendicular planes was extracted using MATLAB (4.2.15). The X, Y, and Z coordinates along with the maximum orthogonal diameters at the defined coordinates served as the input into the Shape-based PCA. Four aortic volumes from the OxAAA cohort, their respective centrelines and orthogonal planes are illustrated.

6.2.4.3 Feature Extraction: Characterization of the Aortic/AAA Surface

In addition to the aortic shape, we hypothesized that surface curvature and structure of the aorta/AAA may play a role in predicting aneurysmal growth. Therefore, the AAA was characterized by constructing curves along the aortic surface. To achieve this visualization, a spline curve was first generated to encompass the aortic/AAA segmentation from each axial slice along the AAA shape. These spline curves were then sampled into 360 equidistant points.

Subsequently, the surface lines were derived by connecting points that are closest together between slices. **Fig. *53*** illustrates this interpolation method utilized for three adjacent axial slices. Surface lines were smoothed using a Savitzky-Golay filter in MATLAB (Polynomial Order: 3, Window Length: 11). **Fig. *54*** displays 4 aortas with AAAs characterized in this format. Five surface lines (at 72° increments) were used for subsequent analysis as we deemed this sufficient to capture the degree of surface curvature, irregularity and asymmetry. Fifty equidistant points were interpolated from the five surface lines (5 surface lines x [50 X-coordinates, 50 Y-coordinates, 50 Z-coordinates] = 750 coordinates). The X, Y and Z coordinates of surface lines at identical starting points (at 72° increments) between patients were normalized by subtracting the mean and dividing by the standard deviation. This matrix (dimensions 750 x 192) served as the input into the PCA.

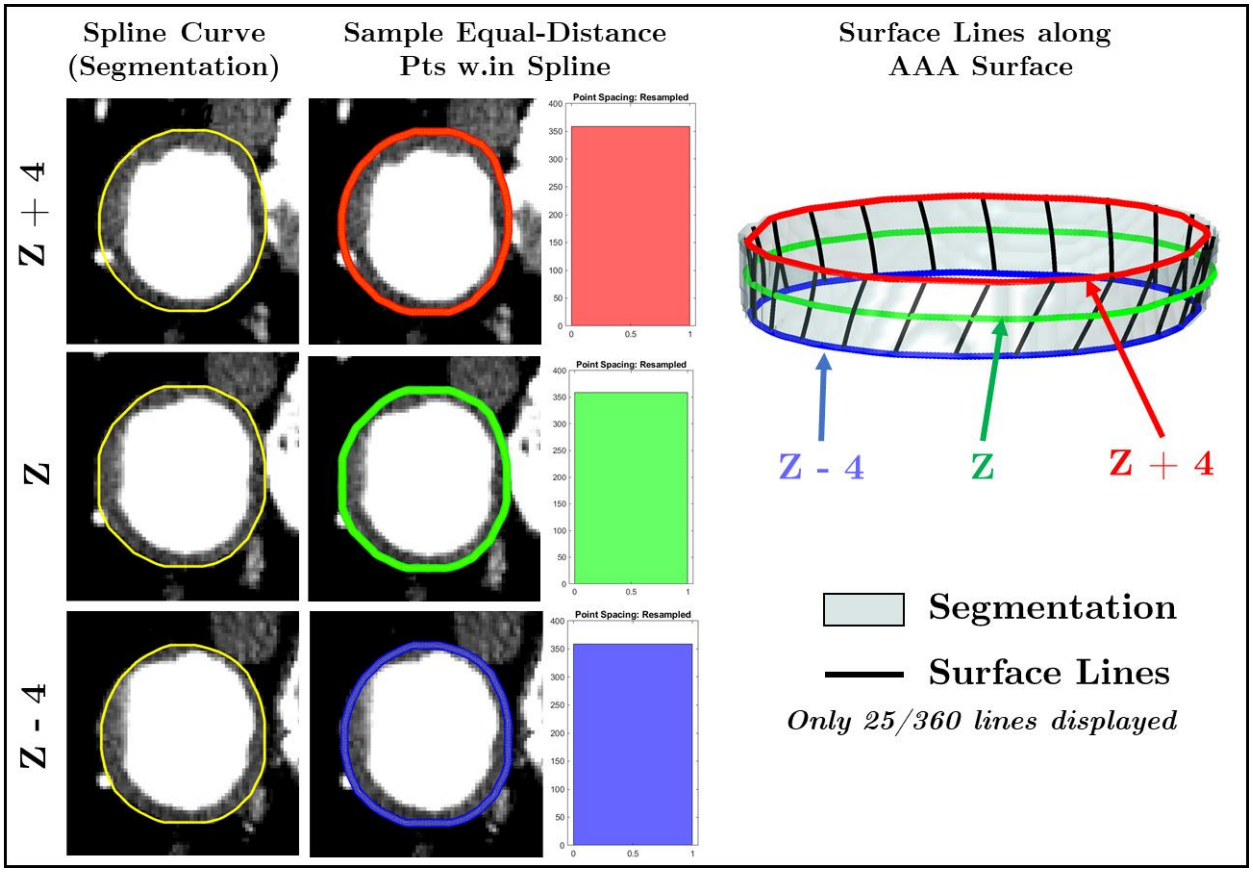


Fig. 53: Method to generate surface lines from the Aortic/AAA segmentations. 360 equidistant points were sampled from each spline curve encompassing the axial (2-D) aortic/AAA segmentation. Spacing between the sampled points are highlighted for each segmentation. Three adjacent slices at axial levels (Z-4, Z and Z+4) and the surface lines connecting these cross-sections are visualized.

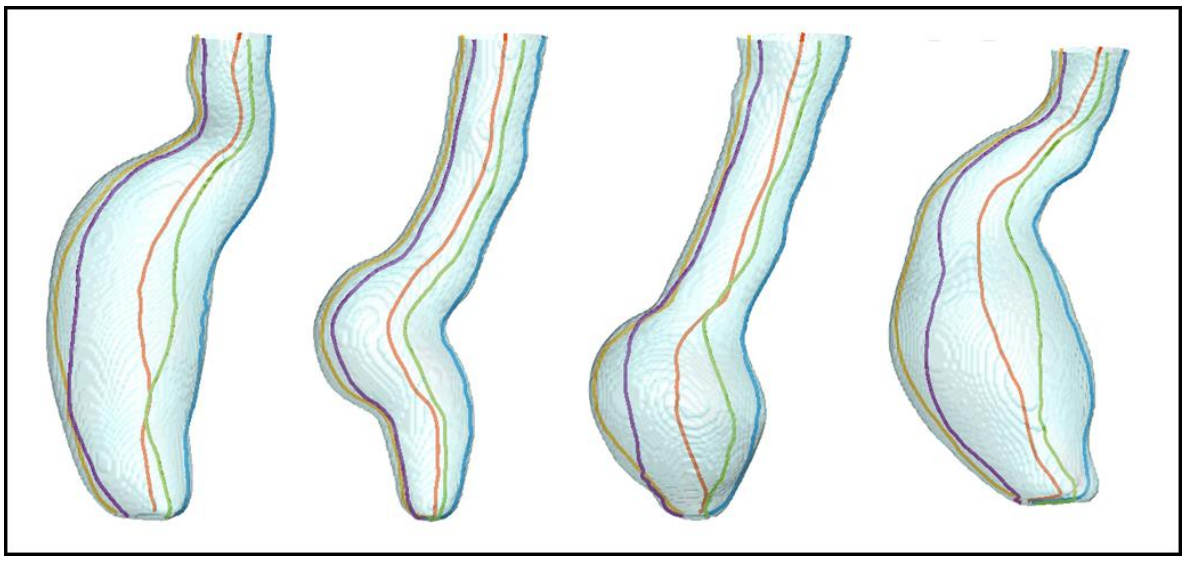


Fig. 54: Characterization of the aneurysmal surface prior to Surface-Based PCA. Four aortic/AAA volumes are characterized by their respective surface lines. Surface lines are generated by connecting points encompassing the aorta/AAA between adjacent slices. For each patient, five surface lines are generated for each volume, each at 72° increments.

6.2.4.4 Feature Extraction: Characterization of Aortic Side Branch Origins

In addition to the aortic shape and surface, we hypothesized that the location of the aortic side branches as they arise from the aorta may play a role in predicting aneurysmal growth secondary to alterations in hemodynamic forces. Aortic/AAA segmentations were expanded to capture the aortic side-branches using methods devised and validated in Section 4.2.12-4.2.13. Additionally, the captured side branches were automatically classified into six categories based on branch artery (1. Celiac Artery, 2. Superior Mesenteric Artery, 3-4. Left/Right Renal artery, 5-6. Left/Right iliac artery). This classification algorithm was based on the locations and spatial relationships between the branches. **Fig. *55a*** illustrates four expanded segmentations with classified branch centrelines. Subsequently, branch origins were defined as the first 5 coordinates (X, Y, Z) of each branch (**Fig. *55b***). Origin points were used instead of the entire branch centreline to focus the analysis at the intersection points and standardize the segment used from each branch. For example, in **Fig. *55*a**, the length of the iliac centrelines were visually different between the patients. This variability is due to the semi-automatic nature of branch segmentation. The X, Y and Z coordinates for each sets of branch origins between patients were normalized by subtracting the mean and dividing by the standard deviation. This matrix (⌊6 branches x 5 origin points/branch⌋ x 192 patients) served as the input into the PCA.

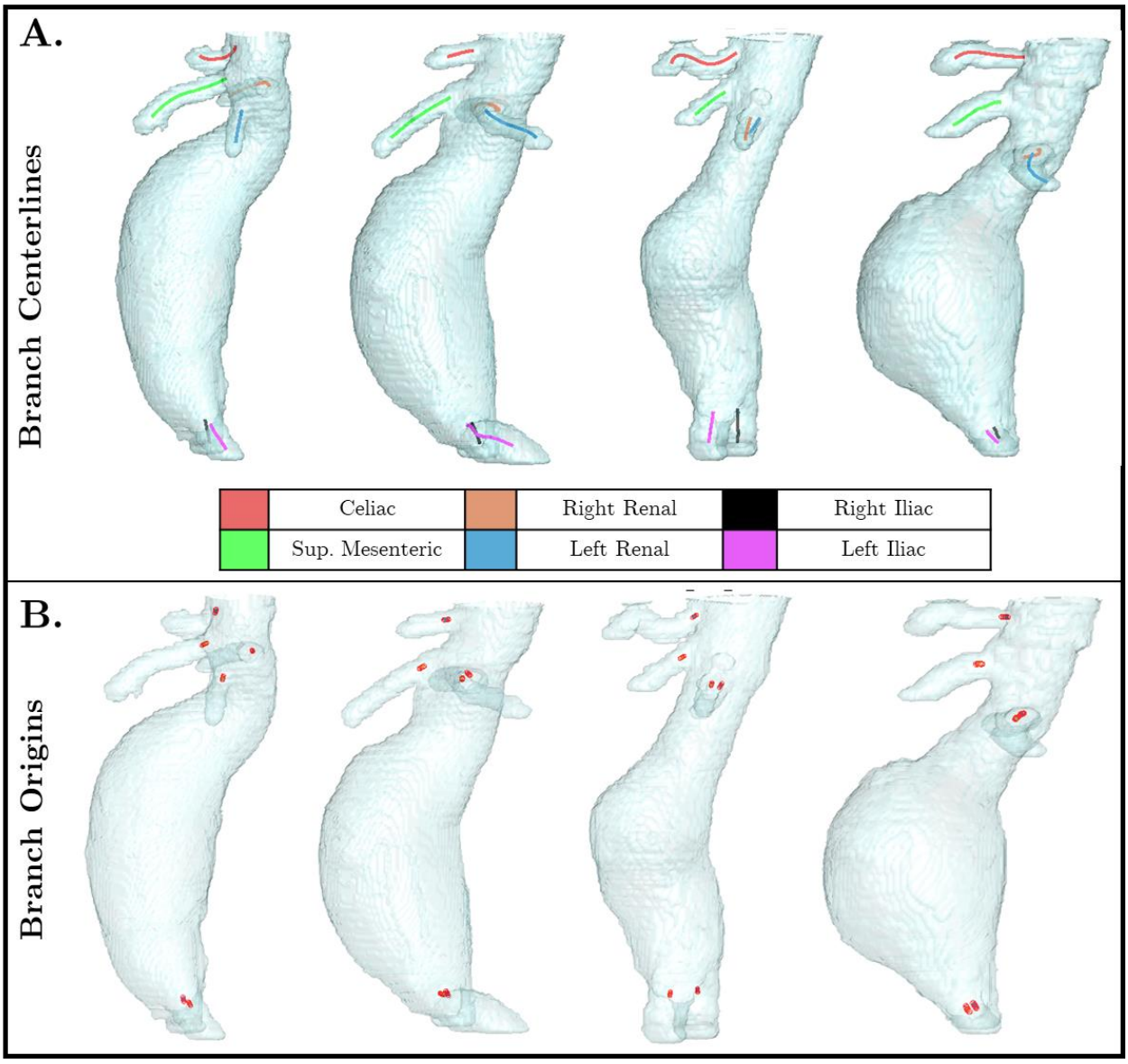


Fig. 55: Origin Point extraction of the aortic side branches from aortic/AAA segmentations. **A.** Segmentation of the aortic side branches and centreline extraction was derived from the deep learning AAA output using methods detailed in Section 4.2.12-4.2.13. Automatic classification of the side branches based on branch artery was performed. The output for four patients within the OxAAA cohort is illustrated. **B.** Branch origins were defined as the first 5 coordinates of each branch. This served as the input into the Branch Point -based PCA.

6.2.4.5 Feature Extraction: Principal Component Analysis

Principal component analysis was performed to assess the correspondences between different components in the shape- (Section 6.2.4.2), surface- (Section 6.2.4.3) and branch point (Section 6.2.4.4) -based input matrices. Here, the primary goal was to investigate and potentially identify additional features that not only characterize the aorta/AAA but also predict aneurysmal progression. PCA was performed separately for each of the input matrices. Coefficients were calculated by using the residuals (Input Data – Mean of Input Data) and singular value decomposition. The obtained principal components were analysed for their variance percentage and component coefficients, to determine their significance. For each of the three PCAs, the principal components that collectively captured 95% of the variance percentage were used for growth prediction analysis.

6.2.5 Developing the growth prediction models using both hypothesis-driven and PCA-derived features

To ascertain if the hypothesis-driven features (APD, UI and RC) were independent of the patient demographic profile for the purpose of AAA growth prediction, we first examined the relationship (spearman correlation) between these extracted geometric features against demographic features in the OxAAA-IMG-III dataset. We then focused the CT image analysis on OxAAA-IMG-IV dataset, using the geometric features extracted from the baseline scan to predict future growth (as recorded by the subsequent pre-operative scan). Summary statistics are described either as average (+/- standard deviations, SD) or median (with interquartile range, IQR). For statistical comparisons, t-tests or ANOVA were used for normally distributed data, whereas Kruskal Wallis test or Spearman correlation were used as non-parametric tests.

Multiple generalized linear models were applied to build the AAA growth prediction models, for the prediction of AAA growth as a categorical (Slow, Medium or Fast growth) or continuous variable. Model training and optimization were performed using iterated 10-fold optimization with shuffling (n = 100 iterations) using the data from the OxAAA-IMG-IV dataset (n =192 patients). Here, the data was partitioned into 10 equal sized subsets (folds). Then, 9-folds are used to train the model and the remaining fold is used for internal evaluation. For each iteration, this process is performed a total of 10 times, with a different fold as the validation data set.

For each of the feature sets, linear discriminant analysis (LDA) was used for the binary classification task of predicting either: 1. Slow or Not-Slow Growth or 2. Fast or Not-Fast Growth. During model optimization, features that were not significant were eliminated. Here, growth rates were categorized into 3 groups: $<$ 2.5 mm/yr (Slow growth), 2.5-5.0 mm/yr (Medium growth), and $>$ 5.0 mm/yr (Fast growth). These thresholds were chosen based on the summary statistics of growth rates observed within this cohort (Slow Growth – Below 1st Quartile, Fast Growth – Above 4th Quartile). Area under the Receiver Operating Characteristic (AuROC) were derived for each feature combination to assess the performance of the regression model in discerning growth against a pre-defined growth rate threshold. Accuracy of the linear regression model was reported using RMSE between the actual annual growth versus predicted growth. We further assessed the prediction accuracy within a 2mm margin as this is the accepted technical variability between measurements of AP diameter in CT images[110,133].

6.2.6 Evaluation of the growth prediction models on an independent validation cohort (MA3RS Study)

An independent set of 148 patients with CT-monitored aneurysmal growth measurements were obtained from the MA3RS dataset (University of Edinburgh) and served as a validation. Aneurysmal and associated side-branch segmentations were automatically-derived from the provided CT images (non-contrast or CECT). Feature extraction was performed using methods highlighted in sections 6.2.4.1 – 6.2.4.4. The optimized models composed of both hypothesis- and PCA-derived features were used to predict aneurysmal growth as both a continuous and categorical outcome. Classification accuracy was evaluated by comparing the AuROC for the independent validation cohort against that obtained with the OxAAA-IMG-IV dataset. Similarly, the accuracy of the linear regression model in predicting AAA growth was obtained by calculating the RMSE difference between model prediction and ground truth.

6.3 Results

6.3.1 Patient demographic features do not influence the geometric features of AAA

102 patients were included in the prospectively recruited OxAAA-IMG-III dataset (Male n= 99; Females n= 3). The demographic information from this dataset are explained in detail in Section 3.1.3 and are presented in **Table 15**. Median AAA size within this cohort was 63.0 mm with an interquartile range from 58.0 to 72.5 mm. No correlations were identified between the demographic parameters and the extracted hypothesis-driven geometric features, indicating that AAA size or APD, UI and RC are independent of patient demographic characteristics.

6.3.2 Hypothesis-driven Geometric Feature Extraction from the OxAAA Growth Prediction Cohort

The second arm of this study included 192 AAA patients with serial CT scans obtained at least 8 months apart (OxAAA-IMG-IV dataset). Median follow-up time between scans was 2.0 years with an interquartile range of 1.0 to 3.7 years. Similarly, median aneurysmal growth was 3.7 mm/yr with an interquartile range of 2.5 to 5.0 mm/yr. There were significant positive correlations between AAA size (Spearman r = 0.15, p = 0.03) and UI (Spearman r = 0.40, p < 0.001) with annual AAA growth rate (**Fig. *56*a-b**). Whereas a significant negative correlation between minimum RC and annual AAA growth rate was observed. (Spearman r = -0.55, p < 0.001, **Fig. *56*c**). **Fig. *57* and Fig. *58*** illustrates six AAAs of similar diameter at baseline with disparate UI and RC. These corresponded to different annual growth rate observed subsequently.

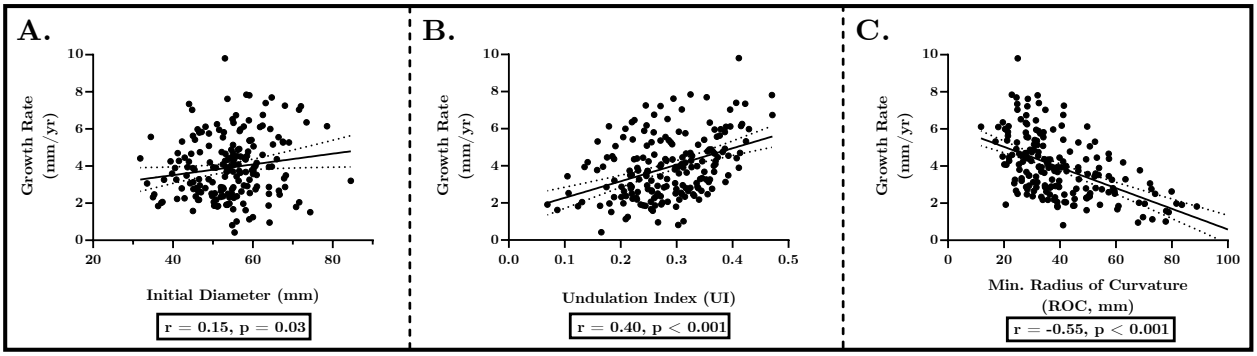


Fig. 56: Linear correlations between extracted geometric featuresfrom AAAs within the OxAAA-IMG-IV dataset. (**A.** Initial Max AP Diameter or AAA size, **B.** Undulation Index (UI) and **C.** Minimum Radius of Curvature (RC) and estimated annual AAA growth rate). Statistical significance is assessed for each correlation and is indicated on each graph[12].

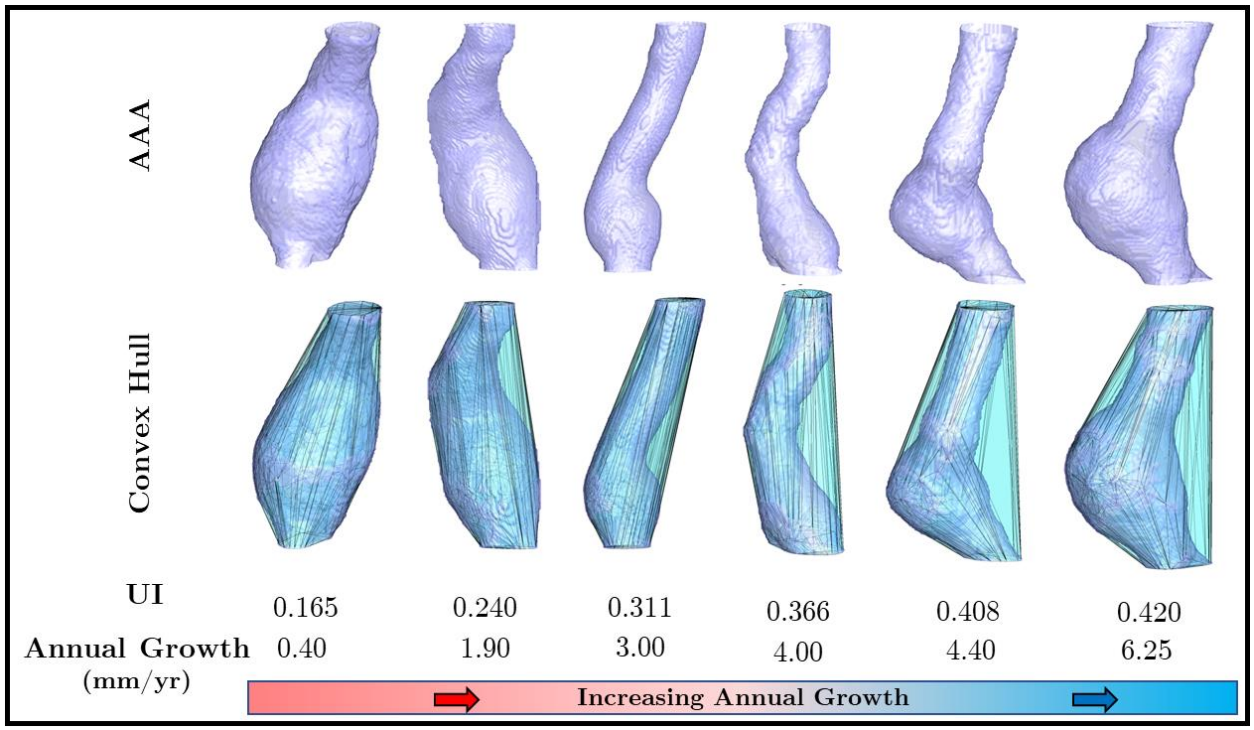


Fig. 57: Six AAAs of similar size displayed alongside their respective convex hull. Undulation indices are calculated for each aneurysmal pair (Section 6.2.3.3). Aneurysms are ordered in terms of increasing UI and positively correlate with increasing annual AAA growth rate[12].

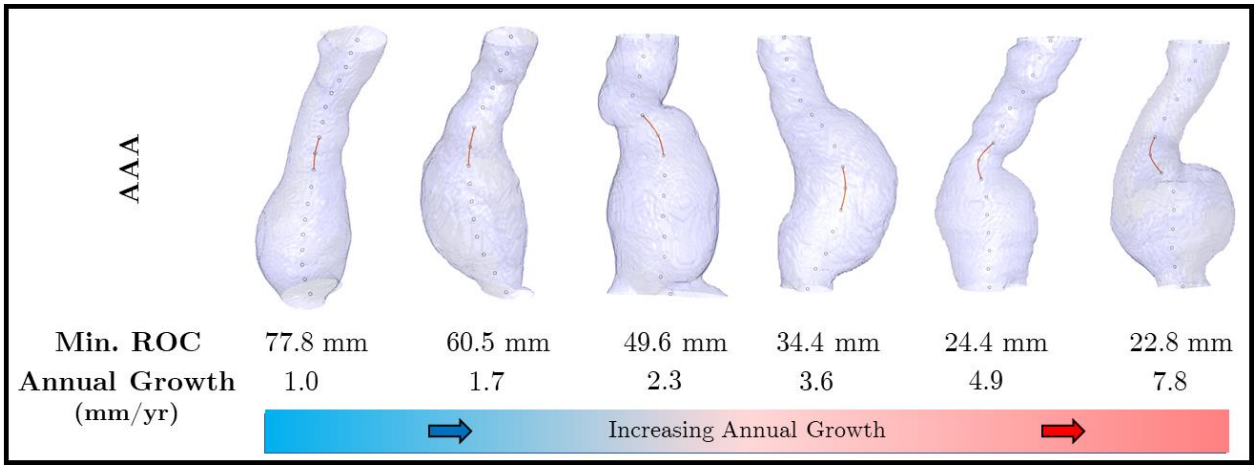


Fig. 58: Six AAAs of similar size displayed with calculated centrelines. Centrelines are calculated using a variation of the homotropic thinning algorithm. Regions with the minimum radius of curvature along each centreline, which correspond to regions of increased curvature, are highlighted in orange. Aneurysms are ordered in terms of decreasing min RC and negatively correlate with increasing annual AAA growth rate[12].

6.3.3 Semi-Supervised Geometric Feature Extraction from the OxAAA Growth Prediction Cohort

Here, we hypothesized that there may be features related to the aneruysmal geometry, in addition to the hypothesis-driven features (UI and RC), that may play a role in AAA growth secondary to changes in hemodynamic forces. As a result, three sets of prinicipal component analysis were independently performed to strategically capture the diverse AAA shapes in forms that were easy to analyze. PCA coefficients were calculated by using the residuals (Input Data – Mean of Input Data) and singular value decomposition.

The first PCA (*AAA-Shape*) was performed using the shape-based AAA characteristics derived using methods outlined in section 6.2.4.2. Each of the 192 patient vectors was composed of 188 points (46 X-coordinates + 46 Y-coordinates + 46 Z-coordinates + 46 maximum orthogonal diameters).The input matrix (192 patients x 188 points) was then normalized per component prior to analysis. **Fig. *59***, which illustrates the cumulative variance captured by the derived modes, suggests that the first 10 modes are sufficient to capture 95% of the variance.

The second PCA (*Surface*) was performed using the surface curves surrounding the AAA. These curves were derived using methods outline in section 6.2.4.3. Each of the 192 patient vectors was composed of 690 points (5 Surface lines x [46 X-Coordinates, 46 Y-coordinates, 46 Z-coordinates] per Surface line). The topmost points of each of the lines (at the level of the Celiac Origin) were equidistant at at 72° increments. The input matrix (192 patients x 690 points) was then normalized prior to analysis. The first 11 modes are sufficient to capture 95% of the vairance (**Fig. *59***).

Finally, the third PCA (*Branch-Point*) was performed to investigate if the location and progression of the aortic side branches have any impact in AAA progression. The branch origins, which consisted of the first 5 coordinates along the branch centreline, were derived using methods outlined in section 6.2.4.4. In this instance, each of the patient vectors was composed of 6 components, one for each branch, and each branch consisted of 5 coordinates (X, Y and Z). The input matrix dimmensions were 192 x 150 (192 patients x [6 branches x 5 coordinates per branch x 3 points per coordinate]). Each coordinate between the patients was normalized prior to analysis. In this instance, the first 17 modes are required to capture 95% of the variance (**Fig. *59***).

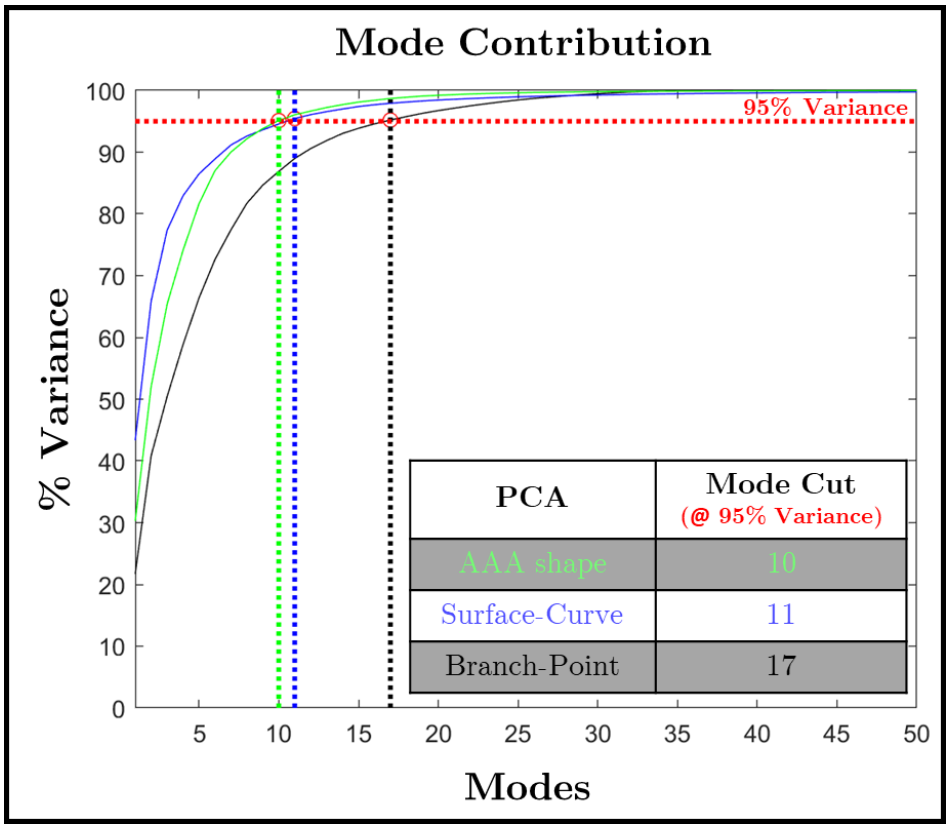


Fig. 59: Cumulative mode contribution for all three sets of PCA. AAA shape/volumes were characterized by 3 methods, with each method focused on capturing a different aspect of the aneurysmal geometry. AAA shape-based characterization (Red) was performed by calculating the aneurysmal centreline and the maximum diameters along planes orthogonal to the centreline. Surface-Curve-based characterization (Blue) was performed by extracting five equidistant curves along the outer surface of the AAA (at 72° increments). Finally, characterization of the AAA branch points (Black) was performed by compiling the first 5 coordinates along each of the 6 side branches (descending aorta). 95% variance was captured using the 10, 11 and 17 modes for the AAA-shape, Surface-Curve and Branch-Point PCAs, respectively

6.3.4 Prediction of AAA growth as a categorical outcome

The baseline characteristics of the subgroups of patients [(slow growth: <2.5mm/year); (some growth: 2.5mm to 5mm/year), (fast growth: >5mm/year)] are summarized in **Table *16***. There was no difference in the starting AAA diameter between the three groups. UI at baseline differed significantly between the three groups (ANOVA, p=0.003). There was also a significant inverse trend of relationship between RC and AAA growth (Kruskal-Willis, $p < 0.0001$).

Table 16: Groups split based on annual AAA growth within the OxAAA Imaging cohort.

	I. < 2.5 mm/yr (n = 44)	**II. 2.5 – 5.0 mm/yr (n = 98)**	**III. > 5.0 mm/yr (n = 50)**	**p**
Time between Scans, yrs (IQR)	2.0 (1.1 – 4.5)	2.2 (1.5– 4.1)	1.6 (1.0 – 3.3)	0.10
AAA Diam⁺, mm (± SD)	51.0 ± 8.8 mm	53.3 ± 7.1 mm	56.4 ± 9.2 mm	0.07
Undulation of Index⁺ (± SD)	0.25 ± 0.06	0.28 ± 0.07	0.32 ± 0.08	**0.003**
Min. Radius of Curvature (IQR)	49.6 (38.5– 58.4)	35.3 (27.6 – 51.2)	29.7 (21.4 – 34.1)	**<0.0001**

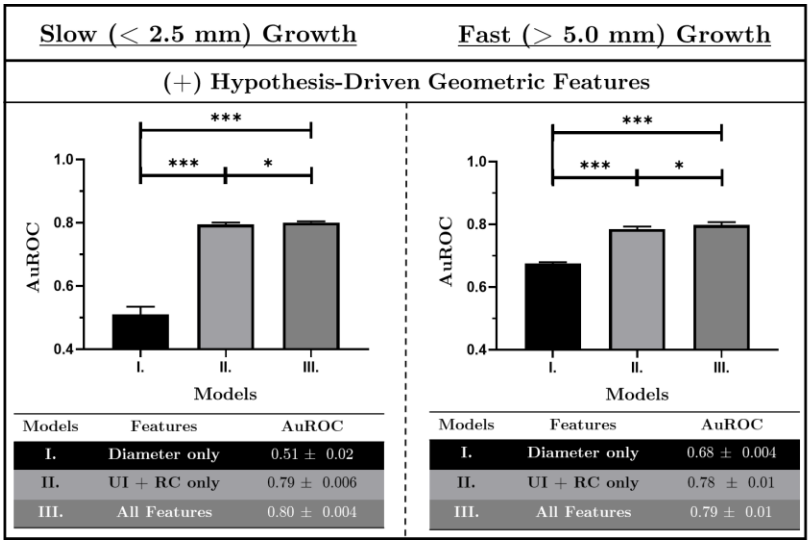
Note: Comparison of subgroups is performed to establish differences between groups. These groups served as the basis for the logistic regression. Variables that follow a Gaussian/Normal Distribution are indicated with a +. For these variables, mean ± SD are presented, and cohort differences are compared using a one-way ANOVA. For variables that don't follow a Gaussian distribution, median and inter-quartile range (IQR) are presented and cohort differences are compared using a Kruskal-Wallis test.

6.3.4.1 Hypothesis-Driven Geometric Features predict AAA growth

Different combinations of the hypothesis-driven features (APD, UI and RC) were used to train multiple logistic regression models. The feature combinations used for each model is indicated in **Fig. *60*** and were trained using an iterated 10-fold cross-validation approach (n = 100 iterations). ROC curves on the validation cohorts were plotted for each feature combination with the threshold of "Slow (< 2.5 mm) Growth" and "Fast (> 5.0mm) Growth". The area under receiver operation curve (AuROC) metric shows good discriminative capacity of AAA growth rate based on all three variables at the predefined thresholds. Using APD, UI and RC as three input variables of the prediction algorithm, the AuROC for predicting slow growth (<2.5mm/year) and prediction fast growth (>5mm/year) was 0.80 and 0.79, respectively. This model comprising of 3 variables significantly outperforms the use of AAA diameter alone as the predictor ($p < 0.001$).

In addition to hypothesis-driven features, PCA-derived features (modes) were used to independently train multiple logistic regression models for the task of predicting AAA growth rate. For each of the three PCAs (1. AAA shape, 2. Surface Curve, 3. Branch Point), an

iterative 10-fold cross-validation approach (n = 100 iterations) was implemented using only the features/modes that captured 95% of the observed variance (**Fig. *59***).



Models	Features	AuROC
I.	Diameter only	0.51 ± 0.02
II.	UI + RC only	0.79 ± 0.006
III.	All Features	0.80 ± 0.004

Models	Features	AuROC
I.	Diameter only	0.68 ± 0.004
II.	UI + RC only	0.78 ± 0.01
III.	All Features	0.79 ± 0.01

Fig. 60: AuROC for each combination of features to assess the performance of the multinomial regression model in discerning AAA growth phenotype (A. > 5.0 mm and B. < 2.5 mm Growth). ROC curves were generated from the evaluation of the validation cohort, following model training (iterative 10-fold cross-validation, n = 100 iterations). Significance testing was performed to compare the AuROCs for each model (n = 100). Significant differences were noted with * or *** indicating $p<0.05$ and $p<0.001$.

6.3.4.2 PCA-derived Geometric Features improve AAA growth prediction

Using the first 10 shape-based modes, the AuROCs for predicting slow and fast growth were 0.69 and 0.78, respectively. LDA highlighted the significance of features/modes 3 ($p < 0.001$) and 7 ($p = 0.008$) in predicting slow growth. Similarly, modes 1 ($p = 0.02$) and 4 ($p < 0.001$) were statistically significant for predicting fast growth. Combining maximum diameter and hypothesis-driven features with these *selected shape-based modes* significantly improved model predictive capacity (Model VI vs Model IV, $p < 0.001$, **Fig. *61*a-b**).

Similarly, using the first 11 surface modes, the AuROCs for predicting slow and fast growth were 0.66 and 0.73, respectively. LDA highlighted the significance of mode 6 ($p < 0.001$) and mode 3 ($p < 0.001$) for predicting slow and fast growth, respectively. Combining maximum diameter and hypothesis-driven features with these *selected surface-line- based modes* improved model predictive capacity (Model IX vs Model VII, $p < 0.001$, **Fig. *61*c-d**).

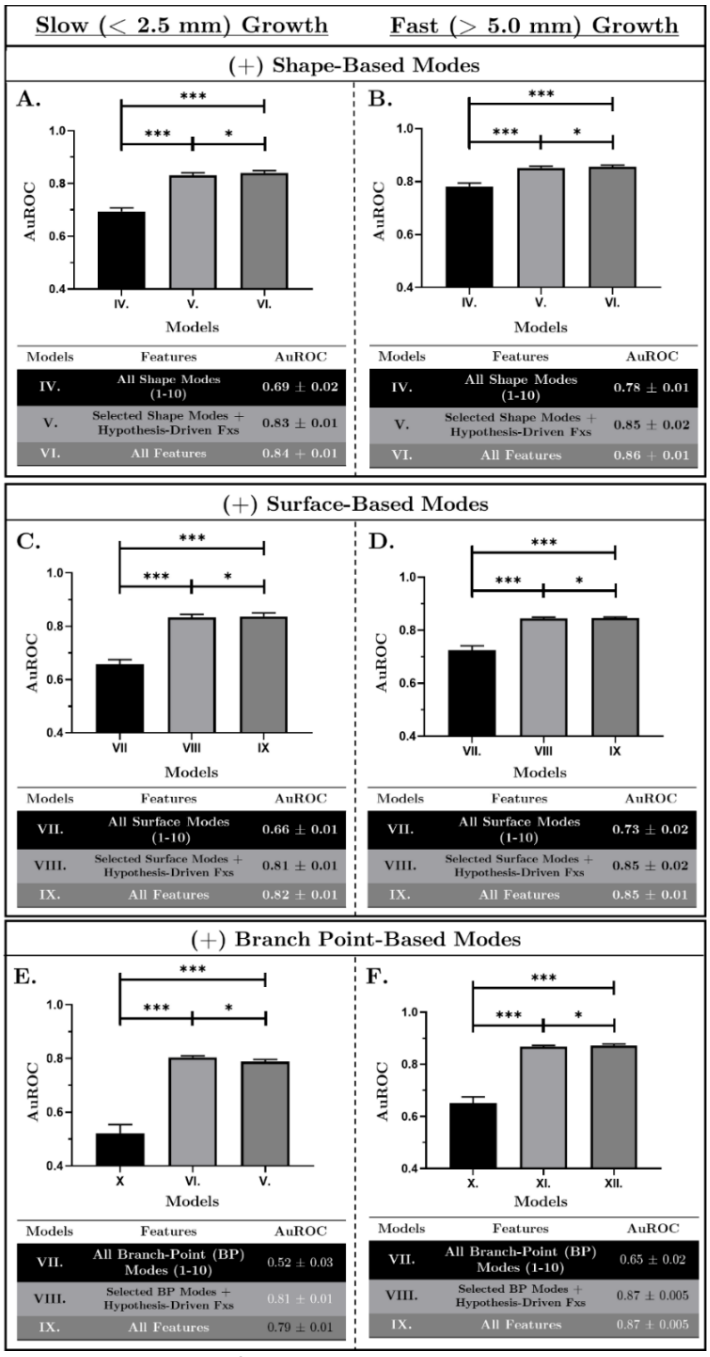
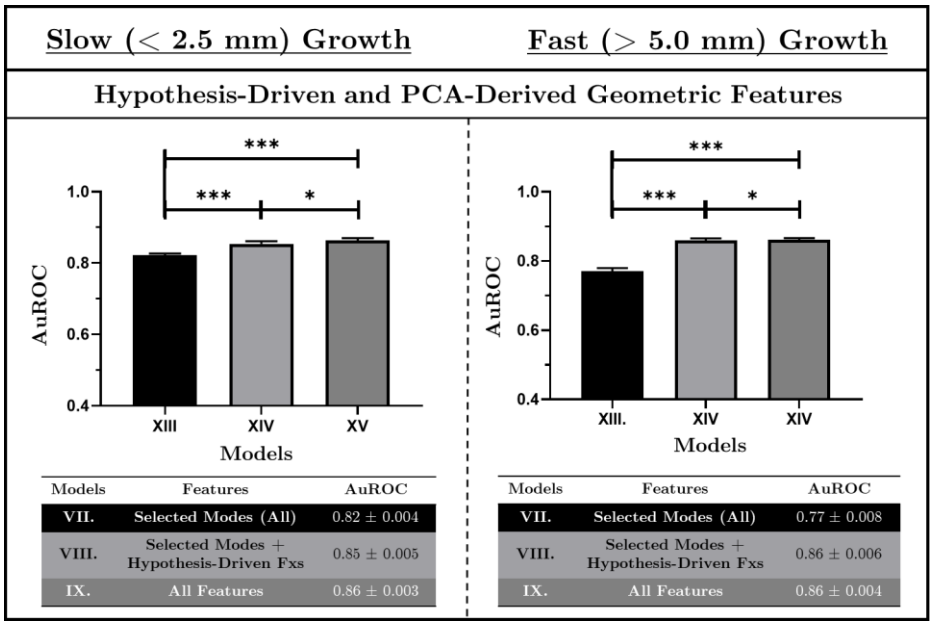


Fig. 61: AuROC for models classifying Slow (< 2.5 mm) and Fast (>5.0 mm) growth trained with PCA-derived features. All models were trained using iterative 10-fold cross-validation (n= 100 iterations). Mode significance was assessed iteratively using linear discriminant analysis. Only significant modes were used for subsequent analysis. A-B. Shape-based modes were derived from the AAA using its centreline and maximum diameters along planes orthogonal to its centreline. Modes [3, 7] and [1, 4] were statistically significant for predicting slow and fast growth, respectively. C-D. Surface-based modes were derived from curves characterizing the aneurysmal outer wall. Modes 6 and 3 were statistically significant for predicting slow and fast growth, respectively. E-F. Branch-Point (BP) modes were obtained from the location/origin of the aortic side branches exiting the descending/abdominal aorta. Here, Modes 10 and 3 were statistically significant for predicting slow and fast growth, respectively. In all cases, the incorporation of the hypothesis-driven features improved model predictive capacity.

The AuROCs for predicting slow and fast growth using the 17 branch-point modes were 0.52 and 0.65, respectively. Mode 10 was significant for predicting slow growth ($p = 0.002$). On the other hand, mode 3 was significant for predicting fast growth ($p < 0.001$). Combining maximum diameter and hypothesis-driven features with these *selected branch point- based modes* improved model predictive capacity (Model XII vs Model X, $p < 0.001$, **Fig. *61*e-f**).

6.3.4.3 Refinement of the individual prediction model for fast/slow growth

Combining the relevant shape-, surface- and branch-point modes improved model predictive capacity for both classification tasks (Models XIII – XIV, **Fig. *62*a-b**). For slow growth, of the four modes, shape modes 3 and 7 as well as branch point mode 10 remained significant (AuROC of Model XV, 0.82). These three modes remain significant when combined with the hypothesis-driven features (Max Diameter, UI and RC). On the other hand, of the three identified modes for the prediction of fast growth, shape mode 4 and branch point mode 3 remained significant (AuROC of Model XV, 0.77). These two modes remain significant when combined with the hypothesis-driven features (APD, UI and RC). For both models, after integration, max diameter and UI were no longer statistically significant and were removed from subsequent models. The coefficients and p-values for each of the variables within the optimized logistic regression models as well as their performance are highlighted in **Table *17*.** These optimized models significantly outperform Models I (APD, $p<0.001$) and III (APD, UI and RC, $p<0.001$).



Models	Features	AuROC
VII.	Selected Modes (All)	0.82 ± 0.004
VIII.	Selected Modes + Hypothesis-Driven Fxs	0.85 ± 0.005
IX.	All Features	0.86 ± 0.003

Models	Features	AuROC
VII.	Selected Modes (All)	0.77 ± 0.008
VIII.	Selected Modes + Hypothesis-Driven Fxs	0.86 ± 0.006
IX.	All Features	0.86 ± 0.004

Fig. 62: AuROC for models classifying Slow (< 2.5 mm) and Fast (>5.0 mm) growth trained with Hypothesis-Driven and PCA-derived features. Hypothesis-Driven Features included maximum diameter, minimum radius of curvature (RC) and undulation index (UI). Selected Modes or PCA-derived geometric features were different for both classification tasks. For slow growth, selected modes (All) included Shape Mode 3 and 7, Surface Mode 6 and Branch Point Mode 10. Of the four selected modes, Shape Mode 3 and 7 and Branch point mode 10 remained significant after integrating with Hypothesis-driven features. For Fast growth, selected modes (All) included Shape Mode 4, Surface Mode 3, and Branch Point Mode 3. Of the three selected modes, Shape Mode 4 and Branch point mode 3 remained significant after integrating with Hypothesis-driven features. Models were trained using iterative 10-fold cross-validation within the OxAAA cohort (n = 192). Model performance was assessed by comparing the AuROC between models for the 100 iterations.

Table 17: Optimized regression model combining both Hypothesis-driven and PCA-derived features to predict slow/fast growth. Model performance is highlighted using AuROC.

Slow (< 2.5 mm) Growth			Fast (> 5.0 mm) Growth		
Features	**Coefficients**	**p**	**Features**	**Coefficients**	**p**
Intercept	-2.653 ± 0.414	-	**Intercept**	-2.951 ± 0.510	-
RC	1.190 ± 0.314	0.0001	**RC**	-2.300 ± 0.588	0.0001
Mode 3 (Shape)	0.944 ± 0.330	0.004	**Mode 4 (Shape)**	-0.843 ± 0.376	0.01
Mode 7 (Shape)	1.145 ± 0.431	0.008	**Mode 3 (BP)**	-0.798 ± 0.409	0.03
Mode 10 (BP)	-1.251 ± 0.325	0.0001			
AuROC	**0.88 ± 0.005**		**AuROC**	**0.87 ± 0.006**	

6.3.4.4 Examination of PCA features provides insight into AAA growth

Shape-based PCA characterization and growth prediction analysis highlighted the significance of *Modes 3 and 7* for the prediction of slow growth and *Mode 4* for the prediction of fast growth. Visualization of the impact of modes 3 and 7 on the average AAA shape (± 2*σ) along with 6 AAAs of similar diameter at baseline with disparate values are show in **Fig. *63*** and **Fig. *64***, respectively. Here, σ represent the deviation from average AAA shape due to the selected mode or feature. The greater the σ, the larger the deviation from the mean and the impact of the selected mode. It is visually apparent from **Fig. *63*b,** that aneurysms, which start adjacent to the location of the renal arteries and are more fusiform in nature, are more likely to present with slow growth patterns. Similarly, **Fig. *64*a**, suggests that aneurysms with a not only a greater degree but also a particular direction of localized tortuosity within aneurysmal neck region are less likely to represent a slow growth phenotype. This is further supported in **Fig. *64*b**, which highlights 6 AAAs with similar diameters at baseline with different growth phenotypes. Mode values for each of these aneurysms are indicated.

Visualization of the impact of mode 4 on the average AAA shape (± 2*σ) along with 6 AAAs of similar diameter at baseline with disparate values are show in **Fig. *65***. Similar to that of mode 4, these visualizations highlight the importance aneurysmal tortuosity type, in addition to degree of curvature. For example, increased centreline tortuosity with a right-sided preference is linked with a rapid-growth phenotype than that with a left-sided preference (**Fig. *65*b**). Additionally, the location of the tortuous region also seems to be linked with AAA progression.

In addition to shape-based modes, certain branch-point modes (mode 3 and 10) played a significant role in the classification of AAA growth phenotype (**Table *17***). **Fig. *66*a** highlights the relevance of Mode 10 in altering the location of the six origin points and its relation to aneurysmal growth (Slow Growth vs Not-Slow Growth). It is apparent that as sigma changes, the greatest change in location is observed within that of the renal (left/right) arteries. Increased vertical placement of the left renal artery origin and medial placement of the right renal artery origin (next to the AAA shape) is associated with the slow growth phenotype (**Fig. *66*b**).

Similarly, **Fig. *67*a** highlights the relevance of Mode 3 in altering the location of the six origin points and its relation to aneurysmal growth. It is apparent that as sigma changes, the greatest change in location is observed within that of the iliac (left/right) arteries. As sigma decreases, which is characteristic of the fast growth phenotype, the origin of the iliac arteries moves inferiorly and laterally towards the right side. This is apparent in the subset of AAAs from the OxAAA-IMG-IV dataset (**Fig. *67*b**).

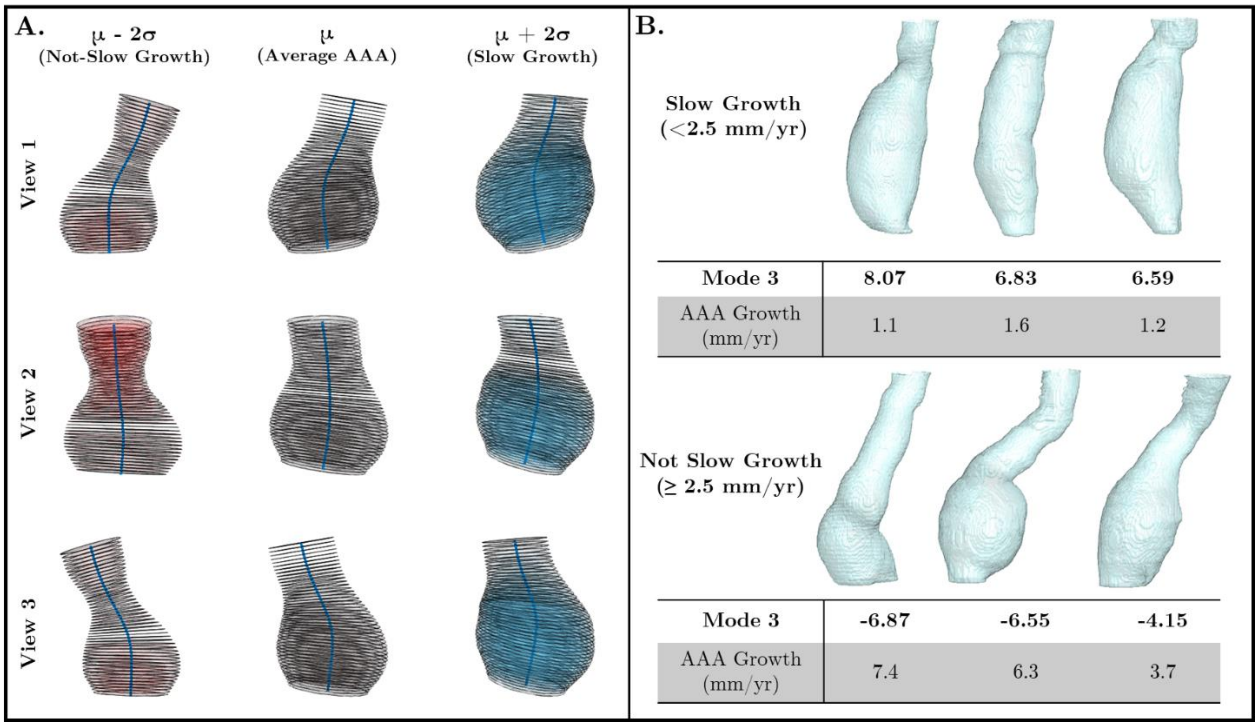
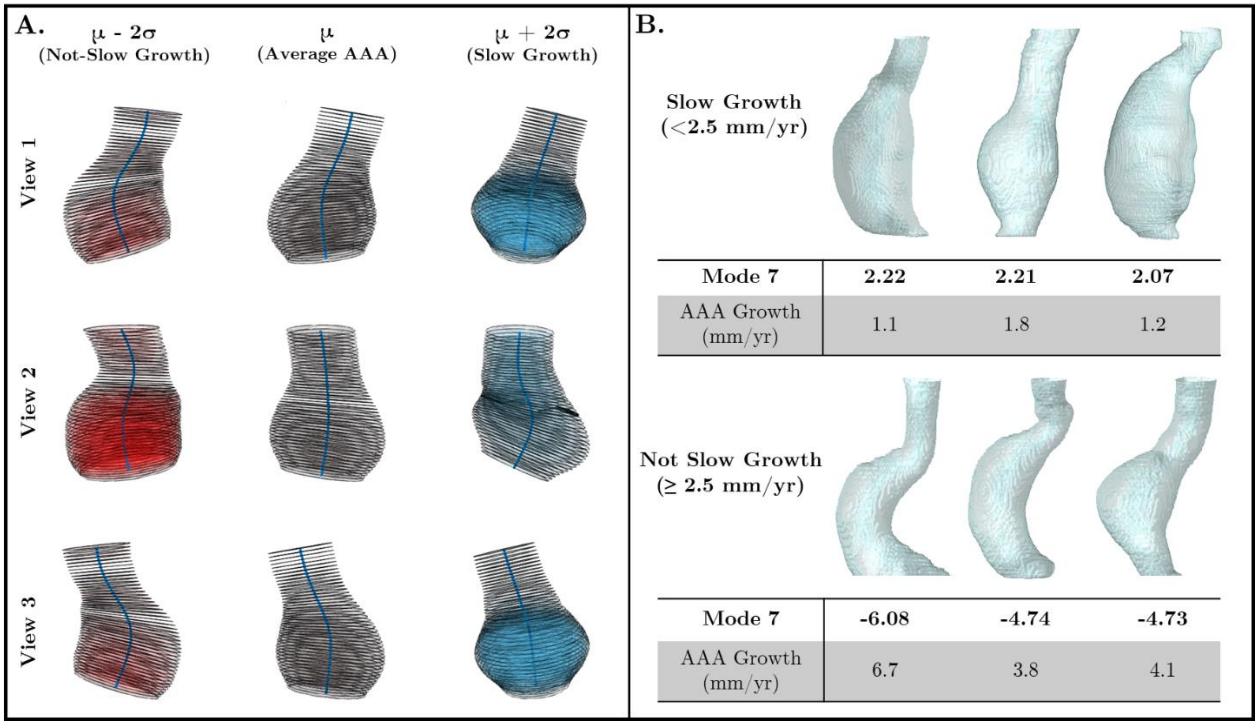


Fig. 63: The relevance of Mode 3 (Shape-based PCA) on the average AAA shape and its impact on aneurysmal growth (Slow vs Not-Slow Growth). A. Average AAA shape (μ) was constructed from the PCA input matrix and consisted of a centreline (46 X,Y,Z coordinates) with planes orthogonal to the centreline (circular planes with defined maximum diameters, 46 max diameters). The influence of the selected mode (mode = 3) was applied to the average shape to obtain the modified aneurysms. Here, the +2σ and -2σ represent the modified aneurysms, 2 standard deviations from the μ. This was visualized to appreciate the influence of the selected mode. Three views were visualized at 90° increments (view 1 -3). B. Six aneurysms with similar max AAA diameter at the starting timepoint were visualized within the OxAAA cohort. Mode 3 is inversely correlated with aneurysmal growth. Aneurysms, which start adjacent to the location of the renal arteries and are more fusiform in nature, are more likely to present with slow growth patterns.



Mode 7	2.22	2.21	2.07
AAA Growth (mm/yr)	1.1	1.8	1.2

Mode 7	-6.08	-4.74	-4.73
AAA Growth (mm/yr)	6.7	3.8	4.1

Fig. 64: The relevance of Mode 7 (Shape-based PCA) on the average AAA shape and its impact on aneurysmal growth (Slow vs Not-Slow Growth). A. Average AAA shape (μ) was constructed from the PCA input matrix and consisted of a centreline (46 X,Y,Z coordinates) with planes orthogonal to the centreline (circular planes with defined maximum diameters, 46 max diameters). The influence of the selected mode (mode = 7) was applied to the average shape to obtain the modified aneurysms. Here, the +2σ and -2σ represent the modified aneurysms, 2 standard deviations from the μ. This was visualized to appreciate the influence of the selected mode. Three views were visualized at 90° increments (view 1 -3). B. Six aneurysms with similar max AAA diameter at the starting timepoint were visualized within the OxAAA cohort. Mode 7 is inversely correlated with aneurysmal growth. Aneurysms with a not only a greater degree but also a particular direction of localized tortuosity are less susceptible to a slow growth phenotype.

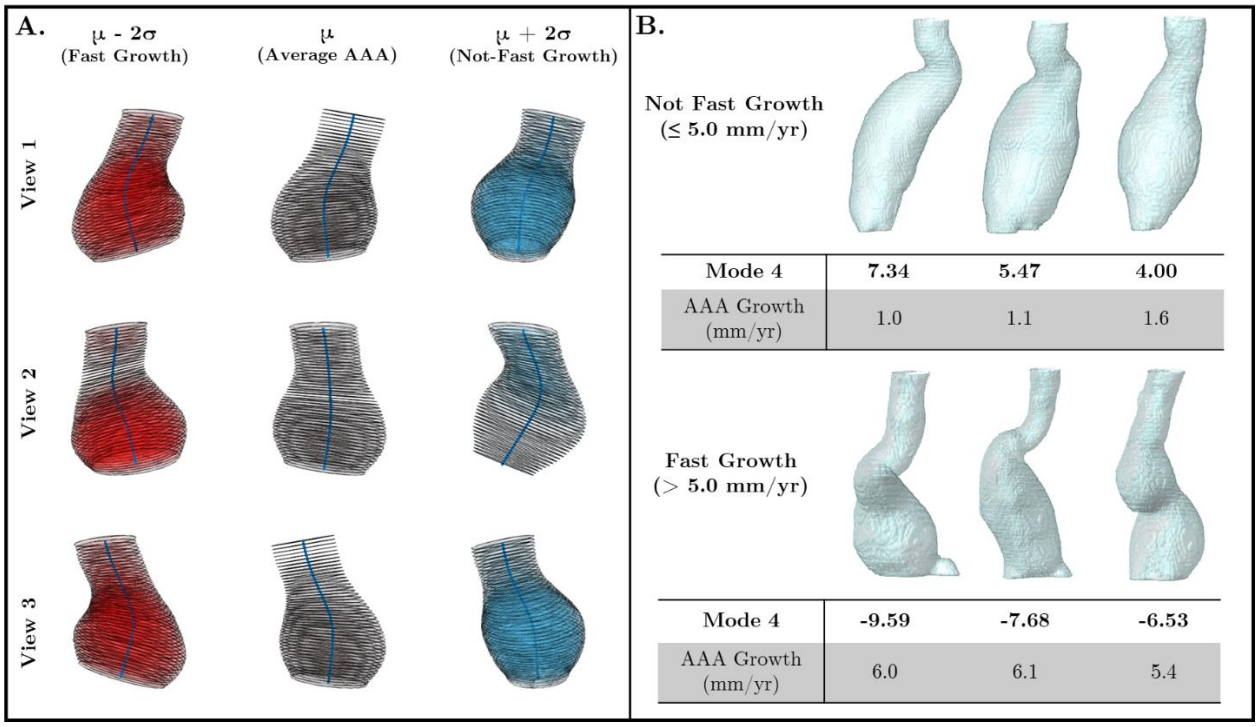


Fig. 65: The relevance of Mode 4 (Shape-based PCA) on the average AAA shape and its impact on aneurysmal growth (Fast vs Not-Fast Growth). A. Average AAA shape (μ) was constructed from the PCA input matrix and consisted of a centreline (46 X,Y,Z coordinates) with planes orthogonal to the centreline (circular planes with defined maximum diameters, 46 max diameters). The influence of the selected mode (mode = 4) was applied to the average shape to obtain the modified aneurysms. Here, the +2σ and -2σ represent the modified aneurysms, 2 standard deviations from the μ. This was visualized to appreciate the influence of the selected mode. Three views were visualized at 90° increments (view 1 -3). B. Six aneurysms with similar max AAA diameter at the starting timepoint were visualized within the OxAAA cohort. Mode 4 is inversely correlated with aneurysmal growth. Aneurysms with a not only a greater degree but also a particular direction of localized tortuosity are more susceptible to fast growth. This is similar to what was observed with the visualization of mode 3.

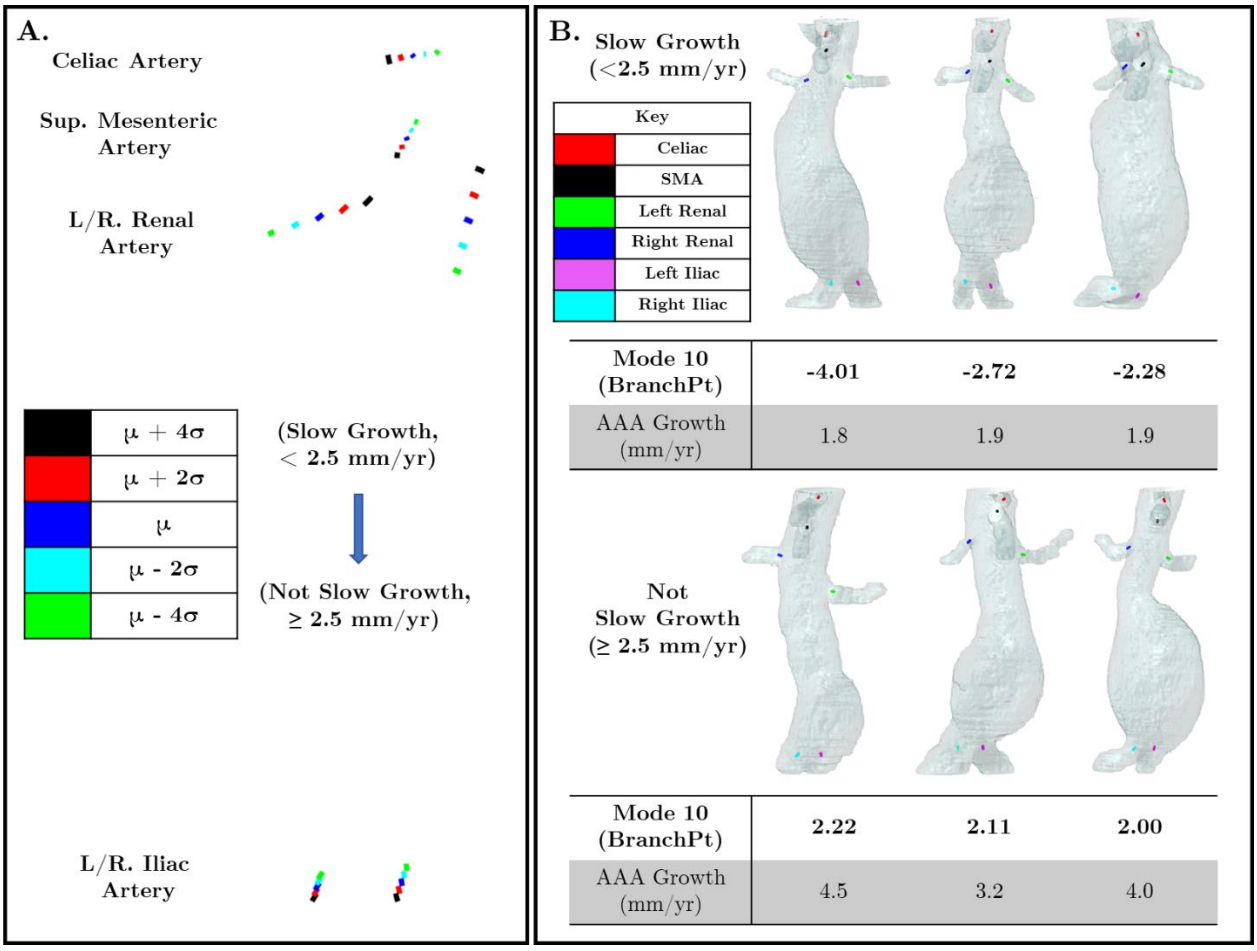


Fig. 66: The influence of Mode 10 (Branch-Point PCA) on average branch locations and its impact on aneurysmal growth (Slow vs Not-Slow Growth). A. Average branch origins (μ, black) were derived from the PCA input matrix and consisted of the first 5 coordinates (X,Y,Z) of each branch centreline. The influence of the selected mode (mode = 10) was applied to the average origins to obtain the modified origins. Here, the origins were calculated at +4σ, +2σ,-2σ and +2σ to visually appreciate the influence of the selected mode. B. Six aneurysms with similar max AAA diameter at the starting timepoint were visualized within the OxAAA cohort. Mode 10 is positively correlated with aneurysmal growth. Increased vertical placement of the left renal artery origin and medial placement of the right renal artery origin (next to the AAA shape) is associated with the slow growth phenotype.

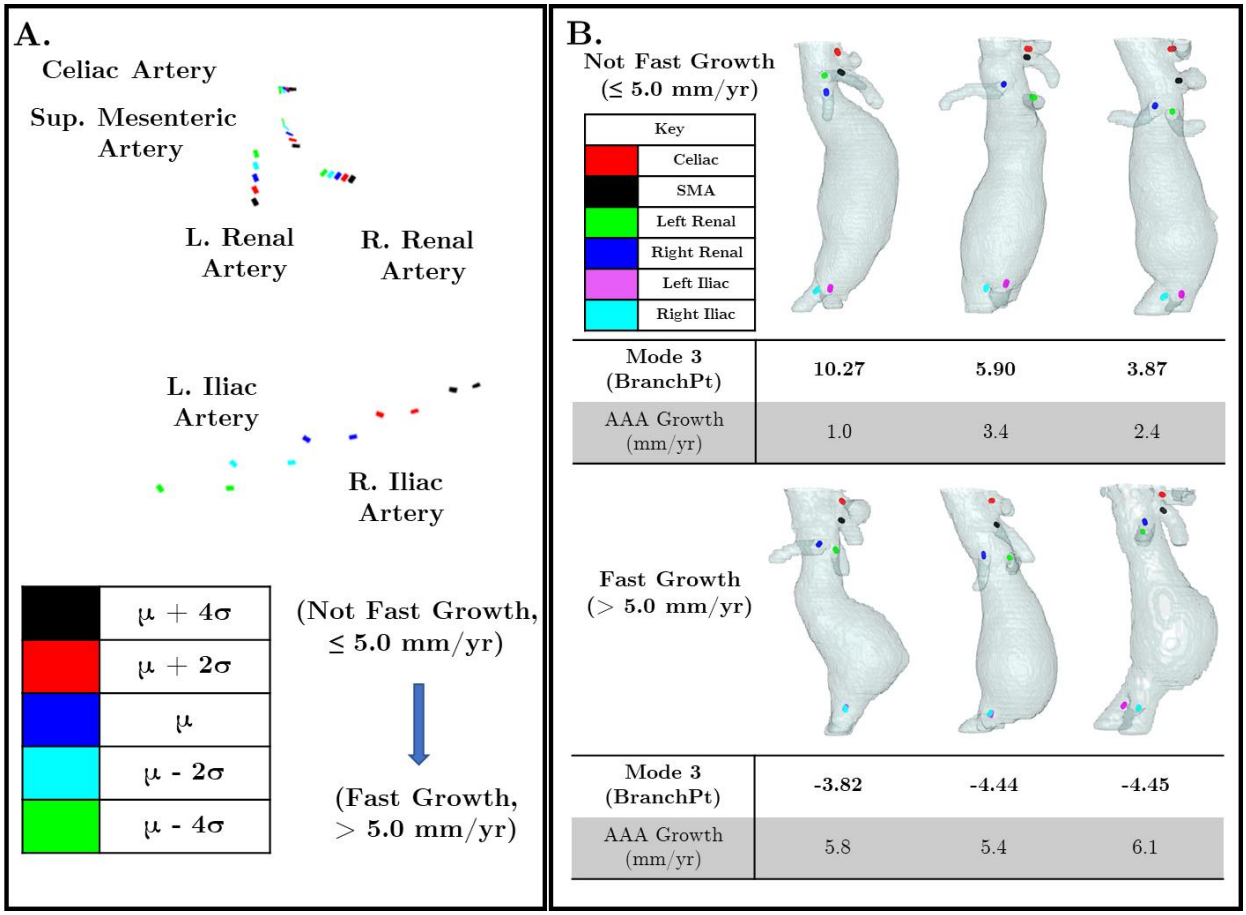


Fig. 67: The influence of Mode 3 (Branch-Point PCA) on average branch locations and its impact on aneurysmal growth (Fast vs Not-Fast Growth). A. Average branch origins (μ, black) were derived from the PCA input matrix and consisted of the first 5 coordinates (X,Y,Z) of each branch centreline. The influence of the selected mode (mode = 3) was applied to the average origins to obtain the modified origins. Here, the origins were calculated at $+4\sigma$, $+2\sigma$, -2σ and $+2\sigma$ to visually appreciate the influence of the selected mode. B. Six aneurysms with similar max AAA diameter at the starting timepoint were visualized within the OxAAA cohort. Mode 3 is negatively correlated with aneurysmal growth. Infero-lateral placement of the iliac artery origin (away from the AAA shape) is associated with the fast growth phenotype.

6.3.5 Prediction of AAA growth rate as a continuous variable

Linear regression models were trained and optimized simultaneously using a 10-fold cross-validation approach to predict AAA growth rate (mm/year) as a continuous variable. With regards to hypothesis-driven features, the linear model trained using Max AP diameter, UI and RC (**Fig. *68*b**) was able to predict annual AAA growth to a greater accuracy than the model trained using only max AP Diameter (**Fig. *68*a**). This was similar to that of the logistic regression model of AAA growth. Growth predictions from this model were significantly correlated ($r = 0.61$, $p < 0.001$) against actual growth measurements. Subsequently, Bland-Altman plots between the hypothesis-driven model output and ground truth measurements indicated more narrow limits of agreements when compared against that of the baseline model (Diameter only). Similarly, the hypothesis-driven model is able to predict annual AAA growth to within 2 mm error in 90% of cases (vs. 69% using the baseline model). Feature coefficients, in addition to its significance within the model, are highlighted in **Table *18***. Of the 3 features evaluated, only UI and RC remained significant. These features were used in subsequent experiments.

The PCA-derived features (Shape-based modes 3,4,7 and Branch-point modes 3 and 10) that were identified through the logistic regression analysis were applied to this problem of predicting AAA growth as a continuous variable. This model was able to predict annual AAA growth to a greater accuracy that trained using only max AP diameter and a similar accuracy to that trained using the hypothesis-driven features (**Fig. *68*c**). Similar limits of agreement were observed between the outputs from the hypothesis-driven and PCA-derived models. Predicted AAA growth was strongly correlated against actual growth measurements ($r = 0.56$, $p < 0.001$) with a mean absolute error between measurements of 1.10 ± 0.78 mm/yr. This model (using PCA-derived features) was able to predict annual AAA growth to within 2 mm error in 88% of cases (vs. 69% using the baseline model). **Table *18*** highlights the significant features within the model. Compared against the shape-based modes, the branch-point modes decreased in significance and were excluded prior to model optimization.

Integrating the selected hypothesis-driven and PCA-derived features further improved model performance as seen in **Fig. *68*d**. The correlation coefficient between predicted and actual AAA growth increased to 0.67 and the limits of agreement reduced when compared against to the

baseline model (Diameter only). The relative importance of each of the features within the model are highlighted in **Table 18**. Finally, this optimized model, can predict annual AAA growth to within 2 mm error in 92.5% of cases (vs. 69% using the baseline model) and within 1 mm error in 73.5% of cases (vs. 52% using the baseline model).

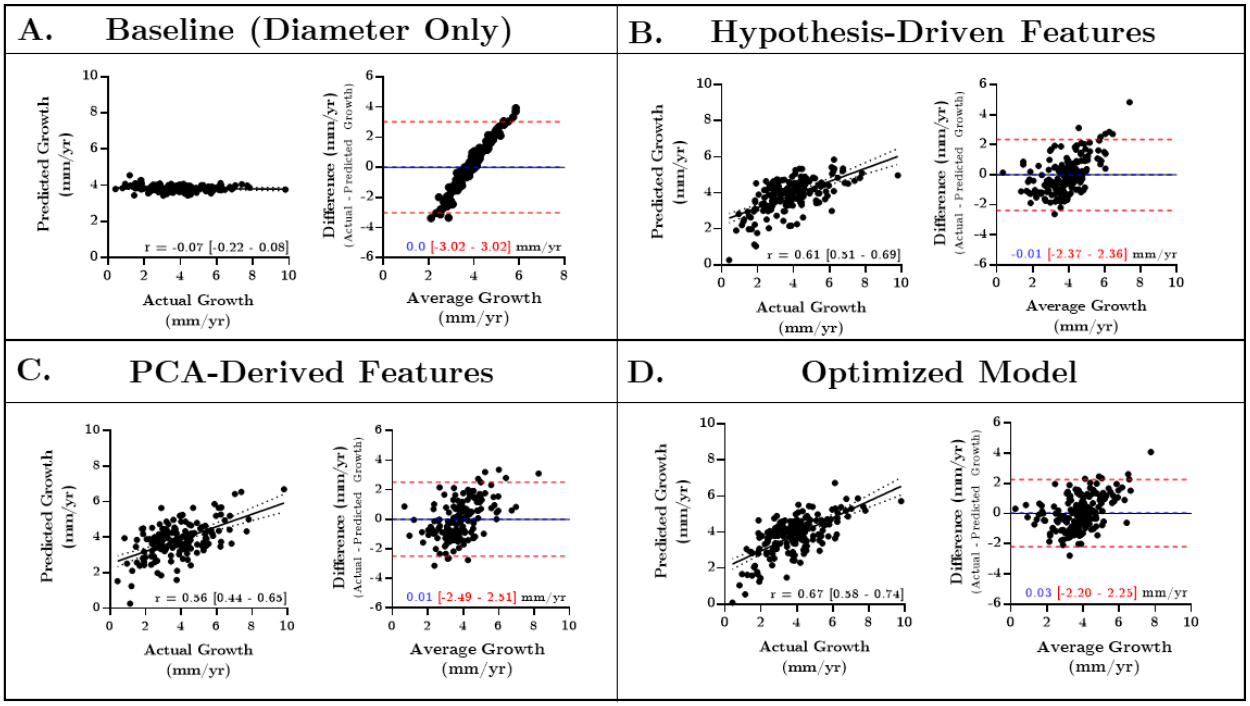


Fig. 68: Outputs of Linear Regression models trained to predict annual AAA growth. Four Linear regression models were trained using 10-fold cross-validation to predict annual AAA growth. The results of each validation fold are reported against actual AAA growth values. The correlation coefficient and its 95% confidence interval are highlighted on each graph (Dotted lines). Bland-Altman analysis was performed for each model to assess the differences between model output and ground truth measurements. Bias (blue) and corresponding 95% limits of agreement (red) are highlighted for each plot. The baseline model consisted of using maximum AP diameter only, as this is a commonly used criteria to monitor AAA progression within the clinic. B-C. Additional models were trained using hypothesis-driven feature (Max Diameter, UI and RC) and PCA-derived features (Shape-based modes 3,4, and 7 and Branch Point-modes 3, and 10). These features were identified to play a significant role in the binary classification of aneurysmal growth. Feature significance was assessed in each model. Significant AAA features were used to train an optimized model that displayed superior performance.

Table 18: Feature coefficients for the four linear regression trained to predict AAA growth.

I. Diameter Only (Fig 19a)		
Features	**Coefficients**	**p**
Max Diameter	0.08 ± 0.08	0.28
MAE (± Std of AE)	1.19 ± 0.97 mm/yr	
RMSE	1.53 mm/yr	

II. Hypothesis-Driven Features (Fig 19b)		
Features	**Coefficients**	**p**
Max Diameter	0.10 ± 0.07	0.16
Undulation Index (UI)	0.28 ± 0.08	**0.001**
Minimum Radius of Curvature (RC)	-0.38 ± 0.08	< **0.001**
MAE (± Std of AE)	1.05 ± 0.75 mm/yr	
RMSE	1.28 mm/yr	

III. PCA-Derived Features (Fig 19c)		
Features	**Coefficients**	**p**
Mode 3 (Shape)	-0.09 ± 0.02	< **0.001**
Mode 4 (Shape)	-0.14 ± 0.03	< **0.001**
Mode 7 (Shape)	-0.18 ± 0.05	< **0.001**
Mode 3 (Branch-Point)	-0.03 ± 0.03	0.44
Mode 10 (Branch-Point)	-0.12 ± 0.03	0.26
MAE (± Std of AE)	1.10 ± 0.78 mm/yr	
RMSE	1.35 mm/yr	

IV. Hypothesis-Driven + PCA Derived Features (Optimized, Fig 19d)		
Features	**Coefficients**	**p**
Undulation Index (UI)	0.17 ± 0.07	**0.02**
Minimum Radius of Curvature (RC)	-0.35 ± 0.06	< **0.001**
Mode 3 (Shape)	-0.07 ± 0.02	**0.001**
Mode 4 (Shape)	-0.10 ± 0.02	< **0.001**
Mode 7 (Shape)	-0.09 ± 0.04	**0.03**
MAE (± Std of AE)	0.93 ± 0.69 mm/yr	
RMSE	1.18 mm/yr	

Note: Significance of each feature is highlighted. Model performance is indicated using mean of the absolute error (MAE) and root-mean-square-error (RMSE) between model prediction and ground truth AAA growth measurements. Full assessment of model performance is highlighted in **Fig. *68.***

6.3.6 Similar AAA Growth predictive accuracy is observed within the independent validation cohort (MA3RS Study)

6.3.6.1 Aneurysmal growth within the MA3RS study dataset

The MA3RS dataset included a total of 148 AAA patients with aneurysmal growth prospectively monitored using serial CT studies. The remaining patients were excluded as 1. AAA follow-up was performed using ultrasound measurements (n = 126), 2. No baseline CT scans were obtained (n = 15) and 3. AAA intervention was performed prior to the 2nd scan (n = 41). Median AAA size as measured by maximum AP diameter for the MA3RS dataset was 53.4 mm with an interquartile range (IQR) of 48.4 to 59.1 mm. This was similar to that observed within the OxAAA cohort (Median AAA Size: 54.5 mm, IQR: 48.0 to 60.0 mm). A secondary CT imaging study was obtained 2 years after the initial study and AAA growth was calculated. Median aneurysmal growth was 2.0 mm/yr with an interquartile range of 1.0 to 3.1 mm/yr. This was significantly less than that observed within the OxAAA cohort (Median AAA growth: 3.7 mm/yr, IQR: 2.5 to 5.0 mm/yr, **Fig. *69*a**). The similarity in AAA size but significant difference in aneurysmal growth patterns highlights a limitation of using AAA size as a sole indicator for growth prediction, in this dataset (**Fig. *69*b**).

6.3.6.2 Geometric feature extraction within the MA3RS study cohort

Extracting UI and RC, highlighted statistically significant differences between the OxAAA and MA3RS study cohorts (***Fig. 69*c-d**). UI was significantly smaller within the MA3RS cohort than in the OxAAA Growth Prediction cohort ($p = 0.02$). On the other hand, RC was significantly greater within the MA3RS cohort ($p < 0.001$).

However, APD, UI and RC of AAAs within the MA3RS dataset display similar relationships with AAA growth as seen with the OxAAA GP Cohort. Within the MA3RS cohort, there were significant positive correlations between AAA size ($r = 0.27$, $p < 0.001$) and UI ($r = 0.45$, $p < 0.001$) with annual AAA growth rate (**Fig. *70*a-b**). A significant negative correlation between minimum RC and annual AAA growth rate was observed ($r = -0.55$, $p < 0.001$, **Fig. *70*c**).

Shape-based and branch point PCA mode extraction was performed on the MA3RS cohort using the methods optimized in section 6.2.4. The relevant modes used for both the logistic and linear regression models, highlighted in section 6.3.4, were isolated from this independent validation cohort and compared against that from the OxAAA-IMG-IV dataset. The results

identified significant differences between shape-based mode 4 (**Fig. *71*a**) and both branch point modes (modes 3 and 10, **Fig. *71*b**). This further highlighted the differences in the aneurysms that were not captured by AAA size alone (max diameter).

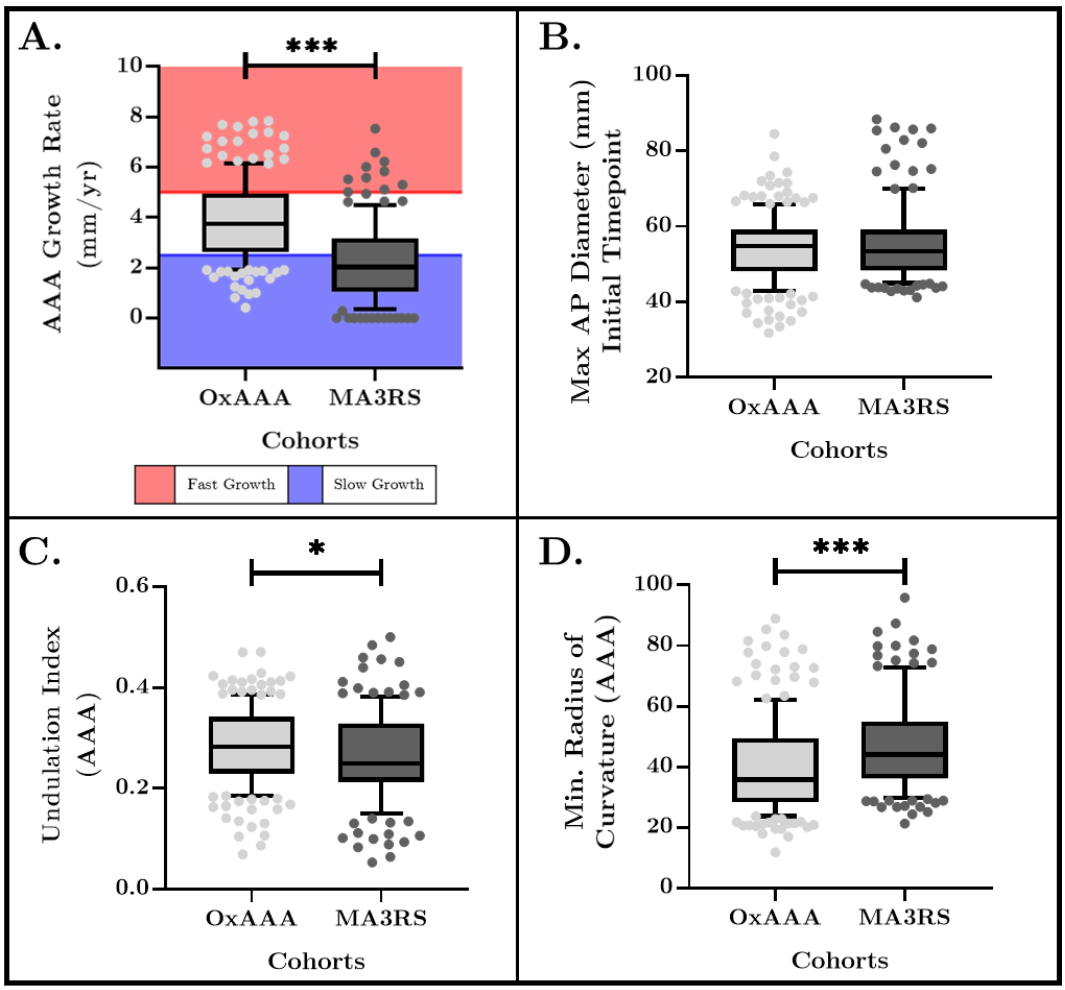


Fig. 69: Comparison of the OxAAA-IMG-IV and MA3RS study cohorts (Hypothesis-driven Features) in preparation for AAA growth prediction. A. AAA growth rates between the two studies were statistically significant ($p < 0.001$). B. Starting Aneurysmal size as measured by maximum AP diameter was similar between the cohorts. C-D. Undulation Index and Minimum radius of curvature (Hypothesis-driven features) were statistically different between the OxAAA and MA3RS cohorts. Aneurysms within the MA3RS study were less asymmetric/undulation (lower UI) and less tortuous (Higher radius of curvature) than those found within the OxAAA cohort.

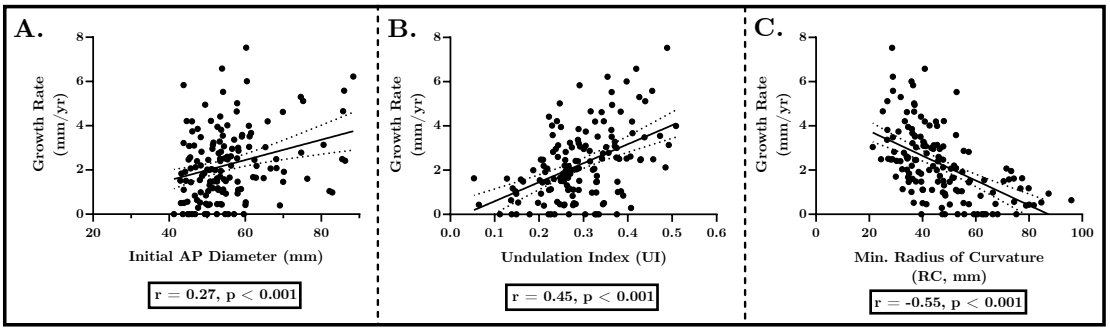


Fig. 70: Linear correlations between extracted geometric featuresfrom AAAs within the MA3RS Dataset **A.** Initial Max AP Diameter or AAA size, **B.** Undulation Index (UI) and **C.** Minimum Radius of Curvature (RC) and estimated annual AAA growth rate). Statistical significance is assessed for each correlation and is indicated on each graph.

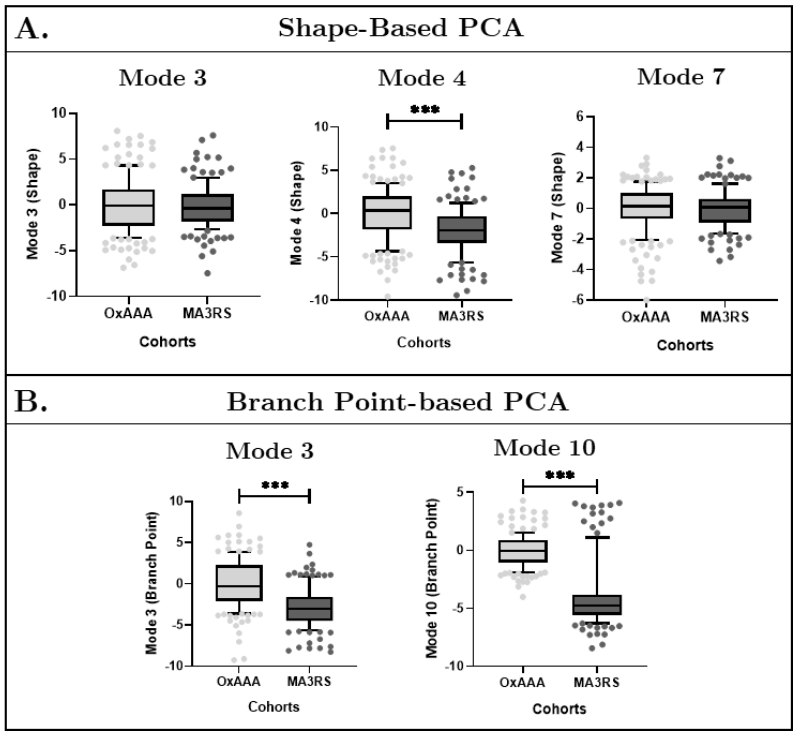


Fig. 71: Comparison of the OxAAA-IMG-IV and MA3RS study datsets (PCA-derived Features) in preparation for AAA growth prediction. A. Shape-based modes were extracted from the MA3RS study cohort using methods optimized in section 6.2.4. Modes 3 and 7 were similar between the OxAAA and MA3RS cohorts. On the other hand, Mode 4 was statistically lower within the MA3RS cohort than in the OxAAA cohort ($p < 0.001$). B. Branch Point modes were extracted from the MA3RS cohort. Both modes were statistically different within the MA3RS cohort than in the OxAAA cohort ($p < 0.001$). This highlights both the diversity of aneurysmal shapes present within the patient population and the limitation of using a AAA size (Max AP Diameter) as the sole indicator to characterize AAA shape.

6.3.6.3 Validation of growth prediction models within the MA3RS dataset

The MA3RS dataset was divided based on AAA growth (Section 6.2.5). The baseline characteristics of the three subgroups of patients [(slow growth: <2.5mm/year); (some growth: 2.5mm to 5mm/year), (fast growth: >5mm/year)] are summarized in **Table *19*.** There was no difference in the starting AAA diameter between the slow (I.) and intermediate (III) groups; however, there was a statistically significant difference between the slow (I.) and fast (III.) groups ($p = 0.01$). UI differed significantly between the three groups (ANOVA, $p < 0.001$). There was also a significant inverse trend of relationship between RC and AAA growth (Kruskal-Willis, $p < 0.001$).

Table 19: Groups split based on annual AAA growth within the MA3RS dataset.

	I. < 2.5 mm/yr (n = 89)	II. 2.5 – 5.0 mm/yr (n = 46)	III. > 5.0 mm/yr (n = 10)	p
AAA Diam$^+$, mm (± SD)	54.3 ± 9.0 mm	55.9 ± 8.8 mm	65.5 ± 14.7 mm	**0.01**
Undulation of Index$^+$ (± SD)	0.26 ± 0.06	0.33 ± 0.07	0.38 ± 0.07	**< 0.001**
Min. Radius of Curvature (IQR)	50.2 (39.5– 63.5)	38.4 (34.1 – 44.5)	33.8 (28.8 – 37.8)	**< 0.001**

Note: Comparison of subgroups is performed to establish differences between groups. Variables that follow a Gaussian/Normal Distribution are indicated with a +. For these variables, mean ± SD are presented, and cohort differences are compared using a one-way ANOVA. For variables that don't follow a Gaussian distribution, median and inter-quartile range (IQR) are presented and cohort differences are compared using a Kruskal-Wallis test.

The trained logistic regression models (6.3.4.3, ***Table 17***) were applied to the MA3RS dataset to predict AAA growth as a categorical outcome. ROC curves on this independent validation cohort were plotted for certain feature combination with the threshold of "Slow (< 2.5 mm) Growth" and "Fast (> 5.0mm) Growth" (**Fig. 72**). The features combinations investigated were identical to that see in ***Table 17***. The area under receiver operation curve (AuROC) metric shows good discriminative capacity of AAA growth rate based on the optimized model (Model IV, Selected hypothesis-driven and PCA-derived features) at the predefined thresholds. The AuROC for predicting slow growth (<2.5mm/year) and prediction fast growth (>5mm/year) using the thresholds defined from the OxAAA-IMG-IV dataset were 0.84 and 0.86, respectively. This optimized model significantly outperforms the use of AAA diameter alone as the predictor for both cases ($p < 0.001$).

Growth predictive accuracy was maintained even after re-defining the slow/fast growth thresholds on the MA3RS dataset based on its bottom (< 1.0 mm/yr) and top quartiles (> 3.0

mm/yr) of AAA growth. The AuROCs for predicting slow (< 1.0 mm/yr) and fast growth (> 3.0 mm/yr) using the optimized models (6.3.4.3, ***Table 17***) were 0.74 and 0.78, respectively.

Similar to the multinomial logistic model, the optimized linear model (Model IV, ***Table 18***) trained using both hypothesis-driven and PCA-derived features, was able to predict annual AAA growth within the MA3RS dataset to a greater accuracy than the model trained using only max AP Diameter (**Fig. 73**

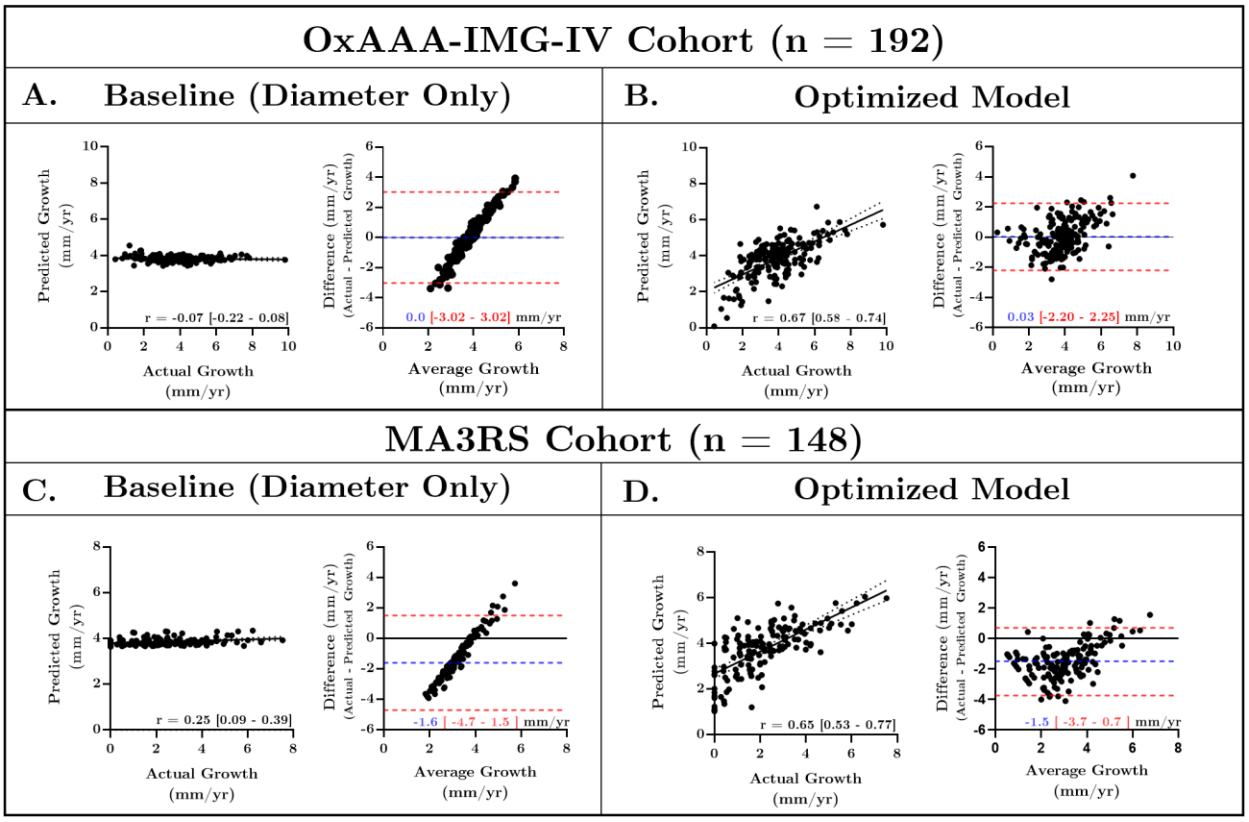


Fig. 73). Predictions from this model were significantly correlated (r = 0.65, p <0.001) and closer (MAE: 1.65 ± 0.94 mm/yr, RMSE: 1.89 mm/yr) to that of observed measurements from the baseline model (MAE: 2.02 ± 1.11 mm/yr, RMSE: 2.32 mm/yr, p =0.002). Although these errors are slightly higher than those observed within the OxAAA-IMG-IV dataset, the results support the ability of this methodology to predict AAA growth using CT-derived aneurysmal geometry.

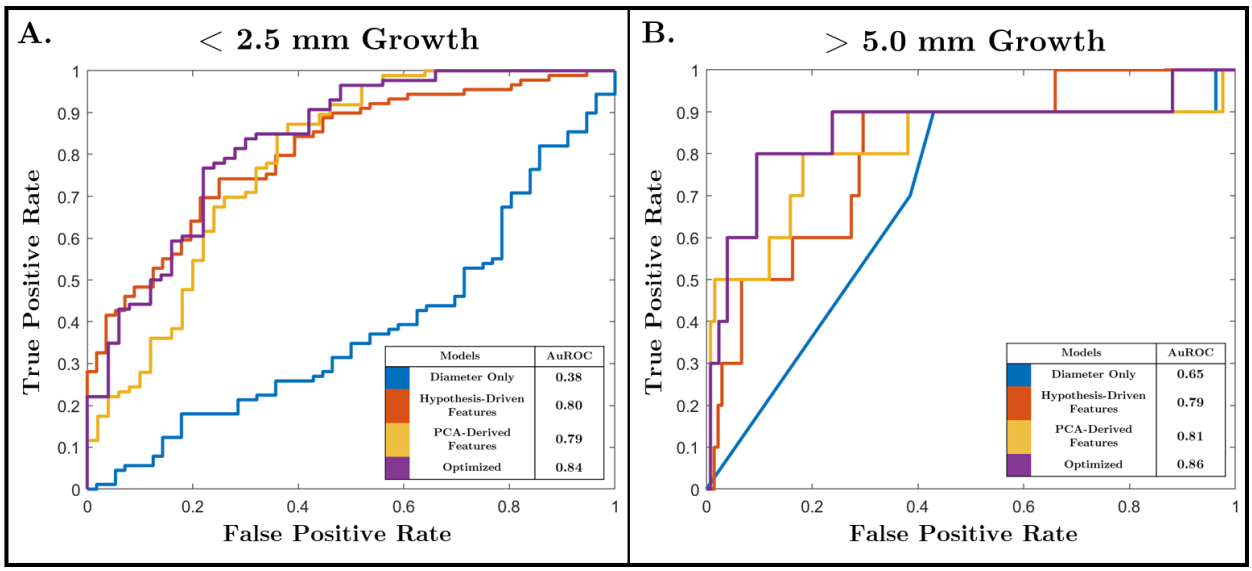


Fig. 72: Receiver operator characteristic (ROC) alongside the area under curve (AUROC) to assess model performance on MA3RS dataset. Models were evaluated in discerning AAA growth phenotype (**A.** > 5.0 mm and **B.** < 2.5 mm Growth**).** ROC curves were generated from the evaluation of the independent validation cohort (MA3RS, n = 148), following model optimization on the OxAAA-IMG-IV dataset. Singificance testing was performed to compare individual ROC curves.

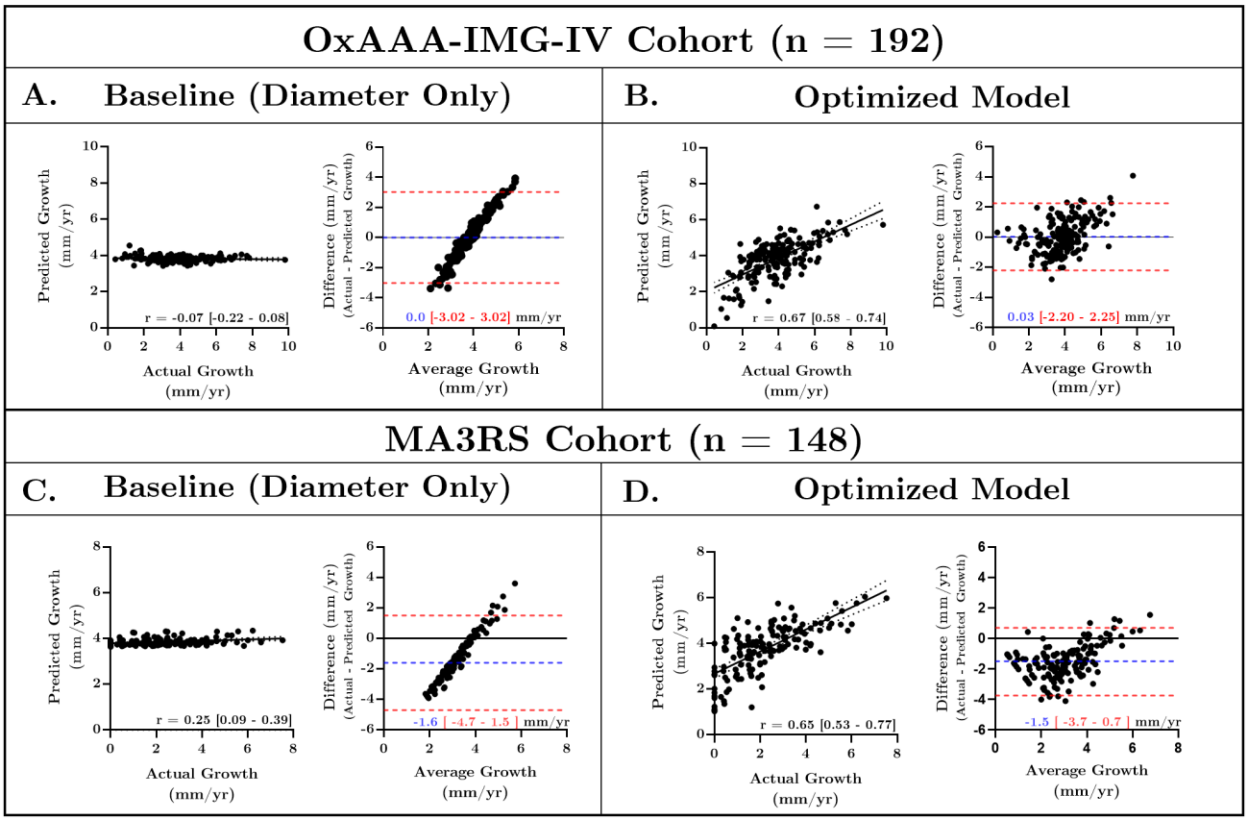


Fig. 73: Outputs of Linear Regression models trained to predict annual AAA growth (MA3RS Dataset). Application of the baseline and optimized linear regression models to predict annual AAA growth on the independent validation cohort (MA3RS dataset) compared against that of the OxAAA-IMG-IV dataset. The correlation coefficient and its 95% confidence interval are highlighted on each graph (Dotted lines). Bland-Altman analysis was performed for both models to assess the differences between model output and ground truth measurements. Bias (blue) and corresponding 95% limits of agreement (red) are highlighted for each plot. A, C. The baseline model consisted of using maximum AP diameter only, as this is a commonly used criteria to monitor AAA progression within the clinic. B, D. The optimized model consisted of a combination of hypothesis-driven (UI and RC) and PCA-derived features (Shape-based modes 3,4, and 7). These features were previously identified to play a significant role in the binary classification and regression of aneurysmal growth on the OxAAA-IMG-IV cohort (A-B).

6.4 Discussion

In the United Kingdom, patients diagnosed with an AAA through the National AAA screening program (NAAASP) are subject to as frequent as 3-monthly surveillance scans[134]. Surveillance of screening-detected or incidentally-diagnosed AAA is also standard clinical practice as recommended by international guidelines[79,107]. Furthermore, in aging populations within developed countries, the health burden of AAA surveillance is expected to rise. As an example, the NAAASP alone incurs a net increase of ~2,000 patients requiring AAA surveillance each year.

Methods to predict future aneurysmal growth are valuable to both clinicians and patients. Such tools can improve the stratification of AAA surveillance frequency in individuals: those with slow growth AAAs should not require as frequent surveillance, whereas intense surveillance (or early intervention) can be justified with an AAA that is likely to exhibit rapid growth. Here, we present a method of AAA growth prediction which utilizes geometric features derived from clinical CT scans.

Recent reports have investigated the use of biomechanical assessments of aortic aneurysms from medical images to predict rupture risk potential. Doyle et al. investigated the use of a ratio between aneurysm wall stress to wall strength, derived from magnetic resonance images (MRI) images, to estimate rupture risk[135]. Their results supported an increase in AAA-related intervention in cases with an elevated ratio after adjusting for AAA diameter and other clinical factors. Additionally, Martufi et al., studied the impact of AAA features (ex. lumen/vessel volume, Intraluminal Thrombus (ILT) thickness, etc) and derived wall stress measurement on AAA outcome prediction[136]. Their results display slight but significant improvements to contemporary methods.

However, such biomechanical models require specific assumptions of physiological / computational variables for each individual patient, which has inherent variability but also may not hold true in real life context[137,138]. For example, the aneurysm biomechanical ratio calculated by Doyle et al is subject to many levels of uncertainty including variability with 1. Image reconstruction accuracy, 2. principal wall stress estimation, and 3. population-based wall strength estimation. This results in significant overlap between results for asymptomatic and symptomatic/ruptured cases and prevents clinical implementation[139]. Therefore, it is difficult to

draw conclusions about this ratio and other similar wall-stress methods for patient-specific prediction of AAA events. This is further supported by Leemans et al., which suggests that biomechanical indices present no added value in the AAA rupture risk assessment[140]. Wall-stress evaluation was not a component of this study and all geometric features were derived directly from the aneurysmal volume.

There is no prior literature on the prediction of AAA growth using complex geometric features extracted from CTs. This has been examined in the context of cerebral aneurysms. Dhar et al., investigated the use of image-based morphological features to predict intracranial aneurysm rupture[125]. Their study highlighted the importance of six features that capture the morphological diversity of intracranial aneurysms and their relation to inflow/outflow vessels. Since then, many of these features have been shown to alter underlying hemodynamics and promote intracranial aneurysmal rupture[125-127]. Although there are many similarities between intracranial and aortic aneurysms, these features have not been investigated in the latter. One such parameter, undulation index captures the degree of surface concavity and increases with surface irregularity[125,126]. A highly asymmetric and/or tortuous infra-renal region would result in an increased UI parameter.

In this study, high undulation seems to arise from a "bent" (correlated with curvature), "bulgy" (a thin cylinder directly attached to a sphere, like the fifth case in Figure 3.) and/or "undulated" shape (two or more spheres attached by cylinders). This parameter is specific to the aneurysmal infra-renal region of the aorta and does not require additional information including inflow/outflow volumes. Similarly, minimum RC is a descriptor of the AAA centreline, which captures the region of maximum curvature. Here, RC is radius of the circular arc that best approximates the curve between a set of adjacent points. The smaller the circular arc, the smaller the RC and the greater degree of local curvature.

The optimized models for "slow" or "fast" AAA growth prediction, highlighted in ***Table 17***, utilize a combination of hypothesis-driven and PCA-derived features. Minimum radius of curvature (RC) is the singular hypothesis-driven feature that is found in both models. The rationale for including RC is included in section 6.2.3.2. Here, we hypothesized, from prior literature, that aneurysms with local regions of increased degree of curvature are more susceptible to rapid aneurysmal growth secondary to alterations in hemodynamic flow/pressure

patterns. This hypothesis is supported by the results obtained from both the OxAAA (**Fig. *56*c**) and MA3RS cohorts (**Fig. *70*c**). Furthermore, we hypothesized that geometric features, in addition to RC, may predispose patients for AAA growth. In this step, PCA was used as the primary method for AAA characterization as it provided 1. a robust method to standardize the apparent aneurysmal diversity and 2. An opportunity to interpret the geometric significance of the significant modes. Multiple methods of aneurysmal characterization were used to maximize the amount of descriptive information captured within the model. The redundant information from each PCA method (ex. Surface-based Modes) was removed during feature reduction/selection prior to model training.

Similarly, the optimized models predict aneurysmal growth as a continuous variable, highlighted in **Table *18***, also integrate hypothesis-driven and PCA-derived features. Interestingly, in this instance, incorporating undulation index improves model information when compared to using RC alone. The predictive capacity of the optimized model is upheld when applied to an independent validation cohort (MA3RS). This underscores the significance of the identified features and their role in AAA growth prediction.

The primary advantage of the geometric measurements described here is that they can be readily extracted from either contrast or non-contrast CT images, without specific adjustment in CT scanning protocols. This analysis is streamlined by our proprietary automated pipeline for high resolution segmentation of blood vessels using deep learning approaches (Section 4)[130]. That no correlation was observed between these geometric features to the patient characteristics (as summarized in **Table 15**) further supports the 'agnostic' nature of these biomarkers as they can be independently deployed as predictive indices without accounting for patients' demographic characteristics. Further validation of our prediction algorithm can therefore be attempted using historic scans already accumulated by vascular surgical units who have an existing clinical image database as part of their routine AAA management practice.

There is emerging literature on the role of best medical therapy (BMT) in reducing overall mortality risk in patients with a AAA, exemplified by the VIVA study long term data[141]. Our method of AAA growth prediction can complement the delivery of BMT in the surveillance pathway. Those predicted to have fast growth should warrant further targeted intensive BMT regime to change their risk profile. Of note, in the prospectively recruited arm of this study, only

a small fraction of patients included were females. We therefore could not rule out an association between these geometric features to the reported demographic features in female patients.

Although AAA surveillance is typically performed using ultrasound scans, many of these patients undergo CT scans for other clinical reasons during the course of their AAA surveillance. These CT scans can be utilised for the added purpose for the prediction AAA growth. (Of note, AAA surveillance is indeed performed using serial CT scans in countries such as Japan.) With the refinement of CT imaging technology and reduction of radiation dose per scan and portable tomographic CTs, it is not implausible for CT scans to replace ultrasound scans as the choice of AAA surveillance in the future. This will facilitate the development and validation of CT image derived prediction algorithm.

6.5 Conclusion

In this chapter, AAA growth prediction models were developed and optimized using geometric features that can be readily extracted from clinically acquired CT scans. As there is well-documented evidence that vascular geometry has a major impact on the development/progression of vascular disease secondary to intravascular hemodynamic forces, we hypothesized that AAA geometry could provide insight into aneurysmal growth. Two hypothesis-driven features, Undulation index (UI) and Radius of curvature (RC), were identified to accurately predict AAA growth as both a categorical and continuous variable when compared to current methods of only using maximum diameter. Here, aneurysms with an increased degree of surface undulation and with local regions of increased curvature are prone to rapid growth. Similarly, principal component analysis (PCA) was used as a semi-supervised method to identify additional relevant geometric features. Growth prediction models using PCA-derived features displayed similar predictive capacity to that using hypothesis-driven features alone. A subset of the PCA-derived features were independent of the hypothesis-driven features and sets the ground for additional exploratory experiments (ex. Hemodynamic impact within the Aneurysm secondary to geometric variations). Integrating the two sets of features improved overall predictive capacity. Finally, the growth prediction models were validated and display similar predictive capacity on an independent cohort (MA3RS study).

6.5.1 Limitations

This study identifies for the first time a set of interpretable geometric features that are able to accurately predict AAA growth in multiple cohorts. However, there are a few limitations that need to be addressed to ensure clinical utility. The accurate extraction of these geometric features relies on the output of the segmentation pipeline detailed in chapter 4. Poor segmentation performance may lead to incorrect AAA shape characterization and poor predictive performance. This underscores the importance of diversifying the data set used for training/optimizing the segmentation pipeline to maximize network performance on unseen data. This may include augmenting the images (aorta/aneurysmal shape or image quality), and increasing the number of annotated images by recruiting additional centres and clinicians. This is described in detail within Section 4.5. Similarly, quality control methods need to be implemented for both model training/optimization and testing. For the former, this includes instituting rigid protocols for personnel training, image acquisition/storage/de-identification, and aorta/AAA annotation. And for the latter, this requires an interactive platform to rapidly evaluate large datasets without requiring time-consuming and labour-intensive 3D-annotations.

An additional limitation is that the CTs within the OxAAA-IMG-IV dataset were retrospectively obtained are devoid of clinical information. Therefore, it was not possible to evaluate the correlation between the geometric features and demographic criteria within this cohort. This comparison was performed on a subset of patients with CT imaging studies within the OxAAA prospective cohort (OxAAA-IMG-III dataset). Furthermore, the number of females was very low, compared to the number of males and no information about environmental factors (diet, smoking habits) and racial variety could be readily extracted from the OxAAA-IMG-IV or MA3RS datasets. This reflects the nature of the disease as AAA is predominantly a disease affecting white males. Racial/environmental imbalances or lack of diversity in the dataset may create bias and limit the generalizability of the results. Similarly, all aneurysms within this cohort were medium-to-large (>50 mm) sized aneurysms at the initial time point and progressed to point of surgery. This is not a common representation of all AAA patients. As a result, a prospective study is required with small-sized AAAs (35-40 mm) to investigate the relevance of these identified features alongside clinical features in predicting aneurysmal growth.

7

Visualization of FDG Uptake without using radioactive tracer enables prediction of clinical outcome in HNSCC

Contents

7.1 Introduction 184
7.2 Methods 185
7.2.1 Patient Population 185
7.2.2 SUV Map Generation from PET Images 186
7.2.3 Patient Contour Segmentation from NCCT/SUV images and Image Registration 187
7.2.4 Tumour Segmentation of the SUV Map for Radiomic Analysis 187
7.2.4.1 Segmentation of tumour and non-tumour tissues (Experiment 1A) 189
7.2.4.2 Regional tumour segmentation based on FDG Uptake (Experiment 1B) 191
7.2.5 Radiomic Feature Extraction from defined segmentations 191
7.2.6 Radiomic Feature Reduction 193
7.2.7 Random Forest classification of metabolic activity based on radiomic signatures 193
7.2.8 Generative Models: Non-Contrast-CT-to-SUV Image Transformation 193
7.2.8.1 Deep Learning Architecture and Model Training 193
7.2.8.2 Model Evaluation: Technical Assessment of Cycle-GAN Simulated SUV-Map Accuracy 194
7.2.8.3 Model Evaluation: Clinical Outcome prediction using simulated SUV maps 195
7.3 Results 196
7.3.1 Patient Population and SUV Map Characteristics 196
7.3.2 NCCT/SUV Registration and Tumour Segmentation 196

7.3.3 Experiment 1A: Radiomic features in NCCT images can differentiate regions of elevated vs negligible FDG uptake 197

7.3.4 Experiment 1B: Radiomics features in NCCT can differentiate high versus low FDG uptake within an individual tumour 199

7.3.5 Experiment 2: Generation of SUV Maps from NCCT 200

7.3.5.1 Technical Assessment of Simulated SUV-Map Accuracy 200

7.3.5.2 Clinical Outcome prediction using Simulated SUV maps 202

7.4 Discussion 204

7.5 Conclusion 207

7.5.1 Limitations 207

7.1 Introduction

Positron emission technology (PET) is an imaging modality that has the ability to visualize abnormal metabolic activity. This is especially important in biological tissues that do not appear pathological based on their morphology[20]. It is a widely adopted clinical tool to differentiate malignant vs benign lesions and in the staging of malignancies. The hallmarks of such malignant tissues are rapid proliferation/angiogenesis, increase in size, local invasion, and distant metastasis[26]. Additionally, PET imaging plays a significant role in the follow-up of patients following chemotherapy and/or surgical resection. It provides clinicians with a semi-quantitative representation of the treatment's impact and can be used to guide further treatment[20,24,25].

At the molecular level, malignant cells have increased glucose utilization due to an upregulation of enzymatic activity. As a result, injection of a glucose-based radionuclide, 2-[fluorine-18]fluoro-deoxy-d-glucose (FDG), can be used to identify these abnormal metabolically-active tissues. The rate of uptake of FDG into malignant tissues has been shown to be proportional to its metabolic activity[20]. However, unlike glucose, FDG is not fully metabolized and becomes trapped within active cells. This accumulation is what is observed in a PET image.

Commonly, PET images are obtained alongside a non-contrast computerized tomography (CT) images to enable the localization of areas of increased metabolic activity with its underlying anatomic structures. Co-registering functional (PET) and anatomic (CT) information has improved clinical confidence in decision making[23,27,28]. The advantage of obtaining a paired PET-CT study has been repeatedly recognized by healthcare professionals for the care of oncology patients. They have been accepted as routine methods for the diagnosis, staging and follow-up for a variety of malignancies (pulmonary nodule[21], melanoma[22], head and neck squamous cell carcinoma[14,23], etc).

Although the advantages for PET/CT imaging are quite striking, this technique has multiple limitations. Following radionuclide injection, patient activity and speech are limited for 20 minutes to minimize physiologic uptake by muscles and imaging is initiated approximately 60 minutes afterwards[20]. Depending on the patient's prior medical history, bowel cleansing[142] and/or bladder catheterization[143] may be required. The CT study takes approximately 60-70 seconds to complete, whereas the PET study takes 30-45 minutes, depending on the

coverage[20,25]. As a result, another limitation is patient motion between the PET and CT imaging studies[144,145]. Motion between imaging studies can prevent proper co-registration and decreases the clinical value of the obtained images. Additional limitations include: 1. additional radiation exposure, and 2. intrinsic patient variability (ex. patient weight, basal metabolic rate, radionuclide dose, duration between injection and imaging, etc.)[146]. The latter is minimized via the calculation of Standard Uptake Value (SUV) maps, which are essentially PET images standardized by patient weight, radiation dose and the interval between injection and imaging. PET/CT imaging is also an expensive and specialist imaging modality that is associated with alarming global inequities. According to the IMAEA Medical imaging and Nuclear Medicine global resources (IMAGINE) database, the population served by 1 PET-CT scanner varies based on the income status of the target country (High income: 601,000 people, Middle income: 3,484,000 people. and Low income: 166,667,000 people)[8].

Malignant tissues at the molecular level are significantly different from healthy tissues, in terms of ultrastructure, tissue organization and metabolic activity[26]. We recently reported the development of a DL pipeline to enable the segmentation and visualisation of blood vessel anatomy/pathology in NCCT without the use of intravenous contrast agent[11,147]. We showed that the raw data captured in a NCCT scan contains sufficient information to differentiate circulating blood from anatomical (vessel wall) or pathological (luminal thrombus) structures.

Here, we hypothesised that the raw data acquired from a NCCT contains sufficient information to differentiate malignant and healthy tissue regions. We further hypothesized that DL generative methods could be used to amplify these subtle differences between tissues in order to obtain a visualization of FDG-uptake without the injection of a radioactive tracer (such as FDG). This enhancement can be used to supplement the NCCT and improve the detection of metabolically-active/malignant lesions without the need to subject patients to PET imaging.

7.2 Methods

7.2.1 Patient Population

In this study, we utilised a collection of paired FDG-PET and CT images of 298 patients with diagnosed head & neck squamous cell carcinoma (HNSCC) prospectively recruited from four different institutions in Quebec, Canada. Vallières et al.[14] utilised this clinical cohort to

investigate the impact of radiomic methods for the risk assessment of tumour progression. Details of this previous study, including patient characteristics and clinical outcomes for each of the patients, are as published and also available through the TCIA repository.

7.2.2 SUV Map Generation from PET Images

Standard uptake value (SUV) is a mathematically derived ratio of tissue radioactivity concentration (A_C) from the PET image that is standardized to the patient's body weight (W), the initial radiation dose (D), half-life of radioactive tracer and the time duration between bolus injection and imaging (Δt)[146].

$$SUV = \frac{A_C}{D * 2^{\left(\frac{-\Delta T}{T_{\frac{1}{2}}}\right)}} * W \qquad \text{(Eq. 1)}$$

Although vulnerable to variability (ex. image noise, low image resolution, region-of-interest-input), this semi-quantitative calculation is a common technique used to standardize PET image comparison between multiple patients and cohorts[20,25]. **Fig. *74***. highlights a paired PET and SUV map image.

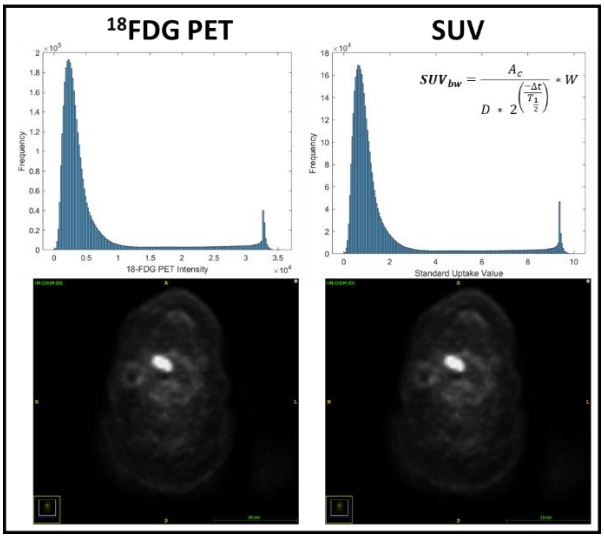


Fig. 74: Conversion of a FDG-PET image to Standard Uptake Value (SUV) map. The formula for this transformation incorporates the patient's body weight (W) in kg, the dose of ^{18}FDG (D), the time duration between ^{18}FDG injection and imaging time (Δt) in seconds and the half-life of ^{18}FDG. The Ac is the pixel intensity within the ^{18}FDG-PET image. The resulting SUV map has the same distribution as that of the original image but is bounded by 0 and 10.

7.2.3 Patient Contour Segmentation from NCCT/SUV images and Image Registration

Direct comparison of NCCT with the PET/SUV images require the two images to be registered and display considerable overlap. This allows for sufficient localization of the PET/SUV map to the anatomical framework provided by the NCCT image. The clinical value of the subsequent experiments relies on the degree of registration between images. Registration accuracy was assessed by comparing the patient contours obtained from the NCCT and PET/SUV images.

Binary segmentation of the SUV map was generated in MATLAB using threshold-based methods (SUV $\geq$ 0.1, 1 [patient]; SUV $<$ 0.1, 0 [background]). The SUV threshold of 0.1 was defined as the smallest positive SUV within the map. Subsequent morphological dilatation and erosion operations using a spherical structuring element within the interior of the patient was used to connect components. 3D-gaussian filtering on the segmented images (sigma, 2) was used to smooth all SUV-map derived patient segmentations (**Fig. *75*a**). Similar threshold-based and dilation/erosion methods were used to generate the segmentation from the NCCT (Hounsfield Unit, HU $\geq$ -500, 1 [patient]; HU $<$ -500, 0 [background], **Fig. *75*b**). The CT threshold of -500 was defined empirically as it is able to capture all soft-tissue regions within the anatomical CT region. Registration accuracy was assessed using the Sorensen-Dice (DICE) coefficient, which is a ratio comparing the similarity between segmentations (2 * elements common to both images/ the total number of elements, **Fig. *75*c**). If the DICE score was less than 90%, a non-rigid b-spline registration algorithm was implemented on the segmentations to ensure registration accuracy. This registration method was used in our previous work[11]. Finally, this method was used to isolate the patient from the underlying table in the NCCT (yellow arrows, **Fig. *75*d**).

7.2.4 Tumour Segmentation of the SUV Map for Radiomic Analysis

The primary focus of this experiment was to characterise and compare the radiomic signatures of highly metabolic tumours with elevated FDG uptake (increased SUV) against regions of low/negligible FDG uptake. Therefore, it was essential to isolate these metabolically-active areas within the patient.

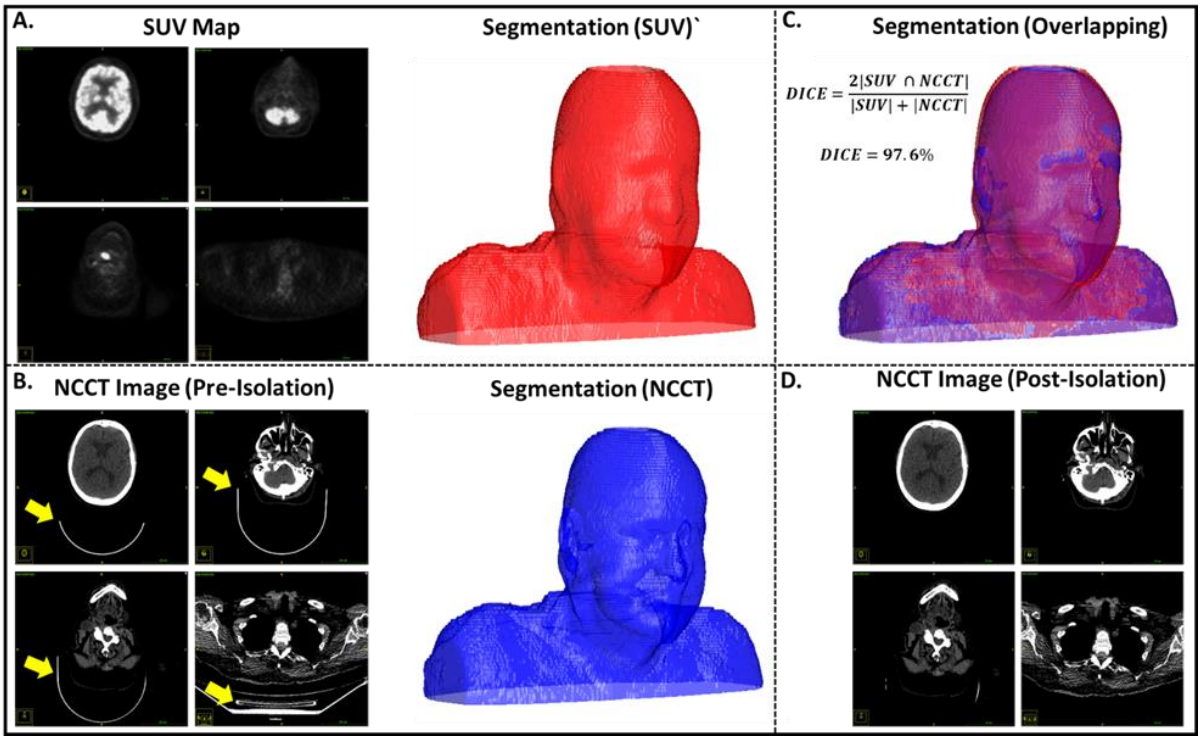


Fig. 75: Patient Surface Contour Segmentation from Non Contrast CT and SUV Images and Image Registration. **A-B.** The surface contour of each patient's head was segmented from both the NCCT and SUV using threshold-based methods followed by morphological dilatation/erosion using a spherical structuring element in MATLAB. Generated segmentations were smoothed using a gaussian filter (sigma = 2). **C.** Registration accuracy was assessed using DICE score overlap. **D.** Additionally, in the NCCT image, the generated segmentation was used to remove the table (yellow arrow). The post-isolated NCCT and SUV images are input for subsequent investigations.

To perform the tumour segmentation using the PET-derived SUV map, the following assumptions were made: 1. the SUV map and NCCT image are registered and display considerable overlap, 2. The largest connected component in the SUV map that displays elevated FDG uptake is the brain, and 3. Tumour areas are highly metabolic and display increased FDG uptake. A patient-specific threshold was empirically-defined based on the maximum SUV (0.35 x Maximum SUV) to isolate regions with increased or elevated SUV. This threshold was defined as it was able to sufficiently differentiate tumour regions from the background noise. A convolution filter of all ones with a 3x3 pixel kernel size was used to smooth the initial segmentation output (Fig. 76a). Subsequently, connected component analysis was used to separate the brain from other regions of elevated FDG uptake (Fig. 76b). All generated segmentations were assessed for manual overlap with the SUV Map and any minor

adjustments were made. This segmentation method was identical to that referenced in Vallieres et al and was rigidly evaluated by a team of radiation oncologists[14].

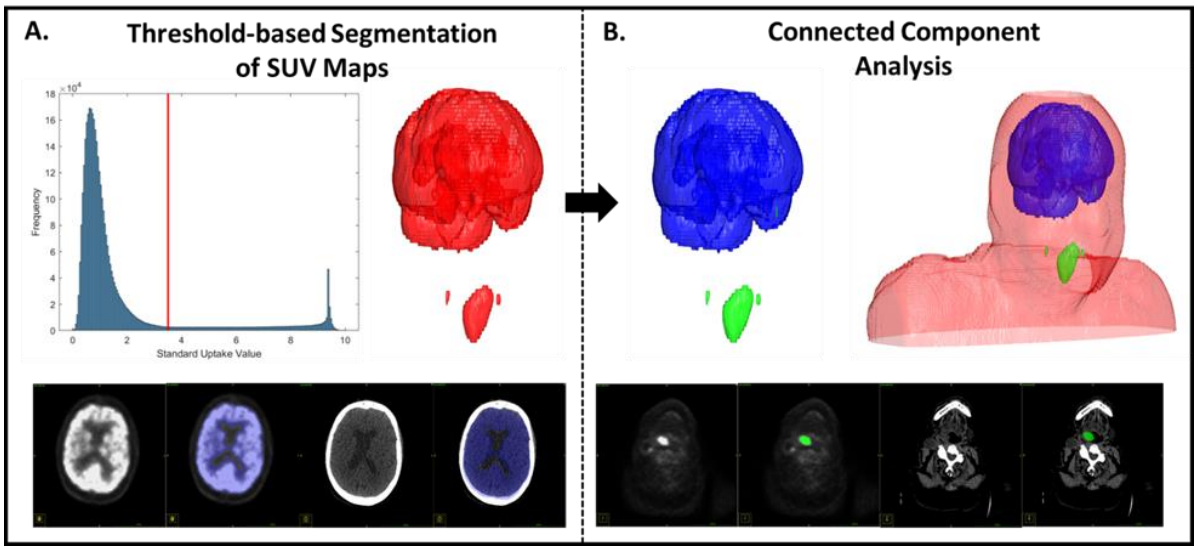


Fig. 76: SUV-Map-based Tumour Segmentation for Radiomic Analysis. A. Threshold Based Segmentation of SUV maps was performed by using an empirically-defined threshold. Segmentations were smoothed using a convolutional filter (kernel size, 3x3 pixels). **B.** Connected component analysis was performed to isolate the brain from other hotspots, which are characteristic of the tumour (primary and/or lymph nodes). The extracted segmentations are visualized along with their respective NCCT image.

7.2.4.1 Segmentation of tumour and non-tumour tissues (Experiment 1A)

Experiment 1A aims to investigate the radiomic differences between regions with elevated FDG uptake (+SUV, ie tumour) and regions of low/negligible FDG uptake (-SUV, ie non-tumor). Regions of negligible FDG uptake include tissues immediately adjacent to the segmented tumour region and thyroid tissue. Thyroid tissue was manually segmented and used as a biological comparison. **Fig. *77*** illustrates thyroid and tumour segmentations from two patients. We reasoned that thyroid tissue has low metabolic activity, in comparison against tumour tissues, and has resemblance to lymph nodes in terms of macroscopic structure and visually on a NCCT image.

To sample adjacent regions to the tumour, the tumour surface was dilated by a factor of 2. The centroid of the expanded segmentation mask was matched to that of the original segmentation. This resulted in two concentric segmentations with equal volumes (**Fig. 78a**). Thyroid segmentation was performed by myself and a trained clinician (ENT clinician, JW - with multiple years experience in reading and annotating Head-and-Neck PET-CTs) directly on

the NCCT image using an open-source segmentation software, ITK-Snap. The overlap between the segmentations were used for subsequent analysis.

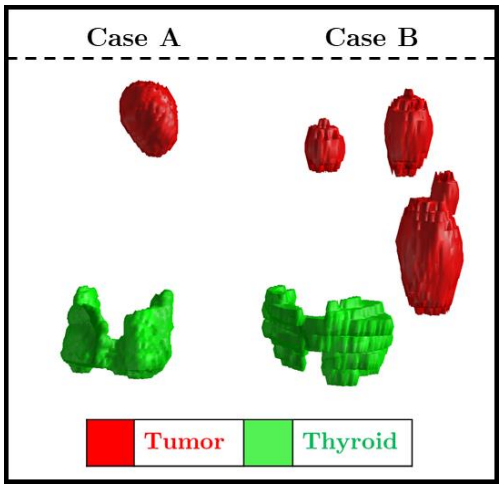


Fig. 77: Visualisation of the thyroid (green) and tumour (red) regions from two patients within the dataset. The thyroid was used as a biological comparison to tumour tissue as it is similar in composition/structure to unaffected/non-tumorous lymph nodes. We hypothesized that these two visually-indistinct regions displayed unique radiomic signatures.

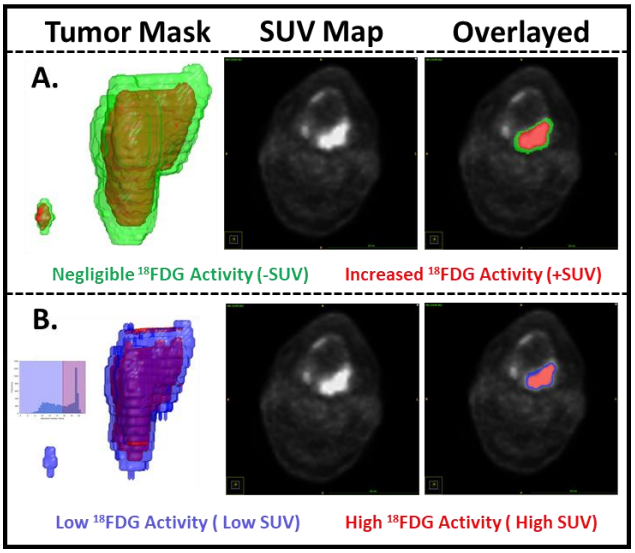


Fig. 78: Regions of different FDG uptake avidity as defined by the PET SUV map. A. For Experiment 1A, the tumour contour (red) was dilated by a factor of 2 at its centroid. The two concentric regions represent areas of elevated FDG uptake (red) and negligible FDG uptake (green), respectively. **B.** For Experiment 1B, the tumour was divided into two sub-regions based on the FDG uptake avidity, using 50th Percentile of SUV (SUV_{50}) as the threshold.

7.2.4.2 Regional tumour segmentation based on FDG Uptake (Experiment 1B)

Experiment 1B aims to characterise the radiomic differences within different regions of a metabolically-active tumour. Tumours were divided into two sub-regions based on the FDG uptake avidity: 1. High FDG uptake ($\geq 50^{th}$ percentile of SUVs [SUV_{50}] within the tumour), and 2. Low FDG uptake ($< SUV_{50}$). The constrained sub-region represents an area of higher FDG uptake within the tumour volume (**Fig. 78b**). The SUV_{50} was specific to each patient and allowed for the differentiation of FDG uptake within the tumour.

7.2.5 Radiomic Feature Extraction from defined segmentations

Anisotropic image and segmentation masks were resampled into isotropic-sized voxels (Isotropic settings: 1mm, 2mm, 3mm, 4mm and 5 mm) in MATLAB. Parameter settings for radiomic feature extraction included 5 pre-defined histogram bin widths (5,10,15,20 and 25). All radiomic features were extracted using Pyradiomics, an open-source python package[55]. For each set of image and parameter settings, 18 first-order, 68 second-order and 1118 filter-based features were calculated. This results in a total of 30,125 features for each region of interest (ROI) ((86 $1^{st}/2^{nd}$ order features + [86 * 13 filtered images]) x 5 Isotropic Settings x 5 Bin-width setting). More information regarding the description of extracted radiomic features can be found within the appendix. The full pipeline for Experiments 1A and 1B are illustrated in **Fig. *79***.

In *Experiment 1A*, radiomic features were extracted from tumours with increased FDG uptake and regions with negligible FDG uptake (1. adjacent to tumour and 2. thyroid tissue). Similarly, in *Experiment 1B*, radiomic features from regions of high and low FDG uptake within metabolically active tumours were extracted. Following feature extraction, patients were divided into training (n = 194) and testing (n = 104) cohorts. This split was identical to that performed by Vallieres et al[14]. Given that each patient may have multiple tumour hotspots, train and test cohorts were divided based on patient to prevent data leakage. This was important to prevent different tumour hot spots from a single patient appearing in both the training and testing cohorts. Feature selection, model training and optimization were performed on the training cohort. The testing cohort was introduced to evaluate model performance.

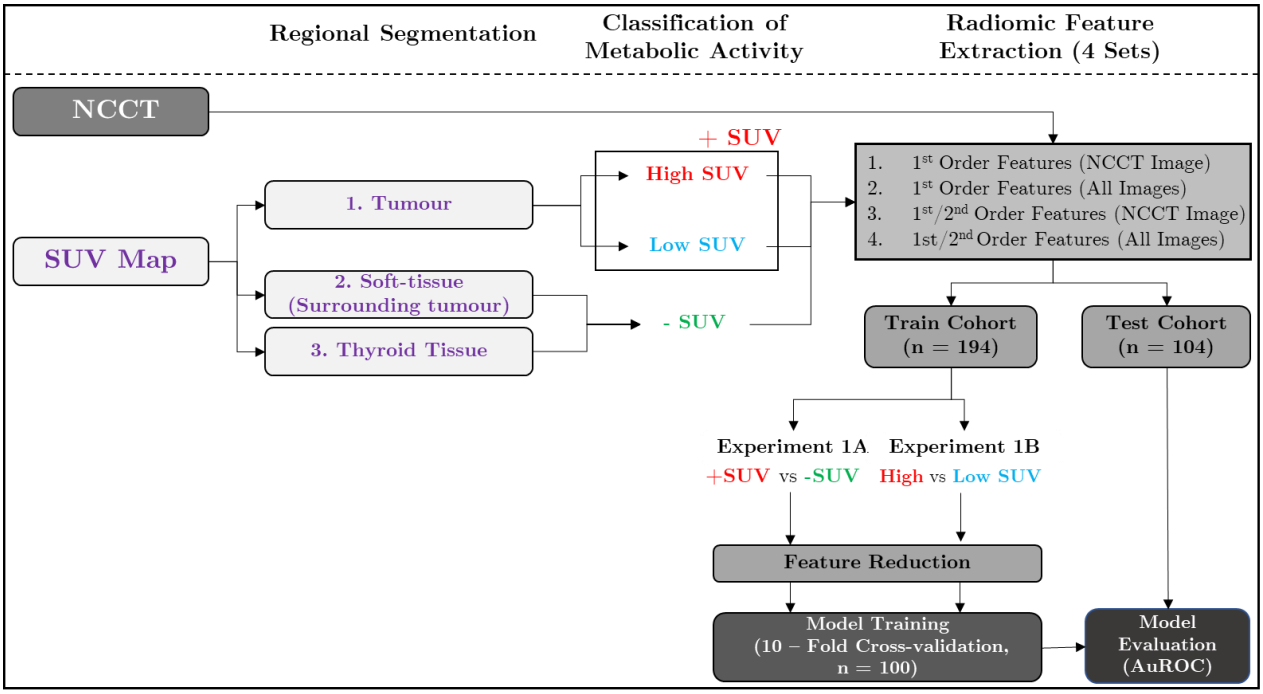


Fig. 79: Workflow for the classification of volumes extracted from NCCT images based on metabolic activity using radiomic signature (Experiments 1A and 1B). Tumour and adjacent soft tissue were segmented from the SUV map using threshold-based methods and the thyroid tissue was manually segmented on the paired/registered NCCT image. Volumes were classified based on metabolic activity into three categories (High SUV, Low SUV and negligible SUV). High/Low SUV were localized within the tumour volume. Regions of negligible SUV included the soft-tissue surrounding the tumour and thyroid tissue. Four sets of radiomic features were extracted within the segmented volumes from either the NCCT image or All Images (NCCT + Filtered Images). Filtered images included the NCCT image with applied Laplacian of Gaussian and Wavelet filters. Full details regarding image filtering can be found within the appendix. Feature reduction was performed in MATLAB using the minimum redundancy, maximum relevance (MRMR) algorithm. 10-fold cross-validation was performed (n= 100) and each of the validated models were applied on the testing cohort.

7.2.6 Radiomic Feature Reduction

For *Experiments 1A and 1B*, four different models were trained using a different combination of radiomic features. These models include:

1. First Order Features from the NCCT Image
2. First Order Features from the NCCT + Filtered NCCT Images
3. First/Second Order Features from the NCCT Image
4. First/Second-Order Features from the NCCT + Filtered NCCT Images

For each model, features were ranked using the minimum redundancy, maximum relevance (MRMR) algorithm in MATLAB. The top 25 features for each model were selected for model training and optimization. Feature extraction specifics for Experiments 1A and 1B can be found within the appendix.

7.2.7 Random Forest classification of metabolic activity based on radiomic signatures

For each experiment, models I to IV were trained on the training cohort of 194 patients with the appropriate feature set using a 10-fold cross-validation approach. Prediction performance was estimated on both the training/validation and previously established testing cohorts using receiver operating characteristic (ROC) curves. Area under the ROC (AuROC) was calculated to compare model performance.

7.2.8 Generative Models: Non-Contrast-CT-to-SUV Image Transformation

7.2.8.1 Deep Learning Architecture and Model Training

A generative adversarial network (GAN) was used for this non-contrast to SUV image transformation task. These networks are a class of DL architectures whereby two neural networks train simultaneously, with one network focused on data generation (generator) and the other focused on data discrimination (discriminator). In this instance, these networks compete against each other to better learn the underlying statistical distribution of the training data. This allows for the generation of new examples from the same distribution. Here, we implement a Cycle-GAN, which can learn transformations between two distributions without the need for direct pairings between samples. We had previously applied the Cycle-GAN architecture for a

similar medical image transformation task as seen in Section 5.3.5 and show its superiority over other GAN architectures (e.g. Con-GAN).[11]

Specifics regarding model architecture and associated training details are described in the appendix. A 3-fold cross-validation paradigm with a training/test data split of 200:98 patients (~8,400: ~3,900 2D axial slices) was employed. Within each training cohort, 50 of the 200 patients were used to internally validate during model training/optimization. The optimized model was then evaluated against the testing cohort. It is important to note that the training/test data split for this experiment is different than previously used. This data split and training paradigm was used to maximize both the training and testing datasets, allowing all NCCT/SUV images to be used for training and testing. For this experiment, SUV maps were inverted as this is the view commonly used by clinicians.

Model performance during training and validation were evaluated using Root-Mean-Square-Error (RMSE) difference between the simulated and the gold standard SUV map. This metric is widely used in image transformation tasks as it evaluates the pixel-to-pixel differences between image pairs.

7.2.8.2 Model Evaluation: Technical Assessment of Cycle-GAN Simulated SUV-Map Accuracy

Tumours within the Cycle-GAN Simulated SUV maps were segmented using the same threshold-based segmentation criterion as used for the ground truth SUV maps as described above. Technical accuracy of the simulated SUV maps were assessed by extracting criteria supported by the PET Response Criteria in Solid Tumours (PERCIST, version 1), which is used to monitor tumour progression and response to treatment[148]. Within each of the tumours defined by the simulated SUV map, four clinically-important metrics were extracted and compared against that of the ground truth: 1. Minimum SUV(SUV_0), 2. SUV at the 50^{th} percentile (SUV_{50}), 3. Maximum SUV (SUV_{Max}) and 4. Tumour burden/volume (in mm^3). Bland Altman plot and correlation coefficient analysis was performed for each testing fold to compare the values obtained from the simulated SUV map and that from the ground truth SUV map.

7.2.8.3 Model Evaluation: Clinical Outcome prediction using simulated SUV maps

Using the Cycle-GAN simulated SUV maps, Random forest models were constructed to predict three clinical outcomes (1. locoregional tumour recurrence, 2. distant metastasis, 3. survival). The primary objective of this experiment was to compare the predictive accuracy using the Cycle-GAN simulated SUV maps against the ground truth. This analysis mirrors that performed by Vallières et al[14] and is visualized in **Fig. *80***.

For each patient, a total of 2,150 radiomic features were extracted. It is important to note that for this experiment, tumour hotspots were grouped by patient and not individually analysed, as was done in Experiments 1A and 1B. Identical training (n = 194) and testing splits (n=104) were implemented for model training and evaluation. The process of integrating the radiomic features into a multivariable model was achieved using the logistic regression utilities of the software DREES[14,149]. Stepwise feature set reduction and selection methods were implemented.

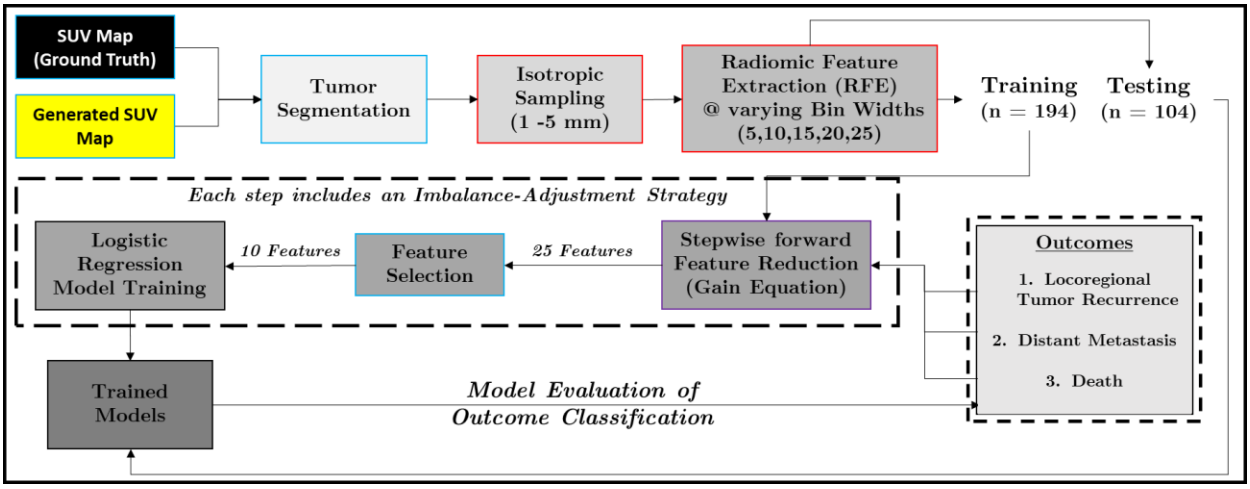


Fig. 80: Pipeline for the Clinical Evaluation of Simulated SUV Maps. This pipeline is based on the work performed by Vallières et al, which focused on the application of SUV maps for the prediction of three clinical outcomes: 1. Locoregional tumour recurrence, 2. Distant Metastasis and 3. Death. The pipeline consists of radiomic feature extraction from the tumour regions within the SUV Map. Feature reduction, selection and model training were performed on the training cohort using an imbalance-adjustment strategy that was identical to Vallieres et al[14]. Optimized models were evaluated on the testing cohort.

Additional information regarding the feature set reduction and selection can be found in the supplement. For each set of features, predictive performance was estimated and the top 3 parsimonious models were chosen for each outcome. The selected parsimonious models for each feature set (1. Ground truth SUV map, and 2. Cycle-GAN Simulated SUV Map) and each of the three outcomes was directly tested on the pre-defined testing set. Model performances between the simulated SUV maps and that of the gold-standard were compared to assess the predictive capacity of the simulated images.

7.3 Results

7.3.1 Patient Population and SUV Map Characteristics

Imaging (PET, CT) data from 298 patients with diagnosed HNSCC were available on TCIA. The patients underwent routine treatment management (radiation – 48, 16%; chemo-radiation – 252, 84%). Imaging was obtained within a median of 18 days (range 6 – 66) prior to the start of treatment[14].

At the time of imaging, the median patient weight was 75 Kg (range – 43 -142 Kg) and the median dosage of FDG-PET injected was 1.65 x 10^8 Bq (range – 3.81 x 10^8 – 31.82 x 10^8). Additionally, the median duration between injection and scan time (Δt) was 1.80 x 10^4 s (range – 1.04 x 10^4 – 3.01 x 10^4). Each patient-specific combination of weight, dose, and Δt, along with the half-life of FDG (6588 s^{-1}) was used to calculate the SUV map. Average SUV within the calculated images is 0.19 ± 0.06. SUV maps were derived from the provided PET images to standardize measurements between patients.

The median follow up time after treatment was 43 months (range – 6 -112 months). Of the 298 patients, 45 patients developed locoregional recurrence, 40 patients developed distant metastasis and 56 patients died. Additional information regarding the patient cohort characteristics can be found within the previously published data documentation[14].

7.3.2 NCCT/SUV Registration and Tumour Segmentation

The accuracy of registration between the NCCT and SUV images was 95.1 ± 1.9%. This was assessed by the DICE score between the surface contour of each patient derived from the NCCT and PET images (**Fig 75**). Tumour segmentations were obtained from the PET-derived SUV map. From 298 patients, 683 hot spots of elevated FDG uptake (elevated SUV, 6.03 ± 1.71)

were isolated, which are characteristic of metabolically active tumours (primary and/or metastatic lymph nodes). The brain, which is highly metabolic was excluded from the derived segmentation. These derived segmentations would serve as the ground truth for subsequent experiments.

7.3.3 Experiment 1A: Radiomic features in NCCT images can differentiate regions of elevated vs negligible FDG uptake

The primary aim of Experiment 1A was to investigate whether first-order radiomic features (ex. Hounsfield Unity Intensity) from NCCT images could be used to define areas that correspond to different FDG uptakes in the paired PET/SUV map. SUVs were significantly higher within the tumour when compared against non-tumour tissue (6.03 ± 1.7 vs 3.21 ± 1.00, $p < 0.001$, **Fig. *81*a**). In the CT images, the average Hounsfield unit intensity within the tumour was less than that of the adjacent non-tumour tissue ($p < 0.01$). This indicates that there may be a difference, albeit subtle, in the radiomics signature between the two regions. Similar results are observed when comparing tumour with that of the thyroid (**Fig. *81*b**).

Random forest models (***Experiment 1A*:** Models I – IV) were trained on a combination of first and second-order radiomic features extracted from the CT to classify regions with increased or negligible SUV (**Fig. *82*a-b**). For the task of classifying tumour vs non-tumour tissue, Model I (First Order – CT) had an AuROC of 0.87 ± 0.1. Model performance improved with the introduction of first order features from filtered images (Model II, First Order – CT + Filter-Based, AuROC – 0.93 ± 0.1, $p < 0.001$). The incorporation of matrix-based radiomic features further improved classification performance (**Fig. *82*a**). With regards to tumour vs thyroid tissue, the AuROC for Model I was 0.94 ± 0.11. Although model performance increases with feature complexity (Models II – IV), the difference in AuROC was not statistically significant (**Fig. *82*b**).

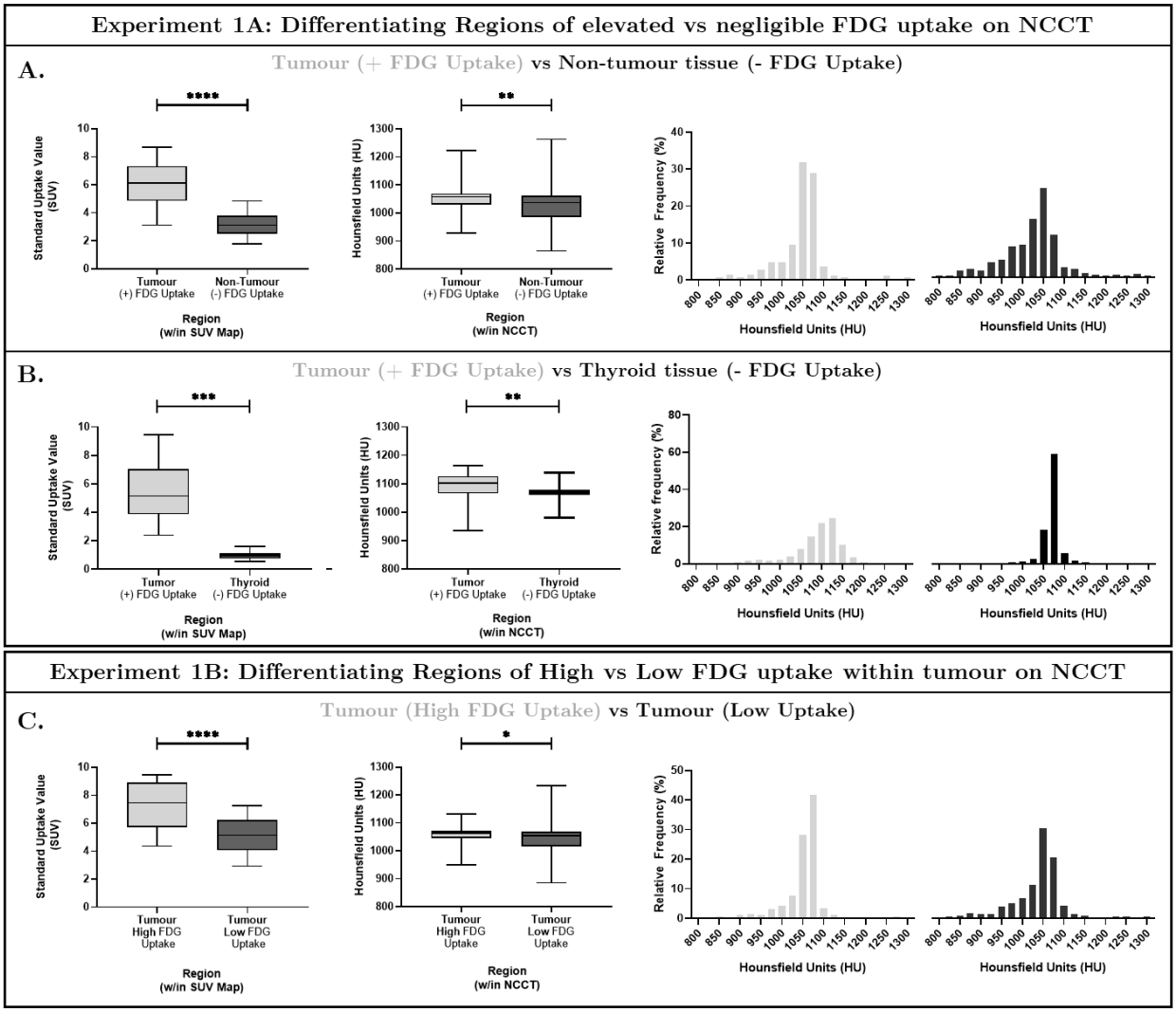


Fig. 81: Differentiating regions of FDG uptake within a CT image. A. Tumour tissues were compared against adjacent non-tumour tissues, as delineated by the PET/SUV map. FDG uptake and HU intensity within tumour tissues were higher than that of non-tumour tissues. **B.** SUV and HU of Tumour tissues were compared against that of thyroid tissues. Thyroid tissue served as a biological comparison to unaffected lymph nodes in terms of macroscopic structure and appearance on CT. FDG uptake and HU intensity within tumour tissues were higher than that of thyroid tissues. **C.** Similarly, segmentations encompassing regions of high and low uptake within the tumour region had statistically different mean SUVs and HU intensities (E, $p<0.05$). This slight HU difference between regions can be appreciated in the adjacent histograms.

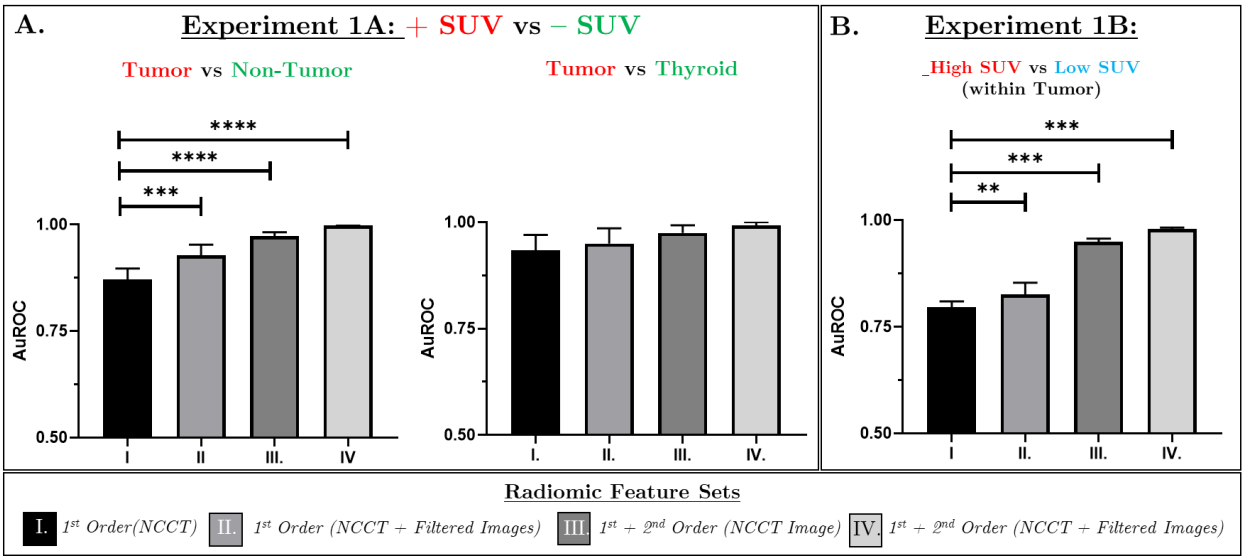


Fig. 82: Area under Receiver Operation Curves for 4 random forest models trained with a combination of radiomic features to classify CT regions based on metabolic activity. A. Experiment 1A compared regions of elevated vs negligible FDG uptake. Regions of negligible FDG uptake included non-tumour tissue (adjacent to the tumour) and thyroid tissue. **B.** Experiment 1B compared regions of High vs Low FDG uptake within the defined tumour segmentation. Each model was trained using a 10-fold cross validation method for 100 iterations on a selected group of 25 radiomic features. Following training, each of the 100 models was applied to the testing cohort to assess model performance. The statistical differences between each model is assessed using a one-way ANOVA. ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$.

7.3.4 Experiment 1B: Radiomics features in NCCT can differentiate high versus low FDG uptake within an individual tumour

The objective of this experiment was to investigate if radiomics features extracted from the NCCT image could differentiate regions with different FDG uptake within the tumour. A patient-specific SUV_{50} threshold (SUV_{50}: 6.62 ± 1.71) was used to further subdivide the tumour into two regions: 1. Regions of high FDG uptake (n = 528, SUV: 7.2 ± 2.0) and 2. Regions of low FDG uptake (n = 683, SUV: 5.1 ± 1.6, **Fig. *81c***). Given that a patient-specific SUV_{50} -threshold was used to differentiate regions, tumour hot spots may either have high FDG uptake, low FDG uptake or a combination of the two. Average tumour volume with SUVs above the 50th percentile ($8.29 \times 10^3 \pm 9.3 \times 10^3$ mm^3) was significantly greater than that below the 50th percentile ($6.39 \times 10^3 \pm 7.8 \times 10^3$ mm^3, $p = 0.009$). In the CT images, significantly lower HU intensity was observed within the tumour region with higher FDG uptake compared to the tumour regions with lower FDG uptake ($p < 0.01$, **Fig. *81c***).

Similarly, four random forest models (***Experiment 1B***: Models I – IV) were trained on a combination of first- and second- order radiomic features extracted from the CT to classify regions with high or low FDG uptake within the tumour (**Fig. *82*b**). Model I (First Order – Image-Based) had an AuROC of 0.79 ± 0.13, which improved with the introduction of first order features from filtered images (Model II, First Order – Image + Filter-Based, AuROC – 0.83 ± 0.14, $p<0.01$). Like that seen for ***Experiment 1A***, the incorporation of matrix-based radiomic features (Models III, IV) further improved classification performance.

7.3.5 Experiment 2: Generation of SUV Maps from NCCT

A 3-fold cross-validation platform was implemented for this CT to SUV map image transformation task. During model training, for each fold, the RMSE between the simulated and ground-truth SUV map images for the training and testing cohort decreased to plateau at 0.30 ± 0.12 and 0.40 ± 0.15, respectively (**Table *20***). **Fig. *83*** illustrates the simulated SUV map alongside their respective gold standards. The visualized error is the difference between the two sets of images and highlights differences in pixel value. The RMSE for each image pair is indicated at the bottom.

7.3.5.1 Technical Assessment of Simulated SUV-Map Accuracy

Mean SUV_0 (2.20 ± 0.78), SUV_{50} (5.95 ± 2.15) and SUV_{Max} (9.89 ± 0.38) within the tumour regions of the simulated maps were significantly less than that of ground truth (SUV_0: 2.40 ± 0.64, SUV_{50}: 6.62 ± 1.71, SUV_{Max}: 9.98 ± 0.15). Subsequently, the bias, as measured by Bland-Altman plot analysis was 11.7% [95% CI: -41.7 – 65.2%], 14.3% [95% CI: -40.5 – 69.2%] and 1.8% [95% CI: -9.7 – 12.1%], respectively (**Fig. *84*a-c**). These values suggest that the simulated SUV map underestimates FDG uptake within the tumour regions. On the other hand, predicted tumour volume/burden per patient ($3.16 \times 10^4 \pm 2.73 \times 10^4$ mm^3) was similar to that of the gold standard ($3.01 \times 10^4 \pm 2.60 \times 10^4$ mm^3, $p = 0.51$). A BA plot comparing the percentage differences in tumour burden between the GAN-simulated and gold standard SUV MAPs is shown in **Fig. *84*d**.

Table 20: RMSE difference between GT- and GAN- SUV Maps.

Test Cohort in each Fold	Patients	Tumour hotpots	2D Slices	RMSE
1	100	253	3,931	0.40 ± 0.14
2	100	227	4,126	0.39 ± 0.16
3	98	202	4,244	0.42 ± 0.15
Total	298	682	12,301	0.40 ± 0.15

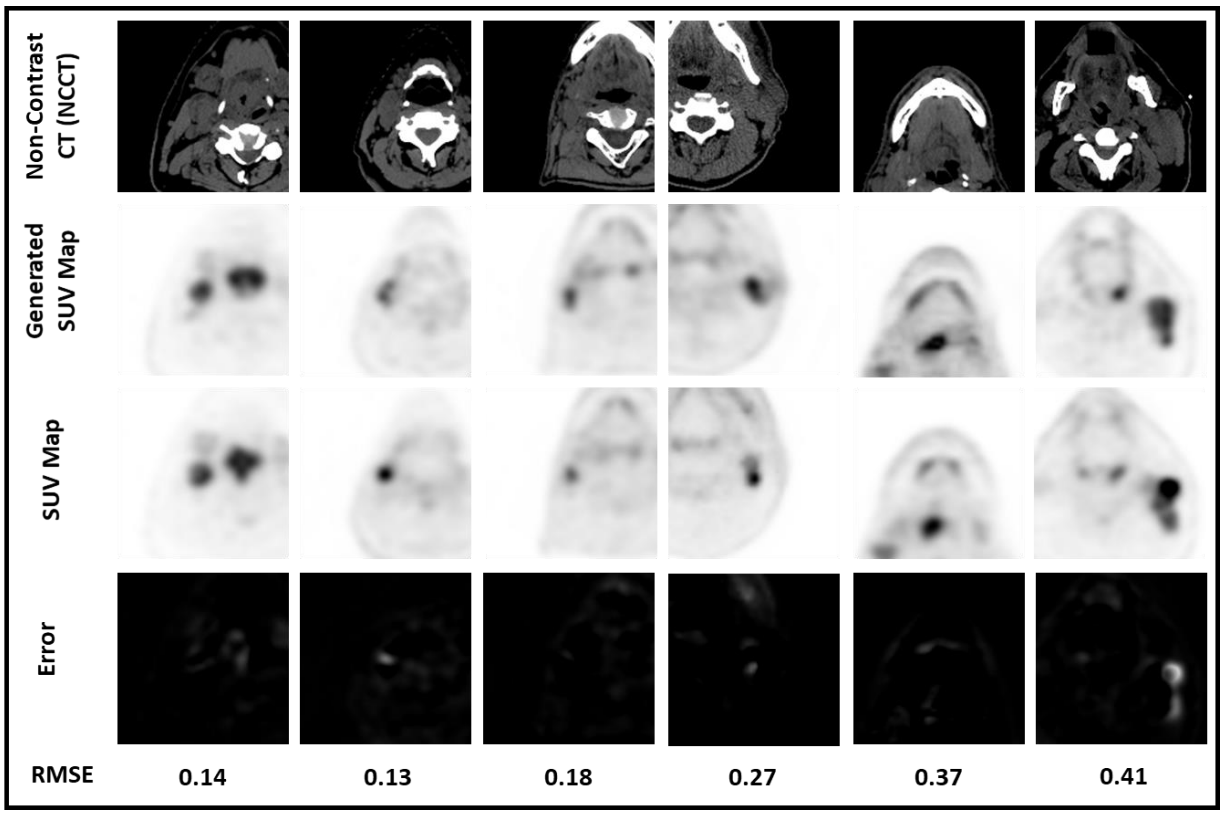


Fig. 83: Simulated SUV Map (Output of Cycle-GAN) displayed alongside its ground truth (Real SUV Map) and Non-Contrast CT axial slice for six patients. The error between the SUV maps is visualized and is represented by the RMSE. It is important to note that these SUV maps are inverted as this is the view commonly used by clinicians.

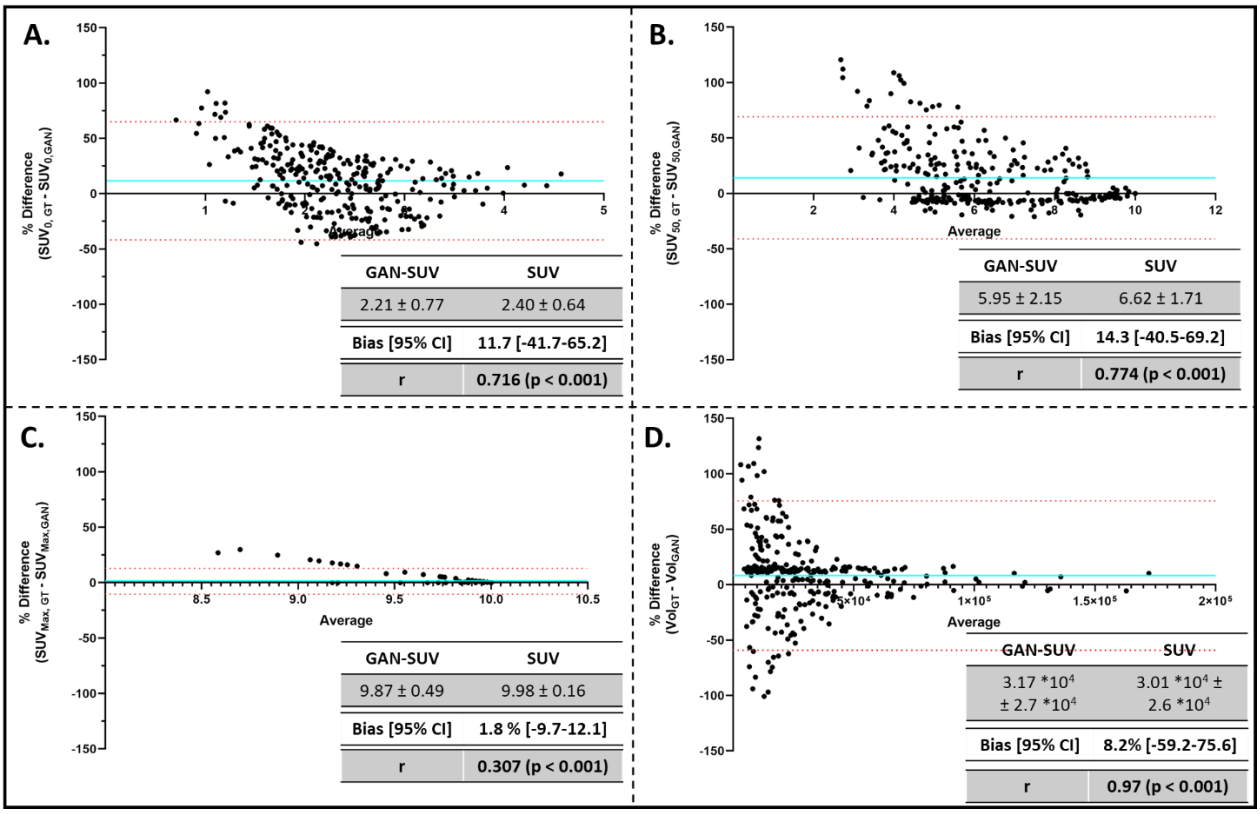


Fig. 84: Technical Assessment of Simulated SUV-Map Accuracy: Bland-Altman plots for the SUV_0 (A), SUV_{50} (B), SUV_{Max} (C) and tumour volume (D) were constructed to assess the percentage difference between the gold-standard and simulated SUV maps. The bias along with the 95% confidence Intervals are indicated each plot. These assessment criteria were adapted from the PERCIST v.1 criteria to characterize and monitor tumour progression using PET images.

7.3.5.2 Clinical Outcome prediction using Simulated SUV maps

Regions of High FDG uptake/SUV were isolated in the Cycle-GAN-simulated SUV (CycleGAN-SUV) map using the threshold-based segmentation method. 86 radiomic features (first + second order features) were extracted from both the Cycle-GAN-SUV and ground truth- SUV maps (GT-SUV) maps for each combination of image parameters (25). Data was separated into training (n = 194) and testing cohorts (n = 104) prior to feature reduction and selection for each outcome (**Fig. *80***). In all three clinical parameters, there was no difference in the outcome prediction using models simulated by GT-SUV maps or Cycle-GAN-SUV maps (**Fig. *85***). For the classification of locoregional recurrence, AuROC was 0.60 ± 0.01 (GT-SUV map) and 0.59 ± 0.02 (Cycle-GAN-SUV map) (p=0.35). For classification of distant metastasis, AuROC was 0.82 ± 0.02 (GT-SUV map) and 0.79 ± 0.01 (Cycle-GAN-SUV map) (p=0.20). For the classification of patient death, AuROC was 0.63 ± 0.01 (GT-SUV map) and 0.62 ± 0.02 (Cycle-

GAN-SUV map) (p= 0.13). The model performances incorporating the GT-SUV maps had similar performance to that observed in the original study by Vallières et al[14].

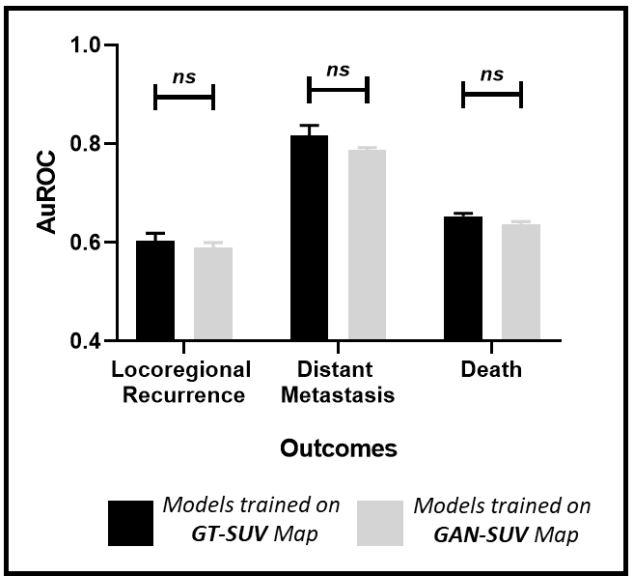


Fig. 85: Area under ROC curve for logistic regression models trained to predict clinical outcomes (1. locoregional tumour recurrence, 2. distant metastasis, and 3. death). Models were trained using selected radiomic features from either the GT- or the CycleGAN- SUV map and evaluated on a fixed testing cohort. The 3 best performing models for each outcome were selected and evaluated. The performance between GT and GAN-SUV maps were analysed for statistical significance for each outcome investigated. Methods for the GT-SUV map analysis were adapted from Vallières et al (2017).

7.4 Discussion

We recently demonstrated the feasibility of simulating contrast enhanced CT images without the injection of IV contrast, using generative DL models. Similar to the workflow described here, we first demonstrated difference in HU intensity and radiomic signature between blood and other soft tissue components[11]. Similarly, abnormal tissues at the molecular level are significantly different from healthy tissues, in terms of ultrastructure, tissue organization and metabolic activity. These altered characteristics have been shown to be present prior to the alteration in morphological structure at the macro- scale and may reflect changes in the tissue's attenuation coefficient. We therefore hypothesized that the raw data acquired from a NCCT can be used to identify region of abnormal metabolic activity. The first objective of this study was to investigate whether there are subtle differences within a NCCT image that can distinguish regions of increased FDG uptake or metabolic activity ('hotspots' on a PET scan) from regions with negligible uptake. This was a necessary preliminary step to ensure that there was sufficient information within the NCCT image for the deep learning method to generate realistic visualizations.

In general, radiomics employs advanced data characterization algorithms to extract pixel-based relationships within a pre-defined region-of-interest. In addition to average HU intensity, the differences between these visually in-distinct regions can be captured using a combination of first- and second- order radiomic features. In this study, we show that there are significant radiomic differences between regions of negligible, low, and high FDG activity in the CT image (Experiments 1A-B). These differences support the validity of this image transformation task. The trained DL generative network likely learns this higher-order information during model training.

The second objective of this study was to investigate if a DL generative network could robustly extract the subtle differences between soft-tissue components in patients diagnosed with HNSCC and generate a visualization of FDG uptake. For this task, standard uptake value (SUV) maps were used for the visualization of FDG uptake. These maps are derived from PET images by standardizing for the patient's weight, radiopharmaceutical dosage (FDG), and the time duration between injection and imaging. These variables account for potential sources of

variation within PET imaging. Following standardization, SUVs for all patients ranged between 0 – 10.

A 3-fold cross-validation approach was employed during training/optimization of the cycle-GAN. There was no data leakage between the training/validation and testing cohorts, ensuring that patients and their respective tumours were either found in the training or testing cohorts. The 2D input data for this DL algorithm was derived from the 3D CT and SUV maps by extracting 2D 144 mm x 144 mm region-of-interests within the larger patient volume. These boundary conditions were defined by the patient contour obtained to evaluate the registration accuracy between the CT and PET images. Additionally, this segmentation was used to remove the underlying table from the patient, especially within the CT image. Given that these scans were obtained from multiple centres, which use PET/CT machines from different manufacturers, the tables that the patients lay on are quite different. Furthermore, as the axial slice moves from the head towards the chest, the 2D axial view of the table significantly changes. Isolating input slices from within the patient volume prevents the generative network from encountering 1. empty slices, 2. slices with a small proportion of the patient and 3. slices with a highly variable table layout. This theoretically should maximize the information learned by the GAN networks for the CT to SUV image transformation task. RMSE between the simulated and ground-truth SUV map was used to optimize training parameters.

Here, we showed that a trained cycle-GAN enables the visualization of a PET-like output from a routine NCCT without the need to obtain a paired PET image. A subset of the PERCIST criteria was used to evaluate the clinical quality of the simulated SUV maps in identifying these tumour hotspots. These guidelines are commonly used to characterize the identified tumour hot spot, monitor tumour progression and its response to treatment. Volume of the tumour hot spot was similar between the simulated and GT-SUV images. This suggests that the generative method is sufficiently able to differentiate healthy tissues from those with altered FDG uptake. On the other hand, it is apparent the generative models underestimate the SUVs within the tumour region as indicated by the statistically significant differences in SUV_0, SUV_{50} and SUV_{Max}.

Given that SUV maps are a semi-quantitative measurement of FDG, the ability of the generative model to identify these altered regions is more important, at least initially, than its internal distribution of SUVs. This study shows for the first time the ability to isolate and segment these altered metabolic areas from the NCCT without the need to obtain a paired PET image. This visualization provides insight into the potential PET visualization and may serve as a method to select if a patient requires subsequent imaging. Similar models can be optimized to further characterize these defined regions. From Experiment 1B, it is apparent that the ability to differentiate regions of low and high FDG activity within the tumour region is possible from the CT image. This suggests that the generative network should be able to properly characterize low/high SUV regions within the tumour. However, the major limitation of this approach is the spatial variability of tumours within the head and neck region. Other instances of generative networks in medical image transformation tasks are constrained to a more regularly occurring phenotype or pathology[11]. This is further amplified by the poor resolution of the input CT images.

Although the simulated SUV maps tended to underestimate the FDG uptake within the tumour region relative to the GT SUV maps, they were able to predict clinical outcomes with the same accuracy as the actual PET scan. The methods and analysis for predicting clinical outcome with the GT SUV maps were adapted from Vallières et al[14,149]. They investigated the impact of radiomic features, extracted from the GT-SUV map, to predict tumour outcome (1. locoregional tumour recurrence, 2. distant metastasis, and 3. patient survival). Their results identified a combination of radiomic features to predict each clinical scenario. Utilizing identical statistical methods for feature reduction and selection, we identify a set of radiomic features that are able to produce similar clinical outcomes within the pre-defined testing cohort. It is important note that our model performances for locoregional recurrence are slightly higher (AuROC: 0.60) than that presented in Vallières et al (AuROC: 0.58). The reason for this slight boost in performance falls possibly to the difference in radiomic feature extraction. Our study extracts a total of 2,150 features for each patient (Primary Tumour + Lymph Nodes) – 18 first order + 68 second-order/texture features for 25 parameter combinations. On the other hand, their study extracts a total of 1,615 features from each patient (10 first-order + 5 shape-based + 40 second-order/texture for 40 parameter combinations). Regardless, these results support the

ability to use this CT to SUV image transformation method to obtain clinically relevant representations of metabolic activity within patients diagnosed with HNSCCs.

7.5 Conclusion

Here, we present an extension of the work presented in Chapter 5 and investigate if NCCT images can be used to monitor and predict the progression of Head and Neck Cancer. Obtaining a PET image is the conventional approach to monitor and manage patients with advanced malignancy as it provides a visualization of its metabolic activity and its local/distant spread. We hypothesized that such information may be inferred directly from a NCCT image. Regions of negligible, low and high tracer uptake (metabolic activity) were isolated on the NCCT image using the paired and registered PET map. Machine learning methods employing random forest classification algorithms on intensity-based radiomic features were able to distinguish these regions of increased tumor uptake on the NCCT image. Subsequently, deep learning generative networks were employed to enhance the quantifiable differences in radiomic signature between regions of different tracer uptake (ex. Tumours/metastatic lymph nodes vs. thyroid tissue). The clinical relevance of the generated images were similar to the GT PET images with regards to predicting tumour progression and patient survival. This visualization can be used to supplement a NCCT for the detection of metabolically-active/malignant lesions without the need to subject patients to PET imaging.

7.5.1 Limitations

This chapter highlights the possibility of simulating PET-like outputs from NCCT images without the use of radioactive tracer. Current limitations of this study are like those outlined in section 5.5.1. Although this data cohort was curated as a part of a multi-centre prospective study, the NCCT acquisition parameters between sites/imaging studies were not standardized. Additionally, it was apparent that CT image resolution was not maximized as axial slice thickness for all NCCT images was 5 mm. This is sufficient when viewing the PET image alongside the NCCT image, but is not ideal when performing radiomic experiments. As seen in section 7.3.3, this data heterogeneity increases the realistic nature of the dataset but may impede the analysis.

Furthermore, this analysis was also limited 2-D axial views within the NCCT/PET image studies. It is possible that a 2.5 dimensional approach via introducing multiple views or an entirely 3-D approach may improve this NCCT to PET image transformation. However, this would require large amounts of data and advanced hardware/computational resources.

8
Conclusion and Future Directions

Contents

8.1 Conclusion 210
8.2 Future Directions 211
8.2.1 Expanding the Data Cohorts 211
8.2.2 Mathematical modelling to simulate AAA shape and morphology 212
8.2.3 Expanding the DL pipelines to focus on other pathologies 213
8.2.4 Investigating the clinical impact and acceptance of DL algorithms 213

8.1 Conclusion

Non-invasive medical imaging is a disruptive technology that has evolved into a platform that is vital for the delivery of personalized precision medicine[2,5]. It is a common tool that allows for the accurate characterization and assessment of disease tissue and is often used by clinicians to aid in decision making. Since its inception, technological advancements have reduced both imaging times and radiation dosage[1]. This has allowed for a seamless integration of medical imaging within health care systems.

As mentioned in Section 2.2, one significant problem with the rapid growth of medical imaging services within clinical practice is the problem of its overutilization[29,30]. It highlights the acquisition of additional medical images that have minimal impact on patient care or patient outcome. This strains health care infrastructures and, more importantly, increases the burden faced by patients. From an imaging perspective, one method to limit image overutilization and reduce the radiation dose delivered to the patient is to maximize the amount of clinically relevant information extracted. This would diminish the need to obtain additional or "unnecessary" scans.

In this thesis, multiple machine learning methods were implemented to maximize the amount of information extracted from non-contrast CT images. Here, the value of NCCT imaging was investigated in two specific pathologies, Abdominal Aortic Aneurysmal (AAA) disease (Chapter 4-6) and Head and Neck Squamous Cell Carcinoma (HNSCC, Chapter 7). These diseases were selected as paired imaging (NCCT and CECT for AAA and NCCT and PET for HNSCC) is obtained for both diseases.

An automatic region-of-interest detection and 3-D segmentation pipeline was developed using DL methods to extract the pathological aorta and AAA from both NCCT and CECT images. This algorithm was able to automatically generate clinically acceptable 3D shapes that could be used for more complex morphological analysis. Subsequently, DL-based image transformation methods were optimized to visualize the pathological aorta and its major side branches directly from NCCT images without the injection of IV contrast. As previous methods to visualize AAA morphology required the use of IV contrast, this novel technique increased the utility of NCCT images within the clinical management pipeline of AAA disease. Aortic shapes

derived from NCCT and CECT images were then used to develop AAA growth prediction models, which were subsequently validated on an independent validation cohort. Finally, the DL-based image transformation methods developed in Chapter 5 were extended to visualize regions of increased radioactive tracer uptake (ex. primary malignancy and metastatic lymph nodes) within a NCCT image in patients diagnosed with HNSCC. The generated PET visualizations had similar predictive capacity for tumour progression and patient survival to that of the GT PET maps.

Ultimately, these investigations highlight the ability to extract higher-order features directly from a NCCT image using machine-learning methods without the addition of IV Contrast for a CECT and radioactive tracer for PET. Subsequently, I show that these features can be useful to clinical practice (i.e stratification of disease). This thesis is poised to disrupt current clinical pathways with a focus on minimizing the number of total imaging studies but maximizing the amount of extracted information.

8.2 Future Directions

This thesis highlights the importance of NCCT imaging in certain clinical scenarios to stratify disease progression and suggests that these benefits could be translatable to similar pathologies. I am fortunate that the majority of this thesis has been published or is currently under peer review (Chapter 7) and therefore, I do not envisage significant additional work to complete my thesis. In this section, we present additional directions that are essential prior to updating the role of NCCT imaging in the clinical management of AAA and HNSCC. Future investigations will be required to validate the current pipeline and to expand the capability of the developeed algorithms to other anatomical sites and pathologies. This will form an ongoing platform of research in the group, and that I am considering post-doctoral fellowship options.

8.2.1 Expanding the Data Cohorts

Most investigations performed in this thesis utilize retrospectively obtained de-identified imaging data from a single centre. This form of data is effective in conducting initial proof-of-concept experiments but is not sufficient when making claims about model robustness and generalizability. The benefits of recruiting more centres and increasing the size of the imaging cohort for training machine/deep learning models include (1) capturing all variations of the

disease phenotype and (2) providing images obtained under different scanners and scanners from different manufacturers. It limits the bias of the DL model to a specific subset of patients/disease types and minimizes the impact of imaging artefact on model output. Similarly, expanding the dataset to include small-sized AAAs will allow for the evaluation and further optimization of our image extraction and growth prediction methods.

Additionally, it is essential to obtain patient demographics and other clinical information. Imaging data represents one aspect of the patient's condition and is commonly coupled with clinical information prior to employing decision making algorithms. By prospectively obtaining imaging studies along with clinical information, we will be able to further optimize our DL image extraction methods and investigate the importance of each component within our prediction algorithms. For example, in Chapter 5, we highlight the role of tube current in upholding non-contrast to contrast image transformation accuracy and suggest that further optimizing scanner settings to patient type may lead to improved model performance. Similarly, in Chapter 6, integrating patient-specific clinical information with AAA-specific geometric measurements may further improve model performance. The unique value of image-specific information can only be determined following the implementation of a prospectively designed study.

8.2.2 Mathematical modelling to simulate AAA shape and morphology

Another future study would be to investigate the impact of additional image augmentation methods to supplement the DL methods trained and optimized in this thesis. In Section 4.2.5, we implement non-linear divergence algorithms to simulate unique AAA shapes to strengthen the segmentation pipeline. We hypothesize that by implementing the methods developed in Section 4.2.3 to simulate AAA shapes, we would be able to systematically generate a diverse array of AAA shapes that could be used to further strengthen the performance of the DL methods trained in this thesis. This would include simulating the presence of ILT within the AAA sac, and augmenting the Lumen-ILT boundary and/or the aortic outer wall.

A direct expansion of this method would be to simulate not only the thoracic aorta but also the major side branches of the aorta. Computational fluid dynamic modelling on the simulated aortic shapes based on the relevant geometric features (Section 6.5), can be used to obtain

insight into the link between alterations in intravascular hemodynamic forces and aneurysmal growth and progression.

8.2.3 Expanding the DL pipelines to focus on other pathologies

The aortic segmentation (Section 4) and the non-contrast to contrast image transformation pipelines (Sections 5) utilize DL algorithms for the ROI detection, segmentation and transformation of the aortic aneurysm. In the aorta alone, AAA disease represents one subset of observable pathological changes. Additional conditions include (1) atherosclerotic plaque/lesions, (2) thoracic aortic aneurysms and (3) aortic dissections. Similar pathological changes can be observed within the neighbouring vasculature (ex. iliac artery aneurysm or carotid artery luminal stenosis/atherosclerosis). Therefore, it is imperative to expand the DL pipelines to automatically capture, segment and transform vasculature (aorta + major side branches) with or without pathological changes. Although the DL pipeline would be similar for model training/optimization, each pathology would require a unique disease-specific data set. Acquisition and curation of these large cohorts would be the rate-limiting step. Similar future investigations are required, albeit with different types of malignancies, for the non-contrast to PET image transformation task detailed in Chapter 7.

8.2.4 Investigating the clinical impact and acceptance of DL algorithms

Overall, the DL algorithms developed over the course of this thesis automatically extract information from CT images and provide guidance on predicting clinical outcome. The true motive underlying these methods are to standardize clinical decision making, enhance the care team efficiency and reduce the burden on the patient. Few of the biggest obstacles in achieving this are the necessity for additional provider training and poor integration within current clinical practice. Therefore, we seek to provide an easy-to-use set of tools that provides clinical assistance in a format that is widely understood. In order to achieve this, it is necessary to engage with radiologists and other clinicians early in the process to guide product development. As a result, future investigations involve establishing prospective blinded studies where radiologists can qualitatively inspect and/or plan surgical intervention/treatment using the derived segmentations and simulated images (ex. Pseudo-Contrast or Pseudo-PET) against that of the GT (CECT and PET). Engaging a variety of clinicians from multiple centres is a time- and labour-intensive task but is essential to bring this technology into the clinic.

9
Appendix

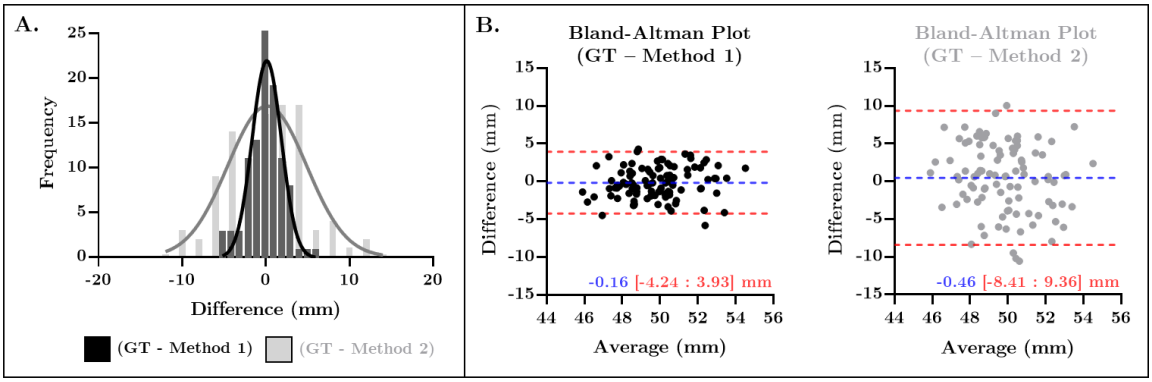
Contents

9.1 Bland-Altman plot analysis to assess measurement accuracy 215
9.2 Radiomic Feature Extraction 216
9.3 Feature Reduction and Selection 217
9.4 Generative Model Training (Non-Contrast-CT to SUV Transformation) 218
9.4.1 CycleGAN Architecture 218
9.4.2 GAN model training 218
9.5 Results 218
9.5.1 Outcome prediction using Simulated SUV maps 218

9.1 Bland-Altman plot analysis to assess measurement accuracy

Bland-Altman plot analysis is a commonly used technique to evaluate the agreement between two sets of measurements. It is valuable in identifying any systematic differences between the measurements (ex. fixed bias) or potential outliers that may confound the analysis. It plots the difference between sets of measurements against their average. The mean difference (y-axis) is the estimated bias, and its standard deviation of the differences captures the fluctuation around the calculated mean. It is commonly presented along with the 95% limits of agreement (LOA), which are the 95% confidence intervals (mean ± 1.96 * standard deviation). The greater the measurement variability, the greater the fluctuation about the mean and the wider the LOA.

In order to demonstrate the utility of a Bland-Altman plot, an example analysis has been illustrated in **Sup Fig. *1***. 100 D_{Max} measurements are randomly sampled from a normal distribution with a mean of 50 mm and a standard deviation of 2 mm. Example Method 1 calculates D_{Max} with a mean difference of 0 mm and a standard deviation of 2 mm. On the other hand, Example Method 2 is less rigorous and calculates D_{Max} with a mean difference of 0 mm but a standard deviation of 5 mm. The differences are normally distributed and are visualized in **Sup Fig. *1***a. Although Method 1 and Method 2 have similar negligible bias, Method 1 has a narrower LOA than that of Method 2. This suggests that Method 1 can more accurately capture the max diameter than that of Method 2. Similar forms of BA plots are performed throughout this thesis.



Sup Fig. 1: Comparative Bland-Altman Analysis with simulated D_{Max} measurements. Methods 1 and 2 represent D_{max} measurements obtained with different levels of accuracy. **A.** Differences between the D_{Max} measurements (1. GT and Method 1 and 2. GT and Method 2) are sampled from a normal distribution with different standard deviations. B. Corresponding Bland-Altman plots are illustrated and highlight that Method 1 has a narrower Limit of Agreement than that of Method 2.

9.2 Radiomic Feature Extraction

All radiomic features were extracted using Pyradiomics, an open-source python package[55]. For each set of image/segmentation and parameter setting, 18 first-order, 68 second-order and 1118 filter-based features were calculated.

1. ***First-Order features:*** These features consist of image-based statistics (ex. minimum, mean, median, maximum, kurtosis, etc), which describe the distribution of voxel intensities within the image region defined by the segmentation mask. These features are not influenced by the shape or volume of the defined mask.

2. ***Second-order features*:** These matrix-based features extract the radiomic "texture" within a defined region of interest shape/volume. These features include those derived from 1. Gray-level Co-occurrence (GLCM), 2. Gray-Level Size Zone (GLSZM), 3. Gray Level Run Length (GLRLM), 4. Neighbouring Gray Tone Difference (NGTDM) and 5. Gray level Dependence (GLDM) matrices.

3. **Filter-based features:** These features represent 1st and 2nd-order features calculated on filtered images. The two filtering methods applied to the images includes 1. Laplacian of gaussian (LOG) and 2. Wavelet filtering. The former is an edge enhancement filter that seeks to emphasize area of gray level change, where a predefined sigma indicates the degree of coarseness in the filtered image. Here, a high sigma favours coarse textures (grey level variations over a large scale), whereas a low sigma favours finer textures (grey level variations over a smaller scale). In this study, 5 levels of sigma were used (σ = 1,2,3,4,5) with 86 1st/2nd order features extracted for each sigma (425 - 3rd order LOG features). Similarly, wavelet filtering produces 8 decompositions for each input image; these decompositions involve applying either a high (H) and/or low (L) pass filter in 3-dimensions (HHH, HHL, HLH, HLL, LLL, LLH, LHL, LHH). For each decomposition, 86 1st and 2nd order features are extracted (680 – 3rd Order Wavelet features). Full documentation for each of the extracted features can be found along with the pyradiomics source code[55].

In Experiments 1A-B, first (n = 18), second-order (n = 68) and filter-based (n = 1118) radiomic features were extracted using all the possible combinations (25) of the following parameters:

1. Isotropic voxel size (5): Sizes of 1mm, 2mm, 3mm, 4mm and 5mm
2. Histogram Bin Widths (5): Widths of 5, 10, 15, 20 and 25

For each region of interest, a total of 30,125 features were extracted ((86 $1^{st}/2^{nd}$ order features + [86 * 13 filtered images]) x 5 Isotropic Settings x 5 Bin-width settings). A subset of these features were used for Models 1-4 (Section 7.2.6). When investigating the clinical relevance of the simulated SUV map (Section 7.3.5.2), only first and second-order features were extracted for each parameter combination (2,150 features per patient, 86 features/combination * 25 combinations). This was done to mimic radiomic features extracted by Vallières et al[14].

9.3 Feature Reduction and Selection

Prior to classification of FDG activity in Experiments 1A/1B, feature reduction was performed using the minimum redundancy, maximum relevance (MRMR) algorithm in MATLAB. The top 25 features (from the 30,125 radiomic features per region) for each model were selected for model training and optimization. This algorithm was selected for feature selection as it tends to efficiently parse through a large array of features to select a subset with a high correlation to an output class and a low correlation between the features[97,98]. By scaling down the feature vector, it prevents overfitting and maximizes model interpretability, which is extremely valuable for clinical problems.

With regards to clinical outcome prediction, feature set reduction was performed on each initial feature set (2,150 features x 194 patients) in a stepwise forward feature selection method utilizing the Gain equation to obtain a subset of 25 features. These selected features maximized both predictive power (via Spearman's rank correlation) and non-redundancy (via the maximal information coefficient). Subsequently, feature selection was performed using a stepwise forward feature selection scheme to maximize the 0.632+ bootstrap AUC (100 samples). The resulted in a combination of features (1-10 features) that optimized model performance. For each of the 10 combinations, predictive performance was estimated and the top 3 parsimonious models were

chosen for each outcome. Additional detail regarding the feature reduction/selection steps can be found within the supplemental materials of Vallières et al[14].

9.4 Generative Model Training (Non-Contrast-CT to SUV Transformation)

9.4.1 CycleGAN Architecture

The generator and discriminator components in the Cycle-GAN model architecture were explicitly defined as least-squares GAN and a 70 x 70 pixel PatchGAN, respectively. The former incorporates an additional least-squares loss function for the discriminator, which in turn, improves the training of the generative model. On the other hand, the discriminator goes through the image pairs, in 70 x 70 pixel patches, and is trained to classify whether the image under question is "real" or "fake". A similar model architecture was used in our prior work for another image medical image transformation task[11].

9.4.2 GAN model training

The Cycle-GAN models were trained with a learning rate of $2.0 * 10^{-5}$ for 200 epochs on overlapping 144 x 144 images located around the patient segmentation, which was derived to assess registration accuracy between the NCCT and SUV images. Four networks (2 generators + 2 discriminators) were trained simultaneously and various loss functions were evaluated at each iteration to document model training. In addition to the loss metrics inherent to the networks, an identity mapping and a cycle consistency loss functions were included to ensure appropriate style transfer and regularization of the generator to allow for image translation, respectively. Model weights were saved every 10 epochs and intermediate model predictions were simulated from the NCCT images within the training cohort. The simulated predictions were independently evaluated against the ground truth CECT images to assess model training. During assessment, overlapping 144 x 144 mm images throughout the patient volume were transformed and a weighted average of the output slices was used to compile the simulated 3D SUV map.

9.5 Results

9.5.1 Outcome prediction using Simulated SUV maps

This section provides the complete description (specific radiomic features with extraction parameters) of the best radiomic models trained, from both the GT and simulated SUV maps,

for each outcome (locoregional recurrence -Appendix Table 1, distant metastasis – Appendix Table 2, and death – Appendix Table 3). Significance of the variable in each of the trained logistic regression models (constructed/optimized from the training dataset - n = 194) was assessed for variable significance using the Wald's test using the DREES software (MATLAB). Appendix Table 4 indicates the performance (AUC, sensitivity, specificity and accuracy) of the best logistic regression models for each of the outcomes on the testing dataset (n = 104). Regression models for the simulated and ground truth SUV maps are similar in performance.

Appendix Table 1: Locoregional Tumour Recurrence (all Variables, p < 0.05)

Fx	Gold-Standard SUV Map	p	Simulated SUV Map	p
1	*GLRLM_ShortRunEmphasis:* Scale = 1mm , binWidth = 25	*0.04*	*GLCM_Imc1:* Scale = 5mm , binWidth = 25	*0.02*
2	*GLSZM_SmallAreaEmphasis:* Scale = 2 mm, binWidth = 15	*0.008*	*GLSZM_ZonePercentage:* Scale = 1 mm, binWidth = 20	*0.005*
3	*GLSZM_GrayLevelNonUniformity-Normalized:* Scale = 3mm , binWidth = 15	*0.02*	*FirstOrder_Skewness* Scale = 4 mm, binWidth = 5	*0.002*
4	*GLCM_Imc1;* Scale = 3mm , binWidth = 10	*0.02*	*GLSZM_SmallAreaEmphasis:* Scale = 1 mm, binWidth = 20	*0.01*
5			*GLSZM_SmallAreaLowGrayLevelEmphasis* Scale = 5mm , binWidth = 20	*0.009*

Appendix Table 2: Distant Metastasis (all Variables, p < 0.05)

Fx	Gold-Standard SUV Map	p	Simulated SUV Map	
1	*GLRLM_ShortRunEmphasis:* Scale = 2mm , binWidth = 20	*0.03*	*FirstOrder_Energy:* Scale = 4mm , binWidth = 5	*0.009*
2	*GLSZM_ZoneEntropy* Scale = 1 mm, binWidth = 20	*0.03*	*GLSZM_SizeZoneNonUniformityNormalized:* Scale = 5mm , binWidth = 25	*0.02*
3	*GLRLM_GrayLevelNonUniformityNormalized:* Scale = 1mm , binWidth = 20	*0.003*	*GLSZM_SmallAreaLowGrayLevelEmphasis* Scale = 3mm , binWidth = 25	*0.014*
4	*GLSZM_GrayLevelNonUniformity;* Scale = 1mm , binWidth = 25	*0.002*	*GLCM_InverseVariance* Scale = 3mm , binWidth = 10	*0.007*
5			*GLSZM_GrayLevelNonUniformity* Scale = 1mm , binWidth = 25	*0.02*

Appendix Table 3: Death (all Variables, p < 0.05)

Fx	Gold-Standard SUV Map	p	Simulated SUV Map	p
1	*GLDM_SmallDependenceEmphasis:* Scale = 1mm , binWidth = 15	*0.01*	*GLDM_GrayLevelNonUniformity:* Scale = 1mm , binWidth = 5	*0.001*
2	*GLRLM_GrayLevelNonUniformityNormalized:* Scale = 1mm , binWidth = 20	*<0.001*	*GLSZM_SizeZoneNonUniformityNormalized:* Scale = 5mm , binWidth = 25	*0.009*
3	*GLCM_Contrast;* Scale = 1mm , binWidth = 15	*0.0023*	*GLSZM_LowGrayLevelZoneEmphasis:* Scale = 1mm , binWidth = 20	*0.04*
4	*GLSZM_GrayLevelNonUniformity;* Scale = 1mm , binWidth = 25	*<0.001*	*FirstOrder_Maximum* Scale = 3mm , binWidth = 5	*0.02*
5	*GLSZM_SizeZoneNonUniformityNormalized:* Scale = 5mm , binWidth = 25	*0.03*	*GLSZM_SmallAreaEmphasis:* Scale = 1 mm, binWidth = 5	*0.03*
6	*GLSZM_SmallAreaEmphasis:* Scale = 1 mm, binWidth = 5	0.04		

Appendix Table 4: Prediction performance of the best radiomic models from GT/GAN SUV maps (logistic regression)

	Outcome	AUC	Sensitivity	Specificity	Accuracy
GT	*Locoregional Recurrence*	0.60	0.67	0.55	0.59
	Distant Metastasis	0.84	0.94	0.54	0.63
	Death	0.65	0.63	0.52	0.67
GAN	*Locoregional Recurrence*	0.59	0.62	0.70	0.59
	Distant Metastasis	0.79	0.83	0.58	0.62
	Death	0.64	0.57	0.68	0.59

10 References

1 Grignon, B., Mainard, L., Delion, M., Hodez, C. & Oldrini, G. Recent advances in medical imaging: anatomical and clinical applications. *Surg Radiol Anat* 34, 675-686, doi:10.1007/s00276-012-0985-0 (2012).

2 Lambin, P. *et al.* Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 48, 441-446, doi:10.1016/j.ejca.2011.11.036 (2012).

3 Brenner, D. J. & Hall, E. J. Computed tomography--an increasing source of radiation exposure. *N Engl J Med* 357, 2277-2284, doi:10.1056/NEJMra072149 (2007).

4 Schuster, A. L. *et al.* Awareness of radiation risks from CT scans among patients and providers and obstacles for informed decision-making. *Emerg Radiol* 25, 41-49, doi:10.1007/s10140-017-1557-8 (2018).

5 Garvey, C. J. & Hanlon, R. Computed tomography in clinical practice. *BMJ* 324, 1077-1080, doi:10.1136/bmj.324.7345.1077 (2002).

6 Khan, U., Yasin, A., Abid, M., Shafi, I. & Khan, S. A. A Methodological Review of 3D Reconstruction Techniques in Tomographic Imaging. *J Med Syst* 42, 190, doi:10.1007/s10916-018-1042-2 (2018).

7 Bae, K. T. Intravenous contrast medium administration and scan timing at CT: considerations and approaches. *Radiology* 256, 32-61, doi:10.1148/radiol.10090908 (2010).

8 Agency, I. A. E. *IMAGINE - IAEA medical imaging and nuclear medicine global resources database.*, <https://humanhealth.iaea.org/HHW/DBStatistics/IMAGINE.html> (2019).

9 Picano, E. Sustainability of medical imaging. *BMJ* 328, 578-580, doi:10.1136/bmj.328.7439.578 (2004).

10 Chandrashekar, A. *et al.* A Deep Learning Pipeline to Automate High-Resolution Arterial Segmentation with or without Intravenous Contrast. *Ann Surg*, doi:10.1097/SLA.0000000000004595 (2020).

11 Chandrashekar, A. *et al.* A Deep Learning Approach to Visualise Aortic Aneurysm Morphology without the Use of Intravenous Contrast Agents. *Ann Surg*, doi:doi: 10.1097/SLA.0000000000004835 (2021).

12 Chandrashekar, A. *et al.* Prediction of Abdominal Aortic Aneurysm Growth Using Geometric Assessment of Computerised Tomography Images Acquired During the Aneurysm Surveillance Period. *Ann Surg*, doi:10.1097/SLA.0000000000004711 (2020).

13 Investigators, M. R. S. Aortic Wall Inflammation Predicts Abdominal Aortic Aneurysm Expansion, Rupture, and Need for Surgical Repair. *Circulation* 136, 787-797, doi:10.1161/CIRCULATIONAHA.117.028433 (2017).

14 Vallieres, M. *et al.* Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. *Sci Rep* 7, 10117, doi:10.1038/s41598-017-10371-5 (2017).

15 Lambin, P. *et al.* Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14, 749-762, doi:10.1038/nrclinonc.2017.141 (2017).

16 Iglehart, J. K. The new era of medical imaging--progress and pitfalls. *N Engl J Med* 354, 2822-2828, doi:10.1056/NEJMhpr061219 (2006).

17 Improvement, N. E. a. N. (2020).

18 Sun, Z., Choo, G. H. & Ng, K. H. Coronary CT angiography: current status and continuing challenges. *The British journal of radiology* 85, 495-510, doi:10.1259/bjr/15296170 (2012).

19 Hinson, J. S. *et al.* Risk of Acute Kidney Injury After Intravenous Contrast Media Administration. *Annals of emergency medicine* 69, 577-586.e574, doi:10.1016/j.annemergmed.2016.11.021 (2017).

20 Kapoor, V., McCook, B. M. & Torok, F. S. An introduction to PET-CT imaging. *Radiographics* 24, 523-543, doi:10.1148/rg.242025724 (2004).

21 Christensen, J. A. *et al.* Characterization of the solitary pulmonary nodule: 18F-FDG PET versus nodule-enhancement CT. *AJR Am J Roentgenol* 187, 1361-1367, doi:10.2214/AJR.05.1166 (2006).

22 Perng, P., Marcus, C. & Subramaniam, R. M. (18)F-FDG PET/CT and Melanoma: Staging, Immune Modulation and Mutation-Targeted Therapy Assessment, and Prognosis. *AJR Am J Roentgenol* 205, 259-270, doi:10.2214/AJR.14.13575 (2015).

23 Yu, H., Caldwell, C., Mah, K. & Mozeg, D. Coregistered FDG PET/CT-based textural characterization of head and neck cancer for radiation treatment planning. *IEEE Trans Med Imaging* 28, 374-383, doi:10.1109/TMI.2008.2004425 (2009).

24 Kitajima, K. *et al.* Present and future roles of FDG-PET/CT imaging in the management of gastrointestinal cancer: an update. *Nagoya J Med Sci* 79, 527-543, doi:10.18999/nagjms.79.4.527 (2017).

25 Basu, S. *et al.* Fundamentals of PET and PET/CT imaging. *Ann N Y Acad Sci* 1228, 1-18, doi:10.1111/j.1749-6632.2011.06077.x (2011).

26 Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* 100, 57-70, doi:10.1016/s0092-8674(00)81683-9 (2000).

27 Gong, S., O'Keefe, G. & Scott, A. Comparison and Evaluation of PET/CT Image Registration. *Conf Proc IEEE Eng Med Biol Soc* 2005, 1599-1603, doi:10.1109/IEMBS.2005.1616743 (2005).

28 Schoenfeld, J. D., Kovalchuk, N., Subramaniam, R. M. & Truong, M. T. PET/CT of cancer patients: part 2, deformable registration imaging before and after chemotherapy for radiation treatment planning in head and neck cancer. *AJR Am J Roentgenol* 199, 968-974, doi:10.2214/AJR.12.8562 (2012).

29 Hendee, W. R. *et al.* Addressing overutilization in medical imaging. *Radiology* 257, 240-245, doi:10.1148/radiol.10100063 (2010).

30 Armao, D., Semelka, R. C. & Elias, J., Jr. Radiology's ethical responsibility for healthcare reform: tempering the overutilization of medical imaging and trimming down a heavyweight. *J Magn Reson Imaging* 35, 512-517, doi:10.1002/jmri.23530 (2012).

31 Fazel, R. *et al.* Exposure to low-dose ionizing radiation from medical imaging procedures. *N Engl J Med* 361, 849-857, doi:10.1056/NEJMoa0901249 (2009).

32 Schauer, D. A. & Linton, O. W. NCRP Report No. 160, Ionizing Radiation Exposure of the Population of the United States, medical exposure--are we doing less with more, and is there a role for health physicists? *Health Phys* 97, 1-5, doi:10.1097/01.HP.0000356672.44380.b7 (2009).

33 Lee, D. W. & Foster, D. A. The association between hospital outcomes and diagnostic imaging: early findings. *J Am Coll Radiol* 6, 780-785, doi:10.1016/j.jacr.2009.08.007 (2009).

34 Gierada, D. S. *et al.* Lung cancer: interobserver agreement on interpretation of pulmonary findings at low-dose CT screening. *Radiology* 246, 265-272, doi:10.1148/radiol.2461062097 (2008).

35 Sistrom, C. L. The appropriateness of imaging: a comprehensive conceptual framework. *Radiology* 251, 637-649, doi:10.1148/radiol.2513080636 (2009).

36 Hess, B. J., Lynn, L. A., Holmboe, E. S. & Lipner, R. S. Toward better care coordination through improved communication with referring physicians. *Acad Med* 84, S109-112, doi:10.1097/ACM.0b013e3181b37ac7 (2009).

37 Lee, T. H. & Brennan, T. A. Direct-to-consumer marketing of high-technology screening tests. *N Engl J Med* 346, 529-531, doi:10.1056/NEJM200202143460715 (2002).

38 Smith-Bindman, R. *et al.* Trends in Use of Medical Imaging in US Health Care Systems and in Ontario, Canada, 2000-2016. *JAMA* 322, 843-856, doi:10.1001/jama.2019.11456 (2019).

39 Seo, H. *et al.* Machine learning techniques for biomedical image segmentation: An overview of technical aspects and introduction to state-of-art applications. *Med Phys* 47, e148-e167, doi:10.1002/mp.13649 (2020).

40 Hesamian, M. H., Jia, W., He, X. & Kennedy, P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *J Digit Imaging* 32, 582-596, doi:10.1007/s10278-019-00227-x (2019).

41 Sharma, N. & Aggarwal, L. M. Automated medical image segmentation techniques. *J Med Phys* 35, 3-14, doi:10.4103/0971-6203.58777 (2010).

42 Kundu, A. Local segmentation of biomedical images. *Comput Med Imaging Graph* 14, 173-183, doi:10.1016/0895-6111(90)90057-i (1990).

43 Maksimovic, R., Stankovic, S. & Milovanovic, D. Computed tomography image analyzer: 3D reconstruction and segmentation applying active contour models--'snakes'. *Int J Med Inform* 58-59, 29-37, doi:10.1016/s1386-5056(00)00073-3 (2000).

44 Currie, G., Hawk, K. E., Rohren, E., Vial, A. & Klein, R. Machine Learning and Deep Learning in Medical Imaging: Intelligent Imaging. *J Med Imaging Radiat Sci* 50, 477-487, doi:10.1016/j.jmir.2019.09.005 (2019).

45 Currie, G., Iqbal, B. & Kiat, H. Intelligent Imaging: Radiomics and Artificial Neural Networks in Heart Failure. *J Med Imaging Radiat Sci* 50, 571-574, doi:10.1016/j.jmir.2019.08.006 (2019).

46 Thrall, J. H. *et al.* Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success. *J Am Coll Radiol* 15, 504-508, doi:10.1016/j.jacr.2017.12.026 (2018).

47 McBee, M. P. *et al.* Deep Learning in Radiology. *Acad Radiol* 25, 1472-1480, doi:10.1016/j.acra.2018.02.018 (2018).

48 Liew, C. The future of radiology augmented with Artificial Intelligence: A strategy for success. *Eur J Radiol* 102, 152-156, doi:10.1016/j.ejrad.2018.03.019 (2018).

49 Shivakumar, N., Chandrashekar, A., Handa, A. I. & Lee, R. Use of deep learning for detection, characterisation and prediction of metastatic disease from computerised

tomography: a systematic review. *Postgrad Med J*, doi:10.1136/postgradmedj-2020-139620 (2021).

50 Saba, L. *et al.* The present and future of deep learning in radiology. *Eur J Radiol* 114, 14-24, doi:10.1016/j.ejrad.2019.02.038 (2019).

51 Ueda, D., Shimazaki, A. & Miki, Y. Technical and clinical overview of deep learning in radiology. *Jpn J Radiol* 37, 15-33, doi:10.1007/s11604-018-0795-3 (2019).

52 Coroller, T. P. *et al.* CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol* 114, 345-350, doi:10.1016/j.radonc.2015.02.015 (2015).

53 Jethanandani, A. *et al.* Exploring Applications of Radiomics in Magnetic Resonance Imaging of Head and Neck Cancer: A Systematic Review. *Front Oncol* 8, 131, doi:10.3389/fonc.2018.00131 (2018).

54 Lv, W. *et al.* Radiomics Analysis of PET and CT Components of PET/CT Imaging Integrated with Clinical Parameters: Application to Prognosis for Nasopharyngeal Carcinoma. *Mol Imaging Biol* 21, 954-964, doi:10.1007/s11307-018-01304-3 (2019).

55 van Griethuysen, J. J. M. *et al.* Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* 77, e104-e107, doi:10.1158/0008-5472.CAN-17-0339 (2017).

56 Aerts, H. J. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. *JAMA Oncol* 2, 1636-1642, doi:10.1001/jamaoncol.2016.2631 (2016).

57 Berenguer, R. *et al.* Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. *Radiology* 288, 407-415, doi:10.1148/radiol.2018172361 (2018).

58 Piccinelli, M., Veneziani, A., Steinman, D. A., Remuzzi, A. & Antiga, L. A framework for geometric analysis of vascular structures: application to cerebral aneurysms. *IEEE Trans Med Imaging* 28, 1141-1155, doi:10.1109/TMI.2009.2021652 (2009).

59 Niu, L. *et al.* Influence of vascular geometry on local hemodynamic parameters: phantom and small rodent study. *Biomed Eng Online* 17, 30, doi:10.1186/s12938-018-0458-8 (2018).

60 Jiang, P. *et al.* Association Between Carotid Bifurcation Geometry and Atherosclerotic Plaque Vulnerability: A Chinese Atherosclerosis Risk Evaluation Study. *Arterioscler Thromb Vasc Biol* 40, 1383-1391, doi:10.1161/ATVBAHA.119.313830 (2020).

61 Detmer, F. J. *et al.* Development and internal validation of an aneurysm rupture probability model based on patient characteristics and aneurysm location, morphology, and hemodynamics. *Int J Comput Assist Radiol Surg* 13, 1767-1779, doi:10.1007/s11548-018-1837-0 (2018).

62 Najafi, M. *et al.* How patient-specific do internal carotid artery inflow rates need to be for computational fluid dynamics of cerebral aneurysms? *J Neurointerv Surg* 13, 459-464, doi:10.1136/neurintsurg-2020-015993 (2021).

63 Lee, R. *et al.* Flow Mediated Dilatation and Progression of Abdominal Aortic Aneurysms. *Eur J Vasc Endovasc Surg* 53, 820-829, doi:10.1016/j.ejvs.2017.03.001 (2017).

64 Richmond, C. Sir Godfrey Hounsfield. doi:10.1136/bmj.329.7467.687 (2004).

65 Whaley, Z. L. *et al.* The Spatial Morphology of Intraluminal Thrombus Influences Type II Endoleak After Endovascular Repair of Abdominal Aortic Aneurysms. *Ann Vasc Surg*, doi:10.1016/j.avsg.2019.05.050 (2019).

66 Hesamian, M. H., Jia, W., He, X. & Kennedy, P. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of digital imaging* 32, 582-596, doi:10.1007/s10278-019-00227-x (2019).

67 Maier, A., Syben, C., Lasser, T. & Riess, C. A Gentle Introduction to Deep Learning in Medical Image Processing. (2018).

68 Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. (2015).

69 Oktay, O. *et al.* Attention U-Net: Learning Where to Look for the Pancreas. (2018).

70 Yoo, T. S. *et al.* Engineering and algorithm design for an image processing Api: a technical report on ITK--the Insight Toolkit. *Stud Health Technol Inform* 85, 586-592 (2002).

71 Finol, E. A. & Amon, C. H. Flow-induced wall shear stress in abdominal aortic aneurysms: Part II--pulsatile flow hemodynamics. *Comput Methods Biomech Biomed Engin* 5, 319-328, doi:10.1080/1025584021000009751 (2002).

72 Bahdanau, D., Cho, K. & Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. (2014).

73 Luong, M.-T., Pham, H. & Manning, C. D. Effective Approaches to Attention-based Neural Machine Translation. (2015).

74 Kollmannsberger, P. *et al.* The small world of osteocytes: connectomics of the lacuno-canalicular network in bone. *New Journal of Physics* 19 (2017).

75 Huttenlocher, D. P., Klanderman, G. A. & Rucklidge, W. J. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15 (1993).

76 Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *arXiv e-prints*, arXiv:1606.06650 (2016). <https://ui.adsabs.harvard.edu/abs/2016arXiv160606650C>.

77 Duquette, A. A., Jodoin, P. M., Bouchot, O. & Lalande, A. 3D segmentation of abdominal aorta from CT-scan and MR images. *Comput Med Imaging Graph* 36, 294-303, doi:10.1016/j.compmedimag.2011.12.001 (2012).

78 Zheng, J. Q., Zhou, X. Y., Li, Q. B., Riga, C. & Yang, G. Z. Abdominal Aortic Aneurysm Segmentation with a Small Number of Training Subjects. *arXiv* (2018).

79 Wanhainen, A. *et al.* Editor's Choice - European Society for Vascular Surgery (ESVS) 2019 Clinical Practice Guidelines on the Management of Abdominal Aorto-iliac Artery Aneurysms. *Eur J Vasc Endovasc Surg* 57, 8-93, doi:10.1016/j.ejvs.2018.09.020 (2019).

80 Chaikof, E. L. *et al.* The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *Journal of vascular surgery* 67, 2-77.e72, doi:10.1016/j.jvs.2017.10.044 (2018).

81 England, A., Niker, A. & Redmond, C. Variability of vascular CT measurement techniques used in the assessment abdominal aortic aneurysms. *Radiography* 16, 173-181 (2010).

82 Haller, S. J. *et al.* Intraluminal thrombus is associated with early rupture of abdominal aortic aneurysm. *Journal of Vascular Surgery* 67, 1051-1058.e1051, doi:10.1016/j.jvs.2017.08.069 (2018).

83 Parr, A. *et al.* Thrombus volume is associated with cardiovascular events and aneurysm growth in patients who have abdominal aortic aneurysms. *J Vasc Surg* 53, 28-35, doi:10.1016/j.jvs.2010.08.013 (2011).

84 Goshima, S. *et al.* Preoperative planning for endovascular aortic repair of abdominal aortic aneurysms: feasibility of nonenhanced MR angiography versus contrast-enhanced CT angiography. *Radiology* 267, 948-955, doi:10.1148/radiol.13121557 (2013).

85 Itagaki, M. W. Using 3D printed models for planning and guidance during endovascular intervention: a technical advance. *Diagn Interv Radiol* 21, 338-341, doi:10.5152/dir.2015.14469 (2015).

86 Lee, K. *et al.* Three-dimensional thrombus segmentation in abdominal aortic aneurysms using graph search based on a triangular mesh. *Computers in Biology and Medicine* 40, 271-278, doi:10.1016/j.compbiomed.2009.12.002 (2010).

87 Egger, J. *et al.* Aorta Segmentation for Stent Simulation. *MICCAI Workshop on Cardiovascular Interventional Imaging and Biophysical Modelling* (2009).

88 Siriapisith, T., Kusakunniran, W. & Haddawy, P. 3D segmentation of exterior wall surface of abdominal aortic aneurysm from CT images using variable neighborhood search. *Comput Biol Med* 107, 73-85, doi:10.1016/j.compbiomed.2019.01.027 (2019).

89 Hong, H. A. & Sheikh, U. U. in *2016 IEEE 12th International Colloquium on Signal Processing & Its Applications (CSPA)* 242-246 (IEEE, 2016).

90 López-Linares, K. *et al.* Fully automatic detection and segmentation of abdominal aortic thrombus in post-operative CTA images using Deep Convolutional Neural Networks. *Medical Image Analysis* 46, 202-214, doi:10.1016/j.media.2018.03.010 (2018).

91 Foley, W. D. & Karcaaltincaba, M. Computed tomography angiography: principles and clinical applications. *Journal of computer assisted tomography* 27 Suppl 1, S23-30, doi:10.1097/00004728-200305001-00006.

92 Aggarwal, S., Qamar, A., Sharma, V. & Sharma, A. Abdominal aortic aneurysm: A comprehensive review. *Exp Clin Cardiol* 16, 11-15 (2011).

93 Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. (2017).

94 Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-Image Translation with Conditional Adversarial Networks. (2016).

95 Whaley, Z. L. *et al.* The Spatial Morphology of Intraluminal Thrombus Influences Type II Endoleak after Endovascular Repair of Abdominal Aortic Aneurysms. *Ann Vasc Surg*, doi:10.1016/j.avsg.2019.05.050 (2019).

96 Rueckert, D. *et al.* Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Trans Med Imaging* 18, 712-721, doi:10.1109/42.796284 (1999).

97 Radovic, M., Ghalwash, M., Filipovic, N. & Obradovic, Z. Minimum redundancy maximum relevance feature selection approach for temporal gene expression data. *BMC Bioinformatics* 18, 9, doi:10.1186/s12859-016-1423-9 (2017).

98 Ding, C. & Peng, H. Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol* 3, 185-205, doi:10.1142/s0219720005001004 (2005).

99 Taha, A. A. & Hanbury, A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging* 15, 29, doi:10.1186/s12880-015-0068-x (2015).

100 MATLAB. version 9.6 (2019a) (The MathWorks Inc., Natick, Massachussets, 2019).

101 Raut, S. S., Chandra, S., Shum, J. & Finol, E. A. The role of geometric and biomechanical factors in abdominal aortic aneurysm rupture risk assessment. *Ann Biomed Eng* 41, 1459-1477, doi:10.1007/s10439-013-0786-6 (2013).

102 Martufi, G., Di Martino, E. S., Amon, C. H., Muluk, S. C. & Finol, E. A. Three-dimensional geometrical characterization of abdominal aortic aneurysms: image-based wall thickness distribution. *J Biomech Eng* 131, 061015, doi:10.1115/1.3127256 (2009).

103 Cayne, N. S. *et al.* Variability of maximal aortic aneurysm diameter measurements on CT scan: significance and methods to minimize. *J Vasc Surg* 39, 811-815, doi:10.1016/j.jvs.2003.11.042 (2004).

104 Lee, S. *et al.* Comparison of image quality and radiation dose between combined automatic tube current modulation and fixed tube current technique in CT of abdomen and pelvis. *Acta Radiol* 52, 1101-1106, doi:10.1258/ar.2011.100295 (2011).

105 Metaxa, E., Kontopodis, N., Tzirakis, K., Ioannou, C. V. & Papaharilaou, Y. Effect of intraluminal thrombus asymmetrical deposition on abdominal aortic aneurysm growth rate. *J Endovasc Ther* 22, 406-412, doi:10.1177/1526602815584018 (2015).

106 Zhu, C. *et al.* Intraluminal Thrombus Predicts Rapid Growth of Abdominal Aortic Aneurysms. *Radiology* 294, 707-713, doi:10.1148/radiol.2020191723 (2020).

107 Chaikof, E. L. *et al.* The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *J Vasc Surg* 67, 2-77 e72, doi:10.1016/j.jvs.2017.10.044 (2018).

108 Lee, R., Jones, A., Cassimjee, I., Handa, A. & Oxford Abdominal Aortic Aneurysm, S. International opinion on priorities in research for small abdominal aortic aneurysms and the potential path for research to impact clinical management. *Int J Cardiol* 245, 253-255, doi:10.1016/j.ijcard.2017.06.058 (2017).

109 Gharahi, H. *et al.* On growth measurements of abdominal aortic aneurysms using maximally inscribed spheres. *Med Eng Phys* 37, 683-691, doi:10.1016/j.medengphy.2015.04.011 (2015).

110 Kontopodis, N. *et al.* Advances in determining abdominal aortic aneurysm size and growth. *World J Radiol* 8, 148-158, doi:10.4329/wjr.v8.i2.148 (2016).

111 Pappu, S., Dardik, A., Tagare, H. & Gusberg, R. J. Beyond fusiform and saccular: a novel quantitative tortuosity index may help classify aneurysm shape and predict aneurysm rupture potential. *Ann Vasc Surg* 22, 88-97, doi:10.1016/j.avsg.2007.09.004 (2008).

112 Bluestein, D. *et al.* Intraluminal thrombus and risk of rupture in patient specific abdominal aortic aneurysm - FSI modelling. *Comput Methods Biomech Biomed Engin* 12, 73-81, doi:10.1080/10255840903077170 (2009).

113 Zambrano, B. A. *et al.* Association of Intraluminal Thrombus, Hemodynamic Forces, and Abdominal Aortic Aneurysm Expansion Using Longitudinal CT Images. *Ann Biomed Eng* 44, 1502-1514, doi:10.1007/s10439-015-1461-x (2016).

114 Lee, A. M. & Chaikof, E. L. Is the abdominal aortic aneurysm rupture rate decreasing? *Adv Surg* 47, 271-286, doi:10.1016/j.yasu.2013.03.004 (2013).

115 Do, H. N. *et al.* Prediction of Abdominal Aortic Aneurysm Growth Using Dynamical Gaussian Process Implicit Surface. *IEEE Trans Biomed Eng* 66, 609-622, doi:10.1109/TBME.2018.2852306 (2019).

116 Lee, R. *et al.* Integrated Physiological and Biochemical Assessments for the Prediction of Growth of Abdominal Aortic Aneurysms in Humans. *Ann Surg*, doi:10.1097/sla.0000000000003154 (2018).

117 Regent, L. *et al.* Integrated Plasma and Tissue Proteomics Reveals Attractin Release by Intraluminal Thrombus of Abdominal Aortic Aneurysms and Improves Aneurysm Growth Prediction in Humans. doi:10.1101/2020.02.28.970483 (2020).

118 Lee, R. *et al.* Applied Machine Learning for the Prediction of Growth of Abdominal Aortic Aneurysm in Humans. *EJVES Short Rep* 39, 24-28, doi:10.1016/j.ejvssr.2018.03.004 (2018).

119 Eurostatistics. *Healthcare resource statistics - technical resources and medical technology - Statistics Explained*, <https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare_resource_statistics_-_technical_resources_and_medical_technology#Use_of_medical_technology> (2020).

120 Research, i. Over 75 Million CT Scans Are Performed Each Year and Growing Despite Radiation Concerns. (2018).

121 Shum, J. *et al.* Quantitative assessment of abdominal aortic aneurysm geometry. *Ann Biomed Eng* 39, 277-286, doi:10.1007/s10439-010-0175-3 (2011).

122 Vorp, D. A., Raghavan, M. L. & Webster, M. W. Mechanical wall stress in abdominal aortic aneurysm: influence of diameter and asymmetry. *J Vasc Surg* 27, 632-639, doi:10.1016/s0741-5214(98)70227-7 (1998).

123 Jalalahmadi, G. *et al.* (Peak) Wall Stress as an Indicator of Abdominal Aortic Aneurysm Severity. *Proc IEEE West N Y Image Signal Process Workshop* 2018, doi:10.1109/WNYIPW.2018.8576453 (2018).

124 Azar, D. *et al.* Mechanical and geometrical determinants of wall stress in abdominal aortic aneurysms: A computational study. *PLoS One* 13, e0192032, doi:10.1371/journal.pone.0192032 (2018).

125 Dhar, S. *et al.* Morphology parameters for intracranial aneurysm rupture risk assessment. *Neurosurgery* 63, 185-196; discussion 196-187, doi:10.1227/01.NEU.0000316847.64140.81 (2008).

126 Zanaty, M. *et al.* Aneurysm geometry in predicting the risk of rupture. A review of the literature. *Neurol Res* 36, 308-313, doi:10.1179/1743132814Y.0000000327 (2014).

127 Yuan, J. *et al.* Hemodynamic and Morphological Analysis of Mirror Aneurysms Prior to Rupture. *Neuropsychiatr Dis Treat* 16, 1339-1347, doi:10.2147/NDT.S254124 (2020).

128 Sharzehee, M., Khalafvand, S. S. & Han, H. C. Fluid-structure interaction modeling of aneurysmal arteries under steady-state and pulsatile blood flow: a stability analysis.

Comput Methods Biomech Biomed Engin 21, 219-231, doi:10.1080/10255842.2018.1439478 (2018).

129 Lauric, A., Hippelheuser, J., Safain, M. G. & Malek, A. M. Curvature effect on hemodynamic conditions at the inner bend of the carotid siphon and its relation to aneurysm formation. *J Biomech* 47, 3018-3027, doi:10.1016/j.jbiomech.2014.06.042 (2014).

130 Chandrashekar, A. *et al.* A Deep Learning Approach to Automate High-Resolution Blood Vessel Reconstruction on Computerized Tomography Images With or Without the Use of Contrast Agent. (2020).

131 Yushkevich, P. A. *et al.* User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage* 31, 1116-1128, doi:10.1016/j.neuroimage.2006.01.015 (2006).

132 Kollmannsberger, P. *et al.* The small world of osteocytes: connectomics of the lacuno-canalicular network in bone - IOPscience. doi:doi:10.1088/1367-2630/aa764b (2017).

133 Manning, B. J., Kristmundsson, T., Sonesson, B. & Resch, T. Abdominal aortic aneurysm diameter: a comparison of ultrasound measurements with those from standard and three-dimensional computed tomography reconstruction. *J Vasc Surg* 50, 263-268, doi:10.1016/j.jvs.2009.02.243 (2009).

134 NHS, E. *Abdominal aortic aneurysm screening - NHS Choices*, <http://www.nhs.uk/conditions/abdominal-aortic-aneurysm-screening/Pages/Introduction.aspx> (2017).

135 Doyle, B. J. *et al.* Biomechanical Assessment Predicts Aneurysm Related Events in Patients with Abdominal Aortic Aneurysm. *Eur J Vasc Endovasc Surg* 60, 365-373, doi:10.1016/j.ejvs.2020.02.023 (2020).

136 Martufi, G. *et al.* Local Diameter, Wall Stress, and Thrombus Thickness Influence the Local Growth of Abdominal Aortic Aneurysms. *J Endovasc Ther* 23, 957-966, doi:10.1177/1526602816657086 (2016).

137 Miller, K. *et al.* Is There a Relationship Between Stress in Walls of Abdominal Aortic Aneurysm and Symptoms? *J Surg Res* 252, 37-46, doi:10.1016/j.jss.2020.01.025 (2020).

138 Malkawi, A. H. *et al.* Patient-specific biomechanical profiling in abdominal aortic aneurysm development and rupture. *J Vasc Surg* 52, 480-488, doi:10.1016/j.jvs.2010.01.029 (2010).

139 Doyle, B. J., Miller, K., Newby, D. E. & Hoskins, P. R. Commentary: Computational Biomechanics-Based Rupture Prediction of Abdominal Aortic Aneurysms. *J Endovasc Ther* 23, 121-124, doi:10.1177/1526602815615821 (2016).

140 Leemans, E. L., Willems, T. P., Slump, C. H., van der Laan, M. J. & Zeebregts, C. J. Additional value of biomechanical indices based on CTa for rupture risk assessment of abdominal aortic aneurysms. *PLoS One* 13, e0202672, doi:10.1371/journal.pone.0202672 (2018).

141 JS, L. & R, S. Population screening and intervention for vascular disease in Danish men (VIVA): a randomised controlled trial. *Lancet (London, England)* 390, doi:10.1016/S0140-6736(17)32250-X (2017).

142 Soyka, J. D. *et al.* Influence of bowel preparation before 18F-FDG PET/CT on physiologic 18F-FDG activity in the intestine. *J Nucl Med* 51, 507-510, doi:10.2967/jnumed.109.071001 (2010).

143 Magota, K. *et al.* Halo artifacts of indwelling urinary catheter by inaccurate scatter correction in (18)F-FDG PET/CT imaging: incidence, mechanism, and solutions. *EJNMMI Phys* 7, 66, doi:10.1186/s40658-020-00333-8 (2020).

144 De Ponti, E. *et al.* Motion Management in PET/CT: Technological Solutions. *Curr Radiopharm* 11, 79-85, doi:10.2174/1874471011666180419150440 (2018).

145 Pepin, A., Daouk, J., Bailly, P., Hapdey, S. & Meyer, M. E. Management of respiratory motion in PET/computed tomography: the state of the art. *Nucl Med Commun* 35, 113-122, doi:10.1097/MNM.0000000000000048 (2014).

146 Kinahan, P. E. & Fletcher, J. W. Positron emission tomography-computed tomography standardized uptake values in clinical practice and assessing response to therapy. *Semin Ultrasound CT MR* 31, 496-505, doi:10.1053/j.sult.2010.10.001 (2010).

147 A, C. *et al.* A Deep Learning Pipeline to Automate High-Resolution Arterial Segmentation with or without Intravenous Contrast. *Annals of surgery*, doi:10.1097/SLA.0000000000004595 (2020).

148 Wahl, R. L., Jacene, H., Kasamon, Y. & Lodge, M. A. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med* 50 Suppl 1, 122S-150S, doi:10.2967/jnumed.108.057307 (2009).

149 Vallieres, M., Freeman, C. R., Skamene, S. R. & El Naqa, I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* 60, 5471-5496, doi:10.1088/0031-9155/60/14/5471 (2015).

11
Publications

Contents

11.1 **A Deep Learning Pipeline to Automate High Resolution Arterial Segmentation with or without Intravenous Contrast**
(Annals of Surgery, 2020)

11.2 **A Deep Learning approach to visualize Aortic Aneurysm Morphology Without the Use of Intravenous Contrast Agents**
(Annals of Surgery, 2021)

11.3 **Prediction of Abdominal Aortic Aneurysm Growth using Geometric Assessment of Computerized Tomography Images Acquired During the Aneurysm Surveillance Period**
(Annals of Surgery, 2020)

11.4 **Beyond the AJR: "Prediction of Abdominal Aortic Aneurysm Growth using Geometric Assessment of Computerized Tomography Images Acquired During the Aneurysm Surveillance Period"**
(Unsolicited Commentary, American Journal of Roentgenology (AJR) 2021)

ORIGINAL ARTICLE

A Deep Learning Pipeline to Automate High-Resolution Arterial Segmentation With or Without Intravenous Contrast

AQ2 *Anirudh Chandrashekar, BE,* Ashok Handa, MBBS, FRCS, MA,* Natesh Shivakumar, MBBS,* Pierfrancesco Lapolla,* Raman Uberoi, BMSc, MBBchir, MRCP,* Vicente Grau, PhD,† and Regent Lee, MBBS, MS(Vasc Surg), DPhil (Oxon), FRCS (Vasc Surg)*✉*

Background: Existing methods to reconstruct vascular structures from a computerized tomography (CT) angiogram rely on contrast injection to enhance the radio-density within the vessel lumen. However, pathological changes in the vasculature may be present that prevent accurate reconstruction. In aortic aneurysmal disease, a thrombus adherent to the aortic wall within the expanding aneurysmal sac is present in >90% of cases. These deformations prevent the automatic extraction of vital clinical information by existing image reconstruction methods.

Aim: In this study, a deep learning architecture consisting of a modified U-Net with attention-gating was implemented to establish a high-throughput and automated segmentation pipeline of pathological blood vessels in CT images acquired with or without the use of a contrast agent.

Methods and Results: Seventy-Five patients with paired noncontrast and contrast-enhanced CT images were randomly selected from an ongoing study (Ethics Ref 13/SC/0250), manually annotated and used for model training and evaluation. Data augmentation was implemented to diversify the training data set in a ratio of 10:1. The performance of our Attention-based U-Net in extracting both the inner (blood flow) lumen and the wall structure of the aortic aneurysm from CT angiograms was compared against a generic 3-D U-Net and displayed superior results. Implementation of this network within the aortic segmentation pipeline for both contrast and noncontrast CT images has allowed for accurate and efficient extraction of the morphological and pathological features of the entire aortic volume.

Conclusions: This extraction method can be used to standardize aneurysmal disease management and sets the foundation for complex geometric and morphological analysis. Furthermore, this pipeline can be extended to other vascular pathologies.

Keywords: aorta, aortic aneurysm, attention-gating, CT angiogram, deep learning, segmentation, U-Net

Abbreviations: AAA, abdominal aortic aneurysm; Attn UNet, Attention U-Net Network; CT, computerized tomography; CTA, computerized tomography angiogram; DL, deep learning; GT, ground truth/manually-derived; HU, hounsfield units; ICC, intraclass correlation coefficient; ILT, intraluminal thrombus; NHS, National Health Service; OUH, Oxford University Hospitals; OxAAA, Oxford Abdominal Aortic Aneurysm Study; WS, wall structure

(*Ann Surg* 2020;xx:xxx–xxx)

A computerized tomography (CT) scan uses multiple X-ray measurements to provide a noninvasive visualization of internal structures. Since the invention of the first commercially-available CT scanner in 1972,[1] the use of CT for the diagnosis and disease management is extensively embedded in modern medicine. Visualization of vasculature on a routine CT is challenging as vessels have similar radio-densities (measured in Hounsfield Unit, HU) to adjacent soft tissues. Injection of intravenous contrast enhances the radio-density within the vessel, enables its visualization and permits rapid segmentation. The produced CT angiogram (CTA) is routinely utilized to for diagnosis. On the other hand, vascular segmentation from noncontrast CT images is a time-intensive and challenging task. Such methods are not readily available to clinicians.

Furthermore, pathological changes, present in the lumen, vessel wall or a combination of both, impede automatic segmentation. In the example of abdominal aortic aneurysms (AAA, abnormal ballooning of the aorta) (Fig. S1A, http://links.lww.com/SLA/C720, red arrow), a thrombus is adherent to the aneurysmal aortic wall (Fig. S1B, http://links.lww.com/SLA/C720, red arrow points toward the AAA) in >90% of cases.[2] Existing methods to segment these CTAs are unable to consistently extract the thrombus and the complex thrombus-lumen interface with accuracy. As such, no automated and standardized methods exist to assess aneurysmal diameter (Fig. S1C, http://links.lww.com/SLA/C720) or thrombus volume. These are vital pieces of clinical information used in the care of AAA patients.

Before the advent of deep learning (DL), vascular segmentation methods incorporated traditional tools including edge detection and/or mathematical models. These methods are complex, difficult to execute and are poorly generalizable. In the early 2000s, image-based DL methods became more approachable, given significant improvements in hardware. Convolutional neural networks, which are the foundation of DL architectures, consist of multiple layers that transform the input using various predefined methods (convolution, nonlinear activation, pooling, etc). The derived high-level abstractions are then extracted by fully connected layers. Finally, the weights of each neural layer and by extension the model are optimized during training.[3,4] In recent years, many groups have strived to identify improvements to this conventional approach.

One well-known architecture for biomedical image segmentation is the U-Net.[5] This model employs skip connections between layers, which serve to integrate the spatial and contextual

AQ3 From the *Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom; and †Department of Engineering Science, University of Oxford, Oxford, United Kingdom.
✉regent.lee@nds.ox.ac.uk.
A.C. and A.H. contributed equally to this work.
Supported by Medical Sciences Division, University of Oxford Medical Research Fund; John Fell Fund, University of Oxford; Academy of Medical Sciences Starter Grant to RL (AMS_SGL013\1015); Clarendon Scholarship to AC.
The authors report no conflicts of interest.
This manuscript was previously published in ArXiv https://arxiv.org/abs/2002.03463
Author Contributions: A.C., R.L and V.G conceived of the presented idea. A.C. developed the theory/workflow and performed the computations with the assistance of A.H. V.G supervised model training and evaluation. R.U and P.L were critical in collecting and organizing the clinical images used in this study. N.S. and P.L. performed the inter-observer variability analysis. All authors discussed the results and contributed to the final manuscript.
The authors declare no conflict of interests.
Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

ISSN: 0003-4932/16/XXXX-0001
DOI: 10.1097/SLA.0000000000004595

CE: D.C.; ANNSURG-D-20-02873; Total nos of Pages: 12;
ANNSURG-D-20-02873

information, to assemble a more precise output. Furthermore, these methods, which were initially limited to 2D, have been applied to 3D images to fully utilize spatial information.[3,5] However, due to memory limitations, many 3D U-Net methods utilize down-sampled input images. This input size may not have enough resolution to represent its diverse anatomical variety. This is especially relevant when evaluating structures with variation that can only be captured at higher resolutions.[3,6] Additionally, most methods are not automatic and require complex user input.

In this study, a modified U-Net architecture was implemented to achieve high-throughput, automated segmentation of pathological vessels (AAA) in CT images acquired with or without the use of IV contrast. In CTA images, our method enables simultaneous segmentation of both the arterial wall and lumen to enable characterization of pathological contents. The model's efficacy was demonstrated by segmenting the thoracic and abdominal aortic regions. Finally, clinical relevance of the trained models was extensively evaluated.

METHODS

Curation of CT Images From a Clinical Cohort

Chest and abdominal CT images were acquired through the Oxford Abdominal Aortic Aneurysm (OxAAA) study. This study received full ethics approval from both Oxford University and Oxford University Hospitals (OUH) NHS Foundation Trust (Ethics Ref 13/SC/0250). As part of the routine preoperative assessment for AAA, a noncontrast CT of the abdomen and a CTA of both the chest and abdomen were performed. CTA images were obtained following contrast injection in helical mode with a predefined slice thickness of 1.25 mm. Noncontrast CT images included only the abdominal aorta and were obtained with a predefined slice thickness of 2.5 mm. Paired images were anonymized within the OUH PACS system before being downloaded onto the secure study drive.

Manual Segmentation of CT Images (Defining the Ground Truth Data)

Seventy-Five patients with paired noncontrast and CTA images were selected. In the CTA, both the aortic lumen and wall structure (WS) were segmented from the aortic root to the iliac bifurcation using the ITK-Snap segmentation software.[7] Semi-automatic segmentation of the aortic lumen was achieved using region-growing by manually delimiting the target intensities between the contrast-enhanced lumen and surrounding tissue. Segmentation of the wall was performed manually by drawing along its boundary using the previously obtained inner lumen as a base. Removing the lumen from the larger segmentation results in a mask highlighting the WS and intra-luminal thrombus (ILT), if present. In the noncontrast CT image, the aorta was manually segmented.

Axial CTA images depicting the ascending thoracic (*yellow arrow*), descending thoracic (*blue arrow*), and abdominal (*red arrow*) aortic regions are shown in Figure 1A-B. The latter is aneurysmal and contains crescentic layers of thrombus. Figure 1C displays the cross-

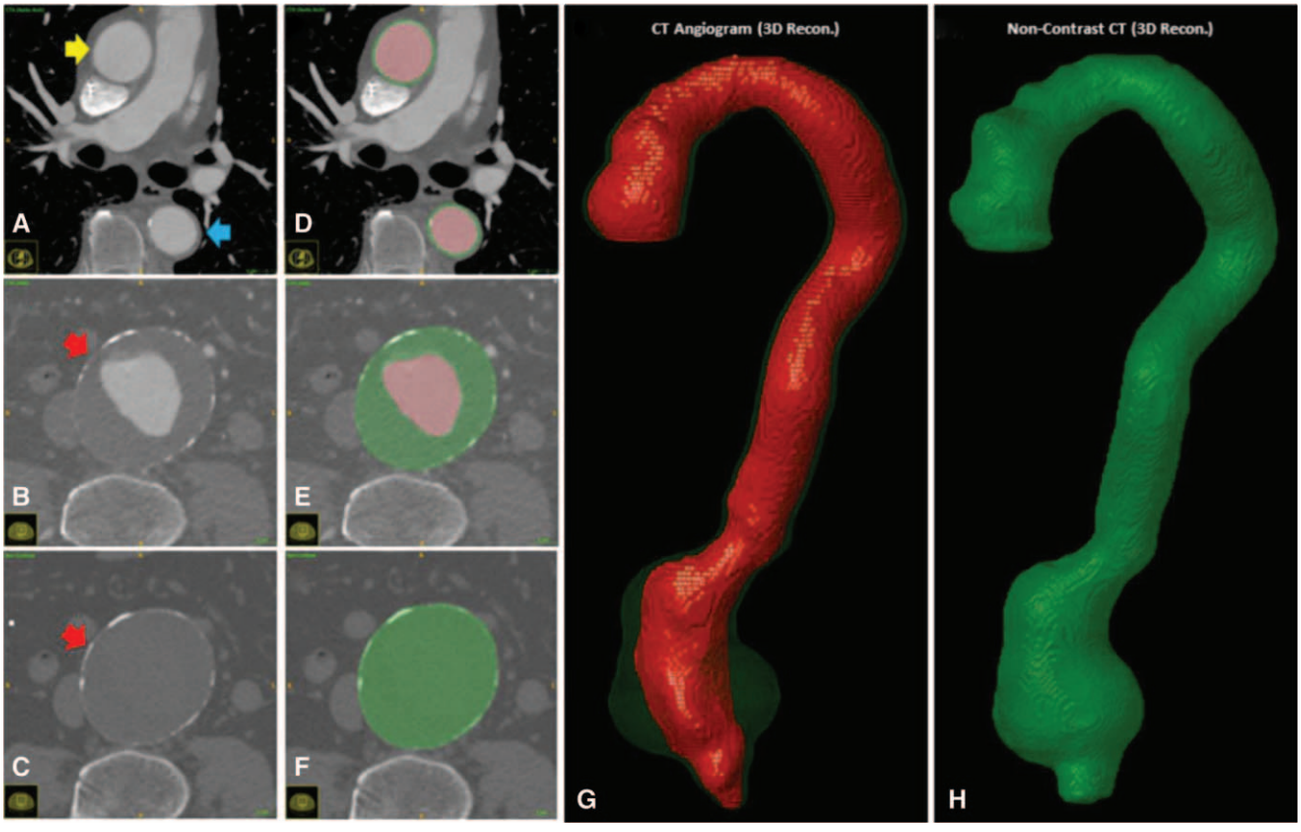


FIGURE 1. A–F. Axial Slices obtained from a CTA and noncontrast CT scan with overlaying manually segmented labels. The lumen is illustrated in red and is typically surrounded by the outer wall in green. In the abdominal region, the green label includes the intra-luminal thrombus, if present. G-H. 3D-reconstructed volumes representing the aortic lumen (red) and wall structure (green), which contains the intra-luminal thrombus, are generated from the masks.

AQ8

section of the abdominal aorta (*red arrow*) in the noncontrast CT scan. Figure 1D–F, show the CT images with the overlying manual segmentations. 3D volumes derived from the manual 2D segmentations are depicted in Figure 1G-H.

Assessment of Intra- and Inter- Observer Variation of Manual Segmentation

Ten patients were selected for intra- and inter-observer variability evaluation. This evaluates the validity of the manual segmentations. For the intra-observer assessment, manual segmentation was performed for the second time by AC after a gap of 2 weeks. For the inter-observer assessment, a trained clinician (NS) performed the segmentations independent of the primary observer. The intraclass correlation coefficient (ICC) was calculated for the intra-/inter-observer analysis to assess the consistency of inner lumen and WS/ILT segmentations.

Data Augmentation

To diversify the training set, CT images and their corresponding segmentations were augmented using divergence transformations. These augmentations employ non-linear warping techniques to manipulate the image in predefined locations. Each image was augmented 10:1 to obtain a total of 825 postaugmented scans. Figure S2, http://links.lww.com/SLA/C720 illustrates an axial slice augmented 10 times. During model training, images were further augmented in 3D using random rotation (0°–15°), translation and scaling (0.7–1.3).

U-Net Architecture

In this study, a variation of the U-Net was used for the aortic segmentation pipeline (Fig. 2A).[5,6] Its general architecture consists of 2 components: the contraction and expansion path (Fig. 2B). The contraction path (red) extracts information to capture the context of the input at the expense of losing spatial information. This is followed by an expansion path (green), where the size of the image increases to produce a predictive binary mask. The lost image detail is restored using skip connections and is merged via concatenation. This integrates the spatial and contextual information to assemble a more precise prediction of the aortic structure.

Attention Gating to Strengthen U-Net Performance

An attention-gated 3D U-Net was evaluated for the segmentation of the aneurysmal aorta. Attention gates utilize information extracted from the coarse scale to filter out irrelevant data exchanged via the skip connections before the concatenation step. The output of each attention gate is the element-wise multiplication of input feature-maps and a learned attention coefficient [0–1]. Given the goal to simultaneously predict the location of the aortic lumen and WS, multidimensional attention coefficients were used. These coefficients were determined using additive addition,[8] which is more

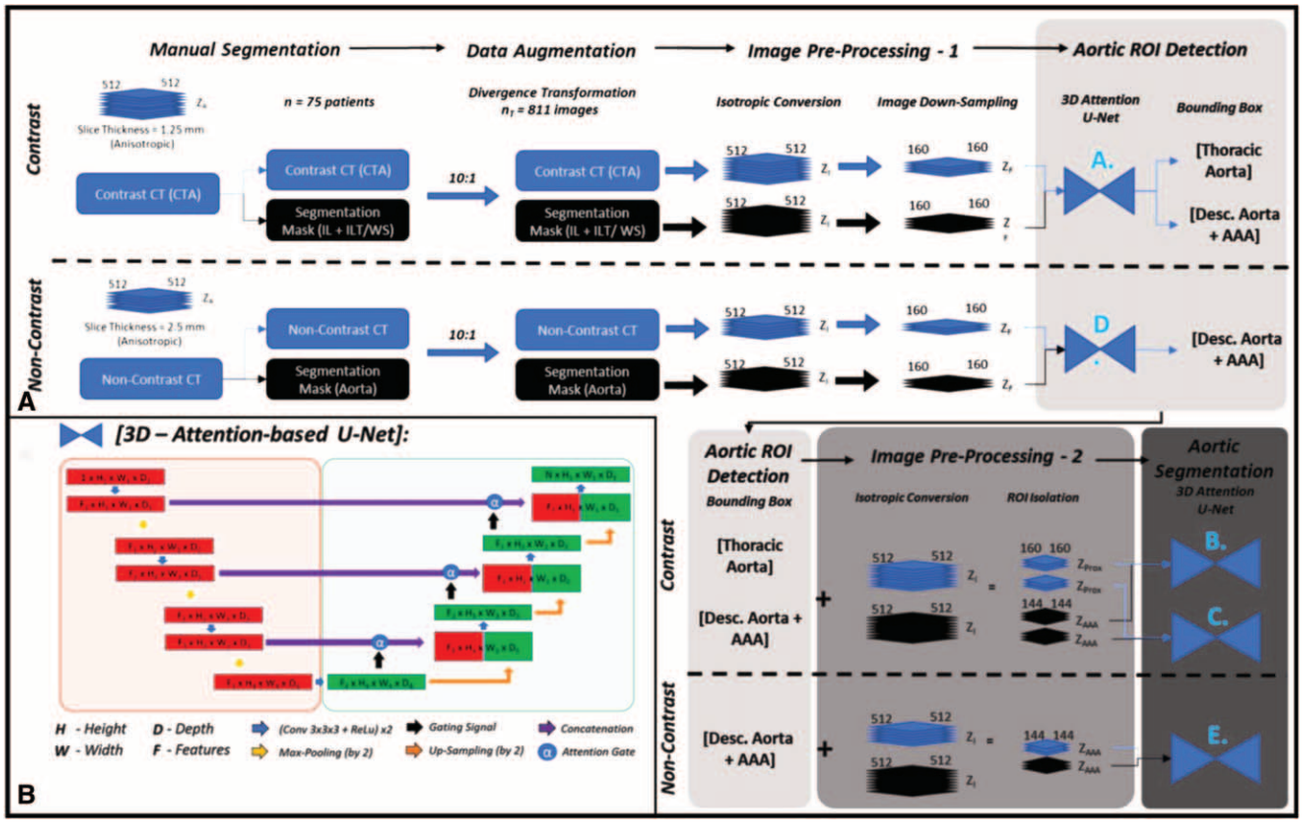


FIGURE 2. A. Automatic aortic segmentation pipeline for the simultaneous detection of the aortic lumen, and intra-luminal thrombus/wall structure. Training required manual segmentation, 2D/3D-data augmentation and preprocessing of both CTA and noncontrast CT images. Aortic ROI detection is coordinated by *U-Net A* for CTA images and *U-Net D* for noncontrast CT images. This is followed by aortic segmentation and is coordinated by *U-Net B+C* for CTA images and *U-Net E* for non-contrast CT images. B. The base architecture for this pipeline is a 3D Attention-based U-Net.

accurate than multiplicative addition.[9] The integration of attention gates for the purpose of pancreatic segmentation has produced superior results when compared to that of prior models.[6] A similar attention mechanism was implemented in this study for aortic segmentation. The performance of this modified U-Net architecture against that of a generic 3D U-Net for segmentation of the aneurysm is included in the supplement. Figure 2B illustrates the 3D U-Net architecture with attention gates utilized in this study.

Loss Function to Evaluate Model Performance

The DICE score was used to quantify model performance at each step. This metric evaluates the similarity between 2 binary images (A and B) and is defined as follows:

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

Here, this index equals twice the number of elements common to both binary images (true positives) divided by the total number of elements in both images (2∗ True Positives + False Positives + False Negatives). This similarity quotient ranges between 0 and 1.

Aortic Segmentation Pipeline: Image Preprocessing

Following data augmentation, all 825 CT images from 75 patients (284,624 CTA axial slices, 145,320 noncontrast CT axial slices) were preprocessed. Preprocessing steps included isotropic voxel conversion and down-sampling by a factor of 3.2 ($512 \times 512 \times Z_{\text{initial/(i)}} \rightarrow 160 \times 160 \times Z_{\text{final/(f)}}$: $Z_f = Z_i$ / 3.2). Here, Z_i and Z_f represents the number of axial slices within the study series before and after pre-processing. The down-sampled images were only used for aortic detection. The higher resolution images were used for aortic segmentation.

Aortic Segmentation Pipeline: Aortic ROI Detection

Attention U-Nets A and D *(Attn U-Net A, D,* refer to Fig. 2A*)* were trained to segment the aorta from these decreased resolution, isotropic CTA and noncontrast CT images, respectively. These architectures were trained and evaluated using 26 patients (Table S1A, http://links.lww.com/SLA/C720). Aortic bounding boxes were generated from the model predictions. Two bounding boxes were generated from the contrast CT image [1. Thoracic (Thor.) and 2. Descending/Abdominal Aorta (AAA)] and one was generated from the non-contrast CT image (1. Descending/ Abdominal Aorta (AAA)]. The ROIs derived from the bounding boxes served as the input data for aortic segmentation. Z_{Thor} or Z_{AAA} represent the number of axial slices within the thoracic aorta and descending aorta/ AAA ROIs, respectively. All subsequent U-Nets (*Attn U-Nets B, C, and E*) were trained using the entire dataset of 75 patients (825 augmented images). This was done to expose the DL models to the diverse and complex aortic/aneurysmal morphology.

Aortic Segmentation Pipeline: Aortic Segmentation

U-Nets B and C *(Attn U-Net B + C,* refer to Fig. 2A*)* were trained on the CTA ROIs to simultaneously segment the aortic lumen and WS regions of the thoracic and WS/ILT of abdominal aorta, respectively. U-Net E was trained on the non-contrast ROIs and was tasked to segment the abdominal aorta. For all 3 U-Nets, 3-fold cross-validation experiments were performed with a data-split of 50:25 patients between training and testing cohorts for each fold (Table S1B, http://links.lww.com/SLA/C720). Each fold consisted of 550 postaugmented images from 50 patients for training. The testing cohort consisted of 25 preaugmented images from 25 patients (testing cohort). There was no overlap between the train/validation and testing cohorts. Table S2, http://links.lww.com/SLA/C720 delineates all the U-Nets trained and evaluated in this study along with their learning parameters. Model training was performed simultaneously on a workstation with 2 11gb NVIDIA RTX 2080 TI graphics cards.

Assessment of Model Accuracy Using Aortic Morphological Features

In addition to the DICE score, 1-,2- and 3-D measurements of aortic morphology were extracted from the aorta. These were used to evaluate the clinical validity of this high-resolution segmentation pipeline. We developed an in-house program in MATLAB to automate the extraction task. We assessed the inter-/intra- observer variation by comparing the algorithm output to the measurements manually extracted (ground truth) from the same CT images. From each patient, measurements were obtained both along the axial plane and the plane orthogonal to the aortic centerline. Six measurements were obtained from three slices [1. slice with the max anteroposterior (AP) diameter, 2–3. 1 cm above and below the slice with the max AP diameter]. Max antero-posterior and transverse diameters were measured in each of the 3 slices. Coefficients of variation between the manual delineation and automatic methods are reported.

Maximum AP diameter (1-D) along the axial plane and axial area (2-D) of the aneurysmal region were automatically extracted from each 3-D image. Finally, 3-D measurements included spatial assessment of the lumen and ILT/WS from CTA images and of the total aortic volume from noncontrast CT images. All metrics were calculated on model predictions and ground truth (GT) segmentations using an in-house program in MATLAB. Bland Altman plots and correlation coefficient analysis assessed bias and the strength of association between the output of the DL models and the GT. Bias for all measurements was reported along with its 95% confidence interval (95% CI).

Second-order features including lumen and ILT/WS centerlines were calculated using an implementation of the homotopic thinning algorithm.[10] Centerline deviation between model predictions and GT annotations was calculated using 1. Average Euclidean distances and 2. Hausdorff distance. The former calculates the distances between 2 closest points in the 2 centerlines. On the other hand, the Hausdorff metric reflects the upper bounds of the former.[11] AQ4
It is the greatest distance between a point in one centerline and the closest point in the adjacent line. Additionally, maximum diameter in planes perpendicular and orthogonal to the generated centerlines were extracted and compared between the ground truth measurements and the model predictions. Root-mean-square-error and percentage deviation were used to assess the similarity between the diameters orthogonal to the AAA centerline. This second-order feature assessment ensures the utility of model predictions for complex geometric/morphological analysis.

RESULTS

CT Image Characteristics

Threefold cross-validation was used during training of the aortic segmentation pipeline. CT image characteristics between the training/validation and testing cohorts were extracted for each fold. Statistical comparison (2-tailed unpaired *t*-tests) between the training and testing cohorts for both CTA and noncontrast CT images across all 3 folds, revealed no significant differences. The CT image characteristics between the groups in fold 1 are summarized in Table S3, http://links.lww.com/SLA/C720.

Intra- and *Inter-* Observer Variability Assessment

There were strong agreements for *inter- and intra-* observer measurements (GT) as measured by DICE (±SD) score and intraclass correlation coefficients from CTA and noncontrast CT images

CE: D.C.; ANNSURG-D-20-02873; Total nos of Pages: 12;
ANNSURG-D-20-02873

(Table S4, http://links.lww.com/SLA/C720). This supports the accuracy of the manual segmentations used for model training.

Aortic Segmentation Pipeline: ROI Selection Accuracy

Attn-U-Nets A and D were trained and tested using the smaller cohort of 26 patients. These networks were tasked to extract the aortic volume from low-resolution isotropic CT images. Both model performances plateaued rapidly after 200 epochs of training. Figure S6, http://links.lww.com/SLA/C720 illustrates the evolving DICE score metric for the validation group during training of these architectures. The segmentation accuracies on the testing cohort for extracting the aortic mask from the CTA and noncontrast CT images were $93.4 \pm 1.2\%$ and $88.7 \pm 4.2\%$, respectively. Implementing this network on the larger cohort allowed for accurate ROI selection of the aortic shape on all images.

Aortic Segmentation Pipeline: Aortic Segmentation Accuracy

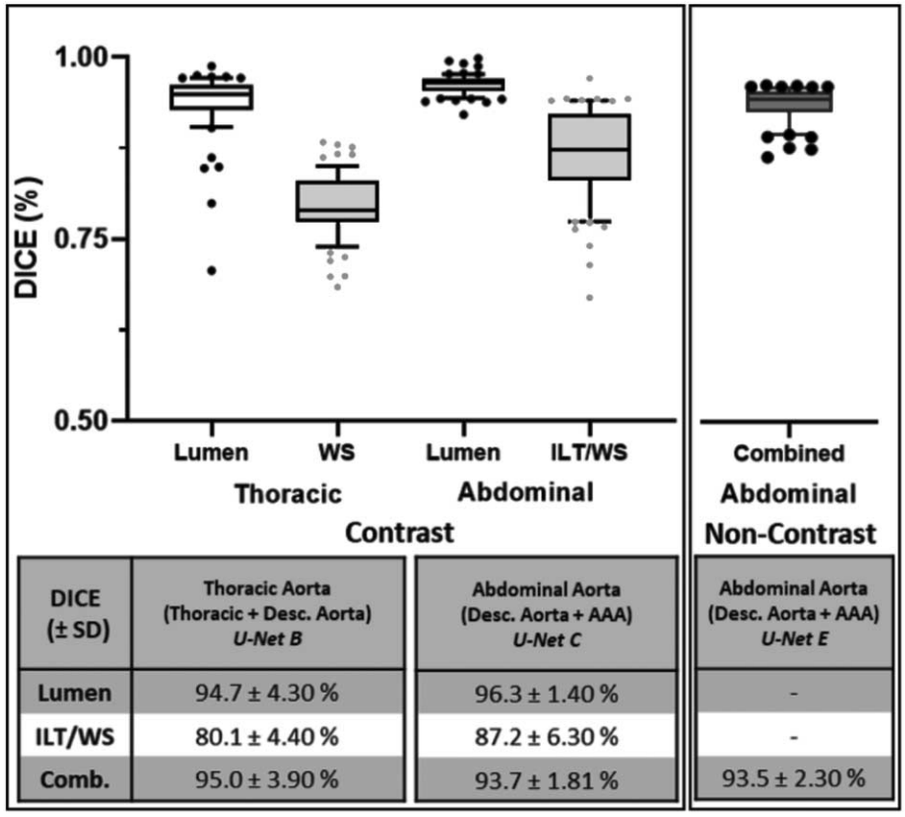
Following ROI selection, threefold cross-validation was used to train the final segmentation models. Figure S7A, http://links.lww.com/SLA/C720 illustrates the evolving DICE score metric for the validation group during training of these architectures. Consequently, Figure 3 displays the performance of *Attn U-Nets B and C* on the ability to segment CTA images via the DICE score metric. The inner lumen DICE accuracy is comparable between the thoracic and the abdominal aorta regions. However, WS DICE accuracy is lower in the thoracic aorta compared to the abdominal aortic region. This is primarily because the thoracic aorta is mostly devoid of ILT and in most cases is a thin circular "ring" surrounding the lumen. Slight differences in this segmentation result in a relatively larger proportion of error, as compared to the abdominal region where WS differences will be proportionally less due to ILT presence.

For noncontrast CT images, threefold cross-validation was utilized to train *Attn U-Net E* (Fig. S7B, http://links.lww.com/SLA/C720). Figure 3B displays the performance of *Attn U-Net E* to segment noncontrast CT images via the DICE metric. The aortic segmentation pipeline for a patient within the test cohort compared against GT annotations is illustrated in Figure S8, http://links.lww.com/SLA/C720.

Aortic Segmentation Pipeline: First and Second Order Assessment of Aortic Morphology

Coefficients of variation (%CV) between manual and automatic measurements for the maximum AP and transverse diameters along the axial plane were $0.7 \pm 0.05\%$ and $1.1 \pm 0.03\%$, respectively. Additionally, %CV between manual and automatic measurements for the maximum AP and transverse diameters along the plane orthogonal to the aortic centerline were $0.9 \pm 0.04\%$ and $1.4 \pm 0.1\%$, respectively. This suggests high concordance between the manual and automatic methods in the calculation of clinical measurements and supports the use of the automatic extraction algorithm for subsequent steps.

Maximal AP diameter ($r_p = 0.99$, $P < 0.001$) and cross-sectional area ($r_p = 0.98$, $P < 0.001$) derived from model predictions of CTA images were very strongly correlated with manually derived measurements (Fig. 4A-B). The variability in diameter measurements was <1.5 mm ($1.2 \pm 0.80\%$). Inner lumen ($r_P = 0.98$, $P < 0.001$) and WS ($r_P = 0.78$, $P < 0.001$) volumes of the thoracic aorta, derived from



DICE (± SD)	Thoracic Aorta (Thoracic + Desc. Aorta) *U-Net B*	Abdominal Aorta (Desc. Aorta + AAA) *U-Net C*	Abdominal Aorta (Desc. Aorta + AAA) *U-Net E*
Lumen	94.7 ± 4.30 %	96.3 ± 1.40 %	-
ILT/WS	80.1 ± 4.40 %	87.2 ± 6.30 %	-
Comb.	95.0 ± 3.90 %	93.7 ± 1.81 %	93.5 ± 2.30 %

FIGURE 3. DICE accuracy of model predictions from CTA/noncontrast CT images compared against GT segmentations. A. DICE scores for the CTA-derived segmentations are divided into nonoverlapping ROIs (Thoracic Aorta [U-Net B] and Abdominal Aorta [U-Net C]). Scores for the lumen, thrombus/wall structure (ILT/WS) only and the combined aortic region are calculated. B. DICE scores for the non-contrast-derived segmentations of the combined aortic region are calculated for the descending aorta/AAA only [U-Net E].

CE: D.C.; ANNSURG-D-20-02873; Total nos of Pages: 12;
ANNSURG-D-20-02873

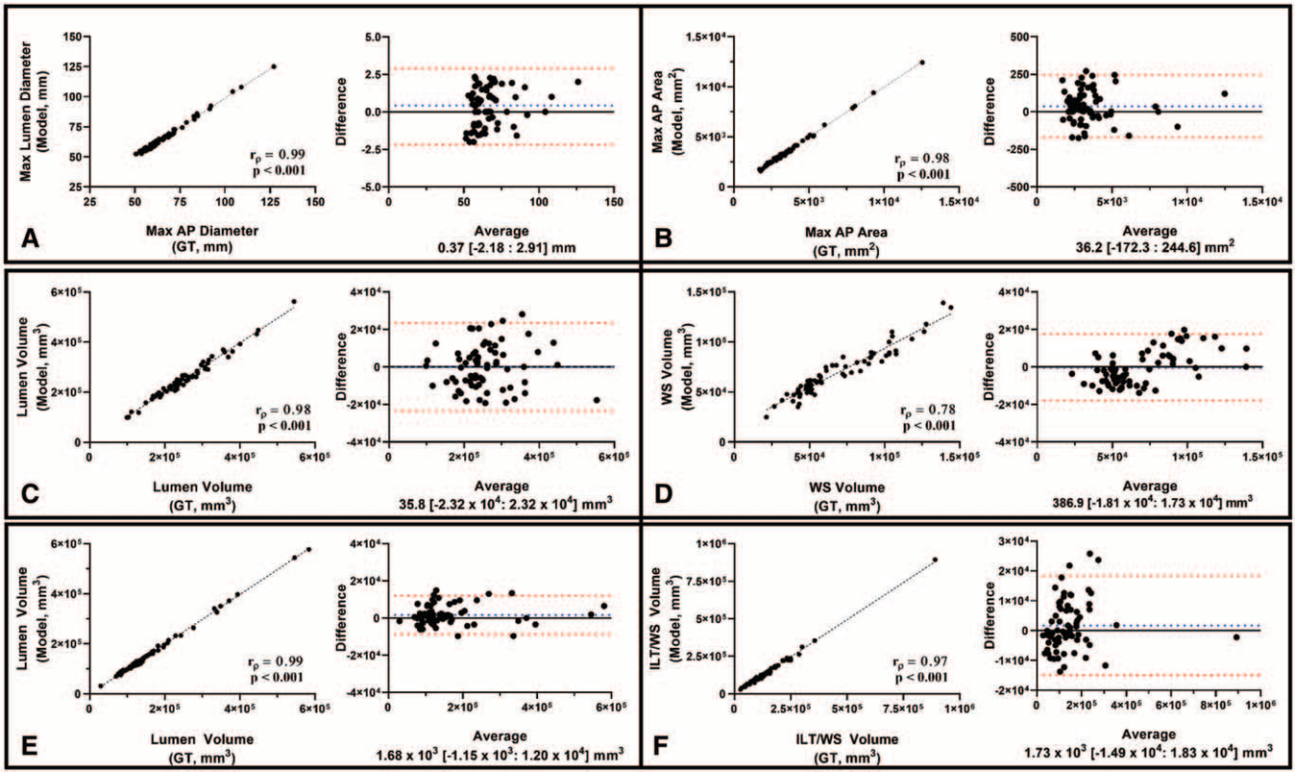


FIGURE 4. Bland-Altman plots and correlation-coefficient analysis comparing the 1-D (Max AP Diameter of AAA - A), 2-D (Max axial area of AAA - B), and 3-D (Lumen/WS volume of Thoracic Aorta – C-D, and lumen/WST/ILT volumes of the Abdominal Aorta - E-F) measurements derived from model predictions compared against those derived from the GT. This analysis was limited to volumes extracted from CTA images. Spearman correlation coefficients (r_ρ) and *P*-values are indicated on the graphs.

the output of U-Net B, were strongly correlated with those obtained from the GT annotations (Fig. 4C-D). The variability in the WS volume measurements (12.40% ± 8.10%) were noticeably greater than for the lumen (3.90% ± 2.64%) in the thoracic aorta. This is inherently linked with its thin circumferential distribution in the thoracic aorta.

Inner lumen ($r_p = 0.99$, $P < 0.001$) and ILT/WS ($r_p = 0.97$, $P < 0.001$) volumes from the abdominal aorta, derived from the output of U-Net C, were very strongly correlated with manually-extracted volumes (Fig. 4E-F). In this case, the ILT/WS variability in the abdominal region (5.50 ± 3.01%) is lower than that in the thoracic aorta (Table 1A). Model predictions of the thoracic (*U-Net B*) and the abdominal aorta (*U-Net C*) from 4 patients in the testing cohort alongside their GT masks are shown in Figure 5A-B.

Furthermore, the similarity in the lumen and ILT/WS centerlines generated from the model predictions and GT annotations is highlighted in Table 1A. Centerline deviations within the thoracic aorta are greater than those observed within the abdominal aorta. This may be due to the difficulty in delineating the border between the aorta and branching arteries within the thoracic region. Model-derived segmentations of these outlets may affect centerline

TABLE 1. Clinical Assessment of Segmented Volumes From CTA and Noncontrast Images.

CTA	% Difference (±SD)		% Difference (±SD)	Noncontrast CT	% Difference (±SD)
Thoracic aorta (*U-Net B*)					
Max AP diameter	—	Abdominal Aorta / AAA (*U-Net C*)	1.20 ± 0.80%	Abdominal Aorta / AAA (*U-Net E*)	Max AP Diameter	1.67 ± 1.10%
Max axial area	—		2.96 ± 2.45%		Max Axial Area	3.60 ± 3.02%
Lumen volume	3.90 ± 2.64%		2.90 ± 2.60%		AAA Volume	1.67 ± 1.10%
WS volume	12.40 ± 8.10%		5.50 ± 3.01%			
	Deviation (Euclidean Distance, ± SD)		Deviation (Euclidean distance, ±SD)			Deviation (Euclidean Distance, ±SD)
Lumen						
Centerline Dev.	1.07 ± 0.67 mm		0.85 ± 0.52 mm		AAA	
Hausdorff dist.	2.60 ± 2.02 mm		2.54 ± 1.98 mm		Centerline deviation	1.94 ± 1.00 mm
WS						
Centerline Dev.	1.64 ± 0.80 mm		1.04 ± 0.57 mm			
Hausdorff dist.	2.84 ± 1.59 mm		2.73 ± 1.73 mm		Hausdorff distance	3.58 ± 2.08 mm

CE: D.C.; ANNSURG-D-20-02873; Total nos of Pages: 12;
ANNSURG-D-20-02873

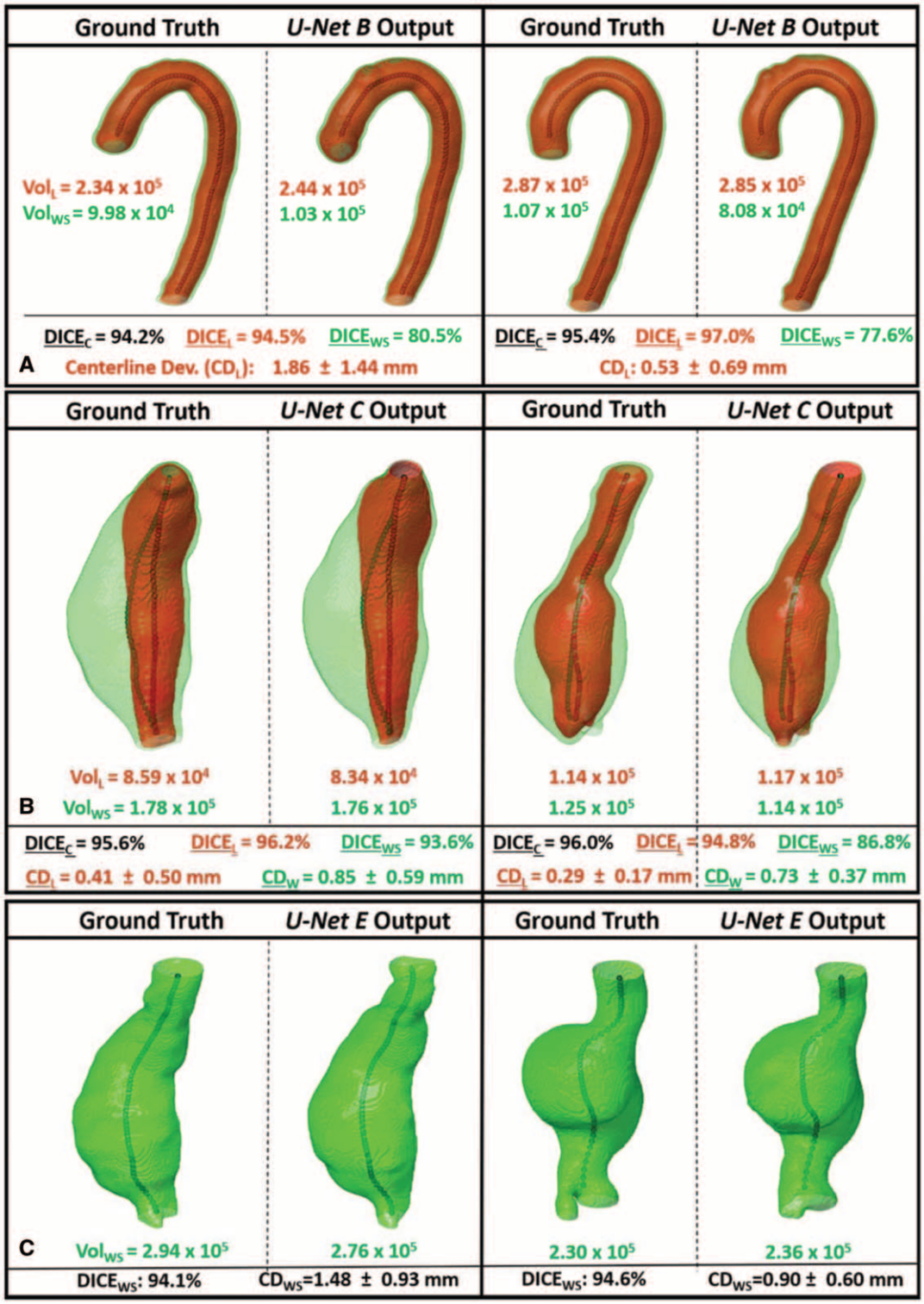


FIGURE 5. Model predictions of the thoracic (A) and abdominal (B) aortic regions from CTA images and the abdominal region from noncontrast CT images (C) within the testing cohort are displayed alongside labelled GT masks. WS and Lumen volumes, when available, are indicated next to each segmentation. DICE scores for the lumen (red), WS (green), and the combined aortic predictions are indicated for each patient. The difference in centerlines derived from the lumen masks are indicated as average Euclidean distance deviation ($CD_L \pm SD$).

properties greater than other metrics. However, the average Euclidean distance deviation is less than 2 mm for the thoracic aorta in 89% of cases (67/75) and for the abdominal aorta in 92% of cases (69/75). The generated centerlines allow for the automatic calculation of max AP diameter along planes orthogonal to the aortic centerline. The resulting diameter profiles between ground truth and predictions were compared and showed a RMSE of 1.45 ± 1.65 mm, which is equivalent to a percentage difference of $2.3 \pm 1.1\%$. Figure 6A

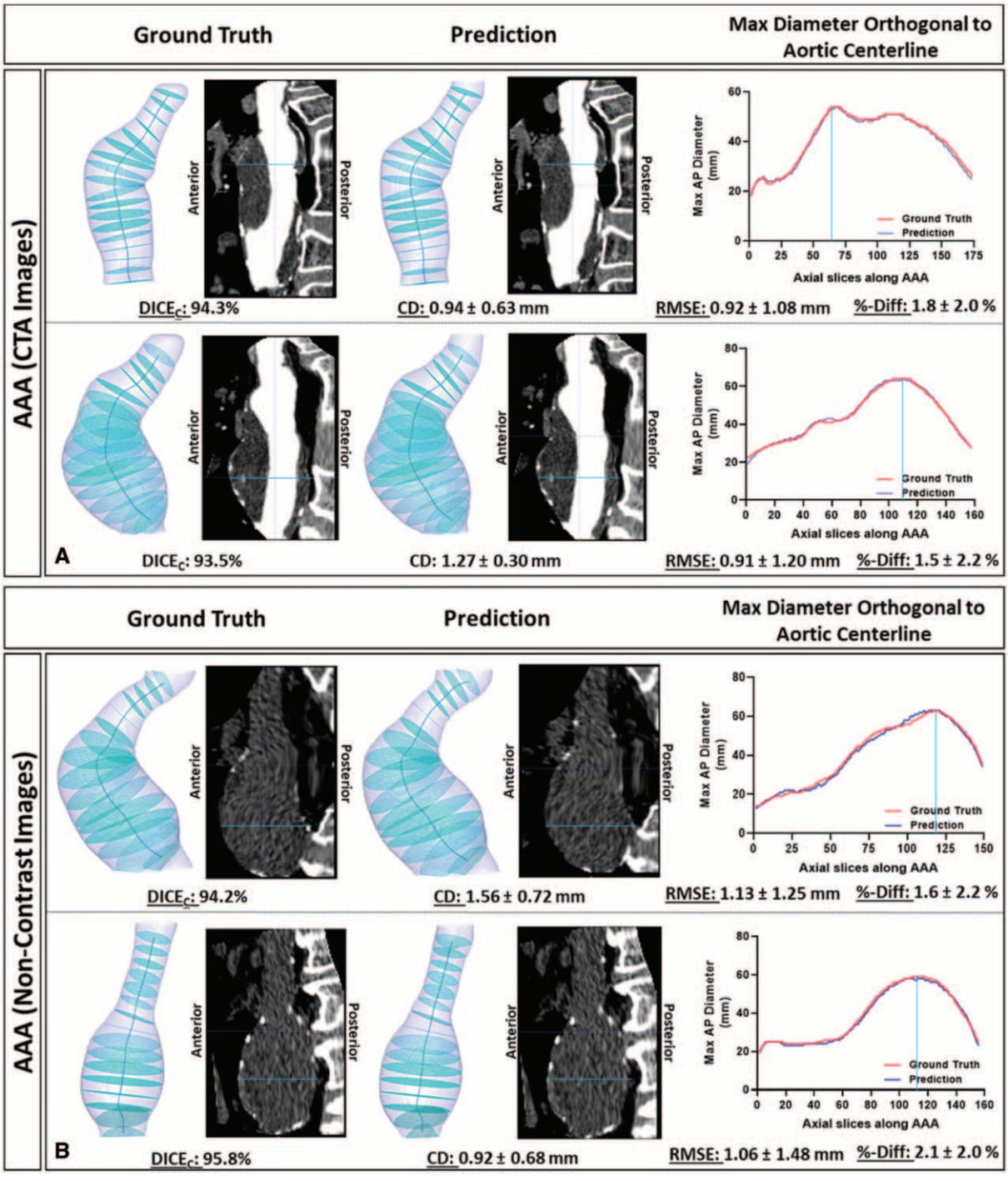


FIGURE 6. Maximum AAA Diameter profiles along planes orthogonal to AAA centerline. Profiles were generated from ground truth and model predictions of AAA volumes derived from contrast-enhanced (Panel A) and noncontrast (Panel B) CT images. The CT images displayed are the straightened views through the sagittal plane (realigned using the centerline). Corresponding DICE scores, average Euclidean distance deviations between centerlines, RMSE and %-difference of diameter profiles are indicated.

CE: D.C.; ANNSURG-D-20-02873; Total nos of Pages: 12;
ANNSURG-D-20-02873

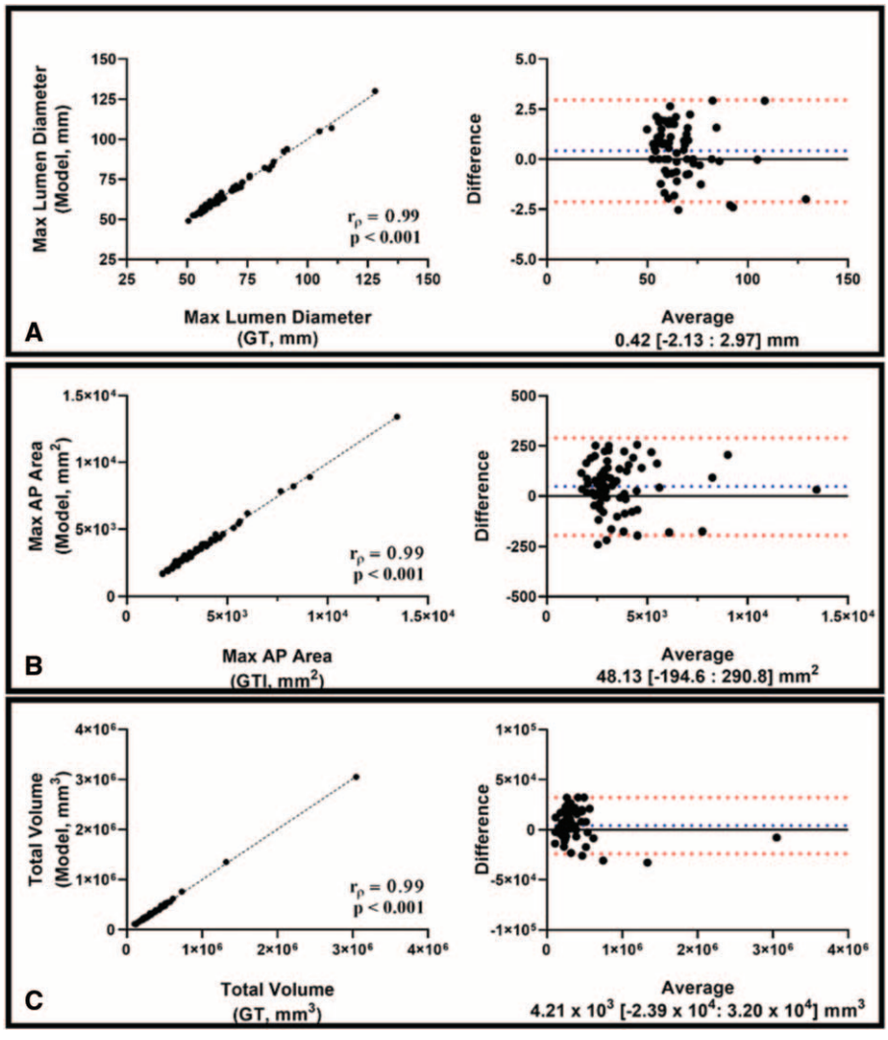


FIGURE 7. Bland-Altman plots and correlation-coefficient analysis comparing the 1-D (Max AP Diameter of AAA - A), 2-D (Max axial area of AAA - B), and 3-D (Total volume of AAA - C) measurements from model predictions compared against those derived from GT. This analysis was limited to volumes extracted from noncontrast CT images. Spearman correlation coefficients (r_ρ) and *P*-values are indicated on the graphs.

illustrates 2 examples of AAAs with planes orthogonal to centerline. These planes were used to generate the re-aligned or straightened view of the AAA. Corresponding DICE scores and average Euclidean distance deviations between centerlines are indicated. Maximum diameter profiles are illustrated for both ground truth and model predictions. These results support the clinical strength of this automatic segmentation platform for CTA images.

Maximal AP diameter ($r_p = 0.99$, $P < 0.001$), cross-sectional area ($r_p = 0.99$, $P < 0.001$) and volume ($r_p = 0.99$, $P < 0.001$) measurements extracted from the model predictions of non-contrast CT images are very strongly correlated with those derived from the GT segmentations (Fig. 7). The variability in extracting these measurements from noncontrast CT-derived segmentations is like that of CTA-derived annotations (Table 1B). Model predictions from 2 patients within the testing cohort are illustrated in Figure 5C. The resulting diameter profiles between ground truth and predictions were compared and showed a RMSE of 2.11 ± 1.32 mm, which is equivalent to a percentage difference of $2.8 \pm 2.1\%$. Figure 6B illustrates 2 examples of AAAs with planes orthogonal to centerline. This study shows for the first time the ability to segment the aneurysmal aorta from noncontrast CT images at a level comparable to a human observer.

DISCUSSION

This study describes a fully automatic and high-resolution algorithm able to extract the aortic volume from both CTA and noncontrast CT images at a level superior to that of other currently published methods.[12,13] High accuracy of our segmentation pipeline was supported by the DICE score metric between model predictions and ground truth annotations for both the thoracic and abdominal aorta. However, this metric, which evaluates the similarity between 2 binary images by evaluating the extent of pixel overlap, has its limitations. In cases of small volumes, minimal changes lead to lower DICE score percentages. This is especially true in the outer wall structure region of the thoracic aorta. This region is usually a thin, circumferential ring around the aortic inner lumen. Small variations within this small volume lead to relatively greater variability and lower DICE scores. Concurrently, if there are small but critical errors in a relatively large volume, the DICE score would remain elevated, but the clinical utility of the image would be diminished. However, this is a common way to compare the performance of segmentation algorithms across methods.

To address the limitations of DICE score metric, clinical utility was demonstrated by comparing first and second-order

measurements, which are important parameters for characterizing AAA progression. Extracting max AP diameter measurements enables the calculation of growth during surveillance and determines timing of surgery.[14,15] Assessing the accuracy of diameter extraction is essential to integrate this DL platform with current methods of AAA management. Cross-sectional area (2-D) has been shown to have the lowest variability in assessing aneurysm change and supplements the 1-D diameter measurements.[16] Evolution of 3-D, especially thrombus volume, and second order indices, including centerline is linked to AAA progression, rupture risk and the incidence of adverse cardiovascular events.[17,18] This automatic method of volume extraction can be used to standardize current methods of aneurysmal disease management and sets the foundation for subsequent complex geometric analysis. Furthermore, the proposed pipeline can be extended to other vascular pathologies.

Before the advent of machine-learning approaches, AAA segmentations were performed using intensity-based semi-automatic algorithms (eg, level-sets, active shape models and graph cut methods).[12,19–21] Their primary drawback was the failure to accurately detect the aneurysm's outer boundary as its intensity is like that of adjacent structures. Although these models may provide good results, there are significant limitations that prevent clinical implementation. These methods are semi-automatic and require significant model optimization. Furthermore, these methods require complex user-input (ex. prior lumen segmentations/centerlines), and are highly data-set dependent.[12,19] The latter limits model robustness and generalizability.

Recently, DL methods on CTAs have been proposed to tackle this problem without encountering many of the limitations of their predecessors. Variations on Deep Belief and U-net based networks have been used to segment the infra-renal region of the aorta.[13,22] Unfortunately, many of these networks are limited to 2-D inputs (axial CT slices), which may fail to appropriately capture the aneurysm's 3D geometry. The accuracy and reproducibility of these models is like that of earlier methods as they are trained and validated on small data sets. Lopez-Linares et al recently proposed a Holistically-Nested Edge Detection network trained in both 2D and 3D that out-performs currently available methods in both pre and post operative AAA segmentation.[23] However, this method is limited to single-class segmentation of the aneurysmal wall and performs poorly with small aneurysms and those with a small thrombus burden.

Current convolutional neural network architectures can capture semantic contextual information by generating a coarse feature-map grid through iterative down-sampling of the input. Features on this coarse map represent location and relationship between structures/tissues at the organ level. However, these architectures struggle to capture small target objects with increased shape variability. This is especially important for pathological vascular cases. Integrating attention gates, which is commonly used in natural image analysis and classification tasks, into this architecture has shown promise in focusing on target structures without the need for additional training.[6] These attention gates can suppress predictions in irrelevant background region and can be trained simultaneously with the underlying network using standard back-propagation methods. The strength of this attention-based U-Net has been documented on the segmentation of abdominal structures[6]; however, its role in aortic segmentation has never before been evaluated. Its superior segmentation performance for aneurysmal segmentation rationalized its incorporation within the full aortic segmentation pipeline.

This is the first time a DL method is used to isolate the aorta/AAA from a noncontrast CT scan. This allows for the extraction of complex morphological information from noncontrast images. Furthermore, the same methodology underpinning this work can be extended to enable automatic segmentation of other structures with or without the use of IV contrast agents.

Although CTAs provide unique insight into aneurysm morphology and the vascular tree, it is not without its disadvantages. Administration of contrast requires needle insertion, which is associated with multiple complications including inadvertent arterial puncture and contrast leak from veins causing skin irritation/damage. Additionally, contrast agents are nephrotoxic and have a 10% incidence of acute kidney injury (contrast-induced nephropathy) after use. This is a problem within the elderly population, who either have decreasing baseline renal function or concomitant chronic kidney disease. Given that a large sub-cohort of patients with aortic aneurysmal disease may have diagnosed renal disease, this study highlights the necessity to re-evaluate the role of noncontrast CT imaging for the management of aneurysmal disease.

CONCLUSIONS

In this study, a novel automated pipeline was developed to enable high resolution segmentation of blood vessels using deep learning approaches. This clinically validated pipeline enables automatic extraction of morphologic features of blood vessels and can be applied for research and potentially for clinical use.

ACKNOWLEDGMENTS

We acknowledge the input by Luke Markham and Hao Xu in the preparation phase of this project. The methods described in this manuscript is subject to a patent filing (PCT/GB2020/052014).

REFERENCES

1. Richmond C. Sir Godfrey Hounsfield. 2004.
2. Whaley ZL, Cassimjee I, Novak Z, et al. The spatial morphology of intraluminal thrombus influences type II Endoleak after endovascular repair of abdominal aortic aneurysms. *Ann Vasc Surg*. 2019;66:77–84.
3. Hesamian MH, Jia W, He X, et al. Deep learning techniques for medical image segmentation: achievements and challenges. *J Digit Imaging*. 2019;32:582–596. AQ5
4. Maier A, Syben C, Lasser T, et al. A Gentle introduction to deep learning in medical image processing. *J Med Phys*. 2018;29:86–101.
5. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015. AQ6
6. Oktay O, Schlemper J, Folgoc LL, et al. Attention U-Net: Learning Where to Look for the Pancreas. 2018.
7. Yoo TS, Ackerman MJ, Lorensen WE, et al. Engineering and algorithm design for an image processing Api: a technical report on ITK–the Insight Toolkit. *Stud Health Technol Inform*. 2002;85:586–592.
8. Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate; 2014.
9. Luong M-T, Pham H, Manning CD. Effective Approaches to Attention-based Neural Machine Translation; 2015.
10. Kollmannsberger P, Kerschnitzki M, Repp F, et al. The small world of osteocytes: connectomics of the lacuno-canalicular network in bone. *New J Phys*. 2017;19. AQ7
11. Huttenlocher DP, Klanderman GA, Rucklidge WJ. Comparing images using the Hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1993;15.
12. Duquette AA, Jodoin PM, Bouchot O, et al. 3D segmentation of abdominal aorta from CT-scan and MR images. *Comput Med Imaging Graph*. 2012;36:294–303.
13. Zheng JQ, Zhou XY, Li QB, et al. Abdominal aortic aneurysm segmentation with a small number of training subjects. arXiv. *XXX*. 2018.
14. Wanhainen A, Verzini F, Van Herzeele I, et al. Editor's choice - European Society for Vascular Surgery (ESVS) 2019 clinical practice guidelines on the management of abdominal aorto-iliac artery aneurysms. *Eur J Vasc Endovasc Surg*. 2019;57:8–93.
15. Chaikof EL, Dalman RL, Eskandari MK, et al. The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *J Vasc Surg*. 2018;67:2.e2–77.e2.
16. England A, Niker A, Redmond C. Variability of vascular CT measurement techniques used in the assessment abdominal aortic aneurysms. *Radiography*. 2010;16:173–181.

17. Haller SJ, Crawford JD, Courchaine KM, et al. Intraluminal thrombus is associated with early rupture of abdominal aortic aneurysm. *J Vasc Surg*. 2018;67:1051–1058.e1.
18. Parr A, McCann M, Bradshaw B, et al. Thrombus volume is associated with cardiovascular events and aneurysm growth in patients who have abdominal aortic aneurysms. *J Vasc Surg*. 2011;53:28–35.
19. Lee K, Johnson RK, Yin Y, et al. Three-dimensional thrombus segmentation in abdominal aortic aneurysms using graph search based on a triangular mesh. *Comput Biol Med*. 2010;40:271–278.
20. Egger J, Freisleben B, Setser R, et al. Aorta segmentation for stent simulation. *MICCAI Workshop on Cardiovascular Interventional Imaging and Biophysical Modelling*. 2009.
21. Siriapisith T, Kusakunniran W, Haddawy P. 3D segmentation of exterior wall surface of abdominal aortic aneurysm from CT images using variable neighborhood search. *Comput Biol Med*. 2019; 107:73–85.
22. Hong HA, Sheikh UU. Automatic detection, segmentation and classification of abdominal aortic aneurysm using deep learning. *2016 IEEE 12th International Colloquium on Signal Processing & Its Applications (CSPA): IEEE*. 2016;242–246.
23. López-Linares K, Aranjuelo N, Kabongo L, et al. Fully automatic detection and segmentation of abdominal aortic thrombus in post-operative CTA images using Deep Convolutional Neural Networks. *Med Image Anal*. 2018;46:202–214.

CE: D.C.; ANNSURG-D-20-03880; Total nos of Pages: 12;
ANNSURG-D-20-03880

ORIGINAL ARTICLE

OPEN

A Deep Learning Approach to Visualize Aortic Aneurysm Morphology Without the Use of Intravenous Contrast Agents

AQ1 *Anirudh Chandrashekar, BE,*‡ Ashok Handa, MBBS, FRCS, MA,*† Pierfrancesco Lapolla, MBBS,* Natesh Shivakumar, BSc, MBChB,* Raman Uberoi, MBBChir, MRCP, FRCR,‡ Vicente Grau, PhD,§ and Regent Lee, MBBS, MS, DPhil, FRCS*†✉*

Background: Intravenous contrast agents are routinely used in CT imaging to enable the visualization of intravascular pathology, such as with abdominal aortic aneurysms. However, the injection is contraindicated in patients with iodine allergy and is associated with renal complications.

Objectives: In this study, we investigate if the raw data acquired from a noncontrast CT image contains sufficient information to differentiate blood and other soft tissue components. A deep learning pipeline underpinned by generative adversarial networks was developed to simulate contrast enhanced CTA images using noncontrast CTs.

Methods and Results: Two generative models (cycle- and conditional) are trained with paired noncontrast and contrast enhanced CTs from seventy-five patients (total of 11,243 pairs of images) with abdominal aortic aneurysms in a 3-fold cross-validation approach with a training/testing split of 50:25 patients. Subsequently, models were evaluated on an independent validation cohort of 200 patients (total of 29,468 pairs of images). Both deep learning generative models are able to perform this image transformation task with the Cycle-generative adversarial network (GAN) model outperforming the Conditional-GAN model as measured by aneurysm lumen segmentation accuracy (Cycle-GAN: 86.1% ± 12.2% vs Con-GAN: 85.7% ± 10.4%) and thrombus spatial morphology classification accuracy (Cycle-GAN: 93.5% vs Con-GAN: 85.7%).

Conclusion: This pipeline implements deep learning methods to generate CTAs from noncontrast images, without the need of contrast injection, that bear strong concordance to the ground truth and enable the assessment of important clinical metrics. Our pipeline is poised to disrupt clinical pathways requiring intravenous contrast.

Keywords: aortic aneurysms, computer vision, computerized tomography, contrast-enhanced computerized tomography, CT angiography, deep learning, generative adversarial network

(*Ann Surg* 2021;xx:xxx–xxx)

In a computed tomography (CT) image, density variations across tissues translate to differences in tissue attenuation and subsequent radiodensities (measured in Hounsfield Units, HU).[1] HU values are generally displayed as a greyscale, with brighter regions corresponding to higher attenuation (eg, bone and calcification) and conversely for darker regions (eg, fat and air). Where treatment of an artery is being considered, a detailed view of the arterial anatomy is required. In the example of abdominal aortic aneurysms (AAA, abnormal ballooning of the abdominal aorta), an intra-luminal thrombus (ILT) adherent to the aortic wall within the enlarging aneurysmal sac is present in 95% of cases[2] (Supplementary Figure S1, http://links.lww.com/SLA/D12). Given the similarities in density, these regions cannot be readily distinguished using a conventional noncontrast CT (NCCT) image.

Clear visualization of these regions can only be achieved by injecting an intravenous (IV) iodinated contrast agent in a CT angiogram (CTA). IV contrast increases the luminal density and attenuation to distinguish the vascular tree from surrounding soft tissues.[1,3] However, CTAs are associated with several disadvantages.[3,4] CTAs are contraindicated in patients with iodine allergies, as most contrast agents are iodine-based. Complications include inadvertent arterial puncture by needle, and contrast leak causing skin irritation/damage.[4] Additionally, contrast agents are nephrotoxic and have up to 12% incidence of reported acute kidney injury following use.[4] This is especially a problem within the elderly population, who either have decreased baseline renal function or concomitant chronic kidney disease. There is a recognized risk of complete kidney failure in these patients, leading to renal dialysis and mortality.

We hypothesize that the raw data acquired from a NCCT can be used to differentiate blood and other soft tissues. Blood is predominantly fluid, with red/white blood cells whereas the adjacent ILT is predominantly fibrinous and collagenous, with red cells/platelets. These individual components vary in ultrastructure and physical density, which should reflect in different (albeit subtle) HUs on a CT scan (either in individual HU values or in their spatial distribution/''texture''). We further hypothesized that using deep learning (DL) generative methods, these subtle differences can be amplified to enable simulation of contrast-enhanced images without the use of contrast agents.

AQ2 From the *Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom; †Department of Vascular Surgery, Oxford University Hospitals, NHS Foundation Trust, United Kingdom; ‡Department of Radiology, Oxford University Hospitals, NHS Foundation Trust, United Kingdom; and §Department of Engineering Science, University of Oxford, Oxford, United Kingdom.

✉regent.lee@nds.ox.ac.uk.

A manuscript describing pilot data of this research was previously published in ArXiv. https://arxiv.org/abs/2003.01223.

The Oxford Abdominal Aortic Aneurysm Study was supported by the following: University of Oxford, Medical Sciences Division Medical Research Fund (MRF/HT2016/2191); University of Oxford, Nuffield Department of Surgical Sciences; John Fell Oxford University Press Research Fund (142/075); National Institute of Health Research (NIHR) Oxford Biomedical Research Centre; RL was supported by the Academy of Medical Science Starter Grant, UK (SGL013/1015). PL was supported by the EU Erasmus+ traineeship studentship. AC is supported by a Clarendon – Sloane-Robinson Scholarship, University of Oxford.

Contents of this manuscript are subject to an international patent filing (PCT/GB2020/052013).

AQ4 The authors report no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

ISSN: 0003-4932/16/XXXX-0001
DOI: 10.1097/SLA.0000000000004835

In this study, we investigate the ability of generative adversarial networks (GANs) for this noncontrast to contrast image transformation task. These networks are a class of DL architectures whereby 2 neural networks train simultaneously, with 1 network focused on data generation (generator) and the other focused on data discrimination (discriminator). Designed to mimic how the human brain operates, these neural networks are sets of algorithms that attempt to understand the underlying relationships in the provided training data for a particular task. In this instance, these networks compete against each other to better learn the underlying statistical distribution of the training data. This allows for the generation of new examples from the same distribution.[5] Many variations of the original GAN have been developed, including conditional GANs (Con-GANs1Conditional GANS are usually abbreviated as CGANs in the machine learning literature. Here, Con-GAN is used to avoid confusion with Cycle-GAN.) and cycle-GANs. The former learns the transformation between 2 paired distributions using a pixel-to-pixel approach.[6] On the other hand, the Cycle-GAN is able to learn transformations between 2 distributions without the need for direct pairings between samples.[5] Here, we assessed the performance of these DL generative models using 2 technical metrics [root-mean-square-error (RMSE), Sørensen–Dice score (DICE)] and 4 clinically important metrics (1. diameter, 2. cross-sectional area, 3. volume of the lumen and outer wall structure, and 4. AAA thrombus morphology).

METHODS

Curation of CT Images From a Clinical Cohort

CT scans were acquired through the Oxford Abdominal Aortic Aneurysm (OxAAA) study. The study received full regulatory and ethics approval from both Oxford University and Oxford University Hospitals National Health Services Foundation Trust (Ethics Ref 13/SC/0250). The OxAAA study was designed to investigate novel biomarkers in the context of AAA disease. Details regarding the OxAAA study have been previously published.[7] The study complies with the principles outlined in the Declaration of Helsinki. Each patient gave consent for the use of clinical and imaging data for research analyses. This research project (contents of this manuscript) was inspired by our attempt to utilize historic CT scans acquired during the AAA surveillance period to discover novel signatures of AAA growth,[8] as many of these historic CT scans were noncontrast scans.

As part of the routine preoperative assessment for aortic aneurysmal disease, a NCCT of the chest/abdomen/pelvis and an arterial phase CTA was performed. CTA images were obtained in helical mode with a predefined slice thickness of 1.25 mm. On the other hand, NCCT images were obtained with a predefined slice thickness of 2.5 mm. Paired contrast and NCCT were anonymized before subsequent analysis. Seventy-five (75) patients with paired NCCT and CTA images (11,243 pairs of images) were randomly selected from the OxAAA cohort and were used for model training. An independent set of 200 patients with paired NCCT and CTA images (29,468 pairs of images) were selected and served as a validation cohort. Paired images were segmented using our proprietary automated DL pipeline[9] and registered/shifted to ensure aortic overlap for subsequent analysis (Supplemental Methods, http://links.lww.com/SLA/D12).

EXPERIMENT 1: HU Sampling Between Different Regions in the AAA

To investigate the regional differences within an aneurysm, 100 paired axial slices were selected from 10 random patients. These slices were visually validated for registration accuracy by 2 blinded reviewers (NS and PL) and were sampled for the underlying HU distribution at the lumen, ILT, and interface locations (Fig. 1). These visually indistinct regions on the NCCT images were identified from their paired CTA images. To account for slight discrepancies in the image registration process, and minimize sampling errors, the thrombus (blue) and lumen (yellow) areas were shrunk by 20% at the adjoining border. The zone between the 2 regions was demarcated as the interface (red). This delineation is clearly indicated in Figure 1. The average HU intensities within each region were compared using a One-way-ANOVA. As a negative control, concentric sampling within the blood lumen in each of the NCCT images was performed.

EXPERIMENT 2: GANs: Cycle-GAN and Con-GAN

In this study, the cycle-GAN and the Con-GAN were used for the noncontrast to contrast image transformation task (Supplementary Figure S2, http://links.lww.com/SLA/D12). Model architectures and training details are described in the Supplemental Methods, http://links.lww.com/SLA/D12.

GAN Models Training and Evaluation

A 3-fold cross-validation paradigm with a training/test data split of 50:25 patients (∼7500: ∼3750 2D axial slices) was employed. Model performances were evaluated using 2 metrics: RMSE and Sorensen-DICE score. The former is a commonly used metric in image transformation tasks to assess the pixel-by-pixel difference between the simulated CTA image and the ground truth (GT) (ie, paired CTA image). The latter is a quantitative similarity assessment of regions within sets of images.[10] Here, $DICE_I$ was calculated to assess overlap in the inner lumen between the simulated pseudo-contrast and GT CTA images. Additionally, $DICE_C$ measures aortic shape overlap between the simulated and GT images. This is a surrogate measure to evaluate both registration accuracy and shape alterations induced by the generative network.

CT Image Quality and GAN Model Performance

The NCCT images used in this study were acquired using a variety of settings (eg, x-ray tube current, rotation time, and pitch factor) that may impact image quality. The combination of x-ray tube current and rotation time determines the amount of incident x-ray photons, which is directly proportional to image quality. Given that these DL platforms utilize these images and the underlying pixel relationships as the basis for transformation, we further hypothesized that NCCT images obtained with greater x-ray tube currents would result in qualitatively and quantitatively better simulated CTA images. This was evaluated by comparing the patients with best and worst image transformations to define an optimal cut-off criterion of image quality and for subsequent analysis.

EXPERIMENT 3: Evaluation of Clinically Important Metrics Using Simulated Pseudo-contrast Images

The following metrics were obtained from the GT CTA and simulated pseudo-contrast images using an in-house algorithm in MATLAB[11]: 1-D measurements: maximum lumen and aortic wall diameters; 2-D measurements: blood lumen and ILT areas; 3-D measurements: blood lumen and ILT volumes. Bland Altman plots and correlation coefficient analysis were performed for each to assess bias and the strength of association between the output of the GAN models and the GTs. Bias measurements are reported along with its 95% confidence interval. In addition, the spatial morphology of the ILT was extracted using an in-house algorithm in MATLAB. This is an important feature that is used to plan surgical intervention and can influence postsurgical outcomes.[7] Based on our previous work, spatial morphology of the ILT can be categorized into 7 categories (Supplementary Figure S3, http://links.lww.com/SLA/D12) as previously detailed.[7]

CE: D.C.; ANNSURG-D-20-03880; Total nos of Pages: 12;
ANNSURG-D-20-03880

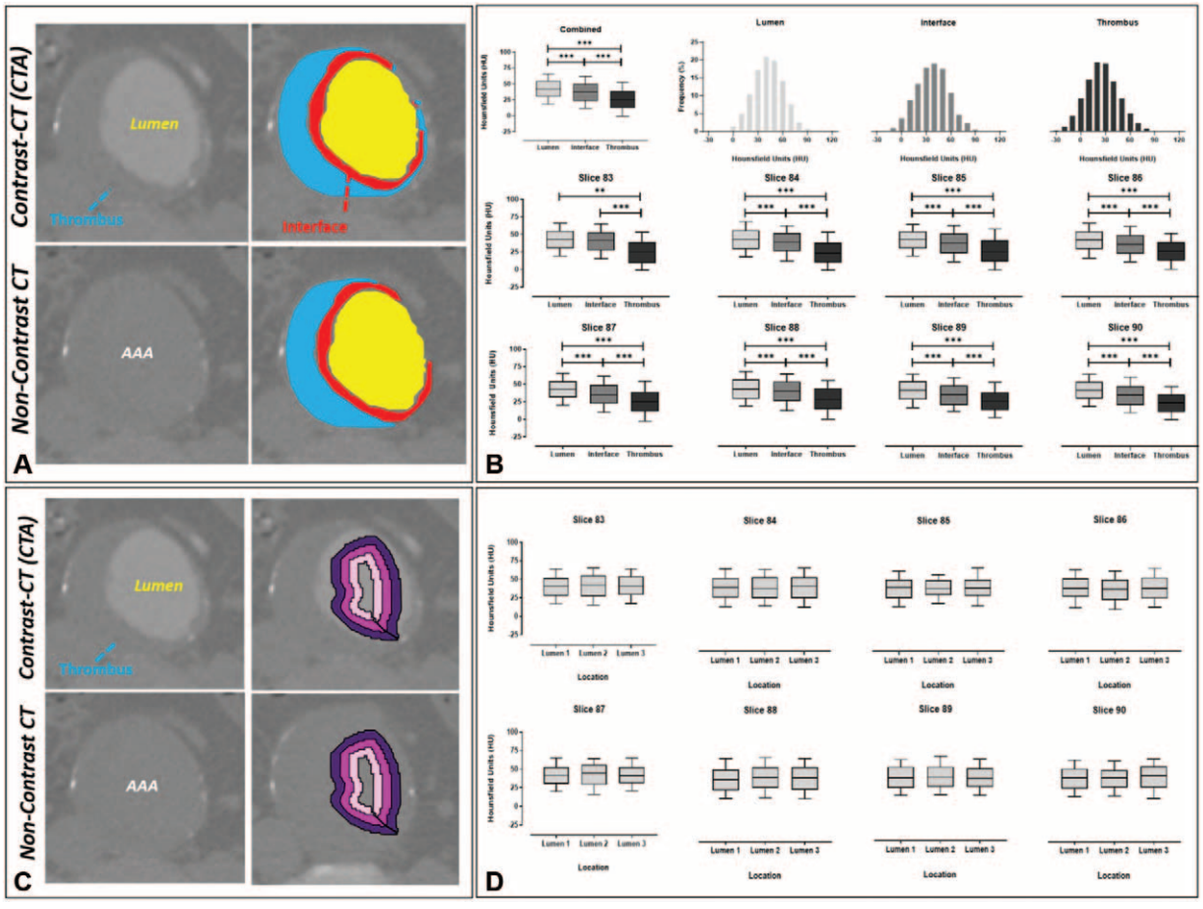


FIGURE 1. Axial slices from both the contrast and noncontrast CT scans. A, Demarcated regions display the thrombus (blue), the lumen (yellow), and the interface between the regions (red). B, Hounsfield unit sampling of the lumen, interface, and thrombus with histograms displaying the frequency of HUs within each region. Analysis at each axial slice was performed and showed differences in HU intensity between the lumen, interface, and thrombus regions. C, Concentric sampling within the lumen as demarcated by the pink, magenta, and purple, was used as a negative control for this experiment. D, Hounsfield unit sampling of the lumen at multiple locations indicated a minimal difference in HU intensity. CT indicates computed tomography; HU, Hounsfield Unit.

AQ9

EXPERIMENT 4: Evaluation of GAN Models on the Independent Validation Cohort

An independent set of 200 patients with paired NCCT and CTA images (total of 29,468 pairs of images) were selected from the OxAAA study and served as a validation cohort. The con- and cycle-GAN models from each validation fold (3) were used to generate pseudo-contrast images of the aorta/aneurysm. The outputs of each model were averaged to generate the final pseudo-contrast images. As a result, for each NCCT image, 2 pseudo-contrast images were generated: 1 from the averaged Con-GAN outputs and the other from the averaged Cycle-GAN outputs. Following image transformation, CTA images were provided to assess transformation accuracy. This was completed using the same methods and metrics as described above. Subsequently, the effect of aneurysmal size (as measured by maximum AAA diameter) and aneurysm shape (fusiform vs saccular) on GAN performance was investigated. The aneurysmal shape was quantified using the non-fusiform index (NFI), which is a 3-D shape index that describes the deviation of the aneurysmal sac from an ideal fusiform shape. This metric was derived from Martufi et al[12] and is explained in detail within the Supplement, http://links.lww.com/SLA/D12.

EXPERIMENT 5: GAN Model Training and Evaluation for the Simulation of Aortic Side Branches

From Experiment 4, we selected the GAN model with superior performance in simulating aortic/aneurysmal-specific features for subsequent analyses. The training data for this experiment consisted of 2-D axial slices centered around the aorta/AAA with surrounding tissue present (aortic region-of-interest, ROI; Supplementary Figure S4, http://links.lww.com/SLA/D12). As a result, this generative network was trained to concurrently simulate both intra- (aortic lumen, thrombus morphology) and extra-aortic/AAA features (main aortic side branch origins: Coeliac artery, superior mesenteric artery, renal arteries, and iliac arteries). Same model training (3-fold cross-validation paradigm, n = 75 patients) and evaluation parameters were utilized for this experiment. Maximum aortic side branch diameter was manually obtained in a blinded fashion from both

CE: D.C.; ANNSURG-D-20-03880; Total nos of Pages: 12;
ANNSURG-D-20-03880

the contrast-enhanced and the GAN-generated pseudo-contrast CT images for the 6 major branches originating from the descending/abdominal aorta, wherever available (1. Coeliac artery, 2. Superior Mesenteric Artery, 3–4. Left/Right Renal Arteries, and 5/6. Left/Right Iliac Arteries). Mean average error between measurements were reported for each branch. Student *t*-testing was performed to assess for statistical significance.

RESULTS

HU Intensities Differ Between Regions of a AAA in NCCT Images

There are discernable differences between the HU frequency distributions of each region (Fig. 1B). Average HU intensity in the NCCT images differed significantly between all 3 regions (blood lumen vs thrombus, blood lumen vs interface, and interface vs thrombus) for all patients assessed. Figure 1B demonstrates the significant differences in HU intensities between the thrombus and lumen for 8 axial slices obtained from 1 patient. Furthermore, the blood lumen/thrombus interface also differed significantly from the other 2 regions, indicating a gradual change from lumen to the thrombus. As a comparison, no significant HU differences were observed following concentric sampling within the lumen (Fig. 1C and D).

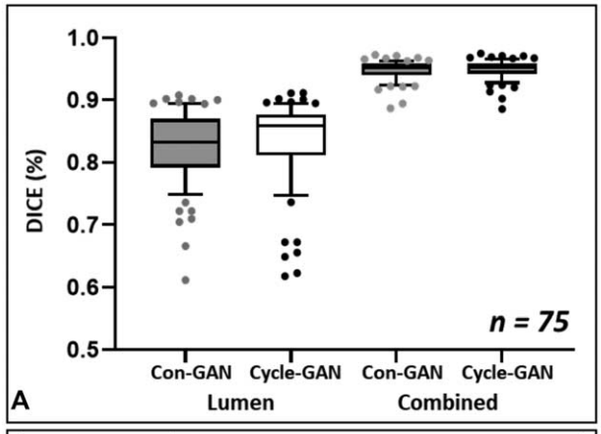
Generative Models Can Simulate Contrast Images Using Noncontrast Images

During model training, the RMSE between the simulated pseudo-contrast images and GT (CTA) images decreased with each epoch (training cycle) to plateau at 3.8 ± 0.8 and 3.9 ± 0.6 for the con-GAN and cycle-GAN, respectively. Similarly, the $DICE_I$ increases with epoch duration to plateau at $91.8\% \pm 0.6\%$ and $92.0\% \pm 0.4\%$ for the con-GAN and cycle-GAN, respectively. Figure 2 indicates the RMSE and DICE scores for the images generated from the testing cohorts across the 3 folds using both GANs (n = 11,243 images, 75 patients). A per-patient transformation accuracy was derived by grouping the 2D-axial images and their respective DICE scores by the patient. With regards to $DICE_I$, the median performance of the cycle-GAN is greater than that of the Con-GAN. Of note, there were multiple overlaps between the outliers below the 10th percentile of cycle-GAN and Con-GAN (6/7) networks. This suggests that there may be image properties inherent to this subgroup of NCCTs leading to decreased transformation accuracy.

CT Image Quality Affects NCCT Transformation Accuracy

Comparing the image properties of the NCCT scans below the 10th percentile and above the 90th percentile highlighted 1 key difference with regards to the x-ray tube current (I_{tube}) used during image acquisition (Supplementary Figure S5, http://links.lww.com/SLA/D12). Scans obtained with lower I_{tube} values produced images with poor transformation accuracy. Subsequently, a threshold criterion of above the 15th percentile of I_{tube} was implemented. This equated to a selection criterion of images obtained with $I_{tube} > 80$ mA. Apart from tube current, there were no observable differences between other CT acquisition parameters (mean ± SD) for the 2 subsets of cases. Notably, no difference was observed in spiral pitch factor (excluded: 1.08 ± 0.17, included: 1.10 ± 0.18, $P = 0.42$), slice thickness (excluded: 2.5 mm, included 2.5 mm, $P = 1.0$), and total collimation width (excluded: 35.9 ± 8.2, included: 34.2 ± 9.3, $P = 0.33$).

Exclusion of these 12 patients ($I_{tube} < 80$ mA), resulted in improvements in RMSE, and in $DICE_I$ for both generative models (Supplementary Figure S6A and B, http://links.lww.com/SLA/D12).



n = 75		Con-GAN	Cycle-GAN
RMSE		4.4 ± 3.7	4.6 ± 4.1
DICE	Inner	85.2 ± 11.4 %	85.6 ± 14.5 %
	Comb.	94.9 ± 4.3 %	95.4 ± 4.0 %

B

FIGURE 2. Transformation accuracy within the testing cohort. A, Box plots of averaged DICE scores per patient within the testing cohort for the lumen ($DICE_I$) and the combined aortic mask ($DICE_C$) segmentations. Segmentations are derived from model (Cycle-GAN/Con-GAN) predictions. Individual data points outside the 10–90th percentile are highlighted. B, RMSE/DICE scores of generated pseudo-contrast axial images. DICE indicates Sørensen–Dice score; RMSE, root-mean-square-error.

$DICE_C$ scores remained unaffected by this exclusion. The blood lumen generated from the Cycle-GAN bears a closer resemblance to the GT image as compared to that generated by the Con-GAN, as reflected by the superior $DICE_I$ scores. On the other hand, $DICE_C$ scores were identical for both sets of model predictions. This is apparent in Figure 3, which illustrates paired CTA/NCCT axial slices alongside their pseudo-contrast images from 6 different patients. The aneurysm volume for each of these patients is illustrated in Figure 4. Correspondingly, the inner lumen areas and thrombus volumes derived from the cycle-GAN model outputs are better approximations to those derived from the GT compared to that of the con-GAN.

Cycle-GAN Outperforms Con-GAN in Simulating Contrast CT Images Based on the Measured Clinical Metrics

Evaluation of aneurysm morphology is useful in defining the biological behavior of an AAA during the natural history of the disease.[7,13–16] Measurements derived from the 2 GAN models' outputs were compared against those obtained from the GT.

1-D Measurements

The Cycle-GAN model is better at approximating the maximum lumen diameter than the Con-GAN model (Fig. 5A). On the other hand, both models have similar strengths in determining the outer vessel wall diameter (Fig. 5B). Maximum inner lumen and outer vessel wall diameters extracted from the model outputs correlated strongly with GT measurements (Cycle-GAN, $\rho = 0.85$ and 0.99, $P < 0.01$; Con-GAN, $\rho = 0.83$ and 0.98, $P < 0.01$).

CE: D.C.; ANNSURG-D-20-03880; Total nos of Pages: 12;
ANNSURG-D-20-03880

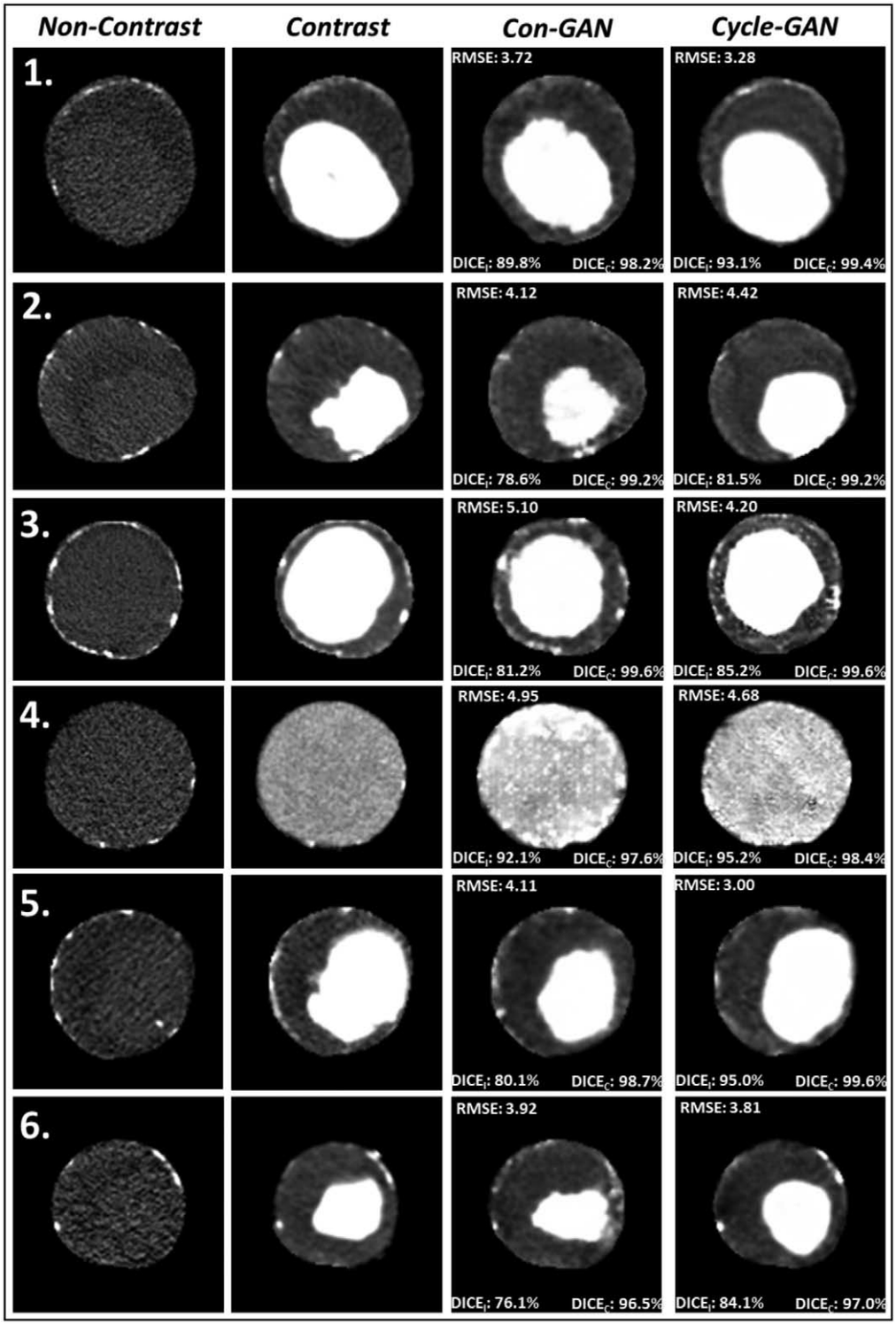


FIGURE 3. Pseudo-contrast images generated using the Con-GAN and Cycle-GAN architectures from 6 different patients alongside their respective NCCT and CTA axial slices. Using the CTA axial images as GT, RMSE, and DICE scores for the inner lumen ($DICE_I$) and combined aorta ($DICE_C$) are displayed. CTA indicates CT angiogram; DICE, Sørensen–Dice score; GAN, generative adversarial network; GT, ground truth; NCCT, noncontrast CT; RMSE, root-mean-square-error.

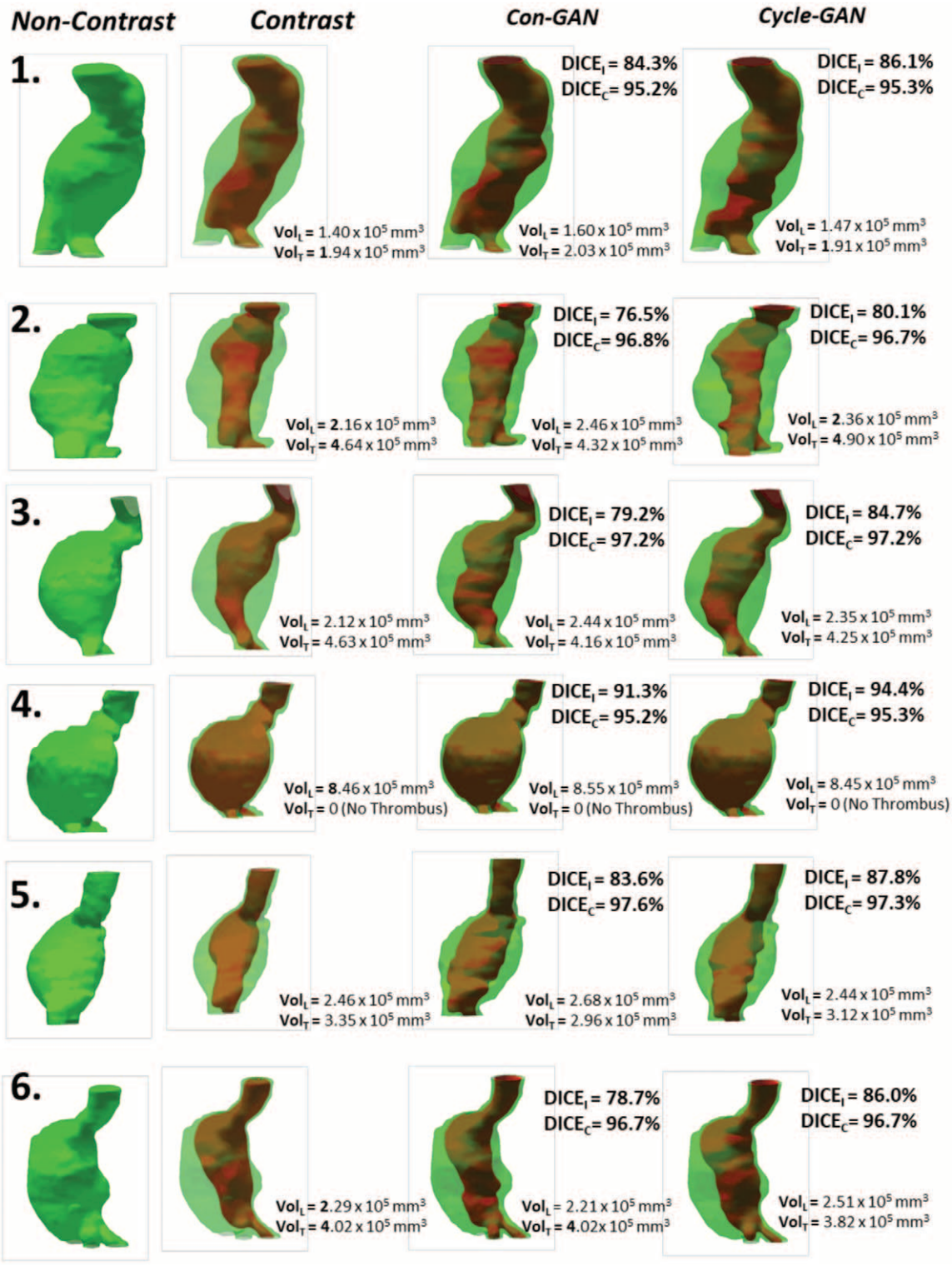


FIGURE 4. Segmented volumes generated from the Con-GAN and Cycle-GAN architectures. Volumes from 6 patients, the same as seen in Figure 6, are displayed alongside their respective NCCT and CTA volumes. $DICE_I$, $DICE_c$ and lumen/thrombus volumes are calculated and displayed for each aneurysmal region. CTA indicates CT angiogram; DICE, Sørensen–Dice score; GAN, generative adversarial network; NCCT, noncontrast CT.

CE: D.C.; ANNSURG-D-20-03880; Total nos of Pages: 12;
ANNSURG-D-20-03880

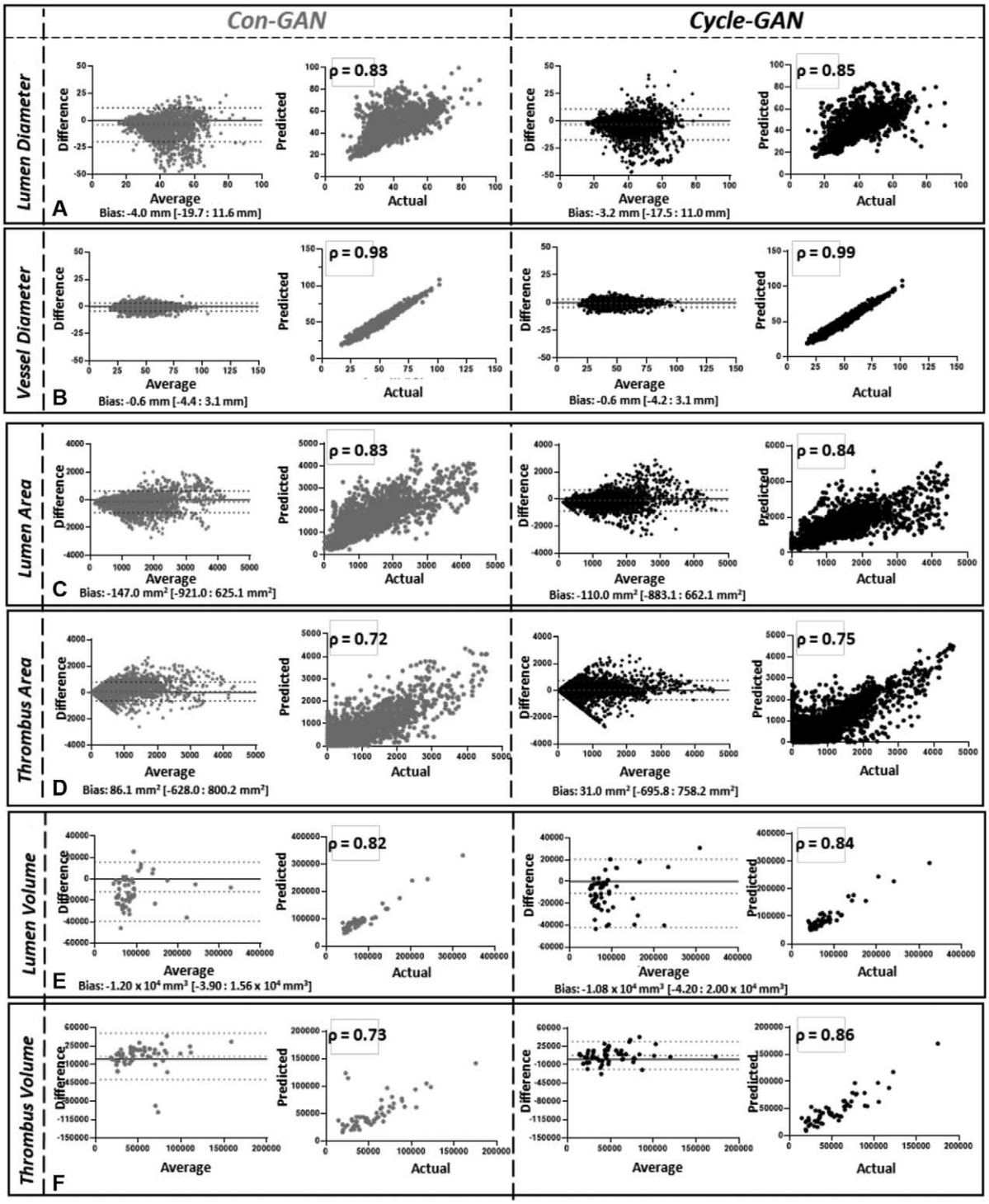


FIGURE 5. One-dimensional diameter (A and B), 2-dimensional area (C and D), and 3-dimensional volume assessment (E and F) of generated images. Bland-Altman plots and correlation-coefficient analysis comparing the measurements of generated images compared against those derived from ground truth segmentations. Measurements derived from Con-GAN and Cycle-GAN outputs are illustrated in grey and black, respectively. Spearman correlation coefficients (ρ) are indicated on the graphs ($P < 0.01$ for all comparisons). GAN indicates generative adversarial network.

CE: D.C.; ANNSURG-D-20-03880; Total nos of Pages: 12;
ANNSURG-D-20-03880

2-D Measurements

The Cycle-GAN model again performed better than the Con-GAN model (Fig. 5C and D). Thrombus area in each axial slice as determined by the Cycle-GAN and Con-GAN models is on average 9.3% ± 11.5% and 9.4% ± 12.2% different from GT measurements.

3-D Measurements

The Cycle-GAN better approximates the 3D-lumen and thrombus volume measurements than the Con-GAN (Fig. 5E and F). Lumen volumes derived from both models (Cycle-GAN: $\rho = 0.84$, $P < 0.01$; Con-GAN: $\rho = 0.82$, $P < 0.01$) and thrombus volumes ($\rho = 0.86$, $P < 0.01$) from the Cycle-GAN correlated strongly with the manually derived measurements.

Pseudo-contrast images within the aneurysmal region produced by Cycle-GAN (Fig. 6A) had an ILT classification accuracy of 93.5%, which outperforms that produced by the generative images of the Con-GAN model (85.7%, Fig. 6B). Examples of the discrepancies between the models are illustrated in Supplementary Figures S7 and S8, http://links.lww.com/SLA/D12.

Similar GAN Performance is Observed Within the Independent Validation Cohort

Of the 200 independent cases, 35 were subsequently excluded as they were obtained at tube currents <80 mA (n = 25) or at unknown tube currents (n = 10). RMSE between the generated pseudo-contrast and CTA images for the 165 patients (Cycle-GAN: 4.2 ± 3.8 and Con-GAN: 4.2 ± 3.7) were similar to that observed in the training set. DICE accuracy of the inner lumen again showed superior performance of the cycle-GAN (84.1% ± 7.2%) when compared against that of the Con-GAN (83.2% ± 7.7%). Extracted 2- and 3-D measurements from this refined testing cohort further supported the finding that the Cycle-GAN is superior at simulating CTA images of the AAA.

As a post-hoc analysis, we compared the GAN performance between the excluded cohort (n = 35) and the rest (n = 165). Area and volume measurements of the AAA lumen/thrombus derived from NCCT images obtained at >80 mA were closer to measurements derived from GT annotations than the excluded cohort ($P < 0.001$ for all comparisons, Supplementary Table S1A and B, http://links.lww.com/SLA/D12). Additionally, pseudo-contrast images within the aneurysmal region produced by Cycle-GAN had an ILT classification accuracy of 90.6%, which outperforms the one produced by the generative images of the Con-GAN model (83.8%). This further supports the use of tube currents as a criterion to assess the quality of DICOM data for analysis.

The role of aneurysmal size (maximum AAA diameter) and shape (NFI) was compared against GAN performance for both models. Supplementary Figure S9A–D, http://links.lww.com/SLA/D12 illustrates 4 aneurysms alongside their respective "ideal" fusiform shape. Aneurysms with a predominantly fusiform shape had lower NFIs compared to those aneurysms with a predominantly saccular shape. This supports the classification of aneurysm shape using NFI. Both AAA size and NFI had no significant impact on transformation accuracy as assessed by DICE score of the inner lumen from both the Cycle- and Con- GANs (Supplementary Figure S9E, http://links.lww.com/SLA/D12). This suggests that trained generative networks do not have an implicit AAA size or shape preference.

Cycle-GAN Can Simulate Extra-aortic/AAA Features Including Aortic Side Branches

Given that the Cycle-GAN generated superior results against that of the Con-GAN, we trained a Cycle-GAN to identify an expanded ROI surrounding the aorta. This concurrently simulates both intra- (aortic lumen, thrombus morphology) and extra-aortic/

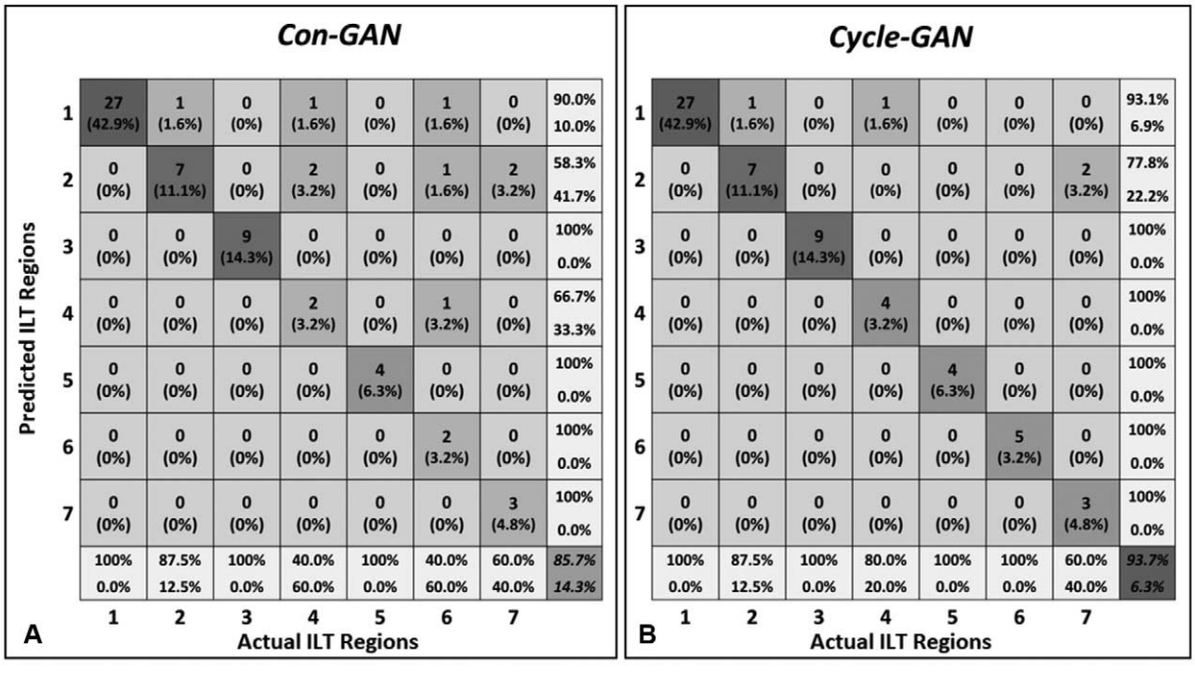


FIGURE 6. Confusion matrices comparing ILT regional classifications between generated images and ground truth segmentations. ILT/WS segmentations derived from model outputs (Con-GAN (A), and Cycle-GAN (B)) were classified into 7 categories as per Fig. S3 and evaluated against that determined by the ground truth segmentations. GAN indicates generative adversarial network; ILT, intra-luminal thrombus.

CE: D.C.; ANNSURG-D-20-03880; Total nos of Pages: 12;
ANNSURG-D-20-03880

AAA features (major abdominal aortic side branches). Despite the expanded input data size (ie, more information to "learn" from), its ability to extract intra-aortic/aneurysmal features are comparable to the previously trained models (smaller ROI/less information to "learn" from) (Supplementary Table S1C, http://links.lww.com/SLA/D12). Figure 7 illustrates generated pseudo-contrast images alongside their respective CTA and NCCT images for 4 patients within the validation cohort. Branch arteries arising from the aorta are visible in all patients. Supplementary Table S2, http://links.lww.com/SLA/D12 highlights the maximum diameters for each branch, obtained in a blinded fashion, from pseudo-contrast CT and its corresponding CTA. There is no difference between the diameter for all 6 major aortic side branches measured using the CTA (GT) or the pseudo-contrast CT (derived from the NCCT images).

DISCUSSION

The primary objective of this study was to investigate whether there are subtle differences between visually indistinguishable regions within the NCCT image. This was required to ensure that the images contained the necessary information for the DL method to generate anatomically-correct visualizations. This was achieved by comparing the HU intensity distributions between regions in the ILT-filled aneurysm. Visually, axial slices within the AAA seem uniform on the NCCT image and the histograms for each of the regions (lumen, interface, and thrombus) display considerable overlap. However, average HU intensity was significantly different between all 3 regions for all patients assessed and a gradual HU change was observed from the aortic lumen to the ILT. This highlights that there are differences, albeit subtle, between the regions that can be exploited and enhanced to estimate the CTA image.

In addition to average HU intensity, the differences between these visually in-distinct regions can be captured using multiple first-order radiomic features (eg, uniformity, kurtosis, and skewness – Supplementary Figure S10, http://links.lww.com/SLA/D12). Radiomics employs advanced data-characterization algorithms to extract underlying pixel relationships and has been used to uncover potential disease features that fail to be appreciated visually. Here, these significant radiomic differences between lumen, thrombus, and its interface in NCCT images strongly support the validity of this image transformation task. The generative networks likely utilize this higher-order information to enhance its performance.

The study's secondary objective was to investigate if DL models (Cycle- and Con-GAN) could extract the subtle differences between soft tissue components in NCCT images within the context of AAA disease and generate CTA images. For both GANs, a 3-fold

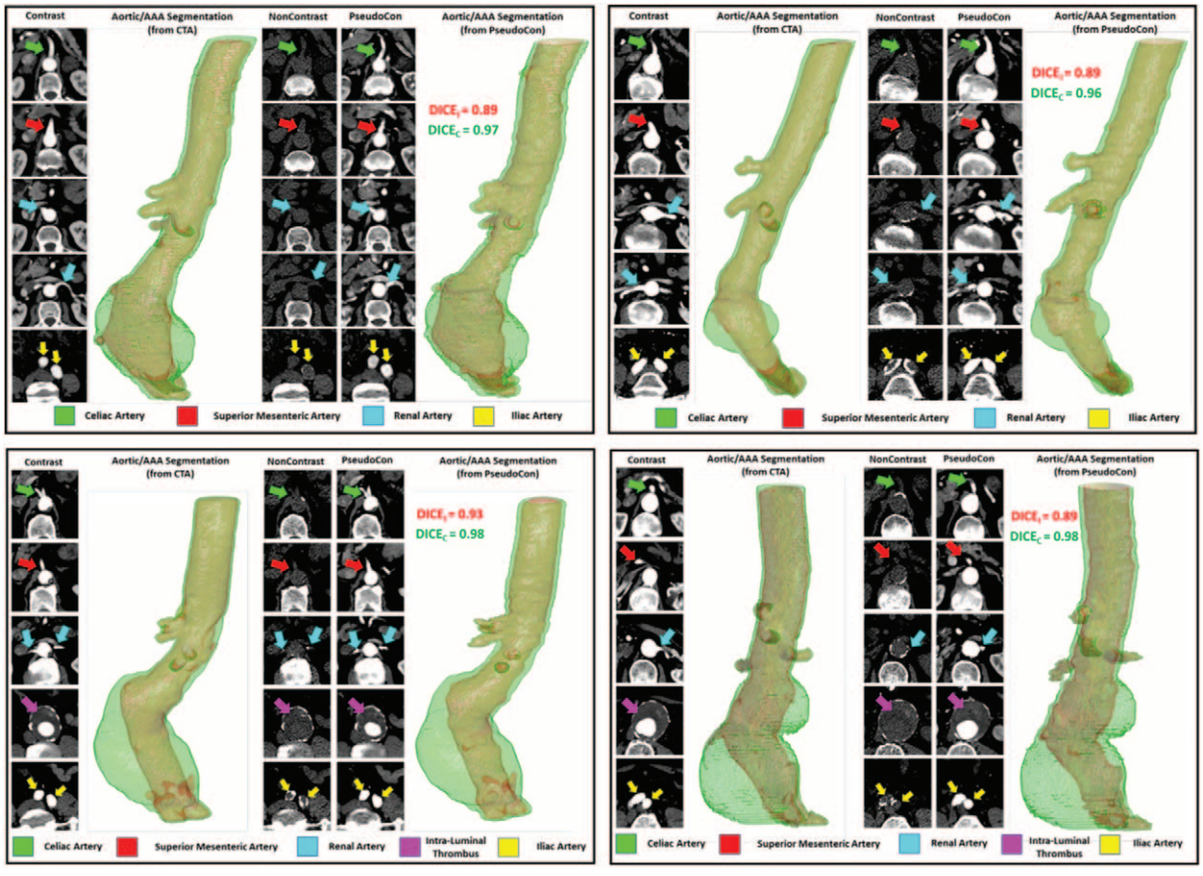


FIGURE 7. Pseudo-contrast CT images are displayed alongside their respective contrast and noncontrast CT Images for 4 patients. Arrows indicate the branch arteries (celiac artery, superior mesenteric artery, renal arteries, and iliac arteries) and the intraluminal thrombus are highlighted wherever visible. Aortic segmentations including the side branches are generated from both the CTA (gold standard) and the pseudo-contrast CT and DICE scores comparing the segmentations are displayed. CTA indicates CT angiogram; DICE, Sørensen–Dice score.

cross-validation approach was employed during training. There was no data leakage between cohorts and the testing cohort for each fold contained a unique set of patients. The 2-D training/testing data was obtained by first isolating the aorta, axial sub-sampling within the volume, and centering the aorta within the extracted 2-D slice. Isolating the aorta by removing surrounding organs and tissues reduces the amount of noise presented to the model. Additionally, centering the aorta in each axial 2D slice reduces the spatial variation of the aorta seen by the generative models. This theoretically should maximize the information learned by the GAN networks for the NCCT to CTA image transformation task. RMSE, a quantitative measure of image difference, and DICE scores for the inner lumen and combined aortic mask were used to optimize the training parameters. DICE accuracy of the inner lumen is an appropriate measure of transformation accuracy as it evaluates the primary goal of the generative models.

It was apparent that CT image quality determines transformation accuracy. Exploring the $DICE_I$ metric within the testing cohort highlighted certain patients with not only decreased transformation accuracies but also greater $DICE_I$ variability within the aortic volume. Here, we hypothesized that poor quality NCCT images have correspondingly poor CTA reconstruction accuracies. Image quality and the extent of distortion (''signal to noise'' ratio) of NCCT images is directly related to the number of x-ray photons incident within the target volume. Accordingly, the number of x-ray photons is directly related to the x-ray tube current (I_{tube}), which is the rate of photon production within the x-ray tube, and CT rotation time. Given that CT rotation time is usually constant, we reasoned that images obtained with decreased I_{tube} have an increased likelihood of generating CTA images with decreased accuracy. Therefore, a stringent selection criterion was imposed to exclude potential outliers with I_{tube} below the 15th percentile (<80 mA). Future studies investigating the role of CT acquisition parameters (eg, slice thickness, kVP, tube current) for this transformation task in greater detail is required to determine the limitations of this technique.

An unexpected observation was that cycle-GAN outperformed the con-GAN. In the original context, we hypothesized that the Con-GAN would have superior performance to the Cycle-GAN as it is able to learn pixel-to-pixel transformations between image pairs. (This is not the case for the Cycle-GAN as data is not introduced as paired images.) This superiority was unexpected but can be rationalized to its underlying network architecture and the multiple training losses (eg, cycle-consistency and identity loss terms). This interesting observation sheds some unique philosophical insight into the nature of ''learning'' by neural network and its similarities of human learning for this particular task. For a cycle-GAN, the neural network learns heuristically from all images of the paired GT data, without being constrained to specific pixel relationships as that of a Con-GAN. It is possible, however, this observation is specific to the dataset utilized here.

We showed that these generative models enable the visualization of aortic aneurysm morphology in CT scans obtained without the use of IV contrast and that transformation accuracy is independent of AAA size or shape. Extracting diameter measurements is required in AAA management as it guides the frequency of aneurysm surveillance and determines the timing of surgery.[13,17] 2-D cross-sectional area measurements of the aneurysm have been shown to complement the 1-D diameter measurements as diameter measurements can be subject to substantial interobserver variability and at times may fail to represent the 3-D growth of the aneurysm. As a result, cross-sectional area measurements have been shown to have the lowest variability in assessing aneurysm size.[16] Evolution of 3-D indices, especially thrombus volume, is linked to AAA progression, rupture risk, and even the incidence of adverse cardiovascular events.[15,18] Assessing ILT spatial morphology is important for surgical planning and has been shown to influence postoperative outcomes. We have previously reported that the spatial morphology of native ILT correlates with the onset of type 2 endoleak, which is an adverse outcome following endovascular surgical repair of aneurysms.[7] This is a task that is not possible from the original NCCT image and if achieved, reinforces the clinical impact of using generative networks for this image transformation task. For the first time, the results highlight the ability to assess the regional location of the ILT from NCCT image with high classification accuracy.

In addition, we showed that a Cycle-GAN trained on $135*135$ mm (144×144 pixels) ROIs surrounding the aorta is able to robustly visualize not only aortic/AAA morphology but also extra-aortic structures, including its side branches. The accuracy of the side branch visualization was assessed by measuring the maximum diameter of each branch from the descending/abdominal aorta. The results support the ability to capture this measurement with a mean average error of $\sim$1.5 mm (1–2 pixels). Using GANs for pseudo-contrast CT visualization of the aorta/AAA, its ILT and its side branches from a NCCT is a novel technique and presents clinicians with a safer alternative to the routinely obtained contrast-enhanced CTA. Future studies are required to determine its clinical utility, such as using this alternative imaging method to plan for endovascular grafting.

This DL approach described here can also be applied to reconstruct other anatomical structures (veins, solid organs, etc) without the need for contrast administration. Beyond the potential clinical utility, our method can be applied for research analysis. There is a growing body of literature on the role of the intraluminal thrombus and AAA growth.[15,19,20] The ability to characterize ILT using noncontrast-enhanced CT scans greatly expands the scope for research in this topic, without the need to subject research participants to contrast injection.

CONCLUSIONS

We present a DL approach to visualize aortic aneurysm morphology and its side branches without the use of IV iodinated contrast agents. This platform technique can be applied to different anatomical structures for research and eventually clinical applications.

ACKNOWLEDGMENTS

The authors acknowledge the support of the clinical team at the Oxford Regional Vascular Service.

REFERENCES

1. Foley WD, Karcaaltincaba M. Computed tomography angiography: principles and clinical applications. *J Comput Assist Tomogr*; 27 Suppl 1:S23–30. AQ5
2. Aggarwal S, Qamar A, Sharma V, et al. Abdominal aortic aneurysm: a comprehensive review. *Exp Clin Cardiol*. 2011;16:11–15.
3. Sun Z, Choo GH, Ng KH. Coronary CT angiography: current status and continuing challenges. *Br J Radiol*. 1013;85:495–510.
4. Hinson JS, Ehmann MR, Fine DM, et al. Risk of acute kidney injury after intravenous contrast media administration. *Ann Emerg Med*. 2017;69:577–586.e4.
5. Zhu J-Y, Park T, Isola P, et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. 2017. AQ6
6. Isola P, Zhu J-Y, Zhou T, et al. Image-to-Image Translation with Conditional Adversarial Networks. 2016.
7. Whaley ZL, Cassimjee I, Novak Z, et al. The spatial morphology of intraluminal thrombus influences type II endoleak after endovascular repair of abdominal aortic aneurysms. *Ann Vasc Surg*. 2019. AQ7
8. Chandrashekar A, Handa A, Lapolla P, et al. Prediction of abdominal aortic aneurysm growth using geometric assessment of computerised tomography images acquired during the aneurysm surveillance period. *Ann Surg*. 2020.

9. Chandrashekar A, Handa A, Shivakumar N, et al. A deep learning pipeline to automate high-resolution arterial segmentation with or without intravenous contrast. *Ann Surg*. 2020.
10. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015;15:29.
11. MATLAB. Version 9.6 (2019a) [Computer Program]. Natick, Massachussets: The MathWorks Inc.; 2019.
12. Martufi G, Di Martino ES, Amon CH, et al. Three-dimensional geometrical characterisation of abdominal aortic aneurysms: image-based wall thickness distribution. *J Biomech Eng*. 2009;131:061015.
13. Wanhainen A, Verzini F, Van Herzeele I, et al. Editor's choice - European Society for Vascular Surgery (ESVS) 2019 clinical practice guidelines on the management of abdominal aorto-iliac artery aneurysms. *Eur J Vasc Endovasc Surg*. 2019;57:8–93.
14. Cayne NS, Veith FJ, Lipsitz EC, et al. Variability of maximal aortic aneurysm diameter measurements on CT scan: significance and methods to minimise. *J Vasc Surg*. 2004;39:811–815.
15. Parr A, McCann M, Bradshaw B, et al. Thrombus volume is associated with cardiovascular events and aneurysm growth in patients who have abdominal aortic aneurysms. *J Vasc Surg*. 2011;53:28–35.
16. England A, Niker A, Redmond C. Variability of vascular CT measurement techniques used in the assessment abdominal aortic aneurysms. *Radiography*. 2010;16:173–181.
17. Chaikof EL, Dalman RL, Eskandari MK, et al. The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *J Vasc Surg*. 2018;67:2–77.e2.
18. Haller SJ, Crawford JD, Courchaine KM, et al. Intraluminal thrombus is associated with early rupture of abdominal aortic aneurysm. *J Vasc Surg*. 2018;67:1051–1058.e1.
19. Metaxa E, Kontopodis N, Tzirakis K, et al. Effect effect of intraluminal thrombus asymmetrical deposition on abdominal aortic aneurysm growth rate. *J Endovasc Ther*. 2015;22:406–412.
20. Zhu C, Leach JR, Wang Y, et al. Intraluminal thrombus predicts rapid growth of abdominal aortic aneurysms. *Radiology*. 2020;294:707–713.

CE: D.C.; ANNSURG-D-20-03756; Total nos of Pages: 10;
ANNSURG-D-20-03756

ORIGINAL ARTICLE

Prediction of Abdominal Aortic Aneurysm Growth Using Geometric Assessment of Computerized Tomography Images Acquired During the Aneurysm Surveillance Period

AQ2 AQ3 *Anirudh Chandrashekar, BE,*† Ashok Handa, MBBS, FRCS, MA,* Pierfrancesco Lapolla,* Natesh Shivakumar, MBBS,* Elisha Ngetich, MBBS,* Vicente Grau, PhD,† and Regent Lee, MBBS, MS, DPhil (Oxon), FRCS (Vasc Surg)*✉*

Objective: We investigated the utility of geometric features for future AAA growth prediction.

Background: Novel methods for growth prediction of AAA are recognized as a research priority. Geometric feature has been applied to predict cerebral aneurysm rupture, but not examined as predictor of AAA growth.

Methods: Computerized tomography (CT) scans from patients with infrarenal AAAs were analyzed. Aortic volumes were segmented using an automated pipeline to extract AAA diameter (APD), undulation index (UI), and radius of curvature (RC). Using a prospectively recruited cohort, we first examined the relation between these geometric measurements to patients' demographic features (n = 102). A separate 192 AAA patients with serial CT scans during AAA surveillance were identified from an ongoing clinical database. Multinomial logistic and multiple linear regression models were trained and optimized to predict future AAA growth in these patients.

Results: There was no correlation between the geometric measurements and patients' demographic features. APD (Spearman $r = 0.25$, $P < 0.05$), UI (Spearman $r = 0.38$, $P < 0.001$) and RC (Spearman $r = -0.53$, $P < 0.001$) significantly correlated with annual AAA growth. Using APD, UI, and RC as 3 input variables, the area under receiver operating characteristics curve for predicting slow growth (<2.5 mm/yr) or fast growth (>5 mm/yr) at 12 months are 0.80 and 0.79, respectively. The prediction or growth rate is within 2 mm error in 87% of cases.

Conclusions: Geometric features of an AAA can predict its future growth. This method can be applied to routine clinical CT scans acquired from patients during their AAA surveillance pathway.

Keywords: abdominal aortic aneurysms, curvature, geometric modeling, undulation

(*Ann Surg* 2020;xx:xxx–xxx)

Abdominal aortic aneurysms (AAA) are an abnormal degenerative condition characterized by pathological dilatations of the abdominal region of the aorta. Clinically, an AAA is defined when the aortic diameter is >50% of the healthy aorta adjacent to the aneurysm. The natural history of an untreated AAA consists of progressive dilatation with eventual rupture and death. The clinical management of AAAs consist of screening/diagnosis, regular surveillance, and timely surgical intervention by open surgical repair or endovascular stent grafting.[1,2]

Methods for the prediction of AAA growth is considered as a priority for research in the opinions of vascular and endovascular surgeons.[3] Accurate prediction of AAA growth in patients can allow for the optimization of surveillance intervals and better inform the timing for surgery. Our prior work has highlighted the feasibility of AAA growth prediction using physiological and biochemical measurements obtained from patients.[4–6] These measurements, however, require additional research steps to the routine clinical care pathway.

Computerized tomography (CT) scans are utilized extensively as diagnostic tests in medicine and surgery. Globally around 150 million CT scans are performed each year.[7,8] In the management pathway of AAAs, each patient requires 1 dedicated CT scan before surgery to plan for the operative approach. During the small AAA surveillance period, a proportion of these patients would have also undertaken CT scan(s) for other medical reasons. In these patients, these historic CT scan(s) performed sometime before AAA surgery could serve as the "baseline" scan. The subsequent pre-operative CT scan would serve as a serial scan and enables growth rate calculation during the time period between the 2. With relevant regulatory approval, these historic baseline CT scans can be anonymized and utilized to discover novel features to predict future AAA growth.

As AAAs enlarge, a variety of geometrical changes are observed including altered aortic tortuosity[9] and increased aneurysmal asymmetry.[10] Several of these changes result in a unique non-uniform distribution of wall stress and have been hypothesized to either favor AAA growth deceleration or increase rupture risk.[11,12] Undulation is a measure of the degree of surface irregularity and asymmetry. Undulation index (UI) of cerebral aneurysms has been utilized to quantify its risk of rupture.[13,14] In this regard, regions of increased curvature [radius of curvature (RC)] within the aneurysm segment contributes to non-laminar fluid flow, and non-uniform wall shear stress.[15,16] No prior literature has examined the utility of UI or RC for AAA growth prediction. Here we hypothesized that aneurysms with an increased degree of surface undulation and with local regions of increased curvature are prone to rapid growth.

METHODS

Patient Cohort

The study was conducted as part of the ongoing Oxford Abdominal Aortic Aneurysm (OxAAA) study (Ethics approval

AQ4 From the *Nuffield Department of Surgical Sciences, University of Oxford, Oxford, United Kingdom; and †Department of Engineering Science, University of Oxford, Oxford, United Kingdom.
✉regent.lee@nds.ox.ac.uk.

The Oxford Abdominal Aortic Aneurysm Study was supported by the following: University of Oxford, Medical Sciences Division Medical Research Fund (MRF/HT2016/2191); University of Oxford, Nuffield Department of Surgical Sciences; John Fell Oxford University Press Research Fund (142/075); National Institute of Health Research (NIHR) Oxford Biomedical Research Centre; AC is funded by the Clarendon-Keble Scholarship, University of Oxford. PL was supported by an EU Erasmus+ traineeship studentship. RL was supported by a Academy of Medical Science Starter Grant, UK (SGL013/1015).

The authors declare no competing interests.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.annalsofsurgery.com).

ISSN: 0003-4932/16/XXXX-0001
DOI: 10.1097/SLA.0000000000004711

TABLE 1. Summary of Participant Demographics at the Pre-surgical Assessment and Significance of Spearman Correlation With Extracted Geometric Parameters

		Significance of Spearman Correlation Coefficient		
	All Participants (n = 102)	**Diam.**	**UI**	**RC**
Male (%)	99 (97)	0.35	0.36	0.09
Age at consent (median/IQR)	72 (67–79)	0.12	0.83	0.70
Height (±SD)	1.75 ± 0.08	0.07	0.45	0.06
Weight (Median/IQR)	81.9 (74–90.2)	0.17	0.15	0.07
BMI (Median/IQR)	26.8 (24.3–28.7)	0.61	0.35	0.18
MAP (±SD)	102.2 ± 12.8	0.13	0.70	0.20
Current smoker (%)	24 (24)	0.57	0.99	0.09
Past smoking Hx (%)	68 (67)	0.97	0.48	0.59
Never smoked (%)	13 (13)	0.65	0.34	0.44
CAD Hx (%)	33 (32)	0.25	0.15	0.56
Coronary intervention (%)	26 (25)	0.63	0.41	0.88
PAOD History (%)	16 (16)	0.85	0.59	0.22
Cerebral art. disease (%)	12 (12)	0.10	0.16	0.34
HTN history (%)	74 (73)	0.53	0.77	1.00
Hypercholesterolemia (%)	61 (60)	0.61	0.67	0.75
Tot. cholesterol (median/IQR)	3.8 (3.2–4.6)	0.23	0.76	0.63
HDL (median/IQR)	1.1 (0.9–1.3)	0.47	0.41	0.06
LDL (median/IQR)	1.5 (0–2.5)	0.26	0.92	0.37
TG (median/IQR)	1.2 (0.8–1.6)	0.49	0.65	0.66
Diabetes (%)	16 (16)	0.88	0.26	0.78
HbA1C (median/IQR)	5.6 (5.4–5.9)	0.19	0.81	0.41
Diabetes - oral/insulin (%)	12 (12)	0.19	0.22	0.17
CKD - eGFR <60 (%)	28 (27)	0.15	1.00	0.96
Creatinine (median/IQR)	86.5 (73.3–101.3)	0.19	0.07	0.74
Beta-blockers (%)	32 (31)	0.68	0.30	0.49
ACEI/ARB (%)	56 (55)	0.27	0.77	0.97
Aspirin (%)	47 (46)	0.83	0.89	1.00
Thienopyridine (%)	9 (9)	0.55	0.17	0.44
Ticragrelor (%)	3 (3)	0.81	0.64	0.72
Anticoagulant (%)	12 (12)	0.93	0.46	0.21
CCBs (%)	43 (42)	0.13	0.64	0.34
Diuretics (%)	22 (22)	0.47	0.38	0.58
Gastro-restraint (%)	31 (30)	0.57	0.68	0.32
Steroids (%)	7 (7)	0.36	0.97	0.15
Statins (%)	74 (73)	0.57	0.76	0.24
AAA Diam (Median/IQR)	63 (58–72.5)			
UI (±SD)	0.23 ± 0.08			
RC (Median/IQR)	35.9 (29.7–46.9)			

Participant demographics were collected at the pre-surgical assessment and were correlated against the extracted geometric parameters of AAA diameter, undulation index, and radius of curvature. Characteristics that follow a Gaussian/Normal Distribution are indicated with a +. For such variables, mean ± SD are presented, and cohort differences are compared using a Student *t*-test. For variables that do not follow a Gaussian distribution, median and inter-quartile range (IQR) are presented and cohort differences are compared using a Mann-Whitney test.

ACEI indicates angiotensin converting enzyme inhibitors; ARB, angiotensin receptor blocker; BMI, body mass index; CAD, coronary artery disease; CCB, calcium channel blockers; CKD, chronic kidney disease; HDL, high density lipoprotein; HTN, arterial hypertension; Hx, history of; IQR, Interquartile range; LDL, low density lipoprotein; MAP, mean arterial pressure; PAD, peripheral arterial occlusive disease; PAOD, peripheral arterial occlusive disease; RC, radius of curvature; SD, standard deviation; TG, triglyceride; UI, undulation index.

Ref: SC/0250/13). The study complies with the principles outlined in the Declaration of Helsinki. Details regarding the OxAAA study cohort and recruitment process have been published.[17] This study cohort consists of 2 arms.

In the first arm, participants were prospectively recruited to the OxAAA study at the time of surgery. Demographic information were recorded from each patient which were matched to the pre-operative CT scan. Height and weight of the patient were measured to calculate their body mass index. The history of coronary artery disease is defined by angina, myocardial infarction, and/or previous coronary interventions (angioplasty or bypass). The history of peripheral arterial occlusive disease is defined by intermittent claudication, critical limb ischemia and/or previous lower limb arterial intervention (angioplasty or bypass). The history of cerebral arterial disease is defined by transient ischaemic attack or stroke. The history of hypertension, diabetes mellitus, and hypercholesterolemia were as diagnosed by the primary care/general practitioner. We further measured the individual's blood pressure, blood cholesterol profile, and HbA1c level to ascertain the effect of their pharmacological therapy. Their current medications were further recorded (Table 1).

In the second arm, we utilized the clinical database (Oxnet Janus), which prospectively registered every patient who underwent elective repair of AAAs at the John Radcliffe Hospital, Oxford, UK. Each of the patients gave written consent for the utilization of clinical images collected during the routine clinical management pathway for research analysis. Radiology records of all non-emergency infra-renal AAA repairs (open surgery or endovascular repair) from February 1st, 2009 to June 30th, 2018 were examined. Only those patients with at least 1 historic CT scan conducted greater than 8 months before the preoperative scan were included for analyses.

Automated Segmentation of CT images

AAA segmentation was performed using a proprietary automated deep learning segmentation platform.[18] The deep learning model generated aortic/aneurysm segmentations which were visually assessed against the source Digital Imaging and Communications in Medicine standard images. Where required, further manual adjustments were performed using the open source ITK-Snap software.[19]

Extraction of Geometric Features

The geometric features included in our analyses were derived from the aortic/aneurysm wall and can therefore be derived from either contrast or non-contrast enhanced CT images. All geometric features were extracted from the above generated volume segmentations using MATLAB. AAA size was measured by calculating the maximum anteroposterior diameter (APD). Annual aneurysmal growth was derived by subtracting the historic/baseline AAA size from pre-operative AAA size and dividing the difference by the time duration (years) between scans (Annual Growth = Δ AAA size in mm/(number of days lapsed between scans/365 days)).

Maximal APD

Maximum APD was automatically extracted from the aneurysmal segmentations using MATLAB. For each axial slice along the aortic volume, the diameter was obtained by measuring the maximum distance between 2 points on the aneurysmal boundary in the sagittal plane. The maximum APD of the AAA was the maximum value from all axial slices.

Radius of Curvature

RC is a centerline-based metric that captures the degree of curvature along the centerline.[16] Here, RC equals the radius of the circular arc that best approximates the curve between a set of adjacent points. The smaller the circular arc, the smaller the RC and the greater the local curvature (Fig. 1A). On the other hand, the larger the circular arc, the greater the RC, and the lower the local curvature (ex. a straight line). AAA centerlines were calculated using an implementation of the homotopic thinning algorithm[20] and were subsampled using b-spline interpolation methods based on the number of axial slices. RC was calculated for adjacent sets of triplet points and the minimum value was obtained. This described the greatest region of curvature within the aneurysmal volume.

Undulation Index

Aneurysmal UI was defined as $\mathrm{UI} = 1 - \left(\frac{V}{V_{\mathrm{CH}}}\right)$.[13,14] Here, V is the volume of the infra-renal abdominal aorta defined as the region between the renal arteries and the iliac bifurcation and V_{CH} is the volume of its convex hull (Fig. 1B). In this instance, the convex hull of AAA is the smallest volume that encompasses the entire region and is convex at all points. It effectively resembles a plastic wrap attached to the inlet and stretched over the entire aneurysmal surface. This parameter captures the degree of surface concavity and increases with surface/shape irregularity. Conversely, a shape that is nonconcave (ex. a perfect sphere) will have a UI of 0.

Developing the Growth Prediction Models

To ascertain if APD, UI, and RC were independent of the patient demographic profile for the purpose of AAA growth prediction, we first examined the relationship (Spearman correlation) between these extracted geometric features against demographic features in the prospectively recruited cohort. We then focused the CT image analysis in the second arm of the cohort, using the geometric features extracted from the baseline scan to predict future growth (as recorded by the subsequent preoperative scan).

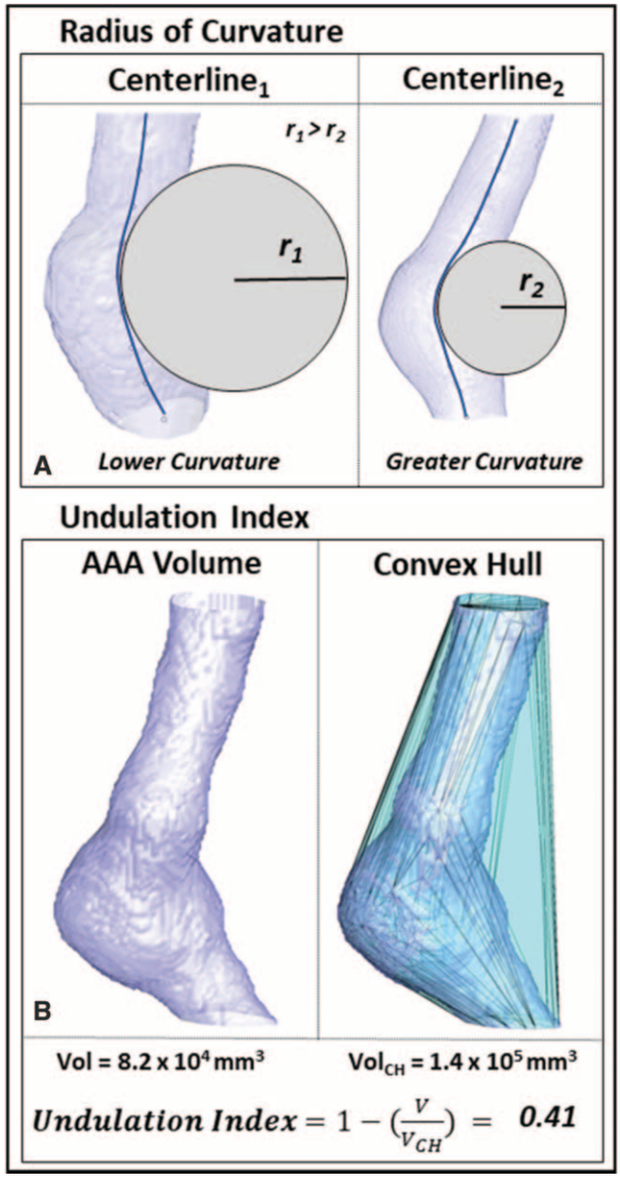


FIGURE 1. Radius of curvature (RC) (A) and undulation index (UI) (B) measurements obtained from abdominal aortic aneurysm (AAA) volumes. AQ13

Summary statistics are described either as average (+/– standard deviations) or median (with interquartile range, IQR). For statistical comparisons, *t*-tests or ANOVA were used for normally distributed data, whereas Kruskal-Wallis test or Spearman correlation were used as nonparametric tests. Multinomial logistic regression and multiple linear regression models were applied to build the AAA growth prediction models, for the prediction of AAA growth as a categorical (slow, some, fast growth) or continuous variable. Data from the second arm of this cohort was randomly split into training (n = 100) and testing (n = 92) sets. Model training and optimization were performed with the training cohort with 10-fold optimization. Here, the data is partitioned into 10 equal sized subsets (folds). Then, 9-folds are used to train the model and the remaining fold is used for internal evaluation. This process is performed a total of 10 times,

with a different fold as the internal validation data set. The optimal parameters were selected, which are then used for the test cohort. Here, the optimized models were used to predict AAA growth on the independent testing cohort (n = 92). Model performance was evaluated using correlation coefficient analysis and root-mean-square-error (RMSE) difference between the predicted and measured AAA growth rates.

For the logistic regression model, growth rates were categorized into 3 groups: <2.5 mm/yr (slow growth), 2.5–5.0 mm/yr (some growth), and >5.0 mm/yr (Fast growth). These thresholds were chosen based on the summary statistics of growth rate observed within this cohort. Receiver operating characteristic curves were plotted for each combination on the testing cohort to assess the performance of the regression model in discerning growth against a pre-defined growth rate threshold. Accuracy of the linear regression model was reported using RMSE between the actual annual growth versus predicted growth. We further assessed the prediction accuracy within a 2 mm margin as this is the accepted technical variability between measurements of AP Diameter in CT images.[21,22]

RESULTS

Patient Demographic Features Do Not Influence the Geometric Features of AAA

One hundred two patients were included the prospectively recruited arm of this cohort (Male n = 99; Females n = 3). The median age at the time of consent was 72 (IQR: 67–79) years old. The majority were ex-smokers (67%) and 24% were current smokers. A history of symptomatic atherosclerotic arterial disease was prevalent in this group (ischaemic heart disease: 32%; peripheral arterial occlusive disease: 16%; cerebral vascular disease: 12%). The majority of participants reported a prior diagnosis of arterial hypertension (73%) and hypercholesterolemia (60%). However, these were well controlled by long term pharmacological therapy [anti-hypertensive(s): 67%, statin: 73%, anti-platelet(s): 46%], as reflected by their controlled mean arterial pressure (102 ± 13 mm Hg) and overall normal cholesterol profiles [median = 3.8 mmol/L (IQR 3.2–4.6), lower than 5.2 mmol/L in 82% of participants] at the time of recruitment. Sixteen percent of the participants reported a history of diabetes mellitus, and 27% had chronic kidney disease with eGFR <60. Baseline demographic data from the prospective cohort (n = 102) are presented in Table 1. Median AAA size within this cohort was 63.0 mm with an interquartile range from 58.0 to 72.5 mm. No correlations were identified between the demographic parameters and the extracted geometric features, indicating that AAA size or APD, UI, and RC are independent of patient demographic characteristics.

Prediction of AAA Growth Using Geometric Features

The second arm of this study included 192 AAA patients with serial CT scans obtained at least 8 months apart. Median follow-up time between scans was 2.0 years with an interquartile range of 1.0–3.7 years. Similarly, median aneurysmal growth was 3.7 mm/yr with an interquartile range of 2.5 to 5.0 mm/yr. These were randomly split into training (n = 100) and validation (n = 92) sets. There were no differences in the follow up duration and geometric indices between the training and validation datasets (Table S1, http://links.lww.com/SLA/C839).

There were significant positive correlations between AAA size (Spearman $r = 0.25$, $P < 0.05$) and UI (Spearman $r = 0.38$, $P < 0.001$) with annual AAA growth rate (Fig S1A and B, http://links.lww.com/SLA/C839). Whereas a significant negative correlation between minimum RC and annual AAA growth rate was observed. (Spearman $r = -0.53$, $P < 0.001$, Fig S1C, http://links.lww.com/SLA/C839). Figures 2 and 3 illustrates 6 AAAs of similar diameter at baseline with disparate UI and RC. These corresponded to different annual growth rate observed subsequently.

Prediction of AAA Growth as a Categorical Outcome

The baseline characteristics of the subgroups of patients [(slow growth: < 2.5 mm/yr); (some growth: 2.5 mm–5 mm/yr), (fast growth: >5 mm/yr)] are summarized in Table S2, http://links.lww.com/SLA/C839. There was no difference in the starting AAA diameter between the 3 groups. UI at baseline differed significantly between the 3 groups (ANOVA, $P = 0.003$). There was also a significant inverse trend of relationship between RC and AAA growth (Kruskal-Willis, $P = 0.003$).

Different combinations of input features (APD, UI, and RC) were used to train multiple logistic regression models. The feature combinations used for each model is indicated in Figure 4 and were trained using a 10-fold cross-validation approach. Receiver operating characteristic curves on the testing cohort were plotted for each feature combination with the threshold of "Slow (<2.5 mm) Growth" and " Fast (>5.0 mm) Growth" (Fig. 4). The area under receiver operation curve metric shows good discriminative capacity of AAA growth rate based on all 3 variables at the predefined thresholds. Using APD, UI, and RC as 3 input variables of the prediction algorithm, the area under receiver operation curve for predicting slow growth (<2.5 mm/yr) and prediction fast growth (>5 mm/yr) is 0.80 and 0.79, respectively. This model comprising of 3 variables significantly outperforms the use of AAA diameter alone as the predictor ($P < 0.01$).

Prediction of AAA Growth Rate as a Continuous Variable

Linear regression models were trained and optimized simultaneously using a 10-fold cross-validation approach to predict AAA growth rate (mm/yr) as a continuous variable. Similar to the multinomial logistic model, the linear model trained using Max AP diameter, UI, and RC was able to predict annual AAA growth to a greater accuracy than the model trained using only max AP Diameter (Fig. 5). Predictions from this model were significantly correlated ($r = 0.61$, $P < 0.001$) and closer (RMSE: 1.32 ± 1.44 mm) to that of observed measurements than that of the other models. Model performances on both train and test cohorts are summarized in Table 2. Similarly, this model is able to predict annual AAA growth to within 2 mm error in 87% of cases.

DISCUSSION

In the United Kingdom, patients diagnosed with an AAA through the National AAA screening program are subject to as frequent as 3-monthly surveillance scans.[23] Surveillance of screening-detected or incidentally-diagnosed AAA is also standard clinical practice as recommended by international guidelines.[1,2] Furthermore, in aging populations within developed countries, the health burden of AAA surveillance is expected to rise. As an example, the National AAA screening program alone incurs a net increase of ~2000 patients requiring AAA surveillance each year.

Methods to predict future aneurysmal growth are valuable to both clinicians and patients. Such tools can improve the stratification of AAA surveillance frequency in individuals: those with slow growth AAAs should not require as frequent surveillance, whereas intense surveillance (or early intervention) can be justified with an AAA that is likely to exhibit rapid growth. Here, we present a method of AAA growth prediction which utilizes geometric features derived from clinical CT scans.

CE: D.C.; ANNSURG-D-20-03756; Total nos of Pages: 10;
ANNSURG-D-20-03756

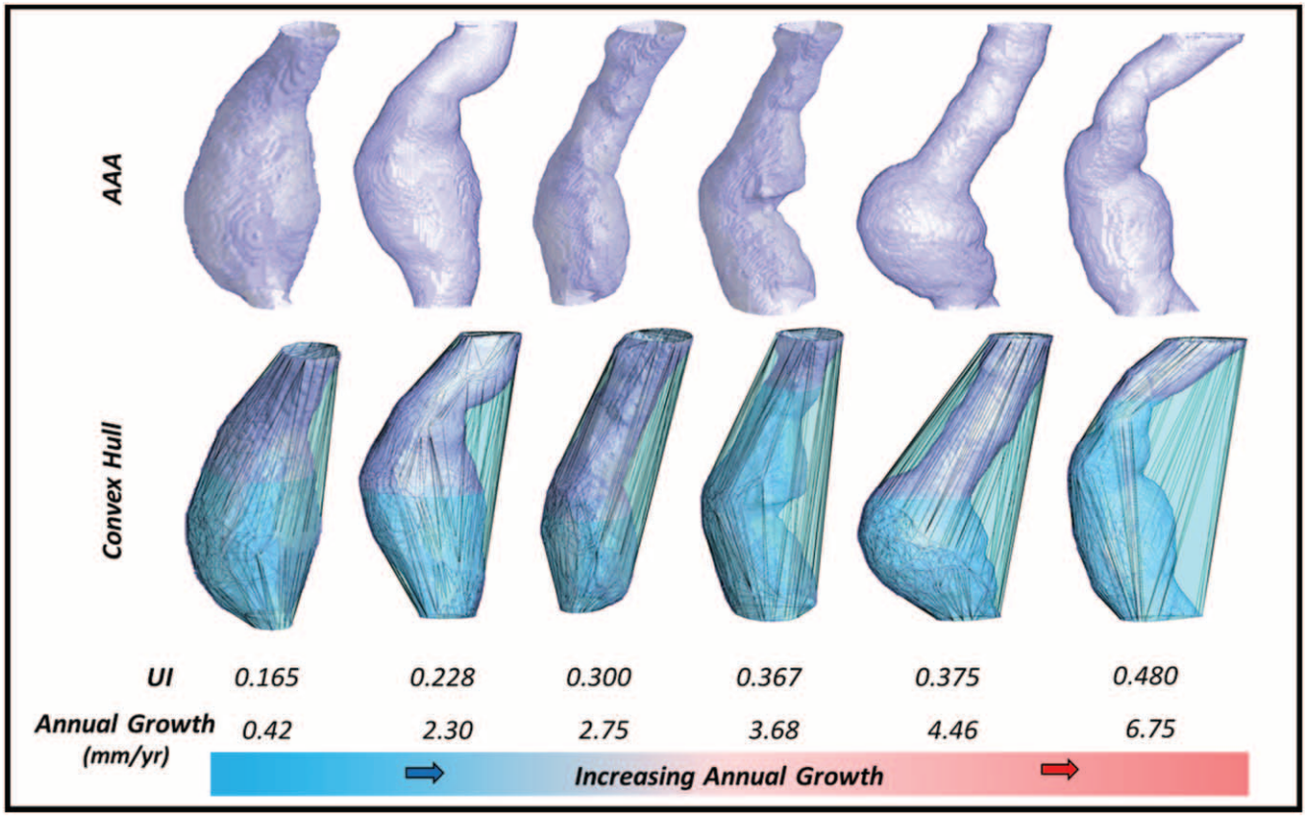


FIGURE 2. Six AAAs of similar size displayed alongside their respective convex hull. Undulation indices are calculated for each aneurysmal pair. Aneurysms are ordered in terms of increasing UI and positively correlate with increasing annual AAA growth rate. AAA indicates abdominal aortic aneurysm; UI, undulation index.

Recent reports have investigated the use of biomechanical assessments of aortic aneurysms from medical images to predict rupture risk potential. Doyle et al. investigated the use of a ratio between aneurysm wall stress to wall strength, derived from magnetic resonance images images, to estimate rupture risk.[24] Their results supported an increase in AAA-related intervention in cases with an elevated ratio after adjusting for AAA diameter and other clinical factors. Additionally, Martufi et al, studied the impact of AAA features (ex. lumen/vessel volume, intraluminal thrombus thickness, etc) and derived wall stress measurement on AAA outcome prediction.[25] Their results display slight but significant improvements to contemporary methods.

However, such biomechanical models require specific assumptions of physiological/computational variables for each individual patient, which has inherent variability but also may not hold true in real life context.[26,27] For example, the aneurysm biomechanical ratio calculated by Doyle et al is subject to many levels of uncertainty including variability with (1) Image reconstruction accuracy, (2) principal wall stress estimation, and (3) population-based wall strength estimation. This results in significant overlap between results for asymptomatic and symptomatic/ruptured cases and prevents clinical implementation.[28] Therefore, it is difficult to draw conclusions about this ratio and other similar wall-stress methods for patient-specific prediction of AAA events. This is further supported by Leemans et al., which suggests that biomechanical indices present no added value in the AAA rupture risk assessment.[29] Wall-stress evaluation was not a component of this study and all geometric features were derived directly from the aneurysmal volume.

There is no prior literature on the prediction of AAA growth using complex geometric features extracted from CTs. This has been examined in the context of cerebral aneurysms. Dhar et al, investigated the use of image-based morphological features to predict intracranial aneurysm rupture.[13] Their study highlighted the importance of 6 features that capture the morphological diversity of intracranial aneurysms and their relation to inflow/outflow vessels. Since then, many of these features have been shown to alter underlying hemodynamics and promote intracranial aneurysmal rupture.[13,14,30] Although there are many similarities between intracranial and aortic aneurysms, these features have not been investigated in the latter. One such parameter, UI captures the degree of surface concavity and increases with surface irregularity.[13,14] A highly asymmetric and/or tortuous infra-renal region would result in an increased UI parameter.

In this study, high undulation seems to arise from a "bent" (correlated with curvature), "bulgy" (a thin cylinder directly attached to a sphere, like the fifth case in Fig. 3) and/or "undulated" shape (2 or more spheres attached by cylinders). This parameter is specific to the aneurysmal infra-renal region of the aorta and does not require additional information including inflow/outflow volumes. Similarly, minimum RC is a descriptor of the AAA centerline, which captures the region of maximum curvature. Here, RC is radius of the circular arc that best approximates the curve between a set of adjacent points. The smaller the circular arc, the smaller the RC and the greater degree of local curvature.

The advantage of the geometric measurements described here (APD, UI, RC) is that they can be readily extracted from either

CE: D.C.; ANNSURG-D-20-03756; Total nos of Pages: 10;
ANNSURG-D-20-03756

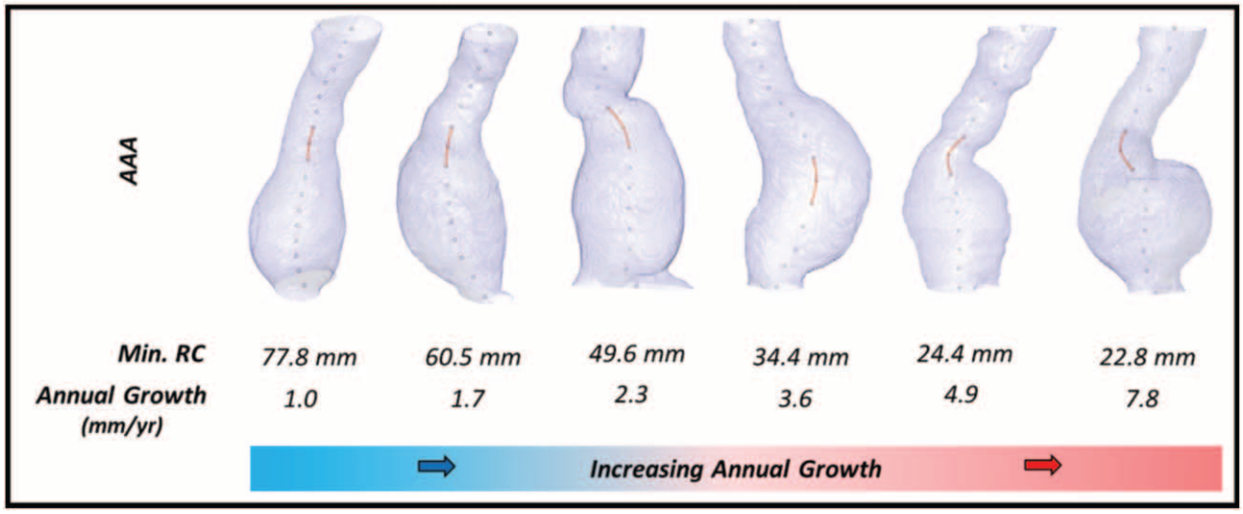


FIGURE 3. Six AAAs of similar size displayed with calculated centerline. Centerline are calculated using a variation of the homotropic thinning algorithm. Regions with the minimum radius of curvature along each centerline are highlighted in orange. These location correspond to region of increased curvature along that respective centerline. Aneurysms are ordered in terms of decreasing min RC and negatively correlate with increasing annual AAA growth rate. AAA indicates abdominal aortic aneurysm; RC, radius of curvature.

contrast or non-contrast CT images, without specific adjustment in CT scanning protocols. This analysis is streamlined by our proprietary automated pipeline for high resolution segmentation of blood vessels using deep learning approaches.[18] That no correlation was observed between these geometric features to the patient characteristics (as summarized in Table 1) further supports our claim that they can be independently deployed as predictive indices without accounting for patients' demographic characteristics. Further validation of our prediction algorithm can therefore be attempted using historic scans already accumulated by vascular surgical units who have an existing clinical image database as part of their routine AAA management practice.

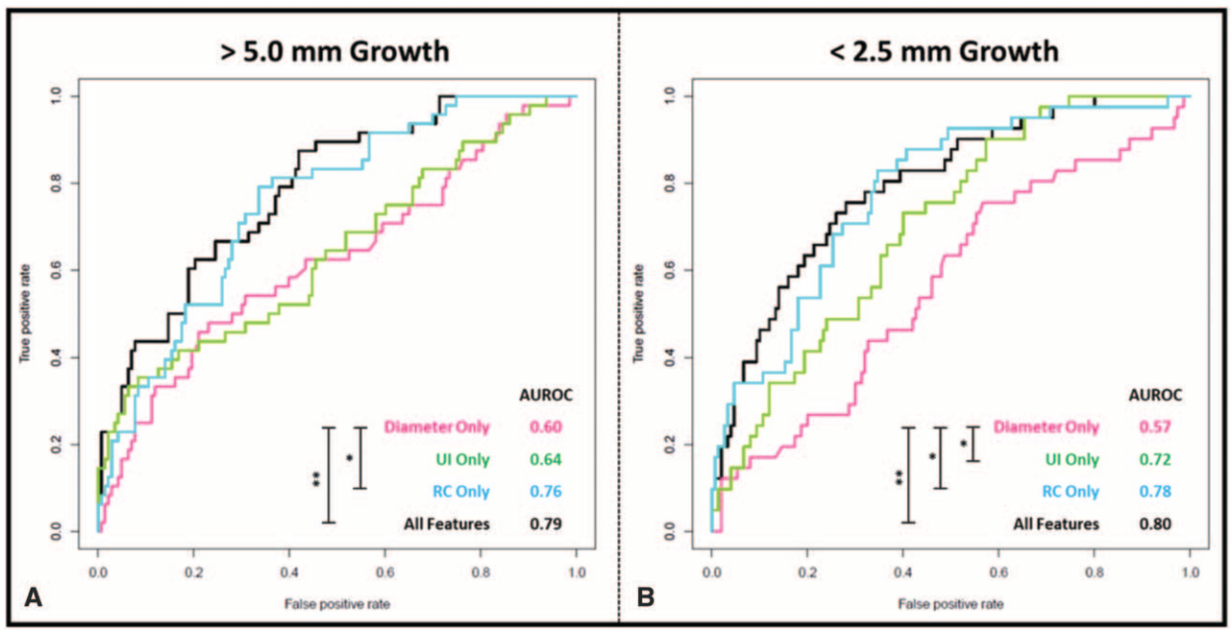


FIGURE 4. Receiver operator characteristic (ROC) curves for each combination of features alongside the area under curve (AUROC) to assess the performance of the multinomial regression model in discerning AAA growth phenotype (A, >5.0 mm and B, <2.5 mm Growth). ROC curves were generated from the evaluation of the test cohort (n = 92), following model training. Significance testing was performed to compare individual ROC curves and significant differences were noted with *or **indicating $P < 0.05$ and $P < 0.01$, respectively. AAA indicates abdominal aortic aneurysm.

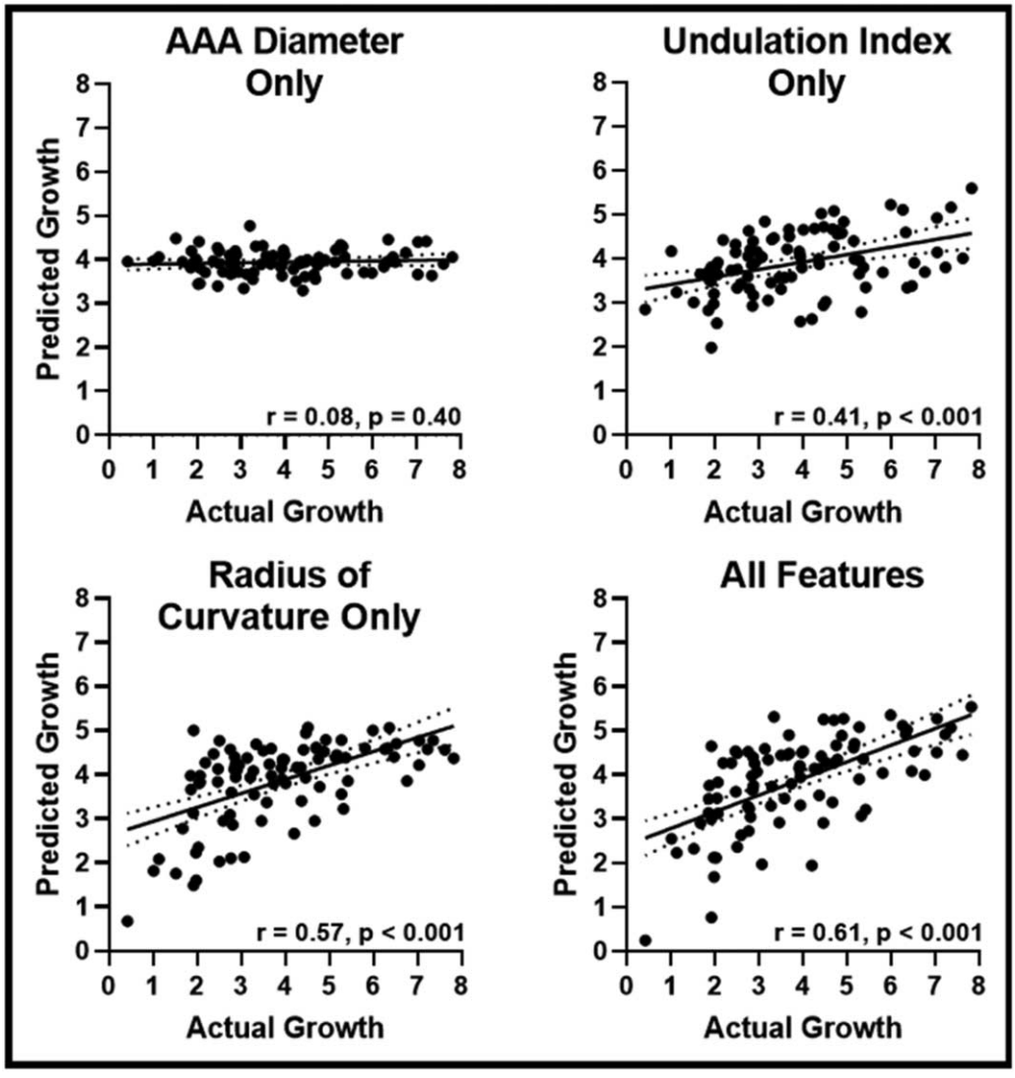


FIGURE 5. Outputs of linear regression models trained to predict annual AAA growth. Four Linear regression models were trained using 10-fold cross-validation to predict annual AAA growth. Models were evaluated using the testing cohort (n = 92) and are reported against actual AAA growth values. Statistical significance is assessed for each correlation and is indicated on each graph. Dotted lines indicate the 95% confidence interval for the regressions. AAA indicates abdominal aortic aneurysm.

There is emerging literature on the role of best medical therapy (BMT) in reducing overall mortality risk in patients with a AAA, exemplified by the VIVA study long term data.[31] Our method of AAA growth prediction can complement the delivery of BMT in the surveillance pathway. Those predicted to have fast growth should warrant further targeted intensive BMT regime to change their risk

TABLE 2. Evaluation of Linear Regression Models to Predict AAA Growth

	Train (n = 100)			Test (n = 92)		
Features	**RMSE**	**Correlation**	***P***	**RMSE**	**Correlation**	***P***
AAA Diam Only	1.52 ± 1.70 mm	0.25	0.03	1.72 ± 1.85 mm	0.08	0.40
Undulation Index Only	1.44 ± 1.66 mm	0.38	<0.001	1.53 ± 1.68 mm	0.41	<0.001
Radius of Curvature Only	1.32 ± 1.50 mm	0.52	<0.001	1.38 ± 1.68 mm	0.57	<0.001
AAA Size + UI + Min. RC	1.26 ± 1.50 mm	0.59	<0.001	1.32 ± 1.44 mm	0.61	<0.001

Four linear regression models were trained using 10-fold cross-validation to predict annual AAA growth. Models were evaluated using the validation folds of the training cohort (n = 100) and testing cohort (n = 92). Model Predictions of AAA growth were compared against observed values using root-mean-square error and correlation coefficient analysis. Significance of correlations are reported.

AAA indicates abdominal aortic aneurysm; RC, radius of curvature; UI, undulation index.

profile. Of note, in the prospectively recruited arm of this study, only a small fraction of patients included were females. We therefore could not rule out an association between these geometric features to the reported demographic features in female patients.

Although AAA surveillance is typically performed using ultrasound scans, many of these patients undergo CT scans for other clinical reasons during the course of their AAA surveillance. These CT scans can be utilized for the added purpose for the prediction AAA growth. (Of note, AAA surveillance is indeed performed using serial CT scans in countries such as Japan.) With the refinement of CT imaging technology and reduction of radiation dose per scan and portable tomographic CTs, it is not implausible for CT scans to replace ultrasound scans as the choice of AAA surveillance in the future. This will facilitate the development and validation of CT image derived prediction algorithm.

CONCLUSIONS

We present an AAA prediction model which utilizes geometric features that can be readily extracted from clinically acquired CT scans. This method can be applied to historic scans acquired during the routine clinical pathways of each AAA patient.

ACKNOWLEDGMENT

The authors acknowledge the support from the clinical team of the Oxford Regional Vascular Services: Chris Darby, Dominic PJ Howard, Patrick Lintott, Tim Magee, Andrew Northeast, Jeremy Perkins, Ed Sideso, and Emma Wilton. The content of this manuscript is subject to a UK patent filing (P304604GB).

REFERENCES

1. Wanhainen A, Verzini F, Van Herzeele I, et al. Editor's choice - European Society for Vascular Surgery (ESVS) 2019 clinical practice guidelines on the management of abdominal aorto-iliac artery aneurysms. *Eur J Vasc Endovasc Surg*. 2019;57:8–93.
2. Chaikof EL, Dalman RL, Eskandari MK, et al. The Society for Vascular Surgery practice guidelines on the care of patients with an abdominal aortic aneurysm. *J Vasc Surg*. 2018;67:2–77.e2.
3. Lee R, Jones A, Cassimjee I, et al. International opinion on priorities in research for small abdominal aortic aneurysms and the potential path for research to impact clinical management. *Int J Cardiol*. 2017;245:253–255.
4. Lee R, Charles PD, Lapolla P, et al. Integrated physiological and biochemical assessments for the prediction of growth of abdominal aortic aneurysms in humans. *Ann Surg*. 2018.
5. Regent L, Cassimee I, Huang H, et al. Integrated plasma and tissue proteomics reveals attractin release by intraluminal thrombus of abdominal aortic aneurysms and improves aneurysm growth prediction in humans. *XXXX*. 2020.
6. Lee R, Jarchi D, Perera R, et al. Applied machine learning for the prediction of growth of abdominal aortic aneurysm in humans. *EJVES Short Rep*. 2018;39:24–28.
7. Eurostatistics. Healthcare Resource Statistics - Technical Resources and Medical Technology - Statistics Explained 2020. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare_resource_statistics_-_technical_resources_and_medical_technology#Use_of_medical_technology.
8. Research i. Over 75 Million CT Scans Are Performed Each Year and Growing Despite Radiation Concerns. 2018.
9. Shum J, Martufi G, Di Martino E, et al. Quantitative assessment of abdominal aortic aneurysm geometry. *Ann Biomed Eng*. 2011;39:277–286.
10. Vorp DA, Raghavan ML, Webster MW. Mechanical wall stress in abdominal aortic aneurysm: influence of diameter and asymmetry. *J Vasc Surg*. 1998;27:632–639.
11. Jalalahmadi G, Helguera M, Mix DS, et al. (Peak) Wall Stress as an indicator of abdominal aortic aneurysm severity. *Proc IEEE West N Y Image Signal Process Workshop*. 2018.
12. Azar D, Ohadi D, Rachev A, et al. Mechanical and geometrical determinants of wall stress in abdominal aortic aneurysms: a computational study. *PLoS One*. 2018;13:e0192032.
13. Dhar S, Tremmel M, Mocco J, et al. Morphology parameters for intracranial aneurysm rupture risk assessment. *Neurosurgery*. 2008;63:185–196. discussion 196–197.
14. Zanaty M, Chalouhi N, Tjoumakaris SI, et al. Aneurysm geometry in predicting the risk of rupture. A review of the literature. *Neurol Res*. 2014;36:308–313.
15. Sharzehee M, Khalafvand SS, Han HC. Fluid-structure interaction modeling of aneurysmal arteries under steady-state and pulsatile blood flow: a stability analysis. *Comput Methods Biomech Biomed Engin*. 2018;21:219–231.
16. Lauric A, Hippelheuser J, Safain MG, et al. Curvature effect on hemodynamic conditions at the inner bend of the carotid siphon and its relation to aneurysm formation. *J Biomech*. 2014;47:3018–3027.
17. Lee R, Bellamkonda K, Jones A, et al. Flow mediated dilatation and progression of abdominal aortic aneurysms. *Eur J Vasc Endovasc Surg*. 2017;53:820–829.
18. Chandrashekar A, Handa A, Shivakumar N, et al. A deep learning approach to automate high-resolution blood vessel reconstruction on computerized tomography images with or without the use of contrast agent. *XXXX*. 2020.
19. Yushkevich PA, Piven J, Hazlett HC, et al. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage*. 2006;31:1116–1128.
20. Kollmannsberger P, Kerschnitzki M, Repp F, et al. The small world of osteocytes: connectomics of the lacuno-canalicular network in bone - IOPscience. *XXXX*. 2017.
21. Kontopodis N, Lioudaki S, Pantidis D, et al. Advances in determining abdominal aortic aneurysm size and growth. *World J Radiol*. 2016;8:148–158.
22. Manning BJ, Kristmundsson T, Sonesson B, et al. Abdominal aortic aneurysm diameter: a comparison of ultrasound measurements with those from standard and three-dimensional computed tomography reconstruction. *J Vasc Surg*. 2009;50:263–268.
23. NHS England. Abdominal Aortic Aneurysm Screening - NHS Choices 2017-06-02, 2017. Available at: http://www.nhs.uk/conditions/abdominal-aortic-aneurysm-screening/Pages/Introduction.aspx.
24. Doyle BJ, Bappoo N, Syed MBJ, et al. Biomechanical assessment predicts aneurysm related events in patients with abdominal aortic aneurysm. *Eur J Vasc Endovasc Surg*. 2020;60:365–373.
25. Martufi G, Lindquist Liljeqvist M, Sakalihasan N, et al. Local diameter, wall stress, and thrombus thickness influence the local growth of abdominal aortic aneurysms. *J Endovasc Ther*. 2016;23:957–966.
26. Miller K, Mufty H, Catlin A, et al. Is there a relationship between stress in walls of abdominal aortic aneurysm and symptoms? *J Surg Res*. 2020;252:37–46.
27. Malkawi AH, Hinchliffe RJ, Xu Y, et al. Patient-specific biomechanical profiling in abdominal aortic aneurysm development and rupture. *J Vasc Surg*. 2010;52:480–488.
28. Doyle BJ, Miller K, Newby DE, et al. Commentary: computational biomechanics-based rupture prediction of abdominal aortic aneurysms. *J Endovasc Ther*. 2016;23:121–124.
29. Leemans EL, Willems TP, Slump CH, et al. Additional value of biomechanical indices based on CTa for rupture risk assessment of abdominal aortic aneurysms. *PLoS One*. 2018;13:e0202672.
30. Yuan J, Huang C, Lai N, et al. Hemodynamic and morphological analysis of mirror aneurysms prior to rupture. *Neuropsychiatr Dis Treat*. 2020;16:1339–1347.
31. Lindhold J, Soggard R. Population screening and intervention for vascular disease in Danish men (VIVA): a randomised controlled trial. *Lancet (London England)*. 2017;390.

AJR Online First Accepted Manuscript

Beyond the *AJR*: "Prediction of Abdominal Aortic Aneurysm Growth Using Geometric Assessment of Computerised Tomography Images Acquired During the Aneurysm Surveillance Period"

Sahar Soleimani, MD, PhD, Linda Chu, MD

https://doi.org/10.2214/AJR.21.25789
Accepted: March 2, 2021
Article Type: Beyond the *AJR*

The complete title page, as provided by the authors, is available at the end of this article.

Recommended citation:
Soleimani S, Chu L. Beyond the *AJR*: "Prediction of Abdominal Aortic Aneurysm Growth Using Geometric Assessment of Computerised Tomography Images Acquired During the Aneurysm Surveillance Period". *AJR* 2021 Mar 17 [published online]. Accepted manuscript. doi:10.2214/AJR.21.25789

Summary of the Investigation

Screening for abdominal aortic aneurysm (AAA) leading to prophylactic interventions and initiation of medical therapy has been shown to reduce all-cause mortality. While the overall gain far exceeds the harm of overdiagnosis, the surveillance frequency and optimal timing for elective surgery in asymptomatic patients remain elusive targets. This goal requires accurate prediction of aneurysm growth and rupture. Current management guidelines are based on AAA diameter, but the pathophysiology of AAA cannot be fully explained by one parameter.

Many patients undergo CT imaging during their course of surveillance. CT imaging can be used to develop geometric and fluid dynamic models with the goal to more accurately predict AAA growth rate. In a recent article by Chandrashekar et al [1], the authors propose to use AAA diameter, undulation index (UI), and radius of curvature (RC) from CT-based volumetric models to predict AAA growth in a total of 192 patients who underwent elective repair of AAA. These geometric features were previously studied in prediction of cerebral aneurysm rupture. In the current study, multinomial logistic and multiple linear regression models were trained, validated, and compared to predict AAA growth in a 2-year median pre-operative interval (average baseline diameter 53.6 mm and 54.7 mm in the training and test cohorts). AAA diameter, UI, and RC demonstrated significant correlation with AAA growth as individual parameters. Three-variant models achieved area under ROC curves of 0.80 and 0.79 for prediction of slow (<2.5 mm/year) and fast (>5 mm/year) growth, respectively. The prediction for growth rate was within 2mm in 87% of cases.

Critical Analysis

The proposed multivariable model outperforms the AAA diameter alone in predicting slow and fast growth phenotypes. The average baseline AAA diameters were not different between the two phenotypes. However, care must be taken in extrapolating the results of this study to the general population under surveillance or with incidentally noted AAA, which tend to be of smaller size and may initially behave differently than the larger counterparts. Additionally, AUCs do not automatically translate into classification accuracy. Larger study is required to evaluate the role of geometric features in prediction of AAA growth and risk of rupture over a longer period of time.

Nonetheless, the findings of this study are a promising step forward towards better characterization of AAA. With the advent and growing availability of automated segmentation and analysis tools in CT imaging literature, these advanced geometric parameters can be incorporated into clinical practice without significant burden on the radiologist. A growing body of literature has shown the utility of volumetric and computational fluid dynamics parameters including aneurysm volume, maximal wall pressure, wall shear stress, and presence of intramural thrombus in evaluation of AAA growth [2, 3]. Moreover, feature analysis and radiomics using standard-of-care CT images have become powerful prognostic and predictive tools, primarily in oncology [4], with new promising data in cardiovascular disease [5]. These multivariable models can be complementary to the presented model for more comprehensive characterization of AAA growth and risk of rupture.

The current study also demonstrates independence of the presented geometric features from patient demographics including smoking history, hypertension, and hyperlipidemia. This independence needs to be further evaluated given well established evidence on the association of these comorbidities with presence of AAA.

Given complexity of aneurysm biology and invasive nature of AAA repair, a multifactorial evaluation of patient risk is required to produce the most precise prediction model for the AAA risk of rupture and improved outcome for patients.

Takeaway Point

Automated multivariable geometric features from standard-of-care CT images can predict AAA growth, particularly in aneurysms larger than 5 cm.

References

1. Chandrashekar A, Handa A, Lapolla P, et al. Prediction of Abdominal Aortic Aneurysm Growth Using Geometric Assessment of Computerised Tomography Images Acquired During the Aneurysm Surveillance Period. *Annals of surgery* 2020;
2. Meyrignac O, Bal L, Zadro C, et al. Combining Volumetric and Wall Shear Stress Analysis from CT to Assess Risk of Abdominal Aortic Aneurysm Progression. *Radiology* 2020; 295:722-729
3. Zhu C, Leach JR, Wang Y, Gasper W, Saloner D, Hope MD. Intraluminal Thrombus Predicts Rapid Growth of Abdominal Aortic Aneurysms. *Radiology* 2020; 294:707-713

4. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nature reviews Clinical oncology* 2017; 14:749-762

5. Kolossváry M, Gerstenblith G, Bluemke DA, et al. Contribution of Risk Factors to the Development of Coronary Atherosclerosis as Confirmed via Coronary CT Angiography: A Longitudinal Radiomics-based Study. *Radiology* 2021:203179

AJR

Title: *Beyond the AJR: "Prediction of Abdominal Aortic Aneurysm Growth Using Geometric Assessment of Computerised Tomography Images Acquired During the Aneurysm Surveillance Period"*

Authors: Sahar Soleimani MD PhD, and Linda C. Chu MD

Affiliations:

The Russell H. Morgan Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, 600 North Wolfe Street, Baltimore, MD, 21287

Corresponding Author: Linda C. Chu, MD, Hal B168, Johns Hopkins Hospital, 600 North Wolfe Street, Baltimore, MD, 21287; phone: 443-831-4342; email: lindachu@jhmi.edu

Author: Sahar Soleimani, MD PhD, Johns Hopkins Hospital, 600 North Wolfe Street, Baltimore, MD, 21287; phone: 410-955-6785; email: sahars@jhmi.edu

Each author substantially contributed to the conception of the work, writing and revision of the manuscript. Each author has approved the final version of the manuscript.

Conflict of interest: None.