

Advanced Automated PET Image Segmentation in Radiation Therapy

Submitted for the degree of Doctor of Philosophy at Cardiff
University

Craig Parkinson

July 10, 2019



Declarations

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed..... (candidate) Date.....

Statement 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed..... (candidate) Date.....

Statement 2

This thesis is the result of my own independent work/investigation, except where otherwise stated. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed..... (candidate) Date.....

Statement 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed..... (candidate) Date.....

Statement 4

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.

Signed..... (candidate) Date.....

Word count: 21,000 words.

In loving memory of Archibald and Janice Chown, Veronica Parkinson and
Paul Denyer (Ted).

Acknowledgements

I would like to give my sincerest thanks to my supervisors Emiliano Spezi, John Staffurth and Christopher Marshall. Thank you for continuously pushing me, challenging me and for all your time reviewing and discussing work. Thank you for being my mentors. I would also like to thank Cardiff School of Engineering, Velindre Cancer Centre and the Wales Research and Diagnostic PET Imaging Centre for supporting this project. Kieran, thank you for your clinical input. Phil, thank you for your technical support. To my parents David and Jackie Parkinson I am grateful for your unwavering support and belief. Thank you to my sister Gemma, brother-in-law Matthew, nephew Elwood Baker-Irons, my fiancé Sarah Lewis and her family. Sarah, thank you for putting up with all of the grumpiness. To both families, I have appreciated all of the coffee and cake; I am looking forward to many more years of it. Thank you all for providing moments of fun and sanity over the past few years, from weddings and births to literally running and flying over mountains.

Abstract

Manual segmentation of the metabolic tumour volume (MTV) in positron emission tomography (PET) imaging is subject to intra and inter-observer variability. Many PET based automatic segmentation algorithms (PET-AS) have been proposed as solutions to this problem with machine-learned techniques showing promise for accurate MTV segmentation. However, no consensus has been reached on the optimal method for radiotherapy (RT) treatment planning, with the current American Association for Physicists in Medicine Task Group 211 and the International Atomic Energy Association advisory committees recommending that not one single PET-AS can be recommended for target volume delineation. This project, therefore, aimed to improve the MTV segmentation of a machine-learned PET-AS methodology called ATLAAS which has been proposed for standardised MTV segmentation by Berthon *et al* in *Radiother Onc* (2016).

Berthon *et al* additionally validated the ATLAAS algorithm on diagnostic PET imaging in *Radiother Onc* (2017). However, it has not been validated externally or for the role of MTV segmentation during treatment. Intra-treatment segmentation is challenging due to reduced metabolic uptake, tu-

mour to background ratio and reduced metabolic volume. Therefore, in this body of work, the performance of ATLAAS and 15¹ other PET-AS chosen from the literature, were evaluated for suitable MTV segmentation in PET imaging acquired after one cycle of neoadjuvant chemotherapy. This research resulted in the development of a new training dataset and demonstrated that ATLAAS can be used as a basis for adaptive radiotherapy and trained on imaging datasets outside of the original training cohort. However, this research still demonstrated that the performance of ATLAAS could be improved. Therefore, this led to an investigation into the inclusion of additional tumour characteristics in the development of the ATLAAS training model, in order to reduce the impact PET image resolution has on MTV segmentation. In this research, derived MTVs were compared to “ground truth” volumes derived from CT imaging. The results presented in this body of work, showed that interpolating PET imaging to the resolution of the CT image improved the performance of PET-AS segmentation and improved ATLAAS MTV segmentation by 19% and inclusion of one of the tumour features compactness one, compactness two or sphericity in the ATLAAS training model improved MTV segmentation by an additional 3%.

As part of this body of work, the requirement for a standardised PET-AS method was demonstrated by developing prognostic models, using standardised imaging and tumour features, from the MTV derived by 9 PET-AS demonstrated by Berthon *et al* in Phys. Med. Biol (2017) to be promising for accurate MTV segmentation. This showed how segmentation of the MTV

¹20-80% Threshold in increments of 10%, Adaptive Thresholding, Region Growing, K-means Clustering with 2 and 3 clusters, Gaussian Fuzzy C-means with 3 and 4 clusters and Fuzzy-C means with 2 clusters

has a subsequent effect on patient risk stratification with patients changing risk stratification quartiles dependent upon the PET-AS used to derive the MTV.

Contents

1	Radiotherapy of Cancer	1
1.1	Thesis Structure	1
1.2	A brief introduction to PET imaging and Segmentation	2
1.2.1	PET imaging	2
1.2.2	Segmentation	3
1.3	Thesis aims	4
1.4	TITAN project novelty & dissemination	6
1.5	Management of Cancer	7
1.5.1	Head and Neck Cancer	8
1.5.2	Oesophageal Cancer	8
1.5.3	Prognostic models	9
1.5.4	Tumour Node and Metastases Staging	10
1.5.5	Radiomic features	11
1.6	Role of external beam radiation therapy in the treatment of Cancer	12
1.6.1	History of radiotherapy	12
1.6.2	Radiotherapy techniques	13
1.7	Challenges in Radiotherapy Treatment and Delivery	14

1.8	Nuclear Medicine Imaging in Radiotherapy Planning	15
1.8.1	PET Imaging physics	18
1.8.2	Attenuation correction and Hybrid PET/CT Scanners	21
1.8.3	PET tracers	22
1.8.4	Sinogram	23
1.8.5	PET Image Reconstruction	23
1.8.6	DICOM Imaging format	25
1.9	Chapter 1 Overview	26
2	Segmentation of PET	33
2.1	Classification of PET-AS methods	34
2.2	PET-AS method implementations	36
2.2.1	Fixed and adaptive threshold algorithms	36
2.2.2	Gradient segmentation	38
2.2.3	Region growing	40
2.2.4	Statistical	40
2.2.5	Consensus Techniques	41
2.2.6	Machine learning	41
2.3	PET-AS Comparison	44
2.4	ATLAAS	47
2.4.1	Training model development	49
2.4.2	Application of ATLAAS training model to FDG PET imaging	50
3	Impact of metabolic tumour volume segmentation on pa- tient overall survival	56

3.1	Materials and methods	58
3.1.1	Patient Cohort	58
3.1.2	PET/CT protocol	58
3.1.3	Treatment Protocols	60
3.1.4	Hardware and software	61
3.1.5	Segmentation algorithms	61
3.1.6	Clinical Data & image analysis	68
3.2	Results	72
3.2.1	Development of Prognostic Models	73
3.2.2	Prognostic Score Calculation	74
3.3	Discussion	76
3.4	Conclusion	79
4	Automated segmentation of the metabolic tumour volume in low tumour to background ratio PET	88
4.1	Materials and Methods	89
4.1.1	Clinical Data	89
4.1.2	¹⁸ F-FDG PET/CT imaging	91
4.1.3	Manual MTV delineation after 1 cycle of chemotherapy	95
4.1.4	Development of ATLAAS _{ICT}	95
4.1.5	PET-AS MTV delineation after 1 cycle of ICT	101
4.1.6	Statistical analysis	101
4.2	Results	105
4.2.1	Manual MTV delineation after 1 cycle of chemotherapy	105

4.2.2	Comparison between PET-AS MTV and clinician derived MTV	106
4.3	Discussion	110
4.4	Conclusion	113
5	Impact of morphological features in the development of machine learned automated segmentation training models	115
5.1	Materials and Methods	116
5.1.1	Hardware and software	116
5.1.2	Training Dataset	117
5.1.3	Validation Dataset	118
5.1.4	Morphological features	121
5.1.5	ATLAAS training model development	127
5.1.6	Interpolation of PET imaging	129
5.1.7	Segmentation algorithms	135
5.1.8	Statistical Analysis	136
5.1.9	Results	137
5.2	Discussion	149
5.3	Conclusion	151
6	Discussion & Conclusion	152
6.1	Discussion	152
6.2	Conclusion	159
	Bibliography	161
	Appendices	192

A Publications	193
A.1 Journal Papers	193
A.2 Presentations	194
A.3 E-Posters and Posters	195
B PET-AS MTV delineation in low tumour to background ratio PET	197
B.1 Correlation of PET-AS methods DSC and Clinician derived radiomic features	197
C Investigation into the inclusion of morphological data in machine learned training models data	224
C.1 Training models DSC	224

List of Figures

1.1	Example PET image and contour	4
1.2	CT Scan of the Liver	27
1.3	ICRU 50 contours	27
1.4	Planar Scinitgraphy	28
1.5	Lung PET/CT	29
1.6	SPECT imaging	29
1.7	Positron emission	30
1.8	LOR and Sinogram	30
1.9	Denoised PET scan	31
1.10	DICOM standard	32
2.1	Watershed segmentation	53
2.2	Active Contouring	54
2.3	ATLAAS training	55
3.1	GE 690 PET/CT scanner	59
3.2	Co-registered PET/CT scan	60
3.3	Software package GUI	63
3.4	MTV contours	64
3.5	Patient risk stratification	83

3.6	Excluded PET-AS risk stratification	84
4.1	FiGaRo clinical pathway	92
4.2	RT immobilisation shell	93
4.3	Low dose and contrast CT	93
4.4	GE 710 PET/CT scanner	94
4.5	Pre and Post-ICT PET imaging	97
4.6	PETSTEP process	99
4.7	MTV and TBR for FiGaRo	102
4.8	MTV and NI for FiGaRo	103
4.9	TBR and NI for FiGaRo	104
4.10	Manually defined MTV	107
4.11	DSC of PET-AS delineated MTV	109
4.12	Percentage increase in MTV	110
4.13	ATLAAS derived MTV	111
5.1	Training contours	117
5.2	Training MTV and NI	119
5.3	Training MTV and TBR	119
5.4	Training NI and TBR	120
5.5	Validation ground truth contours	120
5.6	CT GT contours	130
5.7	Linear interpolation	133
5.8	DSC of ATLAAS	140
5.9	Correlation of compactness one and sphericity	141
5.10	Correlation of compactness one and compactness two	141

5.11 Correlation of compactness two and sphericity	142
5.12 Cases with greater than 200% difference in MTV	142
5.13 Cases with less than 200% difference in MTV	143
5.14 DSC of PET-AS	144
5.15 ATLAAS derived MTV	145
5.16 PET-AS derived MTV	146
5.17 PET-AS derived MTV	147
5.18 KM2 derived MTV	148

List of Tables

1.1	Stage groups based upon TNM Staging	11
2.1	Description of PET-AS methods	34
3.1	Summary of quantitative imaging features	80
3.2	Baseline characteristics of patient cohort	81
3.3	Final output of prognostic models	82
3.4	Prognostic model equations	82
3.5	Number of patients in risk stratification	84
3.6	Patients changing risk stratification	85
3.7	Final Output of Prognostic Models	85
3.8	Prognostic model equations	86
3.9	Summary statistics	86
3.10	Number of patients in each risk stratification	87
3.11	Patients who change risk stratification	87
4.1	Descriptive statistics of PET	97
4.2	Descriptive statistics of PET-AS MTV	114
5.1	Descriptive statistics for ATLAAS MTV	140

List of Equations

1.1	Sensitivity	17
1.2	Specificity	17
1.3	Compton Scattering photon energy	19
1.4	Recoiling electron energy	20
2.1	Thresholding	37
2.2	Standardised Uptake Value	37
2.3	Dice Similarity Coefficient	48
3.1	Adaptive Thresholding	65
3.2	Fuzzy Clustering	67
3.3	Gaussian Clustering	68
5.1	Tetrahedron volume	121
5.2	IBSI tumour volume	121
5.3	IBSI surface area	122
5.4	IBSI surface area	122
5.5	IBSI surface to volume ratio	122
5.6	IBSI compactness one	123
5.7	IBSI compactness two	123
5.8	IBSI spherical disproportion	123
5.9	IBSI sphericity	123

5.10	IBSI asphericity	124
5.11	IBSI centre of mass shift	124
5.12	IBSI Tumour centre of mass	124
5.13	IBSI Intensity centre of mass	125
5.14	IBSI Maximum 3D Diameter	125
5.15	IBSI Major Axis Length	125
5.16	IBSI Minor Axis Length	126
5.17	IBSI Least Axis Length	126
5.18	IBSI Elongation	126
5.19	IBSI Flatness	126
5.20	IBSI volume density	127
5.21	IBSI area density	127
5.22	Linear Interpolation	131
5.23	Linear Interpolation example part 1	131
5.24	Linear Interpolation example part 2	132
5.25	Linear Interpolation example part 3	132
5.26	Linear Interpolation example part 4	132
5.27	Cubic interpolation Part A	134
5.28	Cubic interpolation Part B	134
5.29	Cubic interpolation Part C	134
5.30	Cubic interpolation Part D	134

Chapter 1

Radiotherapy of Cancer

This chapter outlines the aim and structure of this thesis whilst providing background knowledge into the management of cancer and the challenges which exist in the delivery of Radiotherapy (RT) treatment. This chapter also outlines the role of nuclear medicine imaging in RT planning. In the following section the structure of this thesis is described.

1.1 Thesis Structure

In this thesis, Chapters 1 and 2 provide literature reviews into the RT of cancer and the segmentation algorithms proposed for use in RT planning. Chapter 1 described the challenges which exist when delivering RT treatment. However, Chapter 2 describes the numerous algorithms which have been proposed to solve these challenges. In comparison to Chapters 1 and 2, Chapters 3, 4 and 5 contain experimental work conducted throughout this

project. Chapter 3 evaluates the impact that automated tumour segmentation has on the development of prognostic models, Chapter 4 evaluates the performance of 16 automated PET based automatic segmentation (PET-AS) methods in nuclear medicine imaging acquired during treatment and Chapter 5 investigates improving the performance of a machine-learned segmentation methodology by using morphological features in the model development. The final Chapter, Chapter 6, discusses the findings of this project and places them in context to the current literature and state of the art. Chapter 6 also describes the potential future work resulting from this project. In the following section Positron Emission Tomography (PET) imaging and segmentation are introduced.

1.2 A brief introduction to PET imaging and Segmentation

1.2.1 PET imaging

PET is a functional imaging modality which can be used for the imaging of tumours in the body. Patients are injected with a solution labelled with a radioactive source before being placed on a PET scanner. The information obtained from the PET scanner is then reconstructed into an image which is human readable. Figure 1.1a image shows a typical PET image retrieved after patient scanning. The PET image is of the Head and Neck (H&N) of a patient with a tumour of the H&N and the extent of the

disease, due to the increased metabolic rate of the tumour, is visible to the human eye. PET imaging is described in more detail in Section 1.8.1

1.2.2 Segmentation

Segmentation is the process of dividing an image into different regions. For example, dividing an image into its background and foreground. In RT planning PET imaging is segmented in in order to define the tumour as a RT target. Segmentation of the tumour requires a human user to visualise all of the images obtained from the PET scanner. When they detect a tumour on the image, they manually draw a line (contour) around the detectable extent of the disease. Figure 1.1b shows a PET image which has been manually contoured based on visual inspection of a PET scan, with the contour displayed in white around the tumour. However, segmentation of the tumour is limited by inter and intra-observer variability. Inter-observer variability results in two clinicians having differences in the contours they each draw for the same image. Additionally, intra-observer variability results in a singular clinician drawing different contours when the same image is visualised at different time points. Many PET-AS methods have been proposed as solutions to this problem, however, the application of PET-AS methods in RT planning is limited by the current lack of inter-comparison and validation of PET-AS methods [1]. Standardised PET-AS methodologies have been proposed, however they have not been demonstrated or validated externally to the centre that developed the technique. Further, the limited resolution of ^{18}F -Fluorodeoxyglucose (^{18}F -FDG) PET imaging means that

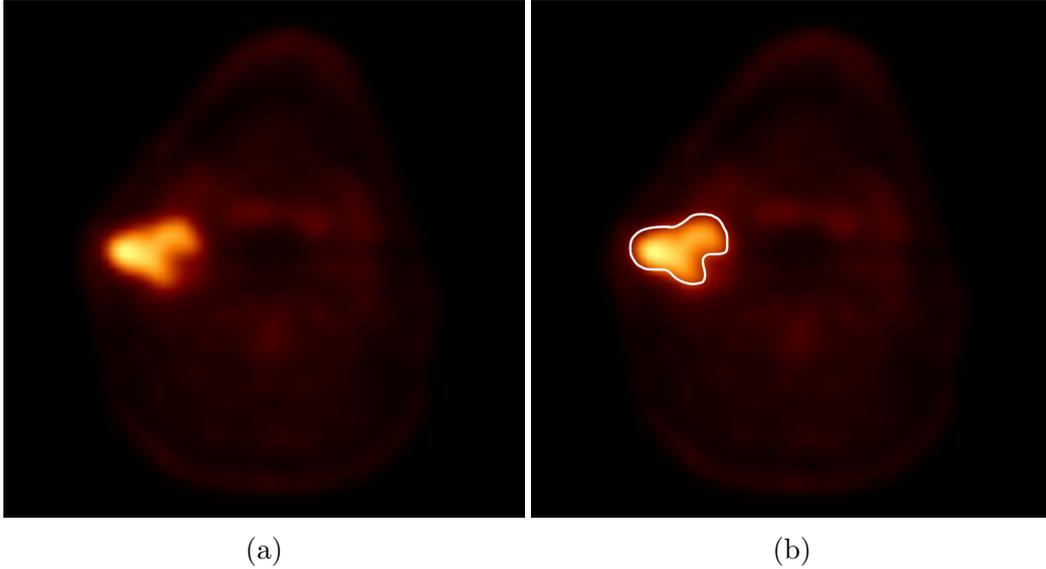


Figure 1.1: (a) An example H&N PET image reconstructed from information obtained from a PET scanner. (b) An example contour drawn in white around the detectable extent of the disease.

the performance of PET-AS methods is potentially limited in comparison to delineation on conventional imaging modalities. Automatic segmentation of PET imaging is described in more detail in Section 2

1.3 Thesis aims

The hypothesis of this project is that the performance of the machine-learned an automatic decision tree-based learning algorithm for advanced image segmentation (ATLAAS) segmentation methodology, for Metabolic Tumour Volume (MTV) delineation, will be higher in comparison to advanced PET-AS methods that have been proposed for accurate MTV delineation.

A project entitled TITAN, was funded by Cardiff School of Engineering and Velindre Cancer Centre. In Greek mythology Atlas, is a Titan that stood on the edge of the Earth holding up the sky. Therefore, this project was entitled TITAN as it aimed to improve the ATLAAS segmentation methodology. The project aimed at addressing the issues briefly introduced in Section 1.2 and was carried out as a collaboration between three institutions:

- The Wales research and diagnostic PET imaging centre (PETIC), which opened in 2010, offers some of the most advanced imaging equipment in the UK, with a high resolution scanner providing high quality images for research and clinical purposes. PETIC is operated by Cardiff University in partnership with Cardiff and Vale University Health Board, and is located at the University Hospital of Wales in Cardiff.
- Velindre Cancer Centre (VCC), located in Cardiff, is one of the largest specialist centres for non-surgical cancer treatment in the UK, with over 5000 new patient referrals every year. It boasts high-end equipment, with electron linear accelerators (linacs) enabling advanced RT procedures, and strong links with the Wales Cancer Trials Unit and the Wales Cancer Bank for conducting world class research through oncology trials.
- Cardiff University, Cardiff School of Engineering conducts world leading research combined with strong links with industry, therefore making it one of the leading engineering schools in the UK.

Therefore, the aims and objectives of this thesis are:

- Improve the performance of the ATLAAS segmentation methodology.
- Externally validate the performance of the ATLAAS segmentation methodology.
- Investigate the effect of PET-AS methods on the development of prognostic models, therefore demonstrating the requirement for a standardised PET-AS methodology in planning RT.
- Investigate the role and robustness of the ATLAAS segmentation methodology for intra-treatment MTV delineation.
- Investigate the impact of morphological features on the performance of ATLAAS.

1.4 TITAN project novelty & dissemination

Throughout the TITAN project, novel research was disseminated throughout the research community. The disseminations from the project were 2 first name authored journal papers and 2 co-authored journal papers. Additionally, research was presented orally at 4 conferences and as posters at 6 conferences. The novelty of the research contained in this body of work is as follows:

- Demonstrated PET-AS MTV delineation influences the development of prognostic models.
- External validation of the performance of the ATLAAS segmentation methodology.

- Applied PET-AS methods to and compared PET-AS method performance in low tumour to background ratio (TBR) scenarios.
- Compared the accuracy of 16 PET-AS methods for intra-treatment MTV delineation.
- Demonstrated interpolation of PET imaging to a higher resolution, before MTV delineation, improves the accuracy of PET-AS methods.
- Improved the performance of the ATLAAS segmentation methodology by including morphological features as classifiers.

1.5 Management of Cancer

Cancer is one of the leading causes of mortality across the world. In 2012, across the world, around 14 million new diagnoses of cancer and 8.2 million cancer related deaths occurred. Europe alone experiences an estimated 24.4% (\approx 341 thousand) of the worlds cancer diagnoses annually and the patients diagnosed with cancer, in Europe, experienced a mortality rate of \approx 21.4% (\approx 175 thousand) [2]. In men, the most common diagnosed cancer sites include the prostate and lungs. In women however, the lungs and breasts are among the most common diagnosed cancer sites. The patient cohorts in this body of work were diagnosed with primary H&N and primary Oesophageal Cancer (OC).

1.5.1 Head and Neck Cancer

H&N cancer is the sixth most common diagnosed cancer type worldwide [3] with a 5 year survival rate of less than 50% [4]. In comparison to the worldwide rates of incidence, in the UK, H&N cancer accounts for 3% of all new cases experienced, with a rate of incidence that has increased by 30% since the early 1990s [5]. H&N cancer is the collective group of cancers consisting of the anatomical sites of the H&N, including the oropharynx, oral cavity, larynx and hypopharynx. In the curative setting, RT is a commonly used treatment for H&N cancer. There is however increasing interest in multimodality therapy including surgery and chemotherapy [6]. H&N cancers are predominately squamous cell carcinoma (SCC) with 60% of tumours being Human Papilloma Virus (HPV) positive [7] with Tumour, Node and Metastasis (TNM) staging predominately determining H&N cancer prognosis [8]. However, risk factors such as HPV status and tobacco usage have been shown to be of significant prognostic importance [9]. Further, predominant risk factors in the diagnosis of H&N cancer are alcohol and tobacco consumption [10].

1.5.2 Oesophageal Cancer

OC is the eighth most common [2] diagnosed cancer worldwide with an estimated 456,000 new incidences every year. Prognosis in OC is extremely poor with rates of mortality approximately matching the rates of incidence, demonstrated by OC having a 1— and 5—year overall survival (OS) rate of 44% and 15%, respectively [11]. Predominant risk factors in OC are alcohol

and tobacco consumption. A further potential risk factor is ingestion of caustic substances [2]. OC typically presents as SCC. However, other types occur including Adenocarcinoma, Neuroendocrine and small cell cancer.

1.5.3 Prognostic models

A variety of techniques are used in the management of cancer. Precision medicine, however, combined with the application of prognostic models aims to ensure each patient is managed with the most appropriate treatment, which may improve patient OS [12–15]. Prognostic models aim to characterise each patient’s risk and OS, when treated with a specific treatment, dependent upon their characteristics. This is known as a patients risk stratification. A patient with a low-risk stratification has a higher OS in comparison to patients with high-risk stratifications. Whereas, patients with intermediate risk-stratification have an OS in between high-risk and low-risk stratifications. By separating patients into risk stratifications, ineffective therapies can be avoided. Therefore, preventing patients being treated with unnecessary and aggressive therapies. The avoidance of these therapies have the potential to improve a patients quality of life. Prognostic models are developed from clinical information including TNM staging and radiomic features [16].

1.5.4 Tumour Node and Metastases Staging

A patient's disease progression and therefore prognosis is determined through the TNM staging of patients. TNM classification aims to separate patients into groups and throughout this thesis patients were staged using the TNM 7th edition [8]. Primary tumours (T) are categorised according to the following:

- T0: No evidence of primary tumour
- T1a: Tumour invades lamina propria or muscularis mucosae
- T1b: Tumour invades submucosa
- T2: Tumour invades muscularis propria
- T3: Tumour invades adventitia
- T4a: Tumour invades pleura
- T4b: Tumour invades adjacent structures

Lymph nodes (N) are categorised according to the following:

- N0: Node lymph node metastases
- N1: Metastases in 1-2 regional lymph nodes
- N2: Metastases in 3-6 regional lymph nodes
- N3: Metastases in 7 or more lymph nodes

Distant Metastases (M) are categorised according to the following:

- M0: No distant metastases

- M1: Distant metastases

Based upon TNM staging patients can be classified into stage groups related to prognosis. Lymph node and distant metastatic disease increases stage group as shown in Table 1.1.

Table 1.1: Stage groups based upon TNM Staging

Stage Group	T	N	M
Stage 0	T0	N0	M0
Stage IA	T1	N0	M0
Stage IB	T2	N0	M0
Stage IIA	T3	N0	M0
Stage IIB	T1, T2	N1	M0
Stage IIIA	T4a	N0	M0
	T3	N1	M0
	T1, T2	N2	M0
Stage IIIB	T3	N2	M0
Stage IIIC	T4a	N1, N2	M0
	T4b	Any N	M0
	Any T	N3	M0
Stage IV	Any T	Any N	M1

1.5.5 Radiomic features

Radiomic features in medical imaging are algorithms applied to an image in order to extract and quantify information not visible to the naked eye. For example, the skewness of the histogram. The extraction of radiomic features is hoped to improve the prediction of a patients OS and therefore patient outcome; however, studies have demonstrated that there are numerous challenges in the extraction of quantitative features [17]. The selection of appropriate and significant features for the development of prognostic models

is of considerable importance because the reproducibility and robustness of radiomic features is influenced by external parameters including the MTV delineation method, image reconstruction parameters and radiomic feature extraction pre-processing steps [18, 19].

1.6 Role of external beam radiation therapy in the treatment of Cancer

1.6.1 History of radiotherapy

Since the discovery of X-rays in 1895, X-rays have been applied in the clinical environment. As early as January 1896, X-rays were used to treat breast cancer and skin lesions [20]. A lack of understanding of the biological effects of radioactivity, however, led to poor cancer control rates and high rates of morbidity [21]. Technological advancements and improvements in the understanding and delivery of RT came from a need for improved disease control and quality of life. Investigations into the delivery of the total radiation dose, showed the application of external beam radiotherapy (EBRT) as a fractionated dosage reduced toxicity in H&N cancer compared to delivery of a single large dose [22]. Further, the development and installation of megavoltage linacs in the 1950s led to improvements in the efficacy of treatments in comparison to the original 200 kilovolt X-rays [21].

1.6.2 Radiotherapy techniques

As our understanding of how X-rays and ionised rays impacted tumour biology, improvements in EBRT, in more recent years, have come from focusing upon the conformality of the delivered dose as well as multi-modality treatment pathways. Improved understanding has also led to the development of multiple EBRT techniques, which are outlined below:

- 3D-Conformal Radiotherapy (3D-CRT) delivers a uniform radiation dose to the tumour, with a specified width, height and depth. However, in comparison to some other EBRT techniques, 3D-CRT encompasses a greater amount of healthy tissue. Increased dosage to the organs at risk (OAR) reduces the efficacy of the treatment and increases the likelihood of local-regional and local tumour recurrences.
- Intensity modulated radiotherapy (IMRT) delivers a greater radiation dose to the tumour and a reduced radiation dose to the surrounding healthy tissue in comparison to 3D-CRT. This improves toxicity levels and improves incidences of local and local-regional reoccurrences. IMRT relies upon the use of collimator leafs to shape the radiation beam and reduce toxicity to the surrounding tissue. However, IMRT is characterised by steep dose gradients which decrease the margin for error in RT planning [23].
- Image guided radiotherapy (IGRT) makes use of imaging at the pre-treatment and treatment delivery stage that can lead to improvements or verify the accuracy of radiotherapy. IGRT encompasses a wide range

of techniques from simple visual field alignment checks, through to more complex volumetric imaging that allows direct visualisation of the radiotherapy target volume and the surrounding anatomy [24].

1.7 Challenges in Radiotherapy Treatment and Delivery

Accurate targeting of the tumour requires careful planning of a RT beams' shape and position. RT planning is a time-consuming process approached with dedicated software. The software calculates the optimal beam arrangement based upon contour information determined by the planning clinician for the tumour and OAR. Contours are drawn using information from anatomical imaging, including Computed Tomography (CT) imaging. CT imaging measures the density of tissue in comparison to water using Hounsfield units. Figure 1.2 shows a CT scan obtained of the Liver demonstrating the different tissue densities in the human body. Reference Hounsfield units for tissue are -1000 for air, 0 for water and +700 for bone tissue. The Liver has Hounsfield values of $\approx 54-60$. Contouring techniques can sometimes be combined through consensus techniques and the registration of imaging. RT planning requires definition of multiple contours as shown in Figure 1.3 and defined in the International Commission on Radiation Units and Measurements (ICRU) report number 50 [25] and more subsequent reports [26–28]:

- Gross Tumour Volume (GTV), corresponding to all of the detectable disease

- Clinical Target Volume (CTV), an extension of the GTV with a margin accounting for microscopic disease extension
- Planning Target Volume (PTV), an extension of the CTV accounting for errors in patient positioning, dose delivery and contouring errors.

Current RT treatments allow for precise RT dose delivery to the target. GTV delineation, however, has been identified as the largest source of error in the delivery of accurate RT treatments [29] with results being difficult to reproduce due to user involvement [30]. Increasingly, PET imaging is being investigated for providing additional and complementary information that can be used to improve the performance of GTV delineation in RT planning [31–38].

1.8 Nuclear Medicine Imaging in Radiotherapy Planning

There are three different imaging modalities in nuclear medicine imaging. These are Planar Scintigraphy (PS), Single Photon Emission Computed Tomography (SPECT) imaging and PET imaging. Figures 1.4, 1.5 and 1.6 show example images obtained from each imaging modality. Figure 1.4 shows an image obtained from a patient scanned using PS. The patient shown in Figure 1.4 was injected with a radiotracer used for the detection and imaging of bone tumours. In PS, areas of increased radioactivity are more visible in the image in comparison to less radioactive areas. Therefore, in Figure

1.4 the skin is seen as a light colour, whereas the bones are the darker regions. Areas of increased radioactivity in comparison to the imaged tissue are therefore potentially cancerous due to a tumours increased metabolic activity. A patient scanned using a PET scanner is shown in Figure 1.5. In conjunction with PS, in PET imaging, areas of increased radioactivity are more visible in comparison to less radioactive areas. In comparison, Figure 1.6 shows an image obtained from a patient scanned using SPECT. The patient shown in Figure 1.6 was injected with a radiotracer into the vascular system in order to monitor blood flow (diffusion) in the brain. The bright (orange) regions of the image indicate increased blood flow compared to the surrounding area.

PS is the most simplistic imaging modality producing a single two dimensional (2D) image and is mostly used for whole body screening for tumours, whereas SPECT & PET imaging produce a series of continuous 2D images [43]. PET imaging, in comparison to SPECT, has between a 100 and 1000 times higher signal to noise ratio (SNR) as well as higher spatial resolution [44]. The higher SNR and spatial resolution of PET imaging offers increased diagnostic power [43] and in comparison to conventional anatomical imaging such as CT, PET imaging discriminates between healthy and tumour tissue with a higher sensitivity (1.1) and specificity (1.2) [34, 35, 45–50]. Figure 1.5 shows a CT image obtained from patient scanned for lung cancer, anatomical tissue including the liver is visible to the human eye. However, due to the homogeneous tissue density within organs, tumours within an organ are difficult to detect.

TP corresponds to the number of true positives, e.g. the number of people correctly identified with tumours. FN corresponds to the number of false negatives, e.g. the number of people with tumours incorrectly identified as being healthy. TN corresponds to the number of true negatives. e.g. the number of healthy people without tumours being identified as healthy. FP corresponds to the number of false positives. eg: the number of healthy people without tumours identified with tumours.

With the use of nuclear medicine imaging it is therefore possible that patients with cancer can be diagnosed earlier, as the extent of the disease is more visible at earlier stages of the disease progression. With earlier diagnoses, it is possible for patient OS to increase due to less aggressive treatment pathways being required [38]. The extent of disease detectable on ^{18}F -FDG PET imaging is known as the MTV.

$$\textit{sensitivity} = \frac{TP}{TP+FN} \quad (1.1)$$

$$\textit{specificity} = \frac{TN}{TN+FP} \quad (1.2)$$

1.8.1 PET Imaging physics

As early as the 1950s, the technology behind PET imaging was being researched [51]. The first applications of PET imaging were the modelling of blood flow and imaging of brain tumours [52, 53]. It was not until the early 2000s, however, that a complete PET/CT scanner was developed [54]. PET imaging requires the injection of ‘radiotracers’, which are analogues of biological molecules labelled with positron-emitting radioisotopes, into the patient. Once patients are injected they typically rest for a period of time, before being scanned, to allow for the maximal uptake of the tracer in the tumour before radioactive decay. After waiting for maximal uptake of the radioisotope labelled tracer in the patient, the patient is placed on the scanner bed.

During the scanning procedure, positrons are emitted from the tumour and travel approximately 2 mm before colliding with electrons. At the point of collision an annihilation event occurs from which two 511 keV gamma rays are produced [55] as shown in Figure 1.7 on page 30. The two gamma rays are detected by scintillation crystals made from bismuth germanate, lutetium oxyorthosilicate or gadolinium silicate [56]. Scintillation crystals are arranged into blocks around a detector ring and convert the energy from the gamma rays into light. As the gamma ray hits the scintillation crystal, the electrons are energised through Compton scatter or photoelectric absorption processes. As the electron travels further through the scintillation crystal, it loses more energy and excites more electrons. As the excited electrons decay into their non-excited state they give off light. Detector blocks are coupled to

photomultiplier tubes (PMT)s which convert the light from the scintillation crystal into a voltage signal [43]. PMTs and scintillation crystals may be interfaced in a couple of ways. A 2D array of crystals may be coupled to 4 PMTs or an array of PMTs may be coupled to a single cut planar crystal [57]. A PET/CT scanner consists of between 20-30 detector rings, each consisting of thousands of scintillation crystals. The spatial resolution of PET imaging is dependent upon a compromise between the number of scintillator crystals and PMTs. Increasing the number of crystals increases the SNR, however, the physical size of the PMT limits the spatial resolution [43].

Compton Scattering

Compton scattering is an interaction between a photon and a loosely bound outer-shell orbital electron of an atom. Where the energy of the incident photon greatly exceeds the binding energy of the electron to the atom, the interaction looks like a collision between a photon and a “free” electron [58].

The photon in Compton scattering is deflected through the scattering angle Θ and part of the photons energy is transferred to the “recoiling” electron. Therefore, the photon loses energy. The scattered photon energy is related to the angle Θ which the photon was deflected by (1.3).

$$E_{sc} = \frac{E_0}{1 + (\frac{E_0}{0.511})(1 - \cos \theta)} \quad (1.3)$$

E_0 is the energy, in MeV, of the incident photon and E_{sc} is the energy, in

MeV, of the scattered photon. The energy of the recoiling electron E_{re} is the difference between the incident and scattered photon energies (1.4).

$$E_{re} = E_0 - E_{sc} \quad (1.4)$$

The energy imparted to the recoiling electron ranges from ≈ 0 for a scattering angle θ of ≈ 0 degrees. The maximum amount of energy transferred to the recoiling electron occurs in backscattering events where the scattering angle θ is 180 degrees. During backscattering events the scattered photons also have a minimum resulting energy due to imparting the maximum amount of energy to the electron.

Photoelectric absorption

Photoelectric absorption occurs at the lower photon energies [59]. In comparison to the energy of the incident photon in Compton Scattering, the energy of the incident photon in photoelectric absorption, is equal to or slightly greater than the binding energy of the electron to the atom. The incident photon interacts with an inner-shell electron which absorbs the energy of the incident photon. The transferred energy is greater than the binding energy of the electron causing it to be ejected from the atom [59]. The remaining energy is converted into kinetic energy allowing the ejected electron to travel through matter.

Ejected electrons are emitted at all angles. However, the angle of emission is

smaller the higher the energy of the incident photon. As the ejected electron travels through matter the kinetic energy is dissipated in the atoms of the matter, until they are brought to rest. The vacancy left in the inner electron shell, by the ejected electron, is filled by an outer-shell orbital electron by electron transition. This effect cascades throughout the outer-shells with the vacancy left in one shell being filled by further outer-shell electrons.

1.8.2 Attenuation correction and Hybrid PET/CT Scanners

Gamma rays produced from concentrations of radiotracer located in the centre of the body have to pass through more tissue compared to gamma rays produced closer to the skin. The gamma rays produced from deeper in the body are attenuated more, therefore attenuation correction is required for accurate quantification of the radiotracer uptake. Before the development of hybrid PET/CT scanners attenuation correction was performed using a transmission based calibration. However, the SNR and spatial resolution of the CT-based attenuation data is superior to that of the transmission based calibration and is faster to acquire [43], which led to the development of hybrid PET/CT scanners. A further advantage of a hybrid PET/CT scanner is that scans are acquired at the same time, with the patient in the same position. Therefore, the registration of functional and anatomical information is possible [56] allowing for improved differentiation between abnormal and normal radiotracer uptake.

1.8.3 PET tracers

PET radiotracers are positron emitters with short half-lives compared to other emitters, including Iodine-124 (^{124}I) and Manganese-52 (^{52}Mn). ^{124}I and ^{52}Mn have half-lives of 102.2 hours and 134.2 hours respectively [60]. Common emitters are Fluorine-18 (^{18}F), Carbon-11 (^{11}C), Oxygen-15 (^{15}O), Nitrogen-13 (^{13}N) and Rubidium-82 (^{82}Rb). The labelling of a radionuclide with analogues of biological molecules to create a radiotracer allows for the depiction of different biological processes to be obtained from a PET scan. ^{18}F -FDG is a glucose analogue which relies upon a tumour's increased metabolic activity, in comparison to surrounding healthy tissue, for the depicting of the tumour tissue. ^{18}F -FDG has a half-life of 110 minutes and is a positron emitter producing a 511 keV gamma ray after annihilation [57]. Patients are typically injected with a fixed dose of ≈ 350 MBq or a dose of ≈ 4 MBq of ^{18}F -FDG/kg and rest for ≈ 60 -90 minutes before being scanned. In comparison ^{18}F -FDG, ^{11}C -Choline (^{11}C -Ch) relies upon a tumour's increased choline transport properties. Therefore, ^{11}C -Ch allows for the depiction of these tumours in areas ^{18}F -FDG is unsuitable, due to naturally high-levels of metabolic activity [61]. ^{11}C -Ch has a half-life of 20.5 minutes and is a positron emitter producing a 511 keV gamma ray after annihilation. Further to taking advantage of the increased properties of tumours, radiotracers such as ^{15}O , ^{82}Rb and ^{13}N can be used for the monitoring of blood flow and also in perfusion studies. ^{15}O has a half-life of 2 minutes, whilst ^{13}N and ^{82}Rb have half-lives of 10 minutes and 75 seconds respectively. ^{13}N and ^{15}O are solely positron emitters producing 511 keV gamma rays; however, ^{82}Rb

is a positron emitter and additionally uses electron capture decay processes which produce 511 keV and 777 keV gamma rays respectively.

1.8.4 Sinogram

Detected annihilation events are stored in a Sinogram. For each event, a line known as the Line Of Response (LOR) can be drawn between the detectors as shown in Figure 1.8a on page 30. Where A, B, C and D are the LOR drawn between pairs of detectors for detected events and X is the centre of the detector ring gantry. Each events LOR is plotted as its angular orientation around the detector ring (from -90° through 0° to 90°) against the LOR displacement from the center of the detector ring to generate the Sinogram as shown in Figure 1.8b on page 30. Where A, B, C and D are the corresponding LOR from Figure 1.8a.

1.8.5 PET Image Reconstruction

PET images are reconstructed into visual representations from the Sinogram obtained from the PET scanner. A variety of techniques exist for reconstructing PET Sinograms [43], including back-projection, filtered back-projection and iterative reconstruction techniques. Additionally Time of Flight (TOF) information recorded during acquisition of the of the Sinogram can be used during the reconstruction process to improve imaging contrast and reduce noise. Figure 1.9 demonstrates noise present in a PET image as well as a PET image denoised using a median filter. Two areas of noise in the PET

image are shown in red and blue. The noise is represented by darker regions in areas of no to little uptake and are typically a singular voxel. When a denoising filter is applied to the image, the heterogeneity is reduced within the red and blue circles.

Back-projection

In back-projection, a reconstruction matrix resolution is chosen and LOR are traced back along through the grid. The counts from each LOR are added to the counts of the preceding back-projected data, resulting in a back-projected image of the original object [57]. Simple back-projection techniques are limited by the blurring of the original object; however, blurring is decreased with increased distance from the original object. Blurring is therefore considered to be spill over from the object into neighbouring voxels.

Filtered Back-projection

Filtered back-projection techniques take advantage of the knowledge that the blurring of a back-projected image decreases with increased distance from the original object by applying filters to the acquired data. Application of a one dimensional convolution filter to the data, before back-projection, results in an image which closely approximates the original object. Additionally, Fourier transform methods complement filtered back-projection techniques by transforming the spatially projected data into the frequency domain. Transforming the data into the frequency domain allows for frequency domain filters to be applied to the image.

Iterative Reconstruction methods

Iterative reconstruction differs from back-projection based methods by using an iterative process and estimating an initial image. The projections for the initial estimated image are computed and compared against the measured projections in the Sinogram obtained from the PET scanner [43]. A correction factor is then applied to projections in order to generate a new Sinogram. The new Sinogram is then back projected to generate a new estimated image for input into the next iteration. Iterative reconstruction using maximum-likelihood expectation maximisation (MLEM) is a computationally expensive process, requiring hundreds of iterations of each projected view before achieving acceptable agreement between the estimated image and the measured projected data. ordered subset expectation maximisation (OSEM), however, reduces the computation time by grouping the angular projections into subsets and MLEM is performed on a subset rather than each projection in the subset.

1.8.6 DICOM Imaging format

Reconstructed PET images are exported from the scanner to the Digital Imaging and Communications in Medicine (DICOM) file format before being transferred to the picture archiving and communications system (PACS). DICOM is the standard for the communication and management of medical imaging information and related data [62]. The DICOM standard facilitates interoperability and compatibility of medical imaging equipment by specifying

ing network communications, the syntax and semantics of commands, media storage services as well as a file format and directory structure. A basic file structure for the DICOM standard is shown in Figure 1.10 on page 32.

1.9 Chapter 1 Overview

H&N & OC are two of the most commonly diagnosed cancers across the world today, with RT treatment being commonly used in the curative and palliative setting. Improvements in the technology behind RT have led to investigations in the con-formality of the delivered RT dose and multi-modality treatment pathways. Delineation of the GTV, however, has been identified as the largest source of error in accurate RT delivery. Therefore, PET imaging has been investigated for providing complementary information to aid in accurate GTV delineation. However, the low spatial resolution of PET imaging combined with complex biological uptake of the radiotracer means delineation on PET imaging is subject to inter and intra-observer variability and is a time consuming process. These combined challenges drive the interest in the need for semi-automated and automated delineation of the tumour on PET imaging. The following chapter introduces PET-AS algorithms and compares the methods that have been proposed in the literature.

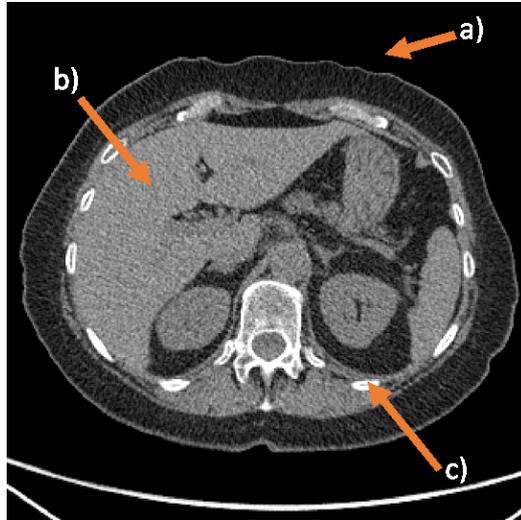


Figure 1.2: CT scan of patient. a) Air has a density of -1000 Hounsfields, b) Liver, has a density of $\approx 54-60$ Hounsfields and c) Bone has Hounsfield values of +700.

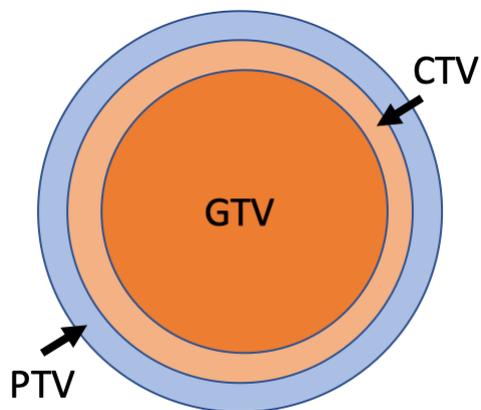


Figure 1.3: The GTV corresponds to all of the detectable disease. Whereas, the CTV is an extension of the GTV, which incorporates all of the GTV whilst accounting for microscopic disease extensions. An additional expansion of the CTV is required to account for errors which occur during radiotherapy planning and delivery.



Figure 1.4: A planar scintigraphy obtained from a patient injected with a radiotracer designed for the imaging and detection of bone tumours [39]. Areas of increased metabolic activity / radioactivity are darker and more visible in comparison to less metabolically active tissue.

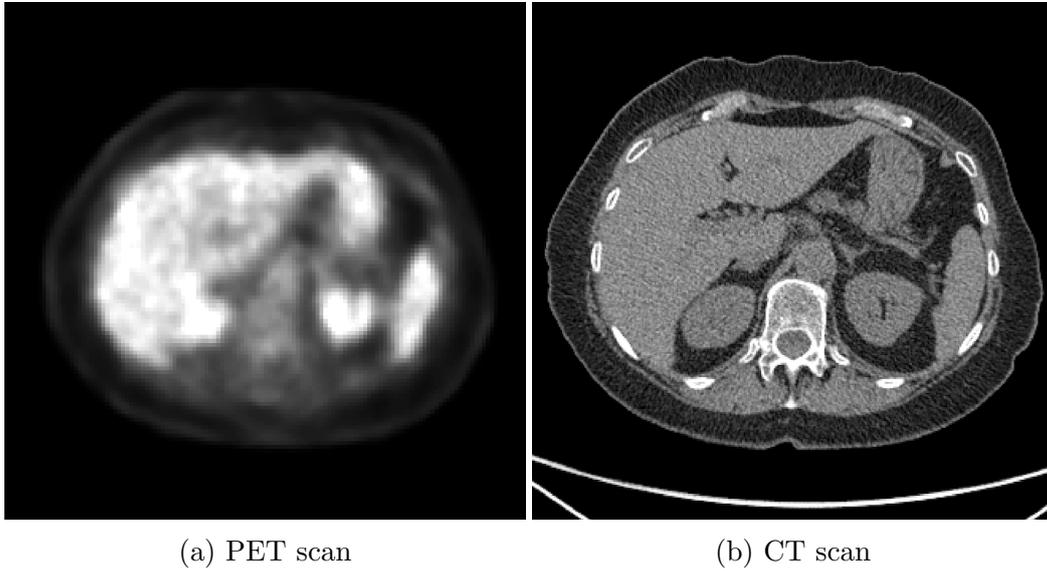


Figure 1.5: (a) PET showing heterogeneous uptake of the radiotracer in the liver. The more visible areas have increased metabolic uptake and are therefore potentially cancerous and b) CT imaging showing homogeneous tissue density values obtained from the same patient [40, 41]

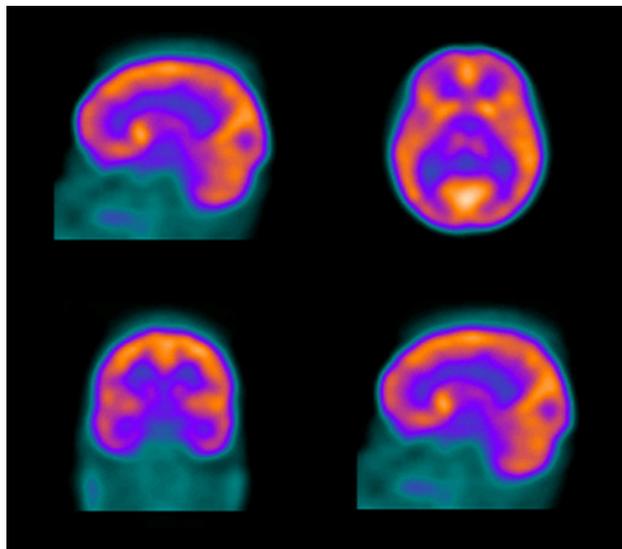


Figure 1.6: SPECT images obtained from a patient injected with a radiotracer designed for the monitoring of diffusion of blood in the brain [42]. Areas which are more visible (orange) indicate increased blood flow.

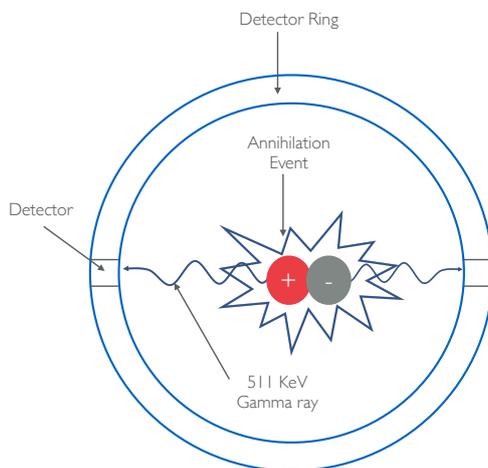
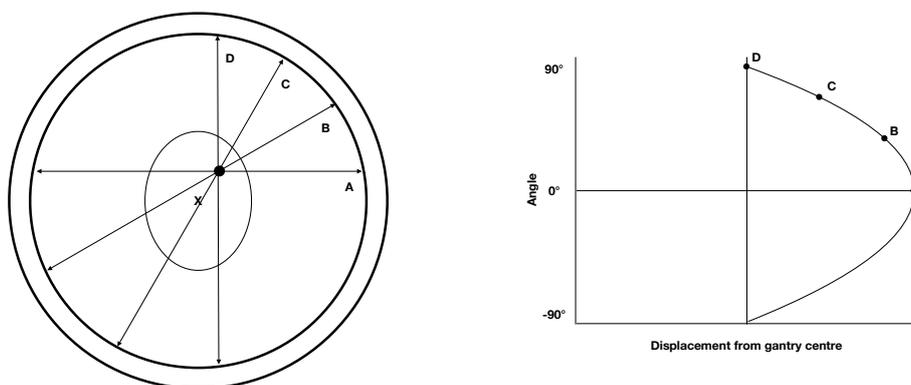
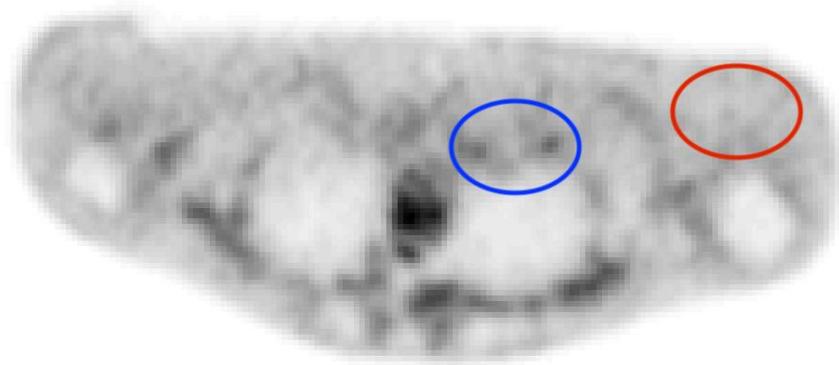


Figure 1.7: The collision of an positron emitted from ^{18}F -FDG and an electron resulting in an annihilation event and generation of gamma rays detected by the PET scanner.

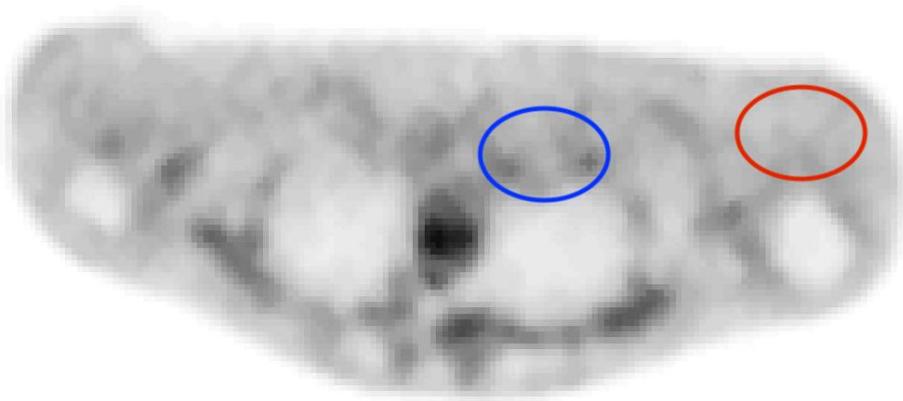


(a) Four lines of response detected by a PET scanner. (b) Sinogram of four lines of response detected by a PET scanner.

Figure 1.8: Lines of response and the resulting generated Sinogram.



(a) Original PET scan



(b) Denoised PET scan

Figure 1.9: (a) PET showing heterogeneous uptake of the radiotracer. Two areas of noise in the PET image are shown in red and blue. The noise is represented by darker regions in areas of no to little uptake and are typically a singular voxel. (b) A median filter is applied to the PET image to reduce noise present in the obtained image.

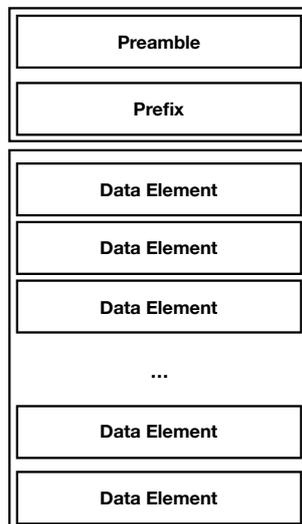


Figure 1.10: DICOM standard file format.

Chapter 2

Segmentation of PET

PET-AS methods potentially offer a more reliable MTV delineation process, which reduces intra-observer and inter-observer variability [37]. Thereby, this allows for the standardisation of MTV delineation across multiple centres [63], which is critical in multi-centre clinical trials. Multiple segmentation algorithms and methodologies have been published and recommended for use in clinical practice [35, 64–72]. There has been no recommendation or consensus, however, on a single segmentation method for use in the clinical environment [1] as the proposed PET-AS methods have been shown to perform differently when applied to PET images with different conditions [33]. The PET-AS methods investigated throughout this body of work are summarised in Table 2.1. The following section defines classifications for the proposed PET-AS methods based upon their implementation, approach and level of automation.

Table 2.1: Name and description of PET-AS methods used in this study, with references of published work using similar segmentation approaches

Algorithm	Description	Key References
AT	3D Adaptive iterative thresholding, using background subtraction	Jentzen <i>et al</i> [67], Drever <i>et al</i> [66]
RG	3D Region-growing with automatic seed finder and stopping criterion	Day <i>et al</i> [68]
KM	3D K-mean iterative clustering with custom stopping criterion	Zaidi and El Naqa [35]
FCM	3D Fuzzy C-mean iterative clustering with custom stopping criterion	Belhassen and Zaidi [71]
GCM	3D Gaussian Mixture Models based clustering with custom stopping criterion	Hatt <i>et al</i> [72]
WT	Watershed Transform-based algorithm, using sobel filter	Geets <i>et al</i> [69], Tylski <i>et al</i> [70]

2.1 Classification of PET-AS methods

PET-AS methods vary in implementation design, from intuitive threshold based segmentation methods [32], which include in the resulting tumour volume all voxels with an intensity higher than a single threshold value, to advanced machine-learned approaches [65]. Further, PET-AS methods differ in levels of automation, from being fully automated to semi-automated requiring user input. Within these, PET-AS method implementations can vary using differing pre and post-processing steps. Therefore, PET-AS methods can be classified in a variety of ways [1]:

- The segmentation algorithm employed and its assumptions and complexity.
- Level of pre and post-processing steps.
- Automation level.

The classification of PET-AS methods based upon image segmentation approach is a commonly used practice. This classification process relies upon comparing the statistical approach, clustering methodology, simplicity or complexity of the PET-AS algorithms.

A second classification approach compares PET-AS methods based upon the pre and/or post-processing steps used in the specific implementation of the PET-AS algorithm. PET-AS algorithms, however, are typically applied to raw PET data that have not been pre-processed. Optionally, de-noising filters may or may not be used in a PET-AS algorithms implementation. Within this classification approach, further classifiers are the use of phantom acquired data to optimise the PET-AS algorithm, as well as the requirement of image databases to develop statistical models for MTV delineation.

Classifying PET-AS algorithms based upon the level of automation requires the division of the MTV delineation into two different steps [73]. These processes are the identification of the tumour location and then the delineation of the MTV. Dependency of MTV delineation on operators, through the identification of the tumour location, requires operators to have specific expert knowledge of ^{18}F -FDG PET, MTV delineation and the diagnosed cancer type. Therefore, the majority of PET-AS algorithms rely upon identification

of the tumour location by a user defining a volume of interest (VOI). The PET-AS method is then applied within this defined VOI. This is classed as the standard delineation process and therefore the majority of PET-AS algorithms are regarded as semi-automated, due to the need for human interaction. Other proposed algorithms rely upon the identification of the tumour location after application of the PET-AS method to the PET image [71] or the manual definition of the area defined as the background uptake in a PET image. Additionally, some PET-AS methods and algorithms require the definition of seed points, within the tumour location, from which the segmentation algorithm is initialised [74].

2.2 PET-AS method implementations

2.2.1 Fixed and adaptive threshold algorithms

In the most simplistic threshold-based PET-AS method, the threshold value is defined by the user of the computer. The threshold value can be defined as a single absolute intensity value, or as a percentage of the maximum intensity within the image (see Equation 2.1). $I_{(i,j)}$ is the value of the image at point i, j and T is the user defined threshold. $Output_{(i,j)}$ is the final value in the output image at point i, j . Defining the threshold as a percentage allows for an image, tumour and patient independent MTV delineation process. This independence is further enhanced when an image voxel value is converted to a Standardised Uptake Value (SUV) (see Eq. 2.2). The injected activity

(ID) corresponds to the amount of activity injected corrected for radioactive decay between injection and image acquisition. The measured activity (C_{img}) is the activity on the acquired PET image, the body weight BW of the patient being expressed in kg.

$$Output_{(i,j)} = \begin{cases} 0, & \text{if } I_{(i,j)} < T \\ 1, & \text{if } I_{(i,j)} \geq T \end{cases} \quad (2.1)$$

$$SUV(t) = \frac{C_{img}(t)}{ID/BW} \quad (2.2)$$

However, the maximum SUV (SUV_{max}) of ^{18}F -FDG PET imaging has been found to be subject to noise [75], therefore thresholding based upon a percentage of the peak SUV (SUV_{peak}) has been proposed as being a more robust methodology [76]. The SUV_{peak} of a PET image is defined as the mean uptake in a 1 cm^3 sphere centred around the SUV_{max} of that PET image. For multiple equivalent SUV_{max} values, the maximum SUV_{peak} is selected.

Adaptive thresholding techniques define the threshold relative to the difference between the SUV_{max} or SUV_{peak} of the MTV and the mean intensity of a region defined as the background. Adaptive thresholding techniques are typically iterative processes that assume the biological uptake of the radio-tracer in comparison to the background uptake is distinctly bi-modal and relatively homogenous. Therefore, an appropriate threshold can be found by

minimising the change in the tumour volume in comparison between one iteration and the previous iteration. The definition of the background, however, is implementation dependent [67, 77–79].

Whilst simple in implementation, thresholding methods have been shown to lack robustness [37] and are subject to the thickness of phantoms walls [77, 80]. The performance of thresholding methods has also been shown to correlate with the SUV_{\max} [32] and the volume of the tumour [81]. These dependencies require operators to have image and patient specific information, combined with expert training to accurately delineate the MTV when using threshold-based PET-AS methods.

2.2.2 Gradient segmentation

Gradient-based segmentation algorithms are based upon finding contours which naturally transition areas with high biological uptake and areas with low biological uptake. If the PET image is visualised as a height map, the changes in color, intensity or texture correspond to the crests and troughs of the image. Qualitatively, edges occur at the boundaries between regions of different color, intensity or texture [82]. The edges can be detected by gradient-based threshold approaches, region-growing techniques using gradient-based thresholding, watershed transform methods which use flooding-based techniques [70] and active contouring [83].

In Watershed segmentation, the image is considered to be a topographical map, for which minima and maxima can be defined. The map is flooded from

the minima until a singular maxima is remains. Figure 2.1 demonstrates the different stages of Watershed segmentation. Local minima are selected in the height map (orange points) and the image is flooded from the minima. The flooding continued until a singular maxima (red points) is left and the resulting watershed (black arrow) is considered to be the final contour.

In active contouring, contours are iteratively deformed and attracted to the crests of an image until a stable contour is reached. In Figure 2.2, an initial loose active contour (red line) is defined around the object to segment (blue T). The active contour is attracted to the change in image intensity between the white background and the object to segment. The final segmentation is reached once the active contour is considered stable. Deformations of the contours can be influenced by image and user guided forces such as anchor points.

An advantage of gradient based methods, in comparison to threshold methods, is that the uptake distribution does not need to be homogeneous along the contour. However, gradient-based methods depend upon the precision of the gradient information, which is influenced by spatial resolution [1]. Additionally, computation of the gradient map amplifies noise in the PET image. Noise is seen as singular voxels with areas of radiotracer uptake, surrounded by low uptake. Therefore, due to the well defined difference between the noise and the surrounding uptake, the boundary between them is determined to be an edge, when computing the gradient map. Therefore, de-noising algorithms should be applied as long as the spatial resolution of the PET image is not decreased.

2.2.3 Region growing

Region growing techniques avoid the generation of disconnected contours on slices by including in the MTV, voxels in the neighbouring region which satisfy a similarity criteria. The first region is grown from the SUV_{\max} , SUV_{peak} of the image or an operator selected seed point. Voxels are included within the grown region based upon statistical properties such as the mean uptake and the Standard Deviation (SD), confidence intervals, textural properties or whether the neighbouring voxels are within a specified threshold range. Voxels included based upon similarity to a threshold range are considered to be connected threshold algorithms. Further to differing inclusion criteria, differing stopping criteria can be implemented. Stopping criteria may include the total number of voxels included as the tumour volume, the number of iterations reached or the difference between the mean background definition and the mean uptake within the tumour volume delineation. Definition of the inclusion criteria and stopping criteria can require the definition of hyper-parameters set by the user; however, optimisation of the hyper-parameter is not possible for all objects [68].

2.2.4 Statistical

Statistical image segmentation aims at classifying voxels within an image into different clusters (unique memberships) and regions based upon the statistical properties of these clusters and regions. This is achieved by probability calculations and estimations. The number of clusters is typically defined

by the operator. Further to this, clustering classification can be binary (background or tumour) as well as probability values. Definition of clustering memberships has been investigated by a number of groups [35, 71, 84]. However, the number of clusters to classify is operator and implementation dependent.

2.2.5 Consensus Techniques

Consensus techniques, including the simultaneous truth and performance level estimation (STAPLE) [85] algorithm and majority vote (MJV) [86] techniques, are based upon the definition of multiple contours by individual PET-AS methods. PET-AS method generated contours are combined based upon statistical probabilities [64] with the aim of minimising the limitations of each segmentation method and maximising the advantages of each one.

2.2.6 Machine learning

Machine-learning approaches to MTV delineation require a learning task consisting of the discrimination of biological tracer uptake within a tumour volume and the background. Learning techniques can be divided into two categories: supervised and un-supervised learning. Supervised learning estimates a mapping from labelled samples which make up the training dataset. Samples are typically labelled manually. In unsupervised learning input samples are used to generate a map; however, their labels are not provided. In

the learning stage, the development of the optimal model with the optimal features is sought. This is followed by a validation stage, which aims to quantify the performance of a model on data outside of the training stage. A variety of machine-learning and classification approaches exist.

Decision Trees

Decision tree (DT) techniques develop a set of rules from a pre-defined set of features by the development of questions. This results in tree-like structures, with leaf nodes representing the final classification of the object being identified. A limitation of DT approaches is that deep and complex trees can be the result of overfitting. Overfitting of the model, to the training data, results in limited performance for the model on data outside of the training model. In complex DTs it is therefore preferable to prune the resulting model, to improve the model performance and for the generalisation of the model.

Random Forest

Random forests are similar to DT approaches. However, they result in the development of multiple DTs from features selected randomly from a set of pre-defined features. The classification result is then the averaged result from all of the DTs that were developed as part of the training step. Due to the development of multiple DTs, random forest approaches are not as easily subject to the overfitting of the model. However, this results in an algorithm which is slow and unsuitable for real-time analysis. The number of DTs

developed as part of the forest are defined as a hyper-parameter. Typically, the more trees developed, the more accurate the model. Further, Random Forests require a larger number of features compared to DTs.

Support Vector Machines

A Support Vector Machine (SVM) is a type of machine learning algorithm which plots data points from provided features into a n-dimensional space where the number of dimensions matches the number of features. The SVM finds the hyper-plane (boundary) which segregates the classes. The hyper-plane segregating the classes is chosen based upon the performance and margin of the hyper-plane to the classes of data. SVMs are robust to outliers, due to hyper-planes being chosen based upon performance. If data classes are not segregated by a hyper-plane an additional feature and dimensionality is added to the model. The hyper-plane is found from the transformation of the data in this additional space and then transformed back to the original number of dimensions. The reliance of hyper-planes in SVM means the performance of the method in high noise data is limited.

Deep Learning

Deep learning requires the training of a model consisting of many layers which use the output from previous layers as input. For example, a convolution neural network consists of fully connected neural networks with alternating layers of convolution and max-pooling layers. Max-pooling layers sample the

data in order to reduce data dimensionality. Therefore, allowing for assumptions to be made about features contained in the sub-regions to be binned and to provide an abstract representation of the data. Investigations into the role of machine-learning based segmentation methodologies and PET imaging are limited due to the higher level of complexity required in comparison to the methods described previously. Additionally, deep learning investigations are typically limited to Magnetic Resonance (MR) or CT-based studies [17]. For more information see the book *Deep Learning* written by Goodfellow *et al* [87].

2.3 PET-AS Comparison

Comparison of the performance, robustness and suitability of the multiple PET-AS methods that have been proposed for MTV delineation is challenging, due to the proposed methods being validated in differing anatomical sites and on differing PET imaging datasets [17]. Comparison studies typically compare the results of PET-AS MTV delineation against the ground truth from spherical phantom inserts or manually defined contours rather than histopathological samples [33, 48, 74, 86, 88]. Tyłski *et al* [74] compared 4 threshold methods and a model-based method in 17 Spheres and 41 non-spherical simulated tumours. However, the results of the PET-AS MTV delineation weren't compared in clinically acquired scans (manual contours or histopathological data). Phantoms are common tools for analysing the sensitivity, noise and spatial resolution of the PET scanner in the clinical

environment [1], therefore are more readily available for comparison studies. A limitation of using phantom based PET scans for the validation of PET-AS methods is that they do not represent realistic tumours, due to using homogeneous uptake on a homogeneous background. Adequate validation of a PET-AS method on realistic uptake distributions can only be achieved using synthetic simulated PET scans or clinically acquired PET scans with histopathological measurement of the tumour. However, the number of studies comparing the result of PET-AS MTV delineation in synthetic simulated tumours with known ground truth contours is limited [65, 74].

Further, it is challenging to compare methods from the literature alone as the proposed PET-AS methods have been developed on and validated in differing PET imaging datasets too. Berthon *et al* [65] proposed a segmentation methodology called ATLAAS, which was based upon the application of machine-learned DTs to select the most appropriate segmentation methodology for the delineation of the MTV. The DTs are developed from known factors which have been shown to effect the delineation of the MTV and include the TBR and volume (mL). They evaluated their method in 85 phantom and printed sub-resolution sandwich phantoms. In an additional study, ATLAAS was validated on 20 clinical diagnostic H&N PET scans [36]. In both studies, ATLAAS was not compared to histopathological data. Geets *et al* [69], Belhassen *et al* [71], Dewalle-Vignon *et al* [89], Abdoli *et al* [83] all validated their proposed PET-AS method in a common cohort of seven patients obtained from a patient cohort of nine patients [46]. In comparison, Day *et al* [68] proposed a 3D region-growing method, which performed bet-

ter than fixed thresholding schemes on 18 rectal and anal cancer patients; however, they cautioned it should be used as a starting point for MTV delineation in order to reduce intra and inter-observer variability. Further to this, Hatt *et al* [72] developed and validated the Fuzzy Locally Adaptive Bayesian (FLAB) method, based on a fuzzy clustering scheme incorporating an expectation maximisation step. The performance of FLAB was evaluated with spherical fillable phantom data and more complex simulated data. FLAB showed high performance compared to thresholding and other clustering methods, especially for small objects. In a comprehensive comparison study, Berthon *et al* [33] evaluated the performance of 8 PET-AS methods in sixteen non-spherical phantom inserts. These methods, however, were not compared in clinical data with or without histopathological specimens or simulated PET data. In the study, it was found that each of the proposed PET-AS methodologies perform differently under different conditions.

The limited number of participants in PET studies and in the validation of the proposed PET-AS algorithms limits the statistical power of the results obtained in studies [36, 46, 68, 69, 71, 83, 89]. This, combined with the knowledge that PET-AS methodologies perform differently under differing conditions (as previously mentioned) resulted in the findings of the American Association for Physicists in Medicine (AAPM) Task Group No 211 report [17], which state that no single PET-AS method can be recommended for realistic target volume delineation in all cases. The AAPM report also states that machine-learned PET-AS methodologies are showing promise for accurate target volume delineation. Further to the AAPM Task Group No

211 report, the current recommendations of the International Atomic Energy Agency (IAEA) is that it is difficult to recommend a single PET-AS method for accurate MTV delineation [31]. The following chapter aims to highlight how the application of PET-AS algorithms in the clinical environment can potentially impact patient OS, thus demonstrating the need for a standardised PET-AS algorithm.

2.4 ATLAAS

ATLAAS [65] is a predictive segmentation model, incorporated into the computational environment for radiotherapy research (CERR) [90] software package developed in Matlab. CERR is open source software that was developed at the university of St Louis (Michigan, USA) and is currently maintained at the Memorial Sloan Kettering Cancer Centre (MSKCC) in New York (USA). ATLAAS is designed to select the most accurate PET-AS method for the optimal segmentation of a given PET image. The most appropriate segmentation method is chosen from a list of advanced PET-AS methods and algorithms which have been built into the system. When ATLAAS is applied to a PET image, ATLAAS computes for each PET-AS algorithm its predicted performance using a number of parameters extracted from the target PET image. Performance is quantified as the predicted dice similarity coefficient (DSC) as shown in Equation 2.3. DSC is defined as twice the intersection of X and Y , divided by the union of X and Y , where X is the ground truth contour and Y is the PET-AS contour. The prediction is done

using a model consisting of DTs built during the training stage of ATLAAS. The training model is built on a large dataset of PET images with tumour parameters varying within a defined range. The work flow for the development of the training model and application of the training model to a PET image is shown in Figure 2.3.

$$DSC = \frac{2|X \cap Y|}{|X| + |Y|} \quad (2.3)$$

ATLAAS is based upon the principle that differing PET-AS methods may be more adequate than others for the delineation of different lesions [65] as clinical lesions can show a variety of patterns, from homogeneous uptake to highly heterogeneous uptake. Further, clinical lesions may incorporate local hot spots or necrotic areas. Previous work by Berthon *et al* [65] identified the following tumour and PET image characteristics as classifiers for PET-AS performance:

- Volume (mL): target object volume.
- Tumour to background ratio peak: Ratio between the target object's SUV_{peak} , calculated as the mean value in a 1 cm^3 sphere centred on the SUV_{max} in the target object, and the background SUV, calculated as the mean intensity in a 0.5 cm thick extension of the object contour.
- Number of discrete intensities (NI): a regional texture feature related to the intensity distribution in the target object.

Additional parameters may be defined by a user, known as hyper-parameters, to inform the development of the statistical model. Potential hyper-parameters for ATLAAS include the lymph node size, the number of involved lymph nodes, the total number of distant metastases as well as patient characteristics including weight and age. ATLAAS is designed to have limited user interaction and with no prior knowledge of the PET other than the primary tumour location, therefore hyper-parameters are not used in the development or application of the statistical model.

The following PET-AS methods with different approaches to segmentation have been included in the ATLAAS training model by Berthon *et al* [65]

- Adaptive iterative thresholding (AT)
- K-means (KM) with 2, 3 and 4 clusters
- Gaussian Fuzzy C-means (GCM) with 3 and 4 clusters
- Watershed Transform (WT)

2.4.1 Training model development

For each of the included PET-AS methods, a DT is developed from a training dataset. The ATLAAS training dataset is developed using the PET Simulator of Tracers via Emission Projection (PETSTEP) simulator [91], which is incorporated into CERR. PETSTEP simulates PET scans using a CT image and a ^{18}F -FDG background uptake map in order to simulate a PET image tumour from contours defined by a user. The ATLAAS training dataset was

generated based on PET/CT data from a fillable phantom. Tumour objects with a range of characteristics relevant to clinical situations were added to the background of the phantom. The training dataset consists of 100 spherical tumour objects modeled for volume and maximum uptake values in the range 0.5 ml-50 ml and 4000 Bq ml⁻¹-40000 Bq ml⁻¹ respectively [65].

2.4.2 Application of ATLAAS training model to FDG PET imaging

The delineation of the MTV on a PET image, for which the ground truth (GT) contour is unknown, requires the estimation of the parameters incorporated into the developed ATLAAS training model. Estimated parameters are acquired from an estimation of the MTV, which is delineated by applying a PET-AS algorithm to the PET image. Estimated tumour characteristics are used as input to the DTs, which output the predicted DSC for each PET-AS algorithm included in the training stage of ATLAAS. The PET-AS method with the highest predicted DSC is then used to delineate the final MTV.

Region of Interest definition

For accurate estimation of the MTV, the definition of a region of interest (ROI) is required. A ROI limits over-contouring of the estimated MTV, allows avoidance of areas of erroneous uptake as well as improving estimation of the training model parameters. A variety of approaches exist for the

definition of the ROI:

- Manual definition requires a user to select which voxels to investigate as potentially being the tumour. This is typically done by “painting” the ROI on the PET image. This process can be time consuming and has the same limitations as manual definition of the MTV.
- Existing contour expansion; if a contour has been pre-defined due to being involved in a retrospective study or RT planning, it is possible to expand this contour by a user defined measurement to use as the ROI.
- Semi-automated definition of the ROI typically requires the user to place a sphere or cube of a user defined size around the area to delineate. More advanced semi-automated processes exist, in which the user defines a limited number of seed points in the sagittal, axial and transverse planes. Seed points are converted to a spheroid around the centre of mass, which AT is then applied to. The resulting delineation is expanded. The advantage of this more advanced process is a reduced time to define the ROI as it requires only the definition of two seed points, and a ROI which is clinically relevant to the data in the PET scan is generated allowing for the avoidance of contouring in areas of erroneous uptake.

This chapter has discussed the multitude of PET-AS algorithms that have been proposed for accurate MTV delineation, from simple threshold-based techniques to more advanced methodologies; including the development of decision trees from a ^{18}F -FDG PET based training dataset. The follow-

ing chapter aims to demonstrate the impact MTV delineation has on a patient's OS and risk stratification, thereby potentially affecting their quality of life.

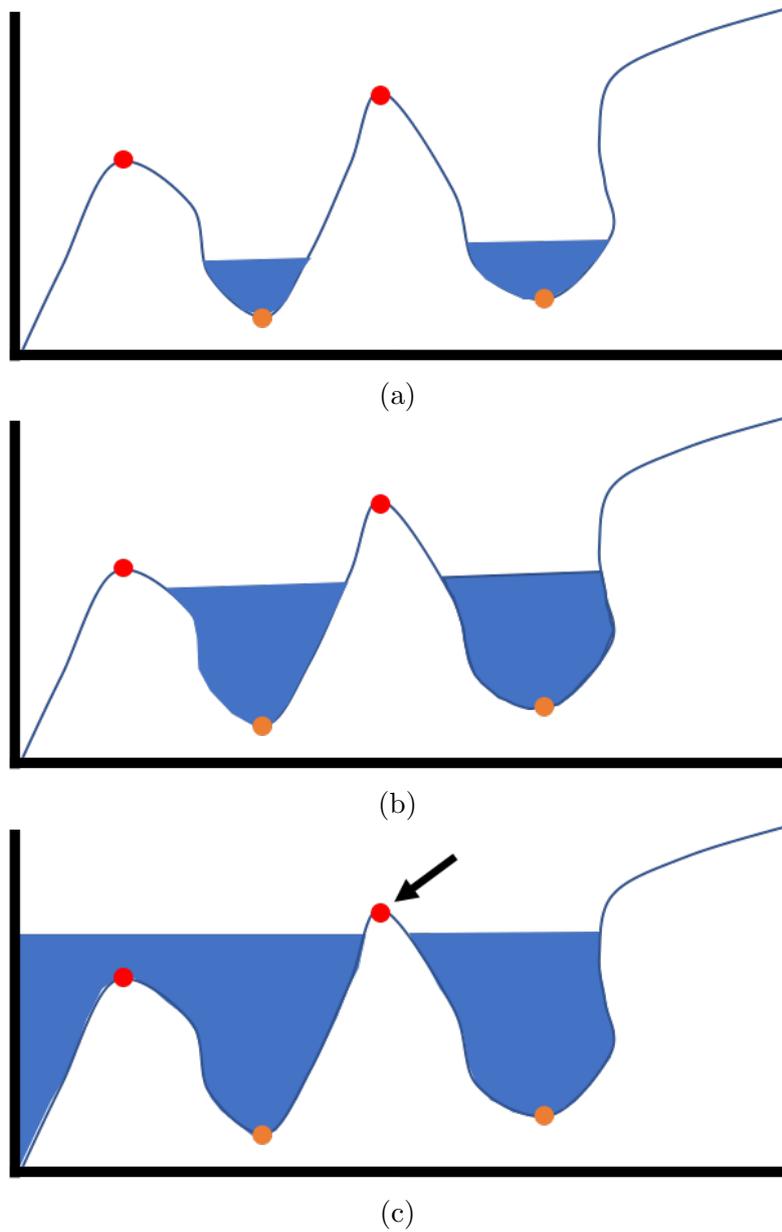


Figure 2.1: a) The first stage of Watershed segmentation. The image is considered as a topographical height map and local minima (orange points) and maxima (red points) are defined from which flooding of the height map starts. b) From the minima selected in the image, the image is flooded until a singular maxima is left. b) The resulting watershed (black arrow) is considered to be the final image contour and segmentation.

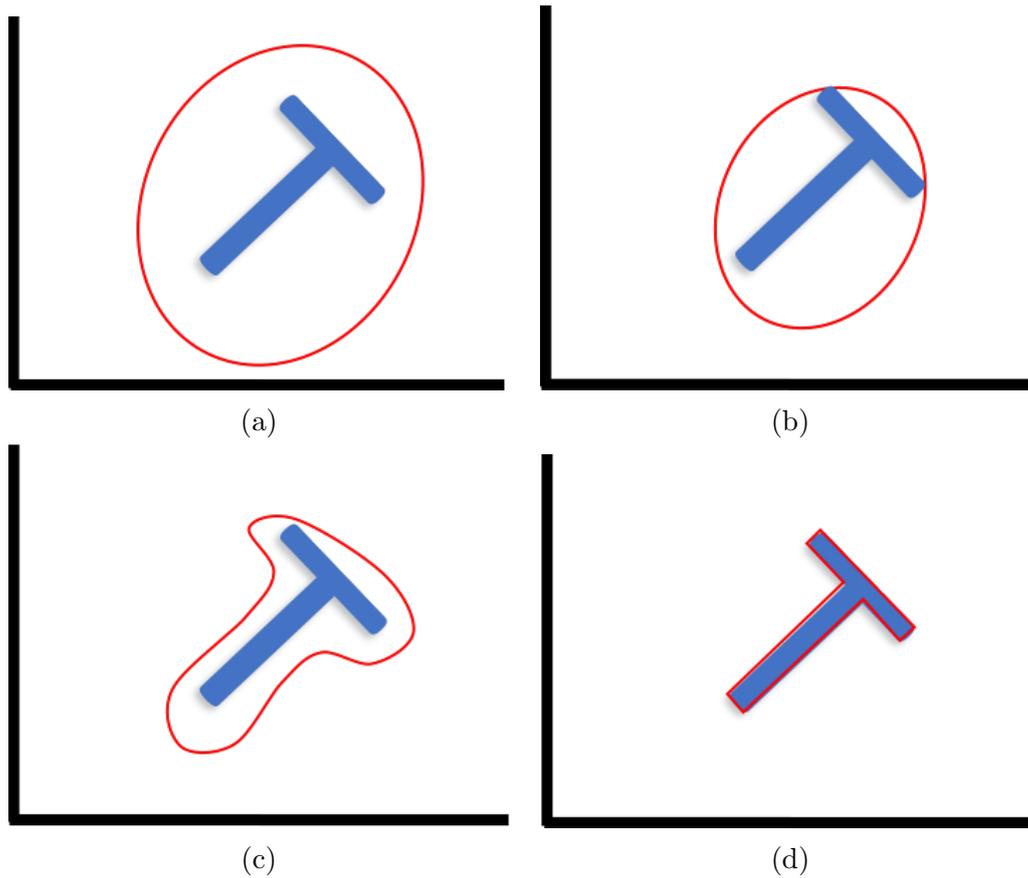


Figure 2.2: a) The first stage of active contouring is to define a loose contour (red line) around the object to segment (blue T). b) The defined contour is attracted to changes in image intensity. c) The active contour is continuously modified until it reaches a stable state. d) The final segmentation once the active contour has reached a stable state.

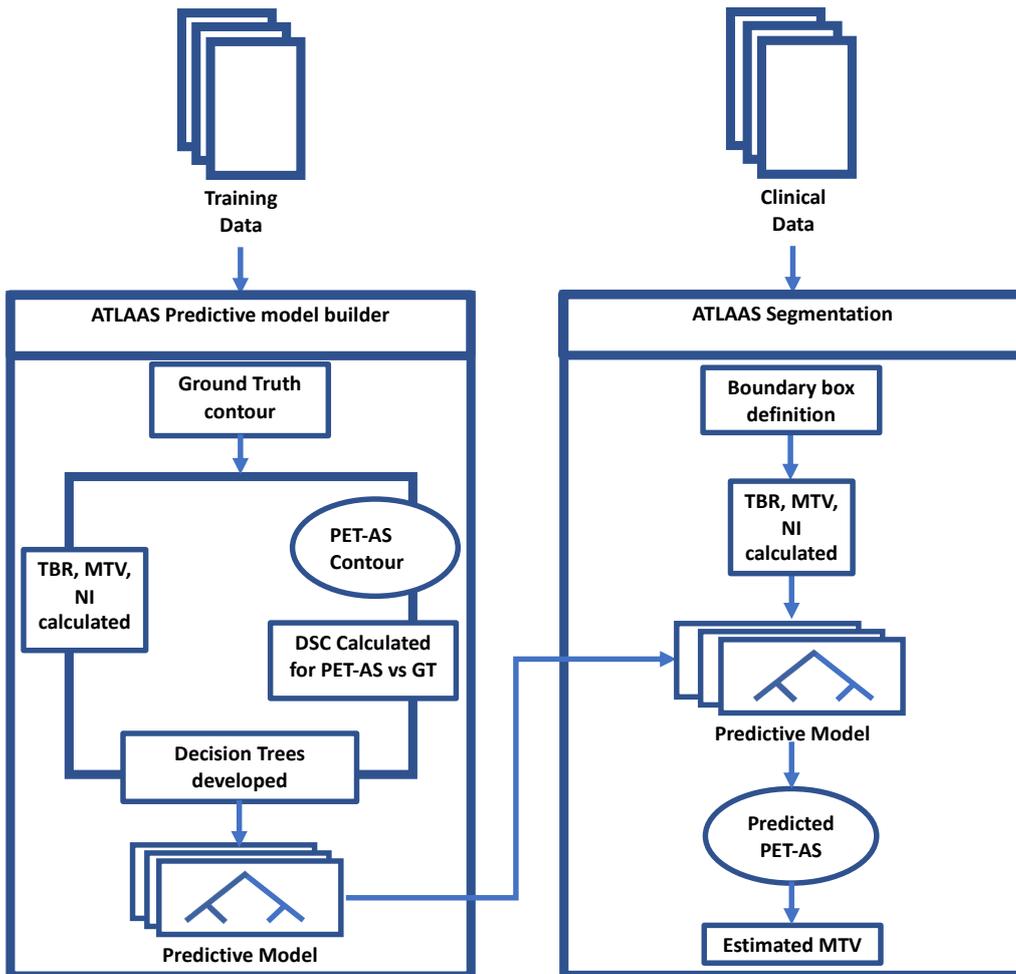


Figure 2.3: The ATLAAS training model and application workflow.

Chapter 3

Impact of metabolic tumour volume segmentation on patient overall survival

OC is the eighth most common [2] diagnosed cancer worldwide, with extremely high rates of mortality. It is hoped that the development of prognostic models, combined with precision medicine, may improve the patient rate of mortality from 1— and 5— year OS rates of 44% and 15% respectively [11]. Traditionally, prognostic models are developed from patient-specific information including age, pathological subtype, molecular characterisation and tumour staging, resulting in a clinical model, which characterises a patient's overall survival and the likelihood of the patient's outcome. By combining the

advanced quantitative analysis of medical imaging and clinically developed prognostic models, it is hoped that the performance of prognostic models can be improved. The advanced quantitative analysis of medical imaging modalities is known as radiomics. Radiomic features aim to identify tumour biomarkers and heterogeneity, through the extraction of high-dimensional data, [15] that can be associated with metastatic growth, recurrence and OS [92]. Radiomic features may also have significant prognostic value in the management of cancer [93]. However, the extraction of radiomic features and the results of radiomic analysis are dependent upon the method used to delineate the MTV [17], with relatively few studies comparing the results of radiomic analysis derived from each PET-AS method (cf. [94] and references therein). Further, few studies have investigated the effect of radiomic analysis from PET-AS methods on patient risk stratification [18,95,96]. Therefore, this chapter aimed to evaluate the influence of PET-AS method MTV delineation on patient risk stratification and the resulting patient OS, in OC, by developing a series of prognostic models in the same patient cohort, with identical clinical data and standardised radiomic features derived from different PET-AS methods. The following sections describe the materials & methods used to achieve the aims of this chapter.

3.1 Materials and methods

3.1.1 Patient Cohort

Four hundred and eighty six patients with biopsy-proven OC, including Gastro-oesophageal junction (GOJ) tumours, were retrospectively identified for inclusion in this chapter. The identified patients were radiologically staged between 16th September 2010 and 31st July 2016 with staging performed according to the International Union Against Cancer (UICC) TNM 7th edition [8] recommendations. The requirement for informed consent from patients was waived by the Institutional Review Board after approval for patient inclusion in research studies (Wales REC 1, UK reference 14/WA/1208). Fourteen of the identified patients were initially excluded due to having missing clinical data. Therefore, following the exclusions, 472 patients were selected for analysis and after contrast-enhanced CT staging investigation, all of the included patients were deemed to have potentially curable disease.

3.1.2 PET/CT protocol

Patients were fasted for at least 6 hours prior to ^{18}F -FDG tracer administration and serum glucose levels were routinely checked and confirmed as less than 7.0 mmol/L prior to PET/CT acquisition. Patients were injected with a dose of 4 MBq of ^{18}F -FDG/kg and rested for 90 minutes before scan acquisition, which is standard practice in our institution. PET/CT scan acquisition



Figure 3.1: A GE 690 PET/CT scanner [97].

was performed using a GE 690 scanner (GE Healthcare, Buckinghamshire, UK). The scanner is shown in Figure 3.1. CT images were acquired in a helical acquisition with a pitch of 0.98 and tube rotation speed of 0.5 seconds with a tube output of 120 kVp. Output modulation was between 20 and 200 mA. The matrix size for the CT acquisition was 512×512 pixels with a 50 cm field of view with no oral or intravenous contrast administered to patients before scanning. PET imaging was acquired at 3 minutes per field of view using the VUE Point time of flight algorithm. The length of the axial field of view was 15.7 cm (skull base to mid-thigh) and PET images were reconstructed with the OSEM algorithm using 24 subsets and 2 iterations. All PET based data was obtained using the same PET/CT scanner and reconstruction method with resulting voxel dimensions of $2.73 \times 2.73 \times 3.27$ mm and an axial slice matrix size of 256×256 voxels. A PET/CT scan demonstrating FDG-avid uptake typical of the patient cohort is shown in Figure 3.2.



Figure 3.2: A co-registered PET/CT scan from the patient cohort demonstrating FDG-avid uptake.

3.1.3 Treatment Protocols

All patients began treatment 2-4 weeks after staging ^{18}F -FDG PET/CT image acquisition. Patients either had Endoscopic Mucosal Resection (EMR), surgery alone, Neo-adjuvant chemotherapy (NACT) or Neo-adjuvant chemoradiotherapy (NACRT) prior to surgery, Definitive chemoradiotherapy (dCRT) or palliative therapy. The optimum treatment strategy was decided by consensus at the MDT. In general, fit patients with tumours pre-operatively staged as T3/T4a, N0/N1 were pre-operatively treated with NACT or NACRT. Less fit patients, or those with T1/2 N0 disease, had surgery alone. TNM staging is defined in Section 1.5. Whilst patients deemed unsuitable for surgery due to co-morbidity and/or performance status, extensive loco-regional

disease, or personal choice received dCRT.

3.1.4 Hardware and software

A 3.2 Ghz Intel Core i5 computer, with 16GB of RAM and a 2GB dedicated AMD Radeon R9 M380 graphics processing unit (GPU) was used to process PET/CT data in this project. PET-AS algorithms were implemented in the Matlab programming language with a Matlab 2016b license (The Mathworks, Natick, USA). Image processing and statistics toolbox licenses were installed as well. Processing of the PET imaging for automated delineation of the MTV was done with the open source software CERR [90].

3.1.5 Segmentation algorithms

The PET-AS methods included in this study were selected for having shown promise for accurate MTV delineation [36]. The PET-AS methods are implemented, as part of a common software package shown in Figure 3.3, as fully automatic methods. In Figure 3.3, a ROI has been defined by the user of the software in pink and the ATLAAS segmentation methodology has been applied to the ROI. The resulting segmentation contour is shown in black. Before definition of the ROI and segmentation with ATLAAS, a reverse gray scale colour map was applied to PET image allowing for intuitive visualisation of the tumour. In a reverse gray scale colour map, areas of higher metabolic uptake are shown as darker regions. In each case, the MTV was defined using AT, Fuzzy C-means (FCM)2, GCM3, GCM4, KM2, KM3,

KM4, Region Growing (RG) and WT PET-AS methods. The parameters and the number of clusters for each PET-AS method were chosen for use in previous studies [77] and having demonstrated acceptable performance in PET imaging [36]. The PET-AS methods were applied to a ROI that was defined manually around the primary tumour by a clinical radiologist with five years' research experience. Contours derived by the nine PET-AS segmentation methods were assessed by the same clinical radiologist subjectively for accurate tumour representation. All tumour contours were visualised using the same software and image settings to ensure consistent methodology. Segmentation methods were considered inadequate for further analysis if less than 90% of contours were non-representative of the primary tumour. This pre-defined value was decided upon prior to image visualisation. Contours were assessed individually and classified as not representative if contours were greatly different from the primary tumour, or included bone, lung or mediastinal tissue. In addition, segmentation methods that had failed or conformed to the boundary of the bounding box were defined as not representative of the primary tumour. Figure 3.4 demonstrates the different contours generated by PET-AS methods and their considered acceptability. Five segmentation methods denoted to have a poor MTV representation were excluded from the study.

Adaptive Thresholding

AT is an iterative method, developed by Drever *et al* [66], that starts with an initial fixed threshold method but it modifies the threshold value on each

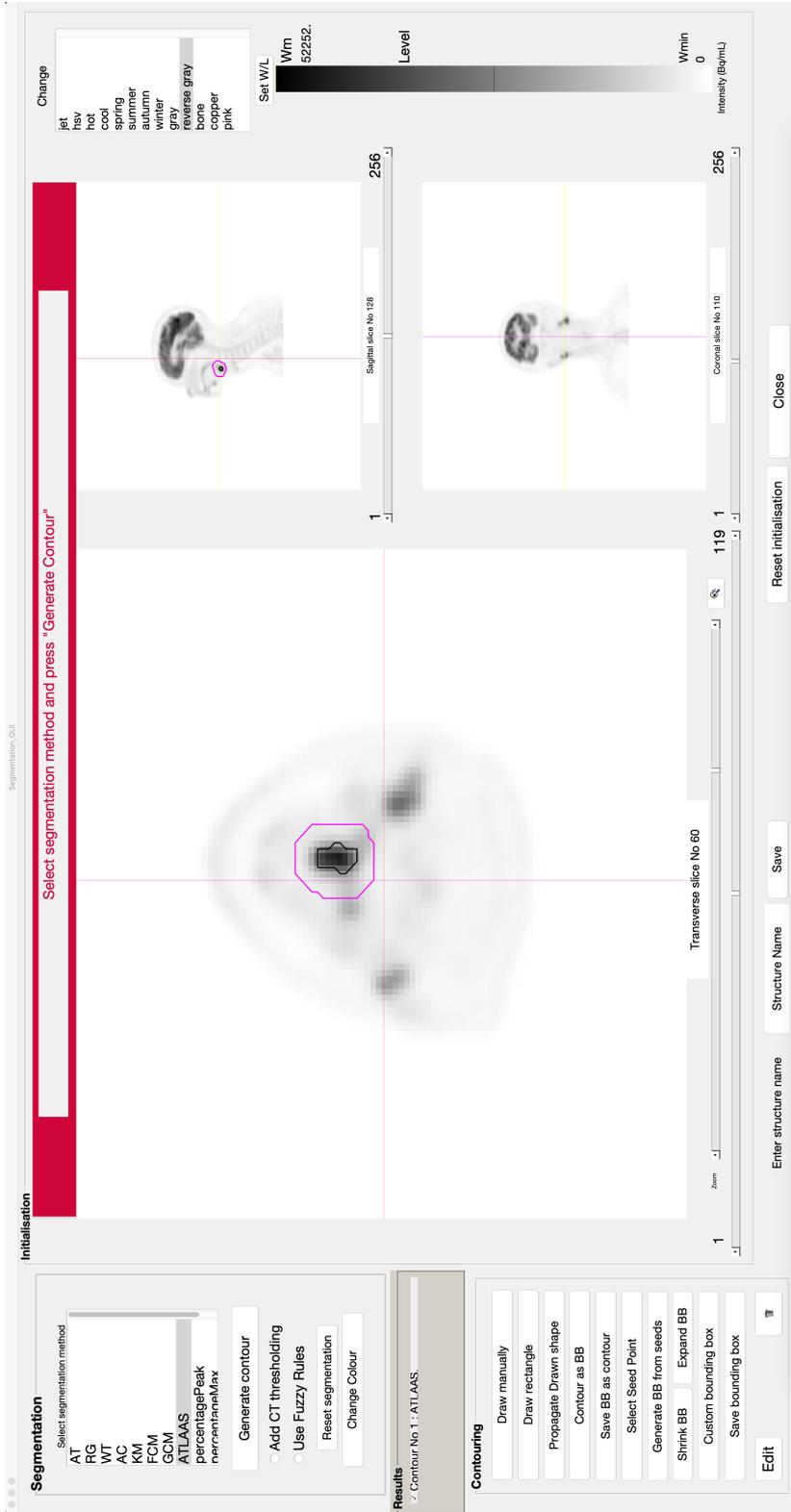


Figure 3.3: Software package containing the automated segmentation algorithms. A ROI has been defined by the user of the software in pink and ATLAAS segmentation has been applied to the ROI. The ATLAAS contour is black. A reverse gray scale colour map was applied to PET image, therefore areas of higher metabolic uptake are shown as darker regions.

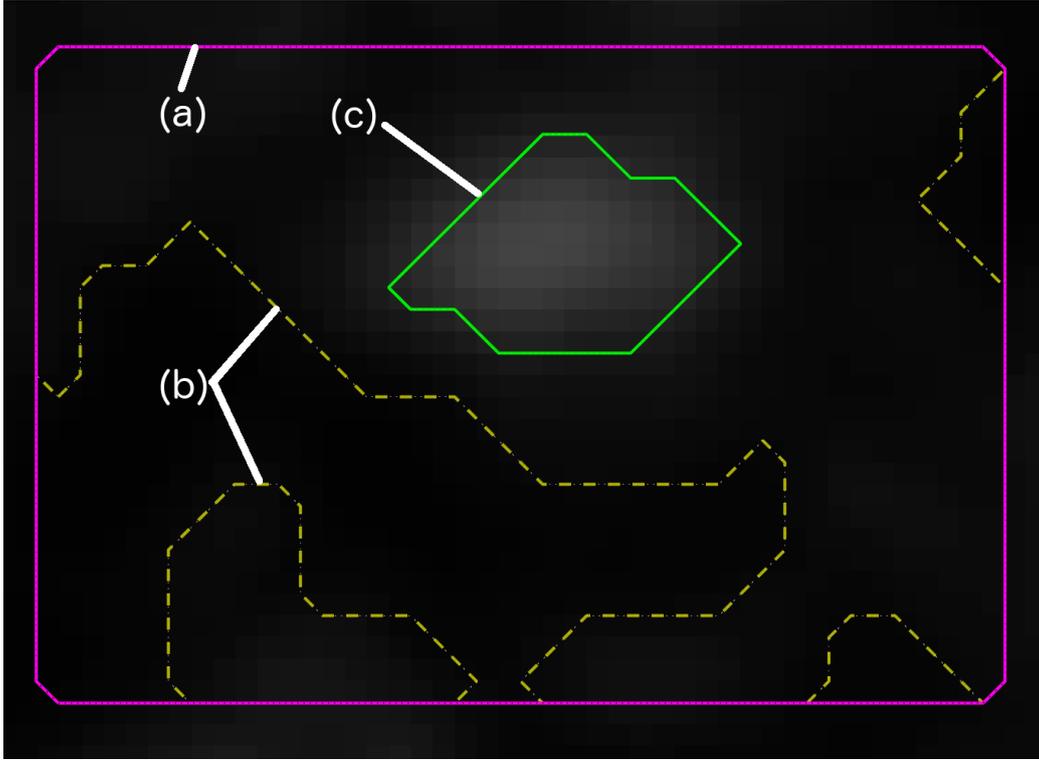


Figure 3.4: Metabolic tumour volume contours derived by PET automatic segmentation methods for a patient that was excluded from the study due to failed PET-based delineations. a) PET automatic segmentation method contour conforming to the boundary region and rejected from the study b) PET automatic segmentation method deriving multiple contours on a singular slice and rejected from the study c) PET automatic segmentation method contour deemed to have acceptable metabolic tumour volume representation and deemed acceptable for inclusion in this study

iteration of the algorithm until the change in volume between one iteration and the next is less than two voxels. The equation for the initial threshold is shown in Eq 3.1

$$T^i = (initThresh \times (SUV_{\max}^i - \mu(background^i))) + \mu(background)^i \quad (3.1)$$

Where *initThresh* is the initial threshold value to be applied to the PET image. In our implementation on iteration $i = 1$ the value is set at 0.45. On subsequent iterations ($i > 1$) an *initThresh* of 0.4 is applied. SUV_{\max} is the maximum SUV value within the region of interest on the PET image for iteration i and $\mu(\textit{background})$ is the mean (μ) uptake defined as the background within the PET image for iteration i . T^i is the final absolute threshold to apply to the PET image on the next iteration. The background uptake of the PET image is initially defined as voxels within the PET image with an intensity of $< 50\%$ of the SUV_{\max} within the image. These parameters were selected from a previous optimisation study performed by Berthon *et al* [77].

Watershed Transform contouring

The implementation of this method is based upon the description in the literature [70]. The algorithm uses Matlab's Sobel filter for the definition of the gradient map. WT finds the crests of the gradient map by simulating water flooding from selected local minima in the gradient map. The flooding of the gradient map continues until only one crest remains. This crest is defined as the contour.

Region Growing

RG is a procedure that groups voxels or subregions into larger regions based on predefined criteria [98]. In this project the implementation of the RG algorithm is based upon the method described by Day *et al* [68]. However, the inclusion criteria, stopping criteria and initial voxels for consideration are implemented differently. RG selects the SUV_{\max} as its seed point and grows a region from that voxel depending on the intensity of the voxel adjacent to it. In our implementation an initial region is grown that considers voxels $> 40\%$ of the SUV_{\max} . From this initial region, for each and every voxel the neighbouring voxels are considered for inclusion as tumour. The criterion for inclusion as tumour is based upon the μ tumour uptake and the SD from the μ . In our implementation voxels are included if they are within twice the SD of the μ . The μ tumour uptake is updated on each iteration. The region stops growing once there is $< 15\%$ change in the tumour volume size from one iteration to the next. These values were chosen due to good performance in a number of phantom scans acquired at PETIC for the optimisation of algorithms. Phantoms consisted of a variety of inserts with a μ volume [range] of 15.27 [0.64 — 80.63] (mL) and μ tumour to background ratios [range] of 10.49 [4.81 — 10.49].

Fuzzy Clustering

FCM was developed to account for the uncertainty arising at tumour boundaries in particular, by using a fuzzy classification instead of a binary one.

It was based on the work described by Belhassen *et al* [71]. In this case, each voxel is assigned a membership value for each cluster, ranging between 0 and 1. The membership value of a voxel x at iteration i is calculated as a probability to belong to the cluster k considered, according to the difference between the voxel intensity and the cluster μ intensity as shown in Eq 3.2. Where u_k^i is the probability of the voxel x belonging to cluster k on iteration i , $I(x)$ is the voxel intensity and M_k^i is the clusters μ intensity. A threshold is applied to resulting probability map to generate the resulting contour. The number of clusters used in this study is 2.

$$u_k^i(x) = \frac{\|I(x) - M_k^i\|}{\sum_j \|I(x) - M_j^i\|} \quad (3.2)$$

Gaussian Clustering

GCM is based on the FCM algorithm, with the difference that each cluster is assumed to have a Gaussian intensity distribution, of which μ and SD are calculated at each step. The cluster membership for each voxel is the probability of the voxel intensity value being generated by the cluster Gaussian distribution as shown in Eq 3.3. Where $(\sigma_k^i)^2$ is the variance of the intensities in cluster k at iteration i . The method is implemented based upon the description by Hatt *et al* [72]. The number of clusters used in this study are 3 and 4.

$$u_k^i(x) = \exp\left(-\frac{\|I(x) - M_k^i\|^2}{2(\sigma_k^i)^2}\right) \quad (3.3)$$

K-means clustering

KM assigns each voxel of the initial image to the cluster with a μ intensity value closest to its own value. This algorithm's implementation is based on the description by Zaidi *et al* [35]. However, it uses a customised initialisation criterion, considering a partition of the image intensity range into the number K of levels chosen by the user. The number of clusters used in this study are 2, 3 and 4.

3.1.6 Clinical Data & image analysis

Only primary tumours were analysed to ensure consistent methodology across all patients. Before quantitative image analysis and texture feature extraction, PET images were re-sampled into fixed bin widths of 0.5 SUVs. A fixed bin width maintains a constant intensity resolution when compared to approaches based on a fixed number of bins [99]. In the development of the prognostic models, age at diagnosis (number of years), radiological stage (stage IA—IV) and treatment (curative vs palliative) were included because these are strong predictors of survival [100]. Curative and palliative treatments were coded as 1 and 2 respectively. Radiological staging for the purposes of prognostic model development was represented categorically. A

patient with Stage IIA cancer was passed to the step-wise cox proportional hazards method as “Stage IIA”, the statistics package then automatically converts the patient’s staging into a numerical value. For example, “Stage IIA” is converted to 2.25 and Stage IIB is converted to 2.75. In cox proportional hazard methods, a reported variable’s hazard ratio is the proportion of how likely an event (death in this body of work) occurs in comparison to the proportion when all reported variables are zero. Therefore, a hazard ratio >1 in this body of work indicates a decrease in OS, whereas a hazard ratio <1 indicates an increase in OS.

Analysis of radiomic features (See Section 1.5) was performed using algorithms implemented as part of the Image Biomarker Standardisation Initiative (IBSI), a multicentre, international collaboration aimed at improving the reproducibility and validation of quantitative medical image analysis studies [15]. The radiomic features selected for inclusion in this study were chosen as they have shown prognostic and predictive significance in other radiomic studies investigating OC [95, 101, 102]. These have been summarised in Table 3.1. Moreover, many radiomic feature implementations have been described [92, 93, 101, 103]. For radiomic feature extraction in this chapter, the MTV was analysed as a three dimensional (3D) volume with no thresholding applied to the MTV mask.

First Order radiomic features

First order statistical metrics summarise the voxel intensity distribution within the segmented MTV, without concern for spatial relationships [104].

First order metrics are typically histogram based and reduce the MTV to singular values describing the μ , minimum, maximum, median, uniformity of the intensities within the MTV. Included in first order statistical analysis is Skewness (asymmetry measure), Kurtosis (pointiness measure) and Entropy (randomness measure). Kurtosis and skewness have been shown to be independent predictors of survival [16], and of prognostic significance in the literature [105].

Higher Order radiomic features

Higher order statistical metrics retain spatial information and are used to quantify inter-voxel intensity relationships. Dissimilarity is the quantification of variation in voxel pairs and is calculated using a Grey Level Co-occurrence Matrix (GLCM) generated for each unique direction and averaged. A low dissimilarity is a result of neighboring voxels having similar values [106]. Zone percentage is calculated from a Grey Level Size Zone matrix (GLSZM) by assessing the fraction of recorded zones compared to the maximum number of possible zones. Heterogeneous MTVs have high zone percentage scores. Grey Level Non-Uniformity (GLNU) is an evaluation of the distribution of zone counts for each intensity value. The feature value is low when the number of zones associated with each intensity value are similar. Coarseness is a Neighborhood Grey-Tone Difference Matrix (NGTDM) feature that gives an indication of the level of spatial rate of change in intensity [107]. GLCM, GLSZM, NGTDM can be computed in 2D or 3D. The matrices in this study were computed in 3D as this may highlight the multi-scale, directional prop-

erties of tumour tissue [108].

Patient's outcome and overall survival data

The primary outcome in this patient cohort was OS, defined as number of months survived from date of diagnosis. Patients were followed-up 3-monthly for the first year, 6-monthly until 5 years then annually thereafter, or until death. All of the included patients were followed-up for at least 12 months. A patient's date of death obtained from the Cancer Network Information System Cymru database (CaNISC, Velindre NHS Trust, Wales).

Statistical Analysis

Categorical variables were described as frequency (percent) and continuous variables as median (range) and differences assessed with appropriate non-parametric tests. Cumulative survival was calculated by the Kaplan-Meier life-table method. Cox regression models with backward conditional method were constructed using identical clinical data and imaging data derived from each of the segmentation methods. An individual prognostic score was calculated from each segmentation method by summation of the products of variables and their corresponding parameter estimate. Using this score, patients were separated into low, intermediate and high-risk groups (higher prognostic score deemed higher risk) and a log-rank test evaluated significant differences in OS. The number of patients that changed risk stratification group depending on the segmentation method used was calculated, and the OS for the different risk groups between models was analysed. The Akaike

information criterion (AIC) statistic evaluated the estimated quality of the models [109]. Given a set of developed models, AIC estimates the quality of a developed model from other developed models by estimating the information lost by the model being assessed. The model with the lowest AIC value and therefore the lowest information loss, is considered the better model. AIC was chosen for the comparison of developed models, instead of the Bayesian information criterion (BIC), due to being designed to select predictive models, whereas BIC is designed to find the true (known) model that represents the data. A p-value of <0.05 was considered statistically significant. Statistical analysis was performed using SAS version 9.4 (SAS, North Carolina, USA) and SPSS version 23.0 (IBM, Chicago, USA).

3.2 Results

The study in this chapter involved four hundred and seventy-two patients, each with 9 MTV contours delineated by AT, FCM2, GCM3, GCM4, KM2, KM3, KM4, RG and WT PET-AS methods. The contours were assessed by a clinical radiologist for accurate tumour representation; after which, forty-five patients and 5 segmentation methods were excluded due to poor MTV delineation. FCM2 failed to delineate an acceptable tumour representation in 145 (30.8%) of cases. KM3 and KM4 failed in 88 (18.6%) and 215 (45.6%) of cases, respectively. RG failed in 389 (82.5%), and GCM4 in 33 (7%) of cases. Therefore, 427 cases with MTVs delineated with KM2, GCM3, AT and WT PET-AS methods were deemed to have accurate tumour representation and

included for further analysis. The 427 cases included for analysis were used to develop the prognostic models for KM2, WT, GCM3 and AT methods. Baseline characteristics of patients are detailed in Table 3.2. The median OS of the cohort was 17.0 months (95% confidence interval (95% CI) 14.8 — 19.2). Median follow-up was 35.0 months (95% CI 28.7 — 41.3). Overall 1— and 2— year survival in the development cohort was 65.3% and 30.1%, respectively.

3.2.1 Development of Prognostic Models

The final steps of each prognostic model are presented in Table 3.3. Three known clinical prognostic factors (age, radiological stage and treatment) remained in each derived model, but there was a difference in the inclusion of texture metrics by segmentation technique. AT and KM2 produced the same model output. In comparison to a study published by Foley *et al* [16], in this chapter radiomic features were not included in the final models for these segmentation methods. However, in this body of work radiomic features were implemented according to IBSI specifications. Additionally, the radiomic features skewness and kurtosis were found to be independently significant for survival using the GCM3 method. Skewness and GLNU were significant using the WT method. Their inclusion in the models illustrates their additional prognostic value compared with current prognostic factors.

3.2.2 Prognostic Score Calculation

The equations for each model, derived from different segmentation methods, were used to calculate the prognostic scores, and are listed in Table 3.4. These calculations were derived using published methods [110].

Figure 3.5 shows the risk stratification (see Section 1.5) for WT, KM2, AT, and GCM3. In Figure 3.5, for the AT and KM2 derived prognostic models, the median OS for the low risk, intermediate risk and high-risk groups was 36.0 months (29.9 — 42.1 months), 18.0 months (15.1 — 20.9 months) and 9.0 months (7.8 — 10.2 months) respectively. In comparison, the median OS for the low risk, intermediate risk and high-risk group in the GCM3 derived prognostic model was 36.0 months (28.8 — 43.2 months), 18.0 months (15.4 — 20.6 months) and 9.0 months (7.7 — 19.2 months). Additionally, in Figure 3.5, the median OS for the WT derived prognostic model low risk, intermediate and high-risk groups was 36 months (27.8 — 44.2 months), 19 months (15.1 — 23 months) and 9 months (7.7 — 10.3 months) respectively.

Table 3.5 shows the number of patients stratified as low, intermediate and high risk for each single prognostic model along with the prognostic score range for each risk stratification group. There was no significant difference in the number of patients stratified as low risk between the prognostic models ($P = 1$). Additionally, no significant difference was found between the number of patients stratified as intermediate risk or high risk ($P = 1$). Therefore, each of the developed prognostic models stratified the patients equally. However, Table 3.6 shows the number of patients who change risk stratification

when risk stratified from the prognostic models. The largest proportion of patients to change risk stratification group was between prognostic models based on GCM3 and on WT (n=73, 17.1%). Patients that change risk stratification have the potential to receive different treatments. However, it can be noted that no patient changed risk stratification group between AT and KM2 because the prognostic models were identical. Whilst patients were found to change risk stratification, there was no overall survival difference between AT, GCM3, KM2 or WT low-risk groups ($\chi^2 = 0.052$, $df = 3$, $p = 0.997$), intermediate-risk groups ($\chi^2 = 0.016$, $df = 3$, $p = 0.999$) or high-risk groups ($\chi^2 = 0.028$, $df = 3$, $p = 0.999$). To estimate the quality and performance of the developed prognostic model, the AIC statistic was calculated, for which the lowest score assumes a higher quality model due to less information loss. The AIC statistics for GCM3, WT and AT/KM2 was 3044, 3048 and 3052 respectively. Therefore, the AIC statistic suggests that GCM3 has the best estimated model performance.

Prognostic models were also developed using the segmentation methods FCM2, GCM4, KM3, KM4 and RG. These segmentation methods were excluded from the main study because the assessing radiologist deemed the contours to not be representative in less than 90% of cases. However, the models are included here for completeness. The final steps of each prognostic model are presented in Table 3.7. Whilst the contours produced using AT and KM2 were considered by the radiologist to be acceptable, the contours produced by the FCM2, KM4 and RG PET-AS methods were considered to be not acceptable. However, the final steps in these prognostic models only used

the clinical variables age, treatment and stage. Additionally, the parameter estimate and hazard ratio for each of the clinical variables was the same. This is demonstrated in Table 3.8, where the equations for each excluded model are presented.

Table 3.9 outlines the summary statistics of median OS (95% CI) for GCM4, FCM2, KM3, KM4 and RG PET-AS methods and Figure 3.6 shows the OS for prognostic models developed from PET-AS methods excluded from the study. The median OS for the low risk, intermediate and high risk groups for the models developed from the excluded PET-AS methods was 36, 18 and 9 months respectively. In addition, Table 3.10 shows the number of patients in each risk stratification group for each of the developed models. There was no significant difference in the number of patients between each risk stratification group ($P = 1$). Additionally for the excluded PET-AS methods, in comparison to the acceptable PET-AS methods, there was no significant difference in the number of patients in each risk stratification group ($P = 1$). Table 3.11 shows the total number of patients and percentage that change risk-stratification group between each prognostic model.

3.3 Discussion

The aim of this chapter was to highlight the need for a standardised MTV delineation method in the development of prognostic models as well as to investigate the impact of MTV delineation on patient risk stratification and therefore a patient's OS. In order to achieve this, first, second and higher-

order radiomic features were extracted from 9 differing PET-AS delineations. The results presented in this chapter demonstrate that the significant variables in the developed prognostic models are dependent upon the MTV delineation method. For example, in the prognostic model developed from the PET-AS method GCM3, the first-order radiomic features Kurtosis and Skewness were found to be significant predictors of a patient's OS. In comparison, in the prognostic model developed from the WT PET-AS method, the higher-order feature GLNU was found to be of significance, as well as Skewness. Whilst radiomic features were significant in the development of the GCM3 and WT based prognostic models, in the AT and KM2 PET-AS method based prognostic models, radiomic features were found not to be significant predictors of a patient's OS. This is in addition to the known clinical predictors, which include a patient's TNM staging and age. Therefore, the results presented in this chapter highlight the dependency of significant PET radiomic variables on the MTV delineation method.

For each unique prognostic model, patients were also found to change risk stratification group and the AIC statistic for each model suggested that the PET-AS method GCM3 has the best prognostic performance, in comparison to the WT and AT/KM2 developed models. The small difference in the AIC statistic between the developed prognostic models, suggests that there is a relatively small difference in the effect of the different PET-AS methods on patient risk stratification between each of the developed prognostic models.

Nine PET-AS delineation methods were used to derive MTVs in this chapter

and the delineated contours were reviewed by a radiologist. After assessment, 5 of the PET-AS methods were excluded from analysis due to poor tumour representation in the majority of cases. However, upon investigation the prognostic models developed by the excluded PET-AS methods (FCM2, KM3, KM4 and RG) developed identical prognostic models to the PET-AS methods AT and KM2. Therefore, whilst radiomic features which are considered significant and associated with a patient's OS are dependent upon the delineation method, this is potentially unrelated to the PET-AS method contours considered acceptable by a radiologist. The presented results demonstrate that patients are potentially assigned to different risk stratification groups depending on the MTV delineation method. This could be significant as sub-groups of patients have the potential to receive more aggressive treatments than is necessary. Therefore, patients have the potential to have a decreased quality of life. In reverse, this has the potential for patients to be denied beneficial treatment.

Morphological features, which describe a tumour's shape, can also be quantified from the MTV. Within this group of radiomic features, the surface to volume ratio, compactness, sphericity and disproportion of the tumour can be characterised. Morphological features were not included in this chapter because the experiment focused upon radiomic features which are potential biomarkers for intra-tumoural heterogeneity. However, studies have investigated the inclusion of shape metrics in prognostic models [111]. The variability in PET-AS performance in any one single clinical case means the standardisation of the delineation of the MTV is critical for the application of

radiomics within OC, especially as the findings in this chapter demonstrate the potential impact of different MTV delineation methods on the development of prognostic models, even when using standardised implementations of radiomic features.

This supports the recommendations of the IAEA [31] and the AAPM Task Group No 211 [1] as they independently reported that they could not recommend a single PET-AS method for MTV delineation. Therefore, an agreed upon PET-AS method should be used for MTV delineation when developing and using prognostic models.

3.4 Conclusion

Prognostic models incorporating quantitative image features are dependent on the method used to delineate the primary tumour. This has a subsequent effect on risk stratification, with patients changing groups depending on the PET-AS method used to delineate the MTV. The findings of this study may have substantial potential impact on clinical management of patients with OC and were published in EJNMMI Res (2018) [38]. This work has shown that the standardisation of PET segmentation is important and should be considered in future prognostic and predictive clinical models. The following chapter aims to externally validate the ATLAAS segmentation methodology, which has been proposed for the standardised delineation of the MTV for RT planning [36].

Table 3.1: Summary of quantitative imaging features

Type / Order of statistics	Feature	Brief Definition
Morphological	Volume	Sum of voxels delineated multiplied by the volume of one voxel
Pre-discretisation	SUVmax	Maximum uptake of FDG in the MTV
	Energy	Sum squared SUV values in the MTV
First order	Skewness	Measures symmetry of intensity histogram
	Kurtosis	Measures flatness of intensity histogram
	Entropy	Measures randomness
Second order	Dissimilarity	Variation of grey level pairs (GLCM). Features were calculated for each unique direction and averaged with a distance setting of 1.
Higher order	Grey-level Non-uniformity	Distribution of zone counts for each intensity value (GLSZM)
	Zone Percentage	Fraction of recorded zones compared to maximum possible
	Coarseness	Measures spatial rate of change in intensity using a distance of 1.

Table 3.2: Baseline characteristics of patient cohort

Median age	67.0 year (range 24 — 84)
Gender	Male 315 (73.8% of included patients): Female 112 (26.2% of included patients)
Histology	number of patients (% of total number patients)
Adenocarcinoma	313 (73.3%)
Squamous Cell Carcinoma	100 (23.4%)
Undifferentiated	5 (1.2%)
High-grade Dysplasia	4 (0.9%)
Neuro-endocrine	3 (0.7%)
Small Cell Carcinoma	1 (0.2%)
Sarcoma	1 (0.2%)
Tumour Location	number of patients (% of total number patients)
Oesophagus	268 (62.8%)
Upper third	14 (5.2%)
Middle third	71 (26.5%)
Lower Third	183 (68.3%)
Gastro-oesophageal junction	159 (37.2%)
Siewert I	67 (42.1%)
Siewert II	42 (26.4%)
Siewert III	50 (31.4%)
Stage Group	number of patients (% of total number patients)
IA	10 (2.3%)
IB	17 (4.0%)
IIA	70 (16.4%)
IIB	13 (3.0%)
IIIA	97 (22.7%)
IIIB	52 (12.2%)
IIIC	76 (17.8%)
IV	92 (21.5%)
Treatment	number of patients (% of total number patients)
Curative	224 (52.5%)
NACT	86 (38.4%)
dCRT	86 (38.4%)
Surgery Alone	31 (13.8%)
NACRT	20 (8.9%)
EMR	1 (0.5%)
Palliative	203 (47.5%)
Mortality	number of patients (% of total number patients)
Alive	132 (30.9%)
Dead	295 (69.1%)

Table 3.3: Final Output of Prognostic Models Derived Using AT, GCM3, KM2 and WT PET Segmentation Methods

AT	Parameter Estimate	p-value	Hazard Ratio	95% CI
Age	0.020	0.001	1.020	1.008 — 1.033
Treatment	-1.075	<0.001	0.341	0.254 — 0.459
Stage	0.144	<0.001	1.155	1.072 — 1.245
GCM3	Parameter Estimate	p-value	Hazard Ratio	95% CI
Age	0.019	0.003	1.019	1.006 — 1.032
Treatment	-1.024	<0.001	0.359	0.266 — 0.485
Stage	0.142	<0.001	1.153	1.068 — 1.245
Kurtosis	0.632	0.002	1.882	1.260 — 2.809
Skewness	-0.789	0.044	0.454	0.211 — 0.980
KM2	Parameter Estimate	p-value	Hazard Ratio	95% CI
Age	0.020	0.001	1.020	1.008 — 1.033
Treatment	-1.075	<0.001	0.341	0.254 — 0.459
Stage	0.144	<0.001	1.155	1.072 — 1.245
WT	Parameter Estimate	p-value	Hazard Ratio	95% CI
Age	0.018	0.004	1.018	1.006 — 1.031
Treatment	-1.063	<0.001	0.345	0.257 — 0.464
Stage	0.140	<0.001	1.150	1.065 — 1.242
GLNU	0.017	0.006	1.017	1.005 — 1.029
Skewness	0.674	0.030	1.962	1.067 — 3.607

Table 3.4: Prognostic model equations

Segmentation Method	Prognostic Model Equation
AT	$(Age \times 0.020) - (Treatment \times 1.075) + (Stage \times 0.144)$
GCM3	$(Age \times 0.019) - (Treatment \times 1.024) + (Stage \times 0.142) - (Skewness \times 0.789) + (Kurtosis \times 0.632)$
KM2	$(Age \times 0.020) - (Treatment \times 1.075) + (Stage \times 0.144)$
WT	$(Age \times 0.018) - (Treatment \times 1.063) + (Stage \times 0.140) + (Skewness \times 0.674) + (GLNU \times 0.017)$

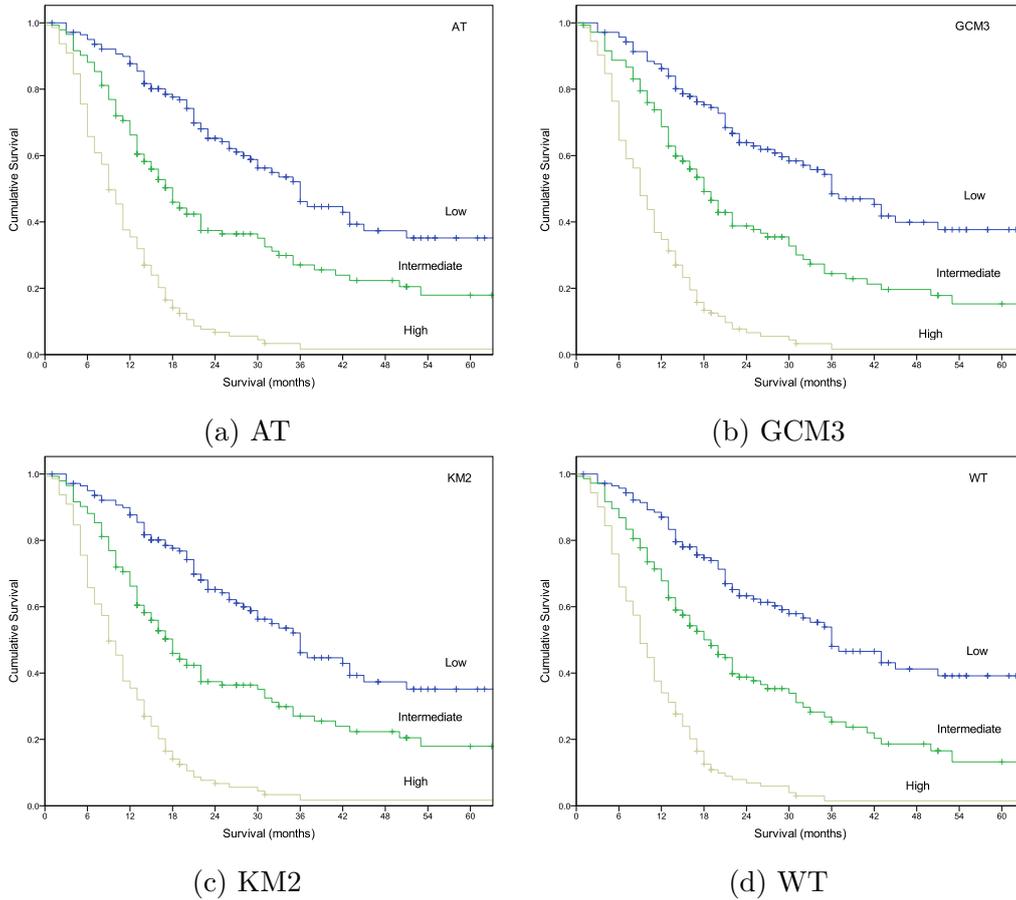


Figure 3.5: Patient risk stratification for AT, KM2, GCM3 and WT PET-AS methods. The median overall survival for the low risk groups in the models developed from AT, GCM3, KM2 and WT was 36 months. The median overall survival for the intermediate risk groups in the models developed from AT, GCM3 and KM2 was 18 months. However, the median overall survival for WT intermediate risk group was 19 months. The median overall survival for the high risk groups in the models developed from AT, GCM3, KM2 and WT was 9 months.

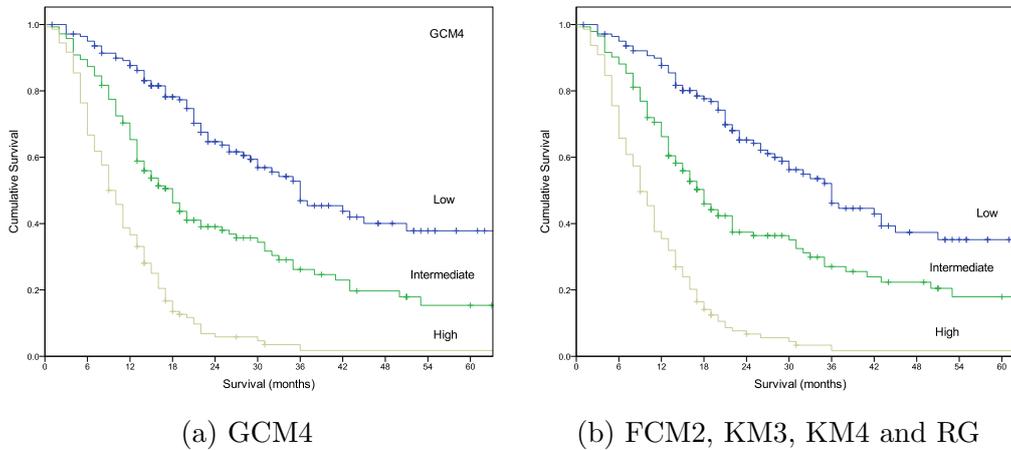


Figure 3.6: Patient risk stratification for excluded PET-AS methods. The median overall survival for the low risk group in the GCM4 derived model was 36 months (27.4 — 44.6) and 36 months (29.9 — 42.1) in the models derived by FCM2, KM3, KM4 and RG. The median overall survival for the intermediate risk group in the GCM4 derived model was 18 months (14.6 — 21.4) and 18 months (15.1 — 20.9) in the models derived by FCM2, KM3, KM4 and RG. The median overall survival for the high risk group in the GCM4 derived model was 9 months (7.7 — 10.3) and 9 months (7.8 — 10.2) in the models derived by FCM2, KM3, KM4 and RG.

Table 3.5: Number of patients in each risk stratification group for each single prognostic model, with the prognostic score range given in brackets

number in risk group	Low Risk (prognostic score range)	Intermediate Risk	High Risk
AT\KM2	141 (-0.45 — 0.98)	143 (0.99 — 2.16)	143 (2.17 — 2.79)
GCM3	140 (-1.13 — 0.36)	143 (0.37 — 1.54)	144 (1.55 — 2.73)
WT	142 (-0.17 — 1.30)	144 (1.31 — 2.48)	141 (2.49 — 3.62)

Table 3.6: Total number of patients and percentage that change risk-stratification group

Number changing group (%)	AT	GCM3	KM2	WT
AT				
GCM3	66 (15.4)			
KM2	0 (0.0)	66 (15.4)		
WT	57 (13.3)	73 (17.1)	57 (13.3)	

Table 3.7: Final Output of Prognostic Models Derived Using FCM2, GCM4, KM3, KM4 and RG PET-AS methods

FCM2	Parameter Estimate	p-value	Hazard Ratio	95% CI
Age	0.020	0.001	1.020	1.008 — 1.033
Treatment	-1.075	<0.001	0.341	0.254 — 0.459
Stage	0.144	<0.001	1.155	1.072 — 1.245
GCM4	Parameter Estimate	p-value	Hazard Ratio	95% CI
Age	0.020	0.001	1.021	1.008 — 1.033
Treatment	-1.055	<0.001	0.348	0.259 — 0.469
Stage	0.159	<0.001	1.172	1.086 — 1.265
Kurtosis	0.207	0.016	1.230	1.039 — 1.455
KM3	Parameter Estimate	p-value	Hazard Ratio	95% CI
Age	0.020	0.001	1.020	1.008 — 1.033
Treatment	-1.075	<0.001	0.341	0.254 — 0.459
Stage	0.144	<0.001	1.155	1.072 — 1.245
KM4	Parameter Estimate	p-value	Hazard Ratio	95% CI
Age	0.020	0.001	1.020	1.008 — 1.033
Treatment	-1.075	<0.001	0.341	0.254 — 0.459
Stage	0.144	<0.001	1.155	1.072 — 1.245
RG	Parameter Estimate	p-value	Hazard Ratio	95% CI
Age	0.020	0.001	1.020	1.008 — 1.033
Treatment	-1.075	<0.001	0.341	0.254 — 0.459
Stage	0.144	<0.001	1.155	1.072 — 1.245

Table 3.8: Prognostic model equations developed from excluded PET-AS methods

Segmentation Method	Prognostic Model Equation
FCM2	$(Age \times 0.020) - (Treatment \times 1.075) + (Stage \times 0.144)$
GCM4	$(Age \times 0.020) - (Treatment \times 1.055) + (Stage \times 0.159) + (Kurtosis \times 0.207)$
KM3	$(Age \times 0.020) - (Treatment \times 1.075) + (Stage \times 0.144)$
KM4	$(Age \times 0.020) - (Treatment \times 1.075) + (Stage \times 0.144)$
RG	$(Age \times 0.020) - (Treatment \times 1.075) + (Stage \times 0.144)$

Table 3.9: Summary statistics of median OS (95% CI) for GCM4, FCM2, KM3, KM4 and RG

95% Confidence Intervals				
Segmentation Method	Risk Group	Median OS	Lower	Upper
GCM4	Low	36.0	27.4	44.6
	Intermediate	18.0	14.6	21.4
	High	9.0	7.7	10.3
FCM2, KM3, KM4, RG	Low	36.0	29.9	42.1
	Intermediate	18.0	15.1	20.9
	High	9.0	7.8	10.2

Table 3.10: Number of patients in each risk stratification group, with the prognostic range in brackets, for GCM4, FCM2, KM3 and RG developed prognostic models

number in risk group	Low Risk (prognostic range)	Intermediate Risk	High Risk
FCM2\KM3\ KM4\RG	141 (-0.45 — 0.98)	143 (0.99 — 2.16)	143 (2.17 — 2.79)
GCM4	141 (-0.13 — 1.02)	142 (1.03 — 2.19)	144 (2.20 — 3.14)

Table 3.11: Total number of patients and percentage that change risk-stratification group for excluded PET-AS methods

Number changing group (%)	AT\FCM2 KM2\KM3 KM4\RG	GCM3	GCM4	WT
GCM3	66 (15.4)			
GCM4	51 (11.9)	65 (15.2)		
WT	57 (13.3)	73 (17.1)	68 (15.9)	

Chapter 4

Automated segmentation of the metabolic tumour volume in low tumour to background ratio PET

H&N cancer is the sixth most common tumour worldwide [3], with rates of incidence in the UK increasing by 30% since the early 1990s [5]. RT is often used to treat Oropharyngeal (OP) SCC and ^{18}F -FDG PET aids MTV delineation in RT planning. Typically, RT planning is performed on CT imaging acquired before RT; however there is increasing interest in multi-modality treatment pathways, as induction chemotherapy (ICT) before RT can lead to tumour downstaging and a reduced MTV [112]. Therefore, RT planning and MTV delineation on pre-ICT PET imaging potentially means the de-

lineated tumour shape, size and volume may be misrepresented when RT treatment starts, leading to inaccuracies in RT planning, which can result in an increased dose to the OAR. Further, MTV delineation on PET imaging acquired after ICT is challenging due to the reduced metabolic activity and MTV. The aim of the study, presented in this chapter therefore, was to evaluate PET-AS methods including the machine-learned method ATLAAS for MTV delineation on PET imaging after ICT. The following section describes the materials & methods used to achieve this aim.

4.1 Materials and Methods

4.1.1 Clinical Data

The patients analysed in this chapter were identified from a phase I, multi-centre, feasibility trial called 18F-FDG-PET Guided Dose-Painting With Intensity Modulated Radiotherapy in Oropharyngeal Tumours (FiGaRO) [113] conducted in the following centres:

- Guy's and St Thomas' Hospitals (London,UK)
- Velindre Cancer Centre (Cardiff, UK)

It investigated dose escalation of residual metabolic uptake on ^{18}F -FDG PET imaging acquired following 1 cycle of ICT, in patients with primary OP SCC, with the aim of improving tumour control rates whilst delivering acceptable toxicity levels. ICT within the trial consisted of a combination of Cisplatin and 5-Fluorouracil (5FU) and ethical approval for the trial, by the research

ethics committee, was granted in July 2012 (REC: 12/LO/1724). All patients provided written informed consent for inclusion in the trial and for research purposes.

Twenty-three patients were enrolled on the trial between October 2013 and March 2017 and were excluded from the trial if they had previously received RT treatment to the H&N region, had a previous malignancy except for non-melanoma skin cancer or had previous/concurrent illness. In total, three patients were excluded from the trial. One patient was excluded from the trial because, after delivery of one cycle of ICT, the residual MTV was considered too small for effective dose escalation. A further two patients were also excluded from the study due to technical and unrelated medical problems. Therefore, twenty patients proceeded to have dose-escalated IMRT. Analysis of the MTVs delineated for planning IMRT showed one PET scan had two disjointed MTVs. Therefore, twenty-one separate MTVs were available for analysis in this chapter.

Patients recruited to the trial had histologically confirmed OP SCC, assessed as either HPV negative by p16 immunohistochemistry and in-situ hybridization for high-risk subtype DNA, or intermediate or high risk HPV positive. Patients were also defined as having a greater than 10 pack/year smoking history, were over 18 years old, staged with at least T2 tumours and had advanced N stage (TNM v7 N2b, N2c, N3) [9]. TNM staging is defined in Section 1.5. All patients were planned for treatment with 2 cycles of ICT followed by primary radical IMRT to the primary and bilateral neck nodes. Radical IMRT occurred concurrently with Cisplatin chemotherapy (chemo-

IMRT). The pathway for patients recruited to the FiGaRO trial is shown in Figure 4.1. Patients were referred from Oncology before consenting to the study and then undergoing the 1st cycle of ICT. Three weeks after the 1st cycle of ICT, PET/CT imaging was acquired for RT planning. The following section describes the PET/CT imaging protocol for each patient.

4.1.2 ^{18}F -FDG PET/CT imaging

For planning IMRT, ^{18}F -FDG PET/CT imaging was acquired in the treatment position using a H&N immobilisation shell. Figure 4.2 shows a typical RT immobilisation shell applied to a dummy patient. The shells are typically made using a mesh structure with lots of small holes and are designed to keep the area that they are applied to completely still. RT masks also ensure repeated (fractionated) RT deliveries are delivered to the same region each time. This is because the RT mask can be marked and then aligned with lasers on the linacs. In addition to acquisition of the planning ^{18}F -FDG PET imaging, a contrast enhanced CT scan and a low dose CT scan were acquired. Figure 4.3 shows an example of a low dose CT scan and contrast enhanced CT scan. A patient who has a low dose CT scan receives a lower radiation dose compared to a normal CT scan. Therefore, a low dose CT scan is typically used for the diagnosis of cancer as patients can receive multiple scans for the same level of radiation exposure as a normal dose CT scan. The low dose CT scan was used for attenuation correction of PET imaging only. All imaging acquired for planning IMRT purposes was acquired 3 weeks following the first cycle of ICT.

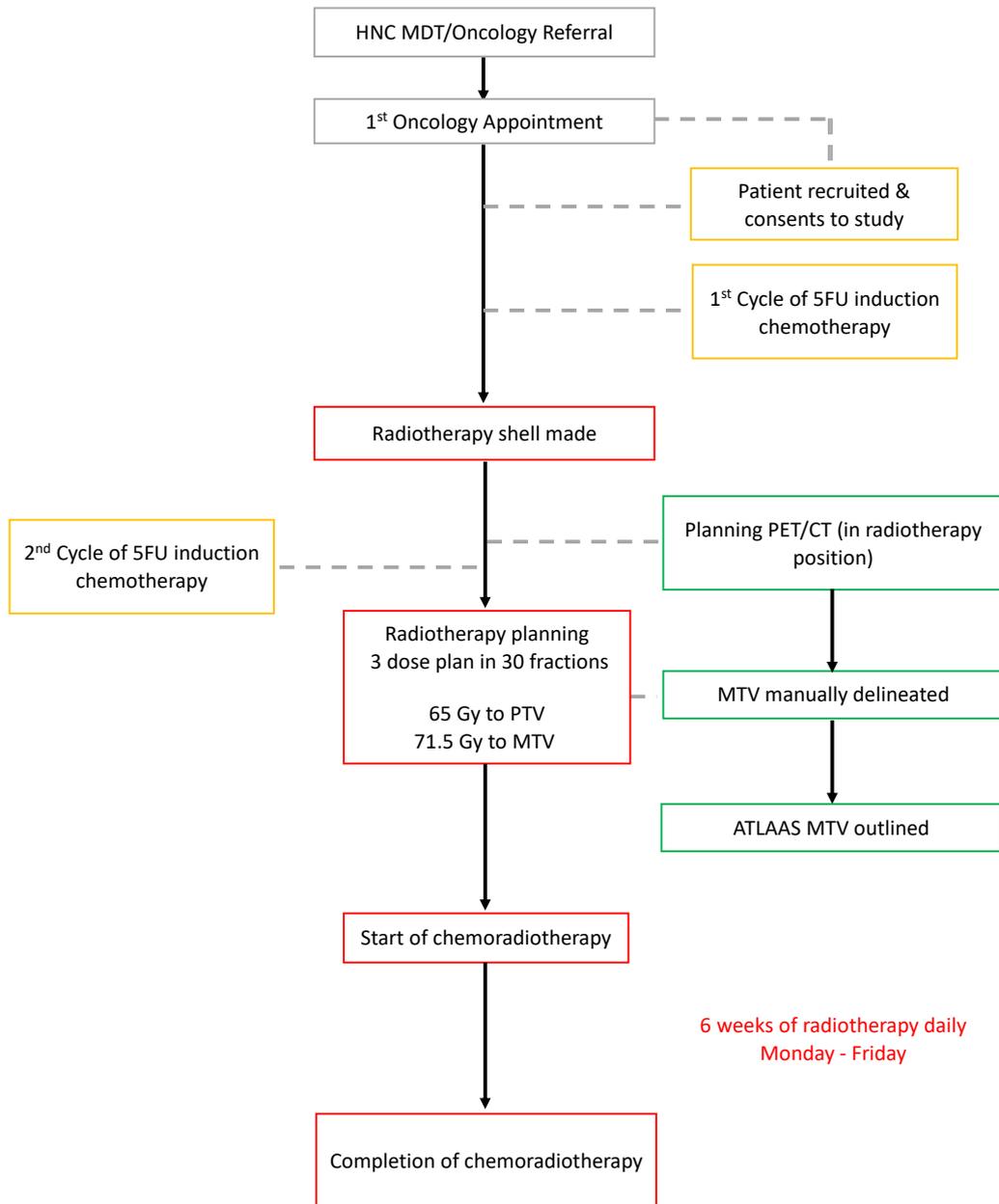
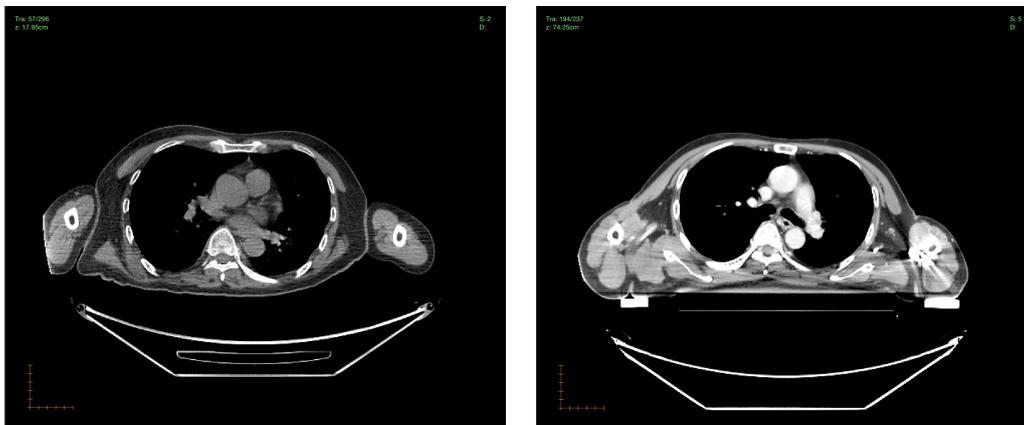


Figure 4.1: Pathway for patients recruited to the FiGaRo trial



Figure 4.2: RT immobilisation shells are used during RT treatment in order to minimise movement from a patient and therefore ensure RT is delivered with a high-degree of accuracy. Immobilisation shells are typically made using a mesh structure. Image reproduced courtesy of Cancer Research UK [114].



(a) Low dose CT scan

(b) Contrast enhanced CT scan

Figure 4.3: (a) Low dose CT scan acquired for attenuation correction of PET imaging used for radiotherapy planning in the FiGaRo trial. (b) Contrast enhanced CT scan acquired after acquisition of PET imaging. Low dose CT scans are acquired with a lower x-ray tube current causing patients to receive a lower radiation dose. Therefore, they are typically used diagnostically as patients can undertake multiple scans for the same level of radiation exposure as a normal dose CT scan.



Figure 4.4: A GE 710 PET/CT scanner [115].

Before PET/CT scanning with either a GE Discovery 710 situated in London or a GE Discovery 690 situated in Cardiff, patients were injected with $350 \pm 10\%$ MBq of ^{18}F -FDG and rested for 90 minutes. Figure 3.1 in Chapter 3 shows a GE 690 PET scanner and Figure 4.4 shows a GE 710 PET scanner. All of the PET images were acquired with a field of view of 700 mm and a matrix size of 256×256 voxels. The resulting PET image voxel dimensions were 2.73×2.73 mm with a slice thickness of 3.27 mm. Further, all of the PET images were acquired using a total of 3 bed positions at 4 minutes per bed position and TOF data, which was used during PET reconstruction. The acquired PET data was reconstructed using OSEM, with 2 iterations and 24 subsets, and a 6.4 mm Full Width at Half Maximum (FWHM) Gaussian post filter was applied to the images after reconstruction. RT planning requires the delineation of the target tumour volume, in the form of the MTV on PET imaging. For the patients included in the FiGaRO trial, the primary MTV for IMRT planning was delineated manually on the PET imaging acquired after one cycle of ICT, according to the process described in the following section.

4.1.3 Manual MTV delineation after 1 cycle of chemotherapy

Manual MTV PET delineation in this study was performed by a nuclear medicine physician and a clinical oncologist jointly, with differences in delineation resolved by consensus; multi-disciplinary approaches have been shown to reduce intra and inter-observer variability [116]. Additionally, it is currently recommended to use fixed windowing and viewing levels, when delineating the MTV [117,118], in order to further reduce intra and inter-observer variability. Therefore, PET images were displayed in SUV and visualised using an inverse linear colour scale with a fixed windowing level and width. All PET data was scaled to a SUV range of 0 to 10. The manually delineated PET MTV was used for RT planning and as the reference MTV within this chapter. MTV delineation was performed using Hermes Hybrid Viewer (Hermes Medical Solutions, Sweden) versions 2.2C and 2.6H at Guys & St Thomas PET Centre and Velocity AI version 2.7 (Varian Medical Systems, Palo Alto, USA) and ProSoma version 3.2 (OSL Oncology Systems Limited, UK) at Velindre Cancer Centre.

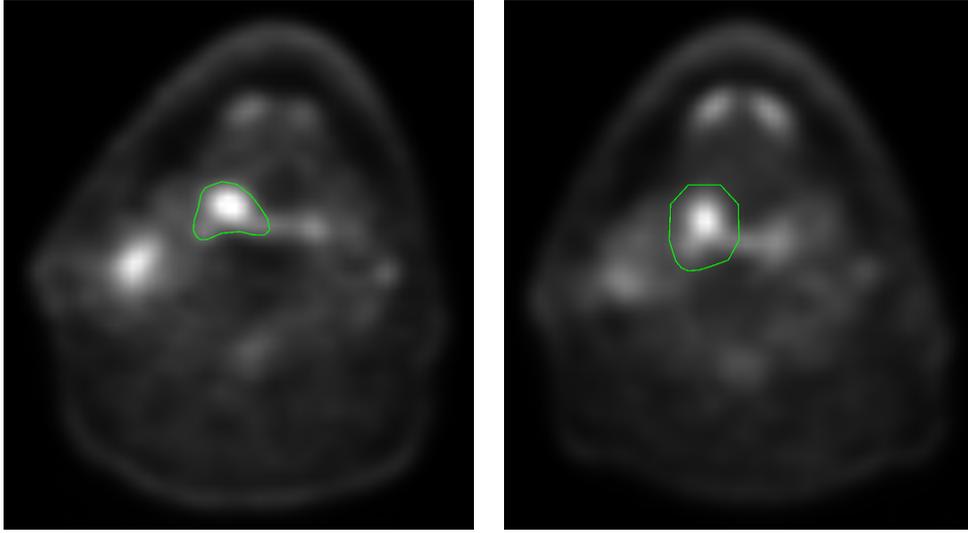
4.1.4 Development of ATLAAS_{ICT}

ATLAAS has been described in detail in Chapter 2 Section 2.4 [65]. The statistical model for ATLAAS was originally developed on pre-treatment H&N ¹⁸F-FDG PET imaging data using the following image and tumour parameters: TBR, MTV and NI. In this chapter, this training dataset is

known as ATLAAS_{ORIG} and therefore, as manual delineation of the MTV and IMRT planning was performed using PET imaging acquired after one cycle of ICT, we compared the characteristics of the PET imaging and MTVs obtained on pre-ICT and post-ICT PET imaging, in order to understand the effect of one cycle of ICT on primary tumour volumes.

Comparison of pre-ICT and post-ICT PET imaging

To standardise the comparison of MTVs on pre- and post-ICT PET imaging, contours were delineated using 42% SUV_{peak} thresholding. Figure 4.5a and Figure 4.5b compares the MTV contours derived by 42% SUV_{peak} thresholding on PET images acquired for diagnosis and RT planning purposes. The 42% SUV_{peak} MTV on PET imaging acquired for diagnostic purposes in Figure 4.5a had a SUV_{max} of 6.69 and SUV_{peak} of 5.03 with a volume of 8.76 mL. In comparison, the 42% SUV_{peak} derived MTV on PET imaging acquired for RT planning in Figure 4.5b had a SUV_{max} of 7.11 and SUV_{peak} of 5.09 with a volume of 10.01 mL. The increased MTV derived on Figure 4.5b is due to the decreased metabolic activity of the MTV, as the appropriate threshold to delineate the MTV is dependent upon the SUV_{max} and SUV_{peak} [32]. However, Table 4.1 shows that ICT reduced the MTV, SUV_{max} , TBR and NI values when contoured using 42% SUV_{peak} thresholding. The results presented demonstrate that there were significant differences in the characteristics of pre-ICT PET imaging, which ATLAAS was trained upon, and the characteristics of post-ICT PET imaging; therefore, simulation and development of new PET scans with characteristics of post-ICT PET scans



(a) 42 % SUV peak thresholding on pre-ICT PET imaging (b) 42 % SUV peak thresholding on post-ICT PET imaging

Figure 4.5: Pre and Post-ICT PET imaging

was required for accurate MTV delineation.

Table 4.1: Mean [Range] MTV, TBR, SUV_{MAX} and NI on ^{18}F -FDG PET imaging acquired before and after ICT when contoured using 42% SUV_{PEAK} fixed thresholding

Parameter	Before ICT ^{18}F -FDG PET	After ICT ^{18}F -FDG PET
Mean MTV (mL)	9.67 [2.79 — 36.18]	7.43 [3.81 — 15.11]
Mean TBR	2.16 [1.77 — 2.69]	1.79 [1.32 — 2.31]
Mean SUV_{max}	16.05 [6.96 — 32.96]	10.93 [4.73 — 25.00]
Mean NI	59.75 [45.00 — 65.00]	54.38 [63.00 — 42.00]

Simulation of ICT PET scans

Following the comparison of pre- and post-ICT PET imaging, synthetic tumours were simulated using PETSTEP [91]. Previously published data

[36, 65] has proven that statistical models, developed from 100 synthetic tumour objects, are suitable for MTV delineation in diagnostic ^{18}F -FDG PET imaging. Therefore, adopting the same approach, in this chapter an additional set of 100 synthetic target tumour objects were simulated, with MTVs, NIs and TBRs covering the range of values measured from 10 of the FiGaRO clinical MTVs, thereby generating a new PET imaging dataset consisting of 200 PET scans, known in this chapter as $\text{ATLAAS}_{\text{ICT}}$. The volumes obtained from the clinical PET scans were in the range 1.59 — 21.25 mL, the range of NIs were 28 — 63 and the range of TBR values was 0.57 — 3.50. The contours obtained from the 10 FiGaRO PET scans, used for tumour characteristic comparison, were also used as a basis for the target tumour simulation. Target tumour objects were simulated using PETSTEP, the simulation process consists of 8 steps [91] as shown in Figure 4.6 and described:

1. A tumour lesion is either added to or replaces the background of a PET image at a location either specified by a user or at the tumour location.
2. The resulting map is blurred to mimic the Point Spread Function (PSF) of a PET scanner.
3. Using a radontransform, data is then forward-projected to generate noise free data.
4. An attenuation map, obtained from a CT image is forward projected to correct the PET image for attenuation. The attenuated data is scaled so that the sum of the intensities is equivalent to the total number of counts being simulated.

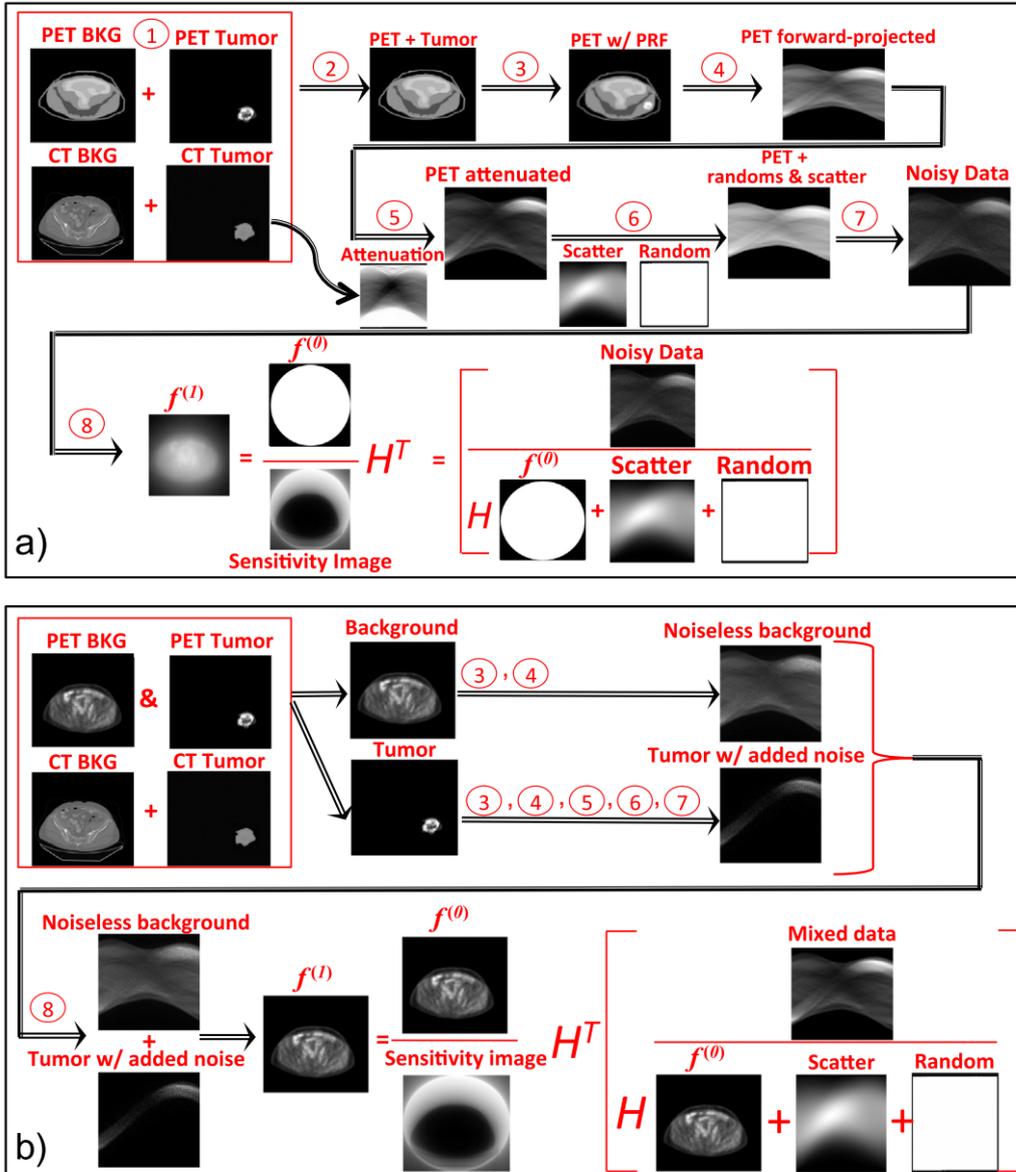


Figure 4.6: The simulation of PET scans using the PETSTEP process. (a) describes the process for pre-existing objects. (b) describes the process for new objects. The data shows the sinograms used in the reconstruction of the PET image and the reconstruction is shown with Poisson noise and initialisation images used in iterative reconstruction. Figure reproduced, with thanks, from Berthon *et al* [91]

5. Scatter and random annihilation events are added to the resulting image. The distribution of the scatter is generated from the original forward projection of the blurred image; however, the distribution of the random annihilation events is generated from a uniform background.
6. Noise experienced in PET images acquired from a PET scanner is added to the simulated data.
7. The produced data can then be reconstructed using an appropriate reconstruction algorithm.
8. After reconstruction filters may be applied to the reconstructed PET image.

Figure 4.7a shows the range of MTV and TBR for the FiGaRO trial PET data and ATLAAS_{ORIG} before simulation of additional data. Visual comparison of the data shows little to no overlap between the characteristics of the FiGaRO data and ATLAAS_{ORIG}. However, Figure 4.7b shows that the simulation of the additional data to create the ATLAAS_{ICT} dataset has improved the overlap between the range of MTV and TBR obtained from FiGaRO trial data and ATLAAS_{ORIG}. Before simulation of additional data, Figure 4.8a shows that the range of MTV and NI obtained from the FiGaRO data and ATLAAS_{ORIG} had little to no overlap and Figure 4.8b shows that the overlap between the two datasets improved after simulation of the additional ATLAAS_{ICT} scans. Additionally, Figure 4.9a shows that the range of NI and TBR for the FiGaRO trial PET data and ATLAAS_{ORIG} had no overlap. Whereas, after simulation of the additional scans for the ATLAAS_{ICT}

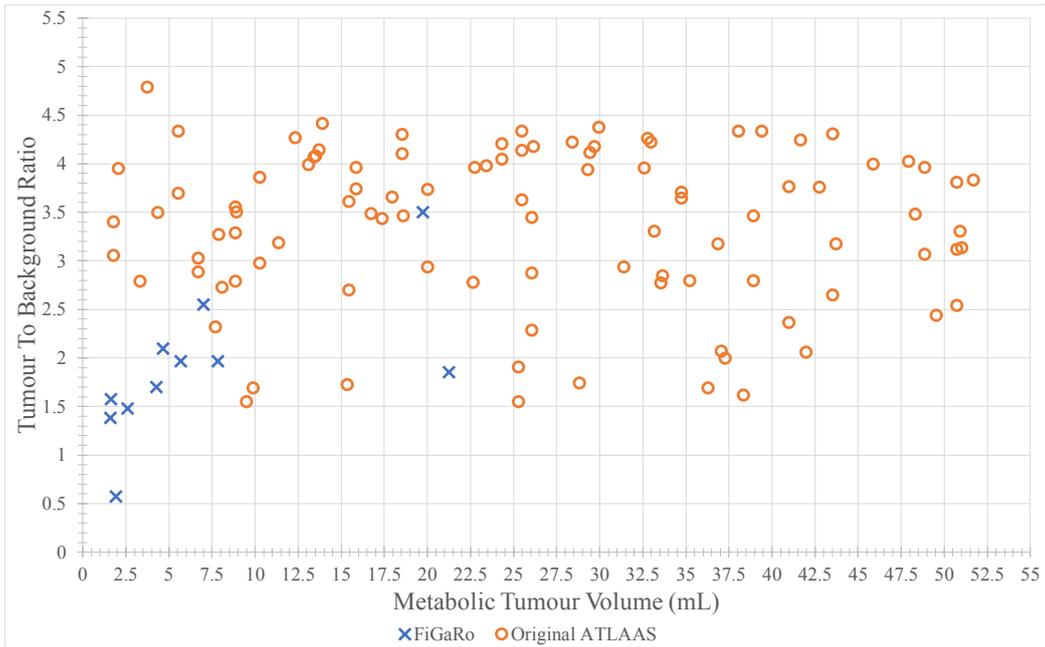
dataset the overlap between the characteristics improved. The ATLAAS_{ICT} dataset was used to develop a new decision tree-based statistical model for ATLAAS.

4.1.5 PET-AS MTV delineation after 1 cycle of ICT

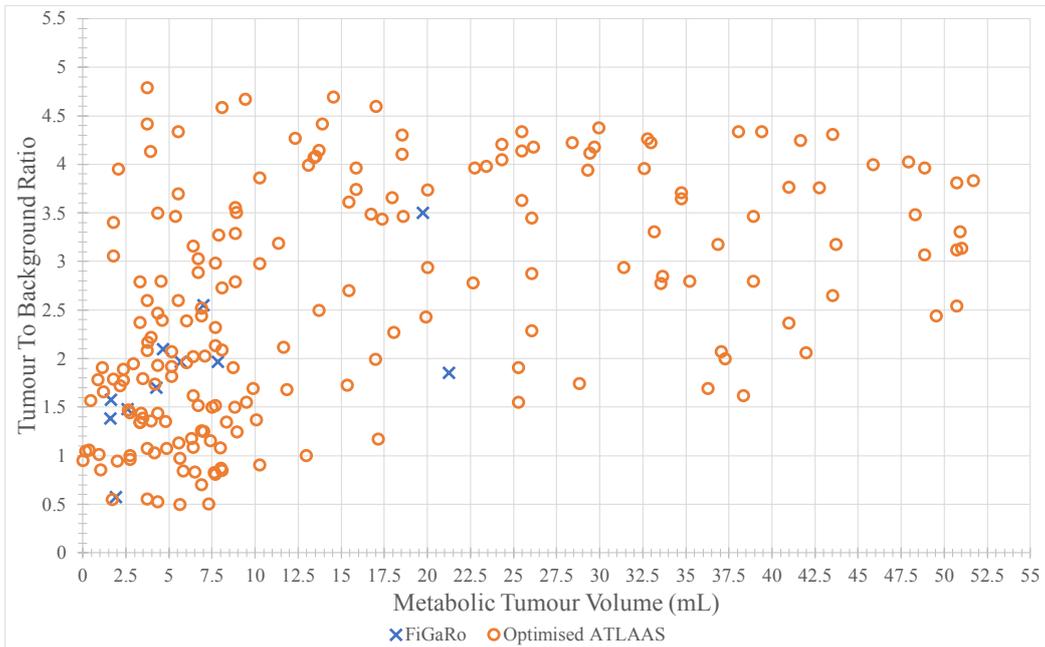
After development of the new statistical model, from the ATLAAS_{ICT} dataset, PET-AS algorithms were applied to the clinically acquired PET scans to derive estimated MTVs. MTVs were defined by the PET-AS methods AT, RG, KM, FCM, GCM, WT and SUV_{peak} PET-AS methods. SUV_{peak} was applied with thresholds ranging from 20% to 80% in increments of 10% (PT20 — PT80). Further, MTVs were delineated with the ATLAAS statistical models ATLAAS_{ORIG} and ATLAAS_{ICT}. Clustering PET-AS methods were applied to the PET images with a different number of clusters. KM was applied with two and three clusters, FCM was applied with two clusters and GCM was applied with three and four clusters. A total of 336 MTVs were delineated by different PET-AS methods, resulting in 16 for each patient.

4.1.6 Statistical analysis

In lieu of histopathological specimens or known ground truths, AAPM Task group 211 [1] recommends reporting DSC when evaluating contours derived by PET-AS methods and clinicians. In this chapter, therefore, the delineated PET-AS MTV contours were compared to the manually delineated MTV using DSC, which was calculated using Matlab 2016b (The MathWorks Inc.,

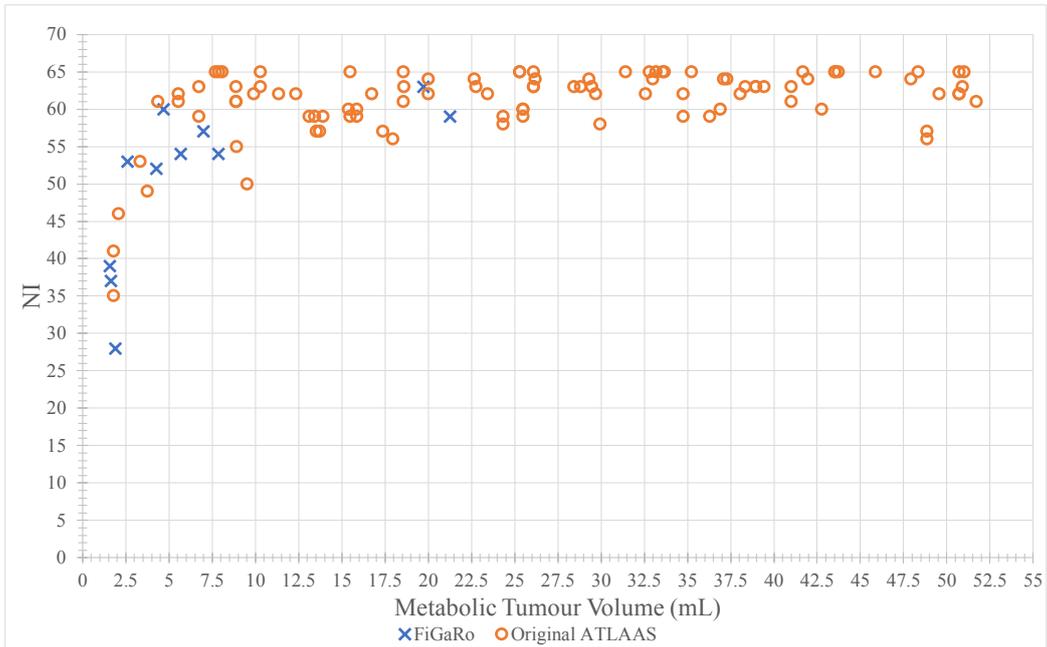


(a) Before PETSTEP simulation of ICT PET scans.

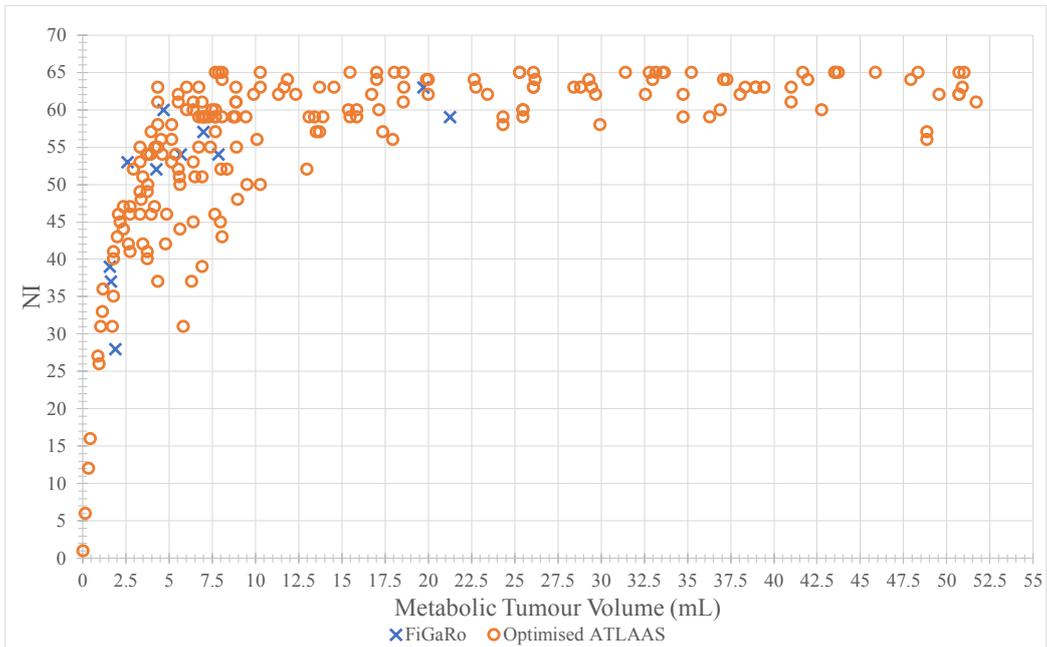


(b) After PETSTEP simulation of ICT PET scans

Figure 4.7: The range of MTV and TBR for the FiGaRO trial PET data and ATLAAS_{ORIG} and ATLAAS_{ICT} datasets. Before simulation, the two datasets had little to no overlap between the characteristics. However, after simulation of the additional scans the overlap between the datasets improved.

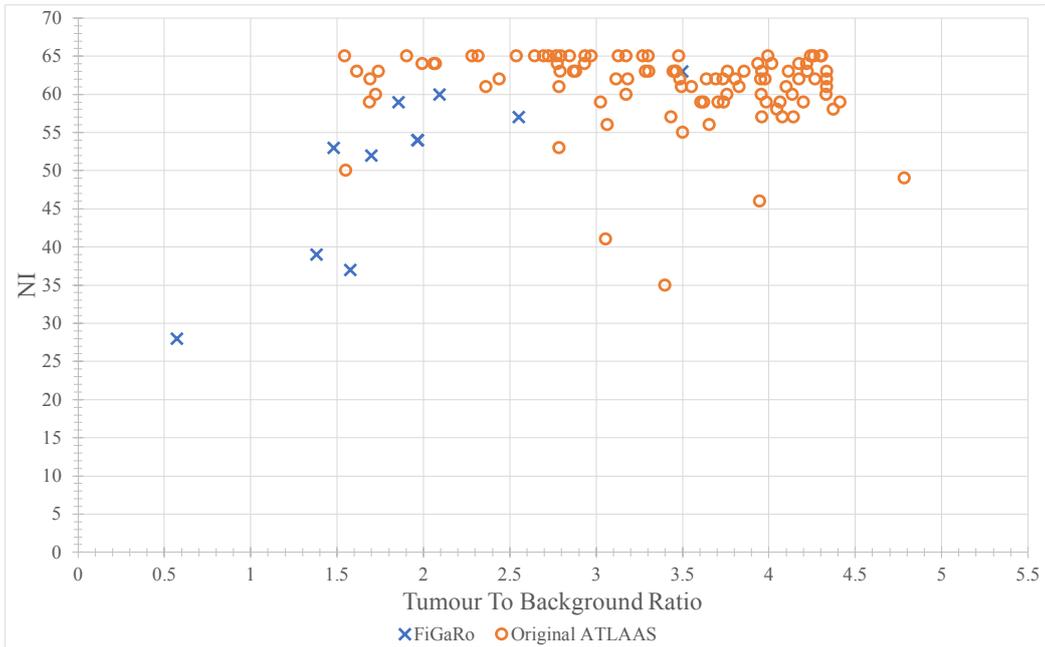


(a) Before PETSTEP simulation of ICT PET scans

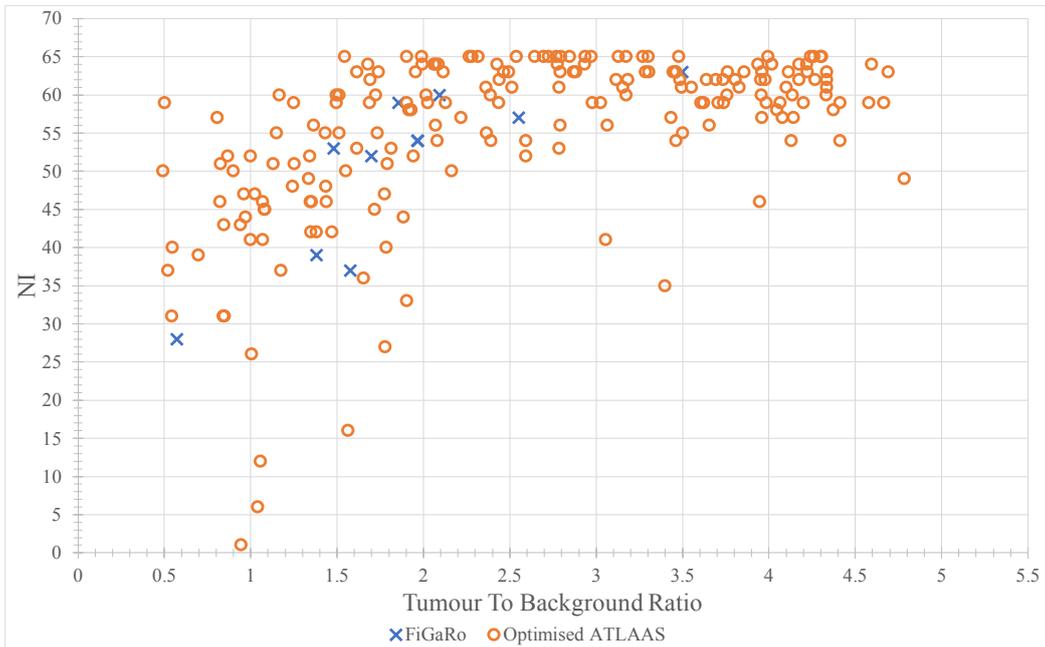


(b) After PETSTEP simulation of ICT PET scans

Figure 4.8: The range of MTV and NI for the FiGaRO trial PET data and ATLAAS_{ORIG} and ATLAAS_{ICT} datasets. Before simulation, the two datasets had little to no overlap between the characteristics. However, after simulation of the additional scans the overlap between the datasets improved.



(a) Before PETSTEP simulation of ICT PET scans



(b) After PETSTEP simulation of ICT PET scans

Figure 4.9: The range of TBR and NI for the FiGaRO trial PET data and $ATLAAS_{\text{ORIG}}$ and $ATLAAS_{\text{ICT}}$ datasets. Before simulation, the two datasets had little to no overlap between the characteristics. However, after simulation of the additional scans the overlap between the datasets improved.

Natick, Massachusetts, US). The mean DSC, range of DSC and the SD were also calculated. The performance (DSC) of each PET-AS method was correlated with the MTV, TBR and SUV_{peak} derived from the clinicians' MTV contours in order to investigate relationships between the derived tumour characteristics and the accuracy of each PET-AS method. The MTV in mL was calculated using in-house software. Statistical differences in DSC distributions were assessed using Kruskal-Wallis and P-values less than 0.05 were considered to be statistically significant.

4.2 Results

4.2.1 Manual MTV delineation after 1 cycle of chemotherapy

The mean MTV and range of MTVs, contoured jointly by the nuclear medicine physician and clinical oncologist, on ^{18}F -FDG PET imaging acquired after one cycle of chemotherapy, was 6.22 [1.12 — 21.25] mL. The SUV_{max} range was 3.51 — 25.00. Figure 4.10 shows contours from four cases, representative of patients, and the different MTV sizes derived by the nuclear medicine physician and clinical oncologist included in the FiGaRO trial. In each case, the user has contoured around the metabolically active region of the tumour, with varying margins of over contouring in each case. The derived MTV contours, when displayed on co-registered PET/CT scans, also encompass areas which anatomically are not able to be metabolically active (air, on the CT

image). Therefore, when deriving the MTV and GTV, functional imaging and anatomical imaging should be taken into account in order to produce clinically relevant target volumes.

4.2.2 Comparison between PET-AS MTV and clinician derived MTV

The manually derived MTV's were compared to the PET-AS derived MTV's using DSC. Figure 4.11 shows the mean DSC (+/- SD) for all PET-AS derived MTV, when compared to the manually delineated MTV. ATLAAS_{ICT} shows significant improvement in accuracy in comparison to ATLAAS_{ORIG}. Additionally, ATLAAS_{ICT} outperformed all of the other PET-AS methods included in this chapter. The mean DSC (+/- SD) for the four best performing PET-AS methods, ATLAAS_{ICT}, PT60, AT, GCM3, and WT was 0.72 (+/- 0.10), 0.61 (+/- 0.20), 0.63 (+/- 0.15), 0.55 (+/- 0.20), 0.60 (+/- 0.21) respectively. A Kruskal Wallace test showed significant difference ($P = 0.0003$) between the PET-AS MTV and the MTVs delineated by ATLAAS_{ICT}, PT60, AT, GCM3 and WT PET-AS methods.

Figure 4.12 shows percentage increase in MTV obtained from the PET-AS methods, AT, ATLAAS_{ICT}, PT60 and WT when compared to the clinician derived MTV. The four best performing PET-AS methods delineated MTVs smaller than the clinician derived MTV in 8 patients and larger in 6 patients. Specifically, PT60 derived MTV larger than all other included PET-AS methods in patients 2, 4, 8, 10 and 19. In these cases, the clinician derived MTV



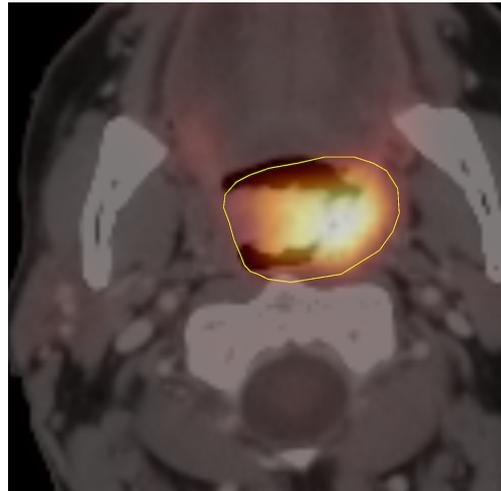
(a) Patient 1 manual MTV



(b) Patient 2 manual MTV



(c) Patient 3 manual MTV



(d) Patient 4 manual MTV

Figure 4.10: Contours delineated by a Nuclear Medicine physician and Clinical Oncologist. In each case, the user has contoured around the metabolically active region of the tumour, with varying margins of over contouring in each case. The derived MTV contours, when displayed on co-registered PET/CT scans, also encompass areas which anatomically are not able to be metabolically active (air, on the CT image). In Figure 4.13 ATLAAS derived contours are compared to the clinicians' MTV.

was less than 2 mL. ATLAAS_{ICT} and AT delineated the equivalent MTV in 19 patients. Figure 4.13 shows contours delineated by the best performing method, ATLAAS_{ICT}, in four patients representative of the patient cohort in comparison to the manually delineated MTV. An enlarged comparison is shown within the white box. In patient 3, ATLAAS under contoured the MTV in comparison to the clinician, whereas in the patient 2 ATLAAS contoured a MTV larger than the clinician. In patients 1 and 4, ATLAAS derived MTVs which are comparable to the the clinician derived MTV. Therefore, whilst ATLAAS was the best performing PET-AS method in this study, it should be used as a guide for target volume delineation which can be adapted and modified according to the treatment planning protocol. ATLAAS has the potential to inform target volume delineation by demonstrating that the manually defined MTV can be reduced or should be expanded to include additional regions of metabolic uptake.

Figures B.1 to B.48 in Appendix B shows the correlations of the PET-AS methods PT20 — PT80, ATLAAS_{ICT}, AT, RG, KM with 2 and 3 clusters, FCM with 2 clusters, GCM with 3 and 4 clusters and WT and the tumour characteristics SUV_{peak} , TBR and volume (mL) obtained from the clinician derived MTV contours. Table 4.2 shows the mean [range] of MTV (mL) and mean SUV_{max} of PET-AS derived MTV.

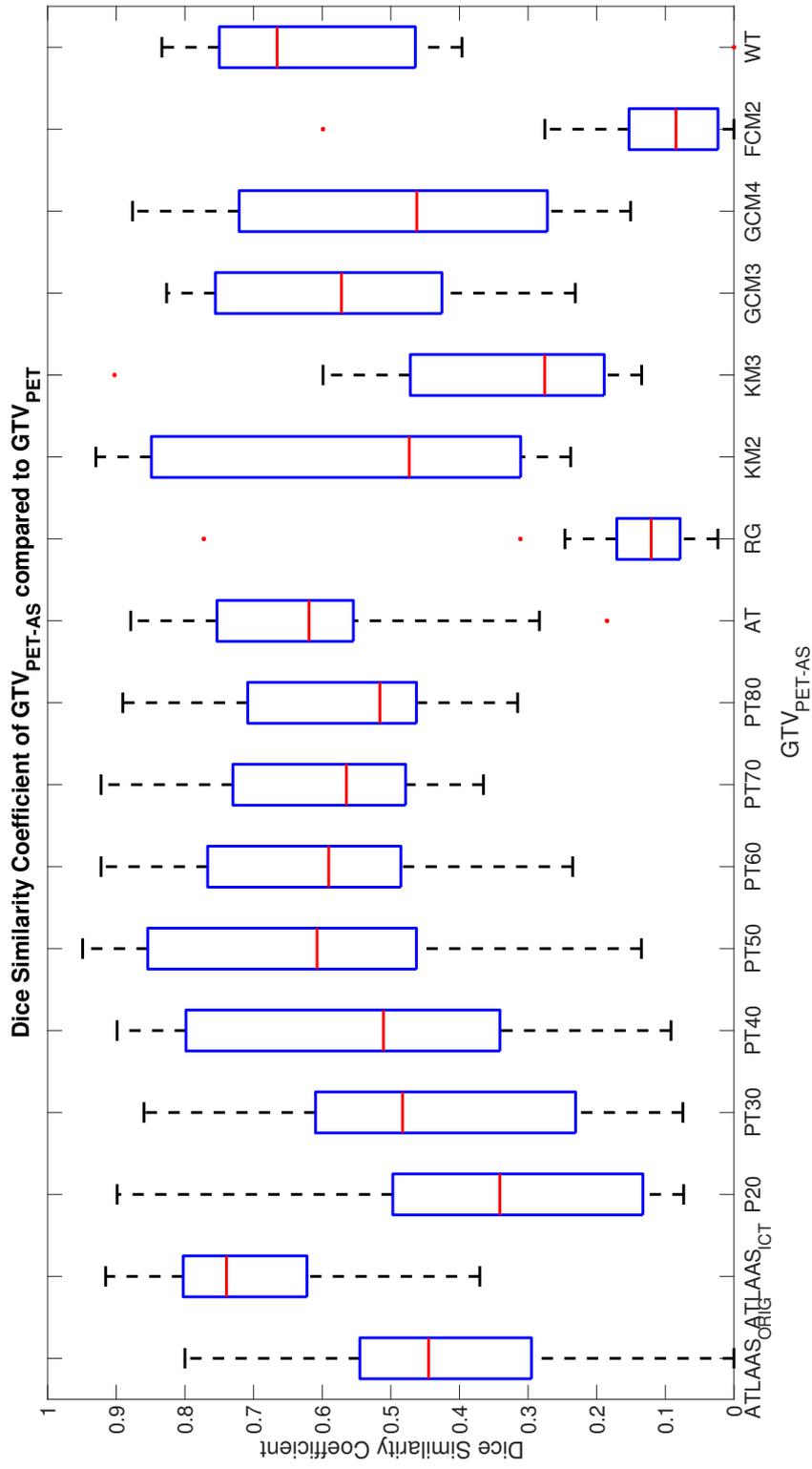


Figure 4.11: DSC of PET-AS delineated MTV when compared to manually delineated MTV

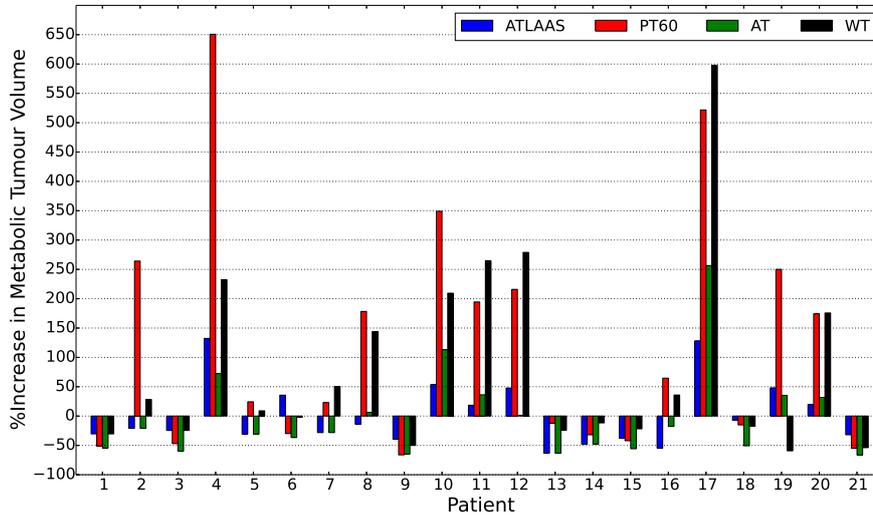


Figure 4.12: The percentage increase in MTV delineated by PET-AS compared to the MTV of the clinicians

4.3 Discussion

MTV delineation for RT planning, in H&N ^{18}F -FDG PET imaging acquired after one cycle of ICT is challenging. In IMRT planning especially, errors in delineation may increase geographical miss [23, 119, 120] because of steep dose gradients resulting in smaller margins of error. Our analysis of ^{18}F -FDG tumour characteristics, including the MTV before ICT and after ICT, demonstrated that tumour characteristics decrease after one cycle of ICT. Therefore, MTV delineation on imaging acquired before ICT can further increase the errors in RT planning and therefore can potentially lead to increased dosage to the OAR. The current recommendations to reduce intra and inter-observer variability, in MTV delineation, are to use fixed windowing and viewing levels when delineating the MTV [117, 118]. However, the

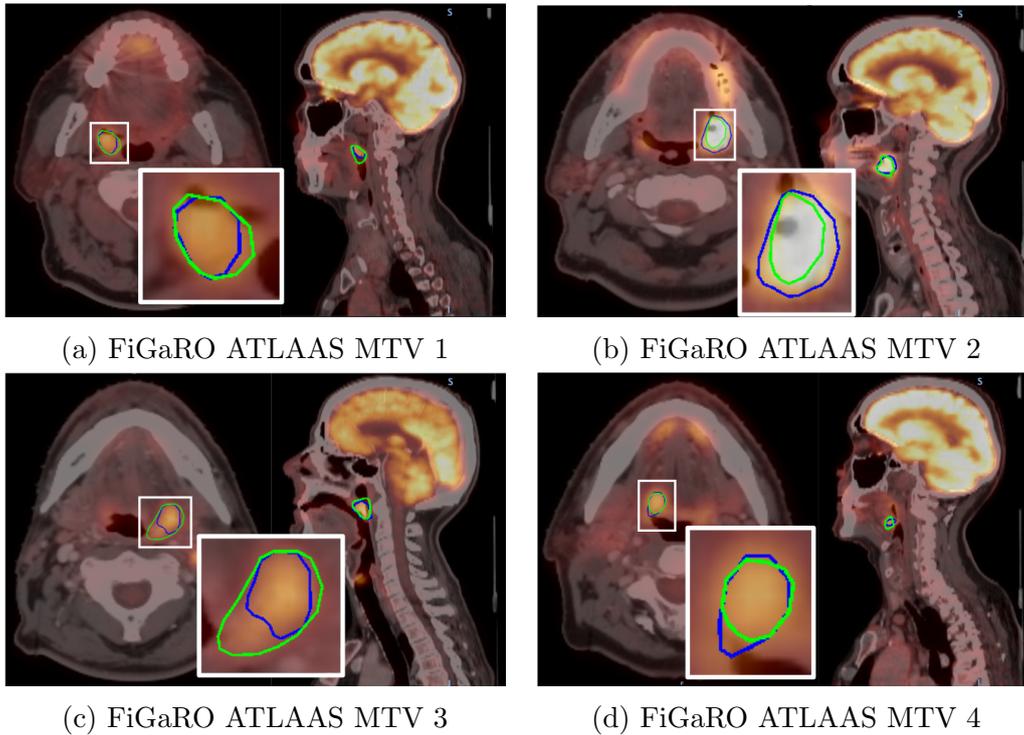


Figure 4.13: ATLAAS derived MTV (blue) compared to manually delineated MTV (green). A zoomed in view of the derived contours is shown in the white box. In patient 3, ATLAAS under contoured the MTV in comparison to the clinician, whereas in the patient 2 ATLAAS contoured a MTV larger than the clinician. In patients 1 and 4, ATLAAS derived MTV which are comparable to the the clinician derived MTV.

reduced MTV and metabolic uptake after one cycle of ICT, combined with standardised viewing parameters, can limit the identification of the tumour disease extension because of reduced tumour contrast compared to the background uptake.

One of the aims of this chapter was to externally validate the ATLAAS segmentation methodology; to achieve this aim, twenty patients recruited to a phase I feasibility study were analysed. The MTVs derived from sixteen PET-AS methods, including the ATLAAS segmentation methodology,

were compared to twenty-one manually delineated MTVs by calculating the DSC, the percentage differences in MTV and the derived contour SUV_{\max} values. Analysis of the obtained results showed a significant difference, statistically, in the performance of the MTVs delineated by the PET-AS methods; with the highest performing PET-AS method in low TBR scenarios, after the development of a new training dataset, being the ATLAAS segmentation methodology. Our analysis also demonstrates that AT techniques were found to be more robust for MTV delineation in comparison to fixed thresholding techniques. Additionally, tumour characteristics obtained from the manually delineated MTV and the performance of the PET-AS methods were correlated. The results follow that of other published studies [32, 33, 77], in that the performance of individual PET-AS methods is dependent upon tumour and imaging characteristics. However, these results also demonstrate that the performance of the ATLAAS segmentation methodology is independent of the same tumour and imaging characteristics and that the performance of individual automated segmentation methodologies can be enhanced through the use of machine-learning techniques. Development of the $ATLAAS_{ICT}$ statistical model demonstrates that it is possible to optimise / adapt the ATLAAS segmentation methodology for new treatments and therefore potentially differing PET tracers and anatomical sites.

The results presented in this chapter are limited by a relatively small cohort ($n = 20$) of patients, although all recruited as part of the same clinical trial in OP SCC. Additionally, patients also underwent only one cycle of chemotherapy; therefore, it was not possible to evaluate the accuracy and robustness of

the PET-AS methods included in this study against the remaining biological tissue after multiple cycles of chemotherapy or after fractions of radiotherapy. This work has been published in *Phys Med* (2019) [121].

4.4 Conclusion

MTV delineation on ^{18}F -FDG PET imaging acquired after one cycle of ICT is challenging due to reduced tumour contrast in comparison to background uptake; however, the machine-learned PET-AS methodology ATLAAS, when optimised, was accurate for MTV delineation in this low TBR scenario. Further, ATLAAS was found to be suitable for MTV delineation on ^{18}F -FDG PET imaging acquired in centres external to VCC and PETIC. Whilst the performance of the PET-AS methods were correlated with the tumour characteristics volume (mL), TBR, NI and SUV_{max} , additional factors and tumour characteristics could influence PET-AS method delineation accuracy. Therefore, the following chapter investigates the role of morphological features extracted from the MTV as classifiers in the development of the ATLAAS machine-learned training models.

Table 4.2: Mean [Range] of Volume (mL), SUV_{MAX} and DSC of the PET-AS derived MTV

PET-AS	Mean MTV (mL)	Mean SUV_{MAX}	Mean DSC (+/- SD)
ATLAAS _{ORIG}	12.47 [0.22 — 56.48]	10.49 [4.74 — 25.65]	0.42 (+/- 0.21)
ATLAAS _{ICT}	6.01 [1.30 — 24.00]	10.19 [3.88 — 25.00]	0.72 (+/- 0.10)
PT20	33.67 [4.65 — 98.85]	10.24 [4.23 — 25.00]	0.33 (+/- 0.21)
PT30	21.48 [4.65 — 53.84]	10.24 [4.23 — 25.00]	0.42 (+/- 0.23)
PT40	14.24 [4.65 — 33.06]	10.24 [4.23 — 25.00]	0.53 (+/- 0.27)
PT50	9.61 [4.47 — 21.98]	10.24 [4.23 — 25.00]	0.60 (+/- 0.26)
PT60	8.66 [3.28 — 44.01]	10.17 [3.18 — 25.00]	0.61 (+/- 0.20)
PT70	5.49 [1.49 — 12.13]	10.24 [4.23 — 25.00]	0.60 (+/- 0.16)
PT80	4.32 [1.15 — 12.13]	10.24 [4.23 — 25.00]	0.55 (+/- 0.16)
AT	3.85 [1.20 — 8.75]	10.28 [4.12 — 25.00]	0.63 (+/- 0.15)
RG	69.03 [1.32 — 160.58]	10.21 [4.62 — 25.00]	0.13 (+/- 0.07)
KM2	12.70 [6.01 — 25.16]	10.32 [4.62 — 25.00]	0.54 (+/- 0.28)
KM3	25.38 [11.91 — 6.26]	10.36 [4.62 — 25.00]	0.34 (+/- 0.19)
GCM3	7.96 [1.78 — 19.73]	10.26 [3.98 — 25.00]	0.55 (+/- 0.20)
GCM4	14.12 [4.50 — 32.91]	10.35 [4.62 — 25.00]	0.48 (+/- 0.25)
FCM2	40.36 [4.71 — 96.40]	6.16 [2.51 — 15.83]	0.11 (+/- 0.14)
WT	7.20 [0.54 — 27.55]	10.20 [3.39 — 25.00]	0.60 (+/- 0.21)

Chapter 5

Impact of morphological features in the development of machine learned automated segmentation training models

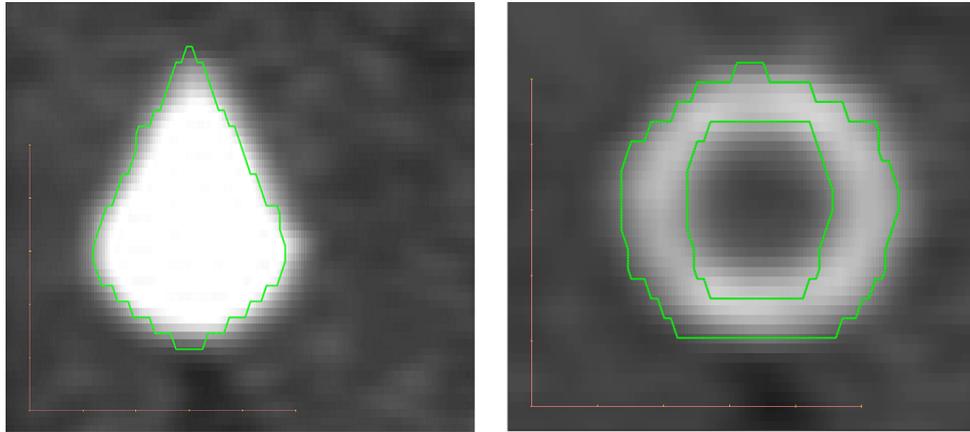
Accurate MTV delineation is crucial for the management of cancer, with inaccuracies leading to increased dose to the OAR [13]. This is especially critical for IMRT where the margin for error is decreased [23]. As described in Section 2.4, the accuracy of PET-AS delineation is influenced by MTV characteristics, including patterns of heterogeneity in the tumour tissue. However, morphological features, which aim to quantify the MTV shape and surface, may also be classifiers due to PET-AS algorithms performing differently in complex shapes. There are currently two techniques used to assess a tumour's

response to therapy, one is to compute the longest diameter of the tumour (Response Evaluation Criteria In Solid Tumours (RECIST) [76]) and the other is to compute the product of the longest diameter of the tumour and the maximum perpendicular diameter of the tumour (World Health Organisation (WHO) [122]). For tumours with equivalent MTVs, it is possible that different longest diameters may be observed and therefore two tumours with the same volume (mL) have the potential to be classified differently; as a responder or non-responder. Using this principle, morphological features may be additional classifiers in the development of training models due to providing an additional feature dimension when equivalent volumes are observed. The inclusion of additional feature dimensions is a technique observed in the development of SVMs to improve hyper-plane definition; therefore, this study aims to investigate the inclusion of morphological data in a machine learned training model to improve PET-AS MTV accuracy and to provide an additional classifier to the ATLAAS segmentation methodology. The materials & methods to achieve this aim are described in the following sections.

5.1 Materials and Methods

5.1.1 Hardware and software

The hardware and software for the generation and processing of data is described in Section 3.1.4.



(a) 512×512 matrix Drop contour on 1 mm PET (b) 512×512 matrix Tori contour on 1 mm PET

Figure 5.1: Contours obtained from the training dataset

5.1.2 Training Dataset

The training dataset in this study consists of 211 simulated PET scans with a total of 260 MTVs. GTs were obtained during the simulation process with a matrix size of 512×512 . The resulting simulated PET image had a matrix size of 256×256 with voxel dimensions of 2.7×2.7 mm. The training dataset MTV ranged from 0.02 mL — 184.80 mL with a mean volume of 54.34 mL. The NI ranged from 2 to 65 with a mean NI of 59.44, whereas TBR ranged from 0.10 — 4.21 with a mean TBR of 1.98. Figure 5.1 shows contours, representative of the different tumour complexities and therefore differing morphological features, within the training dataset with a 512×512 matrix. Figure 5.1a shows a Drop insert and Figure 5.1b shows a Tori insert. These shapes were chosen for this study as they had complex geometries and were identified from previous work [33].

5.1.3 Validation Dataset

The validation dataset in this study consists of a combination of phantom and simulated PET scans with 96 MTVs in total. GT contours were obtained from CT imaging, for simulated data GTs were obtained during the simulation process with a matrix size of 512×512 . The validation dataset MTV ranged from 1.39 mL — 174.96 mL with a mean volume of 32.24 mL. The NI ranged from 58 to 65 with a mean of 48.01 whereas TBR ranged from 1.18 to 3.59 with a mean of 1.95. Figure 5.2 to Figure 5.4 show the overlap between the characteristics of the training and validation dataset. The MTV and NI of the two datasets overlap in Figure 5.2. Additionally the MTV and TBR of the two datasets overlaps in Figure 5.3. Furthermore, the TBR and NI overlap in Figure 5.4. The overlap between the training and validation dataset characteristics ensures ATLAAS was only evaluated in scenarios which it had been trained for. In Chapter 4 a difference in training and validation dataset characteristics was shown to impact the accuracy of ATLAAS. Figure 5.5 shows GT contours obtained from CT imaging in the phantom PET scans in the validation dataset, representative of the different tumour complexities and therefore differing morphological features in the validation dataset. Figure 5.5a shows a GT Pear contour and Figure 5.5b shows a Tube GT contour. Whilst, similar to to the contours included in the training dataset (Figure 5.1), the validation dataset contours have distinctly different geometries and therefore morphological features.

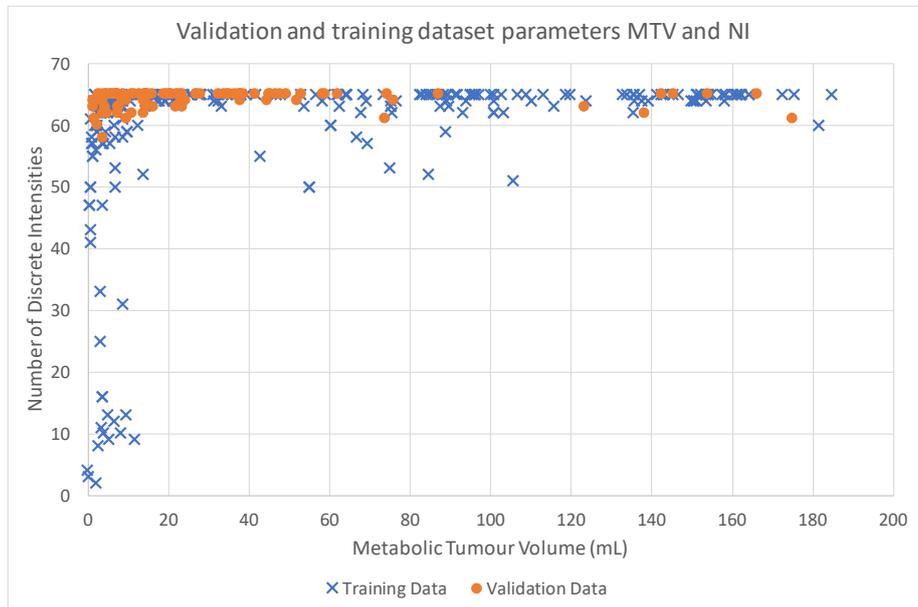


Figure 5.2: Training and validation dataset MTV and NI. There is good visual overlap between the training and validation dataset characteristics.

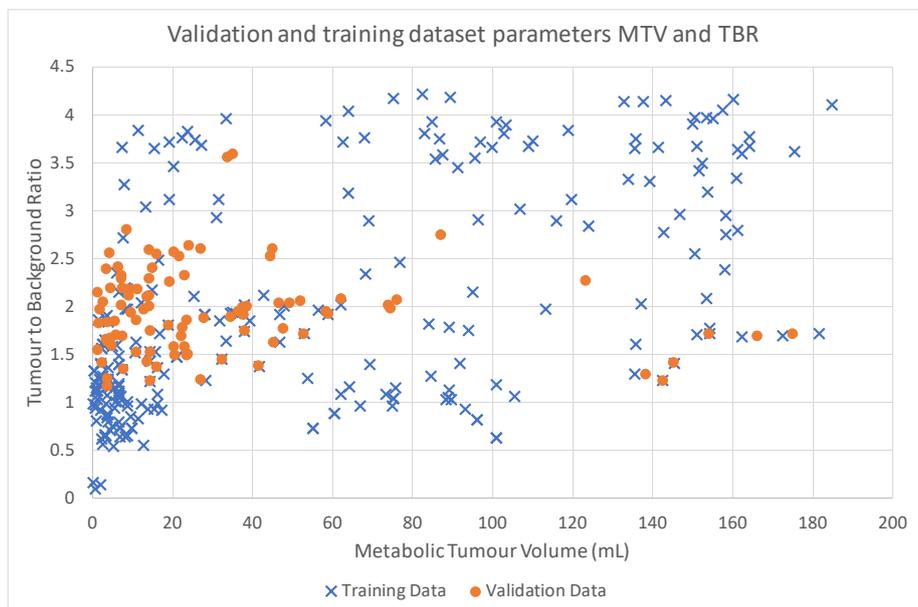


Figure 5.3: Training and validation dataset MTV and TBR. There is good visual overlap between the training and validation dataset characteristics.

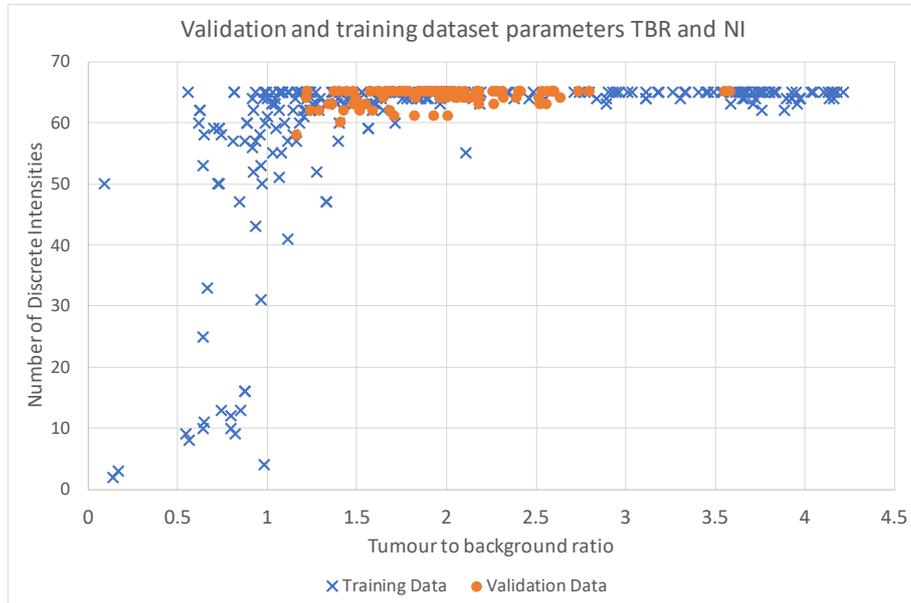


Figure 5.4: Training and validation dataset NI and TBR. There is good visual overlap between the training and validation dataset characteristics.

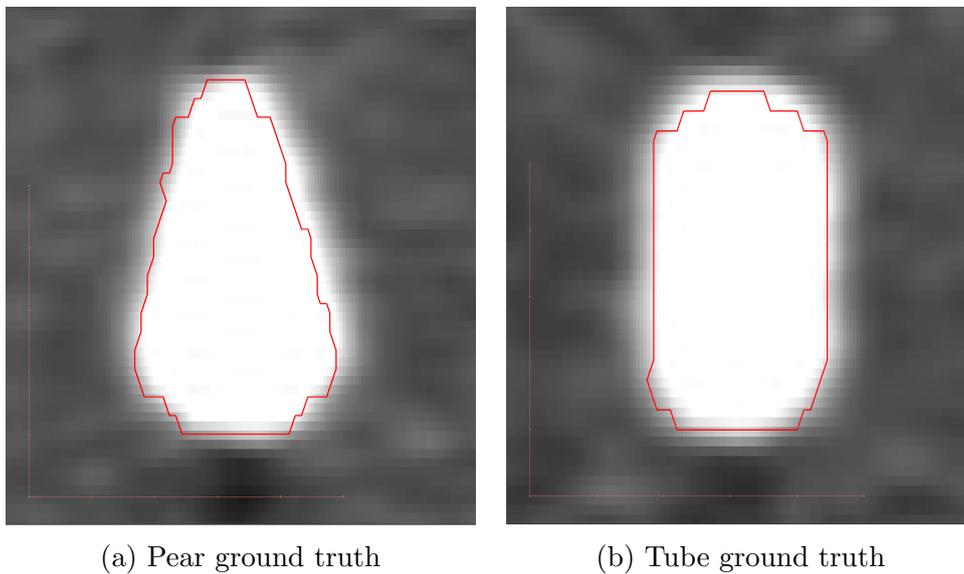


Figure 5.5: Validation ground truth contours

5.1.4 Morphological features

Morphological features aim to quantify the surface and shape of a MTV. Several morphological features (compactness one, compactness two, spherical disproportion, sphericity and asphericity) quantify the deviation of the defined volume from a spheroid of the same volume. These features can be calculated from one another [123] and are therefore highly correlated. However, they are included in this study for completeness.

Volume

The volume of the tumour calculated from a mesh. Equation 5.1 shows the IBSI definition for volume, where V_k is the volume of the tetrahedron formed from the origin at coordinate points $(0, 0, 0)$ with the face k and a, b, c are the vertex points of the face. Volume is then defined as Equation 5.2 where V is the MTV defined as the summation of V_k .

$$V_k = \frac{a \cdot (b \times c)}{6} \quad (5.1)$$

$$V = \sum_{k=1}^{N_{fc}} V_k \quad (5.2)$$

Surface area

The surface area of the tumour. Equation 5.3 shows the IBSI definition for surface area, where A is the surface area calculated by summation of the surface area of the triangular face area, which is shown in Equation 5.4.

$$A = \sum_{k=1}^{N_{fc}} A_k \quad (5.3)$$

$$A_k = \frac{ab \times ac}{2} \quad (5.4)$$

Surface to volume ratio

The surface area divided by the volume of the tumour as defined in Equation 5.5 where A is the surface area and V is the tumour volume.

$$S2V = \frac{A}{V} \quad (5.5)$$

Compactness one, Compactness two, Spherical disproportion, Sphericity and Asphericity

These features quantify the deviation of the tumour volume from a representative spheroid. Compactness one is defined in Equation 5.6, where V is volume and A is the surface area. Compactness two is defined in Equation 5.7. Spherical Disproportion is defined in Equation 5.8. Sphericity is defined in Equation 5.9 and Asphericity is defined in Equation 5.10.

$$\text{compactnessone} = \frac{V}{\pi^{1/2}A^{3/2}} \quad (5.6)$$

$$\text{compactnesstwo} = 36\pi \frac{V^2}{A^3} \quad (5.7)$$

$$\text{sph.disproportion} = \frac{A}{(36\pi V^2)^{1/3}} \quad (5.8)$$

$$\text{sphericity} = \frac{(36\pi V^2)^{1/3}}{A} \quad (5.9)$$

$$\text{asphericity} = \left(\frac{1}{36\pi} \frac{A^3}{V^2} \right)^{1/3} - 1 \quad (5.10)$$

Centre of mass shift

Centre of mass shift (COMShift) is defined as the distance between the tumour volume centroid and the intensity-weighted tumour volume centroid as defined in Equation 5.11. The centre of the tumour mass is defined in Equation 5.12. Where, $N_{v,m}$ is the number of voxels in the mask representing the tumour and X_c is the voxel point set. The intensity weighted centre of mass is defined in 5.13, where the position of the voxel in the intensity mask $X_{c,gl}$ is weighted by it's intensity X_{gl} . $N_{v,gl}$ is the number of voxels in the intensity mask. COMShift aims to quantify how far the tumour areas with a higher metabolic uptake are from the centre of the tumour. A COMShift of 0 means that the highest area of metabolic uptake is at the centre of the tumour mass, the higher the COMShift the further the area of highest metabolic uptake is away from the centre of the tumour mass.

$$\overrightarrow{COMShift} = \left\| \overrightarrow{tumour.centre} - \overrightarrow{intensity.centre} \right\|_2 \quad (5.11)$$

$$\overrightarrow{tumour.centre} = \frac{1}{N_{v,m}} \sum_{k=1}^{N_{v,m}} \vec{X}_{c,k} \quad (5.12)$$

$$\overrightarrow{intensity.centre} = \frac{\sum_{k=1}^{N_{v,gl}} X_{gl,k} \vec{X}_{c,gl,k}}{\sum_{k=1}^{N_{v,gl}} X_{gl,k}} \quad (5.13)$$

Maximum 3D diameter

Distance between the two most distant vertices in the mesh defined in Equation 5.14.

$$Max.3D.Diam = \max \left(\left\| \vec{X}_{vx,k_1} - \vec{X}_{vx,k_2} \right\|_2 \right), \quad k_1 = 1, \dots, N \quad k_2 = 1, \dots, N \quad (5.14)$$

Major, minor and least axis length

Principle component analysis is used to determine the largest, second largest and smallest tumour axis. Major axis length is defined in Equation 5.15. Minor axis length is defined in Equation 5.16 and least axis length is defined in Equation 5.17. λ is the eigenvalue of the eigenvector vector output from the principle component analysis of the tumour.

$$Major = \sqrt[4]{\lambda_{major}} \quad (5.15)$$

$$Minor = \sqrt[4]{\lambda_{minor}} \quad (5.16)$$

$$Least = \sqrt[4]{\lambda_{least}} \quad (5.17)$$

Elongation and Flatness

Elongation is the ratio of the length of the tumour against the width of the tumour as defined in Equation 5.18 and flatness is the ratio of the major axis length and the least axis length as defined in Equation 5.19.

$$Elongation = \sqrt{\frac{\lambda_{minor}}{\lambda_{major}}} \quad (5.18)$$

$$Flatness = \sqrt{\frac{\lambda_{least}}{\lambda_{major}}} \quad (5.19)$$

Volume Density and Area Density

Volume density is the ratio of the tumour volume and the volume of the bounding box, whereas the area density is the ratio of the surface area of the

tumour volume to surface area of the bounding box as defined in Equations 5.20 and 5.21, where bb is the boundary box type. In this study, the boundary box was defined using 3 different approaches, axis alignment, approximation of the enclosing ellipsoid and a convex hull.

$$v.dens = \frac{V}{V_{bb}} \quad (5.20)$$

$$a.dens = \frac{A}{A_{bb}} \quad (5.21)$$

The morphological features were implemented as part of the IBSI initiative for the standardisation of the extraction of radiomic features from medical images [123,124].

5.1.5 ATLAAS training model development

For a given field of view (FOV), a CT image has matrix dimensions of 512×512 voxels, with voxel dimensions of $0.98 \times 0.98 \times 3$ mm. In comparison, for the same FOV, PET image matrices are typically 128×128 voxels or 256×256 voxels dependent upon the scanner and reconstruction settings [125]. PET images with matrix dimensions of 256×256 have voxel dimensions of $2.7 \times 2.7 \times 3$ mm. The resolution (matrix and voxel dimension) of PET imaging, is limited by the physics of the PET scanner; including, the

limited size of the PMTs and size of the detectors. Therefore, due to the low finite spatial resolution, PET imaging is subject to the partial volume effect (PVE) [126–128] which may affect the quantification of morphological features. The PVE consists of two distinct imaging processes, the tissue fractionation effect and 3D image blurring [126]. PVE in PET imaging can be seen in Figure 5.6 as a CT contour is transferred to PET images interpolated to differing resolutions including the resolution of the CT where Figure 5.6a to 5.6e shows the CT contour transferred to 1, 2, 3 and 4 mm interpolated PET imaging. The CT contour is spherical and transferring the contour to lower image resolutions results in the degradation of the contour; degradation increases the number of imaging artefacts, including straight lines (consisting of multiple voxels) and jagged edges with acute changes in contour direction. In comparison, interpolating the PET image to a 1 mm voxel dimension results in the same PET contour, as obtained from the CT imaging. AT-LAAS training models were developed by interpolating the PET images in the training dataset to spatial resolutions of 4 mm, 3 mm and 2 mm isotropic voxels. A further training dataset was developed by interpolating the PET images in the training dataset to a x and y voxel dimensions of 1 mm with a slice thickness of 3 mm. A resolution of $1 \times 1 \times 3$ mm was chosen, instead of a 1 mm isotropic voxel dimension, due to being equivalent to the highest resolution of CT image acquired clinically in our centre ($0.98 \times 0.98 \times 3$ mm voxel dimension). A voxel dimension of $1 \times 1 \times 3$ mm was chosen instead of a $0.98 \times 0.98 \times 3$ mm was chosen to provide equal spacing and distance between the other training datasets (1 mm spacing). Additionally, interpolation to 1 mm isotropic voxel dimension in PET imaging is computationally expensive,

interpolation without an external GPU takes ≈ 30 min — 1 hr dependent upon the number of slices in the PET image and PET-AS delineation at a 1 mm isotropic voxel resolution is time consuming, taking up to ≈ 2 hr dependent upon the PET-AS method and number of slices in the PET image. The specifications for the computer used for 1 mm isotropic voxel dimension interpolation are a 3.2 Ghz Intel Core i5 computer, with 16GB of RAM and a 2GB dedicated AMD Radeon R9 M380 GPU. Timing calculations were performed using the ‘tic’ and ‘toc’ functions available within the MATLAB 2016b programming environment.

The training models were developed for each of the morphological features described in Section 5.1.4 at each voxel resolution ($1 \times 1 \times 3$ mm, $2 \times 2 \times 2$ mm, $3 \times 3 \times 3$ mm and $4 \times 4 \times 4$ mm), resulting in 88 training models in total. These resolutions and morphological features were chosen for completeness.

5.1.6 Interpolation of PET imaging

The Oxford English Dictionary defines the noun interpolation as the action of introducing or inserting among other things or between the members of any series [129]. Further to this, interpolation is defined as the process of determining the value of a function between two points at which it has prescribed values. There are a variety of interpolation techniques including linear, spline, nearest neighbour and cubic interpolation. Interpolation can occur in 2D and 3D directions, nominally known as tri-linear or tri-cubic interpolations in 3D or bi-linear and bi-cubic interpolation in 2D.

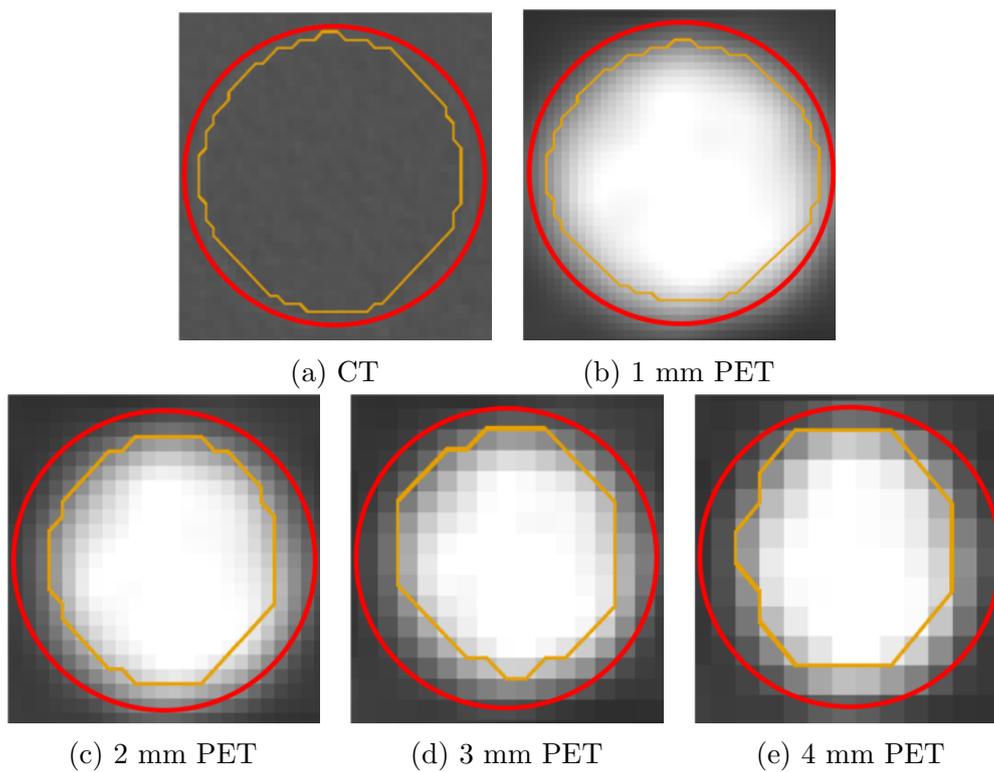


Figure 5.6: CT GT contour (orange) transferred to PET images interpolated to 1, 2, 3 and 4 mm voxel dimensions, demonstrating PVE. For comparison, a spherical contour (red) representing a perfect sphere was overlaid on the PET/CT images.

Linear Interpolation

The estimation of the value of an unknown data point between the values of two known data points as a function of a straight line is known as linear interpolation. Equation 5.22 defines linear interpolation mathematically whilst Figure 5.7 visualises the linear interpolation process. y is the unknown value at point x and y_0 and y_1 are the known values at points x_0 and x_1 . Given the values of $y_0 = 1$, $y_1 = 2$, $x_0 = 0$, $x_1 = 2$ and we want to find the value of y at point $x = 1$. Equation 5.23 to 5.26 shows the calculation of y at point $x = 1$.

$$y = \frac{y_0(x_1 - x) + y_1(x - x_0)}{(x_1 - x_0)} \quad (5.22)$$

$$\begin{aligned} p_1 &= y_0(x_1 - x) \\ &= y_0x_1 - y_0x \\ &= 2 - 1 \\ &= 1 \end{aligned} \quad (5.23)$$

$$\begin{aligned}
p2 &= y1(x - x0) \\
&= y1x - y1x0 \\
&= 2 - 0 \\
&= 2
\end{aligned}
\tag{5.24}$$

$$\begin{aligned}
p3 &= (x1 - x0) \\
&= 2 - 0 \\
&= 2
\end{aligned}
\tag{5.25}$$

$$\begin{aligned}
y &= \frac{p1 + p2}{p3} \\
&= \frac{1 + 2}{2} \\
&= 1.5
\end{aligned}
\tag{5.26}$$

2D Linear interpolation is an extension of linear interpolation. Linear interpolation is applied first in one direction between a set of four rectangular known data points and then applied in the orthogonal direction secondly. Whether linear interpolation is first applied in the x or y direction is respec-

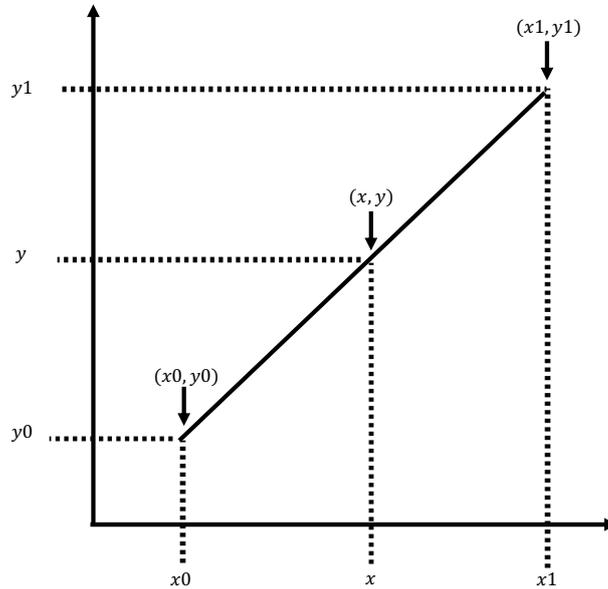


Figure 5.7: Linear interpolation between the value of two known data points to find the estimated value of the unknown data point.

tive of implementation choice. 3D linear interpolation is a further extension of 2D linear interpolation. As with 2D linear interpolation, whether interpolation occurs first in the x or y is respective of implementation design. However, linear interpolation in the z direction occurs once x , y data points have been interpolated.

Cubic Interpolation

Cubic interpolation in comparison to linear, bi-linear and tri-linear interpolation produces a smooth continuous estimation between known data points. Cubic interpolation requires the values of four known data points with two data points outside of two points to interpolate between. Cubic interpola-

tion therefore requires the use of four equations to define the value of the unknown data point x . These are defined in equations 5.27, 5.28, 5.29 with the final value of the unknown data point x shown in equation 5.30. y_0, y_1, y_2, y_3 are the values of the designated four known data points to interpolate between. x is the point of interpolation with a value of 0 equaling the first data point and a value of 1 equaling the second data point to be interpolated between.

$$a_0 = y_3 - y_2 - y_0 + y_1; \quad (5.27)$$

$$a_1 = y_0 - y_1 - a_0; \quad (5.28)$$

$$a_2 = y_2 - y_0; \quad (5.29)$$

$$a_0 \times x \times x^2 + a_1 \times x^2 + a_2 \times x + y_1 \quad (5.30)$$

Nearest Neighbour Interpolation

Nearest neighbour interpolation is the least computationally expensive form of interpolation and the most intuitive to implement. Nearest neighbour interpolation simply assigns the data point x the value of closest data point. The distance between points can be computed using euclidean distance. In 1D Nearest Neighbour interpolation the algorithm takes into account two data points. 2D and 3D nearest neighbour extend this into the x , y and z directions with nearest neighbour being applied in the x or y directions first dependent upon the design of the implementation. In effect, nearest neighbour interpolation up samples an image to a high resolution without estimation of unknown data points.

Spline Interpolation

Spline interpolation fits a smooth curve through all of the known data points.

The interpolation method chosen in this study was 3D linear interpolation (tri-linear), as it is considered to be data consistent and is computationally inexpensive.

5.1.7 Segmentation algorithms

The ATLAAS segmentation methodology incorporating the PET-AS methods, AT, KM clustering with 2, 3 and 4 clusters, GCM clustering with 3 and 4 clusters and WT methods was investigated in this study. For a detailed

description of ATLAAS see Chapter 2 Section 2.4.

5.1.8 Statistical Analysis

The 88 developed ATLAAS training datasets were used to derive MTVs on the validation dataset. The boundary region for applying ATLAAS to the validation dataset was defined as a 1 cm expansion of the GTV. The validation dataset was interpolated to the resolution of the training dataset. Therefore, a training dataset developed on a 4 mm isotropic voxel dimension was applied to a validation dataset with a 4 mm isotropic voxel dimension and a training dataset developed on a 1 mm XY voxel dimension with a 3 mm slice thickness was applied to a validation dataset with a 1 mm XY voxel dimension and 3 mm slice thickness. All training datasets consisted of the same PET MTVs and all validation datasets consisted of the same PET MTVs. Further to comparing the developed ATLAAS training models, the PET-AS methods AT, WT, KM with 2, 3 and 4 clusters and GCM with 3 and 4 clusters were applied to the validation dataset GT volumes. The resulting ATLAAS and PET-AS derived MTVs were compared to the highest-resolution GT volume available using DSC, as defined in Equation 2.3. Significant differences in DSC were assessed using Kruskal-Wallis ($P < 0.05$).

For each morphological feature included in the training dataset a one-tailed mann-whitney U test was used to test for significant improvement in the accuracy of the training model from a low-spatial resolution PET (4 mm isotropic voxel) to an increased spatial resolution PET (3 mm, 2 mm isotropic voxel

resolution and a 1 mm XY voxel resolution with a 3 mm slice thickness). A Kruskal-Wallis test was used to test for significant differences in the accuracy of the training models when additional morphological features were included in the training model. P values less than 0.05 were considered statistically significant. Statistical analysis was performed using a Matlab 2016b license. Image processing and statistics toolbox licenses were installed as well.

5.1.9 Results

A training dataset of 211 ^{18}F -FDG PET scans was interpolated to four differing voxel resolutions. 21 morphological features were included as additional parameters to the ATLAAS training model, providing in total 88 ATLAAS training models including the original ATLAAS training model developed using NI, TBR and volume (mL). The 88 ATLAAS training models were used to delineate an estimated MTV on the validation dataset consisting of 260 PET GTV's. ATLAAS derived contours were assessed visually and by calculating the DSC between the ATLAAS contour and the GT. Figure 5.8 on page 140 shows the median and range of DSC for ATLAAS applied to the validation dataset when trained with the parameters NI, MTV and TBR. Interpolating PET images in the training dataset and validation dataset improved the mean DSC for the parameters, NI, TBR and MTV by 19%, with a mean DSC for 4 mm PET imaging of 0.59 [0.16 — 0.86] and a mean DSC for 1 mm PET imaging of 0.78 [0.46 — 0.93].

In all of the developed models, interpolating PET imaging from a 4 mm isotropic voxel to a 1 mm XY, 3 mm slice thickness resulted in a mean im-

provement in DSC across all of the developed models of 20 %. Box plots showing the median, minimum and maximum DSC, for all of the developed models, are in Appendix C C.1. The best performing training model included the tumour and PET characteristics, NI, TBR, MTV and one of the following morphological features: compactness one, compactness two or sphericity. The models developed with compactness one, compactness two and sphericity each had a mean DSC of 0.81 with a range of 0.46 — 0.93. The correlation of the features compactness one, compactness two and sphericity are shown in Figures 5.9, 5.10 and 5.11, with the p-values and spearman rank correlation values of 0.00 and 1.00 respectively for each comparison. P-values of less than 0.05 reject the hypothesis that there is no correlation between the compared feature values.

Figure 5.12 on page 142 shows cases which had a greater than 200% difference in the MTV derived by ATLAAS and the MTV of the GT contour on the validation dataset and Figure 5.13 on page 143 shows cases which had a less than 200% difference in the MTV derived by ATLAAS and the MTV of the GT contour on the validation dataset. Descriptive statistics for the MTV of the contours derived by ATLAAS with the parameters NI, TBR and volume (mL) and the GT contours are shown in Table 5.1 on page 140. The mean MTV of the GT was 32.24 mL [1.39 — 174.96 mL].

Figure 5.14 on page 144 shows the box plots for the median, minimum and maximum DSC for the PET-AS methods AT, KM clustering with 2, 3 and 4 clusters, GCM clustering with 3 and 4 clusters and WT PET-AS methods. Independent Kruskal-Wallis tests showed a significant difference at the 5 %

significance level in the DSC of AT, WT, KM2, KM3, KM4, GCM3 and GCM4 PET-AS methods when applied to PET images at 4, 3 and 2 mm isotropic voxel dimensions and 1 mm XY voxel dimension with a 3 mm slice thickness (P-values < 0.05).

Figure 5.15 on page 145 shows ATLAAS applied to PET images with a 1 mm XY with a 3 mm slice thickness and 2 mm, 3 mm and 4 mm isotropic voxel dimensions compared to the GT contour obtained from the CT image. Figure 5.15a shows that the ATLAAS contour derived on 1 mm PET imaging was up to 2.90 mm under-contoured when compared to the GT. Figure 5.15b shows that the ATLAAS contour derived on 2 mm PET imaging was up to 1.95 mm under-contoured when compared to the GT. Figure 5.15c shows that the ATLAAS contour derived on 3 mm PET imaging was up to 2.70 mm under-contoured when compared to the GT. Figure 5.15d shows that the ATLAAS contour derived on 4 mm PET imaging was up to 2.98 mm over-contoured when compared to the GT.

Figures 5.16 and 5.17 on page 146 and page 147 respectively, show the contours derived by the PET-AS methods, AT, WT, KM2, KM3, KM4, GCM3 and GCM4 PET-AS methods in a pear phantom insert PET image interpolated to a 1 mm XY voxel dimensions with a 3 mm slice thickness and Figure 5.18 on page 148 shows the contours derived by the PET-AS method KM with 2 clusters. KM2 was the highest-performing individual PET-AS method on 1 mm XY voxel dimensions with a 3 mm slice thickness, with a mean DSC of 0.83.

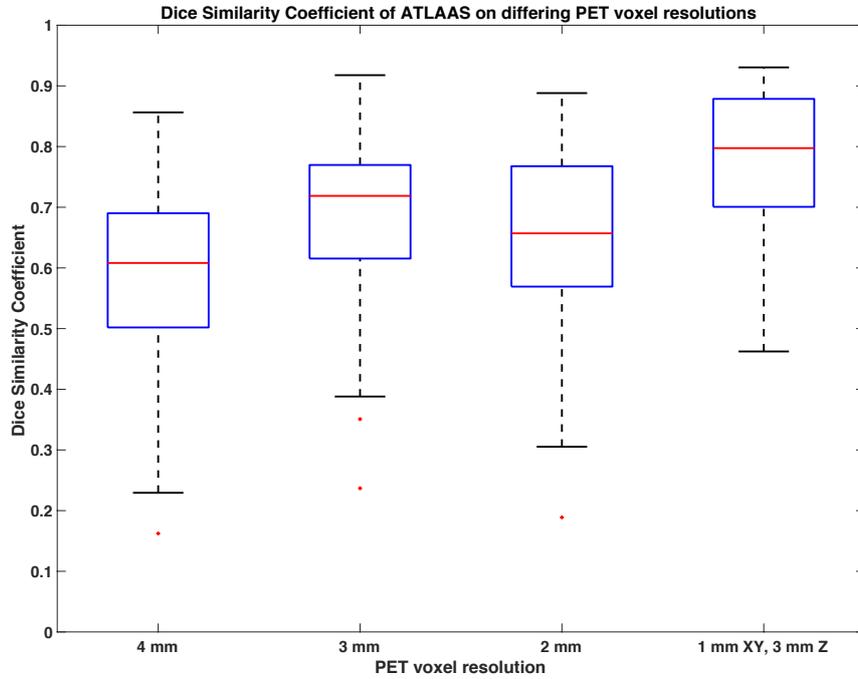


Figure 5.8: DSC of ATLAAS with parameters NI, TBR and MTV applied to the validation dataset

Table 5.1: Descriptive statistics for the volumes derived by ATLAAS with the parameters NI, TBR and volume and the GT contour.

	Ground Truth (mL)	ATLAAS one mm (mL)	ATLAAS two mm (mL)	ATLAAS three mm (mL)	ATLAAS four mm (mL)
Mean	32.24	51.28	52.07	49.83	45.83
Minimum	1.39	2.18	1.73	1.73	3.07
Maximum	174.96	195.05	195.11	175.45	163.07
Standard deviation	38.34	57.76	55.22	52.42	49.25
Median	19.87	33.85	33.96	30.98	26.88

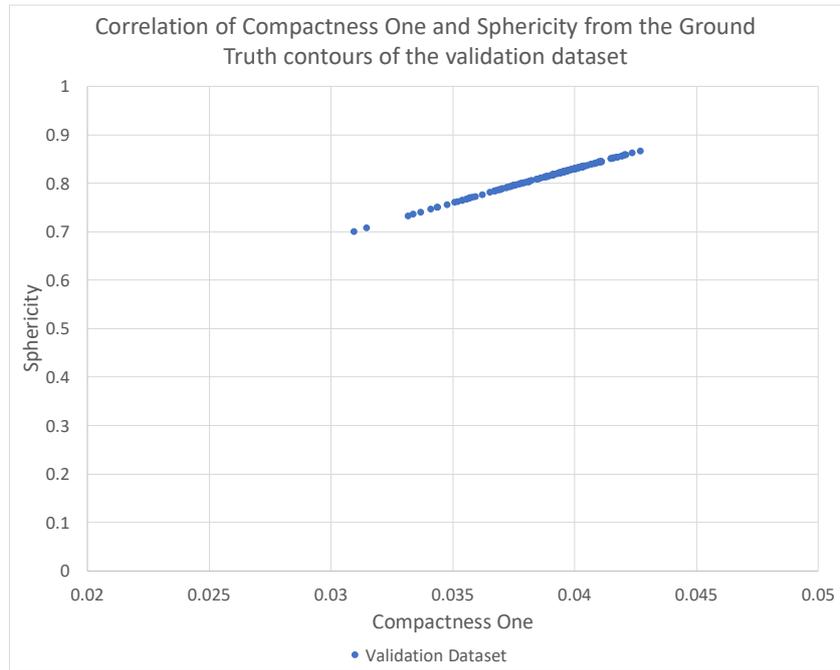


Figure 5.9: The correlation of compactness one and sphericity ground truth feature values in the validation dataset.

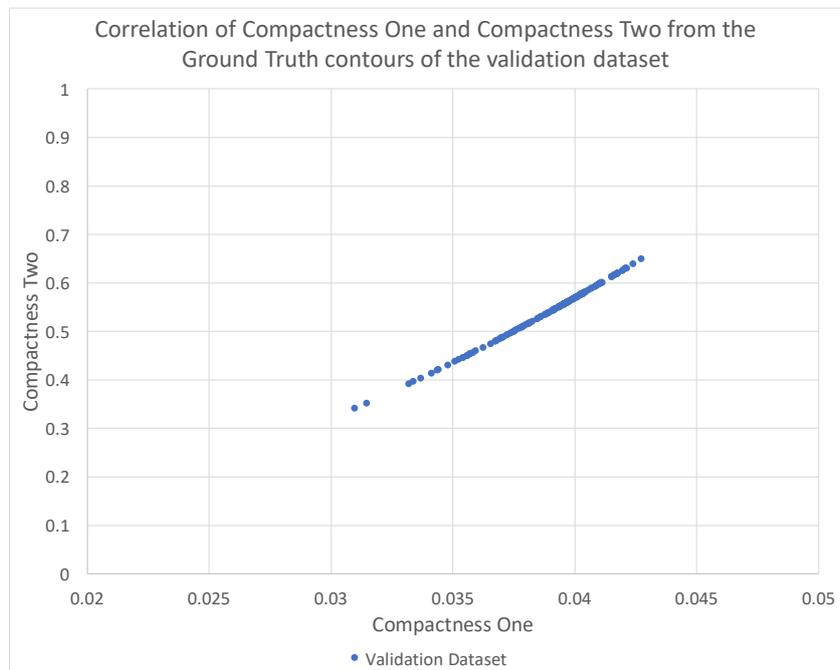


Figure 5.10: The correlation of compactness one and compactness two ground truth feature values in the validation dataset.

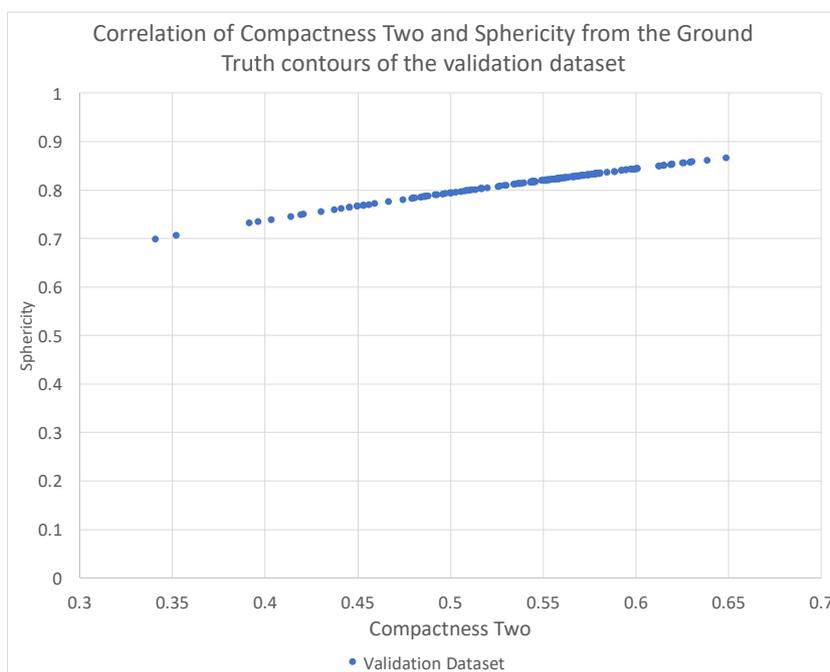


Figure 5.11: The correlation of compactness two and sphericity ground truth feature values in the validation dataset.

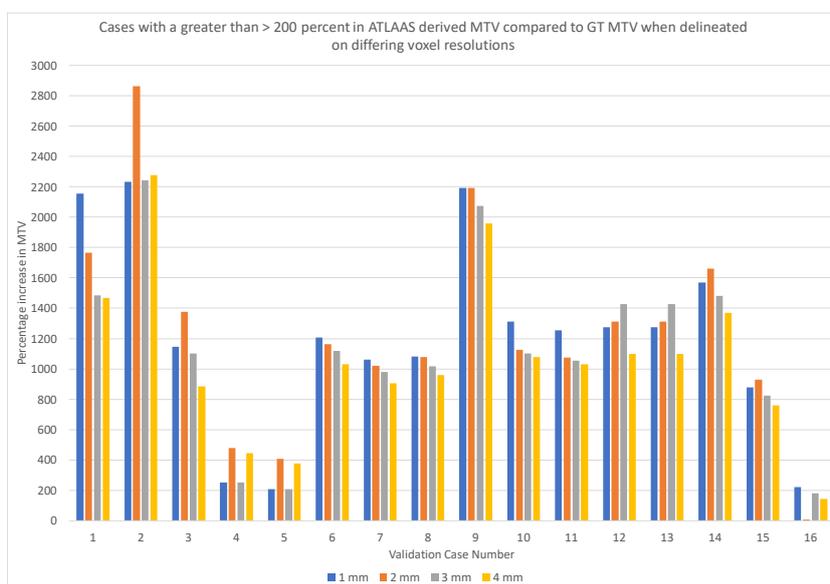


Figure 5.12: Cases with a greater than 200% difference in the volume derived by ATLAAS with parameters NI, TBR and MTV and the volume of the GT contour.

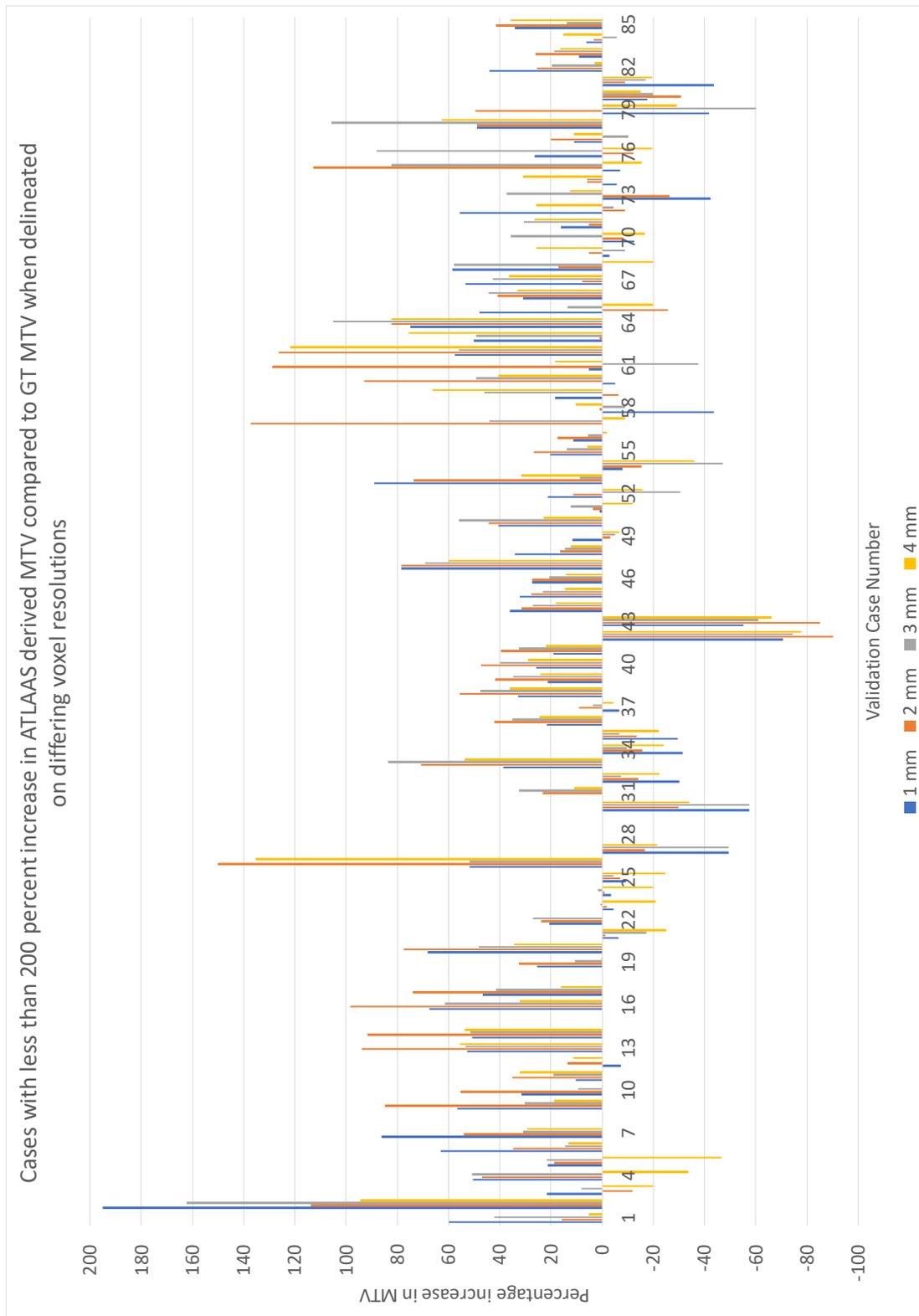
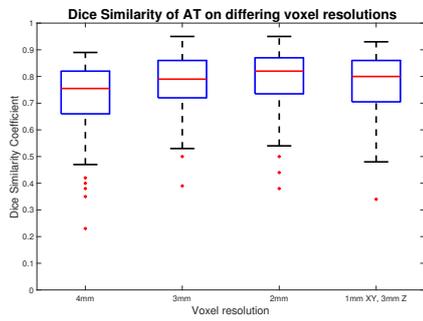
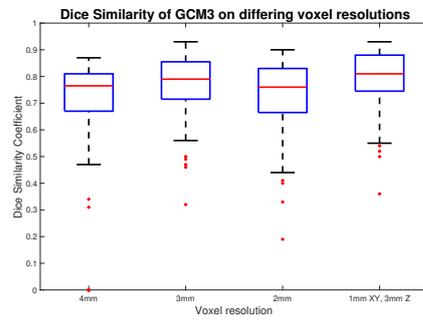


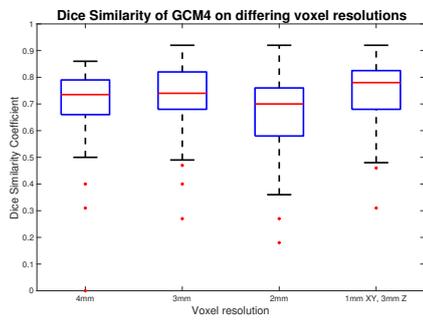
Figure 5.13: Cases with a less than 200% difference in the volume derived by ATLAAS with parameters NI, TBR and MTV and the volume of the GT contour.



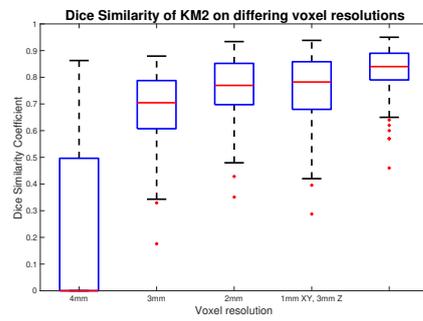
(a) AT



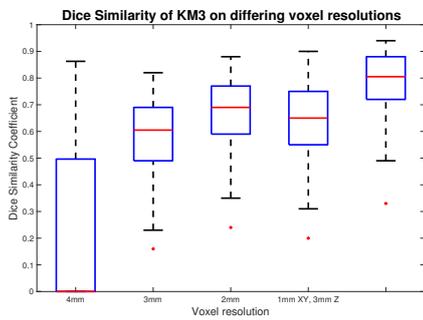
(b) GCM3



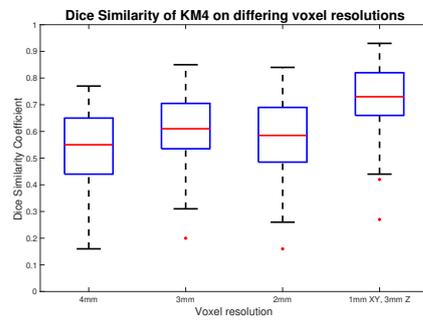
(c) GCM4



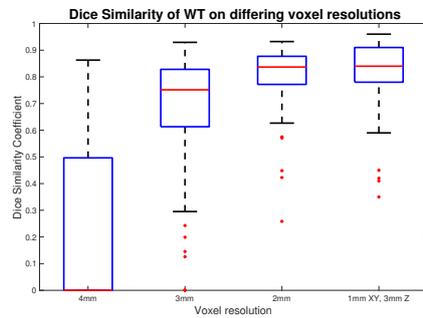
(d) KM2



(e) KM3



(f) KM4



(g) WT

Figure 5.14: DSC of PET-AS applied to PET images with 4 mm, 3 mm, 2 mm and 1 mm XY, 3 mm Z voxel dimensions

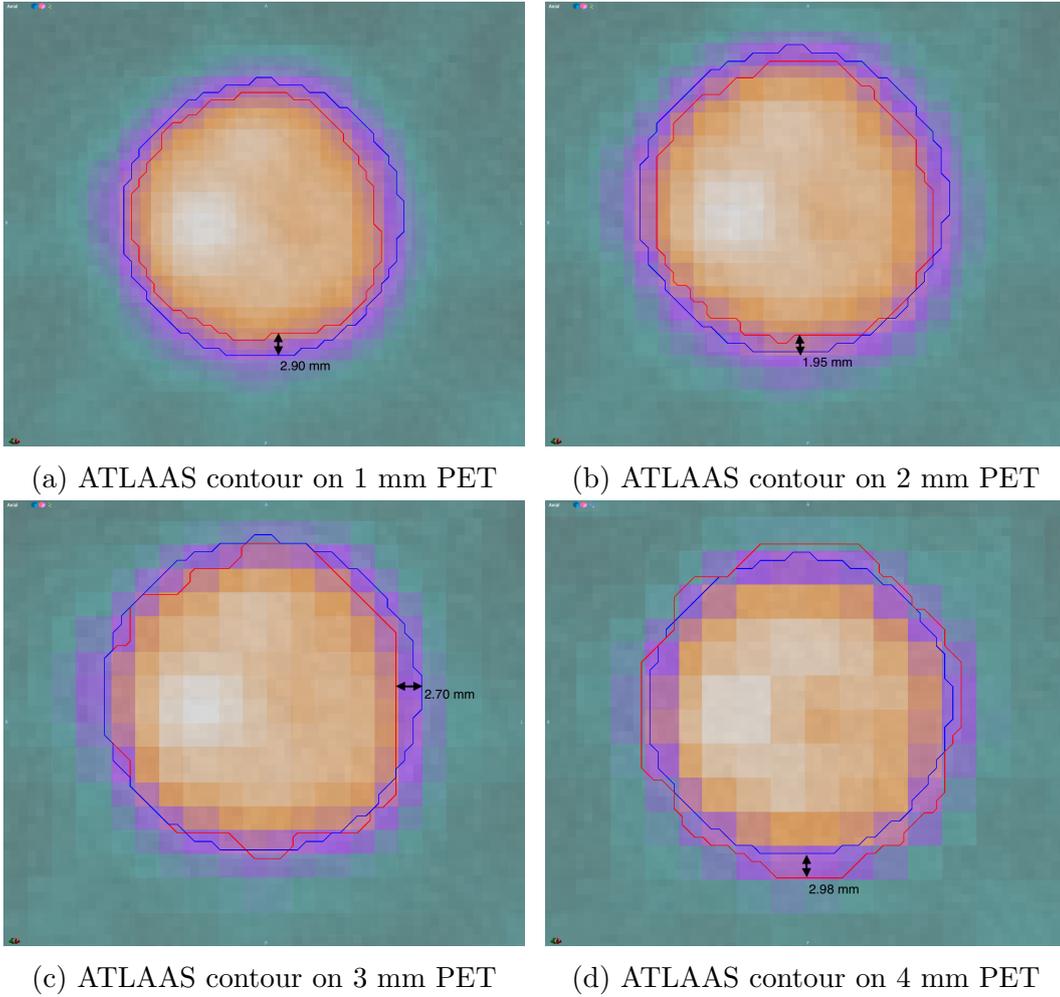
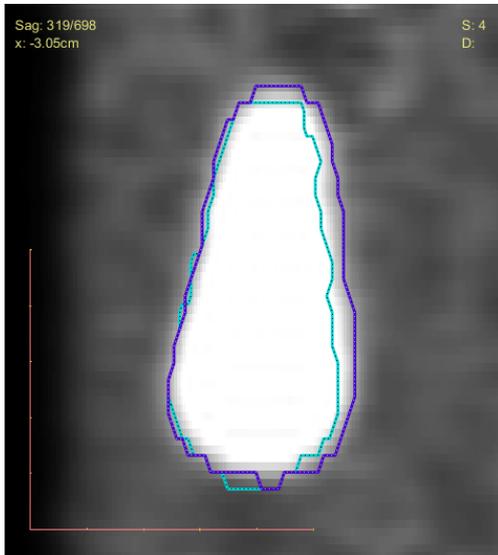


Figure 5.15: ATLAAS derived MTV in comparison to the high-resolution GT contour in a spherical object. ATLAAS contour is red and the GT contour is blue with measured distances between the two contours being 2.90 mm (1 mm PET), 1.95 mm (2 mm PET), 2.70 mm (3 mm PET), 2.98 mm (4 mm PET). Distances between ATLAAS and the GT contour to on 1 mm PET imaging.



(a) GCM3



(b) GCM4

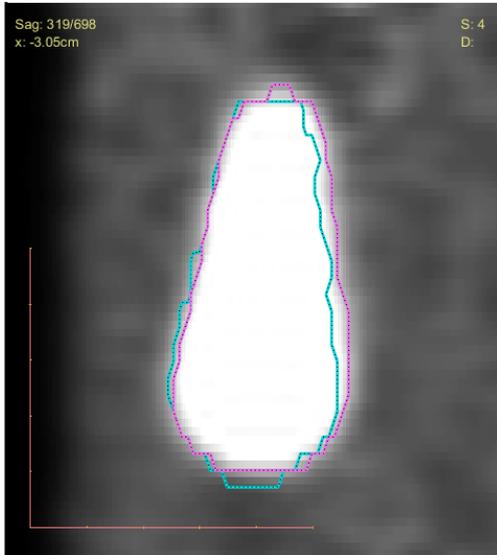


(c) WT

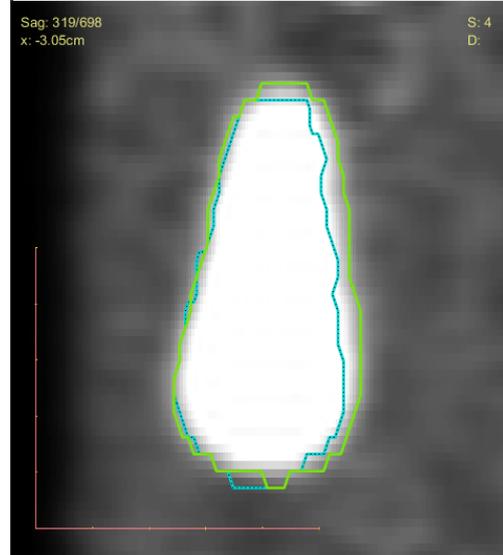


(d) KM2

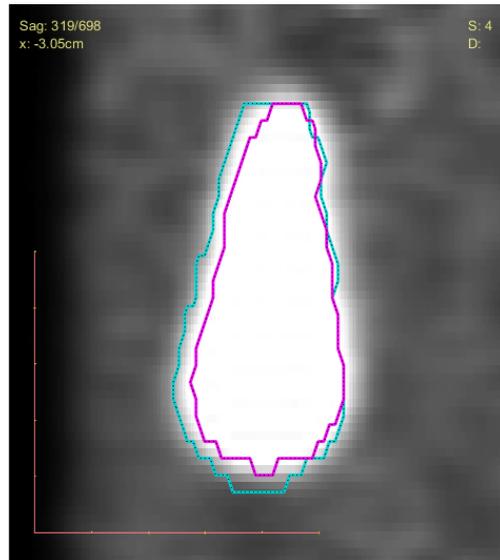
Figure 5.16: PET-AS derived MTV in comparison to the high-resolution GT contour in a Pear PET insert. The GT contour is turquoise, GCM3 (purple), GCM4 (dark green), WT (gold) and KM2 (green)



(a) KM3



(b) KM4



(c) AT

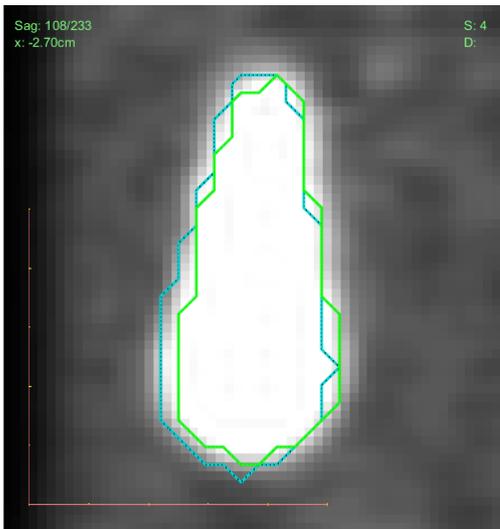
Figure 5.17: PET-AS derived MTV in comparison to the high-resolution GT contour in a Pear PET insert. The GT contour is turquoise, KM3 (pink), KM4 (lime green) and AT (purple)



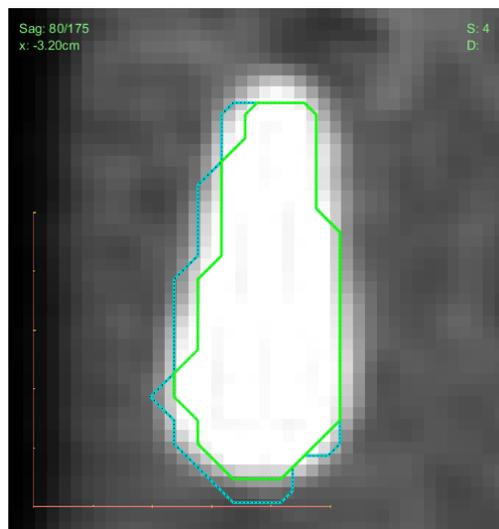
(a) KM2 1 mm



(b) KM2 2 mm



(c) KM2 3 mm



(d) KM2 4 mm

Figure 5.18: KM2 derived MTV in comparison to the high-resolution GT contour in a Pear PET insert. The GT contour is turquoise and KM2 is green.

5.2 Discussion

Investigating the role of additional classifiers in the development of machine-learning PET-AS training models is important for accurate MTV delineation. Additional classifiers, including morphological features, may improve the delineation of complex tumours. Previous work has shown that tumour characteristics and PET imaging characteristics influence the performance of PET-AS methods. Therefore, the use of PET-AS methods in RT planning may have an impact on the dosage to the OAR or patient risk-stratification, or OS. This work aimed to investigate the currently available morphological features which describe a tumours shape for inclusion in machine-learned DTs. Therefore, 22 morphological features were used to develop DTs at four different voxel dimensions to assess the influence of the PVE on the computation of morphological features. The performance of each developed training model was analysed using the DSC.

The results of this experiment show that morphological features are classifiers, in addition to those already known, in the development of machine-learned models for accurate MTV delineation. This is due to individual PET-AS algorithms performing differently in complex shapes. Further, the results show that the resolution of the PET image influences the accuracy of PET-AS delineation. In this study, delineated MTVs were compared against high-resolution GTs derived on a 1 mm XY voxel dimension with a 3 mm slice thickness; however, training models developed with a 2 mm isotropic voxel dimension had a reduced performance in comparison to models developed at a 3 mm isotropic voxel resolution. The reduction in model accuracy is

caused by the comparison of the 2 mm isotropic contour to the GT contour, which requires the estimation of the contour on the 3 mm slice thickness. In comparison, a MTV delineated on a 1 mm slice thickness or a 3 mm slice thickness does not require estimation of the contour at the 3 mm slice thickness. The best performing training model included the tumour and PET characteristics, NI, TBR, MTV and one of the following morphological features: compactness one, compactness two or sphericity. Compactness one, two and sphericity produced the same DSCs in the validation dataset and their feature values were highly correlated (Spearman rank correlation = 1.00, p-value = 0.00), therefore they can be used in the training model interchangeably.

Seven advanced individual PET-AS methods were also included in this study and applied to the validation dataset PET imaging which was interpolated to a 1 mm XY voxel dimension with a 3 mm slice thickness and 2 mm, 3 mm and 4 mm isotropic voxels. Interpolation to a 1 mm XY voxel dimension with a 3 mm slice thickness improved the performance of PET-AS delineation for individual PET-AS methods as well as the ATLAAS segmentation methodology. This work also demonstrates the degradation of a contour transferred from a high resolution CT PET to a lower resolution PET image. Therefore, interpolation of the PET should be to the highest resolution of the imaging modality available at the time of acquisition.

5.3 Conclusion

The morphological features compactness one, two and sphericity are additional classifiers compared to the already known parameters, MTV, TBR and NI. However, MTV delineation accuracy with PET-AS methods is dependent upon the voxel dimensions of the PET image. Interpolation of the PET image to a higher resolution improved PET-AS MTV delineation. In RT planning, improved MTV delineation accuracy is crucial for the standardisation of MTV delineation, as well as to reduce dosage to the OAR. This is especially critical in IMRT where steep dose gradients decrease the margin for error. In the following chapter, the results and conclusion drawn from this chapter, are related to the conclusions drawn throughout this body of work and within the context of the literature.

Chapter 6

Discussion & Conclusion

6.1 Discussion

Chapter 1 provides an introduction to RT of cancer and outlines the hypothesis for this body of work. The hypothesis was, the machine-learned PET-AS methodology called ATLAAS would have a higher level of accuracy in comparison to advanced PET-AS methods, for MTV delineation. In order to investigate this hypothesis fully, the aims and objectives of the TITAN project were:

- Improve the performance of the ATLAAS segmentation methodology.
- Externally validate the performance of the ATLAAS segmentation methodology.
- Investigate the effect of PET-AS methods on the development of prognostic models, therefore demonstrating the requirement for an agreed

upon PET-AS methodology in planning RT.

- Investigate the role and robustness of the ATLAAS segmentation methodology for intra-treatment MTV delineation.
- Investigate the impact of morphological features on the performance of ATLAAS.

In order to identify gaps within current knowledge, a literature review of MTV delineation using PET-AS methods was undertaken and presented in Chapter 2. This identified that whilst PET-AS methods had been investigated for suitable MTV delineation, the impact they have on patient risk stratification when curating large datasets was unknown. Therefore, Chapter 3 aimed to address this issue and established a need for an agreed upon method for MTV delineation. This was accomplished by applying 9 PET-AS methods to the same clinical data, in order to derive different MTVs. Prognostic models were developed from radiomic features extracted from the derived MTVs and were used to risk stratify patients. The results showed that sub-groups of patients have the potential to be denied beneficial treatment and sub-groups of patients may receive unnecessary and futile treatments. These treatments could reduce the quality of life of a patient.

In the literature review (Chapter 2), it was also identified that the role and robustness of PET-AS methods for suitable MTV delineation had not been investigated in intra-treatment PET imaging. Therefore, Chapter 4 was designed to address this issue and the aim was achieved by delineating MTVs on patient's PET scans, which were acquired after one cycle of ICT. The

PET scans were identified from a phase I, multi-centre feasibility trial called FiGaRO. The results of Chapter 4 show that the performance of machine-learning algorithms in scenarios they have not been trained for is limited. However, the performance can be improved by the simulation of additional data. Additionally, the results show that the performance of ATLAAS was higher in-comparison to the advanced and simplistic PET-AS methods included in the chapter. Chapter 4 also achieves the aim of externally validating ATLAAS by using data identified from a multi-centre feasibility trial. Further, the results support the hypothesis that ATLAAS has a higher level of accuracy in comparison to advanced PET-AS methods.

Additionally, in order to meet the aim of improving the performance of the ATLAAS PET-AS methodology, Chapter 5 had the objective of investigating the impact of morphological features on the performance of ATLAAS. To achieve this, 88 training models were developed. The models were developed by individually including 22 morphological features as classifiers. The morphological features were extracted from PET imaging interpolated to 4 different voxel dimensions. The results of Chapter 5 show that interpolating PET to higher resolutions improves the performance of PET-AS methods and that morphological features potentially can improve machine-learned MTV delineation.

A range of PET-AS methods, including the machine-learned PET-AS methodology ATLAAS have been evaluated and compared throughout this project; with the critical analysis techniques and domain knowledge acquired (during the project) being used to optimise the performance of the ATLAAS segmen-

tation methodology. The optimisation of ATLAAS was achieved through the development of new training datasets, which are optimised for low TBR scenarios (Chapter 4). Additionally, in Chapter 5, morphological features were investigated for influencing PET-AS performance and the significant features highlighted during the chapter were included in the development of additional new high-resolution training datasets.

In 2014, the IAEA published a report stating that no single PET-AS method could be recommended for accurate MTV delineation [31]; more recently, a report was published by the AAPM Task Group No 211 [1] stating that no single PET-AS method can be recommended for accurate MTV delineation; further, the report states that machine-learning techniques are showing promise for accurate MTV delineation. The analysis of the results presented throughout this body of work support the joint recommendations of the AAPM Task Group No 211 [1] and the IAEA [31] in that no single PET-AS method is suitable for MTV delineation in all cases. Additionally, this body of work supports the AAPM task group's recommendations that machine-learning techniques are showing promise for accurate MTV delineation by showing that additional (morphological features) classifiers, higher-resolution PET scans and optimisation of training datasets can improve machine-learning PET-AS MTV delineation. This work has shown that in comparison to advanced PET-AS algorithms and simplistic PET-AS algorithms, the machine-learned ATLAAS segmentation methodology was found to be more accurate for MTV delineation (Chapter 4).

Whilst this body of work represents a comprehensive evaluation of the AT-

LAAS segmentation methodology, ATLAAS has not been compared to other machine-learning methodologies including random forest, SVM or deep-learned neural networks. Currently these technologies are being developed in our centre and therefore future work aims to evaluate the efficacy and performance of these methods. Deep-learned neural networks may prove in investigations to have a higher accuracy in comparison to traditional PET-AS techniques and less complex machine-learned PET-AS segmentation methodologies. However, the development and training of the deep-learned neural networks requires extensive hardware requirements in the form of external or dedicated GPUs or multi-core / processor servers. Therefore, these techniques may be of limited use in the clinical environment where the development of new training datasets or neural networks may be required, dependent upon the needs of the clinic. Further, this evaluation is critical as in the 2017 Medical Image Computing and Computer Assisted Intervention (MICCAI) PET tumour segmentation challenge [130], convolution neural networks were the best performing PET-AS method.

The MICCAI challenge in 2017 was the first extensive evaluation of PET-AS methods involving 16 teams and a total of 10 PET-AS methods. However it does not provide a thorough analysis of all available PET-AS methods in all clinical scenarios, anatomical locations or all radiotracers. There is still a need, therefore, for a comprehensive and standardised dataset for the accurate evaluation of PET-AS methods [131], across all anatomical locations and the differing radiotracers in PET imaging. The development of a benchmark dataset will allow for the development of an agreed upon PET-AS method

which is increasingly important as adaptive RT and RT planning after ICT become more popular and increasingly used techniques. An agreed upon PET-AS method for MTV delineation also has implications for prognostic research.

Increasingly, prognostic research is being investigated to ensure patients are treated appropriately and to improve patient OS. However, prognostic research requires the curation of large datasets which are subject to inter and intra-observer variability of the MTV. In Chapter 3, the role of PET-AS methods for the curation of datasets and the development of prognostic models was investigated. The results demonstrate that a agreed upon approach to the delineation of the MTV is required as differences in MTV delineation can cause patients to change risk-stratification and therefore be potentially treated with unnecessary and aggressive therapies. The results of the investigation in Chapter 3 are strengthened by a large patient cohort $n = 472$. In Chapter 3 radiomic features were extracted using SUV bins of 0.5 units. However, a further extension of the study could investigate how different discretisation methods influence the significance of radiomic features in the development of prognostic models and subsequent impact on risk stratification and OS in patients with OC.

With increasing interest in multi-modality treatment pathways it is crucial that the MTV is delineated accurately in planning PET/CT imaging acquired during treatment in order to reduce inter and intra-observer variability, ensure conformal dose to the MTV and to reduce dosage to the OAR. Therefore, 16 PET-AS methods, including ATLAAS were investigated for

accurate MTV delineation in a range of low TBR scenarios and MTVs. Of the 16 PET-AS methods, only ATLAAS showed acceptable performance in the range of scenarios. Therefore, ATLAAS has been proposed for the standardised delineation (agreed upon PET-AS delineation method) of the MTV during treatment. However, the results of the investigation in Chapter 4 are limited by a small patient cohort ($n = 21$). The external validation of the ATLAAS segmentation methodology undertaken as part of this body of work is crucial for acceptance of ATLAAS, as an agreed upon method for MTV delineation in the clinical environment (Chapter 4).

Throughout this thesis additional training datasets were generated using the PETSTEP simulator. When PETSTEP is supplied with a PET scan with specific radiotracer characteristics (for example, ^{18}F -FDG), the resulting simulated PET scans exhibit the same tumour radiotracer uptake characteristics. Therefore, by supplying PETSTEP with a PET scan with differing radiotracer characteristics it is possible to simulate additional radiotracers. Due to the increased tumour characteristics that radiotracers are dependent upon, in order to reduce overfitting of the training model, ATLAAS potentially requires a separate training dataset for each specific radiotracer. This is especially important as the best performing PET-AS methods in ^{18}F -FDG PET imaging may be different in ^{11}C -Ch or ^{15}O based PET scans. In comparison, when applying ATLAAS to different anatomical sites, the anatomical location of the tumour (e.g H&N, OC or prostate) can be included in the development of the ATLAAS decision trees as a categorical classifier in order to improve PET-AS accuracy. Additionally, different treatments may require

different image processing techniques before MTV delineation. For example, inflammation caused by RT has the potential to impact the most appropriate PET-AS method for accurate MTV delineation when RT is re-planned (adaptively) throughout treatment.

6.2 Conclusion

The project described in this thesis aimed to investigate the performance of advanced automated segmentation methods for accurate MTV delineation. The following results were achieved throughout the duration of this project:

- Improved performance of the ATLAAS segmentation methodology in low TBR scenarios by the development of new datasets.
- Improved performance of the ATLAAS segmentation methodology by including additional tumour characteristics as classifiers.
- External validation of the performance of the ATLAAS segmentation methodology.
- Demonstrated improved performance of the ATLAAS segmentation methodology in comparison to 16 PET-AS methods for intra-treatment MTV delineation.
- Demonstrated PET-AS MTV delineation influences the development of prognostic models, therefore demonstrating the requirement for an agreed upon PET-AS methodology in planning RT and in the development of prognostic models.

- Demonstrated interpolation of PET imaging to a higher resolution, before MTV delineation, improves the accuracy of advanced PET-AS and machine-learned PET-AS.

The results presented in Chapters 4 to 5 demonstrate that the performance of the ATLAAS segmentation methodology is greater than the advanced PET-AS methods investigated and the results presented in Chapter 3 highlight the requirement for an agreed upon MTV delineation method. The work presented in this thesis has also paved way for further research into the application of the ATLAAS segmentation methodology in the planning and delivery of radiotherapy treatments at the Velindre Cancer Centre. Specifically, in a multi-centre clinical trial where the ATLAAS segmentation methodology will be used to prospectively delineate the MTV for adaptive radiotherapy treatment planning and delivery [132].

Bibliography

- [1] M. Hatt, J. A. Lee *et al.*, “Classification and evaluation strategies of auto-segmentation approaches for PET: Report of AAPM task group No. 211,” *Medical Physics*, vol. 44, no. 6, pp. e1–e42, may 2017. [Online]. Available: <http://doi.wiley.com/10.1002/mp.12124>
- [2] IARC, “World Cancer Report 2014,” World Health Organization, Tech. Rep., 2014. [Online]. Available: <http://publications.iarc.fr/Non-Series-Publications/World-Cancer-Reports/World-Cancer-Report-2014>
- [3] N. Vigneswaran and M. D. Williams, “Epidemiologic trends in head and neck cancer and aids in diagnosis,” *Oral and Maxillofacial Surgery Clinics of North America*, vol. 26, no. 2, pp. 123–141, 2014.
- [4] C. R. Leemans, B. J. Braakhuis *et al.*, “The molecular biology of head and neck cancer,” *Nat Rev Cancer*, vol. 11, no. 1, pp. 9–22, 2011. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/21160525>
- [5] “Oral cancer statistics — Cancer Research UK.” [Online]. Available: <http://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oral-cancer{\#}heading-One>

- [6] L. Welsh, R. Panek *et al.*, “Prospective, longitudinal, multi-modal functional imaging for radical chemo-IMRT treatment of locally advanced head and neck cancer: the INSIGHT study.” *Radiation oncology (London, England)*, vol. 10, no. 1, p. 112, 2015. [Online]. Available: <http://www.ro-journal.com/content/10/1/112>
- [7] A. R. Kreimer, G. M. Clifford *et al.*, “Human Papillomavirus Types in Head and Neck Squamous Cell Carcinomas Worldwide : A Systematic Review Human Papillomavirus Types in Head and Neck Squamous Cell Carcinomas Worldwide : A Systematic Review,” *Cancer Epidemiology, Biomarkers & Prevention*, vol. 14, no. February, pp. 467–475, 2005.
- [8] L. Sobin and C. Gospodarowicz, MK Wittekind, *UICC. TNM Classification of Malignant Tumours*. New York: Wiley, 2009.
- [9] K. K. Ang, J. Harris *et al.*, “Human Papillomavirus and Survival of Patients with Oropharyngeal Cancer,” *New England Journal of Medicine*, vol. 363, no. 1, pp. 24–35, jul 2010. [Online]. Available: <http://www.nejm.org/doi/abs/10.1056/NEJMoa0912217>
- [10] S. Lambot, E. C. Slob *et al.*, “Modeling of ground-penetrating radar for accurate characterization of subsurface electric properties,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 11, pp. 2555–2568, 2004.
- [11] “Oesophageal cancer survival statistics — Cancer Research UK,” 2014. [Online]. Available: <http://www.cancerresearchuk.org/health-professional/cancer->

statistics/statistics-by-cancer-type/oesophageal-cancer/survi

- [12] A. Y. B. Teoh, P. W. Y. Chiu *et al.*, “Long-term survival outcomes after definitive chemoradiation versus surgery in patients with resectable squamous carcinoma of the esophagus: Results from a randomized controlled trial,” *Annals of Oncology*, vol. 24, no. 1, pp. 165–171, 2013.
- [13] R. Corvò, “Evidence-based radiation oncology in head and neck squamous cell carcinoma,” *Radiotherapy and Oncology*, vol. 85, no. 1, pp. 156–170, 2007.
- [14] K. G. M. Moons, D. G. Altman *et al.*, “Prognosis and prognostic research: application and impact of prognostic models in clinical practice.” *BMJ (Clinical research ed.)*, vol. 338, no. 6, p. b606, 2009. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19502216>
- [15] M. Vallieres, A. Zwanenburg *et al.*, “Responsible Radiomics Research for Faster Clinical Translation,” *Journal of Nuclear Medicine*, p. jnumed.117.200501, 2017. [Online]. Available: <http://jnm.snmjournals.org/lookup/doi/10.2967/jnumed.117.200501>
- [16] K. G. Foley, R. K. Hills *et al.*, “Development and validation of a prognostic model incorporating texture analysis derived from standardised segmentation of PET in patients with oesophageal cancer,” *European Radiology*, vol. 28, no. 1, pp. 428–436, jan 2018. [Online]. Available: <http://link.springer.com/10.1007/s00330-017-4973-y>
- [17] M. Hatt, F. Tixier *et al.*, “Characterization of PET/CT images us-

- ing texture analysis: the past, the present... any future?" *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, no. 1, pp. 151–165, 2017.
- [18] M. Hatt, B. Laurent *et al.*, "Tumour functional sphericity from PET images: prognostic value in NSCLC and impact of delineation method," *European Journal of Nuclear Medicine and Molecular Imaging*, pp. 1–12, nov 2017. [Online]. Available: <http://link.springer.com/10.1007/s00259-017-3865-3>
- [19] I. Apostolova, J. Rogasch *et al.*, "Quantitative assessment of the asphericity of pretherapeutic FDG uptake as an independent predictor of outcome in NSCLC," *BMC Cancer*, vol. 14, no. 1, pp. 1–10, 2014.
- [20] J. M. Slater, *Ion Beam Therapy*, ser. Biological and Medical Physics, Biomedical Engineering, U. Linz, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, vol. 320. [Online]. Available: <http://link.springer.com/10.1007/978-3-642-21414-1>
- [21] H. S. Kaplan, "Basic principles in radiation oncology." *Cancer*, vol. 39, no. 2 Suppl, pp. 689–693, 1977.
- [22] H. Coutard, "Principles of X Ray Therapy of Malignant Diseases," *The Lancet*, vol. 224, no. 5784, pp. 1–8, jul 1934. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0140673600900850>
- [23] A. M. Chen, R. Chin *et al.*, "Inadequate target volume delineation and localregional recurrence after intensity-modulated radiotherapy for human papillomavirus-positive oropharynx cancer: Localregional recur-

- rence after IMRT for HPV-positive oropharynx cancer,” *Radiotherapy and Oncology*, vol. 123, no. 3, pp. 412–418, 2017. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0167814017301561>
- [24] N. C. A. Team, “National Radiotherapy Implementation Group Report: Image Guided Radiotherapy,” NHS, Tech. Rep., 2012. [Online]. Available: <https://www.sor.org/sites/default/files/document-versions/NationalRadiotherapyImplementationGroupReportIGRTFinal.pdf>
- [25] T. Landberg, J. Chavaudra *et al.*, “Report 50,” *Journal of the International Commission on Radiation Units and Measurements*, vol. os26, no. 1, pp. NP–NP, sep 1993. [Online]. Available: <https://academic.oup.com/jicru/article/2923876/Report>
- [26] —, “Report 62,” *Journal of the International Commission on Radiation Units and Measurements*, vol. os32, no. 1, p. NP, nov 1999. [Online]. Available: <https://doi.org/10.1093/jicru/os32.1.Report62>
- [27] P. Deluca, D. Jones *et al.*, “Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT),” *Journal of the ICRU*, vol. 10, no. 1, 2010.
- [28] H.-G. Menzel, “Report 91,” *Journal of the International Commission on Radiation Units and Measurements*, vol. 14, no. 2, pp. 1–160, 2014. [Online]. Available: <https://doi.org/10.1093/jicru/ndx017>
- [29] C. Njeh, “Tumor delineation: The weakest link in the search for accuracy in radiotherapy,” *Journal of Medical Physics*, vol. 33, no. 4,

- p. 136, 2008. [Online]. Available: <http://www.jmp.org.in/text.asp?2008/33/4/136/44472>
- [30] F. P. Vidal, F. Bello *et al.*, “Principles and Applications of Computer Graphics in Medicine,” *Computer Graphics Forum*, vol. 25, no. 1, pp. 113–137, 2006.
- [31] T. Konert, W. Vogel *et al.*, “PET/CT imaging for target volume delineation in curative intent radiotherapy of non-small cell lung cancer: IAEA consensus report 2014,” *Radiotherapy and Oncology*, vol. 116, no. 1, pp. 27–34, jul 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S016781401500153X><http://linkinghub.elsevier.com/retrieve/pii/S016781401500153X>
- [32] W. Xu, S. Yu *et al.*, “Effect of different segmentation algorithms on metabolic tumor volume measured on 18F-FDG PET/CT of cervical primary squamous cell carcinoma,” *Nuclear Medicine Communications*, vol. 38, no. 3, pp. 259–265, 2017. [Online]. Available: <http://insights.ovid.com/crossref?an=00006231-201703000-00010>
- [33] B. Berthon, C. Marshall *et al.*, “Evaluation of advanced automatic PET segmentation methods using nonspherical thin-wall inserts,” *Medical Physics*, vol. 41, no. 2, p. 022502, jan 2014. [Online]. Available: <http://doi.wiley.com/10.1118/1.4863480>
- [34] I. F. Ciernik, E. Dizendorf *et al.*, “Radiation treatment planning with an integrated positron emission and computer tomography (PET/CT): A feasibility study,” *International Journal of Radiation Oncology Bi-*

ology Physics, vol. 57, no. 3, pp. 853–863, 2003.

- [35] H. Zaidi and I. El Naqa, “PET-guided delineation of radiation therapy treatment volumes: A survey of image segmentation techniques,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 11, pp. 2165–2187, 2010.
- [36] B. Berthon, M. Evans *et al.*, “Head and neck target delineation using a novel PET automatic segmentation algorithm,” *Radiotherapy and Oncology*, vol. 122, no. 2, pp. 242–247, 2017. [Online]. Available: <http://dx.doi.org/10.1016/j.radonc.2016.12.008>
- [37] C. Parkinson, J. Chan *et al.*, “EP-1333: Impact of 18 F-Choline PET scan acquisition time on delineation of GTV in Prostate cancer,” *Radiotherapy and Oncology*, vol. 123, pp. S714–S715, may 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167814017317681?{\-}rdoc=1{\&}{_-}fmt=high{\&}{_-}origin=gateway{\&}{_-}docanchor={\&}md5=b8429449ccfc9c30159a5f9aeaa92ffb
- [38] C. Parkinson, K. Foley *et al.*, “Evaluation of prognostic models developed using standardised image features from different PET automated segmentation methods,” *EJNMMI Research*, vol. 8, no. 1, p. 29, dec 2018. [Online]. Available: <https://ejnmmires.springeropen.com/articles/10.1186/s13550-018-0379-3>
- [39] Myohan, “Bone planar scintigraphy.” [Online]. Available: <https://en.wikipedia.org/wiki/Bone{\-}scintigraphy{\#}/>

media/File:NI{_}bone{_}scan2.jpg

- [40] K. Clark, B. Vendt *et al.*, “The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository,” *Journal of Digital Imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.
- [41] National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC), “Radiology Data from the Clinical Proteomic Tumor Analysis Consortium Lung Adenocarcinoma [CPTAC-LUAD] collection,” *The Cancer Imaging Archive*, 2018.
- [42] School of Medicine University of Utah, “SPECT.” [Online]. Available: <https://medicine.utah.edu/radiology/radiology-research/learn/spect.php>
- [43] N. B. Smith and A. Webb, *Introduction to Medical Imaging*. New York, NY: Cambridge University Press, 2011.
- [44] A. Rahmim and H. Zaidi, “PET versus SPECT: strengths, limitations and challenges,” *Nuclear Medicine Communications*, vol. 29, no. 3, pp. 193–207, mar 2008. [Online]. Available: <https://insights.ovid.com/crossref?an=00006231-200803000-00002>
- [45] R. Bar-Shalom, N. Yefremov *et al.*, “Clinical performance of PET/CT in evaluation of cancer: additional value for diagnostic imaging and patient management.” *J Nucl Med*, vol. 44, no. 8, pp. 1200–1209, 2003.
- [46] J.-F. Daisne, T. Duprez *et al.*, “Tumor volume in pharyngolaryngeal squamous cell carcinoma: comparison at CT, MR imaging, and FDG

- PET and validation with surgical specimen.” *Radiology*, vol. 233, no. 1, pp. 93–100, 2004.
- [47] P.-P. Ypsilantis, M. Siddique *et al.*, “Predicting Response to Neoadjuvant Chemotherapy with PET Imaging Using Convolutional Neural Networks.” *PLoS ONE*, vol. 10, no. 9, p. e0137036, 2015. [Online]. Available: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0137036>
- [48] M. Niyazi, S. Landrock *et al.*, “Automated biological target volume delineation for radiotherapy treatment planning using FDG-PET/CT,” *Radiation Oncology*, vol. 8, no. 1, p. 1, 2013. [Online]. Available: RadiationOncology
- [49] J. Bradley, W. L. Thorstad *et al.*, “Impact of FDG-PET on radiation therapy volume delineation in non-small-cell lung cancer,” *International Journal of Radiation Oncology Biology Physics*, vol. 59, no. 1, pp. 78–86, 2004. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0360301603022673>
- [50] D. Lardinois, W. Weder *et al.*, “Staging of non-small-cell lung cancer with integrated positron-emission tomography and computed tomography.” *N Engl J Med*, vol. 348, no. 25, pp. 2500–7, 2003. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/24534208>
- [51] S.-C. Huang, R. E. Carson *et al.*, “Quantitative Measurement of Local Cerebral Blood Flow in Humans by Positron Computed Tomography and ¹⁵O-Water,” *Journal of Cerebral Blood Flow &*

- Metabolism*, vol. 3, no. 2, pp. 141–153, 1983. [Online]. Available: <http://journals.sagepub.com/doi/10.1038/jcbfm.1983.21>
- [52] A. L. Feliu, “The role of chemistry in positron emission tomography,” *Journal of Chemical Education*, vol. 65, no. 8, p. 655, 1988. [Online]. Available: <http://pubs.acs.org/doi/abs/10.1021/ed065p655>
- [53] F. J. Bonte, R. W. Parkey *et al.*, “Radionuclide determination of myocardial blood flow,” *Seminars in Nuclear Medicine*, vol. 3, no. 2, pp. 153–163, 1973.
- [54] D. W. Townsend, “Combined PET / CT : the historical perspective,” *Seminars in Ultrasound, CT, and MRI*, vol. 29, no. 4, pp. 232–235, 2009.
- [55] Penelope Allisy-Roberts and J. Williams, *Farr’s Physics for Medical Imaging*, 2nd ed. Saunders Ltd, 2007.
- [56] V. Kapoor, B. M. McCook *et al.*, “An Introduction to PET-CT Imaging,” *RadioGraphics*, vol. 24, no. 2, pp. 523–543, 2004. [Online]. Available: <http://pubs.rsna.org/doi/10.1148/rg.242025724>
- [57] G. B. Saha, *Basics of PET Imaging*. New York, NY: Springer New York, 2010. [Online]. Available: <http://link.springer.com/10.1007/978-1-4419-0805-6>
- [58] S. Cherry, J. Sorenson *et al.*, “Compton Scattering,” in *Physics in Nuclear Medicine*, 4th ed. Saunders, Elsevier, 2012, ch. 6.
- [59] K. Cooke, “Photoelectric absorption,” in *Practical Radiotherapy*:

Physics and Equipment, P. Cherry and A. Duxbury, Eds. Greenwich Medical Media LTD, 1998, ch. 5, p. 43.

- [60] T. K. Nayak and M. W. Brechbiel, “Radioimmunoimaging with longer-lived positron-emitting radionuclides: potentials and challenges.” *Bioconjugate chemistry*, vol. 20, no. 5, pp. 825–41, may 2009. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19125647><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3397469>
- [61] R. M. Pieterman, T. H. Que *et al.*, “Comparison of ¹¹C-choline and ¹⁸F-FDG PET in primary diagnosis and staging of patients with thoracic cancer,” *Journal of Nuclear Medicine*, vol. 43, no. 2, pp. 167–172, 2002. [Online]. Available: <http://www.embase.com/search/results?subaction=viewrecord{\&}from=export{\&}id=L34151038{\%}5Cnhttp://sfx.library.uu.nl/utrecht?sid=EMBASE{\&}issn=01615505{\&}id=doi:{\&}atitle=Comparison+of+11C-choline+and+18F-FDG+PET+in+primary+diagnosis+and+staging+of+patients+with+thoraci>
- [62] “NEMA PS3 / ISO 12052, Digital Imaging and Communications in Medicine (DICOM) Standard.”
- [63] C. Parkinson, L. Pike *et al.*, “Target volume delineation of FDG PET images post one cycle of induction chemotherapy in oropharyngeal cancer using advanced automated segmentation methods,” in *ESTRO 37*, 2018.
- [64] A. Schaefer, M. Vermandel *et al.*, “Impact of consensus contours from

- multiple PET segmentation methods on the accuracy of functional volume delineation,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 43, no. 5, pp. 911–924, 2016.
- [65] B. Berthon, C. Marshall *et al.*, “ATLAAS: An automatic decision tree-based learning algorithm for advanced image segmentation in positron emission tomography,” *Physics in Medicine and Biology*, vol. 61, no. 13, pp. 4855–4869, 2016. [Online]. Available: <http://stacks.iop.org/0031-9155/61/i=13/a=4855?key=crossref.bbce2194abc5f5075524903049f55627>
- [66] L. Drever, W. Roa *et al.*, “Iterative threshold segmentation for PET target volume delineation,” *Medical Physics*, vol. 34, no. 4, pp. 1253–1265, mar 2007. [Online]. Available: <http://doi.wiley.com/10.1118/1.2712043>
- [67] W. Jentzen, L. Freudenberg *et al.*, “Segmentation of PET volumes by iterative image thresholding.” *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 48, no. 1, pp. 108–114, 2007.
- [68] E. Day, J. Betler *et al.*, “A region growing method for tumor volume segmentation on PET images for rectal and anal cancer patients.” *Medical physics*, vol. 36, no. 10, pp. 4349–4358, sep 2009. [Online]. Available: <http://doi.wiley.com/10.1118/1.3213099>
- [69] X. Geets, J. A. Lee *et al.*, “A gradient-based method for segmenting FDG-PET images: Methodology and validation,” *European Journal of*

Nuclear Medicine and Molecular Imaging, vol. 34, no. 9, pp. 1427–1438, 2007.

- [70] P. Tyłski, G. Bonniaud *et al.*, “18F-FDG PET images segmentation using morphological watershed: a phantom study,” in *2006 IEEE Nuclear Science Symposium Conference Record*, vol. 4. IEEE, 2006, pp. 2063–2067. [Online]. Available: <http://ieeexplore.ieee.org/document/4179433/>
- [71] S. Belhassen and H. Zaidi, “A novel fuzzy C-means algorithm for unsupervised heterogeneous tumor quantification in PET.” *Medical physics*, vol. 37, no. 3, pp. 1309–1324, feb 2010. [Online]. Available: <http://doi.wiley.com/10.1118/1.3301610>
- [72] M. Hatt, C. Cheze le Rest *et al.*, “A Fuzzy Locally Adaptive Bayesian Segmentation Approach for Volume Determination in PET,” *IEEE Transactions on Medical Imaging*, vol. 28, no. 6, pp. 881–893, jun 2009. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/19150782><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2912931><http://ieeexplore.ieee.org/document/4749328/>
- [73] J. K. Udupa, V. R. LeBlanc *et al.*, “A framework for evaluating image segmentation algorithms,” *Computerized Medical Imaging and Graphics*, vol. 30, no. 2, pp. 75–87, 2006.
- [74] P. Tyłski, S. Stute *et al.*, “Comparative Assessment of Methods for Estimating Tumor Volume and Standardized Uptake Value in

- 18F-FDG PET,” *Journal of Nuclear Medicine*, vol. 51, no. 2, pp. 268–276, 2010. [Online]. Available: <http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.109.066241>
- [75] R. Boellaard, N. C. Krak *et al.*, “Effects of Noise, Image Resolution, and ROI Definition on the Accuracy of Standard Uptake Values: A Simulation Study,” *Journal of Nuclear Medicine*, vol. 45, no. 9, pp. 1519–1527, 2004. [Online]. Available: <http://jnm.snmjournals.org/content/45/9/1519.abstractN2->
- [76] R. L. Wahl, H. Jacene *et al.*, “From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors.” *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 50 Suppl 1, no. 5, pp. 122S–50S, 2009. [Online]. Available: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2755245{\&}tool=pmcentrez{\&}rendertype=abstract>
- [77] B. Berthon, C. Marshall *et al.*, “Influence of cold walls on PET image quantification and volume segmentation: A phantom study,” *Medical Physics*, vol. 40, no. 8, p. 082505, jul 2013. [Online]. Available: <http://doi.wiley.com/10.1118/1.4813302>
- [78] J. F. Daisne, M. Sibomana *et al.*, “Tri-dimensional automatic segmentation of PET volumes based on measured source-to-background ratios: Influence of reconstruction algorithms,” *Radiotherapy and Oncology*, vol. 69, no. 3, pp. 247–250, 2003.
- [79] U. Nestle, S. Kremp *et al.*, “Comparison of different methods for delin-

- eation of ^{18}F -FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-Small cell lung cancer,” *The Journal of Nuclear Medicine*, vol. 46, no. 8, pp. 1342–1348, 2005.
- [80] J. A. Lee, “Segmentation of positron emission tomography images: Some recommendations for target delineation in radiation oncology,” *Radiotherapy and Oncology*, vol. 96, no. 3, pp. 302–307, 2010. [Online]. Available: <http://dx.doi.org/10.1016/j.radonc.2010.07.003>
- [81] K. J. Biehl, F.-M. Kong *et al.*, “ ^{18}F -FDG PET definition of gross tumor volume for radiotherapy of non-small cell lung cancer: is a single standardized uptake value threshold approach appropriate?” *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 47, no. 11, pp. 1808–1812, 2006.
- [82] R. Szeliski, *Computer Vision*, ser. Texts in Computer Science. London: Springer London, 2011. [Online]. Available: <http://link.springer.com/10.1007/978-1-84882-935-0>
- [83] M. Abdoli, R. A. J. O. Dierckx *et al.*, “Contourlet-based active contour model for PET image segmentation,” *Medical Physics*, vol. 40, no. 8, p. 082507, jul 2013. [Online]. Available: <http://doi.wiley.com/10.1118/1.4816296>
- [84] M. Hatt, C. Cheze Le Rest *et al.*, “PET functional volume delineation: a robustness and repeatability study,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 38, no. 4, pp. 663–672, apr

2011. [Online]. Available: <http://link.springer.com/10.1007/s00259-010-1688-6>
- [85] S. Warfield, K. Zou *et al.*, “Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation,” *IEEE Transactions on Medical Imaging*, vol. 23, no. 7, pp. 903–921, jul 2004. [Online]. Available: <http://ieeexplore.ieee.org/document/1309714/>
- [86] R. J. McGurk, J. Bowsler *et al.*, “Combining multiple FDG-PET radiotherapy target segmentation methods to reduce the effect of variable performance of individual segmentation methods.” *Medical physics*, vol. 40, no. 4, pp. 1–9, apr 2013. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/23556917>
- [87] I. Goodfellow, Y. Bengio *et al.*, *Deep Learning*. MIT Press, 2016.
- [88] D. A. X. Schinagl, W. V. Vogel *et al.*, “Comparison of Five Segmentation Tools for 18F-Fluoro-Deoxy-Glucose-Positron Emission Tomography-Based Target Volume Definition in Head and Neck Cancer,” *International Journal of Radiation Oncology Biology Physics*, vol. 69, no. 4, pp. 1282–1289, 2007.
- [89] A. S. Dewalle-Vignion, N. Betrouni *et al.*, “A new method for volume segmentation of PET images, based on possibility theory,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 2, pp. 409–423, 2011.
- [90] J. O. Deasy, A. I. Blanco *et al.*, “CERR: a computational environment for radiotherapy research.” *Medical physics*, vol. 30, no. 2003, pp. 979–

985, 2003.

- [91] B. Berthon, I. Häggström *et al.*, “PETSTEP: Generation of synthetic PET lesions for fast evaluation of segmentation methods,” *Physica Medica*, vol. 31, no. 8, pp. 969–980, 2015. [Online]. Available: <http://dx.doi.org/10.1016/j.ejmp.2015.07.139>
- [92] M. Hatt, M. Majdoub *et al.*, “18F-FDG PET Uptake Characterization Through Texture Analysis: Investigating the Complementary Nature of Heterogeneity and Functional Tumor Volume in a MultiCancer Site Patient Cohort,” *The Journal of Nuclear Medicine*, vol. 56, no. 1, pp. 38–44, 2015. [Online]. Available: <http://jnm.snmjournals.org/content/56/1/38.abstractN2>
- [93] Y. Yue, A. Osipov *et al.*, “Identifying prognostic intratumor heterogeneity using pre- and post-radiotherapy 18F-FDG PET images for pancreatic cancer patients,” *Journal of Gastrointestinal Oncology*, vol. 8, no. 1, pp. 127–138, 2017. [Online]. Available: <http://jgo.amegroups.com/article/view/11971/10139>
- [94] S. S. F. Yip and H. J. W. L. Aerts, “Applications and limitations of radiomics.” *Physics in medicine and biology*, vol. 61, no. 13, pp. R150–66, 2016. [Online]. Available: <http://stacks.iop.org/0031-9155/61/i=13/a=R150?key=crossref.134478778713970aff90f16abe110608>{\% }5Cn<http://www.ncbi.nlm.nih.gov/pubmed/27269645>{\% }5Cn<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4927328>
- [95] M. Hatt, F. Tixier *et al.*, “Robustness of intratumour 18F-FDG PET

- uptake heterogeneity quantification for therapy response prediction in oesophageal carcinoma,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 40, no. 11, pp. 1662–1671, 2013.
- [96] F. Ben Bouallègue, Y. A. Tabaa *et al.*, “Association between textural and morphological tumor indices on baseline PET-CT and early metabolic response on interim PET-CT in bulky malignant lymphomas,” *Medical Physics*, vol. 44, no. 9, pp. 4608–4619, sep 2017. [Online]. Available: <http://doi.wiley.com/10.1002/mp.12349>
- [97] G. healthcare, “GE 690 PET/CT.” [Online]. Available: http://www3.gehealthcare.se/sv-se/products/kategorier/pet-ct/pet-ct{_}scanners/discovery{_}pet-ct{_}690
- [98] R. Gonzalez and R. Woods, “Image Segmentation,” in *Digital Image Processing*, second edi ed. New Jersey: Prentice Hall, 2002, ch. 10, pp. 612–613.
- [99] R. T. Leijenaar, G. Nalbantov *et al.*, “The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis,” *Scientific Reports*, vol. 5, no. 1, p. 11075, 2015. [Online]. Available: <http://www.nature.com/articles/srep11075>
- [100] A. R. Davies, J. A. Gossage *et al.*, “Tumor stage after neoadjuvant chemotherapy determines survival after surgery for adenocarcinoma of the esophagus and esophagogastric junction,” *Journal of Clinical Oncology*, vol. 32, no. 27, pp. 2983–2990, 2014.

- [101] F. Tixier, C. C. Le Rest *et al.*, “Intratumor Heterogeneity Characterized by Textural Features on Baseline 18F-FDG PET Images Predicts Response to Concomitant Radiochemotherapy in Esophageal Cancer,” *Journal of Nuclear Medicine*, vol. 52, no. 3, pp. 369–378, 2011. [Online]. Available: <http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.110.082404>
- [102] C. Yip, D. Landau *et al.*, “Primary Esophageal Cancer: Heterogeneity as Potential Prognostic Biomarker in Patients Treated with Definitive Chemotherapy and Radiation Therapy,” *Radiology*, vol. 270, no. 1, pp. 141–148, 2014. [Online]. Available: <http://pubs.rsna.org/doi/10.1148/radiol.13122869>
- [103] G. Doumou, M. Siddique *et al.*, “The precision of textural analysis in 18F-FDG-PET scans of oesophageal cancer,” *European Radiology*, vol. 25, no. 9, pp. 2805–2812, 2015. [Online]. Available: <http://link.springer.com/10.1007/s00330-015-3681-8>
- [104] R. J. Gillies, P. E. Kinahan *et al.*, “Radiomics: Images Are More than Pictures, They Are Data,” *Radiology*, vol. 278, no. 2, pp. 563–577, 2016. [Online]. Available: <http://pubs.rsna.org/doi/10.1148/radiol.2015151169>
- [105] C. Yip, F. Davnall *et al.*, “Assessment of changes in tumor heterogeneity following neoadjuvant chemotherapy in primary esophageal cancer,” *Diseases of the Esophagus*, vol. 28, no. 2, pp. 172–179, feb 2015. [Online]. Available: <https://academic.oup.com/dote/article->

lookup/doi/10.1111/dote.12170

- [106] M. Nakajo, M. Jinguji *et al.*, “Texture analysis of 18F-FDG PET/CT to predict tumour response and prognosis of patients with esophageal cancer treated by chemoradiotherapy,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, no. 2, pp. 206–214, 2017. [Online]. Available: <http://dx.doi.org/10.1007/s00259-016-3506-2>
- [107] M. Amadasun and R. King, “Textural Features Corresponding to Textural Properties,” *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19, no. 5, pp. 1264–1274, 1989. [Online]. Available: <http://ieeexplore.ieee.org/document/44046/>
- [108] A. Depeursinge, A. Foncubierta-Rodriguez *et al.*, “Three-dimensional solid texture analysis in biomedical imaging: Review and opportunities,” *Medical Image Analysis*, vol. 18, no. 1, pp. 176–196, 2014.
- [109] H. Akaike, “A New Look at the Statistical Model Identification,” *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [110] K. G. M. Moons, A. P. Kengne *et al.*, “Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker,” *Heart*, vol. 98, no. 9, pp. 683–690, 2012. [Online]. Available: <http://heart.bmj.com/lookup/doi/10.1136/heartjnl-2011-301246>
- [111] M. C. Desseroit, D. Visvikis *et al.*, “Development of a nomogram combining clinical staging with 18F-FDG PET/CT image features in non-small-cell lung cancer stage III,” *European Journal of Nuclear*

- Medicine and Molecular Imaging*, vol. 43, no. 8, pp. 1477–1485, 2016.
[Online]. Available: <http://dx.doi.org/10.1007/s00259-016-3325-5>
- [112] J. P. Crandall, A. K. Tahari *et al.*, “A comparison of FLT to FDG PET/CT in the early assessment of chemotherapy response in stages IBIIIA resectable NSCLC,” *EJNMMI Research*, vol. 7, no. 1, p. 8, 2017. [Online]. Available: <http://ejnmires.springeropen.com/articles/10.1186/s13550-017-0258-3>
- [113] Guy’s and St Thomas NHS Foundation Trust, “¹⁸F-FDG-PET Guided Dose-Painting With Intensity Modulated Radiotherapy in Oropharyngeal Tumours (FiGaRO), A Phase 1 Pilot Study,” 2016. [Online]. Available: <https://clinicaltrials.gov/ct2/show/NCT02953197>
- [114] Cancer Research UK, “RT shell.” [Online]. Available: <https://www.cancerresearchuk.org/about-cancer/cancer-in-general/treatment/radiotherapy/external/types/conformal>
- [115] Gehealthcare, “GE 710 PET/CT.” [Online]. Available: <http://www3.gehealthcare.in/en/products/categories/pet-ct/pet-ct-scanners/discovery-pet-ct-710>
- [116] N. Lee, P. Xia *et al.*, “Intensity-modulated radiation therapy for head-and-neck cancer: The UCSF experience focusing on target volume delineation,” *International Journal of Radiation Oncology Biology Physics*, vol. 57, no. 1, pp. 49–60, 2003.
- [117] S. K. Vinod, M. Min *et al.*, “A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology,” *Jour-*

- nal of Medical Imaging and Radiation Oncology*, vol. 60, no. 3, pp. 393–406, 2016.
- [118] A. I. J. Arens, E. G. C. Troost *et al.*, “Semiautomatic methods for segmentation of the proliferative tumour volume on sequential FLT PET/CT images in head and neck carcinomas and their relation to clinical outcome,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 41, no. 5, pp. 915–924, 2014.
- [119] A. Eisbruch, L. H. Marsh *et al.*, “Recurrences near base of skull after IMRT for head-and-neck cancer: Implications for target delineation in high neck and for parotid gland sparing,” *International Journal of Radiation Oncology Biology Physics*, vol. 59, no. 1, pp. 28–42, 2004.
- [120] L. A. Dawson, Y. Anzai *et al.*, “Patterns of local-regional recurrence following parotid-sparing conformal and segmental intensity-modulated radiotherapy for head and neck cancer,” *International Journal of Radiation Oncology Biology Physics*, vol. 46, no. 5, pp. 1117–1126, mar 2000. [Online]. Available: <http://linkinghub.elsevier.com/retrieve/pii/S0360301699005507>
- [121] C. Parkinson, M. Evans *et al.*, “Machine-learned target volume delineation of 18 F-FDG PET images after one cycle of induction chemotherapy,” *Physica Medica*, vol. 61, no. April, pp. 85–93, 2019.
- [122] E. A. Eisenhauer, P. Therasse *et al.*, “New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1),” *European Journal of Cancer*, vol. 45, no. 2, pp. 228–247, 2009. [Online].

Available: <http://dx.doi.org/10.1016/j.ejca.2008.10.026>

- [123] A. Zwanenburg, “Image biomarker standardisation initiative.” [Online]. Available: <https://arxiv.org/abs/1612.07003>
- [124] A. Zwanenburg, P. Whybra *et al.*, “Results from the Image Biomarker Standardisation Initiative,” in *ESTRO 37*, 2018.
- [125] C. C. Gray, S. F. Al-Maliki *et al.*, “Data exploration in evolutionary reconstruction of PET images,” *Genetic Programming and Evolvable Machines*, vol. 19, no. 3, pp. 391–419, 2018. [Online]. Available: <http://link.springer.com/10.1007/s10710-018-9330-7>
- [126] M. Soret, S. L. Bacharach *et al.*, “Partial-Volume Effect in PET Tumor Imaging,” *Journal of Nuclear Medicine*, vol. 48, no. 6, pp. 932–945, 2007. [Online]. Available: <http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.106.035774>
- [127] N. J. Hoetjes, F. H. P. Van Velden *et al.*, “Partial volume correction strategies for quantitative FDG PET in oncology,” *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 9, pp. 1679–1687, 2010.
- [128] K. Erlandsson, I. Buvat *et al.*, “A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology,” *Physics in Medicine and Biology*, vol. 57, no. 21, pp. R119–R159, 2012. [Online]. Available: <http://stacks.iop.org/0031-9155/57/i=21/a=R119?key=crossref.4613f99a63010e4337b3a316357415db>

- [129] O. U. Press, “Interpolation, n.” [Online]. Available: <http://www.oed.com/view/Entry/98178>
- [130] M. Hatt, B. Laurent *et al.*, “The first miccai challenge on pet tumor segmentation,” *Medical Image Analysis*, vol. 44, 12 2017.
- [131] B. Berthon, E. Spezi *et al.*, “Towards a standard for the evaluation of pet auto-segmentation methods: requirements and implementation,” *Medical physics*, vol. 44, 05 2017.
- [132] BBC, “PEARL Trial.” [Online]. Available: <https://www.bbc.co.uk/news/uk-wales-39566782>

Abbreviations & Acronyms

μ mean. 65–68, 70

¹¹C Carbon-11. 22

¹¹C-Ch ¹¹C-Choline. 22, 158

¹²⁴I Iodine-124. 22

¹³N Nitrogen-13. 22

¹⁵O Oxygen-15. 22, 158

¹⁸F Flourine-18. 22

¹⁸F-FDG ¹⁸F-Fluorodeoxyglucose. 3, 17, 22, 35, 37, 49, 51, 58, 60, 88, 89,
91, 94, 95, 98, 105, 110, 113, 137, 158, 193

⁵²Mn Manganese-52. 22

⁸²Rb Rubidium-82. 22

2D two dimensional. 16, 70, 129, 132, 133, 135

3D three dimensional. 69, 70, 128, 129, 133, 135

3D-CRT 3D-Conformal Radiotherapy. 13

5FU 5-Fluorouracil. 89

AAPM American Association for Physicists in Medicine. 46, 79, 101, 155

AIC Akaike information criterion. 71, 72, 75, 77

AT Adaptive iterative thresholding. 49, 51, 61, 62, 72–75, 77, 78, 101, 106, 108, 112, 135, 136, 138, 139

ATLAAS an automatic decision tree-based learning algorithm for advanced image segmentation. 4–7, 45, 47–50, 61, 79, 89, 95, 96, 98, 100, 101, 106, 108, 111–113, 116, 118, 128, 135–139, 150, 152–160

BIC Bayesian information criterion. 72

CERR computational environment for radiotherapy research. 47, 49, 61

COMShift Centre of mass shift. 124

CT Computed Tomography. 14, 16, 18, 19, 21, 44, 49, 50, 58–61, 88, 91, 94, 98, 105, 118, 127, 128, 139, 150, 157

CTV Clinical Target Volume. 15

dCRT Definitive chemoradiotherapy. 60, 61

DICOM Digital Imaging and Communications in Medicine. 25, 26

DSC dice similarity coefficient. 47, 50, 101, 105, 106, 112, 136–139, 149, 150

DT Decision tree. 42, 43, 45, 48–50, 149

EBRT external beam radiotherapy. 12, 13

EMR Endoscopic Mucosal Resection. 60

FCM Fuzzy C-means. 61, 66, 72, 75, 76, 78, 101, 108

FiGaRO 18F-FDG-PET Guided Dose-Painting With Intensity Modulated
Radiotherapy in Oropharyngeal Tumours. 89, 91, 94, 98, 100, 105, 154

FLAB Fuzzy Locally Adaptive Bayesian. 46

FOV field of view. 127

FWHM Full Width at Half Maximum. 94

GCM Gaussian Fuzzy C-means. 49, 61, 67, 72–77, 101, 106, 108, 135, 136,
138, 139

GLCM Grey Level Co-occurrence Matrix. 70

GLNU Grey Level Non-Uniformity. 70, 73, 77

GLSZM Grey Level Size Zone matrix. 70

GOJ Gastro-oesophageal junction. 58

GPU graphics processing unit. 61, 129, 156

GT ground truth. 50, 117, 118, 136–139, 149, 150

GTV Gross Tumour Volume. 14, 15, 26, 106, 136, 137

HPV Human Papilloma Virus. 8, 90

H&N Head and Neck. 2, 7, 8, 12, 26, 45, 88, 90, 91, 95, 110, 158

IAEA International Atomic Energy Agency. 47, 79, 155

IBSI Image Biomarker Standardisation Initiative. 69, 73, 121, 122, 127

ICRU International Commission on Radiation Units and Measurements. 14

ICT induction chemotherapy. 88–91, 94, 96, 97, 101, 110–113, 153, 157

IGRT Image guided radiotherapy. 13

IMRT Intensity modulated radiotherapy. 13, 90, 91, 94, 96, 110, 115, 151

KM K-means. 49, 61, 62, 68, 72–78, 101, 108, 135, 136, 138, 139

linacs electron linear accelerators. 5, 12, 91

LOR Line Of Response. 23, 24

MICCAI Medical Image Computing and Computer Assisted Intervention.
156

MJV majority vote. 41

MLEM maximum-likelihood expectation maximisation. 25

MR Magnetic Resonance. 44

MTV Metabolic Tumour Volume. 4, 6, 7, 12, 17, 33, 35–38, 40, 41, 44–47,
50–52, 57, 61, 69, 70, 72, 76–79, 88–90, 94–98, 100, 101, 105, 106, 108,
110–113, 115–118, 121, 136–138, 149–155, 157–160

NACRT Neo-adjuvant chemoradiotherapy. 60

NACT Neo-adjuvant chemotherapy. 60

NGTDM Neighborhood Grey-Tone Difference Matrix. 70

NI Number of discrete intensities. 48, 95, 96, 98, 100, 113, 117, 118, 137, 138, 150, 151

OAR organs at risk. 13, 14, 89, 110, 115, 149, 151, 157

OC Oesophageal Cancer. 7–9, 26, 56–58, 69, 79, 157, 158

OP Oropharyngeal. 88–90, 112

OS overall survival. 8, 9, 11, 17, 47, 52, 56, 57, 69, 71, 73, 74, 76–78, 149, 157

OSEM ordered subset expectation maximisation. 25, 59, 94

PACS picture archiving and communications system. 25

PET Positron Emission Tomography. 2–4, 7, 15–19, 21–23, 25, 26, 33, 35–39, 44–51, 58–62, 65, 68, 77, 79, 88–91, 94–98, 100, 101, 105, 110, 112, 113, 117, 118, 127–129, 136–139, 149–151, 153–158, 160, 193

PET-AS PET based automatic segmentation. 2–4, 6, 7, 26, 33–36, 38, 41, 44–51, 57, 61, 62, 72, 75–79, 89, 101, 105, 106, 108, 111–113, 115, 116, 129, 135, 136, 138, 139, 149–160

PETIC PET imaging centre. 5, 66, 113

PETSTEP PET Simulator of Tracers via Emission Projection. 49, 97, 98, 158

PMT photomultiplier tubes. 19, 128

PS Planar Scinitgraphy. 15, 16

PSF Point Spread Function. 98

PTV Planning Target Volume. 15

PVE partial volume effect. 128, 149

RECIST Response Evaluation Criteria In Solid Tumours. 116

RG Region Growing. 62, 66, 72, 75, 76, 78, 101, 108

ROI region of interest. 50, 51, 61, 62

RT Radiotherapy. 1, 3, 5, 6, 8, 12–15, 26, 51, 79, 88–91, 94–96, 110, 149, 151–153, 157, 159

SCC squamous cell carcinoma. 8, 9, 88–90, 112

SD Standard Deviation. 40, 66, 67, 105, 106

SNR signal to noise ratio. 16, 19, 21

SPECT Single Photon Emission Computed Tomography. 15, 16

STAPLE simultaneous truth and performance level estimation. 41

SUV Standardised Uptake Value. 36, 48, 65, 68, 95, 157

SUV_{max} maximum SUV. 37, 38, 40, 48, 64–66, 96, 105, 108, 112, 113

SUV_{peak} peak SUV. 37, 40, 48, 96, 101, 105, 108

SVM Support Vector Machine. 43, 116, 156

TBR tumour to background ratio. 7, 45, 95, 96, 98, 100, 105, 108, 112, 113, 117, 118, 137, 138, 150, 151, 155, 158, 159

TNM Tumour, Node and Metastasis. 8–11, 58, 60, 77, 90

TOF Time of Flight. 23, 94

UICC International Union Against Cancer. 58

VCC Velindre Cancer Centre. 5, 113

VOI volume of interest. 36

WHO World Health Organisation. 116

WT Watershed Transform. 49, 62, 65, 72–75, 77, 101, 106, 108, 135, 136, 138, 139

Appendices

Appendix A

Publications

A.1 Journal Papers

- Parkinson C, Evans M, Guerrero-Urbano T, Michaelidou A, Pike L, Barrington S, Jayaprakasam V, Rackley T, Palaniappan N, Staffurth J, Marshall C, Spezi E (2019) Machine-learned target volume delineation of ^{18}F -FDG PET images after one cycle of induction chemotherapy. *Physica Medica* Volume 61, Pages 85-93. (DOI: 10.1016/j.ejmp.2019.04.020).
- Foley K, Shi Z, Whybra P, Kalendralis P, Larue R, Berbee M, Sosef M, Parkinson C, Staffurth J, Crosby T, Roberts S, Dekker A, Wee L, Spezi E (2018) External validation of a prognostic model incorporating quantitative PET image features in oesophageal cancer. *Radiotherapy and Oncology*. (DOI: 10.1016/j.radonc.2018.10.033).
- Foley K, Hills R, Berthon B, Marshall C, Parkinson C, Lewis W, Crosby

T, Spezi E, Roberts S (2018) Development and validation of a prognostic model incorporating texture analysis derived from standardised segmentation of PET in patients with oesophageal cancer. *European Radiology* Volume 28 Issue 1. (DOI: 10.1007/s00330-017-4973-y).

- Parkinson C, Foley K, Whybra P, Hills R, Roberts A, Marshall C, Staffurth J, Spezi E (2018) Evaluation of Prognostic models developed using standardised image features from different PET automated segmentation methods. *EJNNMI research* Volume 8 Issue 29. (DOI: 10.1186/s13550-018-0379-3).

A.2 Presentations

- Parkinson C, Whybra P, Marshall C, Staffurth J, Spezi E (2018) AT-LAAS - Investigation into the Incorporation of Morphological Data on Automated Segmentation, *European Journal of Nuclear Medicine and Molecular Imaging*, Volume 45, Pages S72-S73.
- Parkinson C, Marshall C, Staffurth J, Spezi E (2018) Evaluation of the inclusion of morphological features in a machine-learned model: A phantom study, All Wales Medical Physics and Clinical Engineering summer meeting, Cardiff.
- Parkinson C, Evans M, Guerrero-Urbano T, Michaelidou A, Pike L, Barrington S, Jayaprakasam V, Rackley T, Palaniappan N, Staffurth J, Marshall C, Spezi E (2017) Tumour characteristics following one cycle

of induction 5-fluorouracil chemotherapy compared to pre-treatment 18F-FDG PET scans in oropharyngeal tumors: - a FIGARO report, All Wales Medical Physics and Clinical Engineering summer meeting, Swansea.

- Parkinson C, Marshall C, Staffurth J, Spezi E (2016) Hybrid Modality PET auto-segmentation for use in Texture Analysis, All Wales Medical Physics and Clinical Engineering summer meeting, Cardiff.

A.3 E-Posters and Posters

- Whybra P, Parkinson C, Foley K, Staffurth J, Spezi E (2019) A novel normalisation technique for voxel size dependent radiomic features in oesophageal cancer, Radiotherapy and Oncology Volume 133, Supplementary 1.
- Parkinson C, Evans M, Guerrero-Urbano T, Michaelidou A, Pike L, Barrington S, Jayaprakasam V, Rackley T, Palaniappan N, Staffurth J, Marshall C, Spezi E (2018) Target volume delineation of FDG PET images post one cycle of induction chemotherapy in oropharyngeal cancer using advanced automated segmentation methods, Radiotherapy and Oncology. (DOI: 10.1016/s0167-8140(18)31436-1)
- Parkinson C, Foley K, Whybra P, Hills R, Roberts A, Marshall C, Staffurth J, Spezi E (2018) Dependency of patient risk stratification on PET target volume definition in Oesophageal cancer, Radiotherapy

and Oncology. (DOI: 10.1016/s0167-8140(18)31241-6)

- Whybra P, Foley K, Parkinson C, Staffurth J, Spezi E (2018) Effect of interpolation on 3D Texture Analysis of PET Imaging in Oesophageal Cancer, Radiotherapy and Oncology. (DOI: 10.1016/s0167-8140(18)32426-5)
- Parkinson C, Marshall C, Staffurth J, Spezi E (2018) Evaluation of the inclusion of morphological features in a machine-learned model: A phantom study, All Wales Medical Physics and Clinical Engineering summer meeting, Cardiff.
- Parkinson C, Chan J, Syndikus I, Marshall C, Staffurth J, Spezi E. (2017) Impact of 18F-Choline PET scan acquisition time on delineation of GTV in Prostate cancer. Radiotherapy and Oncology (DOI: 10.1016/s0167-8140(17)31768-1).
- Parkinson C, Chan J, Syndikus I, Marshall C, Staffurth J, Spezi E. (2017) Impact of 18F-Choline PET scan acquisition time on delineation of GTV in Prostate cancer, UK radiological and radiation oncology congress, Manchester.

Appendix B

PET-AS MTV delineation in low tumour to background ratio PET

B.1 Correlation of PET-AS methods DSC and Clinician derived radiomic features

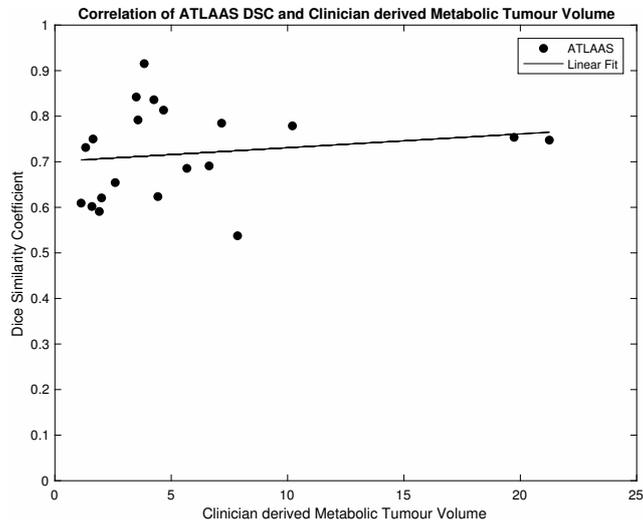


Figure B.1: Correlation of ATLAAS DSC and clinician derived MTV

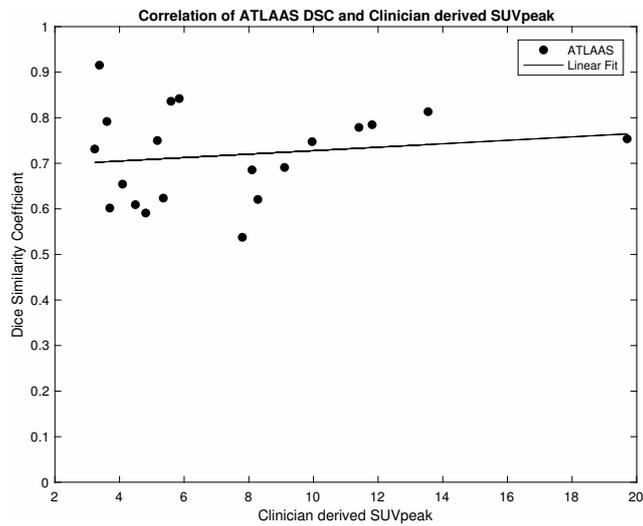


Figure B.2: Correlation of ATLAAS DSC and Clinician derived SUV_{PEAK}

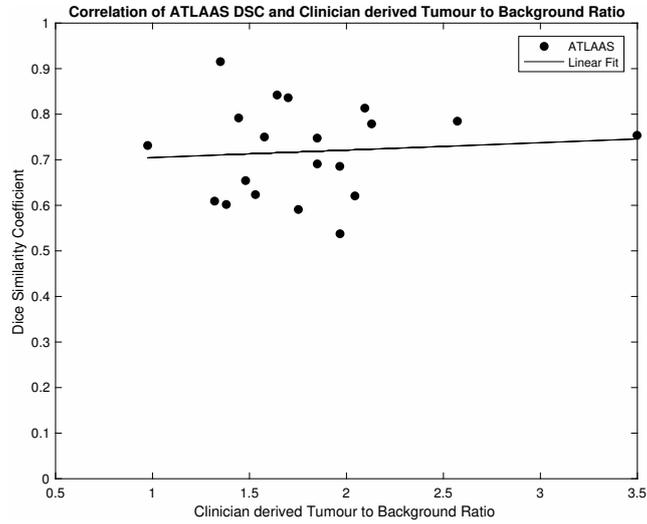


Figure B.3: Correlation of ATLAAS DSC and Clinician derived TBR

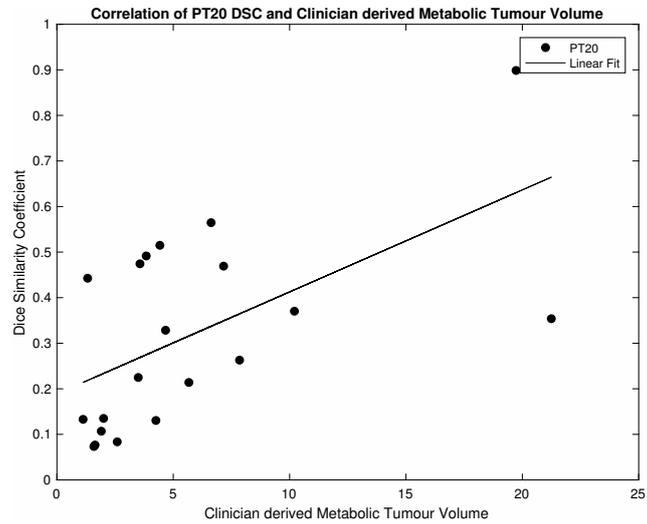


Figure B.4: Correlation of PT20 DSC and Clinician derived MTV

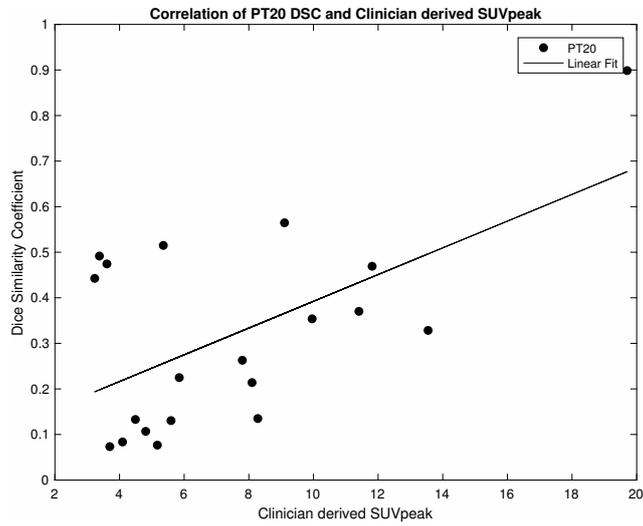


Figure B.5: Correlation of PT20 DSC and Clinician derived SUV_{PEAK}

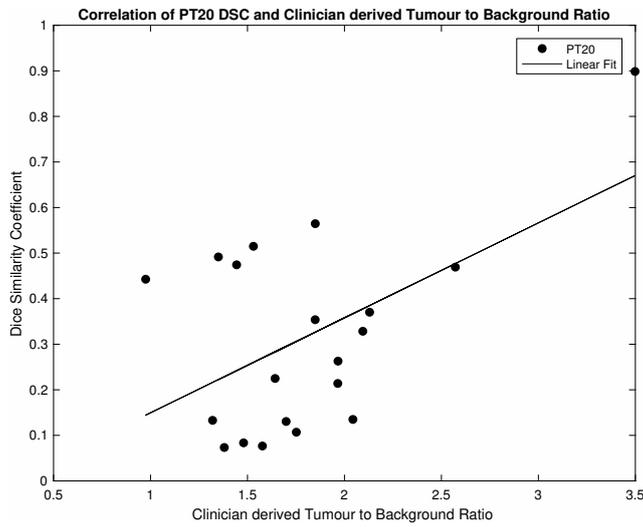


Figure B.6: Correlation of PT20 DSC and Clinician derived TBR

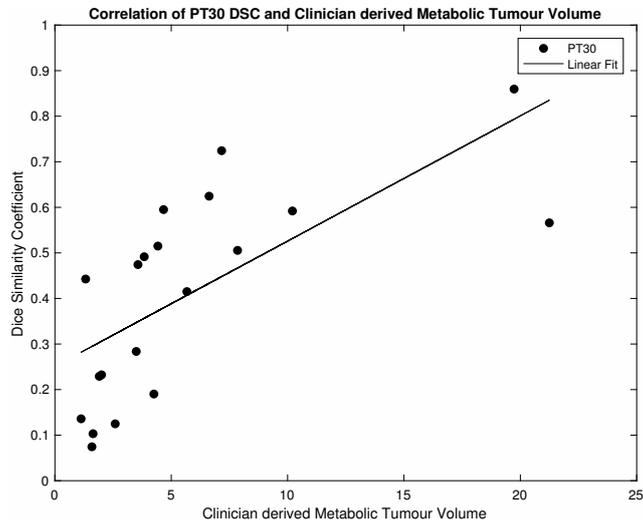


Figure B.7: Correlation of PT30 DSC and Clinician derived MTV

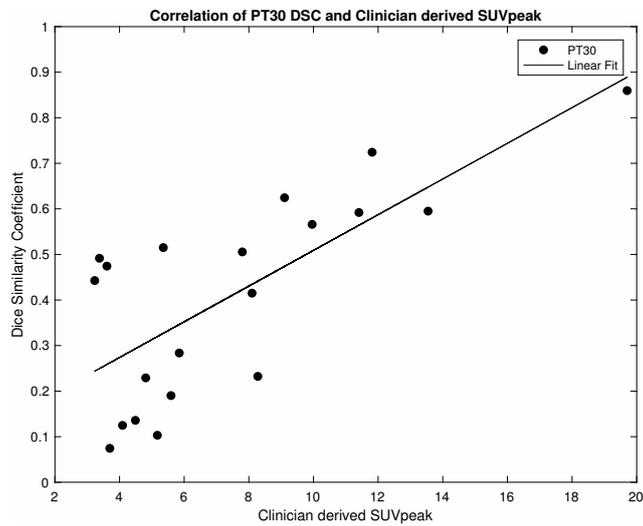


Figure B.8: Correlation of PT30 DSC and Clinician derived SUV_{PEAK}

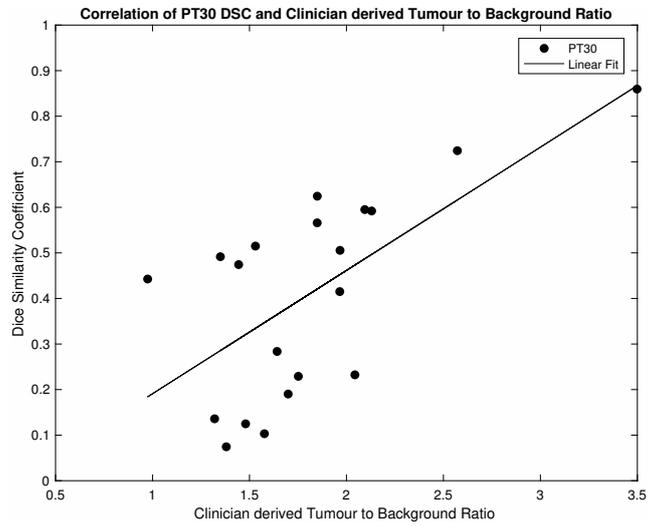


Figure B.9: Correlation of PT30 DSC and Clinician derived TBR

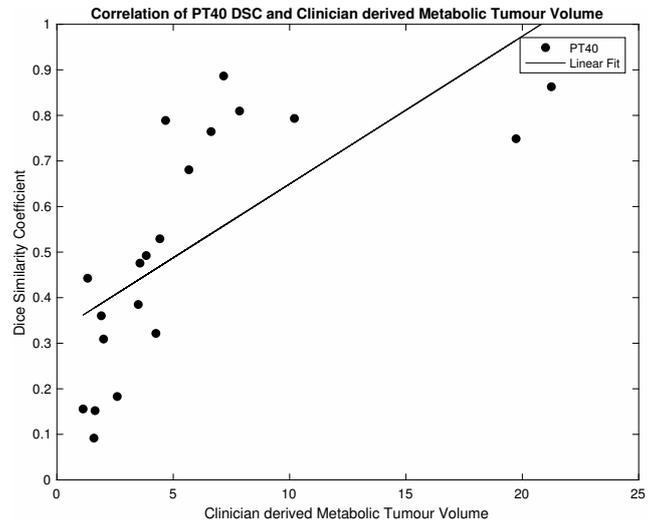


Figure B.10: Correlation of PT40 DSC and Clinician derived MTV

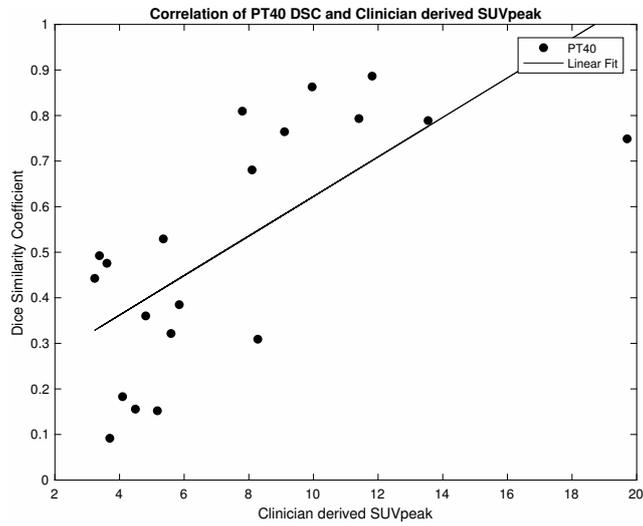


Figure B.11: Correlation of PT40 DSC and Clinician derived SUV_{PEAK}

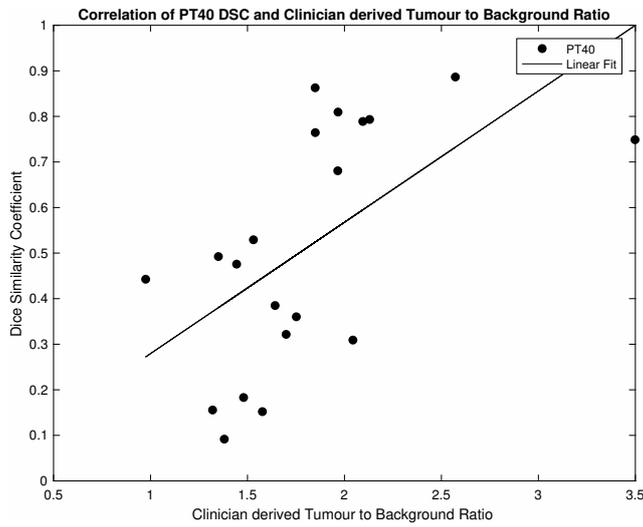


Figure B.12: Correlation of PT40 DSC and Clinician derived TBR

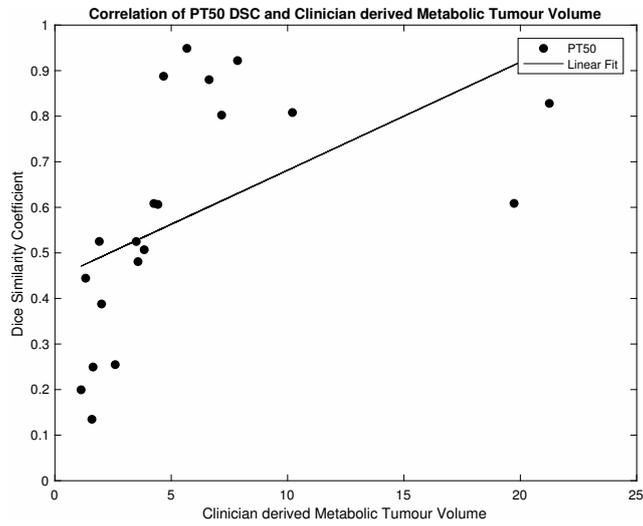


Figure B.13: Correlation of PT50 DSC and Clinician derived MTV

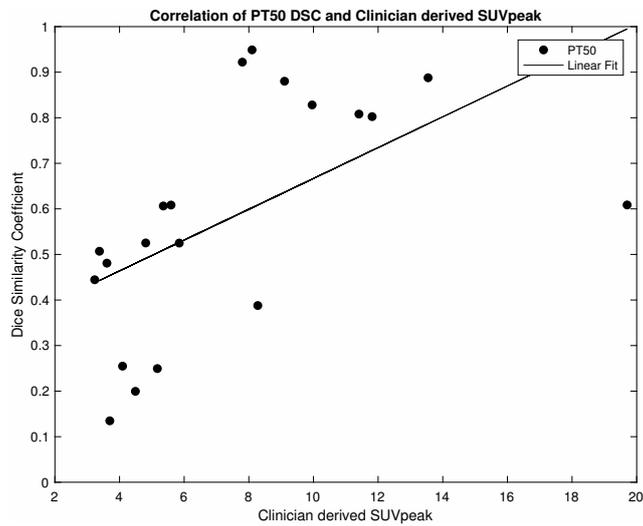


Figure B.14: Correlation of PT50 DSC and Clinician derived SUV_{PEAK}

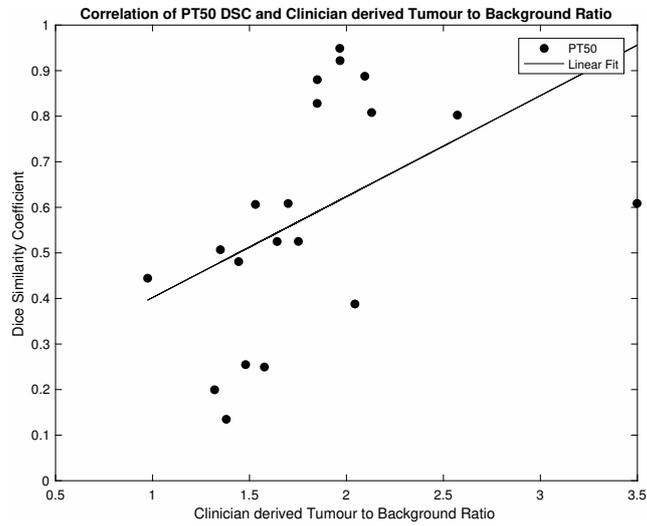


Figure B.15: Correlation of PT50 DSC and Clinician derived TBR

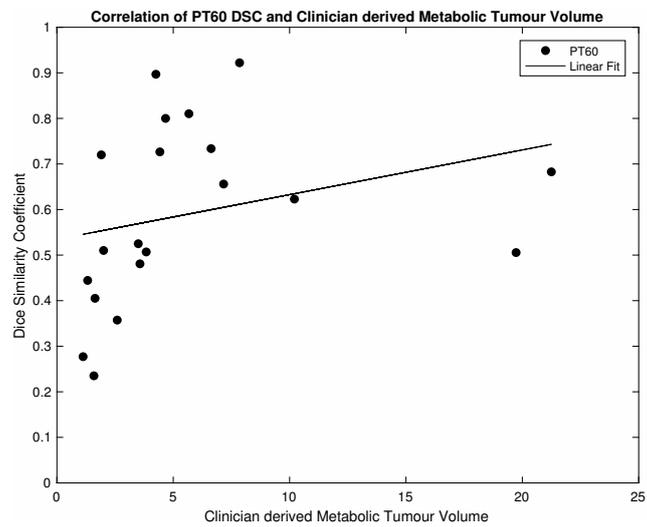


Figure B.16: Correlation of PT60 DSC and Clinician derived MTV

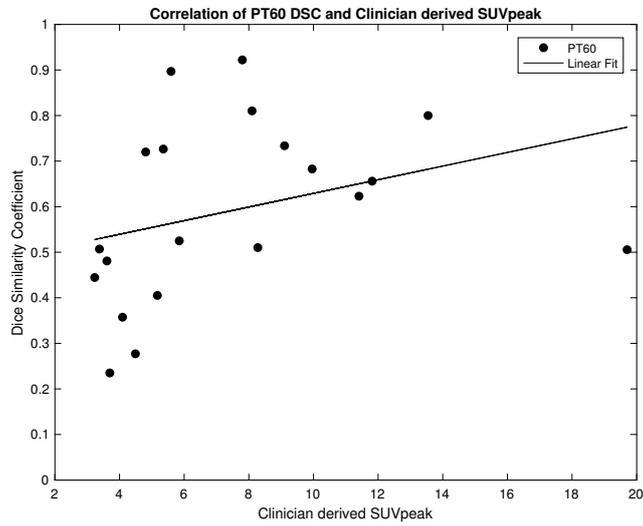


Figure B.17: Correlation of PT60 DSC and Clinician derived SUV_{PEAK}

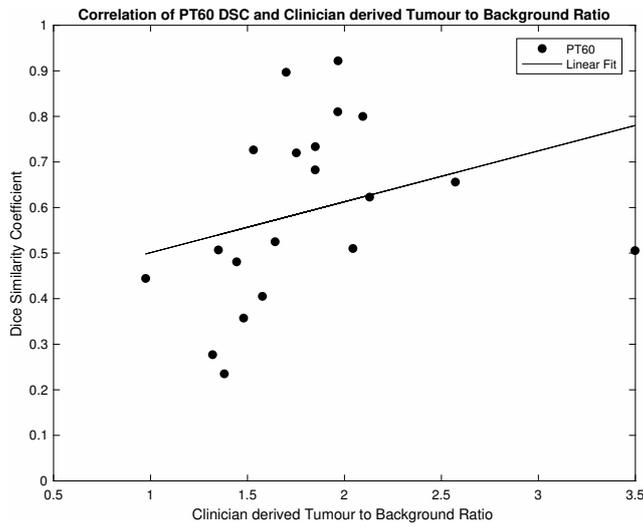


Figure B.18: Correlation of PT60 DSC and Clinician derived TBR

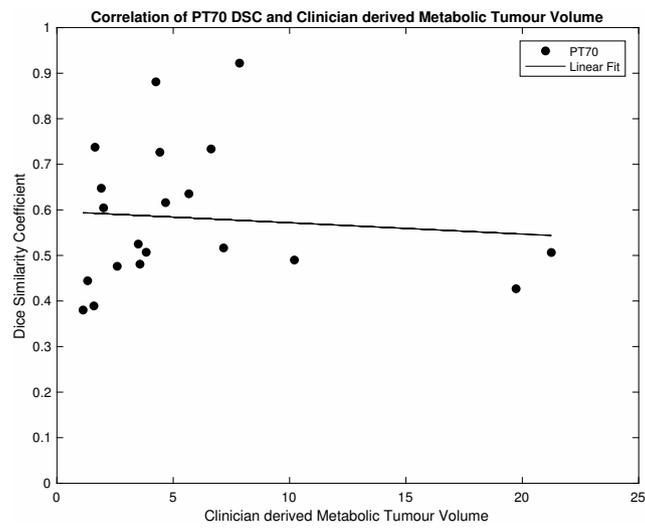


Figure B.19: Correlation of PT70 DSC and Clinician derived MTV

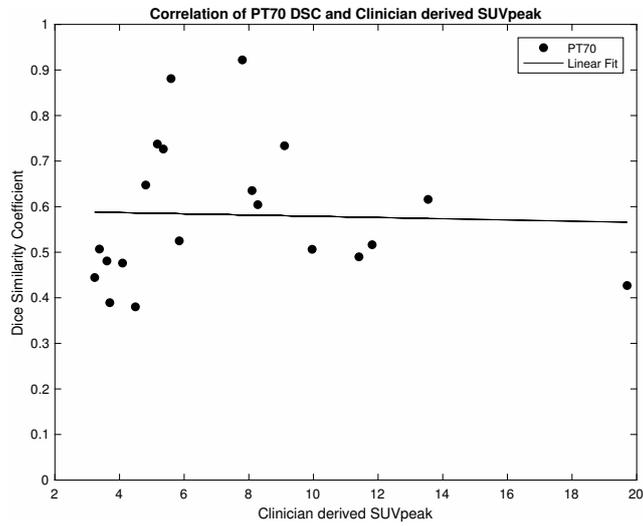


Figure B.20: Correlation of PT70 DSC and Clinician derived SUV_{PEAK}

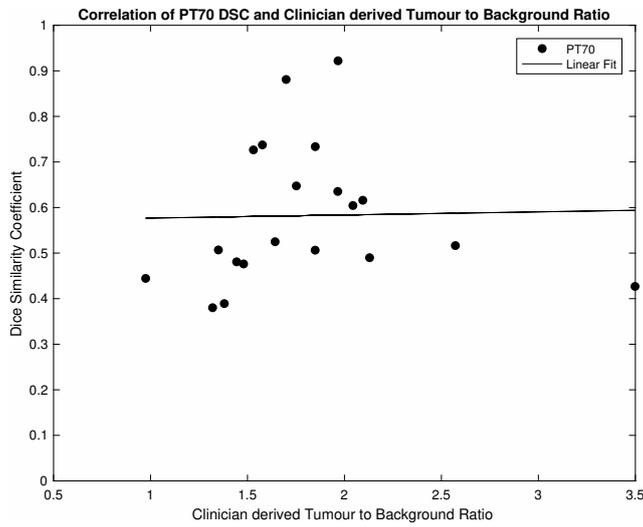


Figure B.21: Correlation of PT70 DSC and Clinician derived TBR

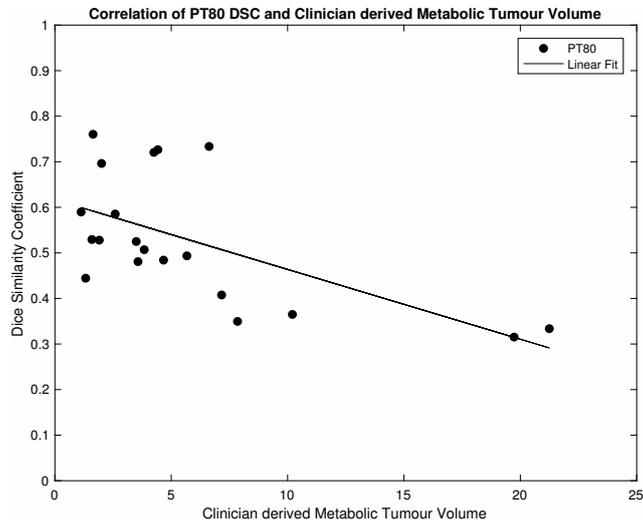


Figure B.22: Correlation of PT80 DSC and Clinician derived MTV

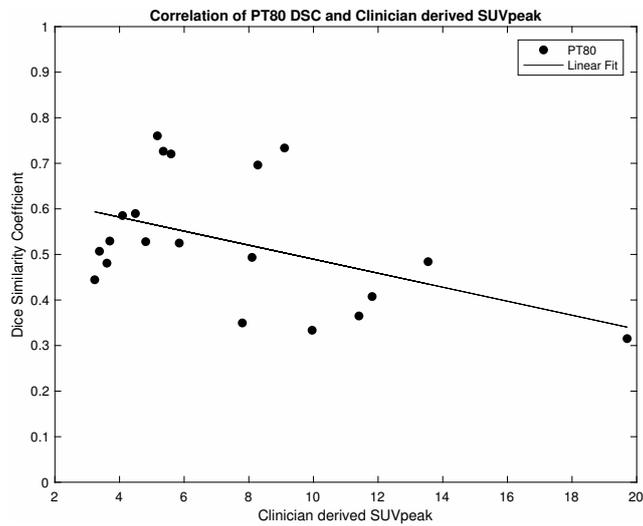


Figure B.23: Correlation of PT80 DSC and Clinician derived SUV_{PEAK}

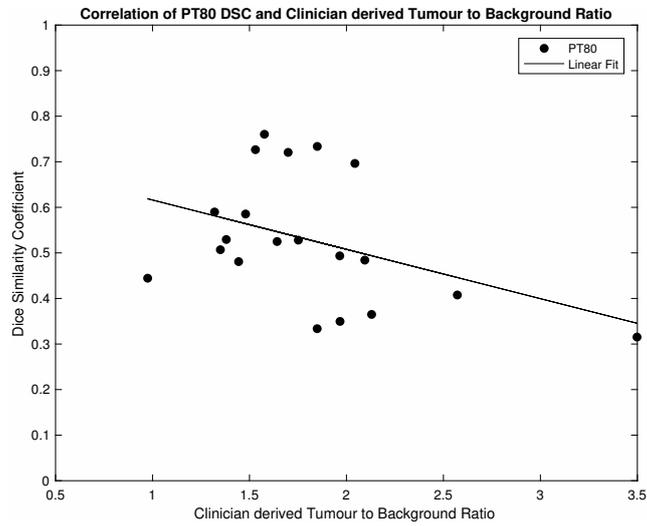


Figure B.24: Correlation of PT80 DSC and Clinician derived TBR

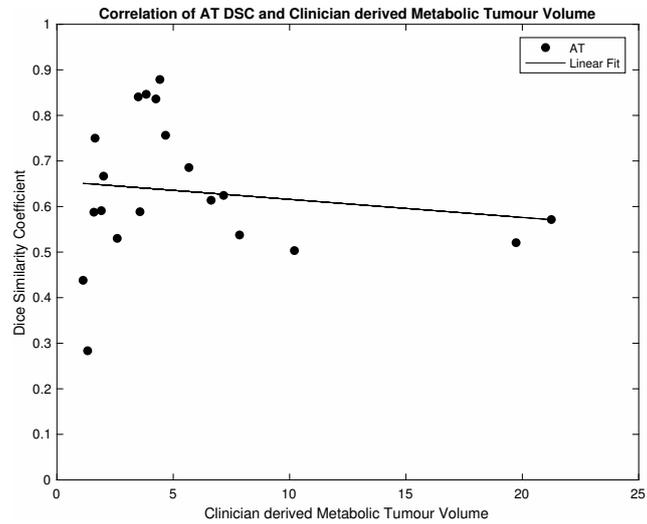


Figure B.25: Correlation of AT DSC and Clinician derived MTV

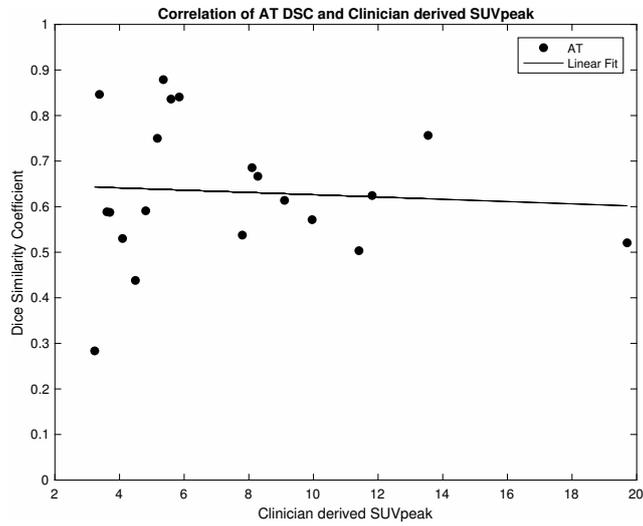


Figure B.26: Correlation of AT DSC and Clinician derived SUV_{PEAK}

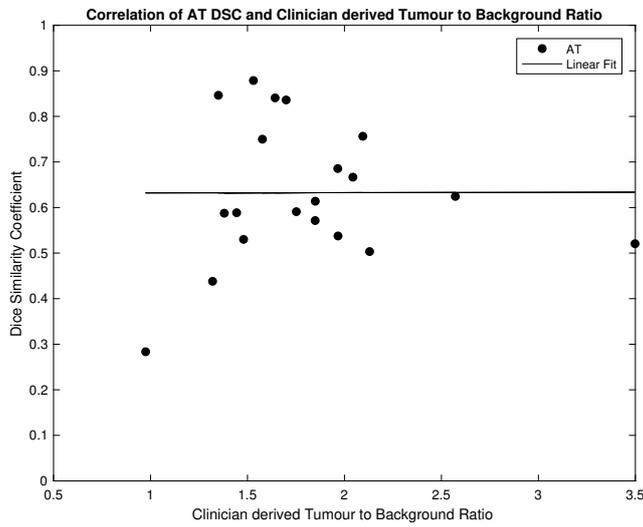


Figure B.27: Correlation of AT DSC and Clinician derived TBR

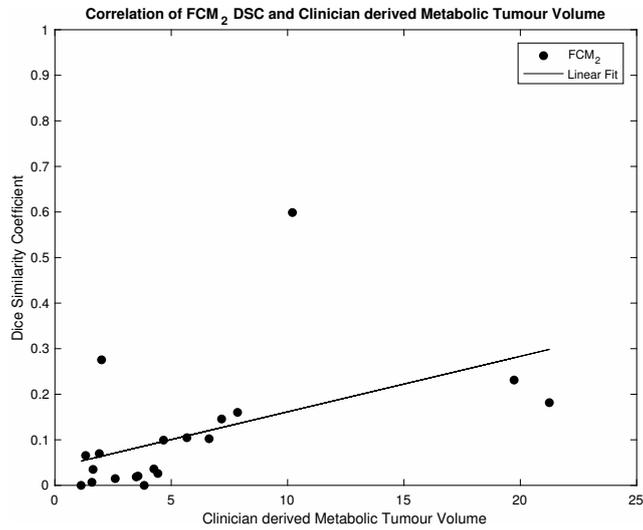


Figure B.28: Correlation of FCM2 DSC and Clinician derived MTV

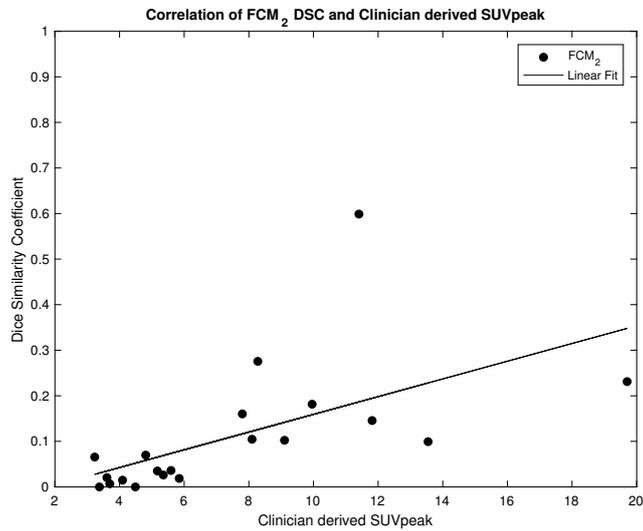


Figure B.29: Correlation of FCM2 DSC and Clinician derived SUV_{PEAK}

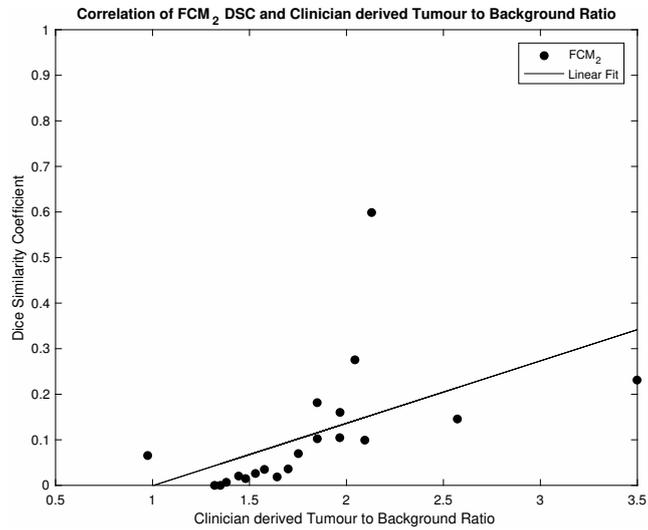


Figure B.30: Correlation of FCM2 DSC and Clinician derived TBR

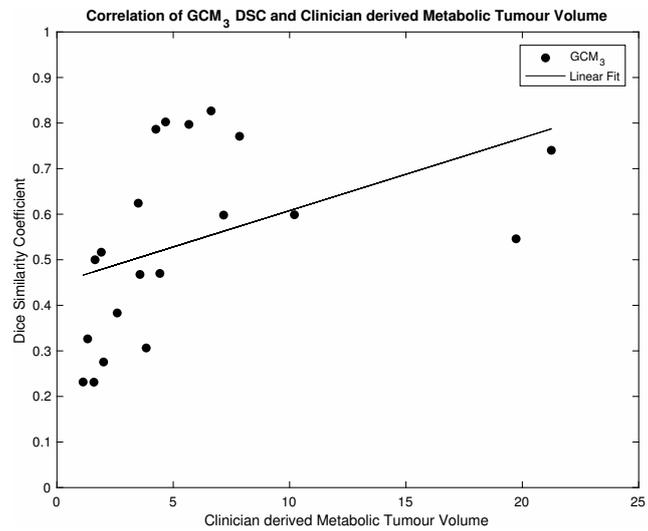


Figure B.31: Correlation of GCM3 DSC and Clinician derived MTV

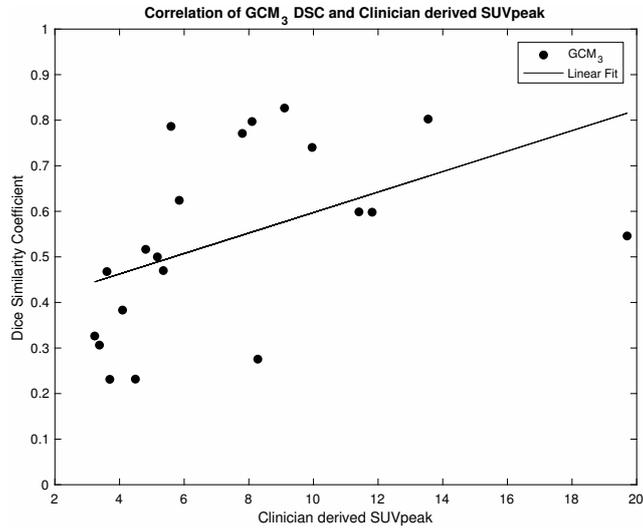


Figure B.32: Correlation of GCM3 DSC and Clinician derived SUV_{PEAK}

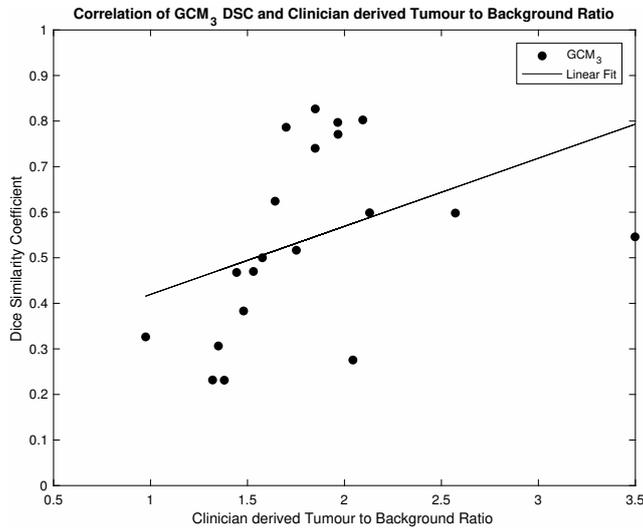


Figure B.33: Correlation of GCM3 DSC and Clinician derived TBR

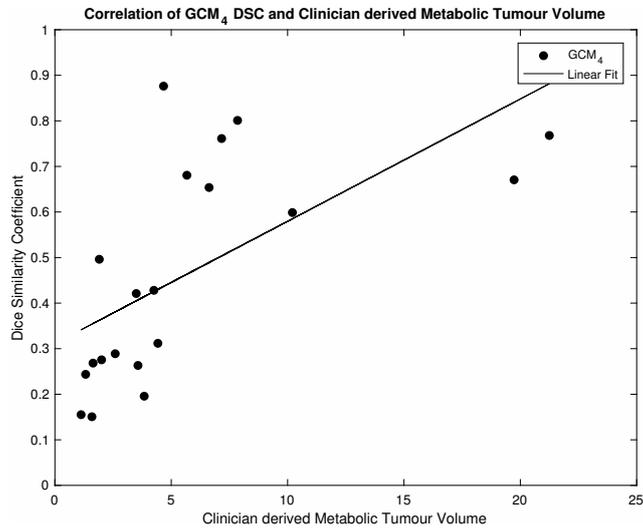


Figure B.34: Correlation of GCM4 DSC and Clinician derived MTV

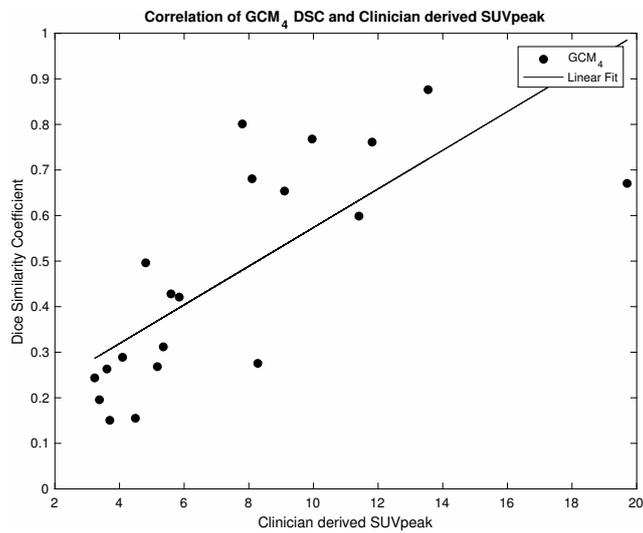


Figure B.35: Correlation of GCM4 DSC and Clinician derived SUV_{PEAK}

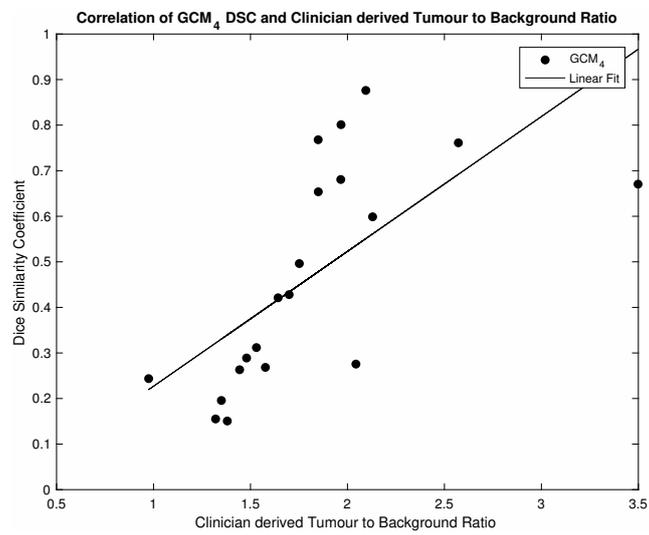


Figure B.36: Correlation of GCM4 DSC and Clinician derived TBR

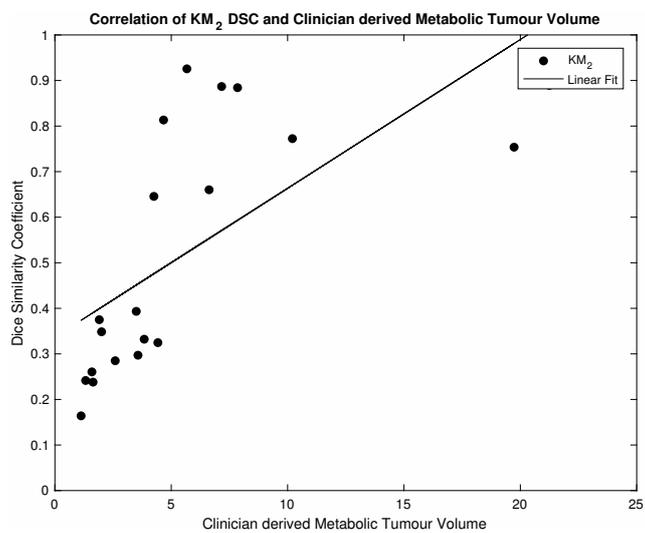


Figure B.37: Correlation of KM2 DSC and Clinician derived MTV

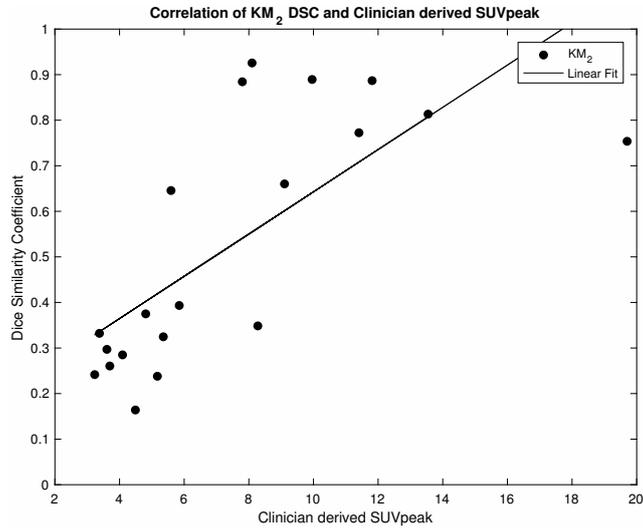


Figure B.38: Correlation of KM₂ DSC and Clinician derived SUV_{PEAK}

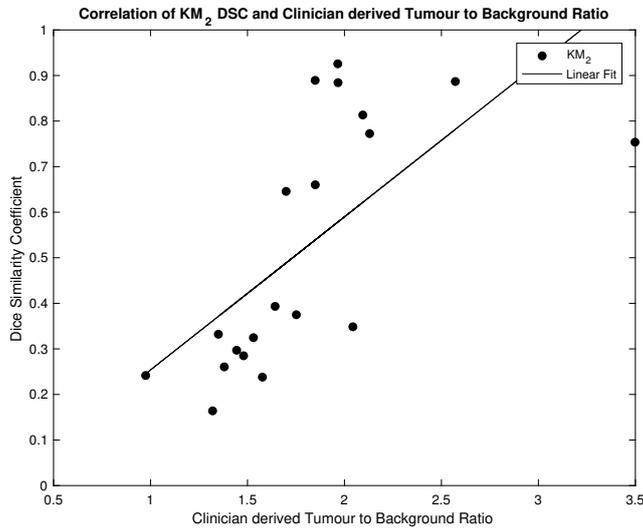


Figure B.39: Correlation of KM₂ DSC and Clinician derived TBR

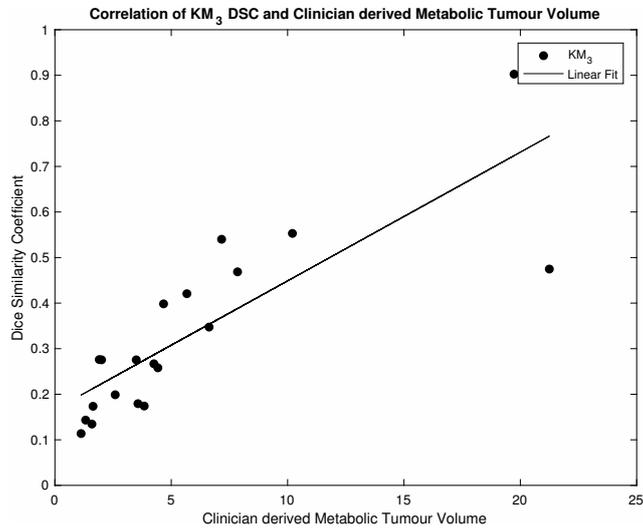


Figure B.40: Correlation of KM3 DSC and Clinician derived MTV

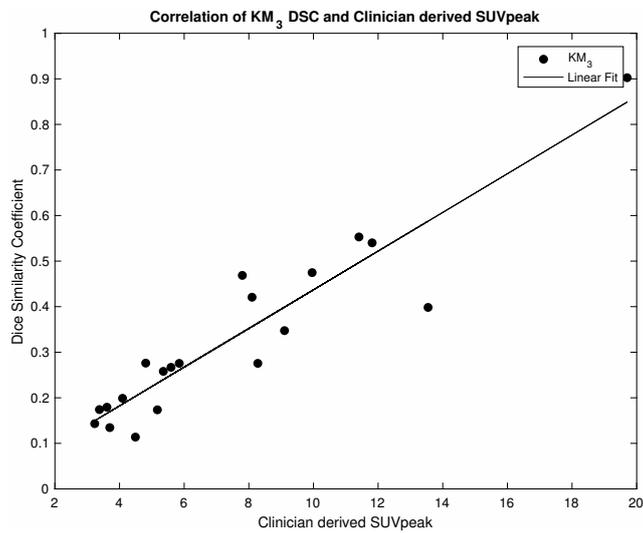


Figure B.41: Correlation of KM3 DSC and Clinician derived SUV_{PEAK}

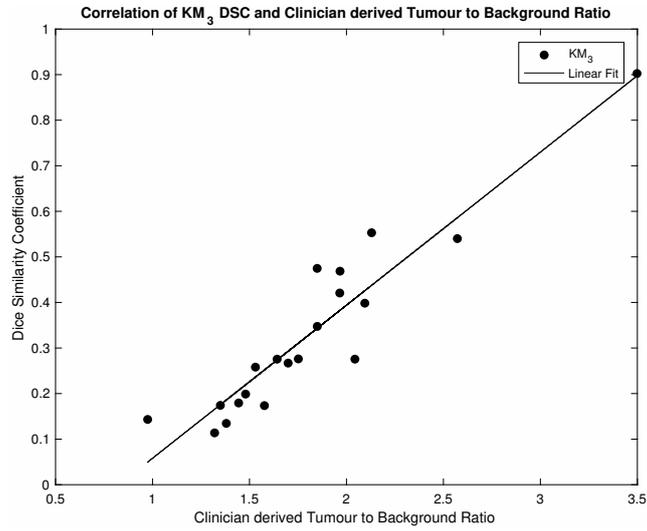


Figure B.42: Correlation of KM3 DSC and Clinician derived TBR

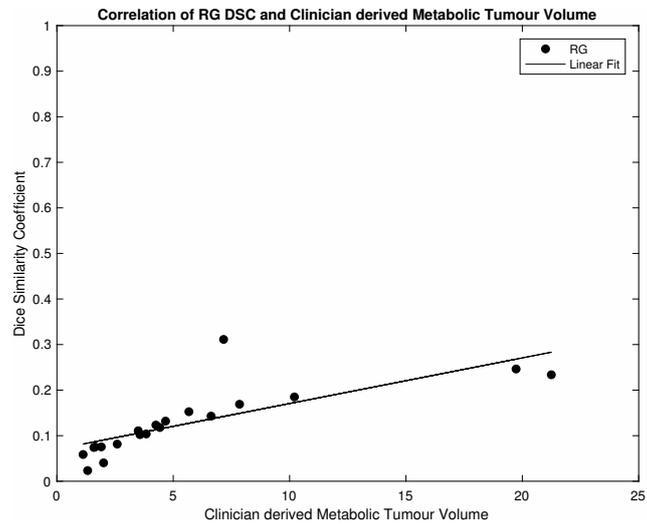


Figure B.43: Correlation of RG DSC and Clinician derived MTV

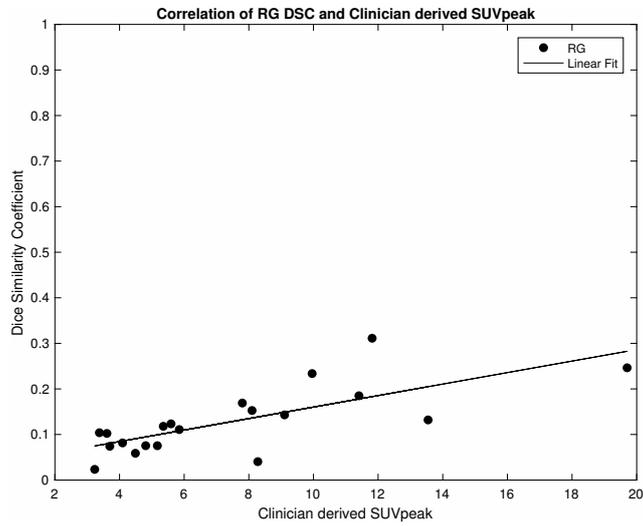


Figure B.44: Correlation of RG DSC and Clinician derived SUV_{PEAK}

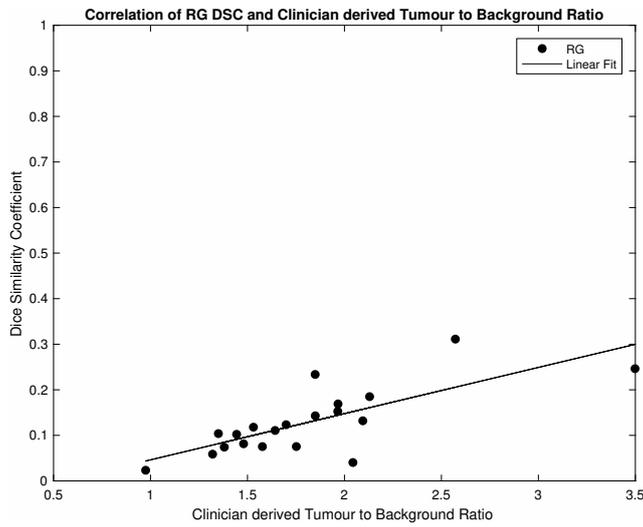


Figure B.45: Correlation of RG DSC and Clinician derived TBR

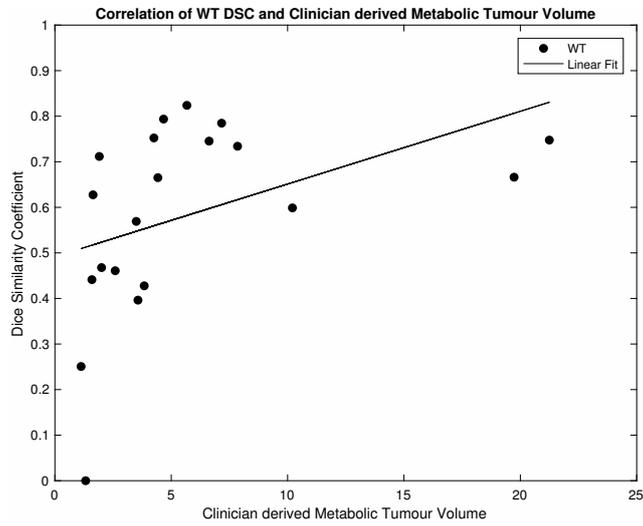


Figure B.46: Correlation of WT DSC and Clinician derived MTV

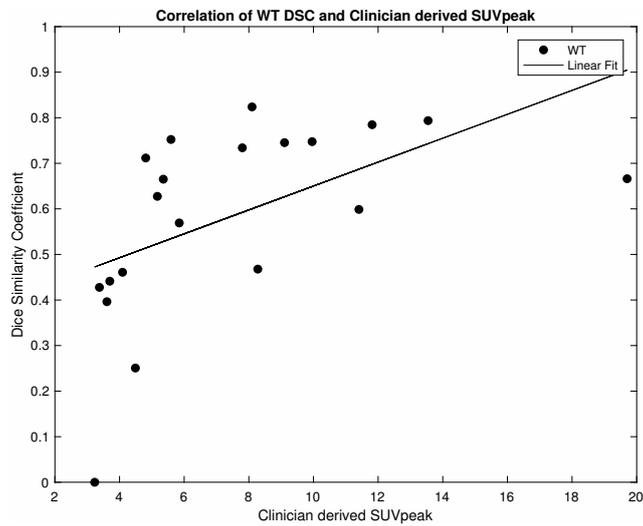


Figure B.47: Correlation of WT DSC and Clinician derived SUV_{PEAK}

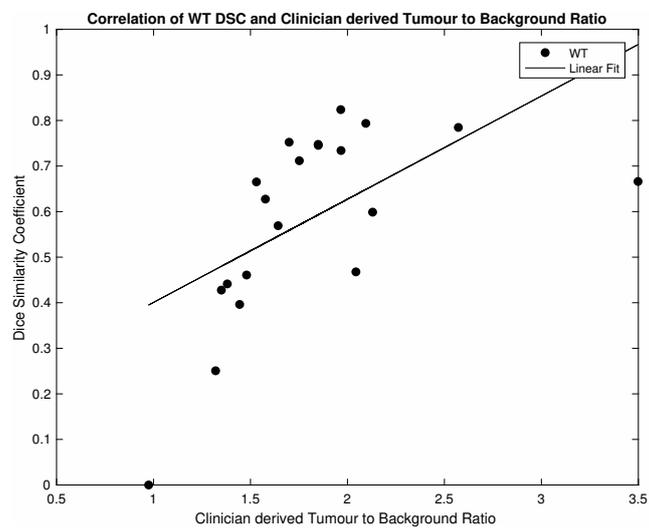


Figure B.48: Correlation of WT DSC and Clinician derived TBR.

Appendix C

Investigation into the inclusion of morphological data in machine learned training models data

C.1 Training models DSC

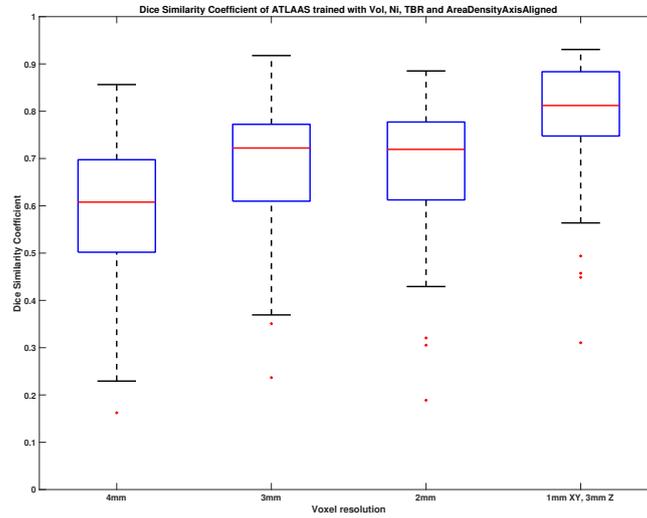


Figure C.1: DSC of ATLAAS using the parameters NI, TBR and MTV and Area Density (Axis Aligned) on 4, 3, 2 and 1 mm PET validation scans.

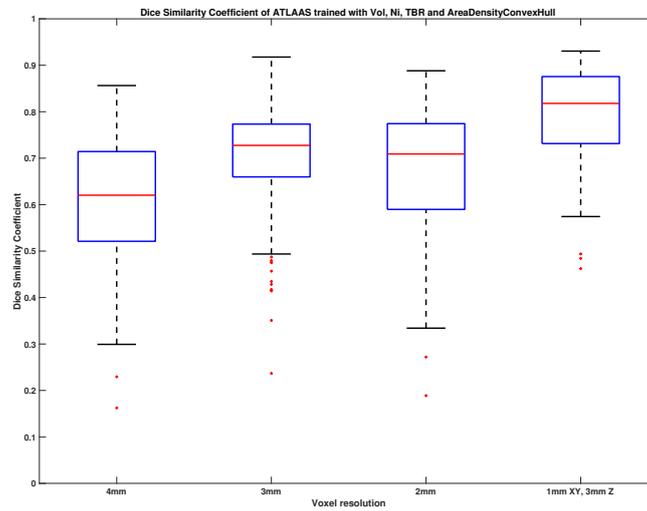


Figure C.2: DSC of ATLAAS using the parameters NI, TBR and MTV and Area Density (Convex Hull) on 4, 3, 2 and 1 mm PET validation scans.

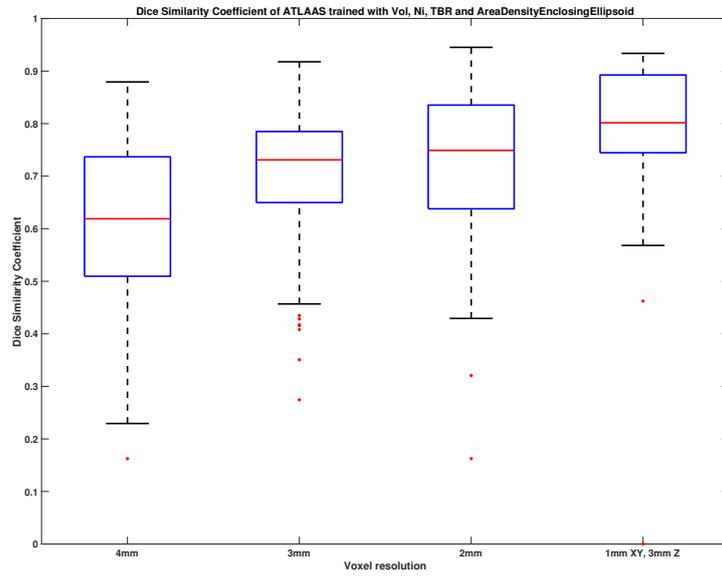


Figure C.3: DSC of ATLAAS using the parameters NI, TBR and MTV and Area Density (Enclosing Ellipsoid) on 4, 3, 2 and 1 mm PET validation scans.

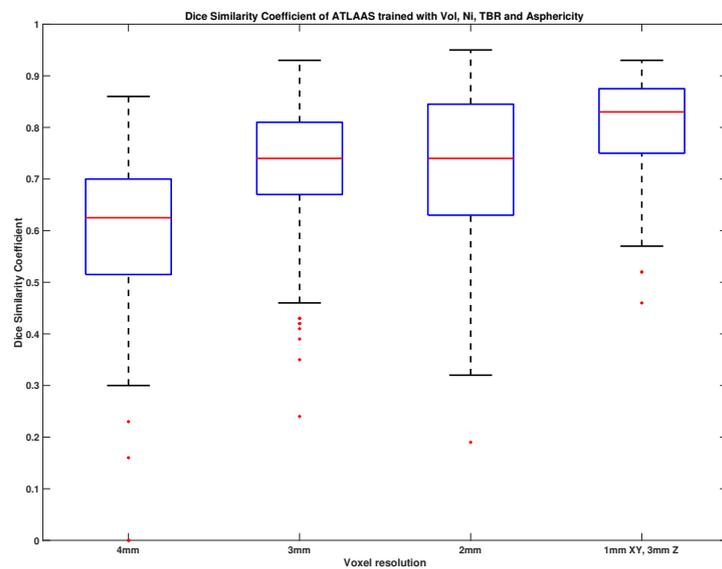


Figure C.4: DSC of ATLAAS using the parameters NI, TBR and MTV and Asphericity on 4, 3, 2 and 1 mm PET validation scans.

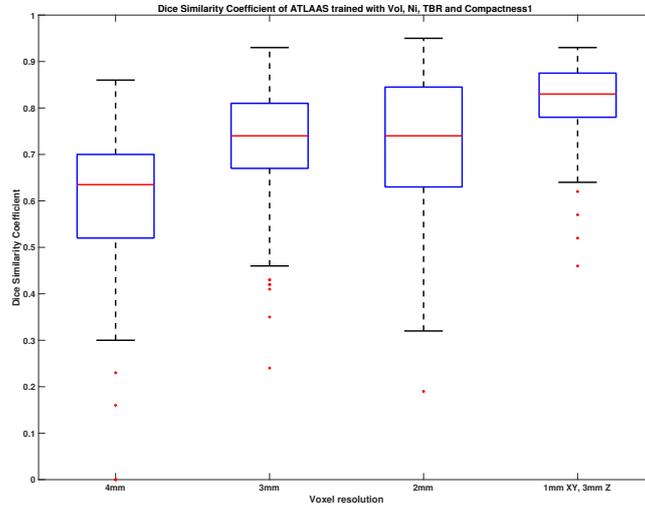


Figure C.5: Dice similarity coefficient of ATLAAS using the parameters NI, TBR and MTV and Compactness one on 4, 3, 2 and 1 mm PET validation scans.

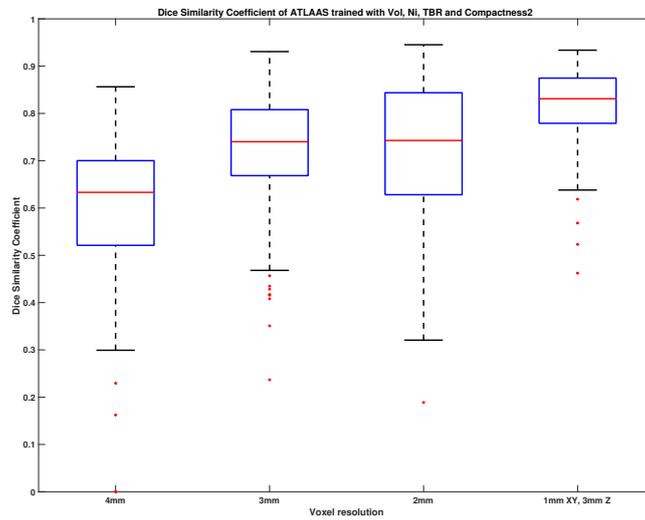


Figure C.6: DSC of ATLAAS using the parameters NI, TBR and MTV and Compactness two on 4, 3, 2 and 1 mm PET validation scans.

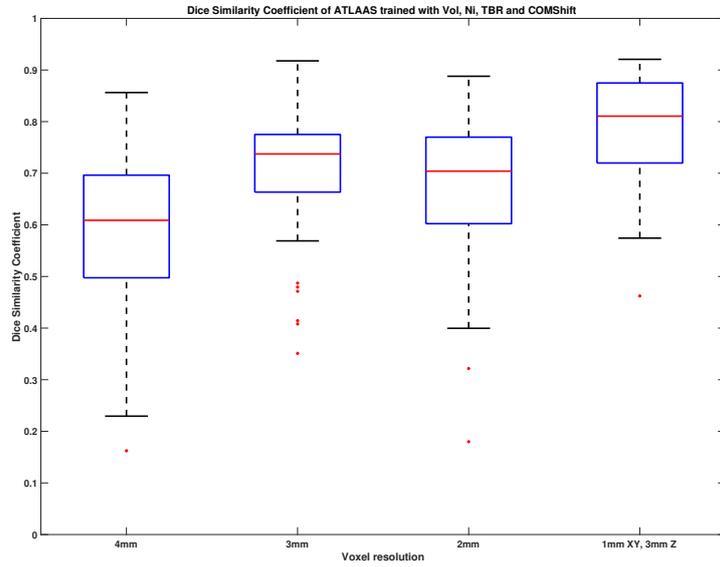


Figure C.7: DSC of ATLAAS using the parameters NI, TBR and MTV and COMShift on 4, 3, 2 and 1 mm PET validation scans.

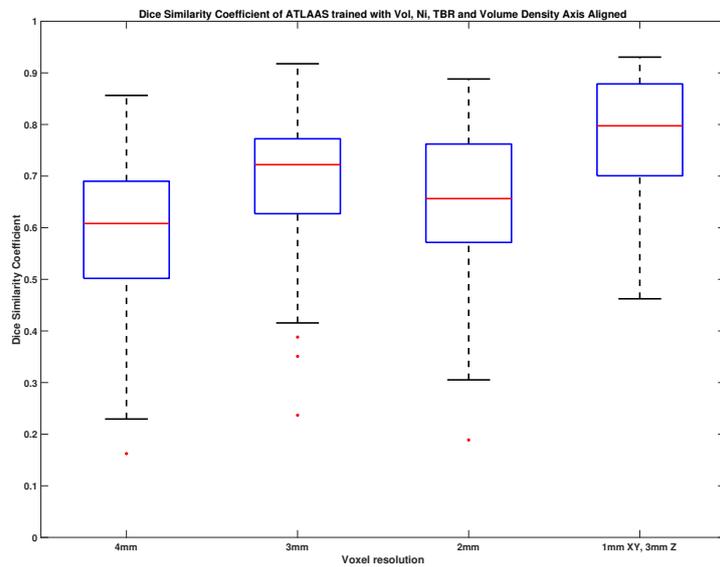


Figure C.8: DSC of ATLAAS using the parameters NI, TBR and MTV and Volume Density (Axis Aligned) on 4, 3, 2 and 1 mm PET validation scans.

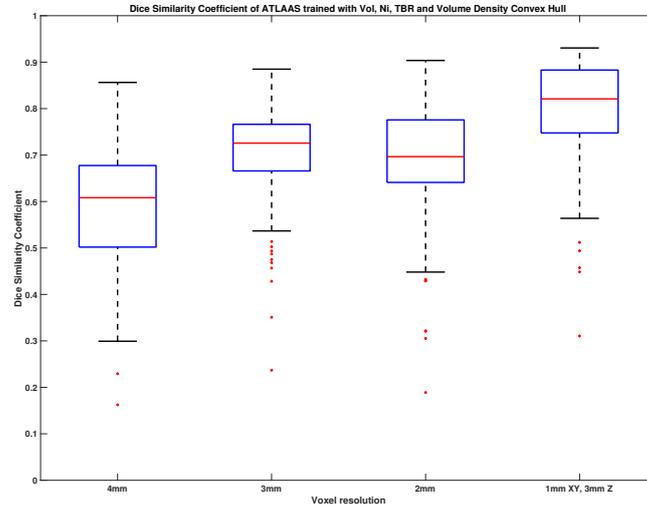


Figure C.9: DSC of ATLAAS using the parameters NI, TBR and MTV and Volume Density (Convex Hull) on 4, 3, 2 and 1 mm PET validation scans.

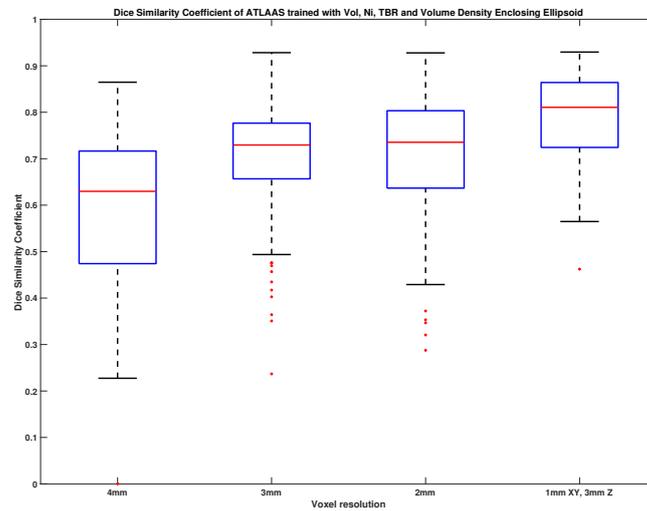


Figure C.10: DSC of ATLAAS using the parameters NI, TBR and MTV and Volume Density (Enclosing Ellipsoid) on 4, 3, 2 and 1 mm PET validation scans.

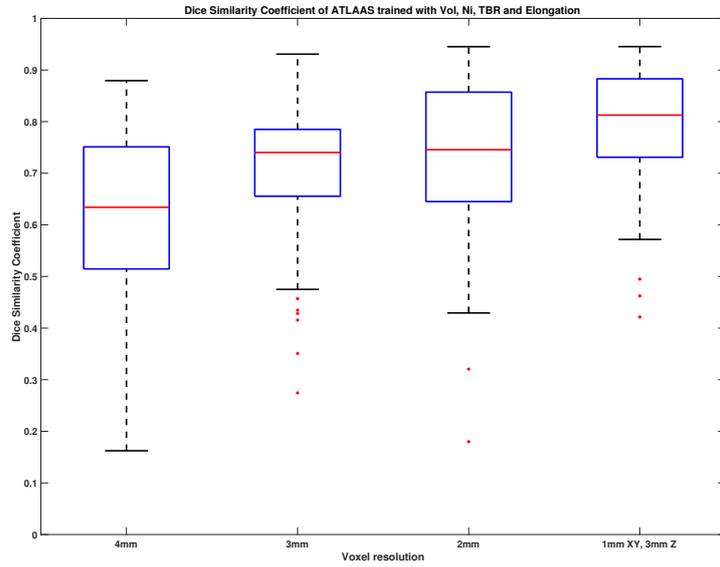


Figure C.11: DSC of ATLAAS using the parameters NI, TBR and MTV and Elongation on 4, 3, 2 and 1 mm PET validation scans.

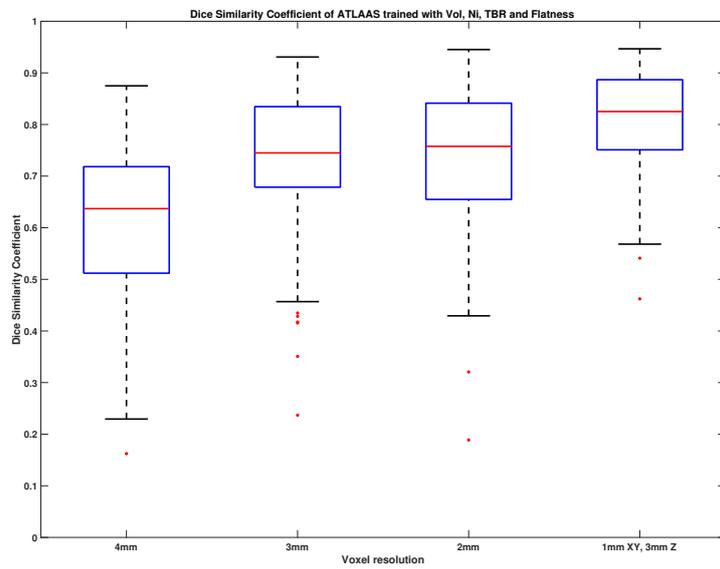


Figure C.12: DSC of ATLAAS using the parameters NI, TBR and MTV and Flatness on 4, 3, 2 and 1 mm PET validation scans.

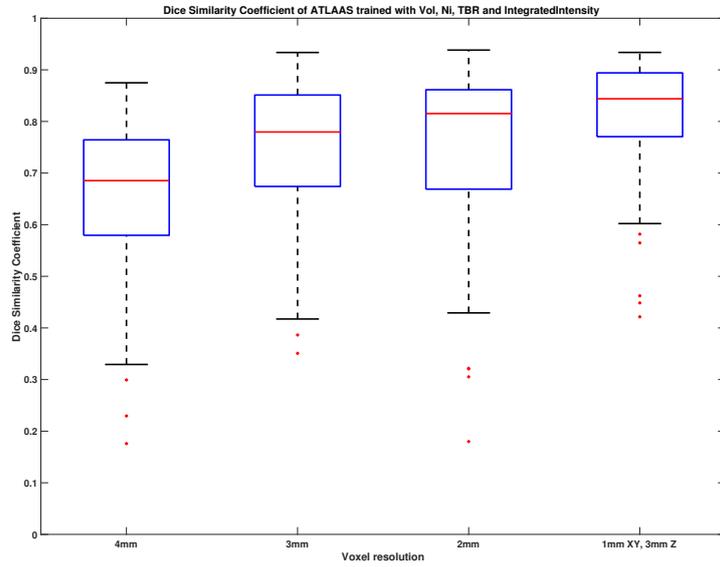


Figure C.13: DSC of ATLAAS using the parameters NI, TBR and MTV and Integrated Intensity on 4, 3, 2 and 1 mm PET validation scans.

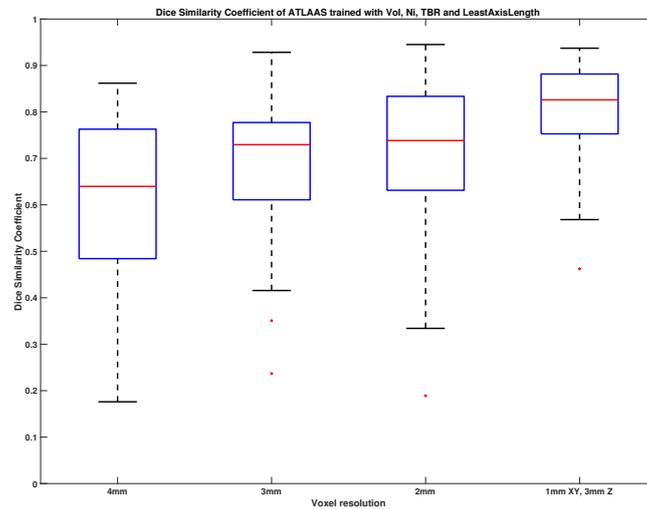


Figure C.14: DSC of ATLAAS using the parameters NI, TBR and MTV and Least Axis Length on 4, 3, 2 and 1 mm PET validation scans.

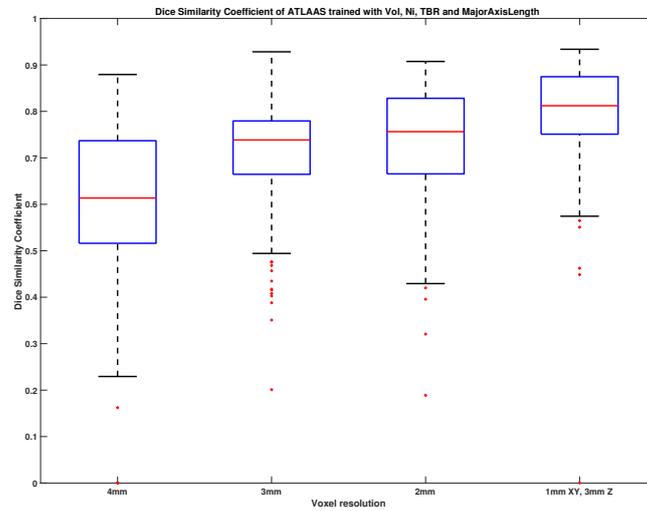


Figure C.15: DSC of ATLAAS using the parameters NI, TBR and MTV and Major Axis Length on 4, 3, 2 and 1 mm PET validation scans.

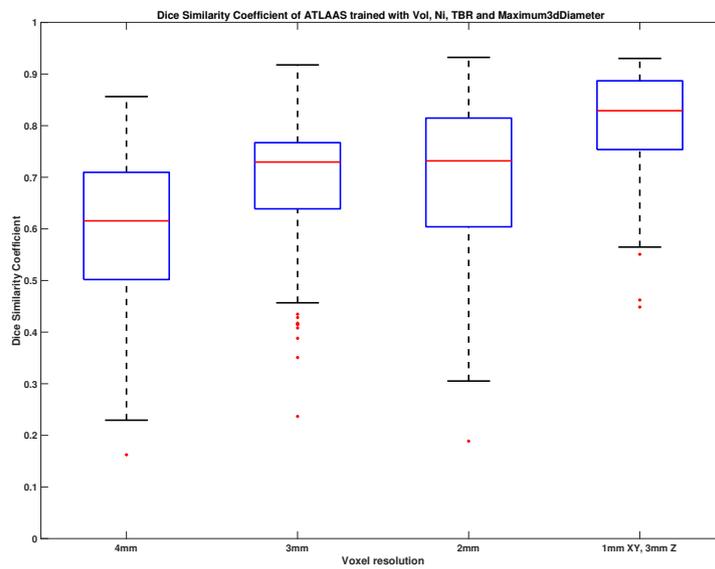


Figure C.16: DSC of ATLAAS using the parameters NI, TBR and MTV and Maximum 3D Diameter on 4, 3, 2 and 1 mm PET validation scans.

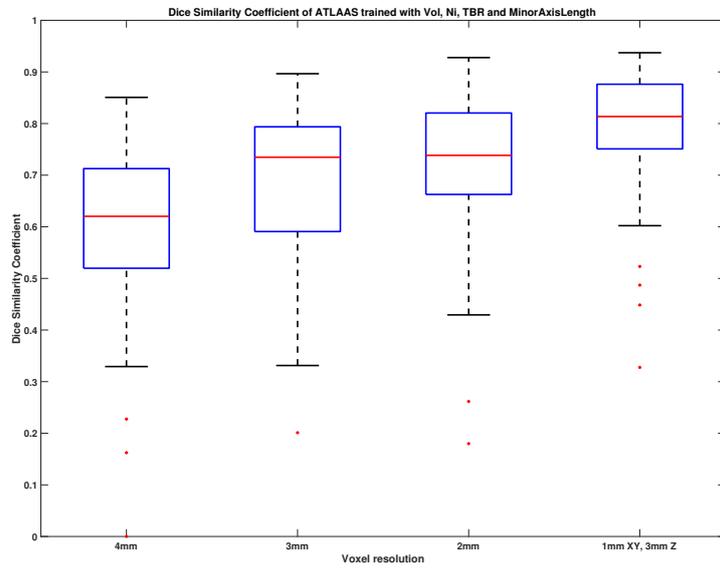


Figure C.17: DSC of ATLAAS using the parameters NI, TBR and MTV and Minor Axis Length on 4, 3, 2 and 1 mm PET validation scans.