POSITRON EMISSION TOMOGRAPHY IMAGE TEXTURE ANALYSIS

Emad Alsyed

School of Engineering, Cardiff University, Cardiff, Wales, UK.

February 2023

Acknowledgements

First and foremost, I would like to praise and thank God, the almighty, who has granted me the knowledge and countless blessing.

I acknowledge the generous financial support from the King Abdulaziz University and Saudi Culture Bureau in London (grant number # KAU1938). I am grateful to Prof Mohammed Aljohani, Head, Department of Nuclear Engineering, King Abdulaziz University, for his valuable suggestions, ever encouraging and motivating guidance.

I would like to thank my primary supervisor Dr Emiliano Spezi. Emiliano has been an ideal mentor and thesis supervisor, offering advice and encouragement with a perfect blend of insight and hardiness. Your feedbacks, assistance and guidance have been definitely useful in my ability to reach the finish line. I will forever be grateful for your support through both academic and personal challenges. I feel I learned a lot of things from you in the last few years, thank you.

Thank you to my secondary supervisor, Prof Christopher Marshall. Your advice and clinical viewpoint were enlightening, especially when I was getting everything rolling. I really respect all of the work you and your group do at PETIC.

I am indebted to my third supervisor Dr Rhodri Smith. Your encouraging words and thoughtful, detailed feedback have been very important to me. I am so proud to work with you and participate on several conferences. Each time we convene I gain

some new insight and knowledge.

Thank you to all of the past and present CIDA team – Craig, Concetta, Emily, Salvo, Ahmed, Nathan, Eszter for all of the kind words and assistance you have provided. Philip and Iona, I really appreciated your valuable comments and suggestions. Thank you so much for the meeting and joking! If I had a dollar for each, I'd be rich! (Seriously, let's put that plan in motion for next year).

To my parents, there are no proper words to convey my deep gratitude for you. I deeply thank you, Mohammed and Sultanah for your prayers, unconditional trust, timely encouragement and endless patience. It was your love that raised me up again and again when I got weary. You have also been generous with your love and encouragement despite the long distance between us. I still remember those moments when you said goodbye with tears in your eyes. That was in 2011 and since that time you were always asking me, when will you finish your study. I can tell you now the dream becomes true and I am finishing the journey and coming back to you with a PhD degree.

My brothers and extended family have all been very wonderful. No words are adequate to express my indebtedness to my uncle Abdulhamid and my mother in-law Aidah for their support, blessings and good wishes. A special thanks to my uncle Abdulaali for his constant contact. Thank you to my brothers Fahad and Yusuf for all of the kind words and assistance you have provided. Special thanks must also go to my brothers Redah, Omar and Ahmed for their unfailing emotional

support, heart-warming kindness, and their help to my family when I was away. To my brothers Abdulrahman and Abdulaziz, our late nights (when I was on annual leave) spent playing PlayStation (mostly, I win. you need to do more training) are especially cherished memories. I am deeply grateful for the years of support and humour my sister Bashayer offered me. Also, thanks must go to my nephew Meshal for his contact every week when I was in USA or UK. I do not forget when you take me from the airport every time I came to KSA.

A special thanks must go to my wife Tahani for her love and constant support, for all the late nights and early mornings. Thank you for being my muse. Sincere thanks to you Tahani for standing by me through all my hardships, and my absences. You gave me support, help, and prevented several wrong turns. You also supported the family during much of my graduate studies in USA and UK. Thank you for being my best friend. I owe you everything. Along with her, I would like to acknowledge my two daughters, Sulaf and Rose. You have never known your dad as anything but a student, it seems. You are great sources of love and relief from the scholarly endeavour.

During the study period, I faced many challenges. covid pandemic was a harrowing experience. I was forced to live alone for a long time and travel from one country to another to reach my family. Also, working remotely was tough. In addition to the covid challenge, I can't forget the horrible days when my baby (Abdullah) spent several weeks in the cardiovascular intensive care unit after performing open-heart

surgery. That time was really terrible and hard for me. So, thank GOD again for countless blessings and for making these hard times a memory.

Abstract

According to the World Health Organisation (WHO), cancer is the second leading cause of death around the world. Cancer is responsible for about 18 percent of deaths worldwide and about 10 million people die from it each year. Multidisciplinary teams participate to manage and deliver effective diagnosis and treatments for cancer patients. Currently, PET images and other medical images are interpreted visually by radiologists and clinicians. However, medical images contain more information than what can be assessed visually. The rapid development of medical image analysis has revolutionised the ability to recognise complex patterns in imaging data and provide a depth of quantitative analysis previously unachievable. Radiomics is defined as extracting quantitative features from medical images which cannot be seen by the naked eye. It is now accepted that further data extraction has the potential to enhance the prognostic and diagnostic power of the radiologist or oncologist. However, despite the promising aspect of radiomics, several challenges remain in the field of radiomics. The major challenges that need to be addressed before radiomics can be applied in the clinic are reproducibility, repeatability, and stability of radiomic features. Several studies have reported that some of PET radiomic features are very sensitive to different sources such as segmentation method, image acquisition and reconstruction protocols. Thus, multiple variables, parameters and condition may cause a variation on radiomic features. For increased confidence in the utilisation of texture features as imaging biomarkers, this thesis intends to determine whether

different confounding factors have an effect on PET image radiomic analysis. In this thesis, preclinical, homogeneous phantom and heterogeneous phantom studies were conducted to assess the impact of different reconstruction settings (TOF, number of iteration, number of subsets, FWHM of the gaussian filter) on PET image radiomic features. In addition, the self organising map (a type of artificial neural network algorithm) were applied to cluster and visualise the resulting data. The results presented in this body of work, indicate that different reconstruction settings have an influence on PET radiomic features and some of the robust features were able to distinguish between regions (phantom inserts). Furthermore, the findings of this thesis showed evidence that suggests self-organising map (SOM) has ability to identify emergent properties that effect their variability, in this case contour size. In addition, the SOM can be used with outcome data to serve as a predictive tool for dependent variables (e.g therapy response, prognosis). In so doing the learnt representations of self-organised features serve as the attributes for prediction which will take into consideration the statistical variability in the underlying dataset.

List of Abbreviations

AI	Articial Intelligence
ANN	Artificial Neural Network
COV	Coecient Of Variation
CT	Computed Tomography
FDG	Fluorodeoxyglucose
FWHM	Full Width at Half Maximum
GLCM	Grey Level Co-occurrence Matrix
GLDZM	grey Level Distance Zone Matrix
GLRLM	grey Level Run-Length Matrix
GLSZM	grey Level Size Zone Matrix
GMP	Good Manufacturing Practice
IBSI	International Biomarker Standardization Initiative
ICC	Intraclass Correlation Coefficient
IQ	Image Quality
LOR	Line Of Response
MLEM	Maximum-Likelihood Expectation-Maximization
NAS	Network Attached Storage
NEMA	National Electrical Manufacturer's Association
NGTDM	Neighbourhood grey Tone Difference Matrix
NHS	National Health Service
NSCLC	Non-Small Cell Lung Cancer
OSEM	Ordered Subsets Expectation Maximization
PA	Percent Agreement
PET	Positron Emission Tomography
PETIC	Positron Emission Tomography Imaging Center
SNR	Signal to Noise Ratio
SOM	Self Organising Map
SPAARC	Spaarc Pipeline for Automated Analysis and Radiomic Computing
SUV	Standardised uptake value
TOF	Time Of Flight
WHO	World Health Organisation

Presentations & Publications

The majority of projects introduced in this thesis have been presented or published as follows.

Presentations

Oral Presentations

- E. Alsyed, R. Smith, S. Paisey, C. Marshall, and E. Spezi, "The stability of PET radiomic features over time: preclinical study," presented at All-Wales Medical Physics and Clinical Engineering Summer Meeting, University Hospital of Wales, Wrexham, UK, JUNE 14, 2019.

- E. Alsyed, R. Smith, S. Paisey, C. Marshall, and E. Spezi, "Positron Emission Tomography (PET) Texture Analysis," presented at PETIC research showcase, University Hospital of Wales, Cardiff, UK, OCT 25, 2019.

- E. Alsyed, R. Smith, L. Bartley, C. Marshall, and E. Spezi, "A Phantom Study to Evaluate the Stability of PET Images Radiomic Features with Time of Acquisition," presented at Annual Congress of the European Association of Nuclear Medicine, Virtual, OCT 22-30, 2020.

- E. Alsyed, R. Smith, L. Bartley, C. Marshall, and E. Spezi, "The Effect of Increasing the Number of Iterations on the Stability of PET Radiomic Features: A Phantom Study," presented at Annual Congress of the European Association

of Nuclear Medicine, Virtual, OCT 20-23, 2021.

- E. Alsyed, R. Smith, L. Bartley, C. Marshall, and E. Spezi, "Phantom with Heterogenous Tumour Inserts to Explore the Impact of Varying Number of OSEM Subsets on PET Radiomic Features," presented at Annual Congress of the European Association of Nuclear Medicine, Virtual, OCT 22-30, 2021.

Poster Presentations

- E. Alsyed, R. Smith, S. Paisey, C. Marshall, and E. Spezi, "The Effects of Reconstruction Time on PET Radiomic Features," presented at:

  ENGIN-Cancer Research Forum, Life Sciences Hub, Cardiff, UK, APRIL 16, 2019.

  United Kingdom Imaging and Oncology congress, Liverpool, UK, JUNE 10-12, 2019.

  AI and Robotics Symposium, School of Engineering, Cardiff University, Cardiff, UK JUNE 27, 2019.

- E. Alsyed, R. Smith, S. Paisey, C. Marshall, and E. Spezi, "Stability of PET Radiomic Features: A Preclinical Study," presented at Annual Congress of the European Association of Nuclear Medicine, Barcelona, Spain, OCTOBER 12-16, 2019.

- E. Alsyed, R. Smith, P. Whybra, S. Paisey, C. Marshall, and E. Spezi, "Evaluating the Stability of PET Radiomic Features to Expectation- Maximization

Reconstruction Iterations," presented at United Kingdom Imaging and Oncology congress, Virtual, JUNE 1-3, 2020.

- E. Alsyed, R. Smith,S. Paisey, C. Marshall, C. Parkinson, P. Whybra, E. Spezi, "The Impact of Varying Number of OSEM Subsets on PET Radiomic Features: A Preclinical Study," presented at European Society for Therapeutic Radiology and Oncology Congress, E20-1258 , PO 1568, Vienna, Austria, NOVEMBER 28 - DECEMBER 01, 2020.

Publications

Conference Paper (Proceedings)

- E. Alsyed, R. Smith, S. Paisey, C. Marshall, C. Parkinson, E. Spezi. "The statistical influence of imaging time and segmentation volume on PET radiomic features: A preclinical study". Presented at: 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Manchester, UK, OCTOBER 31 -NOVEMBER 7, 2019 (DOI: 10.1109/NSS/MIC42101.2019.9059863)

- E. Alsyed, R. Smith, S. Paisey, C. Marshall, C. Parkinson, E. Spezi. "A self organizing map for exploratory analysis of PET radiomic features". Presented at: 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Boston, MA, USA, OCTOBER 31 -NOVEMBER 7, 2020 (DOI: 10.1109/NSS/MIC42677.2020.9507846)

- E. Alsyed, R. Smith, L. Bartley, C. Marshall, C. Parkinson, E. Spezi. "Artifi-

cial Neural Network Algorithm to Cluster and Visualize Phantom Experiment Data". Presented at: 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Boston, MA, USA, 31 October -7 November 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC),"(DOI: 10.1109/NSS/MIC44867.2021.9875826)"

- E. Alsyed, R. Smith, L. Bartley, C. Marshall, C. Parkinson, E. Spezi. "Toward a method of selecting robust heterogeneous PET images radiomic features". This paper will be presented at: 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), Milano, Italy 05-12 November 2022 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC),"(in press)"

Book Chapter

- E. Alsyed, R. Smith, C. Marshall, and E. Spezi (2022). "Texture Analysis Using a Self Organizing Feature Map", in El-Baz, A. (editor) Texture Analysis in Image Processing. CRC Press (Taylor & Francis Group), "(in press)"

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Cancer can start in any tissue or organ and spread to other organs of the body. Cancer begins when the uncontrolled growth of abnormal cells occurs. According to the World Health Organisation (WHO), cancer is the second leading cause of death around the world. Cancer is responsible for about 18 percent of deaths worldwide and about 10 million people die from cancer every year [1]. In 2020, the number of new cases worldwide of breast, colon and cervical cancer were 2.26, 2.21, and 1.93 million, respectively. In Great Britain alone, 375,400 new cases of cancer were reported between the years 2016 and 2018 [2]. Breast, prostate, lung and bowel cancers together were responsible for more than half (53%) of all new cancer cases [2].

In cancer management, the first step is to identify the extent of the disease to achieve an accurate tumour diagnosis. The term "stage" is used to describe the extent

of the cancer [3]. Cancer staging is very important to model an accurate treatment plan. Cancer diagnosis and staging involves a series of clinical observations that review the disease progression. In most cases, this comprises both medical imaging and biopsy (tissue sample) procedures.

Clinicians aim to select the most suitable treatment approaches based on several factors such as disease stage, patient age, and data drawn from previous clinical results. There are different treatment options for different cancer stages. Treatment can involve invasive (e.g. surgical) or non-invasive (e.g. radiotherapy) methods. Thus, multidisciplinary teams participate to manage and deliver effective diagnosis and treatments for cancer patients.

One of the advantages of medical imaging is that it can reveal the hidden structures of the human body for the assessment of patient condition without an invasive surgical procedure. Therefore, medical imaging can play a vital role in many technical aspects in cancer management areas such as screening, diagnosis, treatment and monitoring, and follow up. For this reason, imaging techniques have become indispensable in modern medicine. Different 3D imaging techniques (e.g. CT, PET, MRI) are selected for different cancer types, typically dependent on the type of tissue being evaluated.

Positron Emission Tomography (PET) imaging is one of these techniques which forms an essential part of clinical protocols for many types of cancer. PET image evaluation has been mostly constrained to qualitative assessment and semi quantitative assessment such as standardised uptake value (SUV) measurements. Recent developments in the field of imaging and data analysis have revolutionised the ability to recognise complex patterns in imaging data and provide a depth of quantitative analysis previously unachievable.

PET has potential for quantitative characterisation of disease using radiomic techniques which are defined as extracting quantitative features (information on the image content) from medical images which cannot be seen by the naked eye.

The hypothesis of radiomics is that the characteristic imaging features between various type of disease may serve as a beneficial biomarker for predicting prognosis and therapeutic response. Several aspects of radiomics are described in detail in this body of work (Section 1.2).

Despite the promise of using radiomic features as metrics in prognosis and diagnosis for several cancers, accuracy and stability of radiomic features remains as one of the most challenging aspect for implementing radiomic features as a biomarker.

This chapter provides a brief overview of PET imaging, PET physics, PET radiomics and parameters affecting PET imaging radiomic features. In addition, this chapter highlights the aim and the structure of this thesis.

1.1 PET Imaging

The National Health Services (NHS) has reported more than 34.9 million imaging tests performed in England between April 2020 and March 2021 [4]. X-ray radiographs, MRI, CT and PET scans accounted for about three quarters (73%) of all imaging tests in the England in the year 2020-2021.

Unlike other anatomical imaging technique such as CT scan which give information on anatomical structures, PET is a functional imaging modality which provides images of metabolic activity [5, 6]. Functional imaging is commonly referred to as a medical imaging technique that measures the physiological processes such as blood flow while anatomical imaging provides information on structure and geometry of tissues. The main concept of PET is based on injecting (radiotracer administration may also by inhalation or oral ingestion) the patient's peripheral vein with a short-lived radiotracer, which allows the radiotracer to become part of the vascular circulation and reach the tumour or region of interest. After about 45 to 60 min-

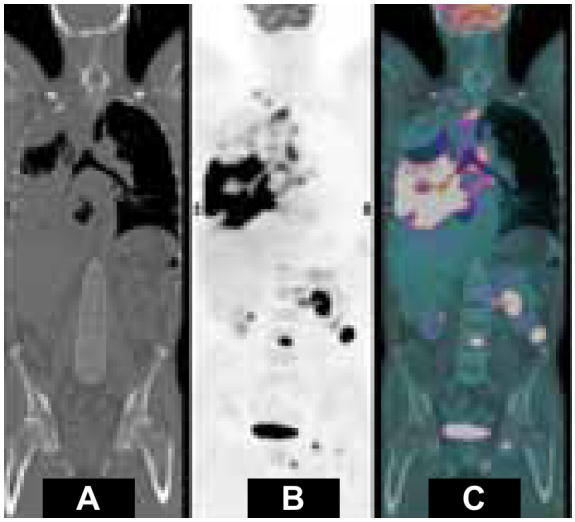

Figure 1.1: A) CT, B) PET and C) PET overlaid over CT image (adopted from [7]) for a patient with NSCLC (Non-small cell lung cancer).

utes (post injection time), patients will be placed on a PET scanner. Data acquired from the PET scanner will be reconstructed into a readable image by human eyes. Figure 1.1 presents an example of 18 F-FDG PET scan for a patient with NSCLC (Non-small cell lung cancer). PET images play an essential role in identifying the abnormalities in living tissue and detecting cancer. In addition, PET can be used to diagnose several health problems, such as Alzheimer's and heart disease [6]. More details of the physics of PET imaging are reported in the next section.

1.1.1 PET imaging physics

The historical backdrop of Positron Emission Tomography imaging began in the second half of the 20th century when the research about PET imaging technology

started. In 1950s, David E. Kuhal and Roy Edwards from the University of Pennsylvania introduced the principle of emission and transmission tomography [8]. In 1977, the first whole-body PET scanner was developed [9]. PET is a sub-modality of nuclear medicine imaging that is used in scanning radio-pharmaceutically targeted biological processes. Typically, to acquire a PET image, patients are injected (radiotracer administration may also by inhalation or oral ingestion) intravenously with a short-lived radiopharmaceutical tracer. This allows the radiotracer to become part of the vascular circulation and reach the tumour or the region of interest. The selection of radiopharmaceutical is dependent on its capability to explore the biochemical function of interest within the human body. These radiopharmaceuticals are also called 'PET tracers'. Various PET tracers have been developed and introduced to enhance the clinical utility of molecular imaging [9, 10]. The most common PET tracer is Fluorine-18 Fluorodeoxyglucose (18FDG) to provide a marker for glucose metabolic rate. The following section will highlight the characteristics of 18FDG as a PET tracer.

1.1.2 18F-FDG

Fluorodeoxyglucose (18F-FDG) an equivalent of glucose, has become a standard radiotracer for cancer patient management. The half life of 18F is 1.83 hours (110

minutes) and it is produced in a cyclotron. 18F is an unstable radioisotope that decays by beta-plus emission or electron capture and emits a neutrino υ and a positron $\beta+$. When the positron annihilates with an electron, the energy is released in the form of coincident photons. Equations below show the 18F annihilation reaction.

$$^{18}_{9}\,\mathrm{F} \longrightarrow {}^{18}_{8}\mathrm{O} + \beta$$

$$\beta + \mathrm{e}^{-} \longrightarrow 511\ \mathrm{KeV}\ \gamma + 511\ \mathrm{KeV}\ \gamma$$

1.1.3 PET imaging reconstruction

After radiotracer injection the patient is placed within the bore of the PET scanner which has scintillation detectors placed around the patient. These detectors are able to detect pairs of anti-parallel gamma (γ) photons nascent from the annihilation process (positron (e+) annihilates with an electron (e-) to release energy in the form of anti-parallel gamma photons). Figure 1.2 shows an illustrative example of the annihilation process. When an incident photon is registered in any detector a timed pulse is generated in that detector. If two pulses are generated by the detectors surrounding the patient within a short time window (3 to 12 nanoseconds), they are counted as coincidence events which are assigned to a line of response (LOR) joining the two relevant detectors.

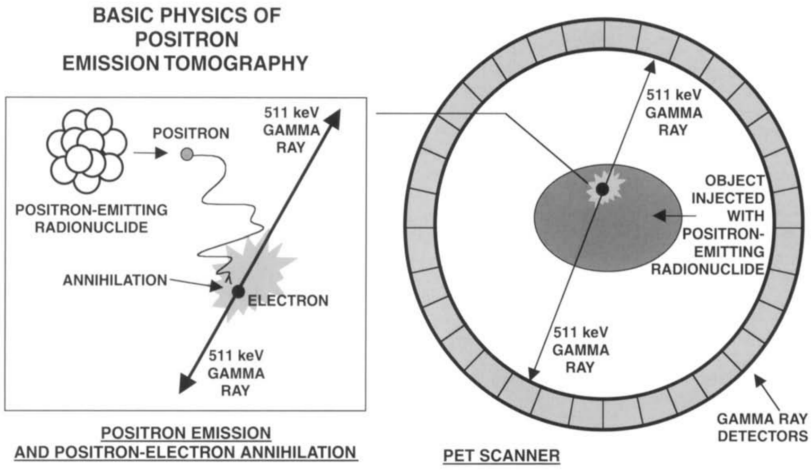

Figure 1.2: An illustrative example (from reference [11]) of the annihilation process demonstrated that two 511 keV photons produced when a positron and an electron annihilate. A circular gamma ray detector array in the PET camera registered the two 511 keV photons.

A sinogram is used to store detected annihilation occurrences. As presented in Figure 1.3, a line called the Line Of Response (LOR) may be formed between detectors for every event. The LOR drawn between pairs of detectors for detected events is A, B, C, and D, and X is the center of the detector ring gantry. The sinogram is generated by plotting each event's LOR as its angular orientation around the detector ring (from -90° to 90°) versus the LOR displacement from the detector ring's center were A, B, C, and D correspond to the LORs. The sinogram generated from the PET scanner is used to rebuild PET pictures into visual representations. Figure 1.3 also illustrates an example of sinograms for brain and the corresponding reconstructed image.

The iterative reconstruction algorithm is a popular technique used to reconstruct PET sinograms and it has become a clinically viable option in PET imaging [12]. Iterative reconstruction involves employing an iterative process and estimating an initial image (and comparing it to real-time measured values) rather than beginning with measured values and back-projects them.

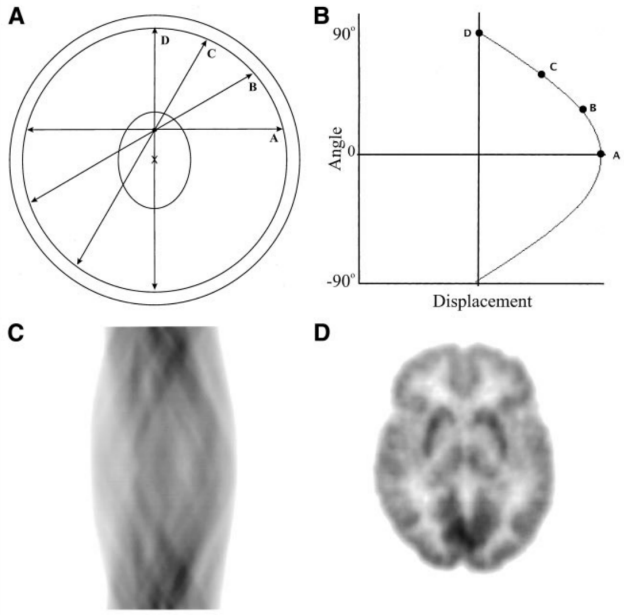

Figure 1.3: (A) PET scanner detects four LOR (lines of response) and (B) the corresponding sinogram. (C) sinograms of more complicated objects (brain) and (D) the corresponding reconstructed image. This image from reference [13].

The predictions for the approximate average picture are calculated and compared to the measured predictions in the sinogram acquired from the PET scanner [14]. Following that, a correction factor (to adjust the initial image) is added to the projections in order to create a new sinogram. The new sinogram is then

back-projected in order to create a new estimated picture, which is then used as input into the next iteration of the algorithm. When employing maximum-likelihood expectation-maximisation (MLEM), it is necessary to do hundreds of iterations of each predicted view before obtaining satisfactory agreement between the projected picture and the measured, predicted information. Figure 1.4 shows a flow chart to demonstrate the procedure of iterative reconstruction. This is a tremendously expensive process computationally. The ordered subset expectation maximisation (OSEM) method, on the other hand, shortens the calculation time by grouping the angular projections into subsets and performing MLEM on the subset as a whole rather than on each projection in the subset. In addition, data obtained during the sinogram's Time of Flight (TOF) collection can be utilised throughout the reconstruction process to enhance the contrast of the image and decrease noise. TOF technique is used to better localise the event along each LOR by measuring the actual time difference between the detection of each coincidence photon.

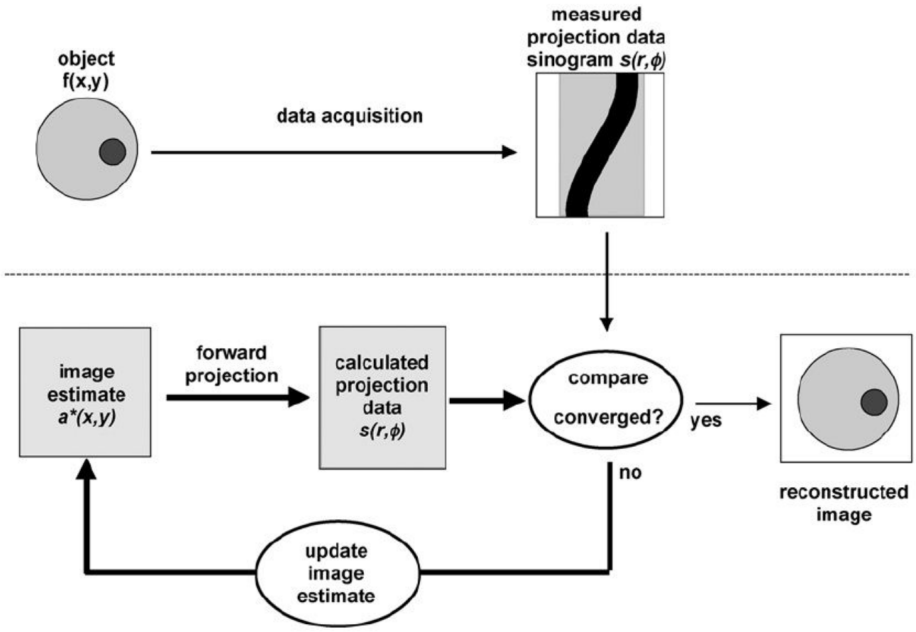

Figure 1.4: Flow chart (adopted from [15]) illustrating the steps of iterative reconstruction of PET images. 1) An estimation of initial image. 2) Calculation of projection data. 3) Comparing the estimated projection data with the measured projection data. 4) Adjusting the initial image based on the differences. 5) If the calculated projection of the estimated image closely match the true measured projection, the reconstructed image will be the final estimated (updated) image.

An increase in the number of iterations is found to cause an increase of noise, which is a random variation in brightness. Different types of image filtering methods are used to improve signal-to-noise ratio (SNR) and minimise background noise. One of the most popular image filtering methods is Gaussian filter. The Gaussian filter generates a kernel and performs a weighted average of surrounding pixels based on the equation of Gaussian (equation 1.1).

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{1.1}$$

Where μ is the mean and σ is the standard deviation which indicates spread of the distribution. The spread of the Gaussian distribution can be determined by the full width at half maximum (FWHM) instead of σ. Figure 1.5 illustrates the relationship between FWHM and σ: FWHM can be calculated using equation 1.2.

$$FWHM = \sigma\sqrt{2ln2} = 2.35\sigma \tag{1.2}$$

1.1.4 PET units

The Becquerel (Bq) is the SI unit of radioactive decay where 1 Bq is equal to 1 disintegration per second. In relation to PET imaging this is normally evaluated as a concentration of radioactivity, namely Becquerel per millilitre (Bq/mL, where 1 ml is equal to 1 cubic centimetre).

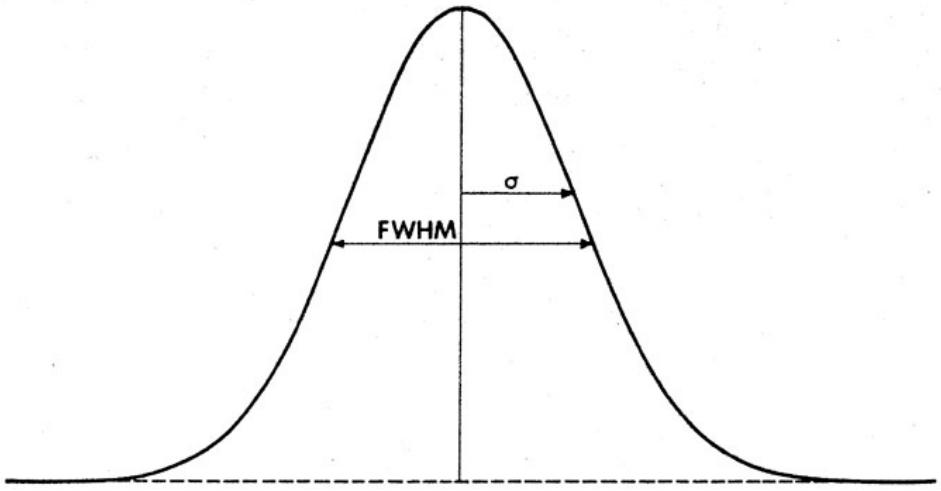

Figure 1.5: Relation between full width at half-maximum (FWHM) and the standard deviation (σ). This image adapted from [16].

In order to remove the variability between patients due to differences in body weight (W) and injected 18F-FDG, a semi-quantitative unit referred to as the standardised uptake value (SUV) can be used. SUV is generally calculated using the following equation:

$$SUV = \frac{\text{Radioactivity concentration } (Bq/ml) \ \times \ \text{Body weight } (g)}{\text{Injected dose } \ \text{(Bq)}} \tag{1.3}$$

Under the assumption that 1 ml of tissue weights 1 gram, SUV will be dimensionless [7].

1.1.5 Imaging data and DICOM format

All digital images are composed of a discrete grid of elements called pixels (picture elements) for 2D or voxels (volume elements) for 3D. Each element in the grid is assigned an intensity value and images are created based on the arrangement of these intensities. In medical imaging, the intensity values are usually referred to as grey levels. Therefore, images visualised in grey scale on a gradient form where the lowest intensity refers to black and the highest to white.

Most often, medical scanners acquire 3D volumetric data at a particular resolution such that each voxel has the same dimensions. Each voxel has a corresponding coordinate position (k) and the coordinate grid of the image is orderly spaced start-

ing from specified origin. The entire image (I) can be represented in matrix notation such that each element describes the intensity (grey value) in defined position. As an example, the following matrix represents the intensity levels for 2D image of size $r \times$c.

$$I = \begin{bmatrix} I_{11} & I_{12} & ... & I_{1c} \\ I_{21} & I_{22} & ... & I_{2c} \\ ... & ... & ... & ... \\ I_{r1} & I_{r2} & ... & I_{rc} \end{bmatrix}$$

Where I_{rc} provides the intensity value (grey level value) at a particular row (r) and column (c). In 3D image, I_{rcs} is utilised to represent the grey level at specific row (r), column (c) and slice (s). This matrix indices are used for feature extraction algorithms by performing computational operations and arriving at a scalar representation for that matrix. Each pixel has two horizontal and two vertical neighbours [17]. If pixel p is at coordinates (x,y), neighbouring pixels would be at coordinates:

$$(x+1,\ y), (x-1,\ y), (x,\ y+1), (x,\ y-1) \tag{1.4}$$

This set of pixels is known as the 4-neighbours of p. In addition to this set of neighbouring pixels, pixel p also has four diagonal neighbours which have coordinates:

$$(x+1,\ y+1), (x+1,\ y-1), (x-1,\ y+1), (x-1,\ y-1) \tag{1.5}$$

These neighbors, together with the 4-neighbors, are known as the 8-neighbors of p and the term neighbourhood of pixel p refer to the set of image locations of the neighbors [17].

Digital images can be stored in various file types. DICOM (Digital Imaging and Communications in Medicine) format is one of the most commonly utilized file formats for medical imaging [18]. Most of the modern scanners use the DICOM standard due to its ability to manage slice-by-slice image acquisition and its integration with Picture Archiving and Communications Systems (PACS). PACS provide the framework for viewing and storing a variety of patient imaging in clinical protocols. The structure of DICOM files consists of a header which is its defining feature and subsequent image data. Each DICOM file may be configured to involve patient information such as name, date of birth, and hospital ID. In addition, the DICOM file can contain acquisition protocols such as equipment, series, study, and patient [18].

1.2 PET Radiomics

Currently, PET images and other medical images are interpreted visually by radiologists and clinicians. However, medical images contain more information than can be assessed visually [19]. The rapid development of computational power has revolutionized the ability to recognize complex patterns in imaging data and provide a depth of quantitative analysis previously unachievable [20]. Radiomics is defined as extracting quantitative features from medical images which cannot be seen by the human naked eye. It is now accepted that further data extraction has the potential to enhance the prognostic and diagnostic power of the radiologist or oncologist [21, 22]. The hypothesis of radiomics is that the characteristic imaging features between various types of disease may serve as a beneficial biomarker for predicting prognosis and treatment response.

A biomarker is a term frequently used in the literature to mean an objective indication of either regular or irregular biological process and the term treatment response is generally understood to mean the likelihood a patient will respond to a given treatment plan [23]. Biomarkers provide value in cancer screening, staging, prognosis, and monitoring [23].

The goal of biomarkers is to mark a clinical end point such as survival or using them as a surrogate endpoint. As an example, a shrinking tumour volume can be a surrogate endpoint to show the effectiveness of the treatment. In general, radiomic studies for imaging biomarkers focus on extracting numerous tumour characteristics from medical images that can be linked to clinical endpoints.

The goal of this is to offer predictive benefit, specially if incorporated into clinical practice [24]. Section 1.2.2 will show examples of radiomic applications and the relationship between radiomic features and clinical assessments. Next section (1.2.1) will describe the concept and workflow of radiomics.

1.2.1 Radiomic workflow

In principle, there is no significant difference between the process of radiomics in PET and other imaging modalities. The radiomic process has four main steps, which are common to all imaging techniques, such as CT, MRI, and PET. The first step is image acquisition on standardised imaging protocols. Acquiring medical images using a standardised protocol may help to eliminate unnecessary confounding variability [25]. Second is the delineation or segmentation of the region of interest, to define the volume for feature extraction. Third is extracting the radiomic features using a

software for image processing. The final step is radiomic analysis to specify suitable features for patient management. All of these stages should be done carefully, since the outcome may vary after changing some parameters. Figure 1.6 shows the general workflow for radiomic processes [26, 27].

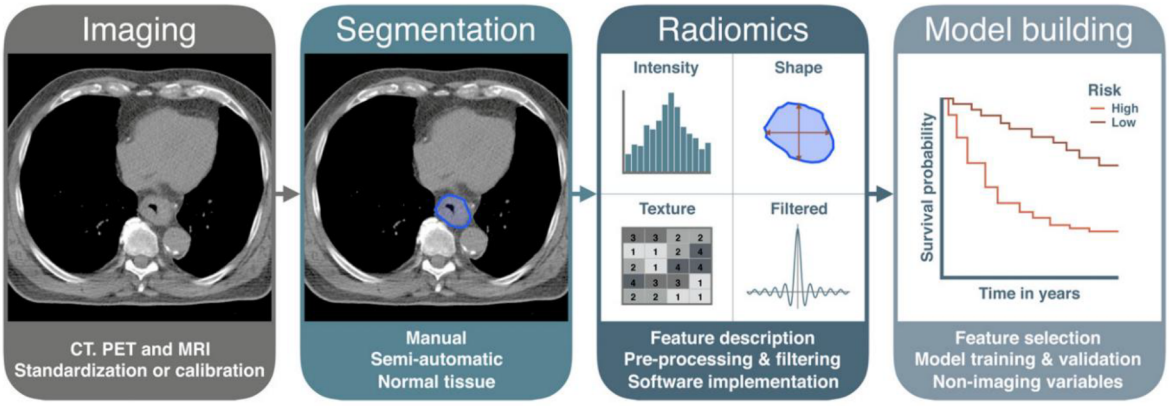

Figure 1.6: A workflow for radiomic processes [27]. Radiomic process generally involves four phases; imaging, segmentation, features extraction and features selection.

1.2.2 Radiomic applications

There is a large volume of published studies describing the role of radiomic features to serve as a complimentary tool of decision making for clinicians in the future. This section will introduce some of radiomic studies, focusing on the clinical application of PET radiomics in several cancers.

According to recent reports, some of PET radiomic features were found to be remarkably correlated with the survival of patients with gliomas [28, 29]. A model can be formed using PET radiomics to enhance prognosis and treatment planning in the future. As an example, Crispin-Ortuzar *et al.* used PET radiomic features in nasopharyngeal cancer (NPC) patients to predict hypoxia status [30].

Previous studies have reported that breast cancer subtype can be predicted utilising PET radiomic features [31, 32]. Wu *et al.* were able to show the possibility of predicting distant metastasis with 101 early-stage Non-small cell lung cancer (NSCLC) patients using PET radiomic features [33]. In a study which set out to assess the treatment response of Erlotinib in NSCLC using features extracted from PET images, Cook *et al.* found that there is a correlation between changes in first-order entropy and overall survival and treatment response [34].

The prominence of PET radiomic features to help prognosis has been reported in a study conducted by Lovinfosse *et al.* In this study various types of PET image texture features were extracted and analysed. They found that textural analysis gave strong independent predictors of survival in patients with locally-advanced rectal cancer (LARC) [35]. The next section provides information on different types of texture features.

1.2.3 Texture features

Quantitative features may be divided into two main categories: (1) first order features and (2) higher order features [36, 26]. First order features are based on the histogram that describes distribution of voxel values (intensities) in an image. Mean, median, maximum, minimum values of the voxel intensities are examples of the first order features. Higher order features which might be called texture features provide information about inter-relationships between neighbouring pixels/voxels [36, 26]. Thus, the term texture is used to mean the spatial variation in pixel/voxel intensity levels [37]. Texture features from five major categories (grey Level Co-occurence Matrix (GLCM), grey Level Run Length Matrix (GLRLM), grey Level Size Zone Matrix (GLSZM), Grey Level Distance Zone Matrix (GLDZM), neighbourhood grey Tone Difference Matrix (NGTDM)) were included in this thesis and described in detail below.

Calculating a matrix which summarises the neighbourhood properties of interest (for each category) is the first step toward computing the different texture features. Then, various mathematical operations on these matrices might be applied to obtain texture features. Texture features were basically designed to assess surface texture in 2D images, but can be extended to 3D images using one of the following methods.

- First: feature values are calculated using the matrix of each slice and then the mean value is calculated over the feature values for all slices.

- Second: merging the different matrices into a single matrix by summing the elements over the slices. Subsequently feature values are calculated on the combined matrix.

In this thesis, we followed the first strategy (averaging the texture features computed in each slice of an imaging volume stack). Illustrative examples of the calculation of texture matrices (GLCM, GLRLM, GLSZM, GLDZM, NGTDM) in 2 dimensional (for the sake of simplicity) will be presented in the following sections.

1.2.3.1 Grey Level Co-occurence Matrix (GLCM)

The grey level co-occurrence matrix (GLCM) is a matrix that expresses how often pairs of pixels with specific values occur in an image. GLCM matrix represents the combinations of grey levels of neighbouring pixels that are distributed along one of the image directions.

Let matrix M ($A \times A$) represents image grey levels, GLCM matrix of size $A \times A$, i and j are pixels of grey level value. Each GLCM (i, j) element will be filled with the occurrences of neighbour pixel pairs with i and j values (i.e. how many grey level

i are neighbours with pixels of grey level j). Figure 1.7 shows a good illustration of how the first three values in a GLCM were calculated. In the input image, there is only one instance where two horizontally adjacent pixels have the values 1 and 1, respectively. As a result (of this), element (1,1) in the output (GLCM) contains the value 1. For more examples of this kind of calculation, element (1,2) in GLCM has the value 2 because two horizontally adjacent pixels have the values 1 and 2. In contrast, there are no instances of two horizontally adjacent pixels with the values 1 and 3. Therefore, GLCM (1,3) has the value 0.

After the GLCM is created, various features can be derived using the different formulas. As an example, the following equation (1.6) , (1.7) and (1.8) will be used to calculate joint maximum (probability of the most common grey level co-occurrence in the GLCM), joint average (sum of joint probabilities) and joint variance (sum of squares) features, respectively.

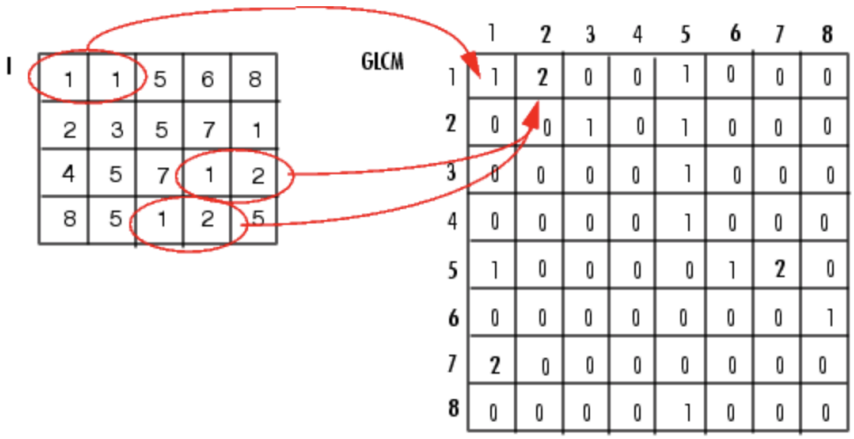

Figure 1.7: An illustrative example of how Grey level co-occurrence matrices created. The input matrix in the left and the calculated GLCM in the rghit. This illustrative example from reference [38].

$$F_{joint.max} = max(p_{i,j}) \tag{1.6}$$

$$F_{joint.avg} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ip_{i,j} \tag{1.7}$$

$$F_{joint.var} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} (i - F_{joint.avg})^2 p_{i,j} \tag{1.8}$$

1.2.3.2 Grey Level Run Length Matrix (GLRLM)

In 1975, Galloway [39] introduced the grey level run length matrix to assess the distribution of discretised grey levels in an image. Unlike GLCM which assesses the combination of levels between neighbouring pixels or voxels, GLRLM assesses grey level run length. The grey level run length can be defined as the frequency of consecutive pixels that have the same grey level value for a given direction. Figure 1.8 exhibits an example of GLRLM. Element (1,1) in GLRLM(0°) contains the value 4

because four pixels have grey level of 1. In addition, element (1,2) in GLRLM(0°) has the value 1 because two continuous pixels have the values of 1.

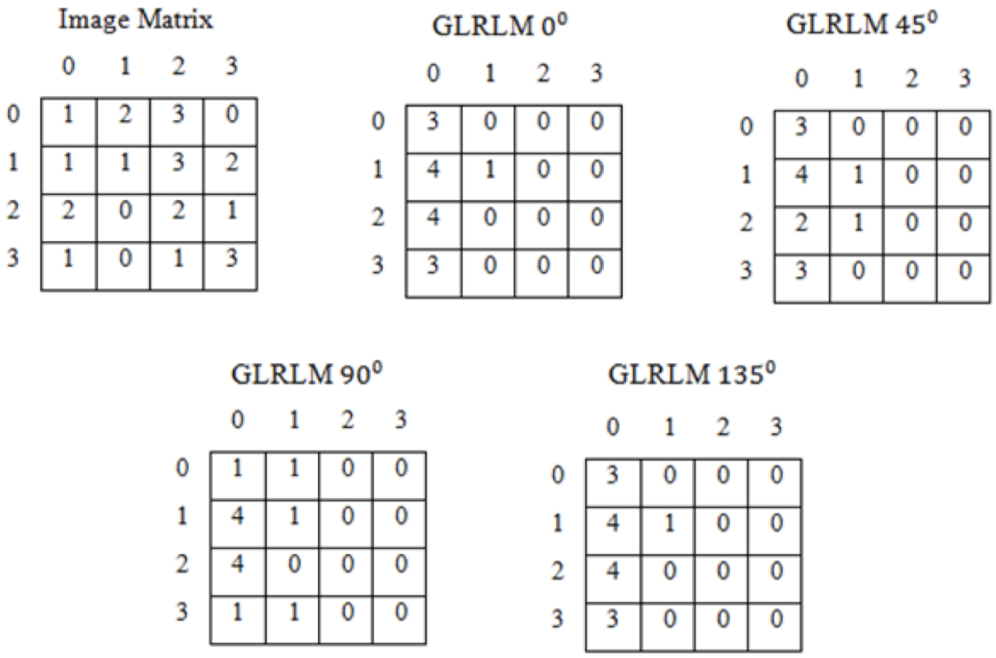

Figure 1.8: An illustrative example of the grey level run length matrices along four different direction (0°, 45° ,90° and 135°) [40].

1.2.3.3 Grey Level Size Zone Matrix (GLSZM)

GLSZM was suggested by Thibault *et al.* in 2014 and computes the number of zones (groups) of connected pixels/voxels with a specific discretised grey level value and size [41]. Pixels/voxels are considerd to be connected if the neighbouring pixles/voxels have similar grey level value. Figure 1.9 shows an example of GLSZM.

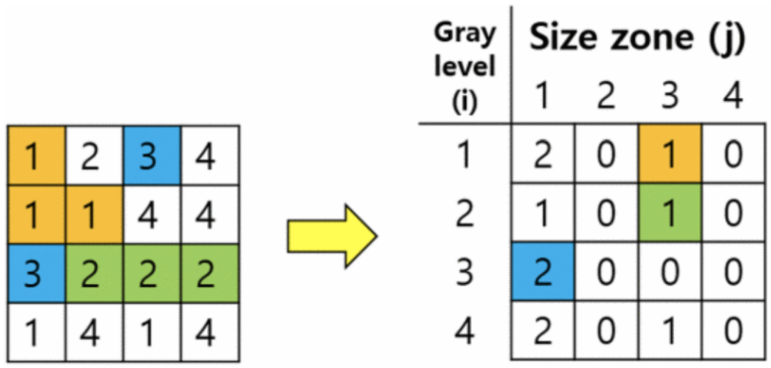

Figure 1.9: An illustrative example of how GLSZM is calculated. Figure adapted from reference [42].

1.2.3.4 Grey Level Distance Zone Matrix (GLDZM)

GLDZM quantifies the number of zones (groups) of pixels/voxels with a certain grey level value and same distance to the predefined ROI edge [41]. GLDZM identifies the relation between grey level and location. To create GLDZM, two maps (grey level grouping map and the distance map) are needed. An illustrative example of how GLDZM is calculated is shown in figure 1.10.

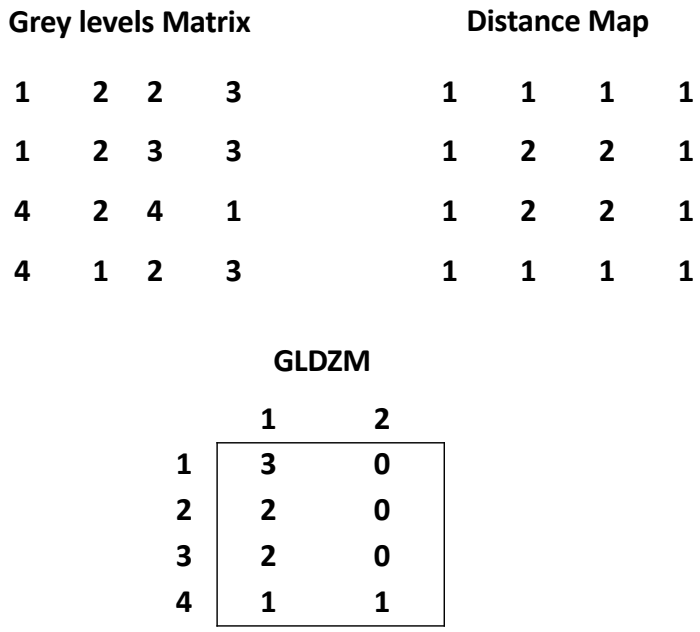

Figure 1.10: An illustrative example of how GLDZM is calculated.

1.2.3.5 Neighbourhood grey Tone Difference Matrix (NGTDM)

In 1989, Amadasun and King introduced the neighbourhood grey tone difference matrix (NGTDM) which includes the sum of differences between the grey level of a pixel/voxel and the mean grey level of its neighbouring pixel/voxel within a predefined distance [43]. An example of creating NGTDM is shown in Figure 1.11.

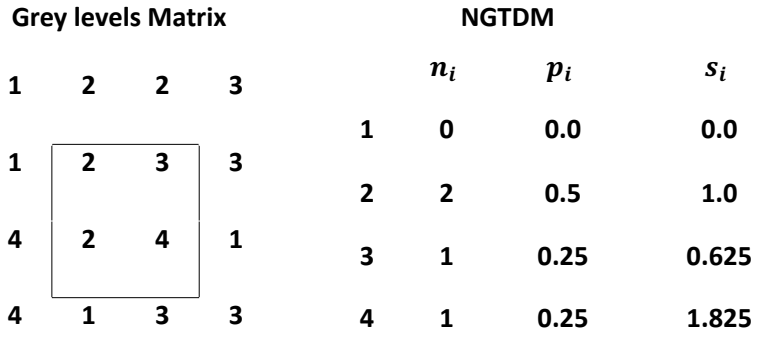

Figure 1.11: An illustrative example of the neighbourhood grey tone difference matrix at d=1. n_i is the grey level count, p_i is the grey level probability and s_i is the neighbourhood average grey level.

1.2.4 Challenges in radiomic

Many studies demonstrated the potential of PET radiomics to serve as a linchpin to personalised cancer treatment with the associated promise of improvements in survival [21, 22]. However, despite the promising aspect of radiomics, several challenges remain in the field of radiomics [44]. A major challenge which needs to be managed before radiomics can be applied in the clinic is the reproducibility, repeatability, and stability of radiomic features.

Several studies have reported that some PET radiomic features are very sensitive to different sources such as segmentation method, image acquisition and reconstruction protocols [21, 45, 46, 47, 48, 49]. Thus, multiple variables, parameters and

conditions may cause a variation in radiomic features. Some of these parameters will be discussed further in the next section.

1.2.5 Parameter affecting PET imaging radiomic features

The effectiveness of implementing radiomic features relies greatly on the robustness of PET radiomics parameters in dealing with inter/intra-scanner image acquisition variability as well as a wide range of uncertainties present in radiomics analysis. Several studies have examined reproducibility and reliability issues, focusing on the effect of PET radiomics characteristics due to variables such as PET image acquisition and reconstruction methods, segmentation volume, intensity discretization scheme, etc [50, 51, 52, 53, 49, 54, 45]. This section will highlight a considerable amount of literature that has been published on the impact of reconstruction settings on PET radiomic features.

In a population of 20 patients with solid tumours, Galavis *et al.* evaluated the impact of reconstruction settings (reconstruction algorithm, iteration, width of the post- reconstruction filter and size of the matrix) on the variability of 50 PET radiomic features (first, second and high order) [46]. 40 out of 50 (80%) features (e.g. GLCM -contrast, dissimilarity and coarseness) showed high variability. However, the

remaining features showed low and moderate variability. Yan *et al.* conducted similar research on a group of 20 patients with primary or secondary lung lesions larger than 5 cm^3, taking into consideration additional variables including time of flight (TOF) and point spread function (PSF) incorporated in the reconstruction [55]. PSF is used in the reconstruction algorithm to improve the spatial resolution and TOF is utilised to better localize the event along each LOR by measuring the actual time difference between the detection of each coincidence photon. In Yan *et al.*'s study, intensity histograms and texture matrices of second and higher order were included. The number of iterations and the size of the post-reconstruction filter have less of an effect than the grid size. Orlhac *et al.* gathered data from 54 patients with breast cancer and 10 spherical phantoms with heterogeneous level of 18F-FDG. This study concluded that the size of the voxels affects the value of certain features. Some features (such as the high grey-level zone emphasis and low grey-level zone emphasis) were more resistant to the examined changes [56].

Lasnon *et al.* assessed the impact of the reconstruction method on the quantitative PET parameters (e.g. SUVmax, SUVmean) and some textural features (e.g. entropy, correlation, dissimilarity, and zone precentage) in a larger series of 71 tumours examined in a population of 60 lung cancer patients [57]. Their study included

three different reconstruction methods (PSF with three iterations and 21 subsets, the PSF with three iterations, a 7 mm Gaussian filter and 21 subsets, and the OSEM type with four iterations and eight subsets). There was no significant difference found between the OSEM and PSF reconstruction for any of the intensity or textural examined features. The filterless PSF (as compared to the OSEM) provided higher values of SUVmax, dissimilarity, SUVmean and correlation.

Van Velden *et al.* examined the impact of the reconstruction technique on the reproducibility of 18F-FDG PET radiomic features in a group of 11 patients with lung cancer [58]. Each patient had two PET scans within three days with new images produced using two different reconstruction settings. In the first setting, EANM recommendations [59] were followed, and Gaussian filter applied. In the second settings, PSF was used in addition to the first methods. TOF, OSEM, 3 iterations and 33 subsets were used in both settings. Researchers found that most of studied radiomic features were reproducible against the two examined reconstruction methods.

PET scans for 64 patients with oesophageal cancer were used by Doumou *et al.* [60] to investigate the effect of image noise smoothing by a Gaussian filter on PET radiomic features. In their study, four different levels of Gaussian filter (2, 2.5,

3, 3.5, and 4 mm) were applied. No significant influence in 70% (37) of features was found when smoothing filter level changed. Sixty patients with lung cancer were recruited for Grootjans *et al.*'s study to examine the effect of varying statistical noise levels on PET texture parameters include zone percentage, entropy, dissimilarity and high-intensity emphasis. This research demonstrated that there are no significant effects on the PET texture parameters when the statistical quality of the images is varied [61].

In a study which set out to assess the robustness of 27 PET image parameters, Forgacs *et al.* found that just four parameters were independent and reproducible against reconstruction settings and delineated volume. These parameters include Coefficient of Variation, Entropy, Correlation, and Contrast [62].

In a large longitudinal study, Shiri *et al.* investigated the stability of PET image radiomic features with changing various type of reconstruction settings (number of subsets, matrix size, scan time per bed position, and the FWHM of the Gaussian filter) [48]. In this study, they used scans for a NEMA phantom and 25 patients. A hundred radiomic features from different feature families (GLCM, GLRLM, GLSZM , NGLD, NGTDM, Intensity, Geometry) were extracted. They found that the reconstruction settings and algorithms had a large impact (COV $>$ 20%) on the GLRLM

(LIRE, LISRE, LILRE), GLSZM (LISZE, LILZE), GLCM (CS), TFC (coarseness) and SUV (SUVvar, SSV1, SsuvV2). Their study showed that the matrix size had a high impact on 56% of all features (COV $>$ 20%), while 6% of all features were robust against matrix size. With changing time per bed position, 52% of features were found to have high stability (COV $\leq$ 5%). Their results showed that features such as GLCM (Entropy, Dissimilarity), GLRLM (LRE, RP), NGLCM (Homogeneity, Entropy), TFC (Homogeneity) were robust against all of the examined reconstruction settings. Variations were similar across the two studies (phantom and patients).

In a phantom study, Pfaehler *et al.* [63] investigated the repeatability of F-FDG PET radiomic features over different reconstruction protocols (SBRs, scan durations, matrix sizes, and FWHMs). They extracted large number (246) of PET radiomic features and intraclass correlation coefficient (ICC) was utilised to assess feature repeatability. The evidence from this study suggests that the highest repeatability occurred when the point-spread-function (PSF) was used. In total, 53% of features were repeatable with PSF, compared to 30% and 32% with OSEM and TOF, respectively. In addition, the results of this study indicated that features extracted from large spheres with high uptake had higher repeatability than small spheres with low uptake.

1.3 Thesis aims

For increased confidence in the utilisation of texture features as imaging biomarkers, this work intends to determine whether different confounding factors have an effect on PET radiomic analysis. A project entitled "Positron Emission Tomography Image Analysis" carried out in cooperation with:

1) The Wales research and diagnostic PET imaging centre (PETIC) which a result of a £16.5M investment by the Welsh Government. PETIC commenced a routine imaging service in 2010 and produced the first good manufacturing practice (GMP) standard radiopharmaceutical in 2011. Today, more than 2,000 patients are scanned in PETIC per year. They provide research opportunities across the schools of Medicine, Bioscience, Chemistry and Engineering.

2) Cardiff University which is one of the top universities in the UK for different branches of engineering. Cardiff School of Engineering carry out world leading research and supportive teaching environment to make it one of the leading engineering schools in the UK.

This thesis aims to address the following research objectives:

(a) Investigate the impact of the post injection time on stability of FDG PET

radiomic features and determine which radiomic features are more stable than others to use in studies that involve multiple time point acquisitions.

(b) Assess the variation of FDG PET radiomic features due to varying reconstruction settings (OSEM, number of subsets, number of iterations and TOF). This assessment aimed to identify which radiomic features are more stable than others to use in studies that involve different reconstruction settings protocols.

(c) Investigate the statistical influence (impact of two combined variables) of imaging time and segmentation volume on PET radiomic features.

(d) Explore, capture and cluster the statistical variability of PET texture parameters using a self-organising map (SOM). SOM is a type of artificial neural network (ANN) that is trained using unsupervised learning to produce a lower dimensional representation of the input data on an underlying manifold.

1.4 Thesis structure

This thesis has been divided into eight chapters. The first chapter deals with a literature review into PET imaging and radiomics. This chapter also provides a background of radiomic PET imaging and the challenges which exist on the implementation of radiomics. In addition, some parameters affecting PET radiomic

features are highlighted in this first chapter. The second chapter is concerned with a pre-clinical study to assess the impact of acquisition time and reconstruction settings on PET radiomic features. The third and fourth chapters present the findings of the research, focusing on the statistical analysis and using self organising map for exploratory analysis of PET radiomic features, respectively. SOM can produce a lower dimensional representation of the input data and cluster the statistical variability of PET texture parameters. The assessment of the stability of the radiomic features against acquisition of time and reconstruction settings using homogeneous inserts will be highlighted in the fifth chapter while chapter six deals with heterogeneous inserts. The seventh chapter presents the findings of the research, focusing on using self organising maps for visualising phantom data. Finally, the eighth chapter provides a discussion and the findings from these studies and the contributions to the current literature. Furthermore, chapter eight highlights the future work and proposed studies in this field.

1.5 Contributions

Chapters within this thesis involve published work in which the author was a lead and the major contributions are as follows:

- Chapter 2 contain material that was published in Radiotherapy and Oncology Journal as an abstract titled "The Impact of Varying Number of OSEM Subsets on PET Radiomic Features:A Preclinical Study" [64]

  – In this work (preliminary analysis of preclinical data), the author examined the effect of changing number of OSEM subsets on PET radiomic features.

  – The author primarily designed the study and performed all the data analysis.

  – The author presented and drafted the original abstract.

- The study in Chapter 3 was published in IEEE Xplore as a manuscript titled "The statistical influence of imaging time and segmentation volume on PET radiomic features: A preclinical study" [53].

  – In this work (preliminary analysis of preclinical data), the author evaluated the impact of two parameters on 18F-FDG PET radiomic features, especially segmentation volume and acquisition time. The determinant of the correlation matrices for each feature whilst varying segmentation contour sizes and image acquisition times was assessed.

  – The author primarily designed the study and performed all the data

analysis.

– The author presented and drafted the original abstract.

- The study in chapter 4 was published in IEEE Xplore as a manuscript titled "A self organising map for exploratory analysis of PET radiomic features" [54].

  – In this present study, the author provided additional evidence that the statistical distribution of the input texture has clear modes which are dependent on contour size.

  – The author mainly performed the analysis and drafted the original manuscript.

- Chapter 6 contains material that was published in European Journal of Nuclear Medicine and Molecular Imaging as an abstract titled "The Effect of Increasing the Number of Iterations on the Stability of PET Radiomic Features: A Phantom Study" [65] and "Phantom with Heterogeneous Tumour Inserts to Explore the Impact of Varying Number of OSEM Subsets on PET Radiomic Features" [66].

  – In this study (phantom data), the author evaluated the impact of changing number of OSEM subsets and number of iterations on the stability of PET radiomic features.

– The author mainly designed and conducted the experiment.

– The author performed all the data analysis and drafted the original manuscript.

In addition, section 6.4 will be published in IEEE Xplore as a manuscript titled "Toward a method of selecting robust heterogeneous PET images radiomic features" [67]. In this study, the author assessed the ability of radiomic features to capture heterogeneity differences between region of interest using the Friedman test.

- The study in chapter 7 was published in IEEE Xplore as a manuscript titled "Artificial Neural Network Algorithm to Cluster and Visualise Phantom Experiment Data" [68].

  – The author used a self organising map (SOM) to visualise and cluster the data derived from phantom experiment.

  – The author mainly performed the analysis and drafted the original manuscript.

Chapter 2

The Impact of Acquisition Time and Reconstruction Settings on PET radiomic features: A Pre-Clinical Study

2.1 Introduction

As mentioned in the literature review (previous chapter), despite the promise of radiomics, some challenges remain in this field. Radiomic features may vary with different conditions and factors [19, 50]. To increase confidence in the utilisation of texture features as imaging biomarkers it is necessary to understand the effect these confounding factors have on radiomics analysis in an attempt to show that these features genuinely match biomarkers. A confounding factor or confounding variable can be defined as an "extra" variable that was not taken into account in radiomics studies. For example, if researchers are examining whether there is a

correlation between overall survival and changes in radiomic features, confounding variables (e.g. post-injection imaging time and reconstruction settings) are any other variables that also have an effect on changes in radiomic features.

Animal models are non-human, living beings that are utilised to examine human illnesses or biological processes without the risk of hurting humans. In the field of oncology, mouse serves as a suitable animal model, whereas some laboratories use other animals such as pigs and dogs which play a significant role in several experiments, such as studying heart diseases and central nervous system problems. Despite that, mouse is considered a good animal model of human disease due to its match with the human genome (99%), fast-breeding cycle and low maintenance cost.

Imaging of animal models like rats and mice is called small animal imaging and the PET scanners dedicated to scan small animals are called small animal scanners or preclinical PET scanners. Small animal PET is an increasingly utilised diagnostic tool to evaluate the effectiveness of therapeutic interventions and to confirm the translation of radiotracers to humans [69]. Even though there are differences in the underlying biology between humans and animals, preclinical research provides a bridge to human clinical translation enabling identical experiments to be carried out across species. Indeed, several studies have found that the patient-derived xenografts

(PDX), human tumour tissue implanted into mice for research objectives, accurately reflect patients' tumours as to gene expression profiles and the histomorphology [70]. Most often, the mouse is the experimental model used in studies for pre-clinical development. It is usually because of its fast breeding cycle and low maintenance costs [69]. Involving pre-clinical research in the field of texture analysis provides an exciting opportunity to advance our knowledge of PET image radiomics.

Even though numerous investigations in the field of radiomics can be done on retrospective clinical data, there are substantial reasons why conducting preclinical research is useful for this area as well. One of these reasons is that there has been an increasing interest in the use of PDX in preclinical PET image radiomics in terms of developing image metrics of response to therapy. An example of this type of research is what has been conducted by Roy *et al.* [71] where they recruited both triple-negative breast cancer (TNBC) patients and subtype-matched PDX into a co-clinical FDG-PET imaging trial to predict response to therapy. They demonstrated that four PDX image features performed better as to predicting response to therapy relative to standard SUV metrics.

Thus, it is necessary to identify whether PDX image features can be impacted by different factors (e.g post-injection imaging time and reconstruction settings) in

an attempt to demonstrate that these features genuinely match biomarkers.

To the best of this author's knowledge, no published work has looked at the effect of acquisition time (post-injection imaging time) and reconstruction settings on pre-clinical PET radiomic features. Therefore, in this chapter (2) we will investigate the effect of these parameters. The retrospective preclinical scans for eight mice with 4T1 (mouse mammary tumour cell line) tumour are available in PETIC in the raw data format enabling researchers to reconstruct the images again but changing the image acquisition and reconstruction parameters (e.g. number of subsets and iterations). For instance, images can be rebinned into different (x number of) time points post injection where each image will have m minutes acquisition period (Equation 2.1). This preclinical (pilot) study may assist researchers in the future to establish a greater degree of planning and modification of the clinical study, especially as to post-injection imaging time.

$$m = \frac{\text{Total acquisition time}}{x} \tag{2.1}$$

2.2 Materials and methods

In this retrospective study, we included eight female BalbC (strain of albino mice) mice imaged after grafting of 4T1 tumour cells. In the following sections, different aspects of the collected pre-clincal data will be described.

2.2.1 Scanner

The pre-clinical scanner available in PETIC was Mediso (Mediso Medical Imaging Systems, Budapest, Hungary) nanoScan PET/CT. This scanner is suitable for dynamic imaging of up to four mice simultaneously with high spatial resolution (down to 700 µm) and 12 cm field of view. The operation system of this scanner requires PET/CT device, acquisition workstation and post-processing workstation. Figure 2.1 shows an illustrative layout of the Mediso PET/CT nano scanner system build-up and a picture for the scanner is presented in Figure 2.2.

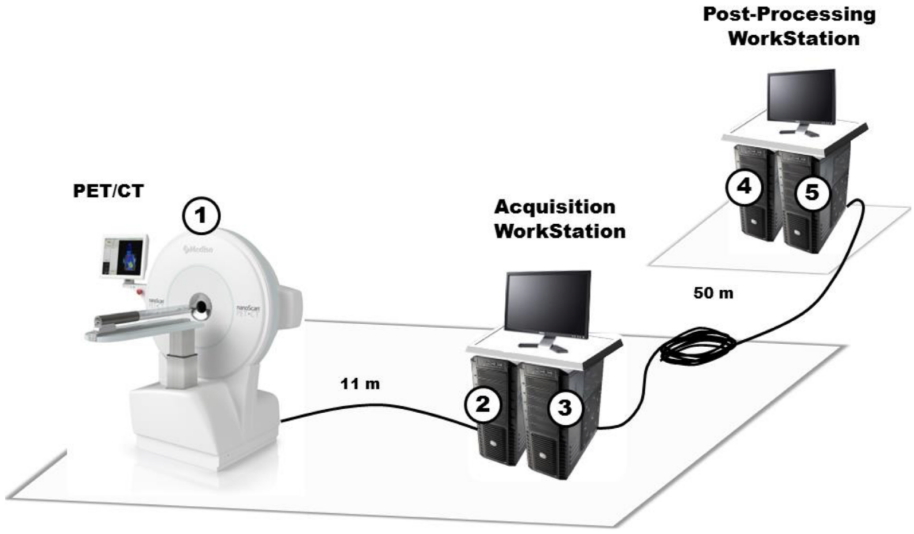

Figure 2.1: An illustrative layout of the system build-up. 1: PET/CT device, 2: Nucline PC for controlling acquisitions, 3: Workstation for PET acquisition and reconstruction, 4: Interview Fusion PC for controlling PET reconstructions and image post-processing, 5: Post workstation for PET reconstruction. This figure from reference [72].

Two software systems were used in this work. First, Nucline (Mediso, 2012) for controlling acquisitions and Interview Fusion (Mediso, 2012) for controlling PET reconstructions and image post-processing.

2.2.2 Animals

Eight female BalbC mice aged 12 to 16 weeks were housed in the conventional, non-specific pathogen-free (SPF) facility of PETIC. Mice were imaged 11 days after grafting of 4T1 tumour cells in matrigel into the mouse mammary glands. All experimental protocols were approved by the Local Ethical Committee. Before PET-CT imaging, each mouse was injected with 100 μL of iopamidol (Niopam 300, Bracco) and

anaesthetised with 3-3.5% isoflurane (Isoflurane, Teva Hospitals) delivered through a nose cone and placed on the heating pad. Mice were then injected via tail vein with 20 MBq of FDG.

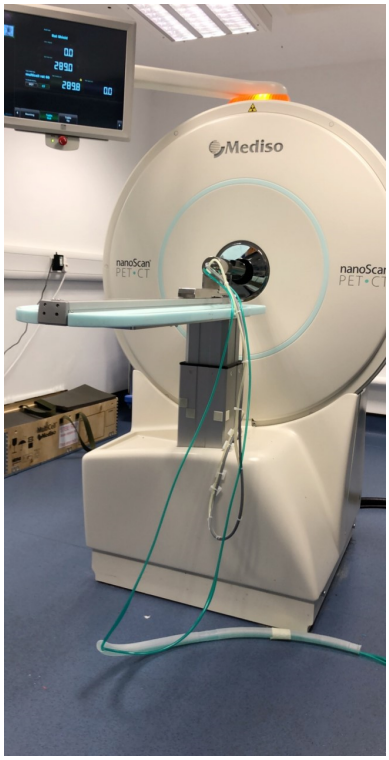

Figure 2.2: A picture of Mediso PET/CT nano scanner.

2.2.3 Data acquisition

At 50 minutes (standard acquisition time point [73]) post injection, mice were placed in the prone position in the scanner with the use of the Mediso Multicell 2 mouse animal bed and scanned for 20 minutes PET and 2.5 minutes CT. The anesthesia

was maintained through entire image acquisition. Respiration was monitored with pressure pad connected to differential pressure transducers for low-range pressure monitoring during entire PET-CT examination. Reconstructed resolution of CT was 0.25 mm. PET scans were reconstructed with attenuation correction at 5 minutes time frames with a spatial resolution of 0.4 or 0.3 mm and energy lower/upper limit was 400/600 keV.

2.2.4 Data transfer

The following scans were sent to Cancer Imaging and Data Analytics group (CIDA) via Network-Attached Storage (NAS):

- Whole body CT scan.
- Whole body PET scan.

The scans were then retrieved from the NAS for analysis on workstations in CIDA, Including Velocity 3.2.1 (Varian Medical Systems, Atlanta, USA) and Spaarc Pipeline for Automated Analysis and Radiomic Computing using an in-house developed tool built on Matlab (SPAARC). Figure 2.3 shows the data transfer workflow.

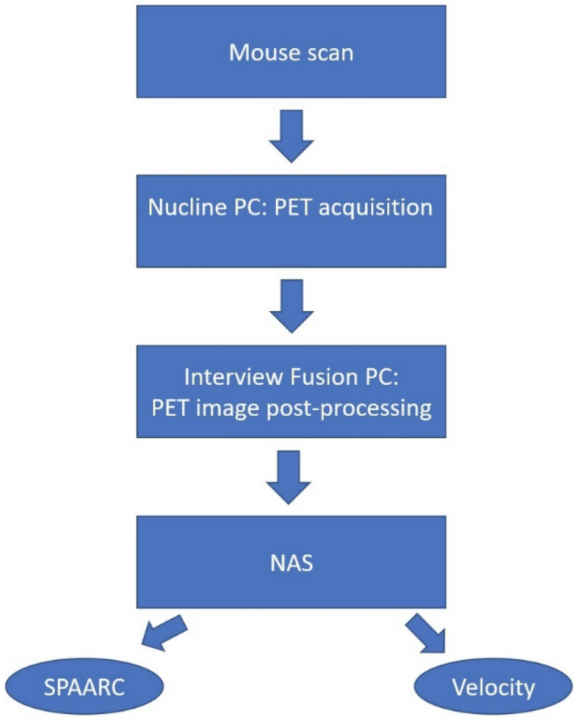

Figure 2.3: Data transfer workflow.

2.2.5 Impact of post injection imaging time

For this study, each mouse was scanned for 20 minutes between 50 and 70 minutes post injection. All images were rebinned into 4×5 (55, 60, 65 and 70) minutes PET scans using InterView™ FUSION and exported to Velocity to perform the segmentation. Table 2.1 shows the time points used for each scan to generate a new data set. All regions of interest (ROI) were segmented in the first time point image and copied to all other subsequent time point images.

Table 2.1: List of different parameters used to generate new images

Parameters	Variations
Number of subsets	1, 2, 4, 6
Number of iterations	1, 3, 6, 8, 10
Post injection imaging time	50-55, 55-60, 60-65, 65-70 minutes

2.2.6 Reconstruction settings

2.2.6.1 Impact of number of OSEM subsets

Eight mice with 4T1 tumours were scanned at PETIC following the protocol described in Section 2.2.3. To evaluate the impact of varying the number of OSEM subsets upon stability of PET radiomic features, scans were reconstructed with four different OSEM subsets (1, 2, 4, 6) as shown in Table 2.1. Figure 2.4 shows coronal view for slice 60 of first mouse with four different OSEM subsets. Tumour contouring was performed manually by an expert on scan with 4 OSEM subsets and copied to all other scans.

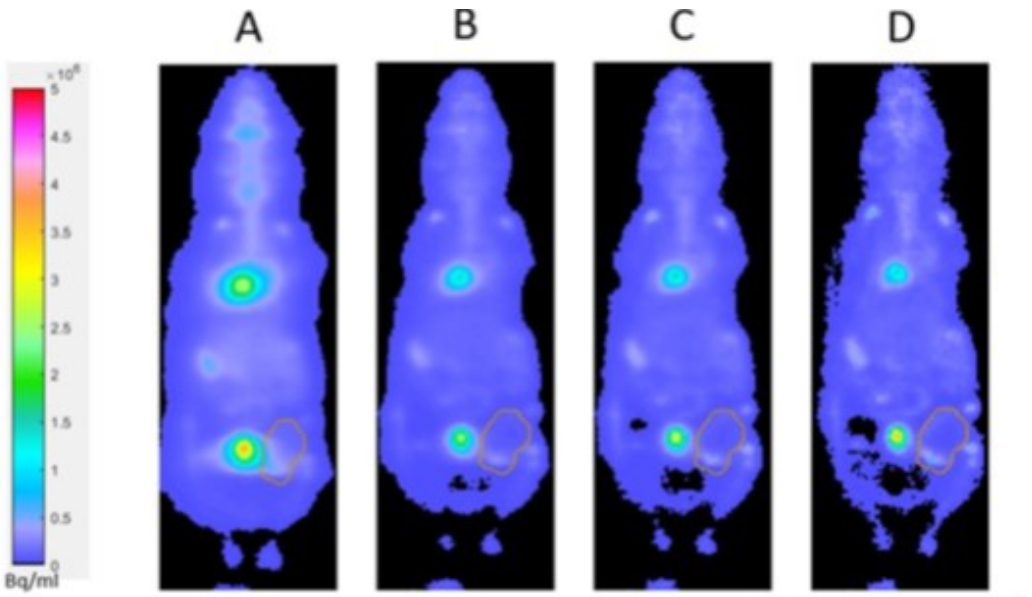

Figure 2.4: Coronal view for slice 60 of the first mouse with 1 (A), 2 (B), 4 (C) and 6 (D) OSEM subsets.

2.2.6.2 Impact of number of EM reconstruction iterations

In order to assess the impact of the number of EM reconstruction iterations upon stability of 18F-FDG PET radiomic features, scans of mice described in section 2.2.3 were reconstructed with five different number of EM reconstruction iterations (1, 3, 6, 8, 10). Contours from previous study (impact of number of OSEM subsets) were copied to all new scans (with different number of iterations).

2.2.7 Feature extraction and data analysis

All scans were imported into SPAARC [74] to extract 78 3D-radiomics features for each region of interest. Features were mainly derived from grey level co-occurrence matrix (GLCM), grey-level run length matrix (GLRLM), grey-level size zone matrix (GLSZM), grey-level distance zone matrix (GLDZM) and neighbourhood grey-tone difference matrix (NGTDM). All radiomic features in SPAARC were compliant with the Image Biomarker Standardisation Initiative (IBSI) [75]. Table 2.2 shows all the extracted features. Throughout this project, we utilised MATLAB and Statistics Toolbox Release R2019a, The MathWorks, Inc., Natick, Massachusetts, United States and the statistical analysis software R version 3.5.3 2019, R Foundation for Statistical Computing, Vienna, Austria.

Table 2.2: List of extracted radiomic features.

Features Group	Features	Features Group	Features
GLCM	Joint maximum	**GLSZM**	Small zone emphasis
	Joint average		Large zone emphasis
	Joint variance		Low grey level zone emphasis
	Joint entropy		High grey level zone emphasis
	Difference average		Small zone low grey level emphasis
	Difference variance		Small zone high grey level emphasis
	Difference entropy		Large zone low grey level emphasis
	Sum average		Large zone high grey level emphasis
	Sum variance		Grey level non-uniformity
	Sum entropy		Grey level non-uniformity normalised
	Angular second moment		Zone size nonuniformity
	Contrast		Zone size non-uniformity normalised
	Dissimilarity		Zone percentage
	Inverse difference		Grey level variance
	Inverse difference normalised		Zone size variance
	Inverse difference moment		Zone size entropy
	Inverse difference moment normalised		
	Inverse variance		
	Correlation		
	Autocorrelation		
	Cluster tendency		
	Cluster shade		
	Cluster prominence		
	First measure of information correlation		
	Second measure of information correlation		
GLRLM	Short runs emphasis	**GLDZM**	Small distance emphasis
	Long runs emphasis		Large distance emphasis
	Low grey level run emphasis		Low grey level zone emphasis
	High grey level run emphasis		High grey level zone emphasis
	Short run low grey level emphasis		Small distance low grey level emphasis
	Short run high grey level emphasis		Small distance high grey level emphasis
	Long run low grey level emphasis		Large distance low grey level emphasis
	Long run high grey level emphasis		Large distance high grey level emphasis
	Grey level nonuniformity		Grey level non-uniformity
	Grey level non-uniformity normalised		Grey level non-uniformity normalised
	Run length non-uniformity		Zone distance non-uniformity
	Run length non-uniformity normalised		Zone distance non-uniformity normalised
	Run percentage		Zone percentage
	Grey level variance		Grey level variance
	Run length variance		Zone distance variance
	Run entropy		Zone distance entropy
NGTDM	Coarseness		
	Contrast		
	Busyness		
	Complexity		
	Strength		

To evaluate feature stability when extracted at different reconstruction settings, coefficient of variation (COV) was calculated for each feature. COV is a practical tool in evaluating the variability of feature measurements and it is one of the most widely used tools in the literature on the stability of radiomic features [55, 48, 76]. COV is the ratio of the standard deviation to the mean and it can be expressed as the following equation (2.2):

$$COV = \frac{\text{Standard Deviation}}{\text{Mean}} \times 100 \tag{2.2}$$

In this study, we categorised features based on their COV values into four groups including stable (COV $\leq$ 5%), moderately stable (5% $<$ COV $\leq$ 10%), poorly stable (10% $<$ COV $\leq$ 20%) and unstable (COV $>$ 20%). The categorisation approach taken in this study is based on Yan *et al.* [55] and Shiri *et al.* [48].

2.3 Results

2.3.1 Impact of post injection imaging time

The results describing the effect of acquisition time on PET radiomic features are shown in the bar chart illustrated in Figure 2.5. Overall, 64 (82%) features showed high (COV $\leq$ 5%), 11 (14%) have moderate (5% $>$ COV $\leq$ 10%) and 2 (2.5%) features have poor (10% $>$ COV $\leq$ 20%) stability against acquisition time (post

injection imaging time). Only 1 feature (GLSZM - large zone low grey level emphasis) were found to be unstable (COV $>$ 20%) when acquisition time changed.

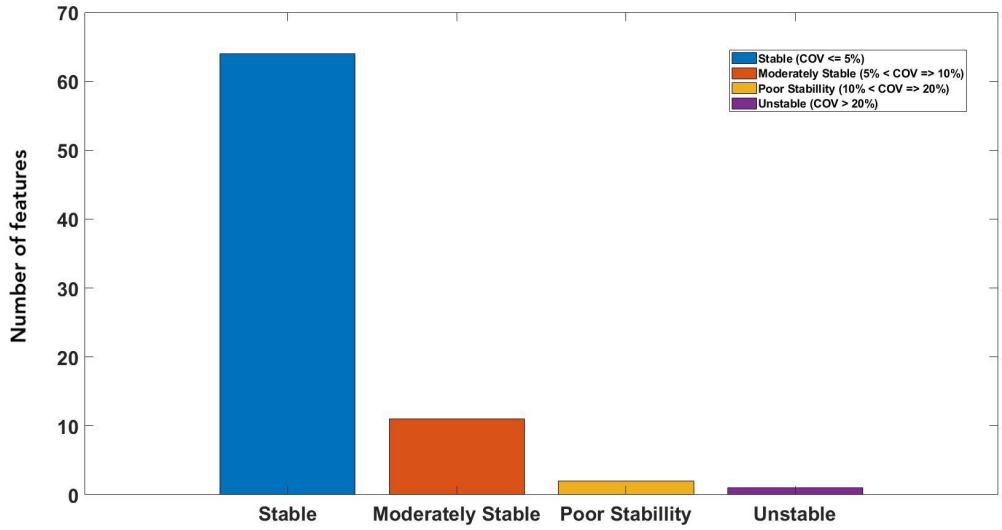

Figure 2.5: Bar chart displaying number of each categories for the stability of extracted feature against post injection imaging time (55, 60, 65 and 70 minutes). Features were classified based on their COV values.

Our results (Table 2.3) showed 22 GLCM features, 15 GLRLM features, 10 GLSZM features, 13 GLDZM features and 4 NGTDM features were found to have high (COV $\leq$ 5%) stability against the acquisition time. No features derived from GLCM, GLRLM, GLDZM and NGTDM showed high variation (COV $>$ 20%). Table 2.3 also depicts number of features that have medium (5% $>$ COV $\leq$ 10%) stability over different acquisition time (features were extracted from each of four time points images). Such features included 3 GLCM features, 1 GLRLM features, 3 GLSZM features, 3 GLDZM features and only one NGTDM feature.

Table 2.3: Total number of features for each category showing the stability level against post injection imaging time

	GLCM	GLRL	GLSZM	GLDZM	NGTDM
Stable	22	15	10	13	4
Moderately stable	3	1	3	3	1
Poor Stability	0	0	2	0	0
Unstable	0	0	1	0	0

2.3.2 Impact of number of OSEM subsets

The results describing the effect of the number of OSEM subsets on pre-clinical PET radiomic features are shown in the bar chart illustrated in Figure 2.6. Overall, only 7% (6) of all features showed high (COV $\leq$ 5%) stability against the number of OSEM subsets. 19 (27%) and 15 (19%) features were found to be moderately (5% $>$ COV $\leq$ 10%) and poorly (10% $>$ COV $\leq$ 20%) stable, respectively. Our result (Table 2.4) showed 4, 1, and 1 features derived from GLCM, GLRLM and GLSZM showed low (COV $\leq$ 5) variation when varying the number of OSEM subsets.

There were no GLDZM and NGTDM features that were highly stable as such. Seven of GLCM, 6 GLRLM, 9 GLSZM, 11 GLDZM and 5 NGTDM features were found to have high (COV $>$ 20%) variation against the number of OSEM subsets.

Table 2.4 also depicts number of features that have medium (5% $>$ COV $\leq$ 10%) stability over different number of OSEM subsets (features were extracted from

each of four OSEM subsets images). Such features included 8 GLCM, 6 GLRLM, 3 GLSZM, 2 GLDZM. There were no NGTDM features that showed medium stability as such.

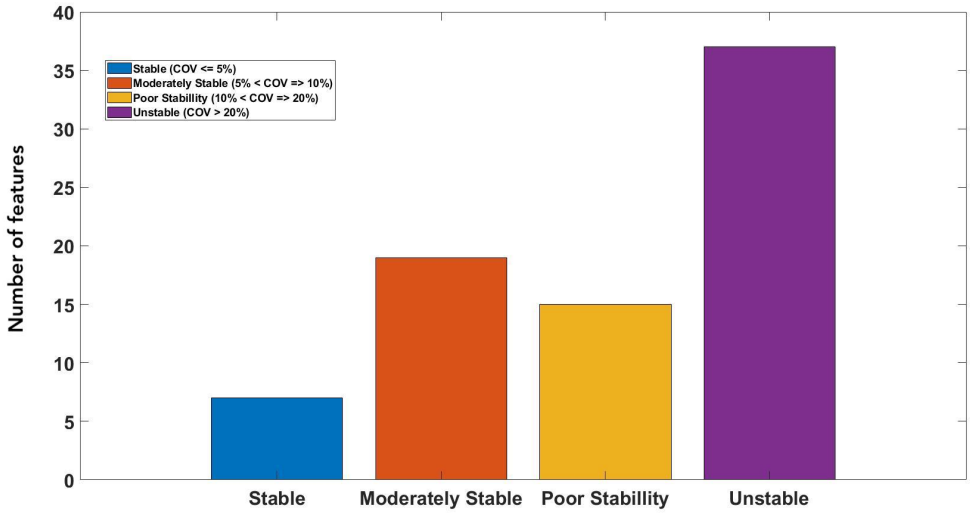

Figure 2.6: Bar chart displaying number of each categories for the stability of extracted feature against number of subsets (1, 2, 4 and 6 OSEM subsets). Features were classified based on their COV values.

2.3.3 Impact of number of EM reconstruction iterations

Figure 2.7 presents the effect of the number of EM reconstruction iterations on pre-clinical PET images radiomic features. Overall, 11 features showed low (COV $\leq$ 5%), 9 features have moderate (5% $>$ COV $\leq$ 10%) and 23 features have poor (10% $>$ COV $\leq$ 20%) stability against the number of EM reconstruction iterations. More than 35 (44%) features were found to be unstable (COV $>$ 20%) when the number of EM reconstruction iterations changed.

Table 2.4: Total number of features for each category showing the stability level against the number of OSEM subsets.

	GLCM	GLRL	GLSZM	GLDZM	NGTDM
Stable	4	1	1	0	0
Moderately stable	8	6	3	2	0
Poor Stability	6	3	3	3	0
Unstable	7	6	9	11	5

Our result (Table 2.5) showed 5 GLCM features, 5 GLRLM features, 1 GLSZM feature (zone size entropy) and 1 GLDZM feature (zone distance entropy) were found to have high (COV $\leq$ 5%) stability against the number of EM reconstruction iterations. In addition, there were no NGTDM features that were highly stable as such.

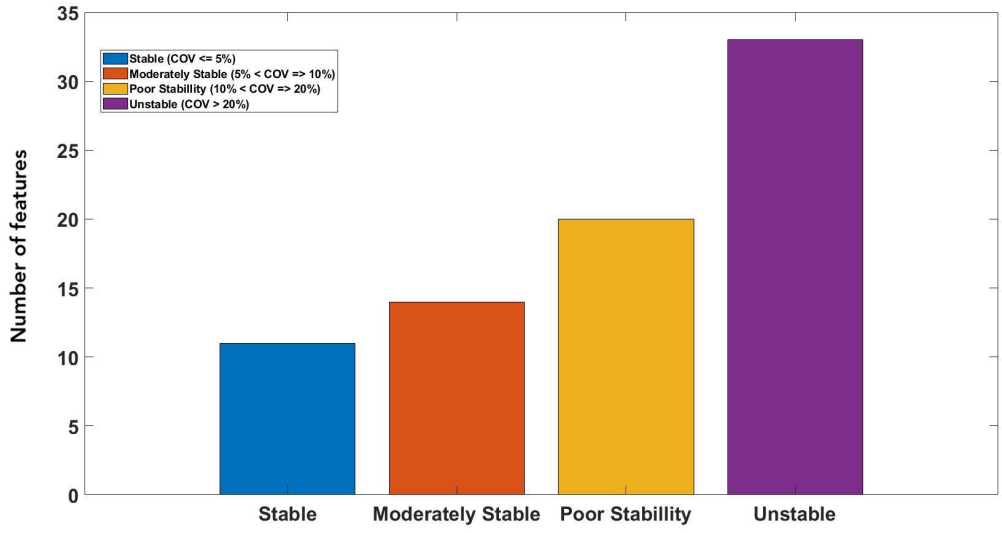

Figure 2.7: Bar chart displaying number of each categories for the stability of extracted feature against number of iterations (1, 3, 6, 8, and 10 EM reconstruction iterations). Features were classified based on their COV values.

Furthermore, Table 2.5 depicts features that have medium (5% > COV ≤ 10%) stability over different number of EM reconstruction iterations. Such features included 5 GLCM features, 2 GLSZM features (small Zone emphasis and grey level non uniformity normalised) and 2 GLDZM features (grey level non Uniformity normalised and zone distance Variance). There were no GLRLM and NGTDM features that showed medium stability as such. 6 features derived from GLCM, 7 features of GLRLM, 9 features of GLSZM, 7 features of GLDZM and all (5) of NGTDM features of were found to be nonstable (COV > 20%) against the number of EM reconstruction iterations.

Table 2.5: Total number of features for each category showing the stability level against the number of EM reconstruction iterations

	GLCM	**GLRL**	**GLSZM**	**GLDZM**	**NGTDM**
Stable	5	5	1	1	0
Moderately stable	5	0	2	2	0
Poor Stability	9	4	4	6	0
Unstable	6	7	9	7	5

2.4 Conclusion

The current research was specifically designed to assess how different reconstruction parameters such as number of subsets and iterations, would cause a variation in the PET radiomic feature stability.

The findings of this study showed that different reconstruction settings have different impact on different features. Stable features (COV $\leq 5\%$) can be categorised as excellent candidates for reproducible tumour assessment. Radiomic features sensitive to changes (COV $> 20\%$) in reconstruction settings should be omitted. A further discussion is provided in Section 8.1.

Chapter 3

The Statistical Influence of Imaging Time and Segmentation Volume on PET Radiomic Features

3.1 Introduction

As mentioned in the literature review, medical images contain more information than what can be assessed visually [20, 62]. As a result, there is an increasing interest in artificial intelligence (AI) for automated image analysis [77], [78] and decision making [79]. Within this paradigm, radiomics may be utilised to build predictive models to assess treatment outcomes.

To increase confidence in the utilisation of texture features as imaging biomarkers, it is necessary to understand the extent these confounding factors have on radiomics analysis. Although extensive research has been carried out on assessing dif-

ferent parameters on PET radiomic features, no single study exists which adequately investigated and compared the impact of two variables (segmentation volume and acquisition time) on the stability or robustness of PET radiomic features. Therefore, this chapter describes the methods used in this evaluation of the impact of two parameters on 18F-FDG PET radiomic features, especially segmentation volume and acquisition time.

To fully assess the influence of the aforementioned confounding variables, the determinant of the correlation matrices for each feature whilst varying segmentation contour sizes and image acquisition times was calculated. The determinant of the correlation matrix is related to the volume of the space occupied by the swarm of the standardised data points (mean = 0, standard deviation =1). When the measures are uncorrelated, this space is a sphere with a volume of 1. When the measures are correlated, the space occupied becomes an ellipsoid whose volume is less than 1. Thus, as the volume approaches 0 the variables are more correlated. If the determinant of the correlation matrix is equal to one this indicates the columns of the correlation matrix are orthonormal. If the determinant of the correlation matrix is equal to zero, this indicates the existence of exact linear dependence amongst the variables with the possibility for predictive analysis using a linear model.

3.2 Materials and methods

3.2.1 Data

The research data in this study is drawn from eight mice with 4T1 tumours with images described in Section 2.2.3. Mice were imaged 50 minutes post injection for a duration of 20 minutes. Images were re-binned into four PET acquisitions each with a 5-minute duration corresponding to 50-55, 55-60, 60-65 and 65-70 minutes post injection. The first time point (50-55 minutes) was used for defining the segmentation contours. Four different systematic 3D-Contour sizes (4, 4.5, 5, 5.5 mm) were generated using the segmentation tool of the Velocity 3.2.1 software (Varian Medical Systems, Palo Alto, CA). Figures 3.1 and 3.2 show the coronal and sagittal images with four contours on the lower flank of the first mouse. Contours defined on the first time point were used on all other images which were re-binned into subsequent time points.

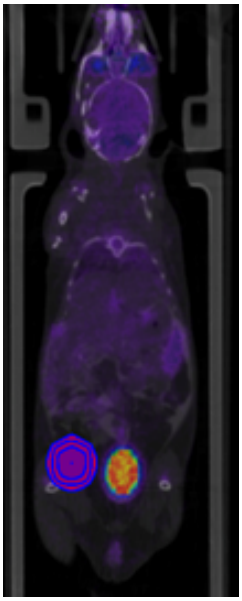

Figure 3.1: Coronal slice of lower right flank (left of image) with four different contours for the first mouse.

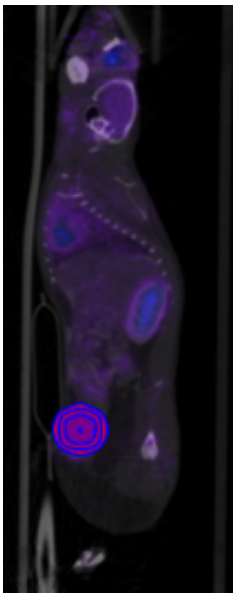

Figure 3.2: Sagital slice of lower right flank (left of image) with four different contours for the first mouse.

3.2.2 Feature extraction and data analysis

SPAARC was used to extract 78 3D-radiomic features (As presented in table 2.2) for each of the four segmented volumes for each of the four time points [74]. Features including grey level co-occurrence matrix (GLCM), grey-level run-length matrix (GLRLM), grey-level size zone matrix (GLSZM), grey-level distance zone matrix (GLDZM) and neighbourhood grey-tone difference matrix (NGTDM) were extracted. All radiomic features were compliant with the IBSI [80].

The correlation between two variables (e.g., n and p) is defined using Equations 1 and 2 (for contour size and imaging time respectively) [81]. In our study, n and p refer to first, second, third or fourth contour size/imaging time point. To fully assess the statistical strength of correlation for each texture feature versus contour size four correlation matrices were constructed at each time point (T1 to T4) (C-T1, C-T2, C-T3, C-T4) using all contour sizes (C1 to C4).

$$C_{n,p} = \frac{\sum_{i=1}^{m=8}(C_n - \overline{C_n})(C_p - \overline{C_p})}{(\sigma_{C_n})\ (\sigma_{C_p})\ (m-1)} \tag{3.1}$$

Where,

- $C_{n,p}$: The correlation coefficient for n^{th} contour size and p^{th} contour size.
- $\overline{C_{n/p}}$: The mean of all features value for $(n/p)^{th}$ contour size.

- $C_{(n/p)}$: Feature value for $(n/p)^{th}$ contour size.

- $\sigma_{C_{n/p}}$: The standard deviation of all the features values for $(n/p)^{th}$ contour size.

Similarly, this was repeated to assess the correlation of texture features versus acquisition time (T-C1, T-C2, T-C3, T-C4) giving four correlation matrices where time is the variable under investigation. The correlation of 78 features with varying time and contour size was investigated resulting in 8 correlation matrices for each feature.

$$T_{n,p} = \frac{\sum_{i=1}^{m=8}(T_n - \bar{T_n})(T_p - \bar{T_p})}{(\sigma_{T_n})\ (\sigma_{T_p})\ (m-1)} \tag{3.2}$$

Where,

- $T_{n,p}$: The correlation coefficient for n^{th} acquisition time and p^{th} acquisition time.

- $T_{(n/p)}$: Feature value for $(n/p)^{th}$ acquisition time.

- $\bar{T_{n/p}}$: The mean of all features value for $(n/p)^{th}$ acquisition time.

- $\sigma_{T_{n/p}}$: The standard deviation of all the features values for $(n/p)^{th}$ acquisition time.

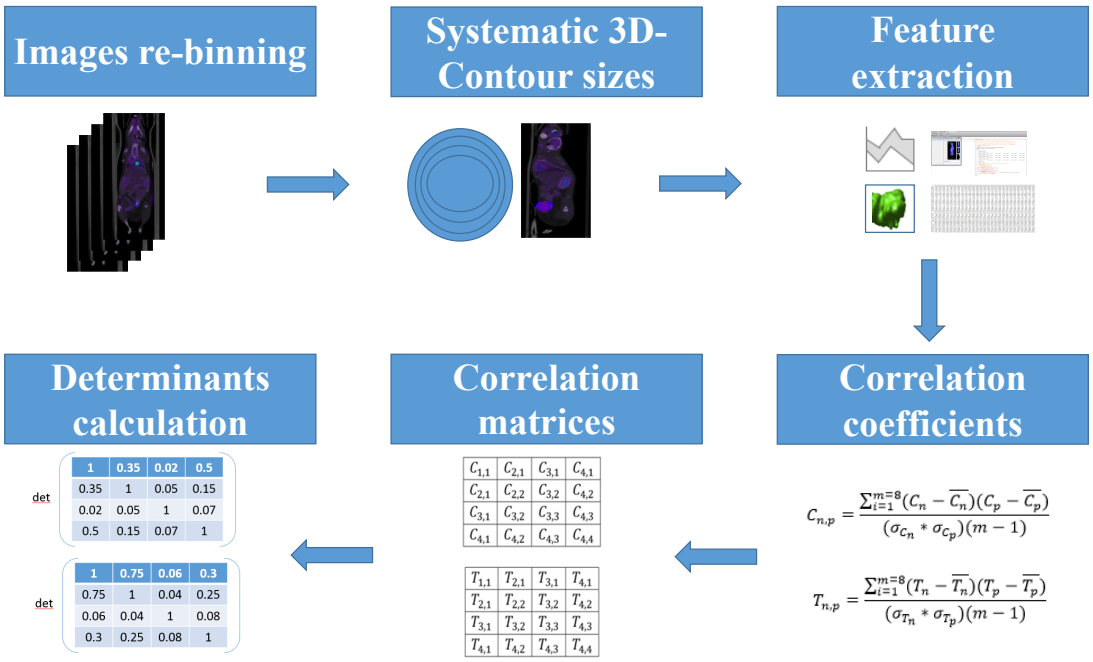

Figure 3.3: Workflow for evaluating the statistical association of texture feature values versus different contour size and time points. Following acquisition images are re-binned into time points 50-55, 55-60, 60-65 and 65-70 minutes post injection. Each image has a 5 minute acquisition period. Four 3D contours were extracted at the 55 minute baseline scan and propagated to subsequent time point images. 78 texture features were extracted for each image. For each feature 8 correlation matrices were constructed considering the correlation of a contour size and image time. The determinant of those matrices allows the robustness of each texture feature to contour size and acquisition time to be determined.

For each feature, the resulting matrices for examination of robustness of texture features with varying contour sizes and imaging acquisition time are presented in Figure 3.4. The determinant of the correlation matrix allows investigation into the linear dependence (or lack thereof) of the variable of interest (acquisition time or contour size) on the texture feature of interest.

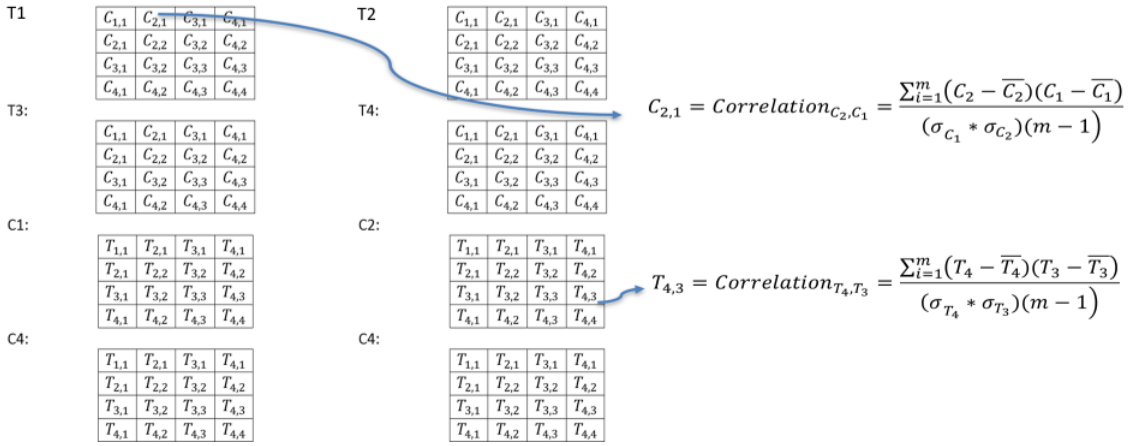

Figure 3.4: An illustrative example of resulting correlation matrices. At each time point (T1 to T4), a correlation matrix was constructed and this matrix contains the correlation coefficients for all contour sizes (C1 to C4). As an example, in the first matrix (at T1), $C_{2,1}$ represents the correlation coefficients for the first and second contour size. Similarly, this was repeated for probing the correlation of texture features versus acquisition time. At each contour size (C1 to C4), a correlation matrix was constructed and this matrix contains the correlation coefficients for all time points (T1 to T4). As an example, in the sixth matrix (at C2), $T_{4,3}$ represents the correlation coefficient for the fourth and third time point.

3.3 Results

Overall, as shown in Figure 3.5, the determinant of correlation when contour size is considered as the variable of interest is much smaller than the determinant of correlation when imaging time is considered as the variable of interest. This is consistent when varying acquisition time for the majority of texture features. The mean determinant of correlation across all texture parameters when considering varying contour size is 0.02378.

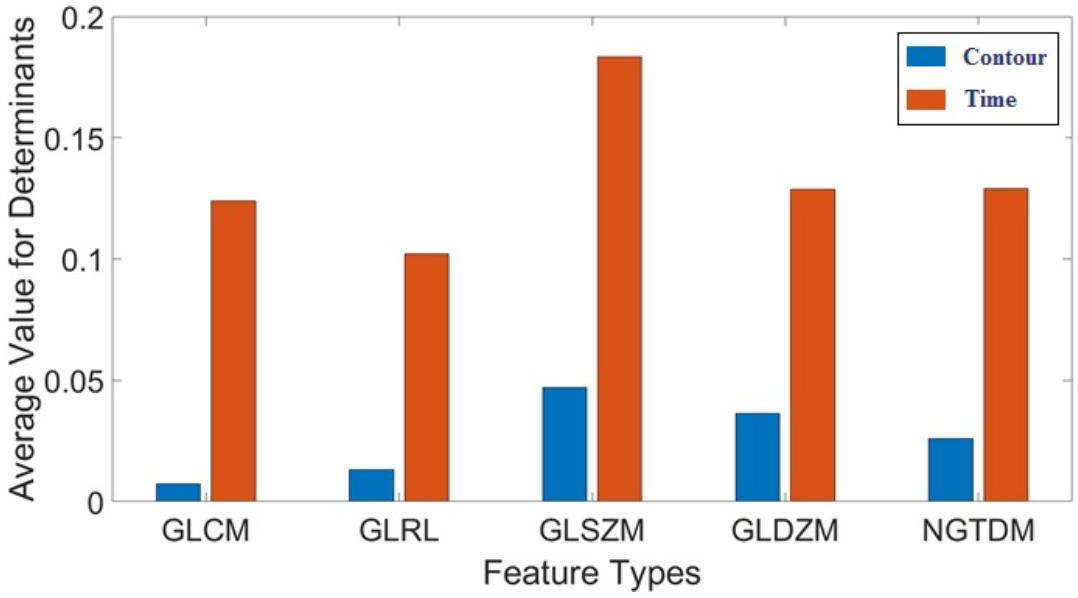

Figure 3.5: Bar chart displaying the mean determinant of correlation for different features types whilst varying acquisition time and contour size.

Features including autocorrelation, information correlation 1, information correlation 2 (GLCM) and coarseness (NGTDM) were found to have 0 value as a mean determinant of correlation when considering varying acquisition time. In addition, there were no GLRLM, GLSZM and GLDZM features that have 0 value as a mean determinant of correlation for both contour and acquisition time.

Table 3.1: Features with highest value of mean determinant of correlation for acquisition time and contour size for each feature families.

	Contour Size - highest	**Acquisition Time - highest**
GLCM	ClusterShade	DifferenceVariance
GLRL	GreyLevelNonUniformity	ShortRunHighGreyLevelEmphasis
GLSZM	ZoneSizeNonUniNormalised	SmallZoneHighGreyLevelEmphasis
GLDZM	ZoneDistanceNonUniNormalised	GreyLevelVariance
NGTDM	Busyness	Contrast

Table 3.1 and 3.2 show the radiomic features with highest and lowest values of mean determinant of correlation for acquisition time and contour size for each feature families. Our result showed that 30 features, including; Joint entropy, Sum entropy, Dissimilarity (GLCM), Short runs emphasis (GLRLM), Strength (GNGTDM) have an increasing determinant of correlation with acquisition time as the contour size increased. Thirteen features including Run entropy (GLRLM) and Sum variance (GLCM), demonstrated a decrease in the determinant of correlation with time when the region of interest volume increased.

Table 3.2: Features with lowest value of mean determinant of correlation for acquisition time and contour size for each feature families.

	Contour Size - lowest	**Acquisition Time - lowest**
GLCM	Information correlation 1	AutoCorrelation
GLRL	LongRunLowGreyLevelEmphasis	LongRunLowGreyLevelEmphasis
GLSZM	LargeZoneLowGreyLevelEmphasis	LargeZoneLowGreyLevelEmphasis
GLDZM	ZoneDistanceNonUni	LargeDistanceEmphasis
NGTDM	Coarseness	Coarseness

The mean determinant of correlation across all GLCM texture parameters when considering varying contour size and acquisition time are 0.00733 and 0.12404, respectively. GLSZM features showed to have the highest mean determinants of correlation at 0.18331 for acquisition time and 0.04688 for contour size.

These findings suggest that more linear dependence of texture features with increasing contour size than increasing imaging time. Less correlation and hence linear dependence is observed when acquisition time is considered as the variable of interest with a mean determinant of correlation across all texture parameters of 0.13296. Increasing contour size however has a variable effect on this metric of correlation.

3.4 Conclusions

The numerous radiomic features proposed may vary with confounding variables such as the time post injection of image acquisition and the accuracy of the delineation of the prescribed segmentation volume. To this avail, we proposed using the determinant of the correlation matrix to analyze radiomic features robustness to confounding variables. For this purpose, dynamic pre-clinical PET images of 8 mice with mammary carcinoma xenografts (4T1) were binned into 5 minutes intervals from 50 to 70 minutes post injection. The effect of variation in segmentation was also explored by incrementally increasing segmentation volume. From each image set, we extracted 78 radiomic features for analysis. The statistical association was measured by the

determinant of the correlation matrix. We have shown that texture features are more correlated with varying segmentation volume than acquisition time (mean determinant of correlation for segmentation volume = 0.02378 vs. mean determinant of correlation for acquisition time = 0.13296) and hence less robust to varying acquisition time. In addition, this work demonstrated that the linear dependence / correlation of texture features with acquisition time is affected by contour size. A linear dependence may be enough to determine absolute margins of error on individual texture features with varying segmentation accuracy which provides scope for future work. Acquisition time however has a variable influence on this correlation and judicious choice of segmentation is required to minimize its effect. An implication of these findings is that both the impact of acquisition time and segmentation volume should be taken into account before radiomics analysis can be applied clinically. This will also have an impact in multi centre studies using radiomics analysis where both image time and contouring will need to be standardised across each centre in order for radiomics analysis to be possible.

Chapter 4

A Self Organizing Map for Exploratory Analysis of PET Radiomic Features

4.1 Introduction

Applying artificial intelligence in the field of radiomics may be an exciting opportunity to advance our knowledge of the usefulness of texture analysis. To date only a limited number of studies have utilised advanced machine learning techniques to investigate the value of texture analysis in PET imaging [44, 82].

Both supervised and unsupervised learning are the two approaches of machine learning. In supervised learning, both input (features) and output (class labels) are required to train the model whereas in unsupervised learning only input data are needed to determine the patterns from the data on its own. Supervised learning can

be used for classification (sorts data into specific categories using a labelled dataset) and unsupervised learning can be used for clustering (partitioning an unlabelled dataset into groups of similar objects). A self-organising feature map (SOM) is a type of artificial neural network (ANN) that is trained using unsupervised learning to produce a lower dimensional representation of the input data on an underlying manifold. Manifold learning has demonstrated its effectiveness in other areas of image processing, such as respiratory motion correction [78] and image segmentation [83, 77].

Although extensive research has been carried out on the impact of different parameters on radiomics, no single study exists which applied SOMs to examine, detect and capture the statistical variability of medical image texture parameters. As shown in Figure 4.1, SOM takes a set of input data, for example, L texture parameters and maps them onto a two dimensional grid of neurons [84, 85]. Each neuron in the grid is assigned an initial weight vector $M = (m_{g1}, m_{g2}, .., m_{gD})$ with the same dimensionality as that of the input data g=(1,2,...L). The training uses competitive learning (nodes compete to respond to input data). Training examples are fed into the network at random, the Euclidean distance to all weight vectors is calculated (Equation 4.1) and is used to update the neuron weights; the neuron

whose vector is similar to the input is known as the best matching unit (BMU).

$$m_{t+1}^{gd} = m_t^{gd} + \eta h(g,q)(x_d - m_{tjd}^t), \qquad for\ 1 \leq d \leq D \tag{4.1}$$

Where,

- η : The learning rate parameter (to controls the size of weight vector).
- $h(g,q)$: The neighbourhood function which has the value 1 at the winning neuron q and decreases as the distance between g and q increases.

Each high dimensional data point is thus embedded onto a single neuron which reproduces its structure. Neuronal weights act as pointers to the input space and form a discrete approximation of the distribution of the training samples. More neurons point to regions with high training sample concentration and fewer where the samples are scarce.

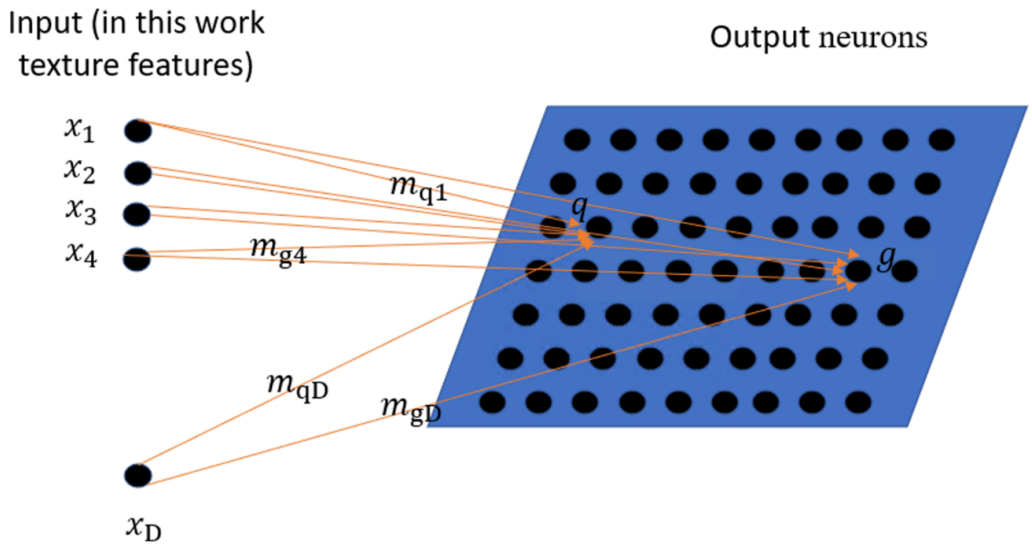

Figure 4.1: Illustrative example of a Self-Organising Map [85]. Each neuron in the grid is assigned an initial weight vector $M = (m_{g1}, m_{g2}, .., m_{gD})$ with the same dimensionality as that of the input data g=(1,2,...L). The training uses competitive learning (nodes compete to respond to input data). Training examples are fed into the network at random, the Euclidean distance to all weight vectors is calculated (Equation 4.1) and is used to update the neuron weights; the neuron whose vector is similar to the input is known as the best matching unit (BMU).

4.2 Materials and methods

4.2.1 Data

Data in this study were gathered from PET images of eight mice with 4T1 tumours described in Section 2.2.3. Mice were dynamically imaged at 50 minutes post injection for 20 minutes, and re-binned into 4 x 5 minutes PET scans (50-55, 55-60, 60-65 and 65-70 minutes post injection). Four different systematic 3D-Contour sizes (4, 4.5, 5, 5.5 mm) were segmented on the first time point (50-55 minutes) using Velocity 3.2.1 software (Varian Medical Systems, Palo Alto, CA). Contours obtained

on the first time point were overlaid on all other images which were re-binned into subsequent time points.

4.2.2 Feature extraction and data analysis

Radiomic features were extracted for each volume at each time points using SPAARC [74]. Therefore, each feature has 128 observable values that resulted from 8 mice, 4 different time points and 4 different contour sizes.

Each texture feature was standardised using Z-score (Z-omic) normalisation across the distribution of texture features acquired for inter-comparison of different textures. The mean Z-omic for 5 pre-defined groups of texture was calculated (GLCM, GLRLM, GLZM, GLDM and NGTDM) and the R software was used to learn the self-organizing map of the averaged Z-omic using:

- 16 organising neural networks.
- A learning rate of 0.05.
- Gaussian neighbourhood function with standard deviation 1.

4.3 Results

From the input dataset each instance of measured texture (Z-omic) is assigned to a single node that best represents its distribution of variability. The standard Kohonen SOM plot produce pie representations of the representative vectors for the grid which are known as codebooks. Thus, the classification of each node and the relative contribution of the grouped features is demonstrated in the codes plot (as shown in Figure 4.2).

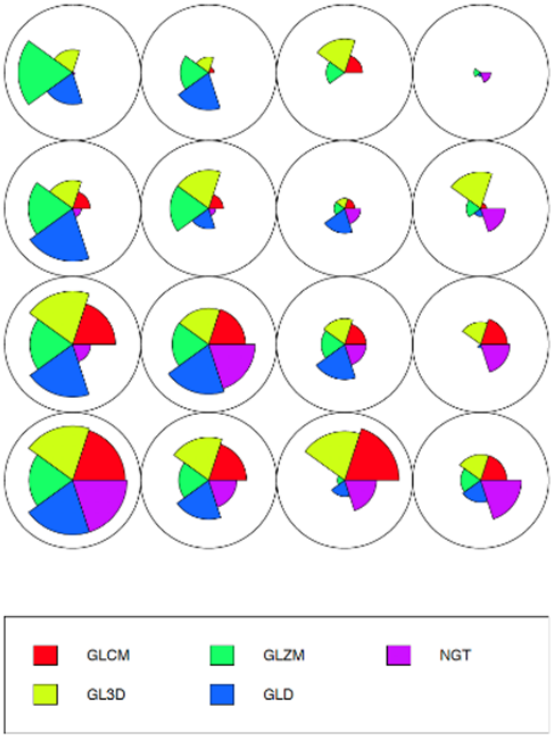

Figure 4.2: Codes plots (pie representations of the representative vectors for the grid) for the texture features. These codes plots represent the clusters (node wieght vectors associated with each node).

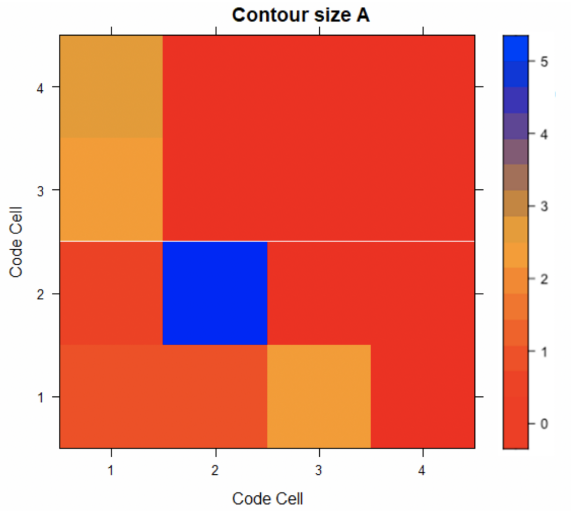

Figure 4.3: The distribution of the first contour size with respect to the self-organized features. This heat map demonstrates the distribution or frequency of the first contour size over the code cells (nodes) of the SOM.

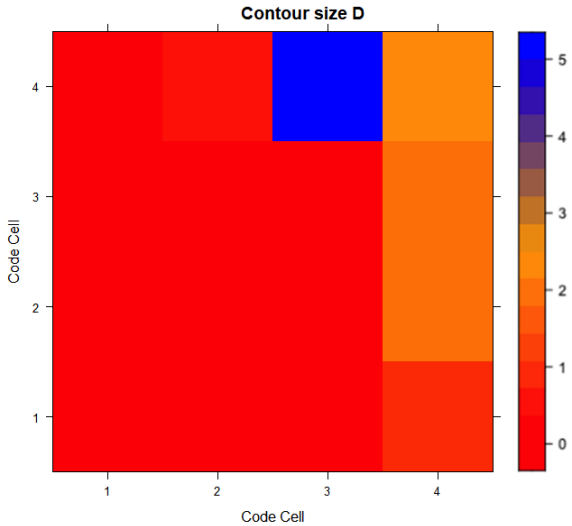

Figure 4.4: The distribution of the fourth contour size with respect to the self-organised features. This heat map demonstrates the distribution or frequency of the fourth contour size over the code cells (nodes) of the SOM.

A heatmap can be visualised to assess the impact of predefined confounding variables. The heat map presents the distribution or frequency of the confounding variables over the nodes of the SOM. The SOM is blind to the suspected confound-

ing variables (Contour Size, Imaging Time), so any visible clusters in the heatmap demonstrate an association between the confounding variable and the input data. Figures 4.3 and 4.4 demonstrated the distribution of the contour size with respect to the self-organised features which examine the distribution of extremes of contour size across the codebook. Two distinct clusters can be visualized in the heatmap allowing us to interpret that contour size acts as a confounding variable when performing texture analysis and changing contour size alters the distribution of standardised (and normalised (z-score)) texture features.

Figures 4.5 and 4.6 indicated that no such relationship was observed when probing post injection imaging time as the confounding variable as no clusters were observed in the heatmap. Taken together, these results suggest that the statistical distribution of the input texture has clear modes which are dependent on contour size.

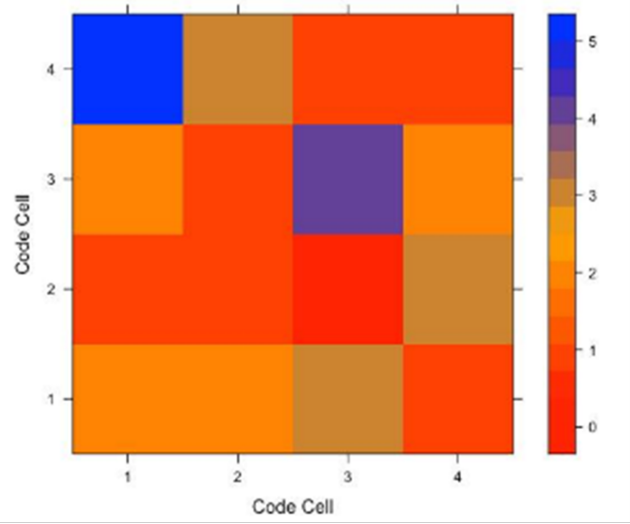

Figure 4.5: The distribution of the first time point with respect to the self-organised features. This heat map demonstrates the distribution or frequency of the first time point over the code cells (nodes) of the SOM.

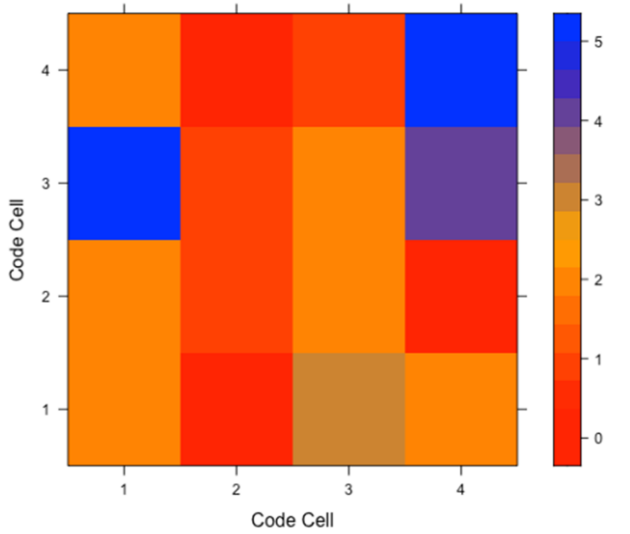

Figure 4.6: The distribution of the fourth time point with respect to the self-organised features. This heat map demonstrates the distribution or frequency of the fourth time point over the code cells (nodes) of the SOM.

4.4 Conclusions

The previous chapter, assessed the influence of two confounding variables (segmentation contour sizes and post injection imaging time) using the determinant of the correlation matrices for each feature. This chapter introduced the innovative application of a self-organising map to texture feature analysis and explore its ability in identifying emergent properties that affect radiomics variability. SOM is a type of artificial neural network (ANN) that is trained utilising unsupervised learning to generate a lower dimensional representation of the input data on an underlying manifold. In this study, the Kohonen SOM took a set of L texture parameters (input data) and mapped them onto a two dimensional grid of neurons.

One of the more significant findings to emerge from this study is that an emergent property of the statistical variability of the grouped textures is the extremes of contour size. However, more research is required to determine the efficacy of applying SOM on radiomic features analysis. For this purpose, phantom data will be involved in Chapter 7.

The SOM may be utilised with outcome data to serve as a predictive tool for dependent variables (e.g therapy response and prognosis). In so doing the learnt

representations of self-organised features serve as the attributes for prediction which will take into consideration the statistical variability in the underlying dataset.

Chapter 5

The Impact of Reconstruction Settings on PET Radiomic Features: A Homogeneous Phantom Study

5.1 Introduction

As mentioned in the literature review, there are several factors that can influence the stability of PET image radiomic features. This indicates a need to understand the various perceptions of the stability of PET image radiomic features that exist among different protocols (e.g. reconstruction settings).

In Chapter 2, the impact of reconstruction settings on PET images radiomic features were assessed in the form of preclinical model. Despite the differences in the underlying biology between humans and animals, the current study shows how different parameters can affect the output of radiomics. However, before implement-

ing radiomics in a clinical setting, more research is needed to better understand the influence of different protocols and settings on PET radiomic features.

Clinical studies are subject to extraneous variables such as biological sensitivity, organ motion and patient size. For this reason, phantom studies can be a reasonable substitute to control for bias relative to biological variability of clinical studies. Although extensive research has been carried out on phantom data, there are limited publications which analyse large numbers of features, compliant with IBSI [80] and different 18F-FDG activity. Therefore, this study was designed to determine the effect of TOF, number of OSEM subsets, number of iterations and filters on 78 PET radiomic features that are compliant with IBSI. The materials and methods used to achieve this purpose are described in the following section of this chapter.

5.2 Materials and methods

5.2.1 Phantom preparation

In this study, the NEMA (National Electrical Manufacturer's Association) [86] IQ (Image Quality) body phantom available at PETIC was used. This phantom consists of a 9.7 cm^3 tank containing six inserts with internal diameters of 10, 13, 17, 22, 28, 37mm. A picture of the used phantom is presented in figure 5.1.

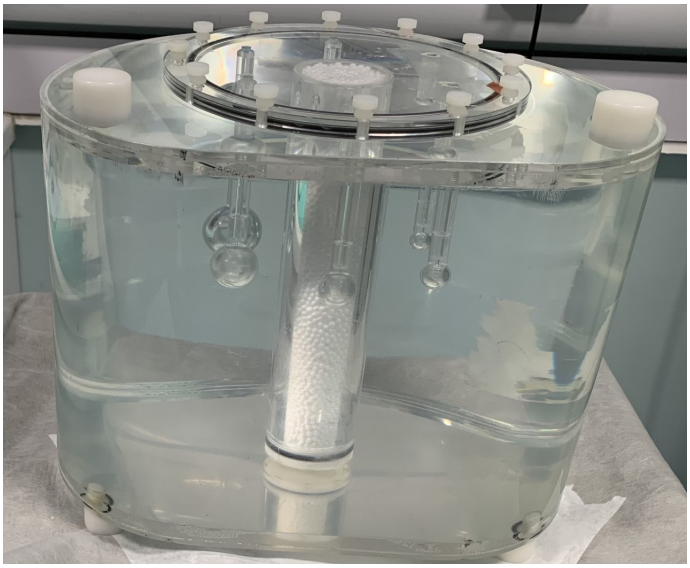

Figure 5.1: Picture of NEMA IQ phantom with unfilled inserts

The spheres were filled with different amounts of 18F-FDG activity. Two inserts were filled with with 20 kBq/ml, two with 40 kBq/ml and two with 80 kBq/ml. An illustrative layout of the filled inserts with their activity is shown in figure 5.2.

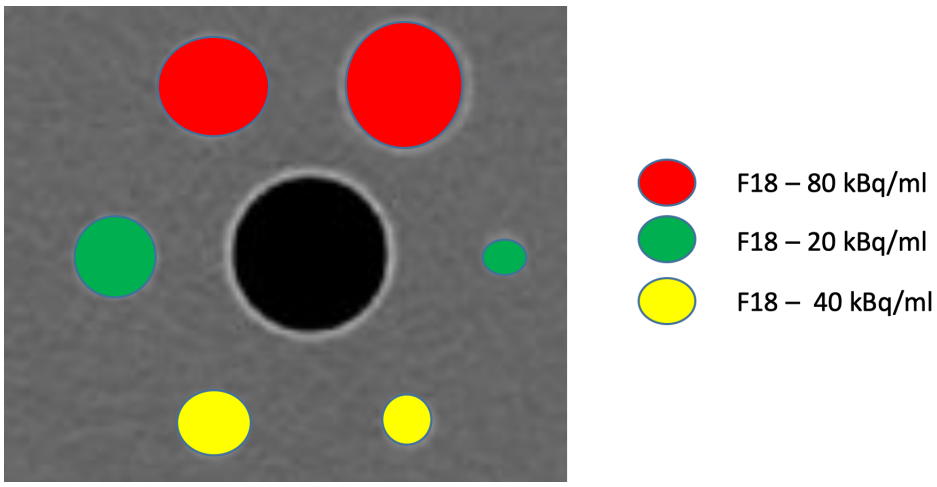

Figure 5.2: An illustrative layout of the filled inserts with 18F-FDG activity

5.2.2 Acquisitions and reconstructions

Experiments were performed using GE Discovery 690 PET/CT scanner (General Electrics Healthcare, Milwaukee, USA). All measurements were carried out with the following acquisition and reconstruction settings:

- Order subset expectation maximization (OSEM)
- 24 subsets
- 2 iterations
- 6.4 mm filter cutoff
- 256 matrix size

The phantom was imaged for 20 minutes. Figure 5.3 shows a picture of the IQ NEMA phantom placed on the scanner. This experiment was repeated three times under similar conditions.

To assess the impact of reconstruction settings on image radiomic features, all images were reconstructed with different reconstruction parameters including: with/without TOF, different numbers of subsets, iterations and full width at half

maximum (FWHM) of Gaussian filter. Table 5.1 shows the reconstruction settings used for each scan to generate its new data set.

Table 5.1: List of reconstruction settings used to generate new images. Default settings: TOF, 24 OSEM subsets, 2 iterations, 6.4 mm filter cutoff.

Reconstruction parameters	Variations
Number of subsets	12, 16, 18, 24, 32
Number of iterations	1, 2, 3, 4, 5, 6
Filter Cut-off	0, 1, 2, 3, 4, 5, 6, 7
TOF	Yes, NO

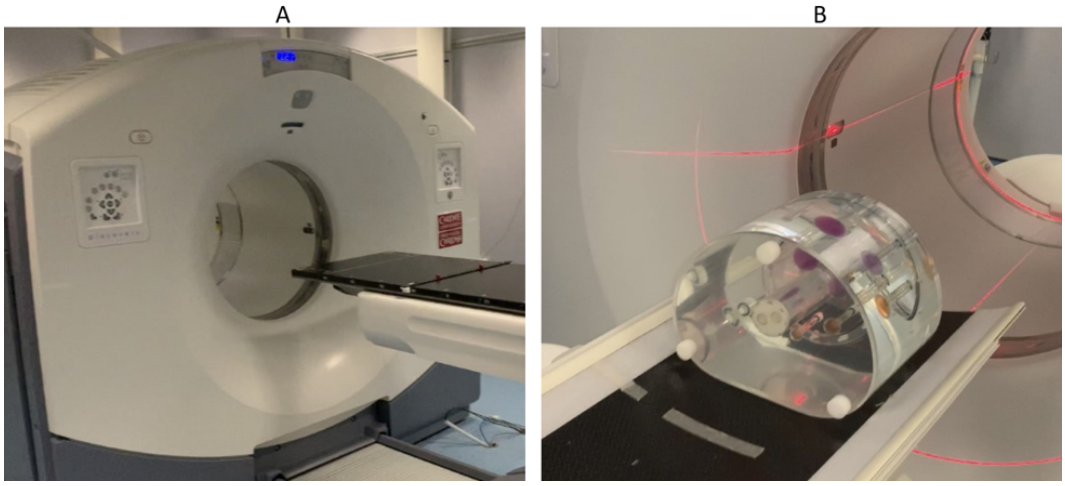

Figure 5.3: Scanner (GE Discovery 690 PET/CT) used in this study (A) and the IQ NEMA phantom after placed on the scanner (B).

5.2.3 Segmentation

For each scan, Velocity 3.2.1 software (Varian Medical Systems, Atlanta, USA) was utilized to obtain the ground truth contours for each insert from the first image (default setting) and contours were then overlaid onto other subsequent image sets produced when varying the reconstruction parameters.

5.2.4 Feature extraction and data analysis

For each ROI, SPAARC [74] was used to extract 78 3D-radiomic features. Features including a 25 grey level co-occurrence matrix (GLCM), 16 grey-level run-length matrix (GLRLM), 16 grey-level size zone matrix (GLSZM), 16 grey-level distance zone matrix (GLDZM) and 5 neighbourhood grey-tone difference matrix (NGTDM) were extracted. SPAARC radiomics analysis was standardised according to the IBSI[75]. All extracted features are listed in Table 2.2.

To evaluate feature stability when extracted at different reconstruction settings, COV was calculated for each feature over the different reconstruction settings in the same way as described in Section 2.2.7. COV is one of the most widely used tools in the literature to assess the stability of radiomic features[55, 48, 76]. Features were categorised based on their COV values into four groups including stable (COV $\leq$ 5%), moderately stable (5% $<$ COV $\leq$ 10%), poorly stable (10% $<$ COV $\leq$ 20%) and unstable (COV $>$ 20%). The categorisation approach taken in this study is based on Yan *et al.* [55] and Shiri *et al.* [48]. For each feature that demonstrated high (COV $\leq$ 5%) or moderate (5% $<$ COV $\leq$ 10%) stability across all reconstruction settings, the percentage of agreement was used to evaluate the degree of agreement between the three phantom experiments.

5.3 Results

5.3.1 Impact of TOF

Sixty two features (79%) were found to have high stability (COV $\leq$ 5%) with TOF. Such stable features included NGTDM (Complexity, Strength), GLSZM (Small zone emphasis, Zone percentage, Zone size entropy), GLDZM (Small distance emphasis, Zone distance variance), GLSZM (Zone percentage, Zone size entropy), GLRLM (short Run emphasis, run percentage) and GLCM (joint entropy, correlation). Eight features include GLDZM- long Run low grey level emphasis, GLSZM- difference Variance, GLRLM- short Run Low grey level emphasis, GLRLM- long run high grey level emphasis showed moderate stability. Five features such as GLRL - low grey level run emphasis and GLSZM - low grey level zone emphasis showed poor stability (10% $<$ COV $\leq$ 20%) against TOF. Only three features, including GLCM- cluster shade, GLSZM - small zone low grey level emphasis and GLSZM- large zone low grey emphasis) were found to be unstable (COV $>$ 20%). The plots of COV values for each feature is shown in Figure 5.4.

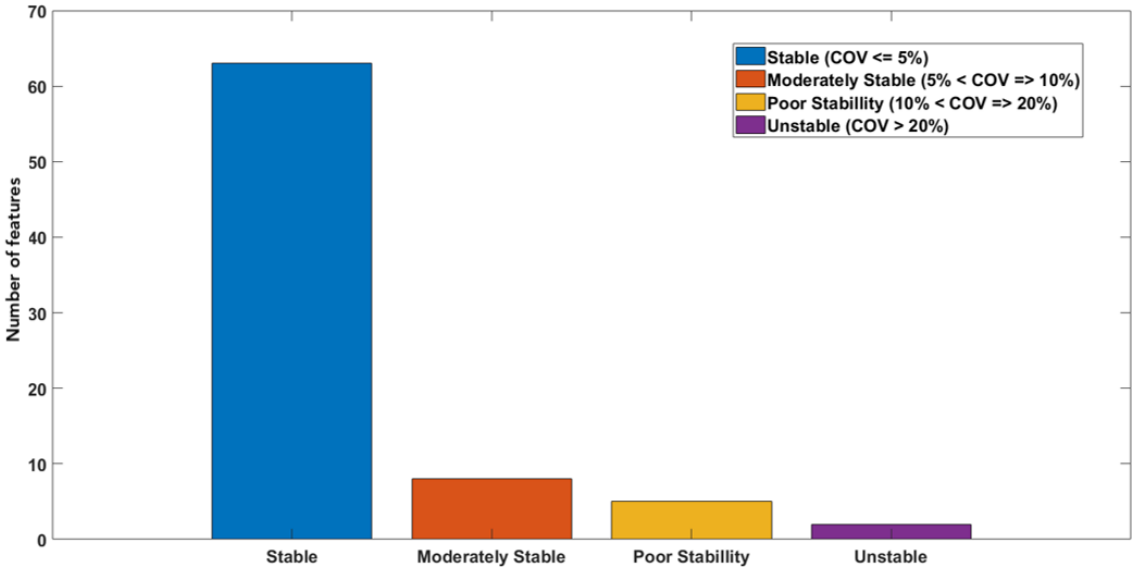

Figure 5.4: Number of features for each category of the stability of PET radiomic features against TOF. Features were classified based on their COV values.

5.3.2 Impact of number of subsets

The impact of the number of subsets on radiomic features was tested with five different numbers of subsets. Sixty two (79%) features showed high stability (COV $\leq$ 5%) with number of OSEM subsets. Eleven (14%) features were found to have moderate stability (5% $<$ COV $\leq$ 10%). Four (5%) features including GLRL (Run length variance), GLSZM (Zone size non uniformity, Small zone low grey level emphasis) and GLDZM (Large distance low grey level emphasis, Small distance low grey level emphasis) were poorly stable.

The remaining feature (GLCM - cluster shade) had high variability (unstable) at different number of subsets. All features from NGTDM were stable (COV $\leq$ 5%) against subsets. Figure 5.5 presents number of features for each category.

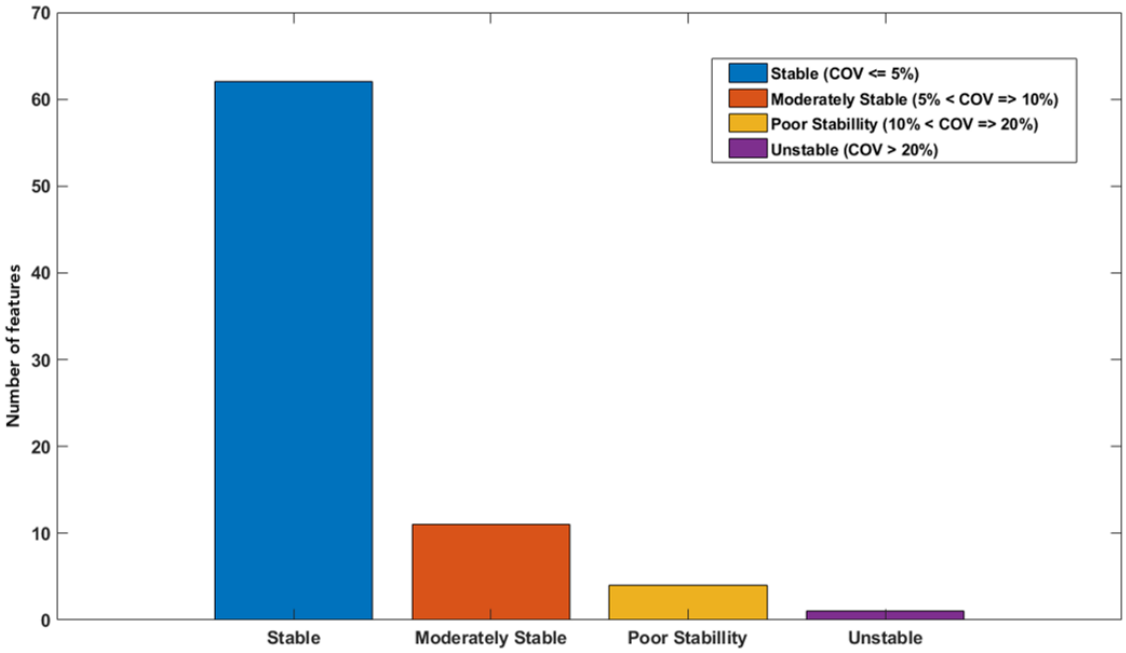

Figure 5.5: Number of features for each category of the stability of PET radiomic features against number of subsets. Features were classified based on their COV values.

5.3.3 Impact of the number of iterations

With changing number of iterations, sixty eight features showed very small variation (COV $\leq$ 5%). Stable features included GLCM (difference average, contrast, dissimilarity, joint entropy), GLRLM (short run emphasis, short run high grey level emphasis), GLSZM (small zone High grey level emphasis, zone size non uniformity), GLDZM (high grey Level zone emphasis, zone percentage) and NGTDM (Coarse-

ness, Busyness, Complexity). 13% (10) of features, were found to have moderate (5% < COV ≤ 10%) stability (Figure 5.6). No features demonstrated poor (10% < COV ≤ 20%) and high variation (COV > 20%).

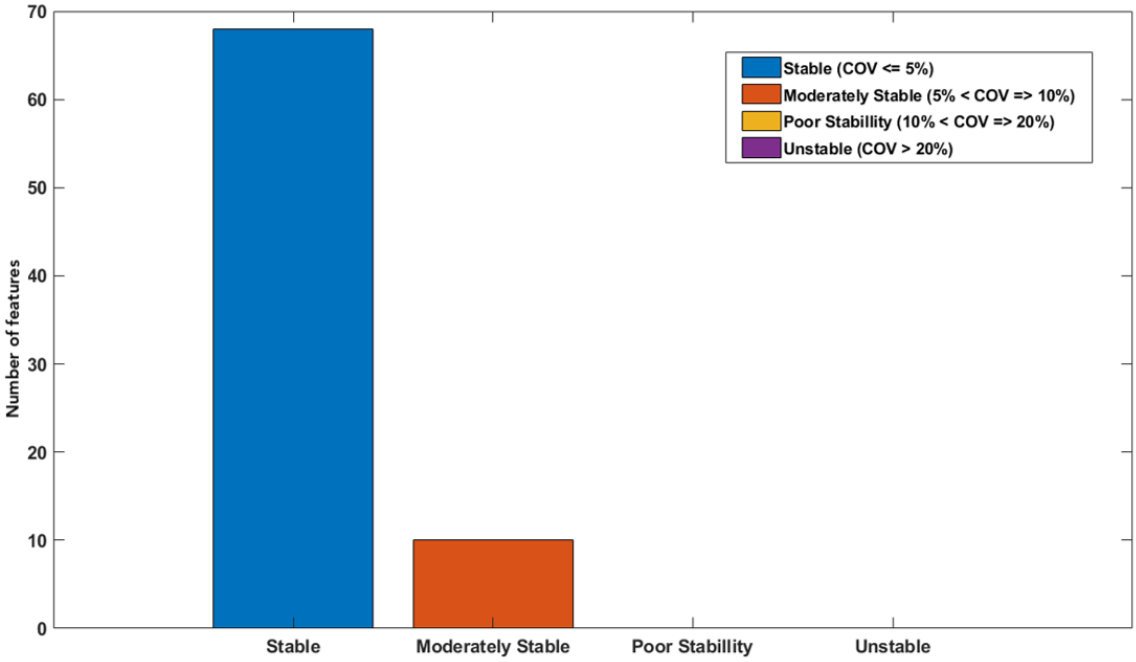

Figure 5.6: Number of features for each category of the stability of PET radiomic features against number of iteration. Features were classified based on their COV values.

5.3.4 Impact of FWHM of the Gaussian filter

More than 66% (52) of features were found to be stable when varying the FWHM of the Gaussian filter. Features with high stability included GLCM (sum average, sum variance, dissimilarity, contrast, inverse difference, inverse difference normalised), GLRLM (run percentage, grey level Variance, run entropy), GLSZM (Small zone

emphasis, High grey level zone emphasis , Small zone high grey level emphasis), GLDZM (Small distance high grey level emphasis, Grey level non-uniformity normalised) and NGTDM (coarseness, contrast, strength).

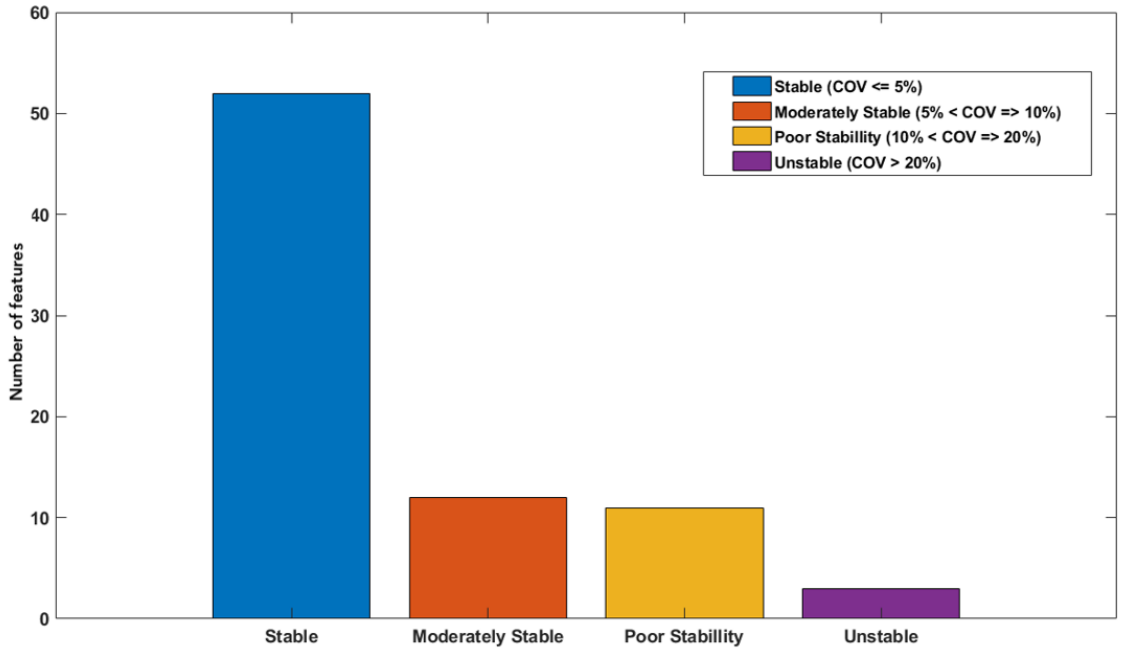

Figure 5.7: Number of features for each category of the stability of PET radiomic features against filters. Features were classified based on their COV values.

Twelve and eleven (included GLCM- joint maximum, GLCM-angular second moment, GLRL-Low grey level run emphasis, GLRL-Short run low grey level emphasis, GLSZM-Small zone low grey level emphasis, GLSZM-Large zone high grey level emphasis, GLDZM-Large distance low grey level emphasis, NGTDM-busyness) features showed moderately stable and poorly stable against FWHM of the Gaussian filter, respectively. Only three features GLCM (cluster shade), GLRLM (Run length

variance) and GLSZM (Large zone low grey level emphasis) showed large variation (COV > 20%) with different filter. Figure 5.7 shows number of features for each category of the stability of radiomic features against FWHM Gaussian filters.

5.3.5 Agreement

To evaluate the reliability of the results, the experiment was repeated two more times by following the same process described in the method section 5.2. The percent agreement (PA) between the variation of features for all experiments across all of the reconstruction settings were calculated. Figure 5.8 shows the PA values of all features across all of reconstruction settings.

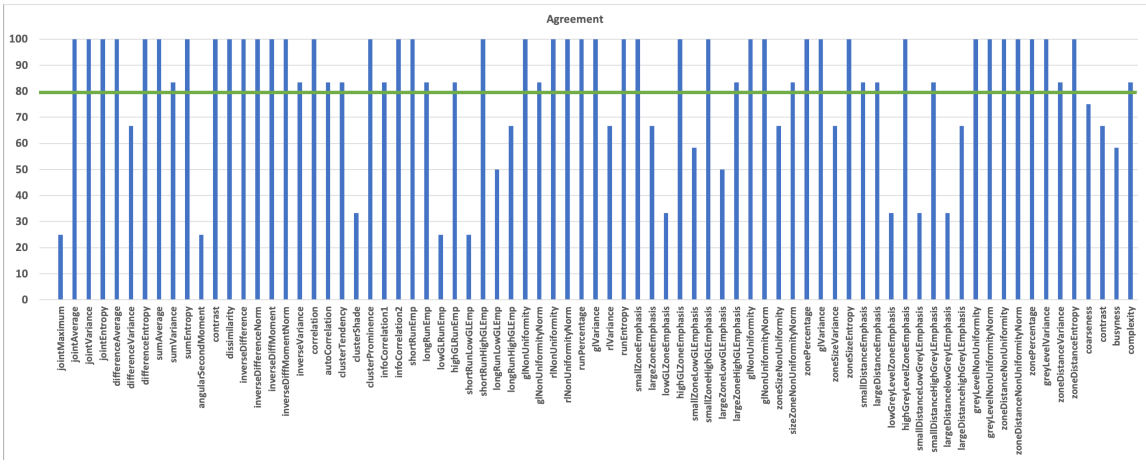

Figure 5.8: Percent agreement (PA) values of all features across all of reconstruction settings. Features with percent agreement (PA) above the green line (80%) were considered to have good agreement

Most features show a high percent agreement with a mean value greater than 82%, especially for GLSZM features, which results in the highest PA with a mean value greater than 86% and a maximum value of 100%. Fifty five (70%) features showed good ($> 80\%$) agreement. In contrast, ten features were found to have 50% or less agreement between the three experiments.

When comparing feature groups, NGTDM features have the lowest percent agreement: an average PA 73% (Figure 5.9). GLRLM and GLSZM were found to have quite similar agreement among all of experiments with a mean value greater than 80%.

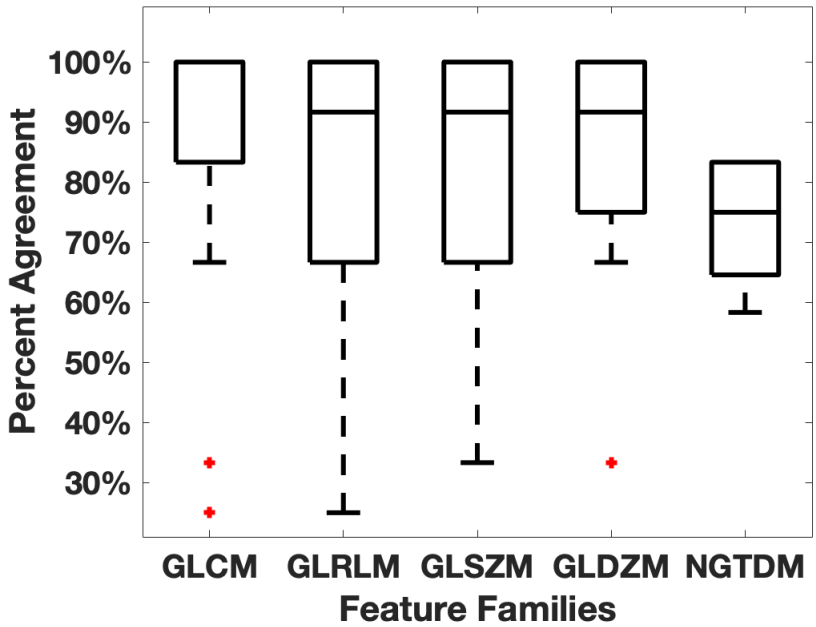

Figure 5.9: Box plot showing the mean of percent agreement among all different texture group.

Table 5.2: List of features that showed high stability and good agreement between experiments on the homogeneous phantom study.

Features Group	Features	Features Group	Features
GLCM	Joint average Joint variance Joint entropy Difference average Difference entropy Sum average Sum variance Sum entropy Contrast Dissimilarity Inverse difference Inverse difference normalised Inverse difference moment Inverse difference moment normalised Inverse variance Correlation Autocorrelation Cluster tendency Cluster prominence Second measure of information correlation	**GLSZM**	Small zone emphasis High grey level zone emphasis Small zone high grey level emphasis Grey level non-uniformity Grey level non-uniformity normalised Zone size non-uniformity normalised Zone percentage Grey level variance Zone size entropy
GLRLM	Short runs emphasis Long runs emphasis High grey level run emphasis Short run high grey level emphasis Grey level nonuniformity Run length non-uniformity Run length non-uniformity normalised Run percentage Grey level variance Run entropy	**GLDZM**	Small distance emphasis Large distance emphasis High grey level zone emphasis Small distance high grey level emphasis Grey level non-uniformity Grey level non-uniformity normalised Zone distance non-uniformity Zone distance non-uniformity normalised Zone percentage Grey level variance Zone distance entropy
NGTDM	Complexity Strength		

5.4 Conclusions

The work described in this chapter was designed to determine the effect of different reconstruction settings (TOF, number of subsets, number of iterations, full width at half maximum (FWHM) of Gaussian filter) on PET images radiomic features using an IQ NEMA phantom with inserts filled with homogeneous radioactivity. In addition, the experiment was repeated two more times and percentage agreement between the three results were calculated. Table 5.2 presented features that showed high stability and good agreement when experiment was repeated. Although the current study is based on a homogeneous radioactivity, the findings allow the removal of features that show high variation. The methods used for this study may be applied to other research to investigate the impact of different reconstruction settings on PET image radiomic features using inserts with heterogeneous radioactivity. This may serve as a future work.

Chapter 6

The Impact of Reconstruction Settings on PET Radiomic Features: An Inhomogeneous Phantom Study

6.1 Introduction

The variation and accuracy of radiomic features remain as one of the most challenging task for implementing radiomic features as biomarkers [50]. Therefore, there is a need for further assessments of the impact of different parameters on PET images radiomics feaures. Clinical studies are subject to extraneous variables such as biological sensitivity, organ motion and patient size. Therefore, phantom studies can be a considerable choice to control for bias relative to biological variability of clinical studies. In Chapter 5, the impact of different reconstruction settings on PET image radiomic features using an IQ NEMA phantom with inserts filled with homogeneous

radioactivity was evaluated. Most of the previous phantom studies of radiomic parameters have utilised homogeneous inserts and very limited research exists which analyse features for heterogeneous phantom images (Forgacs *et al.* [62],Gallivanone *et al.* [76], Presotto *et al.* [87]).

This study set out to assess the effect of reconstruction settings on the stability of PET image radiomic features using homogeneous and heterogeneous inserts. For this purpose, four constructed inserts were placed into a water filled phantom and imaged. Images were reconstructed with different reconstruction settings including number of ordered subsets expectation maximisation (OSEM) subsets, number of iterations, time-of-flight (TOF) and filter cut off. Regions of interest (ROI) were segmented and 78 3D radiomic features for each tumour volume at each setting were extracted. The Coefficient of Variation (COV) was calculated for each feature and stability of features were categorised based on their COV. Methodologically, this study differs from previous research in several ways. Forgacs *et al.* [62] and Presotto *et al.* [87] extracted 27 and 39 radiomic features, respectively while in our study, we extracted 78 radiomic features. Furthermore, this work differs from prior studies on the number of configurations, heterogeneity activity levels and tested reconstruction parameters. As an example, the Forgacs *et al.* [62] investigation was based on only 3 different

numbers of iterations, 2 different subsets and 2 FWHM Gaussian filters whilst our study utilised 6 different numbers of iterations, 5 different subsets and 8 FWHM Gaussian filters. Moreover, all radiomic features in this work were compliant with the IBSI. Hence, this study provides an opportunity to advance our knowledge about the robustness of heterogeneous features against different reconstruction settings.

6.2 Materials and methods

6.2.1 Phantom preparation

Four artificially constructed inserts were placed in a cylindrical uniform water filled phantom. Each insert consists of 7 syringes filled with different radioactivity to model lesions with varying degrees of heterogeneity. Two configurations of homogeneous tumour inserts were constructed by arranging 6 and 7 syringes filled with 40 kBq/ml F-18 activity concentration to mimic tumours ($\approx$ 145 cm^3) with and without necrotic regions, respectively. The two remaining inserts were constructed in a similar way by arranging syringes filled with 3 different F-18 activity concentrations (20,40 and 80 kBq/ml) to mimic heterogeneous tumours with and without necrotic regions, respectively. Figure 6.1 shows an illustrative layout of the four configurations of artificial constructed tumour inserts.

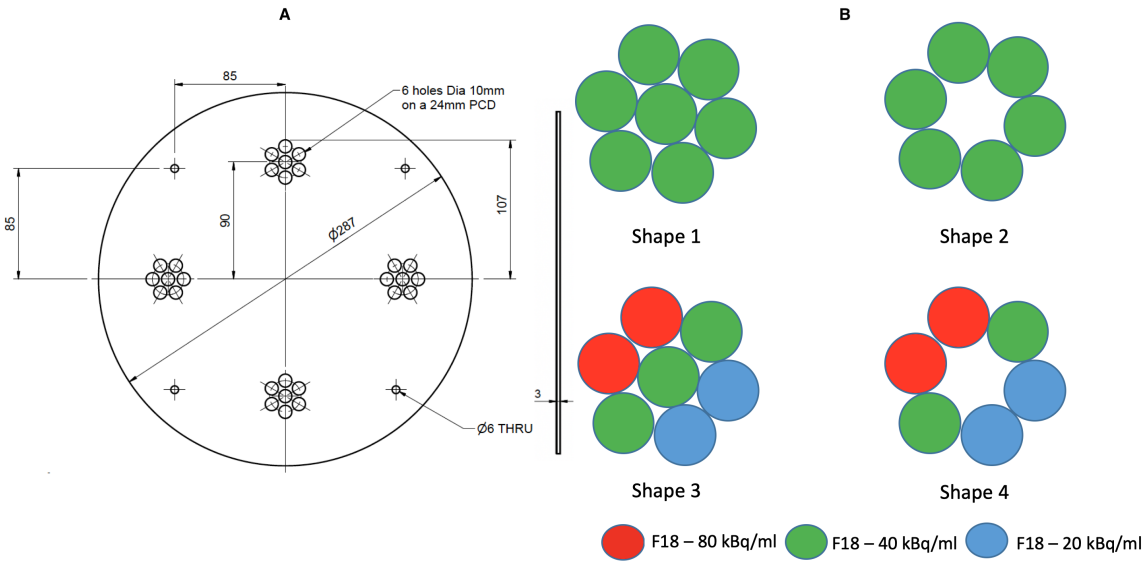

Figure 6.1: An illustrative layout of the syringe mounting plate (A) and four configurations of the artificial tumour inserts (B).

6.2.2 Acquisitions and reconstructions

A GE Discovery 690 PET/CT scanner was used to acquire phantom images. The phantom was imaged for 80 minutes (Figure 6.2). Images were reconstructed using the following default setting (reference image): order subset expectation maximization, PSF, 24 subsets, 2 iterations, 6.4 mm filter cutoff and 256 matrix size. To evaluate the effect of reconstruction settings on image radiomic features, images were reconstructed with different reconstruction settings including: number of subsets, number of iterations, filter cut-off and TOF reconstructions. Table 6.1 shows the reconstruction parameters used to generate new images.

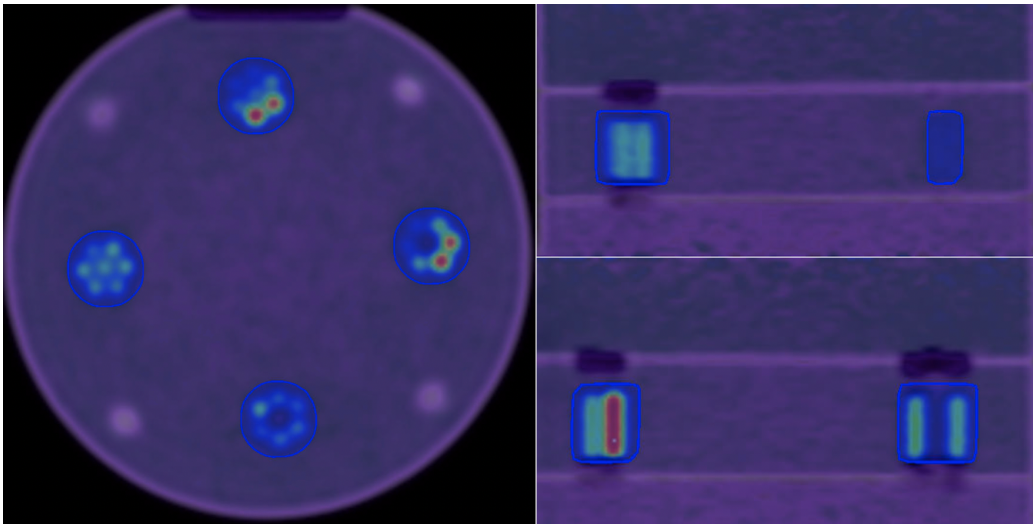

Figure 6.2: Axial, Coronal and Sagittal views for the phantom scan at 80 minutes and default reconstruction settings.

6.2.3 Segmentation

Velocity 3.2.1 software (Varian Medical Systems, Atlanta, USA) was used to obtain the ground truth contour from the first configuration (homogeneous tumour). The manual contouring tools were utilised to draw on new structures in Axial plane using brush and 3D mode. Contour was adjusted in Coronal and Sagittal planes. This contour was then overlaid onto all of the configurations and other subsequent images resulted from changing the reconstruction settings. Figure 6.2 shows Axial, Coronal and Sagittal views for the phantom scan at 80 minutes.

Table 6.1: List of reconstruction settings used to generate new images. Default settings: TOF, 24 OSEM subsets, 2 iterations, 6.4 mm filter cutoff.

Reconstruction parameters	Variations
Number of subsets	12, 16, 18, 24, 32
Number of iterations	1, 2, 3, 4, 5, 6
Filter Cut-off	0, 1, 2, 3, 4, 5, 6, 7
TOF	Yes, No
Total number of parameters	21

6.2.4 Feature extraction and data analysis

For each ROI, SPAARC, an in-house developed tool built on Matlab, was used to extract 78 3D-radiomic features. One third of these features were excluded and 52 features (that showed high stability and high percent agreement in Chapter 5) were included in this study. Features including a 20 grey level co-occurrence matrix (GLCM), 10 grey-level run-length matrix (GLRLM), 9 grey-level size zone matrix (GLSZM), 11 grey-level distance zone matrix (GLDZM) and 2 neighbourhood grey-tone difference matrix (NGTDM) were extracted. SPAARC radiomic analysis is standardized according to IBSI [75]. All examined features are listed in Table 6.2.

Table 6.2: List of examined radiomic features. These features were found to be robust (low variation and high percent agreement) on the homogeneous phantom study.

Features Group	Features	Features Group	Features
GLCM	Joint average Joint variance Joint entropy Difference average Difference entropy Sum average Sum variance Sum entropy Contrast Dissimilarity Inverse difference Inverse difference normalised Inverse difference moment Inverse difference moment normalised Inverse variance Correlation Autocorrelation Cluster tendency Cluster prominence Second measure of information correlation	**GLSZM**	Small zone emphasis High grey level zone emphasis Small zone high grey level emphasis Grey level non-uniformity Grey level non-uniformity normalised Zone size non-uniformity normalised Zone percentage Grey level variance Zone size entropy
GLRLM	Short runs emphasis Long runs emphasis High grey level run emphasis Short run high grey level emphasis Grey level nonuniformity Run length non-uniformity Run length non-uniformity normalised Run percentage Grey level variance Run entropy	**GLDZM**	Small distance emphasis Large distance emphasis High grey level zone emphasis Small distance high grey level emphasis Grey level non-uniformity Grey level non-uniformity normalised Zone distance non-uniformity Zone distance non-uniformity normalised Zone percentage Grey level variance Zone distance entropy
NGTDM	Complexity Strength		

To evaluate feature stability when extracted at different reconstruction settings, COV was calculated for each feature over the different reconstruction settings in the same way that described in Section 2.2.7. COV which is one of the most widely used

tool in the literature on the stability of radiomic features [55, 48, 76]. Features were categorized based on their COV values into four groups including stable (COV $\leq$ 5%), moderately stable (5% $<$ COV $\leq$ 10%), poorly stable (10% $<$ COV $\leq$ 20%) and unstable (COV $>$ 20%).

6.3 Results

As reported in several studies, reconstruction settings have a different influence on PET images [46, 55, 57, 62, 58, 61, 56, 48, 63]. The hierarchical cluster tree of the radiomics features across COV of image reconstruction settings was created as a variability heat map. The results (as shown in Figures 6.3 and 6.4) indicate the effect of reconstruction settings, Time of Flight, number of subsets, number of iterations, and FWHM of a Gaussian filter. The radiomics heatmap (figures 6.3 and 6.4) depicts the variation of features over various parameter settings as quantified utilizing the average of COV for all of lesions. Table 6.3 presents unstable (COV $>$ 20%) features as ranked based on average of COV over all tested reconstruction settings. Unstable features included GLRLM (Long run low grey level emphasis) and GLSZM (Large zone low grey level emphasis). Table 6.3 also showed features including GLCM (Difference entropy, Inverse difference normalised, Inverse differ-

ence moment normalised, Second measure of information correlation), GLRL (Short runs emphasis, Run percentage, Run entropy), GLSZM (Zone size entropy), GLDZM (Zone distance entropy), NGTDM (strenght) to be stable (COV $\leq$ 5%) against all reconstruction settings. Figure 6.5, 6.6 and 6.7 exhibit the COV values for GLCM, GLRLM, GLSZM, GLDZM and NGTDM radiomic features. When comparing feature groups, NGTDM features have the lowest mean COV (figure 6.8). GLSZM was the most sensitive feature type to the reconstruction settings.

6.3.1 Impact of TOF

As shown in Figure 6.9, fifty one features were stable against TOF. Such stable features included GLCM (joint entropy, difference average, sum entropy, correlation, joint maximum), GLRLM (shortRunEmp, Long run high grey level emphasis, Grey level non uniformity, run percentage), GLSZM (Small zone emphasis, Zone percentage, Zone size entropy), GLDZM (Small distance emphasis, Large distance emphasis, Low grey level zone emphasis, Zone distance variance) and NGTDM (Coarseness, Busyness, Complexity, Strength). Only one feature (GLCM-Contrast, GLSZM-Zone Size Variance, GLSZM-Large Zone Emphasis and NGTDM-Contrast) were moderate stable against TOF. No features were poorly stable nor unstable.

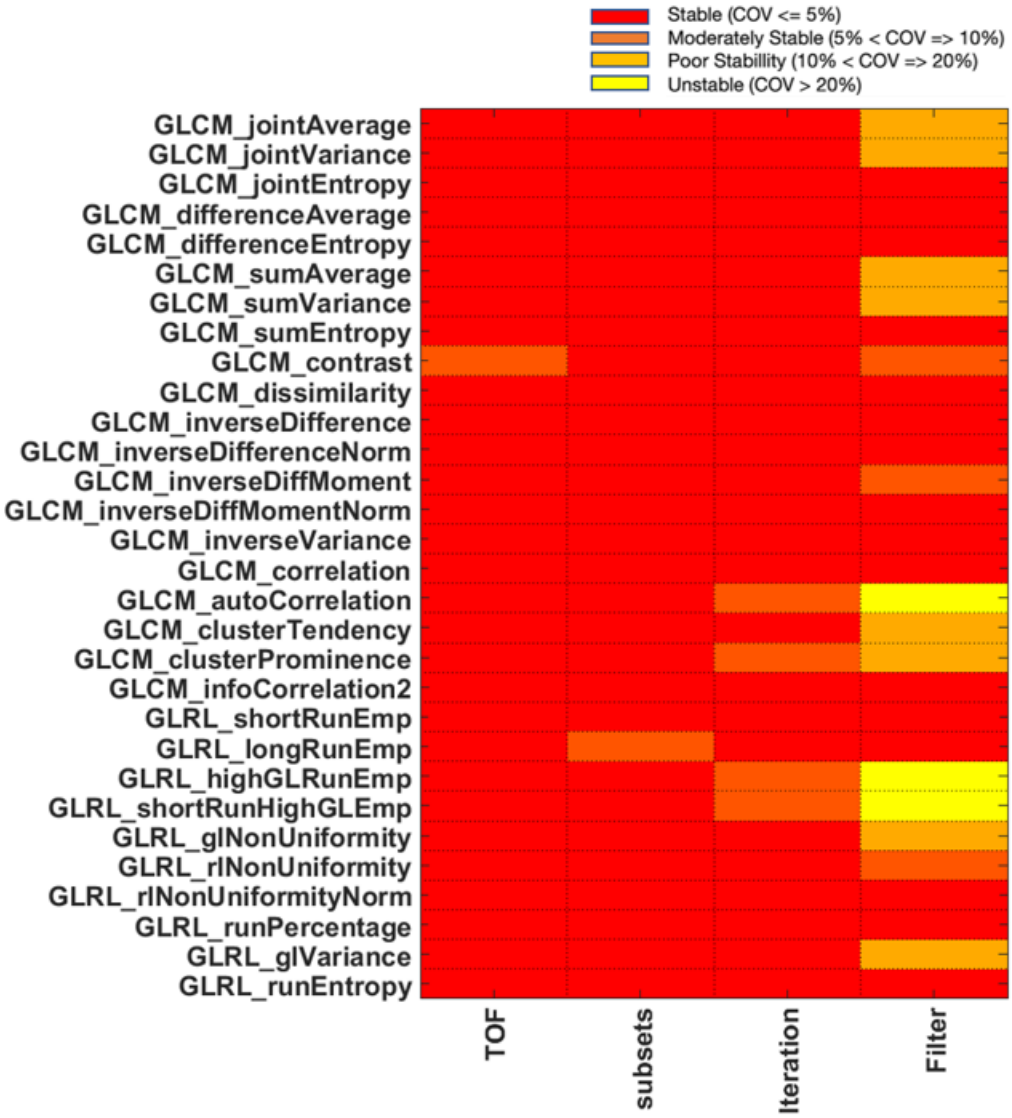

Figure 6.3: Inter-setting coefficient of variation was calculated for each image feature over applying different reconstruction settings (e.g. five different numbers of subsets). Features were categorised into four groups (stable, moderately stable, poorly stable, and unstable) based on their COV values. The hierarchical cluster tree of the radiomics features across COV of image reconstruction settings was created as a variability heat map. This figure is a heat map of the stability of GLCM and GLRL features against different reconstruction settings.

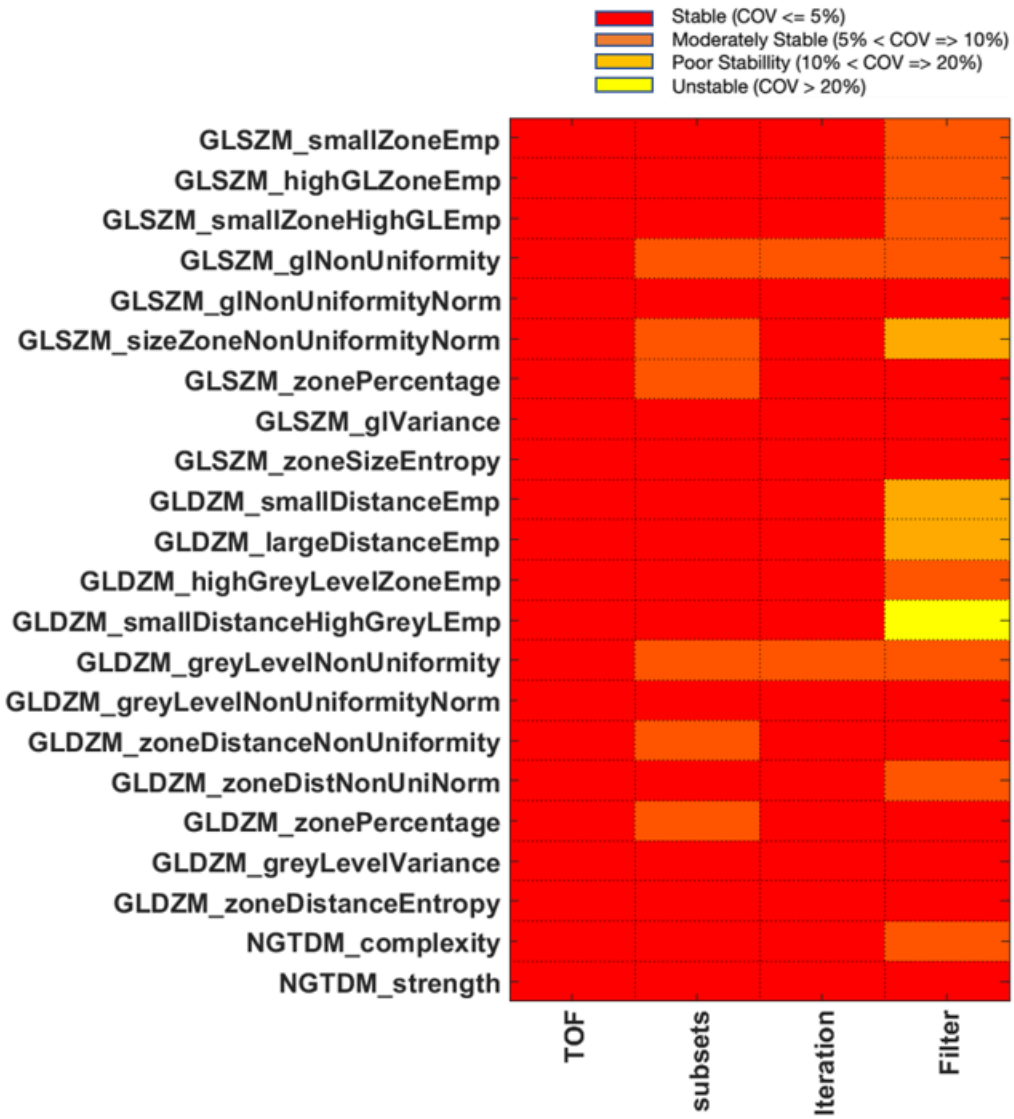

Figure 6.4: Inter-setting coefficient of variation was calculated for each image feature over applying different reconstruction settings (e.g. six different numbers of iterations). Features were categorised into four groups (stable, moderately stable, poorly stable, and unstable) based on their COV values. The hierarchical cluster tree of the radiomics features across COV of image reconstruction settings was created as a variability heat map. This figure is a heat map of the stability of GLSZM, GLDZM and NGTDM features against different reconstruction settings.

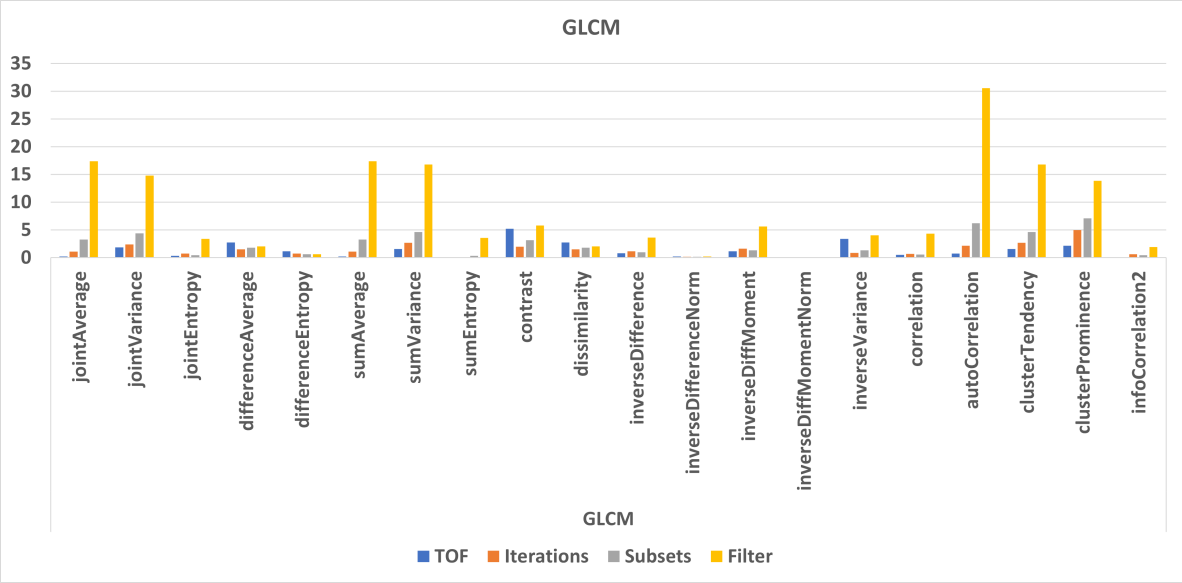

Figure 6.5: Bar chart illustrating the COV values for 20 (GLCM) radiomic features.

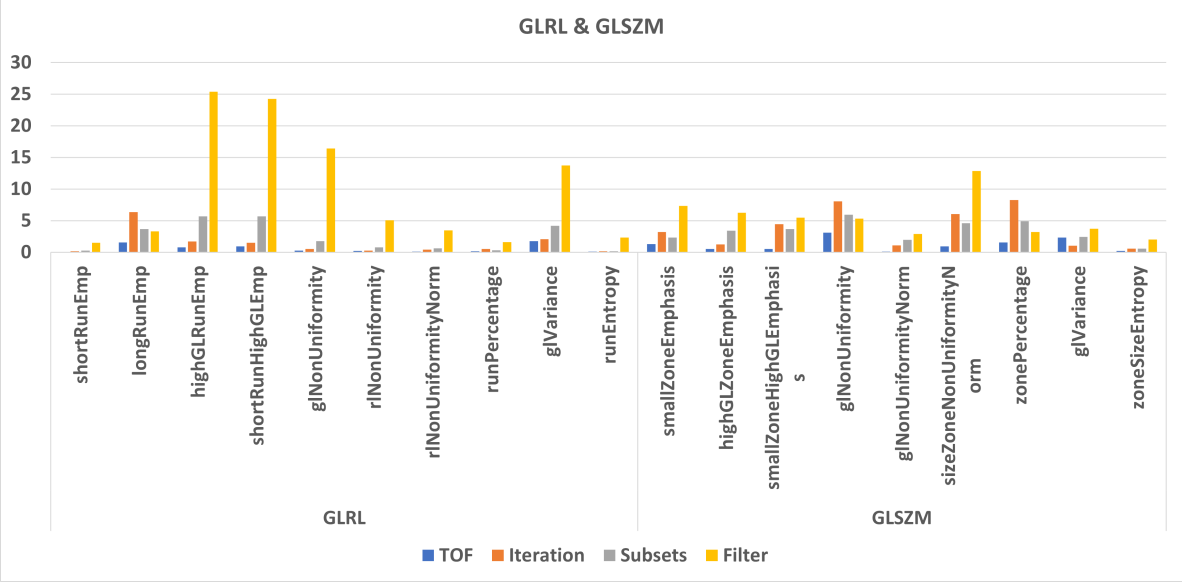

Figure 6.6: Bar chart illustrating the COV values for 19 (GLRLM & GLSZM) radiomic features.

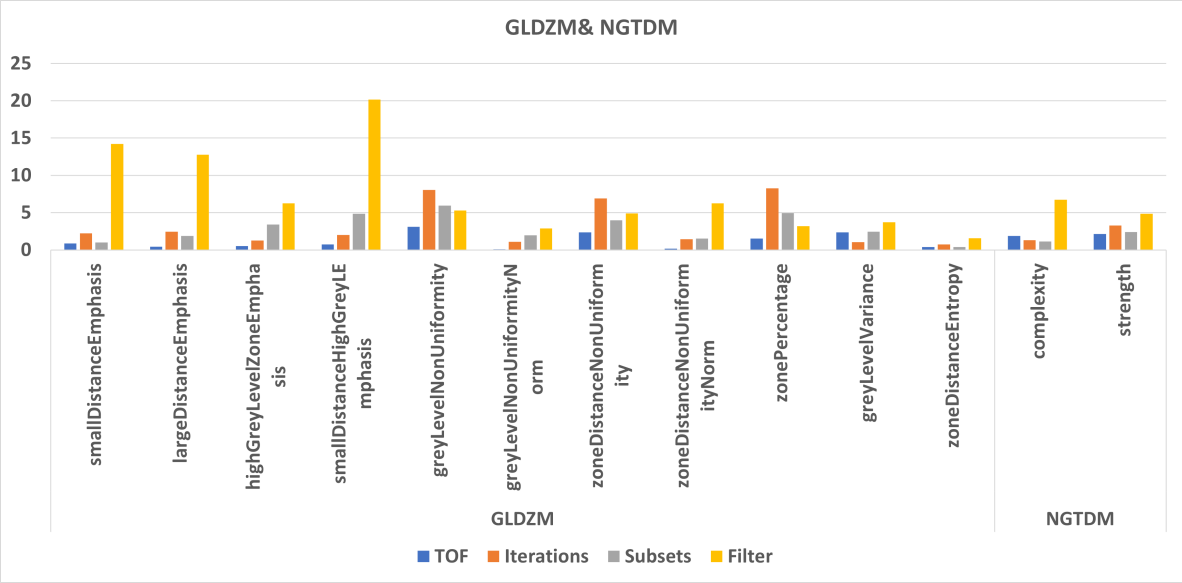

Figure 6.7: Bar chart illustrating the COV values for 13 (GLDZM & NGTD) radiomic features.

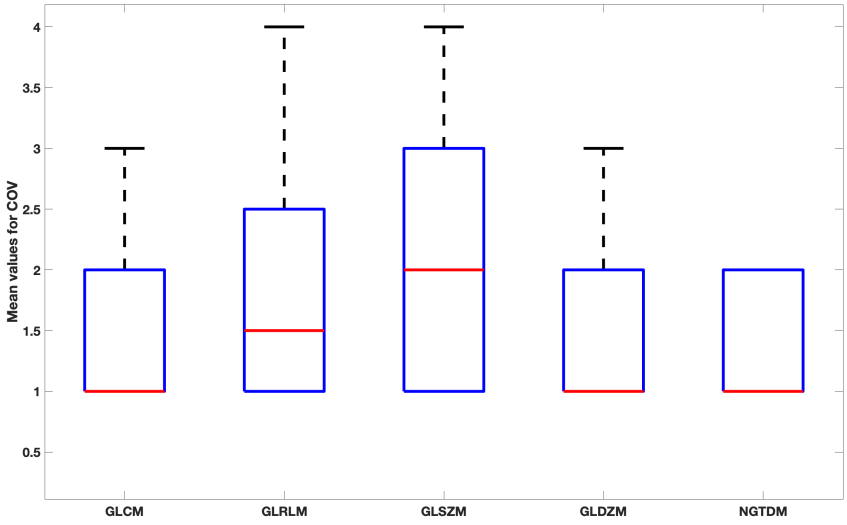

Figure 6.8: Box plot for the mean values of COV for each feature family over all reconstruction settings

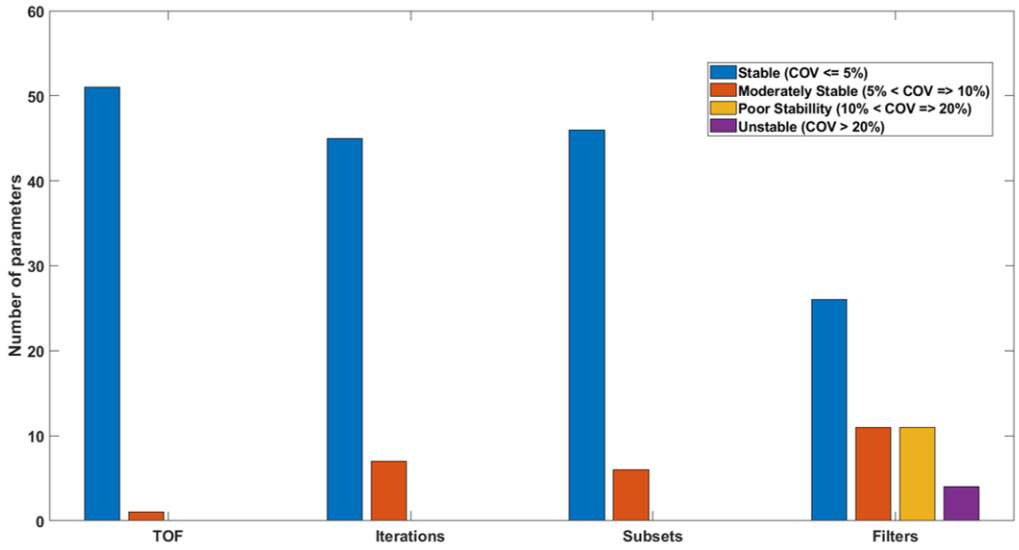

Figure 6.9: Bar chart showing the number of features for each category

6.3.2 Impact of number of subsets

Forty six features were found to have high stability (COV $\leq$ 5%) with number of OSEM subsets. Six features (7%) including GLCM-auto correlation, GLCM-cluster prominence, GLRLM-high grey level run emphasis, GLRLM-short Run high grey level emphasis, GLSZM-grey level non uniformity and GLDZM-grey level non Uniformity were found to have moderate stability (5% $<$ COV $\leq$ 10%). No features showed poor stability nor instability at different number of subsets. All features from NGTDM were stable (COV $\leq$ 5%) against subsets.

Table 6.3: Features for each category of stability over all reconstruction settings

Features	COV < 5%	5% < COV ≤ 10%	10% < COV ≤ 20%	COV > 20%
GLCM	jointEntropy differenceAverage differenceEntropy sumEntropy contrast dissimilarity inverseDifference inverseDifferenceNorm inverseDiffMoment inverseDiffMomentNorm inverseVariance correlation infoCorrelation1 infoCorrelation2	jointAverage jointVariance differenceVariance sumAverage sumVariance autoCorrelation clusterTendency clusterShade clusterProminence	jointMaximum angularSecondMoment	
GLRLM	shortRunEmp longRunEmp glNonUniformity glNonUniformityNorm rlNonUniformity rlNonUniformityNorm runPercentage runEntropy	highGLRunEmp shortRunHighGLEmp glVariance rlVariance	lowGLRunEmp shortRunLowGLEmp longRunHighGLEmp	longRunLowGLEmp
GLSZM	smallZoneEmphasis highGLZoneEmphasis smallZoneHighGLEmphasis glNonUniformityNorm zonePercentage glVariance zoneSizeEntropy	glNonUniformity zoneSizeNonUniformity sizeZoneNonUniformityNorm	largeZoneEmphasis lowGLZoneEmphasis smallZoneLowGLEmphasis largeZoneHighGLEmphasis zoneSizeVariance	largeZoneLowGLEmphasis
GLDZM	smallDistanceEmphasis largeDistanceEmphasis highGreyLevelZoneEmphasis largeDistancehighGreyLEmphasis greyLevelNonUniformityNorm zoneDistanceNonUniformity zoneDistanceNonUniformityNorm zonePercentage greyLevelVariance zoneDistanceVariance zoneDistanceEntropy	smallDistanceHighGreyLEmphasis greyLevelNonUniformity	lowGreyLevelZoneEmphasis smallDistanceLowGreyLEmphasis largeDistancelowGreyLEmphasis	
NGTDM	coarseness complexity strength	contrast busyness		

6.3.3 Impact of the number of iterations

More than 57% (45) of features were found to be stable with different number of iterations. Features with very low varation included GLCM (sum average, sum variance, sum entropy, contrast, dissimilarity, inverse difference, inverse difference normalised), GLRLM (run percentage, grey level Variance, run entropy), GLSZM (Small zone emphasis, High grey level zone emphasis , Small zone low grey level emphasis), GLDZM (Small distance high grey level emphasis, Large distance high grey level emphasis, Grey level non-uniformity normalised) and NGTDM (coarseness, busyness, complexity). Seven (GLRL-Long run emphasis, GLSZM-grey level Nnon uniformity, GLSZM-size zone non uniformity normalised, GLSZM- zone percentage, GLDZM-grey level non uniformity, GLDZM-zone distance non uniformity, GLDZM-zone percentage) features showed moderately stable against number of iteration. No features were found to be poorly stable ($10\% < \text{COV} \leq 20\%$) nor unstable ($\text{COV} > 20\%$) with different numbers of iterations.

6.3.4 Impact of FWHM of the Gaussian filter

With changing FWHM of a Gaussian filter, twenty six features showed very small variation ($\text{COV} \leq 5\%$). About 14% (11) of features were found to be moderately

stable. Eleven features also showed poor stability against FWHM of the Gaussian filter. Four features including GLCM-auto correlation, GLRLM-High grey level run emphasis, GLRLM- short run high grey level emphasis and GLDZM-Small distance high grey level emphasis demonstrated high variation (COV $>$ 20%) (Figure 6.9).

6.4 Towards selection of robust heterogeneous PET image radiomic features

The impact of different reconstruction settings on homogeneous and heterogeneous phantom PET images was investigated in Chapter 5 and this Chapter 6. Features which showed high stability and good agreement in the first phantom study (Chapter 5) using a homogeneous phantom were passed to further assessment with heterogeneous phantom study. Together these two empirical studies provide important insights into the robustness of radiomic features against reconstruction settings. However, the ability of these robust features to identify the differences between regions (inserts) is still unknown. To this end, the present study was designed to determine the capability of PET radiomic features to capture the differences between regions (inserts). Thus, this study will compare the value of features that were found to be robust in Sections 5.3 and 6.3 at six different regions. For this purpose, a further analysis was applied to 38 radiomic features.

To assess the ability of radiomic features to capture heterogeneity differences, the Friedman test [88, 89] can be used. The Friedman test is a statistical tool which was introduced by Milton Friedman in 1937. It involves ranking each row separately and then sums the ranks in each column (group). In our study, rows contain feature values at different reconstruction settings. P values will be small if the sums are very different. In contrast, high P values indicate that there is no significant difference between tested groups.

To determine whether or not there is a statistically significant difference between the means of regions in which the same reconstruction parameter was used in each configuration (shape1 vs shape2, shape1 vs shape3, shape1 vs shape4, shape2 vs shape3, shape2 vs shape4 and shape3 vs shape4), The Friedman test was performed for each feature and the steps of applying the Friedman test can be summarized as follows:

1) Rows (reconstruction parameter) were ranked in ascending order based on features values for each reconstruction parameter.

2) The sum of ranks for each region (column) was calculated

3) The test statistic (Q) was calculated using the following equation (6.1):

$$Q = \frac{12\ x\ \sum R_j^2}{nk(k+1)} - 3n(k+1) \tag{6.1}$$

where:

n = number of reconstruction parameters = 21; k = number of configurations (regions) = 2; R_j^2= sum of ranks for the jth region.

4) Corresponding P value was determined.

The null hypothesis for the Friedman test is that there are no differences between the variables. If the calculated probability is low (P less than the selected significance level) the null-hypothesis is rejected. P value was tested for each feature and if P value was less than 0.05 then the null hypothesis was rejected and stated as false as the texture was showing a difference between those regions within all reconstruction settings. If P value was higher than 0.05 then the null hypothesis was accepted and stated as true as the texture was showing no difference between tested regions.

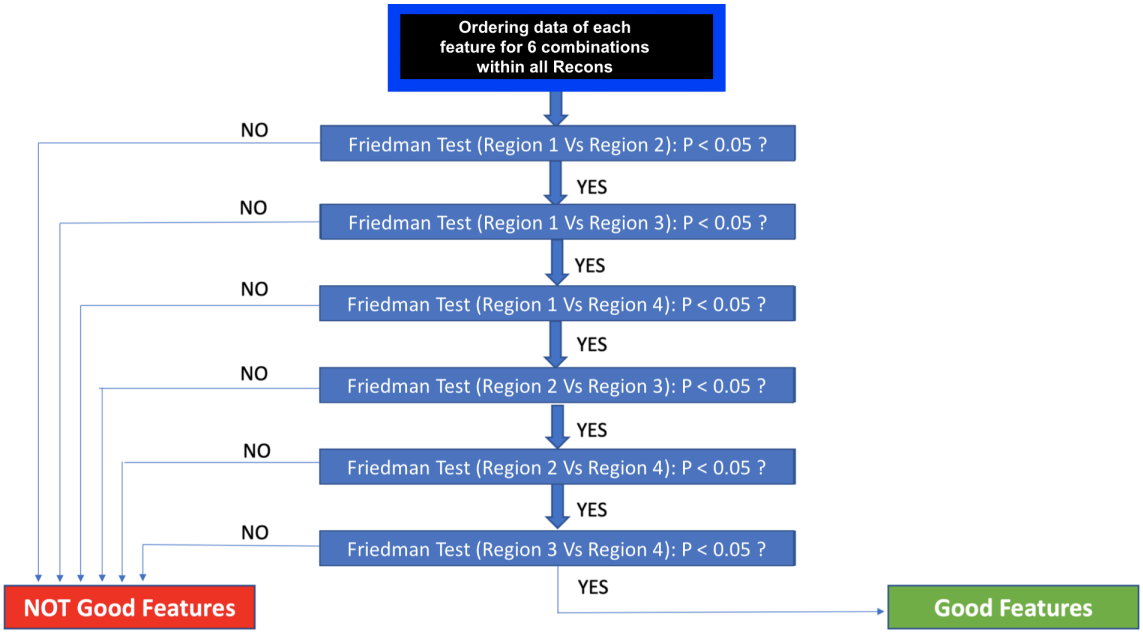

Figure 6.10: Workflow of selecting good features that can detect the differences between regions.

If any feature demonstrates a difference for all of the 6 combinations, the feature will be considered as "good feature" to capture heterogeneity differences. Thus, throughout this thesis, the term good feature is used to refer to feature that showed high stability and has ability to differentiate between heterogeneity levels. Otherwise, that feature is deemed unsuitable to detect the differences between regions. Figures 6.10 and 6.11 show the workflow that was followed in this analysis.

We tested all features (38) with high stability over all reconstruction settings to find out how many of them differed statistically between regions. Fifteen out of 38 (31%) features showed statistically significant difference between each region and

others. Such features included GLCM (differenceAverage, differenceEntropy, dissimilarity and inverseDifference), GLRL (longRunEmp, glNonUniformity and runPercentage) and NGTDM (complexity and strength). More than half (8) of the discriminative features were mainly derived from grey level co-occurrence matrix. It was observed that some features such as GLSZM (glVariance) were statistically different between region 1 vs 3, 2 vs 3 and 2 vs 4, but not between region 1 vs 2, 1 vs 4 and 3 vs 4.

Feature A	Region 1	Region 2
TOF	$X_{1,\ TOF}$	$X_{2,\ TOF}$
Non-TOF	$X_{1,\ Non-TOF}$	$X_{2,\ Non-TOF}$
Number of subsets = 12	$X_{1,\ subsets:12}$	$X_{2,\ subsets:12}$
Number of subsets = 16	$X_{1,\ subsets:16}$	$X_{2,\ subsets:16}$
Number of subsets = 18	$X_{1,\ subsets:18}$	$X_{2,\ subsets:18}$
Number of subsets = 24	$X_{1,\ subsets:24}$	$X_{2,\ subsets:24}$
Number of subsets = 32	$X_{1,\ subsets:32}$	$X_{2,\ subsets:32}$
Number of iterations = 1	$X_{1,\ iteration:1}$	$X_{2,\ iteration:1}$
Number of iterations = 2	$X_{1,\ iteration:2}$	$X_{2,\ iteration:2}$
Number of iterations = 3	$X_{1,\ iteration:3}$	$X_{2,\ iteration:3}$
Number of iterations = 4	$X_{1,\ iteration:4}$	$X_{2,\ iteration:4}$
Number of iterations = 5	$X_{1,\ iteration:5}$	$X_{2,\ iteration:5}$
Number of iterations = 6	$X_{1,\ iteration:6}$	$X_{2,\ iteration:6}$
FWHM filter = 0	$X_{1,\ filter:0}$	$X_{2,\ filter:0}$
FWHM filter = 1	$X_{1,\ filter:1}$	$X_{2,\ filter:1}$
FWHM filter = 2	$X_{1,\ filter:2}$	$X_{2,\ filter:2}$
FWHM filter = 3	$X_{1,\ filter:3}$	$X_{2,\ filter:3}$
FWHM filter = 4	$X_{1,\ filter:4}$	$X_{2,\ filter:4}$
FWHM filter = 5	$X_{1,\ filter:5}$	$X_{2,\ filter:5}$
FWHM filter = 6	$X_{1,\ filter:6}$	$X_{2,\ filter:6}$
FWHM filter = 7	$X_{1,\ filter:7}$	$X_{2,\ filter:7}$

Figure 6.11: An illustrative example showing how the data sorted to perform the Friedman test. The example includes 21 reconstruction settings and two different regions (shape1 vs shape2). This was repeated for each of five other combinations (shape1 vs shape3, shape1 vs shape4, shape2 vs shape3, shape2 vs shape4 and shape3 vs shape4). P values were then calculated for each pair of regions to determine whether or not there is a statistically significant difference between the means of regions.

6.5 Conclusions

This chapter described the work carried out to evaluate the effect of different reconstruction settings on the stability of PET image radiomic features based on tumour inserts with differing levels of heterogeneity. For this purpose, an array of radioactivity filled syringes (7 in total for each tumour) were placed in a phantom and imaged for 80 minutes. Images were reconstructed with different reconstruction settings (TOF, Subsets, Iterations and FWHM Gusion filters). Various types of radiomic features (GLCM, GLRLM, GLSZM, GLDZM and NGTDM) compliant with the IBSI were extracted. The COV and features categorised into 4 groups based on their COV values were then calculated. 32 (41%) features were found to have very low sensitivity against all tested reconstruction parameters. The implications will be explained in the final conclusive discussion (Section 8.1) of the thesis.

The main goal of the study showed in Section 6.4 was to determine the ability of stable features to detect the differences between two or more regions. Further analysis was applied to the findings of Chapter 5 and this Chapter 6. Features extracted from each region (Figure 6.1) at all reconstruction settings were analysed with the Friedman test in order to find out whether the features differed between regions.

This study has shown that 12 features were statistically significantly different ($P < 0.05$) between all regions. Features which were most often statistically significantly different consisted mainly of those derived from the grey level co-occurrence matrix. This research will serve as a base for future studies that focus on implementing radiomic features as biomarkers. Although the current study is based on phantom data, it offers some insight into conducting similar analysis on clinical research. Finally, this study suggests features with high COV should be removed from the list of features that can be utilised as new biomarkers. Further clinical studies are needed to prove an established role of PET images radiomic features in serving as an adjuvant diagnostic tool.

Chapter 7

Artificial Neural Network Algorithm to Cluster Radiomic Data

7.1 Introduction

As described in Section 4.1, self-organizing map (SOM) is a type of artificial neural network which takes a set of texture parameters and maps them onto a two dimensional grid of neurons [84, 85]. Each neuron in the grid is assigned an initial weight vector $M = (m_{g1}, m_{g2}, .., m_{gD})$ with the same dimensionality as that of the input data g=(1,2,...L). The training uses competitive learning (nodes compete to respond to input data). In Chapter 4, SOM was implemented to explore and cluster the statistical variability of PET texture parameters in the pre-clinical model. In that study, a set of texture parameters (input data) have been mapped onto a two dimensional grid of neurons. The findings have demonstrated that the statistical

distribution of the input texture has clear modes which are dependent on contour size. This means that changing contour size alters the distribution of standardised (and normalised (z-score)) texture features.

This chapter is an extension to the previous work (Chapter 4) and attempts to explores applicability of SOM in identifying emergent properties that affect radiomics variability involving heterogeneous phantom study. The methods and results will be highlighted in detail in the following sections.

7.2 Materials and methods

PET images of a phantom containing four inserts were used in this study. The inserts were constructed using an array of radioactivity filled syringes (7 in total for each insert) with varying degrees of heterogeneity (20, 40 and 80 KBq of 18F-FDG). The phantom was imaged for 80 minutes and re-binned into 4 x 20 minute PET scans (20, 40, 60 and 80 minutes). Each tumour was segmented by four different systematic 3D-Contour sizes (9, 10, 11 and 12 mm) using Velocity 3.2.1 software (Varian Medical Systems, Palo Alto, CA). SPAARC was utilised to extract 78 3D-Radiomic features for each volume at each time point. Thus, each feature has 64 observable values that arise from 4 inserts, 4 different contour sizes and different time points. The

Z-score (Z-omic) for each texture measurement was calculated for inter-comparison of different textures. The mean Z-omic was calculated for 5 pre-defined groups of texture including grey level co-occurrence matrix (GLCM), grey-level run-length matrix (GLRLM), grey-level size zone matrix (GLSZM), grey-level distance zone matrix (GLDZM) and neighbourhood grey-tone difference matrix (NGTDM). To learn the self-organising map of the averaged Z-omic using 16 organising neural networks, the R software was used (learning rate of 0.05 and a Gaussian neighbourhood function with standard deviation 1).

7.3 Results

From the input dataset each instance of measured texture (Z-omic) was assigned to a single node that best represents its distribution of variability. The classification of each node and the relative contribution of the grouped features is demonstrated in the codes plot (pie representations of the representative vectors for the grid) shown in figure 7.1. To assess the impact of confounding variables (in this study acquisition time, contour size) on the variability of texture parameters, cluster analysis on the distribution of the contour sizes and acquisition times over the nodes of the SOM can be performed and a heatmap may be visualised.

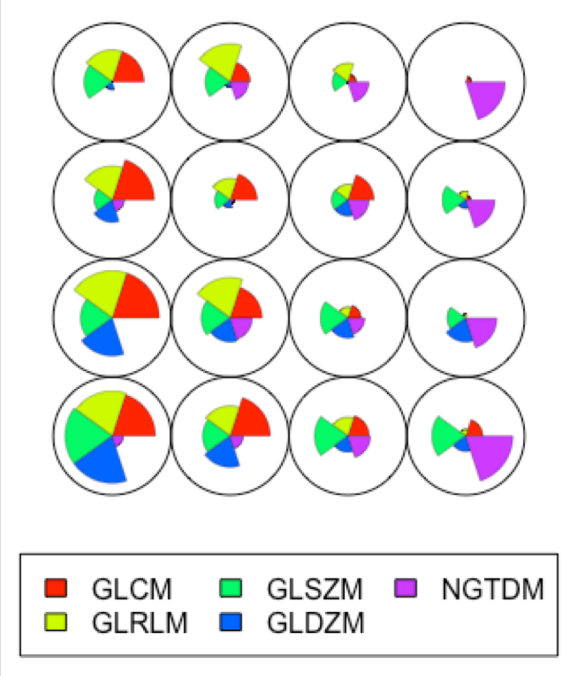

Figure 7.1: Codes plots (pie representations of the representative vectors for the grid) for the texture features. These codes plots represent the clusters (node wieght vectors associated with each node).

The SOM is blind to the considered confounding variables, so any visible clusters in the heatmap, demonstrate an association between the confounding variable and the input data. Figures 7.2 and 7.3 present the distribution of the second and third contour sizes with respect to the self-organised features. From these figures, it can be clearly seen that two distinct clusters can be visualised in the heatmap. This allows us to interpret that contour size acts as a confounding variable when performing texture analysis and varying contour size alters the distribution of standardised (and normalised (z-score) texture features.

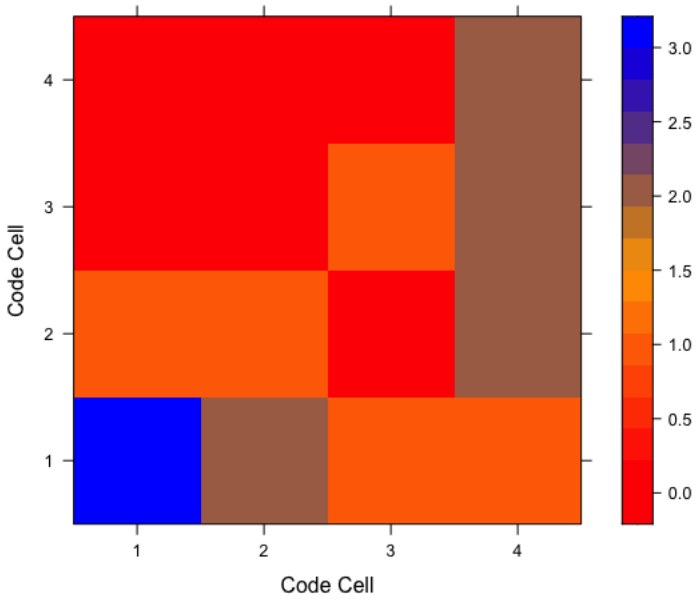

Figure 7.2: The distribution of the second contour size with respect to the self-organised features. This heat map demonstrates the distribution or frequency of the first contour size over the code cells (nodes) of the SOM.

No such relationship was observed when probing post injection imaging time as the confounding variable as no clusters were observed in the heatmap. Therefore, it is evidenced that an emergent property of the statistical variability of the grouped textures is the extremes of contour size. The findings of the current study are consistent with previous study (preclinical study) which suggest that the statistical distribution of the input texture has clear modes which are dependent on contour size.

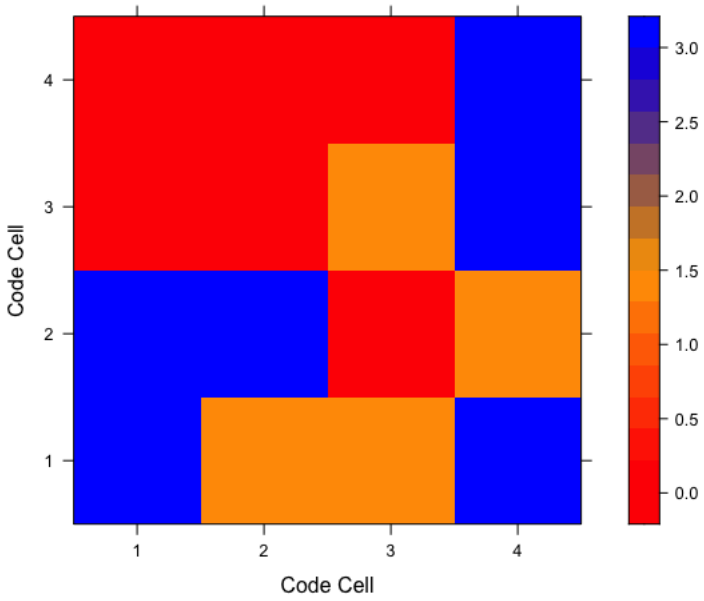

Figure 7.3: The distribution of the third contour size with respect to the self-organized features. This heat map demonstrates the distribution or frequency of the first contour size over the code cells (nodes) of the SOM.

7.4 Conclusions

The work described in this chapter was designed to implement a self-organising map to texture feature analysis in a heterogeneous phantom experiment model. The present study confirms previous findings (described in section 4.3) and contributes additional evidence that suggests SOM has ability to identify emergent properties that affect their variability, in this case contour size.

In addition, the SOM can be used with outcome data to serve as a predictive tool for dependent variables (e.g therapy response, prognosis). In so doing the learnt

representations of self-organised features serve as the attributes for prediction which will take into consideration the statistical variability in the underlying dataset.

Chapter 8

Discussion and conclusions

8.1 Discussion

Chapter 1 provided a literature review and an introduction to PET imaging radiomics. In the literature review, it was identified that the quantitative accuracy of radiomic features is a central issue which determines the success of utilising radiomic features as a biomarker in clinical practice [45, 90]. The accuracy of radiomic features may be affected by different factors such as post injection imaging time and reconstruction settings. More information on the impact of these factors would help us to establish a greater understanding on this matter and increase confidence in the utilisation of texture features as imaging biomarkers [45, 91]. This thesis intended to: (1) determine whether different factors (post injection imaging time and reconstruction settings) have an effect on PET radiomics analysis, (2) investigate the statistical

influence (impact of two combined variables) of imaging time and segmentation volume on PET radiomic features, and (3), explore, capture and cluster the statistical variability of PET texture parameters using a self-organising map (SOM).

In the work addressed in Chapter 2, we aimed to assess the impact of acquisition time and reconstruction settings (number of subsets and iterations) on PET image radiomic features involving a preclinical study. Conducting preclinical studies in the field of radiomics may serve as a bridge to clinical research and trial translation. Our results demonstrated that different reconstruction settings have different impacts on PET radiomic features. For example, grey level non-uniformity from GLRLM and grey level non-uniformity normalised from GLSZM were stable and non-stable against number of iterations, respectively. Zone distance variance from GLDZM had low stability against number of iterations and medium stability against number of subsets. This is in line with previous studies which have demonstrated that the stability of PET radiomic features is changeable over different reconstruction settings.

There are some noticeable differences and similarities between our findings and those of previous studies. For example, Yan *et al.* investigated the impact of reconstruction settings on 61 statistical and heterogeneity features such as GLCM, GLRLM, GLSZM and NGLDM [55]. Their results showed that the iteration number

and full width at half maximum (FWHM) had a similar effect on the stability of features. In our study, in comparison to Yan *et al.*, about 9 features (from 40 common features) have good agreement. For example, the Energy feature had high stability (COV $\leq 5\%$) and the skewness feature was found to have low stability (COV $> 20\%$) against number of iterations in both studies. In another similar study, Bailly *et al.* assessed the variability of 15 textural features against reconstruction parameters (e.g. reconstruction algorithm, number of iterations and matrix size) [92]. Their results demonstrated that 4 (26%) features were very sensitive to changes in the number of iterations. In comparison to Bailly *et al.*, our findings have 23% (3 from 13 common features) agreement with their results. Our results in chapter 2 were based on a preclinical data set while the majority of previous studies were based on either clinical or phantom data and this may explain the differences between our results and others. Thus, a possible explanation for differences between our results and others might be due to biological parameters such as proliferation or metabolism.

In line with the aims laid out in Section 1.3, the studies discussed in Chapter 3 have primarily focused on investigating the influence of a combination of two variables (post injection imaging time and segmentation volume). In this work, the correlation of 78 features with varying time (T1 to T4) and contour size (C1 to

C4) was investigated resulting in 8 correlation matrices (as shown in Figure 3.4) for each feature. The determinant of each correlation matrix was then calculated. Variables were more correlated as the determinant approaches zero and uncorrelated as the determinant approaches one. The results of this study have shown that mean determinant of correlation for post injection imaging time = 0.13296 vs. mean determinant of correlation for segmentation volume = 0.02378. Thus, the key finding drawn from this study is that texture features have greater correlation with contour size than acquisition time.

The research described in Chapter 4 aimed to address the use of artificial neural networks, specifically self-organising maps, to visualise the input data. Each SOM was trained using unsupervised learning to generate a lower dimensional representation of the input data on an underlying manifold. SOM produce heatmaps that can be used to assess the impact of predefined variables. The SOM is blind to the suspected confounding variables, so any visible clusters in the heatmap demonstrate an association between the confounding variable and the input data (radiomic features). Two distinct clusters were visualised in the heatmap allowing us to interpret that contour size acts as a confounding variable when performing texture analysis and changing contour size alters the distribution of standardised (and normalised

(z-score)) texture features. No such relationship was observed when probing post injection imaging time as the confounding variable as no clusters were observed in the heatmap. We expect SOM to be valuable to serve as a predictive tool for dependent variables (e.g therapy response and prognosis). An example of the application of the SOM for predictive analysis will be highlighted at the end of this section.

The purpose of Chapter 5 was to investigate the impact of different reconstruction settings on PET radiomic features. In order to achieve this, an IQ NEMA phantom with 6 inserts filled with homogeneous radioactivity was scanned and images reconstructed with different reconstruction parameters including:

(a) with/without TOF.

(b) Five different numbers of subsets.

(c) Six numbers of iterations.

(d) Eight full widths at half maximum (FWHM) of Gaussian filters.

Seventy eight 3D radiomic features were extracted from each new image. The results presented in this chapter demonstrate that various reconstruction settings have different impacts on PET radiomic features. For example, GLRLM - run length

variance was found to have moderate stability against the number of OSEM subsets and to have instability with changing FWHM of Gaussian filter. In contrast, GLCM - dissimilarity and GLDZM - small distance emphasis were found to be stable with changing all of the reconstruction settings. Whilst the majority of PET radiomic features were stable against various reconstruction protocols, there were several features that were affected when reconstruction settings were varied. The results presented in this chapter suggest that these sensitive features should not be used in the development of models based on radiomics analysis.

Generally, features were found to be more sensitive against FWHM Gaussian filters than other reconstruction parameters. A possible explanation for this might be that increasing FWHM will lead to improved signal to noise and reduce the spatial resolution resulting in smoothing of the intensity distribution.

This experiment was repeated twice under similar conditions and protocols to assess the agreement between the three measurements. Percent agreement was calculated by comparing the variation of each feature with different reconstruction settings. This further analysis demonstrated that the degree of agreement among repeated experiments was high (80% or greater) for nearly three quarters of the tested radiomic features. As an example, GLCM-Joint average was found to have

low COV (high stability) on all of the three measurements. Therefore, the agreement was 100% for this feature. About one quarter of features exhibited low agreement (66% or less) when the impact of various reconstruction settings was evaluated repeatedly. A possible explanation for this disagreement between categories of features may be caused by the fact that COV values are very close to classification thresholds. For instance, joint maximum from GLCM was found to have poor stability against FWHM Gaussian filters in the first experiment as its COV was 19.8%. The same feature was categorized as unstable in the third experiment because its COV was 20.7%. Thus, GLCM showed disagreement between experiments even though the difference in COV was small due to lying adjacent to the classification boundary.

The presented results demonstrate that radiomic features listed in table 8.1 can be considered as robust features for two reasons. First, these radiomic features were found to have a very high stability against all reconstruction settings. Second, they exhibit high percent agreement (nearly the same COV at all of the three measurements) when experiment are repeated.

Table 8.1: List of robust (low variation and high percent agreement) radiomic features.

Features Group	Features	Features Group	Features
GLCM	Joint average Joint variance Joint entropy Difference average Difference entropy Sum average Sum variance Sum entropy Contrast Dissimilarity Inverse difference Inverse difference normalised Inverse difference moment Inverse difference moment normalised (IDMN) Inverse variance Correlation Autocorrelation Cluster tendency Cluster prominence Second measure of information correlation (SMIC)	**GLSZM**	Small zone emphasis High grey level zone emphasis (HGLZE) Small zone high grey level emphasis (SZHGLE) Grey level non-uniformity Grey level non-uniformity normalised (GLNUN) Zone size non-uniformity normalised (ZSNUN) Zone percentage Grey level variance Zone size entropy
GLRLM	Short runs emphasis Long runs emphasis High grey level run emphasis Short run high grey level emphasis (SRHGLE) Grey level nonuniformity Run length non-uniformity Run length non-uniformity normalised (RLNUN) Run percentage Grey level variance Run entropy	**GLDZM**	Small distance emphasis Large distance emphasis High grey level zone emphasis (HGLZE) Small distance high grey level emphasis (SDHGLE) Grey level non-uniformity Grey level non-uniformity normalised (GLNUN) Zone distance non-uniformity Zone distance non-uniformity normalised (ZDNUN) Zone percentage Grey level variance Zone distance entropy
NGTDM	Complexity Strength		

The aim of the Chapter 6 study was to assess the stability of PET image radiomic features with varying reconstruction settings involving different configurations of phantom inserts. This experiment was designed to mimic non uniform distributions. A phantom study to assess the stability of radiomic features removes the complexities of physiologically induced confounding variables if the analysis is performed *in vivo*. Thus, this phantom experiment may be considered analogous to exploring the effects of changing reconstruction settings on the stability of clinically measured radiomic features.

The results of this study indicate that the impact of reconstruction settings on a radiomic feature depends upon the feature in question. For instance, GLCM (Difference entropy, Inverse difference normalised), GLRL (Short runs emphasis, Run entropy), GLSZM (Zone size entropy), GLDZM (Zone distance entropy) were stable against all reconstruction settings, while GLRL (Long run low grey level emphasis) was unstable against most reconstruction settings. NGTDM (Busyness) was moderately stable against subsets and had poor stability against FWHM Gaussian filters.

We have shown that TOF had the lowest impact on radiomic features. TOF improves image quality by locating the position of the annihilation event along the

line of response, improving both spatial resolution and image contrast. This study's results can be explained by the fact that TOF reconstruction produces better (compared to non TOF) contrast for small objects (< 2 cm). Nevertheless, both (TOF and non TOF) produce the same contrast for large objects (> 2 cm) [93]. In our study, the majority of the inserts were bigger than 2 cm.

Applying OSEM methodology enables the reconstruction of PET images to be accelerated. Nevertheless, there is a trade-off between the number of subsets and increasing noise and image quality. There is also a trade-off between the number of iterations and image noise. As the number of iterations increases, image noise will increase [94]. The number of iterations and subsets were found to have quite similar effects on all measured radiomic features. This may be due to the fact that OSEM reconstructions with n iterations and m subsets are equivalent to m iterations and n subsets and increasing any of them will increase sub iterations (subset$\times$iteration) which eventually results in elevated noise level. The largest variation in image features occurred when varying the FWHM of the Gaussian filter. A possible explanation for this might be that smoothing using the Gaussian filter (FWHM) improves signal to noise and decreases spatial resolution, consequently increasing uniformity on the intensity distribution.

This study differs from prior works in several ways. More features, number of lesion configurations, heterogeneity activity levels and reconstruction parameters were used. The present findings are consistent with other research which found that varying reconstruction settings has variable influence on the stability of different PET radiomic features. As an example, Gallivanone *et al.* assessed the impact of different reconstruction settings (i.e. Filters, Iterations and Subsets) on different radiomic features [76]. They concluded that subsets and matrix size had lowest and greatest impact on the stability of features, respectively. In our study, in comparison to Gallivanone *et al.*, about 19, 22, 17 features (from 36 common features) had the same COVs against subsets, Iterations and filter, respectively. In both studies, dissimilarity (GLCM), Short run emphasis (GLRLM), Small zone emphasis (GLSZM), strength (NGTDM) had high stability. Low grey-level run emphasis and Long run low grey-level emphasis (GLRLM), Large zone low grey-level emphasis (GLSZM) were unstable in both studies.

Doumou *et al.* studied the impact of image smoothing (FWHM Gaussian filters), segmentation and quantisation on the stability of 57 heterogeneity features [60]. For the 38 features in common with our study, 12 features had good agreement when varying the FWHM Gaussian filter. As an example, Inverse difference normalised

(GLCM) and strength (NGTDM) were stable and small zone low emphasis and large zone low emphasis (GLSZM) were unstable against varying Gaussian filter in both studies.

In a study by Shiri *et al.*, 100 radiomic features were extracted from patient and phantom images with different reconstruction settings [48]. Our results are consistent with theirs in terms of short run emphasis (GLRLM), zone percentage (GLSZM), correlation, and inverse difference moment (GLCM) having small variability against subsets and FWHM filters.

In another study, Forgacs *et al.* used inhomogeneous tumour inserts (7 syringes) placed in a cylindrical phantom and imaged with different acquisition times and reconstruction settings [62]. According to their strategy, reliable heterogeneity parameters must be volume independent, reproducible, and appropriate for detecting heterogeneity levels. Entropy, Correlation, Homogeneity and Contrast were found to have low variation with varying acquisition times and reconstruction settings [62]. In our study, 3 out of these 4 features were found to have very low COV when varying all of the tested reconstruction settings.

Bailly *et al.* assessed the robustness of 15 features with matrix size, number of iterations, Gaussian post-filtering, noise and the reconstruction algorithm [92]. For

the 13 features in common with our own study, 38% and 54% of them showed the same COVs in number of iterations and FWHM Gaussian filter, respectively. We can expect these common features that showed the same COVs to be particularly useful in multi-centre studies.

There are several causes for the differences between our results and the previous studies. First, the statistical algorithms used to analyze the results were different between studies. Second, the range of categorizations differ from one study to another. For instance, we categorized the features into 4 groups based on the COV values, unlike the Bailly study which used only 3 categorisations. This strongly supports the need for standardisation of the analysis approach, especially in clinical trials and multicentric studies.

In Chapter 6, we also performed further analysis to the findings of Sections 5.3 and 6.3. This analysis was designed to evaluate the ability of radiomic features to distinguish between different synthetic tumour inserts. In addition, this study aimed to select robust heterogeneous PET image radiomic features. For this purpose, forty (out of 78) features investigated in the previous chapters were excluded in this analysis due to their instability. The remaining features (38) were analysed using the Friedman test in order to determine whether these features differed between the

regions. The Friedman test was performed for each combination (shape1 vs shape2, shape1 vs shape3, shape1 vs shape4, shape2 vs shape3, shape2 vs shape4 and shape3 vs shape4) to determine whether or not there was a statistically significant difference between the means of regions in which the same reconstruction parameter was used in each configuration. This study demonstrated that 12 features were statistically significantly different ($p < 0.05$) between all regions. These 12 features may be reasonably considered as good features for two reasons. First, these features showed high stability as indicated in section 5.3 and 6.3. The second reason is that these features are able to distinguish between different tumour regions as mentioned in section 6.4.

In Chapter 7, we implemented self-organising maps to explore and cluster the statistical variability of PET texture parameters in the heterogeneous phantom data. A set of texture parameters (input data) were mapped onto a two dimensional grid of neurons. According to this study, the statistical distribution of input texture has clear modes depending on the contour size, consistent with the results of chapter 4. These two studies (Chapters 4 and 7) introduced the unsupervised SOM for visualisation and clustering of high dimensional data onto a lower dimensional representation. Using this method, we are able to gain insight into the radiomics

dataset and the factors contributing to radiomic distributions and clusterings for interpretation and classification.

SOM can be used in a supervised learning framework; whereby independent variables can be trained with SOM to make classifications / predictions of testing data. The X-Y fused SOM [95] has been proposed for these purposes which uses an additional grid of nodes (Ymap) to map the class information. The procedure for predicting the class membership of new inputs starts with presenting a new input vector to the network. The position of the winning unit in the Xmap is used to look up the class membership of the corresponding unit in the Ymap: the maximum value of this unit's weight vector determines the actual class membership [96]. Constructing a predictive radiomics model using this supervised SOM, would allow prediction of outcome metrics to be coupled with modelling of confounding variables and may prove to be an effective method going forward.

Several limitations need to be considered in this work. First, the current study is based on a relatively small pre-clinical dataset. Larger sample sizes would provide more reliable results with greater precision. With increasing sample size, the standard deviation of the means decreases and, consequently, the coefficient of variation may decrease as well. The future clinical study will involve 50 patients, as described

in section 8.2. In chapter 3, our study aimed to assess the influence of a combination of two factors (segmentation volume and acquisition time). We did not evaluate the impact of a combination of three or more factors. This is a topic that could be investigated in future work. In the work described in chapter 5, all of the phantom inserts were filled with a homogeneous radioactivity as it was not possible to fill these inserts with heterogeneous activity. In the later investigation (chapter 6), it was possible to use different inserts that were capable to be filled with heterogeneous activity.

The impact of interpolation, segmentation and quantization have not been taken into account in this study and these parameters may also have a substantial influence on PET radiomic features. However, we used a fixed isotropic voxel dimension, delineation and bin size with all of reconstructed images to minimize the impact of these parameters. Whybra *et al.* [74] , Leijenaar *et al.* [22] and Lu *et al.* [97] have investigated the effect of these parameters. Furthermore, physiological parameters were absent in the phantom studies and further studies with more focus on clinical data are therefore suggested.

8.2 Future research

Although the general applicability of the current results must be established by future work, the aforementioned studies offered insight into some innovative approaches for radiomics data analysis. Furthermore, the results presented in this thesis have provided a new understanding of the impact of different parameters on PET radiomic features. It would be useful to extend the mentioned approaches in this thesis to include clinical data.

We have prepared a clinical protocol which received research ethics approval and it will be developed further in the future by the PETIC Centre. The proposed research can be briefly summarised as follows:

(a) Study title: Temporal Variation of radiomic features in PET Images.

(b) IRAS project ID: 274533

(c) REC reference: 20/NE/0024

(d) Study Objectives: The main goal of this study is to investigate temporal variation of radiomic features in PET images. Extracted features of the first PET scan will be compared to those extracted from the second (Standard) PET

scan. This will form the main question to assess variability of textural features in PET images due to different acquisition time.

(e) Summary of Study Design: Participants will be of 18-65 years, with no contraindications to undergoing a longer time on PET/CT scanner. Up to 50 participants who have previously agreed to be part in this study will be asked to undergo an additional PET scan. Participants will be asked to stay on the scanner bed for about 50 minutes. Participants will be scanned twice. The first scan will include only PET scan while the second scan include PET/CT scan. After participants done the first scan (PET), they will be scanned again for the second (PET/CT) scan (requested scan). Both scans will be imported into SPAARC to extract 78 3D-radiomic features for each volume. Features include grey level co-occurrence matrix (GLCM), grey-level run-length matrix (GLRLM), grey-level size zone matrix (GLSZM), grey-level distance zone matrix (GLDZM) and neighbourhood grey-tone difference matrix (NGTDM) will be extracted. All radiomic features will be compliant with the IBSI.

In addition to the aforementioned clinical study, several artificial neural network techniques could be applied more broadly. As an example, the self-organizing map may be utilised to cluster and visualise data with three or more dimensions

(confounding variables) instead of two dimensions.

8.3 Conclusions

The studies described in this thesis aimed to:

(a) Investigate the impact of the post injection time on stability of FDG PET radiomic features and determine which radiomic features are more stable than others to use in studies that involve multiple time points acquisitions.

(b) Assess the variation of FDG PET radiomic features due to varying reconstruction settings (OSEM, number of subsets, number of iterations and TOF). This assessment aims to identify which radiomic features are more stable than others to use in studies that involve different reconstruction settings protocols.

(c) Investigate the statistical influence (impact of two combined variables) of imaging time and segmentation volume on PET radiomic features.

(d) Explore, capture and cluster the statistical variability of PET texture parameters using a self-organising map (SOM). SOM is a type of artificial neural network (ANN) that is trained using unsupervised learning to produce a lower dimensional representation of the input data on an underlying manifold.

We have studied the dependence of standardised radiomic features with a multitude of PET image reconstruction parameters and identified those that are unstable. Furthermore, this work specifically explored that some robust heterogeneous PET images radiomic features can detect and capture the differences between inserts. Despite its exploratory nature, this study offers some insight into utilising radiomic features as an adjuvant diagnostic tool along with traditional clinical imaging. These radiomic features may be able to provide prognostic information by correlating tumour heterogeneity with the detrimental biological features of increased tumour stage and patient survival. A benefit of this radiomics analysis technique is that it can be used within existing clinical practice without adding an additional burden to patients, since the technique is an additional post-processing step of standard PET images.

We have introduced the novel application of a self-organising map to radiomics analysis in PET imaging, although the general framework can be applied to all other imaging modalities. We have demonstrated its ability in identifying confounding variables that effect radiomic features. With the extension to greater suspected confounding variables such as image acquisition / reconstruction settings, image pre and post processing variables and varying segmentation algorithms we offer a

flexible method for full interrogation of the radiomics pipeline. We have postulated how the SOM may also be utilised with outcome data to serve as a predictive tool for dependent variables (e.g prognosis, therapy response). In so doing the learnt representations of self-organised features serve as the attributes for prediction which will take into consideration the statistical variability in the underlying dataset. This serves as a promising area for future work.

Bibliography

1. H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.

2. "Cancer Statistics for the UK." [Online]. Available: https://www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk{#}heading-Five

3. S. B. Edge and C. C. Compton, "The american joint committee on cancer: The 7th edition of the AJCC cancer staging manual and the future of TNM," *Annals of Surgical Oncology*, vol. 17, no. 6, pp. 1471–1474, 2010.

4. NHS, "Diagnostic Imaging Dataset Annual Statistical Release 2020/21," Performance Analysis Team, Tech. Rep., 2021. [Online]. Avail-

able: https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2021/11/Annual-Statistical-Release-2020-21-PDF-1.3MB-1.pdf

5. S. Tong, A. M. Alessio, and Paul E Kinahan, "Image reconstruction for PET/CT scanners: past achievements and future challenges," *Imaging in Medicine*, vol. 2, no. 5, pp. 529–545, 2010.

6. K. Najarian and R. Splinter, *Biomedical signal and image processing*, second edi ed. Boca Raton: CRC Press, 2012. [Online]. Available: https://www.crcpress.com/Biomedical-Signal-and-Image-Processing/Najarian-Splinter/p/book/9781439870334

7. R. L. Wahl, H. N. Wagner, and R. S. Beanlands, *Principles and practice of PET and PET/CT*. Lippincott Williams & Wilkins Philadelphia, PA:, 2009.

8. R. Splinter, "Positron emission tomography," in *Handbook of Physics in Medicine and Biology*. CRC Press, 2010, pp. 365–370.

9. R. Nutt, "The history of positron emission tomography," *Molecular Imaging & Biology*, vol. 4, no. 1, pp. 11–26, 2002.

10. T. Jones and D. Townsend, "History and future technical innovation in positron emission tomography," *Journal of Medical Imaging*, vol. 4, no. 1,

p. 011013, 2017. [Online]. Available: http://medicalimaging.spiedigitallibrary.org/article.aspx?doi=10.1117/1.JMI.4.1.011013

11. S. R. Cherry and S. S. Gambhir, "Use of positron emission tomography in animal research," *ILAR Journal*, vol. 42, no. 3, pp. 219–232, 2001.

12. N. B. Smith and A. Webb, *Introduction to medical imaging: physics, engineering and clinical applications*. Cambridge university press, 2010.

13. F. H. Fahey, "Data acquisition in PET imaging," *Journal of Nuclear Medicine Technology*, vol. 30, no. 2, pp. 39–49, 2002.

14. A. Alessio and P. Kinahan, "PET image reconstruction," 2006.

15. Charliex, "Iterative Reconstruction," 2018. [Online]. Available: https://zhuanlan.zhihu.com/p/46207080

16. W. R. Leo, "Techniques for nuclear and particle physics experiments," *Nucl Instrum Methods Phys Res*, vol. 834, p. 290, 1988.

17. R. C. Gonzalez and R. E. Woods, *4TH EDITION Digital image processing*, 2018.

18. M. Onken, M. Eichelberg, J. Riesmeier, and P. Jensch, "Digital imaging and

communications in medicine," in *Biomedical Image Processing*. Springer, 2010, pp. 427–454.

19. G. J. R. Cook, M. Siddique, B. P. Taylor, C. Yip, S. Chicklore, and V. Goh, "Radiomics in PET : principles and applications," *NeuroImage*, pp. 269–276, 2014.

20. V. Parekh and M. A. Jacobs, "Radiomics: a new application from established techniques," *Expert Review of Precision Medicine and Drug Development*, vol. 1, no. 2, pp. 207–226, 2016.

21. I. E. Naqa, P. Grigsby, A. Apte, E. Kidd, E. Donnelyy, W. Thorstad, and J. Deasy, "Exploring feature-based approaches in PET images for predicting cancer treatment outcomes," *Pattern Recogn*, vol. 42, no. 6, pp. 1162–1171, 2009.

22. R. T. H. Leijenaar, S. Carvalho, E. R. Velazquez, W. J. C. Van Elmpt, H. O. Parmar C, O. S. Hoekstra, C. J. Hoekstra, R. Boellaard, A. L. Dekker, R. J. Gillies, H. J. Aerts, and P. Lambin, "Stability of FDG-PET Radiomics features: An integrated analysis of test-retest and inter-observer variability," *Acta Oncologica*, vol. 52, no. 7, pp. 1391–1397, 2013.

23. B. D. W. Group, A. J. Atkinson Jr, W. A. Colburn, V. G. DeGruttola, D. L. DeMets, G. J. Downing, D. F. Hoth, J. A. Oates, C. C. Peck, R. T. Schooley *et al.*, "Biomarkers and surrogate endpoints: preferred definitions and conceptual framework," *Clinical pharmacology & therapeutics*, vol. 69, no. 3, pp. 89–95, 2001.

24. W. Phillip Law and K. A. Miles, "Incorporating prognostic imaging biomarkers into clinical practice," *Cancer Imaging*, vol. 13, no. 3, pp. 332–341, 2013.

25. R. J. Gillies, P. E. Kinahan, and H. Hricak, "radiomics: Images Are More than Pictures, They Are Data," *Radiology*, vol. 278, no. 2, 2016.

26. V. Parekh and M. A. Jacobs, "Radiomics: a new application from established techniques," *Expert Review of Precision Medicine and Drug Development*, 2016.

27. R. T. Larue, G. Defraene, D. De Ruysscher, P. Lambin, and W. Van Elmpt, "Quantitative radiomics studies for tissue characterization: A review of technology and methodological procedures," *British Journal of Radiology*, vol. 90, no. 1070, 2017.

28. L. Papp, N. Pötsch, M. Grahovac, V. Schmidbauer, A. Woehrer, M. Preusser, M. Mitterhauser, B. Kiesel, W. Wadsak, T. Beyer *et al.*, "Glioma survival

prediction with combined analysis of in vivo 11c-met pet features, ex vivo features, and patient features by supervised machine learning," *Journal of Nuclear Medicine*, vol. 59, no. 6, pp. 892–899, 2018.

29. J. Pérez-Beteta, D. Molina-García, J. A. Ortiz-Alhambra, A. Fernández-Romero, B. Luque, E. Arregui, M. Calvo, J. M. Borrás, B. Meléndez, A. Rodriguez de Lope *et al.*, "Tumor surface regularity at mr imaging predicts survival and response to surgery in patients with glioblastoma," *Radiology*, vol. 288, no. 1, pp. 218–225, 2018.

30. M. Crispin-Ortuzar, A. Apte, M. Grkovski, J. H. Oh, N. Y. Lee, H. Schöder, J. L. Humm, and J. O. Deasy, "Predicting hypoxia status using a combination of contrast-enhanced computed tomography and [18f]-fluorodeoxyglucose positron emission tomography radiomics features," *Radiotherapy and Oncology*, vol. 127, no. 1, pp. 36–42, 2018.

31. L. Antunovic, F. Gallivanone, M. Sollini, A. Sagona, A. Invento, G. Manfrinato, M. Kirienko, C. Tinterri, A. Chiti, and I. Castiglioni, "[18 f] fdg pet/ct features for the molecular characterization of primary breast tumors," *European journal of nuclear medicine and molecular imaging*, vol. 44, no. 12, pp. 1945–1954, 2017.

32. S. Ha, S. Park, J.-I. Bang, E.-K. Kim, and H.-Y. Lee, "Metabolic radiomics for pretreatment 18 F-FDG PET/CT to characterize locally advanced breast cancer: histopathologic characteristics, response to neoadjuvant chemotherapy, and prognosis," *Scientific Reports*, vol. 7, no. 1, pp. 1–11, 2017.

33. J. Wu, T. Aguilera, D. Shultz, M. Gudur, D. L. Rubin, B. W. Loo Jr, M. Diehn, and R. Li, "Early-stage non–small cell lung cancer: quantitative imaging characteristics of 18f fluorodeoxyglucose pet/ct allow prediction of distant metastasis," *Radiology*, vol. 281, no. 1, pp. 270–278, 2016.

34. G. J. Cook, M. E. O'Brien, M. Siddique, S. Chicklore, H. Y. Loi, B. Sharma, R. Punwani, P. Bassett, V. Goh, and S. Chua, "Non–small cell lung cancer treated with erlotinib: heterogeneity of 18f-fdg uptake at pet—association with treatment response and prognosis," *Radiology*, vol. 276, no. 3, pp. 883–893, 2015.

35. M. Maas, P. J. Nelemans, V. Valentini, P. Das, C. Rödel, L.-J. Kuo, F. A. Calvo, J. García-Aguilar, R. Glynne-Jones, K. Haustermans *et al.*, "Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data," *The Lancet*

Oncology, vol. 11, no. 9, pp. 835–844, 2010.

36. S. Rizzo, F. Botta, S. Raimondi, D. Origgi, C. Fanciullo, A. G. Morganti, and M. Bellomi, "Radiomics: the facts and the challenges of image analysis," *European Radiology Experimental*, vol. 2, no. 1, 2018.

37. A. Materka, M. Strzelecki *et al.*, "Texture analysis methods–a review," *Technical university of lodz, institute of electronics, COST B11 report, Brussels*, vol. 10, no. 1.97, p. 4968, 1998.

38. MathWorks, "Create a Gray-Level Co-Occurrence Matrix," 2022. [Online]. Available: https://uk.mathworks.com/help/images/create-a-gray-level-co-occurrence-matrix.html

39. M. M. Galloway, "Texture analysis using gray level run lengths," *Computer graphics and image processing*, vol. 4, no. 2, pp. 172–179, 1975.

40. J. Benrabha and F. Meziane, "Automatic roi detection and classification of the achilles tendon ultrasound images," in *Proceedings of the 1st International Conference on Internet of Things and Machine Learning*, 2017, pp. 1–7.

41. G. Thibault, J. Angulo, and F. Meyer, "Advanced statistical matrices for tex-

ture characterization: application to cell classification," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 3, pp. 630–637, 2013.

42. S.-H. Lee and H. Park, "Difference of alzheimer's disease sub-groups using two features from intensity size zone matrix," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 3020–3023.

43. 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.

44. M. Hatt, F. Tixier, L. Pierce, P. E. Kinahan, C. C. Le Rest, and D. Visvikis, "Characterization of PET/CT images using texture analysis: the past, the present... any future?" *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 44, no. 1, pp. 151–165, 2017. [Online]. Available: http://dx.doi.org/10.1007/s00259-016-3427-0

45. G. J. Cook, G. Azad, K. Owczarczyk, M. Siddique, and V. Goh, "Challenges and Promises of PET Radiomics," *International Journal of*

Radiation Oncology Biology Physics, pp. 1–7, 2018. [Online]. Available: https://doi.org/10.1016/j.ijrobp.2017.12.268

46. P. Galavis, C. Hollensen, N. Jallow, and B. P. Al., "Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters," *Acta Oncologica*, vol. 49, no. 7, pp. 12–22, 2010.

47. J. A. Oliver, M. Budzevich, G. G. Zhang, T. J. Dilling, K. Latifi, and E. G. Moros, "Variability of image features computed from conventional and respiratory-gated PET/CT images of lung cancer," *Translational Oncology*, vol. 8, no. 6, pp. 524–534, 2015. [Online]. Available: http://dx.doi.org/10.1016/j.tranon.2015.11.013

48. I. Shiri, A. Rahmim, P. Ghaffarian, P. Geramifar, H. Abdollahi, and A. Bitarafan-Rajabi, "The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies," *European Radiology*, vol. 27, no. 11, pp. 4498–4509, 2017.

49. E. Alsyed, R. Smith, C. Marshall, S. Paisey, and E. Spezi, "Stability of PET Radiomic Features: A Preclinical Study," in *European Journal of Nuclear*

Medicine and Molecular Imaging. NEW YORK, NY: SPRINGER, 2019, p. S759.

50. S. S. Yip and H. J. Aerts, "Applications and limitations of radiomics," *Physics in Medicine and Biology*, vol. 61, no. 13, pp. R150–R166, 2016.

51. A. Traverso, L. Wee, A. Dekker, and R. Gillies, "Repeatability and reproducibility of radiomic features: a systematic review," *International Journal of Radiation Oncology - Biology - Physics*, vol. 102, no. 4, pp. 1143–1158, 2018.

52. M. Hatt, C. Cheze Le Rest, N. Antonorsi, F. Tixier, O. Tankyevych, V. Jaouen, F. Lucia, V. Bourbonne, U. Schick, B. Badic, and D. Visvikis, "Radiomics in PET/CT: Current Status and Future AI-Based Evolutions," *Seminars in Nuclear Medicine*, vol. 51, no. 2, pp. 126–133, 2021.

53. E. Alsyed, R. Smith, C. Marshall, S. Paisey, and E. Spezi, "The Statistical Influence of Imaging Time and Segmentation Volume on PET Radiomic Features: A Preclinical Study," in *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2019*, 2019.

54. E. Alsyed, R. Smith, S. Paisey, C. Marshall, and E. Spezi, "A Self Organizing Map for Exploratory Analysis of PET Radiomic Features," in *2020 IEEE*

Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2020, 2021.

55. J. Yan, J. L. Chu-Shern, H. Y. Loi, L. K. Khor, A. K. Sinha, S. T. Quek, I. W. K. Tham, and D. Townsend, "Impact of Image Reconstruction Settings on Texture Features in 18F-FDG PET," *Journal of Nuclear Medicine*, vol. 56, no. 11, pp. 1667–1673, 2015.

56. F. Orlhac, C. Nioche, M. Soussan, and I. Buvat, "Understanding changes in tumor texture indices in pet: a comparison between visual assessment and index values in simulated and patient data," *Journal of Nuclear Medicine*, vol. 58, no. 3, pp. 387–392, 2017.

57. C. Lasnon, M. Majdoub, B. Lavigne, P. Do, J. Madelaine, D. Visvikis, M. Hatt, and N. Aide, "18F-FDG PET/CT heterogeneity quantification through textural features in the era of harmonisation programs: a focus on lung cancer," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 43, no. 13, pp. 2324–2335, 2016.

58. F. H. P. V. Velden, G. M. Kramer, V. Frings, I. A. Nissen, E. R. Mulder, A. J. D. Langen, O. S. Hoekstra, E. F. Smit, and R. Boellaard, "Repeatability

of Radiomic Features in Non-Small-Cell Lung Cancer [18 F] FDG-PET / CT Studies : Impact of Reconstruction and Delineation," *Molecular Imaging and Biology*, no. February, pp. 788–795, 2016.

59. R. Boellaard, R. Delgado-Bolton, W. J. Oyen, F. Giammarile, K. Tatsch, W. Eschner, F. J. Verzijlbergen, S. F. Barrington, L. C. Pike, W. A. Weber *et al.*, "Fdg pet/ct: Eanm procedure guidelines for tumour imaging: version 2.0," *European journal of nuclear medicine and molecular imaging*, vol. 42, no. 2, pp. 328–354, 2015.

60. G. Doumou, M. Siddique, C. Tsoumpas, V. Goh, and G. J. Cook, "The precision of textural analysis in18F-FDG-PET scans of oesophageal cancer," *European Radiology*, vol. 25, no. 9, pp. 2805–2812, 2015.

61. W. Grootjans, F. Tixier, C. S. van der Vos, D. Vriens, C. C. Le Rest, J. Bussink, W. J. Oyen, L.-F. de Geus-Oei, D. Visvikis, and E. P. Visser, "The impact of optimal respiratory gating and image noise on evaluation of intratumor heterogeneity on 18F-FDG PET imaging of lung cancer," *Journal of nuclear medicine*, vol. 57, no. 11, pp. 1692–1698, 2016.

62. A. Forgacs, H. Pall Jonsson, M. Dahlbom, F. Daver, M. D. Difranco, G. Op-

posits, A. K. Krizsan, I. Garai, J. Czernin, J. Varga, L. Tron, and L. Balkay, "A study on the basic criteria for selecting heterogeneity parameters of F18-FDG PET images," *PLoS ONE*, vol. 11, no. 10, pp. 1–14, 2016.

63. E. Pfaehler, J. R. De Jong, R. A. J. O. Dierckx, F. H. P. van Velden, and R. Boellaard, "SMART (SiMulAtion and ReconsTruction) PET: an efficient PET simulation-reconstruction tool," *EJNMMI Physics*, vol. 5, no. 1, p. 16, 2018.

64. E. Alsyed, R. Smith, S. Paisey, C. Marshall, C. Parkinson, P. Whybra, and E. Spezi, "The Impact of Varying Number of OSEM Subsets on PET Radiomic Features:A Preclinical Study," in *Radiotherapy and Oncology*, 2020, p. 152: S849.

65. E. Alsyed, R. Smith, L. Bartley, C. Marshall, and E. Spezi, "The Efect of Increasing the Number of Iterations on the Stability of PET Radiomic Features: A Phantom Study," in *European Journal of Nuclear Medicine and Molecular Imaging*. NEW YORK, NY: Springer, 2021, pp. S510–S511.

66. E. Alsyed, R. Smith, L. Bartly, C. Marshall, and E. Spezi, "Phantom with Heterogenous Tumour Inserts to Explore the Impact of Varying Number of

OSEM Subsets on PET Radiomic Features," in *European Journal of Nuclear Medicine and Molecular Imaging*. NEW YORK, NY: Springer, 2021, p. S511.

67. E. Alsyed, R. Smith, C. Marshall, L. Bartley, and E. Spezi, "Toward a method of selecting robust heterogeneous PET images radiomic features," in *2022 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2022*, "(in press)".

68. E. Alsyed, R. Smith, L. Bartley, C. Marshall, and E. Spezi, "Artificial Neural Network Algorithm to Cluster and Visualize Phantom Experiment Data," in *2021 IEEE Nuclear Science Symposium and Medical Imaging Conference, NSS/MIC 2021*, 2022 "(in press)".

69. C. Kuntner and D. Stout, "Quantitative preclinical PET imaging: Opportunities and challenges," *Frontiers in Physics*, vol. 2, no. February, pp. 1–12, 2014.

70. Y. S. DeRose, G. Wang, Y.-C. Lin, P. S. Bernard, S. S. Buys, M. T. Ebbert, R. Factor, C. Matsen, B. A. Milash, E. Nelson *et al.*, "Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes," *Nature Medicine*, vol. 17, no. 11, pp. 1514–1520, 2011.

71. S. Roy, T. D. Whitehead, S. Li, F. O. Ademuyiwa, R. L. Wahl, F. Dehdashti, and K. I. Shoghi, "Co-clinical FDG-PET radiomic signature in predicting response to neoadjuvant chemotherapy in triple-negative breast cancer," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 49, no. 2, pp. 550–562, 2022.

72. Operator's Manual, "In-vivo Dual Modality PET / CT Imager O perator ' s Manual," Medical Imaging Systems, Tech. Rep., 2015.

73. C. Müller, A. Singh, C. A. Umbricht, H. R. Kulkarni, K. Johnston, M. Benešová, S. Senftleben, D. Müller, C. Vermeulen, R. Schibli, U. Köster, N. P. van der Meulen, and R. P. Baum, "Preclinical investigations and first-in-human application of 152Tb-PSMA-617 for PET/CT imaging of prostate cancer," *EJNMMI Research*, vol. 9, no. 1, pp. 1–10, 2019.

74. P. Whybra, C. Parkinson, and K. Foley, "Assessing radiomic feature robustness to interpolation in F-FDG PET imaging," *Scientific Reports*, no. December, pp. 0–10, 2019.

75. A. Zwanenburg, M. Valli, and L. Steffen, "image biomarker standardisation initiative," vol. 1, 2016. [Online]. Available: https://arxiv.org/abs/1612.07003

76. F. Gallivanone, M. Interlenghi, D. D. Ambrosio, I. Castiglioni, and G. Trifir, "Parameters Influencing PET Imaging Features : A Phantom Study with Irregular and Heterogeneous Synthetic Lesions," vol. 2018, 2018.

77. R. Smith, S. Paisey, N. Evans, V. Florence, E. Fittock, F. Siebzehnrubl, and C. Marshall, "Deep Learning Pre-Clinical Medical Image Segmentation for Automated Organ-Wise Delineation of PET," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 45, no. 1, p. S290, 2018.

78. R. L. Smith, N. Evans, V. Florence, S. Paisey, E. Fittock, F. Siebzehnrubl, and C. Marshall, "Reinforcement Learning for Automated PET Image Segmentation," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 46, no. 1, p. S952, 2019.

79. R. Forghani, P. Savadjiev, A. Chatterjee, N. Muthukrishnan, C. Reinhold, and B. Forghani, "Radiomics and Artificial Intelligence for Biomarker and Prediction Model Development in Oncology," *Computational and Structural Biotechnology Journal*, vol. 17, pp. 995–1008, 2019.

80. A. Zwanenburg, M. Vallières, M. A. Abdalah, H. J. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R. J. Beukinga, R. Boellaard, M. Bogow-

icz, L. Boldrini, I. Buvat, G. J. Cook, C. Davatzikos, A. Depeursinge, M. C. Desseroit, N. Dinapoli, C. V. Dinh, S. Echegaray, I. El Naqa, A. Y. Fedorov, R. Gatta, R. J. Gillies, V. Goh, M. Götz, M. Guckenberger, S. M. Ha, M. Hatt, F. Isensee, P. Lambin, S. Leger, R. T. Leijenaar, J. Lenkowicz, F. Lippert, A. Losnegård, K. H. Maier-Hein, O. Morin, H. Müller, S. Napel, C. Nioche, F. Orlhac, S. Pati, E. A. Pfaehler, A. Rahmim, A. U. Rao, J. Scherer, M. M. Siddique, N. M. Sijtsema, J. Socarras Fernandez, E. Spezi, R. J. Steenbakkers, S. Tanadini-Lang, D. Thorwarth, E. G. Troost, T. Upadhaya, V. Valentini, L. V. van Dijk, J. van Griethuysen, F. H. van Velden, P. Whybra, C. Richter, and S. Löck, "The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping," *Radiology*, vol. 295, no. 2, pp. 328–338, 2020.

81. A. G. Asuero, A. Sayago, and A. G. Gonz, "The Correlation Coefficient : An Overview The Correlation Coefficient : An Overview," *Critical Reviews in Analytical Chemistry*, vol. 36, no. January, 2006.

82. I. Ackerley, R. Smith, J. Scuffham, M. Halling-Brown, E. Lewis, and E. Spezi, "Using deep learning to detect esophageal lesions in PET-CT scans," in *Biomedical Applications in Molecular, Structural, and Functional Imaging, SPIE*, 2019,

pp. pp–138–146.

83. S. Kadoury, "Manifold learning in medical imaging," in *Manifolds II-Theory and Applications*. IntechOpen, 2018.

84. T. Kohonen, "Self-organized formation of topologically correct feature maps," *Biological Cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.

85. M. Köküer, R. N. G. Naguib, P. Janc, H. B. Younghusband, and R. Green, "Towards Automatic Risk Analysis for Hereditary Non-Polyposis Colorectal Cancer Based on Pedigree Data," pp. 319–337, 2007.

86. National Electrical Manufacturers Association, *NEMA standards publication : stabilized power supplies direct-current output*. New York : The Association, [1972] ©1972. [Online]. Available: https://search.library.wisc.edu/catalog/999916106402121

87. L. Presotto, V. Bettinardi, E. De Bernardi, M. L. Belli, G. M. Cattaneo, S. Broggi, and C. Fiorino, "PET textural features stability and pattern discrimination power for radiomics analysis: An "ad-hoc" phantoms study," *Physica Medica*, vol. 50, no. February, pp. 66–74, 2018.

88. M. Friedman, "The Use of Ranks to Avoid the Assumption of Normality Im-

plicit in the Analysis of Variance," *Journal of the American Statistical Association*, vol. 32, no. 200, pp. 675–701, dec 1937.

89. F. Milton, "A correction: The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the American Statistical Association. American Statistical Association*, vol. 34, no. 205, 1939.

90. A. Ibrahim, S. Primakov, M. Beuque, H. Woodruff, I. Halilaj, G. Wu, T. Refaee, R. Granzier, Y. Widaatalla, R. Hustinx *et al.*, "Radiomics for precision medicine: Current challenges, future prospects, and the proposal of a new framework," *Methods*, vol. 188, pp. 20–29, 2021.

91. I. Buvat, F. Orlhac, and M. Soussan, "Tumor Texture Analysis in PET: Where Do We Stand?" *Journal of Nuclear Medicine*, vol. 56, no. 11, pp. 1642–1644, 2015.

92. C. Bailly, C. Bodet-Milin, S. Couespel, H. Necib, F. Kraeber-Bodéré, C. Ansquer, and T. Carlier, "Revisiting the robustness of PET-based textural features in the context of multi-centric trials," *PLoS ONE*, vol. 11, no. 7, pp. 1–16, 2016.

93. I. Murray, A. Kalemis, J. Glennon, S. Hasan, S. Quraishi, T. Beyer, and N. Avril, "Time-of-flight pet/ct using low-activity protocols: potential impli-

cations for cancer therapy monitoring," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 37, no. 9, pp. 1643–1653, 2010.

94. P. R. Caribé, M. Koole, Y. D'Asseler, B. Van Den Broeck, and S. Vandenberghe, "Noise reduction using a Bayesian penalized-likelihood reconstruction algorithm on a time-of-flight PET-CT scanner," *EJNMMI Physics*, vol. 6, no. 1, 2019.

95. W. Melssen, R. Wehrens, and L. Buydens, "Supervised Kohonen networks for classification problems," *Chemometrics and Intelligent Laboratory Systems*, vol. 83, no. 2, pp. 99–113, 2006.

96. B. Arnrich, C. Kappeler-Setz, G. Tröster, R. La Marca, and U. Ehlert, "Self Organizing Maps for Affective State Detection," *International Workshop on Machine Learning for Assistive Technologies*, pp. 1–8, 2010.

97. L. Lu, W. Lv, J. Jiang, J. Ma, Q. Feng, A. Rahmim, and W. Chen, "Robustness of Radiomic Features in [11C]Choline and [18F]FDG PET/CT Imaging of Nasopharyngeal Carcinoma: Impact of Segmentation and Discretization," *Molecular Imaging and Biology*, vol. 18, no. 6, pp. 935–945, 2016.

Appendix A

Pre-clinical studies

A.1 Post injection imaging time

Table A.1: Coefficient of variation (COV) value for each extracted feature against post injection imaging time. Features were classified based on their COV values. 1= (COV $\leq$ 5%), 2 = (5% $>$ COV $\leq$ 10%), 3 = (10% $>$ COV $\leq$ 20%) and 4 = (COV $>$ 20%).

Feature name	COV value	Category
GLCM - Joint maximum	5.351367	2
GLCM - Joint average	1.856867	1
GLCM - Joint variance	4.021925	1
GLCM - Joint entropy	0.911813	1
GLCM - Difference average	3.019809	1
GLCM - Difference variance	4.854791	1
GLCM - Difference entropy	1.12164	1
GLCM - Sum average	1.856966	1
GLCM - Sum variance	3.810228	1
GLCM - Sum entropy	0.620974	1
GLCM - Angular second moment	2.131694	1
GLCM - Contrast	5.85649	2
GLCM - Dissimilarity	3.019809	1
GLCM - Inverse difference	1.149947	1
GLCM - Inverse difference normalised	0.178889	1
GLCM - Inverse difference moment	1.570165	1
GLCM - Inverse difference moment normalised (IDMN)	0.060077	1
GLCM - Inverse variance	1.591567	1
GLCM - Correlation	0.304526	1
GLCM - Autocorrelation	3.691889	1
GLCM - Cluster tendency	3.810228	1

Continued on next page

Table A.1 – continued from previous page

Feature name	COV value	Category
GLCM - Cluster shade	3.938975	1
GLCM - Cluster prominence	5.07346	2
GLCM - First measure of information correlation (FMIC)	1.133702	1
GLCM - Second measure of information correlation (SMIC)	0.299028	1
GLRLM - Short runs emphasis	0.298638	1
GLRLM - Long runs emphasis	0.518768	1
GLRLM - Low grey level run emphasis	2.10364	1
GLRLM - High grey level run emphasis	3.545783	1
GLRLM - Short run low grey level emphasis (SRLGLE)	1.971981	1
GLRLM - Short run high grey level emphasis (SRHGLE)	3.784629	1
GLRLM - Long run low grey level emphasis (LRLGLE)	6.409901	2
GLRLM - Long run high grey level emphasis (LRHGLE)	2.657251	1
GLRLM - Grey level nonuniformity	0.566826	1
GLRLM - Grey level non-uniformity normalised (GLNUN)	2.000713	1
GLRLM - Run length non-uniformity	0.641193	1
GLRLM - Run length non-uniformity normalised (RLNUN)	0.72873	1
GLRLM - Run percentage	0.396673	1
GLRLM - Grey level variance	4.017431	1
GLRLM - Run length variance	1.266929	1
GLRLM - Run entropy	0.419372	1
GLSZM - Small zone emphasis	0.94284	1
GLSZM - Large zone emphasis	9.93332	2
GLSZM - Low grey level zone emphasis (LGLZE)	8.360394	2
GLSZM - High grey level zone emphasis (HGLZE)	2.941298	1
GLSZM - Small zone low grey level emphasis (SZLGLE)	10.04404	3
GLSZM - Small zone high grey level emphasis (SZHGLE)	4.425984	1
GLSZM - Large zone low grey level emphasis (LZLGLE)	28.20306	4
GLSZM - Large zone high grey level emphasis (LZHGLE)	1.840922	1
GLSZM - Grey level non-uniformity	2.911037	1
GLSZM - Grey level non-uniformity normalised (GLNUN)	1.509611	1
GLSZM - Zone size nonuniformity	4.976107	1
GLSZM - Zone size non-uniformity normalised (ZSNUN)	1.770555	1
GLSZM - Zone percentage	6.433879	2
GLSZM - Grey level variance	2.86255	1
GLSZM - Zone size variance	10.0361	3
GLSZM - Zone size entropy	0.458011	1
GLDZM - Small distance emphasis	3.048988	1
GLDZM - Large distance emphasis	2.691578	1
GLDZM - Low grey level zone emphasis (LGLZE)	8.360394	2

Continued on next page

Table A.1 – continued from previous page

Feature name	COV value	Category
GLDZM - High grey level zone emphasis (HGLZE)	2.941298	1
GLDZM - Small distance low grey level emphasis (SDLGLE)	9.787618	2
GLDZM - Small distance high grey level emphasis (SDHGLE)	4.643458	1
GLDZM - Large distance low grey level emphasis (LDLGLE)	4.975588	1
GLDZM - Large distance high grey level emphasis (LDHGLE)	4.790812	1
GLDZM - Grey level non-uniformity	2.911037	1
GLDZM - Grey level non-uniformity normalised (GLNUN)	1.509611	1
GLDZM - Zone distance non-uniformity	4.853447	1
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	3.483869	1
GLDZM - Zone percentage	6.433879	2
GLDZM - Grey level variance	2.86255	1
GLDZM - Zone distance variance	1.903489	1
GLDZM - Zone distance entropy	0.881776	1
NGTDM - Coarseness	0.858711	1
NGTDM - Contrast	7.417827	2
NGTDM - Busyness	2.563067	1
NGTDM - Complexity	2.916594	1
NGTDM - Strength	4.059501	1

A.2 Numper of Subsets

Table A.2: Coefficient of variation (COV) value for each extracted feature against number of subsets. Features were classified based on their COV values. 1= (COV $\leq$ 5%), 2 = (5% > COV $\leq$ 10%), 3 = (10% > COV $\leq$ 20%) and 4 = (COV > 20%).

Feature name	COV value	Category
GLCM - Joint maximum	40.31893	4
GLCM - Joint average	9.909627	2
GLCM - Joint variance	8.824109	2
GLCM - Joint entropy	6.585163	2
GLCM - Difference average	23.07261	4
GLCM - Difference variance	40.51305	4
GLCM - Difference entropy	14.08076	3
GLCM - Sum average	9.909515	2
GLCM - Sum variance	8.685168	2
GLCM - Sum entropy	2.330588	1
GLCM - Angular second moment	45.57512	4
GLCM - Contrast	38.13272	4
GLCM - Dissimilarity	23.07261	4
GLCM - Inverse difference	10.89228	3
GLCM - Inverse difference normalised	1.0829	1
GLCM - Inverse difference moment	14.50522	3
GLCM - Inverse difference moment normalised (IDMN)	0.234887	1
GLCM - Inverse variance	5.068755	2
GLCM - Correlation	6.625914	2
GLCM - Autocorrelation	17.86604	3
GLCM - Cluster tendency	8.685168	2
GLCM - Cluster shade	12.97639	3
GLCM - Cluster prominence	16.51833	3
GLCM - First measure of information correlation (FMIC)	26.97878	4
GLCM - Second measure of information correlation (SMIC)	3.340462	1
GLRLM - Short runs emphasis	5.278887	2
GLRLM - Long runs emphasis	32.34586	4
GLRLM - Low grey level run emphasis	26.41959	4
GLRLM - High grey level run emphasis	16.24497	3
GLRLM - Short run low grey level emphasis (SRLGLE)	25.97799	4
GLRLM - Short run high grey level emphasis (SRHGLE)	15.03374	3
GLRLM - Long run low grey level emphasis (LRLGLE)	43.85345	4
GLRLM - Long run high grey level emphasis (LRHGLE)	26.98951	4
GLRLM - Grey level nonuniformity	7.190826	2

Continued on next page

Table A.2 – continued from previous page

Feature name	COV value	Category
GLRLM - Grey level non-uniformity normalised (GLNUN)	5.121644	2
GLRLM - Run length non-uniformity	14.71088	3
GLRLM - Run length non-uniformity normalised (RLNUN)	9.595334	2
GLRLM - Run percentage	7.656216	2
GLRLM - Grey level variance	7.572602	2
GLRLM - Run length variance	56.21081	4
GLRLM - Run entropy	3.53098	1
GLSZM - Small zone emphasis	5.12541	2
GLSZM - Large zone emphasis	96.94105	4
GLSZM - Low grey level zone emphasis (LGLZE)	14.72222	3
GLSZM - High grey level zone emphasis (HGLZE)	21.18717	4
GLSZM - Small zone low grey level emphasis (SZLGLE)	18.57148	3
GLSZM - Small zone high grey level emphasis (SZHGLE)	21.74801	4
GLSZM - Large zone low grey level emphasis (LZLGLE)	102.5122	4
GLSZM - Large zone high grey level emphasis (LZHGLE)	88.55754	4
GLSZM - Grey level non-uniformity	43.47209	4
GLSZM - Grey level non-uniformity normalised (GLNUN)	8.614048	2
GLSZM - Zone size nonuniformity	43.91931	4
GLSZM - Zone size non-uniformity normalised (ZSNUN)	9.07466	2
GLSZM - Zone percentage	35.93297	4
GLSZM - Grey level variance	18.76482	3
GLSZM - Zone size variance	94.32406	4
GLSZM - Zone size entropy	2.940183	1
GLDZM - Small distance emphasis	27.32908	4
GLDZM - Large distance emphasis	34.58217	4
GLDZM - Low grey level zone emphasis (LGLZE)	14.72222	3
GLDZM - High grey level zone emphasis (HGLZE)	21.18717	4
GLDZM - Small distance low grey level emphasis (SDLGLE)	20.21362	4
GLDZM - Small distance high grey level emphasis (SDHGLE)	71.28937	4
GLDZM - Large distance low grey level emphasis (LDLGLE)	36.2966	4
GLDZM - Large distance high grey level emphasis (LDHGLE)	25.7348	4
GLDZM - Grey level non-uniformity	43.47209	4
GLDZM - Grey level non-uniformity normalised (GLNUN)	8.614048	2
GLDZM - Zone distance non-uniformity	12.89139	3
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	42.78014	4
GLDZM - Zone percentage	35.93297	4
GLDZM - Grey level variance	18.76482	3
GLDZM - Zone distance variance	32.65053	4
GLDZM - Zone distance entropy	5.977007	2

Continued on next page

Table A.2 – continued from previous page

Feature name	COV value	Category
NGTDM - Coarseness	29.12676	4
NGTDM - Contrast	24.8294	4
NGTDM - Busyness	37.90076	4
NGTDM - Complexity	31.46056	4
NGTDM - Strength	38.60212	4

A.3 Numper of iterations

Table A.3: Coefficient of variation (COV) value for each extracted feature against number of iterations. Features were classified based on their COV values. 1= (COV $\leq$ 5%), 2 = (5% > COV $\leq$ 10%), 3 = (10% > COV $\leq$ 20%) and 4 = (COV > 20%).

Feature name	COV value	Category
GLCM - Joint maximum	23.69427	4
GLCM - Joint average	19.11076	3
GLCM - Joint variance	12.90391	3
GLCM - Joint entropy	2.726672	1
GLCM - Difference average	16.65113	3
GLCM - Difference variance	35.73916	4
GLCM - Difference entropy	9.975299	2
GLCM - Sum average	19.11083	3
GLCM - Sum variance	17.53432	3
GLCM - Sum entropy	3.467981	1
GLCM - Angular second moment	24.34011	4
GLCM - Contrast	31.96637	4
GLCM - Dissimilarity	16.65113	3
GLCM - Inverse difference	6.459106	2
GLCM - Inverse difference normalised	0.883023	1
GLCM - Inverse difference moment	8.733957	2
GLCM - Inverse difference moment normalised (IDMN)	0.265472	1
GLCM - Inverse variance	7.615901	2
GLCM - Correlation	12.41113	3
GLCM - Autocorrelation	36.77697	4

Continued on next page

Table A.3 – continued from previous page

Feature name	COV value	Category
GLCM - Cluster tendency	17.53432	3
GLCM - Cluster shade	17.26093	3
GLCM - Cluster prominence	31.54636	4
GLCM - First measure of information correlation (FMIC)	31.29976	4
GLCM - Second measure of information correlation (SMIC)	6.240388	2
GLRLM - Short runs emphasis	2.329182	1
GLRLM - Long runs emphasis	18.6276	3
GLRLM - Low grey level run emphasis	39.35766	4
GLRLM - High grey level run emphasis	30.87112	4
GLRLM - Short run low grey level emphasis (SRLGLE)	39.30487	4
GLRLM - Short run high grey level emphasis (SRHGLE)	28.42342	4
GLRLM - Long run low grey level emphasis (LRLGLE)	48.91952	4
GLRLM - Long run high grey level emphasis (LRHGLE)	44.278	4
GLRLM - Grey level nonuniformity	13.89243	3
GLRLM - Grey level non-uniformity normalised (GLNUN)	11.39704	3
GLRLM - Run length non-uniformity	4.93188	1
GLRLM - Run length non-uniformity normalised (RLNUN)	3.939353	1
GLRLM - Run percentage	3.16213	1
GLRLM - Grey level variance	14.02952	3
GLRLM - Run length variance	34.85622	4
GLRLM - Run entropy	4.470208	1
GLSZM - Small zone emphasis	7.771379	2
GLSZM - Large zone emphasis	79.16752	4
GLSZM - Low grey level zone emphasis (LGLZE)	17.44922	3
GLSZM - High grey level zone emphasis (HGLZE)	22.95842	4
GLSZM - Small zone low grey level emphasis (SZLGLE)	30.20183	4
GLSZM - Small zone high grey level emphasis (SZHGLE)	12.93995	3
GLSZM - Large zone low grey level emphasis (LZLGLE)	59.66004	4
GLSZM - Large zone high grey level emphasis (LZHGLE)	123.5193	4
GLSZM - Grey level non-uniformity	37.25809	4
GLSZM - Grey level non-uniformity normalised (GLNUN)	9.048696	2
GLSZM - Zone size nonuniformity	42.25676	4
GLSZM - Zone size non-uniformity normalised (ZSNUN)	14.03462	3
GLSZM - Zone percentage	30.56887	4
GLSZM - Grey level variance	18.64365	3
GLSZM - Zone size variance	76.30155	4
GLSZM - Zone size entropy	1.8621	1
GLDZM - Small distance emphasis	20.346	4
GLDZM - Large distance emphasis	14.831	3

Continued on next page

Table A.3 – continued from previous page

Feature name	COV value	Category
GLDZM - Low grey level zone emphasis (LGLZE)	17.44922	3
GLDZM - High grey level zone emphasis (HGLZE)	22.95842	4
GLDZM - Small distance low grey level emphasis (SDLGLE)	15.82816	3
GLDZM - Small distance high grey level emphasis (SDHGLE)	64.48661	4
GLDZM - Large distance low grey level emphasis (LDLGLE)	38.24776	4
GLDZM - Large distance high grey level emphasis (LDHGLE)	10.70088	3
GLDZM - Grey level non-uniformity	37.25809	4
GLDZM - Grey level non-uniformity normalised (GLNUN)	9.048696	2
GLDZM - Zone distance non-uniformity	18.53101	3
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	27.06049	4
GLDZM - Zone percentage	30.56887	4
GLDZM - Grey level variance	18.64365	3
GLDZM - Zone distance variance	7.29593	2
GLDZM - Zone distance entropy	2.702778	1
NGTDM - Coarseness	42.98984	4
NGTDM - Contrast	22.23246	4
NGTDM - Busyness	56.06475	4
NGTDM - Complexity	26.12305	4
NGTDM - Strength	36.5284	4

Appendix B

Statistical Influence of Imaging Time and Segmentation Contour Sizes

Table B.1: The average of determinants of the correlation matrices for each feature whilst varying segmentation contour sizes (Det-C) and post injection times (Det-T).

Feature name	Det-C	Det-T
GLCM - Joint maximum	0.0013	0.0009
GLCM - Joint average	0.0048	0.099
GLCM - Joint variance	0.0059	0.2647
GLCM - Joint entropy	0.0032	0.0829
GLCM - Difference average	0.0059	0.1248
GLCM - Difference variance	0.0312	0.3398
GLCM - Difference entropy	0.0083	0.1362
GLCM - Sum average	0.0048	0.099
GLCM - Sum variance	0.0041	0.1991
GLCM - Sum entropy	0.0044	0.0893
GLCM - Angular second moment	0.0005	0.0117
GLCM - Contrast	0.0133	0.232
GLCM - Dissimilarity	0.0059	0.1248
GLCM - Inverse difference	0.0019	0.0312
GLCM - Inverse difference normalised	0.0049	0.1062
GLCM - Inverse difference moment	0.0018	0.0308
GLCM - Inverse difference moment normalised (IDMN)	0.0121	0.2218
GLCM - Inverse variance	0.0122	0.0811
GLCM - Correlation	0.0007	0
GLCM - Autocorrelation	0.0052	0.1782
GLCM - Cluster tendency	0.0041	0.1991

Continued on next page

Table B.1 – continued from previous page

Feature name	Det-C	Det-T
GLCM - Cluster shade	0.0315	0.2052
GLCM - Cluster prominence	0.014	0.2434
GLCM - First measure of information correlation (FMIC)	0.0002	0
GLCM - Second measure of information correlation (SMIC)	0.0011	0
GLRLM - Short runs emphasis	0.0041	0.0516
GLRLM - Long runs emphasis	0.0005	0.0014
GLRLM - Low grey level run emphasis	0.0051	0.073
GLRLM - High grey level run emphasis	0.0109	0.2872
GLRLM - Short run low grey level emphasis (SRLGLE)	0.0207	0.0505
GLRLM - Short run high grey level emphasis (SRHGLE)	0.0118	0.3036
GLRLM - Long run low grey level emphasis (LRLGLE)	0.0001	0.0002
GLRLM - Long run high grey level emphasis (LRHGLE)	0.008	0.217
GLRLM - Grey level nonuniformity	0.0956	0.0909
GLRLM - Grey level non-uniformity normalised (GLNUN)	0.0079	0.0712
GLRLM - Run length non-uniformity	0.0222	0.0361
GLRLM - Run length non-uniformity normalised (RLNUN)	0.0043	0.0545
GLRLM - Run percentage	0.0013	0.0087
GLRLM - Grey level variance	0.0112	0.2927
GLRLM - Run length variance	0.0002	0.0004
GLRLM - Run entropy	0.0081	0.0935
GLSZM - Small zone emphasis	0.1161	0.1528
GLSZM - Large zone emphasis	0.0086	0.1072
GLSZM - Low grey level zone emphasis (LGLZE)	0.0368	0.0292
GLSZM - High grey level zone emphasis (HGLZE)	0.0199	0.2624
GLSZM - Small zone low grey level emphasis (SZLGLE)	0.0545	0.0733
GLSZM - Small zone high grey level emphasis (SZHGLE)	0.047	0.3655
GLSZM - Large zone low grey level emphasis (LZLGLE)	0.0007	0.0182
GLSZM - Large zone high grey level emphasis (LZHGLE)	0.0314	0.193
GLSZM - Grey level non-uniformity	0.0386	0.1603
GLSZM - Grey level non-uniformity normalised (GLNUN)	0.1053	0.3338
GLSZM - Zone size nonuniformity	0.0105	0.2085
GLSZM - Zone size non-uniformity normalised (ZSNUN)	0.127	0.1848
GLSZM - Zone percentage	0.0154	0.2573
GLSZM - Grey level variance	0.0754	0.3545
GLSZM - Zone size variance	0.0081	0.104
GLSZM - Zone size entropy	0.0548	0.1282
GLDZM - Small distance emphasis	0.0151	0.0012
GLDZM - Large distance emphasis	0.0037	0.0005
GLDZM - Low grey level zone emphasis (LGLZE)	0.0368	0.0292

Continued on next page

Table B.1 – continued from previous page

Feature name	Det-C	Det-T
GLDZM - High grey level zone emphasis (HGLZE)	0.0199	0.2624
GLDZM - Small distance low grey level emphasis (SDLGLE)	0.041	0.0337
GLDZM - Small distance high grey level emphasis (SDHGLE)	0.0742	0.0353
GLDZM - Large distance low grey level emphasis (LDLGLE)	0.059	0.1298
GLDZM - Large distance high grey level emphasis (LDHGLE)	0.0093	0.0334
GLDZM - Grey level non-uniformity	0.0386	0.1603
GLDZM - Grey level non-uniformity normalised (GLNUN)	0.1053	0.3338
GLDZM - Zone distance non-uniformity	0.058	0.3379
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	0.0083	0.0008
GLDZM - Zone percentage	0.0154	0.2573
GLDZM - Grey level variance	0.0754	0.3545
GLDZM - Zone distance variance	0.0015	0.0007
GLDZM - Zone distance entropy	0.0184	0.0882
NGTDM - Coarseness	0.0051	0
NGTDM - Contrast	0.0399	0.3358
NGTDM - Busyness	0.0443	0.0361
NGTDM - Complexity	0.0168	0.2601
NGTDM - Strength	0.0238	0.0134

Appendix C

Homogeneous Phantom Study

C.1 TOF

Table C.1: Coefficient of variation (COV) value for each extracted feature against TOF. Features were classified based on their COV values. 1= (COV $\leq$ 5%), 2 = (5% $>$ COV $\leq$ 10%), 3 = (10% $>$ COV $\leq$ 20%) and 4 = (COV $>$ 20%).

Feature name	COV value	Category
GLCM - Joint maximum	1.209995	1
GLCM - Joint average	2.549334	1
GLCM - Joint variance	1.937466	1
GLCM - Joint entropy	0.059552	1
GLCM - Difference average	0.810374	1
GLCM - Difference variance	5.0369	2
GLCM - Difference entropy	0.669634	1
GLCM - Sum average	2.549265	1
GLCM - Sum variance	1.839564	1
GLCM - Sum entropy	0.176996	1
GLCM - Angular second moment	0.25188	1
GLCM - Contrast	2.342285	1
GLCM - Dissimilarity	0.810374	1
GLCM - Inverse difference	0.513008	1
GLCM - Inverse difference normalised	0.080553	1
GLCM - Inverse difference moment	1.154074	1
GLCM - Inverse difference moment normalised (IDMN)	0.090671	1
GLCM - Inverse variance	0.094908	1
GLCM - Correlation	0.079962	1
GLCM - Autocorrelation	4.622595	1
GLCM - Cluster tendency	1.839564	1

Continued on next page

Table C.1 – continued from previous page

Feature name	COV value	Category
GLCM - Cluster shade	186.0711	4
GLCM - Cluster prominence	2.550577	1
GLCM - First measure of information correlation (FMIC)	1.048508	1
GLCM - Second measure of information correlation (SMIC)	0.081048	1
GLRLM - Short runs emphasis	0.097796	1
GLRLM - Long runs emphasis	0.408435	1
GLRLM - Low grey level run emphasis	10.00919	3
GLRLM - High grey level run emphasis	4.085657	1
GLRLM - Short run low grey level emphasis (SRLGLE)	9.586119	2
GLRLM - Short run high grey level emphasis (SRHGLE)	3.767822	1
GLRLM - Long run low grey level emphasis (LRLGLE)	16.75659	3
GLRLM - Long run high grey level emphasis (LRHGLE)	5.663038	2
GLRLM - Grey level nonuniformity	0.958496	1
GLRLM - Grey level non-uniformity normalised (GLNUN)	0.593215	1
GLRLM - Run length non-uniformity	0.301515	1
GLRLM - Run length non-uniformity normalised (RLNUN)	0.241905	1
GLRLM - Run percentage	0.161003	1
GLRLM - Grey level variance	2.520876	1
GLRLM - Run length variance	1.671803	1
GLRLM - Run entropy	0.03358	1
GLSZM - Small zone emphasis	1.505738	1
GLSZM - Large zone emphasis	5.704243	2
GLSZM - Low grey level zone emphasis (LGLZE)	10.12777	3
GLSZM - High grey level zone emphasis (HGLZE)	1.343503	1
GLSZM - Small zone low grey level emphasis (SZLGLE)	25.454	4
GLSZM - Small zone high grey level emphasis (SZHGLE)	2.853483	1
GLSZM - Large zone low grey level emphasis (LZLGLE)	35.85699	4
GLSZM - Large zone high grey level emphasis (LZHGLE)	4.484486	1
GLSZM - Grey level non-uniformity	1.246538	1
GLSZM - Grey level non-uniformity normalised (GLNUN)	0.648294	1
GLSZM - Zone size nonuniformity	8.212992	2
GLSZM - Zone size non-uniformity normalised (ZSNUN)	2.370851	1
GLSZM - Zone percentage	0.73774	1
GLSZM - Grey level variance	4.350783	1
GLSZM - Zone size variance	5.766627	2
GLSZM - Zone size entropy	0.501403	1
GLDZM - Small distance emphasis	2.199625	1
GLDZM - Large distance emphasis	3.67663	1
GLDZM - Low grey level zone emphasis (LGLZE)	10.12777	3

Continued on next page

Table C.1 – continued from previous page

Feature name	COV value	Category
GLDZM - High grey level zone emphasis (HGLZE)	1.343503	1
GLDZM - Small distance low grey level emphasis (SDLGLE)	10.75978	3
GLDZM - Small distance high grey level emphasis (SDHGLE)	3.590024	1
GLDZM - Large distance low grey level emphasis (LDLGLE)	7.450709	2
GLDZM - Large distance high grey level emphasis (LDHGLE)	1.705133	1
GLDZM - Grey level non-uniformity	1.246538	1
GLDZM - Grey level non-uniformity normalised (GLNUN)	0.648294	1
GLDZM - Zone distance non-uniformity	3.887552	1
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	2.594868	1
GLDZM - Zone percentage	0.73774	1
GLDZM - Grey level variance	4.350783	1
GLDZM - Zone distance variance	4.428949	1
GLDZM - Zone distance entropy	0.476482	1
NGTDM - Coarseness	0.70101	1
NGTDM - Contrast	6.500995	2
NGTDM - Busyness	4.214	1
NGTDM - Complexity	0.441609	1
NGTDM - Strength	0.077947	1

C.2 Number of subsets

Table C.2: Coefficient of variation (COV) value for each extracted feature against number of subsets. Features were classified based on their COV values. 1= (COV $\leq$ 5%), 2 = (5% $>$ COV $\leq$ 10%), 3 = (10% $>$ COV $\leq$ 20%) and 4 = (COV $>$ 20%).

Feature name	COV value	Category
GLCM - Joint maximum	10.28506	3
GLCM - Joint average	1.418932	1
GLCM - Joint variance	0.760987	1
GLCM - Joint entropy	0.150287	1
GLCM - Difference average	0.353159	1
GLCM - Difference variance	2.148313	1
GLCM - Difference entropy	0.391405	1
GLCM - Sum average	1.418975	1
GLCM - Sum variance	0.808427	1
GLCM - Sum entropy	0.169173	1
GLCM - Angular second moment	5.515031	2
GLCM - Contrast	1.097915	1
GLCM - Dissimilarity	0.353159	1
GLCM - Inverse difference	0.389601	1
GLCM - Inverse difference normalised	0.034074	1
GLCM - Inverse difference moment	0.733684	1
GLCM - Inverse difference moment normalised (IDMN)	0.041418	1
GLCM - Inverse variance	1.387875	1
GLCM - Correlation	0.578648	1
GLCM - Autocorrelation	2.570029	1
GLCM - Cluster tendency	0.808427	1
GLCM - Cluster shade	98.67022	4
GLCM - Cluster prominence	1.095071	1
GLCM - First measure of information correlation (FMIC)	0.903235	1
GLCM - Second measure of information correlation (SMIC)	0.072187	1
GLRLM - Short runs emphasis	0.083084	1
GLRLM - Long runs emphasis	1.177121	1
GLRLM - Low grey level run emphasis	9.39097	2
GLRLM - High grey level run emphasis	2.654953	1
GLRLM - Short run low grey level emphasis (SRLGLE)	9.305967	2
GLRLM - Short run high grey level emphasis (SRHGLE)	2.578809	1
GLRLM - Long run low grey level emphasis (LRLGLE)	11.17115	3
GLRLM - Long run high grey level emphasis (LRHGLE)	1.735614	1
GLRLM - Grey level nonuniformity	0.534828	1

Continued on next page

Table C.2 – continued from previous page

Feature name	COV value	Category
GLRLM - Grey level non-uniformity normalised (GLNUN)	0.545988	1
GLRLM - Run length non-uniformity	0.582624	1
GLRLM - Run length non-uniformity normalised (RLNUN)	0.21677	1
GLRLM - Run percentage	0.164443	1
GLRLM - Grey level variance	1.442457	1
GLRLM - Run length variance	9.160833	2
GLRLM - Run entropy	0.172942	1
GLSZM - Small zone emphasis	1.593022	1
GLSZM - Large zone emphasis	6.054601	2
GLSZM - Low grey level zone emphasis (LGLZE)	9.133536	2
GLSZM - High grey level zone emphasis (HGLZE)	0.785225	1
GLSZM - Small zone low grey level emphasis (SZLGLE)	18.45193	3
GLSZM - Small zone high grey level emphasis (SZHGLE)	1.493776	1
GLSZM - Large zone low grey level emphasis (LZLGLE)	16.4033	3
GLSZM - Large zone high grey level emphasis (LZHGLE)	7.524081	2
GLSZM - Grey level non-uniformity	0.959544	1
GLSZM - Grey level non-uniformity normalised (GLNUN)	1.494663	1
GLSZM - Zone size nonuniformity	4.685019	1
GLSZM - Zone size non-uniformity normalised (ZSNUN)	2.684618	1
GLSZM - Zone percentage	1.444865	1
GLSZM - Grey level variance	1.734755	1
GLSZM - Zone size variance	6.69044	2
GLSZM - Zone size entropy	0.233947	1
GLDZM - Small distance emphasis	1.156987	1
GLDZM - Large distance emphasis	2.644105	1
GLDZM - Low grey level zone emphasis (LGLZE)	9.133536	2
GLDZM - High grey level zone emphasis (HGLZE)	0.785225	1
GLDZM - Small distance low grey level emphasis (SDLGLE)	9.49212	2
GLDZM - Small distance high grey level emphasis (SDHGLE)	2.944746	1
GLDZM - Large distance low grey level emphasis (LDLGLE)	7.645567	2
GLDZM - Large distance high grey level emphasis (LDHGLE)	2.433589	1
GLDZM - Grey level non-uniformity	0.959544	1
GLDZM - Grey level non-uniformity normalised (GLNUN)	1.494663	1
GLDZM - Zone distance non-uniformity	2.364321	1
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	1.404652	1
GLDZM - Zone percentage	1.444865	1
GLDZM - Grey level variance	1.734755	1
GLDZM - Zone distance variance	3.803505	1
GLDZM - Zone distance entropy	0.198593	1

Continued on next page

Table C.2 – continued from previous page

Feature name	COV value	Category
NGTDM - Coarseness	1.484466	1
NGTDM - Contrast	2.909557	1
NGTDM - Busyness	4.378581	1
NGTDM - Complexity	0.898151	1
NGTDM - Strength	1.732056	1

C.3 Number of iterations

Table C.3: Coefficient of variation (COV) value for each extracted feature against number of iterations. Features were classified based on their COV values. 1= (COV $\leq$ 5%), 2 = (5% > COV $\leq$ 10%), 3 = (10% > COV $\leq$ 20%) and 4 = (COV > 20%).

Feature name	COV value	Category
GLCM - Joint maximum	9.570726	2
GLCM - Joint average	0.537439	1
GLCM - Joint variance	1.879843	1
GLCM - Joint entropy	0.200842	1
GLCM - Difference average	0.493665	1
GLCM - Difference variance	1.072084	1
GLCM - Difference entropy	0.159247	1
GLCM - Sum average	0.537411	1
GLCM - Sum variance	2.011339	1
GLCM - Sum entropy	0.12298	1
GLCM - Angular second moment	2.937475	1
GLCM - Contrast	1.011021	1
GLCM - Dissimilarity	0.493665	1
GLCM - Inverse difference	0.336833	1
GLCM - Inverse difference normalised	0.047328	1
GLCM - Inverse difference moment	0.421731	1
GLCM - Inverse difference moment normalised (IDMN)	0.03031	1
GLCM - Inverse variance	2.352032	1
GLCM - Correlation	0.326792	1
GLCM - Autocorrelation	0.866789	1

Continued on next page

Table C.3 – continued from previous page

Feature name	COV value	Category
GLCM - Cluster tendency	2.011339	1
GLCM - Cluster shade	1.417484	1
GLCM - Cluster prominence	2.724493	1
GLCM - First measure of information correlation (FMIC)	0.518743	1
GLCM - Second measure of information correlation (SMIC)	0.08609	1
GLRLM - Short runs emphasis	0.147851	1
GLRLM - Long runs emphasis	1.702602	1
GLRLM - Low grey level run emphasis	9.403203	2
GLRLM - High grey level run emphasis	1.329067	1
GLRLM - Short run low grey level emphasis (SRLGLE)	9.301164	2
GLRLM - Short run high grey level emphasis (SRHGLE)	1.629324	1
GLRLM - Long run low grey level emphasis (LRLGLE)	9.507155	2
GLRLM - Long run high grey level emphasis (LRHGLE)	2.492977	1
GLRLM - Grey level nonuniformity	1.29667	1
GLRLM - Grey level non-uniformity normalised (GLNUN)	0.989271	1
GLRLM - Run length non-uniformity	0.735123	1
GLRLM - Run length non-uniformity normalised (RLNUN)	0.338657	1
GLRLM - Run percentage	0.414682	1
GLRLM - Grey level variance	1.053616	1
GLRLM - Run length variance	5.049672	2
GLRLM - Run entropy	0.208568	1
GLSZM - Small zone emphasis	1.012512	1
GLSZM - Large zone emphasis	2.206447	1
GLSZM - Low grey level zone emphasis (LGLZE)	7.703637	2
GLSZM - High grey level zone emphasis (HGLZE)	2.598212	1
GLSZM - Small zone low grey level emphasis (SZLGLE)	7.948516	2
GLSZM - Small zone high grey level emphasis (SZHGLE)	3.914726	1
GLSZM - Large zone low grey level emphasis (LZLGLE)	2.244203	1
GLSZM - Large zone high grey level emphasis (LZHGLE)	1.502802	1
GLSZM - Grey level non-uniformity	0.751124	1
GLSZM - Grey level non-uniformity normalised (GLNUN)	1.110607	1
GLSZM - Zone size nonuniformity	1.880062	1
GLSZM - Zone size non-uniformity normalised (ZSNUN)	1.726127	1
GLSZM - Zone percentage	0.792943	1
GLSZM - Grey level variance	2.760555	1
GLSZM - Zone size variance	2.240038	1
GLSZM - Zone size entropy	0.586374	1
GLDZM - Small distance emphasis	1.131895	1
GLDZM - Large distance emphasis	1.890414	1

Continued on next page

Table C.3 – continued from previous page

Feature name	COV value	Category
GLDZM - Low grey level zone emphasis (LGLZE)	7.703637	2
GLDZM - High grey level zone emphasis (HGLZE)	2.598212	1
GLDZM - Small distance low grey level emphasis (SDLGLE)	7.965627	2
GLDZM - Small distance high grey level emphasis (SDHGLE)	2.431918	1
GLDZM - Large distance low grey level emphasis (LDLGLE)	6.667451	2
GLDZM - Large distance high grey level emphasis (LDHGLE)	3.403854	1
GLDZM - Grey level non-uniformity	0.751124	1
GLDZM - Grey level non-uniformity normalised (GLNUN)	1.110607	1
GLDZM - Zone distance non-uniformity	1.007998	1
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	1.513898	1
GLDZM - Zone percentage	0.792943	1
GLDZM - Grey level variance	2.760555	1
GLDZM - Zone distance variance	2.056709	1
GLDZM - Zone distance entropy	0.612305	1
NGTDM - Coarseness	0.677897	1
NGTDM - Contrast	1.605633	1
NGTDM - Busyness	1.164169	1
NGTDM - Complexity	0.69771	1
NGTDM - Strength	0.639798	1

C.4 FWHM of the Gaussian filter

Table C.4: Coefficient of variation (COV) value for each extracted feature against FWHM of the Gaussian filter. Features were classified based on their COV values. 1= (COV $\leq$ 5%), 2 = (5% $>$ COV $\leq$ 10%), 3 = (10% $>$ COV $\leq$ 20%) and 4 = (COV $>$ 20%).

Feature name	COV value	Category
GLCM - Joint maximum	19.89709	3
GLCM - Joint average	2.221747	1
GLCM - Joint variance	3.5627	1
GLCM - Joint entropy	0.440991	1
GLCM - Difference average	0.471985	1
GLCM - Difference variance	3.534965	1
Continued on next page		

Table C.4 – continued from previous page

Feature name	COV value	Category
GLCM - Difference entropy	1.049624	1
GLCM - Sum average	2.221764	1
GLCM - Sum variance	4.473685	1
GLCM - Sum entropy	0.915924	1
GLCM - Angular second moment	9.317469	2
GLCM - Contrast	1.125414	1
GLCM - Dissimilarity	0.471985	1
GLCM - Inverse difference	2.095014	1
GLCM - Inverse difference normalised	0.071058	1
GLCM - Inverse difference moment	3.851377	1
GLCM - Inverse difference moment normalised (IDMN)	0.043175	1
GLCM - Inverse variance	3.362281	1
GLCM - Correlation	2.403221	1
GLCM - Autocorrelation	4.668182	1
GLCM - Cluster tendency	4.473685	1
GLCM - Cluster shade	35.00537	4
GLCM - Cluster prominence	7.908612	2
GLCM - First measure of information correlation (FMIC)	5.526459	2
GLCM - Second measure of information correlation (SMIC)	1.50439	1
GLRLM - Short runs emphasis	0.352431	1
GLRLM - Long runs emphasis	3.647398	1
GLRLM - Low grey level run emphasis	11.66363	3
GLRLM - High grey level run emphasis	2.765474	1
GLRLM - Short run low grey level emphasis (SRLGLE)	11.44006	3
GLRLM - Short run high grey level emphasis (SRHGLE)	1.631157	1
GLRLM - Long run low grey level emphasis (LRLGLE)	12.01787	3
GLRLM - Long run high grey level emphasis (LRHGLE)	12.66813	3
GLRLM - Grey level nonuniformity	11.03509	3
GLRLM - Grey level non-uniformity normalised (GLNUN)	5.129388	2
GLRLM - Run length non-uniformity	2.983162	1
GLRLM - Run length non-uniformity normalised (RLNUN)	0.775223	1
GLRLM - Run percentage	0.716679	1
GLRLM - Grey level variance	3.165867	1
GLRLM - Run length variance	25.68795	4
GLRLM - Run entropy	0.961359	1
GLSZM - Small zone emphasis	3.527785	1
GLSZM - Large zone emphasis	17.58378	3
GLSZM - Low grey level zone emphasis (LGLZE)	9.713199	2
GLSZM - High grey level zone emphasis (HGLZE)	1.29254	1

Continued on next page

Table C.4 – continued from previous page

Feature name	COV value	Category
GLSZM - Small zone low grey level emphasis (SZLGLE)	14.55137	3
GLSZM - Small zone high grey level emphasis (SZHGLE)	4.864309	1
GLSZM - Large zone low grey level emphasis (LZLGLE)	28.28251	4
GLSZM - Large zone high grey level emphasis (LZHGLE)	18.11073	3
GLSZM - Grey level non-uniformity	1.101546	1
GLSZM - Grey level non-uniformity normalised (GLNUN)	1.709973	1
GLSZM - Zone size nonuniformity	4.49374	1
GLSZM - Zone size non-uniformity normalised (ZSNUN)	6.184505	2
GLSZM - Zone percentage	2.326517	1
GLSZM - Grey level variance	3.208435	1
GLSZM - Zone size variance	19.15134	3
GLSZM - Zone size entropy	0.410451	1
GLDZM - Small distance emphasis	1.107389	1
GLDZM - Large distance emphasis	4.167855	1
GLDZM - Low grey level zone emphasis (LGLZE)	9.713199	2
GLDZM - High grey level zone emphasis (HGLZE)	1.29254	1
GLDZM - Small distance low grey level emphasis (SDLGLE)	9.799667	2
GLDZM - Small distance high grey level emphasis (SDHGLE)	2.589484	1
GLDZM - Large distance low grey level emphasis (LDLGLE)	9.560525	2
GLDZM - Large distance high grey level emphasis (LDHGLE)	6.504616	2
GLDZM - Grey level non-uniformity	1.101546	1
GLDZM - Grey level non-uniformity normalised (GLNUN)	1.709973	1
GLDZM - Zone distance non-uniformity	2.674285	1
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	2.367365	1
GLDZM - Zone percentage	2.326517	1
GLDZM - Grey level variance	3.208435	1
GLDZM - Zone distance variance	10.96196	3
GLDZM - Zone distance entropy	0.88437	1
NGTDM - Coarseness	4.738623	1
NGTDM - Contrast	2.823972	1
NGTDM - Busyness	5.165375	2
NGTDM - Complexity	5.117232	2
NGTDM - Strength	4.300013	1

C.5 Agreement

Table C.5: Percent agreement (PA) values of each features across all of reconstruction settings.

Feature name	Percentages agreement
GLCM - Joint maximum	25 %
GLCM - Joint average	100 %
GLCM - Joint variance	100 %
GLCM - Joint entropy	100 %
GLCM - Difference average	100 %
GLCM - Difference variance	66 %
GLCM - Difference entropy	100 %
GLCM - Sum average	100 %
GLCM - Sum variance	83 %
GLCM - Sum entropy	100 %
GLCM - Angular second moment	25 %
GLCM - Contrast	100 %
GLCM - Dissimilarity	100 %
GLCM - Inverse difference	100 %
GLCM - Inverse difference normalised	100 %
GLCM - Inverse difference moment	100 %
GLCM - Inverse difference moment normalised (IDMN)	100
GLCM - Inverse variance	83 %
GLCM - Correlation	100 %
GLCM - Autocorrelation	83 %
GLCM - Cluster tendency	83 %
GLCM - Cluster shade	33 %
GLCM - Cluster prominence	100 %
GLCM - First measure of information correlation (FMIC)	83 %
GLCM - Second measure of information correlation (SMIC)	100 %
GLRLM - Short runs emphasis	100 %
GLRLM - Long runs emphasis	83 %
GLRLM - Low grey level run emphasis	25 %
GLRLM - High grey level run emphasis	83 %
GLRLM - Short run low grey level emphasis (SRLGLE)	25 %
GLRLM - Short run high grey level emphasis (SRHGLE)	100 %
GLRLM - Long run low grey level emphasis (LRLGLE)	50 %
GLRLM - Long run high grey level emphasis (LRHGLE)	66 %
GLRLM - Grey level nonuniformity	100 %
GLRLM - Grey level non-uniformity normalised (GLNUN)	83 %

Continued on next page

Table C.5 – continued from previous page

Feature name	Percentages agreement
GLRLM - Run length non-uniformity	100 %
GLRLM - Run length non-uniformity normalised (RLNUN)	100 %
GLRLM - Run percentage	100 %
GLRLM - Grey level variance	100 %
GLRLM - Run length variance	66 %
GLRLM - Run entropy	100 %
GLSZM - Small zone emphasis	100 %
GLSZM - Large zone emphasis	66 %
GLSZM - Low grey level zone emphasis (LGLZE)	33 %
GLSZM - High grey level zone emphasis (HGLZE)	100 %
GLSZM - Small zone low grey level emphasis (SZLGLE)	58 %
GLSZM - Small zone high grey level emphasis (SZHGLE)	100 %
GLSZM - Large zone low grey level emphasis (LZLGLE)	50 %
GLSZM - Large zone high grey level emphasis (LZHGLE)	83 %
GLSZM - Grey level non-uniformity	100 %
GLSZM - Grey level non-uniformity normalised (GLNUN)	100 %
GLSZM - Zone size nonuniformity	66 %
GLSZM - Zone size non-uniformity normalised (ZSNUN)	83 %
GLSZM - Zone percentage	100 %
GLSZM - Grey level variance	100 %
GLSZM - Zone size variance	66 %
GLSZM - Zone size entropy	100 %
GLDZM - Small distance emphasis	83 %
GLDZM - Large distance emphasis	83 %
GLDZM - Low grey level zone emphasis (LGLZE)	33 %
GLDZM - High grey level zone emphasis (HGLZE)	100 %
GLDZM - Small distance low grey level emphasis (SDLGLE)	33 %
GLDZM - Small distance high grey level emphasis (SDHGLE)	83 %
GLDZM - Large distance low grey level emphasis (LDLGLE)	33 %
GLDZM - Large distance high grey level emphasis (LDHGLE)	66 %
GLDZM - Grey level non-uniformity	100 %
GLDZM - Grey level non-uniformity normalised (GLNUN)	100 %
GLDZM - Zone distance non-uniformity	100 %
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	100 %
GLDZM - Zone percentage	100 %
GLDZM - Grey level variance	100 %
GLDZM - Zone distance variance	83 %
GLDZM - Zone distance entropy	100 %
NGTDM - Coarseness	75 %

Continued on next page

Table C.5 – continued from previous page

Feature name	Percentages agreement
NGTDM - Contrast	66 %
NGTDM - Busyness	58 %
NGTDM - Complexity	83 %
NGTDM - Strength	83 %
Continued on next page	

Appendix D

Inhomogeneous Phantom Study

D.1 TOF

Table D.1: Coefficient of variation (COV) value for each extracted feature against TOF. Features were classified based on their COV values. 1= (COV $\leq$ 5%), 2 = (5% $>$ COV $\leq$ 10%), 3 = (10% $>$ COV $\leq$ 20%) and 4 = (COV $>$ 20%).

Feature name	COV value	Category
GLCM - Joint variance	1.838211999	1
GLCM - Joint entropy	0.340332167	1
GLCM - Difference average	2.726727804	1
GLCM - Difference entropy	1.147372759	1
GLCM - Sum average	0.224922442	1
GLCM - Sum variance	1.568665995	1
GLCM - Sum entropy	0.080324529	1
GLCM - Contrast	5.176280897	2
GLCM - Dissimilarity	2.726727804	1
GLCM - Inverse difference	0.785846259	1
GLCM - Inverse difference normalised	0.209292705	1
GLCM - Inverse difference moment	1.142224261	1
GLCM - Inverse difference moment normalised (IDMN)	0.083513729	1
GLCM - Inverse variance	3.392467829	1
GLCM - Correlation	0.523308271	1
GLCM - Autocorrelation	0.713534727	1
GLCM - Cluster tendency	1.568665995	1
GLCM - Cluster prominence	2.173427915	1
GLCM - Second measure of information correlation (SMIC)	0.084835519	1
GLRLM - Short runs emphasis	0.030881862	1
GLRLM - Long runs emphasis	1.538080027	1

Continued on next page

Table D.1 – continued from previous page

Feature name	COV value	Category
GLRLM - High grey level run emphasis	0.764578146	1
GLRLM - Short run high grey level emphasis (SRHGLE)	0.92791732	1
GLRLM - Grey level nonuniformity	0.24495383	1
GLRLM - Run length non-uniformity	0.195323554	1
GLRLM - Run length non-uniformity normalised (RLNUN)	0.077768755	1
GLRLM - Run percentage	0.129455373	1
GLRLM - Grey level variance	1.749499736	1
GLRLM - Run entropy	0.082892889	1
GLSZM - Small zone emphasis	1.281863546	1
GLSZM - High grey level zone emphasis (HGLZE)	0.501901469	1
GLSZM - Small zone high grey level emphasis (SZHGLE)	0.497084398	1
GLSZM - Grey level non-uniformity	3.110199198	1
GLSZM - Grey level non-uniformity normalised (GLNUN)	0.088695152	1
GLSZM - Zone size non-uniformity normalised (ZSNUN)	0.916917121	1
GLSZM - Zone percentage	1.545815622	1
GLSZM - Grey level variance	2.342490358	1
GLSZM - Zone size entropy	0.217138061	1
GLDZM - Small distance emphasis	0.872506605	1
GLDZM - Large distance emphasis	0.433151522	1
GLDZM - High grey level zone emphasis (HGLZE)	0.501901469	1
GLDZM - Small distance high grey level emphasis (SDHGLE)	0.718157938	1
GLDZM - Grey level non-uniformity	3.110199198	1
GLDZM - Grey level non-uniformity normalised (GLNUN)	0.088695152	1
GLDZM - Zone distance non-uniformity	2.370001636	1
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	0.183956101	1
GLDZM - Zone percentage	1.545815622	1
GLDZM - Grey level variance	2.342490358	1
GLDZM - Zone distance entropy	0.366710645	1
NGTDM - Complexity	1.859566837	1
NGTDM - Strength	2.122266324	1

D.2 Number of subsets

Table D.2: Coefficient of variation (COV) value for each extracted feature against number of subsets. Features were classified based on their COV values. 1= (COV $\leq$ 5%), 2 = (5% > COV $\leq$ 10%), 3 = (10% > COV $\leq$ 20%) and 4 = (COV > 20%).

Feature name	COV value	Category
GLCM - Joint variance	1.099765407	1
GLCM - Joint entropy	2.358950685	1
GLCM - Difference average	0.708140226	1
GLCM - Difference entropy	1.491282939	1
GLCM - Sum average	0.727449169	1
GLCM - Sum variance	1.099745944	1
GLCM - Sum entropy	2.693300502	1
GLCM - Contrast	0.090500191	1
GLCM - Dissimilarity	1.983821287	1
GLCM - Inverse difference	1.491282939	1
GLCM - Inverse difference normalised	1.168798449	1
GLCM - Inverse difference moment	0.12763064	1
GLCM - Inverse difference moment normalised (IDMN)	1.637147237	1
GLCM - Inverse variance	0.034610497	1
GLCM - Correlation	0.841295017	1
GLCM - Autocorrelation	0.682850821	1
GLCM - Cluster tendency	2.123825618	1
GLCM - Cluster prominence	2.693300502	1
GLCM - Second measure of information correlation (SMIC)	4.942872891	1
GLRLM - Short runs emphasis	0.616499134	1
GLRLM - Long runs emphasis	0.129028329	1
GLRLM - High grey level run emphasis	6.327371522	2
GLRLM - Short run high grey level emphasis (SRHGLE)	1.681415375	1
GLRLM - Grey level nonuniformity	1.509847968	1
GLRLM - Run length non-uniformity	0.514912522	1
GLRLM - Run length non-uniformity normalised (RLNUN)	0.239544216	1
GLRLM - Run percentage	0.387218723	1
GLRLM - Grey level variance	0.506718974	1
GLRLM - Run entropy	2.041231481	1
GLSZM - Small zone emphasis	0.158906554	1
GLSZM - High grey level zone emphasis (HGLZE)	3.213701593	1
GLSZM - Small zone high grey level emphasis (SZHGLE)	1.24566545	1
GLSZM - Grey level non-uniformity	4.419056367	1
GLSZM - Grey level non-uniformity normalised (GLNUN)	8.033040516	2
GLSZM - Zone size non-uniformity normalised (ZSNUN)	1.085380474	1
GLSZM - Zone percentage	6.021638574	2

Continued on next page

Table D.2 – continued from previous page

Feature name	COV value	Category
GLSZM - Grey level variance	8.267507466	2
GLSZM - Zone size entropy	1.047521756	1
GLDZM - Small distance emphasis	0.56910176	1
GLDZM - Large distance emphasis	2.23699304	1
GLDZM - High grey level zone emphasis (HGLZE)	2.440742252	1
GLDZM - Small distance high grey level emphasis (SDHGLE)	1.24566545	1
GLDZM - Grey level non-uniformity	1.986951123	1
GLDZM - Grey level non-uniformity normalised (GLNUN)	8.033040516	2
GLDZM - Zone distance non-uniformity	1.085380474	1
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	6.907546552	2
GLDZM - Zone percentage	1.434992428	1
GLDZM - Grey level variance	8.267507466	2
GLDZM - Zone distance entropy	1.047521756	1
NGTDM - Complexity	0.750278209	1
NGTDM - Strength	1.305635017	1
	3.29036	1

D.3 Number of iterations

Table D.3: Coefficient of variation (COV) value for each extracted feature against number of iterations. Features were classified based on their COV values. 1= (COV $\leq$ 5%), 2 = (5% $>$ COV $\leq$ 10%), 3 = (10% $>$ COV $\leq$ 20%) and 4 = (COV $>$ 20%).

Feature name	COV value	Category
GLCM - Joint variance	3.276727915	1
GLCM - Joint entropy	4.403010125	1
GLCM - Difference average	0.455175554	1
GLCM - Difference entropy	1.764523481	1
GLCM - Sum average	0.612892551	1
GLCM - Sum variance	3.276729999	1
GLCM - Sum entropy	4.608359092	1
GLCM - Contrast	0.341057215	1
GLCM - Dissimilarity	3.17037763	1

Continued on next page

Table D.3 – continued from previous page

Feature name	COV value	Category
GLCM - Inverse difference	1.764523481	1
GLCM - Inverse difference normalised	0.960653016	1
GLCM - Inverse difference moment	0.139134179	1
GLCM - Inverse difference moment normalised (IDMN)	1.345084551	1
GLCM - Inverse variance	0.051078676	1
GLCM - Correlation	1.32228055	1
GLCM - Autocorrelation	0.57693656	1
GLCM - Cluster tendency	6.228228526	2
GLCM - Cluster prominence	4.608359092	1
GLCM - Second measure of information correlation (SMIC)	7.077853975	2
GLRLM - Short runs emphasis	0.427031764	1
GLRLM - Long runs emphasis	0.251850785	1
GLRLM - High grey level run emphasis	3.641990671	1
GLRLM - Short run high grey level emphasis (SRHGLE)	5.689856307	2
GLRLM - Grey level nonuniformity	5.671304315	2
GLRLM - Run length non-uniformity	1.733190256	1
GLRLM - Run length non-uniformity normalised (RLNUN)	0.783906909	1
GLRLM - Run percentage	0.637634952	1
GLRLM - Grey level variance	0.326622638	1
GLRLM - Run entropy	4.167834982	1
GLSZM - Small zone emphasis	0.154024321	1
GLSZM - High grey level zone emphasis (HGLZE)	2.321615055	1
GLSZM - Small zone high grey level emphasis (SZHGLE)	3.413042085	1
GLSZM - Grey level non-uniformity	3.668488619	1
GLSZM - Grey level non-uniformity normalised (GLNUN)	5.950887901	2
GLSZM - Zone size non-uniformity normalised (ZSNUN)	1.946779251	1
GLSZM - Zone percentage	4.59968791	1
GLSZM - Grey level variance	4.922107764	1
GLSZM - Zone size entropy	2.435816206	1
GLDZM - Small distance emphasis	0.561174462	1
GLDZM - Large distance emphasis	1.01887899	1
GLDZM - High grey level zone emphasis (HGLZE)	1.892840122	1
GLDZM - Small distance high grey level emphasis (SDHGLE)	3.413042085	1
GLDZM - Grey level non-uniformity	4.84417675	1
GLDZM - Grey level non-uniformity normalised (GLNUN)	5.950887901	2
GLDZM - Zone distance non-uniformity	1.946779251	1
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	3.967794118	1
GLDZM - Zone percentage	1.528127304	1
GLDZM - Grey level variance	4.922107764	1

Continued on next page

Table D.3 – continued from previous page

Feature name	COV value	Category
GLDZM - Zone distance entropy	2.435816206	1
NGTDM - Complexity	0.390488951	1
NGTDM - Strength	1.128096941	1
	2.40642109	1

D.4 FWHM of the Gaussian filter

Table D.4: Coefficient of variation (COV) value for each extracted feature against FWHM of the Gaussian filter. Features were classified based on their COV values. 1= (COV $\leq$ 5%), 2 = (5% $>$ COV $\leq$ 10%), 3 = (10% $>$ COV $\leq$ 20%) and 4 = (COV $>$ 20%).

Feature name	COV value	Category
GLCM - Joint variance	17.35323668	3
GLCM - Joint entropy	14.81420063	3
GLCM - Difference average	3.393139638	1
GLCM - Difference entropy	2.014348972	1
GLCM - Sum average	0.621729129	1
GLCM - Sum variance	17.35321324	3
GLCM - Sum entropy	16.7950323	3
GLCM - Contrast	3.550719762	1
GLCM - Dissimilarity	5.79200993	2
GLCM - Inverse difference	2.014348972	1
GLCM - Inverse difference normalised	3.642332723	1
GLCM - Inverse difference moment	0.207087748	1
GLCM - Inverse difference moment normalised (IDMN)	5.641649123	2
GLCM - Inverse variance	0.086708274	1
GLCM - Correlation	4.032787974	1
GLCM - Autocorrelation	4.318506169	1
GLCM - Cluster tendency	30.5448642	4
GLCM - Cluster prominence	16.7950323	3
GLCM - Second measure of information correlation (SMIC)	13.82350955	3
GLRLM - Short runs emphasis	1.91838385	1
GLRLM - Long runs emphasis	1.48470735	1
GLRLM - High grey level run emphasis	3.282406447	1
GLRLM - Short run high grey level emphasis (SRHGLE)	25.39841122	4
GLRLM - Grey level nonuniformity	24.26204085	4
GLRLM - Run length non-uniformity	16.38583137	3
GLRLM - Run length non-uniformity normalised (RLNUN)	5.046558563	2
GLRLM - Run percentage	3.474991403	1
GLRLM - Grey level variance	1.592779516	1
GLRLM - Run entropy	13.74265338	3
GLSZM - Small zone emphasis	2.32765943	1
GLSZM - High grey level zone emphasis (HGLZE)	7.334787572	2
GLSZM - Small zone high grey level emphasis (SZHGLE)	6.227976003	2
GLSZM - Grey level non-uniformity	5.480631424	2
GLSZM - Grey level non-uniformity normalised (GLNUN)	5.292866723	2

Continued on next page

Table D.4 – continued from previous page

Feature name	COV value	Category
GLSZM - Zone size non-uniformity normalised (ZSNUN)	2.886425889	1
GLSZM - Zone percentage	12.82895995	3
GLSZM - Grey level variance	3.170003635	1
GLSZM - Zone size entropy	3.690062932	1
GLDZM - Small distance emphasis	2.026089995	1
GLDZM - Large distance emphasis	14.21516415	3
GLDZM - High grey level zone emphasis (HGLZE)	12.78176985	3
GLDZM - Small distance high grey level emphasis (SDHGLE)	6.227976003	2
GLDZM - Grey level non-uniformity	20.13799804	4
GLDZM - Grey level non-uniformity normalised (GLNUN)	5.292866723	2
GLDZM - Zone distance non-uniformity	2.886425889	1
GLDZM - Zone distance non-uniformity normalised (ZDNUN)	4.87532198	1
GLDZM - Zone percentage	6.23002069	2
GLDZM - Grey level variance	3.170003635	1
GLDZM - Zone distance entropy	3.690062932	1
NGTDM - Complexity	1.586002093	1
NGTDM - Strength	6.741349904	2
	4.834648603	1