

Accuracy and stability of radiomic features for characterising tumour heterogeneity using multimodality imaging: a phantom study

K.M. Meijer

Graduation Committee:

Prof. dr. ir. C.H. Slump
Prof. dr. L.F. de Geus-Oei
Dr. W. Grootjans
Drs. R.J. Haarman
T. Boers, MSc

June 11, 2019

	<i>Supervisors</i>
Chairman & Technical supervisor – University of Twente	Prof. dr. ir. C.H. Slump
Medical supervisor – LUMC & University of Twente	Prof. dr. L.F. de Geus-Oei
Medical supervisor - LUMC	Prof. dr. J.L. Bloem
Technical supervisor - LUMC	Dr. W. Grootjans
Professional behaviour supervisor- University of Twente	Drs. R.J. Haarman
External member – University of Twente	T. Boers, MSc

Abstract

Introduction: Characterisation of tumour heterogeneity is becoming increasingly important in the clinical workflow of oncology patients. Radiomics is a promising method for accurate tumour quantification. The aim of this study was to establish the accuracy and variability of radiomic features on a multimodality imaging platform using a multi-compartment heterogeneous phantom.

Materials and Methods: A 3D printed multi-compartment phantom, feasible to be inserted in the NEMA-NU2 positron emitting tomography (PET) phantom, was developed to simulate heterogeneous composition patterns in PET, computed tomography (CT) and magnetic resonance (MR) imaging. Three cylindrical volumes, containing different rectangular compartments (L-shape, stair-shape and U-shape) were filled with different concentrations (ratios of 2:4:8:16 with respect to the background) of 18F-fluorodeoxyglucose (FDG), iodinated contrast, and gadolinium contrast. The rectangular compartments were created at different scales, consisting of cubes with dimensions of 5x5x5 mm, 7.5x7.5x7.5 mm and 10x10x10 mm, respectively. Image acquisition (standard abdominal imaging protocols) was repeated 5 times for variability analysis. Effect of scalability was assessed by comparison of the three scaled compartments. Furthermore, high resolution CT imaging was performed for development of a ground truth reference in order to determine the accuracy of radiomic features. A total of 54 radiomic features were extracted using PyRadiomics.

Results: Regarding the variability analysis, features extracted from MR had the highest reproducibility (N=1 with a variation >20%), followed by CT (N=17 with a variation >20%) and PET (N=28 with a variation >20%). First-Order (FO), Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Neighbouring Gray Tone Difference Matrix (NGTDM) features were reproducible on all imaging modalities, with variations below 20%. Furthermore, multiple features in the GLCM class showed variations below 1% in all modalities. Scalability analysis showed that the size of the phantom structures significantly affected the absolute values of the majority of the radiomic features. PET features were affected to a lesser extent by scale than CT and MR, with the GLCM feature group least affected. Finally, the absolute value of most FO, and GLCM features showed less than 20% deviation from the ground truth reference values, whereas the GLRLM, Gray Level Size Zone Matrix (GLSZM) and NGTDM features showed significant deviations.

Conclusion: This study showed that heterogeneous patterns can be simulated in PET, CT and MR imaging using a multi-compartment imaging phantom. The three imaging modalities showed different, but promising, results regarding variability, scalability and accuracy of radiomic features facilitating a next step towards tumour characterisation.

Contents

Chapter 1: Phantom study and radiomic feature analysis	1
1.1 Introduction	1
1.1.1 Tumour heterogeneity	1
1.1.2 Imaging of tumour heterogeneity	2
Positron emission tomography	2
X-ray computed tomography	2
Magnetic resonance imaging	3
1.1.3 Radiomic features	3
1.1.4 Texture analysis in multiple modalities	4
1.1.5 Heterogeneity phantom	5
1.2 Methods and materials	6
1.2.1 Study design	6
1.2.2 Phantom design	7
1.2.3 Image acquisition	7
1.2.4 Image data processing	9
1.2.5 Feature extraction	11
1.2.6 Statistical analysis	11
1.3 Results	12
1.3.1 Image acquisition	12
1.3.2 Generation ground truth	12
1.3.3 Statistical analysis	14
Variability	14
Scalability	16
Accuracy	16
1.4 Discussion	21
1.4.1 Phantom design	21
1.4.2 Variability	21
1.4.3 Scalability	23
1.4.4 Accuracy	23
1.5 Conclusion	24
Chapter 2: Phantom design and development	25
2.1 NEMA body phantom	25
2.2 Phantom requirements	26
2.3 Phantom case design	27
2.4 Phantom insert design	29
2.4.1 Primary concept: tumour design	30
2.4.2 Final concept: cubical design	31
2.5 Global phantom design	33
2.6 Final design	35
References	36
Appendices	39
Appendix A	39
Appendix B	41
Appendix C	44
Verantwoording	45
Dankwoord	48

Chapter 1: Phantom study for radiomic feature analysis

1.1 Introduction

1.1.1 Tumour heterogeneity

Medical imaging is widespread and routinely used for anatomical and functional imaging of patients in the field of oncology. From diagnosis to tumour staging, treatment planning, treatment response and follow up after treatment, medical imaging has become standard in clinical imaging protocols for many different oncological disease types. In current practice, radiologists use several qualitative features to assess the medical images. Diagnosis and clinical staging of tumour lesions are based on anatomical properties (size and location), disease extent and increasingly functional properties, as determined on medical imaging. Dependent on the anatomical characteristics and clinical question, different imaging modalities are used to examine tumour lesions. Conventional radiography, ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET) are techniques used in the daily clinical practice.

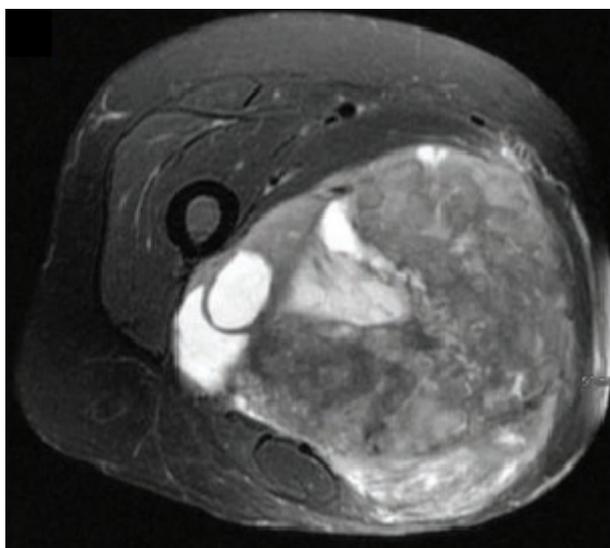


Figure 1: A T2-weighted fat suppressed magnetic resonance image in axial view of the right thigh. A large mass is shown in the medial thigh with heterogeneous presentation.[1]

Tumours are a composition of tissues with different properties and each tumour has its own composition. Therefore, most tumours have an inhomogeneous (unique) micro-environment, defined as tumour heterogeneity (Figure 1). Tumour heterogeneity refers to this variability of many different features in tumour composition (e.g. cells morphology, shape, size, receptors) and their molecular properties (e.g. metabolism, metastatic potential, radiation response, proliferation rate).[2] Recently, intra-tumour heterogeneity has histologically and genetically become more apparent, and a relation between intra-tumour heterogeneity and clinical outcome is established in several research studies. It is known that proliferation, cell density, metabolism and vascularisation of a tumour is important in tumour heterogeneity as well as treatment response and resistance. It has been established that tumours with a highly heterogeneous character lead to a poorer patient prognosis.[3–5]

In different oncological fields there is a focus on personalising treatment, including selection of suitable systemic therapy, guidance of radiation therapy, patient specific dose regulation and improvement of likelihood of patient survival. Furthermore, identification of patients at risk of developing metastases and tumour aggressiveness is important to adequately tailor treatment to the individual patient. This emphasises the importance to adequately capture tumour heterogeneity.[6, 7]

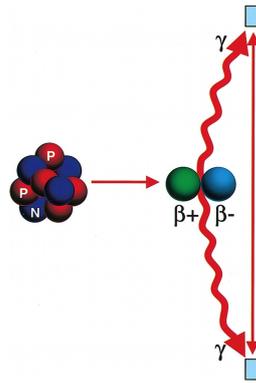


Figure 2: Representation of the annihilation reaction. FDG releases positrons (β^+) from its nucleus which annihilates with electrons β^- . This reaction activates two photons (γ) which are detected by two detectors in opposite directions.[8]

1.1.2. Imaging of tumour heterogeneity

Different imaging modalities have the possibility to measure different physical and biological properties of the tumour. The combined information from these imaging modalities can be used to characterise the tumour. Besides the visual evaluation of a tumour, several quantitative parameters can be extracted from these modalities.

Positron emission tomography

Positron emission tomography (PET) imaging is a functional imaging technique used to assess the metabolic state of tissue. In this regard, PET is commonly performed with the radiopharmaceutical 18F-fluorodeoxyglucose (FDG). FDG is a glucose analogue which is coupled to the positron emitting radionuclide ^{18}F . With this radiopharmaceutical tracer the glucose metabolism of tissue in the whole body can be imaged, as this tracer is preferably taken up in tissues with an elevated metabolism. During positron emitting, a proton is being converted to a neutron in the atomic nucleus. This process generates a positron and an electron neutrino. The positron travels a short distance and annihilates with an electron. The annihilation reaction (Figure 2) causes a formation of two high energy photons which travel in opposite directions. Two detectors at opposite ends of the PET scanner detect these two opposite travelling photons simultaneously, and the radioactivity in the body is localised somewhere along the straight line between the two detectors, the line of response.[8]

PET imaging is commonly used for diagnosis and tumour staging, given its high sensitivity for detecting primary tumours, lymph nodes, and distant metastases. Active tumour lesions usually have an elevated metabolism and subsequently show an increased uptake on FDG-PET images. For many oncological cases, quantitative metrics derived from PET images are correlated with the prognosis of the patient and degree of therapy response. Quantitative metrics in PET are usually based on the standardised uptake value (SUV) where the degree of uptake is corrected for patient's body mass and the amount of injected activity.[9, 10] Volumetric and uptake parameters such as total lesion glycolysis (TLG) and metabolic tumour volume (MTV) are also related to patient outcome.[11]

X-ray computed tomography

X-ray computed tomography (CT) imaging relies on the use of X-rays to reconstruct cross-sectional slices of a scanned patient. The technique is based on the principle that density of tissue material results in different attenuation of the X-rays. The attenuation coefficient quantifies how easily material can be penetrated by X-ray beam and represents the degree of intensity loss of the X-ray beam by the material as it passes through. CT numbers are expressed by the dimensionless quantity of Hounsfield units (HU), which expresses the linear transformation of the measured attenuation coefficients. The transformations are based on a scale of assigned densities of air and pure water.[12] With HU the density of tissue can be examined. In CT imaging, iodinated contrast can be used to generate contrast between different tissues (e.g. quantifying perfusion and vessel leakiness). In tumour imaging vasculature plays a significant role to examine the malignancy and staging of a tumour lesion.[13]

Magnetic resonance imaging

Magnetic resonance (MR) imaging is a different type of imaging relying on different behaviour of molecules in a strong magnetic field. The MR signals that provide the diagnostic information are produced in patient's tissue in response to radiofrequency (RF) pulses, generated by a transmitter coil which surrounds the whole, or a part, of the body. The produced MR signals are detected using a receiver coil. Localisation of the MR signals in the patient's body to reconstruct images is achieved by generation of short-term spatial variations in magnetic field strength across the patient. The use of stronger gradients permits smaller anatomical structures to be visualised, and enables shorter scanning time. The gradient fields are produced by three sets of gradient coils, one in each direction, through which large electrical currents are applied repeatedly in a controlled pulse sequence.[14] Routine MR imaging is commonly used to provide information about tumour shape, size, location and morphology. Imaging regarding cell density is provided using diffusion weighted MR imaging, tumour angiogenesis and interstitial space is examined using dynamic MR imaging, and integrity of the cell membrane is examined with choline MR spectroscopy. The main advantage of MR imaging is its high soft tissue contrast, making it able to clearly depict the size and shape of the tumour. Also, MR imaging can provide information about intra-tumoural hemorrhage, peri-tumoural edema and can even contribute to the prediction of histological tumour grade.[15, 16]

1.1.3. Radiomic features

Multiple functional parameters can be extracted from PET/CT, CT and MR imaging techniques. However, in a clinical setting, medical images are still qualitatively assessed by the radiologist and nuclear medicine physicians. In the recent years research has focused on extracting quantitative tumour information from medical images, providing an objective and quantitative analysis of genetic and histologic heterogeneous tumours.[17, 18] For this matter the extraction of radiomic features from medical imaging is an emerging field in oncology. Radiomics features have been shown to capture spatial and temporal heterogeneity of tumours that can be linked to patient prognosis and treatment response. With the assumption that the textural imaging features of pathologic tissue differs from healthy tissue, radiomic features can contribute to treatment decisions, tumour characterisation, prediction of therapeutic response, prognosis and overall survival. This can eventually contribute to personalise treatment and improve patient outcome.[17, 19]

Radiomics is an area in medical research which focuses on the extraction of a large amount of imaging features from medical images, which are beyond the level of detail assessed by the human eye. Radiomic features (features capturing texture, shape, and uptake/enhancement) are calculated based gray level values in a pre-defined region in an image and the spatial relationship of these values with each other. With radiomics methods, three classes of parameters can be described: first order statistics, second order statistics and higher order statistics, where the most commonly used statistics for radiomic features are the first order- and second order statistics.[4, 17, 20]

First order radiomics observes the likelihood of the presence of a gray value at a randomly chosen location within the image region. First order statistics can be calculated using the histogram of pixel intensities in the image.[20] These calculations only depend on the values of a single pixel, and the



Figure 3: Graphic view of textural second order statistics. Textural features are derived from texture matrices and the distribution of the gray level values in an image.[18] Abbreviations: GLCM = Gray-Level Co-occurrence Matrix, GLRLM = Gray-Level Run Length Matrix, GLSZM = Gray-Level Size Zone Matrix, NGTDM = Neighbouring Gray Tone Difference Matrix

interaction with other pixel values is not taken into account. Examples of first order statistic features are mean intensity, minimum intensity, maximum intensity, uniformity (homogeneity of image array), entropy (level of randomness in the image values) and skewness (asymmetry of the histogram).

Second order radiomics describe the spatial relationship of gray level values between different voxels using different types of algorithms. The algorithms apply different types of matrices in order to calculate radiomic features. The first second order algorithm calculates the occurrence of a pixel with intensity i in relationship to another pixel with intensity j (Figure 3). These statistics are calculated based on a gray level co-occurrence matrix (GLCM) and features such as correlation (gray tone linear dependencies) and entropy (randomness of the gray level distribution in the matrix) can be derived. [19, 20] A GLCM shows the distribution of combinations of discredited gray levels of neighbouring voxels along one of the image directions. A two dimensional (2D) volume has an 8-connected neighbourhood and a three-dimensional (3D) volume has a 26-connected neighbourhood with 13 direction vectors. The GLCM calculates features for every vector direction.

Besides using a co-occurrence matrix, another method for second order radiomics can be used, the run-length matrix (RLM). The run-length matrix analyses the gray level intensities and its relationships in any direction (Figure 3). A "run" is the length of consecutive pixels with the same gray level values in a specific direction. The length of this specific run is defined by a number of image points with a similar gray level value in the run in any direction. The length of the run is related to the fineness of the textures in the image. A fine texture contains more short runs and a coarse structure contains more longer runs.[4, 21] Gray level size zone matrices (GLSZM) describe the gray level distribution in a specific zone of the image. It describes the pixels which are connected by gray level values in the defined zone of the matrix. It takes into account similar gray levels and therefore represents the size of a homogeneous zone in the image (Figure 3).[4] A neighbouring gray level tone different matrix (NGTDM) describes the differences between a voxels by their voxel values (Figure 3). They analyse the gray level and the average gray level of neighbouring voxels within a certain distance defined by the matrix and compare these to the centre voxel.[22, 23]

Radiomic features can give additional quantitative information about tumour phenotype and micro-environment different from information gathered by clinical and laboratory reports. This information can be used in combination with other clinical information and contribute to clinical decisions and care (Figure 4). [24, 25]

1.1.4. Texture analysis in multiple modalities

Texture analysis for the purpose of medical imaging has been studied for decades, starting with conventional radiography and ultrasound.[4] In the last decade, the interest in texture analysis and ra-

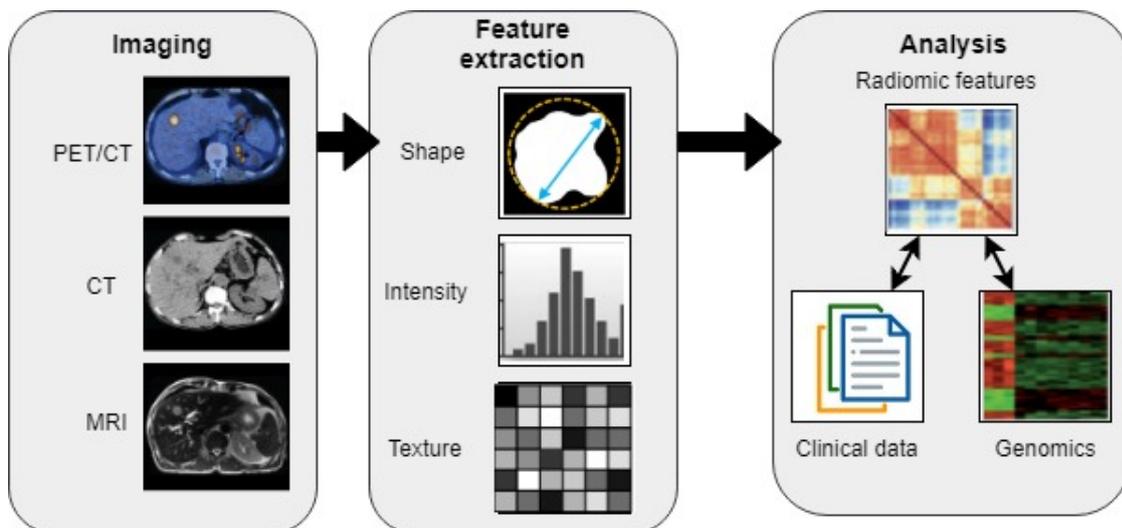


Figure 4: Workflow for extraction of radiomic features from clinical images. The first step contains receiving images of different modalities and delineation of the region of interest (ROI) from the clinical image. The next step is feature extraction within the defined ROI. In the last step the features are analysed, and combined and compared with clinical data and genomics.

radiomics features has grown towards application on FDG-PET, CT and MR imaging. Especially in the field of oncology, textural analysis has a potential advantage in clinical care shown by increased numbers of studies.[10, 19] To date, multiple studies proposed textural analysis as a method for quantification of heterogeneous metabolic activities on PET imaging with promising results. Aerts et al. showed that tumour properties as shape, texture and wavelets were predictive parameters for prognosis of patients with lung cancer.[17] Pyke et al. established that texture analysis of PET images can be used to differentiate between different stages of high-grade gliomas.[26] Also, Reuzé et al. demonstrated the value of radiomic features calculation on PET imaging for the prediction of local recurrence of locally advanced cervical cancer.[27] In a study population of non-small cell lung cancer, Ganeshan et al. found a correlation between texture analysis and FDG uptake which gave the promising result that texture analysis can give additional information about tumour staging. Furthermore, for the same patient population they established an additional value of texture analysis on CT imaging in classification of tumour staging.[28] A second study was performed by Ganeshan et al. on CT images of non-small cell lung cancer population and found association of texture features with biomarkers.[29]

In CT imaging more studies performed various radiomic features for improvement of tumour characterisation. Al-Kadi et al. showed the value of textural features in the differentiation of malignant tumours from benign lesions. The authors found that fractal analysis CT images can give more quantitative information about the aggressiveness of a tumour.[30] Grove et al. described a relation between two quantitative shape and density features and overall survival of patients with lung adenocarcinomas.[31] A study from Cui et al. performed a quantitative texture analysis of lymph nodes in rectal cancer and showed an improvement of the prediction of regional lymph node status.[32]

For imaging with MR techniques, several studies established the additional value of textural features for tissue characterisation. Zhang et al. investigated radiomic features on MR imaging for the prediction of histopathological grades in sarcomas and found significant results suggesting that radiomic features can provide predictive information.[15] Holli et al. have demonstrated that textural analysis of breast tissue can distinguish between healthy breast tissue and malignant breast tissue.[33]

So far, studies have shown the clinical promise of radiomic features. However, there are still technical difficulties to overcome. Galavis et al. demonstrated that a relatively high number of PET features have a strong variability due to differences in acquisition and reconstruction protocols. Moreover, they showed that textural features are affected by iteration numbers, matrix sizes and filter width.[34] Factors related to acquisition and reconstruction of imaging are expected to differ between medical institutions which makes repeatability and consistency of radiomics a difficult task.[18] Selection and segmentation of tumour volumes are also shown to influence the outcome of radiomic features [25, 35]. Within the extraction of radiomic features there are also differences in computation and calculation methods, which makes it difficult to compare radiomics used in different studies.[25] Another critic is that there is no confirmation that the calculated radiomic features represent the actual heterogeneous structures within a tumour because prior knowledge of tumour *in vivo* is not available. Furthermore, each modality has its own technical properties. Factors such as spatial resolution, contrast recovery, scaling, filtering and signal to noise characteristics in the image could be significantly different between different imaging modalities. There is currently limited knowledge regarding the accuracy of the features being calculated on different imaging modalities. This brings different difficulties and challenges for standardisation of clinical radiomic features. With all these variable factors in radiomics analysis, there is a strong need for harmonisation of image acquisition and reconstruction protocols, and standardisation of feature computation before full clinical implementation can be achieved.

1.1.5 Heterogeneity phantom

In order to overcome the technical difficulties of radiomic feature extraction, a primary task is to test its accuracy and variability on different imaging modalities using a phantom. With a suitable phantom heterogeneity can be simulated and used for validation and accuracy analysis. With the knowledge of the properties of the phantom, the ground truth is available and can be used as a reference. With the large amount of differences in representation of heterogeneity in tumour tissue, a phantom study

provides a method to investigate the influence of the imaging technique on the representation of tumour heterogeneity on a medical image.

In the recent years several studies have been performed using phantoms with heterogeneous structures. DiFilippo et al. studied the quantitative analysis of PET and SPECT performance with different phantoms containing multiple fillable chambers using hot spheres and different porous structures. It created different structures and distributions on the PET and SPECT images.[36] Carles et al. performed quantitative analysis on the effect of respiratory motion on PET texture analysis by using heterogeneous phantoms. Their design was composed of the biomaterial alginate and reusable molds were used for the insertion of activity concentrations. Their study established that the use of heterogeneous phantom can contribute to the analysis of image features.[37] More recently, Gallivanone et al. published results of an antropomorphic heterogeneous phantom using different structures and shapes to analyse the reproducibility and variability of radiomic features in FDG-PET imaging.[38] Furthermore, Presotto et al. studied the impact of different acquisition and calculation methods for radiomic features by creating heterogeneity within a phantom.[39] For radiomic feature calculation in MR imaging, Baessler et al. studied the reproducibility and robustness of radiomic features on different MR sequences.[40]

However, these heterogeneous phantoms were mostly used for a single imaging modality and to our knowledge no multimodality heterogeneous phantom has been introduced for the analysis of radiomic features. With the promising results of radiomics in contribution to clinical care and patient survival, there is a need for more research to the added value of radiomic features. In this study we have developed a unique platform for an heterogeneous imaging phantom suitable for PET, CT and MR imaging for quantitative analysis of heterogeneity. The aim of this study was to establish the accuracy and variability of radiomic features for heterogeneous structures on a multimodality imaging platform.

1.2 Methods and materials

1.2.1. Study design

This study was performed using PET, CT and MR imaging in combination with a newly designed and developed heterogeneity phantom. In order to visualise the structures of the phantom on PET, CT and MR imaging, the phantom was filled with ^{18}F -fluorodeoxyglycose, iodinated and gadolinium contrast agents respectively. Image analysis was performed on reconstructed images. Images were registered and segmented before being loaded into software dedicated for radiomic analysis. On each imaging modality, the phantom was scanned five times in similar conditions and scans were processed using a standardised workflow as described in Figure 8. Variability of the radiomic features was studied by examining the variation between repeated scans. The variability was also compared across different imaging modalities. Scalability was assessed by comparing results of different phantom sizes. Furthermore, the geometry of the realised phantom was measured using high resolution CT imaging. These images were used as a reference for accuracy analysis of the radiomics quantification.

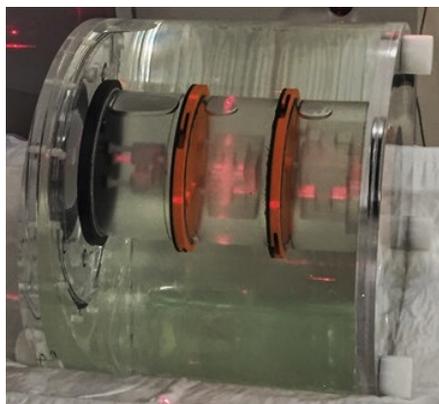


Figure 5: Heterogeneity phantom with the NEMA-NU2 body phantom as casing. The phantom is filled with water and remaining air bubbles are visualised in all three compartments.

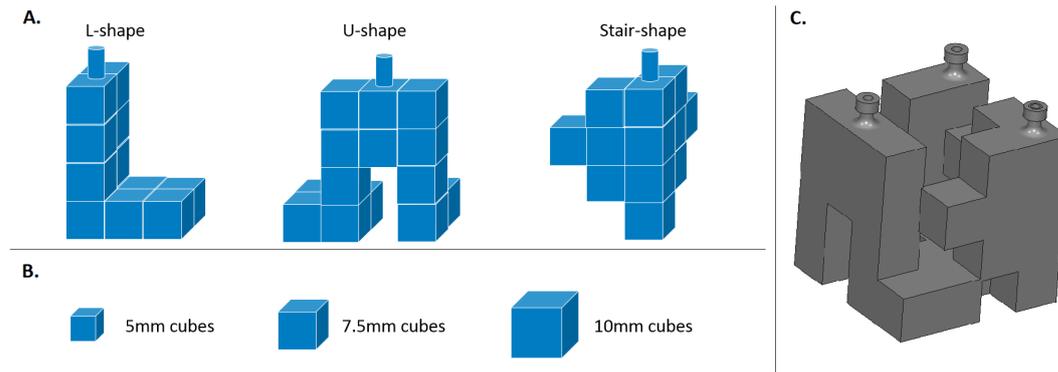


Figure 6: Cubical insertion of each cylindrical compartment of the phantom. A: The different shapes (L-shape, U-shape and stair-shape) of the inserts, all comprised of cubes with constant measures. B: The different sizes of the cubes used to form the three different insert shapes. C: The cubical construction in the cylindrical compartments when all insert shapes are interconnected with each other.

1.2.2 Phantom design

In order to simulate heterogeneous uptake patterns, a cylindrical multi-compartment phantom was designed and developed for multimodality imaging. The phantom is comprised of polymethyl methacrylate (PMMA) (outer casing), in which compartments consisting of polylactic acid (PLA) and photopolymers were inserted. The compartments were created using three-dimensional (3D) printing techniques. All materials were suitable for multimodality imaging and the phantom was designed to be reusable. The phantom was designed to fit in the casing of the NEMA-NU2 body phantom, routinely used for quality assessment of PET images. This body phantom was used as a casing to create a suitable background in which the heterogeneity phantom can be placed. With this requirement the heterogeneity phantom was developed with a maximum height of 202 millimetres (mm) and a cylindrical diameter of 80 mm.

The global design of the phantom is shown in Figure 5. The phantom consists of three separate cylindrical compartments on top of each other (interlocked). Each compartment contains rectangular shaped compartments (L-shape, stair-shape and U-shape), and are closed by specially designed top and bottom plates. The top plate has four fill ports which serve as the entrance to the inner parts of the compartments and can be closed by specially designed caps. Both the top and bottom plate have a specially designed mechanism to interlock the segments with each other.

The design of the inserts was based on a publication of the image biomarker standardisation initiative (IBSI) [41]. This digital phantom concept focused on the design of three differently shaped hollow compartments connected to each other, each with a different homogeneous concentration.[41] With this example, a phantom was created with different cubical shapes and the different structures are referred to as the L-shape, the stair-shape and U-shape (Figure 6A). The inserts are connected to each other by walls of 0.4 mm, and the three inserts together form a larger cubical construction (Figure 6C). The cubical insertion shapes are similar in all three cylindrical compartments, but the size of these shapes are different. The first compartment has inserts shaped by cubes with measures of 5 mm, the second compartment has inserts with 7.5 mm cubes, and the third compartment contains cubes with 10 mm measures (Figure 6B). With these three different sized constructions, a scalable phantom is designed.

1.2.3 Image acquisition

As this study was performed on different imaging modalities, the phantom was filled with different concentrations of contrast agent or radiopharmaceutical, depending on the used imaging modality. To make comparison between imaging modalities possible, similar contrast to background concentrations were used to fill the phantom compartments. Concentrations within the phantom were specified relative to the background concentration. This concentration was appointed as ratio 1 and used to fill the NEMA body phantom. The concentrations of the different compartments were 2 (cylindrical compartments): 4 (L-shape): 8 (stair-shape): 16 (U-shape) with respect to the background concentration,

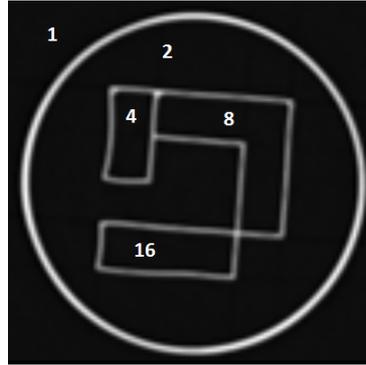


Figure 7: Axial view of a cross section of the phantom within a cylindrical compartment. The numbers represent the concentration ratios in each compartment of the phantom.

shown in Figure 7. Image acquisition was repeated 5 times for each imaging modality. With each acquisition, the phantom was moved slightly in the scanner to simulate re-positioning of the patient.

FDG-PET/CT imaging acquisition

^{18}F -FDG PET/CT imaging was performed using a Philips Vereos Digital PET/CT scanner (Philips Medical Systems). The amount of inserted ^{18}F -FDG was defined by EARL accreditation protocol. Therefore, the NEMA body phantom, the background, was inserted with a concentration of 20 kBq/9.7ml (2.1 kBq/ml) ^{18}F -FDG, and filled up with water. This concentration was defined as ratio 1, and the other compartments were filled with concentrations of 4.1 kBq/ml (ratio 2), 8.2 kBq/ml (ratio 4), 16.5 kBq/ml (ratio 8) and 33.0 kBq/ml (ratio 16). Image acquisition was performed using a standard whole body imaging protocol using 2 bed positions. Image acquisition time was 15 minutes per bed position of the first scan. The next four image acquisitions were performed with a 3 minute image acquisition time (per bed position). Image reconstruction was performed using an EARL reconstruction protocol with 4 mm slice thickness (4x4 mm pixel spacing). Low dose (LD) CT acquisition was performed with 2 mm slice thickness and a pixel spacing of 1.17x1.17 mm. A peak voltage of 120 kilo Volt (kVp) with fixed tube current of 40 milliampere per second (mAs^{-1}) was performed. Reconstructions were performed by the EARL reconstruction protocols, with an ordered-subsets expectation maximisation (3 iterations and 15 subsets) with CT attenuation correction and 5 mm post reconstruction Gaussian filter.

CT acquisition

CT imaging was performed using a Toshiba Aquilion ONE scanner (Canon Medical Systems Europe B.V.). The phantom was filled with water containing different concentration of iodinated contrast agent (Xenetix, 350 mg/ml). The background concentration was defined by clinical protocol for administration of contrast agent in liver abdominal examinations, considering a patient of 70 Kg. For clinical abdominal imaging protocols, the amount of contrast agent used is 1.9 ml/kg, resp. 1.9 ml/L. As the volume of the NEMA body phantom is known to be 9.7 L, the NEMA phantom was inserted with 19 ml contrast agent. The compartments were filled with 3.7 ml/L (cylinder, ratio 2), 7.4 ml/L (L-shape, ratio 4), 14.9 ml/L (stair-shape, ratio 8) and 29.7 ml/L (U-shape, ratio 16). Images were acquired with 5 mm slice thickness (pixel spacing 0.782x0.782 mm), a FC08 convolution kernel, a pitch of 0.8 and a rotation time of 0.5 s. A fixed tube current of 117 mAs^{-1} and fixed voltage of 120 kVp were used. The CT scans were reconstructed with a FC18 convolution kernel and 1 mm slice thickness. Noise reduction was performed using Adaptive Iterative Dose Reduction (AIDR).

MR acquisition

MR imaging was performed using a Philips Ingenia 1.5-Tesla MR scanner (Philips Medical Systems). For MRI acquisition the phantom was filled with a solution of water and Dotarem 0.5 millimoles per millilitre (mmol/ml), a contrast agent for intravenous administration in clinical MR imaging. Approved dose for Dotarem in clinical practice for MRI examinations is 0.1 - 0.3 millimoles per kilogram

(mmol/kg) [42]. With the assumption that a filled NEMA phantom weighs approximately 10 kg (volume 9.7L), 2 ml contrast agent was inserted as background ratio. Compartments were filled with 0.41 ml/L (cylinder, ratio 2), 0.82 ml/L (L-shape, ratio 4), 1.65 ml/L (stair-shape, ratio 8) and 3.3 ml/L (U-shape, ratio 16). Clinical protocol for liver examination was followed to define the sequence used for image acquisition. A standard body coil was used as transmitter and receiver coil. Liver tissue was used as a reference tissue and the relative signal-to-noise ratio (SNR) was 2.35. Five 3D gradient fast field echo (FFE) with T1 weighted images (TR/TE=7.6/4.6 ms; 10° flip angle; 2 mm slice thickness; 1.33x1.33 mm pixel spacing; 320x320x240 mm field of view) were performed with an acquisition time of 3 minutes and 39 seconds per image.

1.2.4 Image data processing

Acquired images were extracted from the clinical system in Digital Imaging and Communications in Medicine (DICOM) format. Before calculation of radiomic features could be performed, multiple data processing steps were required. For the processing of the images, three primary steps were considered: image cropping and registration, image segmentation and creation of the ground truth. The workflow for data processing and extraction of radiomic features is shown in Figure 8.

Image registration

Before registration and segmentation was performed, the acquired data needed to be prepared. Given that the orientation of the imaged phantoms was different during imaging on different imaging modalities, registration was an essential step to align the images. Therefore, the images were cropped in three separated scans with equal sizes, each with one of three compartments depicted. The images were cropped using 3D Slicer (version 4.10.1, Slicer Solutions, Boston, United States). As image acquisition was performed five times on every modality, this resulted in fifteen (N=15) scans for every modality (Figure 8). The cropped scans were used in further image processing and analysis.

As each modality has its own coordinate system and different orientation, registration of the images was performed. The high resolution CT images (in which the phantom casing was imaged filled with air) were set as the reference for co-registration (Figure 8), as these images are used for the ground truth. The cropped images were registered using an Imalytics 3.2 (Philips Technology, Aachen, Germany). Imalytics is a research workstation which allows multimodal registration. The software has multiple registration tools and automatic registration algorithms. The software relies on Elastix, a tool for multimodal image registration tasks which supports multiple registration tools.[43] Registration was applied in two steps; manual registration and automated rigid registration or semi-automated rigid registration. Before automated registration was applied, manual translation and rotation needed to be performed as the automated registration was limited. After manual registration, rigid automated registration of CT- and MR images to the ground truth CT scan was applied. For the PET images this automated registration was not feasible, and semi-automated registration was performed. Registered images were extracted as DICOM files.

Segmentation

Segmentation of the images is an important step in radiomics analysis as features are extracted from the defined segmentation area. By defining a region of interest (ROI) irrelevant structures can be excluded from the feature extraction and more meaningful analyses can be performed.[24] In every image a cylindrical ROI of similar size is drawn. An ROI is drawn in the first and last axial slice of the ROI, and after interpolation a volume of interest (VOI) is created. In the high resolution CT images for the ground truth, three volumes of interest (VOIs) were segmented: one for the 5 mm sized insertion, one for the 7.5 mm sized insertion and one for the 10 mm sized insertion. Because all images were registered to the same CT images, the VOIs were suitable for every scan and were used for all images.

Ground truth

Radiomic features are widely used for quantification of heterogeneous patterns in clinical images. However, with patient data knowledge of the true underlying heterogeneous patterns in a lesion is

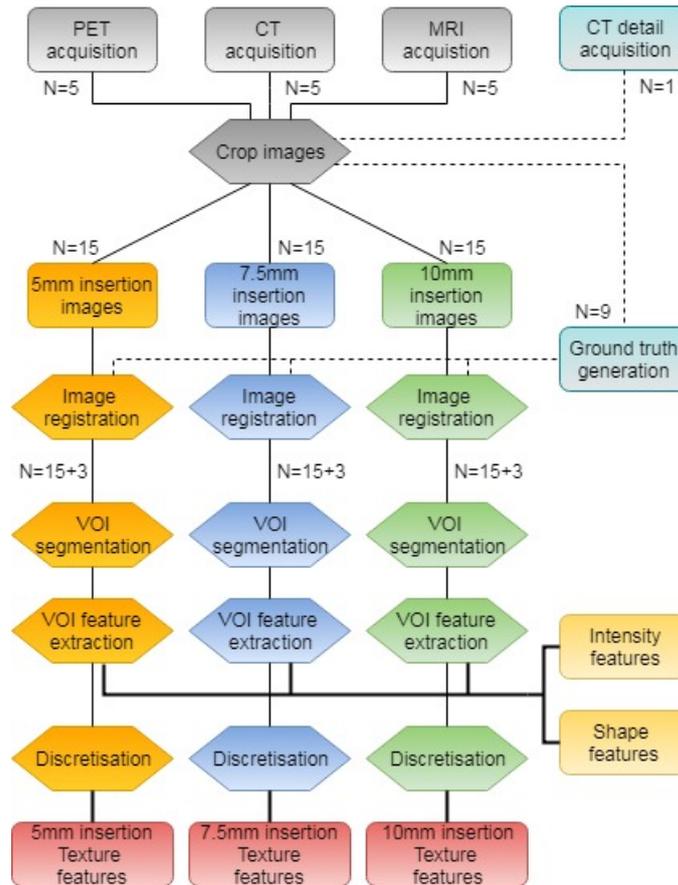


Figure 8: Pipeline of radiomic feature extraction. Images were acquired five times for each modality (N=5). Images are cropped and each scan is divided in three separate images (N=3x5=15). After acquisition of the images for the ground truth, for each modality and each phantom insert size different ground truths are generated (N=9). For each scan a similar workflow continues for the extraction of radiomic features.

usually unavailable. Therefore, there is uncertainty about the imaged structures and computed textural features. This can be solved with the current phantom study by creating a ground truth from high resolution imaging. The ground truth is used to test the accuracy and variability of the radiomic features extracted from different imaging techniques.

The ground truth of this phantom was generated using a highly resolution CT acquisition using a Toshiba Aquilion ONE scanner (Canon Medical Systems Europe B.V.). The images were acquired with 1 mm slice thickness, a fixed tube current of 100 mA and a fixed voltage of 120 kV. The images were reconstructed using 0.5 mm slice thickness with a pixel spacing of 0.206x0.206 mm. The compartments of the phantom were filled with air to optimise the contrast between background (air) and the phantom walls. With the detailed CT imaging accurate depiction of the phantom structure and walls (0.4 mm) was acquired for the generation of the ground truth. After image acquisition the ground truth images were divided in three cropped images: images with the 5 mm insertion, with the 7.5 mm insertion and with the 10 mm insertion. To this end, each modality had three ground truths and a total of nine (N=9) scans was created (Figure 8).

The next step for the generation of the ground truth was to assign voxel values to the different structures. These values are derived from concentrations of the compartments which are different for each modality (PET, CT and MRI). The different ratios are derived from the background value in each modality. For determination of the voxel values for the ground truth, the value of the voxels comprising the background compartment of the NEMA body phantom were measured. In this regard a large ROI was defined in the background and the mean voxel value was recorded. This value was multiplied using the contrast ratios to obtain the voxel values of the different compartments.

The ground truth was generated using specially written scripts implemented in a graphical user interface (GUI) in MATLAB 2018b (Mathworks Inc, Natick, Massachusetts, USA). Cropped CT images

were segmented using a threshold (upper level 1000, lower level -200). After segmentation all different compartments and inserts were distinguished and the voxel values were assigned to each structure. Besides the image values for the structures with concentrations, there was a value assigned to the walls of the phantom. In the PET scans the wall values were assumed to be zero, as the walls contained no activity. For the wall value in CT images, a solid cube of 2x2x2 centimetres (cm) was printed and scanned using the same imaging protocol. The mean value of this area was used as wall value. The wall value in the MRI ground truth was defined using the same method as the background image value, selecting an ROI in an MRI scan. The ground truth images were extracted as DICOM files and processing continued through the workflow shown in Figure 8.

1.2.5 Feature extraction

Image acquisition, pre-processing and radiomic feature extraction were reported according to reporting guidelines described by the Image Biomarker Standardisation Initiative (IBSI) [41]. This guideline can be found in appendix B. The guideline also contains Figure 8, the workflow of the radiomic feature extraction.

After creation of a VOI, feature extraction is the next step. Quantitative radiomic features were extracted from all modality images with the free software PyRadiomics 2.0 using Python 3.6 (Python Software Foundation, Wilmington, Delaware, United States). For every VOI a total of 107 shape-, intensity- and textural features were extracted (Appendix A). Image normalisation and distance weighting were not applied. GLCMs were calculated in thirteen directions (26-connectivity) and were by default symmetrical. Features were divided in seven different classes; shape features, first order features, Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Gray Level Dependence matrix (GLDM) and Neighbouring Gray Tone Difference Matrix (NGTDM). Features were extracted from the modalities and the ground truth using the same parameter file (Appendix C).

For the extraction of texture features, gray values discretisation was applied using a fixed number of 64 bins in all three image modalities.[44] Leijenaar et al. showed that using a fixed bin width may be more appropriate in clinical studies, as features are more stable for change in a fixed bin width than for a change in number of bins [45]. However, this holds true in cases of images with absolute gray values (e.g. HU in CT and SUV in PET). In cases of relative gray values (e.g. signal intensities in MR) this is not evidently the case. Therefore, a fixed bin count is used in this study to ensure radiomic features for all modalities, including MRI. However, it needs to be considered that the range of gray levels can still influence the calculated features.

This study was an explorative study to analyse the radiomic features, and by default a considerably large number (107) of features is extracted. Feature reduction was applied based on correlation, their relevance and literature.[18, 44, 46] Shape features were excluded from analysis as they describe the properties of the entire VOI, e.g. size and 2D/3D diameter, which is equal for each VOI in this study. In clinical cases these features are relevant, however within this study they are mainly useful for confirmation if VOIs are properly extracted. Therefore, they were excluded from further analysis.[24] In feature extraction five different classes were calculated using matrices (GLCM, GLRLM, GLSZM, GLDM, NGTDM), based on different methods. However, some feature classes are strongly correlated. The GLDM is defined as an alternative for the GLCM matrix. It is based on the neighbourhood of a single central voxel, similar to the GLCM method.[23, 41] Therefore, the GLDM class was excluded from analysis. Within the classes, a single feature was selected to represent multiple features which are correlated to one another. For multiple features the normalised value of the feature was used for analysis given that these features are less sensitive to the absolute value of the voxels in the different imaging modalities, which can vary to a large extent.

1.2.6 Statistical analysis

Statistical analysis was performed using SPSS (version 25, IBM Corporation, Armonk, New York, United States) and Microsoft Excel 2016 (Microsoft Corporation, Redmond, Washington, United States). The variability between different image acquisitions (N=5), scalability between different phantom sizes, and accuracy, by comparing to the ground truth, of the features were analysed.

The mean and standard deviation were calculated for each modalities and each insert size. The coefficient of variation (CV), the ratio of the standard deviation to the mean, was used to assess the variability of the five image acquisitions in a single modality. Wahl et al. studied the quantitative assessment of FDG-PET scans with a test-retest study. In a test-retest study they described significant changes in tumour SUV, with a 20% change in SUV in a region with a diameter of 1 cm or larger. Clinical significance was assessed and changes of 30% are described as clinically relevant. As assumed that the percentage for significant difference is largest in PET imaging, a value of 20% or higher is defined as significant variable. In this phantom study radiomic features were all extracted from the PET, CT and MR images using a similar VOI. Therefore, the shape characteristics are not expected to significantly differ from each other after repeating the imaging protocols. The CV of the shape features was established to be less than 1%. Therefore, the features with a CV of 1% or lower were defined as significant robust features.

For the analysis of the scalability using radiomic features, a Kruskal-Wallis H test was performed. The Kruskal-Wallis H test is non-parametric statistical test based on ranking principle. It determines whether there are significant differences between two or more groups. A P-value of less than 0.05 was considered to be statistically significant.[47] Furthermore, significant differences between the insert sizes (5 mm, 7.5 mm and 10 mm phantom sizes) were analysed by the coefficient of variation.

In order to analyse the accuracy of the radiomic features, comparisons between the ground truth and the PET, CT and MR images were made. The value of the extracted features were normalised to the values of the ground truth. Therefore, the value for each feature of the ground truth was defined as 1, with a standard deviation of 20% following the significant level used with the analysis of variability of the phantom. The distribution of the normalised mean value and standard deviation are used for the analysis of the accuracy of the extracted features of each modality towards the extracted features from the ground truth.

1.3 Results

1.3.1 Image acquisition

Image acquisition was performed five times on PET-, CT- and MR imaging. An axial view of image acquisition in each modality is shown in Figure 9. The PET images are shown after fusion with its corresponding low dose CT scan. Each cylindrical compartment with different insert sizes is shown separately. The walls and contours of the phantom are visible in the CT and MRI images. In the PET imaging the phantom walls of the cubical inserts are not visible. The contours of the cylindrical compartments are distinguishable in the PET-, CT- and MR images. Furthermore, the acquisition results of all imaging modalities show a considerable amount of noise, particularly on the CT and PET/CT images. The insert sizes contained different concentration ratios of contrast agent or FDG. In CT and MRI this is shown in different gray window levels. In the PET/CT imaging the window level is defined by a SUV colour bar and differences are shown between background, cylindrical compartments and the cubical insert. However, differences between the interconnected inserts are less distinguishable.

During the procedure of filling the phantom the aim was to fill the phantom completely with fluid. However, air bubbles remained in several compartments. The black ellipses in the upper parts of the cylindrical compartments represent the remained air bubbles. Each cylindrical compartment contained remaining air bubbles, shown in Figure 9. The CT image with the 5mm insert size shows no air bubble but this air bubbles is situated on different slices.

1.3.2 Generation ground truth

The background values (ratio 1) for the ground truth were defined as 2100 Bq/ml for PET, 18 HU for CT and 700 arbitrary units (au) for MR. These values were used for the calculation of the values in the other compartments, together with the absolute ratios. The absolute ratios are defined by the concentrations used during filling procedure of the phantom. The calculated absolute ratios and their values for each modality are shown in table 1.

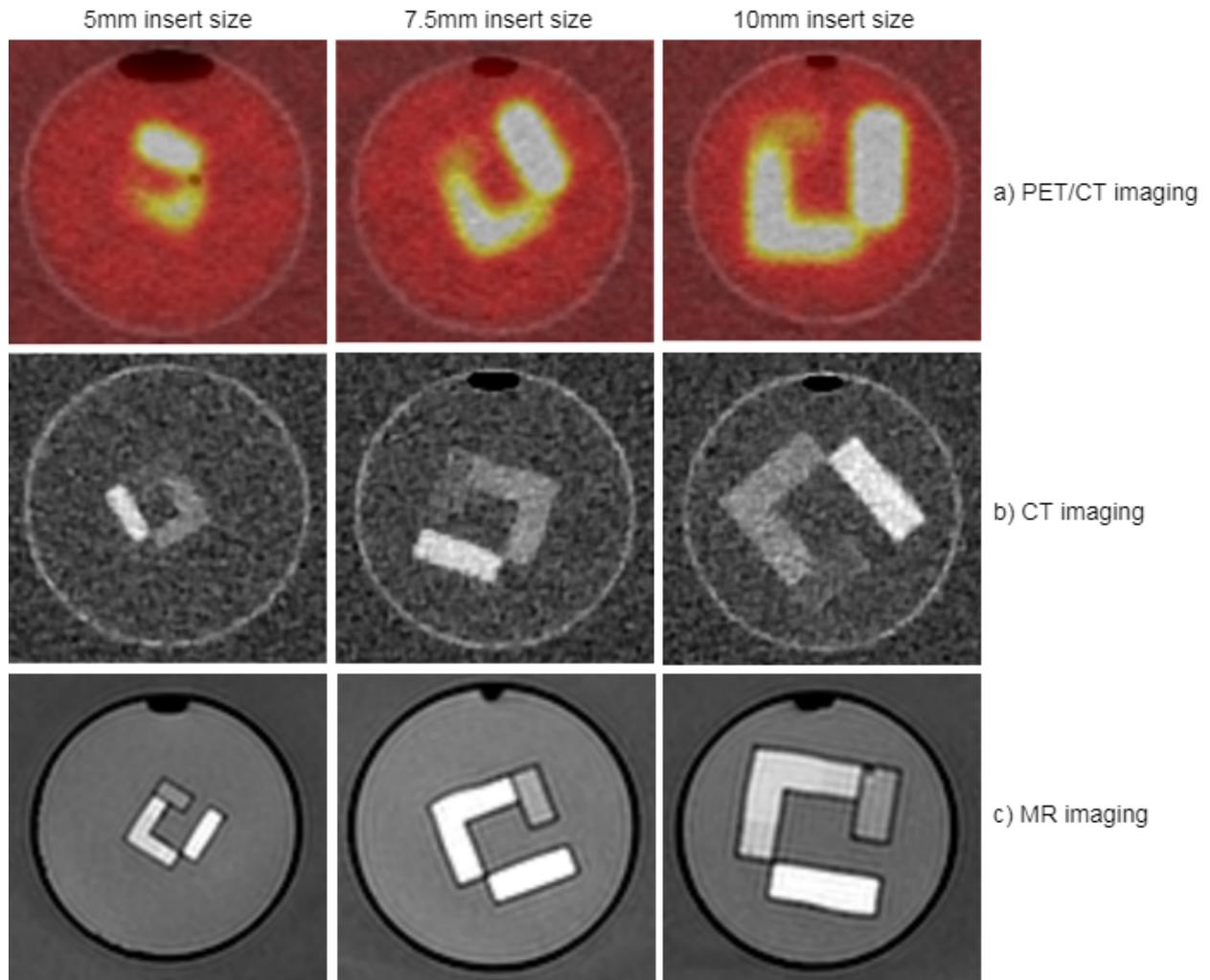


Figure 9: Result of image acquisition for PET/CT, CT and MRI. The black ovals in the upper part of the cylindrical compartments represent air bubbles. a) Results of the PET/CT imaging of the scaled inserts. The results are shown using a 0-40% SUV scaled colour bar. b) CT imaging of the scaled inserts with iodinated contrast shown in gray window levels. c) MR imaging of the scaled inserts with gadolinium contrast. These results are shown in gray window levels.

Table 1: The assigned and absolute ratios for each compartment of the phantom. The ground truth value of the compartments and structures within the phantom are shown in the last column. Values are displayed for PET in Becquerel per millilitre (Bq/ml), for CT in Hounsfield units (HU) and for MR in arbitrary units (au).

Modality	Ratio	Absolute ratio	Ground truth value
PET	1	1.0	2100 Bq/ml
	2	2.0	4200 Bq/ml
	4	3.7	7770 Bq/ml
	8	7.5	15750 Bq/ml
	16	14.9	31290 Bq/ml
	Wall		0 Bq/ml
CT	1	1.0	18.0 HU
	2	1.7	30.6 HU
	4	3.5	63.0 HU
	8	6.9	124.2 HU
	16	13.8	248.4 HU
	Wall		110.0 HU
MRI	1	1.0	700 au
	2	2.1	1470 au
	4	4.3	3010 au
	8	8.5	5950 au
	16	17.0	11900 au
	Wall		150 au

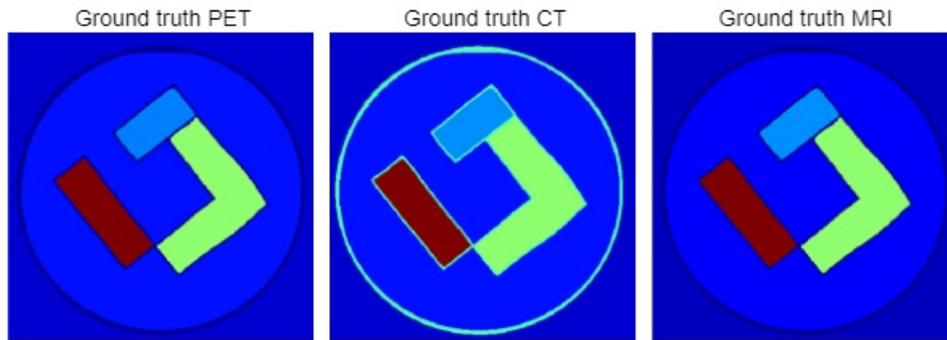


Figure 10: Axial view of the generated ground truths for each modality. For each ground truth a slice of the cylindrical compartment with the 10 mm insert size is visualised. The background (ratio 1), the cylindrical compartments (ratio 2) and the L-shape (ratio 4) are displayed in blue colours. The stair-shape (ratio 8) is visualised in green and the U-shape (ratio 16) is displayed in red. The phantom walls are also visualised in blue.

With the values displayed in table 1, the ground truths of each modality and each insert size were created and saved as DICOM files. An axial slice of the ground truth of each modality, for the 10 mm insert size, is shown in figure 10.

1.3.3 Statistical analysis

With the used radiomics software it was possible to extract 107 features from the images. Before statistical analysis was performed, the amount of features was reduced. The 14 shape features were not included in the analysis because the same VOI is used for all images, which ensures equal sizes and measures. The first order features were reduced to 10 features, and 8 features were excluded based on literature or correlation with each other. The GLCM feature class contained several entropy and energy features, and one of these features was chosen to represent the entropy and energy within this class. The entire GLDM feature class was excluded from analysis given that this feature class is defined as an alternative class for the GLCM features. Furthermore, features (within all feature classes) which also have a normalised value, only the normalised features were included, given that these features are less sensitive to the absolute value of the voxels in different imaging modalities.

The number of features was reduced to 54 features, including the first order class (10), GLCM class (13), GLRLM class (13), GLSZM class (13) and NGTDM class (5). Statistical analysis of the radiomic features was performed with three major factors taken in consideration: variability of the radiomic features, effect of scalability of the phantom and the accuracy of the radiomic features.

Variability

With the calculation of 54 features on every scan, and each scan contained one insert size, this resulted in a CV for a total amount of 162 features on a single modality. Results of CV for every modality and phantom insert size are shown in the table in table 2. The results for PET, including all insert sizes, show 28 features with a CV larger than 20%, for CT there were 17 features with larger than 20% and for MRI there was one feature with a CV larger than 20%. For PET imaging the GLRLM and GLSZM classes had the highest number of features with a CV larger than 20%, 10 and 13 features respectively. The first order class and NGTDM class had both one feature with a CV larger than 20%, and the GLCM class had two features with a CV larger than 20%. Furthermore, there were five features in the GLRLM and GLSZM class with a CV value larger than 20% for all inserts.

Among the calculated CV of the features on the CT scans, 13 of the 17 features with a percentage higher than 20% belonged to the GLSZM class. In this class there was one feature with a CV larger than 20% in all three inserts. Three of the 17 deviating coefficients were from the GLRLM class, one from the GLCM class and one from the first order class. There was one CV among the MRI features with a percentage larger than 20%. This was the cluster prominence feature and belonged to the GLCM class of the 7.5 mm phantom insert size.

Table 2: Results of the coefficient of variation for the test-retest (N=5) radiomic features for every image modality and insert size. (red = very high variability (>30%), yellow = high variability (20-30%), light green = low variability (10-20%) and green = very low variability (0-10%), dark green = robust features (<1%))

Class	Feature	PET imaging			CT imaging			MR imaging		
		5 mm	7.5 mm	10 mm	5 mm	7.5 mm	10 mm	5 mm	7.5 mm	10 mm
		CV (%)	CV (%)	CV (%)	CV (%)	CV (%)	CV (%)	CV (%)	CV (%)	CV (%)
First order	Energy	17.58	17.54	19.85	1.86	1.07	1.25	1.80	2.58	1.51
	Entropy	2.34	0.81	1.81	0.83	1.94	2.25	3.76	1.52	1.34
	Kurtosis	5.41	4.32	1.95	2.16	1.50	2.52	2.07	2.77	0.61
	Maximum	9.54	8.21	11.07	4.33	3.62	1.96	2.72	3.09	3.29
	Mean Absolute Deviation	7.97	9.97	10.01	3.01	1.45	0.98	2.61	1.93	2.39
	Range	9.22	8.24	9.66	2.83	8.56	2.73	6.12	6.64	4.62
	Root Mean Squared	8.76	8.86	10.16	0.94	0.52	0.62	0.89	1.29	0.75
	Skewness	3.10	2.73	2.30	47.91	1.96	3.63	2.26	3.27	0.91
	Uniformity	16.93	9.76	9.66	2.07	6.45	5.57	10.99	4.30	7.89
	Variance	16.17	18.69	20.21	3.87	2.37	1.96	8.02	4.46	4.17
GLCM	Autocorrelation	7.21	2.97	8.45	1.34	13.44	0.49	9.72	6.97	3.23
	Cluster prominence	9.43	7.68	8.19	3.59	30.00	10.01	10.34	22.51	11.62
	Contrast	5.43	6.59	6.15	5.23	12.31	11.64	6.79	8.45	3.02
	Correlation	0.93	1.17	0.80	2.22	1.49	1.17	3.35	2.65	1.18
	Difference Variance	4.84	6.99	7.12	3.11	12.20	9.36	6.70	8.14	3.61
	Inverse Difference Moment Normalized	0.07	0.09	0.08	0.003	0.06	0.01	0.03	0.06	0.02
	Inverse Difference Normalized	0.22	0.18	0.17	0.06	0.29	0.11	0.09	0.17	0.05
	Information Measure Correlation 2	0.24	0.34	0.36	2.14	1.46	0.91	1.92	0.77	0.83
	Joint Energy	38.15	31.76	19.14	5.96	12.74	13.58	13.14	6.81	11.56
	Maximal Correlation Coefficient	1.17	1.05	2.00	0.89	0.64	0.53	1.04	0.82	0.57
	Sum Average	6.01	0.97	6.96	0.67	8.23	0.25	5.00	4.02	1.60
	Sum Entropy	1.78	0.60	0.94	0.33	1.70	1.61	3.11	1.02	1.42
	Sum Squares	6.01	4.96	3.89	2.02	14.80	5.37	4.93	9.65	5.21
	GLRLM	Gray Level Non Uniformity Normalized	9.20	2.45	7.08	1.49	6.12	3.92	8.78	3.79
Gray Level Variance		6.09	5.47	3.39	2.76	14.64	4.93	4.26	9.62	5.34
High Gray Level Run Emphasis		6.73	4.74	6.99	1.36	13.23	0.53	9.37	6.01	3.18
Long Run Emphasis		10.41	7.43	5.62	6.35	2.38	12.14	7.35	9.10	4.50
Long Run High Gray Level Emphasis		7.18	4.29	6.93	7.11	15.02	12.20	14.69	12.33	5.02
Long Run Low Gray Level Emphasis		42.39	20.76	34.12	6.24	20.63	11.74	8.44	13.24	5.97
Low Gray Level Run Emphasis		36.55	12.84	23.51	1.26	21.34	1.32	13.55	10.69	2.23
Run Length Non Uniformity Normalized		3.85	1.21	1.04	3.30	1.43	5.57	3.47	2.13	1.73
Run Percentage		2.78	1.49	1.01	2.30	0.82	4.13	2.53	2.73	1.19
Run Variance		32.67	24.80	20.45	9.92	8.83	18.90	10.21	13.03	8.33
Short Run Emphasis		1.71	0.56	0.46	1.77	0.61	3.02	1.80	1.02	0.80
Short Run High Gray Level Emphasis		6.78	4.91	6.98	1.50	12.81	2.75	9.30	5.43	2.73
Short Run Low Gray Level Emphasis		36.83	13.55	22.58	1.18	21.56	4.13	15.18	11.14	2.42
GLSZM		Gray Level Non Uniformity Normalized	4.69	4.94	2.17	11.29	4.52	8.66	3.69	5.10
	Gray Level Variance	3.14	4.78	3.81	7.12	12.58	4.81	3.27	7.73	4.25
	High Gray Level Zone Emphasis	7.51	5.68	6.19	1.56	11.62	0.33	6.82	3.47	2.62
	Large Area Emphasis	43.41	15.67	48.07	11.68	56.70	21.36	7.36	7.93	8.03
	Large Area High Gray Level Emphasis	37.56	13.61	27.59	12.36	81.04	21.49	15.81	12.89	7.91
	Large Area Low Gray Level Emphasis	62.78	31.95	106.58	11.10	42.46	21.22	7.64	12.04	9.60
	Low Gray Level Zone Emphasis	55.43	19.97	18.26	4.16	22.32	3.42	18.95	11.47	2.98
	Size Zone Non Uniformity Normalized	8.57	4.92	3.38	7.36	2.09	7.40	3.70	1.94	2.70
	Small Area Emphasis	4.48	2.58	1.69	3.93	1.13	4.27	1.93	0.98	1.57
	Small Area High Gray Level Emphasis	10.22	7.71	5.21	3.63	10.92	4.24	9.09	3.31	2.55
	Small Area Low Gray Level Emphasis	43.86	26.57	28.02	21.45	22.17	20.81	12.50	9.81	6.98
	Zone Percentage	3.25	5.22	5.12	9.27	10.54	12.30	6.29	5.92	2.44
	Zone Variance	46.19	16.14	50.42	11.66	56.77	21.35	7.39	7.93	8.05
	NGTDM	Busyness	15.23	5.96	24.15	4.58	6.94	5.62	8.77	7.82
Coarseness		6.40	5.69	9.16	2.91	5.22	3.19	12.22	7.16	6.02
Complexity		8.11	10.45	4.30	7.97	12.51	6.85	6.11	10.50	4.20
Contrast		9.22	4.97	7.51	7.10	12.02	14.73	10.44	9.26	4.27
Strength		4.58	5.99	3.07	9.81	13.76	13.12	10.34	6.88	3.60

Amongst the PET features multiple features have a CV less than 1%. In the first order class the entropy feature of the 7.5 mm insert has a CV less than 1%. The GLCM class has 14 features with a CV less than 1%, 4 features of the 5 mm insert, and 5 features of both the 7.5 mm and 10 mm insert. The GLSZM- and NGTDM class contained no features with a CV less than 1%. For the CT features 5 features in the first order class contained a CV of less than 1%. In the GLCM class 14 features have a CV less than 1%, 5 features of the 5 mm insert, 3 features of the 7.5 mm insert and 6 features of the 10 mm insert. In the GLRLM class the features high gray level run emphasis (10 mm insert), run percentage (7.5 mm) and short run emphasis (7.5 mm) have a CV less than 1%. In the GLSZM class the feature high gray level zone emphasis (10 mm) is the only feature with a CV less than 1%. Among the MRI features the first order class has 4 features with a CV less than 1%, 1 feature in the 5 mm insert and 3 features of the 10 mm insert. In the GLCM class 10 features have a CV less than 1%, 2 features of the 5 mm insert and 4 features of both the 7.5 mm insert and 10 mm insert. The short run emphasis feature of the 10 mm insert in the GLRLM class and the small area emphasis feature of the 7.5 mm insert in the GLSZM class have a CV less than 1%. The NGTDM class contains no features with a CV less than 1%.

Scalability

Table 2 shows a CV percentage larger than 20% for 13 features of the 5mm insert size, 16 features of the 7.5 mm insert size and 16 features of the 10 mm insert size for all modalities. Among the 5 mm insert size there are 12 of the 13 features with a CV larger than 30%. Among the 7.5 mm insert size features there are six of the sixteen features, and among the 10 mm insert size there are 4 of the sixteen features with a CV larger than 30%.

For the 5 mm insert size there is one CV of the CT features and there are 6 of the PET features in the GLSZM class with a CV larger than 20%. In the 7.5 mm insert size six of ten GLSZM features have a CV larger than 20% in CT. For PET this amount is two of five features. Looking at the CV's of the 10 mm insert size, all five CT features with a CV larger than 20% are situated in the GLSZM class. For PET features this number is five of eleven features.

Results of the Kruskal-Wallis test, summarised in table 3, show that most features have a P-value of less than 0.05. For PET imaging this occurred for 70% of all features, and for CT and MR imaging this percentage was 98%. Looking at the features extracted from PET imaging, 16 features show P-values larger than 0.05, in the GLCM- (7), the GLRLM- (3), the GLSZM- (5) and the NGTDM (1) class. In CT imaging there was one feature, in the first order class, which shows a P-value larger than 0.05. The MRI features contained two values with an estimated significance P-value larger than 0.05, both situated in the GLSZM class. The features which have a P-value larger than 0.05 are different for PET, CT and MR, meaning no feature has a significance level larger than 0.05 in more than one modality.

Accuracy

The accuracy is visualised with the values of the features normalised to the values of the ground truth. Graphs in Figure 11A, 12A and 13A show the result for the 5 mm, 7.5 mm and 10 mm insert sizes for the first order and GLCM class for each modality. The other graphs (Figure 11B, 12B, 13B) show these results for the GLRLM-, GLSZM- and NGTDM class.

In all modalities the GLCM class has most features within the normalised range when compared to the ground truth. Of these features, there are four with a value within the range for every modality as well as every insert. These features are the inverse difference moment normalised (IDMN), the inverse difference normalised (IDN), the informational measure of correlation 2 (IMC2) and the maximal correlation coefficient (MCC). Furthermore, the correlation feature falls within the range for almost all modality and insert sizes, except for the 5 mm CT- and 5 mm MR features.

For the PET features a total of 22 normalised features lie in the 80-120% range around the ground truth, concerning all feature classes. For CT features this regards a total of 28 features, and for MRI a total of 24 features. In both PET, CT and MRI most of these features are part of the GLCM class, 20, 15 and 18 respectively. The first order class of the PET features contains three normalised features within the +20% range, and all three are from the 10 mm insert size. In the first order class a total of ten normalised CT features have a value within the normalised range, in the 5 mm insert size (4),

the 7.5 mm insert size (3) and the 10 mm insert size (3). In the MRI features, the first order- and GLSZM class have features with multiple features within the range for all insert sizes. No feature in the NGTDM class, in either PET, CT or MRI, have a normalised value in the +20% range of the ground truth.

Outliers with a larger value compared to the ground truth are situated outside the range of the graphs. Remarkable are the values in GLSZM class with values a factor thousand larger compared to the ground truth. These values appear in all modalities and in every insert size. Furthermore, there are multiple features reaching the horizontal zero axis, meaning the ground truth has larger values compared to the features from the PET, CT and MR images.

Table 3: The results of the Kruskal-Wallis test for each insert size in each modality. A P-value of <0.05 is considered to be statistically significant. The values with a P-value>0.05 are marked.

Feature class	Feature	PET	CT	MRI
		P-value	P-value	P-value
First order	Energy	0.002	0.002	0.002
	Entropy	0.002	0.002	0.002
	Kurtosis	0.011	0.002	0.002
	Maximum	0.002	0.066	0.004
	Mean Absolute Deviation	0.002	0.002	0.002
	Range	0.002	0.002	0.003
	Root Mean Squared	0.002	0.002	0.002
	Skewness	0.007	0.002	0.002
	Uniformity	0.004	0.002	0.008
	Variance	0.002	0.002	0.002
GLCM	Autocorrelation	0.004	0.002	0.008
	Cluster Prominence	0.970	0.009	0.002
	Contrast	1.000	0.004	0.002
	Correlation	0.090	0.004	0.003
	Difference Variance	0.145	0.002	0.002
	Inverse Difference Moment Normalized	0.878	0.004	0.002
	Inverse Difference Normalized	0.007	0.008	0.009
	Informational Measure Correlation 22	0.002	0.002	0.034
	Joint Energy	0.009	0.002	0.006
	Maximal Correlation Coefficient	0.108	0.002	0.004
	Sum Average	0.009	0.002	0.015
	Sum Entropy	0.002	0.002	0.002
Sum Squares	0.395	0.002	0.008	
GLRLM	Gray Level Non Uniformity Normalized	0.002	0.002	0.003
	Gray Level Variance	0.403	0.002	0.002
	High Gray Level Run Emphasis	0.007	0.002	0.013
	Long Run Emphasis	0.125	0.004	0.002
	Long Run High Gray Level Emphasis	0.006	0.004	0.009
	Long Run Low Gray Level Emphasis	0.006	0.009	0.009
	Low Gray Level Run Emphasis	0.008	0.002	0.013
	Run Length Non Uniformity Normalized	0.009	0.004	0.002
	Run Percentage	0.021	0.004	0.002
	Run Variance	0.093	0.004	0.002
	Short Run Emphasis	0.009	0.004	0.002
	Short Run High Gray Level Emphasis	0.010	0.004	0.005
Short Run Low Gray Level Emphasis	0.008	0.002	0.011	
GLSZM	Gray Level Non Uniformity Normalized	0.007	0.004	0.340
	Gray Level Variance	0.065	0.002	0.002
	High Gray Level Zone Emphasis	0.018	0.002	0.185
	Large Area Emphasis	0.054	0.002	0.006
	Large Area High Gray Level Emphasis	0.011	0.002	0.008
	Large Area Low Gray Level Emphasis	0.009	0.002	0.002
	Low Gray Level Zone Emphasis	0.044	0.002	0.031
	Size Zone Non Uniformity Normalized	0.005	0.005	0.009
	Small Area Emphasis	0.005	0.005	0.009
	Small Area High Gray Level Emphasis	0.080	0.006	0.014
	Small Area Low Gray Level Emphasis	0.080	0.003	0.009
	Zone Percentage	0.031	0.004	0.004
	Zone Variance	0.054	0.002	0.006
NGTDM	Busyness	0.009	0.002	0.008
	Coarseness	0.002	0.002	0.003
	Complexity	0.007	0.008	0.009
	Contrast	0.970	0.002	0.008
	Strength	0.002	0.002	0.002

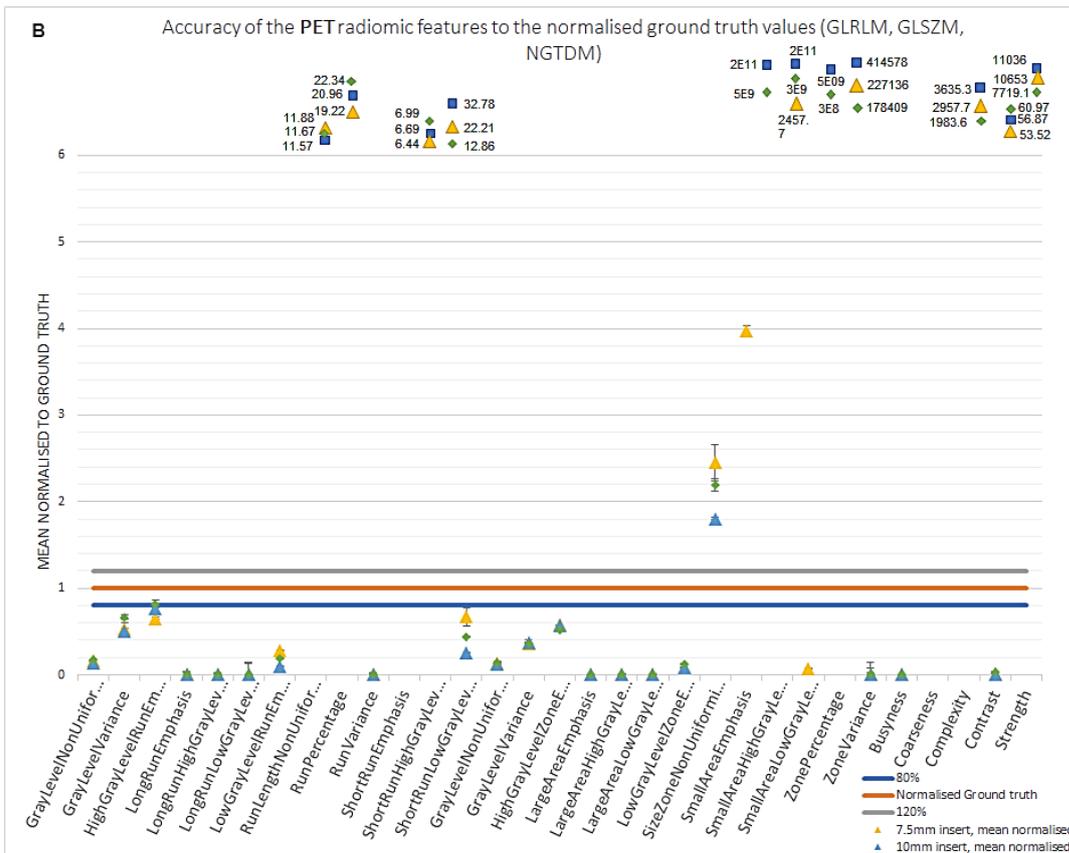
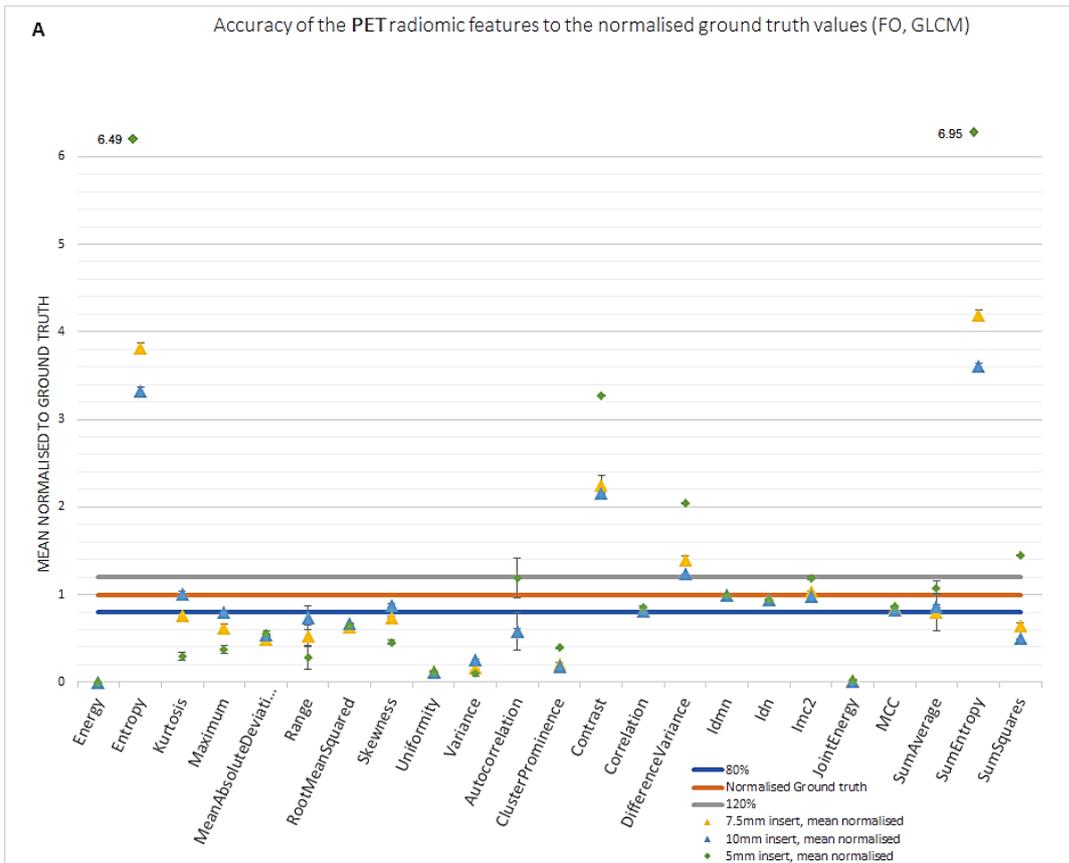


Figure 11: Results of the accuracy of the extracted PET imaging features. The mean values are normalised to the ground truth values. The ground truth is shown as an orange line (=1) with a standard deviation range of 20% (grey and orange lines). Calculated values too large to fit the graphs, are shown above the graph and their values are noted. Figure A shows the result of the first order and GLCM features, and figure B the result of the GLRLM, GLSZM and NGTDM features.

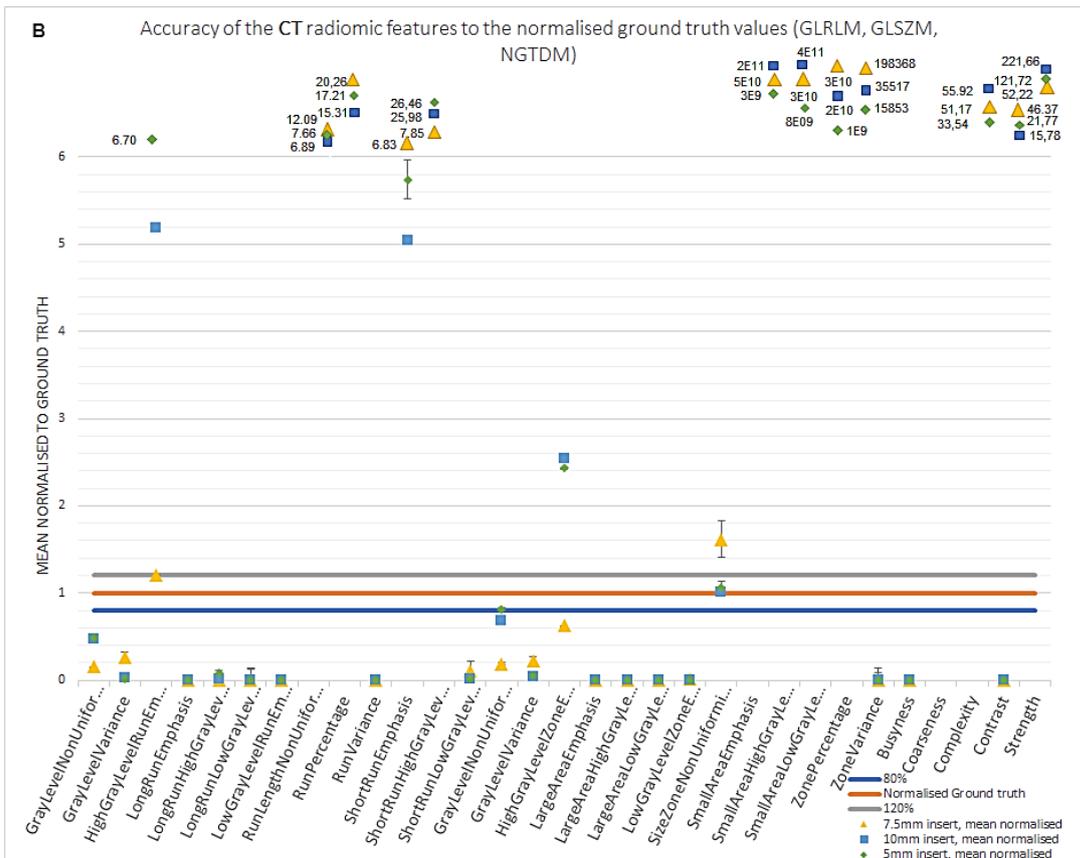
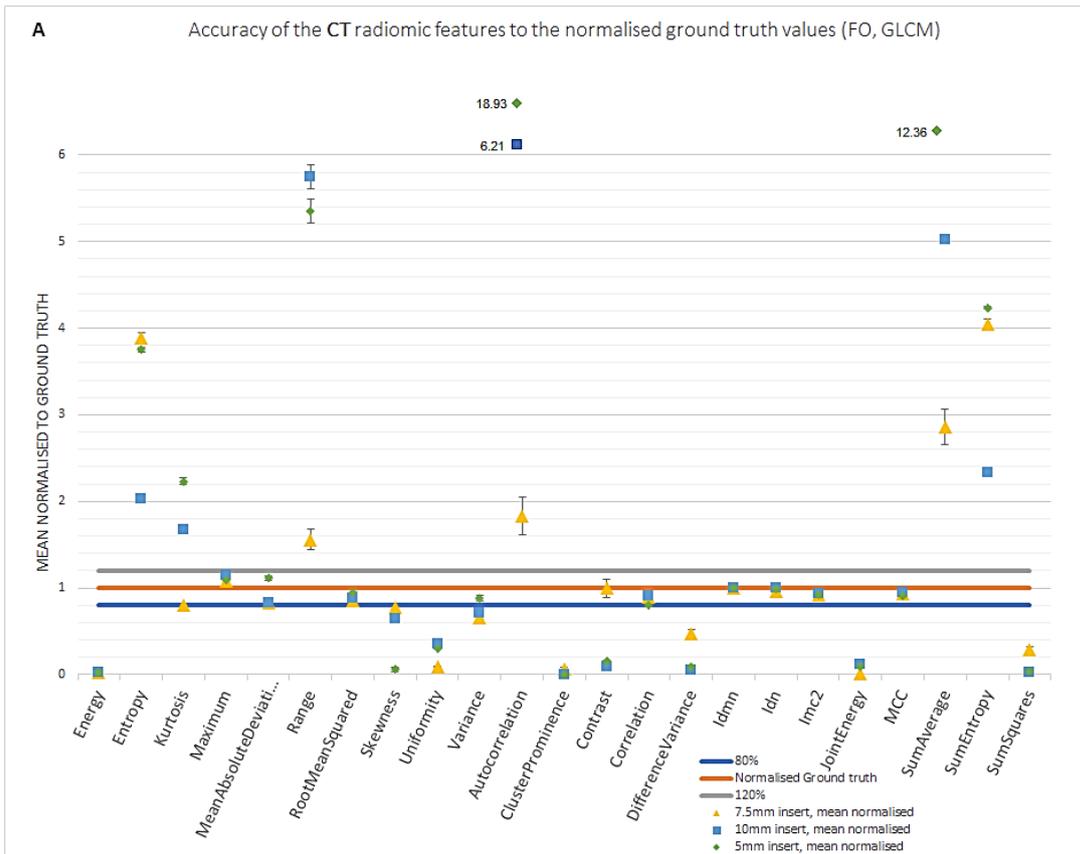


Figure 12: Results of the accuracy of the extracted CT imaging features. The mean values are normalised to the ground truth values. The ground truth is shown as an orange line (=1) with a standard deviation range of 20% (grey and orange lines). Calculated values too large to fit the graphs, are shown above the graph and their values are noted. Figure A shows the result of the first order and GLCM features, and figure B the result of the GLRLM, GLSZM and NGTDM features.

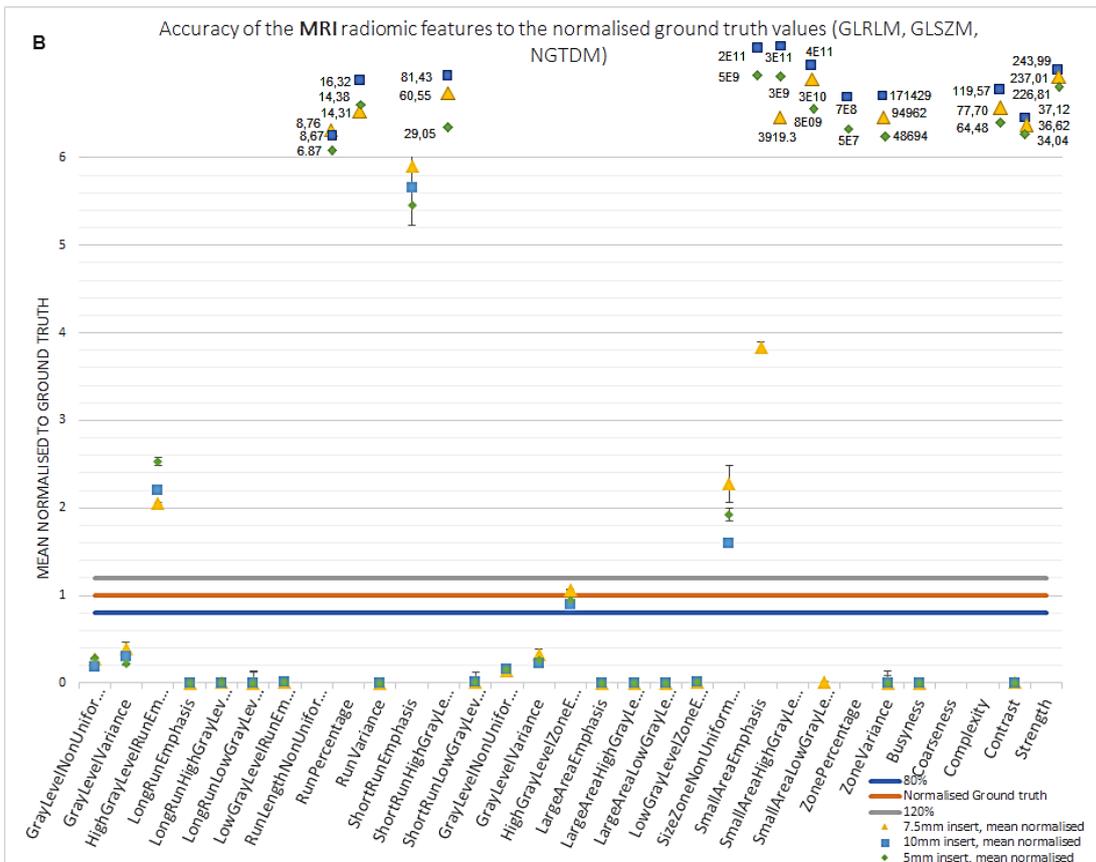
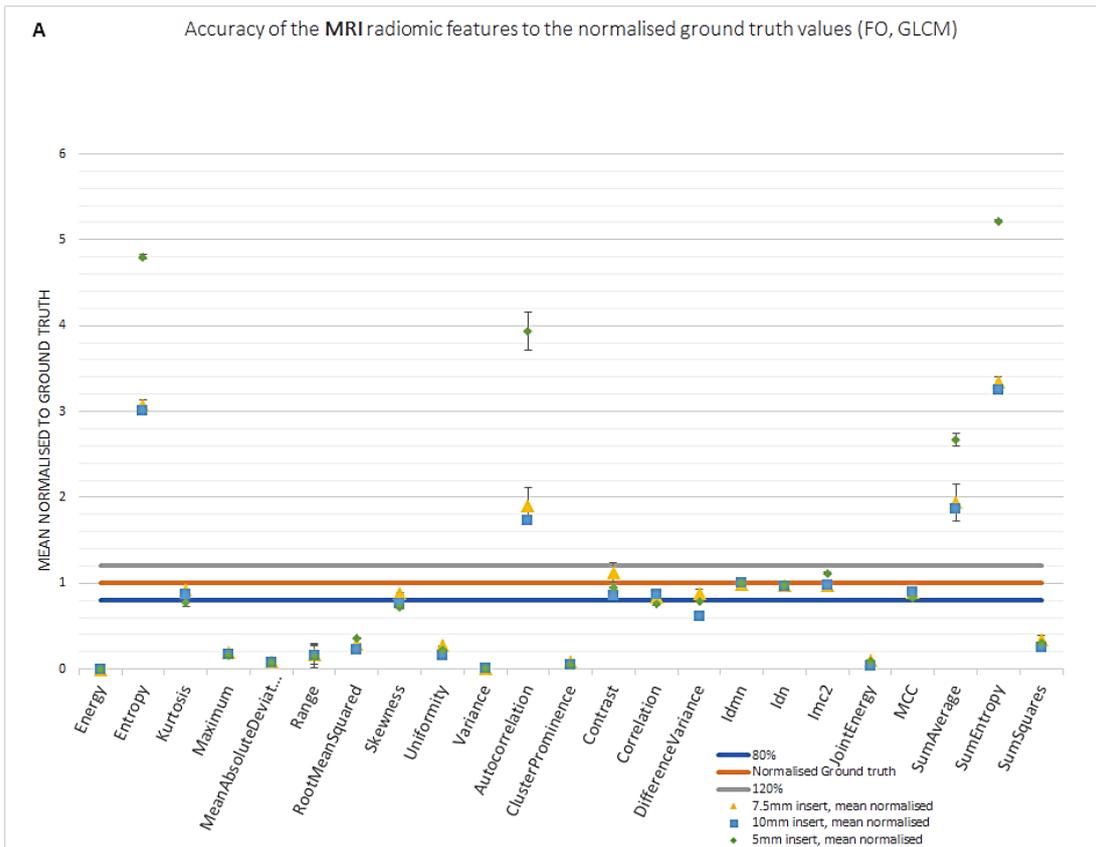


Figure 13: Results of the accuracy of the extracted MR imaging features. The mean values are normalised to the ground truth values. The ground truth is shown as an orange line (=1) with a standard deviation range of 20% (grey and orange lines). Calculated values too large to fit the graphs, are shown above the graph and their values are noted. Figure A shows the result of the first order and GLCM features, and figure B the result of the GLRLM, GLSZM and NGTDM features.

1.4 Discussion

In this study the performance of different imaging modalities for depicting heterogeneous uptake and enhancement was investigated using the extraction of quantitative radiomic features. Accuracy and variability was assessed using a 3D printed multi-compartment heterogeneity phantom.

1.4.1 Phantom design

The design was a result of multiple requirements the phantom needed to fulfill in order to be imaged using PET, CT and MR. One major request was that compartments had very thin walls to avoid a cold wall effect on the images. However, with the 3D printing techniques pushed to the limit, a minimum wall thickness of 0.4 mm was reached before the material would show cracks in the walls causing leakage. With the minimum wall thickness, the wall remains as a substantial structure on the CT and MR images (especially on the MR images) (Figure 9).

With the small and fragile design of the phantom inserts, the design is sensitive for errors and deformations. To this end, every printed element will have its unique shape whereby consistency might be difficult to achieve. Furthermore, with occurrence of these printing faults the stereolithography (STL) files with the original design were not appropriate to be used to create the ground truth as it deviates significantly from the printed design (due to gravity and effects of the techniques used during manufacturing). However, 3D printing errors, technical errors and manufacturing errors during the development of the phantom are inevitable. Multiple studies have used phantoms for the assessment of radiomic features. Most phantom studies are performed for FDG-PET imaging. Shiri et al., Nyflot et al. and Pfaehler et al. used the NEMA body phantom with spheres with varying inner diameters and radioactivity concentrations creating different sphere-to-background ratios. [48–50] However, with the spheres only connected to a homogeneous background and a wall thickness creating undesired cold wall effects, the standard NEMA body phantom seems not suitable for heterogeneity imaging.[51] Baessler et al. created a phantom using different fruit and vegetables for different inner structures. With this phantom they investigated the radiomic features using different MR imaging sequences.[40] With the unknown ground truth of the structures, the inability to administer contrast agents and limited geometry, this phantom is not suitable for standardised imaging protocols. Furthermore, most phantoms are used in a single modality imaging study, and are not suitable for multimodality purposes. Therefore, our study shows a unique study design using the same phantom in different modalities for imaging of heterogeneous structures.

Improvements were made on preliminary designs of the heterogeneity phantom. In the top and bottom plates of the cylindrical compartments, adjustments were made for optimal usage. The adjustments of caps instead of screws and the system to attach the three cylindrical compartments improved the manual handling and the procedure of filling the phantom. However, the single and small insertion holes for each separate compartment came with difficulties. With a single entrance of less than a one centimetre, the work space was limited and filling was only possible with a syringe. To this end, the manual filling was a precise procedure and there was a considerable change air bubbles remained in the inserts. Furthermore, with the single entrance, reusability of the phantom was less convenient as there was not enough space for air to enter the compartment during emptying. Preparation for reusability was therefore a time consuming process.

The advantage of 3D printing techniques is its flexibility for the development of varied structures suitable for heterogeneity imaging. In our study a multi-compartment phantom comprised of cubical inserts was developed and used. This is a selected design suitable for the purpose of this study, but more designs are possible and can be useful for the imaging of heterogeneous structures.

1.4.2 Variability

The assessed radiomic features showed several orders of magnitude of variability in the coefficient of variation. For several radiomic features the magnitude of the CV rises up to more than 50%. These trends could be explained as a function of the class, e.g. the GLSZM class has higher variability than the other classes. However, with the different compositions of the matrices and mathematical methods for feature calculation this direct comparison between two different classes is not feasible. When looking at every imaging modality and phantom size, table 2 shows a trend of low variability in

the first order class and GLCM class, with few features with increased CV values, e.g. the first order skewness feature for the 5 mm insert size. The high value is explained by one scan with a relatively high value for this feature. With a small number of repeated scans (N=5) this high value causes the mean and standard deviation to increase considerably. With a small number of repeated scans a single outlier can have a large impact on the variability. Test-retest analysis could be improved using a larger number of scans.

Furthermore, the GLCM features show most features with a CV less than 1%. Taken into account all features from all three imaging modalities, 56 features had a CV less than 1% and 38 of these features belonged to the GLCM class. Therefore, our results show that the GLCM feature class contains the most robust features compared to the other classes (n=38/90, 42%), followed by the first order features (n=10/90, 11%). The least amount of features with a CV less than 1% were shown in the GLRLM- (n=6/117, 5%), GLSZM- (n=2/117, 1.7%) and NGTDM (0/45, 0%) feature class.

Variability in MR imaging has one single feature with a percentage above the significant value of 20%. Furthermore, with a total of 84% of the features with a CV lower than 10%, the MR features can be described as the modality with the most stable feature in this study. The use of a fixed bin count in the parameter file for feature extraction could be of influence to these results. Multiple studies described that a fixed bin width is more appropriate in the extraction of radiomic features from modalities with absolute values (PET- and CT imaging).[45] However, in MR imaging a fixed bin count was more appropriate as the images contain relative signal intensity values. Furthermore, each imaging modality has its own range of gray values which makes initial comparison between PET, CT and MR imaging not feasible. Therefore, a fixed bin count of 64 was chosen in this study.

Baessler et al. established a good robustness of radiomic features in MR imaging in test-retest study with a fruit-phantom for MR imaging. Multiple shape, GLRLM and GLSZM features showed a concordance correlation coefficient larger than 95%. They used multiple sequences, including a T1 sequence, which show reliable intra- and inter-observer results.[40] Furthermore, Zhang et al. described a promising result for the prediction of tumour grading by texture analysis. Their study used a small study population and only five features were selected, but their result show the value of clinical studies in MR imaging feature extraction.[15] Our results suggest that MR imaging can provide stable feature extraction showing low variability in test-retest analysis.

The size zone features (GLSZM class) show the highest variability in PET- and CT imaging. The composition of the matrices of the GLSZM matrices have a correlation with the GLRLM matrices. However, contrary to the GLCM and GLRLM features, GLSZM matrices are rotation independent and are calculated in all directions with a single matrix. The number of rows of the GLSZM matrix is dependent on the number of different gray levels in the texture, and the (dynamic) amount of columns is dependent on the size of the largest zone. The more homogeneous the structure, the lower the number of rows and the higher the number of columns, resulting in a flatter and wider matrix. As this matrix is calculated in any direction, the matrix is sensitive to slight variations in gray levels in the texture, which can result in varied results in GLSZM class.

In previous studies a low variability between different reconstruction protocols was established in different types of features. Galavis et al. investigated twenty solid tumours for tumour characterisation and found small variation in the first order entropy, energy and maximal correlation coefficient, similar to our results, but also in gray level run emphasis features.[34] Our results show no consistency in variability for the gray level run emphasis features, especially for PET imaging. Cortes et al. found multiple stable run- and size emphasis features in the GLRLM and GLSZM feature classes using a phantom comprised of spheres, including short run emphasis and short zone emphasis.[52] However, our results show the highest variability in these two feature classes among all modalities. Since most studies are performed using different clinical data sets, different patient populations or phantom types, and different reconstruction protocols, it is difficult to compare results. This emphasises the need for standardisation of radiomic feature extraction.

1.4.3 Scalability

The scalability was investigated using a phantom with different insert sizes filled with the same compositions of fluid concentrations. Expected was that the variability decreases when the phantom sizes increase as the impact of noise on the image values is higher on images with a small number of voxels. Table 3 shows that there is a significant difference between the features of the different phantom sizes. Almost all features in CT and MR imaging show a significant differences between the different insert sizes, suggesting that size has an impact on the quantification of radiomic features. However, these significant values only show that there is a difference, but not whether the values increase or decrease as a function of the phantom insert size. Nyflot et al. studied test-retest variability for a phantom with different sized spheres and described differences between the sizes. Corresponding to our results, their results showed no consistent trend in CV values towards the phantom sizes.[49]

The developed phantom contained three compartments, each with a maximum height of 55 mm. Therefore, it is a possibility that the phantom comprised of 5 mm, 7.5 mm and 10 mm cubes, is too small for a proper comparison of scalability. The results show a difference between the phantom sizes, but the data shows no direct relation with phantom scale. Furthermore, compared to the other imaging modalities PET contains the least features with a significant difference between the insert sizes, suggesting that PET is least sensitive for scaling. However, the pixel spacing of PET (4x4 mm) is larger than the difference between the 5 mm and 7.5 mm insert sizes and the 7.5 mm and 10 mm sizes. Therefore, the spatial resolution of PET imaging can influence the results for scalability.

1.4.4 Accuracy

In the graphs shown in Figures 11, 12, 13 a 20% range is shown as a standard for the precision of the system regarding the values of the ground truth. Only few features regarding PET, CT or MR imaging lie within this range, 22, 28 and 24 features of the total of 162 features (54 for each insert size) respectively. Remarkably there are four GLCM features (IDMN, IDN, IMC2, MCC) in every modality and every insert size which have values within the defined range of 20%. These results suggest that technical factors, such as noise, contrast and resolution, have tolerable impact on these specific features. Given the high stability under different imaging conditions, these features can be useful in radiomic feature quantification. Features insensitive for technical and biological factors can be useful for standardisation and harmonisation of heterogeneity studies since comparison between different methods can be performed.[52]

In this study the ground truth is generated using a detailed reconstruction of a CT scan with a slice thickness of 0.5 mm and a pixel spacing of 0.206x0.206 mm. The pixel spacing for the other image acquisitions for CT imaging was a factor 4 higher, for MR imaging a factor 6 higher and for PET imaging a factor 20 higher. With the smaller pixel spacing, the VOI contains a higher number of voxels. Multiple studies established that the number of pixels and pixel size have an impact on the calculation of radiomic features.[53] Cortes et al. found that the signal-to-noise ratio in FDG-PET imaging decreases with the pixel size.[52] In a study of Yan et al. different reconstruction protocols were applied before feature extraction and showed that the spatial resolution influences the derived metrics in FDG-PET imaging.[54] In order to compare the imaging modalities with the ground truth images it is recommended to convert the images to similar spatial resolution. With similar spatial resolution the amount of voxels in the ground truth is equal to the amount of voxels in the imaging modality, and can improve the quantitative analysis and compare between ground truth and imaging modalities.

The ground truth images are composed using the absolute ratios and concentration values in the different compartments of the phantom. The compartments in the ground truths are multiplied by these factors creating new image values to create simulated images. To this end, the ground truth represented an 'ideal noiseless' modality. It is assumed that default imaging technique settings show no effect on image acquisition. This is an assumption which can not hold true for all modalities. Furthermore, with the use of the calculated ratios it is assumed that contrast values are linearly visualised in all modalities. However, with MR imaging the values are scaled and relative values are visualised. In CT and PET images the contrast recovery is not linear and reconstruction protocols, such as point spread function modelling, influence the contrast recovery. Furthermore, the default settings of a scanner, causing signal-to-noise and other effects on the images, were not taken into consideration for the creation of the ground truth images. To make a more appropriate compare of the ground

truth with the imaging modality, these technical factors need to be considered. Furthermore, with a ground truth which holds the technical factors of a modality in addition to the 'ideal noiseless' ground truth, features can be analysed regarding the influence of technical factors. However, it needs to be considered that these technical factors are different for all modalities which makes multimodality comparison a difficult task.

1.5 Conclusion

The results of this study show the possibilities for a heterogeneity phantom available for multimodality imaging. The results of feature extraction showed irregularities in the variability of radiomic features calculated on PET-, CT- and MR imaging. Radiomic features were calculated with identical methods but the variability between modalities was diverse. In every modality the different insert sizes showed significant differences in feature outcome, suggesting that geometry has a considerable impact on feature calculation. Furthermore, the accuracy was investigated using an 'ideal noiseless' ground truth and showed that the majority of the calculated features differs significantly from the ground truth values. Nonetheless, four single features from the GLCM class show to differ not more than 20% from the ground truth values in every modality and every phantom insert size, suggesting that technical factors have a tolerable effect on these features. Quantification of the accuracy and variability of the radiomic features for the imaging modalities can be improved when default settings and noise levels are included in the ground truth images.

The results of this study urge for further analysis towards the quantification of radiomic feature extraction in a multimodality imaging platform. Quantification of radiomic features can lead to standardisation and harmonisation of radiomic features and facilitate the step towards clinical validation and application. The next step towards clinical implementation of radiomics are clinical trials to assess the added value of radiomics in patient care. As these challenges are addressed, heterogeneity analysis using radiomic feature have great potential to play a role in clinical decision making.

Chapter 2: Phantom design and development

The aim of this study was to create a reusable phantom which can simulate heterogeneous uptake and enhancement patterns and is suitable for imaging in multiple modalities, including PET, CT and MR imaging. Such an imaging phantom brings new possibilities for imaging of tumour heterogeneity and testing the performance of different imaging modalities. This chapter describes how the design of the phantom was composed and how the prototype was developed. In previous studies, performed at the Technical University (TU) Delft in cooperation with Leiden University Medical Center (LUMC), several phantoms were designed and developed for tumour heterogeneity in PET/CT imaging. Based on these studies further development of a heterogeneous multi-compartment phantom was performed.

2.1 NEMA body phantom

One of the most commonly used PET phantoms, and also available most institutes, is the NEMA NU-2 body PET phantom. The NEMA body phantom, shown in Figure 14, is an internationally used phantom for image quality assessment of FDG-PET imaging, for NEMA quality control and for the European Association of Nuclear Medicine Research Limited (EARL) accreditation.[9] This phantom consists of fillable spheres and a lung insert. The casing of the phantom is made of transparent polymethyl methacrylate (PMMA). The spheres and insert are placed in a body shaped fillable compartment, the background. The spheres can be filled with radioactivity and the solid ("cold") insert can be comprised of solid material (e.g. air, foam). The individual spheres can only be filled with a single concentration of radioactivity, which makes it difficult to simulate heterogeneous uptake patterns. Flexibility in shape and size is limited, given the fixed geometry of the spheres. Furthermore, the walls of the spheres and inserts have a thickness of 1 mm, where no radioactivity is present, causing a cold wall effect. Cold walls of small structures create artefacts on the imaging such as partial volume effect, which lower the contrast between areas of high activity by underestimating their concentration and spreading their location. Furthermore, the cold walls can also cause Gibbs artefacts during image reconstruction. This leads to inaccuracies on volume and activity estimation, especially for small structures.[51]

The NEMA body phantom case can only be filled with one concentration as background, resulting in a homogeneous compartment. The spheres are all filled with different concentrations, but the inserts are too simplistic to represent realistic heterogeneous (biological) structures. However, the cylindrical NEMA body casing is a suitable structure serving as a homogeneous background compartment, representing a section of a human torso. Therefore, a new design for a heterogeneous structure fitting the NEMA body phantom is proposed, making it possible for the phantom to be easily used at different institutions.

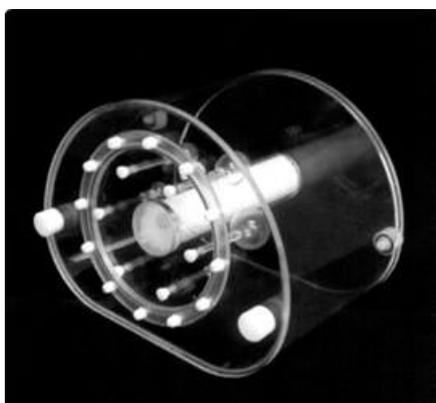


Figure 14: Image of the NU-2 NEMA body PET phantom. This phantom is internationally used for quality control of FDG-PET imaging.[9]

2.2 Phantom requirements

The desire to make a multimodality reusable heterogeneity phantom requires different compartments with complex geometries fillable with different concentrations of contrast agent and radioactivity. Furthermore, the phantom needs to have a reliable ground truth, needs to simulate heterogeneous imaging situations as well as to provide reproducible results. Taking into account the properties and limitations of the existing preliminary phantoms, and the desires for heterogeneity imaging, requirements were defined. Those requirements are described in table 4.

The phantom was designed with three fillable cylindrical segments that can be fitted in the NEMA body PET phantom. Within these segments a construction of three entangled fillable compartments was designed. As the first requirement of the phantom was to fit the NEMA body phantom casing, specific measures and shapes were required. Furthermore, space within the NEMA body phantom is limited, which emphasises the need for small structures. Therefore, three-dimensional (3D) printing techniques were used for development of the phantom as they can provide flexible shapes and sizes. The cylindrical segments were comprised of polymethyl methacrylate (PMMA), and the compartments within these segments were comprised of polylactic acid (PLA) and photopolymer material, suitable to be used in different imaging modalities.

Table 4: Overview of all requirements for the design and development of the multimodality heterogeneity phantom.

Category	Requirement	Approach requirement
Phantom case design	Phantom fits the NEMA body phantom	Maximum height of phantom case is 202 mm.
	Three separate compartments	Three separate and exchangeable compartments which can be separated from and attached to each other.
	Minimal wall thickness	Compartment walls manufactured thin as possible.
	Small insertion holes	Enough space to fit a needle but air needs to escape.
	Minimal set-up time Easily close insertion holes Repeatability	Easy to use. Fast and easy filling and emptying. Screws with fast handling to close the holes. Re-usable and not deformable.
Insert design	Three scalable heterogeneity structures	Three similar shaped and different sized shapes.
	Flexible size and shape	Inserts design suitable for 3D printing.
	Flexible quantitative design	Three cubical hollow shaped inserts based on pixel shape.
	Realistic tumour design	Different hollow shape similar to biological shapes.
	Wall-less inserts No air in inserts	3D printing with thinnest wall possible. Insertion holes at the edge of the design.
Heterogeneity	Heterogeneous inserts	Three different fillable inserts suitable to contain different activity or contrast agent concentrations
	Heterogeneous background	Fillable compartments and background for different activity or contrast agent concentrations
Global design	Transparent	Transparent material for phantom case and 3D printed inserts.
	Watertight	No leaks.
	Waterproof manufactured	Waterproof glue.
	Stable structure	Robust in handling for multiple usage.

2.3 Phantom case design

With the first requirement to fit the NEMA body phantom, the cylindrical phantom was developed with a maximum height of 202 mm. To fit the requirement of three separate heterogeneity structures, the phantom is divided in three separate cylindrical segments. These three segments all have the same measures and can be attached to each other to form a complete phantom. In Figure 15 a sketch of the phantom is shown with the exact measures of the phantom. Each cylindrical segment is made of a hollow PMMA tube with a height of 56.6 mm, an inner diameter of 76 mm and an outer diameter of 80 mm. With the requirement to reduce the amount of material between segments, the wall thickness of the PMMA tube was reduced to the minimum thickness possible before the material would break. With a mould the wall thickness was grinded down to 1 mm. To maintain stability of the structure, the 2 mm wall thickness was preserved at both ends of the cylinder 2 mm for a length of 5 mm from the edges. With the adjustment of the wall thickness, the inner diameter becomes 78 mm. A sketch with the measures for the tube of the cylindrical segment is shown in Figure 16.

Each cylindrical segment was closed off by a solid top and bottom plate. These plates were designed using the 3D computer-aided design (CAD) software Inventor 2018 (Autodesk Inc., San Rafael, California, United States), prepared for the 3D printer using Ultimaker Cura 2 (Ultimaker B.V., Geldermalsen, The Netherlands) and printed of polylactic acid (PLA) material using the Ultimaker 3 (Ultimaker B.V., Geldermalsen, The Netherlands). Both bottom and top plate have a diameter of 92 mm. The capping plates had a specially designed interlocking system to connect the three segments together. The bottom plates have a thickness of 2 mm with the main function of closing the cylindrical segment. The bottom plate of the lowest segment contains an extrusion of three millimetres to secure the phantom in the NEMA body phantom. For the top plate of the upper segment this extrusion was not possible as the insertion holes needed to stay accessible. Therefore, a separate plate was created which could be attached to the top plate of the upper segment. This plate was the outer plate of the phantom and together with the extrusion of the outer bottom plate fixates the phantom in the NEMA body phantom (Figure 17B). Each bottom plate had a circular intrusion in the plate with an outer diameter of 81.5 mm, a width of 3 mm and a depth of 0.8 mm. The design of the bottomplate is

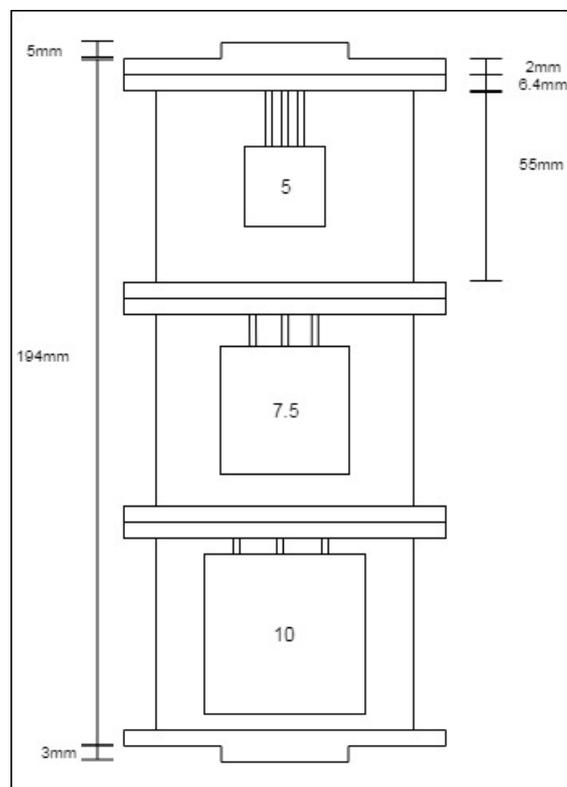


Figure 15: Sketch of the front view of the phantom casing with a simplistic display of the cubical inserts. The measures of the different elements are defined in millimetres (mm).

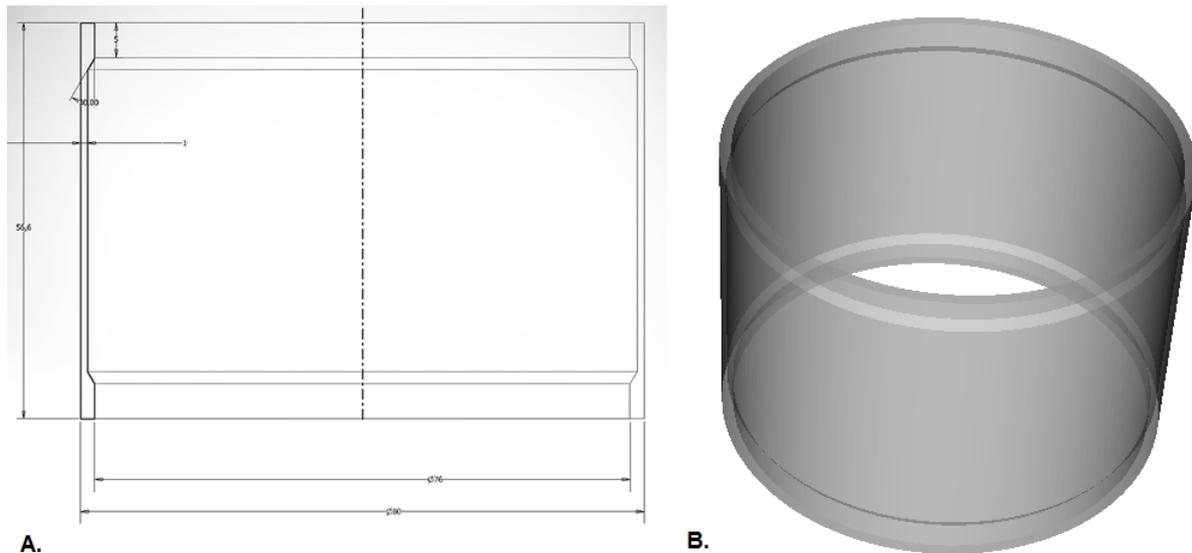


Figure 16: A: Measures of the cylindrical segment, showing the groove in the wall for minimal wall thickness. B: A 3D view of the cylindrical tube made from a PMMA material.

shown in Figure 17A. The intrusion was used to fixate the cylindrical PMMA tube to the plates. With a 0.8 mm intrusion to both ends of the cylindrical tube, the height of the hollow segments will remain 55 mm (Figure 15).

The top plates of each cylindrical segment have a thickness of 6.4 mm. These plates are considerably thicker because they have an extra function as they consist of the only entrance to the hollow inserts within the cylindrical segments. Each top plate has four insertion holes, one for the cylindrical segment and three for the three scalable shapes (Figure 18A and 18B). These insertion holes have a diameter of 2.4 mm, enough space for a syringe used to fill the segment. At the bottom of the plate, each insertion hole has a protrusion for attachment of the insert shapes (Figure 18A).

In previous studies screws were used to close off the insertion holes. However, experience showed that it takes a relative long time to close all insertion holes, which is not user-friendly. Furthermore, the user is exposed to radiation when filling the phantom for PET imaging. Therefore, it was desired to design a new closing system for these insertion holes. As it was preferred to close off the insertion holes with one single action, a different type of screw was designed. A cap which can be pushed into the insertion holes was designed and developed from solid PMMA material (Figure 18C). In order to fixate the caps in the insertion holes, rubber o-rings were used. Each cap consists of an intrusion to secure the o-ring. The size of the o-ring needed to be large enough to make the segment water tight. As the cap is pushed in the insertion hole, the o-ring is compressed and secures a watertight closure of the segments. The phantom had a limited maximum height and it was preferred to make

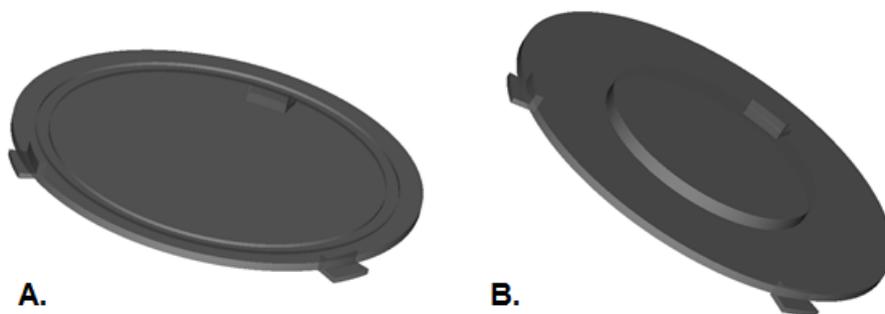


Figure 17: The bottom shapes. A: The bottom plate with the groove for the cylindrical tube to fit. Extrusion at the bottom of this plate are used for the attachment of the separate elements. B: The outer plate of the phantom with its intrusion to fit the NEMA body phantom.

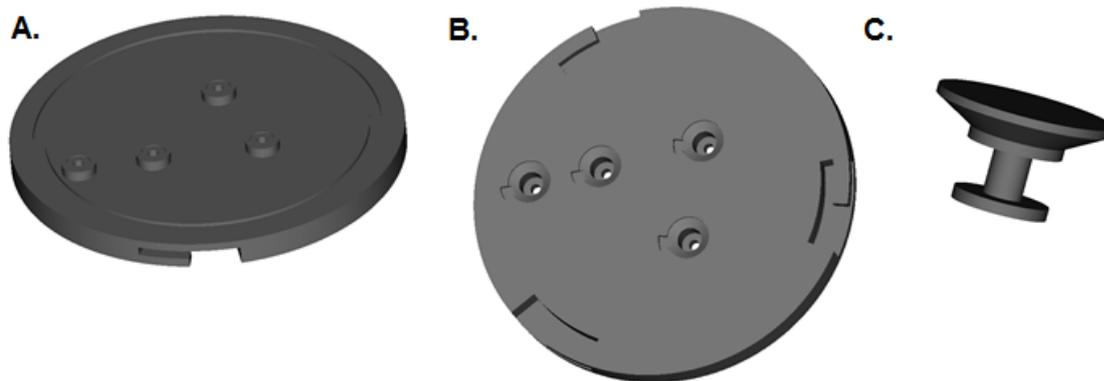


Figure 18: The shapes of the top plates and the cap. A: The bottom of the top plate facing the inner side of the cylindrical segment. Extrusion around the insertion holes serve as surface for the insert shapes to attach to. B: The top plate with the insertion holes with the intrusions for the cap to fit. C: The design of the cap with an intrusion for the o-ring to fit.

the phantom as compact as possible. Therefore, a countersink was created in the capping top plate for the caps to be inserted and equalised with the surface of the top plate. The countersink for the insertion holes in the top plate were matched to the shape of the cap.

Another critical requirement was to design a phantom which could easily be managed, (dis)connected to each other and reused. Most importantly, with the purpose of creating a phantom suitable for imaging with different contrast agents and radiopharmaceuticals, the design needed to be easily refillable for usage in a different modality. Therefore, the separate cylindrical segments need to be easily attached to each other to complete and fixate the phantom structures. In previous studies this was performed using screws in the ridges of the top and bottom plates. However, the system with screws extended assembly time and was not user-friendly, which made it no convenient system. Especially in cases of insertion of radiopharmaceuticals the system was not convenient as radiation exposure was increased. Therefore, a new system for the attachment of the cylindrical segments was designed. This system is based on the click-and-close system used to attach an objective to a camera case. Each bottom plate was designed with a protrusion (Figure 17) shaped like a hook, and each top plate was designed with an intrusion (18A, 18B) for the hook to fit. This way the cylindrical segments can easily be attached to each other, which reduces exposure time and makes the phantom more user-friendly. In Figure 19 a 3D view of the design of the phantom case is shown. All element for the outer case are visualised, including the in- and extrusions of the plates for fixation of the cylindrical segments.

2.4 Phantom insert design

The phantom was designed with three separate cylindrical segments, enabling to simulate three different imaging conditions. With usage of the flexibility of 3D printing techniques, different prototypes were produced before the final shapes were defined and used in multimodality imaging experiments. The phantom inserts were designed using the CAD-software Inventor 2018 (Autodesk Inc., San Rafael, California, United States) and prepared for the 3D printer using the print preparation software PreForm 2.18 (Formlabs, Sommerville, Massachusetts, United States). With the specific and detailed designs of the inserts, the Formlabs Form 2 3D printer was used (Formlabs, Sommerville, Massachusetts, United States) to produce the phantom inserts. With stereolithography (SLA) the printer uses a laser to cause chemical monomers to link together to form polymers. Those polymers are comprised to form the body of a 3D solid structure. With this technique high resolution designs can be manufactured.

One requirement for the insert design was to reduce the wall thickness as much as possible for optimal heterogeneity simulation. However, boundaries are needed to separate different concentrations which makes complete wall-less inserts, whilst adhering to the other design requirements, not achievable. To this end, impact of the walls to imaging and reconstruction (e.g. cold wall effects) are



Figure 19: A 3D view of all elements together forming the phantom case design with the three segments.

inevitable. In order to minimise the effect, the smallest wall thickness is preferred. With the heterogeneous shapes introduced in this study, the minimal wall thickness for the inserts was 0.4 mm to maintain proper 3D printing results.

2.4.1 Primary concept: tumour design

With the main purpose to create a phantom to simulate heterogeneity, the first concept was to create phantom inserts with a shape representing anatomical shapes. Tumours are irregularly shaped and have a varied composition. Fully realistic insertions with this magnitude of complexity are difficult to create at those scales and do not test radiomic features to the fullest extent. Therefore, irregular, spherical and curved shapes are combined to create a simplified tumour design.

The first tumour design was a hollow cylindrical insert shaped like an arch with a radius 5 mm. This shape contained a sphere in the middle of the inflection point of the arch with a radius of 12 mm. The hollow sphere was merged with the cylindrical legs of the arch to form a single compartment. Two spheres were placed at each end of the two legs of the arch, one with a radius of 8 mm and one with a radius of 6 mm. The design of this shape is shown in Figure 20. With the arch shape, including the spheres, there was only a single compartment besides the background and the cylindrical compartment. To this end, another structure was added to the initial design. Within the hollow design of the arch two round curves are placed with a radius of 2.5 mm. These round curves are connected to each other within the centre of the sphere and form one single compartment. The walls of both compartments, the arch and the inner curve, were designed with a wall thickness of 0.4 mm.

As shown in Figure 20, the arch contains three tubes at the upper side of the shape. These three tubes function as insertion tubes and are connected to the insertion holes of the cylindrical segment. The tubes have a diameter of 2.4 mm and with an extra surface at the end of the tube for proper

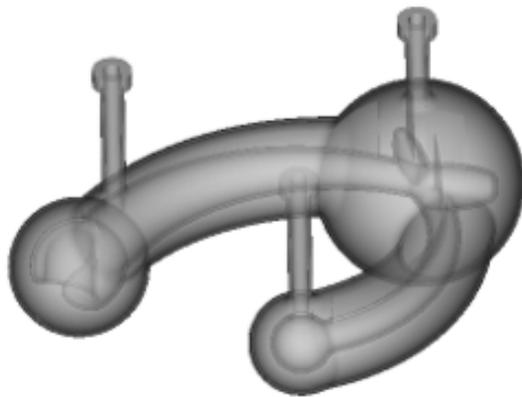


Figure 20: Visualisation of the design of the arch shaped compartments representing irregular tumour shape.

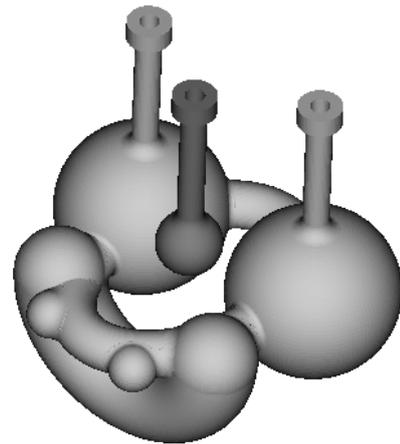


Figure 21: Visualisation of the spherical design representing a lobular tumour structure.

fixation. The tube connected to the large sphere in the middle of the arch, is the single connection to the hollow arch shape. At the end of the legs of the arch there are two insertion tubes located. These two tubes are connected to the inner curve. Furthermore, these two insertion tubes function as a fixation of the inner curve as this shape floats within the hollow arch. With the two separate hollow elements, a two-compartment heterogeneous structure was created.

The concept for a second insert shape design was derived from clinical cases presenting with lobular tumours. The presence of multiple lobules in a single tumour is a characteristic presentation. To simulate this lobular shape, a simplification of this specific structure was designed. The design is comprised of an oval belt connected to multiple spheres (different sizes and distances from each other). The spheres and oval belt are all hollow shapes and interconnect with each other to form a single compartment with a wall thickness of 0.4 mm. A 3D view of this design is shown in Figure 21. There are six spheres with three different sizes. The two large spheres have a radius of 10 mm, the centre spheres have a radius of 5 mm, and the small spheres at the end of the oval belt have a radius of 2.5 mm. The centre spheres are extended and connected to each other by a circular belt perpendicular to the base oval belt. This cylindrical belt has a radius of 5 mm, similar to both centre spheres. All elements of the lobular insert design are connected to each other to form a single compartment. In order to create more heterogeneity within a single compartment, a separate sphere (radius = 4 mm) was added in the centre of the oval belt. With the additional sphere, two different compartments within a single cylindrical segment are created to simulate heterogeneity.

The lobular structure was connected to the plates of the cylindrical compartment by two tubes, located on the top of the two largest spheres. The independent sphere had a single tube connected to the insertion holes of the plate. The tubes function as a fixation point and as an entrance point.

The designs of the simplified tumour shapes were developed and used as inserts for imaging in the phantom. However, with the goal to quantify radiomic features, a voxel-related phantom model was desired. As radiomic feature extraction is based on a voxel-to-voxel analysis, the concept of multiple voxel-shaped compartments was introduced. The next chapter will explain the concept of the cubical shaped compartments.

2.4.2 Final concept: cubical design

For quantification of voxel-to-voxel analysis the concept was brought to an insert design related to the voxel shape. The design of the cubical inserts was derived from the concept of three different fillable compartments connected to each other and filled with three different (radioactive or contrast agent) ratios, as defined in Figure 22A. This concept is introduced by the image biomarker standardisation initiative (IBSI).[41] The cubical composition used in this publication was set as an example for the composition of the heterogeneity phantom inserts. The compositions are comprised of connected cubes with similar measures. The designs of these cubical compositions are all different and

are entangled with each other. The walls of the three different compositions are connected to each other, creating the best possible interconnection for heterogeneity simulation. The compositions of the designs can be visualised using a matrix. The publication of the IBSI uses a 4x5 matrix as a digital phantom by assigning specific values to the cubes (Figure 22A). Each colour represents a different compartment in the digital phantom. In our study a cubical phantom was designed using a 4x4 matrix, shown in Figure 22B. Each colour represents a different compartment and the red coloured cubes are part of one of these three cubical inserts. The three separate, but entangled, shapes can be filled with different ratios of concentrations for the simulation of heterogeneity. The white cubes remain part of the cylindrical segment.

With the concept shown in Figure 22B, three differently shaped cubical compartments were designed, referred to as the L-shape, stair-shape and U-shape (Figure 23A). As no pixel size dimensions were defined in the publication of the IBSI, the cube sizes were defined by own concept with different sizes for investigation of scalability. The choice was to put a different cube size in each cylindrical compartment. Considering the limited amount of space available in each cylindrical compartment and the relative low resolution of the PET scanner (4x4x4 mm), cube sizes of 5x5x5 mm, 7.5x7.5x7.5 mm and 10x10x10 mm were defined (Figure 23B). With these measures the inserts should be visible on PET imaging alone, but are also small enough to fit the phantom compartments. The three similar shaped, but different sized, cubical compartments together form a scalable heterogeneity phantom.

The different cubical compartments were entangled with each other to form a larger cubical structure, shown in Figure 24. With the entangled inserts, the smallest possible wall thickness was preferred to reduce cold wall effects and have a the best connection between heterogeneous structures. Therefore, a wall thickness of 0.4 mm was defined. As the cubical inserts were printed as one structure, the walls of connected planes were reduced to 0.2 mm, and together they form a wall thickness of 0.4 mm feasible to print. With a wall thickness of 0.4 mm, the enlargement of the cubes to measures of 10 mm cubes caused the structure to break. The distances between two borders became too long for the wall and cracks occurred in the wall. Multiple sizes were tested and the minimal wall thickness of the 10x10x10 mm cubical insert design needed to be increased to 0.55 mm.

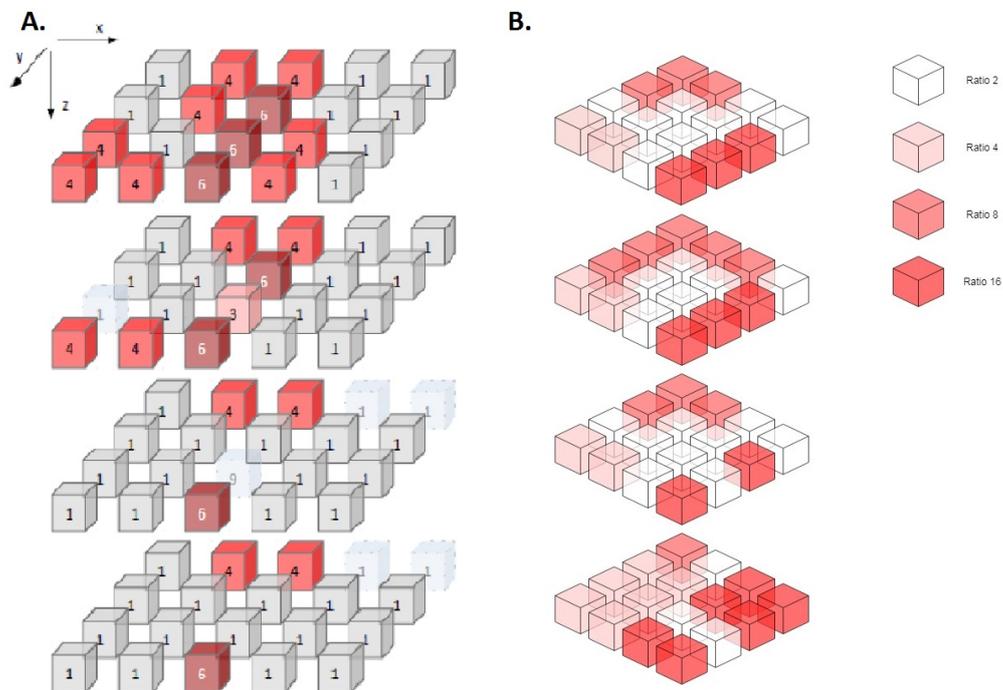


Figure 22: Representation of the structures of the cubical inserts. In both A. and B. the different colours or red represent the different inserts and the numbers give the different ratios in the inserts. The white cubes represent the background. A: A 4x5 matrix which shows the IBSI concept of a digital cubical phantom design. [41] B: This figure shows a 4x4 matrix which represents the design of the cubical compartments for the heterogeneity phantom in this study. The ratios of the compartments are shown right of the matrix.

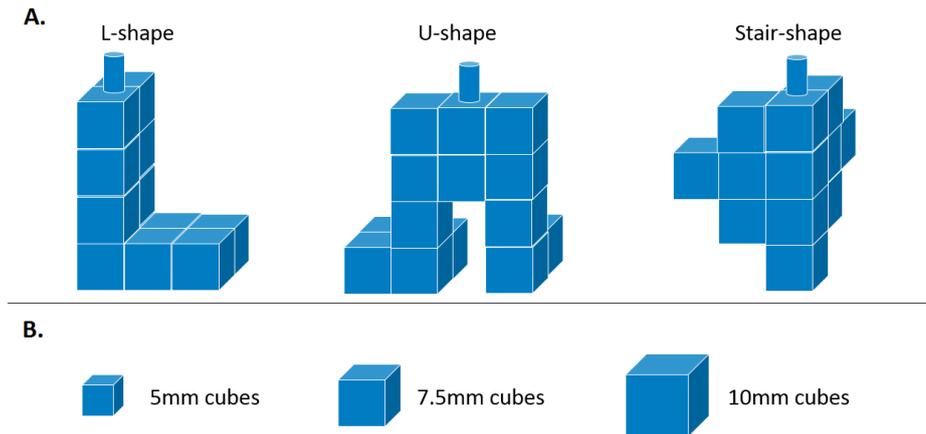


Figure 23: Design of the cubical compartments of each cylindrical segment of the phantom. A: The different shapes (L-shape, U-shape and stair-shape) of the compartments, all comprised of cubes with constant measures. B: The different sizes of the cubes used to form the three different cubical shapes.

In order to create an entrance for the compartments, each compartment had a tube connected to the top plate of the cylindrical segment. This tube had a similar wall thickness and an insertion hole with a diameter of 2.4 mm. The top of this tube had a protrusion which created a larger surface for connection to the top plate. An important requirement was to avoid the presence of air bubbles in the inserts (table 4). Therefore, all insertion tubes were positioned at the top of the phantom, as air travels to the surface of water. Furthermore, with the insertion tubes all at in the same direction, all air bubbles would move in the same direction which can be helpful during the process of filling the phantom. For the cylindrical segments it is expected to be difficult to remove all air bubbles. Therefore, the insertion holes for the cylindrical segments were placed closest to the edge to make it easier to release undesired air bubbles.

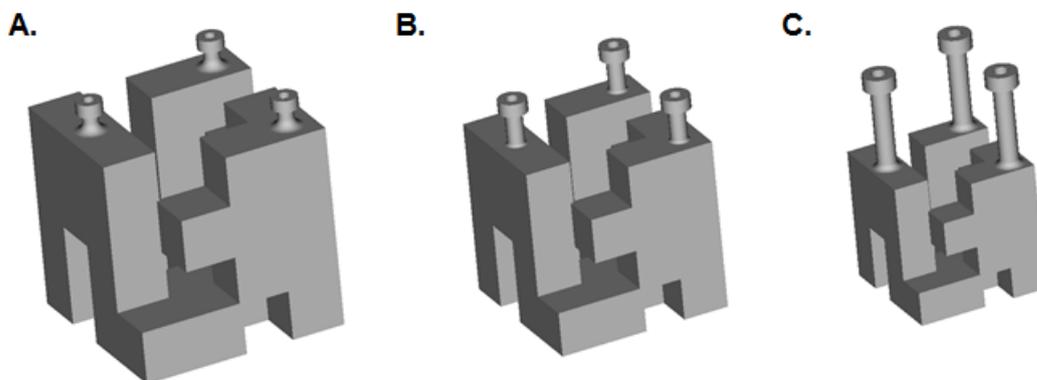


Figure 24: The design of the three cubical compartments entangled to each other. The cubical compartments comprised of 10x10x10 mm, 7.5x7.5x7.5 mm and 5x5x5 mm are shown in Figure A, B and C respectively.

2.5 Global phantom design

Besides the design of the phantom casing and its compartments, more aspects were required before the phantom was ready for usage. In this paragraph multiple general requirements as user-friendly, transparency, waterproof fixation and repeatability will be described, as shown in table 4.

An important requirement was to have transparent inserts and walls. This property makes it easier to work with the phantom. Furthermore, air bubbles can easily be detected and are required to be properly and carefully removed. However, to meet the requirement for minimal wall thickness of all compartments, the walls of the PMMA tubes of the cylindrical compartments were reduced from 2

mm to 1 mm. This adjustment to the material caused transparency to be reduced, but film of fluid over of the material caused transparency to be restored due to change in refraction. As the phantom will be filled with water, the transparency returns completely.

With the aim to fill the phantom inserts with water containing different ratios of activity or contrast agents, it is important to avoid leakages. This was considered during the process of 3D printing, but also when all different phantom elements need to be attached to each other. The cylindrical tubes, the top and bottom plates and the inserts were all developed with different techniques and were all separate elements. With small structures and attachment surfaces, limited options remained in order to make a solid and robust structure. In preliminary phantom studies performed at the institute, the phantom was assembled using different types of glue. However, primary limitation was the loosening of the glue due to humidity. Therefore, a new experiment was performed to identify a proper glue material.

Multiple glue types with different properties were available and could be used to assemble the phantom. For the purposes of the phantom, strength and water resistance are important properties of the glue. Two different and available epoxy glues were tested and compared; Alradite 2014-I and Ergo 5039 Universal. Alradite 2014-I is a grey-coloured two component adhesive epoxy paste used for different purposes in the industry. It is a room temperature curing, thixotropic paste of high strength with good environmental, temperature and chemical resistance. It's described that this material is highly resistant to water and a variety of chemicals. It is commonly used for bonding of metals and electronic components. Ergo 5039 Universal is a colourless super glue from an adhesive thixotropic gel. It has a proper strength for usage on metal and rubber material. With its fast curing in room temperature and temperature resistance range, the properties are suitable to assemble the phantom.

The strength and water resistance of these types of glue were tested. A sample of PLA material, used for the top and bottom plates, and a sample of photopolymer, used for the cubical shapes, were glued together by both Alradite-2012-I and the Ergo 5019 Universal glue. Both constructions cured for 24 hours to receive optimal strength, as described in their manual. After 24 hours the material was placed in water and monitored daily for three weeks. The glue was monitored for change in composition or loss of connection to the material. After three weeks both glue types showed no signs of loosening or strength loss. The PLA and photopolymer material stayed together for the weeks during monitoring. After three weeks the material was removed from the water and the strength was tested. The materials were separated from each other by manually pulling at both ends of the glued material. Both glue types broke after pulling them, but the materials with Ergo 5039 Universal were removed easier from each other than the materials with Alradite 2012-I. Therefore, for assembling the heterogeneity phantom, Alradite 2012-I was used.

The developed phantom elements consists of specific and detailed structures. For proper imaging the required wall thickness was reduced and minimal materials were used to form the different structures. With these properties of the phantom the best imaging circumstances were created, but simultaneously caused the phantom to be rather fragile. Therefore, the phantom needs to be handled with proper knowledge and care.

2.6 Final design

When all possible requirements were met, all the elements of the phantom were assembled and the phantom was inserted in the NEMA body phantom to be used for multimodality imaging. The result of the design and development of the heterogeneity phantom is shown in Figure 25.

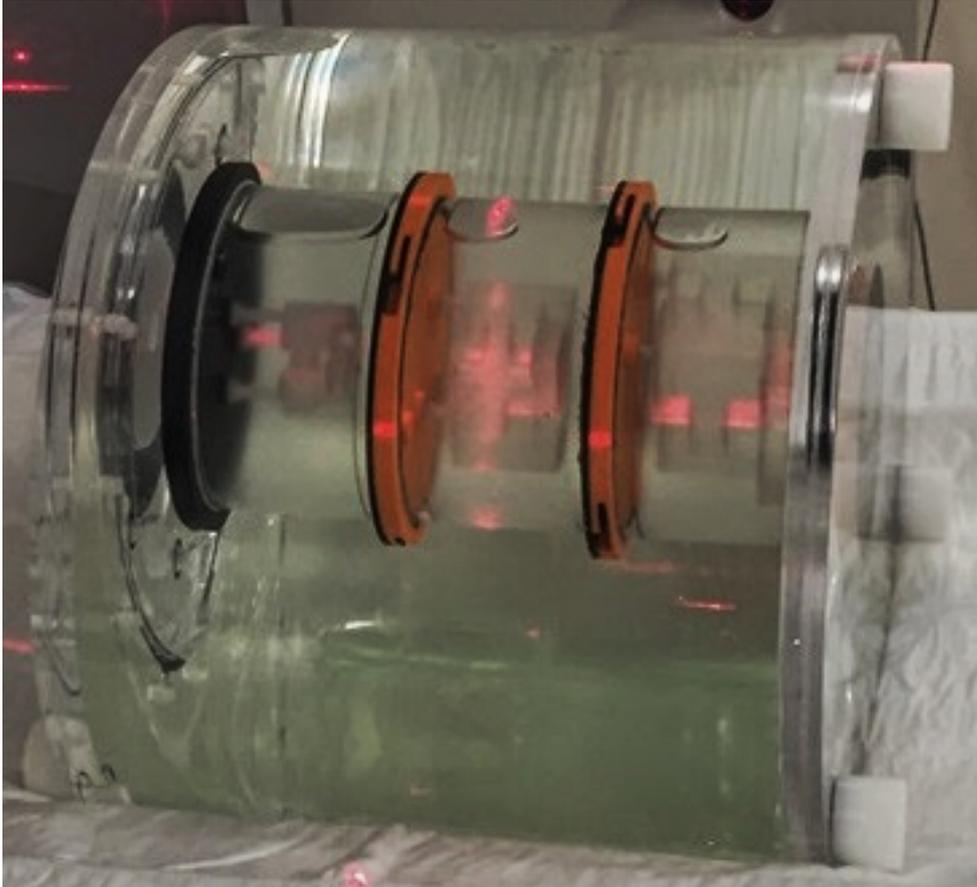


Figure 25: Visualisation of the designed multi-compartment heterogeneity phantom inside the NEMA body phantom. The phantom is filled with water and air bubbles are visible in the three cylindrical segments.

References

- [1] C. O'Neill, J. Moore, A. Kalof, and C. Marroquin, "Management of high-grade pleomorphic sarcoma with colon metastasis," *The Journal of Community and Supportive Oncology*, vol. 15, no. 4, pp. 221–223, 2017.
- [2] A. R. Padhani and K. A. Miles, "Multiparametric imaging of tumor response to therapy," *Radiology*, vol. 256, no. 2, pp. 348–364, 2010.
- [3] J. P. O'Connor, C. J. Rose, J. C. Waterton, R. A. Carano, G. J. Parker, and A. Jackson, "Imaging intratumor heterogeneity: role in therapy response, resistance, and clinical outcome," *Clinical Cancer Research*, vol. 21, no. 2, pp. 249–257, 2015.
- [4] F. Davnall, C. S. Yip, G. Ljungqvist, M. Selmi, F. Ng, B. Sanghera, B. Ganeshan, K. A. Miles, G. J. Cook, and V. Goh, "Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice?," *Insights into Imaging*, vol. 3, no. 6, pp. 573–589, 2012.
- [5] E. Sala, E. Mema, Y. Himoto, H. Veeraraghavan, J. Brenton, A. Snyder, B. Weigelt, and H. Vargas, "Unraveling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging," *Clinical Radiology*, vol. 72, no. 1, pp. 3–10, 2017.
- [6] M. Gerlinger and C. Swanton, "How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine," *British Journal of Cancer*, vol. 103, no. 8, pp. 1139–1143, 2010.
- [7] M. Gerlinger, A. J. Rowan, S. Horswell, J. Larkin, D. Endesfelder, E. Gronroos, P. Martinez, N. Matthews, A. Stewart, P. Tarpey, *et al.*, "Intratumor heterogeneity and branched evolution revealed by multiregion sequencing," *New England Journal of Medicine*, vol. 366, no. 10, pp. 883–892, 2012.
- [8] V. Kapoor, B. M. McCook, and F. S. Torok, "An introduction to PET-CT imaging," *Radiographics*, vol. 24, no. 2, pp. 523–543, 2004.
- [9] 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.
- [10] S. Chicklore, V. Goh, M. Siddique, A. Roy, P. K. Marsden, and G. J. Cook, "Quantifying tumour heterogeneity in 18 F-FDG PET/CT imaging by texture analysis," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 40, no. 1, pp. 133–140, 2013.
- [11] A. Klabatsa, S. Chicklore, S. F. Barrington, V. Goh, L. Lang-Lazdunski, and G. J. Cook, "The association of 18 F-FDG PET/CT parameters with survival in malignant pleural mesothelioma," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 41, no. 2, pp. 276–282, 2014.
- [12] G. N. Hounsfield, "Computed medical imaging. Nobel lecture, December 8, 1979.," *Journal of Computer Assisted Tomography*, vol. 4, no. 5, pp. 665–674, 1980.
- [13] C. M. West, R. A. Cooper, J. A. Loncaster, D. P. Wilks, and M. Bromley, "Tumor vascularity: a histological measure of angiogenesis and hypoxia," *Cancer Research*, vol. 61, no. 7, pp. 2907–2910, 2001.
- [14] E. M. Haacke, R. W. Brown, M. R. Thompson, R. Venkatesan, *et al.*, *Magnetic resonance imaging: physical principles and sequence design*, vol. 82. Wiley-Liss New York, 1999.
- [15] Y. Zhang, Y. Zhu, X. Shi, J. Tao, J. Cui, Y. Dai, M. Zheng, and S. Wang, "Soft Tissue Sarcomas: Preoperative Predictive Histopathological Grading Based on Radiomics of MRI," *Academic Radiology*, 2018.
- [16] 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*, vol. 102, no. 4, pp. 1083–1089, 2018.
- [17] H. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, vol. 5, p. 4006, 2014.
- [18] S. Reuzé, A. Schernberg, F. Orlhac, R. Sun, C. Chargari, L. Dercle, E. Deutsch, I. Buvat, and C. Robert, "Radiomics in nuclear medicine applied to radiation therapy: methods, pitfalls, and challenges," *International Journal of Radiation Oncology Biology Physics*, vol. 102, no. 4, pp. 1117–1142, 2018.
- [19] E. Limkin, R. Sun, L. Dercle, E. Zacharaki, C. Robert, S. Reuzé, A. Schernberg, N. Paragios, E. Deutsch, and C. Fertet, "Promises and challenges for the implementation of computational medical imaging (radiomics) in oncology," *Annals of Oncology*, vol. 28, no. 6, pp. 1191–1206, 2017.

- [20] M. Tuceryan and A. K. Jain, "Texture analysis," in *Handbook of Pattern Recognition and Computer Vision*, pp. 235–276, World Scientific, 1993.
- [21] M. M. Galloway, "Texture analysis using grey level run lengths," *NASA STI/Recon Technical Report N*, vol. 75, 1974.
- [22] R. M. Haralick, K. Shanmugam, *et al.*, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 6, pp. 610–621, 1973.
- [23] C. Sun and W. G. Wee, "Neighboring gray level dependence matrix for texture classification," *Computer Vision, Graphics, and Image Processing*, vol. 23, no. 3, pp. 341–352, 1983.
- [24] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: images are more than pictures, they are data," *Radiology*, vol. 278, no. 2, pp. 563–577, 2015.
- [25] S. S. Yip and H. J. Aerts, "Applications and limitations of radiomics," *Physics in Medicine & Biology*, vol. 61, no. 13, pp. 150–166, 2016.
- [26] T. Pyka, J. Gempt, D. Hiob, F. Ringel, J. Schlegel, S. Bette, H.-J. Wester, B. Meyer, and S. Förster, "Textural analysis of pre-therapeutic [18F]-FET-PET and its correlation with tumor grade and patient survival in high-grade gliomas," *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 43, no. 1, pp. 133–141, 2016.
- [27] S. Reuzé, F. Orhac, C. Chargari, C. Nioche, E. Limkin, F. Riet, A. Escande, C. Haie-Meder, L. Dercle, S. Gouy, *et al.*, "Prediction of cervical cancer recurrence using textural features extracted from 18F-FDG PET images acquired with different scanners," *Oncotarget*, vol. 8, no. 26, pp. 43169–43179, 2017.
- [28] B. Ganeshan, S. Abaleke, R. C. Young, C. R. Chatwin, and K. A. Miles, "Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage," *Cancer Imaging*, vol. 10, no. 1, pp. 137–143, 2010.
- [29] B. Ganeshan, V. Goh, H. C. Mandeville, Q. S. Ng, P. J. Hoskin, and K. A. Miles, "Non-small cell lung cancer: histopathologic correlates for texture parameters at CT," *Radiology*, vol. 266, no. 1, pp. 326–336, 2013.
- [30] O. S. Al-Kadi, D. Watson, *et al.*, "Texture analysis of aggressive and nonaggressive lung tumor CE CT images," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 7, pp. 1822–1830, 2008.
- [31] O. Grove, A. E. Berglund, M. B. Schabath, H. J. Aerts, A. Dekker, H. Wang, E. R. Velazquez, P. Lambin, Y. Gu, Y. Balagurunathan, *et al.*, "Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma," *PLOS One*, vol. 10, no. 3, pp. 1–14, 2015.
- [32] C. Cui, H. Cai, L. Liu, L. Li, H. Tian, and L. Li, "Quantitative analysis and prediction of regional lymph node status in rectal cancer based on computed tomography imaging," *European Radiology*, vol. 21, no. 11, pp. 2318–2325, 2011.
- [33] K. Holli, A.-L. Lääperi, L. Harrison, T. Luukkaala, T. Toivonen, P. Ryymin, P. Dastidar, S. Soimakallio, and H. Eskola, "Characterization of breast cancer types by texture analysis of magnetic resonance images," *Academic Radiology*, vol. 17, no. 2, pp. 135–141, 2010.
- [34] P. E. Galavis, C. Hollensen, N. Jallow, B. Paliwal, and R. Jeraj, "Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters," *Acta Oncologica*, vol. 49, no. 7, pp. 1012–1016, 2010.
- [35] C. Parmar, E. R. Velazquez, R. Leijenaar, M. Jermoumi, S. Carvalho, R. H. Mak, S. Mitra, B. U. Shankar, R. Kikinis, B. Haibe-Kains, *et al.*, "Robust radiomics feature quantification using semiautomatic volumetric segmentation," *PLOS One*, vol. 9, no. 7, pp. 1–8, 2014.
- [36] F. P. DiFilippo, J. P. Price, D. N. Kelsch, and R. F. Muzic Jr, "Porous phantoms for PET and SPECT performance evaluation and quality assurance," *Medical Physics*, vol. 31, no. 5, pp. 1183–1194, 2004.
- [37] M. Carles, I. Torres-Espallardo, A. Alberich-Bayarri, C. Olivas, P. Bello, U. Nestle, and L. Martí-Bonmatí, "Evaluation of PET texture features with heterogeneous phantoms: complementarity and effect of motion and segmentation method," *Physics in Medicine & Biology*, vol. 62, no. 2, pp. 652–668, 2016.
- [38] F. Gallivanone, M. Interlenghi, D. D'Ambrosio, G. Trifirò, and I. Castiglioni, "Parameters Influencing PET Imaging Features: A Phantom Study with Irregular and Heterogeneous Synthetic Lesions," *Contrast Media & Molecular Imaging*, vol. 2018, pp. 1–12, 2018.

- [39] L. Presotto, V. Bettinardi, E. De Bernardi, M. Belli, G. 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, pp. 66–74, 2018.
- [40] B. Baeßler, K. Weiss, and D. D. S. Pinto, "Robustness and Reproducibility of Radiomics in Magnetic Resonance Imaging: A Phantom Study.," *Investigative Radiology*, vol. 54, pp. 221–228, 2018.
- [41] A. Zwanenburg, S. Leger, M. Vallières, S. Löck, *et al.*, "Image biomarker standardisation initiative - feature definitions v1.7," *Computing Research Repository*, 1612.07003, 2016.
- [42] M.-F. Bellin, "MR contrast agents, the old and the new," *European Journal of Radiology*, vol. 60, no. 3, pp. 314–323, 2006.
- [43] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.
- [44] F. Tixier, C. C. Le Rest, M. Hatt, N. Albarghach, O. Pradier, J.-P. Metges, L. Corcos, and D. Visvikis, "Intratumor heterogeneity characterized by textural features on baseline 18F-FDG PET images predicts response to concomitant radiochemotherapy in esophageal cancer," *Journal of Nuclear Medicine*, vol. 52, no. 3, pp. 369–378, 2011.
- [45] R. T. Leijenaar, G. Nalbantov, S. Carvalho, W. J. Van Elmpt, E. G. Troost, R. Boellaard, H. J. Aerts, R. J. Gillies, and P. Lambin, "The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis," *Scientific Reports*, vol. 5, p. 11075, 2015.
- [46] M. Vallières, C. R. Freeman, S. R. Skamene, and I. El Naqa, "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Physics in Medicine & Biology*, vol. 60, no. 14, pp. 5471–5496, 2015.
- [47] S. Greenland, S. J. Senn, K. J. Rothman, J. B. Carlin, C. Poole, S. N. Goodman, and D. G. Altman, "Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations," *European Journal of Epidemiology*, vol. 31, no. 4, pp. 337–350, 2016.
- [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] M. J. Nyflot, F. Yang, D. Byrd, S. R. Bowen, G. A. Sandison, and P. E. Kinahan, "Quantitative radiomics: impact of stochastic effects on textural feature analysis implies the need for standards," *Journal of Medical Imaging*, vol. 2, no. 4, pp. 1–13, 2015.
- [50] E. Pfaehler, R. J. Beukinga, J. R. de Jong, R. H. Slart, C. H. Slump, R. A. Dierckx, and R. Boellaard, "Repeatability of 18F-FDG PET radiomic features: A phantom study to explore sensitivity to image reconstruction settings, noise, and delineation method," *Medical Physics*, vol. 46, no. 2, pp. 665–678, 2019.
- [51] B. Berthon, C. Marshall, A. Edwards, M. Evans, and E. Spezi, "Influence of cold walls on PET image quantification and volume segmentation: a phantom study," *Medical Physics*, vol. 40, no. 8, pp. 1–13, 2013.
- [52] J. Cortes-Rodicio, G. Sanchez-Merino, M. Garcia-Fidalgo, and I. Tobalina-Larrea, "Identification of low variability textural features for heterogeneity quantification of 18F-FDG PET/CT imaging," *Revista Española de Medicina Nuclear e Imagen Molecular (English Edition)*, vol. 35, no. 6, pp. 379–384, 2016.
- [53] 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.
- [54] J. Yan, J. L. Chu-Shern, H. Y. Loi, L. K. Khor, A. K. Sinha, S. T. Quek, I. W. 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.

Appendix A. Radiomic Features

Radiomic features overview

Overview of all radiomic feature classes with its possible features. There are seven different classes:

- Shape
- First order
- Gray-level co-occurrence matrix (GLCM)
- Gray-level run length matrix (GLRLM)
- Gray-level size zone matrix (GLSZM)
- Gray-level dependence matrix (GLDM)
- Neighbouring gray tone difference matrix (NGTDM)

Table A.1 shows all features and feature classes available with the software PyRadiomics 2.0 using Python 3.6 (Python Software Foundation, Wilmington, Delaware, United States).

Table A.1: All feature classes and 107 features available using PyRadiomics and Python for feature calculation and extraction.

Feature Class	Feature	Feature Class	Feature	
Shape	Elongation	GLRLM	GrayLevelNonUniformity	
	Flatness		GrayLevelNonUniformityNormalized	
	LeastAxisLength		GrayLevelVariance	
	MajorAxisLength		HighGrayLevelRunEmphasis	
	Maximum2DDiameterColumn		LongRunEmphasis	
	Maximum2DDiameterRow		LongRunHighGrayLevelEmphasis	
	Maximum2DDiameterSlice		LongRunLowGrayLevelEmphasis	
	Maximum3DDiameter		LowGrayLevelRunEmphasis	
	MeshVolume		RunEntropy	
	MinorAxisLength		RunLengthNonUniformity	
	Sphericity		RunLengthNonUniformityNormalized	
	SurfaceArea		RunPercentage	
	SurfaceVolumeRatio		RunVariance	
	VoxelVolume		ShortRunEmphasis	
	First order		10Percentile	ShortRunHighGrayLevelEmphasis
			90Percentile	ShortRunLowGrayLevelEmphasis
			Energy	GLSZM
Entropy		GrayLevelNonUniformityNormalized		
InterquartileRange		GrayLevelVariance		
Kurtosis		HighGrayLevelZoneEmphasis		
Maximum		LargeAreaEmphasis		
MeanAbsoluteDeviation		LargeAreaHighGrayLevelEmphasis		
Mean		LargeAreaLowGrayLevelEmphasis		
Median		LowGrayLevelZoneEmphasis		
Minimum		SizeZoneNonUniformity		
Range		SizeZoneNonUniformityNormalized		
RobustMeanAbsoluteDeviation		SmallAreaEmphasis		
RootMeanSquared		SmallAreaHighGrayLevelEmphasis		
Skewness		SmallAreaLowGrayLevelEmphasis		
TotalEnergy		ZoneEntropy		
Uniformity		ZonePercentage		
Variance	ZoneVariance			
GLCM	Autocorrelation	GLDM	DependenceEntropy	
	ClusterProminence		DependenceNonUniformity	
	ClusterShade		DependenceNonUniformityNormalized	
	ClusterTendency		DependenceVariance	
	Contrast		GrayLevelNonUniformity	
	Correlation		GrayLevelVariance	
	DifferenceAverage		HighGrayLevelEmphasis	
	DifferenceEntropy		LargeDependenceEmphasis	
	DifferenceVariance		LargeDependenceHighGrayLevelEmphasis	
	Id		LargeDependenceLowGrayLevelEmphasis	
	Idm		LowGrayLevelEmphasis	
	Idmn		SmallDependenceEmphasis	
	Idn		SmallDependenceHighGrayLevelEmphasis	
	Imc1		SmallDependenceLowGrayLevelEmphasis	
	Imc2		NGTDM	Busyness
	InverseVariance			Coarseness
	JointAverage			Complexity
	JointEnergy	Contrast		
	JointEntropy	Strength		
	MCC			
	MaximumProbability			
	SumAverage			
	SumEntropy			
SumSquares				

Appendix B. Radiomics reporting

Reporting by the recommended guidelines for reporting on the image processing sequence for image texture analysis, established by the image biomarker standardisation initiative (IBSI) (table B.1). [41]

Table B.1: Reporting of the image texture analysis according to the recommended IBSI guidelines.

General				
Aim	The aim of this study was to analyse the stability and accuracy of radiomic features in multimodality imaging.			
Acquisition		PET imaging Philips Vereos Digital	CT imaging Toshiba Aquillion ONE	MR imaging Philips Ingenia 1.5T
	18F-FDG activity	Background NEMA 20 MBq (EARL)	-	-
	Xenetix	-	Background NEMA 1.9mL/kg	-
	Gadolinium	-	-	Background NEMA 0.1mmol/kg
	Image acquisition	PET: 15 min (3min) – 2 bed positions. CT: 120 kVp, 40 mAs	120 kVp, 117 mAs FC08 convolution kernel, pitch 0.8, rotation time 0.5 s	3D gradient FFE T1 weighted TR/TE=7.6/6.4 ms 10° flip angle T = 3 min 39 s
	Reconstruction	Ordered-subsets expectation maximisation with CT attenuation correction (3 iterations, 15 subsets)	FC18 convolution kernel, 1 mm slice thickness. Adaptive Iterative Dose Reduction (AIDR)	
	Filter	Gaussian filter 5mm		
	Matrix	144x144	512x512	240x240
	Resolution (mm ²)	4x4	0.782x0.782	1.33x1.33
	Slice thickness (mm)	4	1	2
Slices	66	256	120	
Approach	Image is analysed as a volume (3D).			

<p>Structure</p>		
<p>Software</p>	<p>Image cropping and ROI segmentation: 3D Slicer 4.10.1 (Slicer Solutions) Registration: Imalytics research workstation 3.2 (Philips Technology, Aachen, Germany) Radiomic features: PyRadiomics 3.6</p>	
<p>Data conversion</p>		
<p>Procedure</p>	<p>Images were not converted to SUV but remained in order of Bq/mL.</p>	
<p>Image post-processing</p>		
<p>Procedure</p>	<p>None.</p>	
<p>Segmentation</p>		
<p>Procedure</p>	<p>Segmentation was performed using 3D Slicer 4.10.1 (Slicer Solutions). A circular ROI was drawn at the axial begin and end slice of the VOI. These two ROIs were interpolated to form a segmented VOI.</p>	
<p>Voxel interpolation</p>		
<p>Isotropic voxel dimension</p>	<p>PET acquisition PET reconstruction CT acquisition CT reconstruction MR acquisition MR reconstruction :</p>	<p>4mm x 4mm x 4mm 4mm x 4mm x 4mm 0.782mm x 0.782mm x 1mm 0.782mm x 0.782mm x 1mm 1.33mm x 1.33mm x 2mm 1mm x 1mm x 2mm</p>
<p>Interpolation procedure</p>	<p>Trilinear interpolation was used, grids aligned by centre.</p>	
<p>Grey level rounding</p>	<p>Not applicable.</p>	
<p>Grey level cut-off</p>	<p>Not applicable.</p>	
<p>ROI mask interpolation procedure</p>	<p>Trilinear interpolation was used, grids aligned by centre.</p>	
<p>ROI partial volume</p>	<p>Integer rounding.</p>	

Re-segmentation	
Inclusion/exclusion criteria	Not applicable.
Volume resection	
Bounding box	Volume of the images was resected using 3D slicer 4.10.1. The images were previous to co-registration cropped to three volume boxes of 87x85x65mm, one with 5mm insert compartment, one with 7.5mm insert compartment and one with 10mm insert compartment. Same measures were used for every modality and scan.
Discretisation	
Discretisation algorithm	For feature calculation a fixed bin number was used because MRI imaging contains arbitrary intensities, and therefore a fixed bin size is not recommended. Hatt et al. suggested a limited number of 64 discrete bins for textural extraction on PET-based images. In order to keep consistency in feature extraction, the same bin count was used for each modality.
Discretisation parameters	PET imaging: fixed bin count 64 CT imaging: fixed bin count 64 MR imaging: fixed bin count 64
Feature calculation	
Feature set	<ul style="list-style-type: none"> - First order (10): Energy, Entropy, Kurtosis, Maximum, Mean Absolute Deviation, Range, Root Mean Squared, Skewness, Uniformity, Variance - Gray Level Co-occurrence Matrix (GLCM): Autocorrelation, Cluster Prominence, Contrast, Correlation, Difference Variance, Inverse Difference Moment Normalized, Inverse Difference Normalized, Informational Measure of Correlation 2, Joint Energy, Maximal Correlation Coefficient, Sum Average, Sum Entropy, Sum of Squares - Gray Level Run Length Matrix (GLRLM): Gray Level Non-Uniformity Normalized, Gray Level Variance, High Gray Level Run Emphasis, Long Run Emphasis, Long Run High Gray Level Emphasis, Long Run Low Gray Level Emphasis, Low Gray Level Run Emphasis, Run Length Non-Uniformity Normalized, Run Percentage, Run Variance, Short Run Emphasis, Short Run High Gray Level Emphasis, Short Run Low Gray Level Emphasis - Gray Level Size Zone Matrix (GLSZM): Gray Level Non-Uniformity Normalized, Gray Level Variance, High Gray Level Zone Emphasis, Large Area Emphasis, Large Area High Gray Level Emphasis, Large Area Low Gray Level Emphasis, Low Gray Level Zone Emphasis, Size Zone Non-Uniformity Normalized, Small Area Emphasis, Small Area High Gray Level Emphasis, Small Area Low Gray Level Emphasis, Zone Percentage, Zone Variance - Neighbouring Gray Tone Difference Matrix (NGTDM): Busyness, Coarseness, Complexity, Contrast, Strength
Feature settings	Image normalisation and distance weighting were not applied. GLCMs were calculated in 13 directions and were by default symmetrical. All matrices were combined to one 3D GLCM where features were calculated to.
Standardisation	For the image acquisition the same phantom was used. Feature extraction was performed with the same parameter file and settings.

Appendix C. Parameter file

```
1 # K.M. Meijer
2 # Leiden University Medical Center
3 # Parameter file for feature extraction with Python 3.6
4 # March, 2019
5
6 # This file presents the parameters used for the calculation and
7 # extraction of radiomic features for research purposes.
8 # This parameter file was conducted using the example files of PyRadiomics 2.0
9 # The parameters are used for multimodality imaging scans:
10 # FDG-PET imaging, CT imaging and MR imaging.
11 # Images were extracted from imaging technique using a heterogeneity phantom,
12 # developed in the institute. No clinical data was used.
13
14 ## Extracted using PyRadiomics version: <2.0> ##
15
16 imageType:
17 # No filters were used on the derived images, therefore the original data is used.
18 # Image type original is the default
19 Original: {}
20
21 featureClass:
22 # Enable the classes for calculation
23 # Default is all enabling all feature classes
24 shape:
25 firstorder:
26 glcm:
27 glrlm:
28 glszm:
29 gldm:
30 ngtdm:
31
32 setting:
33 # Usual spacing for CT is often close to 1 or 2 mm, if very large slice thickness is used,
34 # increase the resampled spacing.
35 # On a side note: increasing the resampled spacing forces PyRadiomics to
36 # look at more coarse textures, which may or
37 # may not increase accuracy and stability of your extracted features.
38 interpolator: 'sitkBSpline' # sitkBSpline: (=3)
39 # resampledPixelSpacing: [1, 1, 1]
40 padDistance: 5 # Extra padding for large sigma valued LoG filtered images
41
42 # The ideal number of bins is somewhere in the order of 16-128 bins.
43 # A possible way to define a good binwidth is to extract firstorder:
44 # Range from the dataset to analyze, and choose a binwidth so,
45 # that range/binwidth remains approximately in this range of bins.
46 # Default and recommendation is binWidth: 25
47 # In this study MR imaging is used which contains arbitrary intensities.
48 # Therefore, a fixed bin size is used.
49 binCount: 64
50
51 # Default label value.
52 label: 1
```

Verantwoording

Het afgelopen jaar heb ik mijn afstudeerstage mogen lopen op de afdeling radiologie, sectie nucleaire geneeskunde, in het Leiden Universitair Medisch Centrum (LUMC). Met het voltooien van deze stage is er ook een einde gekomen aan bijna acht jaar studeren en is mijn masterdiploma de kers op de taart. Tijdens de afgelopen jaren heb ik veel geleerd en heb ik mij ontwikkeld als Technisch Geneeskundige, maar ook als persoon. In de laatste twee jaar van mijn master heb ik mijn wetenschappelijke kennis, academische vaardigheden en praktische vaardigheden mogen inzetten en verbreden in de praktijk tijdens verschillende stageperiodes. Daarnaast heb ik mij ook ontwikkeld op professioneel gebied, heb ik mijzelf leren kennen en ben ik gegroeid als persoon.

In mijn thesis is mijn onderzoek beschreven naar de reproduceerbaarheid en betrouwbaarheid van radiomics features voor het karakteriseren van tumor heterogeniteit. Om dit onderzoek uit te kunnen voeren heb ik een fantoom ontworpen en ontwikkeld. Dit heb ik gebruikt om op verschillende beeldvormende technieken experimenten uit te voeren. Ondanks dat dit onderzoek een voorbode is voor het verbeteren van de zorg van oncologische patiënten, was mijn opdracht van een meer technische aard. Ik vond het daarom een mooie uitdaging om het onderzoek dicht bij de kliniek te houden en het klinische aspect niet uit het oog te verliezen. Dit is ook versterkt doordat het onderzoek een onderdeel zou zijn van een groter project, waar onder andere een klinisch wetenschappelijk onderzoek uit voort zou komen.

Tijdens mijn afstudeerstage heb ik mij mogen verdiepen in alle facetten van een wetenschappelijke patiënten studie en heb ik gewerkt aan het opzetten van een dergelijke studie. Voordat een patiënten studie is goedgekeurd en kan beginnen zijn er meerdere onderdelen die afgehandeld moeten worden. Zo moet er een METC aanvraag gedaan worden en moeten er informatiebrieven voor patiënten geschreven worden. Gedurende het begin van mijn afstudeerstage heb ik de verantwoordelijkheid gekregen voor het opstellen van deze documenten. Hierdoor heb ik veel geleerd over het proces rondom de aanvraag hiervan en alles wat komt kijken bij een patiënten studie. In het kader hiervan heb ik ook de Basis cursus Regelgeving en Organisatie voor Klinisch onderzoekers (BROK) gevolgd en het examen hiervan succesvol afgerond en ben ik daarmee bevoegd dergelijk onderzoek uit te voeren. Daarnaast kreeg ik de kans om mijn academische schrijfvaardigheden te trainen en te verbeteren doordat ik de leiding had over het schrijven van de documenten. Voor het opstellen van deze documenten en het uitvoeren van mijn onderzoek heb ik met veel partijen samengewerkt van verschillende afdelingen en instituten. Aangezien er verschillende afdelingen mee zouden werken voor de beoogde patiënten studie, waren er ook meerdere invalshoeken die belicht werden. Vooral tijdens het opstellen van de METC aanvraag heb ik geleerd om klinische te denken en het probleem en de studie via verschillende medische brillen te benaderen, en daarnaast ook om met deze verschillende expertises te communiceren.

Ook voor mijn afstudeeropdracht heb ik verschillende expertises samengebracht. Mijn werkplek was op de nucleaire geneeskunde maar ik heb ook gebruik gemaakt van meerdere technieken en faciliteiten buiten de afdeling en zelfs buiten het ziekenhuis. Hierdoor kwam ik met mensen met verschillende expertises in aanmerking, en ik heb ik mijn kennis kunnen verbreden. Doordat er veel partijen betrokken waren betekende dit ook dat mijn opdracht erg breed was. Dit gaf mij aan de ene kant erg veel vrijheid en leermogelijkheden, maar aan de andere kant was er de valkuil dat de opdracht te breed werd en de essentie uit het oog verloren werd. Vooral in de laatste maanden, na het plannen en uitvoeren van de experimenten, moest ik mijn best doen om bij het beoogde doel te blijven. Er deden zich meer mogelijkheden voor, wat heel positief is, maar dat zou mijn opdracht veel te groot maken. Vooral in het proces van data analyse heb ik goed mijn grenzen aan kunnen geven en vertellen wat ik nog wel en niet kon doen voor mijn afstuderen. Ook zonder deze toevoegingen heb ik nog steeds een heel mooi en compleet onderzoek kunnen neerzetten.

In mijn afstudeerjaar heb ik ook geleerd om me echt verantwoordelijk te voelen voor een groot project. Waar ik in M2 stages een kleine opdracht deed, kreeg ik hier een project met een groter doel. Ik was erg enthousiast over het project en de verschillende taken en werkzaamheden die erbij kwamen vond ik een erg leuke uitdaging. Maar al snel had ik door dat het combineren van mijn afstuderen en het opzetten van een patiënten studie ook een pittige combinatie kon zijn. Plannen en prioriteiten stellen werden erg belangrijk, en daar heb ik ook zeker in geleerd. Ik vond het soms wel moeilijk om het overzicht te bewaren, en de prioriteiten juist te stellen zodat al het werk goed afkwam. Dit is ook

wel eens minder goed uitgepakt maar ik vond dat goede leermomenten die ik kon meenemen in het vervolg. Ik wou graag laten zien wat ik in mijn mars had, maar ik denk dat ik ook niet van mijzelf moest verwachten dat deze combinatie meteen zonder slag of stoot zou verlopen. Het was voor mij ook een geheel nieuwe situatie. Ondanks dat dit project niet is uitgepakt zoals ik had gehoopt, denk ik dat ik veel geleerd heb van deze periode en neem ik deze ervaring mee.

Een bijzonder onderdeel aan mijn afstuderen was de samenwerking met de Dienst Elektronische en Mechanische Ontwikkeling (DEMO) groep aan de Technische Universiteit Delft. In eerdere projecten was er al een samenwerking tot stand gekomen in het ontwerpen en ontwikkelen van fantomen. Ik vond het erg leuk dat ik deze samenwerking mocht voortzetten en een heel nieuw technisch domein mocht leren kennen, ver buiten mijn comfort zone. De mannen van de DEMO groep hebben mij veel geleerd op gebied van ontwerpen, mechanica en 3D printen. Ik heb mij ontwerpprogramma's als Inventor eigen gemaakt en vanuit een heel andere bril naar het probleem kunnen kijken. De leuke uitdaging was om het technische/mechanische samen te brengen met de klinische vraagstukken. Aangezien ik niet eerder met specifieke software en 3D print technieken had gewerkt, moest ik daar mijn eigen weg in zoeken maar vervolgens ook meteen de schakel zijn tussen de DEMO groep en het ziekenhuis. Omdat ik elke week in mijn eentje naar Delft ging, kwam die verantwoordelijkheid er automatisch bij kijken en ik voelde deze ook. Op het begin was alles nieuw voor mij waardoor het nog wat onwennig was, maar op een zeker moment merkte ik dat ik de touwtjes in handen kreeg. Ik wist wat ik wou bereiken in Delft, zorgde ervoor dat ik daarvoor de juiste middelen kon gebruiken, maar zorgde er tegelijkertijd voor dat het op één lijn zat met de relevante aspecten uit de kliniek. Ondanks dat ik vanuit mijzelf niet altijd het zelfvertrouwen heb gehad en vaak bij iemand eerst bevestiging zocht, denk ik dat ik afgelopen jaar hierin wel stappen heb gemaakt. Nadat ik op een zeker moment ook de terugkoppeling had gekregen dat ik de touwtjes goed in handen had met betrekking tot de ontwikkeling van het fantoom, groeide mijn zelfvertrouwen hierin. Ik had in zekere zin ook weinig keus omdat ik alleen naar Delft ging en op mijzelf werd aangewezen als er beslissingen gemaakt moesten worden. En ik denk dat ik zeker tevreden mag zijn met het resultaat, zowel met het ontworpen en ontwikkelde fantoom als met het feit dat ik meer heb durven vertrouwen op mijn eigen inzichten.

Het leuke aan mijn afstudeeropdracht vond ik dat het een brede opdracht was met veel verschillende aspecten, technieken en mogelijkheden. Ik heb echt onderzoek kunnen doen met vallen en opstaan om vervolgens het beste resultaat te creëren. Daarom ben ik ook erg blij en trots dat mijn onderzoek gepubliceerd gaat worden en dat ik een abstract van het onderzoek heb kunnen inleveren voor een congres.

Op klinisch gebied heb ik mij op de afdeling kunnen verdiepen in de verschillende onderzoeken, ziektebeelden en verslaglegging van radiologische beelden. Mijn afstudeeronderzoek richtte zich op kankersoorten met heterogene tumoren, met name in bot- en weke delenkanker. Ik heb veel geleerd over dit diverse en zeldzame ziektebeeld. Hiervoor heb ik veel multidisciplinaire overleggen (MDO) bijgewoond en heb ik kennis opgedaan over de complexe behandeling. Daarnaast heb ik me met name verdiept in het benaderen en verslaan van radiologische beelden. Ik heb geleerd om verschillende klinische factoren samen te brengen en medische beelden systematisch te benaderen. De grote diversiteit aan ziektebeelden vond ik een leuke uitdaging en heeft ervoor gezorgd dat ik mijn kennis heb kunnen verbreden. In de eerste maanden van mijn afstuderen denk ik dat ik de klinische vaardigheden te weinig heb opgepakt. Door het vele werk wat op mij af kwam in het begin is de kliniek in de schaduw komen te staan. Achteraf gezien was dit een goed leermoment voor het stellen van prioriteiten en een andere aanpak. Desondanks heb ik tijdig de draad opgepakt en ervoor gezorgd dat ik regelmatig de kliniek in ging.

In mijn afstudeerjaar ben ik ook meer te weten gekomen over hoe ik mijzelf zie als Technisch Geneeskundige, en welke eigenschappen van mij hierbij passen en waar ik nog aan kan werken. Ik ben iemand die zich betrokken voelt bij het werk wat van mij verwacht wordt. Ik stel me leergierig op en neem de verantwoordelijkheid voor mijn werk. Daarnaast ga ik een uitdaging niet uit de weg en ben ik een harde werker. Afgelopen jaar heb ik dit meegemaakt op meerdere vlakken. De start van mijn afstuderen vond ik een pittige periode. Ik kwam in een nieuwe werkomgeving, nieuwe afdeling en een nieuwe opdracht in combinatie met promotiewerk. Er kwam veel op mij af en ik wou graag laten zien wat mijn capaciteiten waren. Ik heb vanaf het begin hard gewerkt en geprobeerd om mijn plek te vinden. Ik denk dat ik hier wel zoekende in ben geweest, maar ik voelde dat het beter ging en dat ik stappen maakte. Dit gaf mij dan ook steeds meer het vertrouwen dat ik op de goede weg was, en bezig was om mijzelf als onderzoeker en Technisch Geneeskundige te verbeteren.

Toen halverwege mijn afstudeerjaar het promotietraject voor mij wegviel, volgde daarop een lastige periode. Ik had de maanden ervoor hard gewerkt en was nog steeds enthousiast over het gehele project, waardoor de teleurstelling groot was. Tot op heden vind ik het jammer dat het op deze manier is gelopen. Deze gebeurtenis was wel een tekenend moment in mijn afstudeerjaar en heeft veel invloed gehad op de maanden die nog volgden. Het heeft er in eerste instantie voor gezorgd dat mijn motivatie meteen een dieptepunt bereikte. Ik voelde me niet meer op mijn plek en het kostte me erg veel energie om me in te zetten voor mijn afstudeeronderzoek. Toen de directe teleurstelling was afgezwakt, kreeg ik langzamerhand weer wat enthousiasme terug om de opdracht te voltooien. In de opdracht zelf zag ik immers nog steeds potentie en ik vond het onderzoek nog steeds leuk. Het heeft zeker een tijd geduurd om mezelf weer te herpakken en mijn schouders er weer onder te zetten om een mooi afstudeeronderzoek neer te zetten. In deze periode heb ik mijn veerkracht en doorzettingsvermogen laten zien. Ik wou niet opgeven en mijn studie op een mooie manier afsluiten, wat naar mijn idee zeker gelukt is.

Op professioneel gebied ben ik een bescheiden persoon en kan ik mij terughoudend opstellen. Ik heb er soms moeite mee om te vertrouwen op mijn eigen kennen en kunnen, waardoor ik in situaties met andere professionals snel een afwachtende houding aanneem. Afgelopen jaar heb ik op verschillende vlakken geleerd dat ik voor mijzelf mag opkomen en dat mijn mening telt. Op momenten dat het tegen zat ben ik de confrontatie aangegaan en heb ik mijn mening en ervaring kunnen vertellen. Op die momenten heb ik geleerd om mijn gevoelens en ervaringen zo goed mogelijk uit te drukken, voor mijzelf op te komen en de regie te pakken. Ik heb mij hierin kwetsbaar opgesteld, maar achteraf heeft me dit juist meer zelfvertrouwen gegeven. Terugkijkend op deze momenten ben ik erg trots op mijzelf hoe ik met de situatie ben omgegaan. Ik heb veel geleerd van deze periode en ik heb me goed herpakt om mooi resultaat neer te zetten.

Wat mij kenmerkt is dat ik goed zelfstandig kan werken, en dat ik graag zelf naar de oplossingen zoek. Ik zie dat als een positieve eigenschap, maar ik heb tijdens mijn stages ook geleerd om hulp te vragen en samen te werken. Ik denk dat ik hierin gegroeid ben tijdens mijn stages en dat ik een goede teamplayer kan en ook wil zijn. Mijn werkethos, verantwoordelijkheidsgevoel, veerkracht en enthousiasme zijn eigenschappen die mij kenmerken en die ik ook steeds meer ben gaan waarderen. Voor mijn gevoel gingen de laatste jaren TG niet altijd even gemakkelijk, maar ik heb veel bereikt. Ik heb geleerd dat iedereen zijn eigen kwaliteiten heeft en zijn eigen pad bewandeld om een doel te bereiken. Het pad wat ik de afgelopen jaren heb bewandeld heeft me veel geleerd over mijzelf en wie ik wil zijn als Technisch Geneeskundige. Daarbij moet ik zeggen dat een vleugje humor en goede koffie zeker een aandeel hebben gehad om de eindstreep te bereiken. De opgedane kennis en ervaringen neem ik mee en ik ben erg benieuwd wat de toekomst mij zal brengen. Kortom, hier staat een trotse Technisch Geneeskundige!

Karin Meijer

Dankwoord

Na bijna acht jaar studeren mag ik mijzelf dan ook eindelijk Master of Science noemen! Hard werken wordt beloond, daar ben ik het zeker mee eens. Met het behalen van mijn diploma komt er ook een einde aan mijn fantastische studententijd waar ik erg veel leuke dingen heb mogen mee maken en ook erg veel leuke mensen heb ontmoet! Mijn studiejaren ben ik natuurlijk niet zomaar doorgekomen en daarom wil ik nog een aantal mensen bedanken.

Afgelopen jaar heb ik stage gelopen en mijn afstudeeronderzoek gedaan bij de nucleaire geneeskunde in het LUMC. Ik wil graag alle medewerkers van de NuGe bedanken voor het leuke jaar. Er was altijd een fijne sfeer die ervoor gezorgd heeft dat ik me thuis heb gevoeld op C2. In het bijzonder wil ik graag de mensen van de corner office bedanken. Pim, Wyanne, Dennis, Joeri, Willem, Alina en alle M2 studenten die stage hebben gelopen; bedankt voor de gezelligheid, de koffiepauzes, de flauwe grappen, de KB-meetings, de vrimibo's, maar ook zeker jullie hulp afgelopen jaar! In de soms lastige tijden hebben jullie ervoor gezorgd dat de dagen leuk werden en ik heb jullie als collega's erg gewaardeerd! Deze is voor jullie: "Oooh noh!"

Gedurende mijn onderzoek zijn er meerdere mensen van verschillende afdelingen betrokken geraakt bij mijn onderzoek en dankzij hun hulp heb ik mijn resultaten kunnen krijgen. Allereerst wil ik graag prof. Bloem bedanken voor alle hulp, met name voor de klinische aspecten van het onderzoek. Ik heb veel van u geleerd en uw begeleiding heb ik altijd zeer prettig gevonden. Daarnaast wil ik Floris, Irene en Wyger bedanken voor hun hulp bij het uitvoeren van de PET, CT en MRI experimenten. Dankzij jullie is het mogelijk geworden dat ik meerdere experimenten heb kunnen doen met mijn fantoom. Hierbij wil ik ook graag Maryam bedanken voor de input vanuit Philips en de hulp om onze wensen voor het onderzoek voor elkaar te krijgen. Ook wil ik de mensen van de DEMO groep in het RID op de TU Delft bedanken, en in het bijzonder Dimitri, Kevin en Juliana. Ik heb me altijd erg welkom gevoeld in Delft en jullie hulp heeft ervoor gezorgd dat ik een mooi fantoom heb kunnen ontwerpen en ontwikkelen.

Tijdens mijn studiejaren voor mijn afstuderen heb ik mogen genieten van een fantastische studententijd in Enschede. Ik kan hier heel veel mensen voor bedanken maar in het bijzonder wil ik graag Kirsten, Marijn, Robin, Seraya en Suzan bedanken voor de superleuke jaren samen in de collegebanken van TG. De koffie/thee pauzes in de PK, de biertjes in de stad en de (jaarlijkse) tripjes die we samen ondernemen zijn altijd een feest. Gelukkig wordt deze traditie in stand gehouden! Ook wil ik graag de meiden uit Wierden, de Monday-wine-night members, het leukste huis Amaranthos, de liefste jaarclub Florera en de tofste meiden van Leone Rode bedanken voor de toptijden en de vriendschappen die zijn ontstaan. En Florian, bedankt voor de gezellige GV/thee-momenten en je aanmoedigingen. Je bent een topper!

Erik en Imke, bedankt voor onze goede band, jullie gezelligheid en aanmoediging. En Erik, bedankt dat je me als grote broer hebt onderwezen in het studentenleven, bier en muziek!

En natuurlijk last but not least: pap en mam, bedankt voor jullie onvoorwaardelijke liefde, steun en stimulans. Jullie hebben mij geleerd dat hard werken wordt gewaardeerd en beloond, maar dat genieten van de mooie en leuke dingen in het leven voorop staat.

Alleen maar liefde voor jullie!

Karin

