Aus der Universitätsklinik für Radioonkologie mit Poliklinik
Tübingen

Sektion Biomedizinische Physik

CT-radiomics in the Context of Outcome Prediction after Chemoradio Therapy (CRT) in Cancer Patients

Inagural-Dissertation
zur Erlangung des Doktorgrades
der Humanwissenschaften

der Medizinischen Fakulät
der Eberhard Karls Universität
zu Tübingen

vorgelegt von

Socarrás Fernández, Jairo Andrés

2020

Dean: Professor. Dr. B. Pichler

1. Berichterstatter: Professor. Dr. D. Thorwarth
2. Berichterstatter: Professor. Dr. A. Schilling

Tag der Disputation: 18.09.2020

Dedication

A special dedication to who made Tübingen a new home for me: Manu, Bibiana and Corina, as well as to who have always supported me from Bogotá, Munich and Madrid (Torrejón), my incomparable friends: Angélica Abello Cubillos, Leonardo Suárez Sánchez, Paulo Plazas Hurtado, Nicolas Pérez Amaya, Camila Castro Rico and Irene Mayoral Soriano. To all of you guys, long lasting friendship.

Quiero dedicar especialmente este trabajo a mi madre quien gracias a su incondicional apoyo, su dedicación y el ejemplo que me ha dado a lo largo de estos años, me ha dado fortaleza en momentos difíciles y como guía para conseguir lo que me apasiona.

Mein besonderer Dank gilt außerdem meiner geliebten Verbündeten, Kollegin und zukünftige Frau Anna-Lena Peschel für ihre unermüdliche Ermutigung, Hilfe, Liebe und Hingabe in dieser Zeit. Ich danke ihr auch für die Sprachkorrekturen und Anmerkungen inhaltlicher Art.

Acknowledgements

"Our intelligence is what makes us human, and AI is an extension of that quality."

Yann LeCun

Abstract

The World Health Organisation (WHO) has described cancer as the second worldwide cause of death, with Head and Neck Squamous Cell Carcinoma (HNSCC) being the sixth most frequent form of cancer. Different reasons for treatment failure have been discovered, and studies of these propose to personalise cancer treatments as the best strategy to reduce these figures. Radiomics is a relatively new field that uses medical image analysis as a manner to find promising imaging biomarkers to potentially adapt cancer treatments accordingly. Despite the promising results of different studies, no reproducibility has been reached across the scientific community, mainly because of the lack of open and comparable implementations, standardised mathematical definitions of features and imaging pre-processing methods, scan set ups and feature pre-processing. This thesis aims to implement a robust and reproducible radiomic pipeline for studies of risk identifications in patients to assess the feasibility to prospectively personalise CRT in the future.

The radiomic pipeline implemented inside the framework of this thesis was validated in an international collaboration with participation of more than 20 well-recognised institutions across Europe, the United States and Canada. There was addressed the problem of reproducibility due to feature extraction algorithms and definitions as well as imaging pre-processing implementations using a common framework, mathematical definitions of image features and two proposed phantom data for testing purposes.

This radiomic software was customised for clinical research studies inside the necessities of the University Hospital Tübingen. First, correlations between the CT-radiomic signature proposed by Aerts et al. associated to intra-tumour heterogeneity and the somatic mutation of genes associated to aggressiveness in HNSCC were investigated. A CT-radiomic approach was used to propose a substitute identifier of patients at risks of the well-established patient classifier Tumour-to-background ratio $(TBR)_{peak}$ in HNSCC cancer patients. Finally, a CT-radiomics signature was proposed for forecasting tumour regression grades in rectal cancer types with independent validation cohort for organ-preserving strategies in therapies.

This work concludes that variations in medical imaging pre-processing algorithms are suggested to be the main source of unrepeatable radiomic studies and benchmarks are therefore needed, which at the moment of writing this thesis are being obtained. It also remarks that one of the biggest limitations of radiomics is related to the number of patients used for studies which requires larger cohorts for generalisations of radiomic signature performances for outcome predictions, thus prospective

studies and larger number of patients are suggested to confirm all findings inside the framework of this thesis.

Zusammenfassung

Die Weltgesundheitsorganisation (WHO) bezeichnet Krebs als die zweithäufigste Todesursache weltweit und HNSCC als die sechsthäufigste Krebsart. Es wurde nach den verschiedensten Gründe für das Versagen einer Krebstherapie gesucht und herausgefunden, dass die beste Strategie, die Anzahl der Rezidive zu minimieren, eine personalisierte Krebstherapie ist. Radiomics ist ein relativ neues Gebiet, das mit Hilfe von medizinischer Bildanalyse nach Biomarkern sucht, mit deren Hilfe die Krebsbehandlung individuell angepasst und optimiert werden kann. Trotz der vielversprechenden Ergebnisse verschiedenster Studien über diese Methode, konnte bislang keine dieser reproduziert werden. Dies liegt unter anderem an dem Fehlen von öffentlichen, vergleichbaren Implementierungen, an fehlender Standardisierung von mathematischen Merkmalen, an den unterschiedlichsten Vorbearbeitungsverfahren und Vorbearbeitungsmethoden der Bilder sowie an den meist unbekannten CT-Scan Einstellungen. Das Ziel dieser Dissertation ist es, eine robuste und reproduzierbare Pipeline in Radiomics zu implementieren, die eine Risikoidentifizierung eines Patientens ermöglicht und über die Durchführbarkeit einer zukünftigen, personalisierten CRT entscheidet.
Die im Rahmen dieser Dissertation implementierte Radiomic-Pipeline wurde in einer internationalen Zusammenarbeit unter der Beteiligung von mehr als 20 anerkannten Institutionen in Europa, den USA und Kanada validiert. Das Problem der Reproduzierbarkeit aufgrund von Merkmalsextraktionsalogarithmen und -definitionen wurde besprochen, genauso wie die Implementierung von Bildvorverarbeitungen, die Verwendung eines gemeinsamen Rahmens von mathematischer Definitionen, gleiche Bildmerkmalen sowie das Benutzen von zwei Phantomdaten zu Testzwecken.
Diese Software, basierend auf Radiomics, wurde für klinische Forschungsstudien innerhalb des Universitätsklinikums Tübingen entworfen. Im ersten Teil der Arbeit wurden Korrelationen zwischen dem von Aerts et al. (2014) vorgeschlagenen CT-Radiomic-Signature und Intra-Tumor Heterogenitätem bei HNSCC untersucht sowie Korrelationen zwischen der somatischen Mutation von Genen und deren aggresives Zellverhalten. Eine der CT-Radiomic-Ansätze konnte dafür verwendet werden Risikopatienten des HNSCC zu identifizieren. Dieser Ansatz diente als Ersatz zu dem bereits etablierten Tumor-To-Background (TBR_{peak}) Klassifikationsansatz. Schließlich wurde eine CT-Radiomics Methode zur Vorhersage der Tumorregressionsrate bei Rektalkarzinomkrebs benutzt und in einer unabhängigen Validierungskohorte getestet. Hierbei wurden organerhaltende Therapien bevorzugt.

Das Ergebnis der Dissertation ist, dass Variationen der Vorverarbeitungsalgorithmen für die medizinische Bildgebung die Hauptquelle für nicht reproduzierbare radiomische Untersuchungen sind. Daher sind Richtlinien unerlässlich, die zum Zeitpunkt der Erstellung dieser Doktorarbeit zeitgleich entwickelt wurden. Es soll auch an dieser Stelle darauf hingewiesen werden, dass eine der größten Einschränkungen von Radiomics eine zu geringe Anzahl an Patientendaten ist. Eine zuverlässige Verallgemeinerung der Radiomischen Signaturleistung ist daher nur mit einer ausreichend großen Anzahl an Patientendaten möglich und liefert auch erst dann reproduzierbare Ergebnisse. Daher wird für die Bestätigung der Ergebnisse dieser Arbeit, sowie für zukünftigen Studien an diesem Thema, eine Erhöhung der Patientendatensätze empfohlen.

List of Abbreviations

^{18}F	Fluorine
3D	three dimensional
ADC	Apparent Diffusion Coefficient
CRT	Chemoradio Therapy
CT	Computer Tomography
DCE	Dynamic Contrast-Enhanced
DW	Diffusion-Weighted
EES	Extravascular extracellular space
EGFR	Epidermal Growth Factor Receptor
FDG	fluorodeoxyglucose
FMISO	fluoromisonidazole
FT	Fourier transform
GLCM	grey level co-occurrence matrix
GLDZ	grey level distance zone matrix
GLRL	grey level run length matrix
GLSZ	grey level size zone matrix
GTV	Gross tumour volume
HNSCC	Head and Neck Squamous Cell Carcinoma
HVF	Hypoxic volume fraction
IBSI	Image Biomarker Standardisation Initiative
K^{trans}	transfer constant
ML	Machine Learning
MRI	Nuclear Magneto-Resonance Imaging
MTV	Metabolic tumour volume

NGTD   neighbourhood grey tone difference

PCA   Principal component analysis
PET   Positron Emission Tomography

RBF   Radial Basis Function
RECIST   Response Evaluation Criteria in Solid Tumours
RF   radio frequency
ROI   region of interest
RT   Radiation Therapy

SUV   standardised uptake value

TBR   Tumour-to-background ratio
TE   Time echo
TGF   Transforming Growth Factor
TLG   Total lesion glycolysis
TR   Time repetition

UKT   University Hospital Tübingen

v_e extra-vascular extracellular space volume fraction
VEGF   Vascular endothelial growth

WHO   World Health Organization

Contents

Summary of the Thesis

Part I: Introduction

CHAPTER 1: INTRODUCTION

In chapter 1, motivations and state of the art of the topics treated inside the framework of this thesis are supplied.

Part II: Materials and Mehtods

CHAPTER 2: BASICS OF IMAGING MODALITIES AND BIOMARKERS FOR RT

The chapter 2 followed to explain the basic physics and image reconstruction of Computer Tomography (CT), Nuclear Magneto-Resonance Imaging (MRI) and Positron Emission Tomography (PET). Besides it shows the origins of the most common imaging markers from static to dynamic models with their advantages and disadvantages.

CHAPTER 3: THE RADIOMIC HYPOTHESIS AND IMAGE FEATURE ENGINEERING

The chapter 3 provided a compelling introduction to the field of radiomics along with explanations of the most common features for image analysis and how radiomic incorporates them to construct relevant imaging biomarkers.

CHAPTER 4: MACHINE LEARNING IN THE CONTEXT OF RADIOMICS

In chapter 4, a brief description of the most common machine learning algorithms for feature selection and modelling was imparted in regards to radiomics. Moreover, a short discussion about model validation and tuning and selection in this work is also supplied.

CHAPTER 5: PYTHON AS A SOFTWARE DEVELOPMENT LANGUAGE

In chapter 5, a brief description of Python as programming language was provided as well as a short discussion about the reasons to use it and some comparisons with other languages were explained.

Part III: Results

CHAPTER 6: IN-HOUSE SOFTWARE DEVELOPMENT FOR RADIOMICS AND ITS VALIDATION

The chapter 6 yielded in an introductory description of the software implementation and discloses the validation benchmarks of the software inside the international col-

laboration IBSI.

Chapter 7: Radiomics pipeline inside this thesis

The chapter 7 provided a detailed discussion of the radiomics pipeline used in this work as well as the hyperparamenter spaces used for every machine learning model used.

Chapter 8: Correlations of the Aerts signature with somatic mutations in TP53, FAT1 and KMT2D in HNSCC

In chapter 8 summarized an exploratory study to confirm the Aerts signature with heterogeneity in cell pathways mutations, the results are therefore discussed in an adaptation form from the publication of [184].

Chapter 9: CT radiomics: A surrogate predictor of patients at risk as identified by Fluorine (^{18}F)-fluoromisonidazole (FMISO) hypoxia PET?

In chapter 9, the feasibility to use CT-radiomics as a substitute strategy to classify patients at risk in comparison to standard hypoxia-based static model was investigated. The results were thus disclosed.

Chapter 10: An approach for prediction of complete remission of organ-preserving therapy in rectal carcinoma with independent validation)

In chapter 10, the feasibility to use CT-radiomics as a strategy to improve organ-preserving therapies in rectal carcinomas patients was investigated. The results were thus disclosed.

Part IV: Discussion and outlook

Chapter 11: Discussion and outlook

Finally, chapter 11 provides a summary of the observations, discussions and scopes of this work and an outlook to future research.

Part I.

Introduction

1. Introduction

The World Health Organization (WHO) has described cancer as the second worldwide cause of death [9]. The standard treatments are surgery, chemotherapy and RT, which the latter is estimated to profit all cancer patients up to 50% either as a stand-alone treatment or in combination of other cancer treatments [36]. In brief, RT treatment takes advantage of the normally slower grow rate of cell reparation mechanism after exposition to ionising radiation of tumour tissues in comparison with healthy tissues. The consequences are differences in the cancer cell death rates [10] or in other words the tumour shrinking. However, even though RT causes bigger damages in tumour cells, it also produces burdens in healthy tissue. Consequently, fractional RT was proposed in the 1920s as a remedy to maximise targeted tumour damages with minimum effect in healthy tissues [16]. Since then, RT has incorporated advances such as conformal and intensity or volumetric modulated radiation therapy techniques to conform radiation field shape adaptations to tumour shapes, photon and heavy ions therapy to enhance dose depth depositions, minimising skin side effects [85] and three dimensional (3D) image-based guidance (CT or MRI and PET) for accurate tumour distinction and location [24].

Nevertheless, many cancer patients still recur either locally or in distant metastases after treatments. This is the fact for instance in HNSCC where the overall survival rates ranges between 40% and 50% in advanced stages with followups up to 60 months [9] after beginning of CRT. Similar results are seen for other types of tumours. The reasons of these treatment failures are associated to tumour growth (Epidermal Growth Factor Receptor (EGFR) and Transforming Growth Factor (TGF) [61]), proliferation (Cyclin D1 [26,155], Ki-67 [32,47,140]), suppression (p53 and p27 [21,25,44]), angiogenesis and hypoxia (Vascular endothelial growth (VEGF), HIF1α and HIF2α [94,115]), invasive and metastatic potential (MMP [120]), stem cells [135], heterogeneity of cancer cells ([182]), etc. Which with the state-of-the-art clinical cancer diagnosis (tumour type, stage, location and histology) cannot be described. After successfully sequencing the human genome, many research groups have joined efforts to develop the so-called precision medicine (also known as personalised medicine) to adapt treatments according to patients characteristics and thus increase effectiveness [58]. However RT remains difficult to comply with personalised doses due to a lack in assessing individual tumour characteristics non-invasively. Biological properties of tumours may be analysed from the tumour tissue, but practitioners tend to avoid this procedure since it is challenging and highly invasive to patients, which might yield to additional unnecessary risks. Moreover gene sequencing from tumour

samples remains difficult as it is expensive and cannot reflect the full gene expression in the whole tumour in a temporal and spatial level. Subsequently, other techniques have been explored in the recent 30 years to obtain tumour functionality and markers from tomographic data, which could, in their turn, lead to an early dose adaptation in CRT treatment, thus preventing recurrence after initial tumour detection in the applied therapy.

Hypoxia in tumours, which is associated to a feature of low radiation sensitivity [63] and therefore to poor prognosis for treatment outcome [125], can be assessed from changes of image intensities in PET scans along with appropriate radio-tracers [91,183]. This characteristic has been heavily exploited from many researchers to construct hypoxia models or imaging markers to be integrated in RT [137,163]. Nonetheless, researchers have faced difficulties associated to PET reliability [73] and lack of definitive gold standards referring to the period of acquisition of hypoxic information from images or thresholds at the moment classify hypoxic regions [137,183], among others. In spite of the infeasible clinical use, up to date hypoxic models and markers are still promising approaches for dose painting customisation and dose escalations for RT [163].

Changes in the tumour micro-environment, chaotic vasculature and abnormal permeability, can also be tracked by MRI either in modalities such as Dynamic Contrast-Enhanced (DCE) or Diffusion-Weighted (DW). The most promising imaging parameters in these techniques, to predict cancer therapy outcomes, is the pharmacokinetics modelling of gadolinium passage inside the tumour structure. In DCE-MRI parameters such as transfer constant (K^{trans}) and extra-vascular extracellular space volume fraction (v_e) seem to provide prognostic information to therapy outcomes [88]. On the other hand, In DW-MRI parameters such as Apparent Diffusion Coefficient (ADC) are more prognostic to CRT [89]. Despite of the advantages that the use of MRI represents in terms of invasiveness, these parameters lack of validation owe to confounding studies [169]. Besides, clinical routine implementations remain challenging since the need of co-registration between MRI sequences (T1, T2, ADC, etc.) specially in organs at motion, which generate modelling inaccuracy constrains.

On the other hand, Radiomics, which emulates the so-called "Omics" analysis (e.g. genomics, proteomics, transcriptomics, etc). It uses an outstanding number of mathematical measurements from radiological images, also called features, as inputs [98] to find significant numbers of imaging signatures that may be prognostic to cancer treatment outcomes and then those ones are correlated to cancer phenotype expressions [5]. Notwithstanding, it suffers from reproducibility, standardisation and signature significance comprehension across institutes and research groups [57,68,96,178]. The two former are inherited from problems of imaging science, the latter is, however, due to a few studies that investigate the biological meaning of features instead of pre-treatment prognosis power. In recent years, research efforts have been employed

to find correlations between radiomics signatures with biological or clinical information [20, 33, 55], which allow researchers and medical doctors to explain logically the subtle information of targeted tumours. Nonetheless, the number of studies remain sparse. The goal of this thesis is to construct a robust and reproducible radiomic pipeline inside the University Hospital Tübingen (UKT) that allows implementations or incorporations of comprehensive radiomic studies inside the necessities of radio-oncological research and assesses the feasibility to use the pipeline as a standalone or complementary information for CRT treatment customisation.

1.1. Hypothesis

This thesis assumes that tomographic data encodes the biological information of tumours. Therefore, the hypotheses of this work are divided in three as follows:

1. Computers are able to learn from the tomographic data, consequently correlate phenotype features with the underlying biology encoded in tumours, which it should be reproducible independently on the research institute.

2. Radiomics as a technique consists of other imaging markers that should be at least as good as other heavily investigated and potential imaging markers.

3. Radiomics has the potential to improve CRT either in unnecessary treatment applications or stratification of the dose required for successful treatment in cancer patients.

1.2. Description of the Thesis

This thesis is divided in four principal sections, as follows:

1. A compelling discussion of the materials and methods used in this work, such as a description of radiological images and imaging biomarkers (cf. Chapter 2), radiomics (cf. Chapter 3) and machine learning in the context of radiomics (cf. Chapter 4).

2. A collaboration in a common framework reference of radiomics feature definitions for standardised studies (cf. Chapter 6.1), machine learning methods for feature selection and models were included inside our pipeline for robust modelling (cf. Chapter 6).

3. An exploratory study of the correlation of a highly frequent referenced CT-radiomics signature with cell-pathways mutations in HNSCC (cf. Chapter 8).

4. An exploratory study to assess outcome prediction power of a constructed CT-radiomic signature from the significant patients cohort of HNSCC recruited in the last 10 years in the UKT, wherein part of them were used in studies for risk stratification according to an imaging marker derived from PET. The objective is to evaluate whether CT-radiomics could work either as a standalone, complementary or non-conclusive method for risk evaluation in HNSCC patients (Chapter 9).

5. A construction of a CT-radiomics signature to forecast tumour regression grades in rectal carcinoma with fully external validation for improvement of organ-preserving strategies.

Part II.

Materials and Methods

2. Basics of imaging modalities and biomarkers for RT

Since the discovery of X-rays by W. Röntgen in 1895, imaging scans have become a well established guide for treatment planning and adaptations. In more than 100 years, RT has profited from CT, MRI and PET discoveries to obtain either anatomical and/or functional information of imaged tumours and subsequently adapt treatments accordingly. This section aims to discuss the physics and imaging formation from the most common imaging modalities and some imaging markers that have been so far associated either to poor performance after RT.

2.1. Computed Tomography

CT scans are the most routinely used imaging modality in clinics for diagnosis and RT treatment planning, because of the prominent trade-off that CT represents in terms of cost, image resolution, set-up and invasiveness among others. The physics behind CT is based on X-rays, which are photons striking with energies in the range of a few keV to 100 keV [165]. When X-rays interact with matter, photons might be absorbed or/and scattered, resulting in energy attenuation, the responsible phenomena for these are known as photo-electric, Compton and pair-production effect. Photon attenuations after matter interaction are the base of forming X-ray images, since the attenuation coefficient (μ) differs in each body of interaction, μ can be encoded using the Beer-Lambert equation (cf. equation 2.1). $\mu(\vec{x})$ across beam lines are accessible if X-rays intensities in the source (I_0) and detectors (I_d) are measured. In principle μ is associated intrinsically to the atomic number of materials and tissue density.

$$\int_{\vec{x} \in l} \mu(\vec{x}) \, ds = \ln\left(\frac{I_0}{I_d}\right) \tag{2.1}$$

2.1.1. Data Acquisition

A scheme of data acquisition in CT is shown in figure 2.1. The patient is placed in the centre between the collimated X-ray beam and detectors (also called iso-centre). After the X-ray fan beam passes through the patient, attenuation occurs across the beam line. The residual intensity of the beam is then captured by detectors. The process is repeated for different angles of rotation in one (or many) cycles. Finally, μ are recorded computationally for every rotation for purposes of image reconstruction.
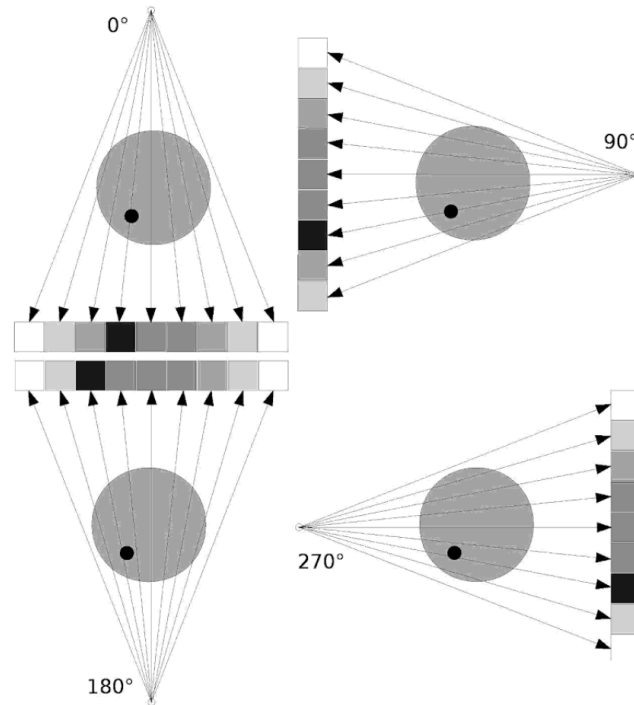
Figure 2.1.: Scheme of the CT principle in two dimensions. Arrows represent X-rays from the source through the body and reaching detectors ordered as an array in different angles. Reprinted from [18] with allowance from Taylor & Francis Group LLC - Books ("Publisher") provided by Copyright Clearance Center ("CCC")

The grey values obtained are proportional to the attenuation (cf. equation 2.2) and given in Hounsfield units (HU), named after the CT inventor. For quality assurance, the value for air corresponds to -1000 HU and 0 HU for water. A typical CT data posses 12 bit precision from -1000 to 3095 (cf. figure 2.2). Modern CT scanners require for a single slide around 1000 irradiations in different angles, which requires several revolutions of the X-ray beam. In the 3D case the beam rotates in spiral form while the radiographic table is translated smoothly (cf. figure 2.3).

$$\mu^* = \frac{\mu - \mu_{water}}{\mu_{water}} \cdot 100 \tag{2.2}$$

2.1.2. Image reconstruction

The image reconstruction methods can be divided in two types, either based on analytic or algebraic solutions. In the former, a continuous model of CT image is constructed to then derive an analytic solution which finally will be discretised in order to yield a numerical solution. The latter rather discretises the CT image
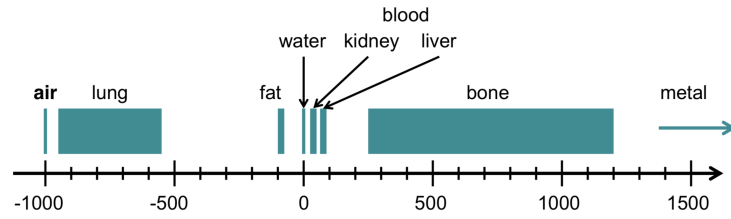
Figure 2.2.: Corresponding $\mu*$ in HU for different region of interest (ROI) in human body. Figure taken from presentation of Wolfgang Wein in the lecture Computer aided medical procedures at TUM.
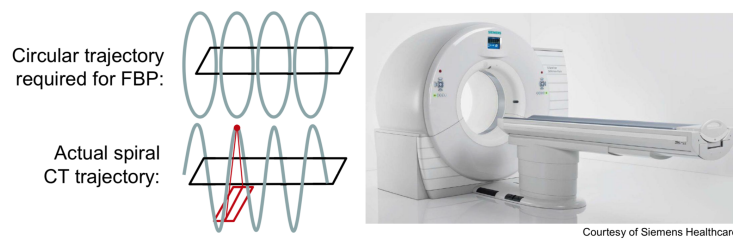
Figure 2.3.: Siemens CT scan set up, and irradiation direction of the X-ray beam when patient insertion. Figure taken from presentation of Wolfgang Wein in the lecture Computer aided medical procedures at TUM.

from the beginning (cf. figure 2.4), where the problem is seen as a large system of linear equations (cf. equation 2.3) to be solved. Algorithms such as back-projection or filtered back-projection are commonly used to solve the Radon transform and obtain the reconstructed image. Since image contrast depends on variations of mean attenuations along the beam paths, tissues with similar attenuation values will be barely distinguished as in the case of soft tissues.
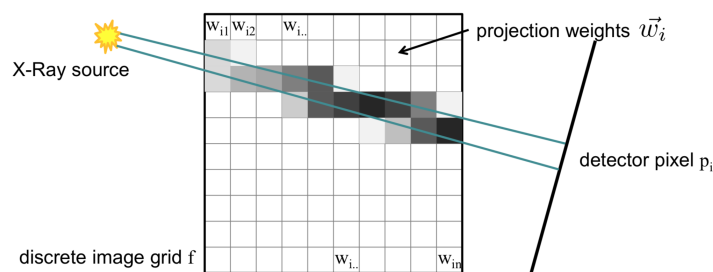
Figure 2.4.: Example of image discretisation for algebraic reconstruction. Figure taken from presentation of Wolfgang Wein in the lecture Computer aided medical procedures at TUM.
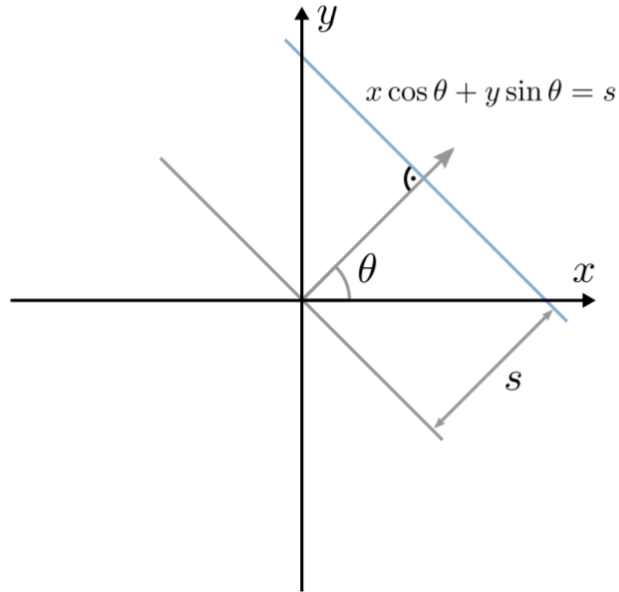
Figure 2.5.: Parametrization for CT image reconstruction. Reprinted from [112] with allowance of Springer ("Publisher") provided by Copyright Clearance Center ("CCC").

$$\sum_{j=i}^{N} w_{ij} f_j = p_i \quad \text{or} \quad W\vec{f} = \vec{p} \tag{2.3}$$

The Radon transform

In 1917, the Austrian mathematician Johann Radon proposed a solution to transform projections ($R : f(L)$) in one dimension to two dimensional functions (cf. equation 2.4).

$$R : f(L) = \int_{L} f(\vec{x}) \, d\vec{x} \tag{2.4}$$

The radon transform yields to density distribution reconstruction (f) from line integrals ($R : f(L)$) under the correct parametrization. Obviously equation 2.4 is similar to equation 2.1. In the CT case of the image a correct parametrization is: $(x, y) \rightarrow (s, \theta)$ (cf. figure 2.5). Schematically the CT image problem can be seen as in figure 2.6, where an inverse function is needed to recover initial density distributions from projections.
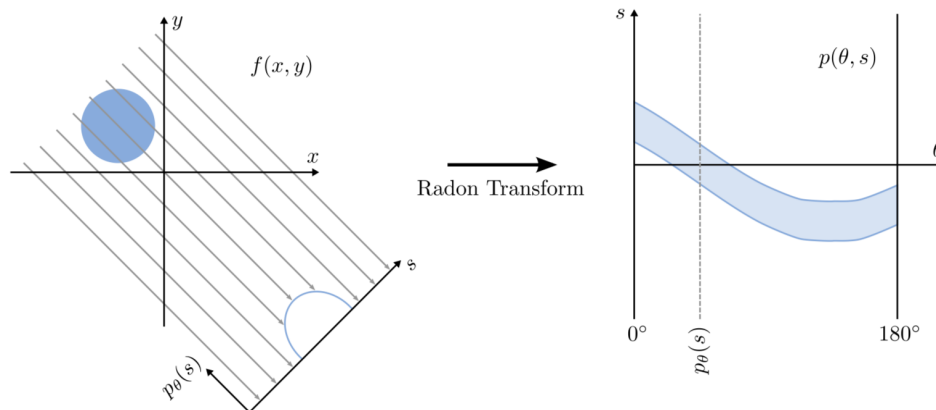
Figure 2.6.: In the left-hand side, the density distribution $\mu(x,y)$ is non-zero in the middle and zero elsewhere for one single rotation. In the right-hand side, the trace profile of density distribution of the objected after rotations can be observed, this is called Sinogram. Reprinted from [112] with allowance of Springer ("Publisher") provided by Copyright Clearance Center ("CCC").

The Fourier slice theorem

Finding a direct inverse function of the Radon transform is not an easy task, yet fortunately a relationship between the Radon transform and the Fourier transform exists, which reduce efforts by only looking for the inverse function of projections in the Fourier space. In a more formal definition the Fourier slice theorem establishes an equivalence that exists between $P_\theta(\omega)$ of the projection $p_\theta(s)$ and a line in the Fourier transform $F(u,v)$ of $f(x,y)$ which runs through the origin and form an angle θ with the u-axis [112]. Figure 2.7 provides a visual idea of the Fourier slice theorem. A proof of the theorem will be not provided in this thesis since it is out of the scope of this thesis. However, the reader can refer to chapter 10 in the book of Birkfellner [18] for the proof.

Direct Fourier inversion or Back-projection

As suggested by the Fourier slice theorem an estimation of $f(x,y)$ by computing all projections in the Sinogram $(S_\theta(\omega))$ to then compute the inverse, is possible (cf. equation 2.5). Nonetheless this approach is computationally disadvantageous, since a map from polar coordinates to Cartesian coordinates (I_θ) are required, which due to irregular sampling for the Fourier space leads to a significant reconstruction errors for high frequency components of the image in the spatial domain (cf. figure 2.8).

$$f(x,y) = F_2^{-1} \circ \int_\theta (I_\theta \circ F_1 \circ P_\theta)(\omega) \tag{2.5}$$
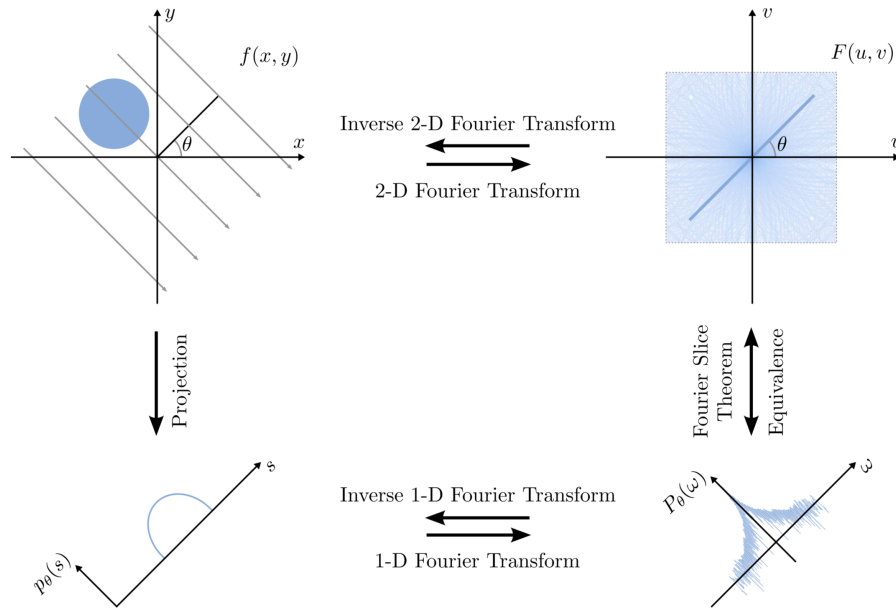
Figure 2.7.: Visualisation of the Fourier slice theorem. Reprinted from [112] with allowance of Springer ("Publisher") provided by Copyright Clearance Center ("CCC").
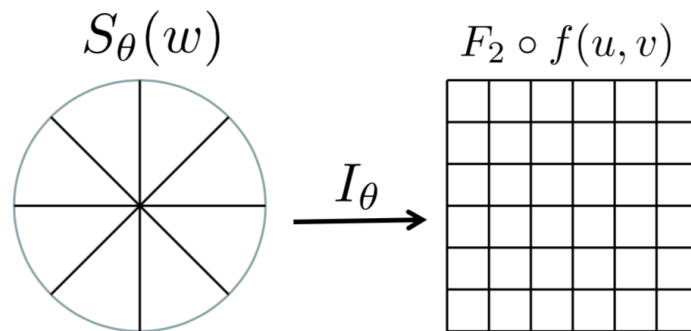
Figure 2.8.: Back-projection problem when no compensation for change in reference coordinates. Figure taken from presentation of Wolfgang Wein in the lecture Computer aided medical procedures at TUM.

Filtered back-projection

In order to compensate for polar sampling in Fourier domain, the inverse function is weighted by the spatial frequency (ω). This is called **Filtering**, since only the highest values of ω do not drop suddenly to zero (cf. equation 2.6).

$$f\left(x, y\right) = \int_{\theta} \left(\int_{\omega} S_{\theta}(\omega) \omega e^{2\pi i \omega t} d\omega \right) d\theta \quad \text{or} \quad f\left(x, y\right) = \int_{\theta} \hat{P}_{\theta}\left(t\right) d\theta \qquad (2.6)$$

The term $\hat{P}_{\theta}\left(t\right)$ is the filtering contribution and can be seen as the inverse Fourier transformation of the projections in the Fourier space ($\hat{P}_{\theta}\left(t\right) = F_1^{-1} \circ [(F_1 \circ P_{\theta}) \omega]$). The choice of a proper filtering term is one of the key problems to speed up reconstruction algorithms assuring good image quality. The filter term (ω) used so far is the so-called ramp filter. It only enhances high frequencies.

2.2. Magnetic Resonance Imaging

Since MRI was introduced in 1980, it has gained popularity as another anatomical imaging modality routinely used in hospitals (cf. figure 2.9). Clinically, MRI is advantageous for RT in comparison to CT because of its (to certain extend) harmlessness to patients. It does not require additional radiation and maintains an excellent biological image contrast, even better than CT for soft tissues in the human body. The physics principle of MRI lies in magnetic moment excitation of hydrogen nuclei, which are bound in the human body in water-based tissues or fat molecules, by a great magnet which provides variations over time of the external magnetisation field (also called B-field).The excitation of magnetic moments comes from the intrinsic quantum property of matter called spin. In absence of external magnetic fields spins are randomly distributed, here the macroscopic magnetisation of matter is zero (non-magnetic matter). On contrary when a (strong) external magnetic field is applied, spins tend to align in the same direction as the external magnetic field which yields to a macroscopic magnetisation of matter ($\vec{M}$). Spins are vectors in 3D, therefore in the presence of a external magnetic field, precession of the spin vectors of the nuclei happen in the orthogonal direction of the exerted B-field. This rate of precession (ω_0) is described by the Lamour frequency (cf. equation 2.7), for every nucleus the inertia to spin precession is different, this is accounted by the gyromagnetic ratio (γ). For the hydrogen nucleus, the gyromagnetic ratio is equal to 42.6 $\frac{MHz}{T}$. The angle of precession between the direction of the external magnetic field and the spin, decreases over time until total alignment and initial position restoration occurs when the B-field is switch-off. The spin precession emits radio frequency (RF) waves which can be measured. The times when the spin precession recovers equilibrium after the B-field is switch-off are called relaxation times, which is the main characteristic to obtain contrast from the different tissues.

Figure 2.9.: MRI scanner in clinical use. Reprinted from [112] with allowance of Springer ("Publisher") provided by Copyright Clearance Center ("CCC").

$$\omega_0 = \gamma \cdot B_0 \qquad (2.7)$$

2.2.1. Data acquisition

In an MRI scanner patients are positioned in the same configuration as in the CT scanner; The data is acquired by measuring the inductive signal generated with certain sequences when applied orthogonal magnetic fields are switched-off in sequence. The idea is to magnetise the patient in the direction of the magnetic field to then apply a second magnetisation field in the transverse direction so that $\mathbf{M}_{trans}$ is maximum and $\mathbf{M}_{long}$ equals to zero and later turn off the transverse magnetic field, relaxation times in the transverse plane (called T_2) are then measured because of the sensing RF waves after transverse magnetisation loss of $1/e \approx 37\%$, which are characteristic for every tissue type. Measuring longitudinal relaxation times are also possible from the same sequence, in this case there is a longitudinal magnetisation gain of the body, the time to recover the $\left(1 - \frac{1}{e}\right) \approx 63\%$ initial magnetisation is called T_1 and depends on the type of tissue (cf. figure 2.10). The density of spins (ρ, also called density of protons) can be directly measured by applying a direct Fourier transform (FT) of the signal received ($S(t)$) (cf. equation 2.8). Further augmentation of contrast can be achieved by weighting the relaxation times or spin density with different acquisition parameters (Time echo (TE) or Time repetition (TR)) or addition of contrast media (cf. figure 2.11) which for sake of document extension are not discussed in this document, yet the reader can refer to the book of Dhawan [37].
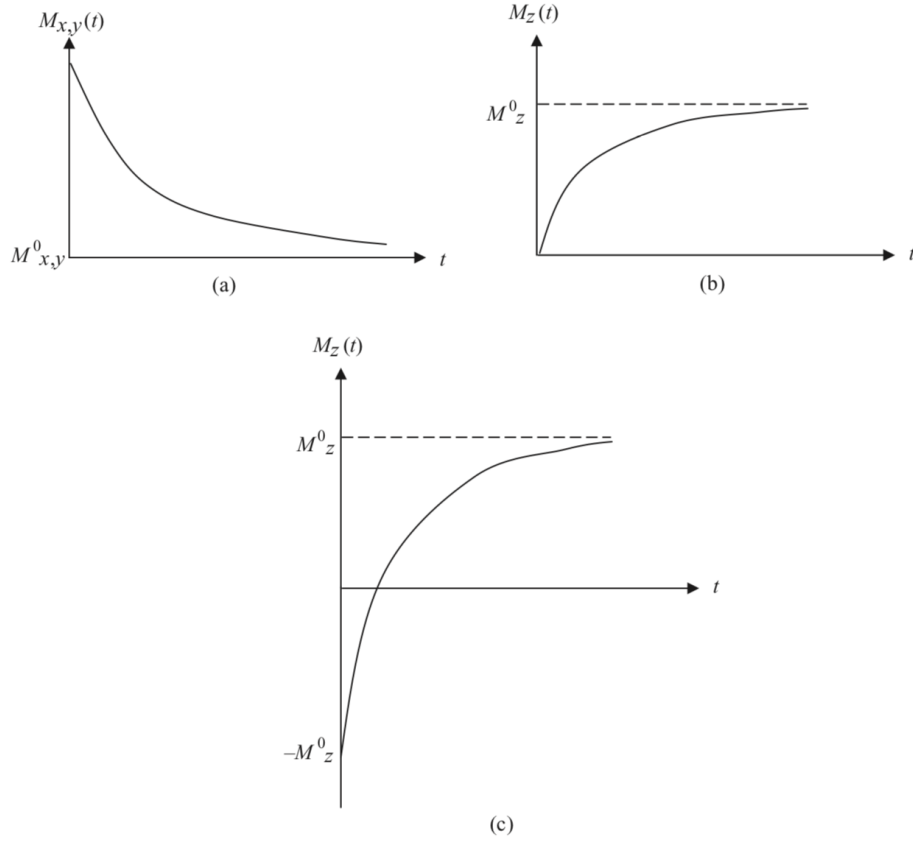
Figure 2.10.: Magnetizations loss after presence of external magnetic field in the transverse (xy) and longitudinal (z) direction. Reprint from [37] with allowance of John Wiley and Sons provided by John Wiley and Sons and Copyright Clearance Center.

$$S\left(t\right) = \int \vec{M}(\mathbf{r},t) d^3\mathbf{r} \quad \overset{\mathbf{FT}}{\longrightarrow} \quad \rho\left(\vec{r}\right) = \vec{M_0} \int S\left(\vec{\omega}\right) e^{i(\vec{\omega}\cdot\vec{r})} d^3 r \qquad (2.8)$$

2.2.2. Image reconstruction

To reconstruct images from the physics principle of MRI, spatial encoding parameters such as T_1-weighted, T_2-weighted and proton density (ρ) are required. For the sake of simplicity, the axes x, y, z are going to be used, z-axis as the in-slice direction (or the direction along the patient) and $x - y$ axes as the transverse section or the in-plane direction. To select the slice to be reconstructed, a linear external magnetisation field is applied in the z-axis ($\vec{B} = B_0 \cdot \vec{z}$). As the Lamour frequency depends on the applied external field (cf. equation 2.7), the spatial encoding of the external field means that the Lamour frequency is spatially encoded too (cf. figure 2.12). The remaining x and y directions can be spatially encoded by measuring the Lamour frequency of

when applying another linear gradient of external magnetic field in their respective directions (this approach is called frequency encoding). A variation of the method can be applied, instead of coding frequency for all of the $x - y$ directions, one of the directions could be spatially encoded by frequency, yet the other in phase (phase encoding), to do so, the frequency encoding in one of the directions (let us say in the x axis) is turned-off while the rate of change of the frequency encoding field with distance (let us assume in the y direction) changes in a defined step-wise, the number of encoded distances are proportional to the number of steps, for instance to produce a slice of 256×256 pixels, the slope of frequency encoding field should be changed 256 times. To recover the spatial distributions of the parameters measured, as in CT (cf. chapter 2.1.2), a reconstruction using filtered back-projection is performed for the frequency encoding signals and for the phase encoding signal, there is no need of filtered back-projection method.
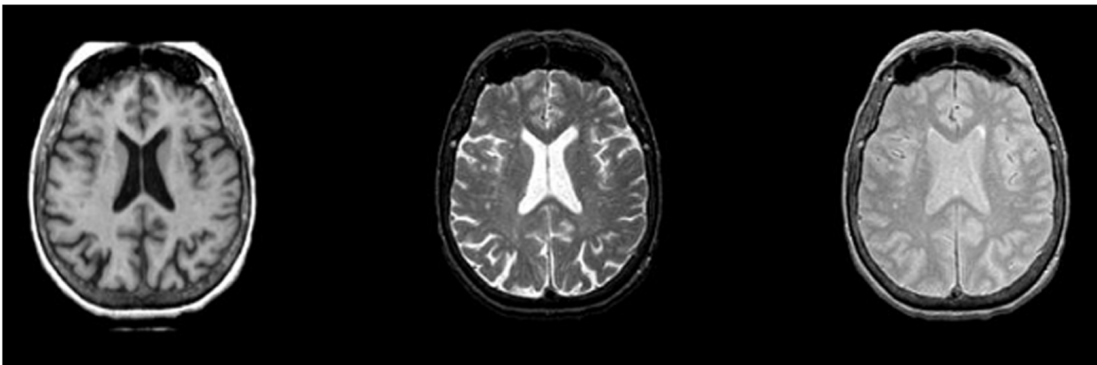
Figure 2.11.: MRI slice of a brain scan using parameters such as T_1-weighted, T_2-weighted and proton density (from left to right respectively) [37].

2.3. Positron Emission Tomography

As opposite to CT and MRI, PET is clinically considered the standard of functional imaging modality, this means that metabolic processes inside the body can be imaged directly using PET in addition to anatomical observations. It relies on the β^+-decay process of radioactive nuclei (Positron emission), which are chemically bound to molecules that agglomerate to metabolic processes that the examiner might be interested in. In PET scans for cancer imaging, the radioactive nucleus ^{18}F is routinely used due to its long half-life (cf. table 2.1), it is chemically bound to a chosen radiopharmaceutical (cf. table 2.2) such as tumour metabolism. For instance for fluorodeoxyglucose (FDG), it agglomerates in high presence of anaerobic glycolysis in tumours. Because of the high accumulation of the radiopharmaceutical molecules in determined metabolic process in tumours and the emission of positrons by the radioactive nuclei, imaging can be performed due to the frequent encounters of positrons
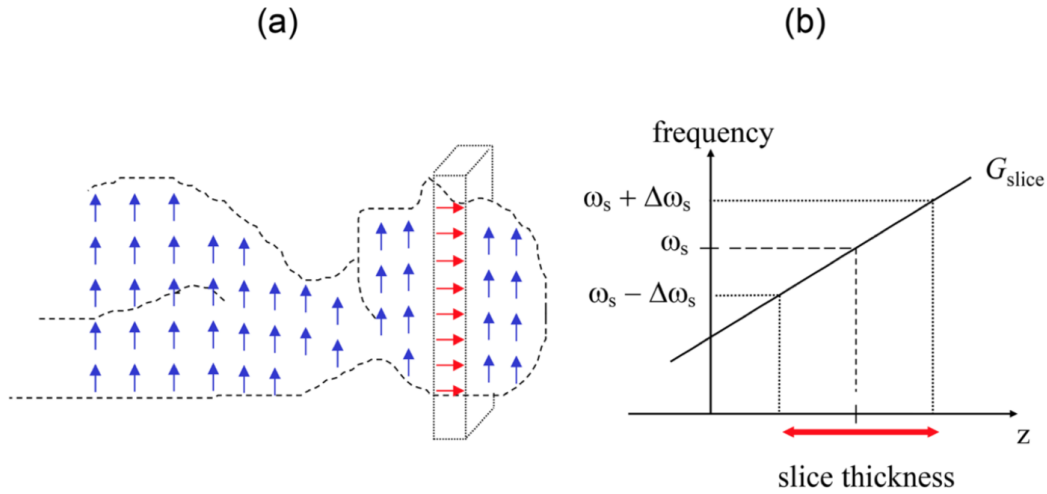
Figure 2.12.: (a)MRI slice selection visualization process. Reprint from [174].

Positron emitting radioactive nuclei	Two photon energy [KeV]	Half-life time [min]
Fluorine ^{18}F	511	109
Oxygen ^{15}O	511	2
Nitrogen ^{13}N	511	10
Carbon ^{11}C	511	20

Table 2.1.: Common positron emitting radioactive nuclei. Reprinted from [2] with allowance of John Wiley and Sons provided by John Wiley and Sons and Copyright Clearance Center.

with electrons which yield to pair productions of photons with energies of 511 keVs travelling in opposite directions (180 degrees). This effect is exploited for imaging purposes.

2.3.1. Data acquisition

To acquire data, the patient must be injected with the desired radiophamaceutical. The time of the scan is adapted to the weight of individuals and the time after injection. Detectors are placed surrounding the patient in a ring form, where coincident photons are then absorbed, resulting in a tracked signal (cf. figure 2.13). the path connecting the two coincident photons is called line of response (LOR). Accounting for all detector signals located a different angles yield to a sinogram, which is then exploited for imaging reconstruction.

Uptake mechanism	Tracer	Tumours imaged	Benefits	Limitations
Anaerobic glycolysis	^{18}F-FDG	NSCLC, Head and Neck tumours, Oral Squamous cell carcinoma, Gastric cancer	Good correlation with tumour aggressiveness and prognosis, Easily reproducible and broad availability	Overlap between uptake in normoxic and hypoxia tumour tissue
Nitroimidazole-like uptake: reduction into RN02 radicals and RNHOH compounds in hypoxic conditions. The covalent binding to macromolecules.	^{18}F-FMISO	Head and neck tumours, Locally advanced HNSCC, Gliobastoma multiform, Breast cancer, NSCLC, Renal cell carcinoma	Good availability. Broadest evidence of value as a hypoxia tracer. Good correlation with immunochemestry and prognosis in most of the cases	Lack of correlation in all tumours. Low tumour-to-background ratio. Variable reproducibility.
	^{18}F-FETNIM (^{18}F-fluoroerythronitroimidazole)	NSCLC. Esophageal cancer	Promising tracer with possible correlation with outcome. Slightly higher tumour-to-background ratio than ^{18}F-FMISO	Limited evidence compared to ^{18}F-FMISO.
	^{18}F-EF3 (^{18}F-2-nitroimidazol-trifluoropropyl acetamide)	Rats bearing syngeneic rhabdomyosarcoma tumours. Head and neck tumours.	Promising tracer	Very limited evidence, mostly preclinical.
	^{18}F-FETA (^{18}F-fluoroetanidazole)	Mice bearing MCF-7, RIF-1, EMT6, HT1080/26.6, and HT1080/1-3C xenografts	Promising tracer with better biodistribution than ^{18}F-FMISO	Preclinical evidence

Table 2.2.: Most used radiopharmaceuticals in PET imaging of tumour hypoxia [111].
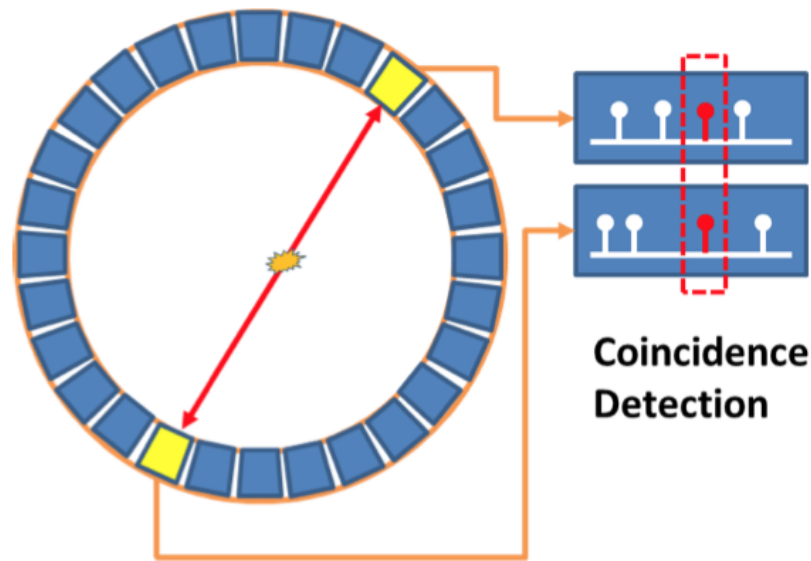
Figure 2.13.: Alignment of photon detectors for PET scanners and illustration of the coincidence detection principle. Reprinted from [147] under the Creative Commons Attribution 4.0 International License.

2.3.2. Image reconstruction

As in CT (cf. chapter 2.1.2), after obtaining the resulting sinogram, the process of image reconstruction can be performed by either using analytic or algebraic methods. Nonetheless, image quality of PET scans in comparison with MRI or CT are of lower quality, since only the accounted signals come from coincident photons. Modern systems can measure the time between the detection of the two photons and subsequently the determination of the position of the pair-production to some extent (time of flight PET, [18]), which increase contrast and resolution.

2.4. Imaging biomarkers

After the non-extensive summary of imaging modalities in medicine (cf. Chapter 2), it is clear that tumour anatomy and function can be tracked, and consequently, tumour changes. However, how to derive reliable measurements of the underlying biology or structure in the scanned tumour from imaging modalities (Imaging biomarkers) is not. The WHO and the new Response Evaluation Criteria in Solid Tumours (RECIST) categorise tumour response after treatment in four types based on tumour size (complete response, partial response, stable response and progressive disease [77, 157]), which information can be easily obtained using single or multiple imaging modalities. The reduction in tumour size is explained for the unavailability of tumours to replicate or proliferate due to the lack of required nutrients and oxygen. Notwith-

standing, reduction of tumour assessments are intra-observer dependent and therefore prone to errors, which can be reduced by semi-automatic segmentation of tumours. Besides, there are many studies which have found that despite tumour reduction in a few patients, overall survival outcomes remain low (under 10%) [83, 118, 159]. More complex imaging biomarkers can be derived which in this section they are going to be divided in static-image-derived and dynamic-image-derived biomarkers. For sake of document extension only the most heavily researched imaging biomarkers are going to be discussed.

2.4.1. Static-image-derived biomarkers

As PET scanners are able to image biological mechanisms of tumours, depending on the tracer, intensity uptakes from images are used to derive (semi) quantitative imaging biomarkers. For instance, by injection of a FDG tracer when PET scans, the assessment of the intensity uptake contrast of the static image of a patient, can be used to follow glucose-transport activity, and consequently the aggressiveness of the targeted tumour (since cancer cells require more consumption of glucose than a healthy tissue.). A more pronounced intensity-uptake contrast in the imaged region, indicates a higher glucose metabolism [84]. To quantify this effect, the **standardised uptake value (SUV)** is derived from the intensity-value of the image when uptake of the injected tracer divided by the average activity in the body of the scanned patient [4] (cf. figure 2.14), which value is seen as semi-quantitative. The challenges to use SUV lie on the perfect static image to derive calculations, duration of the scan, plasma clearance of FDG, reconstruction method, among others that make SUV difficult to replicate results in other studies [150]. Some investigations have used SUV_{mean} to only account for average intensity uptake inside the contoured tumour and others prefer the use of SUV_{max} (maximum intensity uptake value inside the tumour) to reduce intra-observer variability [78, 106]. Other researchers in their wish to make their methods reproducible, they use **Metabolic tumour volume (MTV)** which instead of using intensity-uptake values, it counts the number of voxels or pixels where predominant intensity-uptakes are, to then divide it by the total volume of the tumour. In order to assess volume, an anatomical imaging modality is then used in combination with PET (either MRI or CT) [28]. There are no agreements in research about the right approach to segment or extract MTV, the most commonly used approach is based on absolute or fixed percentage of SUV_{max}, and gradient or adaptive segmentation methods [128]. **Total lesion glycolysis (TLG)** has been introduced as novelty biomarker to profit from advantages of SUV and MTV. TLG multiplies SUV and MTV magnitudes, which theoretically it should extract the intensity of the activity glucose consumption as well as the metabolic activity and subsequently the total tumour damage [69].

If instead of quantifying glycolysis inside tumours, other tumour metabolism characteristics are wished to image, such as growth or hypoxia, the radiopharmaceutical-tracer must be changed. For example, as hypoxia tracer, ^{18}F-FMISO is commonly
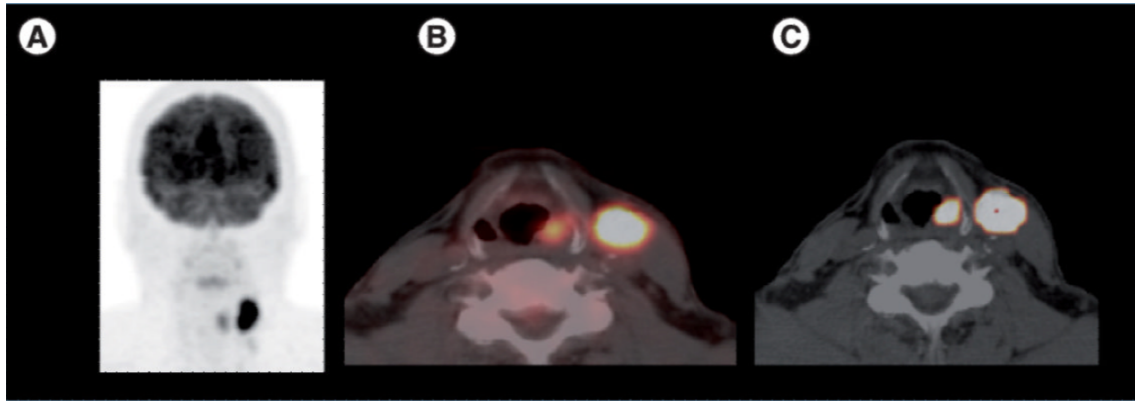
Figure 2.14.: (a) baseline ^{18}F-FDG PET/CT scan of a 53-years-old male patient, the highest contrasts represents the higher glucose consumption regions (b) co-registered axial PET/CT images where two focal intensity regions can be observed, the small one corresponds to a periform sinus squamous cell carcinoma ($SUV_{mean} = 7.3$ and $SUV_{max} = 9.1$), while the largest one corresponds to a ipsiateral nodal metastasis ($SUV_{mean} = 22.9$ and $SUV_{max} = 28.5$) (c) fused CT and volumetric PET images that shows small-size primary tumour (MTV = 1.6 ml) with large node MTV of 13.1 ml. Reprinted from [128].

used (cf. table 2.2), where similar (semi)-quantitative biomarkers can be derived as in ^{18}F-FDG PET, for instance, **TBR** which compares SUV inside tumours with SUV in healthy tissues [117], or **Hypoxic volume fraction (HVF)** which is similar to MTV. Some radiopharmaceutical-tracers can be used even to image molecular pathways, e.g. ^{64}Cu-DOTA-VEGF$_{121}$, which images VEGF (Tumour angiogenesis) cf. figure 2.15. All of the explained imaging biomarkers have been proved to have certain prognosis to treatment outcome in patients, but they still suffer from many challenges, for instance, they have been observed to change in magnitude during treatment, so many confounding studies have revealed different optimal time-to-event thresholds, in order to derive these markers from static images. No agreement has been reached so far either to use baseline static images or during treatment, and if during treatment, when to derive these imaging biomarkers [128]. Besides, establishing a threshold to classify the tumour accordingly is also conflicting, which yields difficulties to compare across studies [15].

2.4.2. Dynamic-image-derived biomarkers

Some tumour characteristics can be only evaluated on a dynamic-basis. meaning they are almost impossible to be described using only one single scan. For instance, assessing tumour vasculature requires that at least two images are taken (before and after contrast agent passage in tumour tissue), such that images are subtracted one another
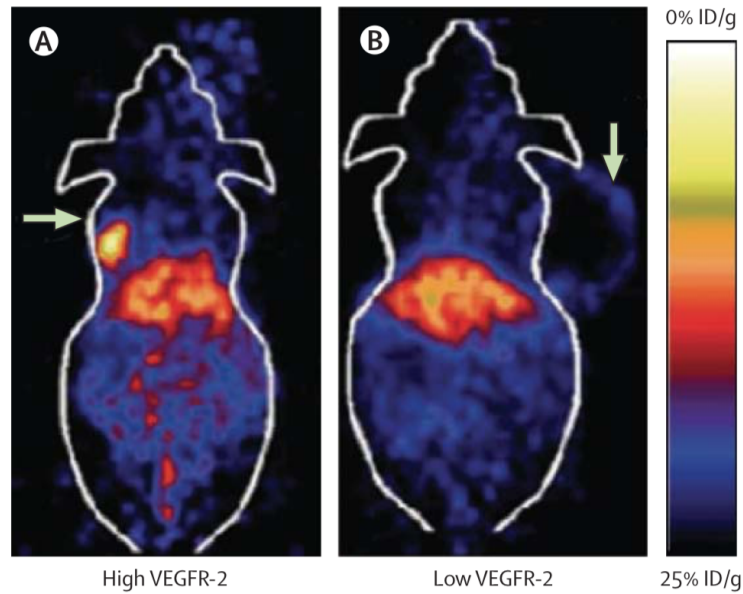
Figure 2.15.: Animal ^{64}Cu-DOTA-VEGF$_{121}$ PET scan, (a) small tumour (arrow) presents signs of higher VEGF than the larger one (arrow in (b)). ID/g represents the percentage of injected dose per gram of tissue. Reprinted from [127] with allowance of Elsevier Science & Technology Journals provided by Copyright Clearance Center ("CCC")..

resulting in the desired region of interest. The analysis for these types of imaging biomarkers are more challenging since they demand expertise of image specialists. A very simple imaging biomarker that can be computed from DCE-CT is the change of hounsfield units after post injection of the iodinated contrast agent in the tumour, since this correlates to tumour blood supply and the ability of tumour blood vessels to concentrate contrast agent (an indicative of tumour angiogenesis). To quantify difference in HU reduction the **tumour perfusion score** is proposed as the average difference of HU (pre and post injection of the contrast agent) during treatment [95] (cf. figure 2.16). Certainly, a similar imaging biomarker can be derived from DCE-MRI (ie, **Enhancing fraction**), however the correlation of the signal produced by relaxation time T_1-weighted with the contrast agent passage (usually gadolinium) is not linear, which makes quantification more complex than in CT [131]. Studies have shown correlations between these biomarkers and outcomes of patients during cancer therapy (cf. figure 2.17).

More complex modelling can be constructed from dynamic images to target biological process of tumour tissues, that being the case of pharmacokinetic modelling which has been used to track gadolinium passage inside tumours when administration to patients from DCE- or DW- (still in exploratory phase [40]) MRI. Brix, Toft, the two compartments exchange or tissue uptake models are frequently applied to search for

MRI-based biomarkers such as $\mathbf{K}^{trans}$, $\mathbf{v}_e$, **rate of contrast between plasma and Extravascular extracellular space (EES)** (k_{ep}), **volume plasma** (v_p), **blood flow** (F_b), **plasma flow** (F_p), **permeability-surface area product** (PS), **relative signal intensity at 80 seconds** (rSI_{80}), **Initial area under concentration-time curve at 60 seconds** ($IAUC_{60}$), among others. Here the treated patient must be scanned in a sequence of 1-5 seconds, usually over a period of 10 minutes, to detect changes in micro-vasculature of the studied tumour from the measured signal (T_1 or T_1-weighted, cf. figure 2.18). As expected, such quick imaging acquisitions come with drawbacks in spatial resolution, hence depending on needs, trade-offs must be evaluated. Derived parameters from dynamic DCE-MRI have shown better results in tracking tumour micro-vasculature abnormalities, which can even be used to detect hypoxic tumours [39, 80, 122], than static-image based DCE-MRI [74, 113] with better reproducibility [49, 49]. The preference of using DCE-MRI over DCE-CT for dynamics studies is supported by the absence of ionised radiation and superior soft tissue contrast of MRI in comparison to CT. Nonetheless, these parameters lack of reference values for interpretation, as well as many of them strongly correlate with the used imaging protocol or generation process which plays an important role at the moment of consideration for multi-centre phase II or III trials.

In a similar approach, dynamic PET parameters can be computed by tracking changes in intensity-uptake values in voxels in a sequence of scans (cf. figure 2.19) for different tracers. As tracers target different biological processes in tumours, derived parameters can explain broader tumour mechanisms than in dynamic DCE-MRI. Notwithstanding, lower spatial resolution and longer image reconstruction times than in DCE-MRI as well as on-site radiopharmaceutical production for imaging, make the modality less attractive to integrate into clinics. The sequence acquisition is about 10 seconds between frames in an interval up to 60 minutes after post injection of the desired tracer. Time activity curves of voxels in the regions of interest are then fitted to obtain quantitative imaging parameters as in dynamic DCE-MRI. Here classical kinetic modelling is more complex since it assumes co-existence of different compartments in the same volume (homogeneous oxygen concentration), which contradicts findings of concentration gradients in tissues [162]. Besides blood input functions are also challenging to obtain from time activity curves. This problem might be solved with help of simulation of biological mechanism for production of PET images [172].
Due to the many limitations, lack of standards and significance, up to date of work publication, none of the discussed imaging bio-markers have been accepted as a surrogate endpoints to treatment.
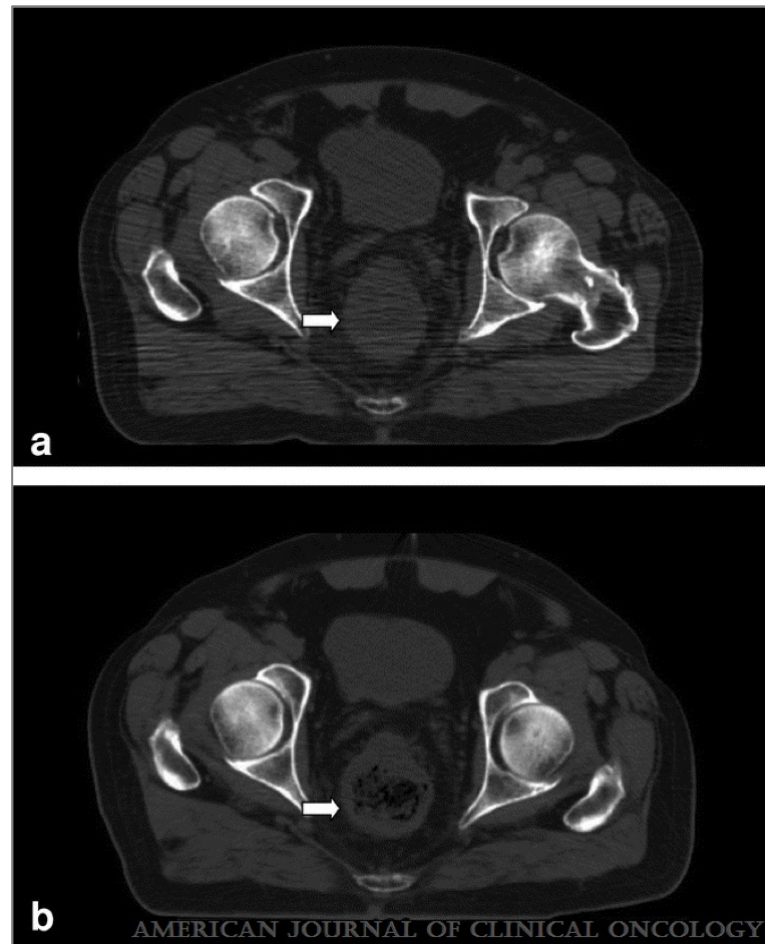
Figure 2.16.: Pelvis CT scan, (a) Average HU in tumour tissue before treatment (38.4 ± 5.9) and after contrast media passage (53.9 ±7.8, cf. white arrow). (b) Average HU in tumour after seven days treatment (36.9 ± 5.6) and contrast media post injection (48.3 ± 6.9, cf. white arrow) [95].
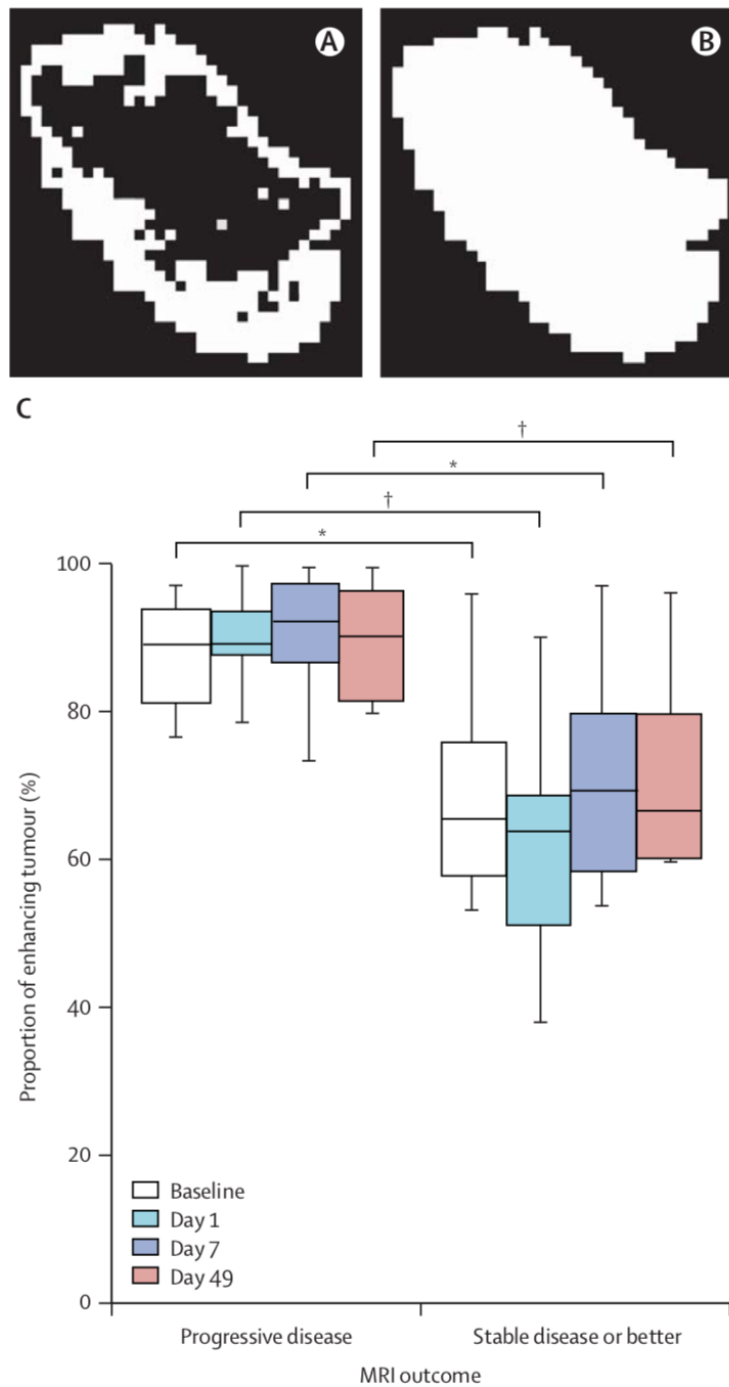
Figure 2.17.: (A) Enhanced tumour voxels (B) total of tumour voxels (C) discrimination between stable and progressive tumour disease according to percentage of enhanced fractions in different stages of therapy. Reprinted from [127] with allowance of Elsevier Science & Technology Journals provided by Copyright Clearance Center ("CCC")

1. Acquire dynamic images

time

T₂w images

Dynamic images

2. Define tumour ROI on high resolution T₂w images and transfer to dynamic series

3. Define AIF ROI on dynamic images

4. Obtain tumour T₁ from saturation-recovery images

5. Extract signal intensity curve for tumour and convert to concentration-time curve using measured T₁

6. Extract signal intensity curve for AIF and convert to concentration-time curve using known blood T₁

F_p, PS, v_e, v_p, K^{trans}

7. Fit model to concentration-time curve to give parameters

Figure 2.18.: Work-flow for analysis of phamacokinetics modelling and parameter acquisitions. AIF: Arterial input function, PS: permeability surface, F_p: plasma flow, v_p: plasma volume. Reprinted from [17] with allowance of Elsevier Science Technology Journals provided by Copyright Clearance Center ("CCC")
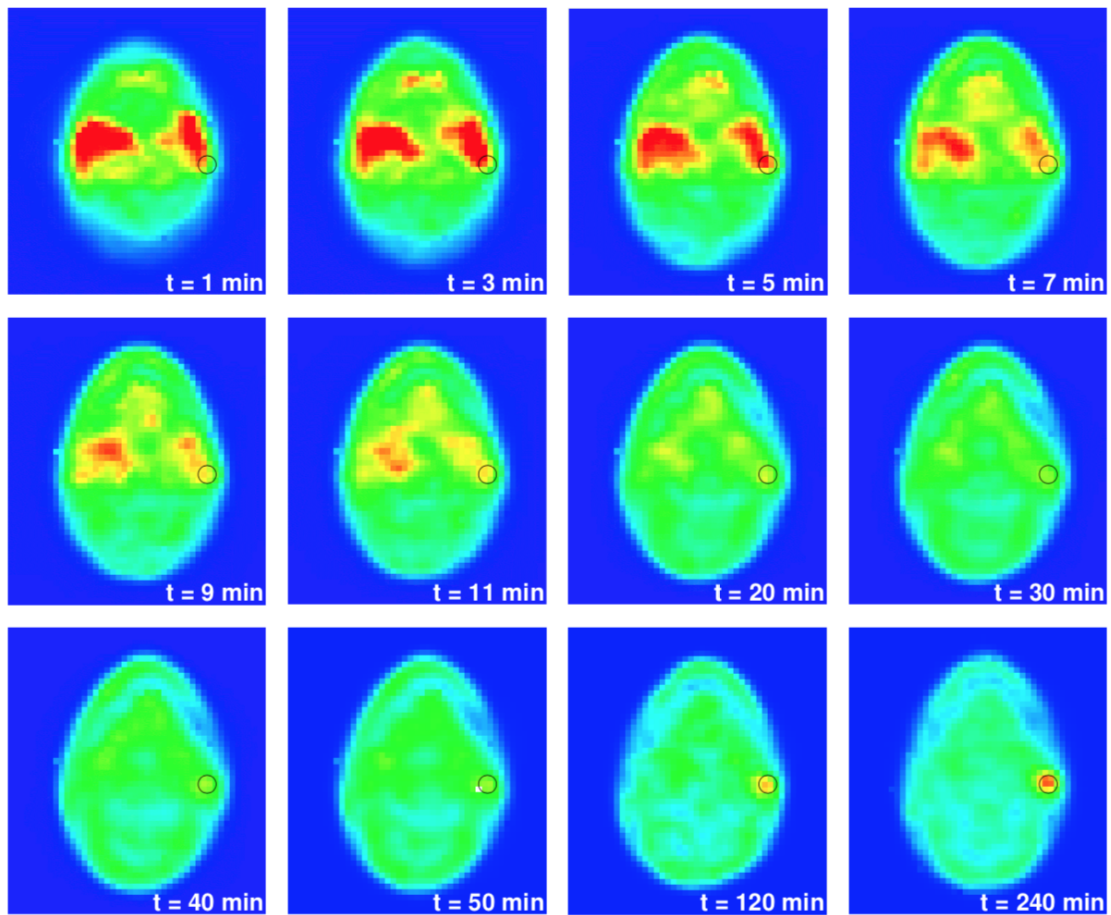
Figure 2.19.: Sequence of dynamic ^{18}F-FMISO PET scans for one slice in a HNSCC cancer patient to study changes in uptake in a voxel (black-border circle) in time. Reprinted from [160].

3. Radiomics

As discussed in chapter 2, very few quantitative markers can be obtained from static images with only RECIST as approved endpoints in treatment. In general terms, medical imaging modalities are used in a qualitative fashion in clinics (diagnosis, treatment planning, and follow-up and treatment guidance). On the other hand, despite the multiple possible parameters which can be obtained from dynamic imaging, clinical integration is cumbersome. This is due to the complexity of models, non-widely accessible medical imaging technologies for many health-care institutions, costs, as well as problems that come from imaging sciences, such as a lack of protocol standardisation. In 2012, a paper published by Lambin et al. [98] proposed a high-throughput number of mathematical parameters that could be obtained from static images of which many are not obvious to medical doctors. This parameters, also called *features* due to their origin in the computer vision field, require high order statistics and signal filtering analysis of images to be derived. They aim to capture the underlying biology of tumours from the tomographic data by assuming that the current state of the medical image modalities already have the power to quantify information about the pathophysiology of tumours [5,57,97]. In this chapter, a summary of the radiomic framework will be provided, as well as an explanation of the most frequently used features in the field.

3.1. The radiomic workflow

The novelty of radiomics is highlighted on the feature extraction process, since a high number of quantitative image parameters are proposed to characterise ROIs. Here a ROI is defined as a contoured or segmented proportion of the tomographic data that is aimed to be studied (e.g. primary tumours, metastatic lesions, lymph nodes or healthy tissue). Many factors need to be considered at the moment of computation of the quantitative image parameters, since variations in values from the same ROI can be obtained due to bias from segmentation algorithms, imaging protocols and software implementations (intensity discretization, imaging filtering, sampling, etc.) [42,68,181]. The reduction of these sources of variations requires full the disclosure of implementations and imaging protocols. The radiomic pipeline consists, broadly speaking, in four processes: image extraction and pre-processing, feature extraction, feature selection and, finally, model selection, assessment and validation (cf. figure 3.1).

After selection of a ROI, the resulting images are mined to obtain high numbers of hand-crafted features which, generally speaking, can be divided in four family types of features: intensity histogram, size and shape, texture and features after image filtering. All radiomic studies start with the selection of desired targets to predict, which can be either clinical endpoints, biological processes in tumours, or survival outcome. After the full radiomic analysis, these features are expected to provide reports on tumour phenotype and impact positively the tumour treatment in patients. Here, a common frame of imaging protocols is necessary to reduce confusing variability, which is well-known in the field [57, 68, 143, 144, 178] but, nevertheless, still common in radiomic studies.
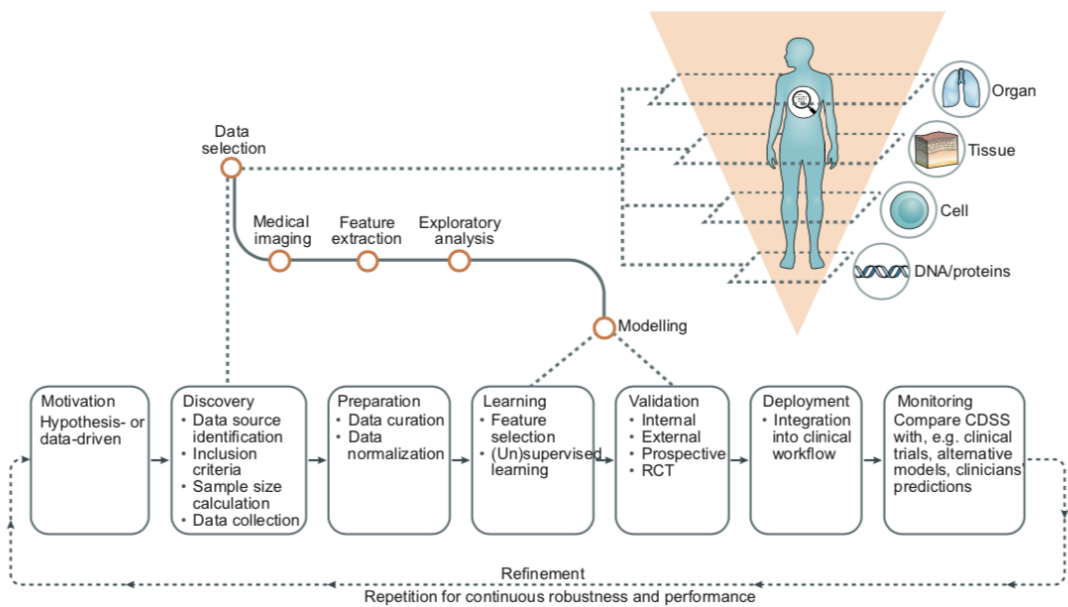
Figure 3.1.: Scheme of the radiomics work flow. Reprinted from [97] with allowance of Nature Publishing Group (Permissions) provided by Copyright Clearance Center ("CCC").

Here, the exploratory analysis comes from the fact that many of the radiomic features computed are highly correlated by either intrinsic correlations (mathematical definitions of parameters) or correlations of intra-sample feature variabilities (which do not add any relevant information). In both cases, redundant features can be reduced by clustering highly correlated features and either merge them in one single feature per cluster (for instance by using PCA) or choosing one of the features as representative inside the clusters.

In the modelling phase, one can identify the feature selection process and model application methodology itself. The former (probably the most important one [101]) is applied due to the high number of features that still remain after the exploratory

analysis. Theoretically, the number of hand-crafted features extracted can be unlimited if different filters are applied, but this risks model performance due to over-fitting since there will be more features than observations (patients). The latter depends on the resources available and preference of the study. For instance, if there are strong reasons to assume independence between features, logistic regression models might be preferred. Model selection affects the performance of features [132]. Therefore, applications of a variety of models are recommended.

Validation phase or holdout test phase is the final stage before reporting findings and can be set in an either internal or external cohort that has not been used in the modelling phase. The validation phase is used to confirm or dismiss findings from the learning phase. Similar performance metrics must be obtained between the learning phase and the validation phase, where the reader can observe the quality of the used methods. The external validation phase is always preferred because real, 'untouched' validation can be better guaranteed.

3.2. Image feature engineering

As discussed in section 3.1, the novelty of radiomics relies on the proposition of high-throughput features to characterise regions/volumes of interest from volumetric images. In this section a description of the most used imaging features used in radiomics will be imparted. These definitions follow the recommendations of the document of the international collaboration for standard definitions of image features in Radiomics which the reader can refer to the manual for more details and mathematical definitions [184].

3.2.1. Intensity-based features

Intensity-based features are assumed to describe the first order statistical distribution of values of the analysed volumetric-image. In this category, as volumetric-images are treated as samples of values, one can measure the centre of the distribution by means of, for instance, the sample **mean**, which is described as the average intensity value inside the ROI. In case of CT, it might provide information of the structure density value of tumours (cf. section 2.1). **median** is required by computation and is ordering the values of the imaging sample from low to high and chooses the value in the middle. In case that the sample value distribution is even, the average of the two values in the middle is taken. Measuring the spread of the distribution is also important in the description of first order statistics. Parameters such as **variance** refer to a level of average difference between the mean intensity value and other intensity values inside the ROI (**mean absolute value**) to the power of two. **Median absolute value** measures the average distance between every value in the image and the median. **Coefficient of variations** describes the ratio between the average distance of values

with respect to the mean and the mean itself of the grey level values of ROI, or, in other words, it measures dispersion. **Range** is defined to account for the distance between the minimum and the maximum grey level. Due to the imperfect delineations of ROI features such as maximum or minimum grey level are sensitive to outlier values, hence, the features such as $\mathbf{10^{th}}$ **percentile** and $\mathbf{90^{th}}$ **percentile** dismiss 90% of the highest values and compute the maximum in the subset. Likewise, $\mathbf{90^{th}}$ **percentile** dismiss 10% of the highest values and compute the maximum intensity values of the distribution inside the subset. **Interquartile range** (IQR) is similar to the range but without extreme values (usually values between 25% and 75% of the distribution). First order **energy** is defined as the square sum of grey level values in the ROI. **Quartile coefficient of dispersion** is similar to the coefficient of dispersion, yet more robust since it does not count on outliers (frequently from 25th to 75th percentile of the intensity values in the ROI). **Skewness** is defined as the degree of value accumulative tendency before or after the mean of grey level value distribution (cf. figure 3.2). The excess of **kurtosis** measures the peakedness or the distance between the most repeated intensity value and the rest in the distribution (cf. figure 3.3). **Maximum grey level** represents the maximum intensity value that can be found in the ROI. In this sense, **maximum grey level** can be also computed as the largest intensity value inside the ROI. Finally, **root mean square**, or also called the quadratic mean, is the square average of the intensity values in the sample.

In general terms, these features provide a broad view of the relationship between values and relative frequency inside the ROI.

3.2.2. Shape-and-size based features

Information from size and shape of the volumetric-image can also be computed by accounting the delineation masks. Here, two representations might be implemented: masks consist on a collection of voxels with certain volume or masks represent a surface mesh containing the volume of the ROI. Feature computations using the two approaches barely differ when masks are sufficiently large (more than 1000 voxels), yet overestimation occurs otherwise because of the lack of optimal considerations of the edges in ROI by the first approach. **Volume** is represented as the count of voxels belonging to the ROI. **Surface Area** is the area occupied by the external surface which contains the ROI. **Surface-to-volume ratio** as its name suggests this is the fraction between the surface area and the volume of the ROI, this feature aims to account differences not only in shape but also size. In this sense many other features might be derived and thus correlated, such as **compactness 1**, **compactness 2**, **spherical disproportion**, **sphericity**, **asphericity** which aim to measure how sphere-like is the ROI, derivatives can be obtained such as **maximum diameter** (distance between the two most distant vertices). One can change comparisons of simple geometric forms, for instance instead of enclosing the ROI in a sphere form, it could be an ellipsoid, then computations of features such as axis lengths, defined as the longest distances inside the enclosing ellipsoid in 3 directions, **major axis**
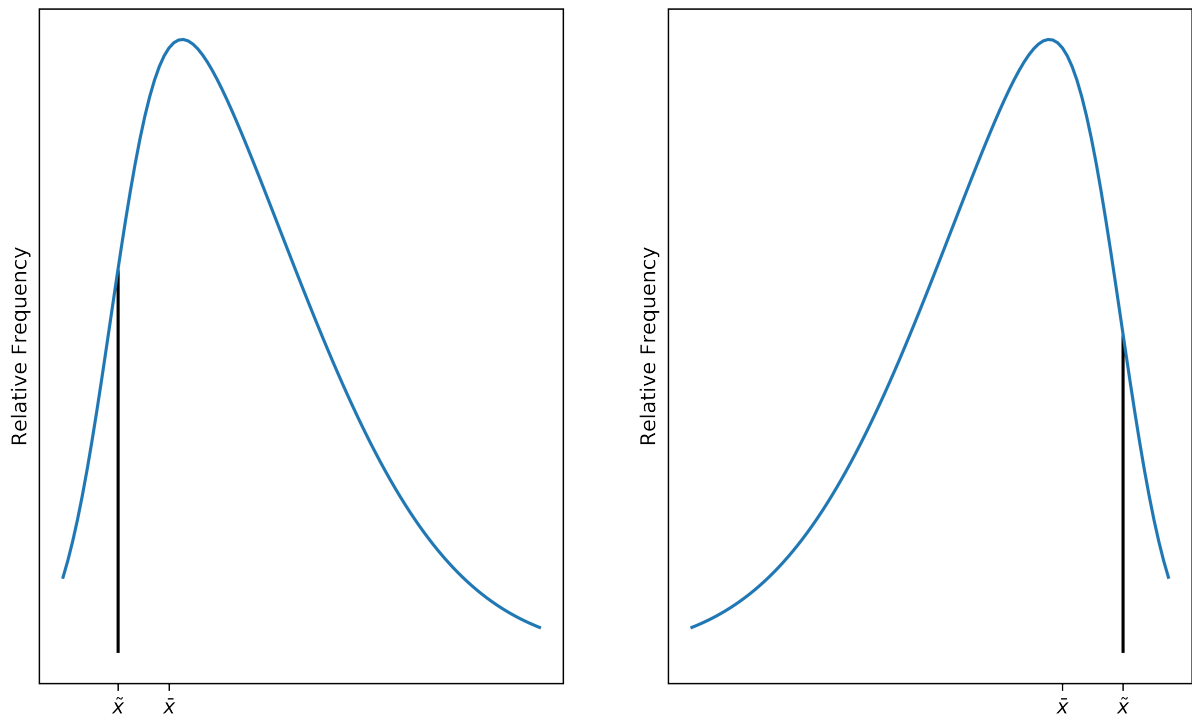
Figure 3.2.: Skewness visualisation: In the values distribution of the left-hand side skewness is positive (a higher frequency of values is accumulated before the mean ($\bar{x}$)), while in the right-hand side it is negative (a higher frequency of values is accumulated after the mean).
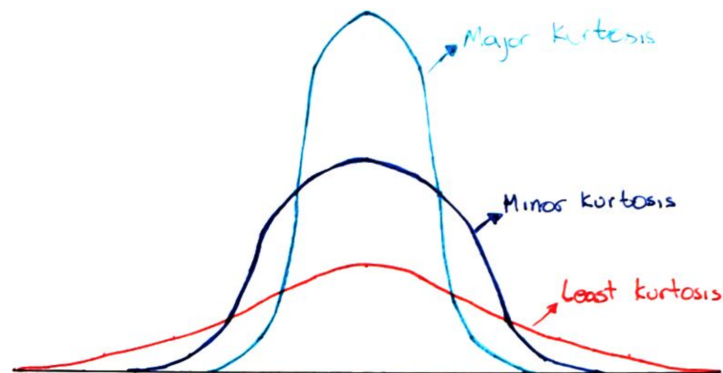
Figure 3.3.: Kurtosis visualisation example: Kurtosis measures the peakedness of a distribution of values.

length minor axis length least axis length, these parameters help to measure the **elongation** and **flatness** of the ROI, one can also compute fraction of the actual

volume of the ROI in comparison to the enclosing ellipsoid (**volume density in a enclosing ellipsoid**), in the same direction **area density in a enclosing ellipsoid**. Measurements of distributions of low or high intensity values inside the volume in the ROI can be captured with **centre of mass shift**.

3.2.3. Texture-based features

Texture features belong to the high order descriptive statistic analysis, since they account for associations of not only single grey level values in the distribution but their vicinity, these features are assumed to quantify volumetric image structures, in the medical field image heterogeneity. To count for the vicinity, auxiliary occurrence matrices speed-up computations of features. Let us consider the case of:

The grey level co-occurrence matrix (GLCM):

The GLCM quantifies repetitions of values across 13 unique directions in case of 3D in the ROI. If considering only the next neighbour: (1, 0, 0), (0,1,0), (0,0,1), (0,1,1), (0,1,-1), (1,0,1), (1,0,-1), (1,1,0), (1,-1,0), (1,1,1), (1,1,-1), (1,-1,1) and (1,-1-1), which yields to 26 connected neighbours. The construction of the GLCM consist of assuming a 3D matrix form, where indexes on every axis (i, j, k) are a sorted discretised value of the intensity distribution of the ROI, then for a specified direction, every combination index is filled by the number of repetitions of index combinations in the specified direction (cf. figure 3.4). A single matrix is obtained if averaging over all possible directions. Finally using this matrix we can compute features such as: **joint maximum probability** which account for the most co-occurrence value in the averaged GLCM [66], **joint average** which describes the value-weighted sum of joint probabilities [166], **joint variance** which computes the variance of the GLCM, **joint entropy** which describes the entropy of the GLCM, **angular second moment** which is also called uniformity or energy of the GLCM [29], **contrast** that describes values of the GLCM yet weighted by the square of differences of discretised values in the image [184], **Dissimilarity** which essentially is the same as contrast yet weights do not strongly bias the result since, they are obtained by only the difference of discretised values, **Inverse difference**, which as its name suggests, quantifies how homogeneous the volumetric image is, by reducing values of the GLCM with large differences [29], **normalised inverse difference** which normalises inverse difference by the number of unique values in the image [66], **correlation** which computes correlations inside the GLCM, or **auto-correlation** [156], **difference average** quantifies dissimilarities for the diagonal probability or in other words an auxiliary matrix derived GLCM that counts for differences between discretised values in the voxels [168], **difference variance** measures the variance for the diagonal probability [67], **difference entropy** computes the statistical entropy of the diagonal probability, **measure of information correlation** which describes the information gain of the GLCM with the values of the volumetric-image [67].

1	2	2	3
1	2	3	3
4	2	4	1
4	1	2	3

(a) Example of pixel values (discretised)

		j		
	0	3	0	0
i	0	1	3	1
	0	0	1	0
	2	1	0	0

(b) GLCM in (1, 0) direction

		j		
	0	0	0	2
i	3	1	0	1
	0	3	1	0
	0	1	0	0

(c) GLCM in (0, -1) direction

		j		
	0	2	0	1
i	2	2	1	2
	0	1	2	1
	1	2	1	0

(d) GLCM in (1, 1) direction

		j		
	0	2	1	1
i	2	2	2	1
	1	2	0	1
	1	1	1	0

(e) GLCM in (-1, 1) direction

		j		
	2	1	2	1
i	1	4	1	1
	2	1	2	1
	1	1	1	2

(f) GLCM in (-1, 1) direction
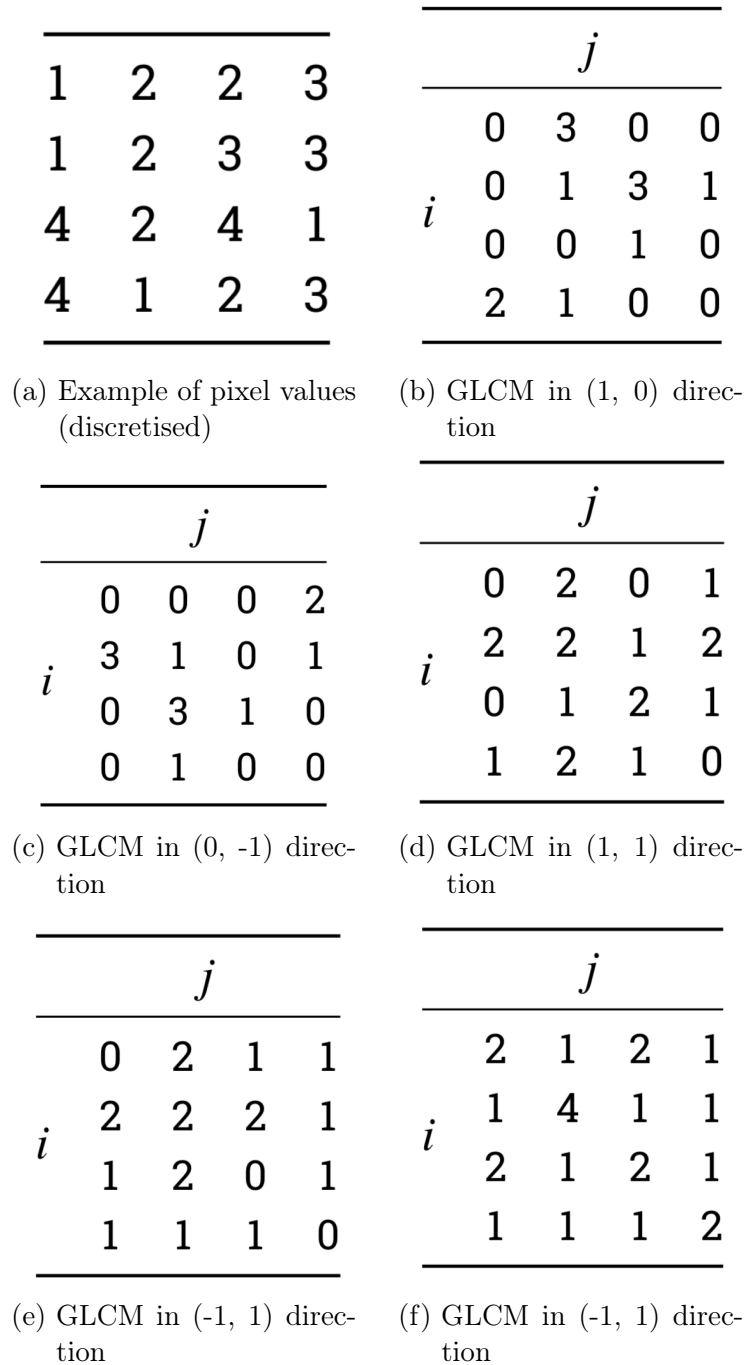
Figure 3.4.: Example of a GLCM construction using first neighbours in an 2D image, this analysis can be extended to the case of 3D. Reprinted from [184]

The grey level run length matrix (GLRL):

As GLCM only accounts for pairwise discretised neighbours voxel values, another auxiliary matrix was introduced by Galloway [50] to account for consecutive discretised

voxel values in different directions. A run length describes the length of repetitions of consecutive discretised voxels. An example of a GLRL construction is shown in figure 3.5. As in GLCM a merged matrix in all possible directions can be obtained to compute features such as **short runs emphasis** which its name suggests it sums up all values inside the GLRL where short runs contribute stronger to the sum, in imaging terms it accounts for micro subvolumes inside the volumetric image, **long runs emphasis** which as opposite to short runs lengths, it computes macro sub-volumes in the volumetric image, **low grey level run emphasis** which describes length of volumes accounted by low values in the image. In a similar direction yet emphasis is in high values in the image is **high grey level run emphasis**, **short run low grey level emphasis** which emphasises low values and micro sub-volumes in the image or in other words the first quadrant of the GLCM, the second, third and forth quadrant can be also emphasised by the **short run high grey level emphasis**, **long run low grey level emphasis** and **long run high grey level emphasis** respectively [34]. Distribution tendency of the GLCM can also be quantify by **grey level non-uniformity**, **run length non-uniformity**, **grey level variance**, **run length variance**, **run entropy** [6], **run percentage** quantify the proportion between the number of runs and the maximum possible runs, which means that for highly uniform volumetic images, run percentage is very low.

The grey level size zone matrix (GLSZ):

Regardless of directions or run-lengths, the GLSZ searches for connected voxels in the volumetric image. Connected voxels means that discretized voxel values are connected to construct a region if neighbour voxel values of the analysed voxel are the same regardless of direction or run-lenghts. In 3D the connectedness is evaluated in 26 neighbour voxels (cf. figure 3.6). From the GLSZ, **small zone emphasis** can be computed, as its name suggests it weights heavily small size-zones (regions). In the same direction **large zone emphasis** can be derived, here the relevance lies on the large volumetric substructures of the ROI. **Low grey level zone emphasis** looks for penalising high discretised voxel values and with the same analysis **High grey level zone emphasis** penalises low voxel values in the GLSZ matrix. The GLSZ matrix can also be divided in four quadrants, the upper-left, upper-right, bottom-left and bottom-right, where **small zone low grey level emphasis**, **large zone low grey level emphasis**, **small zone high grey level emphasis**, **large zone high grey level emphasis** accounts for the quantity of entries in these quadrants respectively. As in GLRL tendency distributions across size zones and grey levels can be accounted by **grey level non-uniformity**, **size zone non-uniformity**, **grey level variance**, **size zone variance**, **size zone entropy**. The fraction of realised zones divided by the largest number of potential zones can be also derived (**Zone percentage**), which as in GLRL the lower the number, the more homogeneous the ROI.

$$
\begin{array}{cccc}
1 & 2 & 2 & 3 \\
1 & 2 & 3 & 3 \\
4 & 2 & 4 & 1 \\
4 & 1 & 2 & 3
\end{array}
$$

(a) Example of pixel values (discretised)

		Run length j			
		1	2	3	4
	1	4	0	0	0
i	2	3	1	0	0
	3	2	1	0	0
	4	3	0	0	0

(b) GLRL in (0, 1) direction

		Run length j			
		1	2	3	4
	1	4	0	0	0
i	2	3	1	0	0
	3	2	1	0	0
	4	3	0	0	0

(c) GLRL in (1, 1) direction

		Run length j			
		1	2	3	4
	1	4	0	0	0
i	2	3	1	0	0
	3	4	0	0	0
	4	3	0	0	0

(d) GLRL in (-1, 1) direction

		Run length j			
		1	2	3	4
	1	2	1	0	0
i	2	2	0	1	0
	3	2	1	0	0
	4	1	1	0	0

(e) GLRL in (1, 0) direction

Figure 3.5.: Example of a GLRL construction in an 2D array of values, this analysis can be extended to the case of 3D. Reprinted from [184]

The grey level distance zone matrix (GLDZ):

Here, the size-zone, voxel value and location relation is targeted, thus two matrices are actually computed, the first one as in GLSZ and the second one maps the minimum number of voxels where the edge of the ROI is reached (cf. figure 3.7b) and from there the GLDZ computation can be performed (cf. figure 3.7) [158]. This accounts

						Zone size j				
1	2	2	3			1	2	3	4	5
1	2	3	3		1	2	1	0	0	0
4	2	4	1	i	2	0	0	0	0	1
4	1	2	3		3	1	0	1	0	0
					4	1	1	0	0	0

(a) Example of pixel values (discretised)

(b) GLSZ in 8-connected neighbour voxels for the 2D case.

Figure 3.6.: Example of a GLSZ construction in an 2D array of values, this analysis can be extended to the case of 3D Reprinted from [184].

for the location of the volumetric substructures in the image. From the GLDZ matrix features which measure different parts of the matrix can be obtained such as **small distance emphasis**, which heavily weights low distances (low j-th indexes) in the GLDZ, in less abstract terms it emphasises volume substructures that are located closer to the edge of the ROI, **large distance emphasis** emphasises for volume structures that are at the core of the ROI, in an analogue way, **low grey level**- and **high grey level zone emphasis** can be derived from the GLDZ to emphasis in low and high voxel values. As in the GLSZ, the GLDZ can be constructed to emphasise different regions (**small distance low grey level**, **small distance high grey level**, **large distance low grey level** and **large distance high grey level** emphasis). Tendency distributions in the GLDZ are accounted by **grey level non-uniformity**, **zone distance non-uniformity**, **grey level variance**, **zone distance variance**, **zone distance entropy** and finally the fraction between the realised zones and the maximum number of potential zones (**Zone percentage**).

The neighbourhood grey tone difference (NGTD):

The NGTD accounts for occurrence distributions inside a defined neighbourhood, voxel-value probability and sum of the magnitude of differences between voxel-values and the average voxel value within the neighbourhood [7]. An example of a NGTD construction can be observed for the case of 2D in figure 3.8, for detailed computations the reader might refer to the manual of the IBSI collaboration [184]. Here features such as **coarseness** provides an indication of the range of change across the ROI [7], as well as **contrast** yet here voxel values are also weighted. Penalisation of small changes in voxel values in neighbouring voxels are captured by **busyness**. **Complexity** measures rapid changes of voxel values weighted by large differences of them. Finally, **strength** provides a measure of slow changes of voxel values weighted

1	2	2	3
1	2	3	3
4	2	4	1
4	1	2	3

(a) Example of pixel values (discretised)

1	1	1	1
1	2	2	1
1	2	2	1
1	1	1	1

(b) Distance map of the ROI to edge for the 2D case.

	j	
	1	2
i		
1	3	0
2	2	0
3	2	0
4	1	1

(c) GLDZ for the 2D case.
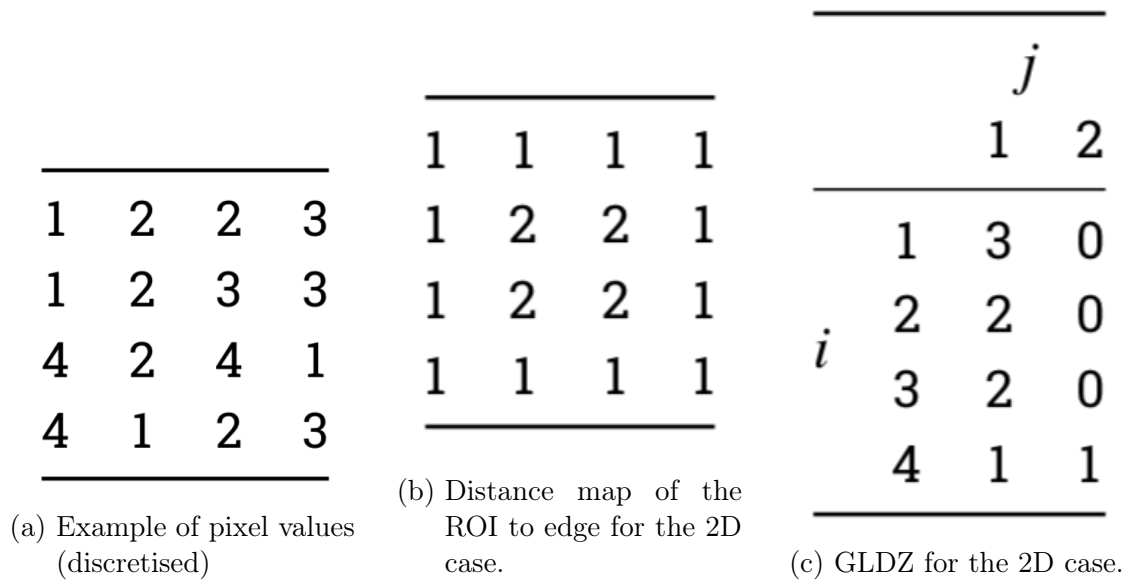
Figure 3.7.: Example of a GLDZ construction in an 2D array of values, this analysis can be extended to the case of 3D. Reprinted from [184]

by large differences in voxel values.

1	2	2	3
1	2	3	3
4	2	4	1
4	1	2	3

(a) Example of pixel values (discretized)

	n_i	p_i	s_i
i			
1	0	0.00	0.000
2	2	0.50	1.000
3	1	0.25	0.625
4	1	0.25	1.825

(b) NGTD in 1 pixel distance and defined neighbourhood for the 2D case.
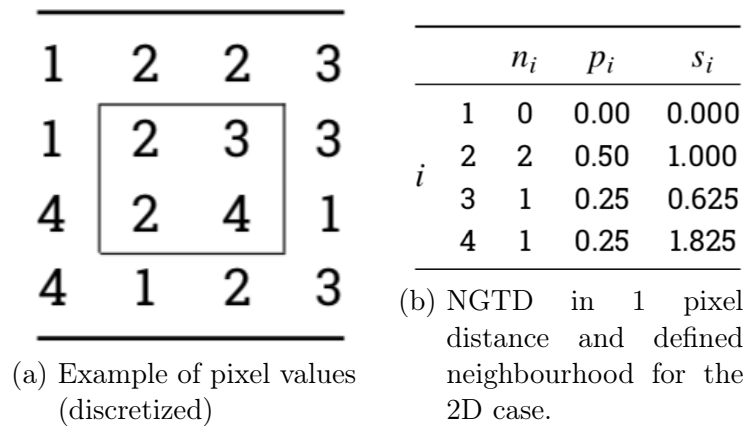
Figure 3.8.: Example of a NGTD construction in an 2D array of values, this analysis can be extended to the case of 3D. Reprinted from [184].

3.2.4. Filter-based features

Filter-based features are derived when enhancing images either in spatial or frequency domain. Many filters (such as Laplacian filter, Gabor filter, Wavelets filters) can be applied to images depending on the pattern of the image to be enhanced and the domain. For instance in the spatial domain, the Laplacian filter is very useful to highlight borders within the image (cf. figure 3.9). Here, after the application of the laplacian filter mask to the image, a new image, where only the borders can be seen, is generated or in other words the image displays regions where strong changes of pixel/voxel values appeared. The idea is to analyse or quantify this images using the features that have been already explained in previous sections.

As all images can be represented in the frequency domain with no loss of information (via Fourier transform), it is natural to think about pattern variations enhancement in this domain. Although it is difficult to make associations between the image in the spatial domain with its frequency domain, some statements can be deduced from frequency domain representations of the image. For instance frequency values correspond to changes to pattern variations of the image in the spatial domain (cf. figure 3.10). For the complete explanation of image filtering, the reader might refer to the book of Gonzales et al [59].

In the Fourier representation of images, enhancement of low changes of pattern variations can be acquired by applying a threshold function or mask, this is called low pass filter. The opposite can also be done (highlighting of extreme changes of pattern variations) which is called high pass filter. Due to the dimensionality of radiological images, continuous separations of pattern variations are applied with high and low pass filters in frequency domain image (cf. figure 3.11). Here quantification of the yielded images can be obtained by features from intensity and texture categories of radiomic features.
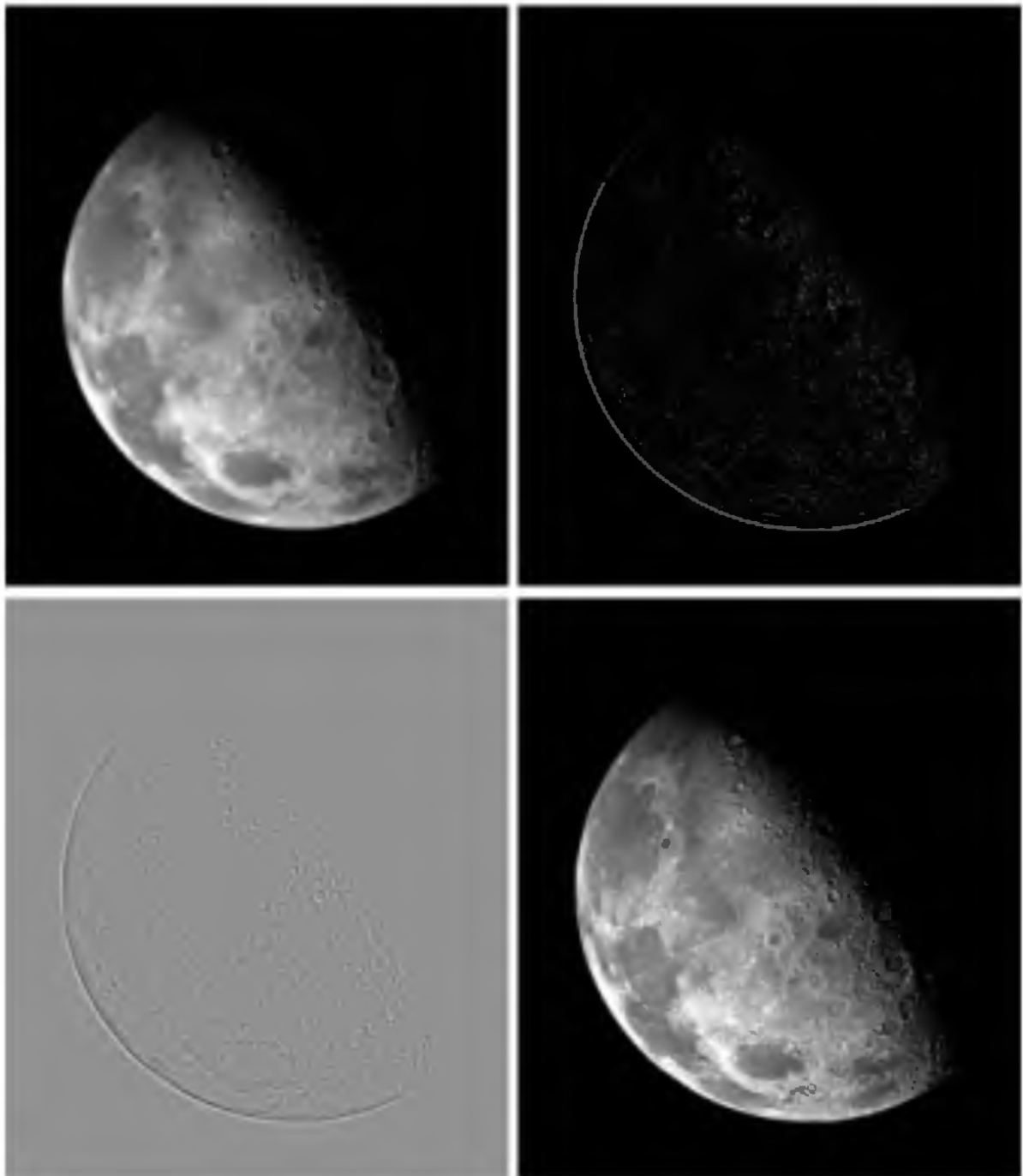
Figure 3.9.: Upper-left figure is an original image from the north pole of the moon, the upper-right is the image obtained after applying a laplacian filter of the north pole of the moon, the bottom-left image is the image of the north pole of the moon after applying the laplacian filter but shifted in order to provide better visualisation and the bottom-right image is the original image with addition of the laplacian filter image, such that borders can be highlighted. Reprinted from [59] with allowance of The Institute of Electrical and Electronics Engineers, Incorporated (IEEE) provided by Copyright Clearance Center ("CCC").
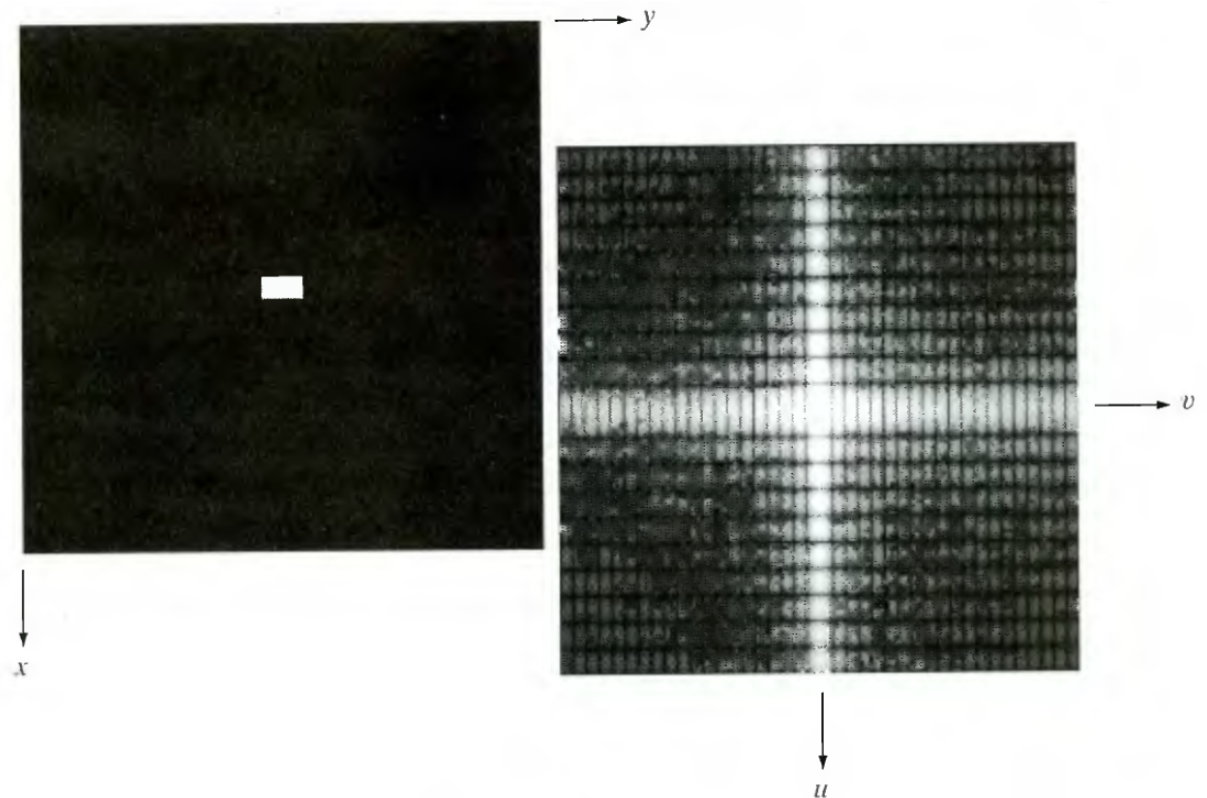
Figure 3.10.: Image and Fourier domain representation of the image, x and y are the coordinates of pixel in spatial domain while u and v in the frequency domain. Reprinted from [59] with allowance of The Institute of Electrical and Electronics Engineers, Incorporated (IEEE) provided by Copyright Clearance Center ("CCC").
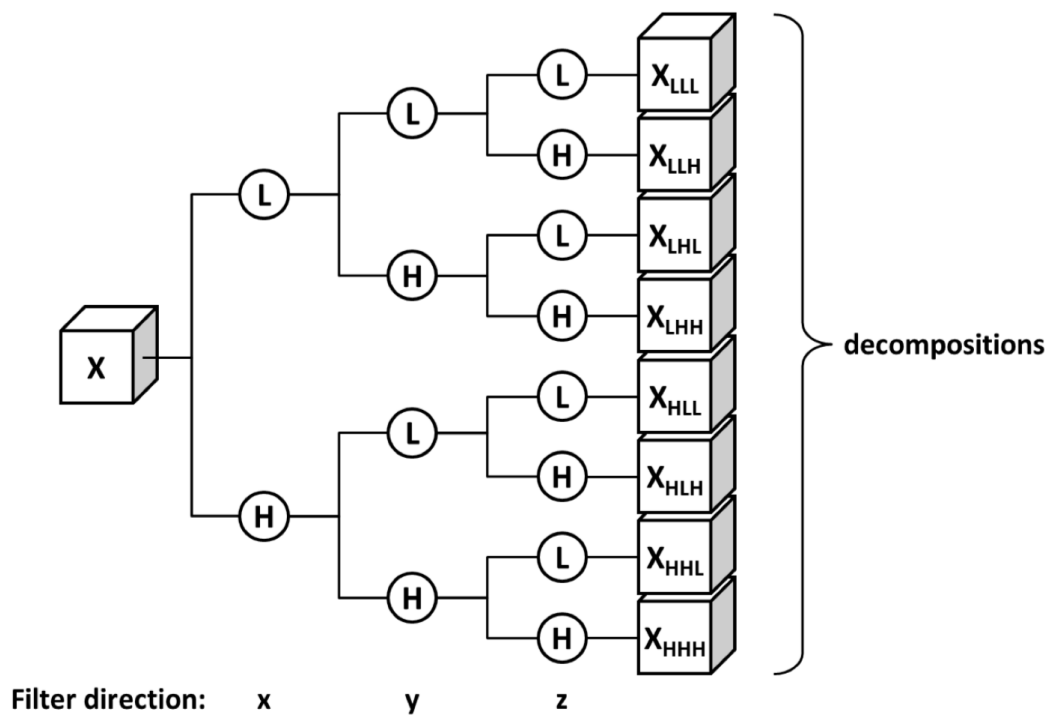
Figure 3.11.: Image and decomposition of the image in low and high filters subsequently. Reprinted from [5] supplementary material with allowance of Springer Nature provided by CCC.

4. Machine learning in the context of radiomics

The number of possible image biomarkers that can be obtained from radiomics are theoretically infinite, problems appear when this number (feature space) is so large that the data available becomes sparse to find statistical significance (the so called curse of dimensionality). Fortunately, in modern days that is not only an issue of radiomics but in many other fields. Mathematicians, statisticians, physicists and computer scientists have developed algorithms and methods to deal with large amounts of data and obtain profits out of them since the eighties, this field is called Machine Learning (ML). In this chapter an overview of the most common machine learning methods will be provided, a complete reference of machine learning is out of the scope of this thesis, nonetheless the reader might refer to the book of Mohri et al [116].

The general definition of ML is, as its name suggests, to teach machines (computers) to learn from experienced data to foresee outcomes of new input data. A machine learning pipeline consists of data mining, which aims to extract data out of the problem to be solved, data cleaning, which means to process and prepare the data in a format that can be read by a computer or a programming language, feature selection or reduction, which intends to select or reduce features to only the relevant ones for the aimed task, mathematical modelling, which selects the best model and tune it for prediction tasks and finally validation and/or holdout test, which aims to confirm predictive power of the chosen mathematical modelling (cf. figure 4.1). Usually in radiomics the learning data (also called experienced data) is labelled which means that the outcome target (such as clinical status, stage, recurrence, among others) is known for every sample. Therefore the pipeline that will be explained refers to the supervised machine learning scheme. A compelling description of the different ML stages will be provided in the following sections.

4.1. Feature exploration

One of the problems in radiomics is that many of the proposed features are correlated by either definition or variability, avoiding these features improves results since they do not add new information and lead to over-fitting. Implementations of strategies to deal with correlated features are needed such as PCA, kernel PCA and clustering. The first step is to measure correlations of features, which can be captured by Pearson
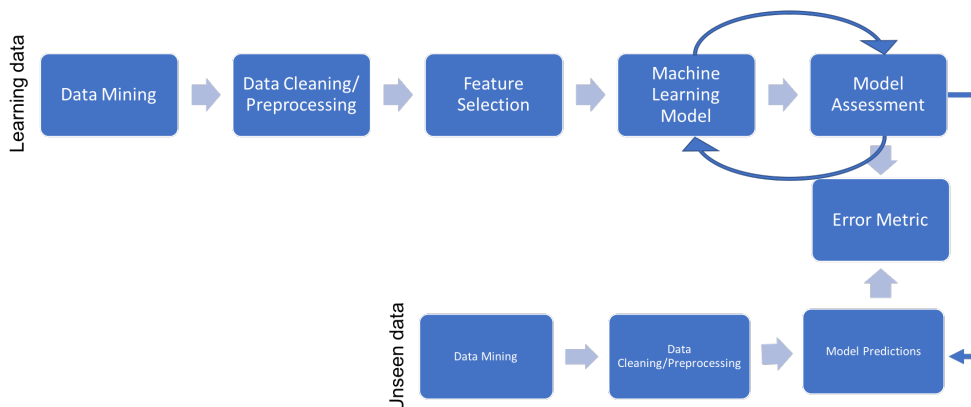
Figure 4.1.: Supervised machine learning pipeline.

product-moment correlation score (cf. equation 4.1). In short the Pearson score (r) is a pair-wise measurement of linear dependence between features (x and y). The range of the Pearson score is between -1, when features are perfectly negative correlated, and 1, when features are perfectly correlated. The value 0 means no linear correlation.

$$r = \frac{\sum_{i=1}^{n} \left[(x^i - \mu_x)(y^i - \mu_y) \right]}{\sqrt{\sum_{i=1}^{n} (x^i - \mu_x)^2} \sqrt{\sum_{i=1}^{n} (y^i - \mu_y)^2}} \tag{4.1}$$

An assessment for pair-wise monotonic correlation between features can be captured by the Spearman correlation score (r_s, cf. equation 4.2), which tends to be a generalisation of the Pearson score [1]. Here a ranking (d) is needed to be constructed across the number of entries (n) from features (Usually from higher to lower values). There are many approaches to account for pairwise correlations, however as the topic is out of the scope of this thesis, the reader might refer to the book of Dodge [38].

$$r_s = 1 - \frac{6 \sum d_i}{n(n^2 - 1)} \tag{4.2}$$

4.2. Feature reduction algorithms

After assessing pairwise correlations of features, clusters of highly correlated features can be constructed. The scientist can choose how to proceed with highly correlated features, one of the most intuitive practices is to choose only one feature inside the cluster to represent the information captured in it. Notwithstanding, as clustering features implies an arbitrary correlation threshold to group features, choosing one feature as representative of the cluster might yield into information loss. Alternatively, other strategies for feature reduction such as PCA are preferred since it constructs new features inside clusters with maximum variance (minimum information loss) inside clusters as representatives, here the trade-off comes from the loss of feature meaning,

which it can be reduced by choosing a high correlation threshold (usually more than 0.90 and less than -0.90). Feature reduction means to project the original feature space onto a new feature space and compress data that keep most of the relevant information.

4.2.1. PCA

PCA is an unsupervised linear transformation method which aims to find directions of maximum variance inside the original feature space and project them onto a new feature space with lower or equal dimensions than the original feature space (cf. figure 4.2) [138]. Let us suppose that the original space is composed from a d-dimensional feature vector $\mathbf{x} = [x_1, x_2, ..., x_d]$ with $\mathbf{x} \in \mathbb{R}^d$, a transformation matrix $\mathbf{W}$ is needed such that the projections $\mathbf{z} = [z_1, z_2, ..., z_k] = \mathbf{x}W$, with $\mathbf{W} \in \mathbb{R}^{d \times k}$ and $\mathbf{z} \in \mathbb{R}^k$ (commonly $k \ll d$). As the reader might wonder, features with different scales alter the variance direction, therefore scaling is required when PCA is applied. Here, the idea is to find $\mathbf{W}$ by means of linear algebra methods of calculating eigenvalues and vectors, where the largest eigenvalues represents the importance of the feature onto the composition of the eigenvector.
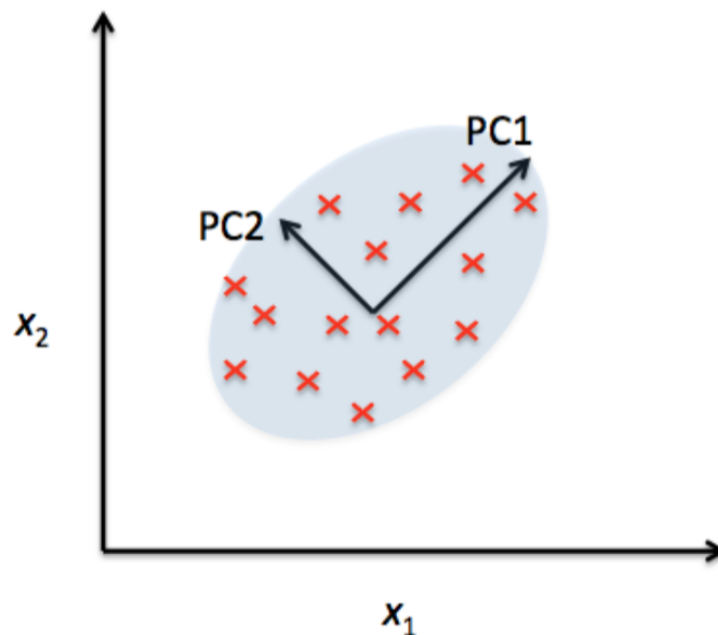
Figure 4.2.: Let us assume that x_1 and x_2 are original features, $PC1$ and $PC2$ are orthogonal projections onto the original feature space, $PC1$ captures the highest variance direction in the original feature space. Reprinted from [138] with allowance of Packt Publishing Limited provided by Copyright Clearance Center ("CCC").

4.2.2. Kernel PCA

When correlations between features are not linear, PCA fails to project the feature space into high-variance-lower-dimension feature space. As an alternative, the use of a kernel to increase dimensionality of feature space into a new feature space where linear correlations are feasible and then applied PCA is the core of kernel PCA. In mathematical terms this means to find $\mathbf{z} = \phi(\mathbf{x})$, where $\phi : \mathbb{R}^d \mapsto \mathbb{R}^k$ and $k \gg d$. Here the selection of the function ϕ derives the so-called kernel, in literature the following kernels are proposed:

The polynomial kernel:

$$k\left(\mathbf{x}^i, \mathbf{x}^j\right) = \phi\left(\mathbf{x}^i\right)^T \phi\left(\mathbf{x}^j\right) = \left(\left(\mathbf{x}^i\right)^T \left(\mathbf{x}^j\right) + \theta\right)^p \tag{4.3}$$

Where θ is an arbitrary constant and p is the order of the polynomial.

The hyperbolic tangent kernel:

$$k\left(\mathbf{x}^i, \mathbf{x}^j\right) = \tanh\left(\eta\left(\mathbf{x}^i\right)^T \left(\mathbf{x}^j\right) + \theta\right) \tag{4.4}$$

Here η and *theta* are arbitrary constants.

The Radial Basis Function (RBF) kernel:

$$k\left(\mathbf{x}^i, \mathbf{x}^j\right) = \exp\left(-\frac{\|\mathbf{x}^i - \mathbf{x}^j\|^2}{2\sigma^2}\right) \tag{4.5}$$

After choosing the right kernel for the feature space problem, the usual PCA strategy can be implemented. For the complete development of the strategy the reader might refer to the publication of B. Scholkopf et al [151].

4.3. Feature selection algorithms

In the context of radiomics, the course of dimensionality is a latent problem, due to the amount of inputs (features) extracted from images. Fortunately not all of them are relevant to the outcome prediction it is targeted, however the choice of the right subset of inputs to maximise the performance of the selected model plays an even more important role in radiomics than model selection [101]. Broadly speaking feature selection algorithms can be classified in three categories, filter-based methods, which is independent of the classifier, wrapper and embedded methods, which are classifier dependent. In **wrapper methods** the utility of the candidate subset of features is assessed by the error metric of a particular classifier in the training/test (sometimes also called validation) phase. As the method is based on searching an adequate subset of features inside the feature space for a particular classifier, it results to high cost in computation as well as the tendency to be overly specific for the classifier, which

may yield into lack of generalisation. In **embedded methods** the feature search is reduced to only the training phase, which reduces computation costs. Nonetheless, limitations are still present since the feature subset still relies on the selected classifier. On the other hand **filter methods** are constructed over the assessment of the candidate feature or subset of features by a heuristic scoring criterion (usually based on statistical tests, variance or information among others). Filter methods are computationally efficient and independent of classifiers [23]. The selection of the approach to follow depends on the number of events/observations and features/inputs, which for instance for embedded methods are superior for large number of observations, while wrapper methods become impossible for more than forty features [45] (cf. figure 4.3). In radiomic studies, the number of features are often rather large with very few data points (patients), therefore filter methods are frequently the most suitable approach as a first feature selection algorithm. The determinant element in filter methods is the heuristic scoring criterion which is translated into a relevance index (ri) with aim to maximise "relevancy" with minimum redundancy. Here, some of the most used relevance indexes used in radiomics are presented [132]:

4.3.1. Fisher score

The fisher score aims to select features that maximise separation of values between class labels (m) and minimise the within class label distance [41]. In other words it favours features with larger distances between the average label ($\mu_{d,m}$) and feature average (μ_d) with lower variance within the label ($\sigma_{d,m}^2$) as follows:

$$\mathrm{ri}(x_d) = \frac{\sum_m n_m \left(\mu_{d,m} - \mu_d\right)^2}{\sum_m n_m \sigma_{d,m}^2} \tag{4.6}$$

Here, n_m accounts for the number of samples or observations (patients).

4.3.2. T-test score

The T-test score searches for features that could largely separate classes with maximum stability or low noise (variability across observations) as follows:

$$\mathrm{ri}(x_d) = \frac{\mu_1 - \sum_{m=2} \mu_m}{\sqrt{\sum_m \left(\frac{\sigma_m^2}{n_m}\right)}} \tag{4.7}$$

The reader might notice that m refers to the label class and n_m the number of observations/samples inside the label class. The T-test score provides outstanding results when the number of samples are low (under 40) with significantly high number of features [79].
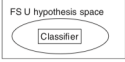
Model search	Advantages	Disadvantages	Examples
Filter	Univariate		
	Fast	Ignores feature dependencies	χ^2
	Scalable	Ignores interaction with	Euclidean distance
	Independent of the classifier	the classifier	*i*-test
			Information gain,
			Gain ratio (Ben-Bassat, 1982)
	Multivariate		
	Models feature dependencies	Slower than univariate techniques	Correlation-based feature
	Independent of the classifier	Less scalable than univariate	selection (CFS) (Hall, 1999)
	Better computational complexity	techniques	Markov blanket filter (MBF)
	than wrapper methods	Ignores interaction	(Koller and Sahami, 1996)
		with the classifier	Fast correlation-based
			feature selection (FCBF)
			(Yu and Liu, 2004)
Wrapper	Deterministic		
	Simple	Risk of over fitting	Sequential forward selection
	Interacts with the classifier	More prone than randomized	(SFS) (Kittler, 1978)
	Models feature dependencies	algorithms to getting stuck in a	Sequential backward elimination
	Less computationally	local optimum (greedy search)	(SBE) (Kittler, 1978)
	intensive than randomized methods	Classifier dependent selection	Plus q take-away r
			(Ferri *et al.*, 1994)
			Beam search (Siedelecky
	Randomized		and Sklansky, 1988)
	Less prone to local optima	Computationally intensive	Simulated annealing
	Interacts with the classifier	Classifier dependent selection	Randomized hill climbing
	Models feature dependencies	Higher risk of overfitting	(Skalak, 1994)
		than deterministic algorithms	Genetic algorithms
			(Holland, 1975)
			Estimation of distribution
			algorithms (Inza *et al.*, 2000)
Embedded	Interacts with the classifier	Classifier dependent selection	Decision trees
	Better computational		Weighted naive Bayes
	complexity than wrapper methods		(Duda *et al.*, 2001)
	Models feature dependencies		Feature selection using
			the weight vector of SVM
			(Guyon *et al.*, 2002;
			Weston *et al.*, 2003)

Figure 4.3.: Summary of the advantages and disadvantages of feature selection approaches. Reprinted from [146] with allowance of the Oxford University Press provided by the Copyright Clearance Center ("CCC").

4.3.3. χ^2 score

On contrary to the previously defined criterions, the χ^2 score is histogram-based of samples grouped in r feature values across classes.

$$\text{ri}(x_d) = \sum_r \sum_m \frac{(n_{r,m} - \mu_{r,m})^2}{\mu_{r,m}} \tag{4.8}$$

Here $n_{r,m}$ represents the number of samples that have the r-feature value in the m class. $\mu_{r,m}$ is defined as follows:

$$\mu_{r,m} = \frac{n_i \cdot n_m}{N} \tag{4.9}$$

where n_r stands for the number of samples with r-feature value and n_m the number of samples belonging to the m class. N is the total number of samples [132].

4.3.4. Wilcoxon rank test score

The Wilcoxon rank test ranks features according to separations of the median population between classes as follows:

$$\text{ri}(x_d) = (N-1)\frac{\sum_m n_m \left(\mu_{r_m} - \mu_r\right)^2}{\sum_m \sum_{n_m} (r_{m,n_m} - \mu r)^2} \tag{4.10}$$

The r_{m,n_m} is the rank value of the sample in class m, μ_{r_m} stands for the average of the rank of samples of m class, and μ_r is the mean rank of all samples [132].

4.3.5. Gini index score

The Gini score is also accounted as a measure of impurity and can be computed as follows:

$$\text{ri}(x_d) = 1 - \sum_m \left(p(m \mid x_d)\right)^2 \tag{4.11}$$

Where $p(m \mid x_d)$ is derived from the probability of having m given the feature x_d. For smaller gini index score, the subset of features can predict better the classes.

4.3.6. Mutual Information Maximisation

Another impurity measure which comes from information theory of Shannon [153]. As follows:

$$I(x_d \mid m) = \sum_{x_d} \sum_m p(m \mid x_d) \log \left(\frac{p(m \mid x_d)}{p(m)p(x_d)}\right) \tag{4.12}$$

In other words it measures the amount of entropy loss caused by the introduction of a x_d feature.

4.3.7. Mutual information feature selection

Here the relevancy score (ri) penalises redundant features (x_d with x_k) and favours features with mutual correlation to outcome class (m) as proposed by Battiti [11]:

$$\text{ri}(x_d) = I(x_d \mid m) - \beta I(x_d \mid x_k) \tag{4.13}$$

Battiti proved that $\beta = 1$ is frequently optimal, yet if $\beta = \frac{1}{d}$ where d is the number of features, minimises the relevancy-redundancy trade-off as proved by Peng et al [134].

4.3.8. Joint mutual information

The relevancy score favours features (x_d) that in combination with another feature (x_k) can correlate with the outcome classes as follows [108]:

$$\mathrm{ri}(x_d) = I\left(x_d, x_k \mid m\right) \tag{4.14}$$

4.3.9. Conditional infomax feature selection

The relevancy score, similar to the joint mutual information relevancy score, rewards features x_d that in combination with other subset of features x_k inside the feature space correlate with outcome class m, yet features x_d correlate with the outcome label at the same time. Besides it penalises features that correlates with any other one inside the feature space as follows [107]:

$$\mathrm{ri}(x_d) = I\left(x_d \mid m\right) - I\left(x_d \mid x_k\right) + I\left(x_d, x_k \mid m\right) \tag{4.15}$$

For more relevancy scores, the reader might refer to the work of Koller and Sahami [93] and Yu and Liu [179]. The complete description of wrapper methods, please refer to Pudil et al [136] and Siedlecky and Sklansky [154]. For embedded methods, please refer to Duda et al [41] and Guyon et al [64].

4.4. Machine learning models

As a part of the radiomics pipeline, selection of a model that is able to predict outcomes from a proposed subset of features is a decision of the researcher based probably in its field of expertise, since ML models come from multiple disciplines inside computer science, mathematics, neuroscience, physics and biology, the decision is mostly biased on the knowledge of models and not the suitability of models to data, which may yield into lack of generalisation in unseen data. A good practice of machine learning in the context of clinical studies relies on testing features with a handset of models and perform analysis on them [31]. The ML methods in this work were implemented inside the binary and multi-label classification tasks (supervised machine learning) as follows:

4.4.1. k-Nearest neighbours (KNN)

The KNN model is one of the most intuitive models, it does not learn any decision function from data but it rather clusters new input based on similarity to the experienced data (cf. figure 4.4). The algorithm can be summarised in:

1. Select the number of neighbour data-points k and a distant metric

2. Find the k-closest data-points to the new data point
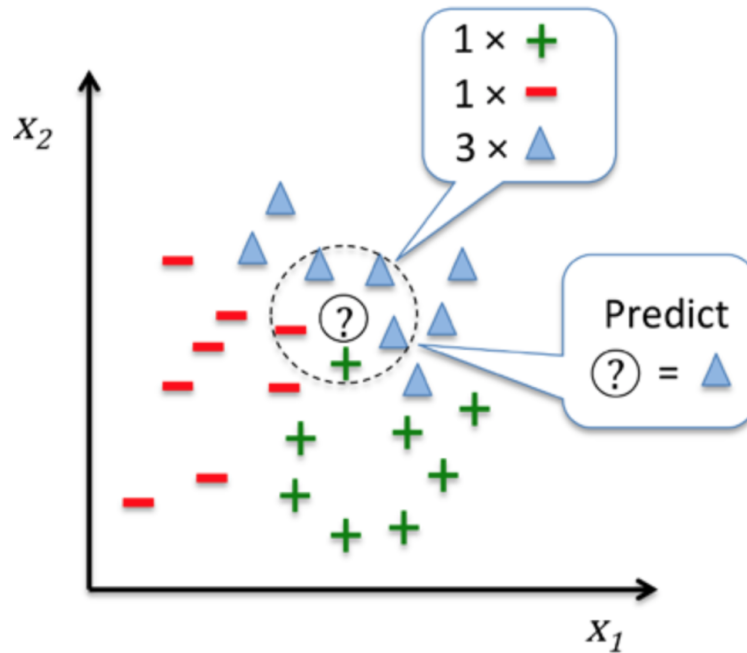
Figure 4.4.: Visualisation of of how new data-points are labelled according to one of the classes based on majority vote in KNN, x_1 and x_2 represents the features in the hypothetical problem. Reprinted from [138] with allowance of Packt Publishing Limited provided by Copyright Clearance Center ("CCC").

3. Assign the label to the new data point based on voting of the k-nearest data-points

The main advantage of a KNN is the adaptability to collection of new data, however removing samples burdens the performance of the classifier. Besides the complexity problem of KNN grows linearly as the number of samples increases [46].

4.4.2. Logistic regression (LogR)

Despite of the name this model is very popular among clinical studies to establish thresholds for decision making (classification). The decision function is the sigmoid function (cf. equation 4.16) which due to the S shape (cf. figure 4.5), it can be fitted to separate outcome classes across feature values. Let us assume z as the net combination of inputs ($\mathbf{x}$) with adjusted weights ($\mathbf{W}$), such that $z = \mathbf{W}^T\mathbf{x} = w_0 + w_1x_1 + w_2x_2 + ... + w_dx_d$ being d the cardinality of $\mathbf{x}$ or in other words, the number of features.

$$\phi(z) = \frac{1}{1 - e^{-z}} \tag{4.16}$$
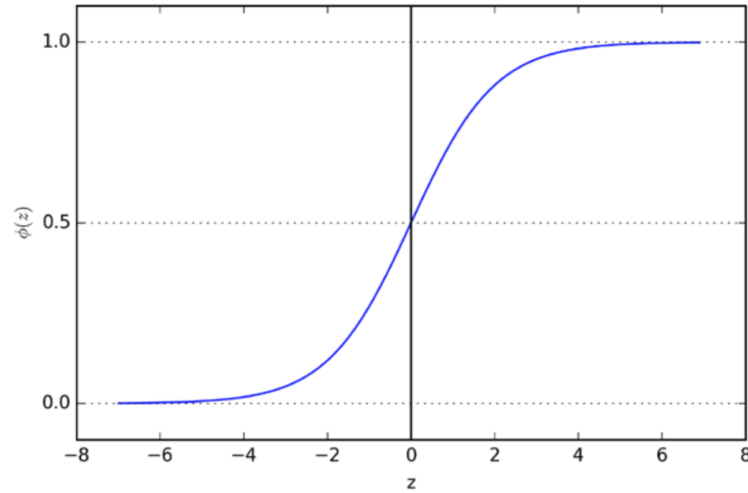
Figure 4.5.: Sigmoid function visualisation. As $z \to \infty$, $\phi(z) \to 1$, when $z = 0$, $\phi(z) = 0.5$ and if $z \to -\infty$, $\phi(z) = 0$. Reprinted from [138] with allowance of Packt Publishing Limited provided by Copyright Clearance Center ("CCC").

To fit the $\mathbf{W}$ according to features and label, a construction of a cost function is required as follows:

$$J(\mathbf{W}) = C \left[\sum_i \left(-\log \left(\phi(z^i) \right) + \left(1 - y^i \right) \left(-\log \left(1 - \phi(z^i) \right) \right) \right] + \frac{1}{2} \|w\|^2 \quad (4.17)$$

Which accounts for the regularisation error of the logistic regression function in comparison to the actual classification labels. Here to find the adequate $\mathbf{W}$, optimisation of the function $J(\mathbf{W})$ with respect to $\mathbf{W}$ is needed. The constant C accounts the strength of the constraint to the samples. The success of the model depends on the linear separability of the samples in the feature space with the label classes. For the detailed deduction of the weights, the reader might refer to the book of Friedman, Tibshirani and Hastie [45].

4.4.3. Support vector machines (SVM)

The logistic regression algorithm intends to maximise the conditional likelihood of the training data, whereas the idea behind SVM is to maximise the distance between the closest data points to the decision boundary (also called support vectors, cf. figure 4.6). In mathematical terms, the equation

$$\frac{\mathbf{w}^T \left(x_{pos} - x_{neg} \right)}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|} \quad (4.18)$$
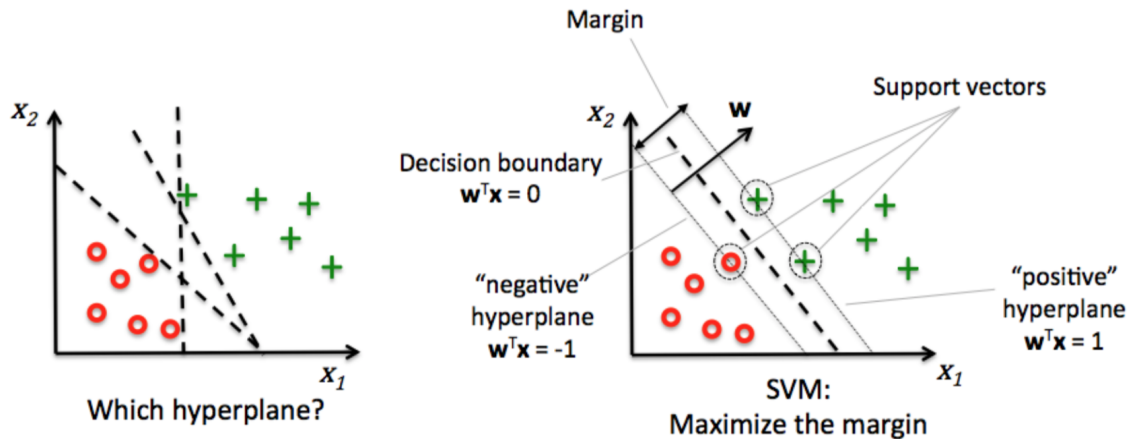
Figure 4.6.: Selection of an optimal decision boundary inside a training cohort. Here margins means as the distance between the separating hyper-planes. Reprinted from [138] with allowance of Packt Publishing Limited provided by Copyright Clearance Center ("CCC").

is required to be maximised, yet computationally results much easier to minimise $\frac{\|w\|^2}{2}$. For full description of the method, the reader might refer to the book of Vapniks [170]. One of the advantages of this approach with respect to logistic regression is that SVM is insensitive to outliers since it only accounts for the support vectors and not the complete training data-points as in logistic regression. However, finding the best hyper-plane yields in high computational costs or no convergence at all. For these cases, a relax constant (ζ) was introduced such that:

$$
\begin{aligned}
\mathbf{w}^T\mathbf{x}^i \leq 1 \quad &if \quad y^i = 1 - \zeta^i \\
\mathbf{w}^T\mathbf{x}^i < -1 \quad &if \quad y^i = 1 + \zeta^i
\end{aligned}
\tag{4.19}
$$

Yielding into a new equation $\frac{1}{2}\|\mathbf{w}\|^2 + C\left(\sum_i \zeta^i\right)$ to be minimised. The choice of the constant C adapts the penalty strength for misclassification (cf. figure 4.7).
One advantage of the SVM over the logistic regression model is that it can be generalised for classification problems that are not linearly separable by using the kernel trick, as mentioned in section 4.2.2, which means to increase the feature space in one dimension more where the problem can be linearly separable and retrieve the margin to the original feature space (cf. figure 4.8).

4.4.4. Gaussian Naïve Bayes (GNB)

The Naïve-Bayes classifier, predict probabilities of belonging to class m having features $X = [x_1, x_2, ..., x_D]$, based on independent probabilities across the feature space (cf. equation 4.20). Despite this is not always the case and a rather naive assumption, it often provides outstanding results.
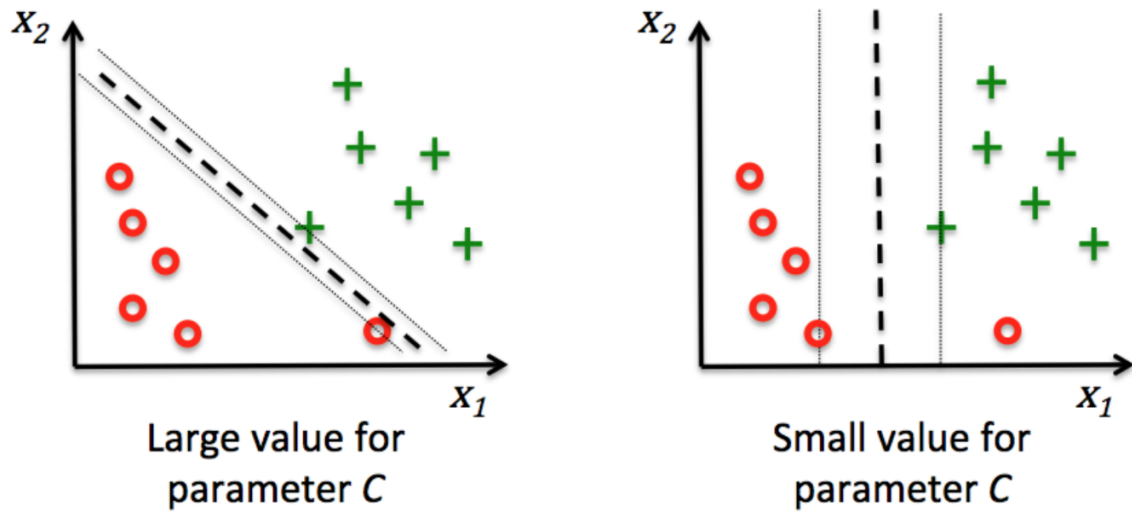
Figure 4.7.: Difference in margins for different C in a hypothetical classification problem. Reprinted from [138] with allowance of Packt Publishing Limited provided by Copyright Clearance Center ("CCC").

$$p(m \mid X) = \prod_{d=1}^{D} p_d(m \mid x_d) \tag{4.20}$$

4.4.5. Decision Trees (DT)

The construction of a model based on decision tree relies on the separation of data based on questions or thresholds across the feature space (cf. figure 4.9). Here the main component is the measurement of separability due to features in the classes across samples, for which there are usually three types of impurity measures *Entropy*, *Gini index* and *classification error*. The *entropy* is a measure of the disorder of features values inside the classes, thus the lower the entropy, the better the separability of classes according to feature. Mathematically as follow:

$$E(t) = - \sum_{c} p(c \mid t) \log_2 p(c \mid t) \tag{4.21}$$

$p(c \mid t)$ is the probability of having the class c inside the a defined t node (question/threshold). As in equation 4.11 the *Gini index* intends to minimise misclassification probabilities in t nodes. The *classification error* is less sensitive to changes as it is not differentiable (cf. equation 4.22). The decision tree performance using the different impurity measurements yields to similar results (cf. figure 4.10).

$$CE(t) = 1 - p(c \mid t) \tag{4.22}$$

In short the construction of the decision tree algorithm is as follows:
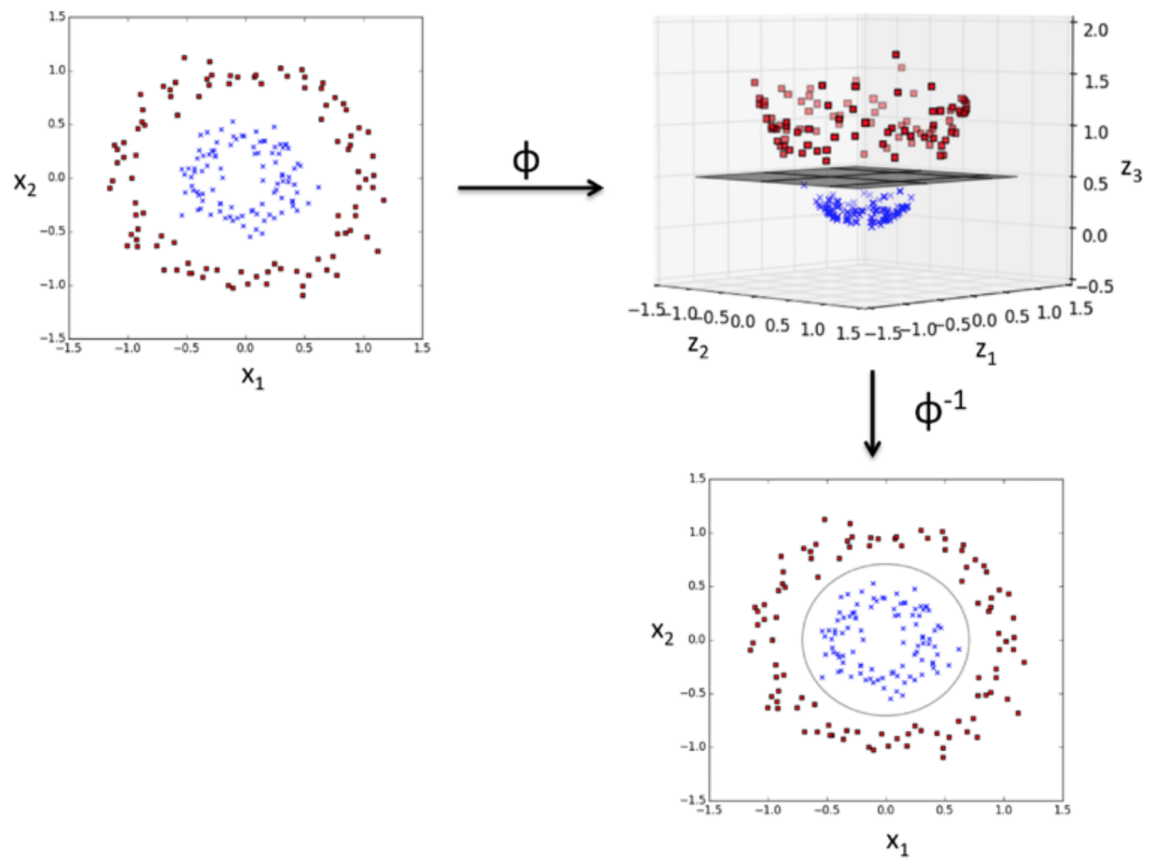
Figure 4.8.: Margin construction using a RBF kernel for a non-linearly separable hypothetical classification problem. Reprinted from [138] with allowance of Packt Publishing Limited provided by Copyright Clearance Center ("CCC").
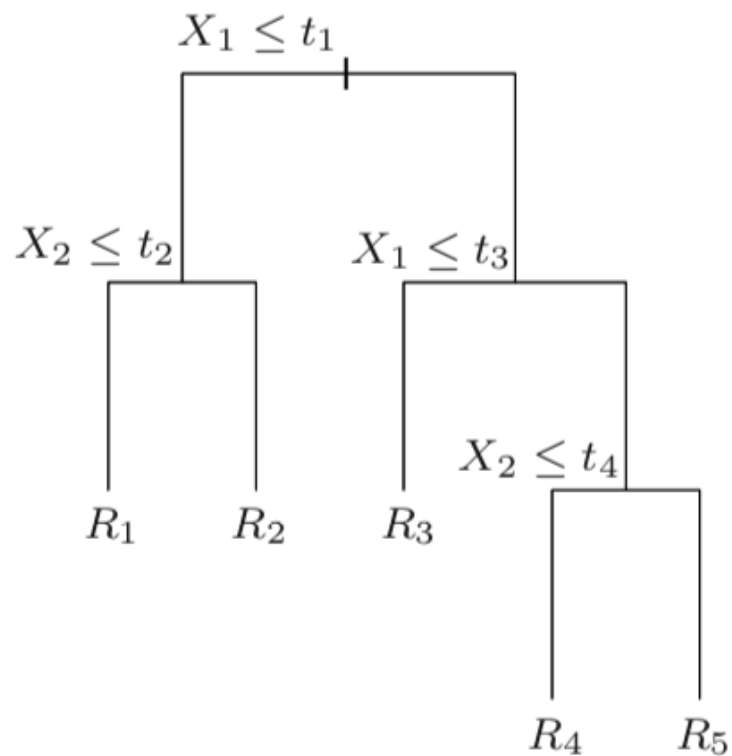
Figure 4.9.: Example of a construction of a tree according to two feature space X_1 and X_2. Reprinted from [45] with allowance of Springer Nature provided by the Copyright Clearance Center.
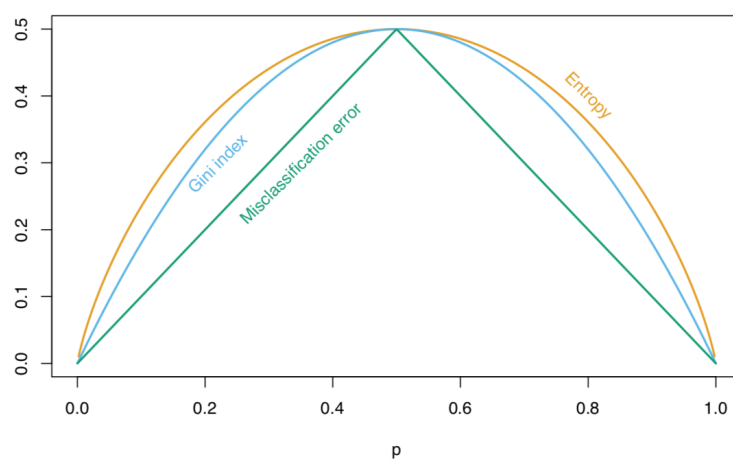
Figure 4.10.: Impurity measure values for a defined t node [45].

1. Select an impurity measure

2. Select a feature that minimises the impurity measure according to classes

3. Create a node and select a threshold that minimises the impurity measurement

4. repeat 2 and 3 until desired minimisation of the impurity measure

The deeper the tree, the higher the risk of overshooting the decision function.

4.4.6. Random forest (RF)

The random forest algorithm profits from assembling decision trees that even in case they do not provide good performances as standalone, the combination of them does. The reasons are that due to a partial capturing of the complexity of the task by every built decision tree. The random forest algorithm is described as follows:

1. Choose randomly n samples from the training cohort with replacement

2. From the bootstrap sample construct a decision tree with the best subset features that minimises the impurity measurement

3. Repeat 1 and 2 k-times.

4. Predict via majority votes of decision trees

As the decision function comes via majority vote of decision trees, overly complicated decision trees are balanced by simple decision trees, which it avoids model overfitting of course at expenses of computational costs if the number of decision trees are increased. As we increase the number of DT, the interpretation of the model is also cumbersome.

4.5. Model selection and tuning

One of the challenges of machine learning is to select a highly predictive model from a learning/experienced data to correctly predict outcomes in an unseen/"untouched" new data set for generalisations. To do so, the model should find an optimal solution between capturing the complexity of the data (variance) and the under fitting (bias) the model. All machine learning models excel at learning data since they are design to be able capture the complexity of data. Therefore, assessing the performance of a model based on a metric error in training data as whole does not provide an excellent indicative for later performances in unseen/new data sets. Nevertheless, from the learning cohort an indicative or expected performance metric might be obtained via statistical methods for re-sampling. Let us start with the choice of an error metric to assess model performance in classification problems (regression problems were not implemented in this thesis and therefore are out of the scope of it, the reader might refer to the book of Friedman et al [45] for further readings):

4.5.1. Model performance metrics

The most intuitive model performance metric for classification problems is to account for the number of correct predictions or true positives (TP) plus true negatives (TN) divided by the total number of observations (N), usually called **accuracy** or in mathematical words:

$$Acc = \frac{TP + TN}{N} \tag{4.23}$$

Notwithstanding, this metric fails to capture real model performance in an unseen dataset when the label outcomes are imbalanced (one of the classes are repeated more than the others), as it is very often the case in clinical studies. As an example, let us assume a cohort of patients with different stages of tumours, the task will be to stratify advanced from non advanced stages, however one of the classes (let us say non-advanced) are repeated 75% more than the other one. If a model that predicts that all patients will have non-advanced stage of tumours, the accuracy will be already 75% for a very uninformative model, which can not predict correctly one of the classes. Better error metrics for classification are based on the so-called **true positive rates** (TPR) and **false positive rates** (FPR) which are ratios between the number of TP and total positives (T) predicted by the model and false positives (FP) divided by the total number of negatives (N) predicted by the classifier, in mathematical terms:

$$TPR = \frac{TP}{T}$$
$$FPR = \frac{FP}{N} \tag{4.24}$$

In the clinical context and therefore in radiomics, the TPR is much more important than the FPR. Using these two ratios, error metrics such as **precision** (PRE), **ROC-AUC** and **F1** score can be derived as:

$$PRE = \frac{TP}{TP + FP}$$
$$F1 = 2\frac{PRE \times TPR}{PRE + TPR} \tag{4.25}$$

The **ROC-AUC** score comes from the area under the curve of plotting TPR and FPR. A perfect classifier would identify 100% (or 1) of TPR for 0% (or 0) of FPR.

Despite all the error metrics explained in this section were discussed under the binary classification problem assumption, they can be extended to the multi-class problem via **one vs all** algorithm [141].

4.5.2. Detection of over- or under-fitting in models

After selection of a model performance metric the next step to follow is to estimate performances of models. The classic strategy to analyse this is to split the learning
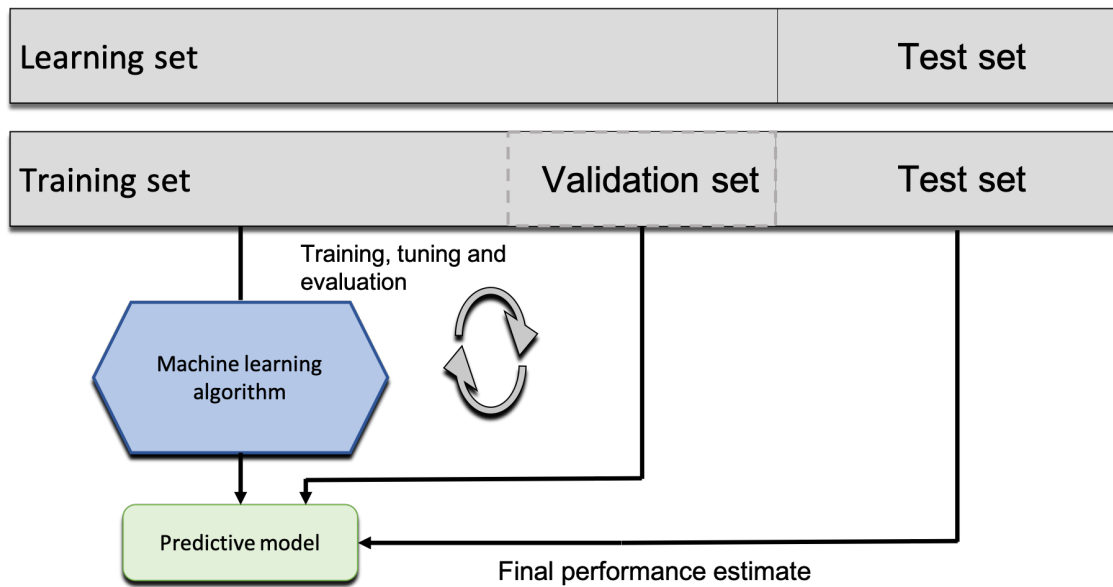
Figure 4.11.: Holdout workflow for model validation. Reprinted from [138] with allowance of Packt Publishing Limited provided by Copyright Clearance Center ("CCC").

cohort into a training and validation cohort and confirm performance in a holdout test cohort (also called the **holdout** method, cf figure 4.11), if it is performed just once, no generalisation can be expected since the model performance metric might be influenced by the way the learning cohort was split. A much more robust approach is to split the learning cohort in k-arbitrary folds where k-1 folds are used to train and the k-fold used to test, k times randomly (also called **cross validation**, cf. figure 4.12).

The cross validation approach is challenging when the imbalanced class problem is treated. The study of Kohavi et al [92] showed that by using stratified k-cross validation there is a reduction of the variance - bias trade-off. The choice of k-folds depends on the extension of the dataset, for large dataset is recommended to use small k-folds since it reduces computational costs with robust performance. On the other hand, for small datasets, large k-folds estimate better the expected performance metric. In literature 10 folds is used as a standard choice. Finally another very useful approach for estimating performance metrics in independent data is the **bootstrapping** method, which in general terms is similar to cross-validation yet with replacement, this means that for every iteration, the training and test samples are re-sampled in a way that it might include already used samples in previous iterations.

The selection of the approaches to estimate a model performance metric depends on the extension of the data set, computer power and preferences of the researcher. Let us go back to the definition of under- and over-fitting, a model is considered
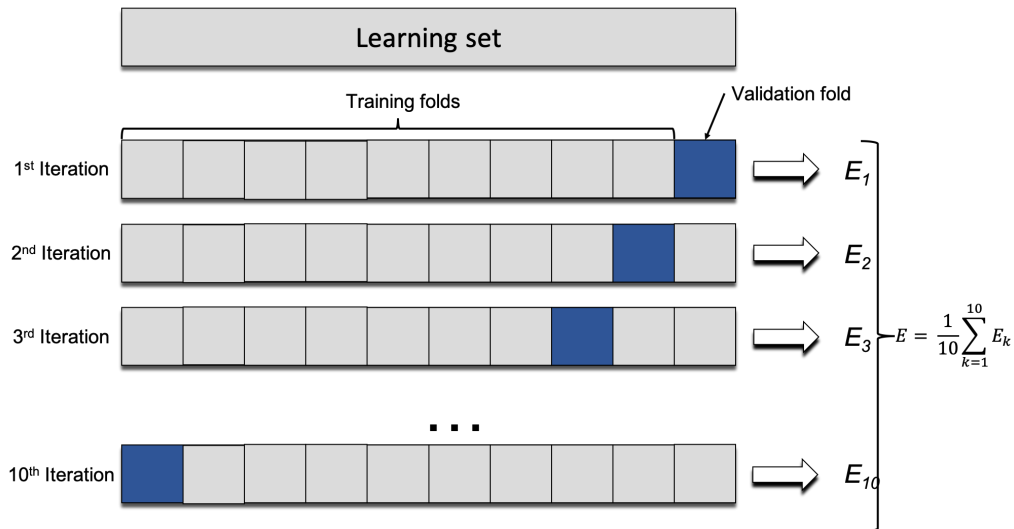
Figure 4.12.: 10 cross-validation workflow for model validation, E stands for error metric average over the different error metric in test fold E_k. Reprinted from [138] with allowance of Packt Publishing Limited provided by Copyright Clearance Center ("CCC").

under-fitted (high biased) if the model performance metric in the training and validation cohort are similar but far away of the desired model performance metric or if the model performance metric in the test cohort outperforms the model performance in the learning cohort, this might be because of the simplicity of the model (or limited number of predictive features). On contrary, a model is considered over-fitted (with high variance) if the metric performance model in the training and validation cohorts are very different, with the model performance metric in the training cohort superior than the validation cohort, this means that the model can capture the complexity of the data in the training phase but unable to predict similar performance in unseen data. Learning curves helps to detect those issues by plotting the model performance metric against a parameter that is considered to produce the issue, for instance number of samples, number of features, a determined parameter in the model among others (cf. figure 4.13).

4.5.3. Tuning machine learning models to address over-/under-fitting

Very few alternatives are available when the problem of over-/under-fitting comes from the fact that the modelling is performed under few data samples. To maximise and generalise performance of models the best solution is to increase the number of observations. This in clinical studies is often very expensive or even infeasible. Nonetheless, model optimisation parameters can be always performed. This is called model hyper-parametrization, which means to find the optimal model parameters
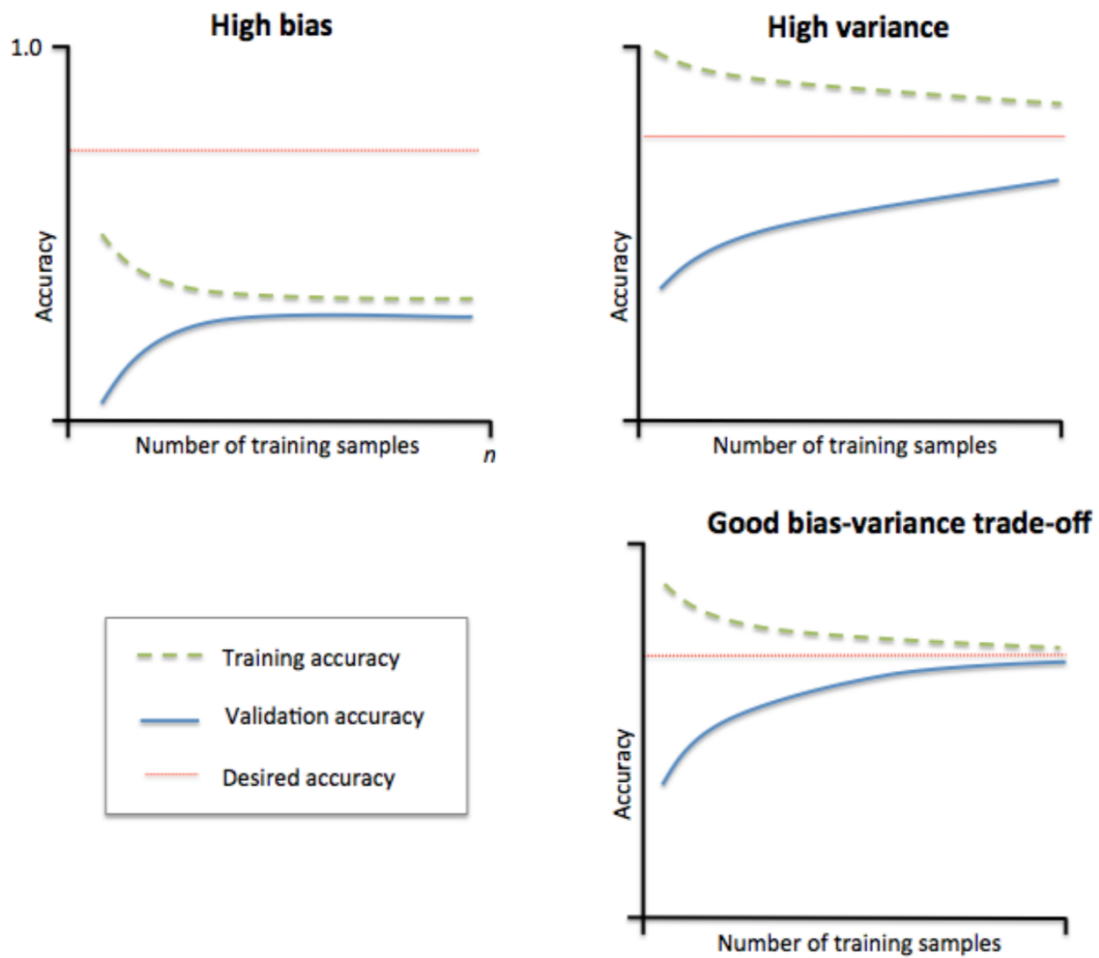
Figure 4.13.: Learning curve for different bias-variance models according to number of training samples and model performance accuracy. Reprinted from [138] with allowance of Packt Publishing Limited provided by Copyright Clearance Center ("CCC").

that perform the best in training and validation cohorts. In this work, a large set of hyper-parameters were considered and the best subset of them was found via grid search (Looping over the whole hyper-parameter space to find the optimal subset of them).

5. Python as Software Development Language

5.1. Pyhton

Python is an open source oriented object programming language created and designed by Guido van Rossum. Due to open source it has a vast community of developers who construct applications in many fields, from Video Games to Scientific computations. Python is currently used in two versions, the version 2.x and 3.x.

5.1.1. Why Python?

The implementations of this work for image pre-processing and feature extraction were carried under the programming language Python 3.6. The reasons for the choice of Python 3.6 over any other open source programming languages, lies on the the trade-off between stability, readability, simple syntax, extensibility and a big community of developers that could help to shape codes for better implementation, which allows easy comparisons. The selection of Python 3.6 over the 2.7 version lies on the fact that the 2.7 version in the near future will not longer support new packages, which in terms of software durability inside UKT is a big downside. Python 3.6 has reached a maturity which allows developers to profit for an extensive amount of scientific packages to perform multiple studies, from mathematical utilities to imaging analysis in the framework of this thesis. One of the downsides of Python is related to the memory allowance, Python as being a dynamic-memory programming language does not fix memory size allocations for different types of variables (such as integers, floats and so on.). However with the fast development of computer capacities such problems have been relegated as a minor issue, which in term of the framework of this work are easily handled. As the production code for image pre-processing and feature extraction was implemented in Python 3.6, the model building and prediction phase was also coded in the same language to reduce latency and track easily issues. ML libraries for python are relatively new inside the environment in comparison with the R software and therefore relatively limited. However for purposes of ML modelling and predictions in the context of radiomics inside the framework of this thesis, libraries such as *Pandas*, *Scikit-learn* and *Mlxtend* as well as *Matplotlib* for visualization are enough for the software production.

5.1.2. Python libraries inside the framework of the thesis

The number of scientific libraries are very broad. Nonetheless inside this work, *NumPy*, *Nibabel*, *SciPy*, *Scikit-image*, *PyWavelets*, *Pandas*, *Scikit-learn*, *Mlxtend* and *Matplotlib* were mainly used.

Numerical Python (NumPy):

NumPy is the highly efficient standard library to handle numeric operations inside the Python environment. It provides the core multidimensional array object that are necessary for scientific purposes [27]. NumPy integrates algorithms and data structures from C/C++ and Fortran that have been developed over the years and are already known to work time and memory efficiently. The data structure of NumPy is the object ndarray.

Scientific Python (SciPy):

SciPy is a library that is built on top of NumPy, which in combination with NumPy allows manipulations of ndarrays for scientific and engineering purposes. The SciPy library contains mathematical and statistical algorithms that were implemented in the framework of this thesis, such as Pearson, Kendall and Spearman correlation coefficient and many other statistical tools.

Nibabel:

Nibabel is a scientific library for Python that translates many medical image formats to NumPy array objects [22]. DICOM is the standard format for medical images, but due to data protection affairs Nifti format sometimes it is preferred since it only provides the imaging data without extra information (fully anonymous).

Scikit-Image:

Scikit-image provides a vast number of algorithms and utilities for imaging processing in 2D and 3D, such as segmentation, registration, transformation, rendering, object recognition and so on [167].

PyWavelets:

PyWavelts library for Python allows to transform 3D array objects in the Fourier space via wavelets functions of preference [173]. In the framework of this thesis it was implemented to obtain Wavelets-family features.

Pandas:

Pandas, which name come from **pa**nel **da**ta, is a scientific library for data analysis in Python which is built on top of NumPy, SciPy and Matplotlib (for visualisation purposes). The data structure object of pandas is the DataFrame and Series. It gives the ability to Python to work with spreadsheet-like data for fast loading, manipulation, visualisation and handling [114].

Scikit-learn:

Scikit-learn is a highly optimised library for Python that provides a vast number of algorithms and utilities for machine learning purposes [133].

Mlxtend:

The Mlxtend library for Python provides machine learning algorithms and utilities that were not included in the Scikit-learn library, such as the sequential feature selection algorithm [139].

Matplotlib:

Matplotlib is a highly optimised library for purposes of visualisation in the Python environment, it posses all the advantages of Matlab and IDL for free [176].

Part III.

Results

6. In-house software development for Radiomics and its validation

Since the genesis of radiomics in 2012 [98], few widely accessible radiomic implementations have been developed to carry complete radiomic studies. However, they still lack on validation [168, 181]. Therefore an in-house software development approach was carried with the aim of integrating a customised radiomic pipeline inside the UKT radio-oncological research necessities. In this chapter, a description of the radiomic implementation inside this work will be provided as well as the its performance inside the IBSI collaboration [184], for the sake of completeness, the reader might refer to the appendices 12 and 13 for full software disclosure.

6.1. Software validation inside the IBSI collaboration

One of the biggest challenges inside the field of radiomics is to be able to reproduce findings across institutes and patient data and if not possible, argue why otherwise, many studies have proved problems of this nature [57, 68, 178], however few collaborations have been achieved to tackle the problem so far. The IBSI were born as an open collaboration initiative with aim to make radiomics a reproducible science [184] with participation of more than 20 institutions inside Europe, USA and Canada (cf. table 6.1). As an open collaboration initiative, the radiomic software developed in the framework of this thesis also participated in it. Nonetheless, the reach of the IBSI initiative is limited to common definitions and nomenclature of features and methods to extract them, the validity of feature stability and machine learning algorithms, remains for further investigations. In this chapter contributions to the IBSI initiative will be disclosed for benchmark acquisitions in the customised radiomic software.

IBSI contributors by institution	Location
Department of cancer imaging and metabolism, Moffitt Cancer Center	Tampa (FL), USA
Dana-Farber Cancer Institute, Brigham and Women's Hospital, and Harvard Medical School, Harvard University	Boston (MA), USA
Institute of Information Systems, University of Applied Sciences Western Switzerland (HES-SO)	Switzerland
Department of medical physics, Memorial Sloan Kettering Cancer Center	New York (NY), USA

Department of electrical and computer engineering, Whiting School of Engineering, Johns Hopkins University	Baltimore (MD), USA
Russell H. Morgan department of radiology and radiological science, School of medicine, Johns Hopkins University	Baltimore (MD), USA
Department of nuclear medicine and molecular imaging, University of Groningen, University Medical Center Groningen (UMCG)	Groningen, the Netherlands
Department of Radiation Oncology, University Hospital Zurich, University of Zurich	Zurich, Switzerland
Polo Scienze Oncologiche ed Ematologiche, Istituto di Radiologia, Università Cattolica del Sacro Cuore, Fondazione Policlinico Universitario Agostino Gemelli	Rome, Italy
Biomedical Imaging Group, Ecole polytechnique federale de Lausanne (EPFL)	Lausanne, Switzerland
Laboratory of medical information processing (LaTIM) - team ACTION (image-guided therapeutic action in oncology), INSERM, UMR 1101, IBSAM, UBO, UBL, Brest	France
Imaging technology for radiation therapy group, the Netherlands Cancer Institute (NKI)	Amsterdam, the Netherlands
Department of Radiation Oncology, Physics Division, University of Michigan and Medical Physics Unit, McGill University	Montréal, Québec, Canada
Division of radiation oncology, MD Anderson Cancer Center	Houston (TX), USA
Surgical Planning Laboratory, Brigham and Women's Hospital and Harvard Medical School, Harvard University	Boston (MA), USA
Department of medical image computing, German Cancer Research Center (DKFZ)	Heidelberg, Germany
Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital (MGH) and Harvard Medical School, Harvard University	Cambridge (MA), USA
Department of Radiation Oncology (The D-lab), GROW-School for Oncology and Developmental Biology, Maastricht University Medical Centre+	Maastricht, The Netherlands
Section for Biomedical Physics, Department of Radiation Oncology, Universitätsklinikum Tübingen, Eberhard Karls Universität Tübingen	Tübingen, Germany

Department of Clinical Medicine, University of Bergen	Bergen, Norway
OncoRay – National Center for Radiation Research in Oncology, Faculty of Medicine and University Hospital Carl Gustav Carus, Technische Universität Dresden, Helmholtz-Zentrum	Dresden - Rossendorf, Germany
Department of Radiation Physics, University of Texas MD Anderson Cancer Center	Houston (TX), USA
Institute of Information Systems, University of Geneva	Geneva, Switzerland
Cardiff School of Engineering, Cardiff University	Cardiff, United Kingdom
Imaging technology for radiation therapy group, the Netherlands Cancer Institute (NKI)	Amsterdam, the Netherlands
Department of radiology, Leiden University Medical Center (LUMC)	Leiden, the Netherlands

Table 6.1.: List of contributors of the IBSI collaboration by institution and location [184].

6.1.1. Digital imaging phantom

To preliminary test definitions and algorithms for feature extraction two small digital numerical arrays were proposed, one of the arrays was proposed as exemplary image and the second one as a mask to obtain a region of interest. Here, intensity, shape and texture features were only considered for extractions, due to filter-family features do no add reproducibility value since they depend on the chosen filter (Imaging preprocessing step). The phantom characteristics are as follow:

- The voxel size of the array is assumed to be $2 \times 2 \times 2$ mm

- The number of voxels are $5 \times 4 \times 4$ (x, y, z)

- The voxels included in the region of interest are given by a mask which is an array of values, where 1 means that the voxel belongs to the ROI and 0 otherwise, cf. figure 6.1.

- Some intensities are not included in the phantom. The minimum grey level is 1 whereas 6 is the maximum.

Since, the image representation does not require preprocessing (it can be directly represented as a *NumPy* array object), the feature calculations were performed directly as showed in subsection 6.1.3.

6.1.2. Radiomic phantom data

A publicly available patient data set was proposed to construct benchmarks for imaging pre-processing methods and radiomic feature extraction (`DOI:10.17195/candat.2016.08.1`), inside the cohort of the patient images, *PAT1* was used. The images are stored as a stack of slices of *DICOM* and *NifTI* format. A ROI mask of the Gross tumour volume (GTV) was also provided for calculations. Despite the large variation of possible configurations inside the framework of IBSI, in this work only configuration C (features extraction in a 3D approach, merging texture matrix and 64 bins) was considered since the implemented software followed these specifications and therefore only these results are presented (cf. figure 6.2).

6.1.3. Bechmarks

After more than two years work inside the IBSI collaboration, benchmarks for the digital phantom were achieved. In the case of the radiomics phantom, although some consensus have been reached in many of the analysed features, further investigations remain necessary. Here, benchmarks are presented for the different family features as well as the results of the software constructed in the frame of this thesis as follows:

Intensity-based statistical features:

feature	dig. phantom	conf. C	**Tübingen** (dig./ conf.)
mean	2.15	-49 ± 2.7	(Match, Match)
variance	3.05	$(5.06 \pm 0.13) \cdot 10^4$	(Match, Match)
skewness	1.08	-2.14 ± 0.04	(Match, Match)
kurtosis	-0.355	3.53 ± 0.22	(Match, Match)
median	1	40 ± 0.3	(Match, Match)
minimum	1	-939 ± 3	(Match, No match)
10th percentile	1	(-424 ± 13)	(Match, Match)
90th percentile	4	86 ± 0.1	(Match, Match)
maximum	6	393 ± 10	(Match, No match)
interquartile range	3	67 ± 4.7	(Match, Match)
range	5	$(1.33 \pm 0.01) \cdot 10^3$	(Match, Match)
mean absolute variation	1.55	158 ± 3	(Match, Match)
robust mean absolute deviation	1.11	66.8 ± 3.4	(Match, Match)
median absolute deviation	1.15	119 ± 3	(Match, Match)
coefficient of variation	0.812	-4.59 ± 0.28	(Match, Match)

quartile coefficient of dispersion	0.6	1.03 ± 0.46	(Match, Match)
energy	567	$(2.44 \pm 0.11) \cdot 10^9$	(Match, Match)
root mean square	2.77	230 ± 3	(Match, Match)

Table 6.2.: Comparisons between the software constructed for radiomics inside the framework of this thesis and benchmark of the IBSI collaboration for intensity-based radiomic features. [184].

The radiomic software does not agree with the IBSI collaboration due to the fact that it does not apply mask re-segmentation accordingly, since it is not always necessary. Thus, some of the voxels presented as outside the ROI are considered for the software inside, yielding in disagreements between comparisons.

Intensity histogram features

feature	dig. phantom	conf. C	Tübingen (dig./ conf.)
mean	2.15	-38.5 ± 0.1	(Match,)
variance	3.05	$(81 \pm 2) \cdot 10^4$	(Match,)
skewness	1.08	-2.14 ± 0.04	(Match,)
kurtosis	-0.355	3.52 ± 0.22	(Match,)
median	1	42	(Match,)
minimum	1	3 ± 0.15	(Match,)
10th percentile	1	(24.0 ± 0.6)	(Match,)
90th percentile	4	44	(Match,)
maximum	6	56.0 ± 0.4	(Match,)
mode	1	43	(Match,)
interquartile range	3	3.0 ± 0.2	(Match,)
range	5	53.0 ± 0.6	(Match,)
mean absolute variation	1.55	6.32 ± 0.13	(Match,)
robust mean absolute deviation	1.11	2.58 ± 0.13	(Match,)
median absolute deviation	1.15	4.75 ± 0.11	(Match,)
coefficient of variation	0.812	0.234 ± 0.004	(Match,)
quartile coefficient of dispersion	0.6	$(3.61 \pm 0.26) \cdot 10^{-2}$	(Match,)
entropy	1.27	3.73 ± 0.03	(Match,)
uniformity	0.512	0.139 ± 0.002	(Match,)
maximum histogram gradient	8	$(4.76 \pm 0.02) \cdot 10^3$	(Match,)

maximum gradient grey level	3	41	(No match,)
minimum histogram gradient	-50	$(-4.68 \pm 0.05) \cdot 10^3$	(Match,)
minimum gradient grey level	1	44	(Match,)

Table 6.3.: Comparisons between the software constructed for radiomics inside the framework of this thesis and benchmark of the IBSI collaboration for Intensity-histogram-based-radiomic features. [184].

Intensity-volume histogram features

feature	dig. phantom	conf. C	**Tübingen** (dig./ conf.)
volume fraction at 10% intensity	0.324	0.977	(Match,)
volume fraction at 90% intensity	$9.46 \cdot 10^{-2}$	$(1.52 \pm 0.19) \cdot 10^{-4}$	(Match,)
intensity at 10% volume	5	88	(Match,)
intensity at 90% volume	2	-421 ± 13	(Match,)
volume fraction difference between 10% and 90% intensity	2	0.997	(Match,)
intensity difference between 10% and 90%	3	510 ± 13	(Match,)
Area under the IVH curve	0.32	(0.680 ± 0.002)	(Match,)

Table 6.4.: Comparisons between the software constructed for radiomics inside the framework of this thesis and benchmark of the IBSI collaboration for Intensity-volume-radiomic features. [184].

Morphological features

Here, surface area and volume was computed in an approximate manner. Therefore disagreements are present with the IBSI collaboration since they used a mesh generated approach. In practical terms, there are not differences since great difference appears when ROIs have less than 1000 voxels, which are rare.

feature	dig. phantom	conf. C	Tübingen (dig./ conf.)
approximate volume	592	$(3.68 \pm 0.05) \cdot 10^5$	(Match, Partial match)
surface area	388	$(3.43 \pm 0.02) \cdot 10^4$	(No match, No match)
surface to volume ratio	5	88	(No match, No match)
compactness 1	$4.11 \cdot 10^{-2}$	$(3.26 \pm 0.01) \cdot 10^{-2}$	(No match, No match)
compactness 2	0.599	(0.378 ± 0.003)	(No match, No match)
spherical disproportion	1.19	1.38	(No match, No match)
sphericity	0.843	(0.723 ± 0.002)	(No match, No match)
asphericity	0.186	0.383 ± 0.004	(No match, No match)
centre of mass shift	0.672	45.6 ± 2.6	(Match, Match)
maximum 3d diameter	13.1	125	(No match, No match)
major axis length	11.4	93.3 ± 0.4	(Match, Partial match)
minor axis length	9.31	82.0 ± 0.4	(Match, Partial match)
least axis length	8.54	70.9 ± 0.4	(Match, Partial match)
elongation	0.816	0.879	(Match, Partial match)
flatness	0.749	0.76	(Match, Partial match)
volume density (AABB)	0.869	0.478 ± 0.002	(No match, No match)
area density (AABB)	0.866	0.678 ± 0.002	(Match, Match)
volume density (OMBB)	0.869	NS	(No match, NA)
Area density (OMBB)	0.866	NS	(Match, NA)
Integrated intensity	$1.2 \cdot 10^3$	$(-1.80 \pm 0.13) \cdot 10^7$	(Match, Match)
Moran's I index	$3.97 \cdot 10^{-2}$	$(8.24 \pm 0.01) \cdot 10^{-2}$	(Match,)
Geary's C measure	0.974	0.846	(Match,)

Table 6.5.: Comparisons between the software constructed for radiomics inside the framework of this thesis and benchmark of the IBSI collaboration for morphological-radiomic features. [184].

3D GLCM, with merging features

feature	dig. phantom	conf. C	Tübingen (dig./ conf.)
joint maximum	0.509	NS	(Match,)
joint average	2.15	NS	(Match,)

joint variance	3.13	NS	(Match,)
joint entropy	2.57	NS	(Match,)
difference average	1.38	NS	(Match,)
difference variance	3.21	NS	(No match,)
difference entropy	1.64	NS	(Match,)
sum average	4.3	NS	(Match,)
sum variance	7.41	NS	(No match,)
sum entropy	2.11	NS	(Match,)
angular second moment	0.291	NS	(Match,)
contrast	5.12	NS	(Match,)
dissimilarity	1.38	NS	(Match,)
inverse difference	0.688	NS	(Match,)
inverse difference normalised	0.856	NS	(Match,)
inverse difference moment	0.631	NS	(Match,)
inverse difference moment normalised	0.902	0.994	(Match,)
inverse variance	$5.74 \cdot 10^{-2}$	NS	(Match,)
correlation	0.183	NS	(Match,)
autocorrelation	5.19	NS	(Match,)
cluster tendency	7.41	276 ± 7	(Match,)
cluster shade	17.4	NS	(Match,)
cluster prominence	147	NS	(Match,)
information correlation 1	$-2.88 \cdot 10^{-2}$	NS	(No match,)
information correlation 2	0.269	NS	(Match,)

Table 6.6.: Comparisons between the software constructed for radiomics inside the framework of this thesis and benchmark of the IBSI collaboration for GLCM-radiomic features. [184].

3D GLRL, with merging features

feature	dig. phantom	conf. C	**Tübingen** (dig./ conf.)
short runs emphasis	0.729	NS	(Match,)
long runs emphasis	2.76	NS	(Match,)
low grey level run emphasis	0.607	NS	(Match,)

high grey level run emphasis	9.64	NS	(Match,)
short run low grey level emphasis	0.372	NS	(Match,)
short run high grey level emphasis	8.67	NS	(Match,)
long run low grey level emphasis	2.16	NS	(Match,)
long run high grey level emphasis	15.6	NS	(Match,)
grey level non-uniformity	281	NS	(Match,)
grey level non-uniformity normalised	0.43	NS	(Match,)
run percentage	0.68	NS	(Match,)
grey level variance	3.48	NS	(Match,)
run length variance	0.598	NS	(Match,)
run entropy	2.62	NS	(Match,)

Table 6.7.: Comparisons between the software constructed for radiomics inside the framework of this thesis and benchmark of the IBSI collaboration for GLRL-radiomic features. [184].

3D GLSZ features

feature	dig. phantom	conf. C	Tübingen (dig./ conf.)
small zone emphasis	0.255	NS	(Match,)
large zone emphasis	550	NS	(Match,)
low grey level emphasis	0.253	NS	(Match,)
high grey level emphasis	15.6	NS	(Match,)
small zone low grey level emphasis	$2.56 \cdot 10^{-2}$	NS	(Match,)
small zone high grey level emphasis	2.76	NS	(Match,)
large zone low grey level emphasis	503	NS	(Match,)
large zone high grey level emphasis	$1.5 \cdot 10^{3}$	NS	(Match,)

feature	dig. phantom	conf. C	Tübingen (dig./ conf.)
grey level non-uniformity	1.4	NS	(Match,)
grey level non-uniformity normalised	0.28	NS	(Match,)
zone size non-uniformity	1	NS	(Match,)
zone size non-uniformity normalised	0.2	NS	(Match,)
zone percentage	$6.76 \cdot 10^{-2}$	NS	(Match,)
grey level variance	2.64	NS	(Match,)
zone size variance	331	NS	(Match,)
zone size entropy	2.32	NS	(Match,)

Table 6.8.: Comparisons between the software constructed for radiomics inside the framework of this thesis and benchmark of the IBSI collaboration for GLSZ-radiomic features. [184].

3D GLDZ features

feature	dig. phantom	conf. C	Tübingen (dig./ conf.)
small distance emphasis	1	NS	(No match,)
large distance emphasis	1	NS	(No match,)
low grey level emphasis	0.253	NS	(Match,)
high grey level emphasis	15.6	NS	(Match,)
small distance low grey level emphasis	0.253	NS	(No match,)
small distance high grey level emphasis	15.6	NS	(No match,)
large distance low grey level emphasis	0.253	NS	(No match,)
large distance high grey level emphasis	15.6	NS	(No match,)
grey level non-uniformity	1.4	NS	(Match,)
grey level non-uniformity normalised	0.28	NS	(Match,)

zone distance non-uniformity	5	NS	(No match,)
zone distance non-uniformity normalised	1	NS	(No match,)
zone percentage	$6.76 \cdot 10^{-2}$	NS	(Match,)
grey level variance	2.64	NS	(Match,)
zone distance variance	0	NS	(Match,)
zone distance entropy	1.92	NS	(Match,)

Table 6.9.: Comparisons between the software constructed for radiomics inside the framework of this thesis and benchmark of the IBSI collaboration for GLDZ-radiomic features. [184].

3D NGTD features

feature	dig. phantom	conf. C	**Tübingen** (dig./ conf.)
coarseness	$2.96 \cdot 10^{-2}$	NS	(Match,)
contrast	0.584	NS	(No match,)
busyness	6.54	NS	(Match,)
complexity	13.5	NS	(No match,)
strength	0.763	NS	(Match,)

Table 6.10.: Comparisons between the software constructed for radiomics inside the framework of this thesis and benchmark of the IBSI collaboration for GLDZ-radiomic features. [184].
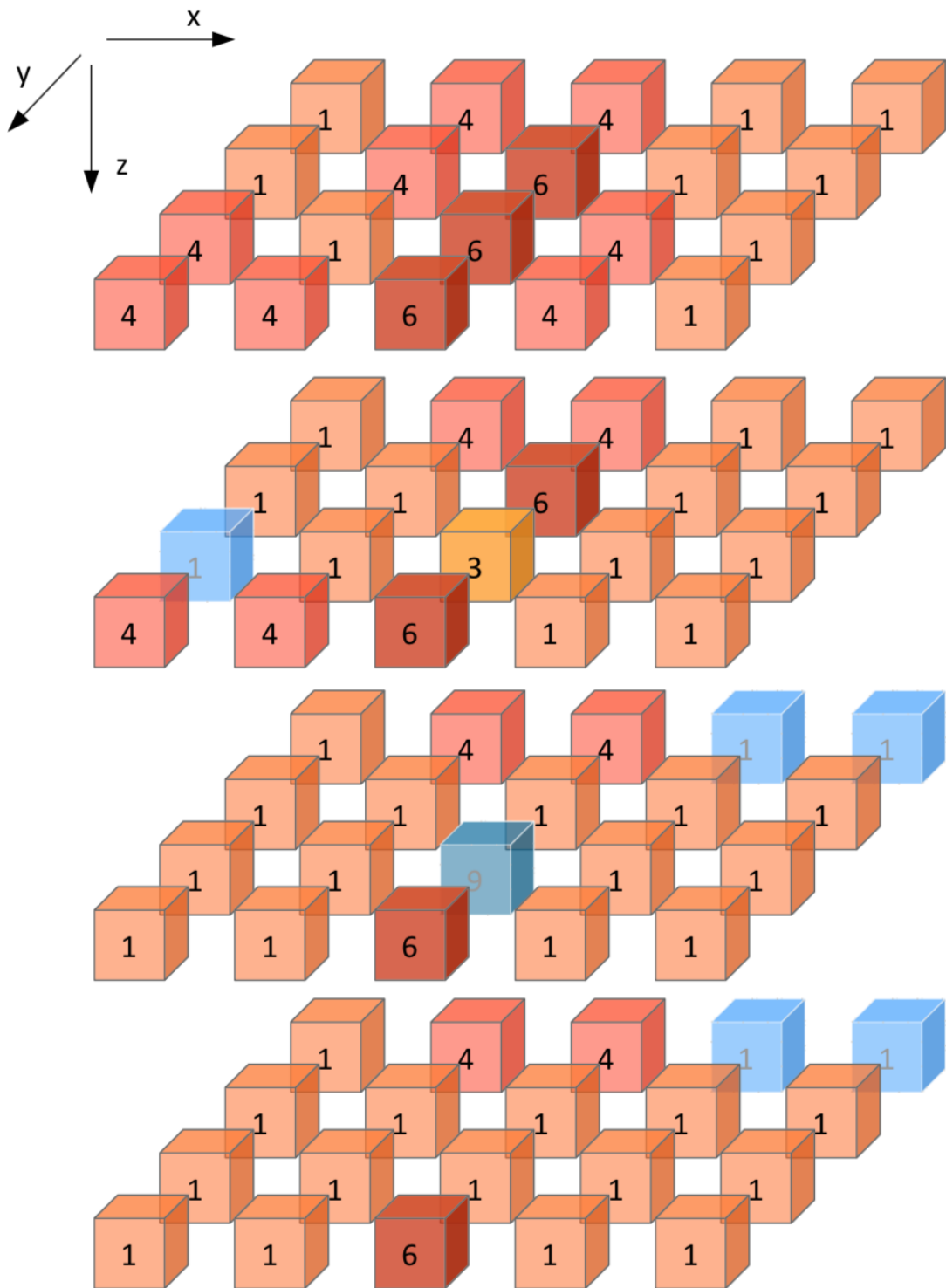
Figure 6.1.: Digital phantom array visualization. Blue voxels are excluded from the region of interest. The coordinate system is such that x increases from left to right, y increases from back to front and z increases from top to bottom, as is indicated by the axis definition in the top-left. Reprinted from [184].

	config. A	config. B	config. C	config. D	config. E
Sample	PAT1	PAT1	PAT1	PAT1	PAT1
ROI	GTV-1	GTV-1	GTV-1	GTV-1	GTV-1
Approach	2D	2D	3D	3D	3D
Interpolation	no	yes	yes	yes	yes
Voxel dimension (mm)		2×2	$2 \times 2 \times 2$	$2 \times 2 \times 2$	$2 \times 2 \times 2$
Interpolation method		bilinear	trilinear	trilinear	tricubic spline
Grey level rounding		nearest integer	nearest integer	nearest integer	nearest integer
ROI interp. method		bilinear	trilinear	trilinear	trilinear
ROI partial volume		0.5	0.5	0.5	0.5
Re-segmentation					
Range (HU)	$[-500, 400]$	$[-500, 400]$	$[-1000, 400]$	no	$[-500, 400]$
Outliers	no	no	no	3σ	3σ
Discretisation					
Texture and IH	FBS: 25 HU	FBN: 32 bins	FBS: 25 HU	FBN: 32 bins	FBN: 32 bins
IVH	no	no	FBS: 2.5 HU	no	FBN: 1000 bins

Figure 6.2.: Radiomic phantom configuration characteristics for imaging preprocessing. IH: Intensity histogram; FBS: Fixed bin size; FBN: Fixed bin number; IVH: Intensity volume histogram; NA: not applicable. Reprinted from [184].

7. Radiomics pipeline inside this thesis

7.1. Imaging pre-processing and feature extraction

For the radiomics analysis, CT images are used without voxel resizing, in order to avoid inclusion of artificial information that might cause noise at the moment of feature calculations. Only soft tissue voxels with values between -250 and 120 HU are considered in order to exclude air, and other non-tissue elements. 64 bins were used to group voxel values for texture feature calculations.

Features definitions are obtained from the Imaging Biomarker Standardisation Initiative (IBSI) [184]. For the texture features, we used the grey-level co-occurrence (GLCM), grey-level run length (GRLM), neighbourhood grey tone difference (NGTDM), grey-level size zone (GLZSM) and grey-level distance zone (GLDZM) matrix. They are computed in 3 dimensions regardless of differences between in-plane and in-slice voxel dimensions. One level undecimated wavelet features are obtained as follows. Firstly, the original images are filtered using high (H) or low-pass (L) "Coiflet 1" filter in every image (x, y, z) direction. Different filter combinations resulted in 8 filtered images. Subsequently, intensity and texture features are computed for each filtered image [60]. In summary, we extract 1150 radiomics features from GTV regions contoured in the planning CT scans. All filtering and feature computations were implemented in-house in Python 3.6.

7.2. Feature pre-processing

Several of the radiomics features described by the IBSI are highly correlated and therefore redundant. Hence, in the training phase, we cluster correlated features (more than 95% correlation in Pearson correlation coefficient), in order to optimise the feature selection process. To do so, features are first scaled according to the quartile range (interquartile range, IQR), which ranges between the first quartile (25% quantile) and the third quartile (75% quantile). Then, they are clustered hierarchically according to Pearson correlation coefficient [170]. Finally, every cluster is reduced to one single feature using principal component analysis (PCA) to conserve the maximum possible variance inside the cluster [81]. Moreover, all features with variance lower than 0.3 are excluded from the final feature set.

7.3. Feature selection & model tuning

According to Leger et al. [101] feature selection methods play a more important role in predicting outcomes than the models themselves. Therefore, a four-step feature selection method was implemented as follows:

Step 1: The training cohort is randomly subsampled with replacement in a balanced fashion so that each subsample contains the same number of samples with one label class (i.e. poor prognosis) as the samples of patients with the other label class (i.e. good prognosis). This was repeated 100 times, thus creating a set of 100 subsamples.

Step 2: Within each subsample, variable importance was determined using:

- correlation measures (Pearson [14], Kendall [3], Spearman [1]),

- mutual information (mutual information maximisation [134]),

- univariate significance test scores (Fischer, χ^2 [76]),

- multivariate forward selection using classification models (decision trees (DT), k-nearest neighbours (KNN), logistic regression (LogR), random forest (RF), naïve-Bayes (GNB), support vector machines (SVM) [19]) based on the model Receiver Operating Characteristic Curve - Area under the curve (ROC-AUC) score [75].

For all methods above, up to twenty most important features are kept, and the remaining features are discarded. These feature subsets are then aggregated across the different methods to form a final subset of the five most commonly occurring features for each subsample.

Step 3: The features in the final subset of each of the 100 subsamples are then aggregated and heuristically ranked using the following scoring:

$$\text{RS} = \frac{n_a}{100} \frac{1}{\mu_r(\sigma_r + 1)}$$

The scoring favours the number of appearances (n_a) of a feature in the 100 subsets, and penalises its mean rank μ_r together with the standard deviation of its rank μ_r in the subsets. The five most highly ranked features are subsequently selected.

Step 4: We determine the signature for each of the classifiers using a sequential forward feature selection method [136]. For this purpose, we perform 5-fold cross-validation using the training data set. For each classifier, the set of features that produced the model with the highest average AUC on the validation folds was used as a signature.

After feature selection, model hyperparameters such as the number of neighbours for KNN are optimised using grid search (cf. Table 7.1) and 5-fold cross validation. All methods and algorithms are implemented in-house in Python 3.6 using the packages Pandas, Scikit-learn and mlxtend for machine learning. Figure 7.1 presents a schematic overview of the algorithmic workflow used in this thesis.
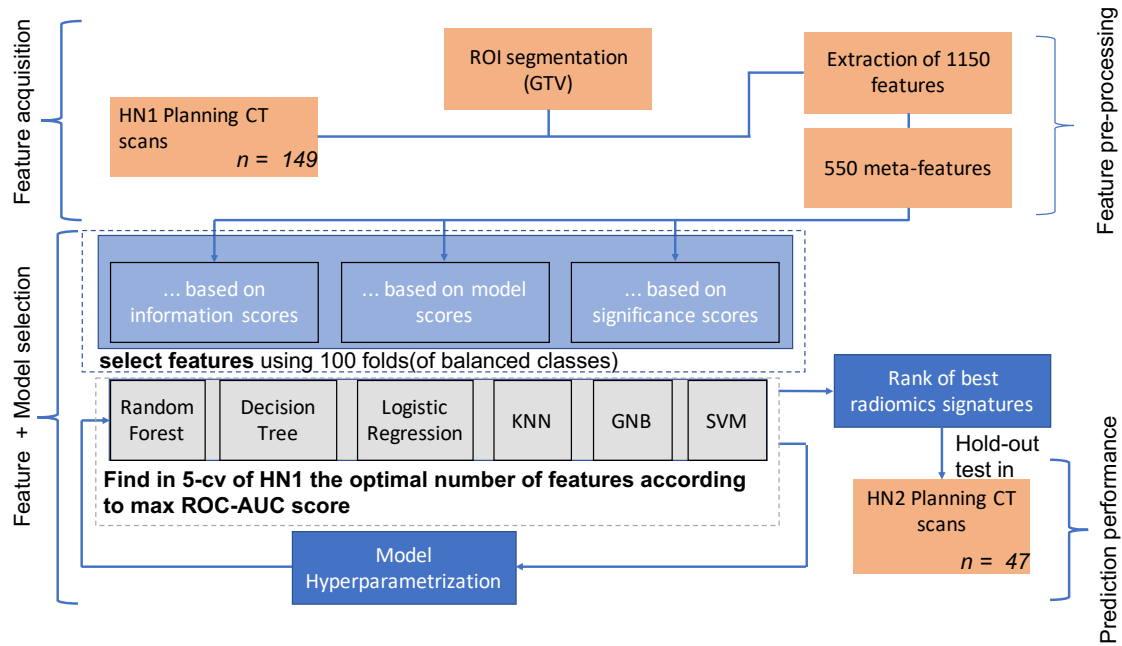
Figure 7.1.: Workflow of the CT radiomics model construction. The first part consists of feature extraction process which after acquisition, the features requires an exploratory analysis for searching redundancies and subsequently, obtain the final set of meta features (cf. section CT radiomics). Then, the training cohort is subsampled to be integrated in the process of feature and model selection where features are ranked based on information, model performance and significance scores using 100 subsamples of the training cohort, a heuristic score is then set to choose the best and most stable features. Finally, the best CT radiomics signatures are chosen according to ROC-AUC score, then hyperparameters chosen to improve ROC-AUC in training and tested in the test cohort to prove findings.

Model	Hyperparameter	Hyperparameter space
Logistic Regression	Penalty	'L1', 'L2'
	C	10-3, 10-2, 10-1, 1, 10, 102, 103, 104
	Class Weight	{0: 0.3}, {0: 0.4}, {0: 0.5}, None
Nearest Neighbours	Number of Estimators	3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25
	Weights	Uniform, Distance
Support Vector Machine	C	10-3, 10-2, 10-1, 1, 10, 102, 103, 104
	Kernel	RBF, Polynomial, Sigmoid
Random Forest	Criterion	Gini, Entropy
	Number of Estimators	3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25
	Class Weight	{0: 0.3}, {0: 0.4}, {0: 0.5}, None
	Max Depth	8, 9, 10, None
Gaussian Naïve Bayes	None	None
Decision Tree	Criterion	Gini, Entropy
	Class Weight	{0: 0.3}, {0: 0.4}, {0: 0.5}, None
	Max Depth	8, 9, 10, None

Table 7.1.: Description of the hyperparameter space used for each Machine Learning model. In class weight the number 0 represents the negative class (no recurrence event), as the negative class is less important to predict as the positive class, low values, meaning prediction importance, were given for the negative class in the hyper-parametrization.

8. Correlations of the Aerts signature with somatic mutations in TP53, FAT1 and KMT2D in HNSCC

The publication of Aerts et al [5] investigated 440 CT based radiomic features including intensity, shape, texture and multiscale wavelet in lung cancer and HNSCC for prognostic value [5]. After training and validation, the best performing features of each category were extracted for lung cancer and validated in HNSCC, in a second step, the identified CT radiomic signature was correlated with gene expressions in HNSCC associated to tumour phenotype heterogeneity. An association between enriched cell-cycling pathways indicating proliferation and increased intra-tumour heterogeneity features (texture and wavelet feature) has been found.

The impact of such findings implies the potential to determine functional changes and risk groups of HNSCC patients [164]. On the other hand, the Cancer Genome Atlas (TCGA) characterised several frequent somatic variants in HNSCC including TP53 (cell cycle control and survival), KMT2D (chromatin modification) and FAT1 (Wnt / β-catenin signalling, cell-cell contacts, cell orientation, cell fate) [100, 121]. As a cross-link to clinical features, variants in TP53 and FAT1 were predominantly found in human papillomavirus (HPV) negative tumours [121]. In HPV negative patients, variants in FAT1 were reported to be associated with beneficial outcome in surgically treated HNSCC patients [87]. In this chapter, we investigated the reproducibility between FAT1 and other gene-expression mutations of tumour cells found International Cancer Genome Consortium (ICGC) with the radiomic signature of the study of Aerts et al (2014) [5].

In the aim of reproduce the work of Aerts et al [5] for HNSCC patients, an hypothesis driven strategy was used, in which the identified CT radiomics signature in the mentioned study was explored to find correlations with the genetic aberration expressions associated to be highly prognostic to categorised heterogeneity in HNSCC types of tumours of the TCGA [186].

8.1. Patients and diagnostics

Twenty patients with locally advanced HNSCC were recruited for this prospective biomarker study. All declared their written informed consent and the study was approved by the local ethics committee (reference number 577/2014BO2). All patients were treated with definitive radiochemotherapy up to 70 – 77Gy. Human papillomavirus (HPV) association was investigated by immunohistochemical staining for p16 or PCR-based assays. Clinical data was extracted from the medical reports [186].

8.2. CT-radiomic feature extraction

Based on the report by Aerts et al. (2014) [5], our first tested hypothesis postulated that cell cycle alterations by somatic mutations (driver gene variants; TP53) might correspond with the CT-radiomic features associated in the study with poor prognosis. As a second hypotheses, we investigated if other frequently mutated driver genes

correlate with heterogeneity features of the tumour.

Based on our unenhanced planning CT scans (Somatom Sensation Open, Siemens Healthineers, Erlangen, Germany; slice thickness of 3mm, in-plane pixel size of 1.27mm, ordered subset expectation maximization (OSEM) 3D (four iterations, eight subsets) with a 3D Gaussian filtering for imaging reconstruction), we analysed the two correlated radiomic features for measuring intra-tumour heterogeneity that were described by Aerts et al. (2014) [5], namely "Run Length Nonuniformity" and "wavelet Grey Level Nonuniformity HLH". Furthermore, we included "Grey Level Nonuniformity" (Aerts et al.: Texture Feature 47), as a complementary feature, as the authors also reported on this feature in the referenced publication. Therefore, in total, three particular heterogeneity features were investigated and the features were calculated following the previous report of Aerts et al. [5] for confirmation and standardisation. The GTV of the primary tumours were delineated for treatment planning by experienced radiation oncologists. These delineations were subsequently used for radiomic analyses. Due to concerns regarding the influence of dental artefacts [103], we investigated both, the data of all 20 patients and, as a subgroup, the patients that had no CT artefacts in the area of interest. Texture features were preprocessed in a 3D fashion regardless of the in-plane, in-slice length differences and we grouped the intensity values in 64 different bins due to the sparse range of intensity values (between -250 to 120 Hounsfield units) across the GTV for texture feature acquisition purposes. For wavelet estimations we used the undecimated wavelet filter (coiflet) [62]. If air or bony structures were included in the GTV, the delineations were adapted and extreme Hounsfield units were excluded for radiomic feature extraction. Thereby, solely in one patient, the GTV was considerably modified due to massive air and bone involvement (oropharyngeal HNSCC with infiltration of the maxillary sinus) [186].

8.3. Statistic analysis

For the statistic analyses, we used R and SPSS (IBM Corp., Armonk, NY). The Mann-Whitney U test and robust linear regression (M-estimator) from MASS R package (logarithmic scale) were used for calculations. Significance estimations of regression coefficients were calculated by the robust F-test (Wald test, sfsmisc R package). A p-value < 0.05 was considered significant [186].

8.4. Correlation analysis

No definitive evident of the Aerts signature was found to indicate tumour heterogeneity, that means that low correlation was obtained with variants in TP53 or KMT2D (data not shown). Nevertheless, a significant association with FAT1 was encountered, as variants in FAT1 corresponded with reduced expressions of the mentioned CT radiomic features of the primary tumour (Grey Level Nonuniformity: $p = 0.019$; Fig.

Age		
Median	60	49 - 75
Gender		
Male	18	90%
Female	2	10%
HPV Status (n,%)		
Positive	5	25%
Negative	14	70%
Unknown	1	5%
Tumour localisation (n, %)		
Oral cavity	2	10%
Oropharynx	14	70%
Hypopharynx	4	20%
Smoking History (n, %)		
Yes	17	85%
No	3	15%

Table 8.1.: Clinical characteristics of patients.

1 A of the study of Zwirker [186]; Run Length Nonuniformity: $p = 0.046$; Fig. 1 B of [186] and wavelet Grey Level Nonuniformity HLH: $p = 0.035$; Fig. 1 C of [186]). This association was found in all three selected CT-radiomic features and the observation remained significant in two features, when patients were excluded, who had dental artefacts in the area of the primary tumour.The association of somatic mutations in FAT1 and smaller GTVs of the primary tumour was not significant, yet prognostic ($p = 0.059$; Fig. 2 A of [186]). However, smaller GTVs of the primary tumour corresponded with reduced expressions of the investigated radiomic features.

We found an association between FAT1 mutated tumours and reduced heterogeneity of the primary tumours according to radiomic features. Reduced heterogeneity corresponded with smaller primary tumour volumes and in FAT1 mutated tumours, a trend towards reduced primary tumour volumes was observed, though significance levels were not reached. Postulating FAT1 being a tumour suppressor in HNSCC [100], inactivating mutations would be expected to cause rather extended volumes. However, the influence of FAT1 on proliferation in oral squamous cell carcinomas was described to be rather limited [124]. In this way, other oncogenes and tumour suppressors might have comparably stronger influence on tumour growth and GTV extent.
As discussed above, inactivating / missense variants of FAT1 might result in reduced invasiveness, attenuated migration and looser cell-cell contacts. One could speculate that this translates into less expression of the radiomic features discussed across this study and smaller tumour volumes as indicated in our cohort, One can also notice that by definition, the discussed radiomic features are correlated with volume despite

they do not necessarily mean the same measurement, since definitions on texture features refer to counting the number of patterns, whose probability for large volume are higher, nevertheless small yet highly heterogeneous tumours translate it into more expression in the investigated radiomic features. As increased heterogeneity correlated with poor outcome [5] and smaller GTVs [90] are associated with a good prognosis, our findings might support to a recent publication reporting favourable outcomes of HPV negative, surgically treated HNSCC patients, if they presented with mutant FAT1 [87]. Thus, our data suggest a possible interrelation between FAT1 mutations, reduced heterogeneity and smaller GTVs.

Our study revealed interesting preliminary findings in HNSCC patients. The limitation of our study is the small cohort and resulting limited effect size. Therefore, our radiogenomic observations remain solely descriptive and need to be confirmed in larger studies. For validation and prognostic models, bigger prospective cohorts would be essential. A further limitation is the localisation of the biopsy. We sequenced FFPE material that was collected at primary diagnosis. The exact localisation within the primary tumour is therefore not known. In case of intra-tumour genetic heterogeneity, some variants might be missed and others overestimated. However, we chose only the most frequently mutated driver genes for correlations. As these are well in line with previously reported drivers of HNSCC, these mutations are thought to determine relevant biological and functional variations that might translate into radiomic features.

9. CT radiomics is not a surrogate predictor of patients at risk as identified by [^{18}F]FMISO hypoxia PET

Tumour hypoxia has been shown to be prognostic for poor outcome after CRT in HNSCC by several studies [110, 175, 180, 183]. However, biological heterogeneity as identified by radiomics and others factors have also been linked to poor outcome after chemoradiation therapy [71, 72, 177]. Hypoxia can be measured invasively using probes, or assessed non-invasively using specific radiotracers in PET imaging [161]. In clinical research, the most commonly used hypoxia imaging tracer is ^{18}FFMISO [48, 91]. Tumour-to-muscle ratio (TMR) is a simple but very robust metric that is commonly used to derive the hypoxic status from PET data, though other methods exist as well [73, 161, 183]. Different studies have shown that TMR assessed 2 to 4 hours after tracer injection enables differentiation between hypoxic and normoxic tumours based on a threshold value (e.g. $TMR_{peak} \geq 1.4$) [91, 110, 117], and consequently to distinguish patients at increased risk of loco-regional failure (LRF) at different time points of CRT [110, 183].

Radiomics, which is a technique for quantitative analysis of medical images, hypothesises that imaging features capture anatomical or functional tumour heterogeneity in solid tumours without the need for additional diagnostic interventions such as biopsies [98]. Different research teams have shown not only a significant prognostic power of radiomics features and signatures in the task of patient stratification for LRF in patients after CRT but also correlations with gene expression in different forms of cancer [13, 86], [5, 55, 82, 104, 145, 171]. Therefore, some authors hypothesised that CT radiomics captures tissue heterogeneity caused by tumour hypoxia [5, 56, 98, 102].

If the hypothesis is true, CT-radiomics should be able to identify similar at-risk patient subgroups as identified by [18F]FMISO TMR up to certain extend, considering that poor prognosis after cancer treatments is a confluence of many different factors. However, this hypothesis might yield to more affordable and eventually less complicated strategies to adapt medical treatments in tumours, since it would allow the exchangeability of [18F]FMISO imaging markers and CT-radiomics signatures for treatment outcome prediction.
The underlying hypothesis for the current study was that an independently trained CT radiomics model might serve as surrogate for hypoxia PET imaging to stratify patients into risk groups according to outcome after RCT of HNSCC. Ideally, a CT radiomics signature might be able to capture similar risk profiles as hypoxia imaging using FMISO PET. To investigate this hypothesis, a CT radiomics model was first developed based on n=149 HNSCC and subsequently validated with an independent, bi-institutional data set of n=47 patients for whom FMISO PET data were also available to compare the potential of CT radiomics versus FMISO PET imaging for patient stratification.

	H&N1	H&N2
Number of patients	149	47
Age (mean, range)	62 (39 - 87)	58 (45 - 76)
GTV-CT volume (mean, range)	61.6(1.4 - 326.7) cm^3	62.7 (10.4 - 238.8) cm^3
Gender (female/male)	25/124 (16.8%/83.2%)	7/40(14.9%/85.1%)
Number of locoregional failures	50 (34%)	15(32%)
Median follow-up time (median, range)	12 (0 - 82) months	17 (1 - 75) months
Distant Metastases	26 (17%)	7 (15%)
T-stage (Tis/T1/T2/T3/T4)	1/1/17/46/84	0/0/2/19/26
N-stage (N0/N1/N2a/N2b/N2c/N3)	20/14/46/3/55/11	5/4/7/16/13/2
Radiation dose (mean, range)	70 (66-72) Gy	71(69 -72) Gy
Chemotherapy		
5-FU/MMC	116 (77.8%)	25 (53.2%)
Cisplatin	16 (10.7%)	1 (2.1%)
Cisplatin/5-FU	3 (2.0%)	21 (44.7%)
Other	14 (9.4%)	0

Table 9.1.: Patient population characteristics

9.1. Data sets

The data set consisted of 196 patients in total with HNSCC in advanced stages scheduled for definitive CRT who had been recruited in a period of 10 years (from 2005 to 2015) at the University Hospital Tübingen (UHT, n=171) and the University Hospital Dresden (UHD, n=25). The patient cohort was regrouped into two cohorts: HN1 (n=149 all from UHT), where only native radiotherapy (RT) planning CT data with delineations of gross tumour volumes (GTV) by an experienced radiation oncologist was available. For HN2 (n=47), in addition to native planning CT images and GTV delineations also FMISO PET/CT data were available at baseline before the start of treatment [175, 183]. At both hospitals, patients were treated with definitive CRT with a mean radiation dose of 70 Gy, in addition to fluorouracil (5-FU) and mitomycin (MMC) or concomitant weekly cisplatin. For further details refer to table 9.1.

9.2. Imaging data

For all patients of subgroup HN1, native planning CT scans were acquired using a Somatom Sensation Open (Siemens Healthineers, Erlangen, Germany). In subgroup HN2, patients also received a planning CT. In addition, [18F]FMISO PET/CT scans were acquired using a Siemens Biograph 16 (UHD, UHT) or a Siemens Biograph mCT. PET data were reconstructed using OSEM 3D (four iterations, eight subsets) with a 5-mm 3D Gaussian filtering the [18F]FMISO PET/CT acquisition protocol consisted of static scans acquired four hours post injection with injected activities of 250 – 444 MBq. For further details see table 9.2. For the subsequent analysis, original GTV delineations from RT treatment planning were used.

9.3. CT-radiomics signature

In this study we implemented the pipeline described in chapter **??** for feature selection, where we used HN1 as our training cohort and HN2 as our test cohort. We tested the models created using the signature obtained in our training cohort (HN1) for each classifier in the HN2 cohort as (see table **??**bsequently the model in our training phase was used to stratify patients into high and low risk groups at a 0.5 risk threshold probability.

9.4. FMISO PET/CT Imaging

9.4.1. Tumour-to-muscle ratio extraction

TMR$_{\text{peak}}$ values were extracted from ^{18}F-FMISO PET/CT scans according to:

$$TMR_{max} = \frac{SUV_{max}}{SUV_{muscle-mean}} \tag{9.1}$$

Peak values of FMISO standardised uptake values in the GTV were determined by averaging voxels represented in 0.5 cm3 sphere of highest tumour uptake (SUV$_{\text{peak}}$) as described in Zips et al [183], Löck et al [110] and Mönnich et al. [117]. The mean muscle standardised intensity uptake value (SUV$_{\text{muscle}}$) was obtained from manually contoured regions of deep neck muscles.

9.4.2. Model validation for TMR$_{\text{peak}}$

Model validation was performed in HN2 where the TMRpeak was used to classify tumours into high and low risk groups based on the 1.6 threshold obtained in the study of Löck et al. [110].

Modality	Scanning parameters	HN1	HN2
CT	Scanners	Siemens Somatom Sensation Open	Siemens Biograph (n = 36) Siemens Biograph mCT (n = 11)
	Slice thickness [mm]	3	3 (n = 22), 5 (n = 25)
	In-plane resolution [mm]	1.27	1.27 (n = 22), 1.38 (n = 25)
	Tube Voltage [kVP]	120	120
	Tube Current [mA]	40	100 (n = 25), 40 (n = 22)
	Reconstruction Kernel	Convolution kernel B40S filtered back projection	Convolution kernel B40S filtered back projection
PET	Scanners		Siemens Biograph (n = 36) Siemens Biograph mCT (n = 11)
	Slice thickness [mm]		5
	In-plane resolution [mm]		1.38 (n = 25), 2.42 (n = 22)
	Administrated [18F]FMISO activity [MBq]		250 − 300 (n = 25), 315 − 444 (n = 22)
	Reconstruction kernel		5-mm Gaussian filter OSEM3D 4 integration 8 subsets
	Scan duration time		15-min bed position (n = 22), 12-min bed position (n = 25)
	Attenuation correction		Based on CT
	Standard Uptake Value (SUV) normalisation		Body weight

Table 9.2.: Details of CT and PET imaging data

F. selection criteria	Model	Num. meta features	Name of the associated features in clusters	Hyperparam.	AUC HN1	AUC HN2
RF	KNN	2	Size Zone: LZHGLE', 'LHH Minimum Histogramm Gradient'	number of neighbors: 25	0.76 ±0.09	0.59
				weights: distance		
RF	RF	4	Area under IVH curve', 'HHH RL: LGLRE', 'HHL Intensity histogram median', 'LLH SZ: LZLGLE'	Class weight: {0: 0.5}	0.75 ±0.07	0.56
				Criterion: 'Entropy'		
				Max depth: 10		
				Number of estimators: 9		
DT	RF	3	NGTD: Busyness', 'NGTD: Busyness', 'LHL Median'	Bootstrap: False	0.75 ±0.10	0.59
				Class weight: None		
				Criterion: 'Gini'		
				Max depth: None		
				Number of estimators: 10		
χ^2	KNN	5	'HLH Energy', 'LLH Energy', 'LLL SZ: LZHGLE', 'LLL SZ: LZE', 'LHH SZ: ZS non-uniformity'	number of neighbors: 23	0.74 ±0.10	0.52
				weights: distance		
KNN	LR	4	'LLL LZE', 'HHL Intensity histogram median', 'LLH Range', 'HLH Energy'	C: 1000	0.71 ±0.10	0.53
				Class weight: {0: 0.5}		
DT	LR	3	'HHL Intensity histogram median', 'LLL SZ: LZHGLE', 'LHL DZ: ZD non-uniformity'	C: 1.0	0.70 0.09	0.52
				Class weight: None		

Table 9.3.: Description of the best performing CT radiomics signatures and models according to ROC-AUC score in HN1 and HN2, Random Forest (RF), Decision Trees (DT), k-nearest neighbours (KNN), Logistic Regression (LR).

9.5. Model Comparison

In order to assess whether the patients at risk classified by the best CT radiomics signature are similar to the classified patients at risk based on TMRpeak, the following simple matching score (MS) was used:

$$MS = \frac{TP + TN}{NT}$$

MS measures the ratio between the number of patients that both models predict either as patients at high risk (TP) or as low risk patients (TN) divided by the total number of patients (NT) in HN2. If MS equals 1, it means that both modalities predict the same treatment outcome for a patient, whereas 0 indicates complete disagreement.

9.6. Statistical Analysis

Stratification of patients into risk groups for LRC was assessed using Kaplan-Meier curves and the log-rank test. All statistical analyses were performed using the *lifelines* package implemented in Python 3.6. A p-value < 0.05 was considered as significant.

9.7. Are there correlations?

Following model training in the HN1 cohort, the six best models had AUC-ROC values ranging between 0.70 ± 0.09 and 0.76 ± 0.09. The best performing CT radiomics model was a 25-Nearest Neighbours model based on two radiomics meta-features associated to 'LLL Size Zone (SZ): Large Zone High Grey Level Emphasis' and 'LHH Minimum histogram gradient' [184] (cf. table 9.3). However, in the HN2 validation cohort the AUC-ROC of the models decreased to a range between 0.52 and 0.59 using a 0.5 threshold for risk classification. Stratification of the validation cohort HN2 into high and low risk patients failed according to this CT radiomics model, underlined by a p-value=0.18 in the log-rank test (cf. figure 9.1 (a)).
On the contrary, in the same HN2 cohort, the FMISO/PET TBRpeak imaging marker resulted in a ROC-AUC score of 0.66 using a threshold of 1.6, as proposed for the exploratory cohort (n = 25) in the study of Löck et al. 2017 [110]. Likewise, in the same cohort a better stratification was achieved by TBR_{peak} using the log-rank test (p-value=0.02, cf. figure 9.1 (a)). The threshold proposed by Mönnich et al. 2015 ($TBR_{peak} = 1.88$) was also used, which yielded a ROC-AUC score of 0.62 similar to the threshold of Löck et al. 2017. Nevertheless, the stratification of the HN2 cohort into high and low risk patients failed (p-value = 0.12, log-rank test).

In summary, the CT radiomics model does not perform better than TBR_{peak} in stratifying the HN2 cohort according to LRF. An accuracy of MS = 55.3% was obtained between the two models, which suggests that there is a weak correlation between the

CT radiomics signature classification and the risk classification of patients by FMISO TMRpeak (cf. figure 9.1 b). The CT radiomics features encountered in this study, are assumed to quantify pattern-variation-values of heterogeneity in a LLL and LHH frequency filtered tumour in a volumetric image [184]. As an example, in figure 9.2, an irregular pattern-structure variation is distributed homogeneously across the ROI, while in figure 3a, the pattern-structure variation in the ROI is rather low, but equally distributed within the ROI.
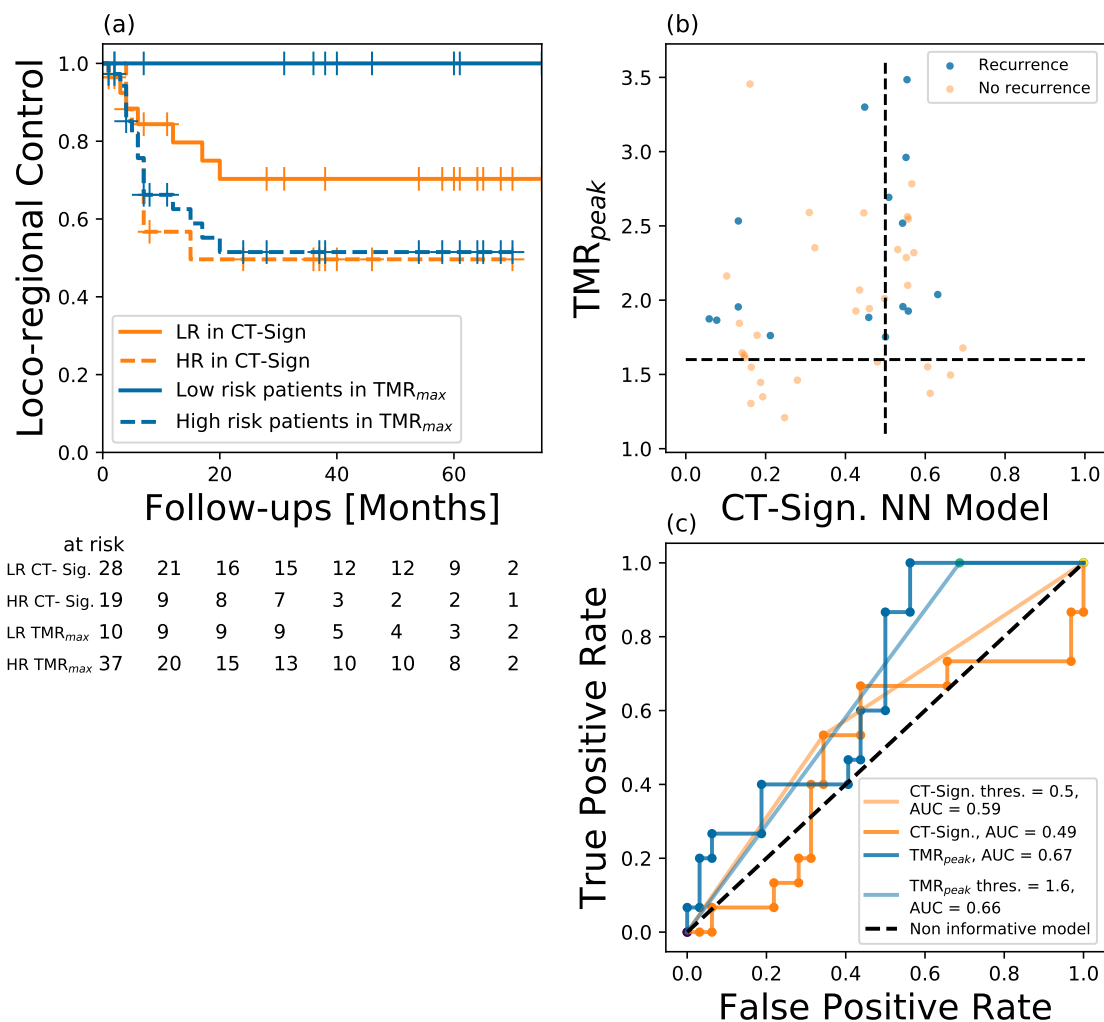
Figure 9.1.: (a) Kaplan-Maier curves for $\text{TMR}_{\text{peak}} > 1.6$ (p-value=0.02) in comparison to the best-performing CT radiomics signature using the 0.5 threshold to stratify patients at risk (p=0.18). (b) Patient classification according to CT radiomics signature score (x-axis) and TMR_{peak} (y-axis), yielding a matching score of 0.553. (c) Receiver Operator Characteristic (ROC) for predicting recurrence and non-recurrence with both models. CT radiomics signature ROC-AUC = 0.59 and 0.49 using the 0.50 threshold and for floating thresholds respectively, TMR_{peak} ROC-AUC = 0.66 and 0.67 in HN2 using 1.6 threshold for floating thresholds respectively at-risk classification tasks.
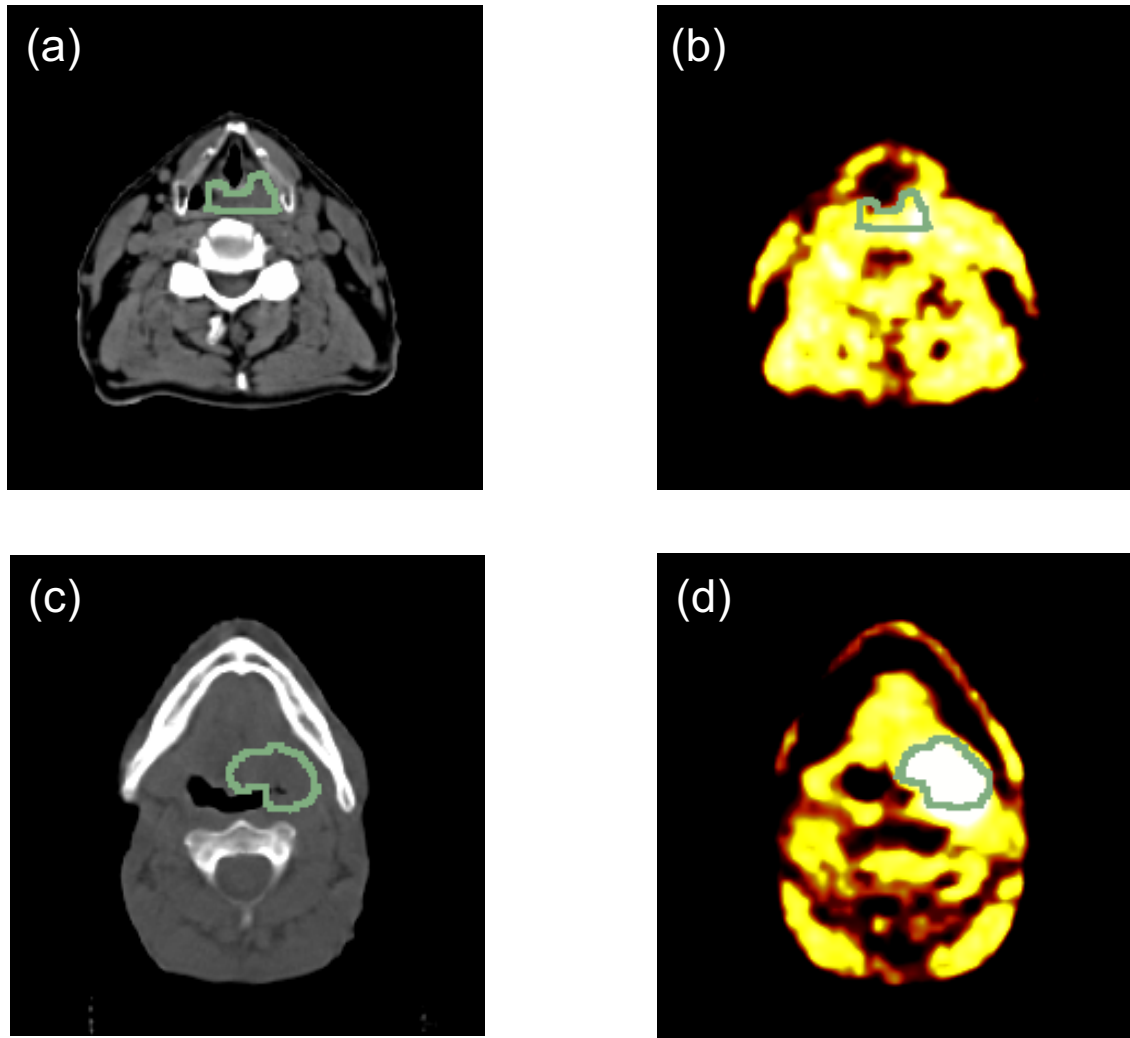
Figure 9.2.: Image (a) is a planning CT scan with (b) the [18F]FMISO PET scan after 4h post injection and their ROIs of a patient who did not recur after chemo-radiotherapy, Images (c) and (d) are the planning CT scan and the [18F]FMISO PET scan after 4h post injection and their ROIs respectively of a patient who had a recurring tumour after chemo-radiotherapy, in the ROIs of the PET scans, the low risk patient shows lower TBR$_{peak}$ (1.44) and CT radiomics signature model probability (0.18) than the high risk patient (1.96 and 0.54 respectively).

10. A radiomic approach for prediction of complete remission of organ-preserving therapy in rectal carcinoma with independent validation

Recently, locally advanced rectal carcinoma treatments have achieve great results in terms of local control rates in phase II (up to 98%) [142]. However those figures are only achieved with very aggressive therapies in surgery [70,129], and in chemoradiotherapy [54,149]. The drawback of such strategies jeopardizes considerably the organs in the neighbourhood of the colon [12,30]. The prediction of patients who might be candidates for less aggressive therapies remains as one of the biggest challenges in the field of organ-preserving strategy of therapies [51,65,130,148]. Two main approaches have been addressed, firstly, either the omission or limitation of radiation therapy based on pretreatment MRI parameters and secondly, the omission or limitation of surgery after response of chemoradiotherapy.

Radiomics can potentially improve organ-preserving therapies [51], due to correlations of imaging features with the underlying biology in tumours [5,97,103], which could allow better subgroups of therapy responders. Up to date, there are limited but promising approaches on the field [35,123] for instance Ke Nie et al, achieved and ROC-AUC of 0.89 for predictions of good responders in a cohort of 48 rectal cancer [123]. This study aims to assess prediction performance of good vs bad responders of radiomics for potential use in organ-preserving therapies by using a significant cohort of patients (TUE, n = 136) from the University Hospital Tübingen with fully independent validation cohort from University Hospital Florence (FLO, n = 76).

10.1. Patients and diagnostics

Consecutive patients treated for histologically-confirmed, locally advanced adenocarcinoma of the rectum at the Universities of Tübingen (TUE) and Florence (FLO) in two subsequent time frames (1/1/07 to 31/12/10 and 1/1/11 to 31/12/17 in TUE and FLO, respectively) were considered for our analysis. In general, staging included Gadolinium-enhanced pelvic MR, iodinated contrast-enhanced CT of the chest and abdomen, and colonoscopy. Clinical stage was defined according to UICC/TNM 7th edition. After multidisciplinary discussion, all patients with T2-T3, N positive rectal cancer deemed amenable to undergo a full course of pre-operative radiation-based treatment followed by curatively-intended surgery could be included in our study. No tumour upper distance limit from the anal verge was specified. No upper age limit was defined. Neoadjuvant chemotherapy, unresectable primary tumour (T4), previous RT to the pelvis or previous surgical manipulation of the rectum were exclusion criteria. In addition, patients with unrecognisable rectal GTV on the planning CT or with image artifacts induced by hip prosthesis or rectal stent could not be included. As per local practice, standard of care for neoadjuvant treatment differed between the two centres. In TU, 50.4 Gy were delivered in 28 fractions of 1.8 Gy each (5 fractions per week). Radio-sensitizing chemotherapy consisted of 120-hour continuous infusion of 5-fluorouracil during the first and fifth weeks of radiation (daily dose of 1000 mg/m2 on days 1 through 5 and 29 through 33, respectively). Selected

patients received also deep regional hyperthermia, which was administered with a Sigma Eye or Sigma-60 applicator once or twice weekly for at least 60 minutes to a target temperature of 40.5° Celsius, as previously described [152], [52]. In FL, a total dose of 45 Gy was delivered with standard fractionation over 5 weeks (25 fractions of 1.8 Gy per day). Chrono-modulated capecitabine was prescribed for the whole RT course at a daily dose of 825 mg/m2 BID. After restaging and due time interval (usually 6 to 10 weeks from the end of CRT), surgery was performed. Abdomino-perineal resection or rectal anterior resection (RAR) with TME were the procedures of choice. Pathologic response evaluation was assessed in accordance with Dworak's tumor regression grade40 in both institutions. Dworak's 5-point scale was as follows: 0, 1, 2, 3 and 4 scores were indicative of no regression, predominantly tumor with significant fibrosis and/or vasculopathy, predominantly fibrosis with scattered tumor cells, only scattered tumor cells in the space of fibrosis with/without acellular mucin, and no vital tumor cells detectable, respectively. Capecitabine or 5-fluorouracil (plus folinic acid) - based, adjuvant chemotherapy was offered to selected patients in case of unfavourable pathologic findings.

10.2. Imaging analysis and protocols

In terms of CT acquisition, treatment planning and delivery, the following procedures were performed, according to local standard of practice. A CT scan (Big Bore, Philips Medical Systems, Cleveland, OH, USA) was acquired at 3 mm slice width for planning purpose. Most patients were immobilized in the prone position with an ankle-holder. In order to displace the small bowel loops from the irradiation field, a belly board device was used. RT was delivered by a linear accelerator (Elekta, Crawley, UK) with standard 3-field box technique or intensity modulated radiotherapy. In terms of delineation, the same procedures were followed in both institutions. The following organs at risk (OAR's) were contoured: femoral heads, bladder, small bowel, penile bulb and anal canal (if not infiltrated). Typically, the clinical target volume (CTV) consisted of the mesorectum and internal iliac, pre-sacral and obturatory lymph nodes. For the purpose of this study all primary tumours were manually segmented by two experienced radiation oncologists (CG and PB) in a blinded fashion. Staging MR T2-weighted sequences could be used to aid GTV definition.

10.3. CT-radiomics Analysis

The extraction, feature and model selection was performed as explained in 7 using the models discussed in chapter 4, we used as our training cohort the TUE cohort and FLO as an independent test cohort. The ROC-AUC metric was implemented as a metric in order to avoid imbalanced class problems.
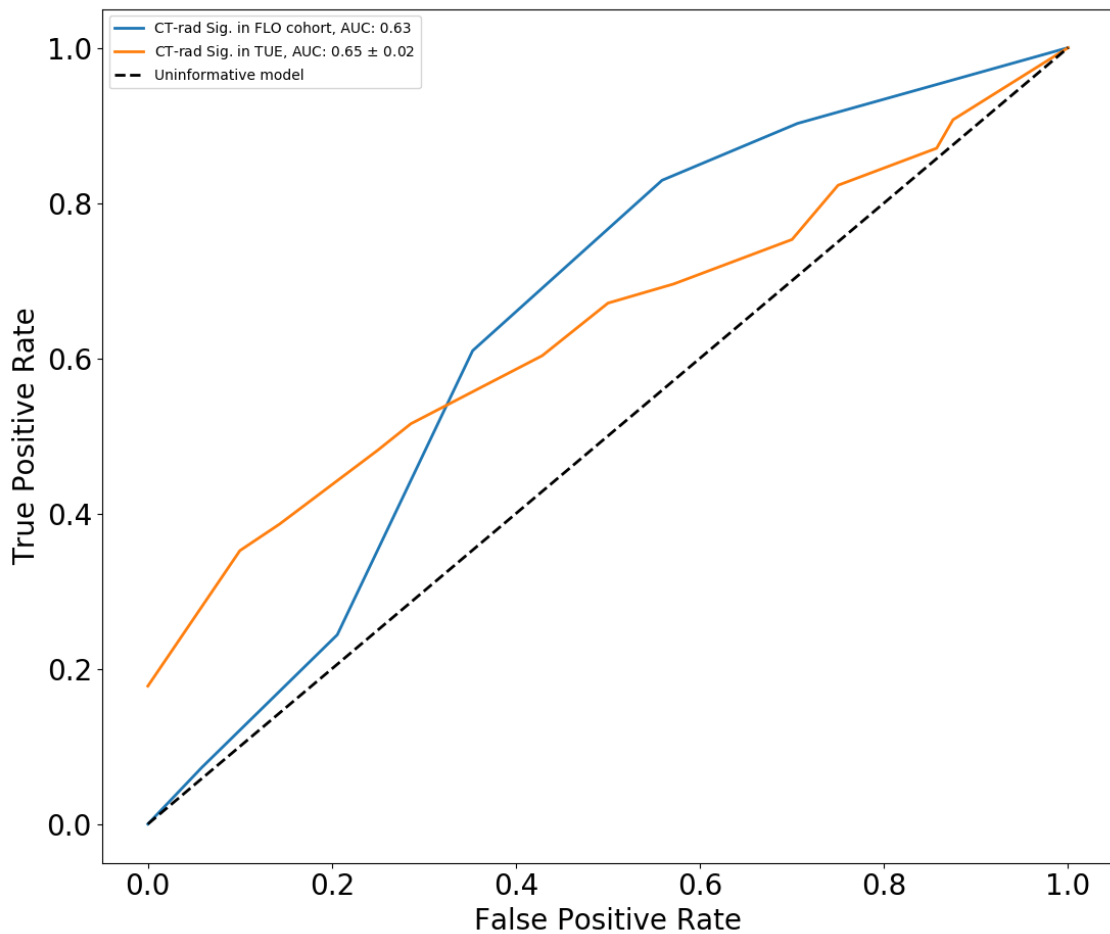
Figure 10.1.: Receiver operator characteristic curves for the learning and validation cohorts.

10.4. Predictions

A model was identified to have a similar ROC-AUC in the learning cohort (TUE) as well as in the validation cohort (FLO). The model consisted on an 5-nearest neighbours algorithm with PCA-feature cluster associated to "Range" intensity feature. In the learning cohort the computed ROC-AUC is about 0.65 ± 0.02 while for the independent validation cohort (FLO) the performance was 0.63, cf. figure 10.1.

Part IV.

Discussion and Outlook

11. Discussion and outlook

Medical imaging modalities have become the cornerstone technology to diagnose cancer status and potentially the cornerstone technology to personalise CRT with low invasiveness in comparison with other techniques. The plethora of proposed and researched image biomarkers promise better treatment adaptations. Radiomics is a fast growing research field, that due to the immense studies which provides promising results in the matter, has become heavily used to find new correlations and predictions in the context of clinical research. Yet sometimes at expenses of non-standardised methodology or reproducibility [97]. Under the course of this thesis only two widely-accessible implementations are available for radiomic studies [168,181]. However, two problems should be mainly addressed in the implementations:

- Reproducibility of radiomic features

- Robust modelling for outcome predictions

In this thesis a customized radiomic software was implemented for the necessities of clinical research studies of the Radiation oncology department of the University Hospital Tübingen, with the aim to identify and validate imaging biomarkers related to heterogeneity and other biological phenotype expressions of aggressiveness of tumours (cf. chapter 6). The imaging processing and feature extraction part of the software was validated inside the IBSI international collaboration (more than 20 groups participated in it to reach consensus) for reproducibility purposes. Robust modelling was built on top of state-of-the-art python libraries for machine learning purposes. Model evaluations were cautiously carried to address problems of over-fitting and under-fitting from highly correlated features to limitations in the number of samples.

The implementations developed for image preprocessing and feature extraction (cf. chapter 6) were validated for most of the features proposed by the IBSI international collaboration. Nonetheless, some features remain not standardised since internal decisions were more convenient for the performed clinical research inside the framework of this thesis, such as volume and surface area computations, which differs from the IBSI collaboration and thus features that depend on these parameters. For the research applications of this thesis, CT images had more than 1000 voxels, hence generations mesh approaches, which imply more computational implementations, debugging and testing at no much higher accuracy than voxel counting, were disregarded. Inside the IBSI collaboration, consensus was reached across institutes for mathematical definitions of features. Nonetheless variability in imaging preprocessing implementations

were identified as the main source of lack of reproducibility of features computations inside the collaboration, further optimization work is being performed inside the collaboration to establish benchmarks at the time of the writing phase of this thesis and therefore not definitive results can be obtained. In this thesis was decided to limit grey levels values in the range of tissue grey levels for CT images, which could be extended to MRI and PET, no interpolation or down-sampling was applied in spite of differences in voxel-axis lengths, texture features were computed in 3D fashion, and neighbourhood was consider for 1 voxel distance vicinity for all texture matrices. Despite those issues, radiomics studies can be still performed since feature selection process and ML modelling are influenced by feature variability inside training samples of patients and performance over test samples of patients, not over an specific meaning of feature value. Intra-institution reproducibility is affected however by the issue, therefore participation in the initiative should be continued in the future. Inside the IBSI collaboration no filter image preprocessing is discussed, thus consensus could not be reached. The selection of a filter is specific for the pattern variations that the researcher would like to highlight or enhance from images. Nonetheless, as variations of feature values are mostly originated in the image preprocessing part, filtered image features lack of reproducibility could also come from this issue. In the research inside of this work, wavelets features were usually picked by most of the feature selection algorithms implemented as highly predictive for the required prediction tasks, hence benchmarks are also needed in this regard at least for the most common image filters.

All clinical research results inside the framework of this thesis should not be seen as definitive, since much optimization should still be implemented from methodological points of views up to number of patients included in cohorts. These optimizations should be carried out in the future, which due to time limitations these were not implemented. In the following, a summary of conclusions, limitations and study extends are discussed.

In the first research application of this thesis (cf. chapter 8) correlations between CT-radiomic features and gene mutation expressions associated to heterogeneity phenotype were tested in 20 patients of HNSCC driven by the study of Aerts et al. [5]. Some somatic mutations of FAT1 and tumour volume were correlated with the so called Aerts-signature for intra-tumour phenotype expression heterogeneity, The features inside the Aerts signature are broader than only capturing volume owe to they compute number of patterns inside the ROI and thus the larger the volume, the larger the possibility to find more patterns in the ROI. Mutations of TP53 and KMT2D expressions did not show correlations with the proposed radiomic signature, although it does not mean that generally there is no correlation with intra-tumour heterogeneity, as it can be suggested by the study, the effect of TP53 mutations is assumed to be broader on tumour development than the responsibility of structural heterogeneity in tumours. This study is aligned to lack of definitive replications from other studies of correlations of intra-tumour heterogeneity with the Aerts-signature [57, 68], [178],

these findings requires larger cohorts to validate this exploratory data and should be performed in the future. The scientific innovation of this work remains in the trial to confirm correlations between radiomic features with genetic expressions which are rather sparse in clinical research, despite the advantage of finding positive conclusive results [186].

As a second research application (cf. chapter 9), the goal was to find a CT-radiomic signature which could be a substitute identifier of patients at risk of the hypoxia static ^{18}F-FMISO PET imaging biomarker TBR_{peak} in HNSCC.

Two features, out of 550 radiomics meta-features, along with the KNN model were identified as the best-performing radiomics signature from the training cohort (HN1). We assumed them to be associated to phenotypical expressions of heterogeneity in tumours, since these features are defined to quantify pattern-variations of grey-levels in medical images [5, 56, 97, 184]. As we had a retrospective data set and therefore could not access genetic information of the tumours, a direct proof of this assumption is lacking.

As previously indicated [91,117,183], the results of our study show that pre-treatment FMISO PET TMR_{peak} has a significant prognostic power to discriminate between patients with high and low risk of LRF following CRT. This is aligned to the study of Zips et al. [183] which found a significant prognostic power of the TMR_{median} feature in the baseline and at the second week after the start of the CRT treatment. The approach in our study is based on the results of Löck et al. 2017 [110]. A better discriminative power of the TMR_{peak} threshold may however be reached in second-week images after the start of CRT. We did not explore this approach, because we were limited by the data, as our study was performed retrospectively and we did not have weekly [18F]FMISO PET and CT scans for all patients. In the case of CT radiomics, features extracted from imaging during treatment resulted in a higher prognostic power compared to features from the start of CRT, as shown by Fave et al. 2016 [43] and Leger et al. 2019 [102]. In the publication of Löck et al. [110], TMR_{peak} was not found to be significantly related to LRF for their exploratory cohort (n = 25). However, in Mönnich et al. [117] TMR_{peak} was found significant for 22 out of the HN2 cohort. The two studies [110, 117] showed some methodological differences. The approach of Löck et al. [110] might be a more robust method because it used an exploratory cohort for assessing TBR_{peak} thresholds at different time points during the course of CRT in addition to an independent validation cohort for testing. Whereas in the study of Mönnich et al. [117], the derivation of the TMR_{peak} threshold consisted of the median-value in the exploratory cohort, which was not independently validated [185]. The ROC-AUC was lowered from 0.77 in Mönnich et al. to 0.66 in this study. A possible explanation for these results might be the lack of standardisation for the determination of SUV_{muscle} and SUV_{peak} in 0.5 cm^3 of tumour or muscle tissue, which depends strongly on manual delineation procedures.

Moreover, the difference in AUC results may be an effect of the increased sample size.

Larger, more heterogeneous solid tumours often develop hypoxia and therefore increase risk of LRF [55, 82, 119, 145]. The hypothesis of this study was that CT radiomics, which is assumed to quantify heterogeneity in tumours, could be used to provide a prognostic model that significantly correlates with LRF after CRT and thus also up to a significant extent with an imaging metric for hypoxia, such as TBR_{peak}. However, hypoxia may not be the only cause of LRF. Different factors such as patient characteristics, tumour biology and also treatment related issues contribute to the observed outcome which may not be captured by CT radiomics. The weak overlap in matching predictions between both modalities (55.3 %) might be a consequence of this. A more robust approach to test our hypothesis would be directly targeting hypoxia gene expressions, hypoxia imaging biomarkers [109] or FMISO image distributions via a deep learning architecture such as a Convolutional Neural Network (CNN) based approach [8,53], instead of targeting loco-regional outcomes of tumours. However, we could not pursue these approaches because of our small cohort size to train and test findings.

In this study, model training is performed using LRF data. The binary nature of the response variable introduces limitations to this study as censored events as well as the time to recurrence is neglected. Other studies have presented radiomics models which include time-to-event data [102] and might thus be considered more accurate in terms of event modelling.
Another possible limitation of our study is the application of wavelet filters in the 3D image space. Voxel lengths were not interpolated and thus no equal voxel spacing was used leading to larger voxel dimensions in slice direction compared to in-plane voxel spacing. This may affect the generation of new filtered images and subsequently the corresponding feature values. However, interpolation operations might also introduce additional artefacts to the data. To date, it is unknown to which extend this might affect the process of feature selection and machine learning modelling.
The chosen CT radiomics signature is based on the best performing signature inside the training phase. This is not always the safest choice according to Leger S. et al. 2017 [101]. As a result, we tested the 6 best CT radiomics signatures from the training phase in our validation cohort, where similar results were obtained (cf. table 9.3). We therefore did not see any impediment to compare simply the best signature from our training phase with the results obtained for TBR_{peak} as a matter of consistency. In this study, no direct correlation between FMISO PET TBR_{peak} and the best performing CT radiomics model was found. CT radiomics may pick tumour phenotypic heterogeneity from CT data which might be linked to tumour hypoxia, but indirectly. However, direct assessment of tumour hypoxia with specific imaging techniques and radiotracers is suggested to have a more powerful prediction power.

Finally, a radiomic approach was proposed as an organ-preserving strategy for ra-

diochemotherapy, modest figures were found in comparison with the study of Nie et al [123]. Imaging biomarkers are considered as promising features to inform personalised therapeutic decisions. However, reports on the use of them for organ-preserving strategies remain sparse.

Regarding results obtained in this study, better figures were obtained in previous studies [35,123]. However the success of this study comprehend the size of the cohort of patients involved as well as the inclusion of a validation cohort from another institution with a similar but still different imaging set-up, which confirms the need for standardised imaging protocols across institutions and studies suggested by the IBSI collaboration [184]. Similar ROC-AUC scores were found in the two cohorts, which suggests that findings can be replicated at other institutions. The feature (PCA-range) by definition captures the difference between the maximum and the minimum grey level of the region of interest, which is an indication of the density distribution inside the analysed region or in other words the contrast in the region. It seems to be that the larger the PCA-range, the larger the probability of failure for complete remission of the tumour.

Our study posses limitations in different aspects, for instance the deep learning (DL) approach was never used due to lack of understanding of the features that a DL approach might encounter, which usually tends to be abstract for clinical implementations. This approach is suggested to be researched in the future. Another possible limitation is associated to the target used for machine learning purposes, as the target was aimed to predict good vs poor regression grades (a binary problem), a multi-label approach might be more suitable for better prediction of clinical status, for instance a one vs all approach, the approach was not integrated in this study due to time constrains. Nonetheless, further research in this direction is encouraged.

As CT is used in the clinical routine for diagnosed purposes, the result of this work might stimulate the research of the use of advanced imaging analysis to spare the use of aggressive therapies in rectal carcinomas and be used as a strategy for organ-preserving therapies.

The ML approach used inside the framework of this thesis was performed under the construction of hand-crafted, yet meaningful, features from medical images and a classical machine learning pipeline for feature and model selection. More advanced approaches use deep learning algorithms [99,105,126], where the algorithm learns from images as inputs the features to be accounted for prediction. The choice for a more traditional radiomic pipeline was based on the fact that some hard treatable issues arise from this approach. First, the sizes of the samples inside this work did not super-passed the order of hundreds of patients (which is far too small for purposes of implementing deep learning architectures). Besides, as a deep learning algorithms propose to itself the features or combination of features to be prognostic

without a direct connection to biological definitions, they are barely interpretable in their meanings and associations to well established biological process in tumours. Therefore, the translations of such eventually encountered features seem much harder for clinical applications. Finally, in order to make the deep learning approach computationally effective, it requires the use of high performance GPU graphic cards for image analysis (recursive and convolutional neuronal networks), which in the moment of embarking in the development of this thesis, they were not available. The approach should anyway be considered for future studies as it broadens the spectrum of imaging biomarkers and up to date of publication, applications to HNSCC patients have not been performed yet.

Only static radiographic images were used inside the radiomic studies of this thesis, no dynamic approaches were undertaken, despite interesting clinical research applications as proposed by Fave et al [43] for tumour treatment responses due to time limitations and lack of data samples (patients and scans), further investigations on this direction should also be consider in the future.

This work presents a full implementation of a radiomic software, which were applied to different clinical studies. This promising results suggest that radiomics can be used as a either substitute or complementary information for CRT adaptations, however still further optimizations should be performed to validate findings, such as software validation, training and test cohort size and longitudinal studies.

Part V.

Summary

Summary of the Thesis

Part I: Introduction

Chapter 1: Introduction

In chapter 1, motivations and state of the art of the topics treated inside the framework of this thesis are supplied.

Part II: Materials and Mehtods

Chapter 2: Basics of imaging modalities and biomarkers for RT

The chapter 2 followed to explain the basic physics and image reconstruction of CT, MRI and PET. Besides it shows the origins of the most common imaging markers from static to dynamic models with their advantages and disadvantages.

Chapter 3: The radiomic hypothesis and image feature engineering

The chapter 3 provided a compelling introduction to the field of radiomics along with explanations of the most common features for image analysis and how radiomic incorporates them to construct relevant imaging biomarkers.

Chapter 4: Machine learning in the context of radiomics

In chapter 4, a brief description of the most common machine learning algorithms for feature selection and modelling was imparted in regards to radiomics. Moreover, a short discussion about model validation and tuning and selection in this work is also supplied.

Chapter 5: Python as a Software Development Language

In chapter 5, a brief description of Python as programming language was provided as well as a short discussion about the reasons to use it and some comparisons with other languages were explained.

Part III: Results

Chapter 6: In-house software development for radiomics and its validation

The chapter 6 yielded in an introductory description of the software implementation and discloses the validation benchmarks of the software inside the international col-

laboration IBSI.

CHAPTER 7: RADIOMICS PIPELINE INSIDE THIS THESIS

The chapter 7 provided a detailed discussion of the radiomics pipeline used in this work as well as the hyperparamenter spaces used for every machine learning model used.

CHAPTER 8: CORRELATIONS OF THE AERTS SIGNATURE WITH SOMATIC MUTATIONS IN TP53, FAT1 AND KMT2D IN HNSCC

In chapter 8 summarized an exploratory study to confirm the Aerts signature with heterogeneity in cell pathways mutations, the results are therefore discussed in an adaptation form from the publication of [184].

CHAPTER 9: CT RADIOMICS: A SURROGATE PREDICTOR OF PATIENTS AT RISK AS IDENTIFIED BY ^{18}F-FMISO HYPOXIA PET?

In chapter 9, the feasibility to use CT-radiomics as a substitute strategy to classify patients at risk in comparison to standard hypoxia-based static model was investigated. The results were thus disclosed.

CHAPTER 10: AN APPROACH FOR PREDICTION OF COMPLETE REMISSION OF ORGAN-PRESERVING THERAPY IN RECTAL CARCINOMA WITH INDEPENDENT VALIDATION)

In chapter 10, the feasibility to use CT-radiomics as a strategy to improve organ-preserving therapies in rectal carcinomas patients was investigated. The results were thus disclosed.

Part IV: Discussion and outlook

CHAPTER 11: DISCUSSION AND OUTLOOK

Finally, chapter 11 provides a summary of the observations, discussions and scopes of this work and an outlook to future research.

Appendices

12. Software documentation: Image preprocessing and feature extraction

In this chapter full disclosure of the radiomic implementations for feature extraction and image preprocessing are discussed.

12.1. preprocessing.

The preprocessing module provides functions for imaging processing tasks, such as sampling, grey level quantizations and segmentations.

- `adjust_grey_levels(volume, new_min, new_max)`:
  Voxels of a given volume are mapped to a new volume where the voxels are between a desired new minimum (`new_min`) and maximum (`new_max`).

- `apply_segmentation_mask(volume, segmentation_mask)`:
  Apply segmentation to a given `volume` based on a `ndarray` of ones and zeros (`segmentation_mask`), where ones means that the region is inside of the volume of interest and 0 otherwise.

12.2. read_img_data.py

This module provides functions to read and process medical images in *nifti* format. The module `preprocessing.py` and the built-in library `nibabel` were used to translate nifti images to a `ndarray` object from *NumPy*.

- `process_patient_data(folder_path, image_path, seg_path)`:
  reads and transform the desired medical image data located at the address `folder_path/image_path` and segmentation mask at the address folder `folder_path/seg_path` to an *ndarray* object. Here relevant information from the used scan parameter are extracted from the *nifti*-header to prepare the data for texture analysis, shape features and filters applications in the original *nifti*-image.

- `get_processed_folder(folder_path)`:
    Reads the saved `ndarray` objects obtained from `process_patient_data` module for feature extraction located at `folder_path`.

12.3. Stats_features.py

This module provides functions to compute first order statistics features (Intensity-family) from the loaded `ndarray` objects. Here the built-in library *SciPy* was used for desired statistical operations.

- `compute_stats_features(volume)`:
    computes 18 statistical features in a `volume` based on the definition of IBSI for Intensity features for radiomics [184].

- `get_stats_feature_names()`:
    obtains the names of the 18 statistical features computed from `compute_stats_features`.

12.4. Intensity_histogram_features.py

This module contains a series of computational fictions to compute intensity features based on histogram of grey levels as proposed by the IBSI collaboration [184].

- `histo_gen(volume)`:
    generates a histogram based on a defined `volume` array.

Listing 12.1: Python code for histogram generation of an ndarray object, Intensity_histogram_features.py.

```python
def histo_gen(volume):

    # generation of bins according to volume
    l_bins = np.arange(np.nanmin(volume), np.nanmax(volume) + 2)
    # resampling ndarray to 1-darray
    X_d = np.hstack(np.round(volume))
    # histogram generation
    histo = np.histogram(X_d[np.isnan(X_d) == False], bins = l_bins)
    # frequency computations
    H = histo[0]
    # Number of voxels
    N_v = float(np.size(volume[np.isnan(volume) != True]))
    # Number of grey-levels
    N_g = float(histo[1][-2])
    # Proportion computations
    p = np.zeros(np.size(H))
    for i in range(np.size(p)):
```

```
    p[i] = H[i]/float(N_v)

return X_d, H.copy(), N_v, N_g, p
```

- `compute_ih_features(volume)`:
  computes 28 features based on the generated histogram from `histo_gen` based on definitions of the IBSI collaboration [184].

- `get_ih_feature_names()`:
  obtains the names of the computed 28 features from `compute_ih_features`.

12.5. Intensity_volume_histogram_features.py

This module delivers utilities to compute intensity volume histograms features defined in the IBSI manuscript [184].

- `IVH_gen(volume)`:
  generates a volume intensity histogram from a given `ndarray` object (`volume`)

Listing 12.2: Python code for a volume intensity histogram generation of an ndarray object, Intensity_volume_histogram_features.py.

```python
def IVH_gen(volume):
        # Number of voxels belonging to the ROI
    N_v = float(np.size(volume[np.isnan(volume) != True]))
    # initialize the frequency of voxels that are under a defined
        threshold
    v = np.array([])
    # initialize bins
    gamma = np.array([])

    X_min = np.nanmin(volume)
    X_max = np.nanmax(volume)

    # computes proportion of voxels under the different number of grey
        levels and bins
    I = np.arange(int(np.nanmin(volume)),int(np.nanmax(volume) + 1))
    for i in I:
        if i == np.nanmin(volume):
            delta = 0
            item = 1
        else:
            delta = (volume < i)
            delta = delta.astype(int)
            item = 1 - delta.sum()/N_v
        v = np.append(v, item)
        item2 = (i - X_min)/(X_max - X_min)
        gamma = np.append(gamma, item2)
```

```
    return I, v, gamma, N_v
```

- `compute_ivh_features(volume)`:
    computes 7 features based on the generated volume intensity histogram from
    `IVH_gen` based on definitions of the IBSI collaboration [184].

- `get_ivh_feature_names()`:
    obtains the names of the computed 7 features from `compute_ivh_features`.

12.6. Morph_features.py

The module contains functions and utilities to compute shape-family radiomic features as proposed by the IBSI collaboration [184]. For mesh generations the module `spatial` from the built-in library *SciPy* is used.

- `compute_aabb(X_c)`:
    computes volume and area of an axis-aligned bounding box given a list of mesh
    vectors `X_c`.

Listing 12.3: Python code for computation of axis-aligned bounding box, Morph_features.py.

```python
def compute_aabb(X_c):
    # Extract vector points from the mesh vector
    points = X_c
    # Construct a convex hull from the mesh vectors and largest
        distances inside the convex hull
    hull_points = points[ConvexHull(points).vertices]
    min_x = np.nanmin(hull_points[:,0], axis=0)
    max_x = np.nanmax(hull_points[:,0], axis=0)
    min_y = np.nanmin(hull_points[:,1], axis=0)
    max_y = np.nanmax(hull_points[:,1], axis=0)
    min_z = np.nanmin(hull_points[:,2], axis=0)
    max_z = np.nanmax(hull_points[:,2], axis=0)
    # Computes volume of the axis-aligned bounding box
    vol = (max_x - min_x)*(max_y - min_y)*(max_z-min_z)
    # Computes area of the axis-aligned bounding box
    area = 2*(max_x - min_x)*(max_y - min_y) + 2*(max_x - min_x)*(max_z
        - min_z) + 2*(max_y - min_y)*(max_z - min_z)

    return vol, area
```

- `compute_ombb(X_c)`:
    calculates the volume and the area of a oriented minimum bounding box for a
    given set of mesh vectors `X_c`

Listing 12.4: Python code for computation of oriented minimum bounding box, Morph_features.py.

```python
def compute_ombb(X_c):
    points = X_c
    pi2 = np.pi/2.0
    # generates vectors for the convex hull generated from the given
        mesh vectors of the image
    hull_points = points[ConvexHull(points).vertices]
    edges = np.zeros((len(hull_points), 3))
    # generates edges for the convex hull
    edges = hull_points[1:] - hull_points[:-1]
    for i in range(len(edges)):
        edges[i,:] = edges[i,:]/np.linalg.norm(edges[i,:], axis= 0)
    # maps convex hull edges into a oriented bounding box space
    thetas = np.zeros((len(edges),3))
    for element in range(len(edges)):
        thetas[element,0] = np.arctan2(edges[element,1], edges[element
            ,0])
        thetas[element,1] = np.arctan2(edges[element,2], edges[element
            ,0])
        thetas[element,2] = np.arctan2(edges[element,2], edges[element
            ,1])
    thetas = np.abs(thetas*90.0/pi2)
    vol = np.zeros(len(thetas))
    surface_area = np.zeros(len(thetas))
    for item in range(len(thetas)):
        R_x = np.array([[1,                           0,                           0],
                        [0, np.cos(thetas[item,0]), -np.sin(thetas[item
                            ,0])],
                        [0, np.sin(thetas[item,0]),  np.cos(thetas[item
                            ,0])]])

        R_y = np.array([[np.cos(thetas[item,1]), 0,  np.sin(thetas[item
            ,1])],
                        [0,                           1,                           0],
                        [-np.sin(thetas[item,1]),0,  np.cos(thetas[item
                            ,1])]])

        R_z = np.array([[np.cos(thetas[item,2]), -np.sin(thetas[item,2])
            , 0],
                        [np.sin(thetas[item,2]),  np.cos(thetas[item,2])
                            , 0],
                        [0                     ,                     0,  1]])

        rotations = R_x.dot(R_y).dot(R_z)
        rot_points = np.dot(rotations, hull_points.T)

        min_x = np.nanmin(rot_points[0,:], axis=0)
        max_x = np.nanmax(rot_points[0,:], axis=0)
        min_y = np.nanmin(rot_points[1,:], axis=0)
        max_y = np.nanmax(rot_points[1,:], axis=0)
```

```
    min_z = np.nanmin(rot_points[2,:], axis=0)
    max_z = np.nanmax(rot_points[2,:], axis=0)
# computes volume and surface area of the oriented − minimum
    bounding box
    vol[item] = (max_x − min_x)*(max_y − min_y)*(max_z−min_z)
    surface_area[item] = 2*(max_x − min_x)*(max_y − min_y) + 2*(
        max_x − min_x)*(max_z − min_z) + 2*(max_y − min_y)*(max_z −
        min_z)
vol_ombb = np.nanmin(vol)
sa_ombb = np.nanmin(surface_area)

return vol_ombb, sa_ombb
```

- `apply_pca(X_c)`:
  computes principal components of a given set of mesh vectors `X_c`

Listing 12.5: Python code for computation of a PCA, Morph_features.py.

```
def apply_pca(X_c):
    # computes covariance of the mesh points
    cov = np.cov(X_c)
    # computes eigen values and eigen vectors of the covariance matrix
    evals, evecs = np.linalg.eig(cov)
    # sorts eigen values from larger to smaller
    evals_ord = evals[::−1]
    return evals_ord
```

- `morph_matices(volume, vox_vol)`:
  creates matrices that will be utilised for feature computations.

Listing 12.6: Python code for morphological matrix generation, Morph_features.py.

```
def morph_matices(volume, vox_vol):
    # Extraction of grey level values that belongs to the region of
        interest
    X_gl = volume[np.isnan(volume) == False].flatten()
    # Generation of component locations of the voxels belonging to the
        region of interest
    ix, iy, iz = np.where(np.isnan(volume) == False)
    ix = (ix + 0.5)*vox_vol[0]
    iy = (iy + 0.5)*vox_vol[1]
    iz = (iz + 0.5)*vox_vol[2]
    # Mesh generation matrix
    X_c = np.array([ix, iy, iz])
    # Individual voxel volume computation
    vox_volume = vox_vol[0]*vox_vol[1]*vox_vol[2]
    # Matrix generation of volume entries
    X_v = np.ones(np.shape(X_gl))*(vox_volume)

    return X_gl, X_c.T, X_v
```

- `compute_CoM_shift(X_c, X_gl)`:
  acquires centre mas shift from a given mesh vectors array `X_c` and a matrix grey level `X_gl`.

- `max_diam_3D(X_c)`:
  calculates the distance between the two furthest apart vectors in the mesh vector array `X_c`.

Listing 12.7: Python code for computation of maximum 3D diameter, Morph_features.py.

```
def max_diam_3D(X_c):
    vectors = X_c
    # Convex hull generation from the X_c vector array
    hull = ConvexHull(vectors)
    set_largest_vectors = [vectors[hull.vertices]]
    # Extracts the furthest apart vectors
    item_max = np.amax(set_largest_vectors, axis = 1)
    item_min = np.amin(set_largest_vectors, axis = 1)
    # Computes the distance between the furthest vectors
    return np.linalg.norm(item_max - item_min)
```

- `Surface_Area(volume, vox_vol)`: computes approximate surface area of a given 3D-array by counting external voxel faces.

Listing 12.8: Python code for approximate surface area computation, Morph_features.py.

```
def Surface_Area(volume, vox_vol):
    # Initializes by copying a temporal volume
    vol_temp = volume.copy()
    # Extracts size of the 3d dimension
    dim_k = np.shape(volume)[2]
    # maps volume that belongs to the region of interests
    vol_temp[~np.isnan(volume)] = 1
    # initialize counting of voxel faces that belongs to outer face
    surface_slice = np.zeros(dim_k + 2)
    c_top = 0;
    c_buttom = 0;
    for k in range(dim_k):
        c = vol_temp[:,:,k]
        cx = 0;

        dim_x ,dim_y = np.shape(c)

        b = np.zeros(dim_x);
        b[b==0] = np.nan

        for i in range(dim_x):
```

```
            for j in range(dim_y):
                if ~np.isnan(c[i,j]) and k == 0:
                    c_top += 1
                elif ~np.isnan(c[i,j]) and k == dim_k-1:
                    c_buttom += 1
        # computes surface area in the top face of the region of
            interest
        area_top = c_top*(vox_vol[0]*vox_vol[1])
        # computes surface area in the buttom face of the region of
            interest
        area_buttom = c_buttom*(vox_vol[0]*vox_vol[1])

        # repeats process for side faces of the region of interests
        c_comp = c[0,:][:,np.newaxis]
        if nan_equal(c_comp, b) == True:
            c_new = c
        else:
            c_new = np.concatenate((b.reshape(-1,1),c), axis=1)


        if nan_equal(c[:, dim_y-1][:, np.newaxis],b) == True:
            c_new = c_new
        else:
            c_new = np.concatenate((c_new, b.reshape(-1,1)), axis=1)

        dim_x_new, dim_y_new = np.shape(c_new)
        d = np.zeros(dim_y_new)
        d[d==0] = np.nan

        if nan_equal(c_new[0,:],d) == True:
            c_new = c_new
        else:
            c_new = np.concatenate((d.reshape(1,-1),c_new), axis = 0)

        dim_x_new, dim_y_new = np.shape(c_new)
        d = np.zeros((1, dim_y_new))
        d[d==0] = np.nan

        if nan_equal(c_new[dim_x_new-1,:], d):
            c_new = c_new
        else:
            c_new = np.concatenate((c_new,d.reshape(1,-1)), axis = 0)

#        B = np.concatenate(c_new, axis=2)
        cx = 0
        dim_x_new , dim_y_new = np.shape(c_new)

        for j in range(dim_y_new):
            for i in range(dim_x_new):
                if np.isnan(c_new[i,j]) != 1 and np.isnan(c_new[i,j+1])
                    == 1:
```

```
                    cx += 1
                if np.isnan(c_new[i,j]) !=1 and np.isnan(c_new[i,j-1])
                    == 1:
                    cx += 1

        cy = 0
        for j in range(dim_y_new):
            for i in range(dim_x_new):
                if np.isnan(c_new[i,j]) != 1 and np.isnan(c_new[i+1,j])
                    == 1:
                    cy += 1
                if np.isnan(c_new[i,j]) !=1 and np.isnan(c_new[i-1,j])
                    == 1:
                    cy += 1
        borders = cx + cy
        surface_slice[k] = borders*vox_vol[0]*vox_vol[2]
        surface_slice[dim_k] = area_top
        surface_slice[dim_k + 1] = area_buttom


    return surface_slice
```

- `Area_ellip(morph_pca_major, morph_pca_minor, morph_pca_least, n)`:
  obtains area of the minimum enclosing ellipsoid in the 3D-image. Inputs: major, minor and least principal component `morph_pca_major`, `morph_pca_minor` and `morph_pca_least` of the X_c., since the computation needs the Legendre polynomials the `n` is set to 10 for high-accuracy.

Listing 12.9: Python code for surface area of a minimum ellipsoid contained in the X_c matrix, Morph_features.py.

```
def Area_ellip(morph_pca_major, morph_pca_minor, morph_pca_least, n):
    # initializes coefficients for legendre polynomials
    alpha = np.sqrt(1 - ((morph_pca_minor**2)/(morph_pca_major**2)))
    beta = np.sqrt(1 - ((morph_pca_least**2)/(morph_pca_major**2)))
    nu = np.arange(0,n)
    Amplitude = ((alpha*beta)**nu)/(1-(4*nu**2))
    varia = (alpha**2 + beta**2)/(2*alpha*beta)
    # computes legendre polynomials values
    pol = np.zeros(len(nu))
    for i in nu:
        pol[i] = sp.special.legendre(i)(varia)

    return 4*np.pi*morph_pca_major*morph_pca_minor*Amplitude.dot(pol)
```

- `get_I_moran_index(X_gl, X_c)`:
  computes moran index from a given grey level matrix `X_gl` and vector matrix array `X_c` according to [184].

- `get_C_geary_measure(X_gl, X_c)`:
    computes C geary index from a given grey level matrix `X_gl` and vector matrix array `X_c` according to [184].

- `compute_morph_features(volume, vox_vol)`:
    calculates all 24 morphology-family features proposed for the IBSI collaboration [184].

- `get_morph_feature_names()`:
    acquires the name of the features computed in `compute_morph_features(volume, vox_vol)`

12.7. GLCM.py

This module contains the utilities necessaries to compute the GLCM matrix (cf. section 3.2.3).

- `compute_GLCM_3D(volume, N_g, direction)`:
    constructs an angular dependent GLCM for a given direction in 3D from an input `volume` and a maximal grey level `N_g`.

Listing 12.10: Python code for generation of an GLCM from a 3d-array object, GLCM.py.

```python
def compute_GLCM_3D(volume, N_g, direction):

    # create gray level co-occurrence matrix
    GLCM = np.zeros((int(N_g), int(N_g)))

    # get volume size
    d1, d2, d3 = volume.shape[0], volume.shape[1], volume.shape[2]

    # one of the 13 3D-directions
    separating_vector = directions_3D[direction]

    # calculate cooccurrences
    for x,y,z in itertools.product(range(d1), range(d2), range(d3)):

        neighbor = np.array([x,y,z]) + separating_vector
        if (neighbor>=0).all() and (neighbor<np.asarray(volume.shape)).\
           all():
            i = volume[x,y,z]
            j = volume[neighbor[0], neighbor[1], neighbor[2]]

            # check for missing voxel values
            if np.isfinite(i) and np.isfinite(j):
                GLCM[int(i)-1,int(j)-1] += 1
```

```
    # add cooccurrences of the inverse direction
    GLCM += np.transpose(GLCM)

    return GLCM
```

- `get_normalized_GLCM(GLCM):`
    converts entries of the constructed `GLCM` to probabilities.

- `get_merged_GLCM(GLCM_list):`
    averages over the whole set of possible GLCMs (i.e GLCM in all directions) contained in the `GLCM_list`, the output is a 2d-array object.

12.8. GLCM_features.py

This module provides functions to compute radiomic features from the generated GLCM as proposed by IBSI collaboration [184].

- `get_p_i(GLCM):`
    obtains row and column marginal probability of the given `GLCM`

- `get_p_i_minus_j(GLCM):`
    computes diagonal probabilities of the given `GLCM`

- `get_p_i_plus_j(GLCM):`
    acquires cross-diagonal probabilities of a given `GLCM`

- `compute_GLCM_features(GLCM):`
    computes all 25 `GLCM` features defined according to IBSI collaboration [184]

12.9. GLRLM.py

The module contains the basic utilities for GLRL computations. it uses the module `sequences`, which includes methods to extract one-dimensional grey level sequences from a 3D image, in all directions.

- `compute_GLRLM_3D(volume, N_g, direction):`
    builds from a `volume` and a maximal grey level `N_g` an angular dependent GLRL for a given 3D `direction`. Runs lengths are acquired by computing all sequences in the given 3D `direction` individually, which is done by the function `get_run_lengths(sequence, N_g, max_run_length)`.

- acquires all run lengths within a one-dimensional `sequence` and creates the respective GLRL with size $N\_g \times$ `max_run_length`.

Listing 12.11: Python code for generation of an GLRL from a 3d-array object and sequences, GLRLM.py.

```python
def get_run_lengths(sequence, N_g, max_run_length):

    # matrix of run lengths
    run_lengths = np.zeros((int(N_g), max_run_length))

    gray_level = sequence[0]
    count = int(1)
    for i in range(1,len(sequence)):
        # if grey level is followed by the same grey level, count up
        if sequence[i] == gray_level:
            count += 1

        # if previous grey level was a valid one, save run length by
        #   incrementing matrix entry
        else:
            if np.isfinite(gray_level):
                run_lengths[int(gray_level)-1, count-1] += 1
            gray_level = sequence[i]
            count = 1

    # if last grey level is valid, save last run length
    if np.isfinite(gray_level):
        run_lengths[int(gray_level)-1,count-1] += 1

    return run_lengths

def compute_GLRLM_3D(volume, N_g, direction):

    # maximal run length along the given direction, determines width of
    #   GLRLM
    N_r = get_max_run_length_3D(volume, direction)

    # create GLRLM
    GLRLM = np.zeros((int(N_g), N_r), dtype=float)

    # get all sequences of gray levels within the volume, along the
    #   given direction
    if np.array_equal(direction, 0):
        seq = sequences.get_x_axis_seqs_3D(volume)
    elif np.array_equal(direction, 1):
        seq = sequences.get_y_axis_seqs_3D(volume)
    elif np.array_equal(direction, 2):
        seq = sequences.get_z_axis_seqs_3D(volume)
    elif np.array_equal(direction, 3):
        seq = sequences.get_in_plane_diag_seqs_3D(volume,135)
    elif np.array_equal(direction, 4):
        seq = sequences.get_in_plane_diag_seqs_3D(volume,45)
    elif np.array_equal(direction, 5):
        seq = sequences.get_in_plane_diag_seqs_3D(np.swapaxes(volume
```

```
            ,1 ,0) ,135)
    elif np.array_equal(direction , 6):
        seq = sequences.get_in_plane_diag_seqs_3D(np.swapaxes(volume
            ,1 ,0) ,45)
    elif np.array_equal(direction , 7):
        seq = sequences.get_in_plane_diag_seqs_3D(np.transpose(volume
            ,(2 ,0 ,1)) ,135)
    elif np.array_equal(direction , 8):
        seq = sequences.get_in_plane_diag_seqs_3D(np.transpose(volume
            ,(2 ,0 ,1)) ,45)
    elif np.array_equal(direction , 9):
        seq = sequences.get_xyz_diag_seqs_3D(volume)
    elif np.array_equal(direction , 10):
        seq = sequences.get_xyz_diag_seqs_3D(np.rot90(volume ,2))
    elif np.array_equal(direction , 11):
        seq = sequences.get_xyz_diag_seqs_3D(np.rot90(volume))
    elif np.array_equal(direction , 12):
        seq = sequences.get_xyz_diag_seqs_3D(np.rot90(volume ,3))

    # insert all run lengths into GLRLM
    for s in seq:
        GLRLM += get_run_lengths(s , N_g, N_r)

    return GLRLM
```

- `get_merged_GLRLM(GLRLM_list)`:
  averages all GLRL in all possible directions contained in `GRLR_list`. The output is an invariant GLRL

12.10. GLRLM_features.py

The module provides functions to acquire all 16 GLRL features discussed inside the IBSI collaboration [184].

- `compute_GLRLM_features(GLRLM, N_v)`:
  calculates all the 16 `GLRLM` features and the number of voxel of images `N_v`.

- `combine_GLRLM_features(GLRLM_list, N_v_list)`:
  computes all 16 GLRL features for a list of GLRLMs and average over them

- `get_GLRLM_feature_names()`:
  obtain the names of the 16 GLRL features.

12.11. GLSZM.py

The module includes utilities to build a GLSZ. It uses the built-in library *Scikit-image*, specifically the *measure* module to evaluate connectivity in 2D and 3D.

- `compute_GLSZM(img, N_g)`:
    assembles a GLSZ of an 3D- or 2D-array object `img` with a maximal gray level `N_g`.

Listing 12.12: Python code for generation of an GLSZ from a 3D- or 2D-array object, GLSZM.py.

```python
def compute_GLSZM(img, N_g):

    s = img.shape

    # maximal zone size in the volume, determines width of the GLSZM
    # discriminates between 2D and 3D inputs
    max_zone_size = s[0]*s[1]*s[2] if len(s)==3 else s[0]*s[1]

    # create GLSZM
    GLSZM = np.zeros((int(N_g), max_zone_size), dtype=float)

    # determine regions (in the sense of 8-connectedness) of voxels
    # with the same gray level

    labels = measure.label(img, connectivity=3) if len(s)==3 else measure
        .label(img, connectivity=2)
    props = measure.regionprops(labels, intensity_image=img)
    for prop in props:
        size = prop.area
        gray_level = prop.mean_intensity

        # ignore regions that are assigned to nan
        if np.isfinite(gray_level):
            GLSZM[int(gray_level)-1, int(size)-1] += 1

    return GLSZM
```

12.12. GLSZM_features.py

The module provides utilities for GLSZ feature computations as proposed by the IBSI collaboration [184].

- `compute_GLSZM_features(GLSZM, N_v)`:
    calculates all 16 `GLSZM` features for radiomics accounting the number of voxels that are assigned to valid grey levels `N_v`.

- `combine_GLSZM_features(GLSZM_list, N_v_list)`:
    averages features values obtained from the all the GLSZ `GLSZM_list`.

- `get_GLSZM_feature_names()`:
    obtains the names of the computed features from the function `combine_GLSZM_features`

12.13. GLDZM.py

The module provides functions and utilities to construct the GLDZ. It uses the module `measure` from the built-in library *Scikit-image* for labelling connectivity in 3D.

- `compute_GLDZM(img, N_g)`:
    constructs a GLDZ from a 3D NumPy array `img` with maximum grey level `N_g` in the `img`

Listing 12.13: Python code for generation of an GLDZ from a 3D- or 2D-array object, GLDZM.py.

```python
def compute_GLDZM(img, N_g):

    s = img.shape

    # maximal zone size in the volume, determines width of the GLDZM
    # discriminates between 2D and 3D inputs
    # create distance map
    D = np.zeros(s, dtype=float)


    ran = np.min(s)
    if len(s)==3:
        if ran%2 == 0:
            for i in range(int(ran/2)):
                D[0+i,(0+i):(s[1]-i),(0+i):(s[2]-i)] = 1+i
                D[(s[0]-1-i),(0+i):(s[1]-i),(0+i):(s[2]-i)] = 1+i
                D[(0+i):(s[0]-i),0+i,(0+i):(s[2]-i)] = 1+i
                D[(0+i):(s[0]-i),s[1]-1-i,(0+i):(s[2]-i)] = 1+i
                D[(0+i):(s[0]-i),(0+i):(s[1]-i), 0+i] = 1+i
                D[(0+i):(s[0]-i),(0+i):(s[1]-i), s[2]-1-i] = 1+i
        else:
            for i in range(int(ran/2) + 1):
                D[0+i,(0+i):(s[1]-i),(0+i):(s[2]-i)] = 1+i
                D[(s[0]-1-i),(0+i):(s[1]-i),(0+i):(s[2]-i)] = 1+i
                D[(0+i):(s[0]-i),0+i,(0+i):(s[2]-i)] = 1+i
                D[(0+i):(s[0]-i),s[1]-1-i,(0+i):(s[2]-i)] = 1+i
                D[(0+i):(s[0]-i),(0+i):(s[1]-i), 0+i] = 1+i
                D[(0+i):(s[0]-i),(0+i):(s[1]-i), s[2]-1-i] = 1+i
    else:
        if ran%2 == 0:
            for i in range(int(ran/2)):
                D[0+i,(0+i):(s[1]-i)] = 1+i
                D[(s[0]-1-i), (0+i):(s[1]-i)] = 1+i
                D[(0+i):(s[0]-i),0+i] = 1+i
                D[(0+i):(s[0]-i), (s[1]-1-i)] = 1+i
        else:
            for i in range(int(ran/2) + 1):
                D[0+i,(0+i):(s[1]-i)] = 1+i
                D[(s[0]-1-i), (0+i):(s[1]-i)] = 1+i
```

```
                    D[(0+i ):( s [0]−i ) ,0+i ] = 1+i
                    D[(0+i ):( s [0]−i ) , ( s[1]−1−i )] = 1+i

        # determine regions (in the sense of 8−connectedness) of voxels
        # with the same gray level
#      labels = measure.label(img,neighbors=8);
        N_d = np.nanmax(D)
        GLDZM = np.zeros(( int (N_g), int (N_d)), dtype=float )
        labels = measure.label(img,connectivity=3) if len(s)==3 else measure
            .label(img,connectivity=2)
        props = measure.regionprops(labels, intensity_image=img)
        for prop in props:
            size = np.min(D[labels == prop.label])
            gray_level = prop.mean_intensity

            # ignore regions that are assigned to nan
            if np.isfinite(gray_level):
                GLDZM[ int (gray_level)−1,int (size)−1] += 1

        return GLDZM
```

12.14. GLDZM_features.py

- `compute_GLDZM_features(GLDZM, N_v)`:
  acquires all 16 feature values proposed for `GLDZM` accounting for the number of voxels in the respective image volume that are assigned to valid gray levels [184].

- `get_GLDZM_feature_names()`:
  obtains names of the feature values calculated in `compute_GLDZM_features`.

12.15. NGTDM.py

The module includes functions that allow to construct the NGTD for a 2D and a 3D array object.

- `compute_NGTDM(img, N_g, distance)`:
  assembles the NGTD for a given 2D- or 3D-array array object `img`, the `N_g` input is the max gray level in the volume and determines the height of the NGTD. The radius of the neighbourhood is accounted by `distance`.

- `get_neighborhood_average_difference_3D(volume, voxel_position, distance)`:
  averages differences between voxel grey level and its neighbourhood at determined `voxel_position` with radius `distance` within the given 3D-array object

Listing 12.14: Python code for generation of an NGTD from a 3D- object and average difference computations, NGTDM.py.

```python
def get_neighborhood_average_difference_3D(volume, voxel_position,
    distance):

    w = 0 # to determine the neighborhood size
    av = 0 # average gray level of neighboring voxels

    # regard all positions within the defined neighborhood and
    # consider only positions within the volume that are not assigned to
        nan
    for k in itertools.product(range(-distance, distance+1), repeat=3):
        v_array = np.asarray(voxel_position)
        voxel = (v_array[0], v_array[1], v_array[2])
        neighbour = (voxel[0]+k[0], voxel[1]+k[1], voxel[2]+k[2])
        nb_array = np.asarray(neighbour)
        # check if neighbor position lies within volume and if voxel is
            valid
        if (nb_array >= 0).all() and (nb_array < np.asarray(volume.shape
            )).all() and not(np.array_equal(nb_array, v_array)) and np.
            isfinite(volume[neighbour]):
            w += 1
            av += volume[neighbour]

    if w != 0:
        return np.abs(volume[voxel] - av/w)
    else: return 0

def compute_NGTDM(img, N_g, distance):

    N_g = int(N_g)

    # for each gray level: get average differences to  the respective
        neighborhood
    # discriminate between 3D and 2D inputs
    if len(img.shape)==3:
        s_i = [[get_neighborhood_average_difference_3D(img, j, distance)
            for j in np.transpose(np.where(img == i+1))] for i in range
            (N_g)]
    else:
        s_i = [[get_neighborhood_average_difference_2D(img, j, distance)
            for j in np.transpose(np.where(img == i+1))] for i in range
            (N_g)]

    # gray level counts
    n_i = [len(s_i[i]) for i in range(N_g)]
    # sum up all average differences of the same gray level
    s_i = [np.sum(s_i[i]) for i in range(N_g)]
    N_v = np.sum(n_i)
    # handle slices where no gray levels are assigned (due to specific
        shapes of masks)
```

```
    if N_v == 0:
        p_i = [0 for i in range(N_g)]
    else:
        # calculate gray level probabilies
        p_i = [n_i[i]/float(N_v) for i in range(N_g)]

    NGTDM = np.zeros((N_g,3), dtype=float)
    NGTDM[:,0] = n_i
    NGTDM[:,1] = p_i
    NGTDM[:,2] = s_i

    return NGTDM
```

12.16. NGTDM_features.py

The module includes utilities to compute NGTD features as proposed by the IBSI collaboration [184].

- `compute_NGTDM_features(NGTDM)`:
    acquires all 5 features based on the `NGTDM` 2D-array object.

- `get_NGTDM_feature_names()`:
    obtains the name of the 5 features computed in `compute_NGTDM_features`.

12.17. matrices.py

The module includes the modules `GLCM`, `GLRLM`, `GLSZM`, `GLDZM` and `NGTDM` to compute final texture matrices. It also confirms that only 2D and 3D array objects are being used as inputs for the construction of matrices.

- `all_cooccurrences(img, N_g)`:
    calculates direction-dependent GLCMs for given 3D -array obejct `img`, averages them into a single invariant GLCM, which is normalized subsequently to obtain probability values.

- `all_run_lengths(img, N_g)`:
    calculates direction-dependent GLRLMs for 3D -array obejct `img` and averages them into a single invariant GLRLM.

- `all_size_zones(img, N_g)`:
    computes the GLSZ for a given 3D-array object `img`.

- `all_distance_zones(img, N_g)`:
    computes the GLDZ for a given 3D-array object `img`.

- `all_neighbourhood_differences(img, N_g, distance):`
  calculates the NGTDM for given 3D-array object `img` and neighbourhood radius `distance`.

- `combined_matrix(matrix_list):`
  merges all matrices in `matrix_list` into a single matrix of the same shape by summing entries over all matrices.

- `average_matrix(matrix_list):`
  merges all matrices in `matrix_list` into a single matrix of the same shape by averaging entries over all matrices.

- `combined_NGTDM(matrix_list):`
  merges all NGTDMs in `matrix_list` into a single NGTDM of the same shape by summing all entries, and calculating subsequently proportion entries.

12.18. Wavelet_transforms.py

The module includes utilities to transform a 3D-array object into its wavelet representation via one-level undecimated wavelet transform as proposed by Aerts et al [5]. The module is supported on the built-in library *PyWavelets* which module `swtn` is used.

- `compute_uwt3D(img):`
  computes Undecimated Wavelet transformed of the given 3D-array object.

Listing 12.15: Python code for generations of a wavelet representation of a 3D- object, Wavelet_transform.py.

```python
def compute_uwt3D(img):
    img = img.copy()
    s = np.shape(img)
    # assures that 3D-array objects do not have nan entries
    if np.nanmin(img) > 0:
        img[np.isnan(img) == True] = 0.0

    # assures that original input img has a quadratic shape, otherwise
        computation impossible
    if s[0]%2 == 1:
        img_n = img[:s[0]-1,:,:]
    else:
        img_n = img.copy()

    if s[1]%2 == 1:
        img_n = img_n[:,:s[1]-1,:]

    if s[2]%2 == 1:
```

```
        img_n = img_n[:,:,:s[2]−1]

    # computes Wavelets transformation of input
    WT = swtn(img_n, wavelet='coif1', level=1)
    return WT[0]
```

12.19. feature_extraction.py

The module includes methods for feature extraction methods based on the module `matices`, `Stats_features`, `Morph_features` for extraction of Intensity-, Shape/Morphology- and Texture- family features.

- `compute_3D_texture(volume)`:
    implements the 3D approach. Constructs all texture matrices of the given volume and extracts all 78 texture features from the resulting matrices.

- `compute_3D_int(volume)`:
    extracts or summarize in a matrix-like form output the intensity features.

- `compute_3D_morph(volume, vox_vol)`:
    extracts or summarize in a matrix-like form output the shape features.

12.20. computes_CTs.py

The file contains main instructions to run calculations.

- `preprocess_data()`:
    generates the instruction to adapt and process data for the build-in libraries and the python environment.

Listing 12.16: Python code for image data preparation and processing into the Python environment, compute_CTs.py.

```
def preprocess_data():
#    obtains directory folder where the data is stored
    dirs = os.listdir(path_to_database)
    # loop initialization for every image of patients
    for d in sorted(dirs):
        path = os.path.join(path_to_database,d)
        if os.path.isdir(path):
            name = os.path.basename(path)
            print(name)
            # prepare data for intensity and shape feature acquisition
            data_int = readimg.process_patient_data_for_int(path,
                rel_path_to_CT_data, rel_path_to_seg_data)
            # prepare data for texture feature acquisition
```

```
data_tex = readimg.process_patient_data_for_tex(path,
    rel_path_to_CT_data, rel_path_to_seg_data)
# prepare data for wavelet transformation and intensity WT
    feature acquisition
data_wt, filters = readimg.process_patient_data_for_wt(path,
    rel_path_to_CT_data, rel_path_to_seg_data)
# prepare data for texture wavelet acquisition
data_wt_tex, filters_tex = readimg.
    process_patient_data_for_wt_tex(path,
    rel_path_to_CT_data, rel_path_to_seg_data)
# store preprocessed data in a numpy object format for
    further calculations
np.save(os.path.join(path_to_processed_folder_tex, name),
    data_tex[0])
np.save(os.path.join(path_to_processed_folder_vox_spa, name)
    , data_tex[1])
np.save(os.path.join(path_to_processed_folder_int, name),
    data_int[0])
np.save(os.path.join(path_to_processed_folder_int_wt, name),
    data_wt)
np.save(os.path.join(path_to_processed_folder_filters, name)
    , filters)
np.save(os.path.join(path_to_processed_folder_tex_wt, name),
    data_wt_tex)
np.save(os.path.join(path_to_processed_folder_filters_text,
    name), filters_tex)
```

- `compute_img_features(data, dim)`:
  generates instructions for first-order-statistic feature computations of all data images `data` and preferred dimension approach `dim`

Listing 12.17: Python code for set of instructions of intensity family feature acquisition, compute_CTs.py.

```
def compute_img_features(data, dim):
        # acquires intensity feature names
    feature_names = f_stat.get_stats_feature_names()+ f_ih.
        get_ih_feature_names() + f_ivh.get_ivh_feature_names()
        # initialize acquisition of features values and present them in
            a matrix-like form
    all_features = np.zeros((len(data),len(feature_names)))

    # compute all features for each image file
    os.chdir(path_to_result_folder)
    for i in range(len(data)):

        CT_img = data[i]

        # differs between 2D and 3D
        if dim == 2:
```

```
        feature_values = feature_extraction.compute_2D(CT_img)
    else:
        feature_values = feature_extraction.compute_3D_int(CT_img)

    all_features[i,:] = feature_values


    print(str(i+1) +'_finished')

return feature_names, all_features
```

- `compute_morph_features(data, vox_vol, dim)`:
  generates instructions for shape feature computations of all data images `data`,
  accounting for their voxel size `vox_vol` and preferred dimension approach `dim`.

Listing 12.18: Python code for set of instructions of morphology family feature acquisition, compute_CTs.py.

```
def compute_morph_features(data, vox_vol, dim):

    feature_names = f_morph.get_morph_feature_names()

    all_features = np.zeros((len(data),len(feature_names)))

    # compute all features for each image file
    os.chdir(path_to_result_folder)
    for i in range(len(data)):

        CT_img = data[i]
        vox_vol_item = vox_vol[i]

        # discriminate approach
        if dim == 2:
            feature_values = feature_extraction.compute_2D(CT_img)
        else:
            feature_values = feature_extraction.compute_3D_morph(CT_img,
                vox_vol_item)

        all_features[i,:] = feature_values

        print(str(i+1) +'_finished')

    return feature_names, all_features
```

- `compute_texture_features(data, dim)`:
  generates instructions for texture feature computations of all data images `data`
  and preferred dimension approach `dim`.

- `compute_img_WT_features(img, filters, dim)`:
  generates instructions for wavelet intensity feature computations of all data images `img`, accounting for their filter `filters` and preferred dimension approach `dim`.

Listing 12.19: Python code for set of instructions of wavelets intensity family feature acquisition, compute_CTs.py.

```python
def compute_img_WT_features(img, filters, dim):

    # load processed images of all patients


    names = f_stat.get_stats_feature_names()+ f_ih.get_ih_feature_names
        () + f_ivh.get_ivh_feature_names()
    # 8 due to 8 filters HHH, HHL, HLL, LLL, LHL, LHH, HLH, LLH

    feature_names = []

    for item in filters[0]:
        feature_names.append([item + '_' + x for x in names])

    feature_names = np.array(feature_names)
    feature_names = feature_names.flatten()


    all_features = np.zeros((len(img),len(feature_names)))
    for i in range(len(img)):


        data = img[i]
        # Run over the different filters in the data

        feature_values = np.array([])
        for item in range(len(data)):

            CT_img = data[item].copy()

            # apply the selected dimension analysis approach
            if dim == 2:
                feature_values = np.append(feature_values,
                    feature_extraction.compute_2D(CT_img))
            else:
                feature_values = np.append(feature_values,
                    feature_extraction.compute_3D_int(CT_img))

        all_features[i,:] = feature_values
        print(str(i+1) +'_finished')

    return feature_names, all_features
```

- `compute_text_WT_features(img, filters ,dim)`:
  generates instructions for wavelet texture feature computations of all data images `img`, accounting for their filter `filters` and preferred dimension approach `dim`.

- `core()`: runs the programme and store computations in a pandas `DataFrame` object for further analysis of different features.

Listing 12.20: Python code for set of instructions to run the programme, compute_CTs.py.

```python
def core():
    if preprocess:
        preprocess_data()

    # acquire folder files and patients anonymization name
    files = [f for f in os.listdir(path_to_processed_folder_int) if os.
        path.isfile(os.path.join(path_to_processed_folder_int, f))]
    files_1 = sorted(files)
    Patients = [f.replace('.npy','') for f in files_1]

# get numpy arrays from patients (for intensity and morphology)
    data_no_proc = readimg.get_processed_folder(
        path_to_processed_folder_int)
# get voxels size from patients
    vox_vol = readimg.get_processed_folder(
        path_to_processed_folder_vox_spa)
# get numpy arrays from patients (for texture)
    data_proc = readimg.get_processed_folder(
        path_to_processed_folder_tex)
# gets numpy arrays for int WT features
    data_wt_no_proc = readimg.get_processed_folder(
        path_to_processed_folder_int_wt)
# gets numpy arrays of filters
    filters = readimg.get_processed_folder(
        path_to_processed_folder_filters)
# gets numpy array for texture wt features
    data_tex_wt = readimg.get_processed_folder(
        path_to_processed_folder_tex_wt)

    filters_tex = readimg.get_processed_folder(
        path_to_processed_folder_filters_text)

# Obtain names and values of Intesity family of Patients
    int_names, intensity_features = compute_img_features(data_no_proc,
        3)
# Transforms intensity features and save numpy array object of features
    to a DataSeries object
    Frame = pd.DataFrame(intensity_features)
    Frame.columns = int_names
```

```
        Frame.index = Patients

        os.chdir(path_to_result_folder)
        Frame.to_csv('Intensity_based_Rad_features.csv')

# Transforms shape features and save numpy array object of features to a
    DataSeries object
        morph_names, morph_features = compute_morph_features(data_no_proc,
            vox_vol, 3)

        Morph_frame = pd.DataFrame(morph_features)
        Morph_frame.columns = morph_names
        Morph_frame.index = Patients

        os.chdir(path_to_result_folder)
        Morph_frame.to_csv('Morph_based_rad_features.csv')

# Transforms texture features and save numpy array object of features to
    a DataSeries object
        text_names, text_features = compute_texture_features(data_proc, 3)

        Text_frame = pd.DataFrame(text_features)
        Text_frame.columns = text_names
        Text_frame.index = Patients

        os.chdir(path_to_result_folder)
        Text_frame.to_csv('Texture_based_rad_features.csv')

# Transforms intensity wavelets features and save numpy array object of
    features to a DataSeries object
        int_wt_names, int_wt_features = compute_img_WT_features(
            data_wt_no_proc, filters, 3)

        WT_int_frame = pd.DataFrame(int_wt_features)
        WT_int_frame.columns = int_wt_names
        WT_int_frame.index = Patients

        os.chdir(path_to_result_folder)
        WT_int_frame.to_csv('Wavelets_intensity_based_rad_features.csv')

# Transforms texture wavelets features and save numpy array object of
    features to a DataSeries object
        text_wt_names, text_wt_features = compute_text_WT_features(
            data_tex_wt, filters_tex, 3)

        WT_tex_frame = pd.DataFrame(text_wt_features)
        WT_tex_frame.columns = text_wt_names
        WT_tex_frame.index = Patients

        os.chdir(path_to_result_folder)
        WT_tex_frame.to_csv('Wavelets_texture_based_rad_features.csv')
```

12.21. Imaging preprocessing and feature extraction software summary

As a summary, the figure 12.1 provides the module dependency inside the architecture of the software for imaging processing and feature extraction.

"path/to/nifti/data"

compute_CTs

Wavelets_transform

read_img_data

preprocessing

GLCM

GLRLM

sequences

matrices

GLSZM

Stats_features

GLDZM

Intensity_histogram_features

NGTDM

feature_extraction

Intensity_volume_histogram_features

Morph_features

GLCM_features

GLRLM_features

GLSZM_features

GLDZM_features

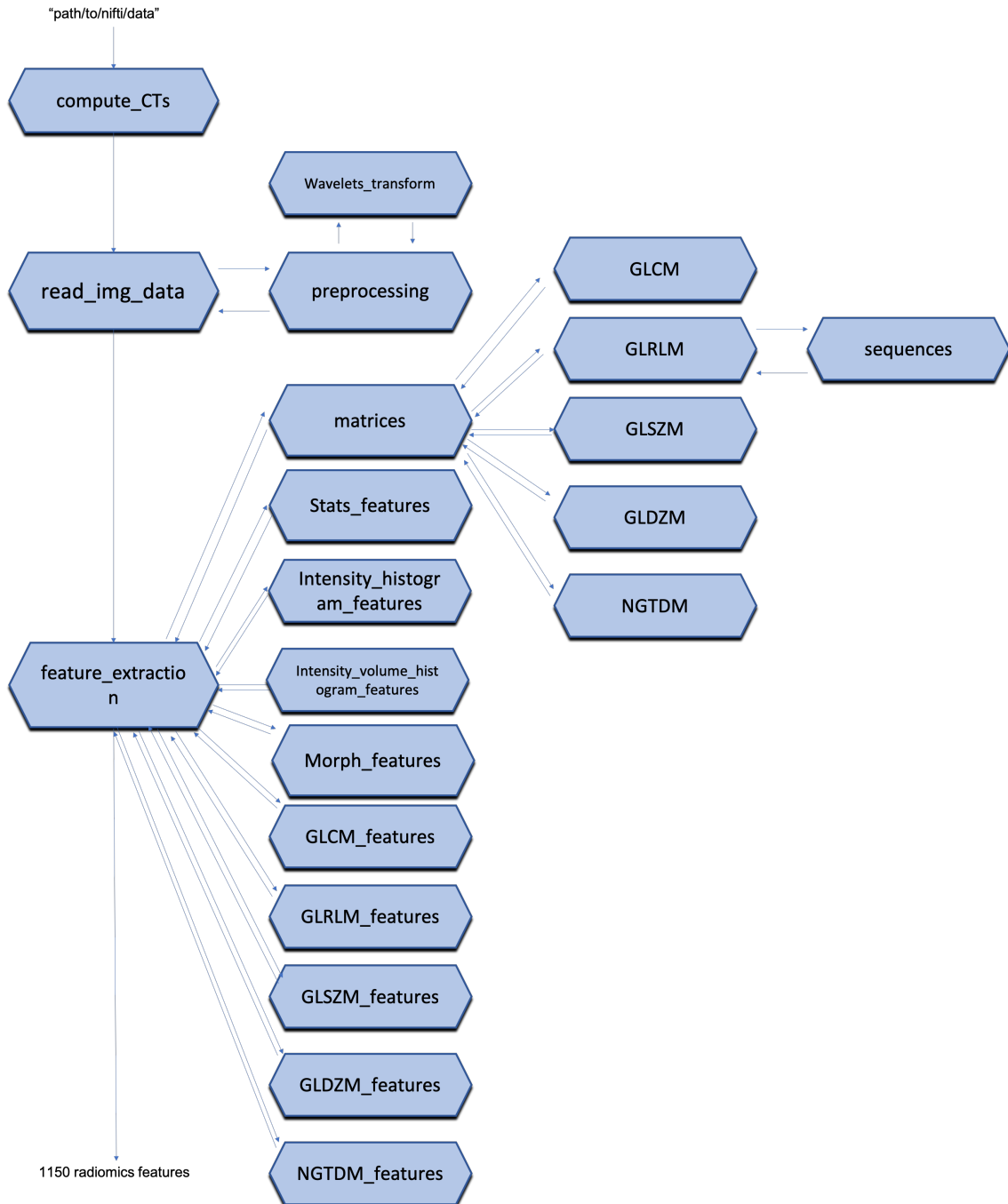1150 radiomics features

NGTDM_features

Figure 12.1.: Diagram of modules dependency architecture.

13. Software documentation: Feature selection and ML models

In this chapter full disclosure of the radiomic implementations for feature selection and ML modelling are discussed. The implementations follows the theory exposed in section 4.

13.1. radiomics_functions.py

This module is a class object that provides functions for feature preprocessing tasks, such as PCA applications and feature correlation assessments. it uses the built-in library *Scikit-learn* and *Pandas* and *Numpy*.

- `f_corr_assestment(self, df, d=1.0):`
  assesses correlations between features from the DataFrame object input data `df`, accounting by the threshold correlation `d`, which by default is set to 1 or 100% pearson correlation coefficient.

Listing 13.1: Python code for correlation assessment of an DataFrame object, radiomics_functions.py.

```python
def f_corr_assestment(self, df, d=1.0):
        # extract correlation matrix from the pandas method.
    corrMatrix = df.corr()
    corrMatrix.loc[:,:] = np.tril(corrMatrix, k=-1)

    already_in = set()
    result = []
    for col in corrMatrix:
        perfect_corr = corrMatrix[col][corrMatrix[col] >= d].index.
            tolist()
        if perfect_corr and col not in already_in:
            already_in.update(set(perfect_corr))
            perfect_corr.append(col)
            result.append(perfect_corr)
    return result
```

- `apply_PCA_in_f(self, X_train, X_test, cluster):`
  applies the principal component algorithm imported from *Scikit-learn* to training (`X_train`) and test (`X_test`) DataFrame objects using the highly correlated feature lists `cluster`.

Listing 13.2: Python code for PCA applications of DataFrame objects, radiomics_functions.py.

```python
def apply_PCA_in_f(self, X_train, X_test, cluster):
        # Initialize storage of pca features
    X_train_pca = np.zeros((X_train.shape[0], len(cluster)))
    X_test_pca = np.zeros((X_test.shape[0], len(cluster)))
        # Initialize PCA method from scikit learn to combine
            correlating feature to one single component
    pca = PCA(n_components = 1)

    # for each patient in 'data': apply PCA to each of the clusters
        in 'clustered_features'
    for i in range(len(cluster)):
        comp1, comp2 = [X_train[cluster[i]], X_test[cluster[i]]]
        comp1 = pca.fit_transform(comp1)
        comp2 = pca.transform(comp2)
        X_train_pca[:,i] = comp1[:,0]
        X_test_pca[:,i] = comp2[:,0]

    # Translate the numpy object to the original pandas object
    X_train_pca = pd.DataFrame(X_train_pca)
    # Rename features
    names = []
    for j in range(len(cluster)):
        names.append(cluster[j][-1])
    X_train_pca.columns = names

    X_test_pca = pd.DataFrame(X_test_pca)
    X_test_pca.columns = names

    return X_train_pca, X_test_pca
```

13.2. filter_selection.py

The module contains different utilities for filter feature selection methods (cf. section 4.3). It uses the functions *chi2*, *f_classif* and *mutual_info_classif* from the built-in module *Scikit-learn.feature_selection* as well as *mutual_info_score* from *Scikit-learn.fmetrics* and *ScyPy.stats*.

- `pearson_scorer(X, y)`:
    computes pearson correlation coefficient between two `ndarrays`, X and y.

- `kendall_score(X, y)`:
    computes kendall correlation coefficient between two `ndarrays`, X and y.

- `spearman_score(X, y)`:
    computes spearman correlation coefficient between two `ndarrays`, X and y.

- `mutual_info_feature_selection(X, y):`

    computes mutual information feature selection relevance coefficient between two `ndarrays`, `X` and `y`.

Listing 13.3: Python code for computation of mutual information feature selection coefficient of DataFrame objects, filter_selection.py.

```python
def mutual_info_feature_selection(X, y):
# initialize relevance score per feature
    P = np.array([])
    for i in range(X.shape[1]):
        temp = 0
        for j in range(X.shape[1]):
            if j != i:
                # computes mutual information coefficient across feature
                    space
                temp += mutual_info_score(X[:, [i]], X[:, [j]])
        # computes relevance score accordingly
        P = np.append(P, mutual_info_score(X[:, [i]], y.reshape(-1, 1)) -
            temp)
    return P
```

- `conditional_mutual_information(X, Y, Z):`

    calculates conditional mutual information coefficient relevance coefficient between three `ndarrays`, `X`, `Y` and `Z`.

Listing 13.4: Python code for computation of conditional mutual information coefficient of DataFrame objects, filter_selection.py.

```python
def conditional_mutual_information(X, Y, Z):
        # Initialize score per feature
    P = np.array([])
    # initialize number of bins for histogram in target label classes
    n_bins_z = len(np.unique(Z))
    for i in range(X.shape[1]):
        I = 0
        for j in range(Y.shape[1]):
            if j != i:
                n_bins_x = len(Z)
                n_bins_y = len(Z)
                argument_xyz = np.concatenate((X[:, [i]], Y[:, [j]], Z.
                    reshape(-1,1)), axis = 1)
                argument_xz = np.concatenate((X[:, [i]], Z.reshape(-1,1)
                    ), axis = 1)
                argument_yz = np.concatenate((Y[:, [j]], Z.reshape(-1,1)
                    ), axis = 1)
                # construct conditional joint probability distribution
                p_xyz, _ = np.histogramdd(argument_xyz, bins = [n_bins_x
                    , n_bins_y, n_bins_z])
```

```
                     # contruct conditional probability distribution
                     p_xz, _, _ = np.histogram2d(argument_xz[:, 0],
                          argument_xz[:, 1], bins = [n_bins_x, n_bins_z])
                     p_yz, _, _ = np.histogram2d(argument_yz[:, 0],
                          argument_yz[:, 1], bins = [n_bins_y, n_bins_z])
                     # construct probability distribution
                     p_z, _ = np.histogram(Z, bins = n_bins_z)
                     p_xyz = p_xyz/np.sum(p_xyz)
                     p_xz = p_xz/np.sum(p_xz)
                     p_yz = p_yz/np.sum(p_yz)
                     p_z = p_z/np.sum(p_z)
                     # computes argument of the equation for conditional
                          mutual information coefficient
                     p_xy_z = p_z*p_xyz
                     log_p_xy_z = p_xyz*np.log10(p_xy_z)
                     temp = p_xz*p_yz
                     log_temp = p_xyz*np.log10(temp)
                     a = log_p_xy_z - log_temp
                     # computes coefficient for numerical values
                     I += np.sum(a[np.isfinite(a)])
           P = np.append(P, I)
     return P
```

- `joint_mutual_information(X, y)`:
  calculates the joint mutual information coefficient based on the
  `conditional_mutual_information` function.

- `conditional_infomax(X, y)`:
  computes the conditional infomax feature extraction coefficient based on the
  `conditional_mutual_information` and `mutual_info_feature_selection` functions.

13.3. feature_selection.py

The module compile algorithms to stratify features according to feature selection
approaches. It is built up on top of *Scikit-learn* functions and utilities.

- `select_features_corr(X_train, y_train, feature_names, n_features, \`
  `training_size, n_rounds, stat_test, class_weight = None)`:
  selects features according to chosen filter method (`stat_test`) with bootstrap-
  ping for feature stability, one can select the extend of samples `training_size`,
  the number of resample times `n_rounds` and the maximum number of features
  `n_features`

Listing 13.5: Python code for selection of features according to relevance indexes of
DataFrame objects, feature_selection.py.

```python
def select_features_corr(X_train, y_train, feature_names, \
 n_features, training_size, n_rounds, stat_test, class_weight = None):
        # Chi-square works only for positive values, so escalation
            accordingly is needed.
        if stat_test == chi2:
        slc = MinMaxScaler()
        X_train = slc.fit_transform(X_train)

    # Initialize feature relevance matrix per resample iteration
    total_selected_features_per_round = np.empty([n_features, n_rounds],
        dtype='|S600')

    for j in range(n_rounds):
        # Activates balanced or unbalanced lable classes for binary
            problems
        if (class_weight == 'balanced') & (len(y_train < 0.5) > len(
            y_train > 0.5)):
            X_positive, y_positive = X_train[y_train == 1], y_train[
                y_train == 1]
            X_negative, y_negative = shuffle(X_train[y_train == 0],
                y_train[y_train == 0], n_samples = len(y_positive))
            X_train = np.append(X_positive, X_negative, axis=0)
            y_train = np.append(y_positive, y_negative)
        elif (class_weight == 'balanced') & (len(y_train > 0.5) > len(
            y_train < 0.5)):
            X_negative, y_negative = X_train[y_train == 0], y_train[
                y_train == 0]
            X_positive, y_positive = shuffle(X_train[y_train == 1],
                y_train[y_train == 1], n_samples = len(y_negative))
            X_train = np.append(X_positive, X_negative, axis = 0)
            y_train = np.append(y_positive, y_negative)

        # resample train and test data sets randomly
        X, y = shuffle(X_train, y_train)
        # Initialize feature selection object according to filter
            function and maximum number of features allowed
        feature_selector = GenericUnivariateSelect(score_func=stat_test,
            mode='k_best', param=n_features)
        # Chooses the number of samples to train
        number_of_samples = int(round(training_size*len(y)) + 1)
        X_train_boots, y_train_boots = X[:number_of_samples], y[:
            number_of_samples]
        feature_selector.fit(X_train_boots, y_train_boots)
        # Obtain indexes of the most important features according to
            filter function
        indexes_fs = np.flipud(np.argsort(feature_selector.scores_[:,0])
            )
        indexes_fs = indexes_fs[:n_features]
        # translate indexes to feature names
        selected_features = np.array([])
```

```
        for i in indexes_fs:
            selected_features = np.append(selected_features,
                feature_names[i])
        scores = np.flipud(np.sort(feature_selector.scores_[:,0]))
        scores = scores[:n_features]
        total_selected_features_per_round[:,j] = selected_features
        # Excludes feature repetition in resampling iteration
    all_relevant_features = np.unique(total_selected_features_per_round)
    # Computes feature stability relevance
    feature_relevance = {}
    for item in all_relevant_features:
        rank, round_of_appereance = np.where(
            total_selected_features_per_round == item)
        n_of_appereance_feature = len(rank)
        relevance_score = n_of_appereance_feature*(1.0/np.mean(rank + np
            .ones(len(rank))))*(1.0/(np.std(rank) + 1.0))/n_rounds
        feature_relevance[item] = relevance_score

        # Initialise computation of most stable and relevant features
    relevant_features = []
    for key, value in sorted(feature_relevance.items(), key = lambda k:
        (k[1], k[0])):
        key = key.decode('UTF-8')
        relevant_features.append(key)

    relevant_features = relevant_features[::-1]
    return relevant_features[:n_features]
```

- `select_features_model(X_train, y_train, n_features, cv, model, \`
    `model_name):`
    selects features according to chosen model object (`model`) and name (`model_name`)
    with cross validation (`cv`) for feature stability. `n_features` refers to the maximum number allowed for features.

Listing 13.6: Python code for selection of features according to model of DataFrame objects, feature_selection.py.

```
def select_features_model(X_train, y_train, n_features, cv, mode,
    model_name):
        # Initialise sequetial forward selection instance
        sfs = SFS(estimator = model, k_features = n_features, forward =
            True, floating = False, scoring = 'roc_auc', cv= cv)
        # fit the object class to the train feature space and outcome
        sfs.fit(X_train, y_train)
        # returns feature name imporances
        return np.array(sfs.k_feature_names_)
```

13.4. ML_models.py

The module includes methods to optimise models according to desired number of features and hyper-parameters from models discussed in section 4.4. It uses utilities from *Scikit-learn.model_ selection* and *Mlxtend.feature_ selection.*

- `model_optimization_inputs(model, train_x, train_y, n_features, \str_params, cv)`:
  obtains combinations of number of features and proposed model hyper-parameters (`str_param`) that performs the best in the desired `model`, using the training feature space `train_x` and the training target `train_y`.

Listing 13.7: Python code for model selection according to model of DataFrame objects, ML_models.py.

```python
def model_optimization_inputs(model, train_x, train_y, n_features,
    str_param, cv):

        # Proposed hyper-parameter search according to ML model
        param_grid = {'LR': {'estimator__C': 10.0*np.arange(-3,4,1), '
            estimator__class_weight': [{0: 1, 1: 2}, {0: 1, 1: 3}, {0:
            1, 1: 4}, {0: 1, 1: 5}],}, 'KNN': {'estimator__n_neighbors':
             np.arange(3, 27, 2)}, 'SVM': {'estimator__C': 10.0*np.
            arange(-3,4,1),     'estimator__class_weight': [{0: 1, 1: 2},
             {0: 1, 1: 3}, {0: 1, 1: 4}, {0: 1, 1: 5}],}, 'RF': {'
            estimator__n_estimators': np.arange(3, 11, 2), '
            estimator__max_depth': np.arange(5,11),  '
            estimator__class_weight': [{0: 1, 1: 2}, {0: 1, 1: 3}, {0:
            1, 1: 4}]}    'GNB': {},        'DT': {'estimator__max_depth':
             [7, 8, 9, 10], 'estimator__class_weight': [{0: 1, 1: 2},
            {0: 1, 1: 3}, {0: 1, 1: 4}, {0: 1, 1: 5}]},}

        # Initialise storage of best performing inputs
        optimization_inputs = []

        # Grid search construction of hyper parameters
        grid = ParameterGrid(param_grid[str_param])
        # Initialise sequetial forward selection instance with cross
            validation and performance metric roc-auc
        sfs = SFS(estimator = model, k_features = n_features, forward =
            True, floating = False, scoring = 'roc_auc', cv= cv)
        # Iterates over possible parameters in the grid
        for params in grid:
                # Changes model hyperparameters accordingly
                sfs.set_params(**params)
                # fit the object class to the train feature space and
                    outcome
                sfs.fit(X_train, y_train)
                scores = sfs.subsets_
                new_scores = dict()
```

```
                    for i in range(1, n_features + 1):
                        i = int(i)
                        new_scores[i] = scores[i]
                        new_scores[i]['std_score'] = np.std(scores[i]['
                            cv_scores'])

                new_scores = pd.DataFrame(new_scores)
                new_scores = new_scores.T
                ordered_new_scores = new_scores.sort_values(['avg_score'
                    , 'std_score'], ascending = [False, True])
                extraction_best = ordered_new_scores.iloc[0]
                extraction_best['parameters'] = params
                optimization_inputs = pd.concat([optimization_inputs,
                    extraction_best.T], axis = 0)

        # returns feature name imporances
        return optimization_inputs
```

13.5. Rad_analysis_ML.py

The file contains mean instruction to run radiomics ML analysis. As a summary, figure 13.1 exemplify module dependency inside the architecture of the software for feature selection and machine learning modelling.
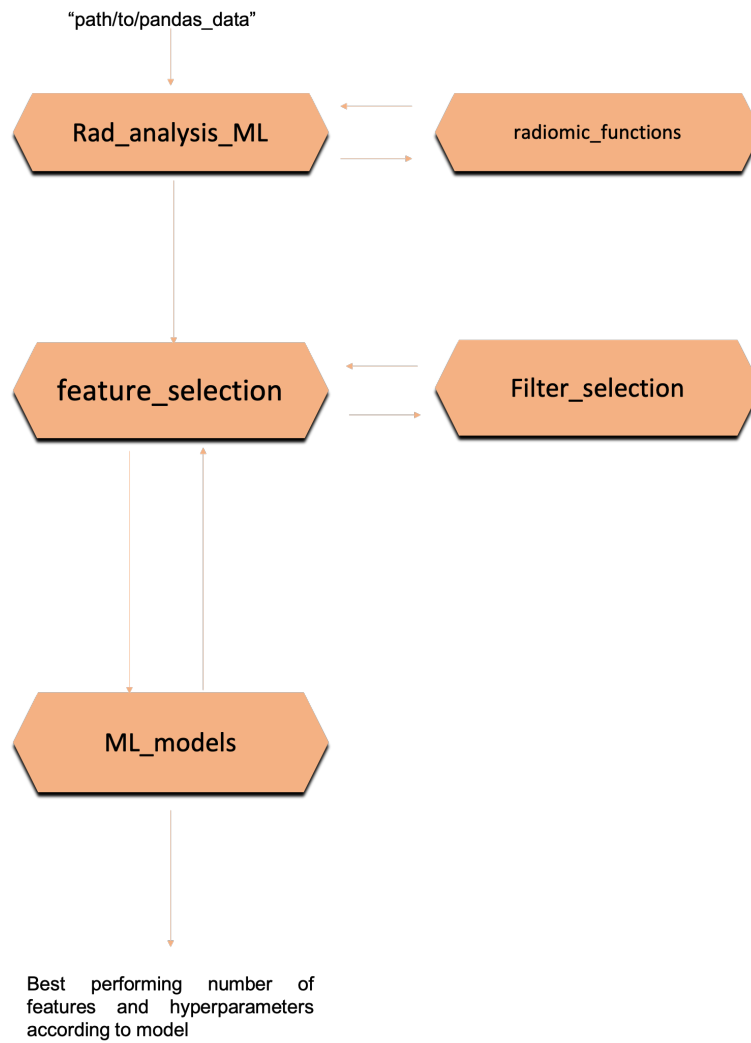
Figure 13.1.: Diagram of modules dependency architecture for ML purposes.

Declaration of the framework of the collective work

Parts of this thesis are published or pending for publication in different peer-to-peer review journals. In the following, the contributions of each author are specified.

I. Radiogenomics in head and neck cancer: correlation of radiomic heterogeneity and somatic mutations in TP53, FAT1 and KMT2D

Kerstin Zwirner	first author, conception and execution of the project, production of the manuscript.
Franz J. Hilke	genetic data acquisition and analysis.
German Demidov	genetic data acquisition and analysis.
Jairo Socarras Fernandez	image preprocessing and radiomic feature extraction.
Stephan Ossowski	genetic data acquisition and analysis.
Cihan Gani	critical discussion of the results, proofreading of the manuscript.
Daniela Thorwarth	supervision of the radiomic analysis of the project.
Olaf Riess	supervision of the genetic analysis of the project.
Daniel Zips	critical discussion of the results, proofreading of the manuscript
Christopher Schroeder	supervision of the genetic analysis of the project
Stefan Welz	supervision of the project, conception and design of the project.

II. Comparison of patient stratification by CT radiomics and hypoxia PET in head-and-neck cancer RT

Jairo Andrés Socarrás Fernández	first author, conception and execution of the project, production of the manuscript.
David Mönnich	18F-FMISO-PET/CT data acquisition and analysis.
Sara Leibfarth	critical discussion of the results, proof-reading of the manuscript
Alex Zwanenburg	guidelines for medical image analysis and ML.
Stefan Leger	guidelines for medical image analysis and ML.
Steffen Löck	guidelines for medical image analysis and ML.
Stefan Welz	18F-FMISO-PET/CT data acquisition
Christina Pfannenberg	18F-FMISO-PET/CT data acquisition
Christian la Fougère	18F-FMISO-PET/CT data acquisition
Gerald Reischl	18F-FMISO-PET/CT data acquisition
Michael Baumann	critical discussion of the results, proof-reading of the manuscript
Daniel Zips	critical discussion of the results, proof-reading of the manuscript
Daniela Thorwarth	supervision of the project, conception and design of the project.

III. A radiomic approach for prediction of complete remission of organ-preserving therapy in rectal carcinoma with independent validation

Pierluigi Bonomo	first author, conception and execution of the project, data acquisition.
Cihan Gani	data acquisition, conception and execution of the project
Jairo Andrés Socarrás Fernández	image analysis, radiomic pipeline implementation and production of the manuscript.
Daniela Thorwarth	critical discussion of the results, proof-reading of the manuscript
Daniel Zips	supervision of the project, conception and design of the project.

Bibliography

[1] *Spearman Rank Correlation Coefficient*, pages 502–505. Springer New York, New York, NY, 2008.

[2] *Nuclear Medicine Imaging Modalities*, chapter 6, pages 139–155. John Wiley Sons, Ltd, 2010.

[3] Hervé Abdi. The Kendall rank correlation coefficient. *Encyclopedia of Measurement and Statistics. Sage, Thousand Oaks, CA*, pages 508–510, 2007.

[4] Michael C Adams, Timothy G Turkington, Joshua M Wilson, and Terence Z Wong. A systematic review of the factors affecting accuracy of suv measurements. *American Journal of Roentgenology*, 195(2):310–320, 2010.

[5] Hugo JWL Aerts, Emmanuel Rios Velazquez, Ralph TH Leijenaar, Chintan Parmar, Patrick Grossmann, Sara Carvalho, Johan Bussink, René Monshouwer, Benjamin Haibe-Kains, Derek Rietveld, et al. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature communications*, 5:4006, 2014.

[6] Fritz Albregtsen, Birgitte Nielsen, and Håvard E Danielsen. Adaptive gray level run length features from class distance matrices. In *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, volume 3, pages 738–741. IEEE, 2000.

[7] Moses Amadasun and Robert King. Textural features corresponding to textural properties. *IEEE Transactions on systems, man, and Cybernetics*, 19(5):1264–1274, 1989.

[8] Amine Amyar, Su Ruan, Isabelle Gardin, Romain Herault, Chatelain Clement, Pierre Decazes, and Romain Modzelewski. Radiomics-net: Convolutional neural networks on fdg pet images for predicting cancer treatment response. *Journal of Nuclear Medicine*, 59(supplement 1):324–324, 2018.

[9] Athanassios Argiris, Michalis V Karamouzis, David Raben, and Robert L Ferris. Head and neck cancer. *Lancet (London, England)*, 371(9625):1695–709, 2008.

[10] Rajamanickam Baskar, Kuo Ann Lee, Richard Yeo, and Kheng-Wei Yeoh. Cancer and radiation therapy: current advances and future directions. *International journal of medical sciences*, 9(3):193, 2012.

[11] Roberto Battiti. Using mutual information for selecting features in supervised neural net learning. *IEEE Transactions on neural networks*, 5(4):537–550, 1994.

[12] Regina G H Beets-Tan, Jeroen Leijtens, Laurentius Ziekenhuis, and Geerard L Beets. Wait-and-see policy for clinical complete responders after chemoradiation for rectal cancer. 2011.

[13] Niha Beig, Jay Patel, Prateek Prasanna, Sasan Partovi, Vinay Varadan, Anant Madabhushi, and Pallavi Tiwari. Radiogenomic analysis of hypoxia pathway reveals computerized mri descriptors predictive of overall survival in glioblastoma. In *Medical Imaging 2017: Computer-Aided Diagnosis*, volume 10134, page 101341U. International Society for Optics and Photonics, 2017.

[14] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. *Noise reduction in speech processing*, volume 2. Springer Science & Business Media, 2009.

[15] Thierry Berghmans, Michele Dusart, Marianne Paesmans, Claude Hossein-Foucher, Irene Buvat, Catherine Castaigne, Arnaud Scherpereel, Céline Mascaux, Michel Moreau, Martine Roelandts, et al. Primary tumor standardized uptake value (suvmax) measured on fluorodeoxyglucose positron emission tomography (fdg-pet) is of

prognostic value for survival in non-small cell lung cancer (nsclc): a systematic review and meta-analysis (ma) by the european lung cancer working party for the iaslc lung cancer staging project. *Journal of Thoracic Oncology*, 3(1):6–12, 2008.

[16] Jacques Bernier, Eric J. Hall, and Amato Giaccia. Radiation oncology: a century of achievements. *Nature Reviews Cancer*, 4:737 EP –, 09 2004.

[17] Jonathan M Bernstein, Jarrod J Homer, and Catharine M West. Dynamic contrast-enhanced magnetic resonance imaging biomarkers in head and neck cancer: potential to guide treatment? a systematic review. *Oral oncology*, 50(10):963–970, 2014.

[18] Wolfgang Birkfellner. *Applied medical image processing: a basic course.* CRC Press, 2016.

[19] Christopher M Bishop. *Pattern recognition and machine learning.* springer, 2006.

[20] Marta Bogowicz, Oliver Riesterer, Luisa Sabrina Stark, Gabriela Studer, Jan Unkelbach, Matthias Guckenberger, and Stephanie Tanadini-Lang. Comparison of pet and ct radiomics for prediction of local tumor control in head and neck squamous cell carcinoma. *Acta oncologica*, 56(11):1531–1536, 2017.

[21] Carol R. Bradford, Shaobo Zhu, Haruko Ogawa, Tetsuya Ogawa, Matthew Ubell, Ajita Narayan, Garfield Johnson, Gregory T. Wolf, Susan G. Fisher, and Thomas E. Carey. P53 Mutation Correlates With Cisplatin Sensitivity in Head and Neck Squamous Cell Carcinoma Lines. *Head and Neck*, 25(8):654–661, 2003.

[22] Matthew Brett, Michael Hanke, Ben Cipollini, Marc-Alexandre Côté, Chris Markiewicz, Stephan Gerhard, Eric Larson, Gregory R Lee, Yaroslav Halchenko, Erik Kastman, et al. nibabel: 2.1. 0. *Zenodo*, 2016.

[23] Gavin Brown, Adam Pocock, Ming-Jie Zhao, and Mikel Luján. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *Journal of machine learning research*, 13(Jan):27–66, 2012.

[24] Jerrold T Bushberg and John M Boone. *The essential physics of medical imaging.* Lippincott Williams & Wilkins, 2011.

[25] By Arnauld Cabelguenne, Anne-marie Houllier, Thierry Soussi, Daniel Brasnu, Phillipe Beaune, Ollivier Laccourreye, Pierre Laurent-puig, and Purpose The. p53 alterations predict tumor response to neoadjuvant chemotherapy in head and neck squamous cell carcinoma : A prospective series. 18(7):1465–1473, 2000.

[26] P Capaccio, G Pruneri, N Carboni, a V Pagliari, M Quatela, B M Cesana, and L Pignataro. Cyclin D1 expression is predictive of occult metastases in head and neck cancer patients with clinically negative cervical lymph nodes. *Head & neck*, 22(3):234–40, 2000.

[27] Leo Chin and Tanmay Dutta. *NumPy essentials: boost your scientific and analytic capabilities in no time at all by discovering how to build real-world applications with NumPy.* Packt Publ., 2016.

[28] Man Ki Chung, Han-Sin Jeong, Sang Gyu Park, Jeon Yeob Jang, Young-Ik Son, Joon Young Choi, Seung Hyup Hyun, Keunchil Park, Myung-Ju Ahn, Yong Chan Ahn, et al. Metabolic tumor volume of [18f]-fluorodeoxyglucose positron emission tomography/computed tomography predicts short-term outcome to radiotherapy with or without chemotherapy in pharyngeal cancer. *Clinical cancer research*, pages 1078–0432, 2009.

[29] David A Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing*, 28(1):45–62, 2002.

[30] C Coco, V Valentini, A Manno, G Rizzo, MA Gambacorta, C Mattana, A Verbo, and A Picciocchi. Functional results after ra-

diochemotherapy and total mesorectal excision for rectal cancer. *International journal of colorectal disease*, 22(8):903–910, 2007.

[31] Gary S Collins, Johannes B Reitsma, Douglas G Altman, and Karel GM Moons. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (tripod): the tripod statement. *BMC medicine*, 13(1):1, 2015.

[32] Christian Couture, Hélène Raybaud-Diogène, Bernard Têtu, Isabelle Bairati, Danielle Murry, Josée Allard, and André Fortin. p53 and Ki-67 as markers of radioresistance in head and neck carcinoma. *Cancer*, 94(3):713–722, 2002.

[33] Mireia Crispin-Ortuzar, Aditya Apte, Milan Grkovski, Jung Hun Oh, Nancy Y Lee, Heiko Schöder, John L Humm, and Joseph O Deasy. Predicting hypoxia status using a combination of contrast-enhanced computed tomography and [18f]-fluorodeoxyglucose positron emission tomography radiomics features. *Radiotherapy and Oncology*, 127(1):36–42, 2018.

[34] Belur V Dasarathy and Edwin B Holder. Image characterizations based on joint gray level—run length distributions. *Pattern Recognition Letters*, 12(8):497–502, 1991.

[35] Carlo N De Cecco, Maria Ciolina, Damiano Caruso, Marco Rengo, Balaji Ganeshan, Felix G Meinel, Daniela Musio, Francesca De Felice, Vincenzo Tombolini, and Andrea Laghi. Performance of diffusion-weighted imaging, perfusion imaging, and texture analysis in predicting tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3t mr: initial experience. *Abdominal Radiology*, 41(9):1728–1735, 2016.

[36] Geoff Delaney, Susannah Jacob, Carolyn Featherstone, and Michael Barton. The role of radiotherapy in cancer treatment: estimating optimal utilization from a review of evidence-based clinical guidelines. *Cancer: Interdisciplinary International Journal of the American Cancer Society*, 104(6):1129–1137, 2005.

[37] Atam P Dhawan. Medical imaging modalities: magnetic resonance imaging.

[38] Yadolah Dodge. *The concise encyclopedia of statistics*. Springer Science & Business Media, 2008.

[39] Stephanie B Donaldson, Guy Betts, Suzanne C Bonington, Jarrod J Homer, Nick J Slevin, Lucy E Kershaw, Helen Valentine, Catharine ML West, and David L Buckley. Perfusion estimated with rapid dynamic contrast-enhanced magnetic resonance imaging correlates inversely with vascular endothelial growth factor expression and pimonidazole staining in head-and-neck cancer: a pilot study. *International Journal of Radiation Oncology* Biology* Physics*, 81(4):1176–1183, 2011.

[40] Juliette P Driessen, Paulina MW van Kempen, Geert J van der Heijden, Marielle EP Philippens, Frank A Pameijer, Inge Stegeman, Chris HJ Terhaard, Luuk M Janssen, and Wilko Grolman. Diffusion-weighted imaging in head and neck squamous cell carcinomas: A systematic review. *Head & neck*, 37(3):440–448, 2015.

[41] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[42] Yu-Hua Dean Fang, Chien-Yu Lin, Meng-Jung Shih, Hung-Ming Wang, Tsung-Ying Ho, Chun-Ta Liao, and Tzu-Chen Yen. Development and evaluation of an open-source software package "cgita" for quantifying tumor heterogeneity with molecular images. *BioMed research international*, 2014, 2014.

[43] Xenia Fave, Lifei Zhang, Jinzhong Yang, Dennis Mackin, Peter Balter, Daniel Gomez, David Followill, Aaron Kyle Jones, Francesco Stingo, Zhongxing Liao, et al. Delta-radiomics features for the prediction of patient outcomes in non–small cell lung cancer. *Scientific reports*, 7(1):588, 2017.

[44] Antoine Flahault, Sophie Pe, Guy Monceaux, Florence Coulet, and Patrice Callard. p53 Gene Status as a Predictor of Tumor Re-

sponse to Induction Chemotherapy of Patients With Locoregionally Advanced Squamous Cell Carcinomas of the Head and Neck. 18(2):385–394, 2000.

[45] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, NY, USA:, 2001.

[46] Jerome H Friedman, Jon Louis Bentley, and Raphael Ari Finkel. An algorithm for finding best matches in logarithmic time. *ACM Trans. Math. Software*, 3(SLAC-PUB-1549-REV. 2):209–226, 1976.

[47] Lidija Fumic-Dunkic, Vladimir Katic, Sasa Janjanin, Ivica Klapan, Anamarija Simuncic, and Aleksandar Vcev. Retrospective analysis of Ki-67 antigen expression in paraffin tissue blocks of laryngeal squamous cell carcinoma. *American Journal of Otolaryngology - Head and Neck Medicine and Surgery*, 24(2):106–110, 2003.

[48] A Fyles, M Milosevic, D Hedley, M Pintilie, W Levin, L Manchul, and RP Hill. Tumor hypoxia has independent predictor impact only in patients with node-negative cervix cancer. *Journal of Clinical Oncology*, 20(3):680–687, 2002.

[49] Susan M Galbraith, Martin A Lodge, N Jane Taylor, Gordon JS Rustin, Søren Bentzen, J James Stirling, and Anwar R Padhani. Reproducibility of dynamic contrast-enhanced mri in human muscle and tumours: comparison of quantitative and semi-quantitative analysis. *NMR in Biomedicine*, 15(2):132–142, 2002.

[50] MM Gallowy. Texture analysis using gray level run length. *Computer Graphics and Image Processing*, 4:172–179, 1975.

[51] C Gani, P Bonomo, K Zwirner, C Schroeder, A Menegakis, C Rödel, and D Zips. Organ preservation in rectal cancer–challenges and future strategies. *Clinical and translational radiation oncology*, 3:9–15, 2017.

[52] Cihan Gani, Christopher Schroeder, Vanessa Heinrich, Patrick Spillner, Ulf Lamprecht, Bernhard Berger, and Daniel Zips. Long-term local control and survival after preoperative radiochemotherapy in combination with deep regional hyperthermia in locally advanced rectal cancer. *International Journal of Hyperthermia*, 32(2):187–192, 2016.

[53] Fei Gao and Sven Zuehlsdorff. An ai system using hybrid cnn and radiomics model to improve pet image quality. *Journal of Nuclear Medicine*, 60(supplement 1):1216–1216, 2019.

[54] Jean-Pierre Gérard, Thierry Conroy, Franck Bonnetain, Olivier Bouché, Olivier Chapet, Marie-Thérèse Closon-Dejardin, Michel Untereiner, Bernard Leduc, Éric Francois, Jean Maurel, et al. Pre-operative radiotherapy with or without concurrent fluorouracil and leucovorin in t3-4 rectal cancers: results of ffcd 9203. *Journal of clinical oncology*, 24(28):4620–4625, 2006.

[55] Olivier Gevaert, Sebastian Echegaray, Amanda Khuong, Chuong D Hoang, Joseph B Shrager, Kirstin C Jensen, Gerald J Berry, H Henry Guo, Charles Lau, Sylvia K Plevritis, et al. Predictive radiogenomics modeling of egfr mutation status in lung cancer. *Scientific reports*, 7:41674, 2017.

[56] Olivier Gevaert, Lex A Mitchell, Achal S Achrol, Jiajing Xu, Sebastian Echegaray, Gary K Steinberg, Samuel H Cheshier, Sandy Napel, Greg Zaharchuk, and Sylvia K Plevritis. Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features. *Radiology*, 273(1):168–174, 2014.

[57] Robert J Gillies, Paul E Kinahan, and Hedvig Hricak. Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2015.

[58] Geoffrey S Ginsburg and Huntington F Willard. *Essentials of genomic and personalized medicine*. Academic Press, 2009.

[59] Rafael C.. Gonzalez, Richard E.. Woods, and Steven L.. Eddins. *Digital Image Processing Using MATLAB®*. Gatesmark Publishing, 2009.

[60] RC Gonzalez, RE Woods, and SL Eddins. Digital image processing using matlab" gatesmark publishing, 2009.

[61] Jennifer Rubin Grandis, Mona F Melhem, William E Gooding, Roger Day, Valerie A Holst, Marilyn M Wagener, Stephanie D Drenning, and David J Tweardy. Levels of TGF-α and EGFR Protein in Head and Neck Squamous Cell Carcinoma and Neck Squamous Cell Carcinoma and Patient Survival. 90(11), 1998.

[62] Amara Graps. An introduction to wavelets. *IEEE computational science and engineering*, 2(2):50–61, 1995.

[63] Louis Harold Gray, A_D Conger, Michael Ebert, S Hornsey, and OCA Scott. The concentration of oxygen dissolved in tissues at the time of irradiation as a factor in radiotherapy. *The British journal of radiology*, 26(312):638–648, 1953.

[64] Isabelle Guyon, Jason Weston, Stephen Barnhill, and Vladimir Vapnik. Gene selection for cancer classification using support vector machines. *Machine learning*, 46(1-3):389–422, 2002.

[65] Angelita Habr-Gama, Joaquim Gama-Rodrigues, Guilherme P São Julião, Igor Proscurshim, Charles Sabbagh, Patricio B Lynn, and Rodrigo O Perez. Local recurrence after complete clinical response and watch and wait in rectal cancer after neoadjuvant chemoradiation: impact of salvage therapy on local disease control. *International Journal of Radiation Oncology* Biology* Physics*, 88(4):822–828, 2014.

[66] Robert M Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804, 1979.

[67] Robert M Haralick, K Shanmugam, Its'Hak Dinstein, et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, 3(6):610–621, 1973.

[68] Mathieu Hatt, Florent Tixier, Larry Pierce, Paul E Kinahan, Catherine Cheze Le Rest, and Dimitris Visvikis. Characterization of pet/ct images using texture analysis: the past, the present... any future? *European journal of nuclear medicine and molecular imaging*, 44(1):151–165, 2017.

[69] Mathieu Hatt, Dimitris Visvikis, Olivier Pradier, and Catherine Cheze-Le Rest. Baseline 18 f-fdg pet image-derived parameters for therapy response prediction in oesophageal cancer. *European journal of nuclear medicine and molecular imaging*, 38(9):1595–1606, 2011.

[70] RJ Heald. The 'holy plane'of rectal surgery. *Journal of the Royal Society of Medicine*, 81(9):503–508, 1988.

[71] Michael Höckel, Claudia Knoop, Karlheinz Schlenger, Birgit Vorndran, Evmarie Bauβnann, Margarete Mitze, Paul Georg Knapstein, and Peter Vaupel. Intratumoral po2 predicts survival in advanced cancer of the uterine cervix. *Radiotherapy and Oncology*, 26(1):45–50, 1993.

[72] Michael Höckel, Karlheinz Schlenger, Billur Aral, Margarete Mitze, Uwe Schäffer, and Peter Vaupel. Association between tumor hypoxia and malignant progression in advanced cancer of the uterine cervix. *Cancer research*, 56(19):4509–4515, 1996.

[73] Michael R Horsman, Lise Saksø Mortensen, Jørgen B Petersen, Morten Busk, and Jens Overgaard. Imaging hypoxia to improve radiotherapy outcome. *Nature reviews Clinical oncology*, 9(12):674, 2012.

[74] PJ Hoskin, MI Saunders, K Goodchild, ME Powell, NJ Taylor, and H Baddeley. Dynamic contrast enhanced magnetic resonance scanning as a predictor of response to accelerated radiotherapy for advanced head and neck cancer. *The British journal of radiology*, 72(863):1093–1098, 1999.

[75] Jin Huang and Charles X Ling. Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering*, 17(3):299–310, 2005.

[76] Carl J Huberty and John D Morris. Multivariate analysis versus multiple univariate analyses. *Psychological bulletin*, 105(2):302, 1989.

[77] RD Hunter. World health organization (who) handbook for reporting results of cancer treatment. *WHO: Geneva, Switzerland*, 1979.

[78] Tatianie Jackson, Margaret K Chung, Gustavo Mercier, Al Ozonoff, and Rathan M Subramaniam. Fdg pet/ct interobserver agreement in head and neck cancer: Fdg and ct measurements of the primary tumor site. *Nuclear medicine communications*, 33(3):305–312, 2012.

[79] Peyman Jafari and Francisco Azuaje. An assessment of recently published gene expression data analyses: reporting experimental design and statistical factors. *BMC Medical Informatics and Decision Making*, 6(1):27, 2006.

[80] Jacobus FA Jansen, Diane L Carlson, Yonggang Lu, Hilda E Stambuk, Andre L Moreira, Bhuvanesh Singh, Snehal G Patel, Dennis H Kraus, Richard J Wong, Ashok R Shaha, et al. Correlation of a priori dce-mri and 1h-mrs data with molecular markers in neck nodal metastases: Initial analysis. *Oral oncology*, 48(8):717–722, 2012.

[81] Nandakishore Kambhatla and Todd K Leen. Dimension reduction by local principal component analysis. *Neural computation*, 9(7):1493–1516, 1997.

[82] Christoph A Karlo, Pier Luigi Di Paolo, Joshua Chaim, A Ari Hakimi, Irina Ostrovnaya, Paul Russo, Hedvig Hricak, Robert Motzer, James J Hsieh, and Oguz Akin. Radiogenomics of clear cell renal cell carcinoma: associations between ct imaging features and mutations. *Radiology*, 270(2):464–471, 2014.

[83] Theodore G Karrison, Michael L Maitland, Walter M Stadler, and Mark J Ratain. Design of phase ii cancer trials using a continuous

endpoint of change in tumor size: application to a study of sorafenib and erlotinib in non–small-cell lung cancer. *Journal of the National Cancer Institute*, 99(19):1455–1461, 2007.

[84] Gary J Kelloff, Kenneth A Krohn, Steven M Larson, Ralph Weissleder, David A Mankoff, John M Hoffman, Jeanne M Link, Kathryn Z Guyton, William C Eckelman, Howard I Scher, et al. The progress and promise of molecular imaging probes in oncologic drug development. *Clinical Cancer Research*, 11(22):7967–7985, 2005.

[85] Faiz M Khan and John P Gibbons. *Khan's the physics of radiation therapy*. Lippincott Williams & Wilkins, 2014.

[86] Philipp Kickingereder, Michael Götz, John Muschelli, Antje Wick, Ulf Neuberger, Russell T Shinohara, Martin Sill, Martha Nowosielski, Heinz-Peter Schlemmer, Alexander Radbruch, et al. Large-scale radiomic profiling of recurrent glioblastoma identifies an imaging predictor for stratifying anti-angiogenic treatment response. *Clinical Cancer Research*, 2016.

[87] Ki Tae Kim, Bo-Sung Kim, and Ju Han Kim. Association between fat1 mutation and overall survival in patients with human papillomavirus–negative head and neck squamous cell carcinoma. *Head & neck*, 38(S1):E2021–E2029, 2016.

[88] Sungheon Kim, LA Loevner, H Quon, A Kilger, E Sherman, G Weinstein, A Chalian, and H Poptani. Prediction of response to chemoradiation therapy in squamous cell carcinomas of the head and neck using dynamic contrast-enhanced mr imaging. *American journal of neuroradiology*, 31(2):262–268, 2010.

[89] Sungheon Kim, Laurie Loevner, Harry Quon, Eric Sherman, Gregory Weinstein, Alex Kilger, and Harish Poptani. Diffusion-weighted magnetic resonance imaging for predicting and detecting early response to chemoradiation therapy of squamous cell carcinomas of the head and neck. *Clinical Cancer Research*, 15(3):986–994, 2009.

[90] Joost L Knegjens, Michael Hauptmann, Frank A Pameijer, Alfons J Balm, Frank J Hoebers, Josien A de Bois, Johannes H Kaanders, Carla M van Herpen, Cornelia G Verhoef, Oda B Wijers, et al. Tumor volume as prognostic factor in chemoradiation for advanced head and neck cancer. *Head & neck*, 33(3):375–382, 2011.

[91] Wui-Jin Koh, Janet S Rasey, Margaret L Evans, John R Grierson, Thomas K Lewellen, Michael M Graham, Kenneth A Krohn, and Thomas W Griffin. Imaging of hypoxia in human tumors with [f-18] fluoromisonidazole. *International Journal of Radiation Oncology* Biology* Physics*, 22(1):199–212, 1992.

[92] Ron Kohavi et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Ijcai*, volume 14, pages 1137–1145. Montreal, Canada, 1995.

[93] Daphne Koller and Mehran Sahami. Toward optimal feature selection. Technical report, Stanford InfoLab, 1996.

[94] Michael I. Koukourakis, Alexandra Giatromanolaki, Efthimios Sivridis, Constantinos Simopoulos, Helen Turley, Kate Talks, Kevin C. Gatter, and Adrian L. Harris. Hypoxia-inducible factor (HIF1A and HIF2A), angiogenesis, and chemoradiotherapy outcome of squamous cell head-and-neck cancer. *International Journal of Radiation Oncology Biology Physics*, 53(5):1192–1202, 2002.

[95] Michael I Koukourakis, Ioannis Mavanis, George Kouklakis, Michael Pitiakoudis, George Minopoulos, Costantinos Manolas, and Costantinos Simopoulos. Early antivascular effects of bevacizumab anti-vegf monoclonal antibody on colorectal carcinomas assessed with functional ct imaging. *American journal of clinical oncology*, 30(3):315–318, 2007.

[96] Virendra Kumar, Yuhua Gu, Satrajit Basu, Anders Berglund, Steven A Eschrich, Matthew B Schabath, Kenneth Forster, Hugo JWL Aerts, Andre Dekker, David Fenstermacher, et al. Radiomics: the process and the challenges. *Magnetic resonance imaging*, 30(9):1234–1248, 2012.

[97] Philippe Lambin, Ralph TH Leijenaar, Timo M Deist, Jurgen Peerlings, Evelyn EC de Jong, Janita van Timmeren, Sebastian Sanduleanu, Ruben THM Larue, Aniek JG Even, Arthur Jochems, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nature Reviews Clinical Oncology*, 14(12):749, 2017.

[98] Philippe Lambin, Emmanuel Rios-Velazquez, Ralph Leijenaar, Sara Carvalho, Ruud GPM van Stiphout, Patrick Granton, Catharina ML Zegers, Robert Gillies, Ronald Boellard, André Dekker, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *European journal of cancer*, 48(4):441–446, 2012.

[99] Jiangwei Lao, Yinsheng Chen, Zhi-Cheng Li, Qihua Li, Ji Zhang, Jing Liu, and Guangtao Zhai. A deep learning-based radiomics model for prediction of survival in glioblastoma multiforme. *Scientific reports*, 7(1):10353, 2017.

[100] C René Leemans, Peter JF Snijders, and Ruud H Brakenhoff. The molecular landscape of head and neck cancer. *Nature Reviews Cancer*, 2018.

[101] Stefan Leger, Alex Zwanenburg, Karoline Pilz, Fabian Lohaus, Annett Linge, Klaus Zöphel, Jörg Kotzerke, Andreas Schreiber, Inge Tinhofer, Volker Budach, et al. A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. *Scientific reports*, 7(1):13206, 2017.

[102] Stefan Leger, Alex Zwanenburg, Karoline Pilz, Sebastian Zschaeck, Klaus Zöphel, Jörg Kotzerke, Andreas Schreiber, Daniel Zips, Mechthild Krause, Michael Baumann, et al. Ct imaging during treatment improves radiomic models for patients with locally advanced head and neck cancer. *Radiotherapy and Oncology*, 130:10–17, 2019.

[103] Ralph TH Leijenaar, Sara Carvalho, Frank JP Hoebers, Hugo JWL Aerts, Wouter JC Van Elmpt, Shao Hui Huang, Biu Chan, John N

Waldron, Brian O'sullivan, and Philippe Lambin. External valida-tion of a prognostic ct-based radiomic signature in oropharyngeal squamous cell carcinoma. *Acta oncologica*, 54(9):1423–1429, 2015.

[104] Hui Li, Yitan Zhu, Elizabeth S Burnside, Karen Drukker, Kather-ine A Hoadley, Cheng Fan, Suzanne D Conzen, Gary J Whitman, Elizabeth J Sutton, Jose M Net, et al. Mr imaging radiomics sig-natures for predicting the risk of breast cancer recurrence as given by research versions of mammaprint, oncotype dx, and pam50 gene assays. *Radiology*, 281(2):382–391, 2016.

[105] Zeju Li, Yuanyuan Wang, Jinhua Yu, Yi Guo, and Wei Cao. Deep learning based radiomics (dlr) and its usage in noninvasive idh1 prediction for low grade glioma. *Scientific reports*, 7(1):5467, 2017.

[106] Chun-Ta Liao, Joseph Tung-Chieh Chang, Hung-Ming Wang, Shu-Hang Ng, Chuen Hsueh, Li-Yu Lee, Chih-Hung Lin, I-How Chen, Shiang-Fu Huang, Ann-Joy Cheng, et al. Pretreatment primary tumor suvmax measured by fdg-pet and pathologic tumor depth predict for poor outcomes in patients with oral cavity squamous cell carcinoma and pathologically positive lymph nodes. *International Journal of Radiation Oncology* Biology* Physics*, 73(3):764–771, 2009.

[107] Dahua Lin and Xiaoou Tang. Conditional infomax learning: an integrated framework for feature extraction and fusion. In *European Conference on Computer Vision*, pages 68–82. Springer, 2006.

[108] Hongzhi Liu, Zhonghai Wu, and Xing Zhang. Feature selection based on data clustering. In *International Conference on Intelligent Computing*, pages 227–236. Springer, 2015.

[109] Steffen Löck, Annett Linge, Annekatrin Seidlitz, Anna Bandurska-Luque, Alexander Nowak, Volker Gudziol, Frank Buchholz, Daniela E Aust, Gustavo B Baretton, Klaus Zöphel, et al. Re-peat fmiso-pet imaging weakly correlates with hypoxia-associated gene expressions for locally advanced hnscc treated by primary ra-diochemotherapy. *Radiotherapy and Oncology*, 135:43–50, 2019.

[110] Steffen Löck, Rosalind Perrin, Annekatrin Seidlitz, Anna Bandurska-Luque, Sebastian Zschaeck, Klaus Zöphel, Mechthild Krause, Jörg Steinbach, Jörg Kotzerke, Daniel Zips, et al. Residual tumour hypoxia in head-and-neck cancer patients undergoing primary radiochemotherapy, final results of a prospective trial on repeat fmiso-pet imaging. *Radiotherapy and Oncology*, 124(3):533–540, 2017.

[111] Egesta Lopci, Ilaria Grassi, Arturo Chiti, Cristina Nanni, Gianfranco Cicoria, Luca Toschi, Cristina Fonti, Filippo Lodi, Sandro Mattioli, and Stefano Fanti. Pet radiopharmaceuticals for imaging of tumor hypoxia: a review of the evidence. *American journal of nuclear medicine and molecular imaging*, 4(4):365, 2014.

[112] Andreas Maier, Stefan Steidl, Vincent Christlein, and Joachim Hornegger. *Medical Imaging Systems: An Introductory Guide*, volume 11111. Springer, 2018.

[113] Nina A Mayr, William TC Yuh, Vincent A Magnotta, James C Ehrhardt, James A Wheeler, Joel I Sorosky, Charles S Davis, B-Chen Wen, Douglas D Martin, Retta E Pelsang, et al. Tumor perfusion studies using fast magnetic resonance imaging technique in advanced cervical cancer: a new noninvasive predictive assay. *International Journal of Radiation Oncology* Biology* Physics*, 36(3):623–633, 1996.

[114] Wes McKinney. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython.* " O'Reilly Media, Inc.", 2012.

[115] H. Mineta, K. Miura, T. Ogino, S. Takebayashi, K. Misawa, Y. Ueda, I. Suzuki, M. Dictor, Å Borg, and J. Wennerberg. Prognostic value of vascular endothelial growth factor (VEGF) in head and neck squamous cell carcinomas. *British Journal of Cancer*, 83(6):775–781, 2000.

[116] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of machine learning.* MIT press, 2012.

[117] David Mönnich, Stefan Welz, Daniela Thorwarth, Christina Pfannenberg, Gerald Reischl, Paul-Stefan Mauz, Konstantin Nikolaou, Christian la Fougère, and Daniel Zips. Robustness of quantitative hypoxia pet image analysis for predicting local tumor control. *Acta Oncologica*, 54(9):1364–1369, 2015.

[118] Bruno Morgan, Anne L Thomas, Joachim Drevs, Juergen Hennig, Martin Buchert, Asvina Jivan, Mark A Horsfield, Klauss Mross, Howard A Ball, Lucy Lee, et al. Dynamic contrast-enhanced magnetic resonance imaging as a biomarker for the pharmacological response of ptk787/zk 222584, an inhibitor of the vascular endothelial growth factor receptor tyrosine kinases, in patients with advanced colorectal cancer and liver metastases: results from two phase i studies. *Journal of Clinical Oncology*, 21(21):3955–3964, 2003.

[119] Edmund A Mroz, Aaron D Tward, Curtis R Pickering, Jeffrey N Myers, Robert L Ferris, and James W Rocco. High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma. *Cancer*, 119(16):3034–3042, 2013.

[120] Cherie-Ann O. Nathan, Nazanin Amirghahri, Cliff Rice, Fleurette W. Abreo, Runhua Shi, and Fred J. Stucker. Molecular Analysis of Surgical Margins in Head and Neck Squamous Cell Carcinoma Patients. *The Laryngoscope*, 112(12):2129–2140, 2002.

[121] Cancer Genome Atlas Network et al. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature*, 517(7536):576, 2015.

[122] Kate Newbold, Isabel Castellano, Elizabeth Charles-Edwards, Dorothy Mears, Aslam Sohaib, Martin Leach, Peter Rhys-Evans, Peter Clarke, Cyril Fisher, Kevin Harrington, et al. An exploratory study into the role of dynamic contrast-enhanced magnetic resonance imaging or perfusion computed tomography for detection of intratumoral hypoxia in head-and-neck cancer. *International Journal of Radiation Oncology* Biology* Physics*, 74(1):29–37, 2009.

[123] Ke Nie, Liming Shi, Qin Chen, Xi Hu, Salma Jabbour, Ning Yue, Tianye Niu, and Xiaonan Sun. Rectal cancer: assessment of neoadjuvant chemo-radiation outcome based on radiomics of multi-parametric mri. *Clinical cancer research*, pages clincanres–2997, 2016.

[124] Yukiko Nishikawa, Tatsuhiko Miyazaki, Koh-Ichi Nakashiro, Hidehisa Yamagata, Mayumi Isokane, Hiroyuki Goda, Hiroshi Tanaka, Ryota Oka, and Hiroyuki Hamakawa. Human fat1 cadherin controls cell migration and invasion of oral squamous cell carcinoma through the localization of β-catenin. *Oncology reports*, 26(3):587–592, 2011.

[125] Marianne Nordsmark, Søren M Bentzen, Volker Rudat, David Brizel, Eric Lartigau, Peter Stadler, Axel Becker, Markus Adam, Michael Molls, Juergen Dunst, et al. Prognostic value of tumor oxygenation in 397 head and neck tumors after primary radiation therapy. an international multi-center study. *Radiotherapy and Oncology*, 77(1):18–24, 2005.

[126] Luke Oakden-Rayner, Gustavo Carneiro, Taryn Bessen, Jacinto C Nascimento, Andrew P Bradley, and Lyle J Palmer. Precision radiology: predicting longevity using feature engineering and deep learning methods in a radiomics framework. *Scientific Reports*, 7(1):1648, 2017.

[127] James PB O'Connor, Alan Jackson, Marie-Claude Asselin, David L Buckley, Geoff JM Parker, and Gordon C Jayson. Quantitative imaging biomarkers in the clinical development of targeted therapeutics: current and future perspectives. *The lancet oncology*, 9(8):766–776, 2008.

[128] Vasavi Paidpally, Alin Chirindel, Stella Lam, Nishant Agrawal, Harry Quon, and Rathan M Subramaniam. Fdg-pet/ct imaging biomarkers in head and neck squamous cell carcinoma. *Imaging in medicine*, 4(6):633, 2012.

[129] G Palmer, C Anderin, A Martling, and T Holm. Local control and survival after extralevator abdominoperineal excision for locally advanced or low rectal cancer. *Colorectal Disease*, 16(7):527–532, 2014.

[130] In Ja Park, Y Nancy You, Atin Agarwal, John M Skibber, Miguel A Rodriguez-Bigas, Cathy Eng, Barry W Feig, Prajnan Das, Sunil Krishnan, Christopher H Crane, et al. Neoadjuvant treatment response as an early response indicator for patients with rectal cancer. *Journal of clinical oncology*, 30(15):1770, 2012.

[131] Geoffrey JM Parker and David L Buckley. Tracer kinetic modelling for t 1-weighted dce-mri. In *Dynamic contrast-enhanced magnetic resonance imaging in oncology*, pages 81–92. Springer, 2005.

[132] Chintan Parmar, Patrick Grossmann, Johan Bussink, Philippe Lambin, and Hugo JWL Aerts. Machine learning methods for quantitative radiomic biomarkers. *Scientific reports*, 5:13087, 2015.

[133] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12(Oct):2825–2830, 2011.

[134] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (8):1226–1238, 2005.

[135] M. E. Prince, R. Sivanandan, A. Kaczorowski, G. T. Wolf, M. J. Kaplan, P. Dalerba, I. L. Weissman, M. F. Clarke, and L. E. Ailles. Identification of a subpopulation of cells with cancer stem cell properties in head and neck squamous cell carcinoma. *Proceedings of the National Academy of Sciences*, 104(3):973–978, 2007.

[136] Pavel Pudil, Jana Novovičová, and Josef Kittler. Floating

search methods in feature selection. *Pattern recognition letters*, 15(11):1119–1125, 1994.

[137] Joseph G Rajendran, David L Schwartz, Janet O'Sullivan, Lanell M Peterson, Patrick Ng, Jeffrey Scharnhorst, John R Grierson, and Kenneth A Krohn. Tumor hypoxia imaging with [f-18] fluoromisonidazole positron emission tomography in head and neck cancer. *Clinical Cancer Research*, 12(18):5435–5441, 2006.

[138] Sebastian Raschka. *Python machine learning.* Packt Publishing Ltd, 2015.

[139] Sebastian Raschka. Mlxtend: Providing machine learning and data science utilities and extensions to python's scientific computing stack. *The Journal of Open Source Software*, 3(24), 2018.

[140] H Raybaud-Diogène, a Fortin, R Morency, J Roy, R a Monteil, and B Têtu. Markers of radioresistance in squamous cell carcinomas of the head and neck: a clinicopathologic and immunohistochemical study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 15(3):1030–8, 1997.

[141] Ryan Rifkin and Aldebaro Klautau. In defense of one-vs-all classification. *Journal of machine learning research*, 5(Jan):101–141, 2004.

[142] Claus Rödel, Ullrich Graeven, Rainer Fietkau, Werner Hohenberger, Torsten Hothorn, Dirk Arnold, Ralf-Dieter Hofheinz, Michael Ghadimi, Hendrik A Wolff, Marga Lang-Welzenbach, et al. Oxaliplatin added to fluorouracil-based preoperative chemoradiotherapy and postoperative chemotherapy of locally advanced rectal cancer (the german cao/aro/aio-04 study): final results of the multicentre, open-label, randomised, phase 3 trial. *The Lancet Oncology*, 16(8):979–989, 2015.

[143] Erik Roelofs, André Dekker, Elisa Meldolesi, Ruud GPM van Stiphout, Vincenzo Valentini, and Philippe Lambin. International

data-sharing for radiotherapy research: an open-source based infrastructure for multicentric clinical data mining. *Radiotherapy and Oncology*, 110(2):370–374, 2014.

[144] Erik Roelofs, Lucas Persoon, Sebastiaan Nijsten, Wolfgang Wiessler, André Dekker, and Philippe Lambin. Benefits of a clinical data warehouse with data mining tools to collect data for a radiotherapy trial. *Radiotherapy and Oncology*, 108(1):174–179, 2013.

[145] Aaron M Rutman and Michael D Kuo. Radiogenomics: creating a link between molecular diagnostics and diagnostic imaging. *European journal of radiology*, 70(2):232–241, 2009.

[146] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. A review of feature selection techniques in bioinformatics. *bioinformatics*, 23(19):2507–2517, 2007.

[147] James Sanders. Emission tomography. In *Medical Imaging Systems*, pages 207–236. Springer, 2018.

[148] Alfred M Sassler, Ramon M Esclamado, and Gregory T Wolf. Surgery after organ preservation therapy: analysis of wound complications. *Archives of Otolaryngology–Head & Neck Surgery*, 121(2):162–165, 1995.

[149] Rolf Sauer, Heinz Becker, Werner Hohenberger, Claus Rödel, Christian Wittekind, Rainer Fietkau, Peter Martus, Jörg Tschmelitsch, Eva Hager, Clemens F Hess, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. *New England Journal of Medicine*, 351(17):1731–1740, 2004.

[150] Heiko Schoder, Yusuf E Erdi, Kenneth Chao, Mithat Gonen, et al. Clinical implications of different image reconstruction parameters for interpretation of whole-body pet studies in cancer patients. *The Journal of Nuclear Medicine*, 45(4):559, 2004.

[151] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Kernel principal component analysis. In *International Conference on Artificial Neural Networks*, pages 583–588. Springer, 1997.

[152] Christopher Schroeder, Cihan Gani, Ulf Lamprecht, Claus Hann von Weyhern, Martin Weinmann, Michael Bamberg, and Bernhard Berger. Pathological complete response and sphincter-sparing surgery after neoadjuvant radiochemotherapy with regional hyperthermia for locally advanced rectal cancer compared with radiochemotherapy alone. *International Journal of Hyperthermia*, 28(8):707–714, 2012.

[153] Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[154] Wojciech Siedlecki and Jack Sklansky. On automatic feature selection. *International Journal of Pattern Recognition and Artificial Intelligence*, 2(02):197–220, 1988.

[155] B D Smith and B G Haffty. Molecular markers as prognostic factors for local recurrence and radioresistance in head and neck squamous cell carcinoma. *Radiation oncology investigations*, 7(3):125–44, 1999.

[156] Leen-Kiat Soh and Costas Tsatsoulis. Texture analysis of sar sea ice imagery using gray level co-occurrence matrices. *CSE Journal Articles*, page 47, 1999.

[157] Patrick Therasse, Susan G Arbuck, Elizabeth A Eisenhauer, Jantien Wanders, Richard S Kaplan, Larry Rubinstein, Jaap Verweij, Martine Van Glabbeke, Allan T van Oosterom, Michaele C Christian, et al. New guidelines to evaluate the response to treatment in solid tumors. *Journal of the National Cancer Institute*, 92(3):205–216, 2000.

[158] Guillaume Thibault, Jesus Angulo, and Fernand Meyer. Advanced statistical matrices for texture characterization: application to cell classification. *IEEE Transactions on Biomedical Engineering*, 61(3):630–637, 2014.

[159] Anne L Thomas, Bruno Morgan, Mark A Horsfield, Anthony Higginson, Andrea Kay, Lucy Lee, Eric Masson, Marie Puccio-Pick,

Dirk Laurent, and William P Steward. Phase i study of the safety, tolerability, pharmacokinetics, and pharmacodynamics of ptk787/zk 222584 administered twice daily in patients with advanced cancer. *Journal of Clinical Oncology*, 23(18):4162–4171, 2005.

[160] D Thorwarth and M Alber. Biologically adapted radiotherapy on the basis of hypoxia imaging with fmiso pet. In *RADIOTHERAPY AND ONCOLOGY*, volume 84, pages S87–S87. ELSEVIER IRELAND LTD ELSEVIER HOUSE, BROOKVALE PLAZA, EAST PARK SHANNON, CO, CLARE, 00000, IRELAND, 2007.

[161] Daniela Thorwarth and Markus Alber. Implementation of hypoxia imaging into treatment planning and delivery. *Radiotherapy and Oncology*, 97(2):172–175, 2010.

[162] Daniela Thorwarth, Susanne M Eschmann, Frank Paulsen, and Markus Alber. A kinetic model for dynamic [18f]-fmiso pet data to analyse tumour hypoxia. *Physics in Medicine & Biology*, 50(10):2209, 2005.

[163] Daniela Thorwarth, Susanne-Martina Eschmann, Frank Paulsen, and Markus Alber. Hypoxia dose painting by numbers: a planning study. *International Journal of Radiation Oncology* Biology* Physics*, 68(1):291–300, 2007.

[164] I Tinhofer, A Stenzinger, T Eder, R Konschak, F Niehr, V Endris, L Distel, MG Hautmann, R Mandic, C Stromberger, et al. Targeted next-generation sequencing identifies molecular subgroups in squamous cell carcinoma of the head and neck with distinct outcome after concurrent chemoradiation. *Annals of Oncology*, 27(12):2262–2268, 2016.

[165] Kouichi Tsuji, Jasna Injuk, and René Van Grieken. *X-ray spectrometry: recent technological advances.* John Wiley & Sons, 2005.

[166] Michael Unser. Sum and difference histograms for texture classifi-

cation. *IEEE transactions on pattern analysis and machine intelligence*, (1):118–125, 1986.

[167] Stefan Van der Walt, Johannes L Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.

[168] Joost JM van Griethuysen, Andriy Fedorov, Chintan Parmar, Ahmed Hosny, Nicole Aucoin, Vivek Narayan, Regina GH Beets-Tan, Jean-Christophe Fillion-Robin, Steve Pieper, and Hugo JWL Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer research*, 77(21):e104–e107, 2017.

[169] HWM Van Laarhoven, DWJ Klomp, M Rijpkema, YLM Kamm, DJ Th Wagener, JO Barentsz, CJA Punt, and A Heerschap. Prediction of chemotherapeutic response of colorectal liver metastases with dynamic gadolinium-dtpa-enhanced mri and localized 19f mrs pharmacokinetic studies of 5-fluorouracil. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In vivo*, 20(2):128–140, 2007.

[170] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013.

[171] A Vignati, S Mazzetti, V Giannini, F Russo, E Bollito, F Porpiglia, M Stasi, and D Regge. Texture features on t2-weighted magnetic resonance imaging: new potential biomarkers for prostate cancer aggressiveness. *Physics in Medicine & Biology*, 60(7):2685, 2015.

[172] Linda J Wack, David Mönnich, Wouter Van Elmpt, Catharina ML Zegers, Esther GC Troost, Daniel Zips, and Daniela Thorwarth. Comparison of [18f]-fmiso,[18f]-faza and [18f]-hx4 for pet imaging of hypoxia–a simulation study. *Acta Oncologica*, 54(9):1370–1377, 2015.

[173] F Wasilewski. Pywavelets: Discrete wavelet transform in python, 2010.

[174] Andrew G Webb. *Magnetic Resonance Technology: Hardware and System Component Design*. Royal Society of Chemistry, 2016.

[175] Stefan Welz, David Mönnich, Christina Pfannenberg, Konstantin Nikolaou, Mathias Reimold, Christian La Fougère, Gerald Reischl, Paul-Stefan Mauz, Frank Paulsen, Markus Alber, et al. Prognostic value of dynamic hypoxia pet in head and neck cancer: results from a planned interim analysis of a randomized phase ii hypoxia-image guided dose escalation trial. *Radiotherapy and Oncology*, 124(3):526–532, 2017.

[176] Matt A Wood. *Python and Matplotlib Essentials for Scientists and Engineers*. Morgan & Claypool Publishers, 2015.

[177] Zhaohai Yang, Laura H Tang, and David S Klimstra. Effect of tumor heterogeneity on the assessment of ki67 labeling index in well-differentiated neuroendocrine tumors metastatic to the liver: implications for prognostic stratification. *The American journal of surgical pathology*, 35(6):853–860, 2011.

[178] Stephen SF Yip and Hugo JWL Aerts. Applications and limitations of radiomics. *Physics in Medicine & Biology*, 61(13):R150, 2016.

[179] Lei Yu and Huan Liu. Efficient feature selection via analysis of relevance and redundancy. *Journal of machine learning research*, 5(Oct):1205–1224, 2004.

[180] Catharina ML Zegers, Wouter Van Elmpt, Bart Reymen, Aniek JG Even, Esther GC Troost, Michel C Öllers, Frank JP Hoebers, Ruud MA Houben, Jonas Eriksson, Albert D Windhorst, et al. In vivo quantification of hypoxic and metabolic status of nsclc tumors using [18f] hx4 and [18f] fdg-pet/ct imaging. *Clinical Cancer Research*, 2014.

[181] Lifei Zhang, David V Fried, Xenia J Fave, Luke A Hunter, Jinzhong Yang, and Laurence E Court. Ibex: an open infrastructure software platform to facilitate collaborative work in radiomics. *Medical physics*, 42(3):1341–1353, 2015.

[182] Pan Zhang, Neena Mirani, Ada Baisre, and Helen Fernandes. Molecular heterogeneity of head and neck squamous cell carcinoma defined by next-generation sequencing. *American Journal of Pathology*, 184(5):1323–1330, 2014.

[183] Daniel Zips, Klaus Zöphel, Nasreddin Abolmaali, Rosalind Perrin, Andrij Abramyuk, Robert Haase, Steffen Appold, Jörg Steinbach, Jörg Kotzerke, and Michael Baumann. Exploratory prospective trial of hypoxia-specific pet imaging during radiochemotherapy in patients with locally advanced head-and-neck cancer. *Radiotherapy and Oncology*, 105(1):21–28, 2012.

[184] Alex Zwanenburg, Stefan Leger, Martin Vallières, Steffen Löck, et al. Image biomarker standardisation initiative. *arXiv preprint arXiv:1612.07003*, 2016.

[185] Alex Zwanenburg and Steffen Löck. Why validation of prognostic models matters? *Radiotherapy and Oncology*, 127(3):370–373, 2018.

[186] Kerstin Zwirner, Franz J Hilke, German Demidov, Jairo Socarras Fernandez, Stephan Ossowski, Cihan Gani, Daniela Thorwarth, Olaf Riess, Daniel Zips, Christopher Schroeder, et al. Radiogenomics in head and neck cancer: correlation of radiomic heterogeneity and somatic mutations in tp53, fat1 and kmt2d. *Strahlentherapie und Onkologie*, pages 1–9, 2019.