



UiT The Arctic University of Norway

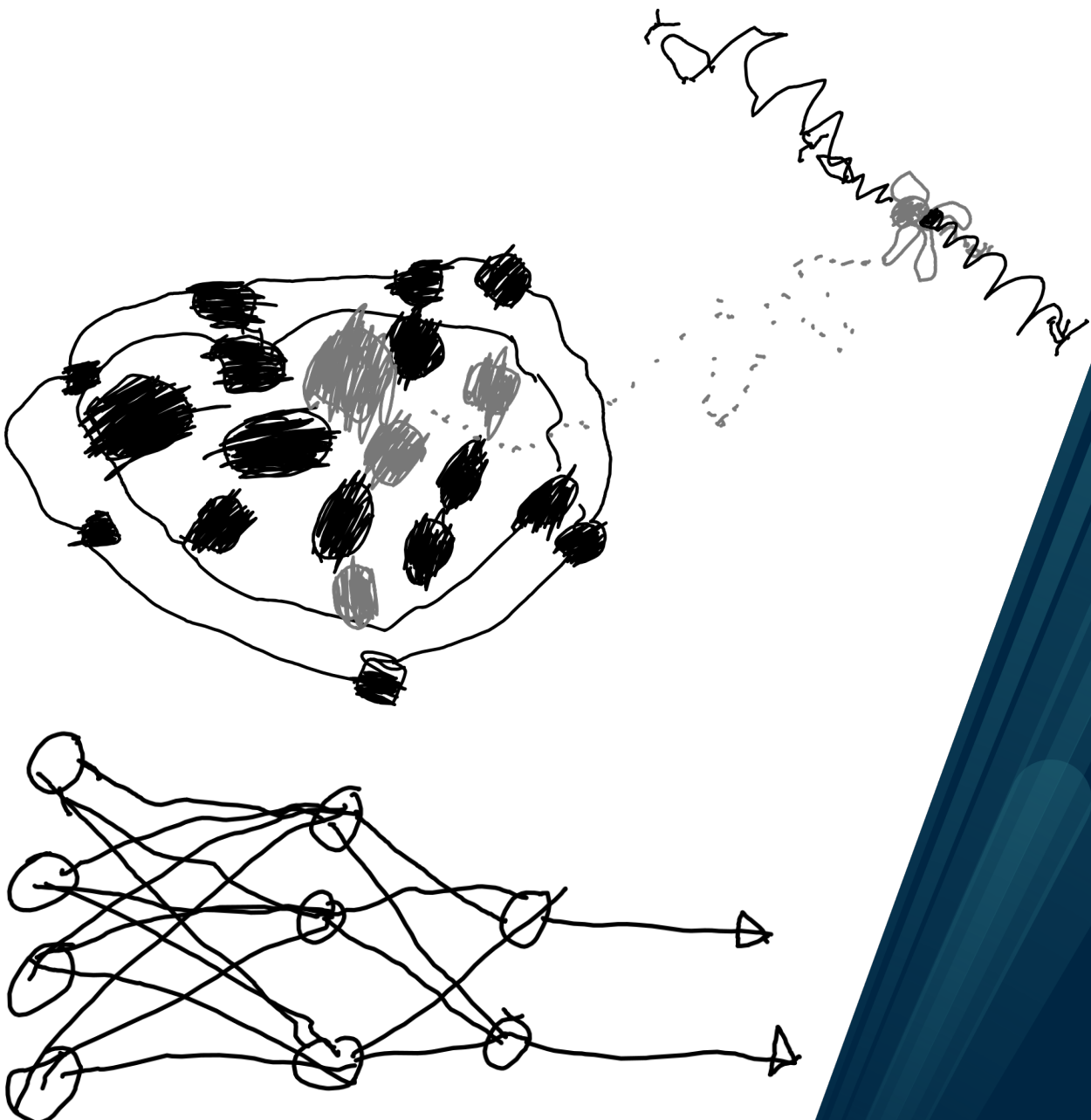
Faculty of Health Sciences

Department of Clinical Medicine

Advancing Quantitative PET Imaging with Machine Learning

Samuel Kuttner

A dissertation for the degree of Philosophiae Doctor - June 2021



Front cover:

Top: Drawing of an atom, positron emission, recombination and annihilation process.
Ola Kuttner, 6 years old.

Bottom: Drawing of a multilayer perceptron. Eskil Bruun, 8 years old.

Abstract

Medical imaging with positron emission tomography (PET) plays an important role in the detection, staging, and treatment response assessment of many diseases, including cancer, neurological and cardiovascular conditions, inflammation and infection. PET imaging is based on measuring the distribution of injected radioactive tracers, designed to follow specific biological pathways. One of the main advantages, compared to other imaging modalities, such as computed tomography or magnetic resonance imaging (MRI), is that PET allows, not only visualization of regional tracer uptake, but also quantification of the underlying biological process.

There are many challenges associated with PET imaging, which, unless accounted for, may reduce accuracy and precision in PET-based quantification. This thesis addresses the impact of imaging artifacts and subject motion on static PET-based tumor quantification and on machine-learning-based prediction models. Furthermore, the challenge of arterial blood sampling, required for quantification in dynamic PET is addressed. To this end, four papers are presented, suggesting methodology for improved PET-based quantification.

In the first two papers, the impact of imaging artifacts and respiratory motion on tumor quantification is investigated in two lung cancer PET/MRI cohorts. **Paper I** demonstrates that specific type of imaging artifacts may have clinical implications for patients undergoing serial imaging for tumor therapy response assessment. This is important both for PET-based quantification on the patient level, and as a pre-processing step for machine-learning-based prediction models.

Paper II investigates the impact of motion on PET-based tumor quantification. This main outcome of this study is a list of motion-invariant features which may be extracted from standard free-breathing PET examinations, to build predictive models for disease state or survival prediction. In this way, time-consuming motion compensation acquisitions are not needed, which simplifies the imaging workflow for patients, and allows for collection of data in retrospect, where respiratory gating or motion correction was not performed.

In the next two papers, a non-invasive machine-learning-based input function (MLIF) is proposed to replace the arterial input function (AIF), required for tracer kinetic modelling in dynamic PET applications. In **Paper III**, the MLIF method is evaluated in a small-animal dynamic ^{18}F -fluorodeoxyglucose PET cohort. By using image-derived input features, it is demonstrated that MLIF is a feasible approach for non-invasive prediction of a reference AIF in mice.

In **Paper IV**, the MLIF method is further developed and evaluated in a human clinical brain-PET cohort using the ^{15}O -water tracer. It is shown that the MLIF method is feasible in human brain PET applications by using three image-derived curves as input into the machine learning models. The proposed machine-learning-based approach to AIF estimation may considerably simplify the acquisition and analysis workflow in future pre-clinical and clinical dynamic PET studies, by avoiding the need for invasive blood sampling.

Acknowledgements

This thesis is the result of many years of work, even before it actually started. Already in 2011, me and my supervisors wrote the first grant applications that, several years later, in 2016 had been further developed and refined into a more concrete project that was finally granted funding. Throughout the journey with this thesis I have been privileged to get support and encouragement from many skilled people and I am thankful to everyone who contributed in their unique way.

First of all, I would like to thank my supervisors. Jan Axelsson, thank you for being an invaluable resource throughout all of this work. Your patience and guidance made several of the works of this thesis possible. You have an ability to always see the thin red line that is missing from the story. Rune Sundset, thank you for always supporting me, throughout all the initial grant applications and finally throughout the work with this thesis. I highly appreciate your guidance and collaboration in the clinical study that we initiated together, as part of this thesis. Robert Jenssen, thank you for supporting and encouraging me, first through several of the grant applications, and finally also all the way through the realization of this project. You have always been available for discussing small and large problems. Also, thank you also for encouraging me to join the UiT Machine Learning Group.

I would like to thank all my co-authors. Martin Lyngby Lassen, thank you for an unforgettable week in Vienna and for showing me how the analysis of PET data *really* should be done! Thomas Beyer, thank you for welcoming me to an inspiring visit your research group during this week. Live Eikenes and Silje Kjærnes Øen, thank you for sharing the lung-PET/MRI dataset with me that really ignited this project, and for your guidance and many discussions that, in the end, resulted in the first paper. I would like to thank you, Erna-Elise Paulsen, for being highly supportive in the planning and conduction of the clinical study, eventually leading to Paper II. I would also like to thank the co-workers at the PET imaging center, specifically, Anneli Johansen, Gunn Karin Skoglund, Katharina K. Lindberg, and Kent Johansen for helping me to conduct the clinical study. Kristoffer Knutsen Wickstrøm, thank you for showing me the way through the machine learning jungle. I have truly enjoyed our discussions that eventually led to the last two papers of this thesis. Ana Oteiza, Montserrat Martin-Armas, Esmaeil Dorraji and Kristin

Fenton, I would like to thank you all for the team effort and great time we had together in the lab during the intensive PET-scanning sessions for Paper III. Gustav Kalda, thanks for interesting collaboration and discussion during your summer project, and in the following year with Paper III. I would like to thank the team at Uppsala University Hospital, Mark Lubberink, Lieuwe Appel, Andreas Tolf and Joachim Burman for providing me the data for Paper IV. Special thanks to Mark for interesting physics-related discussions during the data analysis for this paper, and to Lieuwe for his time and effort in the manuscript writing process. Tom-Vegard Markussen, thank you for helping me with the delineation of the MRI and CT images. I learned a lot from our many hours at the drawing screen. Stine Hansen, thanks for your critical view and for the many discussions we had during the time we shared office. Mari Kleiveland Opsal and Stian Bakkevoll, it was highly interesting to follow your PET/MRI experiments during your Master's theses, and I appreciated all the discussions we had during your projects.

During the thesis work, I have been privileged to become integrated with the UiT Machine Learning group. I have really appreciated the kindness and hospitality of you all, and it was never a long way to get help with any kind of problem. I really enjoyed the many lunch discussions we had of various work and non-work-related topics. Special thanks to Sara Björk, Kristoffer Knutsen Wickstrøm and Luigi Luppino for proof reading the thesis.

I would like to bring special attention to Ole-Petter Rekvig. I am grateful that you took me under your wings back in 2014 and shared with me some of your valuable knowledge in grant application writing. Your guidance was the key contribution which led to the two grant applications that were funded in 2016. I would also like to thank Cecilia Marie Futsæther and Eirik Malinen for the initial inspiring meetings and discussions at the beginning of this thesis work.

I would like to thank my parents for always encouraging me to learn and try new things in life, seek problems, find solutions and to always stay hungry and curious. To my dear wife, Inga, thank you for bearing with me, listening and enduring all my work-related problems and discussions that I have brought home in the afternoons. And finally, I would like to thank my wonderful kids, Eskil, Ola and Lotta, because you are the three most important meanings of my life.



Samuel Kuttner
Tromsø, June 2021

List of publications

Included papers

This thesis is based on the following original journal papers:

- I Samuel Kuttner, Martin Lyngby Lassen, Silje Kjærnes Øen, Rune Sundset, Thomas Beyer and Live Eikenes. ”**Quantitative PET/MR imaging of lung cancer in the presence of artifacts in the MR-based attenuation correction maps**”, *Acta Radiologica*, Apr. 2020, Vol. 61, pp 11-20
- II Samuel Kuttner, Erna-Elise Paulsen, Robert Jenssen, Rune Sundset, Jan Axelsson, ”**Motion-robust radiomic features for image classification in ^{18}F -FDG PET/MRI imaging of lung cancer**”, *Manuscript*
- III Samuel Kuttner, Kristoffer Knutsen Wickstrøm, Gustav Kalda, S Esmail Dorraji, Montserrat Martin-Armas, Ana Oteiza, Robert Jenssen, Kristin Fenton, Rune Sundset and Jan Axelsson, ”**Machine learning derived input-function in a dynamic ^{18}F -FDG PET study of mice**”, *Biomedical Physics & Engineering Express*, Jan. 2020, Vol. 6, p 015020
- IV Samuel Kuttner, Kristoffer Knutsen Wickstrøm, Mark Lubberink, Andreas Tolf, Joachim Burman, Rune Sundset, Robert Jenssen, Lieuwe Appel and Jan Axelsson, ”**Cerebral blood flow measurements with ^{15}O -water PET using a non-invasive machine-learning-derived arterial input function**”, *Journal of Cerebral Blood Flow & Metabolism*, Jan. 2021, [published online ahead of print, Feb. 9, 2021]

Other papers

The following original journal paper, oral presentations and poster presentation also contribute to the thesis, but are not included:

- 5 Samuel Kuttner, Martin Lyngby Lassen, Silje Kjærnes Øen, Rune Sundset, Thomas Beyer and Live Eikenes. ”**The impact of artefacts in MR based attenuation correction maps on PET/MR images of lung cancer patients**”, oral presentation at *Tromsø PET-seminar*, Tromsø, Norway, Sep. 2017.
- 6 Samuel Kuttner, Martin Lyngby Lassen, Silje Kjærnes Øen, Rune Sundset, Thomas Beyer and Live Eikenes. ”**The impact of artefacts in MR-based attenuation correction maps on PET/MR images of lung cancer**”, oral presentation at *MedFys 2018*, Kvitfjell, Norway, Feb. 2018.
- 7 Samuel Kuttner, Kristoffer Knutsen Wickstrøm, Gustav Kalda, S Esmail Dorraji, Montserrat Martin-Armas, Ana Oteiza, Robert Jenssen, Kristin Fenton, Rune Sundset and Jan Axelsson, ”**Machine learning derived input function in dynamic ^{18}F -FDG PET**”, poster presentation at *Northern Lights Deep Learning Workshop (NLDL) 2019*, Tromsø, Norway, Jan. 2019.
- 8 Stine Hansen, Samuel Kuttner, Michael Kampffmeyer, Tom-Vegard Markussen, Rune Sundset, Silje Kjærnes Øen, Live Eikenes, Robert Jenssen, ”**Un-supervised supervoxel-based lung tumor segmentation across patient scans in hybrid PET/MRI**”, *Expert Systems With Applications*, Nov. 2020, Vol. 167, p 114244,

List of abbreviations

AI	artificial intelligence
AIF	arterial input function
CBF	cerebral blood flow
CT	computed tomography
FDG	fluorodeoxyglucose
GPs	Gaussian processes
HU	Hounsfield units
IDIF	image-derived input function
IFT	inverse Fourier-transform
LC	lumped constant
LSTM	long-short-term memory
MLIF	machine-learning-based input function
MLP	multilayer perceptron
MRI	magnetic resonance imaging
MTV	metabolic tumor volume
PET	positron emission tomography
RD	relative difference
RF	radio-frequency
SNR	signal-to-noise ratio
SUV	standardized uptake value
TE	echo-time
TR	repetition-time
VOI	volume of interest

Contents

Abstract	i
Acknowledgements	iii
List of publications	v
List of abbreviations	vii
1 Introduction	3
1.1 Challenges	6
1.2 Aims of the thesis	8
1.3 Approaches	9
1.4 Brief summary of papers	10
1.5 Datasets and ethical approvals	11
1.6 Organization of the thesis	16
2 Medical imaging with PET, CT and MRI	17
2.1 Principles of PET	18
2.2 Principles of CT	37
2.3 Principles of MRI	37
2.4 Hybrid imaging	43
2.5 Tracer kinetic modeling	51
2.6 PET applications in clinic and research	62
3 Machine learning for medical imaging	69
3.1 Basic concepts and terminology	69
3.2 Neural networks	74
3.3 Gaussian processes	81
3.4 Radiomics	88
3.5 Error estimation and validation methods	91
4 Summary of research	93
4.1 Paper I	94
4.2 Paper II	95

Contents

4.3	Paper III	96
4.4	Paper IV	97
5	Results and discussion	99
5.1	Paper I	100
5.2	Paper II	103
5.3	Paper III	107
5.4	Paper IV	109
6	Conclusion	111
7	Limitations and future directions	113
7.1	Explainable artificial intelligence	114
8	Paper I	115
9	Paper II	127
10	Paper III	161
11	Paper IV	175
	Bibliography	195

1 Introduction

Positron emission tomography (PET) is a medical imaging technique that visualizes the distribution of an injected radioactive tracer in living subjects. The tracer consists of a molecule which has been designed to follow a specific biological pathway in the body, and where one of the atoms has been replaced by a radionuclide. By detecting photons emitted from the radioactive decay with a PET scanner, it is possible to follow the physical and chemical path of the tracer inside the body.

Medical imaging with PET plays an important role in the detection, staging, and treatment response assessment of many diseases, including cancer, neurological and cardiovascular conditions, as well as inflammation and infection. The most widespread application of clinical PET is within oncology. Many diseases, including cancer, introduce alterations in the glucose metabolism, which can be measured with the glucose analogue ^{18}F -fluorodeoxyglucose (FDG) [1–3].

PET is quantitative in the sense that it allows, not only visualization, but also quantification of regional tracer uptake non-invasively. This is accomplished by calibrating the voxel-values of the reconstructed images into units of radioactivity concentration (Bq/cm^3). Measurements using standardized uptake value (SUV) is the main quantitative approach in static PET. On the other hand, with dynamic PET imaging, it is possible to fully assess the time-dependent tracer distribution in the body. This allows *in vivo* quantification of biological processes, such as glucose metabolism or blood flow, by using tracer kinetic modelling [4].

The quantitative accuracy in PET depends on the proper application of a number of corrections during image reconstruction, the most important being attenuation correction. The clinical work-horse for many oncological applications for the last 20 years has been hybrid PET/computed tomography (CT) imaging, where the CT information is used for both anatomical localization, and for attenuation correction [5, 6]. As clinical hybrid PET/magnetic resonance imaging (MRI) systems were introduced, around 10 years ago, the implementation of attenuation correction faced many new challenges which had to be solved before quantification with PET/MRI could become comparable to that of PET/CT [7, 8]. Some of these challenges, for instance the presence of artifacts in MRI-based attenuation correction maps,

1 Introduction

remains to date, and may contribute in increasing the quantification errors in PET/MRI imaging [9–13].

Lung cancer is the second most frequent cancer type and the leading cause of cancer-related death in the world [14]. Early and accurate staging of the disease is important when deciding which treatment should be given, for instance surgery, chemotherapy or radiotherapy [15]. Hybrid imaging with ^{18}F -FDG PET/CT is the standard of care today for staging and treatment response assessment of lung cancer [16–18]. While ^{18}F -FDG PET is a commonly accepted surrogate for metabolic activity, CT merely visualizes anatomical structures, and it involves ionizing radiation, that could be harmful [19–21]. On the other hand, MRI may provide anatomical images without the use of ionizing radiation, and in addition, visualize biological aspects of the soft tissue not shown with PET or CT [22]. Thus, the introduction of integrated PET/MRI systems has opened new possibilities for tumor characterization by adding excellent soft-tissue contrast, provided by MRI, to the functional information from PET. In this way, simultaneous, multiparametric images can be acquired, that facilitate precision medicine and personalized treatment of the disease [7, 23].

Visual inspection of the vast amount of images generated from a hybrid PET/MRI scan is a tedious and time-consuming task. Rather, the analysis of large image-sets can be automated using computer algorithms, based on *machine learning*. This is a field of science where a computer-model learns patterns from data in order to perform a given task, for instance classification or regression. One common machine learning approach is to first allow the models to learn the patterns of the data using training samples. Then, the trained model is applied to new, unseen samples, to perform the given task [24–26]. For instance, image-classification in the medical domain is the process of assigning a tumor image to a specific histological subgroup, or predicting the probability of survival, using machine learning. Such models could be integrated in the clinical workflow to allow for computer-aided diagnosis, or decision support for the reviewing clinician.

Machine learning algorithms require some measured input quantities, referred to as *features*. These are real numbers derived from the images, or from other measurements, that describes the data such that, for instance, classification or regression is possible. Although simple regional SUV-based quantitative measurements could be good feature candidates, the use of more comprehensive textural features, so-called *radiomics*, has been proposed as prognostic biomarkers in PET [27–32]. Radiomics refers to the process of extracting various mathematical features from medical images. In this way, a tumor may be characterized by a large number of quantitative descriptors, in addition to the commonly-used SUV.

Patterns of features may match different disease states, or be related to patient survival. This allows a multi-dimensional radiomic feature-set to be combined into a machine-learning-based model with the purpose of predicting these end points. Radiomics may, in this way, provide a high-dimensional tumor-characterization based on static PET images. On the other hand, dynamic PET imaging with tracer kinetic modelling allows to truly quantify an underlying biological process, and thus yield a more detailed tissue quantification than possible by static PET [33]. Quantitative kinetic parameters derived from dynamic ^{18}F -FDG-PET imaging has shown clinical potential in tumor diagnosis and staging, as well as in therapy-monitoring research studies [33]. Furthermore, in neurological applications, dynamic PET imaging of ^{15}O -water is the reference-standard for perfusion (blood flow) measurements [34–36]. Also, dynamic PET is an essential tool in pharmaco-kinetic characterization of new tracers, commonly performed with small-animal pre-clinical PET imaging, prior to injection in humans [33, 37].

1.1 Challenges

There are many challenges in the imaging chain that affects PET-based quantitative measurements [38–40]. Figure 1.1 gives an overview of some of the main challenges. These can be grouped into four main categories:

1. Physics, aiming at the underlying physical events of the radioactive nucleus and the resulting annihilation photons.
2. Biology, referring to factors related to the subject (human or animal), or the object (lesion or healthy tissue) inside the subject
3. Acquisition, including factors related to the tracer, timing, scanner, reconstruction and artifacts.
4. Analysis, referring to the analysis and processing of images and data.

The challenges addressed in this thesis, indicated in yellow in Figure 1.1, are presented next.

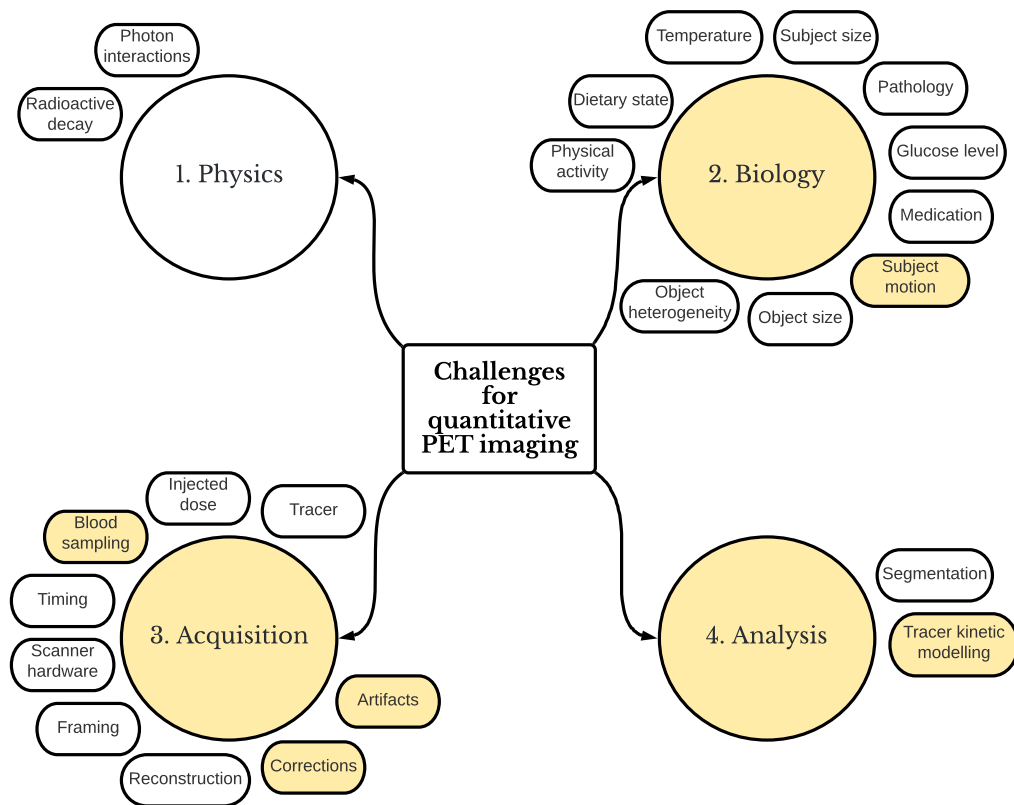


Figure 1.1: Overview of the challenges that are faced in quantitative PET imaging. Challenges that are addressed in this thesis are indicated in yellow.

Attenuation correction and artifacts. The single most important correction for obtaining quantitative PET images is attenuation correction. The annihilation photons emitted from the radioactive PET tracer inside the subject will be attenuated by the surrounding tissue. The longer the path through the tissue, the more attenuation. Unless corrected for, this will show as reduced tracer uptake in central parts of the imaged subject. In hybrid PET/CT imaging, attenuation correction is performed on the basis of an acquired low-dose CT scan. The voxel-values of a CT image are proportional to the amount of attenuation at the effective energy of the particular x-ray beam used during CT scanning. However, the CT image must be converted into attenuation at 511 keV, which is the energy of the annihilation photons, in order to derive the attenuation correction map valid for PET [6]. However, in hybrid PET/MRI imaging, the voxels of the MRI images are unrelated to photon attenuation and can thus not be directly transformed into attenuation values valid for PET. This was one of the main challenges to be solved before hybrid PET/MRI could be clinically introduced. One common solution is the segmentation approach that first separates air, water, fat tissue, and lung tissue, and then assigns each of these tissues with tabulated attenuation values. However, some tissues, for instance bone, are not visible in MRI images. In addition, MRI-based attenuation correction maps are prone to imaging artifacts, such as truncation and susceptibility. Such artifacts may affect the diagnostic quality, hinder accurate quantification of the tracer-distribution [11, 12], and lead to non-consistent SUV quantification in serial examinations [13]. Therefore, accurate MRI-based attenuation correction is critical for accurate and reproducible PET-based quantitative measurements in tumor response assessment studies [7, 9, 41–43].

Subject motion. While certain CT and MRI imaging sequences are short and allow human patients to follow breath-hold instructions, the length of a typical PET scan is in the order of 10-60 minutes. Subject motion due to breathing or muscle relaxation is inevitable during such long acquisitions. A PET image represents the integrated counts of annihilation events over a fixed time-frame, emitted from discrete voxels in the subject. Therefore, motion within a time-frame causes unwanted blurring, which degrades the quality and the quantitative accuracy of the PET image. Motion-induced image blur may have a large impact on the measured SUV values and other extracted radiomic feature values [44, 45]. Breathing-induced motion may also introduce misalignment between the attenuation correction map (commonly acquired during end-expiration breath-hold) and the PET data (acquired in free-breathing), which may lead to additional image artifacts [46].

Blood sampling for tracer kinetic modelling. Tracer kinetic modelling requires accurate determination of an arterial input function (AIF), representing the tracer time-activity curve in arterial blood. The AIF is obtained by measuring the

1 Introduction

time-dependent radioactivity concentration of the tracer in arterial blood through invasive blood sampling. In small-animal PET imaging of rodents, such a procedure is hampered by the limited blood volume that can be withdrawn without altering animal physiology, the complex surgery required for inserting an arterial catheter into the blood vessel, and the terminal end-point of such a procedure [4]. Also in clinical PET imaging, arterial cannulation is an invasive, laborious and time-consuming procedure. Due to induction of pain and risk for complications, arterial cannulation may discourage patients and volunteers from participating in research PET studies. Furthermore, in order to obtain a useful AIF curve, careful cross calibration of the blood measurement detector and the PET scanner must be performed. In addition, because the blood is most commonly sampled from the radial artery in humans, additional corrections for dispersion and delay of the tracer must be applied, to obtain a valid AIF for the site of interest, for instance in the brain [35, 47–49].

1.2 Aims of the thesis

From Figure 1.1, it is evident that there are many challenges associated with quantitative PET imaging. Nevertheless, as quantitative PET dates back at least 40 years, many of the mentioned challenges have already been addressed. However, as new scanner technology is introduced clinically, such as PET/MRI, new challenges are brought along. Also, new approaches, such as machine learning, may provide increased efficiency, accuracy and precision in the solution of existing problems.

The overall aim of this thesis is to:

Develop methodology for improved quantification with PET.

All four included papers are dealing with this aim.

The following specific aims are addressed:

1. Investigate how artifacts in the attenuation correction maps as well as respiratory motion affects quantification in PET/MRI of lung cancer.
2. Develop a non-invasive AIF prediction model for tracer kinetic modelling in dynamic PET studies.

1.3 Approaches

In **Paper I** and **Paper II**, the focus is on the first specific aim, namely to investigate how imaging artifacts and respiratory motion affects tumor-based image features in two independent lung cancer PET/MRI cohorts. The focus of these papers are primarily on the pre-processing of the PET data prior to inclusion into larger databases of a clinical study, or into training of a predictive machine learning model. More specifically, these works investigate the impact of imaging artifacts (**Paper I**) and respiratory motion (**Paper II**) on tumor quantification in two lung cancer PET/MRI cohorts. With this knowledge, it is possible to correct for the affected artifacts, and select the features that are least affected by motion, to describe the pathology in the images. This is important both for PET-based quantification on the patient level, and as a pre-processing step for machine-learning-based prediction models.

In **Paper III** and **Paper IV**, the second specific aim is addressed, namely that of non-invasive arterial input function prediction in dynamic PET. In these two studies, a machine-learning-based input function (MLIF) is proposed to replace the AIF, required for tracer kinetic modelling in dynamic PET applications. In **Paper III**, two different machine learning approaches are evaluated in a pre-clinical mouse PET cohort, to predict an image-based reference AIF. Here, the impact of different sets of input features on the generated AIFs are evaluated. In **Paper IV**, the models are further developed for a clinical brain-PET cohort and evaluated using a blood-sampled AIF. The use of a machine-learning-based input function may considerably simplify the acquisition and analysis workflow in pre-clinical and clinical dynamic PET studies, by avoiding the need for invasive blood sampling.

1.4 Brief summary of papers

Paper I investigates the frequency and the test-retest reproducibility of artifacts in MRI-based attenuation correction maps in a PET/MRI cohort of non-small-cell lung cancer patients. Further, artifacts in the attenuation correction maps are corrected and impact of the corrections on PET-based quantification are evaluated in the test-retest scans. The results showed that specific type of artifacts may have clinical implications for patients undergoing serial imaging for tumor therapy response assessment. This is important for PET-based quantification on the patient level, and as a pre-processing step for machine-learning-based prediction models.

In **Paper II** the motion variability of radiomic features in a non-small-cell lung cancer PET/MRI cohort is investigated. By comparing features extracted from free-breathing and motion corrected PET images, and by using correlation-based feature-selection, a list of motion-robust features with low correlation is presented. The results indicate that these features may be extracted from standard free-breathing PET examinations, to build predictive models for disease state or survival prediction. In this way, time-consuming motion compensation acquisitions are not needed, which simplifies the imaging workflow for patients, and allows for collection of data in retrospect, where respiratory gating or motion correction was not performed.

Paper III presents the MLIF approach for machine-learning-based AIF prediction in a small-animal dynamic ^{18}F -FDG PET cohort. By using image-derived input features, it is demonstrated that machine learning is a feasible approach for accurate and non-invasive prediction of a reference AIF in mice. Furthermore, it is shown that the input features may not necessarily need to contain the myocardium for accurate AIF predictions, which is relevant when the heart is outside the PET field-of-view during scanning. The MLIF approach to AIF estimation could significantly simplify the workflow for pre-clinical dynamic PET imaging of rodents, by evading the need for invasive arterial blood sampling.

Paper IV further develops the MLIF approach from **Paper III**, here in a human clinical brain-PET cohort using the ^{15}O -water tracer. In this work, three automatically segmented image-derived time-activity-curves are used as input into the MLIF model, to predict the AIF. The results indicate that non-invasive AIF prediction, using the MLIF approach, is also feasible in human brain PET applications. Further, it is demonstrated that tracer kinetic modelling using an MLIF could successfully predict clinically significant changes cerebral blood flow (CBF) induced from acetazolamide medication. Our proposed MLIF method shows potential to replace the AIF obtained from blood sampling for CBF measurements using ^{15}O -water PET and kinetic modelling. This could minimize the risk for complications and, at the same time, simplify the clinical dynamic PET imaging workflow.

1.5 Datasets and ethical approvals

Paper I and **Paper II** are based on human clinical PET/MRI scans of patients with non-small-cell lung cancer; the former obtained from collaborative partners at St. Olavs Hospital, Trondheim and The Norwegian University of Science and Technology; the latter was collected by the PhD candidate at the University Hospital of North Norway. **Paper III** is based on data from the pre-clinical PET core facility at UiT The Arctic University of Norway, where the PhD candidate, together with a collaborative reserach group, collected pre-clinical mouse PET scans for a parent study. The data for **Paper IV** was based on a human clinical PET/CT trial, obtained from collaborators at Uppsala University Hospital, Sweden.

1.5.1 Paper I

The data for this study were obtained in retrospect from a completed clinical study performed by a collaborative research group at St. Olavs Hospital, Trondheim and The Norwegian University of Science and Technology.

Patient population

The single-injection dual-time point imaging study included 25 patients with histologically confirmed non-small-cell lung cancer, having totally 26 lung lesions. Patients fasted $15\text{h} \pm 4\text{h}$ before the injection of $281\text{ MBq} \pm 41\text{ MBq}$ ^{18}F -FDG. PET/MRI assessment started $113\text{ min} \pm 10\text{ min}$ post ^{18}F -FDG injection. PET/CT imaging was conducted prior to the PET/MRI scan, but this data were not used in the current study.

PET/MRI imaging

The PET/MRI acquisitions were performed using a Siemens Biograph mMR (software version VB20P) (Siemens Healthineers, Erlangen, Germany). All patients had the same anatomical region scanned twice without repositioning, using a free-breathing and arms down scan protocol. First, a 10 minute, 1-bed position scan, centered over mediastinum, was performed, immediately followed by another 10 minute, 2-bed position scan of the whole thorax (Figure 1.2). A standard DIXON-based MRI attenuation correction map was acquired for each scan.

1 Introduction

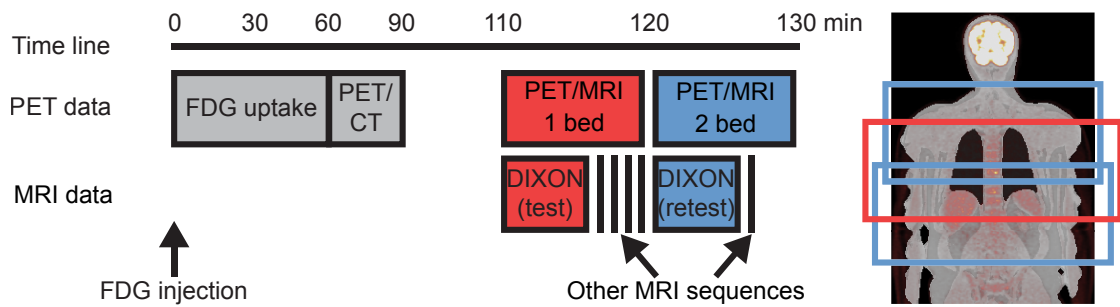


Figure 1.2: Time line of the single-injection dual-time point imaging study of **Paper I**. Patients were injected with ^{18}F -FDG, followed by 60 minutes rest, before undergoing PET/CT and PET/MRI imaging. PET data and MRI DIXON attenuation correction data that were included in this study are indicated in red and blue color, for the 1-bed position scan centered over mediastinum (test), and the 2-bed position scan of the whole thorax (retest), respectively. The corresponding anatomical regions are indicated approximately on the coronal overview scan to the right.

Ethical approval

This study was approved by the Norwegian Regional Committees for Medical and Health Research Ethics (REC reference 2017/915). All patients signed written informed consent.

1.5.2 Paper II

This prospective clinical PET/MRI study was planned and conducted by the PhD candidate at the PET Imaging Center, University Hospital of North Norway, Tromsø.

Patient population

The study included 18 patients with histologically confirmed non-small-cell lung cancer. Patients fasted $15\text{h} \pm 6\text{h}$ before the injection of $277\text{ MBq} \pm 66\text{ MBq}$ ^{18}F -FDG. PET/MRI scanning started 2 hours post injection.

PET/MRI imaging

The PET/MRI acquisitions were performed on a 3T integrated PET/MRI system (Biograph mMR; Siemens Healthineers; software version E11). Patients were positioned on the scanner bed with the arms along the torso. A 20-minute list-mode PET acquisition of the thorax was conducted in free-breathing for a single bed-position. MRI scanning was performed simultaneously with PET acquisition. Standard DIXON-based attenuation-correction maps were acquired in end-expiration breath-hold for each scan. Also MRI imaging of five respiratory phases were conducted, to allow for MRI-based motion correction (See section 2.4.6).

Ethical approval

This study was approved by the Norwegian Regional Committees for Medical and Health Research Ethics (REC reference 2017/1952). All patients signed written informed consent.

1 Introduction

1.5.3 Paper III

The pre-clinical PET/CT data for this study were collected in retrospect from a completed animal study focusing on PET imaging of tertiary lymphoid structures in two different mouse strains [50]. The PhD candidate participated actively in the planning and conduction of this parent study, which was performed by a collaborative research group at UiT The Arctic University of Norway.

Animals

Thirty-six female mice from two strains (NZBWF1, Jax stock #10008 (n=24) and BALB/ cAnNCrl (n=12)), purchased from The Jackson Laboratory and Charles River Laboratories, respectively, were included in the study. To minimize the effect of dietary state and anaesthesia on the ^{18}F -FDG uptake in the mice, a strict fasting and anaesthesia protocol was followed prior to PET imaging. In short, the mice were fasted for 3 h 50 min \pm 20 min, weighed and anesthetized for 1 h 17 min \pm 19 min prior to tracer injection. An oxygen-isoflurane mixture (4% and 2% isoflurane for induction and maintenance, respectively) was used for anesthesia.

PET/CT imaging

PET/CT imaging of totally 68 mouse scans was performed using a TriumphTM LabPET-8TM small-animal PET/CT scanner (TriFoil Imaging Inc.). Each mouse was scanned between 1-5 times at different ages (range 7-37 weeks). The anesthetized mice were centered in the field-of-view of the PET/CT scanner, while lying on a 35°C heated bed inside an animal imaging cell (Equipment Veterinaire Minerve), with sensors monitoring heart and breathing rate. Tracer administration was conducted by the injection of 10.5 ± 1.8 MBq of ^{18}F -FDG in 100 μl sterile saline through a tail-vein catheter during 30 s, with an infusion pump (56 scans), or by manual injection followed by 20 μl flush of sterile saline (12 scans). A 60 minute list-mode PET acquisition was started at injection time, followed by CT imaging for PET attenuation correction. The PET images were reconstructed into 44 time-frames (24 \times 5, 9 \times 20 and 11 \times 300 s).

Ethical approval

This animal study was approved by The Competent Authority on Animal Research, the Norwegian Food Safety Authority; FOTS id 6676/2015.

1.5.4 Paper IV

The data for this study were obtained in retrospect from a completed clinical research study performed by a collaborative research group at Uppsala University Hospital.

Subjects

The study population comprised 25 subjects, both patients with multiple sclerosis and healthy volunteers (mean age (range) in years: 40 (23–56); female-male-ratio 15:10). In this methodological study, there was no differentiation between the two groups, as the subject's health status was not considered to have an impact on the proposed methods. Therefore, all authors were blinded to the health status of each subject, and thus, no comparisons were made between multiple sclerosis patients and healthy subjects.

PET/CT and MRI imaging

All subjects underwent two 10 min dynamic brain PET scans on either an ECAT Exact HR+ stand-alone PET scanner (Siemens, Knoxville, TN; n=9) or a Discovery MI PET/CT scanner (GE Healthcare, Waukesha, MI; n=16). The scans started simultaneously with an automated bolus injection of 5 MBq/kg ^{15}O -water. Each subject underwent one scan at baseline and one scan 15-30 min after intravenous administration of acetazolamide medication. Attenuation correction was based on a 10 min transmission scan with rotating ^{68}Ge rod sources (ECAT) or an ultra-low-dose CT scan (Discovery MI). The PET images were reconstructed into 26 time-frames (1×10, 8×5, 4×10, 2×15, 3×20, 2×30 and 6×60 s).

Continuous arterial blood sampling was performed during 10 min for each scan using either an ABSS V3 (Allogg, Mariefred, Sweden; subjects scanned on ECAT) or PBS-100 (Veenstra-Comecer, Joure, The Netherlands; subjects scanned on Discovery MI).

In addition, all subjects underwent MRI on a 3T MRI scanner (Achieva, Philips Healthcare, Best, The Netherlands) with a 32-channel head coil.

1 Introduction

Ethical approval

The parent study was approved by the Swedish Ethical Review Authority (reference 2014/453). All subjects signed written informed consent. Since the present work was purely an image analysis methodology study using pseudonymized data, it was not covered by the Swedish or Norwegian regulations on medical research in humans and as such, no additional ethical approval was necessary.

1.6 Organization of the thesis

The remainder of this thesis is organized as follows: Chapter 2 and Chapter 3 presents the relevant background knowledge, including medical imaging with PET, CT and MRI, the data analysis methods, and the relevant machine learning approaches. Chapter 4 provides a summary of each of the four included research papers. Chapter 5 and Chapter 6 discuss the findings in the papers and concludes the work. Chapter 7 presents limitations of the thesis and proposes possible future directions and outlook. The four included research papers are presented in Chapter 8 through Chapter 11.

2 Medical imaging with PET, CT and MRI

Figure 2.1 shows one transversal slice of a lung cancer patient, imaged with three different medical imaging modalities, PET, CT and MRI. These imaging techniques are based on fundamentally different underlying principles.

PET is based on the injection of radioactive tracers into the blood stream, tailored to follow specific biological pathways in the body. By detecting photons emitted from the radioactive decay, it is possible to reconstruct three-dimensional images of the spatial and temporal tracer distribution in the body. In order to make PET images quantitative, several corrections must be applied during the reconstruction process, to calibrate the image voxels into units of radioactivity concentration.

PET images display physiological or functional information with little anatomical information. Therefore, it is common to couple PET scanners with CT, and recently also MRI systems. Such hybrid imaging modalities allow not only for co-registered anatomical images, but also for attenuation correction, which is one of the most important corrections required for quantitative PET imaging. There are a number of limitations with hybrid PET/MRI systems, especially concerning attenuation correction, which must be considered to preserve the quantitative accuracy in PET/MRI applications. Possible advantages of PET/MRI compared to PET/CT include superior soft-tissue contrast provided by MRI, reduced radiation dose to the patient, and MRI-based elastic motion correction of the PET data, the latter which may provide improved image quality and quantification in thoracic imaging.

Among the included works in this thesis, **Paper I** and **II** are based on clinical ^{18}F -FDG PET data from static hybrid PET/MRI scanning of lung cancer patients. **Paper III** includes dynamic ^{18}F -FDG PET images of mice from a pre-clinical PET/CT study, while in **Paper IV**, dynamic brain ^{15}O -water PET images of patients from a clinical PET/CT study are used. Therefore, in the following sections, it is relevant to describe the basic principles and underlying physics behind each of the imaging techniques of Figure 2.1. The main emphasis will be on PET imaging and how supporting modalities of CT and MRI are used in combination with PET in hybrid imaging. Quantitative analysis of dynamic PET

2 Medical imaging with PET, CT and MRI

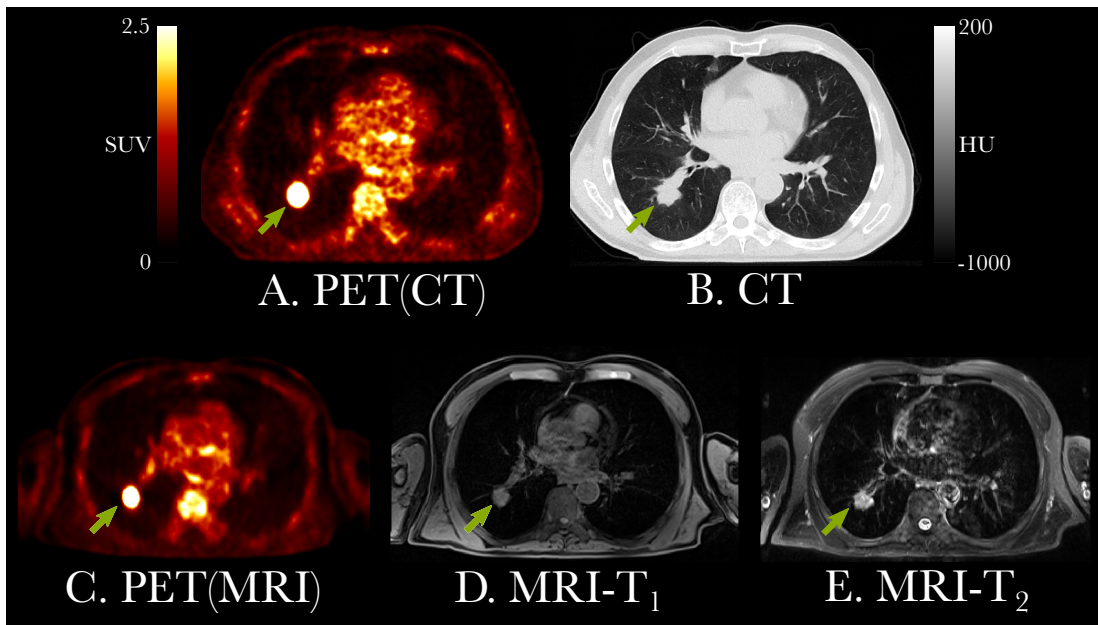


Figure 2.1: PET/CT (top row) and PET/MRI (bottom row) scans of a patient with non-small-cell lung cancer using the ^{18}F -FDG tracer. A. PET from the PET/CT scan. B. CT. C. PET from the PET/MRI scan. D. T_1 -weighted MRI. E. T_2 -weighted MRI. The PET/CT scan was performed 60 minutes post-injection of ^{18}F -FDG, while PET/MRI started 120 minutes after injection. Geometrical differences between PET/CT and PET/MRI are attributed to arms-up protocol for PET/CT and arms-down protocol for PET/MRI. The lung tumor is indicated with green arrow in each image.

images using tracer kinetic modeling is also included, followed by a section on clinical and research applications with PET.

2.1 Principles of PET

PET is a medical imaging technique that visualizes the distribution of injected radioactive tracers in living subjects. By detecting annihilation photons from positron-emitting nuclei, mathematical methods allow the reconstruction of three-dimensional images of the tracer uptake. The following sections presents the underlying theory and physics of PET imaging.

2.1.1 Radioactive decay

PET imaging is based on radioactive, positron-emitting atoms. When an unstable parent nucleus undergoes radioactive decay, it is transformed into a stable state. The process is spontaneous, meaning that the exact moment of the decay cannot be predicted. Nevertheless, the number of nuclei in a radioactive sample at a given time point can be described mathematically, in terms of average decay rates. The average decay rate, or *activity*, A , for a sample of N radioactive atoms over time Δt , is defined as:

$$A = \frac{\Delta N}{\Delta t} = -\lambda N \quad [s^{-1}], \quad (2.1)$$

where λ is the *decay constant* for the radionuclide. From Equation 2.1, the number of atoms, N , after decay time, t in a sample of N_0 initial atoms, is given by:

$$N(t) = N_0 e^{-\lambda t} \quad (2.2)$$

Following from Equation 2.2, an expression for the decay constant can be derived:

$$\lambda = \frac{\ln 2}{T_{1/2}} \quad [s^{-1}]. \quad (2.3)$$

$T_{1/2}$ is the so-called *half-life*, a characteristic time for each radionuclide, at which precisely half of the initial number of radioactive atoms in a given sample remains (Figure 2.2).

As the activity is proportional to the number of atoms in a sample, Equation 2.2 can be expressed in terms of A :

$$A(t) = A_0 e^{-\lambda t} \quad [s^{-1}], \quad (2.4)$$

where $A(t)$ is the activity of the radioactive sample at time t , and A_0 is the initial activity at $t = 0$. The derived SI unit for activity is Becquerel ($1 \text{ Bq} = 1 \text{ s}^{-1}$).

2 Medical imaging with PET, CT and MRI

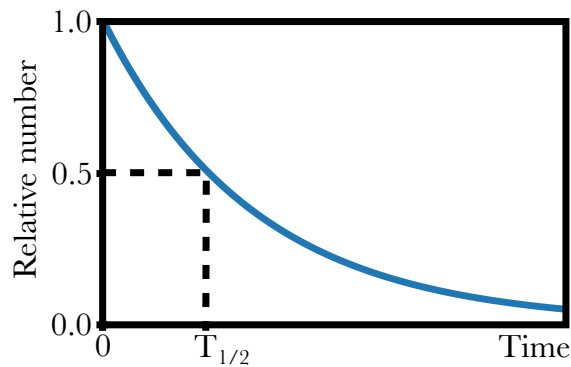


Figure 2.2: Illustration of exponential decay. The relative number of atoms at time $t = T_{1/2}$, is half the initial number of atoms at time $t = 0$.

2.1.2 Positron emission and annihilation

Nuclei which have an imbalance in the ratio of protons and neutrons are unstable and will, at some time point, undergo radioactive decay, to be transformed into another, more stable nucleus. In nuclei which are neutron deficient, positron-emission is one of the possible modes of decay. In such nuclear radioactive decay, one proton (p^+) is converted into a neutron (n) by the ejection of a positively charged electron, called positron (β^+), and a neutrino (ν), according to the reaction [6]:



The ejected positron traverses the surrounding material and loses energy in atomic collisions. Within a distance of a few mm, when the particle is close to rest, it combines with a free electron in an annihilation reaction. Both particles disappear, and their masses are converted into two anti-parallel photons according to $E = mc^2$. These two 511 keV, so-called *annihilation photons*, are the basis of the image-forming signal in PET. The positron emission and annihilation processes are depicted in Figure 2.3.

Positron-emitting radionuclides are commonly produced in a cyclotron by bombarding a target material with accelerated charged particles, such as protons or deuterons (${}^2_1\text{H}$ nuclei). The target material undergoes nuclear reactions and is transformed into an unstable nucleus that is prone to radioactive decay [6]. Several of the positron-emitting radionuclides have short half-life, thus requiring a cyclotron, in close proximity to the PET scanner. Some commonly-used cyclotron-based positron-emitting radionuclides are presented in Table 2.1.

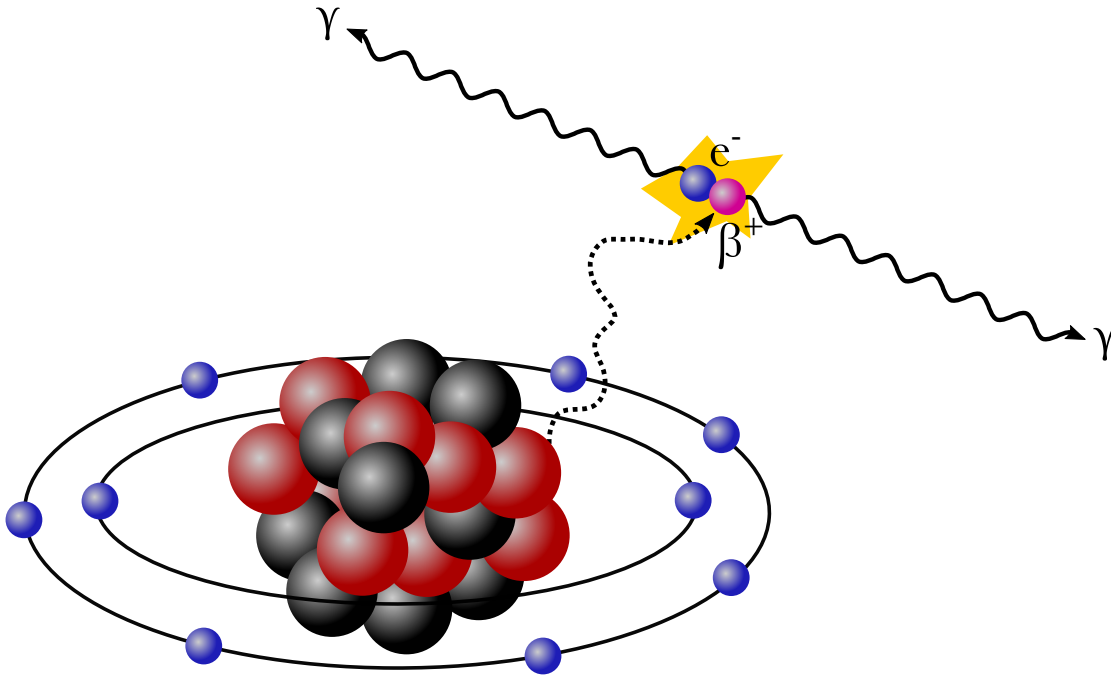


Figure 2.3: A radioactive nucleus undergoes radioactive decay by the emission of a positron (β^+). The positron, travels a few mm from the nuclei before it recombines with a free electron (e^-) by the emission of two anti-parallel annihilation photons (γ).

2.1.3 Photon attenuation

The photons emerging from the annihilation reaction interact with the surrounding medium through mainly two processes: Compton scatter, which is predominant at higher photon energies, and photoelectric effect, which is dominant at lower energies. These interactions essentially results in a reduction of the photon flux, a process called *attenuation*. The ability of a material to attenuate photons is characterized by the *linear attenuation coefficient*, μ , of the material. The number of photons, N , remaining in a photon beam of initially N_0 photons, after passing the distance x through a medium, is described by an exponential equation [51]:

$$N(x) = N_0 e^{-\mu x} \quad (2.6)$$

Both Compton scatter and photoelectric effect result in the ejection of orbital electrons, thus, causing ion pairs (a free, negatively charged electron and a positively charged atom) along the photon path.

2 Medical imaging with PET, CT and MRI

Table 2.1: Cyclotron-based positron emitting radionuclides commonly used in PET with their respective half-life [6].

Radionuclide	Half-life [min]
^{11}C	20.4
^{13}N	9.97
^{15}O	2.03
^{18}F	110

2.1.4 Photon detection

The orbital shell vacancy in each ionized atom, caused by either Compton scatter or photoelectric effect interactions, will eventually be filled with a free electron. In this so-called recombination process, energy is released by the emission of a photon [51]. PET detectors are made of special scintillating materials, in which the recombination energy is liberated as visible light. PET scintillation crystals should ideally have high photon attenuation, high light output and fast timing properties. Commonly used scintillation materials include bismuth germanate (BGO), cerium-doped lutetium oxyorthosilicate (LSO(Ce)), lutetium yttrium orthosilicate (LYSO) or thallium-doped sodium iodine (NaI(Tl)) [6].

The light produced in the recombination event is typically very small and must necessarily be amplified. In PET applications, this is accomplished by coupling the crystal to a photomultiplier tube or a light-sensitive semiconductor detector. Photomultiplier tubes are commonly used in hybrid PET/CT scanners. They are, however, incompatible with magnetic fields, making them unsuitable in hybrid PET/MRI scanners. In these applications, avalanche photodiodes (APDs) or silicon photomultipliers (SiPMs) are utilized as light amplifiers [6].

2.1.5 PET scanner design

In a PET scanner, multiple detector elements, each comprising a scintillator crystal coupled with either a photodiode or a photomultiplier tube, are placed adjacent to each other in a cylindrical geometry to form the *detector ring*. This allows near-simultaneous detection of the two opposite annihilation photons in a so-called *coincidence event*. Each two opposite detector pair defines a physical volume in which the annihilation event occurred. This volume is referred to as the *line-of-response*. Thus, by analyzing the signals from opposing detectors, a list of coincidence events along each valid line-of-response can be obtained [6]. Different types of coincidence events are discussed in Section 2.1.7.

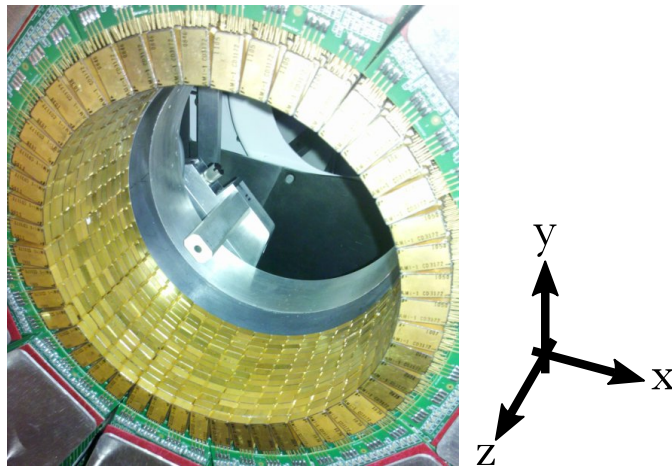


Figure 2.4: Photograph of the PET detectors of a 16-ring pre-clinical PET scanner with 15 cm ring diameter. Each gold-block contains the scintillator coupled with an avalanche photodiode. The holder for the rotating linear ^{68}Ge rod is visible in the background. The coordinate system for the PET geometry is shown to the right.

Some scintillation crystals, for instance LSO or LYSO, have a timing accuracy in the order of picoseconds, thus allowing to measure the tiny time-difference between the arrival of the two annihilation photons. In this way the annihilation event can be localized along the line-of-response to within a few centimeters. This concept, known as *time-of-flight* PET, may provide increased signal-to-noise ratio compared to non-time-of-flight PET, because information about the spatial localization of the event can be included in the reconstruction process [6].

Modern PET scanners are build up from multiple detector rings, placed adjacent to each other. In this way the sensitivity of the scanner is increased by allowing for detection of annihilation photons from oblique angles. Typically 8-24 detector rings are common, resulting in an axial field-of-view of 15-25 cm for clinical scanners [52, 53], or 5-15 cm for pre-clinical systems [54–56] (Figure 2.4). If larger axial coverage is needed, for instance in whole-body PET imaging, the images from multiple bed-positions are stitched together to form an extended-field-of-view PET image. The radial field-of-view is commonly in the range of 60-70 cm in diameter for human PET [52, 53], or around 8-10 cm in pre-clinical PET scanners [55, 56].

2.1.6 Raw data storage

The PET raw data can be stored in either sinogram-mode or list-mode format, which will be described next.

2 Medical imaging with PET, CT and MRI

Sinogram-mode

A sinogram is a two-dimensional matrix, in which the total counts for all line-of-responses are registered. The axes of the sinogram represents the distance, r , from the PET scanner center, and projection angle, θ (Figure 2.5). If only static PET images are needed, the raw data, consisting of the projection profiles are commonly stored as sinograms. For dynamic PET, multiple sinograms can be stored, each measuring the average number of PET events in a pre-defined time-interval. After acquisition, it is not possible to create new time-frames [6].

List-mode

When the PET raw data is stored in list-mode format, each detected coincidence event is recorded sequentially in a *list-mode table* together with a time stamp, at which the event occurred. Thus, the list-mode table consists of a list of coincidence events, in the order they were registered. This allows post-acquisition division of events into sub-tables with arbitrary-length, also called *binning*, and thus generation of arbitrary-length sinograms. These can, in-turn, be reconstructed into PET images of corresponding length, for instance as a dynamic PET sequence [6].

2.1.7 Coincidence event types

The photons originating from an annihilation event in the central part of the scanner reaches the detectors within a few nanoseconds. Three types of coincidence event may occur during PET acquisition, as depicted in Figure 2.6. If both photons are detected within a given *coincidence timing window* (typically in the order of 5-10 ns), and within a given energy window (typically 350-650 keV), the event is recorded as a *true coincidence event* (Figure 2.6A). *Random coincidences* occur when the annihilation photons from two unrelated annihilation events are detected within the coincidence timing window, and thus registered as an event. This results in a false line-of-response for this event (Figure 2.6B). *Scattered coincidences* occur when one or both of the annihilation photons are scattered by an angle and detected as an event. Also here, the result will be a false line-of-response (Figure 2.6C) [6].

The sum of true, random and scattered events are called *prompt coincidences*. These are all registered in the raw-data sinogram or list-mode table. However, as the random and scattered events have lost the spatial relation to their origin of emission, they are of no value in image formation, and would only contribute with noise in the reconstructed PET image. Therefore, they must be subtracted from

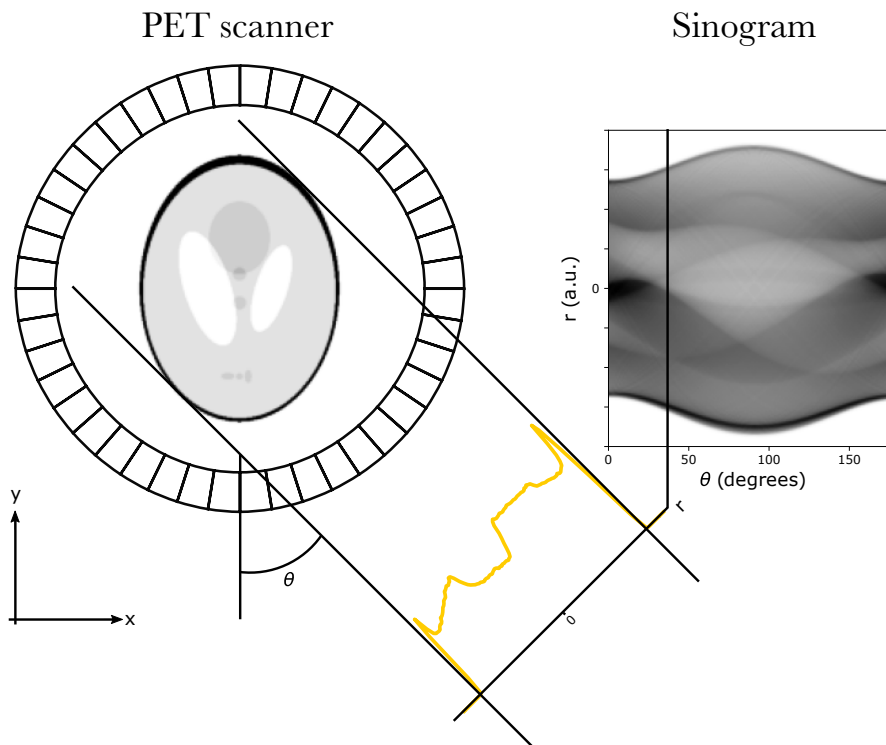


Figure 2.5: Illustration of the PET scanner geometry in image space (x - y) and the corresponding sinogram space (r - θ) in two dimensions. The sinogram is a representation of the projection profiles over all angles, θ , and radial distances, r , in image space. A point in image space will trace out a curve in sinogram space.

the prompt counts [6]. In fact, random and scatter correction are just two of several corrections applied to PET raw data, in order to make the PET image voxels have the units of radioactivity concentration. These corrections will be described next.

2.1.8 PET raw data corrections

Random-coincidence corrections

A random coincidence occurs when the photons from two unrelated annihilation events are registered by two opposing detectors, within the coincidence timing window. The random coincidence rate increases with the square of the amount of activity in the PET scanner. It results in reduced image contrast by adding a uniform background to the images and thereby altering the relationship between voxel intensity and activity.

2 Medical imaging with PET, CT and MRI

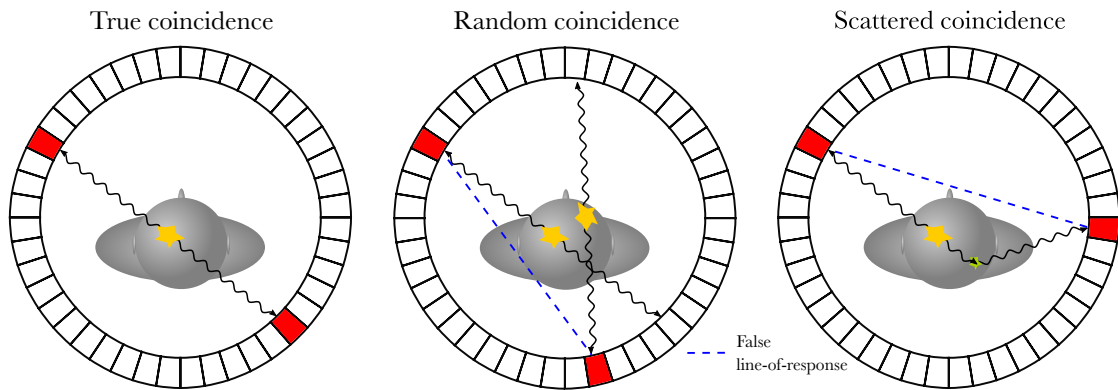


Figure 2.6: True, random and scattered coincidence events.

Random coincidences may be estimated from the *singles count-rate* of each of the two detectors connected by a line-of-response, as outlined in reference [5]. This requires recording of the single-events for each detector, which is the count-rate registered by each detector element independent of any possible detected coincidence events. Alternatively, the random coincidences may be estimated by acquiring a sinogram with a delayed time window. With this method, a delayed sinogram is acquired where only events that have arrival times separated by a pre-defined delay (for instance 50 ns) are recorded. Two annihilation photons from the same decay will hit the detectors within around 2 ns. Therefore, a 50 ns delay will reject all true (and scattered) events, and thus, only contain random events. The random coincidence sinogram is subtracted from the prompt sinogram, to result in a random-corrected sinogram [6].

Scatter correction

Scattered coincidences occur when one or both of the annihilation photons are scattered by an angle and hit a non-opposing detector, to result in a false line-of-response for this event. The false line-of-responses are manifested as induced haze and reduced contrast and quantitative accuracy in the resulting PET images. In human PET applications, the fraction of scattered events can be as high as 60%-70%. In theory, scattered events could be discarded by rejecting coincidence events below 511 keV. However, in practice, the scintillation crystal has limited energy resolution, and therefore, a finite window is needed, typically in the order of 350-650 keV. Consequently, only a small fraction of the scattered events are removed by the finite energy window [5].

To remove further scatter contributions, the scatter component can be estimated

from simultaneously acquired anatomical data (such as CT or MRI, see Section 2.4). One method for scatter correction is to register the counts outside the border of the object, which are mainly caused by the scattered component. By fitting a Gaussian function to the projection profile, it is possible to estimate the scatter contribution inside the object by extrapolation. The scatter component is then subtracted from the prompt counts [6]. Another method for scatter correction is to simulate the scatter contribution on the basis of a non-scatter-corrected reconstruction of the PET image and the attenuation correction map (see section 2.1.8). The estimated scatter fraction is then subtracted from the prompt counts and scaled to match the acquired PET data [5]. The scatter is estimated from either a simple single scatter model [57], or an analytical model [58]. This method is the most common scatter correction method implemented in clinical PET systems today [59].

Attenuation correction

The single most important correction for obtaining fully quantitative PET images is attenuation correction. As described in Section 2.1.3, a photon flux originating from a source inside a medium will be attenuated. As evident from Equation 2.6, the longer traversing path, the more attenuation. This will show as reduced tracer uptake in central parts of the imaged object. The attenuation correction can be derived from Equation 2.6. If an annihilation event occurs at depth x inside an object of thickness T , the probability, P_1 , that a photon will reach detector 1 along a line-of-response, is given by:

$$P_1 = \frac{N(x)}{N_0} = e^{-\mu x}. \quad (2.7)$$

Assuming that the other annihilation photon from this event was emitted in the direction of detector 2, it has to traverse a distance of $T - x$ through the object, to reach the detector. The probability that it will reach the detector is given by:

$$P_2 = e^{-\mu(T-x)}. \quad (2.8)$$

The probability, P , that both photons will reach the detectors is given by the product:

$$P = P_1 \cdot P_2 = e^{-\mu T}. \quad (2.9)$$

2 Medical imaging with PET, CT and MRI

Evidently, the total probability to detect both photons are only dependent on the total object thickness along the line-of-response and thus independent of the event depth. Thus, if the linear attenuation coefficient, μ , and object thickness, T , are known for a line-of response in the sinogram, the attenuation correction factor, A is given by:

$$A = \frac{1}{P} = e^{\mu T}. \quad (2.10)$$

In most situations μ will vary along T , and the integral over x is used to calculate the attenuation probability for a line-of-response:

$$A = e^{\int_{-\infty}^{\infty} \mu(x) dx}. \quad (2.11)$$

In this way, each line-of-response in the sinogram can be corrected by multiplication with A prior to reconstruction. In practice, the distribution of all linear attenuation correction values, called *attenuation correction map*, is obtained from a rotating transmission-source, or derived from either CT or MRI images (see sections 2.4.1 and 2.4.4). The attenuation correction map is then transformed into the sinogram domain and multiplied with the PET-sinogram to obtain an attenuation corrected PET-sinogram [5, 6].

Normalization

Because of small variations in crystal size, or varying coupling-efficiency with the light-amplifier, the individual detector elements of a PET scanner have varying photon-detection efficiency. To generate a uniform output from all detector elements, *normalization* is necessary. This is usually accomplished with a ^{68}Ge source. Briefly, if all detector elements are exposed to the same photon flux, the registered counts from each element will vary, depending on the efficiency. A normalization factor for each individual detector element is obtained by dividing its counts with the average counts from all elements directly in the sinogram domain [6].

Dead-time correction

When a detector element is hit by a photon, a finite amount of time is required for the reaction-chain of photon-interactions in the scintillator, light-amplification, and determination of the position and energy of the photon. During this processing-time, the detector element cannot process new photon hits. This is called *dead-time*. If

another photon hits the detector during its dead-time, the two signals may be erroneously added together, thus destroying both photon-signals. The effect of dead-time is therefore an underestimation of the true radioactivity concentration, and to avoid this, dead-time correction must be applied. This commonly consists of an empirical model, in the form of a look-up-table, where the relative underestimation of radioactivity concentration has been measured for different radioactivity concentrations and geometries. The factor from the look-up-table can be applied directly in the sinogram domain for any given geometry and concentration.

2.1.9 Image reconstruction

With the sinogram corrected for the effects discussed above, the PET image can be reconstructed by employing either analytical, or iterative methods, as described next.

Analytical methods

Analytical methods employ an invertible mathematical model to reconstruct the images. Filtered back-projection is the most commonly used analytical method because it is fast and computationally efficient. It is, however, rarely used in clinical practice, due to noise amplification in the reconstruction process. [6].

Iterative methods

Iterative reconstruction methods are commonly used in PET applications because of reduced noise and thus improved image quality compared to analytical reconstruction. The aim with iterative reconstruction is to approach the true image $f(x, y)$, by successive approximations, $f^*(x, y)$, as summarized in the following steps: 1) Start with an initial estimate of $f^*(x, y)$, being either a blank, or uniform image, 2) compute the forward-projections of the estimated image to obtain an estimated sinogram, 3) compare the estimated sinogram with the actual measured sinogram using a cost function, and use the difference between these to update the estimated image $f^*(x, y)$. Steps 2) and 3) are repeated until $f^*(x, y)$ converges towards $f(x, y)$ [6].

The *system matrix*, \mathbf{M} , is central in the iterative reconstruction process. Each element in the system matrix, j , represents the probability that a photon-pair is emitted from voxel i in the image and detected in line-of-response j . Thus, the system matrix relates the projection profiles to the underlying radioactive source

2 Medical imaging with PET, CT and MRI

distribution in the subject. \mathbf{M} is included in the iterative reconstruction loop to account for scanner-specific, or statistical properties of the data and can be estimated from the measured (or simulated) activity distribution of a point source placed at all locations inside the PET scanner [6].

Once the system matrix has been determined, one of several iterative algorithms may be applied to reconstruct images from the acquired projection profiles. One commonly known iterative algorithm is the maximum-likelihood expectation-maximization algorithm, in which the voxel activities are estimated through a series of iterations. This algorithm is computationally intensive, so the most common reconstruction algorithm in clinical PET applications is therefore ordered-subset expectation-maximization, in which only a subset of the projections are used in the first iterations. This will speed up the initial iterations, as fewer forward-projections have to be calculated. As $f^*(x, y)$ approaches $f(x, y)$, more projections are included to obtain a more accurate estimate [6].

Reconstruction of three- or four-dimensional data

A two-dimensional sinogram represents the projection profiles in one transversal slice (x-y-plane, see Figure 2.4) of the object inside the PET scanner. In analytical reconstruction methods, each slice is reconstructed separately from the respective sinogram. In iterative methods, the system matrix can as well include a three-dimensional relation between the voxels and projection elements, by simply expanding the indices i and j to three dimensions. Thus, a three-dimensional volume is reconstructed in each iteration [6].

For dynamic PET data, the sinogram of each time frame is either recorded directly (sinogram-mode), or is created post-acquisition from the list-mode data. Each sinogram then represents one time-frame of the dynamic sequence, and during reconstruction, each time-frame is reconstructed independently of all others. The final dynamic PET image can then be represented as a four-dimensional matrix (x, y, z, t) with three spatial dimensions, x, y, z and one temporal dimension, t [6].

In multi-ring PET systems, where oblique coincidence-angles are also accepted, each sinogram will contain counts from non-perpendicular line-of-responses. Thus, there is not a one-to-one relation between transversal slices and registered counts in the sinogram. One solution to this is to re-bin the oblique data into corresponding non-oblique transversal slices, for instance by assigning each oblique angle with its average axial position. This procedure is called *single-slice rebinning* [6].

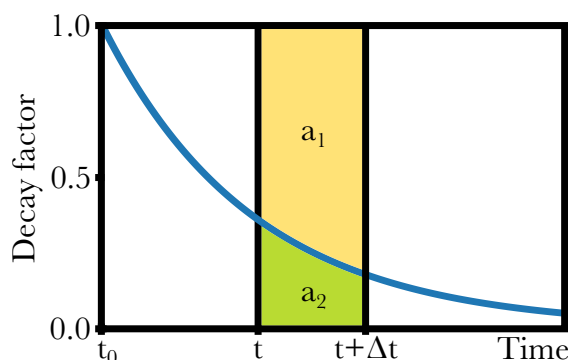


Figure 2.7: The concept for calculating the effective decay factor in a PET frame of length Δt , starting at time t . In the absence of decay, the counts during Δt are proportional to the total area $a_1 + a_2$. Due to radioactive decay, the counts during Δt will be proportional to a_2 . The effective decay factor is given by the ratio $a_2/(a_1 + a_2)$.

Decay correction

PET tracers have a short half-life in comparison with the PET image acquisition times (See Table 2.1). Therefore, it is necessary to perform decay correction of the acquired PET frames. The concept of *effective decay factor* is illustrated in Figure 2.7. For a time frame starting at time point t , with a duration of Δt , the effective decay factor, DF_{eff} , can be derived by the ratio of the area under the decay curve, a_2 (obtained by integration of Equation 2.4), to the total area with no decay, $a_1 + a_2$:

$$DF_{\text{eff}}(t, \Delta t) = \frac{a_2}{a_1 + a_2} = e^{-\lambda t} \cdot \frac{1 - e^{-\lambda \Delta t}}{\lambda \Delta t} \quad (2.12)$$

In practice, for a static PET acquisition, all image voxels are multiplied with the factor $1/DF_{\text{eff}}$ from Equation 2.12 in order to decay correct the registered counts within the static PET acquisition to $t = 0$. For a dynamic PET sequence, the $1/DF_{\text{eff}}$ factor is applied to the voxels in each time frame, using the specific starting point, t , for each time frame. In this sense, all dynamic PET time frames will be decay corrected to $t = 0$. It is common to define $t = 0$ as the injection time of the tracer, although this time point can be freely chosen [6].

Quantitative calibration

If appropriate sinogram-based corrections, as well as image reconstructions have been applied to the PET raw data, the voxel-values in the reconstructed image will

2 Medical imaging with PET, CT and MRI

be proportional to the radioactivity concentration of that voxel in the imaged object. A global quantitative calibration factor is then applied to all voxel-values, to convert the reconstructed image into the units of radioactivity concentration (kBq/cm³ or equally kBq/ml). The calibration factor is usually obtained by scanning a uniform phantom with known radioactivity concentration of a positron-emitting nuclide [6]. To reduce potential bias, the radioactivity concentration injected into the phantom should preferably be measured using the same measurement device (ion-chamber or well-counter) as used for the patient-injection syringes. This correction is therefore sometimes referred to as *well-counter correction*.

Standardized uptake value

Measurements of lesions or other high-uptake regions in a PET image are commonly performed by first segmenting the region, either manually or using automated methods, and then performing statistical measurements on the voxel-values inside the segmented region. In order to compare quantitative measurements between patients and scans, it is common to normalize the voxel-values, $C_{x,y}$ [kBq/ml] to the amount of injected radioactivity, D [MBq] per body weight, w [kg], to obtain an SUV image [60]:

$$SUV_{x,y} = \frac{C_{x,y}}{D/w} [g/ml] \quad (2.13)$$

Under the assumption that one ml of tissue weights one gram, SUV becomes dimensionless. Note that both $C_{x,y}$ and D in Equation 2.13 must be decay corrected to the same time-point. The SUV can be interpreted as the tracer concentration of a voxel (nominator of Equation 2.13) relative to a uniform whole-body concentration (denominator of Equation 2.13). For instance, an SUV of 1.0 is the equivalent uptake as if all tracer was uniformly distributed in the body [60]. As an illustration, the tumor in Figure 2.1A and C shows an increased SUV, thus represents an increased ¹⁸F-FDG-uptake compared to the surrounding tissue.

2.1.10 PET tracers

In clinical applications, positron-emitting radionuclides (Table 2.1) are often attached to carrying molecules. Such labelled molecules are called *radiopharmaceuticals*, *radiotracers*, or simply *tracers*. There exists a large variety of labeled compounds that allows imaging of different physiological and biological processes,

Table 2.2: Examples of PET radiopharmaceuticals and their application for imaging different biological processes [62, 63].

Radiopharmaceutical	Biological process
^{18}F -FDG	Glucose metabolism
^{18}F -FLT	Cell proliferation
^{18}F -fluoride	Bone metabolism
^{18}F -FMISO	Hypoxia
^{68}Ga -DOTATOC	Somatostatin receptor
^{68}Ga -DOTATATE	
^{15}O -water	Perfusion

such as glucose metabolism, blood flow, hypoxia, receptor-binding etc. Some commonly encountered radiotracers are listed in Table 2.2. The clinical work-horse in oncological PET is the glucose analogue ^{18}F -FDG, which allows non-invasive measurement of glucose metabolism [61].

PET tracers are typically injected in very small concentrations that do not alter the biological processes of interest. For instance, the intra-tumor ^{18}F -FDG concentration in a typical clinical PET scan is in the order of 10^{-15} mol/l in a high-uptake tumor. Thus, the PET scanner is a highly sensitive instrument for non-invasive imaging of these small tracer amounts [62].

2.1.11 Tracer uptake

The PET tracer is injected intravenously into the blood stream of the subject. It will then distribute into cells according to pathways specific for the tracer. Figure 2.8 displays the temporal distribution of ^{18}F -FDG over one hour in three tissues of a mouse. Myocardium shows an early peak due to the first-pass of blood with high tracer concentration directly after injection, while at later time points, the tracer uptake is increasing due to the constant glucose consumption of the beating heart. The kidney also shows an initial peak followed by decrease of activity as the tracer is excreted from blood, via the kidney, into the urinary bladder. Soft tissue and tumors mostly follow the dynamic pattern similar to that of the brain region. At early times, there is an initial rise of the activity, while after some time, there is a stable state plateau, or slow decrease of the activity. In static PET scanning of tumors, it is important that imaging takes place at this late stable state in order to achieve comparable scans within and between subjects. Therefore, it is common to let the subject rest for some time after injection, before imaging starts. This, so-called *uptake-time*, may vary according to tracer and scan protocol. In

2 Medical imaging with PET, CT and MRI

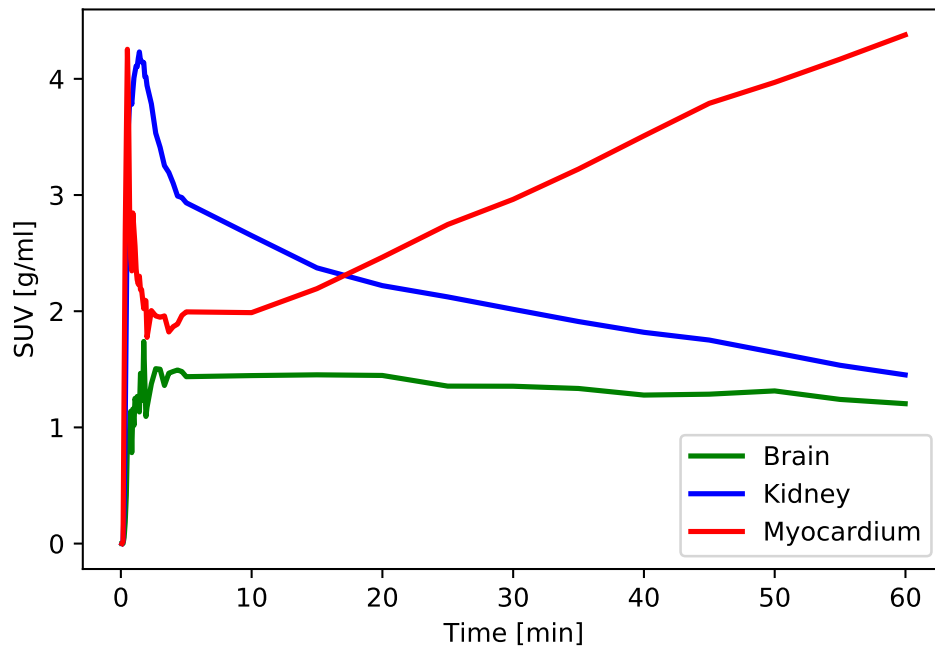


Figure 2.8: ^{18}F -FDG uptake over 60 minutes in brain, kidney and myocardium tissues of a mouse.

clinical routine static ^{18}F -FDG scans, 60 minutes \pm 10 minutes uptake-time is recommended before scan start. Also, as the glucose level affects the ^{18}F -FDG uptake in normal tissue, it is important that the subject has fasted at least 4 hours before scanning [64].

2.1.12 PET acquisition

Depending on the sensitivity of the PET system, the amount of injected tracer and the desired counting statistics, the duration of acquisition at each bed-position may vary in length. Imaging times of around 5-20 minutes per bed position are common in static PET imaging, while in dynamic PET imaging, acquisition-times between 10-60 minutes of one single bed position are common [40].

Recently, with improved sensitivity of modern PET scanners, it is possible to obtain high quality images in as short as 30 s per bed position. This allows each bed position to be scanned at multiple times during acquisition, and as such, allow for whole-body dynamic PET imaging [65].

2.1.13 Image quality in PET

Image quality in medical imaging is generally quantified by contrast, noise and spatial resolution. Compared to other imaging modalities, such as CT and MRI, PET has relatively high noise level and low spatial resolution, while the contrast depends on the tracer and examined body part.

Contrast

Contrast in PET images refers to the intensity of an object relative to the surrounding background voxels. For instance, in tumor PET imaging, it is desired to have a high tumor-to-background ratio, meaning that it should be easy to visually spot the tumor when reviewing the images. Contrast, thus, heavily depends on the tracer kinetics and the body region examined. For instance, a lung tumor may have high contrast, as the background lung activity is low (Figure 2.1), while a brain tumor may have lower contrast, since the background tracer uptake in normal brain tissue is high. Noise may also limit the image contrast, especially for low-contrast objects. Thus, it is desired to have a high contrast-to-noise ratio [6].

Noise

Noise in PET images originate from several sources. The random nature of radioactive nuclear decay follow Poisson statistics, such that the number of registered counts, N , during a time interval has a variance of \sqrt{N} [6]. This implies that short time frames, in for instance dynamic PET studies, have a lower signal-to-noise ratio (SNR) compared to longer time frames. If 100 coincidence events are registered during a short time frame, $\text{SNR} = N/\sqrt{N} = 100/10 = 10$, while for a longer frame duration, if 10 000 events are registered, $\text{SNR} = 10000/100 = 100$ [6].

The random and scatter corrections are unable to correct for all random and scattered events. Thus, there will exist false line-of-responses in the raw data that will introduce additional noise in the reconstructed images [6].

The reconstruction process itself will also affect the noise level in the images, depending on the number of iterations and subsets, filtering method etc. [6].

Lastly, when imaging short-lived positron-emitting radionuclides, the relatively long acquisition times will result in significant decayed tracer activity in late time frames. This will reduce the counting statistics, and as such, increase the noise level in late parts of the scan [6].

2 Medical imaging with PET, CT and MRI

Spatial resolution

Spatial resolution in PET is fundamentally limited by two physical effects. First, the finite range of the positron before it recombines with a free electron, results in a few mm distance between the site of decay and the site of annihilation. Moreover, because of the residual energy of the positron and the free electron at recombination, the two annihilation photons will have a small deviation from the 180° emission angle. This is called *non-collinearity effect*. These two effects combined results in a slightly shifted line-of-response for the origin of the radioactive decay, which cannot be corrected for. The physical resolution limitation depends on the distance travelled by the two non-collinear photons, thus for clinical (human) PET scanners with around 80 cm separation between opposite detectors, the lower bound on resolution is around 1.8 mm. On the other hand, small-bore pre-clinical PET systems have a lower bound physical resolution of around 0.7 mm [6, 66].

In addition to the physical limitations, the finite size of the PET detector elements, depth of interaction in the crystals, reconstruction method and filter, also affects the final spatial resolution. As a result of these additional effects, clinical PET scanners have a practical resolution limitation of around 2.4 mm, while pre-clinical may offer 0.8 mm resolution [6, 66].

Partial-volume effect

The partial-volume effect refers to two phenomena that diminishes the voxel-values of a PET image. First, the limited spatial resolution, due to the factors described above, introduces image blurring, which causes spill-over from high- to low-intensity regions in the image. This is equivalent to convolution of the source distribution in the image with the *point spread function* of the PET system. Due to these effects, in a PET image of a small point source, there will be apparent activity outside the physical boundary of the source [67].

The second contribution to partial-volume effect is the finite size of the 3D image voxel grid. The cubic voxels have discrete side-lengths that do not perfectly match the contours of the imaged object. For instance, the region covered by a voxel may contain both soft-tissue and blood, both of which have tracer uptake. Thus, the value will inevitably represent an average of the different underlying tracer-signals. This effect is therefore referred to as the *tissue-fraction effect*, and may have to be taken into account, for instance during kinetic modeling (Section 2.5) [67, 68].

2.2 Principles of CT

CT imaging is based on a rotating x-ray tube coupled with a radiation detector on the opposite side. A fan-beam of photons irradiates the subject, while the detector registers the attenuation profile on the opposite side. Simultaneously with tube-detector rotation, the patient table moves through the radiation beam, until the desired axial length has been irradiated. Similarly as for PET, the acquired projection profiles are stored as sinograms for each axial slice (see Section 2.1.6). The sinograms are reconstructed into a three-dimensional image matrix, using either filtered-backprojection or iterative methods, similar to those described for PET (see Section 2.1.9) [6, 69].

The voxels of a CT image represent the linear attenuation coefficient, μ , specific for the particular photon spectrum and filtration in the x-ray tube. These are commonly normalized to the linear attenuation coefficient of water, into so-called Hounsfield units (HU), according to [6, 69]:

$$\text{HU}(x, y, z) = \frac{\mu(x, y, z) - \mu_{\text{water}}}{\mu_{\text{water}}} \times 1000 \quad (2.14)$$

By applying HU normalization, the water voxels in a CT images will have a value of 0 HU, air will be -1000 HU while bone may approach values of 3000 HU. A common HU-range to visualize the lung tissues in a CT image is [-1000,200], as exemplified in Figure 2.1B.

2.3 Principles of MRI

MRI image formation is based on the spatial encoding of radio-frequency signals emitted from excited hydrogen nuclei (protons) which are relaxing in a stationary magnetic field. In general, any nuclei with an odd number of protons and/or odd number of neutrons have an associated nuclear spin. When placed in an external magnetic field, \mathbf{B}_0 , these nuclei will align along the field and start precessing around the field axis with a frequency referred to as the *Larmor frequency*, ω_0 , given by:

$$\omega_0 = \gamma B_0 \quad (2.15)$$

Here, B_0 is the magnetic field strength and γ is the gyromagnetic ratio, a constant which depends on the type of nucleus. A typical clinical magnet has a field-strength of 1.5-3 T, while in pre-clinical MRI, field-strengths may be in the range of 5-15 T.

2 Medical imaging with PET, CT and MRI

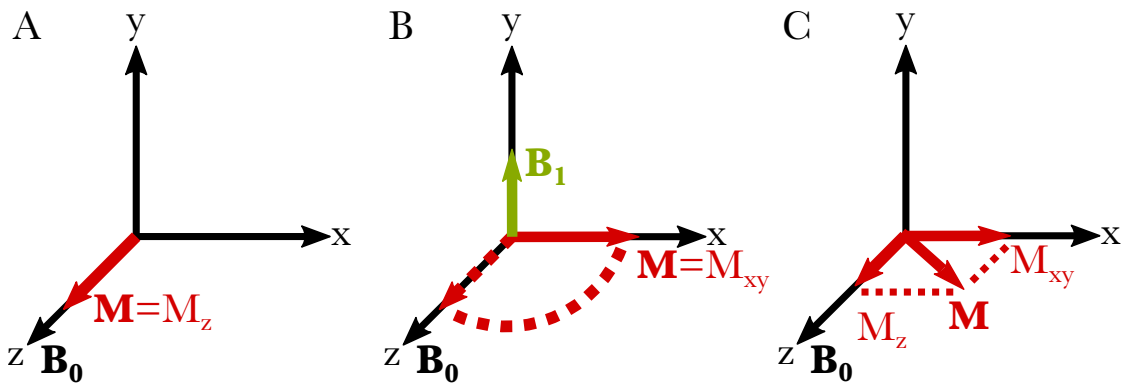


Figure 2.9: Illustration of the MRI coordinate system. A rotating frame-of-reference around the z-axis has been chosen, with the angular frequency equal to the Larmor frequency, ω_0 . A. The external magnetic field, \mathbf{B}_0 , is applied along z, causing the net magnetization, \mathbf{M} , to also align along the z-axis. B. A second magnetic field, \mathbf{B}_1 , is applied with the Larmor frequency during time t_p , perpendicular to \mathbf{B}_0 . This causes \mathbf{M} to be flipped. In this example, the time, t_p is chosen such that \mathbf{M} is flipped into the x-y-plane. In the rotating coordinate system, both M_{xy} and \mathbf{B}_1 appear stationary. C. When \mathbf{B}_1 is terminated, relaxation of M_{xy} begins, such that the x-y-component is reduced exponentially, with characteristic time, T_2 , and M_z is increased exponentially, with characteristic time T_1 .

Moreover, in medical MRI, the image-forming signals originate from the hydrogen atoms of the water molecules in the tissue [70]. Thus, with knowledge of the field strength and the type of nucleus, the Larmor frequency can be calculated, which is important for MRI image formation, as discussed later.

A hydrogen nuclei may exist in either of two spin-states, spin-up, or spin-down. In a given tissue sample in a magnetic field, around half of the hydrogen nuclei will align parallel to the magnetic field, while the other half will align anti-parallel. The spin-up state has slightly lower energy state than spin-down, and thus, the number of nuclei with spin-up will be slightly greater than the number of nuclei with spin-down. For instance, in a sample of 10^{23} nuclei, the difference in the number of nuclei in the different states is around 10^{17} [69]. This creates a *net magnetization*, \mathbf{M} , which is aligned parallel with \mathbf{B}_0 (Figure 2.9A).

2.3.1 Excitation and relaxation

If a second magnetic field, \mathbf{B}_1 , is applied with the Larmor frequency during time t_p , perpendicular to \mathbf{B}_0 , the individual spins will begin a precession around \mathbf{B}_1 . This will imply that \mathbf{M} is flipped by an angle α , given by [70]:

$$\alpha = \gamma B_1 t_p. \quad (2.16)$$

If \mathbf{B}_1 is applied for long enough time to flip \mathbf{M} by $\alpha = 90^\circ$, this is called a *90° pulse* (Figure 2.9B). Similarly, if the pulse is applied for double the time, it is called a *180° pulse*. These so-called radio-frequency (RF) pulses are emitted from coils wound in a loop around the main magnetic field \mathbf{B}_0 . Their operation are based on Faraday's law, namely that a sinusoidal alternating current through the loop at the Larmor frequency will generate an oscillating magnetic field perpendicular \mathbf{B}_1 . When the RF pulse is stopped after time, t_p , the excited magnetization will begin to relax, and start to align with \mathbf{B}_0 again (Figure 2.9C). The rate at which \mathbf{M} returns to \mathbf{B}_0 is characterized by the T_1 and T_2 relaxation-times, respectively. T_1 describes the exponential recovery of \mathbf{M} along \mathbf{B}_0 , while T_2 describes the exponential loss of transverse magnetization.

The exponential recovery and relaxation of the magnetization vectors follow the form of Equation 2.2, with λ being equal to $-T_1$ and $+T_2$, respectively. In practice, the magnetization in the transverse plane disappears within a few ms, while it may take up to 15 s to recover \mathbf{M} along \mathbf{B}_0 in water (tissue dependent) [69].

Faraday's law also works in reverse, namely that an alternating magnetic field around a coil will induce a current in the coil. Thus, when the magnetization returns to ground state, a current is induced in the RF coil, called *free induction decay*. Different tissues have different T_1 and T_2 relaxation-times. The difference in relaxation-properties of the different tissues may provide contrast between tissues in the resulting MRI-image, as will be described Section 2.3.3 [69].

2.3.2 Spatial encoding

In a clinical MRI scanner, the entire patient, or the part of the patient to be scanned, lies in the magnetic bore of the scanner, within a homogeneous \mathbf{B}_0 . Any RF pulse, \mathbf{B}_1 , emitted from the RF coil, will flip the magnetization in the entire section of the patient inside the scanner, and the recorded relaxation-signal will not represent any meaningful anatomy. Therefore, in order to form a three-dimensional image, spatial encoding of the RF signals are required. This is commonly accomplished by applying z, y, and x-directional gradients through special *gradient coils*. As an example, the spatial encoding for the spin-echo pulse sequence is outlined here.

First, a gradient coil defines the field-of-view along the z-direction, by introducing a gradient, G_z , along \mathbf{B}_0 . Because the Larmor frequency depends on the magnitude of \mathbf{B}_0 , with the gradient coil applied, it will vary along the z-direction, allowing the

2 Medical imaging with PET, CT and MRI

activation of distinct slices by the emission of RF-pulses with a Larmor frequency matching that of a specific slice. Thus, the frequency of the RF-pulse selects a specific slice along the z-direction, while the bandwidth of the RF pulse determines the slice thickness. [70].

Localization in the y-direction is performed using another set of gradient coils, namely the *y-gradient coils*. These change the magnetization along the y-direction by applying a gradient, G_y , in the activated slice, subsequent to the slice selection gradient. The protons located in high y-positions will precess faster than protons located in low y-positions. When the y-gradient is switched off, the precession returns to that of ω_0 , but with the y-position of each voxel encoded by different phases of the relaxation signal [70].

Finally, another set of gradient coils, positioned at 90° relative to the y-gradient coils, encode the spatial positions in the x-direction. This is accomplished by applying a magnetic field gradient, G_x , in the x-direction subsequent to the y-gradient during readout. This results in a varying frequency of the received signals along the x-direction [70].

The three gradients, G_z , G_y and G_x , are commonly referred to as *slice*, *phase* and *frequency encoding gradients*. These may be applied in any given spatial direction to encode the relaxation-signal, not only in z, y, and x, as exemplified here.

2.3.3 Pulse sequences

By arranging the various excitation pulses and gradients in so-called *pulse-sequences*, it is possible to spatially differentiate tissues with different relaxation-times. A basic spin-echo sequence is visualized in Figure 2.10. This sequence consists of a 90° pulse, at time t_1 , that flips the magnetization into the transversal plane, followed by a 180° pulse, applied at time t_2 , that flips all magnetization by 180° . During the time between t_1 and t_2 , some of the transversal magnetization has started to de-phase (the free induction decay), such that at a specific echo-time (TE), $TE = 2 \cdot (t_2 - t_1)$, there will be a re-phasing *echo* peak. This echo is the image-forming signal in the spin-echo sequence [70].

Another 90° pulse is applied at repetition-time (TR), $t = TR$, after the initial 90° pulse. If this pulse is applied before all magnetization, \mathbf{M} is rebuilt (depending on T_1 for each specific tissue), the next echo will have lower amplitude. Thus, by choosing TR short enough, before all magnetization has recovered along the z-axis, it is possible to obtain MRI images where the contrast depends on T_1 properties of the tissue. Such images are called *T_1 -weighted*. Tissues with short T_1 (such as fat or

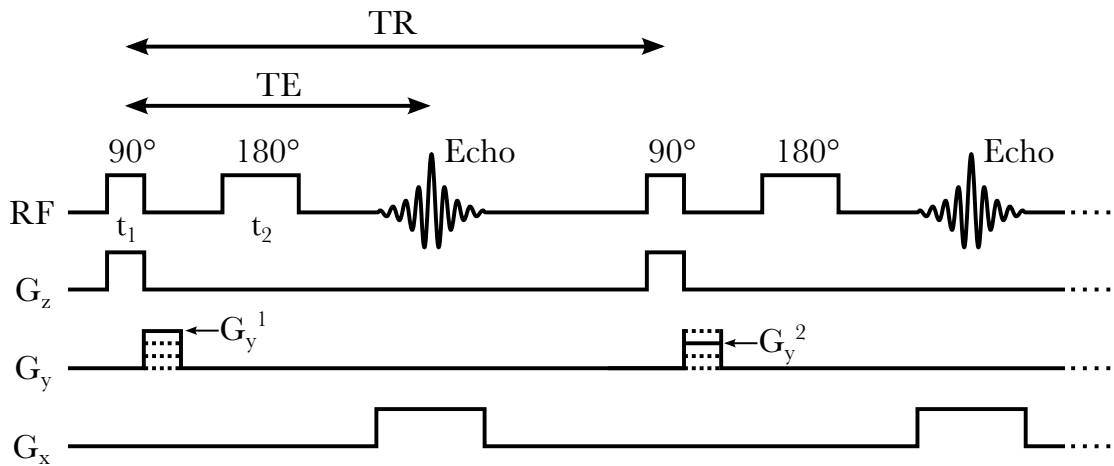


Figure 2.10: The spin-echo pulse sequence. A 90° RF pulse is applied at time t_1 together with the slice selection gradient, G_z . Subsequently, the first phase-encoding gradient, G_y^1 is applied, followed by a 180° pulse at time t_2 . An echo is formed at *echo-time* TE. The pulse sequence is repeated after *repetition-time* TR, but this time with phase-encoding gradient G_y^2 . When the sequence has been repeated for all phase-encoding gradients, the next slice can be imaged by repeating all phase-encoding steps for another slice-selection gradient.

gray matter) will produce larger signals, and thus brighter voxel-values, compared to tissues with long T_1 (such as water or cerebrospinal fluid) [70].

Similarly, for long TR values, by adjusting TE, it is possible to generate T_2 -weighted images. Here, the recorded signal depends on the amount of magnetization still present in the transversal plane, which is a characteristic of T_2 . Thus, tissues with short T_2 relative to TE (such as fat and gray matter) will produce a weak signal, because most of the signal has decayed before the echo occurs. On the other hand, tissues with long relative T_2 will decay less until the 180° pulse, and thus yield a higher relative signal [70]. For instance, the tumor in the T_2 -weighted image in Figure 2.1D appears brighter than the surrounding tissue.

One final possibility can occur for the spin-echo sequence. With long TR and short TE times, the signal will be dependent on the initial magnetization in the transversal plane, which depends on the proton density of the tissue. Such images are therefore *proton-density weighted* [69, 70]. An overview of the three different effects from varying TR and TE is shown in Figure 2.11.

Sequences may also be build up from other combinations of flip-angles and read-out-gradients. For instance, the *gradient-echo* sequence uses flip-angles below 90° , and application of two opposite gradients: the first to de-focus the transversal spins, followed by a second read-out gradient, to re-focus the spins, and at the

2 Medical imaging with PET, CT and MRI

		TR	
		Short	Long
TE	Short	T ₁ -weighted	Proton density
	Long	-	T ₂ -weighted

Figure 2.11: Effects of various combinations of repetition-time (TR) and echo-time (TE) in the spin-echo MRI sequence.

same time perform read-out encoding [69]. Because of the partial flip-angle and immediate read-out, the gradient-echo sequence allows for very short TR, and thus short overall imaging times.

2.3.4 Signal detection and k-space

The RF-echo emitted during spin re-phasing is detected as an induced current in two orthogonal receiver coils. The current amplitude as a function of time is amplified, digitized and stored in a discrete, two-dimensional matrix, referred to as *k-space*. As the two coils produce the same signal (but with 90° phase-shift), they are stored in k-space as a complex number, with one of the two signals being the real part, while the other signal being the imaginary part [70].

For the spin-echo sequence (and many other pulse sequences), k-space is filled row-by-row. Each row corresponds to the received signal from one specific application of the phase-encoding gradient. Each column represents the signal amplitudes at different times during the echo. Lastly, each two-dimensional k-space matrix corresponds to one specific axial slice in the patient (that of the current slice-selection gradient). The inverse Fourier-transform (IFT) of the k-space matrix from one specific z-slice results in a two-dimensional MRI image of that z-slice. Figure 2.12 illustrates the concept of k-space and MRI image reconstruction in two dimensions [70]. The final three-dimensional MRI image is the result IFTs of a three-dimensional k-space matrix. Note that the IFT of a complex k-space results in a complex image space with both real (Re) and imaginary (Im) parts. In most common applications, only the magnitude image is stored, where magnitude = $\sqrt{Re^2 + Im^2}$.

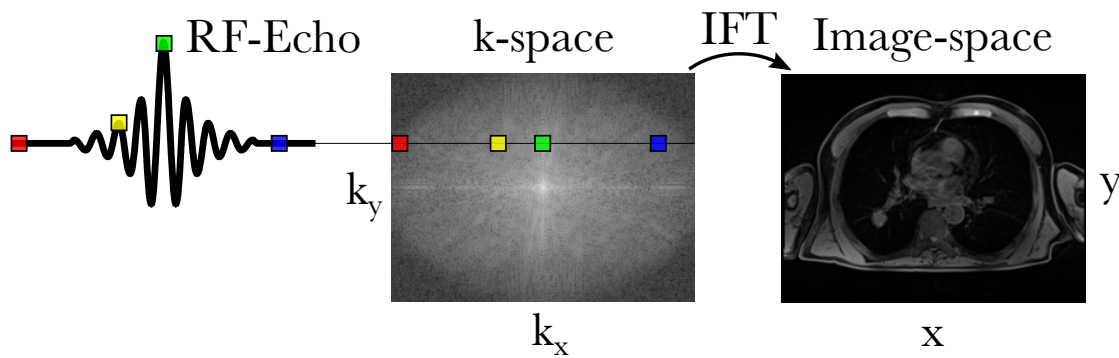


Figure 2.12: Illustration of signal detection and image reconstruction in MRI for one axial z-slice. The RF-Echo (see figure 2.10) emitted during spin re-phasing is received by the induction of a current in two orthogonal coils. The current amplitude as a function of time is digitized and stored as a complex number in a discrete matrix, referred to as *k-space*. In the spin-echo sequence, k-space is filled row-by-row, where each row corresponds to one application of the phase-encoding gradient. The inverse Fourier-transform (IFT) of k-space results in the final MRI image of the current z-slice. Note that only the magnitude of the real (Re) and imaginary (Im) parts of the k-space matrix is shown in the illustration above, where $\text{magnitude} = \sqrt{\text{Re}^2 + \text{Im}^2}$.

2.4 Hybrid imaging

While PET provides functional images of biological processes in the body, most tracers provide no direct anatomical localization of the tracer uptake. Therefore, in practice all commercially available PET systems, both clinical and pre-clinical, are integrated with an anatomical imaging scanner, such as CT or MRI. Two such hybrid imaging systems, PET/CT and PET/MRI will be described next.

2.4.1 PET/CT systems

In hybrid PET/CT imaging, the CT component serves two purposes, namely that of anatomical localization, and that of attenuation correction. The Hounsfield units of the CT image are proportional to the linear attenuation coefficient at the effective energy of the particular x-ray beam and filtration used during CT scanning. For the CT image to be valid at the energies of the PET annihilation photons, the voxel-values must be converted into the linear attenuation coefficient at 511 keV for each specific tissue in the CT image. The energy dependence of the linear attenuation coefficient is significantly different for soft tissue and for bone. Therefore, the conversion is usually done with a bi-linear mapping, where voxel-values with $\text{HU} < 0$ (such as soft-tissue, water and air) are scaled differently compared to voxels with $\text{HU} > 0$ (such as different types of bone) [6].

2 Medical imaging with PET, CT and MRI

2.4.2 Radiation dose

Both PET and CT are based on physical emission of high-energy photons (511 keV in PET and a spectrum of photon energies with maximum energy of typically 120 keV for CT). When these photons undergo interactions with the surrounding tissue (Compton scattering or photoelectric effect, Section 2.1.3), electrons are liberated. These electrons, in-turn, undergo many charged-particle-interactions with the surrounding tissue, transferring their kinetic energy to the surrounding material in many small steps. *Absorbed dose* is the total energy deposited in a material per unit mass, and it is measured in the units of *gray* (Gy), where $1 \text{ Gy} = 1 \text{ J/kg}$ [51].

The *effective dose* may be calculated by taking into account the sum of absorbed dose in all organs, the type of particles (photons or electrons), and the radiation-sensitivity of each organ. Effective dose has the unit of Sievert (Sv) and is intended for use as a radiation protection quantity, rather than for individual risk-estimation purposes. By incorporating radiation- and tissue-specific weight factors, it serves as a measure of population-based risk-estimate for cancer, with 5.5%/Sv [71].

The radiation dose from a typical ^{18}F -FDG PET scan is in the order of 6 mSv, while the CT contributes with an additional of 10-20 mSv, depending on scan protocol and body region scanned [21, 72, 73]. For comparison, the annual background radiation in Norway is around 3 mSv [74].

2.4.3 PET/MRI systems

Hybrid clinical PET/MRI systems were introduced commercially in 2008, even though the idea to combine excellent soft-tissue contrast from MRI, with functional imaging of PET was born already before the first PET/CT was released [75]. Nevertheless, several technological challenges had to be solved before a hybrid PET/MRI system could be developed; most importantly designing PET detectors that could operate in strong magnetic fields, as well as deriving the attenuation correction map for PET, based on MRI data [53]. The potential advantages with hybrid PET/MRI over PET/CT include the excellent soft-tissue contrast of MRI, compared to CT, the true simultaneous acquisition of PET and MRI images, as opposed to sequential acquisition in PET/CT, and the possibility of MRI-based motion correction. In addition, MRI, as opposed to CT (and PET) does not involve ionizing radiation, and therefore, does not contribute to the radiation dose to the patient.

Table 2.3: Linear attenuation coefficient values assigned to specific tissues in the DIXON attenuation correction sequence in PET/MRI.

Tissue	μ [cm^{-1}]
Background	0
Lung	0.0224
Fat	0.0884
Soft-tissue	0.1

2.4.4 MRI-based attenuation correction

The accuracy of attenuation correction is critical for quantitative PET imaging. This has been one of the more challenging tasks to solve during integration of PET and MRI systems, mainly because the small bore-size and high magnetic field-strength of MRI scanners did not allow integration of a CT-system or rotating transmission sources for attenuation correction purposes [42]. Current solutions include the conversion of MRI images to attenuation maps. This is not as straightforward as in CT attenuation correction because the MRI signal is related to proton density and spin-relaxation-times of tissue and does therefore not, as opposed to CT, reflect any attenuation information. Thus, the bi-linear conversion approach outlined for PET/CT systems in Section 2.4.1, is not valid for PET/MRI [42]. Instead, it is common to derive the attenuation correction maps by segmenting the MRI images into distinct tissue classes. The DIXON MRI sequence is well-suited for this task, because it consists of consecutive scans that provide images of water and fat separately. By applying image post-processing, this allows for segmentation into four tissue classes (air, fat tissue, lung tissue and soft tissue), and subsequently assigning each class the corresponding known linear attenuation coefficient (Table 2.3) [8].

Recently, deep-learning-based methods have been proposed for the direct transformation of MRI images to CT attenuation correction maps [76, 77], although such solutions are not yet clinically available.

2.4.5 Image artifacts in PET/MRI

Since PET/CT imaging has been available for many years, several well-known artifacts for this hybrid modality have been described. For instance, both metallic implants and intravenous contrast agents may affect the CT-based attenuation correction map, resulting in overestimation of the PET activity in the affected region [78, 79]. On the other hand, with the introduction of hybrid PET/MRI,

2 Medical imaging with PET, CT and MRI

new types of problems and artifacts has emerged. These will be described in the following sections, and are relevant for **Paper I**.

Truncation artifacts

The field-of-view of clinical PET/MRI systems is often too small to completely cover the subject in the lateral (x) direction, leading to truncation artifacts, and consequently, inability to correct for attenuation outside the MRI image (Figure 2.13A) [10].

Truncation of the field-of view may be solved by estimating the missing parts of the attenuation map based on PET uptake in these regions [80]. This works well for non-specific PET-tracers, with a high background uptake in soft-tissue, such as ^{18}F -FDG (Figure 2.13D). Alternatively, a pure MRI-based approach may be applied, where the magnetic field at off-center positions are corrected for inhomogeneities. This allows for obtaining two additional scans where the field-of-view can be shifted in the lateral direction [81].

Absence of bone signal

The bone-signal in most MRI sequences is small, due to very short relaxation-times for bone. Therefore, while bone is usually not visible in MRI images, it is highly attenuating the annihilation photons. Inability to compensate for the lack of bone may lead to underestimations in the reconstructed PET images, especially in regions close to osseous tissue [9].

The absence of bone signal is commonly solved for whole-body imaging by using an atlas-based bone models, which consists of major bones, such as skull, spine and pelvis (Figure 2.13F) [82]. These models, however, lack smaller bones such as ribs and may be prone to misalignment errors, especially in the lung region, due to respiratory motion [82, 83].

Atlas-based bone models are, obviously, not capable of taking into account abnormal patient anatomy. Ultra-short echo-time sequences have sufficient timing properties to detect the very rapid T_2 decay of bone. These gradient-echo-based MRI sequences are therefore commonly employed in brain imaging to derive the bone signal of the skull [84].

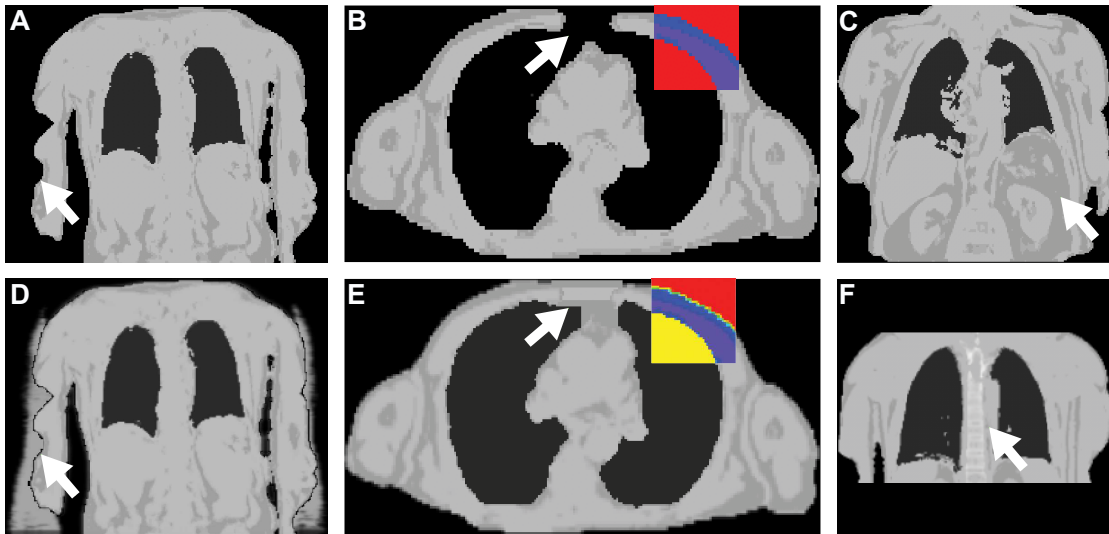


Figure 2.13: Examples of artifacts and corrections in the MRI-based attenuation correction maps. A. Truncation artifact, where the arms of the patient are outside the MR field-of-view. B. Susceptibility artifact caused by surgical sternal wires. Note the failed lung segmentation, where lung attenuation values have been incorrectly assigned to background (red color in the box insert). C. Tissue inversion artifact with a soft-fat tissue swap. D. Correction of truncated arms using the background tracer activity [80]. E. Susceptibility artifact correction using a post-processing algorithm [12]. Note that the lungs have been "filled" with lung attenuation values (yellow color in the box insert). F. Correction of missing bone by adding an atlas-based bone model of the spine [82].

Susceptibility artifacts

Local magnetic field inhomogeneities, caused by, for instance, metallic implants, such as endoprotheses, sternal ceclages and dental fillings, may result in signal voids in MRI images, and thus erroneous tissue classification of the attenuation maps [85]. In regions where tissue is falsely being classified as air (Figure 2.13B), this may lead to severe underestimation of the attenuation coefficient, and thus underestimation of the tracer activity concentration [86]. Several approaches have been proposed to overcome the issue, suggesting either variations of pulse-sequence parameters to minimize signal void [87–90], or by post-processing of the attenuation correction map (Figure 2.13E) [86, 91–93].

Tissue inversion artifacts

During segmentation of the water and fat images from the DIXON-sequence, the water and fat classes may sometimes be incorrectly assigned, leading to a swap of

2 Medical imaging with PET, CT and MRI

the water and fat classes, and consequently incorrect attenuation values for these tissues (Figure 2.13C). This may result in under- or over-estimations of the tracer activity concentration in the affected tissue [11, 46]. There is no straight-forward solution to tissue-inversion artifacts, other than manual post-processing, or repeated examination [46].

2.4.6 Motion correction in PET

Patient motion due to breathing or muscle relaxation is inevitable during a typical PET acquisition. Certain imaging sequences, such as the 19-second DIXON MRI sequence, are short and allow patients to follow breath-hold instructions. Nevertheless, some patients are unable to comply with these due to their ongoing disease, while some imaging sequences, such as a 20 minute PET acquisition, obviously are too long to allow for breath-hold. In these cases, motion causes unwanted blurring, which degrades the quality and the quantitative accuracy of the PET images (Figure 2.14A). Breathing-induced motion may also introduce misalignment in the co-registered PET and MRI images. Specifically, any mismatch between the attenuation correction map (acquired during end-expiration breath-hold) and the PET data (acquired in free-breathing) may lead to image artifacts [46].

In **Paper II**, two methods to correct for motion in PET images have been studied, based on end-expiration gating and elastic motion correction. These two methods will be described next.

End-expiration gating

To reduce breathing-induced motion-blur and artifacts, end-expiration gating can be utilized. With an external respiration sensor, usually an air-filled cushion positioned on the thorax of the patient, the breathing amplitude is recorded as a function of time, along-side with the list-mode PET-data. In this way, the list-mode table can be binned in distinct phases, corresponding to the different breathing phases of the respiratory cycle. The coincidence events from each respiratory phase are separated and used to reconstruct a PET image for each respiratory phase [94, 95]. Most commonly, only the end-expiratory phase is of interest, as this corresponds to the most stationary part of the breathing cycle, and consequently, the PET image with least motion-blur.

Figure 2.14B displays the reconstructed PET image of the end-expiratory phase for a lung cancer patient. This resulted in less apparent motion-blur, but due to

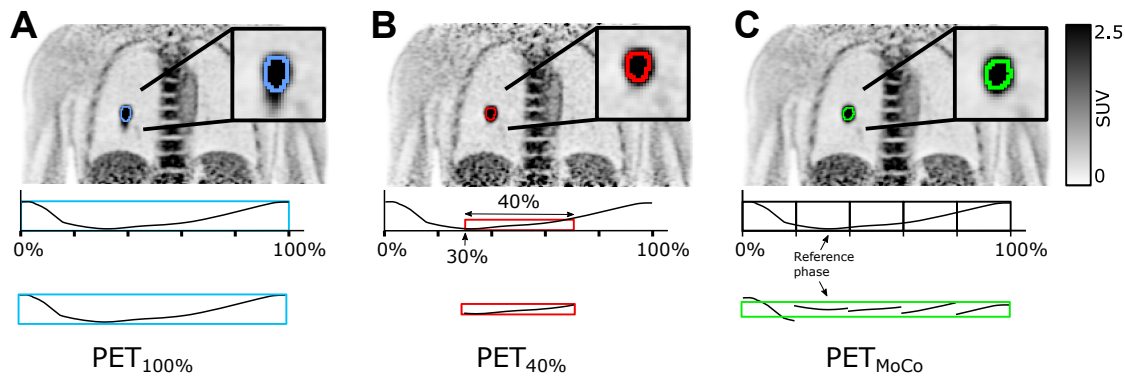


Figure 2.14: PET image of a patient with a lung tumor (top row), respiration curve (middle row) and an illustration of the amount of PET data used for reconstruction and its variability per breathing cycle (bottom row). A. Free-breathing PET, where all (100%) of the PET data is used for reconstruction ($\text{PET}_{100\%}$). B. End-expiration gating, where only 40% of the PET data, from the stationary end-expiration part of the breathing-cycle, are used for reconstruction ($\text{PET}_{40\%}$). C. MRI-based motion corrected PET data (PET_{MoCo}). The primary tumor and corresponding 41% SUV_{max} delineation is shown in the right lung and in the insert.

the reduced number of included counts, end-expiration gating also results in an increased relative noise level, as visible when comparing Figure 2.14A and B.

Elastic motion correction

To overcome the increased noise level in end-expiration gated PET, elastic-motion correction can be performed. Here, the breathing cycle is divided into a finite number of phases (for instance 5 phases), and PET images of each phase are reconstructed, each representing one part of the breathing cycle. The elastic motion vectors that transforms each phase into a reference phase (commonly the end-expiration phase) are then calculated and applied to the gated PET images. This will transform each phase image to the reference phase. The transformed images are then summed to provide a single, motion corrected PET image including all acquired PET counts, thus having an improved signal-to-noise ratio (Figure 2.14C) compared to end-expiration gating (Figure 2.14B).

The motion vectors used for transformation of the images from one phase to the reference phase are derived from the *image velocity* between two image frames [96], which will be described next.

2 Medical imaging with PET, CT and MRI

Estimation of image velocity

Image velocity refers to the motion vectors of individual image voxels that describe the movement of intensities between time frames. Briefly, a voxel at location (x, y, z, t) , with intensity, $I(x, y, z, t)$, will have moved $\Delta x, \Delta y, \Delta z$, during time between two image frames, Δt , such that the following equation holds true [97]:

$$I(x, y, z, t) = I(x + \Delta x, y + \Delta y, z + \Delta z, t + \Delta t) \quad (2.17)$$

The image velocity, \vec{V} , for this voxel is defined as the distance, $\Delta x, \Delta y, \Delta z$, it has moved during time Δt :

$$\vec{V} = \left(\frac{\Delta x}{\Delta t}, \frac{\Delta y}{\Delta t}, \frac{\Delta z}{\Delta t} \right) \quad (2.18)$$

The components of \vec{V} can be estimated from the derivatives of the image in all four dimensions, x, y, z, t , using a three-dimensional expansion of the Lucas-Kanade algorithm, as described in detail in reference [97].

Once \vec{V} has been estimated, the voxels in a given time frame $t + \Delta t$ can be transformed into time frame t by inverting the displacement vectors $\Delta x, \Delta y$ and Δz [97].

In **Paper II**, the image velocities in Equation 2.18 were derived from dynamic MRI data, which will be described next.

Elastic motion correction in PET/MRI

In PET/MRI applications, the motion vectors may be derived using a rapid radial stack-of-stars spoiled three-dimensional gradient echo sequence, which samples the center of k-space with a frequency of 5-10 Hz. This sequence allows to extract both a respiratory signal from the k-space center as well as MRI images of each phase of the breathing cycle. The MRI phase images can, in-turn, be used to derive the motion vector fields between the breathing phases, as described above.

In many situations, it is desired to use the MRI system to acquire several other sequences simultaneously with PET. Therefore, a respiratory sensor may be used (the same cushion as for respiratory gated acquisitions). The MRI-based respiratory signal can then be correlated to the cushion-based signal to allow for motion correction of the entire PET scan, even after the radial MRI sequence has finished.

Subsequent to PET acquisition, the list-mode PET data is binned into sinograms corresponding to the respiratory phases of the MRI-derived motion vector fields. The motion vectors are then transformed into sinogram-space and applied to the gated PET sinograms to transform all PET phases to the reference phase. The motion corrected sinograms are then summed and reconstructed into motion corrected PET images using the end-expiration MRI-based attenuation correction map (Section 2.1.9) [98]. Figure 2.14C illustrates an MRI-based motion corrected PET image.

The end-expiration attenuation correction map from the DIXON sequence can be compared with the MRI images from all breathing-phases. In this way, patients who were unable to comply with breathing instructions (which is not uncommon in lung cancer imaging) may be identified. In these cases, the acquired attenuation correction map (in, for instance end-inspiration) can be corrected by transforming it into the end-expiration phase with the corresponding MRI-derived motion vectors. The corrected attenuation correction map can, in-turn, be used during PET-image reconstruction.

2.5 Tracer kinetic modeling

When a PET tracer is injected into the blood stream, it circulates to all cells in the body through the vascular system. The tracer may be taken up in extra- or intra-cellular space or may be passed back into the blood. Thus, the tracer concentration in both blood and tissue will be time-varying and depend on specific biological properties, such as cell metabolism, cell-surface receptors or blood flow. For instance, Figure 2.15 illustrates the time-dependent blood and brain-tissue activity following injection of ^{15}O -water in a human subject.

The underlying biological processes which control the tracer distribution can be quantified by using the rate-of-change of the radioactivity concentrations in tissue, measured by PET, as well as the measured arterial blood concentration, the so-called AIF. This is accomplished by constructing mathematical models which can be fit to the measured PET time-activity curves and to the measured AIF. Such models are called *tracer kinetic models*, and the estimated parameters of these models are rate constants quantifying a specific biological process. For instance, dynamic ^{18}F -FDG-PET imaging allows the quantification of glucose metabolism in a tissue (**Paper III**), and dynamic ^{15}O -water-PET scans can be used to measure tissue perfusion (blood flow) in the brain (**Paper IV**).

2 Medical imaging with PET, CT and MRI

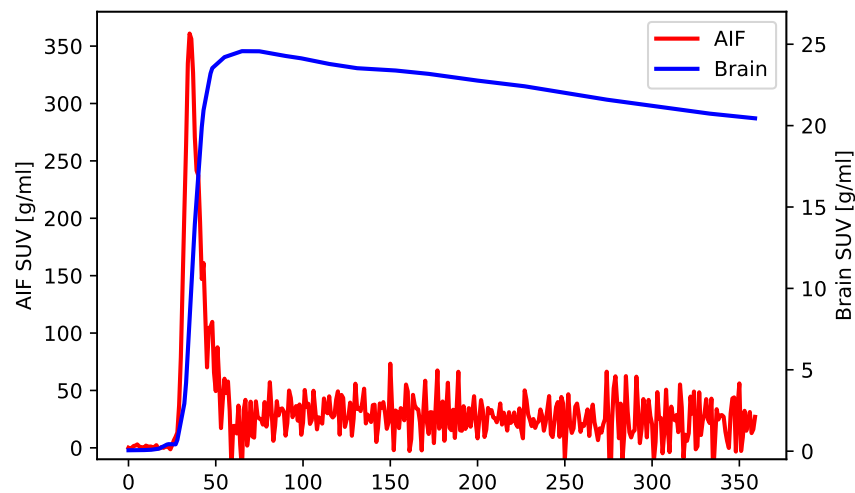


Figure 2.15: Illustration of an arterial input function (AIF) measured in blood and the whole-brain radioactivity concentration measured with PET, following the injection of ^{15}O -water in a human subject. Note the different scaling on the y-axis for the two curves.

2.5.1 Compartments

An assumption in tracer kinetic modeling is that the PET tracer is available in physiologically separated pools, or *compartments*. A compartment may represent a specific spatial region, such as extra- or intra-cellular space, or it may represent one specific state of the tracer, for instance metabolised state. A compartment, i , is solely characterized by its time-dependent, homogeneous tracer concentration, $C_i(t)$. Consequently, any time-varying concentration in a compartment is caused by tracer-transport into or out from the compartment. The *rate constants* quantify the rates at which tracer transport between compartments take place. They are denoted with the letter k and have the unit of inverse time, $[\text{min}^{-1}]$ ¹ [6, 99].

Compartment models may have an arbitrary number of compartments, although the model equations can often be derived from simple, first-order differential equations. In the following, the compartment models valid for ^{15}O -water and ^{18}F -FDG will be described, which are relevant for **Paper IV** and **Paper III**, respectively.

¹The unit of the rate constant K_1 is $\left[\frac{\text{ml}(\text{blood})}{\text{min}} \frac{1}{\text{ml}(\text{tissue})} \right]$, representing the blood volume whose tracer content is completely extracted in one minute, and transferred to one ml of tissue. Cancellation of the volume units is formally not done, and subsequently, this rate constant is denoted with a capital letter [4].

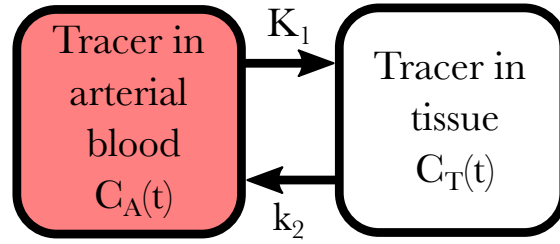


Figure 2.16: The one-tissue compartment model.

2.5.2 One-tissue compartment model

The simplest form of a compartment model is the *one-tissue compartment model*, illustrated in Figure 2.16. This model is commonly used for tracers that can diffuse freely between vascular and tissue space, such as ^{15}O -water.

The tracer diffusion between arterial blood and tissue can be described by the following state equation [4, 6, 99, 100]:

$$\frac{dC_T(t)}{dt} = K_1 \cdot C_A(t) - k_2 \cdot C_T(t), \quad (2.19)$$

where $C_T(t)$ is the time-activity curve in tissue and $C_A(t)$ is the time-activity curve in arterial blood, also known as the AIF. The solution to Equation 2.19 is given by [101]:

$$C_T(t) = K_1 \cdot C_A(t) \otimes e^{-k_2 \cdot t}, \quad (2.20)$$

where \otimes denotes mathematical convolution. As mentioned in Section 2.1.13, due to the tissue-fraction effect, the PET signal is composed of both tissue and blood components. Therefore, the radioactivity concentration measured with PET, $C_{PET}(t)$, is modeled as the sum of the fractional tissue and blood components, such that:

$$C_{PET}(t) = (1 - V_A) \cdot C_T(t) + V_A \cdot C_A(t). \quad (2.21)$$

Here V_A is the fractional arterial blood volume in the tissue. In addition to the measured PET signal, also the AIF must be measured. The challenges with AIF measurements are discussed in Section 2.5.4. For the ^{15}O -water tracer, CBF can be closely approximated by K_1 [101].

2 Medical imaging with PET, CT and MRI



Figure 2.17: The uptake pathways of glucose and fluorodeoxyglucose (FDG) from blood into tissue. Both glucose and FDG are phosphorylated into glucose-6-phosphate and FDG-6-phosphate, respectively. Glucose is metabolised further by glucolysis, while FDG-6-phosphate is not, and consequently is trapped in the cell.

Glucose and FDG uptake

The uptake pathways of FDG is closely related to the uptake pattern of glucose (Figure 2.17). Both FDG and glucose are taken up into the cells by glucose transporters and then phosphorylated (metabolised) into respectively glucose-6-phosphate and FDG-6-phosphate by the enzyme hexokinase. FDG-6-phosphate is however, as opposed to glucose-6-phosphate not metabolized further and can therefore, as opposed to normal glucose, not exit the cell. In many tumor cells, dephosphorylation is also down-regulated, and consequently, FDG-6-phosphate is trapped in the cell. This results in accumulation of tracer in tumor which will therefore have high contrast in the resulting PET images [4].

Two-tissue compartment model

The compartment model commonly used for ^{18}F -FDG is the *irreversible two-tissue compartment model*, illustrated in Figure 2.18. This model assumes ^{18}F -FDG to be either in a free state, or bound in the tissue cells, with activity concentrations C_1 and C_2 , respectively. The two state equations are [4, 6, 99, 100]:

$$\frac{dC_1(t)}{dt} = K_1 \cdot C_A(t) - (k_2 + k_3) \cdot C_1(t) \quad (2.22)$$

$$\frac{dC_2(t)}{dt} = k_3 \cdot C_1(t) \quad (2.23)$$

As previously, $C_A(t)$ is the AIF.

PET cannot differentiate the two signals $C_1(t)$ and $C_2(t)$, and consequently only the total radioactivity concentration in tissue, $C_T(t)$, can be measured:

$$C_T(t) = C_1(t) + C_2(t). \quad (2.24)$$

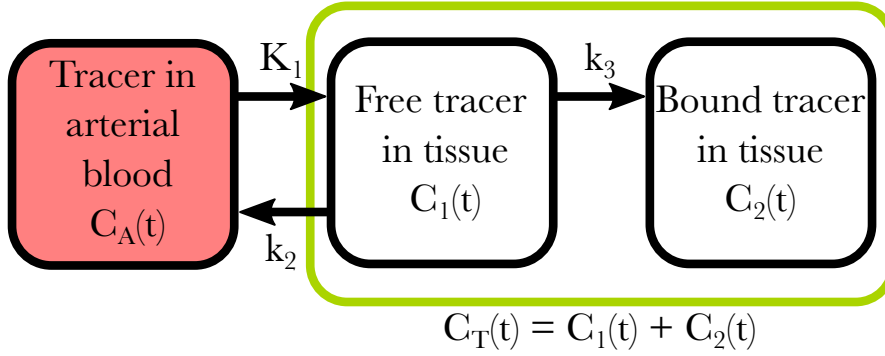


Figure 2.18: The irreversible two-tissue compartment model. The green box indicate the tissue concentration measured with PET.

$C_T(t)$ is indicated by the green box in Figure 2.18.

By using Equations 2.22-2.24, $C_T(t)$ can be expressed as [102]:

$$C_T(t) = \frac{K_1}{k_2 + k_3} [k_3 + k_2 \cdot e^{-(k_2+k_3) \cdot t}] \otimes C_A(t). \quad (2.25)$$

As for the one-tissue compartment model, the tissue-fraction effect has to be taken into account, according to Equation 2.21.

At late time-points, as $t \rightarrow \infty$, the exponential term gets sufficiently small. The tissue activity is then given by:

$$C_t(t) = \frac{K_1 \cdot k_3}{k_2 + k_3} \cdot C_A(t) \quad (2.26)$$

Here, the *net-influx rate constant*, K_i , is defined as [4]:

$$K_i \equiv \frac{K_1 \cdot k_3}{k_2 + k_3} \quad (2.27)$$

Subsequently, the metabolic rate of glucose, MR_{glu} , can be estimated from K_i as [103]:

$$MR_{glu} = K_i \cdot \frac{C_{glu}}{LC}. \quad (2.28)$$

Here, C_{glu} is the glucose concentration in blood, and the lumped constant (LC) accounts for the difference in metabolic rate of glucose, to that of ^{18}F -FDG. C_{glu} should be measured just before PET scanning with an ordinary glucose meter,

2 Medical imaging with PET, CT and MRI

while the LC can be found in the literature. For instance, the LC in normal brain tissue has been measured to 0.9 in humans and to 0.7 in rats [104, 105].

2.5.3 Graphical analysis with Patlak plot

In most practical situations, only K_i is of interest, and the individual rate constants, $K_1 - k_3$ are rarely used. For these cases, the analysis of a full compartment model may be simplified by using graphical methods, such as the *Patlak plot*, in which the data are normalized to follow a linear model at late time points.

The Patlak plot is valid for any underlying compartment model, as long as there is one final compartment where the tracer is trapped, such as the irreversible ^{18}F -FDG model (Figure 2.18). After sufficient time, the tracer concentration in blood will be in steady-state with the irreversible compartment, and consequently, any change in the total tissue concentration normalized to the blood concentration, represents a change in the irreversible compartment. The normalized tissue-to-blood concentration is then described by a linear curve, which can be derived from Equation 2.25 [106]:

$$\frac{C_T(t)}{C_A(t)} = V_0 + K_i \left(\frac{\int_0^t C_A(\tau) d\tau}{C_A(t)} \right) \quad (2.29)$$

Here, V_0 is the initial *volume of distribution*, that is, the initial value of the tissue-to-blood concentration ratio. If the $C_T(t)/C_A(t)$ ratio is plotted versus the normalized time $\left(\frac{\int_0^t C_A(\tau) d\tau}{C_A(t)} \right)$, V_0 and K_i can be estimated from linear regression at late time points [106].

2.5.4 Input function

As mentioned in Section 2.5.2, the time-dependent AIF, $C_A(t)$, must be known in order to estimate the rate constants from Equations 2.20 and 2.25. Several different approaches to obtain the AIF will be discussed next.

Arterial blood sampling

The gold standard method to measure the AIF is through arterial blood sampling. Arterial sampling must start simultaneously with tracer injection and PET imaging, and must therefore necessarily be performed while the subject lies in the PET scanner. An artery is cannulated and arterial blood is sampled throughout the entire imaging session by a continuous arterial line connected to a scintillation-detector in the PET scanning room. Such, so-called *in-line* blood concentration measurements may have a sampling frequency of, for instance, 1 Hz. One, or a few late manual blood samples should be extracted from the continuous line and measured in a gamma counter cross-calibrated to the PET scanner, so that the AIF curve can be calibrated and used together with the image-derived PET-data [40].

Arterial blood sampling is feasible in human PET imaging by sampling from the radial artery [107–109]. It is also possible in pre-clinical studies of rats [110–112] and mice [111, 113, 114] by cannulating the femoral, tail, or carotid artery [110, 115, 116]. However, for mouse-PET studies, blood sampling is far more challenging compared to humans or rats, due to the requirement of invasive surgery for arterial cannulation [110, 116], as well as the limited blood volume that can be withdrawn without altering the animal physiology [117].

The AIF at the sampling site (such as the wrist) will differ from a local AIF at the site of interest (such as the brain), because of dispersion in the blood vessels and measurement tubes (in case of in-line sampling) [48]. Furthermore, due to the finite withdrawal speed during continuous arterial blood sampling, there will be a delay between the true and measured AIFs [35]. Appropriate corrections for dispersion and delay must therefore be applied (see Section 2.5.5) [118].

Some tracers, for instance ^{18}F -FLT, are metabolized in the body, resulting in radio-labelled metabolites circulating in the blood [40]. Thus, for such tracers, appropriate metabolite-correction must be applied to the measured AIF data before compartment modeling. Also, some tracers are taken up by the red blood-cells. This fraction is not available for tissue-uptake, and must also be subtracted from the AIF before compartment modeling. The metabolite and plasma fractions are typically assessed by separating plasma, red blood cells and metabolites in the manually drawn blood samples collected during the PET scan, and measuring each fraction separately in a cross-calibrated gamma counter. Time dependence of the plasma and metabolite fractions must be taken into account by collecting multiple samples throughout the duration of the PET scan and applying corrections for all time points [116, 119, 120].

2 Medical imaging with PET, CT and MRI

Population-based input function

Due to the invasive, and time-consuming procedures required for arterial blood sampling, alternative methods may be employed to estimate the AIF. Population-based input function is one example of such method. For a specific tracer, injection protocol and subject population, it is possible to calculate a population-based average AIF from arterial blood sampling of a few subjects. This enables estimation of an AIF for a new subject, by scaling the population-based input function with the radioactivity concentration measured in one, or few late venous blood samples. Although, invasive arterial blood sampling can be avoided, obviously, this method neglects individual physiological differences as well as scan-dependent variations, and still, it requires the sampling and measurement of at least one blood sample for curve scaling [40, 121, 122].

Image-derived input function

Another approach to overcome the limitations with arterial blood sampling is the use of an image-derived input function (IDIF) [40]. An IDIF can be measured inside a suitable arterial blood pool directly in the reconstructed PET images. Depending on which body part is inside the field-of-view of the dynamic PET sequence, different blood pools may be available. For instance, in human PET imaging of the thorax, the left ventricle or ascending or descending aorta may be used, while in the abdomen, the abdominal aorta is an optional blood pool [123]. In brain imaging, the carotid arteries or the venous sinuses can be employed for the IDIF [124–126]. In pre-clinical PET imaging of rodents, if the heart is inside the field-of-view, left ventricle can be used, while to avoid spill-over from the surrounding myocardium, vena cava is a suitable alternative [127].

Due to the distance between the IDIF sampling site and the tissue of interest, correction for tracer dispersion must be applied also with image-derived methods [127]. Furthermore, due to the limited spatial and temporal resolution of the PET imaging system, image-derived methods suffer from both partial volume effects and increased image noise. These limitations require standardized methods for partial volume correction and artery delineation, which may be difficult to achieve in practice. Consequently, image-derived methods are not very common in clinical or research studies [40, 128].

Simultaneous estimation

Simultaneous estimation is a method which does not require an AIF for estimating the kinetic parameters. Instead, it is applied to image data to estimate both the AIF and kinetic parameters simultaneously [129–131]. One drawback of this method is that it assumes a known mathematical AIF model and requires at least one late blood sample for parameter estimation. Recently, non-invasive simultaneous estimation methods were developed that obviate the need for the single late blood sample. Instead, additional input variables from electronic health records were used as input into the models [132, 133]. The limitation of this method is that a large set of clinical variables must be collected and handled for each patient. These may not necessarily be available in the health records for all patients and healthy volunteers in research studies.

2.5.5 Arterial input function corrections

As mentioned in Section 2.5.4, a measured AIF must be corrected for delay and dispersion effects, which will be described next.

Delay correction

Delay correction is required to account for the systematic time-delay between the tracer arrival times at the site of blood sampling (e.g. the wrist) compared to the tissue of interest (such as the brain). For instance, for ^{15}O -water brain studies, Equation 2.20 can be modified to include a time-delay term, Δt , as follows [134]:

$$C_T(t) = K_1 \cdot C_A(t + \Delta t) \otimes e^{-k_2 \cdot t}. \quad (2.30)$$

With this modified model equation, delay correction may be included as a free parameter during model fitting. It can also be estimated separately by shifting the arterial blood curve to match the slope of the total PET counts [135], or on a voxel-to-voxel basis, to account for delay-time heterogeneity in the tissue [136]. Delay correction should be estimated before, or simultaneously with dispersion correction of the AIF [118].

2 Medical imaging with PET, CT and MRI

Dispersion correction

If the arterial blood signal, $C_A(t)$, in Equation 2.20 and 2.25 originates from external blood sample measurements (either discrete or in-line measurements), the measured blood signal, $g(t)$, will be affected by dispersion in the blood-vessels and in the detector system tubes. This can be modeled as a convolution of the true AIF, $C_A(t)$, and a dispersion function, $d(t)$ as [48]:

$$g(t) = C_A(t) \otimes d(t). \quad (2.31)$$

A mono-exponential dispersion model may be assumed [48]:

$$d(t) = \frac{1}{\tau} e^{-\frac{t}{\tau}}, \quad (2.32)$$

where τ is the dispersion constant. An expression for the true AIF, $C_A(t)$, can be obtained by the Laplace transform [118]:

$$C_A(t) = g(t) + \tau \frac{dg}{dt}. \quad (2.33)$$

The dispersion constant may either be fixed to a specific value, such as 15 seconds as recommended in [48], or it can be included as a free parameter during model fitting [118]. For instance, for ^{15}O -water brain studies, Equation 2.30 can be modified to include the dispersion constant, τ as follows [118]:

$$C_T(t) = \tau K_1 \cdot g(t + \Delta t) + (1 - \tau k_2) K_1 \cdot g(t + \Delta t) \otimes e^{-k_2 \cdot t}, \quad (2.34)$$

where $g(t + \Delta t)$ is the uncorrected measured arterial blood curve from Equation 2.31, including both delay and dispersion errors.

Figure 2.19 shows an example of a measured and dispersion-corrected AIF using Equation 2.33. Evidently, this correction amplifies the noise as a consequence of the gradient-term in Equation 2.33.

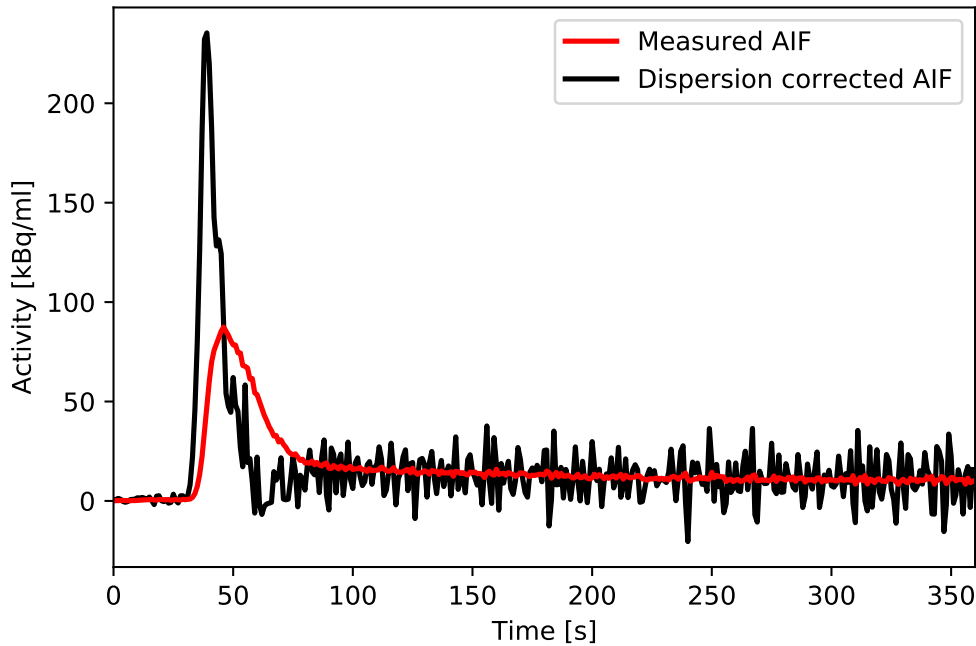


Figure 2.19: Example of an AIF from a subject injected with ^{15}O -water. The red curve is the measured AIF using in-line continuous arterial blood sampling. The black curve is the dispersion-corrected AIF using Equation 2.33 with $\tau = 15$ s.

2.5.6 Parameter estimation

The aim with kinetic modeling is to estimate the rate constants $K_1 - k_4$ from Equations 2.20, 2.25, 2.30 or 2.34, implicitly defined through Equation 2.21. This is commonly done using *non-linear least square fitting*. This must be done in an iterative manner, by starting with an initial guess of the parameters and comparing the measured and estimated tissue-curves, $C_{PET}(t)$ and $C_{Model}(t)$. The objective is then to find the parameters that minimize the sum of squared differences between the curves [137].

2.5.7 Quantification error in PET

As outlined in the preceding sections, there are many sources of error in the imaging chain that affects the final result of a PET-based quantitative measurement. The quantitative error in a PET measurement may be characterized by *systematic* and *random* errors. Systematic errors, or *biases*, cause the mean value over repeated

2 Medical imaging with PET, CT and MRI

PET measurements under identical conditions to deviate from the true value, thus being a measure of *accuracy*. On the other hand, random errors affects the variance of repeated measurements, and thus affecting the *precision* of the measurement.

2.6 PET applications in clinic and research

In the current thesis, clinical PET applications within lung cancer are studied in **Paper I** and **Paper II** while **Paper IV** focuses on brain perfusion. **Paper III** is based on a pre-clinical mouse PET study. These applications are described in the following sections.

2.6.1 Clinical PET applications

The most widespread application of clinical PET is within oncology. Many diseases introduce alterations in the glucose metabolism which can be measured with the glucose analogue tracer ^{18}F -FDG. The major clinical application of this tracer is therefore detection, staging and assessing treatment response of cancer, for which PET is an important tool in the patient management decision chain [4]. Common applications within oncology include lung, esophageal, colorectal, head and neck, and breast cancer, as well as melanoma and lymphoma [3]. Clinical PET is also commonly used in neurology, to diagnose Alzheimer's disease, Parkinson's disease, epilepsy or psychiatric disorders. Furthermore, cardiovascular applications of PET involve the evaluation of myocardial viability and to diagnose patients with coronary artery disease [6]. ^{18}F -FDG-PET may also be used to diagnose infection and inflammation, by detecting granulocytes and macrophages with increased glucose metabolism [3].

2.6.2 Lung cancer

Prevalance

In 2020, there were more than 19 million new cancer cases in the world, with around 10 million deaths from the disease. Of all cancers, lung cancer is, following breast cancer, the second most frequent cancer type (11% of all cases) and the world-wide leading cause of cancer-related death (18% of all deaths) [14]. The most common causal risk factors for lung cancer are tobacco smoking, radon and air pollution, with smoking being the leading cause, with 965 500 lung cancer deaths world-wide

Table 2.4: TNM classification according to the 8th edition of lung cancer stage classification. The class definitions serve here for illustrative purposes only, and have been somewhat simplified. Please refer to reference [140] for the complete definitions.

T (Primary tumor)	
T0	No evidence of primary tumour
T1	Tumour ≤ 3 cm
T1a	Tumour ≤ 1 cm
T1b	1 cm < Tumour ≤ 2 cm
T1c	2 cm < Tumour ≤ 3 cm
T2	3 cm < Tumour ≤ 5 cm
T2a	3 cm < Tumour ≤ 4 cm
T2b	4 cm < Tumour ≤ 5 cm
T3	5 cm < Tumour ≤ 7 cm
T4	Tumour > 7 cm
N (Regional Lymph Nodes)	
N0	No regional lymph node metastasis
N1	Metastasis in ipsilateral pulmonary or hilar nodes
N2	Metastasis in ipsilateral mediastinal/subcarinal nodes
N3	Metastasis in contralateral mediastinal/hilar, or supraclavicular nodes
M (Distant Metastasis)	
M0	No distant metastasis
M1	Distant metastasis
M1a	Separate tumour nodule(s) in a contralateral lobe; tumour with pleural/pericardial nodules or malignant pleural/pericardial effusion
M1b	Single extrathoracic metastasis in a single organ
M1c	Multiple extrathoracic metastases in one or several organs

in 2010. A higher risk is associated with increased duration of smoking, and with increased number of cigarettes smoked daily [138].

Histological subtypes

Lung cancer is traditionally classified into two main types: small cell carcinoma and non-small cell carcinoma, with prevalances of around 10-15%, and 85-90%, respectively. Non-small cell carcinoma is further divided into three main histological types: adenocarcinoma, squamous cell carcinoma and large cell carcinoma. Adenocarcinoma and squamous cell carcinoma are the predominant types, with the incidence of squamous cell carcinoma falling and adenocarcinoma rising and surpassing that of squamous cell carcinomas [139].

Classification and staging

Lung cancer tumors are staged according to standard TNM categories, currently following the 8th edition of lung cancer stage classification, which is the world-wide standard as of January 1, 2017 [140]. Each tumor is described according to three components: T for the physical extent of the primary tumor, N for regional lymph nodes, and M for distant metastases. The current TNM classification categories are shown in Table 2.4. From specific combinations of T, N and M, the tumor is staged into one of four stage groups: 0, I, II, III or IV, according to Table 2.5 [140].

2 Medical imaging with PET, CT and MRI

Table 2.5: Lung cancer stage grouping. Adopted from [138].

T	N	M	Stage
T1a	N0	M0	IA1
T1b	N0	M0	IA2
T1c	N0	M0	IA3
T2a	N0	M0	IB
T2b	N0	M0	IIA
T1a	N1	M0	IIB
T1b	N1	M0	
T1c	N1	M0	
T2a	N1	M0	
T2b	N1	M0	
T3	N0	M0	
T1a	N2	M0	IIIA
T1b	N2	M0	
T1c	N2	M0	
T2a	N2	M0	
T2b	N2	M0	
T3	N1	M0	
T4	N0	M0	
T4	N1	M0	
T1a	N3	M0	IIIB
T1b	N3	M0	
T1c	N3	M0	
T2a	N3	M0	
T2b	N3	M0	
T3	N2	M0	
T4	N2	M0	
T3	N3	M0	IIIC
T4	N3	M0	
Any T	Any N	M1a	IVA
Any T	Any N	M1b	IVA
Any T	Any N	M1c	IVB

The role of PET/CT for lung cancer diagnosis

PET/CT imaging plays an important role in the diagnosis of lung cancer, especially for small pulmonary nodules less than 3 cm in diameter. In these small-volume cancers, metabolic changes can be detected with ^{18}F -FDG PET with much greater sensitivity compared to the anatomical changes visible in chest radiography or CT. False-negative results may be obtained for small nodules less than 8–10 mm in diameter or with low-grade malignancies. Also for larger tumors, with SUV greater than 2.5, PET/CT may increase the likelihood of malignancy detection, with worse prognosis with increasing SUV. Furthermore, if the PET/CT examination reveals hilar or mediastinal tumor infiltration, or mediastinal lymph node involvement, surgery may be avoided. Likewise, if PET/CT shows non-infiltrating tumors, surgery is associated with increased prognosis [141].

In the clinical workflow, PET/CT is an integral part in the staging of lung cancer. For T-staging, the tumor size is most often adequately determined with CT. However, in some cases, where lung atelectasis or obstructive pneumonia is present, or when the primary tumor has infiltrated the pleura or mediastinum, PET is more sensitive compared to CT for T-staging [142]. PET/CT is also the single-most sensitive imaging modality for detecting hilar and mediastinal lymph node infiltration, which is important for N-staging. In addition, PET/CT is superior compared to other imaging modalities in detecting local, regional and distant

2.6 PET applications in clinic and research

metastases in, for instance, bone, liver or adrenal glands, which is critical information for M-staging [142]. For the detection of brain-metastases however, MRI is superior to ^{18}F -FDG PET/CT, due to the high background activity of ^{18}F -FDG in the brain [141]. Nevertheless, PET/CT may be used to guide a biopsy to viable (non-hypoxic and non-necrotic) regions of the tumor, or alternatively, towards a mediastinal lymph node instead of the primary tumor. In this way, PET/CT-guided biopsy may provide additional histological information, allowing for more accurate staging than possible without PET/CT [142]. With the added staging information obtained from PET/CT, patient management is changed in up to 40% of the cases, compared to pre-PET diagnosis [141].

PET/CT is also important for radiation therapy treatment planning. For instance, the detection of previously unknown distant metastasis may change the planned outcome of the therapy from curative to palliative treatment. Also, PET/CT is commonly used for the delineation of the lung cancer disease, to allow radiation therapy treatment planning [141].

Lastly, PET/CT may be useful for evaluating treatment response, during or after completed radiation-, chemo-, or combined radio-chemo-therapy. After completed treatment, PET/CT has shown to be highly sensitive in detecting residual or recurrent lung cancer disease [141].

The role of PET/MRI for lung cancer diagnosis

PET/MRI has gained interest in the diagnosis of several cancers, including lung cancer, with the potential advantage to combine the excellent soft-tissue contrast of MRI with the functional information from PET. However, it is difficult to obtain high quality diagnostic MRI of the lungs due to the low density of protons in lung tissue. Cardiac- and respiratory motion further complicates the acquisition of lengthy MRI scans of the thorax [19]. Nevertheless, a recent review found that PET/MRI is a robust alternative to PET/CT for T- and N-staging. A few distant metastases were found in liver and brain on PET/MRI, which were overlooked on PET/CT, but there were not enough data to provide any general statistical evidence for which modality was best for M-staging. PET/CT was more sensitive in detection of smaller lung nodules (<5 mm), compared to PET/MRI. However, the differences found between PET/CT and PET/MRI rarely changed patient management [143].

Artifacts may frequently occur in the MRI-based attenuation correction maps. These may introduce inconsistent SUV measurements in serial scans, which might affect the reliability of therapy assessment studies [13].

2 Medical imaging with PET, CT and MRI

Altogether, there is currently no evidence of PET/MRI being a candidate to replace PET/CT in the diagnosis and staging of lung cancer patients [143]. Nevertheless, one potential advantage with PET/MRI is simultaneous motion correction using MRI-derived motion vector fields (See section 2.4.6). This may reduce motion-induced blurring of the images, and provide increased sensitivity for small lesions [144]. Future research is needed to provide evidence for this hypothesis. Another advantage is that MRI does not contribute to the radiation dose to the patient.

2.6.3 Brain perfusion

Non-invasive quantification of CBF allows to identify patients with different neurological disorders, such as chronic internal carotid artery occlusion, brain tumors and epilepsy. There exists several imaging modalities to measure CBF, for instance using single photon emission computed tomography, CT, MRI or ultrasound [34]. However, PET imaging, using radio-labeled water (^{15}O -water), is considered the reference standard, because it provides accurate three-dimensional quantitative CBF measurements on a voxel-level in the whole brain [34–36]. The tracer is administered as an intravenous bolus injection, starting simultaneously with the start of a dynamic PET sequence. By simultaneous sampling the AIF, and by the assumption of the freely-diffusive tracer model, CBF is obtained for each brain region from kinetic modeling (see Section 2.5.2).

The acetazolamide challenge is commonly used to assess critically reduced perfusion by paired CBF measurements (baseline and post-acetazolamide). Acetazolamide medication dilates the vascular system by increasing carbon dioxide levels in the blood-stream, and thereby increasing the cerebral arterial blood flow velocity [145]. Vasodilation occurs over 3-4 minutes, with maximal effect at 10-15 minutes post-acetazolamide injection. In healthy patients, CBF increases by more than 30% in the post-acetazolamide scan, compared to baseline [146, 147]. On the other hand, in patients with chronic internal carotid artery occlusion, the vessels are already dilated at baseline. This occurs due to autoregulatory mechanisms that compensate for the occluded vessels to retain the perfusion in the brain. The already dilated vessels are unable to further expand, and thus, under the acetazolamide condition, perfusion is equal, or less, compared to baseline [145, 146]. The rapid radioactive decay of ^{15}O (2 minute half-life, Table 2.2) allows paired PET measurements to be performed in a sequence, without patient re-positioning between the baseline and acetazolamide scan.

2.6.4 Pre-clinical imaging

Many pre-clinical PET applications has emerged from human, clinical PET imaging, by simply translating protocols and tracers from the hospital clinic. In this way, medications and interventions can be studied in mice or rats using similar PET protocols and tracers as for humans. One benefit of animal models is that repetitive PET scans can be performed over time, to study the temporal evolution of diseases. For instance, by implanting tumor cells under the skin of an animal, and allowing the tumor to grow *in vivo*, in so-called *xenograft* models, the tumors can be imaged with PET over time. After the final imaging session, the tumor can be removed from the animal by surgery and investigated with other, *ex-vivo* high-resolution imaging techniques, such as microscopy or autoradiography, to validate the lower-resolution PET images. In this way, novel therapeutic drugs, or other cancer-treatments can be evaluated in animal models prior to clinical studies in humans [4].

There is a fundamental size-difference between humans and mice, with an underlying weight-ratio in the order of 3500 (70kg/0.02kg). Thus, the technological demands on pre-clinical PET systems, in terms of spatial resolution and sensitivity are much higher, compared to clinical PET. As discussed in Section 2.1.13, the theoretical lower bound of resolution is around 1.8 mm and 0.7 mm for clinical and pre-clinical PET systems, respectively. In terms of volumetric voxel resolution, this corresponds to a ratio of $1.8^3/0.7^3 = 17$. Thus, even with dedicated small-animal PET scanners, the spatial resolution is inferior compared to human clinical systems [148].

2.6.5 Tracer development

Although the majority of clinical PET applications are performed with ^{18}F -FDG, this tracer also has limitations. For instance, it is non-specific, meaning that it is taken up in all living cells, and not only in cancer cells. This may be a drawback in, for instance, brain imaging, where the high natural background uptake of ^{18}F -FDG in normal brain tissue may mask possible tumor uptake. Furthermore, certain tumors have low ^{18}F -FDG-uptake, making them hard to visualize with PET. Thus, there is a demand for novel, more specific tracers to be used for PET applications.

The development of a new tracer is a complex process involving many steps. One of the roles of pre-clinical imaging in tracer developments is to test early promising tracers in animal models that mimic a specific human disease [37]. One obvious benefit of imaging an unknown tracer in mice is that whole-body dynamic imaging is possible. Thus, it is possible to map the *in-vivo* whole-body biodistribution using PET, although such measurements must also be validated using *ex-vivo* biodistribution. Nevertheless, biological differences between humans and rodents

2 Medical imaging with PET, CT and MRI

may introduce variations of the applicability of a certain tracer in humans, compared to rodents. Results from pre-clinical studies must therefore be used with caution when extrapolated to humans [37].

3 Machine learning for medical imaging

Machine learning is a field of research where a computer model can learn patterns in data in order to perform a given task, such as classification or regression. The trained model can then be applied to new, unseen data. Machine learning is a sub-field of the wider concept of artificial intelligence (AI), which refers more broadly to technology that can simulate human intelligence. Thus, an AI technology may consist of many specific machine learning models in order to accomplish a given task without human instruction or interaction.

One potential application of machine learning in the field of medical imaging is to replace invasive biopsy. For instance, a database of tumor images may be used to train a classification model to separate two tumor classes, based on PET images. When a new patient arrives to the hospital and undergoes PET imaging, using the same protocol as for the images in the training database, the trained model can be used to classify the tumor into one of the two tumor classes. Thus, the patient may avoid invasive biopsy with accompanied risk for complications [26].

In **Paper I** and **Paper II**, the impact of imaging artifact and patient motion on tumor-based image features are studied. By proper pre-processing of medical images, and selection of motion-robust features, robust predictive models can be built in future studies.

In **Paper III** and **Paper IV**, machine learning is used to predict the AIF in dynamic PET studies. This may simplify future pre-clinical and clinical dynamic PET imaging by avoiding the need for invasive blood samples.

3.1 Basic concepts and terminology

Figure 3.1 illustrates the different sub-fields of machine learning. The topics relevant for this thesis are indicated by the green square in the figure. *Supervised learning*, refers to situations where the labels of the example data are known. Two common

3 Machine learning for medical imaging

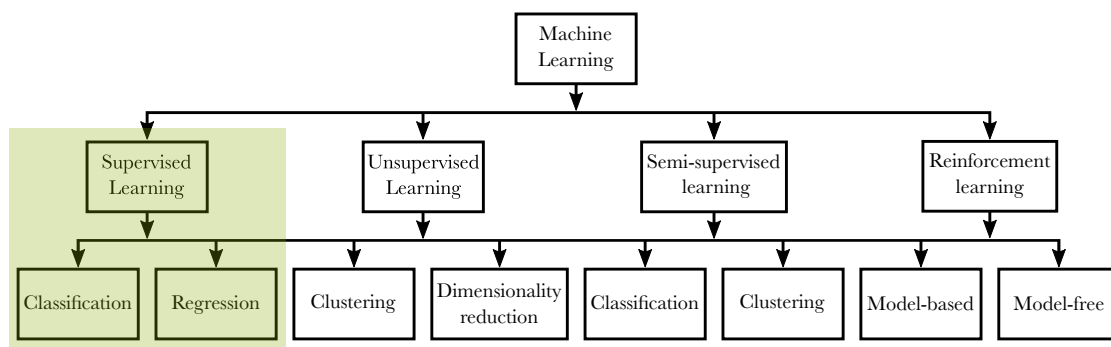


Figure 3.1: Taxonomy of different machine learning approaches. The topics relevant for this these are indicated by the green square.

tasks of supervised learning are *classification* and *regression*. A typical classification task could be to train a model that separates the two classes best, while in regression, the task is to predict a real number [149]. The concepts of supervised techniques will be further described in the sections that follow, with main focus on classification, but may with some minor changes to the models also be applied to regression problems.

3.1.1 Features

A machine learning model is commonly trained on some measurable properties of the data, so-called *features*. A feature should capture characteristic properties of the data, that allow the model to learn the given task. Multiple features, x_1, x_2, \dots, x_p , may be extracted from each data sample, to form a *feature vector*, \mathbf{x} [149]:

$$\mathbf{x} = [x_1, x_2, \dots, x_p]^T. \quad (3.1)$$

Figure 3.2 illustrates a diagram of two features, x_1 and x_2 extracted from a hypothetical dataset with known labels. Once the model has been trained, a new data point may be classified using the model. The classification result may be different depending on the chosen model, as illustrated in the figure. Returning to the example in the introduction of this chapter, the two extracted features could be, for instance, average SUV and volume of a segmented tumor in a PET image, and the classification task could be to separate two histological subtypes of tumors. In practice a higher-order feature set is often used, possibly with several hundreds of extracted features from each segmented tumor (See section 3.4).

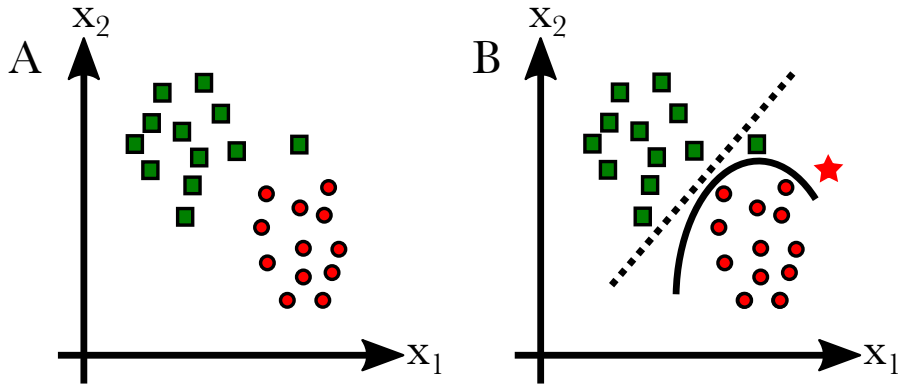


Figure 3.2: Illustration of classification of a labelled dataset with two features, x_1 and x_2 . A. Dataset with labels indicated by green squares and red circles. B. Classification result by linear (dashed line) and non-linear model (solid line). One data point from the square class is falsely classified to the circle class by the linear model. A new data point, illustrated with a star, is classified to the circle class by the linear model and to the square class by the non-linear model.

3.1.2 Discriminant functions

The line and curve that separates the two classes in Figure 3.2 are generally referred to as *discriminant functions*, or equivalently for linear models, *decision hyperplanes*, denoted by $f(\mathbf{x})$. The aim of a typical classification task is to find the discriminant function, that separates the classes. Thus, $f(\mathbf{x})$ takes a p -dimensional feature value, \mathbf{x} , as input, and outputs the class-relationship, for instance $+1$ or -1 , for class w_1 and w_2 , respectively. The mapping function can be either *linear* or *non-linear*. For instance, a linear mapping function has the form of [149]:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b, \quad (3.2)$$

where $\mathbf{w} = [w_1, w_2, \dots, w_p]$ is a weight vector and b is the bias term.

For a two-class problem, the discriminant function is given by $f(\mathbf{x}) = 0$, such that the output of the mapping function for two classes, w_1 and w_2 , is given by [149]:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b > 0 \quad \forall \mathbf{x} \in w_1 \quad (3.3)$$

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b < 0 \quad \forall \mathbf{x} \in w_2 \quad (3.4)$$

The problem to solve is how to find a suitable set of weights and bias that correctly classifies all training data points into the two classes w_1 and w_2 . One algorithm to accomplish this is the *perceptron algorithm*, which will be described next.

3 Machine learning for medical imaging

3.1.3 The perceptron

The perceptron algorithm, is a method to find the weights of a linear classifier [150]. First, the perceptron *loss function* is defined as [149]:

$$J(\mathbf{w}, b) = \sum_{x \in Y} \delta_x (\mathbf{w}^T \mathbf{x} + b), \quad (3.5)$$

where Y is the subset of feature vectors belonging to training samples that are misclassified by the hyperplane $\mathbf{w}^T \mathbf{x} + b$. The parameter $\delta_x = -1$ if $\mathbf{x} \in w_1$ and $\delta_x = +1$ if $\mathbf{x} \in w_2$. In this way, as long as there are misclassified points according to the current hyperplane, the perceptron loss will be larger than zero. For instance, if $\mathbf{x} \in w_1$ is misclassified, then $\mathbf{w}^T \mathbf{x} < 0$ (according to Equation 3.3) and $\delta_x < 0$, thus $J(\mathbf{w}, b) > 0$ (and vice versa for a misclassified $\mathbf{x} \in w_2$). When all points have been correctly classified, $J(\mathbf{w}, b) = 0$, because Y will be an empty set [149].

With the loss function specified, it is possible to define a weight-update regime, for instance by the iterative *stochastic gradient descent* method, such that [149]:

$$\mathbf{w}(s+1) = \mathbf{w}(s) - \rho_s \left. \frac{\partial J(\mathbf{w})}{\partial \mathbf{w}} \right|_{\mathbf{w}=\mathbf{w}(s)}. \quad (3.6)$$

Here, $\mathbf{w}(s)$ are the estimated weights at iteration step s , and ρ_s is a positive number called *learning rate*. Intuitively, Equation 3.6 may be understood as subtracting a positive number from the current weight estimate, in the direction of decreasing loss with respect to the weights. The concept of stochastic gradient descent is illustrated in Figure 3.3. Substituting Equation 3.5 into 3.6 gives [149]:

$$\mathbf{w}(s+1) = \mathbf{w}(s) - \rho_s \sum_{x \in Y} \delta_x \mathbf{x}, \quad (3.7)$$

which concludes the perceptron algorithm. By selecting an initial estimate of the weights, $\mathbf{w}(0)$, and an appropriate learning rate, ρ_s (which may vary with iteration number, s), it can be shown that the perceptron algorithm converges in a finite number of iterations [149].

Once the perceptron algorithm has converged, that is, for a given set of training data, there are no more misclassified samples, the learned set of weights and bias, also known as the *trained model*, may be applied to a new, unseen sample. This is done by calculating the weighted sum of the inputs and adding the bias term, according to Equation 3.2, and then applying the classification criteria [149]:

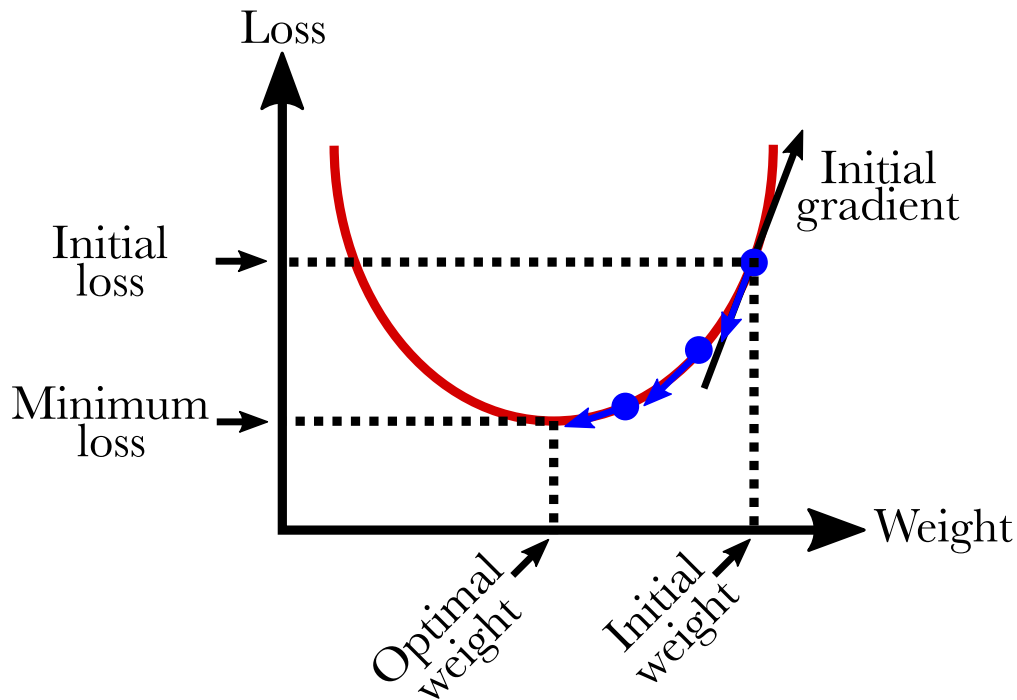


Figure 3.3: Illustration of the stochastic gradient descent algorithm. The red bell-shaped curve illustrates the loss-weight space, that is, for a given set of weights, the loss reaches a global minimum. The weights are initialized and the gradient of the loss function is calculated. The weights are updated along the negative direction of the gradient, towards the minimum value, with a step size equal to the learning rate. Blue arrows indicate intermediate weight updates with decreasing (adaptive) learning rate, to reduce the risk of overshooting the global minimum.

$$\text{If } f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b > 0 \text{ assign } \mathbf{x} \text{ to } w_1 \quad (3.8)$$

$$\text{If } f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b < 0 \text{ assign } \mathbf{x} \text{ to } w_2 \quad (3.9)$$

This procedure of linear classification is illustrated as a basic network unit, or *perceptron* in Figure 3.4. Here the operation of assigning a class label to the output of $f(\mathbf{x})$ by Equations 3.8-3.9 is performed by the so-called *activation function*, $g(z)$. In this case, the activation function is a non-continuous *step function*:

$$g(z) = \begin{cases} 1 & z > 0 \\ -1 & z < 0 \end{cases} \quad (3.10)$$

To guarantee convergence, the perceptron algorithm assumes linearly separable classes. In many physical problems however, this is a non-realistic assumption.

3 Machine learning for medical imaging

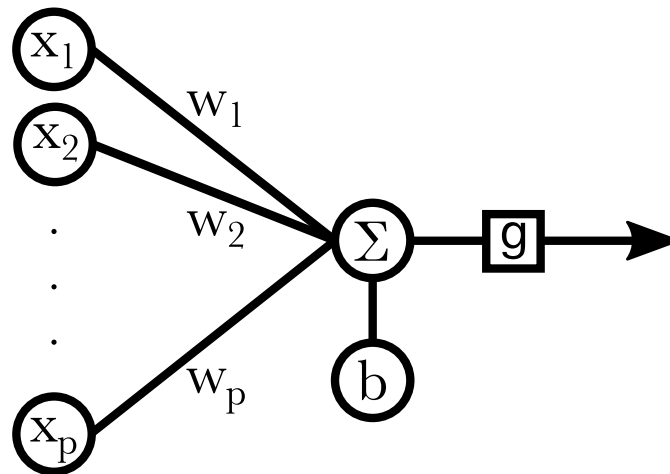


Figure 3.4: Illustration of the perceptron. Each input feature x_1, x_2, \dots, x_p is multiplied by a weight, w_1, w_2, \dots, w_p , summed and added to the bias term, b . The output number then goes through a non-linear activation function, g , to generate the class label for the input, for instance -1 if $\mathbf{w}^T \mathbf{x} + b < 1$.

Therefore, the perceptron is rarely used for any real-life problems. Nevertheless, it is fundamental for understanding more complex neural networks, which will be described next.

3.2 Neural networks

Linear classification algorithms, such as the perceptron, only have a guaranteed solution for linearly separable classes. For non-linearly separable data, linear classifiers may still be used, but by other means of optimizing the decision surface, for instance by the minimizing the mean squared error. Nevertheless, in most situations, non-linear classifiers have superior performance to linear classifiers. As will be shown, neural networks may be formed by combining several perceptron units, to perform, not only non-linear classification, but also regression. The neural network approach to regression is relevant for **Paper III**, where it is used for AIF prediction.

3.2.1 Multilayer perceptron

The multilayer perceptron (MLP), or *feed-forward neural network* is an example of a non-linear classifier, which is build up by combining several single perceptron units, as shown in Figure 3.5.

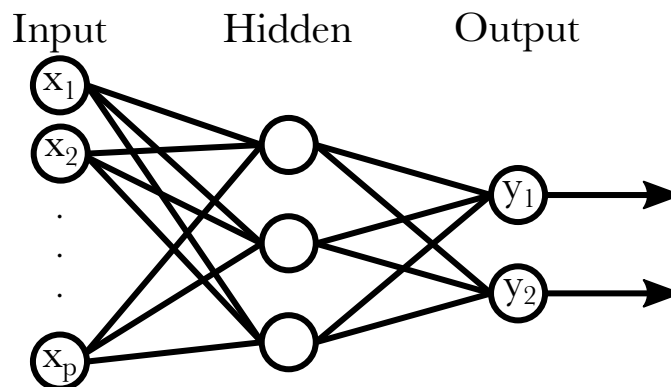


Figure 3.5: Illustration of a multilayer perceptron, or feed-forward neural network. Circles in the diagram represent the *nodes* of the network. The input features x_1, x_2, \dots, x_p are presented to the network at the input layer. The line from each input node represents multiplication by a weight. Each node in the hidden and output layer represents summation of the weighted output from the previous layer, addition of the bias term, and applying the activation function to the result. Note that the bias term and the non-linear activation function at each node have been omitted in the illustration.

As for the perceptron, the input features x_1, x_2, \dots, x_p enter at the input layer. However, there are some other key differences between the MLP and the perceptron. First, an intermediate, so-called *hidden layer*, has been added after the input layer. This layer serves for transformation of the input vector into a new space. Second, several perceptron units, so-called *nodes*, have been added to the hidden layer. Each node may be thought of as one hyperplane that can separate the classes in the intermediate space. In Figure 3.5, the hidden layer has been exemplified with three neurons, thus the p -dimensional input vector is transformed into a three-dimensional intermediate space. Lastly, the output layer, which for the single-unit perceptron was only one number, is here a two-dimensional vector, $\mathbf{y} = [y_1, y_2]$, where $\sum_i y_i = 1$. Classification into two classes is now performed by encoding the labels of the data as $\mathbf{y} = [1, 0]$ for $\mathbf{x} \in w_1$ and $\mathbf{y} = [0, 1]$ for $\mathbf{x} \in w_2$. In this way classification into more than two classes is possible, by simply increasing the number of output nodes [149].

3.2.2 Architectures and activation functions

The *architecture* of a neural network, that is, the specific combination of number of hidden layers and the number of nodes in hidden and output layers, determines the capabilities of the network, and must be pre-defined by the user. For instance, complex classification tasks may be solved by including a large number of hidden layers, in so-called *deep neural networks*. The activation function at each node may

3 Machine learning for medical imaging

also be chosen freely. However, the step-function used for the perceptron has zero gradient for all inputs (except for $z = 0$, where it is undefined), and can not be used in the gradient descent-based training of neural networks (as will be outlined in the next section). Instead, alternative activation functions are commonly used, for instance the sigmoid function or the rectified linear unit (ReLU) inside the network, and the *softmax* function at the output layer [149, 151]:

$$\text{Sigmoid : } g(z) = \frac{1}{1 + e^{-z}} \quad (3.11)$$

$$\text{ReLU : } g(z) = \max(0, z) \quad (3.12)$$

$$\text{Softmax : } g(z)_j = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad (3.13)$$

The activation function, $g(z)$, takes the output, z , from the preceding node, and maps it to a new number. For instance, the sigmoid function in Equation 3.11, outputs a number in the range 0 to 1, while the ReLU in Equation 3.12 has linear output only for positive input numbers. For the softmax function in Equation 3.13, indices j and k refers to output nodes and K is the total number of output nodes (classes). This function forces the output of node i in a K -class problem to the range 0 to 1 and ensures that all output nodes sum to 1. In this way the output at each node can be interpreted as a probability of that class assignment [149, 151].

3.2.3 Training of neural networks

Training of a neural networks refers to the iterative procedure of finding the weights and bias parameters that optimally solves a desired task. When some features, \mathbf{x} , are presented at the input layer of a trained network, the calculated output at node k , \hat{y}_k , should become similar to the desired output of that node, y_k . This is achieved by optimizing a loss function, as introduced for the perceptron in Section 3.1.3. The loss function measures the performance of the neural network for the current set of weights by comparing the network output to the desired output (label). Examples of commonly-used loss functions are the mean squared error and the cross-entropy loss [149]:

$$\text{Mean squared error : } J = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^K (\hat{y}_k - y_k)^2 \quad (3.14)$$

$$\text{Cross entropy : } J = - \sum_{n=1}^N \sum_{k=1}^K y_n^k \log[\hat{y}_n^k]. \quad (3.15)$$

Here, the first summation is over N data points and the second summation is over K output nodes. The mean squared error compares two real numbers, and is therefore ideal to use in regression settings (Section 3.2.4). For K -class classification with softmax activation (Equation 3.13), the cross entropy is preferred, as it sums the log of the probability of class assignment at each output node for each class separately.

Similar to the perceptron example in Section 3.1.3, the task of network training is to minimize the loss function. This is done by the stochastic gradient descent algorithm, as outlined for the perceptron in the preceding section. In short, the weights of the network are first set to some initial values. A data sample is fed through the network, resulting in a computed output at each node. Subsequently, the prediction error is calculated by evaluating the loss function. Based on the chain-rule, the derivative of the loss for the i th weight, $\frac{\partial J}{\partial w_i}$, can be calculated using the so-called *backpropagation algorithm*. The update of each weight, w_i , is then performed according to Equation 3.6. The process is repeated for next sample in the dataset. One so-called *epoch* is completed when all training samples have been fed through the network. Usually, training is repeated for several tens to hundreds of epochs. By monitoring the value of the loss function, training can be stopped when there is no further decrease of the loss function [149].

In practice, performing a weight update for each data sample may result in slow computations, and fluctuations of the weight estimates and the loss function. Therefore, it is common to feed *mini-batches* of data samples through the network. The gradients at each weight are then averaged over all samples in the mini-batch before weight update is performed. This speeds up the computation times and reduces the fluctuations of the weights. The typical size of a mini-batch may vary in the range of tens to hundreds, depending on the size of the dataset [152].

The learning rate in Equation 3.6 must be carefully chosen. A too small learning rate will lead to slow convergence, while a too large learning rate may cause fluctuation around the optimal solution. To this end, several alternative gradient descent-based optimization methods has been developed, that calculates adaptive learning rates for each parameter, such as Adagrad or the *Adaptive Moment Estimation (ADAM)* algorithm [152].

3 Machine learning for medical imaging

3.2.4 Regression

MLPs used for classification outputs probabilities of class memberships based on the input data. It is straight-forward to turn an MLP into a *regression model*, which, based on the input data, predicts a real-valued number, instead of a class probability, as output. First, only one output node is required, instead of K output classes for classification. Next, the softmax function in the output layer must be replaced by a linear activation function, i.e. $g(z) = z$. Lastly, as the cross entropy loss function is intended for use with probability distributions, alternatives, such as mean squared error loss function (Equation 3.14) should be considered.

With these modifications, an MLP may be used to predict a real valued number instead of classes membership. For instance, if the input are a set of tumor features, the network can be trained to predict some clinically relevant quantitative measure, based on the training data, such as expected number of months of survival.

3.2.5 Recurrent neural networks

The feed-forward neural network, introduced above, processes a fixed-length input vector and outputs a class assignment probability or a real number, in the cases of classification and regression, respectively. Such methods work for time-independent input features, such as those originating from a static PET image. However, sequential input data, for instance from a dynamic PET sequence, cannot be easily handled by ordinary feed-forward networks. For these data, where the input features are a time-series with τ time steps, $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(\tau)}$, another type, so-called *recurrent neural networks* have been developed.

Recurrent neural networks are based on *parameter sharing* across the time steps. This means that the same weight is used at different time points in the sequence, and that the output at each time-step is a function of all previous time-steps. This concept is illustrated in Figure 3.6, where the same parameter, w_1 is used for all time steps. In addition, a recurrent connection between the hidden layers has been introduced by the weight w_2 . This serves for passing information from previous time steps forward, and therefore allows the hidden layer to function as a memory state for the network. In recurrent neural networks, the hidden layer is therefore named *hidden state*. Thus, for each time step, t , the prediction, $y^{(t)}$ is modelled as $y^{(t)} = f(\mathbf{x}^{(t)}, \mathbf{h}^{(t-1)})$, where $\mathbf{x}^{(t)}$ is the current time step input, $\mathbf{h}^{(t-1)}$ is the previous time step hidden state, and f is parametrized by a neural network [153, 154].

The example network in Figure 3.6 has one output node for each time step, thus it calculates an output time series with the same length as the input. Other

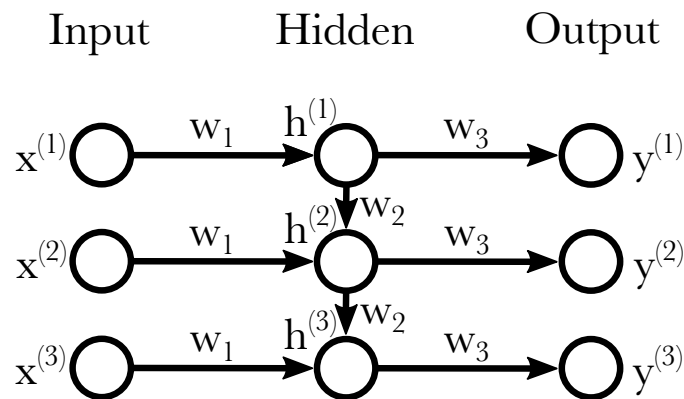


Figure 3.6: Illustration of a recurrent neural network with three time steps. Circles in the diagram represent nodes of the network. The input features $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, $\mathbf{x}^{(3)}$ represent different temporal instances of \mathbf{x} . The line from each input node represents multiplication by a weight, where the same weight is used for all time steps between the input and hidden layer, as well as between the hidden and output layer. Each node in the hidden and output layer represents summation of the weighted output from the previous layer, addition of the bias term, and applying the activation function to the result. Note that the bias term and the non-linear activation function at each node have been omitted in the illustration.

architectures are also possible, for instance where a class probability, or a real number is predicted at the end of the last time step [154].

Similar to feed-forward neural networks, recurrent neural networks are trained with the backpropagation algorithm, but because of the time-dimension, this process is referred to as *backpropagation through time*. Unfortunately, standard recurrent neural network architectures have a limited ability to learn long-term dependencies because of so-called *vanishing* or *exploding gradients*. These arise from the repetitive multiplication of the same weights in the network during backpropagation. An activation of the hidden state at an initial time-step may therefore disappear and be over-written with a new activation at later time steps [153, 155].

3.2.6 Long short-term memory networks

To overcome the problem of learning long-term dependencies, *gated* recurrent neural network architectures have been introduced. One example of such a network is the long-short-term memory (LSTM) network [156], illustrated in Figure 3.7. In addition to the hidden state, an LSTM also contains a cell state that passes information forward from previous time steps. The information that is stored in the cell state is carefully regulated by three serial gates: an input gate, a forget

3 Machine learning for medical imaging

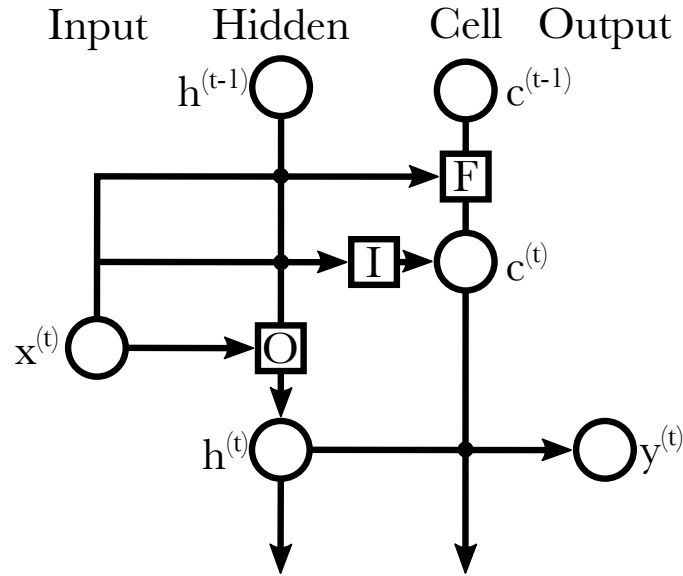


Figure 3.7: Illustration of one time-step of a long short-term memory network. Circles in the diagram represent nodes of the network. The cell state works as the long-term memory in this network. First, the input feature at the current time step, $\mathbf{x}^{(t)}$, and the hidden state from the previous time step, $\mathbf{h}^{(t-1)}$, are combined to allow the *forget gate*, F , to control what is removed from the cell state from the previous time step, $\mathbf{c}^{(t-1)}$. Next, $\mathbf{x}^{(t)}$ and $\mathbf{h}^{(t-1)}$ are combined to let the *input gate*, I , control what is added to the cell state to form $\mathbf{c}^{(t)}$. In addition, $\mathbf{x}^{(t)}$, and $\mathbf{h}^{(t-1)}$ are also used at the *output gate*, O , to generate the hidden state representation for the current time step, $\mathbf{h}^{(t)}$, which is then used together with the current cell state, $\mathbf{c}^{(t)}$ to generate the output at the current time step, $\mathbf{y}^{(t)}$. The vectors $\mathbf{h}^{(t)}$ and $\mathbf{c}^{(t)}$ are passed on to the next time step, where the process is repeated for with input $\mathbf{x}^{(t+1)}$. The line from each node represents multiplication by a weight, where the same weight is used for all time steps. Each node represents summation of the weighted inputs to that node, addition of the bias term, and application of an activation function. Note that the bias term and the activation functions have been omitted in the illustration. Adopted from [154].

gate and an output gate decide what information will be added to, removed from, or carried on by the cell state at each time step [156]. In practice, the gates consist of trainable weights and activation functions. For instance, if the input gate allows to add some information to the cell state at the first time step, this is carried on to subsequent time steps as long as the forget gate does not remove it. The information from the cell state can be combined with the hidden state to generate an output at any subsequent time step, as controlled by the output gate [153].

The LSTM approach is used for AIF prediction in **Paper III**.

3.3 Gaussian processes

In Equation 3.2 of Section 3.1.2, the mapping function of a classification system, $f(\mathbf{x})$, was specified as a linear function with parameters \mathbf{w} and b . Similarly, the neural network approach (Section 3.2) was described as a system of connected linear mapping functions with a finite set of learnable parameters. Gaussian processes (GPs) introduce a different approach for estimating such mapping functions.

GPs represent a *non-parametric* method, such that there are no assumptions on a particular set of parameters. It is based on *Bayesian statistics*, meaning that it starts by assuming a prior distribution over all possible mapping functions, before observing the training data. Samples from the prior follow a Gaussian distribution, which should reflect general properties of the expected training data in terms of mean value and variance. With *Bayes rule*, the prior distribution over functions can be conditioned on the training data points, such that functions passing through training data points are given increased weight, resulting in a posterior distribution. By sampling functions from the posterior it is possible to estimate the underlying mapping function, $f(\mathbf{x})$, as a GP, specified by its mean function. The estimated mapping function can then be used to infer function values for new, unseen, test data. In addition, the variance of the posterior distribution can be calculated, and thus provide an uncertainty measure of the predictions [157].

The main advantage of GPs over the neural network approach is the possibility for uncertainty estimation directly from the input training data. Furthermore, GPs use *covariance functions*, referred to as *kernels*, to calculate similarities between data points. Such, so-called *kernel methods*, are commonly known to work well with sparse training datasets, as opposed to neural networks [158].

Although GPs can be used for classification, for this thesis it is relevant to describe the methods for function estimation (regression). GP regression is used for AIF prediction in **Paper III** and **Paper IV**.

3.3.1 Definition

A GP is, by definition, a collection of any finite number of random variables that follow a joint multivariate Gaussian distribution. Given an input vector, \mathbf{x} , a GP describes the probability distribution of the mapping function value, $f(\mathbf{x})$, at the location of the input. A GP is completely defined by its *mean function*, $m(\mathbf{x})$, and *covariance function* (or *kernel function*), $k(\mathbf{x}_i, \mathbf{x}_j)$, where the mean function is commonly set to zero [157].

3 Machine learning for medical imaging

3.3.2 Covariance functions

A covariance function is a function that takes two vectors as input and returns a real number which is small if the vectors are close in feature space, that is, if they are *similar*. Likewise, a large real number is returned if the input vectors are far apart, or *dissimilar*. Covariance functions are often denoted as $k(\mathbf{x}_i, \mathbf{x}_j)$, where \mathbf{x}_i and \mathbf{x}_j are two p -dimensional input vectors. A *kernel matrix* may be generated as a square matrix, $K_{i,j}$, with entries $k(\mathbf{x}_i, \mathbf{x}_j)$. Per definition, covariance functions must generate kernel matrices that are *symmetric* and *positive semidefinite* (eigenvalues ≥ 0) [157].

There exists many different covariance functions from which two will be presented next. The *squared exponential* covariance function is intuitive and simple to explain and is commonly used in many kernel applications. However, in **Paper III** and **Paper IV**, a covariance function from the *Matérn* class was used for regression.

Squared exponential covariance function

One commonly used covariance function is the *squared exponential* covariance function, defined for two p -dimensional input vectors, \mathbf{x}_i and \mathbf{x}_j , as [157]:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 e^{-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^\top L(\mathbf{x}_i - \mathbf{x}_j)}. \quad (3.16)$$

Here, $L = \ell^{-2}I$ for isotropic kernels, where ℓ is a scalar, or $L = \text{diag}(\boldsymbol{\ell})$ for anisotropic kernels, where $\boldsymbol{\ell}$ is a vector of length p , thus with different values for each dimension. In Equation 3.16, σ_f^2 is a hyperparameter that controls the overall variance of the distributions, while the ℓ is the *length scale*, controlling the scaling of the inputs dimensions. Intuitively, the length scale affects the distance at which two points are considered to be close (or *similar*), along a specific dimension. In this way, the covariance function returns a positive number that is large if the input vectors are dissimilar, while for similar input, the exponential term approaches unity [157].

Matérn class of covariance functions

The squared exponential covariance function has been suggested to be too smooth for modeling many physical processes. As an alternative, the *Matérn* class of covariance functions has been proposed. In addition to the overall variance and length scale, this covariance function has another hyperparameter, ν , which control the smoothness

of the generated functions. The scaling of the Matérn covariance function has been chosen such that when $\nu \rightarrow \infty$, it approaches the squared exponential. Also, the expression for the Matérn class covariance functions simplifies considerably when choosing ν as a half-integer. It is argued in reference [157] that one of the most relevant cases for machine learning is when $\nu = 5/2$. In this case, the Matérn covariance function is a product between a polynomial and an exponential function, as follows [157]:

$$k_{\nu=5/2}(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \left(1 + \sqrt{5}r + \frac{\sqrt{5}}{3}r^2 \right) e^{-\sqrt{5}r}, \quad (3.17)$$

where $r = (\mathbf{x}_i - \mathbf{x}_j)^\top L(\mathbf{x}_i - \mathbf{x}_j)$, with L and σ_f^2 being defined equally as in Equation 3.16.

3.3.3 Prior and posterior distributions

Once the covariance function has been defined, it is possible to generate samples from the prior distribution, \mathbf{f}_* . Assuming N_* test locations at which the prior should be generated, the prior distribution is defined by a random Gaussian vector:

$$\mathbf{f}_* \sim \mathcal{N}(0, K(X_*, X_*)). \quad (3.18)$$

Here, a zero mean function is assumed. Further, K is the symmetric $N_* \times N_*$ covariance matrix where the entries are the result of the kernel function evaluated for each pair of input points, $\mathbf{x}_i, \mathbf{x}_j | i, j = 1, \dots, N_*$, in a p -dimensional input data matrix, X_* [157]. Figure 3.8A illustrates three realizations of the prior distribution for a one-dimensional input vector.

The prior distribution in Figure 3.8A has equal variance for all input points, thus, all predicted curves are equally probable. To restrict the number of possible predictions, training data must be introduced. First, assuming N training samples, the joint prior distribution of the training outputs, \mathbf{f} and the test outputs, \mathbf{f}_* are formed by [157]:

$$\begin{bmatrix} \mathbf{f} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N} \left(0, \begin{bmatrix} K(X, X) & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix} \right). \quad (3.19)$$

Note that as kernel matrices are symmetric, $K(X, X_*) = K(X_*, X)$. Thus, $K(X, X_*)$ and $K(X_*, X)$ results in an $N \times N_*$ and $N_* \times N$ matrix, respectively, while $K(X, X)$ in an $N \times N$ matrix.

3 Machine learning for medical imaging

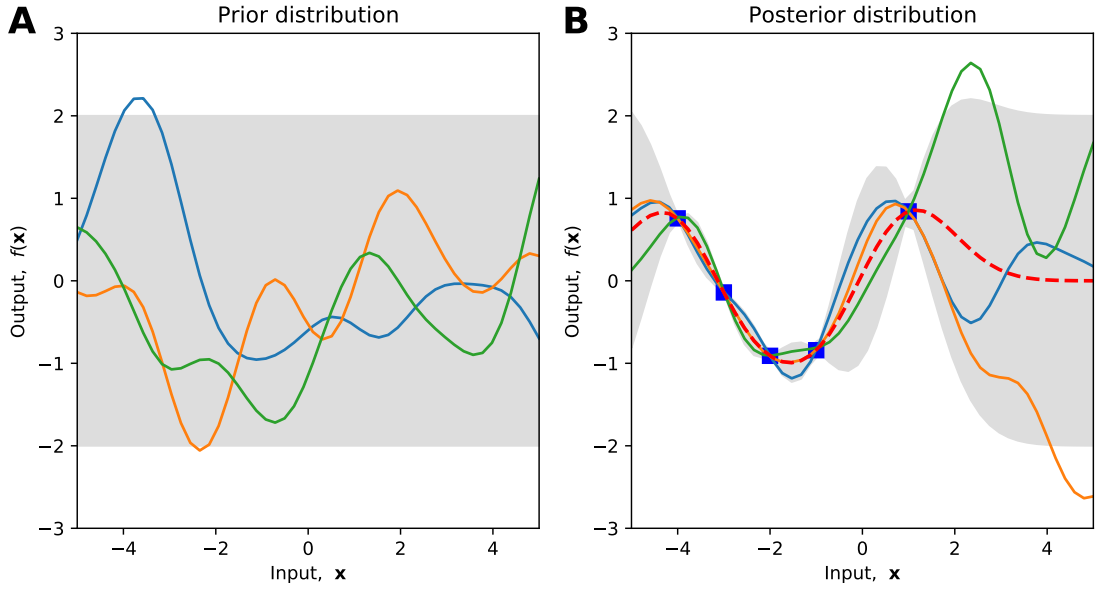


Figure 3.8: Illustration of a GP's prior and posterior distribution using the squared exponential covariance function (Equation 3.16) with unit variance, $\sigma_f^2 = 1$ and unit length scale, $\ell = 1$. A. Three realizations of the prior distribution for a one-dimensional input vector, consisting of $N = 50$ evenly distributed data points in the range between -5 and 5. B. The posterior distribution is conditioned on the five training data points, effectively resulting in a rejection of all functions that do not pass through the training data. In both A and B, the shaded area represents the mean plus and minus two standard deviations for each input point. Note that the 50 discrete data points are here represented by continuous lines.

By conditioning the joint prior distribution on the training data points, the posterior distribution is obtained. The posterior distribution does, therefore, only contain functions that agree with the observed training data [157]. This is illustrated for the one-dimensional case in Figure 3.8B. Evidently, the variances at the locations of the training observations are zero, and thus all the sampled functions, and obviously also the mean function, passes through the training points. Regression can now be performed at any desired test location along the input dimension.

The posterior distribution may be derived by conditioning the joint Gaussian distribution on the observations, again assuming zero mean function [157]:

$$\mathbf{f}_* | \mathbf{f} \sim \mathcal{N}(K(X_*, X)K(X, X)^{-1}\mathbf{f}, K(X_*, X_*)K(X, X)^{-1}K(X, X_*)) \quad (3.20)$$

Here, the first term describes the predictive mean, while the second term is the variance.

3.3.4 Noisy observations

In reality, the observations are usually corrupted by noise, ϵ . Consequently, the mapping function is modified to $y = f(\mathbf{x}) + \epsilon$. It is common to assume a Gaussian noise model, such that $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2)$. Furthermore, noise is assumed to be uncorrelated between samples. This allows the noise term to be incorporated into the covariance function, such that, for the squared exponential covariance function, Equation 3.16 becomes [157]:

$$k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 e^{-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^\top L(\mathbf{x}_i - \mathbf{x}_j)} + \sigma_\epsilon^2 \delta_{ij}, \quad (3.21)$$

where δ_{ij} is the Kronecker delta, equaling one if $i = j$ and zero elsewhere. Similarly, the prior on the observations becomes $y \sim \mathcal{N}(0, K(X, X) + \sigma_\epsilon^2 I)$, where I is the identity matrix, thus affecting only the diagonal terms of the covariance matrix. A similar expression can be derived also for the Matérn covariance function in Equation 3.17. With these modifications, the equations for the predictive mean, $\mathbb{E}[f_*]$, and the variance, $\mathbb{V}[f_*]$ becomes [157]:

$$\mathbb{E}[f_*] = K(X_*, X)[K(X, X) + \sigma_\epsilon^2 I]^{-1} \mathbf{y}, \quad (3.22)$$

$$\mathbb{V}[f_*] = K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_\epsilon^2 I]^{-1} K(X, X_*) \quad (3.23)$$

Note that, in Equation 3.22, the mean function is a linear combination of the observations, \mathbf{y} . On the other hand, the variance, in Equation 3.23 is independent of the observations and thus only depending on the input training and test data. Also of note is that the first term in Equation 3.23, $K(X_*, X_*)$, is the prior covariance (from Equation 3.18), from which a positive term is subtracted, representing the information obtained from the observations. This explains the observed reduction of the variance close to the training data, in Figure 3.8B [157].

As \mathbf{y} is an $N \times p$ matrix, where p is the number of dimensions, the output of Equation 3.22 will be a corresponding $N_* \times p$ matrix for all test samples. Similarly, the output of the predicted variance, in Equation 3.23, will be an $N_* \times N_*$ matrix.

Both Equation 3.22 and 3.23 involves inversion of an $N \times N$ matrix, which may be computationally demanding for large datasets. Practical implementations, therefore, calculate the so-called *Cholesky decomposition* (or *matrix square root*), M , of the covariance matrix, such that $K = MM^\top$, where M is a lower triangular matrix. This decomposition may significantly simplify the matrix inversion by a two-step forward and backward substitution, as outlined in reference [157].

3.3.5 Hyperparameter optimization

By the introduction of the noise term in Equation 3.21, the hyperparameter space is now parametrized by $\boldsymbol{\theta} = \{\sigma_f^2, \ell, \sigma_\epsilon^2\}$ ¹. Figure 3.9 indicates how varying hyperparameters affects the prior and posterior distributions. Evidently, the variance term, σ_f^2 , affects the overall variations of the generated functions along the output dimension (Figure 3.9C-D); the length scale, ℓ , alters the variations along the input dimension (Figure 3.9E-F); and the noise term, σ_ϵ^2 , introduces noise within the generated functions, which consequently implies a non-zero variance at the test points (Figure 3.9G-H).

The hyperparameters are commonly optimized on the basis of the observed training data, \mathbf{y} . The objective function used for this purpose, is the so-called *marginal likelihood*, as this expression contains both the observations and the hyperparameters. The derivation of this expression involves an exponential term originating from the covariance function. To simplify the expressions, it is common to compute the logarithm of the marginal likelihood. It can be shown that the log marginal likelihood is given by [157]:

$$U = \log[p(\mathbf{y}|X, \boldsymbol{\theta})] = -\frac{1}{2}\mathbf{y}^\top (K(X, X) + \sigma_\epsilon^2 I)^{-1} \mathbf{y} - \frac{1}{2} \log[|K(X, X) + \sigma_\epsilon^2 I|] - \frac{N}{2} \log[2\pi] \quad (3.24)$$

The optimal solution is found by maximizing this objective function using, for instance, gradient descent. Alternatively, minimizing the negative log marginal likelihood would yield the same result. In short, the partial derivatives with respect to the hyperparameters, $\boldsymbol{\theta}$, are calculated and solved for $\frac{\partial U}{\partial \theta_i} = 0$, where θ_i represents each of the hyperparameters of the covariance function. Once the optimal hyperparameters have been found, they can be used together with Equations 3.22 and 3.23 to calculate the GPs predictive mean and variance, respectively [157].

¹This parametrization is valid for the squared exponential and the Matérn covariance functions. For other covariance functions, additional hyperparameters may be required.

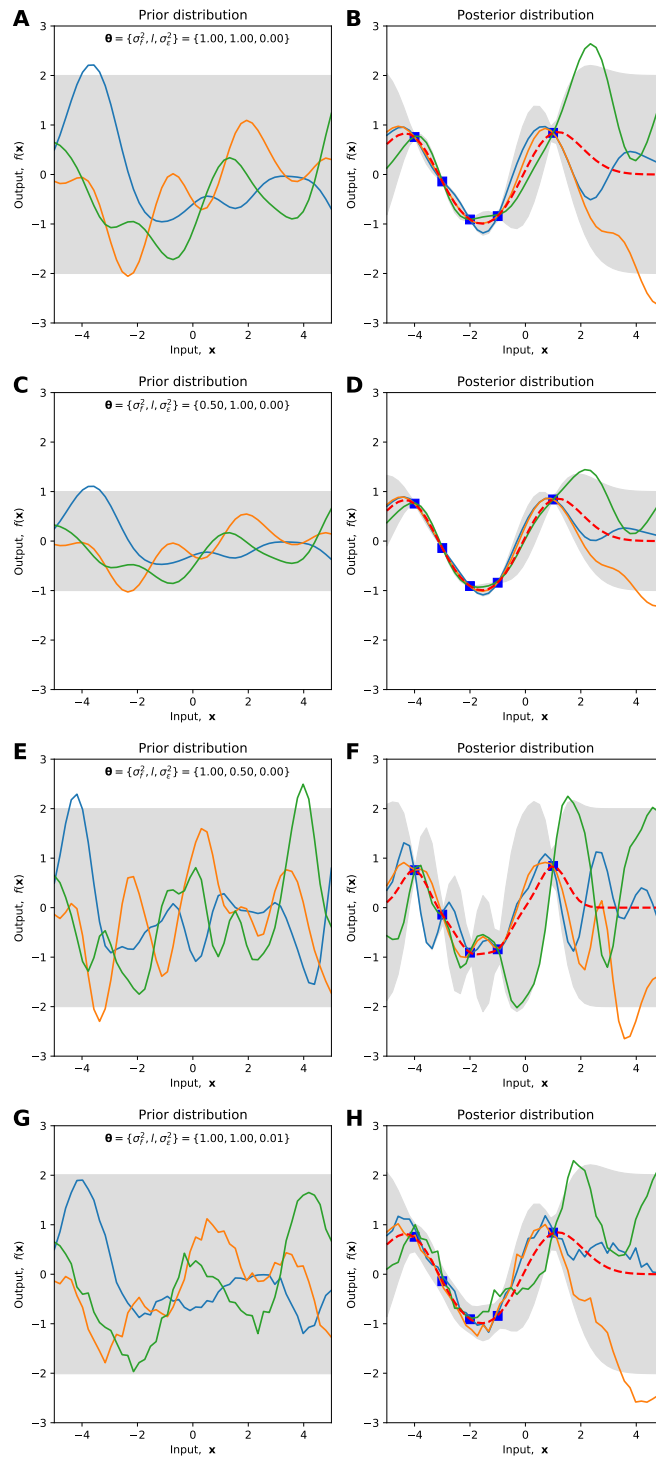


Figure 3.9: Illustration of a GP's prior (left column) and posterior (right column) distribution using the squared exponential covariance function (Equation 3.16) with varying hyperparameters, variance, length scale, and noise, $\theta = \{\sigma_f^2, l, \sigma_\epsilon^2\}$, as indicated in each subfigure. A-B. Unit overall variance, unit length scale and zero noise variance (equal to Figure 3.8). C-D. Decreased overall variance yields decreased overall variations of the generated functions. E-F. Decreased length scale results in increased variations along the input dimension of the generated functions. G-H. Increased noise variance yields within-function variations and introduces an uncertainty around each training point. Please refer to the legend of Figure 3.8 for a full description of the generation of the prior and posterior distributions.

3.4 Radiomics

Radiomics refers to the concept of high-throughput quantitative image analysis based on extracted medical imaging features. It relies on standardized definitions of mathematical features that are extracted from medical images obtained from standardized imaging protocols. The underlying hypothesis is that an automated analysis of the extracted features, by means of for instance machine learning, is superior to manual, visual analysis by a human observer [159].

Radiomics has been used to build predictive models in CT [29, 31], MRI [160, 161], and PET [162]. Specifically, radiomic features extracted from lung-tumor PET images have been used for the prediction of histological subtypes [163], epidermal growth factor receptor (EGFR) mutation status [28], anaplastic lymphoma kinase (ALK) gene expression [164], survival [165–171], local control [172] and recurrence [171, 173].

The major challenges in the field of radiomics has been the lack of standardized feature definitions and processing workflow that allows generalization of results between imaging centers [174]. Ideally, to build robust prognostic models, radiomic biomarkers should have low variability to factors which are unrelated to the disease. Therefore, standardized imaging and processing protocols are important to minimize the variability due to scanner type [175], reconstruction settings [176–178], segmentation method [176, 178], and gray scale discretization method [179]. The Image Biomarker Standardisation Initiative (IBSI) is a recent attempt to provide procedures for standardized processing and feature extraction, as well as definitions of hundreds of mathematical imaging features [180].

Radiomics is relevant for **Paper II**, where the motion-robustness of PET-derived radiomic features is investigated. In the following, the most important key concepts in the radiomics pipeline are discussed.

3.4.1 Image processing

It is important that the images are processed to enhance the quality for subsequent feature extraction. This may include noise reduction, artifact reduction and normalization. For instance, PET images should be converted to the units of SUV (Section 2.1.9), and relevant artifacts should be corrected prior to further analysis (Section 2.4.5). If the voxel-dimensions are non-isotropic, they should also be interpolated into isotropic resolution, to make features rotationally invariant [180].

3.4.2 Segmentation

Extraction of radiomic features requires the definition of a volume of interest (VOI) in the image. This is the region in image space, for instance a tumor, within which the features are calculated. A standardized and reproducible VOI-definition method should be used within a study to reduce the variability. The VOI can be segmented either manually slice-by-slice, or by automatic, or semi-automatic segmentation methods [181]. In ^{18}F -FDG-PET, a semi-automated threshold method based on, for instance, 41% SUV_{max} can be employed to segment the tumor [64]. For the segmentation of normal brain anatomy, due to the rigid geometry of the skull, co-registration with a standardized VOI template is a plausible alternative to manual segmentation [182]. The segmented region is commonly represented as a binary mask with the same size as the original image, but with the voxel-values equal to zero outside the VOI and one inside the VOI.

3.4.3 Gray scale discretization

The range of intensity values within the VOI must be discretized to a limited range in order to extract meaningful radiomic features. Two main methods can be used for gray scale discretization. Either, the number of bins or the bin-width can be fixed. With the fixed bin-number method, the voxels are first normalized to the range 0 to 1 by the relative difference between maximum and minimum intensity values within the VOI. Next, they are re-sampled into a desired number of bins, by multiplication of the number of bins, most commonly 8, 16, 32, 64, 128 or 256 bins [179, 183]. The resulting bin number is then rounded down to the nearest integer bin, which implies that the normalized intensities no longer are related to the original image units. This method is therefore most suitable for images without absolute units, but where contrast differences are important, such as MRI. However, for PET and CT, the fixed bin-width method is recommended, where the image intensities are arranged into a new bin after each pre-defined bin-width. In this way, the intensities maintain the original units [180].

3.4.4 Feature extraction

When the images have been preprocessed, segmented and intensity-discretized, features can be extracted from each VOI. The features are commonly subdivided into voxel-based and shape features. Voxel-based features are calculated on the intensity values inside the VOI, while shape features are derived from the binary segmentation mask. Voxel-features, in-turn, may be subdivided into first order (or

3 Machine learning for medical imaging

intensity-based) statistical features, histogram features, and higher-order statistical (or texture) features. The latter group may include gray level co-occurrence matrix (GLCM) features, gray level dependence matrix (GLDM) features, gray level run length matrix (GLRLM) features, gray level size zone matrix (GLSZM) features, and neighboring gray tone difference matrix (NGTDM) features. These are all based on secondary matrices calculated from the discretized intensity values of the original images, each representing a different encoding of the spatial intensity variations of the original image. Details on the calculation of many standardized voxel- and shape-features are given in reference [180].

3.4.5 Image transformations

Features may be derived from original as well as transformed images. Specifically, features from wavelet filtered images have been commonly used for building predictive models in medical imaging [31, 184]. Wavelet filtering decomposes the original image into coefficients with varying level of detail by convolution with *wavelet functions* of varying scale. At each scale, directional low-pass (L) and high-pass (H) filtering of the image is performed with pre-defined wavelet functions. While the wavelet function and number of scales can be chosen arbitrarily, the so-called *coiflet wavelet*, applied at three scales, has been common within medical imaging [31, 184]. This results in eight wavelet coefficients from which features can be extracted using the original segmentation mask. The decompositions are commonly named LLL, LLH, LHL, LHH, HLL, HLH, HHL and HHH, according to the x, y, z order of directional filtering [31].

3.4.6 Feature selection

The features extracted from both original and transformed images may form a feature vector of several hundreds of features for each subject, many of which may be correlated or irrelevant for the prediction task. Furthermore, training a model with limited amount of data from a high-dimensional feature space may result in overfitting, meaning that the model will not generalize well to new data [185]. Also, as discussed earlier, features should be robust to study-specific factors, such as scanner and image processing settings. Feature selection is therefore an important step prior to, or simultaneously with model training.

Several feature selection methods, both supervised and unsupervised, exist [185]. For instance, correlated features add minor value to the models and may be removed [29]. Moreover, with so-called *wrapper methods*, subsets of features from a

3.5 Error estimation and validation methods

high-dimensional feature space are used to train a model. Evaluation is performed and training is repeated for a different feature subset, to find the optimal feature combination. This method is useful for low-dimensional feature sets, but is computationally intensive and thus not practical when dimensionality is increased. A comprehensive overview of feature selection methods is out of scope for this thesis but can be found in references [185, 186].

3.5 Error estimation and validation methods

When a machine learning model has been trained using the available training data, it is desired to evaluate its performance. This should be carried out on an external test dataset which has not been part of the training. In many practical situations only a single dataset is available for both training and testing and the dataset must therefore be split into a training and test set prior to training. This is referred to as the *holdout method*. It can be shown that the error probability decreases with increasing size on the training dataset. On the other hand, the variance of the model evaluation increases with decreasing size of the test dataset. Hence, for limited dataset sizes, the number of samples used for training and testing must be carefully considered [149].

It is common to use 80% of the data to train the models, while 20% are held out for subsequent testing. In this example, the dataset is split into five parts, four parts used for training and one part used for testing. By systematically excluding each of the five parts for test set, and repeating model training with the remaining four parts, it is possible to calculate an error-estimate of the model performance, by averaging the performance measurements from all test sets. This method is referred to as *K-fold cross validation*, with K being the number of splits.

One drawback of the holdout method is that, for small datasets, retracting 20% of the data from model training may decrease the accuracy of the trained model. In these situations, the *leave-one-out method* may be used. Here, one sample is left out for testing, while the remaining samples are used in the training set. The process is repeated with each of the samples being in the test set once. Thus, this is equivalent to K -fold cross validation with K being equal to the number of samples. One drawback of the leave-one-out method is that it is highly computationally demanding, as a new model must be trained for each sample in the dataset.

In **Paper III** and **Paper IV**, the leave-one-out method is used for cross validation, due to the limited size of the available datasets.

4 Summary of research

4.1 Paper I

Quantitative PET/MR imaging of lung cancer in the presence of artifacts in the MR-based attenuation correction maps

Samuel Kuttner, Martin Lyngby Lassen, Silje Kjærnes Øen, Rune Sundset, Thomas Beyer and Live Eikenes, *Acta Radiologica*, Apr. 2020, Vol. 61, pp 11-20

The aim of this paper was two-fold. First the frequency and test-retest reproducibility of artifacts in MRI-based attenuation correction maps in a PET/MRI lung cancer patient cohort was investigated. Second, the impact of artifact corrections on PET-based tumor quantification was studied. More specifically, truncation of arms outside the field-of-view, susceptibility and tissue inversion artifacts, as well as the absence of bone signal were investigated. Truncation artifacts were present in all of the acquisitions (100%), while susceptibility and tissue inversion artifacts were observed in 12%, and 52% of the scans, respectively. Furthermore, bone is not visible in standard MRI, due to the rapid spin-relaxation rates.

The most important result from this study was that up to 20% relative differences were introduced after susceptibility artifact correction, with large inconsistencies between test-retest scans. On the other hand, bone and truncation corrections affected the quantitative accuracy in lung tumor SUV with less than 5%, and also, tissue inversion was anticipated to have a small effect.

This study has demonstrated that specific type of imaging artifacts may have clinical implications for patients undergoing serial imaging for tumor therapy response assessment. Moreover, the study highlights the importance of correcting for imaging artifacts prior to training of predictive machine learning models, for instance in tumor classification studies. In this way, extracted tumor features will be the least affected by PET/MRI-imaging-related limitations, which will result in increased robustness of the predictive models.

Contributions by the author: The study concept and design were developed by me, in collaboration with the other co-authors. I performed the image and data analysis and discussed the results with the other co-authors. I wrote the main draft of the manuscript.

4.2 Paper II

Motion-robust radiomic features for image classification in ^{18}F -FDG PET/MRI imaging of lung cancer

Samuel Kuttner, Erna-Elise Paulsen, Robert Jenssen, Rune Sundset, Jan Axelsson, *Manuscript*

This paper aimed at studying motion variability of radiomic features in PET images of lung tumors. In addition to the artifacts studied in **Paper I**, breathing-induced motion may also introduce variability in the radiomic features, which reduce the accuracy of predictive machine learning models. Motion was characterized by comparing radiomic features extracted from tumor segmentations in three different PET reconstructions: 1) Free-breathing PET, 2) end-expiration PET and 3) MRI-based motion corrected PET. In each reconstruction, the primary tumor was segmented using a threshold algorithm, followed by the extraction of 834 standardized radiomic features from each segmentation.

This main outcome of this study was a list of 43 motion-invariant features, 14 of which had low correlation to the other included features. These features may be extracted from any of the three PET image reconstructions, without introducing reconstruction-bias in the feature values. The free-breathing PET scan protocol is simpler and more time-efficient to perform, compared to the motion correction sequences. Thus, by employing motion-robust features in machine learning prediction models, gating and motion correction acquisitions are not needed, which simplifies the PET workflow for patients, and allows for collection of data in retrospect, where respiratory gating or motion correction was not performed. Furthermore, hybrid PET/MRI is still an emerging modality, thus limiting the use of MRI-based motion correction at all PET imaging centers.

Contributions by the author: I developed the study concept and planned and coordinated the clinical PET/MRI study, in collaboration with the other co-authors. The image processing and data analysis was performed by me, and the results were discussed with the co-authors. I wrote the main draft of the manuscript.

4.3 Paper III

Machine learning derived input-function in a dynamic ^{18}F -FDG PET study of mice

Samuel Kuttner, Kristoffer Knutsen Wickstrøm, Gustav Kalda, S Esmail Dorraji, Montserrat Martin-Armas, Ana Oteiza, Robert Jenssen, Kristin Fenton, Rune Sundset and Jan Axelsson, *Biomedical Physics & Engineering Express*, Jan. 2020, Vol. 6, p 015020

The aim of this study was to explore machine-learning-based prediction of the AIF, using the MLIF approach, in a small-animal dynamic ^{18}F -FDG PET study. The AIF is required for the measurement of glucose metabolism by tracer kinetic modelling of dynamic ^{18}F -FDG-PET data. However, arterial blood sampling in rodents is challenging due to limited blood volume and complex surgery.

Two MLIF models, based on GPs and an LSTM network, respectively, were trained for AIF prediction using image-derived tissue regions as input. A reference AIF was formed by fitting an established AIF model to image-derived data of two blood-rich tissues (vena cava and left ventricle). The MLIF models were evaluated by first comparing the predicted MLIF curves with the reference AIF curves directly. Next, rate constants in different tissues, obtained from compartment modelling using both the predicted MLIF and the reference AIF curves, were compared. Also, the impact of different subsets of input tissue regions on AIF prediction was studied.

Both GPs and LSTM models generated MLIF curves similar to the reference AIF. The rate constants from both models agreed well with those obtained from the reference AIF. Myocardium was found to be important for MLIF prediction, but MLIF curves with similar error was obtained also without the myocardium in the input data. This study demonstrated that machine learning is feasible for accurate and non-invasive prediction of a reference AIF in ^{18}F -FDG studies of mice.

The main limitation of this study was that a blood-sampled AIF was unavailable for training. Instead, an image-based reference AIF was used. **Paper IV** demonstrates that the proposed methods are also feasible for predicting an AIF obtained from blood samples in a human PET cohort.

Contributions by the author: The idea was conceived by me and further developed in collaboration with the co-authors. I participated in the planning and conduction of the animal experiments, image acquisition and processing, together with several of the co-authors. Machine learning and compartment modeling was performed by me, and the results were discussed with the co-authors. I wrote the draft of the manuscript.

4.4 Paper IV

Cerebral blood flow measurements with ^{15}O -water PET using a non-invasive machine-learning-derived arterial input function

Samuel Kuttner, Kristoffer Knutsen Wickstrøm, Mark Lubberink, Andreas Tolf, Joachim Burman, Rune Sundset, Robert Jenssen, Lieuwe Appel and Jan Axelsson, *Journal of Cerebral Blood Flow & Metabolism*, Jan. 2021, [published online ahead of print, Feb. 9, 2021]

CBF can be measured with dynamic PET of ^{15}O -water by using tracer kinetic modelling. However, an AIF obtained from arterial blood sampling is required for such measurements. Arterial cannulation is an invasive, laborious and time-consuming procedure, associated with pain and risk for complications. The aim of this study was to further develop the MLIF methodology from **Paper III**, and to investigate if the non-invasive, machine-learning-based approach to AIF estimation was feasible in a clinical brain-PET study, evaluated with blood-based AIF data.

An MLIF model based on GP was employed in this dual-injection, dual-scan study. Each subject was scanned both before (baseline) and following acetazolamide medication. Acetazolamide dilates the vascular system and thereby increases the CBF. Three different image-derived time-activity curves were automatically segmented from the carotid arteries and used as input into the AIF prediction model. Two training situations were evaluated: in *case 1*, training and testing was performed on baseline and acetazolamide scans separately; in *case 2*, baseline scans were used for training followed by testing on acetazolamide scans. The idea with *case 2* was to resemble a more realistic clinical situation compared to *case 1*, namely that of having a database with baseline scans for training, followed by applying the trained model on subjects with an altered condition, in this case the acetazolamide scans. The approaches were evaluated by comparing AIF and MLIF curves, as well as whole-brain grey matter CBF values estimated by kinetic modelling using either AIF or MLIF.

The results showed that when training and testing was performed on baseline and acetazolamide scans separately (*case 1*), the AIF and predicted MLIF curves were found to be similar. Subsequent kinetic modelling resulted in similar CBF values when using AIF and MLIF, as well as similar CBF increase between baseline and acetazolamide scans. However, when using baseline scans for training followed by testing on acetazolamide scans (*case 2*), the CBF increase from baseline to post-medication conditions was reduced. It is suggested that the difference in measured acetazolamide provocation was caused by the different local input functions in the brain (image-derived input for *case 2*) and in the wrist (blood-based AIF sampling

4 Summary of research

for *case 1*). Although the relative changes between baseline and acetazolamide scans were different for *case 1* and *case 2*, both were significant and could, hence, be used in clinical practice to differentiate baseline from acetazolamide conditions.

The non-invasive MLIF method shows potential to replace the AIF obtained from blood sampling for CBF measurements using ^{15}O -water PET and kinetic modelling, thus, minimizing the risk for complications and, at the same time, simplifying the dynamic PET imaging workflow.

Contributions by the author: The idea was conceived by me and further developed in collaboration with the co-authors. I implemented the machine learning and compartment models. The results were analyzed and discussed in close collaboration with the other co-authors. I wrote the draft of the manuscript.

5 Results and discussion

There are many challenges associated with quantitative PET imaging. Over the course of more than 40 years since the clinical introduction of PET, many of the challenges have been solved. However, by the recent introduction of PET/MRI, new challenges have emerged, mainly related to limitations with MRI-based attenuation correction. Unless accounted for, these limitations may reduce accuracy and precision in PET/MRI-based quantification. Furthermore, subject motion introduces image blurring and possibly also misalignment between the attenuation correction map and the PET data, both of which affects PET-based quantitative measurements. Also, dynamic PET-based quantification requires additional steps during acquisition and analysis, the most important being that of arterial blood sampling, which is a challenging and time-consuming procedure in both humans and rodents.

The overall aim of this thesis is to address these challenges in order to allow for improved PET-based quantification. This is accomplished by studying two specific aims. First, it was investigated how artifacts in the attenuation correction maps as well as respiratory motion affected quantification in PET/MRI of lung lesions. Next, an MLIF method was developed, as a non-invasive alternative to the AIF, required for tracer kinetic modelling in dynamic PET studies.

The combined effect of artifact reduction and motion correction, may improve PET/MRI-based tumor quantification, and allow accurate and reproducible PET-based quantitative measurements to be performed. Also, by extracting motion-robust quantitative features from a lung tumor, in a radiomic framework, it is possible to train machine-learning-based models for disease state prediction using data collected without the use of gating or motion correction. This may considerably simplify the PET imaging protocols, shorten the image acquisition times and allow for inclusion of data in retrospect, where respiratory gating or motion correction was not performed. Also, quantitative measurements from dynamic PET studies, using tracer kinetic modelling, may be considerably simplified by using the MLIF approach, rather than performing full arterial blood sampling. This may allow easier clinical and pre-clinical implementation of dynamic PET protocols, and in this way advance the outcome of quantitative PET studies.

5.1 Paper I

In **Paper I**, the frequency and the test-retest reproducibility of artifacts in MRI-based attenuation correction maps in a PET/MRI lung cancer patient cohort was investigated. Individual and combined corrections of three types of artifacts (absence of bone, truncation of arms, and susceptibility artifacts) in standard DIXON-based attenuation correction maps were studied by investigating their impact on lung lesion quantification.

The most important result from this study were that up to 20% relative differences was introduced after susceptibility artifact correction, with large inconsistencies between test-retest scans. On the other hand, bone and truncation corrections affected the quantitative accuracy in lung tumor SUV with less than 5%.

Among the included patients, susceptibility artifacts were caused by a void of MRI signal due to the local magnetic field distortions induced by the surgical sternal wires after thoracotomy. Three scans with susceptibility artifacts were found in the dataset. In two of these acquisitions the susceptibility artifacts in the sternum resulted in a failure of the lung segmentation algorithm and the following invalid assignment of air attenuation values to lung tissue. For these scans, the artifact correction, and subsequent filling of the lungs with correct attenuation values, had a large impact on the corresponding lung tumor SUVs (-5.2% to -22.1%). However, the algorithm failure was inconsistent between test-retest scans, thus introducing scan-dependent variations in SUV measurements. For acquisitions with susceptibility artifacts where the lung segmentation was successful, the impact of the corrections on the SUV values were lower (-0.3% to -2.8%). This demonstrated that the accuracy in the attenuation values of the tissue surrounding a lung lesion was important for accurate tumor quantification, as also reported by others [82].

The observed test-retest inconsistency of specific type of imaging artifacts may have clinical implications for patients undergoing serial imaging for tumor therapy response assessment. Therefore, careful inspection of the MRI-based attenuation correction maps is required, preferably while the patient is still inside the scanner. In this way, a simple re-scan of the DIXON sequence may, in some cases, provide an artifact-free attenuation correction map. However, if the artifacts persists, specifically susceptibility artifacts must be corrected, to ensure highest possible diagnostic accuracy of a lung PET/MRI examination.

This study highlights the importance of correcting for imaging artifacts prior to training of predictive machine learning models, for instance in tumor classification

studies. In this way, extracted tumor features will be the least affected by PET/MRI-imaging-related factors, which will result in increased robustness of the predictive models.

In **Paper I**, there were only three patients with susceptibility artifacts in the study population. In other words, it was not possible to estimate any population average effects from the susceptibility artifact corrections. Therefore, in a follow-up simulation study, synthetic artifacts in 19 of the patients from **Paper I** were introduced. Briefly, the lung tissue attenuation values were replaced with the corresponding values for air in the attenuation correction maps of all included patients. Subsequently, these modified attenuation correction maps were used during PET image reconstruction, and the difference in SUV- and volume-based tumor features were compared to the case with correct attenuation correction maps [187]. The general reconstruction and image analysis methods were similar to that of **Paper I**.

Figure 5.1 shows the result of the follow-up simulation study [187]. Significant median relative differences of around 10% were found in SUV-based features when comparing original and artifact-induced images. Individual relative differences of up to 40% was observed in the SUV-features. On the other hand, tumor-size-metrics were not affected by the simulated artifacts.

The results from the simulation study strengthens the conclusions from **Paper I**, namely that susceptibility artifacts may cause clinically relevant biases in lung tumor SUV measurements.

In **Paper I**, the susceptibility artifacts were corrected by manual in-painting of the affected region in the attenuation correction maps. Other approaches for metal artifact reduction in PET/MRI have been suggested. For instance, dedicated MRI sequences, may reduce the size of metal artifacts, although no clinical study has yet shown the potential of this approach applied to whole-body PET/MRI [188]. Also, artifact correction methods have been proposed using deep learning approaches. This has shown potential for automated corrections of arm-truncation and dental metal-artifacts in the MRI attenuation correction maps [189, 190]. However, this approach has not yet been evaluated for metal artifacts in sternum, which is relevant for this thesis. One reason may be that the thorax anatomy is more complex compared to the head and neck anatomy, and therefore require more complex models and more training data.

A different promising approach is to estimate a pseudo-CT attenuation correction map directly from the non-attenuation-corrected PET data using a deep learning algorithm based on generative adversarial networks [191]. The pseudo-CT can then be used for attenuation correction. This approach has been evaluated for

5 Results and discussion

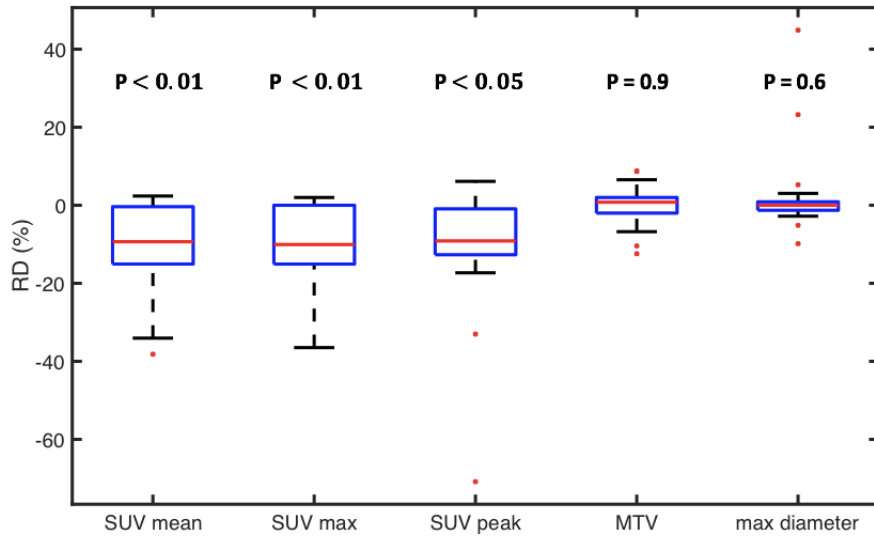


Figure 5.1: Box plot showing the relative difference (RD) in five quantitative measurements obtained from the segmented lung lesions in PET images reconstructed with artifact-free and artifact-induced MRI-based attenuation correction maps. Nineteen patients were included in this simulation study. The artifact considered in this study was that of falsely assigning lung tissue with the attenuation value of air. Note that this figure is inverted compared to Figure 4 of **Paper I**. This is because here, the reference image is the artifact-free, while in **Paper I**, the reference image was the artifact image. The tumors were segmented using the 41% SUV_{max} threshold. The metabolic tumor volume (MTV) is the volume of this segmentation. SUV_{mean} and SUV_{max} refers to the average and maximum voxel-value inside the MTV, respectively. SUV_{peak} is defined as the average SUV within the 1 cm^3 volume with maximal average value inside the tumor [64]. Maximum diameter refers to the longest straight line that can be drawn in three-dimensions inside the tumor. P values indicate statistical significance based on a paired t-test.

PET/CT, while for PET/MRI applications, the method would require co-registered CT images for the neural network training. Another limitation with this method is that it only works for non-specific tracers, such as ^{18}F -FDG, as the pseudo-CT generation is based on PET-input data with a high content of anatomical details.

5.2 Paper II

One specific limitation that was not addressed in **Paper I** was the impact of respiratory motion on PET quantification. Previous studies have shown that breathing motion has a large impact on the extracted SUV values, the metabolic tumor volumes, as well as on extracted radiomic features [44, 45, 192–194]. These studies evaluated the impact of motion based on end-expiration-gated PET [44, 45]. This acquisition mode has an optimal duty cycle in the order of 35% and thus have an increased noise level compared to free-breathing PET [195].

In **Paper II**, the impact of breathing-induced motion on PET quantification was evaluated by an MRI-based elastic motion correction algorithm (BodyCompass, Siemens Healthineers, Erlangen), recently made available at our PET imaging center. With this technique the motion vector fields are calculated over the phases of the breathing cycle using MRI, and used for motion correction of the simultaneously acquired PET data. In this way, a stationary and motion corrected PET image can be obtained without rejection of acquired PET events, thus allowing for motion corrected images with similar noise level as free-breathing PET [98, 196].

A prospective clinical study was initialized, where patients with suspected non-small-cell lung cancer, referred to clinical PET/CT at our institution, were asked to participate in a research PET/MRI scan after their clinical scan. The overall aim with the study was to generate data for training a machine-learning-based model for PET-based disease classification. One approach to do this is by extracting image features as input into the models. To build robust prognostic models, the input features should have low variability to factors which are unrelated to the disease, such as scanner type, reconstruction settings, tumor segmentation method, gray scale discretization method and motion. While many of these factors have been extensively studied in previous literature [175–179], only a few studies have addressed the impact of motion on radiomic features, and in these studies, end-expiration gated PET scans were used [44, 45]. Therefore, in **Paper II**, the impact of MRI-based motion correction on PET-based radiomic features was investigated. The aim was to generate a list of motion-robust features, which may be extracted from standard free-breathing PET examinations, to build machine-learning-based models for disease state or survival prediction.

Breathing-induced motion was studied by using two different motion correction acquisitions: classical end-expiration gating ($PET_{40\%}$) as well as the new MRI-based elastic motion correction (PET_{MoCo}). These were compared to a reference free-breathing PET acquisition ($PET_{100\%}$). The rationale was that any feature that remained equal (within a tolerance) in all three scans was independent of the scan mode, and thus also of motion, as characterized by PET.

5 Results and discussion

This main outcome of this study was a list of 43 motion-invariant features. These features may be extracted from any of the three PET acquisition modes, without introducing a scan-type-dependent bias in the feature values. Furthermore, 14 of the 43 features were found to have low correlation to the other included features, which could therefore represent one possible low-dimensional feature set to be used as input into subsequent prediction models. However, correlation-based feature elimination is just one of many possible feature selection methods [29, 186, 197]. With the 43 motion-robust features found in this study, it is possible to explore both supervised or unsupervised feature selection methods, to find the most optimal features for the prediction being investigated.

The results from this study have two important implications for PET-based quantification and machine-learning-based prediction models. First, it represents a *feature selection* step in the machine learning pipeline; second, it may simplify the acquisitions by requiring only free-breathing PET scanning. These two aspects will be discussed next.

It is well known that machine learning models trained on a high-dimensional feature space using too few training examples (referred to as $p \gg N$ problems), may result in overfitting of the models. In other words, the number of samples, N , must be large enough with respect to the number of features, p . For many classification problems, it can be observed empirically that the classification error will initially decrease as the number of features are increased. At some critical number of features, the error will start to increase again. This is referred to as the *peaking phenomenon*. Thus, for a given problem there is an optimal number of features, which, only up to a given number, is increasing with the number of trainable model parameters. [149]. This implies a necessity for feature selection.

In **Paper II**, a constraint was imposed on the selected features, namely that they should be independent of motion. For instance, a feature that is dependent on the breathing of the patient may not aid in the separation of tumor classes. The clinical implication of the results is that in future studies, the motion-robust features could be extracted from either of the scans and used in a classification model. The free-breathing PET scan protocol is simpler and more time-efficient to perform, compared to the motion correction sequences. Thus, by employing motion-robust features in machine learning prediction models, gating and motion correction-based acquisitions are not needed, which simplifies the PET workflow for patients. Moreover, hybrid PET/MRI is still an emerging modality, thus limiting the use of MRI-based motion correction at all PET imaging centers. This may allow to generate datasets from multi-center trials, where access to a PET/MRI scanner is not necessarily required. Lastly, it may also allow collection of data in retrospect, where respiratory gating or motion correction was not performed.

The main limitation with this study was the limited sample size and that the end-points of all patients were not yet known. Nevertheless, the sample size was comparable to similar published studies with the aim to investigate motion variability of image-derived features [44, 45]. Once the end-points have been reached, a future follow-up study may allow to investigate the impact of the motion-robust features on disease state or survival predictions.

In the present work, the motion vectors used for elastic motion correction were derived from MRI data. To simplify the imaging protocol even further, the motion vectors can also be derived from gated PET data. In short, this is accomplished by dividing the breathing cycle into distinct phases, and reconstructing PET images of each phase, similar to conventional gating. To avoid mismatch between the attenuation correction map, commonly acquired during end-expiration breath hold, and the end-inspiration breathing phases, no attenuation correction is performed for these gated image reconstructions. The image velocities are then calculated between each gate and the reference image, as described in Section 2.4.6, and used to reconstruct motion corrected PET images, in a similar way as for PET/MRI (Section 2.4.6 and 2.4.6) [97].

One limitation with PET-based estimation of the motion vectors is that it is only applicable with tracers that have a high background uptake, such as ^{18}F -FDG. For tracers with more specific uptake, such as many ^{11}C , ^{68}Ga and ^{18}F -based tracers (other than ^{18}F -FDG) [198, 199], there may not be enough motion information between the gated images to calculate the motion vectors for all voxels and consequently PET-based elastic motion correction may not work.

Another limitation is that PET-based elastic motion correction does not allow correction of the attenuation correction map for patients who did not comply with breathing instructions. The pseudo-CT approach for generating attenuation correction maps of several breathing phases from gated PET images, discussed in Section 5.1, may be feasible in these cases. Alternatively, with the use of MRI-based motion correction, the end-expiration attenuation correction map from the DIXON sequence can be compared with the MRI images from all breathing-phases. In this way, cases where the patient was unable to comply with breathing instructions (which is not uncommon in lung cancer imaging) may be identified and corrected. In these cases, the acquired attenuation correction map (in, for instance end-inspiration) can be transformed into end-expiration by using the MRI-derived motion vectors, and, in-turn, be used during PET-image reconstruction. In the case of PET/CT, it is non-trivial to obtain phase information from the CT-based attenuation correction map. This would require four-dimensional CT images, which would considerably increase the radiation dose to the patient, and further prolong the imaging time [200]. On the other hand, with MRI-based motion correction, the

5 Results and discussion

attenuation correction map can be transformed into the correct breathing phase using the motion vector fields, thus reducing the image artifacts in the reconstructed PET images [98]. MRI-based elastic motion correction may, therefore, be considered as the most feasible elastic motion correction strategy for lung cancer patients, although the deep-learning-based pseudo-CT approach could also be highly relevant for this problem.

Recently, a deep-learning-based approach for estimating the motion vector fields was proposed [201], which could be a viable alternative to MRI-based derivation of the motion vectors. However, this method also suffers the above-mentioned limitations of tracer application and attenuation correction, so it is not obvious where this method could provide advantages over, for instance MRI, or PET-derived motion vector fields.

5.3 Paper III

Paper III addresses one of the main challenges in quantitative, pre-clinical, dynamic PET imaging, namely that of obtaining a valid AIF. In rodents, this is hampered by the limited allowed blood withdrawal volume and complex and terminal surgical procedures for inserting an arterial catheter [4]. In this work, the MLIF approach is proposed as a non-invasive alternative for AIF estimation in a dynamic PET study of mice.

The main result of this study was that an MLIF, predicted from image-derived tissue regions, could be used as a feasible alternative to a reference AIF estimated from vena cava and left ventricle. In addition, the MLIF could be predicted using different combinations of input tissue time-activity curves, not necessarily including the myocardium wall, which is a blood-rich tissue that closely resembles the AIF. This could be relevant in studies where the heart is outside the PET field-of-view.

Two AIF prediction models were evaluated, based on GPs and an LSTM neural network, respectively. Mathematically, these two methods represent two vastly different approaches for solving the problem of AIF estimation. GPs is a non-parametric approach based on the conditioning of a Gaussian distributed prior, on the training data, where the prior distribution is specified by a kernel matrix. On the other hand, the LSTM model in this study was parameterized by roughly 2100 weights, each of which have to be updated iteratively during training.

The LSTM model provided slightly better agreement between predicted and reference AIF curve shapes, compared to GPs, although the result from compartment modelling showed similar performance between the models. In **Paper III**, the use of the LSTM method was recommended, due to the slightly lower errors in AIF prediction. One aspect, which was not discussed in the paper, was training time. With N training samples, the GPs approach relies on the inversion of an $N \times N$ covariance matrix (Equation 3.22). Although this operation may become challenging for very large training data sets (in the order of 1000 samples or more), for the data set sizes of the current thesis, this matrix inversion was computationally very fast. On the other hand, the LSTM method involves iterative calculation of gradients, and weight-updates, which was considerably more time-consuming (calculations were performed on a desktop CPU). Although no timing data was recorded in this work, typical training times for the leave-one-out cross validation approach of this paper was in the order of minutes for the GPs approach, and in the order of hours for the LSTM approach. In the experimental setup of the current work, this was not an issue, and therefore, this discussion was not emphasized in **Paper III**. However, at a later stage, for clinical implementation, computational time may be of significant interest, which could affect the choice of method for

5 Results and discussion

AIF estimation. This was also the motivation for choosing the GPs approach in **Paper IV**.

One advantage with pre-clinical PET imaging of mice is that the whole mouse body fits inside the field-of-view of the PET scanner. This allows dynamic imaging and subsequent extraction of time-activity curves for readily all organs. From a machine learning point of view, this means that the number of input features may be large, which can be an advantage, as discussed in Section 5.2. A wrapper method was implemented to evaluate different combinations of input features on AIF prediction. Not surprisingly, the myocardium showed to be important for accurate predictions. This region inevitably contains spill-in from the blood pool, thus inherently including a strong component that reflects the AIF.

Both the left ventricle and vena cava regions were used in the generation of the reference AIF, as there were no blood-sample-based AIFs available in this study. Therefore, they had to be excluded from the input features for model training. Obviously, the MLIF method must be validated with real blood-based AIFs in a future pre-clinical PET study. In this case, it is conceivable that the inclusion of the left ventricle and vena cava regions will contribute positively to AIF prediction, as the information in these regions are directly related to blood. The MLIF model will then represent the transformation that maps image-derived input data, essentially by performing partial volume correction, to a ready-to-use AIF.

Another interesting and related approach in future whole-body mouse AIF prediction studies, would be to increase input feature space even further. Hypothetically, if all mice were coregistered to a reference anatomy template, each individual voxel could represent an input feature. This approach would be highly sensitive to abnormal anatomy, positioning of the animal, as well as statistical noise in individual voxels, and would therefore not be practical. A potentially more feasible approach would be to utilize a *convolutional neural network* for feature extraction [202]. This method is based on multiple layers of trainable convolution filters and subsequent pooling of voxels, to reduce the dimensions of the input image to subsequent layers. Before the output layer, the reduced-size image is flattened to a one-dimensional feature vector. In the next step, the feature vector from each time step could be passed on to an LSTM network, to predict the AIF. With this method, manual segmentation of different tissues in the PET images would not be necessary, because the network effectively is trained to become a feature extractor by the convolutional layers. However, as the number of parameters of such a network increases even further, compared to the LSTM approach, it is hypothesized that a larger training dataset would be necessary.

5.4 Paper IV

The aim of **Paper IV** was to leverage the results from **Paper III** in order to investigate if the MLIF approach to AIF estimation was feasible in a human clinical brain-PET cohort for CBF measurements using the ^{15}O -water tracer. The reference AIF in this study was based on continuous arterial blood sampling, and thus representing a more realistic clinical situation, compared to **Paper III**. Also, the acetazolamide challenge was performed on each subject. This allowed to study, not only if CBF could be correctly estimated using an MLIF, but also if a clinically relevant CBF change could be detected, between baseline and acetazolamide scans.

The proposed, non-invasive, MLIF method showed potential to replace the AIF obtained from blood sampling for CBF measurements using ^{15}O -water PET and kinetic modelling. This could minimize the risk for complications and, at the same time, simplify the clinical dynamic PET imaging workflow. Three automatically segmented image-derived curves were used as input into the MLIF model. In short, the three input regions were generated, representing blood-uptake with different partial-volume contents around the carotid arteries (10 voxels are within the artery, 100 voxels mostly within artery, and 1000 voxels also sampling the partial-volume generated spill-out to pixels outside the artery). The median voxel-value was derived for each time-frame and region, resulting in three IDIF time-activity curves. The three regions were used in a similar GPs framework as the one proposed in **Paper III**, but here it was used for the prediction of a real blood AIF. It is not unlikely that the approach described in Section 5.3, using a convolutional neural network as feature extractor, could also work in this case, taking the rigid brain anatomy into consideration. This should be investigated in a future study.

It was further demonstrated that the MLIF approach could successfully predict clinically significant changes in CBF, induced from acetazolamide medication. One key finding, which is thoroughly discussed in **Paper IV**, is that a lower CBF change was observed for MLIF compared to AIF, when comparing baseline and acetazolamide scans. However, as the increase was significant compared to baseline, the MLIF method still shows a clinical potential to differentiate baseline and acetazolamide scans.

Another important observation in the study was that when comparing CBF measurements obtained from AIF to those obtained from MLIF, some outliers were found. At first, it was assumed that these outliers were due to an inaccurate prediction of the MLIF, and a subsequent error in the CBF value, following compartment modelling. However, at closer inspection, it seemed to be the AIF causing the CBF values to become outliers. This was realized when comparing the CBF values to

5 Results and discussion

population averages, not only for predicted, but also for the reference CBF values. Therefore, future methods should be developed to quality control the input data.

In this study a fixed dispersion constant of $\tau = 15$ s was assumed, as recommended in [48]. This dispersion constant explains both external dispersion in the measurement tubes, as well as internal dispersion in the vessels, from the artery sampling point to the brain region studied. External dispersion is approximately constant in all scans, as the detector tubes have a fixed length. However, internal dispersion may vary in the range between 4-6 s [48], and could therefore affect the results.

Individually fitted dispersion constants can be estimated from Equation 2.34 by performing an initial fit of the kinetic parameters and dispersion and delay constants, to the measured PET data. The AIFs can thereafter be corrected with the individual dispersion constant for each subject, before included in MLIF training. Individually fitted dispersion constants may introduce additional between-subject variations of the corrected AIF curves, not seen in the current data, and should therefore be evaluated in future studies, preferably using larger datasets.

Another limitations of this study was that two populations (healthy subjects and multiple sclerosis patients) and two scanners were present in the data, potentially introducing biases. Such variations in the input data could make the prediction models more robust, because they will learn natural variations of the data. However, considering the small size of the dataset, these variations could also result in noise and thus reduce the accuracy of the trained models. Therefore, in future studies, a homogeneous dataset should preferably be investigated.

The GPs variance in Equation 3.23 can be interpreted as a measure of whether a test sample under consideration is close or far away from the available training data points in input feature space. Thus, it could potentially be used to detect outliers. However, a high variance does not necessarily imply that the predicted AIF is incorrect, only that is based on uncertain grounds [157]. In the present study, no relationship was found between the CBF_{MLIF}/CBF_{AIF} ratios and the GPs variance. Further research is needed to develop methods for uncertainty estimation and possible outlier identification using, for instance, the GPs variance.

6 Conclusion

In this thesis, several challenges associated with quantitative PET imaging of humans and small-animals have been addressed. Methodologies have been developed that could potentially advance quantification in both static and dynamic PET imaging.

Attenuation correction related artifacts have been identified, that affect the quantitative accuracy in PET/MRI imaging of lung lesions (**Paper I**). Here, it was found that susceptibility artifacts, caused by metal objects, may cause clinically relevant variations in lung tumor SUV measurements, which might affect the accuracy in serial studies of the same patient. These findings warrant careful inspection of MRI-based attenuation correction maps, and subsequent correction of susceptibility artifacts, to ensure highest possible diagnostic accuracy of the PET/MR examination. This is also an important pre-processing step before including the imaging data in the training of machine-learning-based models.

Moreover, a set of motion-robust PET-based tumor features have been identified (**Paper II**), that can be extracted from simple free-breathing PET reconstructions, without employing gating or motion correction during image acquisition. This may considerably simplify and shorten clinical scan protocols, and reduce the need for patient compliance to breathing instructions, in imaging studies used for acquiring training data for machine learning models. It may also allow inclusion of retrospective imaging data, or data from other imaging centers, where respiratory gating or motion correction was not applied.

For quantification in dynamic PET studies, methodology for non-invasive estimation of the AIF have been developed, for both pre-clinical ^{18}F -FDG-PET (**Paper III**) as well as for clinical ^{15}O -water-PET applications (**Paper IV**). The proposed MLIF approach uses solely image-derived input data into the machine learning models, which in practice implies evading of arterial cannulation. This may considerably simplify quantification in small-animal dynamic PET imaging, as arterial cannulation in rodents is challenged by limited blood volume and a complex and terminal surgical procedure. Also, in human dynamic PET imaging, non-invasive AIF estimation may reduce pain and risk for complications, and significantly simplify the dynamic PET imaging workflow.

6 Conclusion

In conclusion, the four research works presented in this thesis have contributed to the field of quantitative PET imaging, mainly by addressing challenges related to imaging artifacts and subject motion in static PET quantification, and to challenges related to blood sampling required for tracer kinetic modelling in dynamic PET quantification.

7 Limitations and future directions

This chapter indicates some of the limitations with this thesis, and points out possible directions of future research in the continuation of this thesis.

In **Paper I**, the inspection of the attenuation correction maps was performed visually, and subsequent artifacts were corrected manually. Both these aspects could be automated. For instance, a machine-learning-based classifier can be trained to detect cases of susceptibility artifacts where the lung segmentation has failed. This could be implemented as a quality-assurance step in the imaging workflow to reduce the risk for attenuation-correction-related artifacts being undiscovered. If artifacts are detected, they could also be corrected automatically using deep convolutional neural networks [189, 190].

In **Paper II**, a set of features that were robust to the acquisition mode (free-breathing, end-expiration gating or motion correction) were derived. Because of the limited sample size, and that the end-points were not yet known, it was not possible to build predictive models to test the impact of the motion-robust features on disease or survival predictions. The clinical parent study, from which the data to this paper was extracted, is still ongoing, so when sufficient number of patients have been included, and when the clinical end-points have been reached, it will be possible to evaluate the potential of these motion-robust features with a machine learning model for tumor histology classification, or survival prediction.

In **Paper III** the use of an MLIF to predict the AIF was evaluated in a pre-clinical dynamic PET study using ^{18}F -FDG. The main limitation with the validation of the method was the lack of a blood-based AIF. In **Paper IV** the MLIF approach was evaluated in a human brain study using ^{15}O -water. However, the MLIF method remains to be evaluated for ^{18}F -FDG in mice. Also, it would be highly interesting to investigate if the MLIF approach could be adopted to other animal models and tracers, possibly also those requiring metabolite correction, by merely retraining the models. If validated correctly, this will give a foundation for a simplified MLIF-based approach to dynamic PET-based quantification in future clinical and research applications.

7.1 Explainable artificial intelligence


In **Paper III** and **Paper IV**, the predictive MLIF models were trained by using manually extracted input time-activity curves. As mentioned in Section 5.3, convolutional neural networks may work as trainable feature extractors that take an input image, and output a feature vector. Such an approach could be evaluated in future research, to allow for automatic feature extraction, thus, evading the need for manual segmentation of the input images [202].

Inevitably, more complex models requires larger amounts of training data. Moreover, as the model complexity increases, the rationale behind the predictions becomes harder to interpret, and consequently more difficult to trust, especially in the field of medicine and healthcare. There has been a growing expectation that machine learning models should generate, not only a *prediction*, but also an *explanation*. This has lead to a boost in the research field of *explainable artificial intelligence* in recent years, leading to the development of a vast number of methods for interpreting the predictions from deep neural networks [203–205]. Interpretability models may explain which features at the input layer are important for a specific prediction at the output layer. For instance, in case a convolutional neural network is used for AIF prediction, an interpretability approach could highlight which voxels in the input PET image are important for a given prediction at the output layer. This could also work as a quality assurance step for the AIF predictions in the sense that, if non-blood-related regions are highlighted, the predicted MLIF-curve may not be valid.

Clearly, interpretability models represent a highly interesting and relevant approach, and an obvious *next-step* in future machine learning research. This will allow to use machine learning, not only for classification or regression, but also for a deeper understanding of the learning aspects of the models. This may become highly relevant for presenting the machine-learning-based predictions to researchers, clinicians and patients in order to gain trust and belief in the automated decisions.

8 Paper I

Quantitative PET/MR imaging of lung cancer in the presence of artifacts in the MR-based attenuation correction maps

Acta Radiologica
0(0) 1–10
© The Foundation Acta Radiologica
2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0284185119848118
journals.sagepub.com/home/acr


Samuel Kuttner^{1,2,*} , Martin Lyngby Lassen^{3,5,*},
Silje Kjærnes Øen⁴, Rune Sundset^{1,2}, Thomas Beyer³ and
Live Eikenes⁴

Abstract

Background: Positron emission tomography (PET)/magnetic resonance (MR) imaging may become increasingly important for assessing tumor therapy response. A prerequisite for quantitative PET/MR imaging is reliable and repeatable MR-based attenuation correction (AC).

Purpose: To investigate the frequency and test–retest reproducibility of artifacts in MR-AC maps in a lung cancer patient cohort and to study the impact of artifact corrections on PET-based tumor quantification.

Material and Methods: Twenty-five lung cancer patients underwent single-day, test–retest, ¹⁸F-fluorodeoxyglucose (FDG) PET/MR imaging. The acquired MR-AC maps were inspected for truncation, susceptibility, and tissue inversion artifacts. An anatomy-based bone template and a PET-based estimation of truncated arms were employed, while susceptibility artifacts were corrected manually. We report the frequencies of artifacts and the relative difference (RD) on standardized uptake value (SUV) based quantification in PET images reconstructed with the corrected AC maps.

Results: Truncation artifacts were found in all 50 acquisitions (100%), while susceptibility and tissue inversion artifacts were observed in six (12%) and 26 (52%) of the scans, respectively. The RD in lung tumor SUV was < 5% from bone and truncation corrections, while up to 20% RD was introduced after susceptibility artifact correction, with large inconsistencies between test–retest scans.

Conclusion: The absence of bone and truncation artifacts have limited effect on the PET quantification of lung lesions. In contrast, susceptibility artifacts caused significant and inconsistent underestimations of the lung tumor SUVs, between test–retest scans. This may have clinical implications for patients undergoing serial imaging for tumor therapy response assessment.

Keywords

PET/MR, lung cancer, attenuation correction, artifacts

Date received: 20 November 2018; accepted: 13 April 2019

Introduction

Lung cancer is the most frequent cancer type and the leading cause of cancer-related death in the world (1). Positron emission tomography (PET)/computed tomography (CT) with ¹⁸F-fluorodeoxyglucose (FDG) is the standard of care today for lung cancer staging and is also increasingly used to aid in radiotherapy treatment planning and for tumor therapy response assessment (2–6).

The recent introduction of integrated PET/magnetic resonance (MR) systems has opened new possibilities

¹Nuclear Medicine and Radiation Biology Research Group, Department of Clinical Medicine, University of Tromsø - The Arctic University of Norway, Norway

²The PET Imaging Center, University Hospital of North Norway, Norway

³QIMP Team, Center for Medical Physics and Biomedical Engineering, Medical University of Vienna, Austria

⁴Department of Circulation and Medical Imaging, Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology, Norway

⁵Cedars-Sinai Medical Center, Los Angeles, California

*Equal contributors.

Corresponding author:

Samuel Kuttner, University of Tromsø, Tromsø 9019, Norway.

Email: samuel.kuttner@uit.no

for tumor characterization by adding excellent soft-tissue contrast, provided by MR imaging (MRI), to the functional information from PET. In this way, simultaneous, multiparametric images can be acquired, that facilitate precision medicine and personalized treatment of the disease (7,8). The use of non-ionizing MRI is another benefit for the use of PET/MR, rather than PET/CT, in the follow-up assessment of patients undergoing radiation or chemotherapy (9).

The reproducibility of the standardized uptake values (SUV) in PET/MR imaging is important for a reliable assessment of therapy response, and it is strongly linked to accurate MR-based attenuation correction (AC) during PET image reconstruction (7,10–13). In fully integrated PET/MR systems, standard AC maps are calculated from the segmentation of MR images into a number of tissue classes (e.g. air, lung, fat, and soft tissue), thereby assigning a specific attenuation coefficient to each tissue (7,14). A known limitation of many MR sequences used for generating AC maps is the low bone signal, which may cause an underestimation of quantitative measurements in the resulting PET images in close vicinity to osseous tissue (11). Furthermore, the reduced transaxial field of view (FOV) of the MR, in comparison with the PET FOV, has been reported to result in truncation artifacts of distal body parts (15). In addition, distortion of the magnetic field, caused by metallic implants, such as surgical sternal wires, may cause susceptibility artifacts in the AC maps (16,17). Lastly, soft-fat tissue inversions in the AC map may occur (16). These effects have been proven to affect both diagnostic quality and hinder accurate quantification of the tracer-distribution in brain, head/neck, and cardiac PET studies (16,17). Further, a recent study reported that MR-AC related artifacts occurred frequently and inconsistently in test–retest scans of lung cancer patients, leading to non-consistent SUV quantification in serial examinations (18).

The aim of the current study is to investigate the frequency and the test–retest reproducibility of artifacts observed in standard MR-based AC maps in a lung cancer patient cohort undergoing PET/MR imaging. Further, we intend to correct for artifacts in the AC maps and evaluate the impact of the corrections on the PET-based quantification in the test-retest setup.

Material and Methods

Ethical approval

This study was approved by the Norwegian Regional Committees for Medical and Health Research Ethics (REC reference 2017/915). All patients signed written informed consent.

Patient population, preparation, and imaging overview

Twenty-five lung cancer patients with a total of 26 lung lesions were included in this single-injection dual-time point PET/MR imaging study. Patients fasted 15 h \pm 4 h before the injection of 281 MBq \pm 41 MBq FDG. PET/MR assessment started 113 min \pm 10 min post FDG injection (Fig. 1).

PET/MR image acquisition

The PET/MR acquisitions were performed in a Siemens Biograph mMR (software version VB20P) (Siemens Healthineers, Erlangen, Germany) using a free-breathing and arms-down scan protocol during the entire simultaneous image acquisition. All patients had the same anatomical region scanned twice, without repositioning and reinjection between the scans, thus, allowing for the assessment of the test–retest reproducibility of the images. First, a 10-min, one-bed position scan, centered over the mediastinum, was performed, immediately followed by another 10-min, two-bed

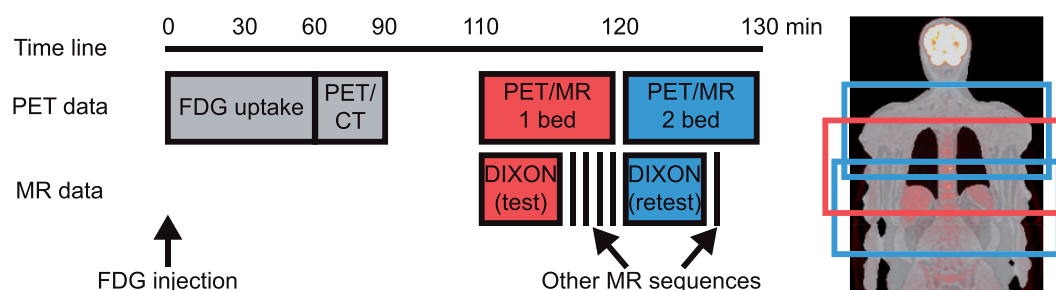


Fig. 1. Time line of the single-injection dual-time point imaging study. Patients were injected with FDG, followed by 60 min rest, before undergoing PET/CT and PET/MR imaging. PET and MR DIXON data that were included in this study are indicated in red and blue, for the one-bed position scan centered over the mediastinum (test) and the two-bed position scan of the whole thorax (retest), respectively. The corresponding anatomical regions are indicated approximately on the coronal overview scan to the right. FDG: F-fluorodeoxyglucose; PET: Positron emission tomography; CT: computer tomography; MR: magnetic resonance.

position scan of the whole thorax (Fig. 1). A standard DIXON-based MR-AC map was acquired for each scan, resulting in a total of 50 MR-AC maps for the 25 patients.

Inspection of the attenuation correction maps

All 50 MR-AC maps were inspected for artifacts by two experienced imaging physicists. We report the frequency of the observed artifacts from this test-retest assessment, divided into three categories: truncation artifacts; susceptibility artifacts in the sternum; and tissue inversion (Fig. 2a–c). For the second category, we specify the number of acquisitions for which lung segmentation in the MR-AC map failed.

Corrections of the attenuation correction maps

For each patient, up to four additional sets of test-retest AC maps were formed by correcting for the artifacts that were found in each original AC map (Fig. 2d–f): (i) missing bone was corrected by adding an anatomy-based bone template of the spine (19); (ii) truncation artifacts, if found, were corrected using the MLAA algorithm (20). Both these corrections were performed using vendor-provided software, thus, updating the AC maps in DICOM-format; (iii) susceptibility artifacts, causing misclassified

voxels in the sternum and lungs, if found, were corrected using an in-house developed algorithm implemented in Matlab (Mathworks, MA, USA) (17). This correction was applied on the interfiles used for reconstructions in a vendor-provided reconstruction tool (JSRecon (e7-tools), Siemens Healthineers, Erlangen, Germany); (iv) a fully corrected AC map was formed by applying all the above-mentioned corrections. The different AC maps are denoted $AC_{Original}$, AC_{Bone} , AC_{Trunc} , AC_{Susc} , and AC_{All} , respectively, throughout this paper. Thus, for each patient, a total of eight (no AC_{Susc}) or 10 (with AC_{Susc}) different AC maps were included for subsequent analyses.

PET image reconstruction

Static PET images were reconstructed for all patients by employing each of the generated AC map, to form $PET_{Original}$, PET_{Bone} , PET_{Trunc} , PET_{Susc} , and PET_{All} (Fig. 3). This resulted in a total of 214 PET image series included for further evaluation. All reconstructions were performed using ordered-subset expectation-maximization (OSEM) algorithm with three iterations, 21 subsets, and 4 mm Gaussian smoothing. The matrix size of the reconstructed PET images was $344 \times 344 \times 127$ and $344 \times 344 \times 224$ for test (one-bed) and retest (two-bed) scans, respectively, with a voxel size of $2.1 \times 2.1 \times 2.0$ mm.

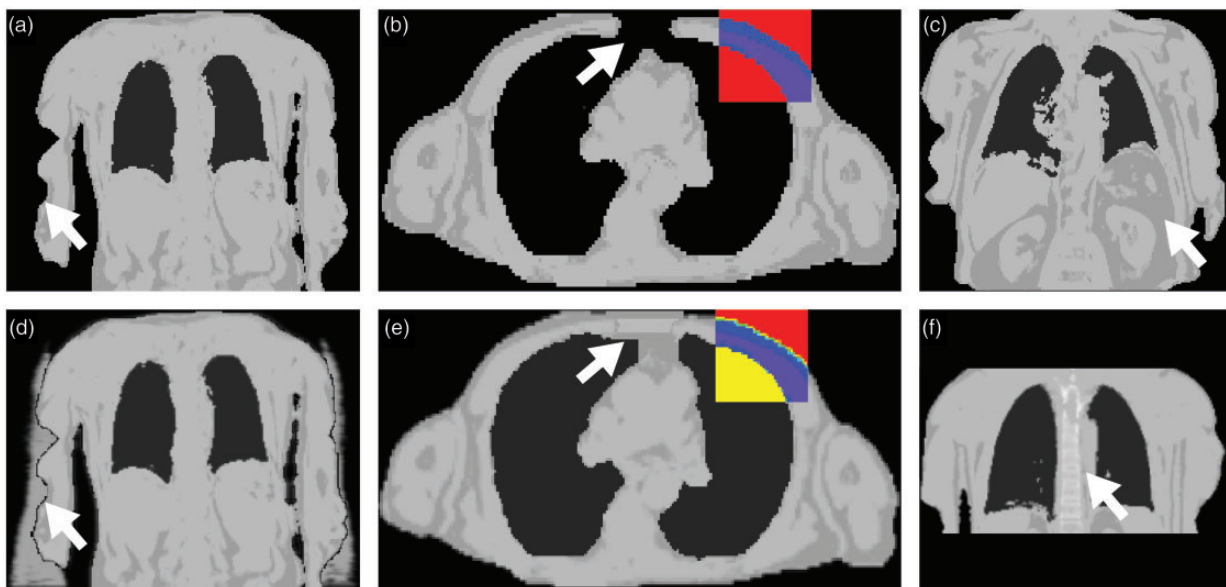


Fig. 2. Examples of the three types of artifacts reported in this study and their corrections in the MR-AC maps: (a) Truncation artifact, where the arms of the patient are outside the MR FOV. (b) Susceptibility artifact caused by surgical sternal wires. Note the failed lung segmentation, where lung attenuation values have been incorrectly assigned to background (red color in the box insert). (c) Tissue inversion artifact with a soft-fat tissue swap. This artifact was not corrected in this study. (d) Correction of truncated arms using the MLAA algorithm (20). (e) Susceptibility artifact correction using an in-house developed algorithm (17). Note that the lungs have been “filled” with lung attenuation values (yellow color in the box insert). (f) Correction of missing bone by adding an anatomy-based bone template of the spine (19). MR-AC: magnetic resonance attenuation correction; MR FOV: magnetic resonance field of view.

PET image evaluation

FDG-avid lesions were delineated using a 41% SUV_{max} threshold (21). For each lesion, the mean, maximum, and peak SUV values were measured (SUV_{mean} , SUV_{max} , and SUV_{peak}). Relative difference (RD) was calculated using Eq. 1:

$$RD = \left[\frac{SUV_{Corrected}}{SUV_{Original}} - 1 \right] \times 100\% \quad (1)$$

where $SUV_{Corrected}$ and $SUV_{Original}$ represent measurements obtained from the PET images reconstructed with the corrected and original AC maps, respectively (Fig. 3). Statistical significance was calculated using Wilcoxon signed-rank test, after square root transformation of the data, to correct for asymmetric distributions.

The impact of the corrections on tumor size was evaluated by measuring the largest diameter and the volume of the delineated lesions in all reconstructed images.

Results

Frequency of artifacts

Truncation artifacts were reported in all 50 MR-AC maps (100%), whereas susceptibility artifacts, caused by sternal wires, and tissue inversion was observed in six (12%) and 26 (52%) of the MR-AC maps,

respectively. All susceptibility artifacts were reproducible between test–retest scans. As a consequence of susceptibility artifacts in the sternum, lung tissue segmentation failed for two of the patients in the test scan (Table 1).

Tissue inversion artifacts were observed in a total of 19 (76%) patients, which was reproduced in the retest scan in seven (37%) of the patients. For 12 (63%) of the patients, tissue inversion artifacts were present in only one of either the test or retest scans (Table 1).

Correction of artifacts

The relative differences of the lung tumor SUVs after applying bone, truncation, and all corrections are shown in Fig. 4. Inclusion of the anatomy-based bone template caused an increased tumor SUV_{mean} , SUV_{max} , and $SUV_{peak} < 0.5\%$ ($P < 0.001$) in both test and retest scans, while correction of truncation and all artifacts yielded a corresponding increase $< 3\%$ ($P < 0.01$) (Fig. 4).

Table 2 shows the relative difference in the lung tumor SUV for the three patients with susceptibility artifacts in the AC maps after applying the different corrections. The susceptibility artifacts introduced an acquisition-varying effect on the tumor SUV values, with test–retest variations of up to 22% (Patient 21).

To visualize the effect of the different AC map corrections on the reconstructed PET images, relative difference images of a representative patient is shown in Fig. 5 and Supplementary Figures S1 and S2.

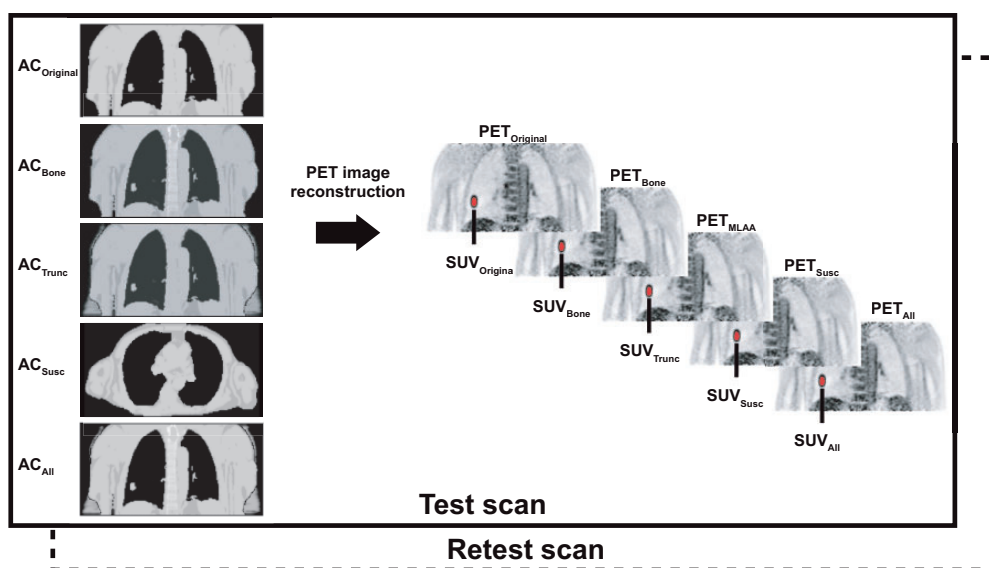


Fig. 3. Schematic overview of MR-AC map correction, PET image reconstruction, and subsequent quantitative SUV extraction. For each patient, up to five different sets of test–retest PET images were reconstructed, using the original AC map ($AC_{Original}$), and each of the corrected AC maps (AC_{Bone} , AC_{Trunc} , AC_{Susc} , and AC_{All}). From each PET image, the lung tumor SUV_{mean} , SUV_{max} , and SUV_{peak} was extracted from the delineated tumor (41% of SUV_{max} threshold). The analysis was performed for both test and retest scans. MR-AC: magnetic resonance attenuation correction; PET: Positron emission tomography; SUV: standardized uptake values; AC: attenuation correction.

Table 1. Frequency of artifacts observed among the test–retest AC maps in the dataset.

Patient no.	Truncation		Susceptibility		Failed lung seg.		Tissue inversion	
	Test	Retest	Test	Retest	Test	Retest	Test	Retest
1	X	X						
2	X	X					X	X
3	X	X						
4	X	X						
5	X	X						
6	X	X					X	X
7	X	X	X	X			X	
8	X	X					X	
9	X	X					X	
10	X	X						
11	X	X					X	
12	X	X					X	X
13	X	X					X	
14	X	X					X	
15	X	X					X	X
16	X	X					X	X
17	X	X					X	
18	X	X					X	
19	X	X					X	
20	X	X	X	X	X		X	X
21	X	X	X	X	X		X	
22	X	X					X	
23	X	X						
24	X	X					X	
25	X	X					X	X
Test–Retest sum	25 (100)	25 (100)	3 (12)	3 (12)	2 (8)	0 (0)	8 (32)	18 (72)
Total sum	50 (100)		6 (12)		2 (4)		26 (52)	
Test \cup Retest	25 (100)		3 (12)		2 (4)		19 (76)	
Test \cap Retest	25 (100)		3 (12)		0 (0)		7 (37)	
(Test \cup Retest) \ (Test \cap Retest)	0 (0)		0 (0)		2 (4)		12 (63)	

Values are presented as n (%).

Positive observations are indicated with X. AC: attenuation correction.

Fig. 6 shows the association between the SUV and lesion diameter, without and with corrections for artifacts in the AC maps. While the corrections affected the SUV values, they did not introduce alterations in the tumor diameters (Table 3). Similar results were obtained for the tumor volumes (Table 3 and Fig. 7).

Discussion

We have studied how individual and combined corrections of the three most frequent artifacts (absence of bone, truncation of arms, and susceptibility artifacts) in standard DIXON-MR-AC maps affect lesion quantification in a PET/MR imaging cohort of lung cancer patients. Our main finding is that artifacts in the MR-AC maps introduce variability in the lung lesion SUVs of up to 22% in a test–retest set-up without patient repositioning. Absence of bone is caused by

known limitations in the MR DIXON scan sequence and, therefore, is present in all scans (11). The impact of adding the anatomy-based bone template on the lung tumor SUVs was low ($< 0.5\%$) (Fig. 4) and the effect is limited to the close vicinity of the spine (Fig. 5). This agrees with earlier studies that showed only local effect close to the bones and no clinically relevant effects for lung lesions, when omitting bone from the AC map (14,22,23).

Truncation artifacts were observed in all AC maps. These artifacts arise from fixed geometrical limitations in the MR FOV and, thus, are not expected to affect test–retest examinations. This is also in agreement with an earlier study (17). We found that truncation artifacts introduced up to 3% relative differences in the lung tumor SUV values (Fig. 4) and even larger effects inside the truncated regions (Fig. 5), which is in accordance with previous studies (24,25). This result indicates that accurate truncation correction of the AC

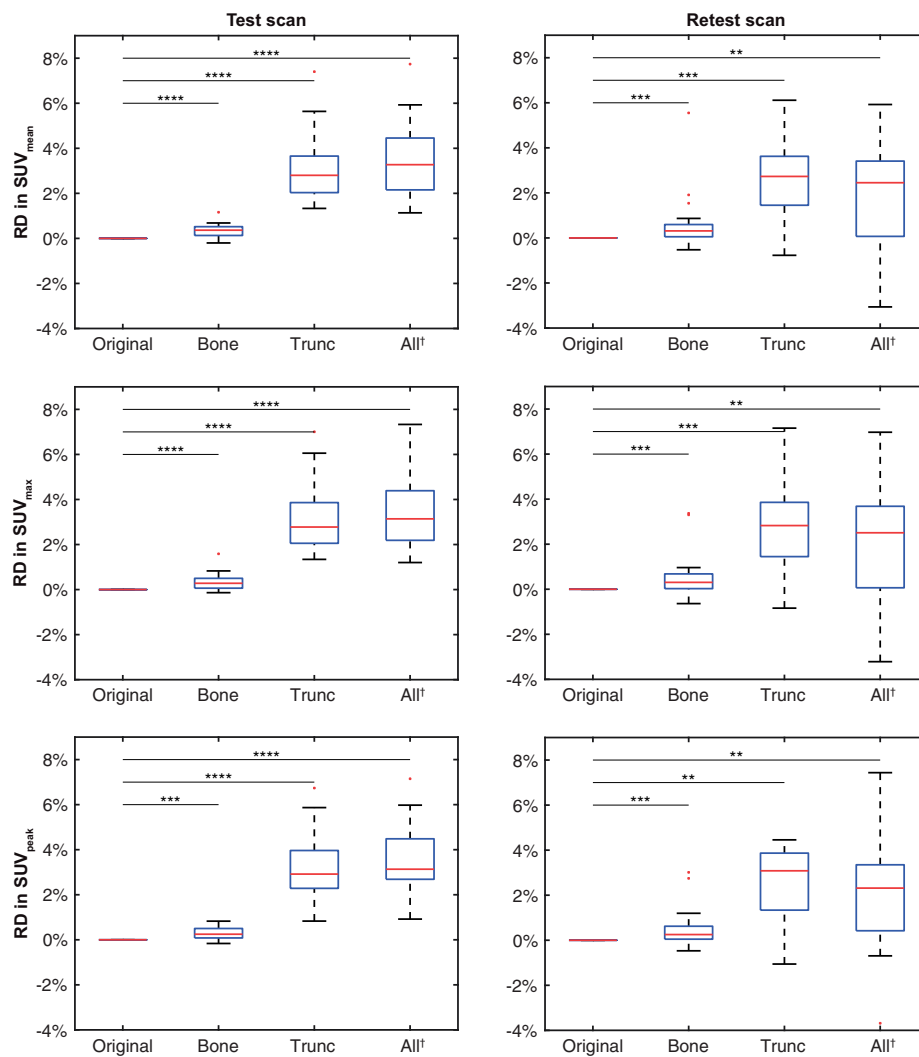


Fig. 4. Box plots showing the relative difference in lung lesion SUV_{mean} (top row), SUV_{max} (middle row), and SUV_{peak} (bottom row), obtained from the 41% SUV_{max} threshold, after correcting for bone, truncation, and all effects, compared to the original, uncorrected case, for test (left) and retest (right) scans, respectively. Asterisks (*) indicate significance level. †Corrections for susceptibility artifacts of the three patients from Table 2 are not included in this column. SUV: standardized uptake values.

Table 2. Relative difference in SUV before and after correction of susceptibility, and all artifacts, respectively, for the three patients with susceptibility artifacts in the AC maps.

Patient no.	Susceptibility artifacts corrected (%)						All artifacts corrected (%)					
	SUV_{mean}		SUV_{max}		SUV_{peak}		SUV_{mean}		SUV_{max}		SUV_{peak}	
	Test	Retest	Test	Retest	Test	Retest	Test	Retest	Test	Retest	Test	Retest
7	2.2	1.6	2.3	1.5	2.8	2.3	8.3	5.1	7.8	4.0	8.3	7.2
20	6.5*	1.2	7.3*	0.4	5.2*	0.3	10.6*	0.8	11.6*	0.2	9.4*	-0.2
21	20.0*	1.4	22.1*	1.4	16.1*	1.2	25.9*	4.4	28.7*	4.5	20.7*	3.8

*Values derived from acquisitions where lung segmentation failed and was subsequently corrected. SUV: standardized uptake values; AC: attenuation correction.

maps is not necessarily critical for clinically relevant lung tumor quantification (6).

The frequency of susceptibility artifacts was reproducible between test–retest scans (Table 1). This agrees

with a previous study focusing on artifacts in MR-AC maps for cardiac PET/MR applications (17). However, in the present study we report a reduced frequency of susceptibility artifacts (12%), compared to the previous

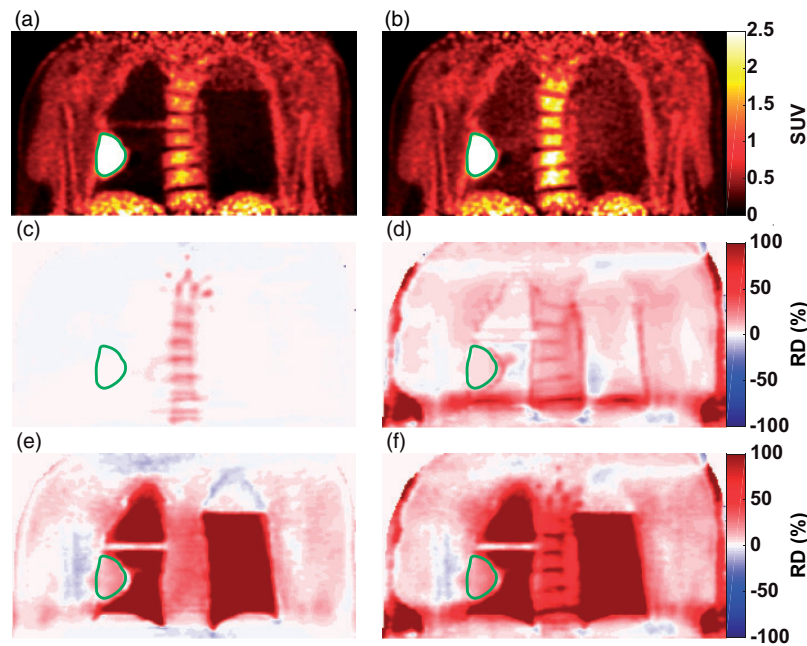


Fig. 5. The effect of the different corrections in the AC maps on the reconstructed PET images for a representative coronal slice of patient 21. (a) The original PET image. (b) The fully corrected PET image. (c–f) RD images between original and corrected PET images, where bone correction (c), truncation correction (d), susceptibility correction (e), and all corrections (f), have been applied, respectively. A schematic circumference of the lung lesion for this patient has been indicated with a green line. Of note, the horizontal stripe in the central part of the right lung and the top part of the left lung was caused by transitions between correct and incorrect lung segmentation regions in the MR-AC map. AC: attenuation correction; PET: Positron emission tomography; RD: relative difference; MR-AC: magnetic resonance attenuation correction.

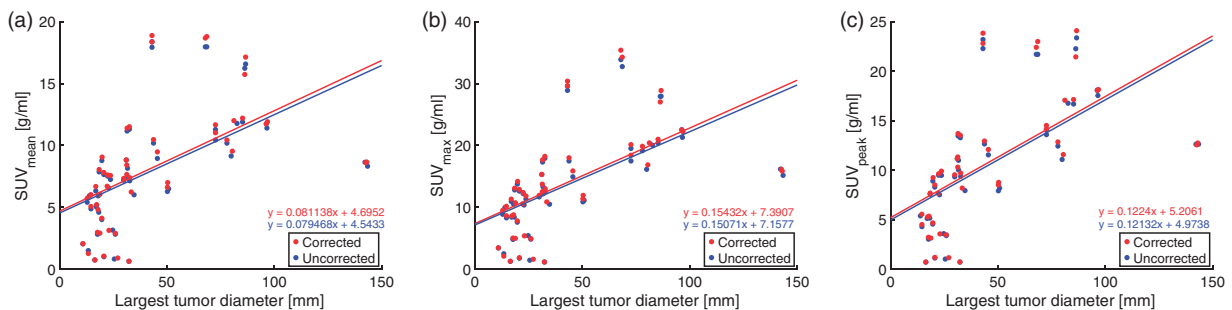


Fig. 6. Lesion diameter without and with all corrections for artifacts in the AC maps. Susceptibility artifact corrections of the three patients from Table 2 are not included here. The corrections did not introduce any relevant alterations in the tumor diameters when using SUV_{mean} (a), SUV_{max} (b), nor SUV_{peak} (c) assessments. Note that because SUV_{peak} is not defined for lesions with volume $< 1 \text{ cm}^3$, 5 data points were excluded from (c). AC: attenuation correction; SUV: standardized uptake values.

study (25%). This is due to the fact that cardiac patients have a higher frequency of sternal wires following from cardiac surgeries, compared to lung cancer patients. For two of the acquisitions in our study (Patients 20 and 21, test scan), the susceptibility artifacts in the sternum caused misclassified voxels in parts of the lungs, due to failure of the lung segmentation algorithm. For these scans, the artifact correction, and subsequent filling of the lungs with correct attenuation values, had a large impact ($>100\%$) on the resulting

PET SUV values in the lungs (Fig. 5 and Supplementary Figure S2) and on the corresponding lung tumor SUVs ((-5.2) – (-22.1%)) (Table 2). For acquisitions with susceptibility artifacts where the lung segmentation was successful, the impact of the corrections on the SUVs were lower ((-0.3) – (-2.8%)) (Table 2 and Supplementary Figure S1). This demonstrates that the accuracy in the attenuation values of the tissue surrounding a lung lesion is important for accurate tumor quantification, as also reported

Table 3. Tumor diameter and volume without and with corrections for artifacts in the AC maps^a. The corrections did not introduce any relevant alterations in the tumor diameters or volumes.

	Original	Absolute difference		
		Bone	Trunc	Bone + Trunc
Volume ≤ 10 cm³				
Diameter (mm)	22.5 \pm 7.0	0.0 \pm 0.1 ($P = 0.10$)	0.1 \pm 0.5 ($P = 0.45$)	0.0 \pm 0.6 ($P = 0.98$)
Volume (cm ³)	2.8 \pm 2.0	0.0 \pm 0.0 ($P = 0.01$)	0.0 \pm 0.0 ($P = 0.19$)	0.0 \pm 0.1 ($P = 0.60$)
Volume > 10 cm³				
Diameter (mm)	76.7 \pm 28.8	0.0 \pm 0.0 ($P = 0.33$)	0.1 \pm 0.3 ($P = 0.07$)	0.1 \pm 0.4 ($P = 0.55$)
Volume (cm ³)	54.6 \pm 47.0	0.3 \pm 0.8 ($P = 0.10$)	0.6 \pm 1.1 ($P = 0.02$)	0.0 \pm 1.2 ($P = 0.87$)
All volumes				
Diameter (mm)	43.3 \pm 32.4	0.0 \pm 0.1 ($P = 0.07$)	0.1 \pm 0.4 ($P = 0.13$)	0.0 \pm 0.5 ($P = 0.75$)
Volume (cm ³)	22.7 \pm 38.4	0.1 \pm 0.5 ($P = 0.09$)	0.3 \pm 0.8 ($P = 0.02$)	0.0 \pm 0.7 ($P = 0.82$)

^aSusceptibility artifact corrections of the three patients from Table 2 are not included here. AC: attenuation correction.

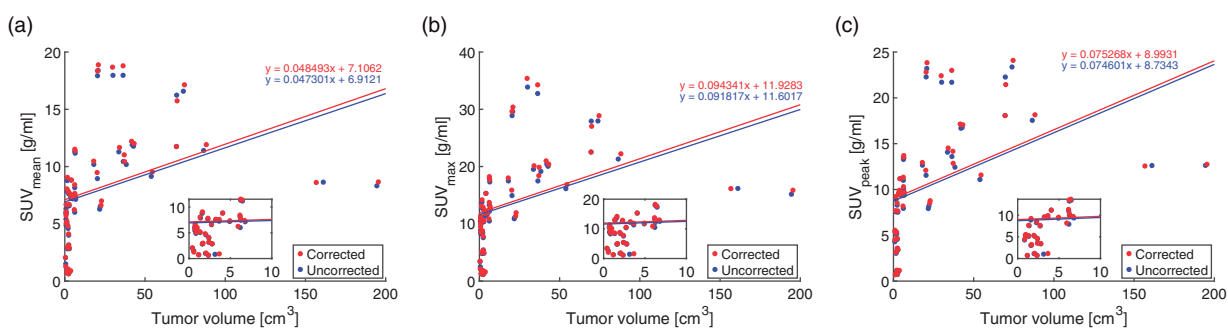


Fig. 7. Tumor volume without and with all corrections for artifacts in the AC maps. Susceptibility artifact corrections of the three patients from Table 2 are not included here. Data points with volumes < 10 cm³ are shown in the insert. The corrections did not introduce any relevant alterations in the lesion volumes when using SUV_{mean} (a), SUV_{max} (b), nor SUV_{peak} (c) assessments. Note that because SUV_{peak} is not defined for lesions with volume < 1 cm³, 5 data points were excluded from (c). AC: attenuation correction. SUV: standardized uptake values.

in an earlier study (19). Furthermore, the importance of correct attenuation values in the sternum, for lesion quantification, was shown in another recent study (26).

The combined effect of the applied corrections for the absence of bone and truncation had only limited effect ($RD < 3\%$) on the lung tumors (Fig. 4). However, correcting for susceptibility artifacts, introduced by sternal wires, caused a RD in the lesions of up 28.7% (Table 2), a clinically relevant measure (6), which may affect the treatment assessment of the patient, as the effects from these artifacts were not reproducible between test–retest scans.

We report tissue inversion artifacts in 26/50 (52%) acquisitions, with low reproducibility (37%) between test–retest scans (Table 1). Our incidence rate is higher than previously reported for other disease groups (16,17). Tissue inversion artifacts in the AC maps may introduce up to 35% quantification errors in the affected areas of the PET image (16). In the current study, tissue inversion was observed exclusively in the abdomen of the patient. This explains the higher

frequency of tissue inversion in retest scans (two-bed), compared to test scans (one-bed). Because all lung lesions were far away from the affected areas, the impact on tumor quantification was anticipated to be low; therefore, no tissue inversion correction was employed in the current study.

In response assessment studies, both lesion size and SUV can be used as measures of therapy response, where one or both may decrease after successful therapy (6). In this study, it was found that the AC map corrections had no impact on the measured lesion diameter or volume (Table 3, Figs. 6 and 7), thus, lesion size-based assessment of therapy response is independent of AC map artifact corrections. Furthermore, the increase in SUV values due to the corrections were independent of tumor size (Figs. 6 and 7); however, because the SUV values were increased by the corrections (Table 2 and Figs. 4, 6, and 7 and Table 2), careful inspection and correction of the AC maps, before PET image reconstruction, is still critical in SUV-based response assessment studies.

Our study has several limitations. Bone artifacts were corrected using an anatomy-based bone template (19), which does not model the ribs. However, the effect of the ribs is expected to be non-significant, due to their breathing-induced shifting during acquisition and the fact that they are small and mainly consist of trabecular bone (27).

Further, susceptibility artifacts in the MR-AC maps were corrected with a simple, manual method in this study. More advanced and fully automatic correction methods have been proposed earlier (28,29). However, with the introduction of novel artifact reducing MR sequences (e.g. MAVRIC, VAT, WARP, etc.), susceptibility artifacts could be minimized already at acquisition, thus potentially eliminating the need for manual retrospective correction of the AC maps. Unfortunately, such sequences are currently time-consuming, which hampers the implementation into clinical routine (30–33).

While the PET/MR system evaluated in the current study used a DIXON-based sequence for AC, other systems, using non-DIXON-based sequences, may cause other types of errors. However, systems using, for example, T1-weighted fast gradient-echo sequences, are expected to be prone to similar types of artefacts as the DIXON-based methods.

Another limitation was that different FOVs were used between test and retest scans. However, this is assumed to have minimal effect on the resulting SUV_{mean} and SUV_{peak} , because of their robustness to noise, and because the same lung segmentation algorithm was applied in both one-bed and two-bed scans (14,34).

In addition, we did not investigate the effect of respiratory motion on the MR-AC maps, which is known to introduce large biases in thoracic PET imaging (35–37).

Lastly, there was a relatively small number of patients included in this study.

In conclusion, the majority of artifacts found in the MR-AC maps, in this clinical PET/MR lung cancer study, did not affect the quantitative accuracy of SUV or tumor size measurements. However, susceptibility artifacts may cause clinically relevant variations in lung tumor SUV measurements, which might affect the accuracy in serial studies of the same patient. Therefore, our findings warrant careful inspection of the MR-AC maps and subsequent correction of susceptibility artifacts, to ensure the highest possible diagnostic accuracy of the PET/MR examination.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Northern Norway Regional Health Authority (grant no. HNF1349-17) and the Central Norway Regional Health Authority (grant no. 46056912).

ORCID iD

Samuel Kuttner  <https://orcid.org/0000-0001-7747-9003>

References

1. Pasic A, Postmus PE, Sutedja TG. What is early lung cancer? A review of the literature. *Lung Cancer* 2004;45:267–277.
2. Van Elmpt W, Zegers CML, Das M, et al. Imaging techniques for tumour delineation and heterogeneity quantification of lung cancer: Overview of current possibilities. *J Thorac Dis* 2014;6:319–327.
3. Sauter AW, Schwenzer N, Divine MR, et al. Image-derived biomarkers and multimodal imaging strategies for lung cancer management. *Eur J Nucl Med Mol Imaging* 2015;42:634–643.
4. Ladrón De Guevara Hernández D. The role of PET/CT imaging in lung cancer. *J Cancer Ther* 2015;6:690–700.
5. Greenspan SB. Role of PET/CT for precision medicine in lung cancer: perspective of the Society of Nuclear Medicine and Molecular Imaging. *Transl Lung Cancer Res* 2017;6:617–620.
6. Pinker K, Riedl C, Weber WA. Evaluating tumor response with FDG PET: updates on PERCIST, comparison with EORTC criteria and clues to future developments. *Eur J Nucl Med Mol Imaging* 2017;44:55–66.
7. Bezrukov I, Mantlik F, Schmidt H, et al. MR-based PET attenuation correction for PET/MR imaging. *Semin Nucl Med* 2013;43:45–59.
8. Huo E, Wilson DM, Eisenmenger L, et al. The role of PET/MR imaging in precision medicine. *PET Clin* 2017;12:489–501.
9. Lee SM, Goo JM, Park CM, et al. Preoperative staging of non-small cell lung cancer: prospective comparison of PET/MR and PET/CT. *Eur Radiol* 2016;26:3850–3857.
10. Keller SH, Holm S, Hansen AE, et al. Image artifacts from MR-based attenuation correction in clinical, whole-body PET/MRI. *Magn Reson Mater Physics, Biol Med* 2013;26:173–181.
11. Bezrukov I, Schmidt H, Mantlik F, et al. MR-based attenuation correction methods for improved PET quantification in lesions within bone and susceptibility artifact regions. *J Nucl Med* 2013;54:1768–1774.
12. Wagenknecht G, Kaiser HJ, Mottaghy FM, et al. MRI for attenuation correction in PET: Methods and challenges. *Magn Reson Mater Physics, Biol Med* 2013;26:99–113.
13. Beyer T, Lassen ML, Boellaard R, et al. Investigating the state-of-the-art in whole-body MR-based attenuation correction: an intra-individual, inter-system, inventory

- study on three clinical PET/MR systems. *Magn Reson Mater Physics, Biol Med* 2016;29:75–87.
14. Martinez-Moller A, Souvatzoglou M, Delso G, et al. Tissue classification as a potential approach for attenuation correction in whole-body PET/MRI: evaluation with PET/CT data. *J Nucl Med* 2009;50:520–526.
 15. Schramm G, Langner J, Hofheinz F, et al. Influence and compensation of truncation artifacts in MR-based attenuation correction in PET/MR. *IEEE Trans Med Imaging* 2013;32:2056–2063.
 16. Ladefoged CN, Hansen AE, Keller SH, et al. Impact of incorrect tissue classification in Dixon-based MR-AC: fat-water tissue inversion. *EJNMMI Phys* 2014;1:101.
 17. Lassen ML, Rasul S, Beitzke D, et al. Assessment of attenuation correction for myocardial PET imaging using combined PET/MRI. *J Nucl Cardiol* 2017. doi: 10.1007/s12350-017-1118-2.
 18. Olin A, Ladefoged CN, Langer NH, et al. Reproducibility of MR-based attenuation maps in PET/MRI and the impact on PET quantification in lung cancer. *J Nucl Med* 2018;59:999–1004.
 19. Paulus DH, Quick HH, Geppert C, et al. Whole-body PET/MR imaging: quantitative evaluation of a novel model-based MR attenuation correction method including bone. *J Nucl Med* 2015;56:1061–1066.
 20. Nuyts J, Bal G, Kehren F, et al. Completion of a truncated attenuation image from the attenuated PET emission data. *IEEE Trans Med Imaging* 2013;32:237–246.
 21. Boellaard R, Delgado-Bolton R, Oyen WJG, et al. FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0. *Eur J Nucl Med Mol Imaging* 2015;42:328–354.
 22. Samarin A, Burger C, Wollenweber SD, et al. PET/MR imaging of bone lesions - Implications for PET quantification from imperfect attenuation correction. *Eur J Nucl Med Mol Imaging* 2012;39:1154–1160.
 23. Aznar MC, Sersar R, Saabye J, et al. Whole-body PET/MRI: The effect of bone attenuation during MR-based attenuation correction in oncology imaging. *Eur J Radiol* 2014;83:1177–1183.
 24. Beyer T, Bockisch A, Kühl H, et al. Whole-body 18F-FDG PET/CT in the presence of truncation artifacts. *J Nucl Med* 2006;47:91–99.
 25. Mawlawi O, Erasmus JJ, Pan T, et al. Truncation artifact on PET/CT: Impact on measurements of activity concentration and assessment of a correction algorithm. *Am J Roentgenol* 2006;186:1458–1467.
 26. Sviriydenka H, Delso G, De Galiza Barbosa F, et al. The effect of susceptibility artifacts related to metallic implants on adjacent-lesion assessment in simultaneous TOF PET/MR. *J Nucl Med* 2017;58:1167–1173.
 27. Lassen ML, Hacker M, Rausch I, et al. Standard MR-based attenuation correction does not cause significant uptake errors in the myocardium. Poster presented at the Annual Congress of the European Association of Nuclear Medicine (EANM); 2014 Oct 18–22; Gothenburg, Sweden.
 28. Schramm G, Maus J, Hofheinz F, et al. Evaluation and automatic correction of metal-implant-induced artifacts in MR-based attenuation correction in whole-body PET/MR imaging. *Phys Med Biol* 2014;59:2713–2726.
 29. Ladefoged CN, Andersen FL, Keller SH, et al. Automatic correction of dental artifacts in PET/MRI. *J Med Imaging* 2015;2:024009.
 30. Sutter R, Ulbrich EJ, Jellus V, et al. Reduction of metal artifacts in patients with total hip arthroplasty with slice-encoding metal artifact correction and view-angle tilting MR imaging. *Radiology* 2012;265:204–214.
 31. Talbot BS, Weinberg EP. MR Imaging with metal-suppression sequences for evaluation of total joint arthroplasty. *Radiographics* 2015;36:209–225.
 32. Dillenseger JP, Moliere S, Choquet P, et al. An illustrative review to understand and manage metal-induced artifacts in musculoskeletal MRI: a primer and updates. *Skeletal Radiol* 2016;45:677–688.
 33. Jungmann PM, Agten CA, Pfirrmann CW, et al. Advances in MRI around metal. *J Magn Reson Imaging* 2017;46:972–991.
 34. Lodge MA, Chaudhry MA, Wahl RL. Noise Considerations for PET quantification using maximum and peak standardized uptake value. *J Nucl Med* 2012;53:1041–1047.
 35. Werner MK, Parker JA, Kolodny GM, et al. Respiratory gating enhances imaging of pulmonary nodules and measurement of tracer uptake in FDG PET/CT. *Am J Roentgenol* 2009;193:1640–1645.
 36. Liu C, Pierce II LA, Alessio AM, et al. The impact of respiratory motion on tumor quantification and delineation in static PET/CT imaging. *Phys Med Biol* 2009;54:7345–7362.
 37. Apostolova I, Wiemker R, Paulus T, et al. Combined correction of recovery effect and motion blur for SUV quantification of solitary pulmonary nodules in FDG PET/CT. *Eur Radiol* 2010;20:1868–1877.

9 Paper II

Motion-robust radiomic features for image classification in ^{18}F -FDG PET/MRI imaging of lung cancer

Samuel Kuttner^{1,2,3*}, Erna-Elise Paulsen^{4,5}, Robert Jenssen², Rune Sundset^{1,3}, Jan Axelsson⁶

¹Nuclear Medicine and Radiation Biology Research Group, Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway.

²UiT Machine Learning Group, Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway.

³The PET Imaging Center, University Hospital of North Norway, Tromsø, Norway.

⁴Translational Cancer Research Group, Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway.

⁵Department of pulmonology, University Hospital of North Norway, Tromsø, Norway.

⁶Department of Radiation Sciences, Umeå University, Umeå, Sweden.

*Corresponding and first author:

Samuel Kuttner, Nuclear Medicine and Radiation Biology Research Group, Department of Clinical Medicine, UiT The Arctic University of Norway, 9037 Tromsø, Norway,

Phone: +47-77 66 99 53, e-mail: samuel.kuttner@uit.no

Abstract

This study investigates motion variability of radiomic features in positron emission tomography (PET) images of lung tumors. This resulted in a list of motion-invariant features which can be extracted from free-breathing PET examinations. Furthermore, a condensed list of features useful to build robust predictive models is suggested.

Radiomic features are becoming increasingly important in non-invasive, machine-learning-based models for prediction of disease state or survival. In lung cancer PET imaging, breathing-induced motion introduces variability in the radiomic features, which reduce the accuracy of a prediction model.

Eighteen non-small-cell lung cancer patients underwent a 20 minute, single-bed, PET/magnetic resonance imaging (MRI) acquisition in free-breathing. Three PET images were reconstructed for each patient: 1) Free-breathing PET (PET_{100%}), 2) end-expiration PET (PET_{40%}) and 3) MRI-based motion-corrected PET (PET_{MoCo}). In each reconstruction, the primary tumor was segmented using a threshold method, and 834 standardized radiomic features were extracted from each segmentation. Motion robust features were selected by comparing feature values from PET_{100%} to the corresponding values in PET_{40%} and PET_{MoCo}. The following three criteria were used to select motion robust features: relative difference less than 5%, P-value larger than 0.05 and concordance correlation coefficient larger than 0.95. A condensed list of features with low correlation was created from the motion-robust features, based on a correlation-matrix threshold of 0.95.

We found 43 of the 834 features to be motion robust. Fourteen of these had low correlation, according to the correlation-matrix threshold, and may be used as one possible feature-set for predictive model training. PET_{MoCo} features showed lower differences relative PET_{100%}, than did PET_{40%}.

In conclusion, we have identified 43 motion-robust lung-tumor PET features which can be extracted from free-breathing PET acquisitions, thus omitting the need for gating and motion correction.

Keywords

PET, Radiomics, features, variability, motion correction, lung cancer

Introduction

Lung cancer is the most frequent cancer type and the world-wide leading cause of cancer-related death (Wild et al. 2020). Medical imaging with ^{18}F -fluorodeoxyglucose (FDG) positron emission tomography (PET) plays an important role in the staging of the disease, and visual review and standardized uptake value (SUV)-based metrics are commonly used for clinical lung tumor characterization (Sauter et al. 2015; Ladrón De Guevara Hernández 2015; Greenspan et al. 2017; Pinker et al. 2017). Numerous studies have investigated the use of textural features, so-called radiomics, as prognostic biomarkers (He et al. 2018; S. S. Yip et al. 2017; Wu et al. 2016; Lovinfosse et al. 2016; Aerts et al. 2014; Cook et al. 2013). Radiomics refers to the process of extracting a large number of mathematical features from medical images, and thus, with a larger number of descriptors than with just SUV, characterize a tumor non-invasively.

Multiple radiomic features can be combined into predictive models for disease state or survival. Examples of traditional algorithms are Cox's regression and Kaplan-Meier analysis (Simon et al. 2011; Bianconi, Palumbo, Spanu, et al. 2020), while more recently, the use of machine-learning-based classifiers, for instance deep neural networks, have been introduced (Ahn et al. 2019; Avanzo et al. 2020). Radiomic features extracted from lung-tumor PET images have been used to build prognostic models for histological subtypes (Bianconi, Palumbo, Fravolini, et al. 2019), EGFR mutation status (S. S. Yip et al. 2017), ALK gene expression (Yoon et al. 2015), survival (Tixier et al. 2014; Pyka et al. 2015; Ohri et al. 2016; Desseroit et al. 2016; Hatt et al. 2018; Kirienko et al. 2018; Oikonomou et al. 2018), local control (Dissaux et al. 2020) and recurrence (Oikonomou et al. 2018; Ahn et al. 2019). A requirement of these models is that the number of features must be smaller than the number of patients to avoid overfitting (Simon et al. 2011). With a large number of features, the number of patients is normally the limitation in clinical studies. Therefore a selection of the most relevant features must be performed (Blum et al. 1997; Urbanowicz et al. 2018). For instance, correlated features add minor value to the models and may be removed (Wu et al. 2016).

To build robust prognostic models, radiomic biomarkers should have low variability to factors which are unrelated to the disease. Therefore, the stability of radiomic feature to various effects have been extensively studied in recent years. For example, it has been shown that features extracted from PET images are affected by the scanner type (Fried et al. 2016), reconstruction settings (Altazi et al. 2017; Shiri et al. 2017; Velden et al. 2016), the tumor segmentation method (Altazi et al. 2017; Velden et al. 2016), and the gray scale discretization method (Leijenaar et al. 2015) used. In addition, one of the main challenges in thoracic PET is image blurring due to respiratory motion. Previous studies have shown that breathing motion has a large impact on the extracted SUV values, the metabolic tumor volumes, as well as on the radiomic features (Oliver, Budzevich, Zhang, et al. 2015; S. Yip et al. 2014).

To reduce breathing-induced motion-blur, end-expiration-based PET is commonly utilized, where only PET data from the stationary phase of the breathing cycle are used during image reconstruction. Thus, compared to free-breathing PET, it reduces motion-induced blurring, enabling PET-based features to be better resolved (S. Yip et al. 2014). On the other hand, end-expiration

PET has an optimal duty cycle in the order of 35% (Van Elmpt et al. 2011) and thus have an increased noise level compared to free-breathing PET, which might increase variability of many radiomic features (Oliver, Budzevich, Hunt, et al. 2017; S. Yip et al. 2014).

The recent introduction of integrated PET/magnetic resonance imaging (MRI) systems has opened new possibilities for tumor characterization, not only by providing excellent soft-tissue contrast by MRI, but also by allowing image-based motion correction of the PET data. With this technique the motion vector fields are calculated over the phases of the breathing cycle, and used for motion correction of the simultaneously acquired PET data. In this way, a stationary and motion corrected PET image can be obtained without rejection of acquired PET events, thus allowing for motion corrected images with similar noise level as free-breathing PET (Furst et al. 2015; Grimm et al. 2015). Furthermore, breathing induced motion introduces mismatch between the attenuation-correction map and PET image data. This commonly introduces artifacts in free-breathing and end-expiration PET images (Keller et al. 2013). With MRI-based motion correction, the attenuation-correction map can be transformed into the correct breathing phase using the motion vector-fields, thus reducing the image artefacts in the reconstructed PET images (Furst et al. 2015). To the best of our knowledge, no study has, to date, investigated the impact of MRI based motion correction on radiomic features.

In this study we investigate the motion variability of radiomic features in a non-small-cell lung cancer PET cohort. Our aim was to generate a list of motion-invariant features that have the potential to be used in predictive models for disease state or survival prediction. These features may be extracted from free-breathing, end-expiration or motion corrected PET images without introducing reconstruction-bias in the feature values. Specifically, the free-breathing PET scan protocol is less complex and time-consuming to perform, compared to the motion correction sequences. Thus, by employing motion-robust features in machine learning prediction models, end-expiration gating and motion correction acquisitions are not needed, which simplifies the PET workflow for patients. Furthermore, hybrid PET/MRI is still an emerging modality, thus limiting the use of MRI-based motion correction at all PET imaging centers.

Materials and Methods

Ethical approval

This study was approved by the Norwegian Regional Committees for Medical and Health Research Ethics (REC reference 2017/1952). All patients signed written informed consent.

Patient population and imaging preparations

Eighteen patients with histologically confirmed non-small-cell lung cancer were included in this prospective PET/MRI study. Patients fasted $15\text{h} \pm 6\text{h}$ before the injection of $277\text{ MBq} \pm 66\text{ MBq}$ FDG. PET/MRI scanning started 2 hours post injection. The patient characteristics are summarized in Table 1.

PET/MRI acquisition

The PET/MRI acquisitions were performed on a 3T integrated PET/MRI system (Biograph mMR; Siemens Healthineers; software version E11). Before scanning, patients were positioned on the scanner bed with the arms along the torso. MRI acquisitions were performed employing a flexible 6-element surface coil (Body Matrix Coil; Siemens Healthineers) positioned over the thorax. Respiratory phase was monitored during the scan with a respiratory cushion between the surface coil and the thorax, positioned between the costal arch and sternum. This region was then centered in the field-of-view of the scanner. A 20-minute list-mode PET acquisition was conducted in free breathing for a single bed-position. MRI scanning was performed simultaneously with PET acquisition. Standard DIXON-based attenuation-correction maps with 4 tissue classes (air, fat, lung tissue, soft tissue) were acquired in end-expiration breath-hold for each scan, including a dedicated scan to account for arms outside the field-of-view (Blumhagen et al. 2014). A bone-model template was added to the segmented attenuation-correction maps (Kuttner et al. 2020).

A self-gated T1-weighted radial "stack-of-stars" spoiled 3-dimensional gradient echo sequence with fat suppression (StarVIBE, Siemens Healthineers) was used to acquire MRI images of five respiratory phases, followed by calculation of corresponding motion vector fields that allowed non-rigid registration of each phase to the end-expiration reference phase (Furst et al. 2015; Grimm et al. 2015).

Image reconstruction, gating and motion correction

Static PET images ($PET_{100\%}$) were reconstructed from the entire 20 minute PET scan (Figure 1A), by employing the end-expiration attenuation-correction map.

End-expiration PET images ($PET_{40\%}$) were created from the 20 minute list-mode data by using 40% of the breathing phase. The end-expiration phase was set to be from 30% - 70% of the breathing cycle, as measured from peak-inspiration with the respiratory cushion (Figure 1B). $PET_{40\%}$ was reconstructed using the end-expiration attenuation-correction map.

Motion corrected PET images (PET_{MoCo}) were reconstructed by binning the 20 minute list-mode PET events into sinograms for five respiratory phases corresponding to the phases for the MRI-derived motion vector fields. The five motion vector fields were transformed into the sinogram domain and applied to each PET sinogram to result in five motion-compensated PET sinograms in end-expiration (reference phase). These five sinograms were then summed and reconstructed into the motion corrected PET_{MoCo} image, which resulted in 100% duty cycle (Figure 1C).

In the PET_{MoCo} workflow, a visual comparison between end-expiration and end-inspiration StarVIBE images and the acquired attenuation-correction map is performed. For PET_{MoCo} reconstructions, attenuation-correction maps in end-expiration (reference phase) are required. In some patients, where the attenuation-correction maps were found to be in end-inspiration, the attenuation-correction maps were transformed into end-expiration using the MRI-derived motion vector fields, prior to PET image reconstruction.

All PET image reconstructions were performed using an ordered-subset expectation-maximization algorithm with 4 iterations, 21 subsets and 3 mm Gaussian smoothing. The matrix size of the

reconstructed PET images was $344 \times 344 \times 128$ with a voxel size of $2.1 \times 2.1 \times 2.0$ mm. All PET images were converted to SUV prior to further processing. End-expiration PET reconstruction (manual gating) and motion correction (BodyCOMPASS, Siemens Healthineers) are commercially available and were installed on the clinical PET workstation by the vendor.

Tumor delineation and feature extraction

The primary tumor was delineated for each patient and reconstruction ($PET_{100\%}$, $PET_{40\%}$ and PET_{MoCo}) using a 41% SUV_{max} threshold (Boellaard et al. 2015). From each segmented tumor, 834 radiomic features were extracted, out of which 106 features were derived from the original PET images, and 728 features were derived from wavelet transforms of the original PET image.

The 106 features from the original PET images could be subdivided into 91 voxel-based features and 14 shape features, and were defined according to The Image Biomarker Standardisation Initiative (IBSI) (Zwanenburg et al. 2020). In addition one clinically relevant feature, SUV_{peak} , was also extracted from the original PET images (Boellaard et al. 2015). The 91 voxel-based features could be subdivided into 7 classes, including 18 first-order statistical features, 22 Gray Level Co-occurrence matrix (GLCM) features, 14 Gray Level Dependence Matrix (GLDM) features, 16 Gray Level Run Length Matrix (GLRLM) features, 16 Gray Level Size Zone Matrix (GLSZM) features, and 5 Neighboring Gray Tone Difference Matrix (NGTDM) features.

Voxel-based features were also extracted from 8 wavelet transforms of the original PET image, as described in the literature (Aerts et al. 2014). In short, directional low-pass (L) and high-pass (H) filtering was applied to the image, resulting in 8 transformed images from which features could be extracted using the original tumor delineations. The decompositions are named LLL, LLH, LHL, LHH, HLL, HLH, HHL and HHH, according to the x, y, z order of directional filtering.

Radiomic features were extracted using PyRadiomics 3.0 (Griethuysen et al. 2017).

Data analysis and statistics

Individual tumor movement for $PET_{100\%}$ was quantified as center-of-mass displacement relative $PET_{40\%}$ and PET_{MoCo} , and described as median and 75-percentile deviation. The 75-percentile was included because of the realization that only a small fraction of the patients showed substantial respiratory movement.

The relative difference in feature-values for $PET_{40\%}$ and PET_{MoCo} compared to $PET_{100\%}$ was calculated for all features and reported as mean and 95% confidence interval across patients. Feature-values with a relative difference lower than 5% was considered to be motion robust in this work.

The t-distribution was used for calculation of the confidence intervals. Inspection of quantile-quantile plots revealed that normality could not be assumed. Consequently, statistical significance was calculated using Wilcoxon signed-rank test. P-values lower than 0.05 were considered significant.

The concordance correlation coefficient (CCC) was calculated for all features to determine the correlation between $PET_{40\%}$ and $PET_{100\%}$, and between PET_{MoCo} and $PET_{100\%}$. In this work, motion robust features were defined to have a CCC larger than 0.95, corresponding to a "substantial" or "high" strength-of-agreement level (McBride 2007).

Selection of motion robust features

In this work we define a feature to be motion robust if it has the same value in PET_{100%}, PET_{40%} and PET_{MoCo} reconstructions. Wilcoxon signed-rank test measured if the differences were significant. The CCC was included as a third measure to evaluate the agreement between features, which may handle situations where other statistical tests fail (Lin 1989). Thus, all three criteria (relative difference, P-values, and CCC) was employed to find motion robust features.

First, from the 834 initially extracted features, all features with relative difference lower than 5%, P-value above 0.05 and CCC larger than 0.95 in both PET_{40%} and PET_{MoCo} comparisons were selected as motion robust features. In this three-wise combined comparison, the features matching these criteria have feature-values that are robust to the PET reconstruction type (PET_{100%}, PET_{40%} and PET_{MoCo}), and, consequently, independent of the breathing pattern of the patient.

A similar pair-wise selection was also performed comparing PET_{100%} to PET_{40%} features, and PET_{100%} to PET_{MoCo} features separately, to identify features that were pair-wise motion robust.

To thoroughly investigate which features were truly motion robust, the patients with large tumor movement were evaluated separately. In this analysis, patients with tumor center-of-mass displacement larger than the median displacement in PET_{100%} relative to PET_{MoCo} were included. The three-wise selection criteria was re-evaluated on this reduced data-set. Displacement was quantified as the distance between the center-of-mass distance of PET_{100%} relative that of PET_{MoCo}.

Selection of uncorrelated features

Among the selected motion robust features, there could still be redundancy if features were highly correlated. Therefore, feature selection is an important pre-processing step in classification tasks, to reduce the number of redundant features in order to improve the performance of the classifier (Blum et al. 1997). In this work, to generate a set of features with low correlation for PET_{40%} and PET_{MoCo}, one example of a feature selection method was applied: The correlation matrix was calculated between all motion robust features (Figure 3A), and feature-pairs with correlation coefficient larger than 0.95 were identified as candidates for removal. For each candidate pair, i, j , with correlation coefficient c_{ij} , the feature with highest average correlation, C , across all (n) features, $C = \frac{1}{n} \sum_j c_{ij}$, was removed (Wu et al. 2016). The features remaining after the selection process could be treated as both motion robust and with low correlation (Figure 3B).

As a final step, we compared the features from recent radiomics publications with our results. None of the publications found were performed using respiratory-motion information.

Results

One patient was identified as an outlier and removed from the subsequent analyses due to abnormal feature-values, possibly due to infiltration of the lung tumor into vertebral bone and the presence of a rare autoimmune disease.

Visual inspection of Figure 1 reveals that the PET_{40%} and PET_{MoCo} image have smaller tumor volume compared to PET_{100%}. As expected, the noise level in PET_{40%} is visually higher than for PET_{100%}, while the noise level for PET_{MoCo} is comparable to PET_{100%}.

The deviation in center-of-mass relative $PET_{100\%}$ was 0.96 mm (median) and 1.99 mm (75-percentile) for $PET_{40\%}$. For PET_{MoCo} the corresponding deviation was 2.6 mm (median) and 3.7 mm (75-percentile) relative $PET_{100\%}$.

Figure 2 shows the difference, relative $PET_{100\%}$, for all 834 features for $PET_{40\%}$ and PET_{MoCo} . In general, there were more features with low relative difference values for PET_{MoCo} , compared to $PET_{40\%}$.

The number of features per feature class fulfilling each criteria in the three-wise comparison is summarized in Table 2. From the 834 initially extracted features, 126 features had a relative difference lower than 5%, 169 had a P-value above 0.05 and, 193 had a CCC larger than 0.95, and 43 features fulfilled all three criteria, in the three-wise $PET_{40\%}$ and PET_{MoCo} comparisons.

Table 3 lists all 43 of the 834 evaluated features remaining after applying the three-wise selection criteria. These features are considered stable and independent of the breathing pattern of the patient and can thus be extracted from non-gated or non-motion-corrected PET-examinations. Six of these features were previously reported to be useful for prediction of survival or disease state for non-small-cell lung cancer patients, in recent radiomics publications, as indicated in Table 3. Supplementary Table S1 and S2 lists, respectively, the 80 and 213 features remaining after applying the three selection criteria pair-wise to the $PET_{40\%}$ and PET_{MoCo} comparisons. Supplementary Table S3 shows the resulting 38 features remaining after applying the three-wise criteria on the nine patients with largest tumor displacement.

One example of a feature reduction method, to remove highly correlated features, was applied, to reduce the number of features even further. This example is illustrated in Figure 3 for $PET_{40\%}$. Figure 3A, shows the Pearson correlation matrix for all pairs of the 43 motion-robust features from Table 3. Figure 3B displays the sparse matrix with the remaining features after performing the correlation-based feature elimination procedure (correlation coefficient > 0.95). The equivalent figure for PET_{MoCo} is shown in Supplementary Figure S1. This elimination process reduced the 43 motion robust features to 14 motion robust features with low inherent correlation for both $PET_{40\%}$ and PET_{MoCo} .

Table 4 displays the features with low between-feature correlation remaining after the correlation-based elimination process. The majority of the features were the same between $PET_{40\%}$ (Table 4A) and PET_{MoCo} (Table 4B). Two features for $PET_{40\%}$ and two features for PET_{MoCo} were unique. As indicated in Table 4, the two features "GLCM JointEnergy" and "NGDTM Coarseness" have been previously reported to be diagnostically predictive in PET-radiomics research (Ohri et al. 2016; Kirienko et al. 2018).

Table 5 shows 24 other PET-radiomic features which were not motion robust, but which have been reported to be predictive for health status, using non-gated PET.

Discussion

Respiratory motion introduces artefacts and blurring in thoracic PET images. This results in breathing-amplitude dependent variability of PET-based radiomic tumor-features, yielding less robust predictive models for disease classification. In this study, we identify radiomic features that are robust to motion and, thus, can be extracted from simple free-breathing PET (PET_{100%}) without gating or motion correction, which both require additional time and special attention during scanning. These features are candidates for building robust predictive models for disease state or survival classification, that can be implemented with standard clinical PET acquisition protocols.

We extracted 834 standardized radiomic features from 18 non-small-cell lung cancer patients within a clinical PET/MRI trial. Features from PET_{100%} were compared to two motion-compensation PET methods: end-expiration (PET_{40%}) and MRI-based motion corrected (PET_{MoCo}) reconstruction. A feature was classified as motion robust if the following three criteria were fulfilled: relative difference < 5%, P-value > 0.05 and CCC > 0.95.

We found that 43 of the 834 extracted features were motion robust (Table 2), while 94% of the extracted features exhibited variability to motion. Shape features were generally more robust compared to voxel-based features, which is in agreement with several previous publications (Oliver, Budzevich, Zhang, et al. 2015; Oliver, Budzevich, Hunt, et al. 2017; Du et al. 2019; Peerlings et al. 2019).

As illustrated in Figure 2, the spread of the relative difference values (relative to PET_{100%}) were higher for PET_{40%} than for PET_{MoCo}. This coincides with the observed larger image noise for PET_{40%} (Figure 1). One plausible explanation may be that some first and higher-order voxel-based features are known to be highly noise dependent (Oliver, Budzevich, Hunt, et al. 2017), resulting in the observed higher variability of the PET_{40%}-based feature values.

Motion amplitude

A requirement when studying motion variability of radiomic features is that motion is present in the data. In this study, the presence of motion was characterized by the difference in tumor center-of-mass between free-breathing PET_{100%} and the two motion compensated PET reconstructions (PET_{40%} and PET_{MoCo}). We found average center-of-mass differences of 1-2 mm. This is of similar magnitude as the reported 1.5 mm displacement for lung lesions located in the upper lung lobe, which are known to be less affected by motion, than central or lower located lesions (Grootjans et al. 2016). To accommodate for some patients having small tumor motion, we investigated the 75-percentile displacement, which was found to be 2-5 mm. A smaller fraction of the patients was, thus, affected by larger tumor motion than the average patient, which could affect the results, as discussed below.

Motion robust features

In this work we define a feature to be motion robust if it has the same value in the three PET reconstructions (PET_{100%}, PET_{40%} and PET_{MoCo}). Initially, we were inspired by previous

studies using only a CCC-based selection criterion to classify features as robust to motion (Oliver, Budzevich, Zhang, et al. 2015; Oliver, Budzevich, Hunt, et al. 2017; Timmeren et al. 2016; Traverso et al. 2018). With this selection method, we found features with high CCC values with large deviations, which, thus, were not motion robust. Instead, we employed three criteria, based on relative difference, P-value and CCC, which all had to be fulfilled in pair-wise, or three-wise comparisons, to classify a feature as motion robust.

Forty-three features remained after applying the three-wise selection criteria (Table 3). These 43 features can be considered independent of the PET reconstruction type (PET_{100%}, PET_{40%} and PET_{MoCo}), and may be extracted from a static, free-breathing PET, without the need for gating or motion correction.

The majority of motion robust features were based on wavelet transformed PET-images (Table 3, feature 21-43). Wavelet features have been successfully included in radiomics models in computed tomography (CT) (Aerts et al. 2014; Wu et al. 2016) and MRI (Zhou et al. 2020), but to the best of our knowledge, to date, not been used in any published predictive PET study. As our list of motion robust features is dominated by wavelet-based features, we believe that incorporating wavelet transforms would be an important factor to successful PET-radiomics prediction-studies of lung cancer.

Above, we have discussed the 43 features that were found invariant to all three reconstructions (PET_{100%}, PET_{40%} and PET_{MoCo}). To further understand the differences between PET_{40%} and PET_{MoCo}, we compared the reconstructions pair-wise. Comparing PET_{100%} to PET_{40%} (using the same three selection criteria as above) resulted in 82 features being motion robust (Supplementary Table S1). Likewise, the same comparison between PET_{100%} and PET_{MoCo} resulted in 219 motion robust features (Supplementary Table S2). As discussed earlier, PET_{40%} has increased noise level compared to both PET_{100%} and PET_{MoCo}. In addition, both PET_{40%} and PET_{MoCo} has reduced motion, compared to PET_{100%} (Figure 1). The difference in motion compensation methodology between PET_{40%} (using only end-expiration PET counts) and PET_{MoCo} (non-rigid registration of five breathing phases), may have introduced differences in tumor texture, and thus, resulted in different sets of motion robust features. This may explain the differences between Table 3 (43 features) and Table S1 (82 features).

To investigate the impact of motion amplitude on the selected features, we performed the analysis on the half of patients with largest tumor motion, as measured with the center-of-mass difference in PET_{100%} relative PET_{MoCo}. This resulted in a list of 38 features (Supplementary Table S3). Half of these features coincided with our suggested list of motion robust features in Table 3. Shape features was the feature class with most overlap, where all shape features from Table 3 were present in Supplementary Table S3. Around half of the wavelet-based features remained equal, while no first-order features were overlapping between the tables. The different set of features between Table 3 and Supplementary Table S3 indicates that the remaining list of features after applying the three selection criteria is highly dependent on the fraction of tumors with small and large motion amplitude in the data set. As clinical studies consists of patients with varying motion amplitude,

we believe that our list of 43 features in Table 3 is a more representative feature-set for a normal cohort, representing all types of breathing patterns.

Eliminating correlated features

To reduce the number of correlated features, an unsupervised, correlation-matrix based, elimination process was applied. This approach reduced the 43 motion robust (but correlated) features (Table 3), to 14 features with low correlation, for $PET_{40\%}$ and PET_{MoCo} (Figure 3, Table 4). Correlation-based feature elimination is just one of many possible feature selection methods (Blum et al. 1997; Wu et al. 2016; Urbanowicz et al. 2018). With the 43 motion-robust features in Table 3 as a starting point, it is possible to explore supervised or unsupervised feature selection methods, to find the most optimal features for the prediction being investigated.

Features predictive of health status

Among the 43 motion robust features in Table 3, we found six features (Entropy, Uniformity, GLCM JointEnergy, GLCM JointEntropy, GLCM MaximumProbability and NGTDM Coarseness) previously used in disease-related prediction models and cited in recent PET radiomics publications.

Earlier research have identified features predictive of health status using non-gated PET, but which did not fulfill our criteria for motion robustness (Table 5). Interestingly, these features could still differentiate between clinical end points, even though the images were blurred by motion. It is possible that these prediction models could be further improved by using motion robust features, as suggested in a recent CT-based radiomics study (Du et al. 2019).

Related work

Three of the motion robust features found in our study (Firstorder Entropy, GLCM Joint Entropy and GLCM SumEntropy) were also reported as motion-invariant by Oliver, Budzevich, Zhang, et al. 2015. They found six additional features with low variability, which we did not find to be motion robust (Shape Sphericity, GLCM Imc2, GLRLM ShortRunEmphasis, GLRLM LongRunEmphasis, and GLRLM RunPercentage). One plausible explanation to this may be that the total number of radioactive decays within the acquisition time was lower, compared to our study. As discussed above, GLCM and GLRLM features are highly noise sensitive, thus, it is likely that the different noise levels in the two studies affected the results. In addition, Oliver, Budzevich, Zhang, et al. 2015 used solely a CCC-based criterion to classify features, which leaves the possibility that features fulfilling the CCC-criterion still are not motion robust when also relative differences and P-values are considered.

We found NGTDM Coarseness to be motion robust while NGTDM Contrast was non-motion robust. This is contradictory to the study by S. Yip et al. 2014 who found motion-variability in the former feature, while the latter was motion-robust. The PET counts per bed position was approximately double in S. Yip et al. 2014, compared to our study, suggesting that the noise properties, in combination with motion, could possibly affect the variability of feature-values. Another possible explanation to the different results is that S. Yip et al. 2014 used relative difference and P-value criteria for feature selection, while we, in addition, employed CCC, to select motion robust features.

Limitations

Our sample size was comparable to similar published studies (Oliver, Budzevich, Zhang, et al. 2015; S. Yip et al. 2014) and we hypothesize that this did not affect our results. Because of the limited sample size, and that the end-points are not yet known, we could not build predictive models to test the impact of the motion robust features on disease or survival predictions. We will investigate this in a future follow-up study.

Conclusion

In conclusion, we have identified 43 motion robust tumor PET tumor features, of which 37 have not been previously reported. These features can be extracted from simple free-breathing PET reconstructions, without employing gating or motion-correction.

Acknowledgements

This research was supported by grant HNF1349-17 from the Northern Norway Regional Health Authority. We would like to thank the radiographers Anneli Johansen, Katharina K. Lindberg and Kent Johansen, at The PET Imaging Center, University Hospital of North Norway, for performing the PET/MRI scanning.

References

- Aerts, Hugo J. W. L. et al. (Sept. 2014). “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach”. In: *Nat. Commun.* 5.1, pp. 1–8.
- Ahn, H.K. et al. (June 2019). “Pre-treatment 18F-FDG PET-based radiomics predict survival in resected non-small cell lung cancer”. In: *Clin. Radiol.* 74.6, pp. 467–473.
- Altazi, Baderaldeen A. et al. (2017). “Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms”. In: *J. Appl. Clin. Med. Phys.* 18.6, pp. 32–48.
- Avanzo, Michele et al. (Oct. 2020). “Radiomics and deep learning in lung cancer”. In: *Strahlentherapie und Onkol.* 196.10, pp. 879–887.
- Bianconi, Francesco, Isabella Palumbo, Mario Luca Fravolini, et al. (2019). “Texture Analysis on [18F]FDG PET/CT in Non-Small-Cell Lung Cancer: Correlations Between PET Features, CT Features, and Histological Types”. In: *Mol. Imaging Biol.* 21.6, pp. 1200–1209.
- Bianconi, Francesco, Isabella Palumbo, Angela Spanu, et al. (2020). “PET/CT radiomics in lung cancer: An overview”. In: *Appl. Sci.* 10.5, pp. 1–11.
- Blum, Avrim L. and Pat Langley (Dec. 1997). “Selection of relevant features and examples in machine learning”. In: *Artif. Intell.* 97.1-2, pp. 245–271.
- Blumhagen, Jan O. et al. (Jan. 2014). “Field of view extension and truncation correction for MR-based human attenuation correction in simultaneous MR/PET imaging”. In: *Med. Phys.* 41.2, p. 022303.
- Boellaard, Ronald et al. (Feb. 2015). “FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0”. In: *Eur. J. Nucl. Med. Mol. Imaging* 42.2, pp. 328–354.
- Cook, Gary J.R. et al. (2013). “Are pretreatment 18F-FDG PET tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy?” In: *J. Nucl. Med.* 54.1, pp. 19–26.
- Desseroit, Marie Charlotte et al. (2016). “Development of a nomogram combining clinical staging with 18F-FDG PET/CT image features in non-small-cell lung cancer stage I-III”. In: *Eur. J. Nucl. Med. Mol. Imaging* 43.8, pp. 1477–1485.
- Dissaux, Gurvan et al. (2020). “Pretreatment 18F-FDG PET/CT Radiomics Predict Local Recurrence in Patients Treated with Stereotactic Body Radiotherapy for Early-Stage Non-Small Cell Lung Cancer: A Multicentric Study”. In: *J. Nucl. Med.* 61.6, pp. 814–820.
- Du, Qian et al. (2019). “Radiomic feature stability across 4D respiratory phases and its impact on lung tumor prognosis prediction”. In: *PLoS One* 14.5, pp. 1–16.
- Fried, D et al. (June 2016). “MO-DE-207B-07: Assessment of Reproducibility Of FDG-PET-Based Radiomics Features Across Scanners Using Phantom Imaging”. In: *Med. Phys.* 43, pp. 3705–3706.
- Furst, S. et al. (2015). “Motion Correction Strategies for Integrated PET/MR”. In: *J. Nucl. Med.* 56.2, pp. 261–269.
- Greenspan and Bennett S. (2017). “Role of PET/CT for precision medicine in lung cancer: perspective of the Society of Nuclear Medicine and Molecular Imaging”. In: *Transl. Lung Cancer Res.* 6.6, pp. 617–620.

- Griethuysen, Joost J.M. van et al. (Nov. 2017). “Computational Radiomics System to Decode the Radiographic Phenotype”. In: *Cancer Res.* 77.21, e104–e107.
- Grimm, Robert et al. (Jan. 2015). “Self-gated MRI motion modeling for respiratory motion compensation in integrated PET/MRI”. In: *Med. Image Anal.* 19.1, pp. 110–120.
- Grootjans, Willem et al. (Nov. 2016). “The impact of optimal respiratory gating and image noise on evaluation of intratumor heterogeneity on 18F-FDG pet imaging of lung cancer”. In: *J. Nucl. Med.* 57.11, pp. 1692–1698.
- Hatt, Mathieu et al. (2018). “Tumour functional sphericity from PET images: prognostic value in NSCLC and impact of delineation method”. In: *Eur. J. Nucl. Med. Mol. Imaging* 45.4, pp. 630–641.
- He, Bo et al. (2018). “A biomarker basing on radiomics for the prediction of overall survival in non-small cell lung cancer patients”. In: *Respir. Res.* 19.1, pp. 1–8.
- Keller, Sune H. et al. (Feb. 2013). “Image artifacts from MR-based attenuation correction in clinical, whole-body PET/MRI”. In: *Magn. Reson. Mater. Physics, Biol. Med.* 26.1, pp. 173–181.
- Kirienko, Margarita et al. (2018). “Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery”. In: *Eur. J. Nucl. Med. Mol. Imaging* 45.2, pp. 207–217.
- Kuttner, Samuel et al. (Jan. 2020). “Quantitative PET/MR imaging of lung cancer in the presence of artifacts in the MR-based attenuation correction maps”. In: *Acta radiol.* 61.1, pp. 11–20.
- Ladrón De Guevara Hernández, David (2015). “The Role of PET/CT Imaging in Lung Cancer”. In: *J. Cancer Ther.* 6.6, pp. 690–700.
- Leijenaar, Ralph T.H. et al. (2015). “The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis”. In: *Sci. Rep.* 5.1, pp. 1–10.
- Lin, Lawrence I-kuei (Mar. 1989). “A Concordance Correlation Coefficient to Evaluate Reproducibility”. In: *Biometrics* 45.1, pp. 255–268.
- Lovinfosse, Pierre et al. (2016). “FDG PET/CT texture analysis for predicting the outcome of lung cancer treated by stereotactic body radiation therapy”. In: *Eur. J. Nucl. Med. Mol. Imaging* 43.8, pp. 1453–1460.
- McBride, GB (2007). “Equivalence measures for comparing the performance of alternative methods for the analysis of water quality variables”. In: *Auckland, New Zealand: National Institute of Water & Atmospheric Research Ltd.*
- Ohri, Nitin et al. (2016). “Pretreatment 18F-FDG PET textural features in locally advanced non-small cell lung cancer: Secondary analysis of ACRIN 6668/RTOG 0235”. In: *J. Nucl. Med.* 57.6, pp. 843–848.
- Oikonomou, Anastasia et al. (2018). “Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy”. In: *Sci. Rep.* 8.1, pp. 1–11.

- Oliver, Jasmine A., Mikalai Budzevich, Dylan Hunt, et al. (Oct. 2017). “Sensitivity of Image Features to Noise in Conventional and Respiratory-Gated PET/CT Images of Lung Cancer: Uncorrelated Noise Effects”. In: *Technol. Cancer Res. Treat.* 16.5, pp. 595–608.
- Oliver, Jasmine A., Mikalai Budzevich, Geoffrey G. Zhang, et al. (2015). “Variability of image features computed from conventional and respiratory-gated PET/CT images of lung cancer”. In: *Transl. Oncol.* 8.6, pp. 524–534.
- Peerlings, Jurgen et al. (Dec. 2019). “Stability of radiomics features in apparent diffusion coefficient maps from a multi-centre test-retest trial”. In: *Sci. Rep.* 9.1, pp. 1–10.
- Pinker, Katja, Christopher Riedl, and Wolfgang A. Weber (Aug. 2017). “Evaluating tumor response with FDG PET: updates on PERCIST, comparison with EORTC criteria and clues to future developments”. In: *Eur. J. Nucl. Med. Mol. Imaging* 44.S1, pp. 55–66.
- Pyka, Thomas et al. (2015). “Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy”. In: *Radiat. Oncol.* 10.1, pp. 1–9.
- Sauter, Alexander W. et al. (2015). “Image-derived biomarkers and multimodal imaging strategies for lung cancer management”. In: *Eur. J. Nucl. Med. Mol. Imaging* 42.4, pp. 634–643.
- Shiri, Isaac et al. (2017). “The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies”. In: *Eur. Radiol.* 27.11, pp. 4498–4509.
- Simon, Richard M. et al. (May 2011). “Using cross-validation to evaluate predictive accuracy of survival risk classifiers based on high-dimensional data”. In: *Brief. Bioinform.* 12.3, pp. 203–214.
- Timmeren, Janna E Van et al. (Dec. 2016). “Test–Retest Data for Radiomics Feature Stability Analysis: Generalizable or Study-Specific?” In: *Tomography* 2.4, pp. 361–365.
- Tixier, Florent et al. (2014). “Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: Prognostic value in non-small cell lung cancer”. In: *J. Nucl. Med.* 55.8, pp. 1235–1241.
- Traverso, Alberto et al. (Nov. 2018). “Repeatability and Reproducibility of Radiomic Features: A Systematic Review”. In: *Int. J. Radiat. Oncol. Biol. Phys.* 102.4, pp. 1143–1158.
- Urbanowicz, Ryan J. et al. (Sept. 2018). “Relief-based feature selection: Introduction and review”. In: *J. Biomed. Inform.* 85, pp. 189–203.
- Van Elmpt, Wouter et al. (2011). “Optimal gating compared to 3D and 4D PET reconstruction for characterization of lung tumours”. In: *Eur. J. Nucl. Med. Mol. Imaging* 38.5, pp. 843–855.
- Velden, Floris H. P. van et al. (Oct. 2016). “Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [18F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation”. In: *Mol. Imaging Biol.* 18.5, pp. 788–795.
- Wild, CP, E Weiderpass, and BW Stewart (2020). “World cancer report: cancer research for cancer prevention”. In: *Lyon: International Agency for Research on Cancer.*
- Wu, Weimiao et al. (2016). “Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology”. In: *Front. Oncol.* 6.March, pp. 1–11.

- Yip, Stephen et al. (Dec. 2014). “Comparison of Texture Features Derived from Static and Respiratory-Gated PET Images in Non-Small Cell Lung Cancer”. In: *PLoS One* 9.12. Ed. by Olga Y. Gorlova, pp. 1–14.
- Yip, Stephen S.F. et al. (2017). “Associations Between Somatic Mutations and Metabolic Imaging Phenotypes in Non-Small Cell Lung Cancer”. In: *J. Nucl. Med.* 58.4, pp. 569–576.
- Yoon, Hyun Jung et al. (Oct. 2015). “Decoding Tumor Phenotypes for ALK, ROS1, and RET Fusions in Lung Adenocarcinoma Using a Radiomics Approach”. In: *Medicine* 94.41, pp. 1–8.
- Zhou, Jiali et al. (Dec. 2020). “Predicting the response to neoadjuvant chemotherapy for breast cancer: wavelet transforming radiomics in MRI”. In: *BMC Cancer* 20.1, pp. 1–10.
- Zwanenburg, Alex et al. (May 2020). “The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping”. In: *Radiology* 295.2, pp. 328–338.

Figures

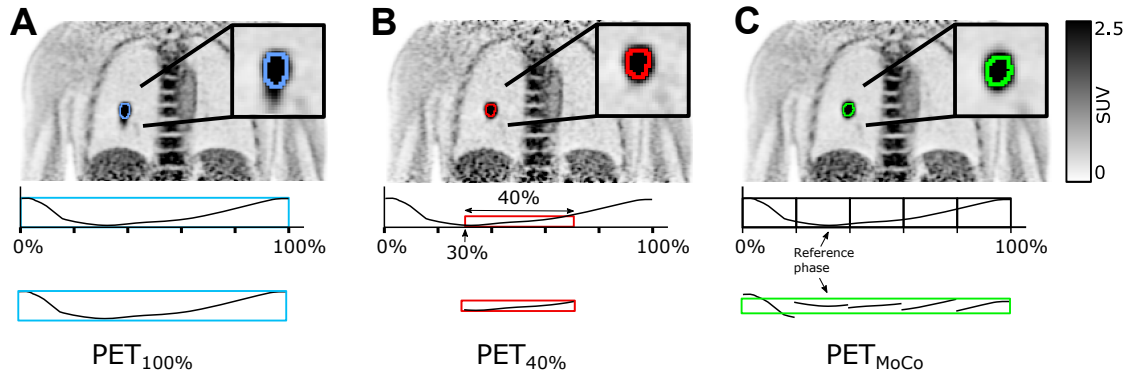


Figure 1. PET image of one representative patient (top row), respiration curve (middle row) and an illustration of the amount of PET data and its variability per breathing cycle, used for reconstruction (bottom row). A. PET_{100%}. B. PET_{40%}. C. PET_{MoCo}. The primary tumor and corresponding 41% SUV_{max} delineation is shown in the right lung and in the insert.

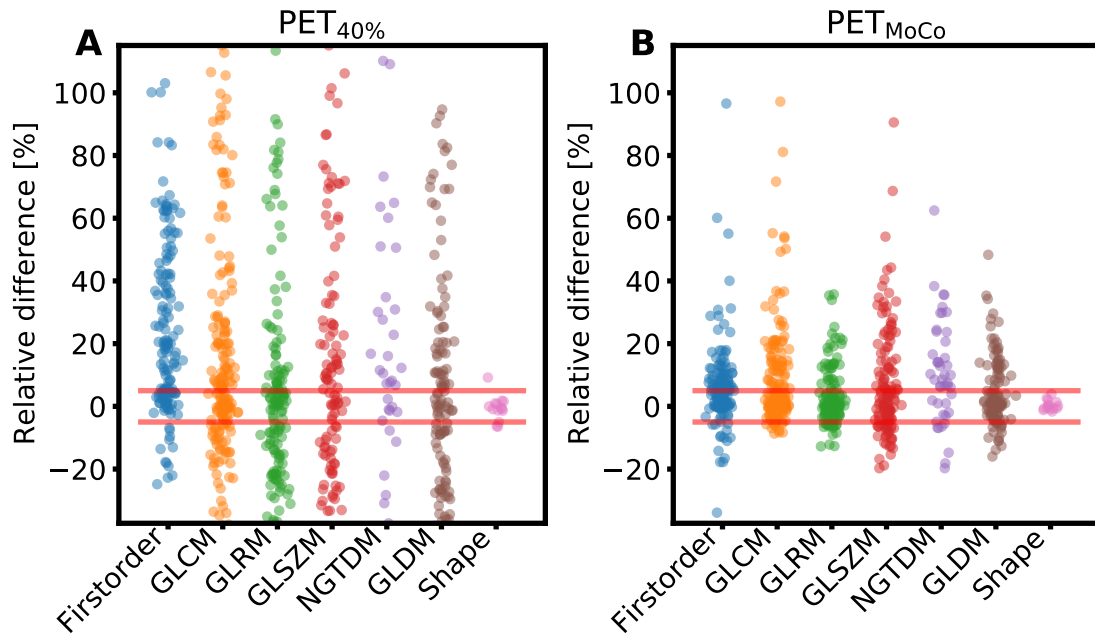


Figure 2. Relative difference (RD) of all features in different feature classes. A. RD of PET_{40%} to PET_{100%}. B. RD of PET_{MoCo} to PET_{100%}. In A and B, horizontal red line indicate $\pm 5\%$ RD. The y-axis limits were chosen to include 97% of the data points.

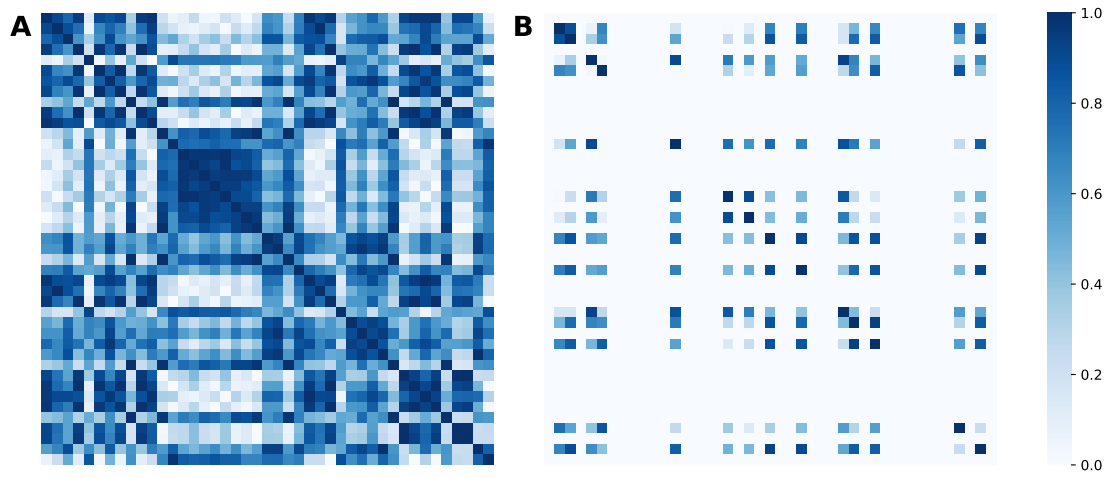


Figure 3. Plots of the absolute value of pair-wise Pearson correlation coefficients for the 43 motion robust features from Table 3. A. Before removal of correlated features. B. After removal of features with correlation coefficient > 0.95 .

Tables

Table 1. Patient characteristics.

Characteristics	Value
Female	7
Male	11
Age	71 (52-81) [†] years
Weight	80 (51-132) [†] kg
Histological subtype	
Adenocarcinoma	9
Squamous cell carcinoma	9
Clinical disease stage*	
I	5
II	3
III	7
IV	3

[†]Value presented as Mean (range)

*TNM 8th edition

Table 2. Number of features remaining per selection criterion. "None" indicates the distribution of features without any selection criterion applied. "All" indicates that all three selection criteria were fulfilled.

Feature class	Number of features				
	None	RD < 5%	P > 0.05	CCC > 0.95	All
Firstorder	163	14	35	46	7
GLCM	198	44	29	41	8
GLRLM	144	31	22	37	7
GLSZM	144	11	31	19	6
NGTDM	45	2	17	12	2
GLDM	126	14	26	27	6
Shape	14	10	9	11	7
Sum	834	126	169	193	43

RD = Relative difference

CCC = concordance correlation coefficient

Table 3. All features with $PET_{40\%}$ and PET_{MoCo} relative difference $< 5\%$, P-value > 0.05 , and $CCC > 0.99$. Features that are coinciding with Supplementary Table S3 have been marked with a dagger (†). Features that have been used in recent radiomics literature for disease state prediction are marked with the corresponding reference. Relative difference is shown with the $\pm 95\%$ confidence interval.

	Feature name	$PET_{40\%}$			PET_{MoCo}			Reference
		Relative difference [%]	P-value	CCC	Relative difference [%]	P-value	CCC	
1	Original Firstorder Entropy	0.69 ± 1.36	0.62	0.991	1.08 ± 1.54	0.14	0.988	
2	Original Firstorder Median	4.49 ± 5.20	0.12	0.986	5.00 ± 4.71	0.06	0.984	(Bianconi, Palumbo, Fravelini, et al. 2019)
3	Original Firstorder SUV peak	2.24 ± 4.84	0.80	0.993	4.98 ± 4.52	0.08	0.987	
4	Original Firstorder Uniformity	-0.92 ± 4.47	0.59	0.994	-3.34 ± 4.97	0.83	0.994	(Bianconi, Palumbo, Fravelini, et al. 2019)
5	Original GLCM Imc1	-2.21 ± 9.63	0.36	0.974	4.01 ± 7.09	0.27	0.972	
6	Original GLCM JointEntropy	-3.83 ± 5.08	0.14	0.989	-4.93 ± 6.40	0.49	0.991	
7	Original GLCM JointEntropy	0.89 ± 1.18	0.10	0.997	0.73 ± 1.18	0.15	0.995	(Kirienko et al. 2018)
8	Original GLCM MaximumProbability	-2.86 ± 7.85	0.36	0.995	-0.69 ± 7.89	0.72	0.981	(Pyka et al. 2015; Tixier et al. 2014; Oikonomou et al. 2018)
9	Original GLDM DependenceNonUniformity	-1.65 ± 5.06	0.29	0.995	-0.46 ± 7.71	0.55	0.994	(S. S. Yip et al. 2017)
10	Original GLRLM GrayLevelNonUniformityNormalized	-1.10 ± 4.40	0.62	0.995	-3.27 ± 4.93	0.79	0.993	
11	Original GLRLM RunEntropy	-0.07 ± 1.55	0.65	0.982	1.15 ± 1.19	0.16	0.983	
12	Original GLRLM RunLengthNonUniformity	-4.04 ± 4.85	0.06	0.990	-1.01 ± 7.73	0.08	0.992	
13	Original NGTDM Coarseness	-1.37 ± 3.63	0.29	0.991	-1.98 ± 5.42	0.27	0.980	
14	Original Shape MajorAxisLength	-1.39 ± 2.00	0.10	0.999	-1.66 ± 3.16	0.15	0.998	
15	Original Shape Maximum2DDiameterColumn	-0.46 ± 2.25	0.55	0.998	-0.38 ± 4.03	0.57	0.995	
16	Original Shape Maximum2DDiameterRow	1.79 ± 4.17	0.43	0.992	0.15 ± 3.68	0.79	0.993	
17	Original Shape Maximum3DDiameterSlice	-1.32 ± 2.89	0.33	0.998	-1.22 ± 3.66	0.33	0.996	
18	Original Shape Maximum3DDiameter	-0.89 ± 1.99	0.12	0.999	-0.91 ± 2.82	0.47	0.998	
19	Original Shape MinorAxisLength	-0.01 ± 2.42	0.62	0.999	2.05 ± 3.84	0.41	0.992	
20	Original Shape SurfaceArea	0.97 ± 4.25	0.83	0.994	0.13 ± 5.91	0.23	0.996	
21	Wavelet-LLL GLRLM RunLengthNonUniformity	4.11 ± 5.74	0.87	0.995	-1.09 ± 7.71	0.15	0.991	
22	Wavelet-LHH GLDM DependenceEntropy	0.33 ± 1.35	0.69	0.954	0.90 ± 0.84	0.06	0.975	
23	Wavelet-LHL GLDM DependenceEntropy	-0.30 ± 0.85	0.12	0.984	0.67 ± 0.64	0.06	0.990	
24	Wavelet-LHL GLRLM RunLengthNonUniformity	0.79 ± 5.43	0.52	0.992	-0.90 ± 8.10	0.08	0.992	
25	Wavelet-LHL GLSZM ZoneEntropy	0.59 ± 1.61	0.72	0.971	0.14 ± 1.32	0.27	0.986	
26	Wavelet-LLL Firstorder Entropy	0.44 ± 0.97	0.65	0.993	0.73 ± 0.85	0.12	0.994	
27	Wavelet-LLL Firstorder Median	4.10 ± 5.18	0.15	0.986	4.79 ± 4.78	0.06	0.984	
28	Wavelet-LLL Firstorder Uniformity	-1.37 ± 4.14	0.59	0.996	-3.08 ± 3.86	0.55	0.996	
29	Wavelet-LLL GLCM Imc1	0.17 ± 5.84	0.59	0.985	2.48 ± 3.13	0.19	0.987	
30	Wavelet-LLL GLCM JointEntropy	1.87 ± 5.57	0.98	0.977	-0.90 ± 6.06	0.52	0.992	
31	Wavelet-LLL GLCM JointEntropy	-0.17 ± 1.06	0.41	0.997	0.22 ± 1.08	0.52	0.996	
32	Wavelet-LLL GLCM MaximumProbability	-1.92 ± 7.52	0.46	0.983	-4.03 ± 6.08	0.49	0.993	
33	Wavelet-LLL GLDM DependenceEntropy	-0.65 ± 0.99	0.12	0.983	0.40 ± 0.82	0.49	0.989	
34	Wavelet-LLL GLDM DependenceNonUniformity	4.26 ± 6.06	0.09	0.994	0.67 ± 7.33	0.11	0.992	
35	Wavelet-LLL GLDM SmallDependenceEmphasis	2.68 ± 6.06	0.06	0.982	-0.13 ± 1.97	0.55	0.995	
36	Wavelet-LLL GLRLM GrayLevelNonUniformityNormalized	-1.39 ± 4.13	0.76	0.996	-3.06 ± 3.87	0.65	0.996	
37	Wavelet-LLL GLRLM RunEntropy	0.13 ± 1.06	0.98	0.991	0.72 ± 0.79	0.14	0.993	
38	Wavelet-LLL GLSZM GrayLevelNonUniformityNormalized	-2.21 ± 4.05	0.38	0.994	-3.22 ± 3.92	0.52	0.996	
39	Wavelet-LLL GLSZM SizeZoneNonUniformity	-1.47 ± 5.65	0.16	0.994	-1.31 ± 8.47	0.11	0.993	
40	Wavelet-LLL GLSZM SizeZoneNonUniformityNormalized	0.59 ± 2.81	0.38	0.973	-0.46 ± 2.51	0.79	0.988	
41	Wavelet-LLL GLSZM SmallAreaEmphasis	0.16 ± 1.28	0.46	0.985	-0.37 ± 1.39	0.87	0.989	
42	Wavelet-LLL GLSZM ZoneEntropy	0.33 ± 0.79	0.69	0.990	0.82 ± 0.86	0.08	0.986	
43	Wavelet-LLL NGTDM Coarseness	2.32 ± 3.33	0.49	0.998	-2.68 ± 5.22	0.11	0.990	

Table 4. List of motion robust ($RD < \pm 5\%$, $P > 0.05$, $CCC > 0.95$) features with low correlation (correlation coefficient < 0.95). Features that are unique in subtable A and B are indicated in bold. Relative difference is shown with the $\pm 95\%$ confidence interval.

(A) $PET_{40\%}$

	Feature name	Relative difference [%]	P-value	CCC	Reference
1	Original Firstorder Median	4.49 ± 5.20	0.12	0.986	
2	Original Firstorder SUV peak	2.24 ± 4.84	0.80	0.993	
3	Original GLCM Imc1	-2.21 ± 9.63	0.36	0.974	
4	Original GLCM JointEnergy	-3.83 ± 5.08	0.14	0.989	(Kirienko et al. 2018)
5	Original NGTDM Coarseness	-1.37 ± 3.63	0.29	0.991	(Pyka et al. 2015)
6	Original Shape Maximum3DDiameter	-0.89 ± 1.99	0.12	0.999	
7	Original Shape SurfaceArea	0.97 ± 4.25	0.83	0.994	
8	Wavelet-LHH GLDM DependenceEntropy	0.33 ± 1.35	0.69	0.954	
9	Wavelet-LHL GLSZM ZoneEntropy	0.59 ± 1.61	0.72	0.971	
10	Wavelet-LLL GLCM Imc1	0.17 ± 5.84	0.59	0.985	
11	Wavelet-LLL GLCM JointEnergy	1.87 ± 5.57	0.98	0.977	
12	Wavelet-LLL GLCM MaximumProbability	-1.92 ± 7.52	0.46	0.983	
13	Wavelet-LLL GLSZM SizeZoneNonUniformityNormalized	0.59 ± 2.81	0.38	0.973	
14	Wavelet-LLL GLSZM ZoneEntropy	0.33 ± 0.79	0.69	0.990	

(B) PET_{MoCo}

	Feature name	Relative difference [%]	P-value	CCC	Reference
1	Original Firstorder SUV peak	4.98 ± 4.52	0.08	0.987	
2	Original GLCM Imc1	4.01 ± 7.09	0.27	0.972	
3	Original GLCM JointEnergy	-4.93 ± 6.40	0.49	0.991	(Kirienko et al. 2018)
4	Original NGTDM Coarseness	-1.98 ± 5.42	0.27	0.980	(Pyka et al. 2015)
5	Original Shape Maximum3DDiameter	-0.91 ± 2.82	0.47	0.998	
6	Original Shape SurfaceArea	0.13 ± 5.91	0.23	0.996	
7	Wavelet-LHH GLDM DependenceEntropy	0.90 ± 0.84	0.06	0.975	
8	Wavelet-LHL GLSZM ZoneEntropy	0.14 ± 1.32	0.27	0.986	
9	Wavelet-LLL Firstorder Median	4.79 ± 4.78	0.06	0.984	
10	Wavelet-LLL GLCM Imc1	2.48 ± 3.13	0.19	0.987	
11	Wavelet-LLL GLCM JointEnergy	-0.90 ± 6.06	0.52	0.992	
12	Wavelet-LLL GLCM MaximumProbability	-4.03 ± 6.08	0.49	0.993	
13	Wavelet-LLL GLDM DependenceEntropy	0.40 ± 0.82	0.49	0.989	
14	Wavelet-LLL GLSZM SizeZoneNonUniformityNormalized	-0.46 ± 2.51	0.79	0.988	

Table 5. Reported features from recent radiomics literature. Relative difference, P-values and CCC are from this study. Features are ordered by the mean relative difference (over PET_{40%} and PET_{MoCo}). Relative difference is shown with the $\pm 95\%$ confidence interval.

Feature name	PET _{40%}			PET _{MoCo}			Reference
	Relative difference [%]	P-value	CCC	Relative difference [%]	P-value	CCC	
1 Original GLCM ldmn	-0.26 \pm 0.19	0.01	0.984	-0.25 \pm 0.36	0.18	0.961	(S. S. Yip et al. 2017)
2 Original GLCM lmc2	-1.74 \pm 2.43	0.23	0.914	0.37 \pm 0.85	0.25	0.982	(Dissaux et al. 2020)
3 Original Shape Sphericity	-5.06 \pm 3.94	0.00	0.964	-0.62 \pm 1.46	0.62	0.996	(S. S. Yip et al. 2017; Hatt et al. 2018)
4 Original Firstorder Kurtosis	5.59 \pm 5.27	0.04	0.934	1.17 \pm 6.57	0.62	0.911	(Oikonomou et al. 2018)
5 Original GLCM Id	-4.51 \pm 3.73	0.00	0.984	-2.47 \pm 3.54	0.23	0.993	(Oikonomou et al. 2018; Pyka et al. 2015; Tixier et al. 2014)
6 Original Shape VoxelVolume	-6.34 \pm 5.68	0.02	0.977	-0.69 \pm 7.74	0.07	0.992	(S. S. Yip et al. 2017; Bianconi, Palumbo, Fravolini, et al. 2019)
7 Original GLCM JointAverage	2.95 \pm 5.68	0.65	0.976	5.10 \pm 6.32	0.08	0.971	(Ohri et al. 2016)
8 Original GLCM Correlation	-6.42 \pm 3.33	0.00	0.925	-2.58 \pm 4.47	0.11	0.967	(Pyka et al. 2015)
9 Original Firstorder Mean	4.44 \pm 5.23	0.14	0.986	5.64 \pm 4.79	0.02	0.983	(Bianconi, Palumbo, Fravolini, et al. 2019)
10 Original Firstorder Maximum	5.60 \pm 5.83	0.11	0.980	5.85 \pm 5.15	0.04	0.980	(S. S. Yip et al. 2017; Yoon et al. 2015; Kirienko et al. 2018)
11 Original NGTDM Busyness	-4.57 \pm 9.02	0.19	0.992	-6.94 \pm 10.72	0.41	0.976	(S. S. Yip et al. 2017; Pyka et al. 2015)
12 Original Firstorder Minimum	5.68 \pm 5.80	0.11	0.980	5.92 \pm 5.19	0.04	0.980	(S. S. Yip et al. 2017; Kirienko et al. 2018)
13 Original GLRLM LongRunLowGrayLevelEmphasis	-9.76 \pm 19.87	0.05	0.923	-3.00 \pm 28.03	0.14	0.701	(S. S. Yip et al. 2017)
14 Original GLSZM SizeZoneNonUniformity	8.63 \pm 9.99	0.59	0.995	-4.34 \pm 13.78	0.83	0.994	(S. S. Yip et al. 2017)
15 Original GLSZM ZonePercentage	13.56 \pm 11.79	0.01	0.972	-0.98 \pm 10.02	0.14	0.980	(Tixier et al. 2014)
16 Original GLRLM LongRunHighGrayLevelEmphasis	2.88 \pm 12.75	0.87	0.979	11.78 \pm 12.29	0.02	0.971	(Kirienko et al. 2018)
17 Original GLCM DifferenceAverage	10.27 \pm 6.47	0.01	0.969	7.31 \pm 6.61	0.02	0.964	(Tixier et al. 2014)
18 Original GLDM HighGrayLevelEmphasis	6.63 \pm 12.62	0.83	0.980	11.99 \pm 13.14	0.03	0.969	(Tixier et al. 2014)
19 Original GLRLM HighGrayLevelRunEmphasis	7.06 \pm 12.69	0.79	0.981	12.15 \pm 13.10	0.03	0.970	(Kirienko et al. 2018)
20 Original Firstorder Variance	8.00 \pm 13.01	0.83	0.975	13.99 \pm 13.80	0.05	0.928	(Tixier et al. 2014)
21 Original GLSZM LowGrayLevelZoneEmphasis	-12.98 \pm 18.14	0.01	0.960	-16.79 \pm 18.82	0.09	0.692	(Bianconi, Palumbo, Fravolini, et al. 2019)
22 Original NGTDM Strength	16.04 \pm 17.97	0.18	0.960	16.57 \pm 17.77	0.38	0.939	(S. S. Yip et al. 2017)
23 Original GLSZM SmallAreaHighGrayLevelEmphasis	19.53 \pm 22.14	0.41	0.978	23.71 \pm 29.08	0.05	0.957	(Dissaux et al. 2020)
24 Original Firstorder Skewness	-66.31 \pm 154.80	0.02	0.896	7.56 \pm 17.26	0.98	0.824	(Kirienko et al. 2018)

Supplementary material

Motion-robust radiomic features for image classification in ^{18}F -FDG PET/MRI imaging of lung cancer

Samuel Kuttner^{1,2,3*}, Erna-Elise Paulsen^{4,5}, Robert Jenssen², Rune Sundset^{1,3}, Jan Axelsson⁶

¹Nuclear Medicine and Radiation Biology Research Group, Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway.

²UiT Machine Learning Group, Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway.

³The PET Imaging Center, University Hospital of North Norway, Tromsø, Norway.

⁴Translational Cancer Research Group, Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway.

⁵Department of pulmonology, University Hospital of North Norway, Tromsø, Norway.

⁶Department of Radiation Sciences, Umeå University, Umeå, Sweden.

*Corresponding and first author:

Samuel Kuttner, Nuclear Medicine and Radiation Biology Research Group, Department of Clinical Medicine, UiT The Arctic University of Norway, 9037 Tromsø, Norway,

Phone: +47-77 66 99 53, e-mail: samuel.kuttner@uit.no

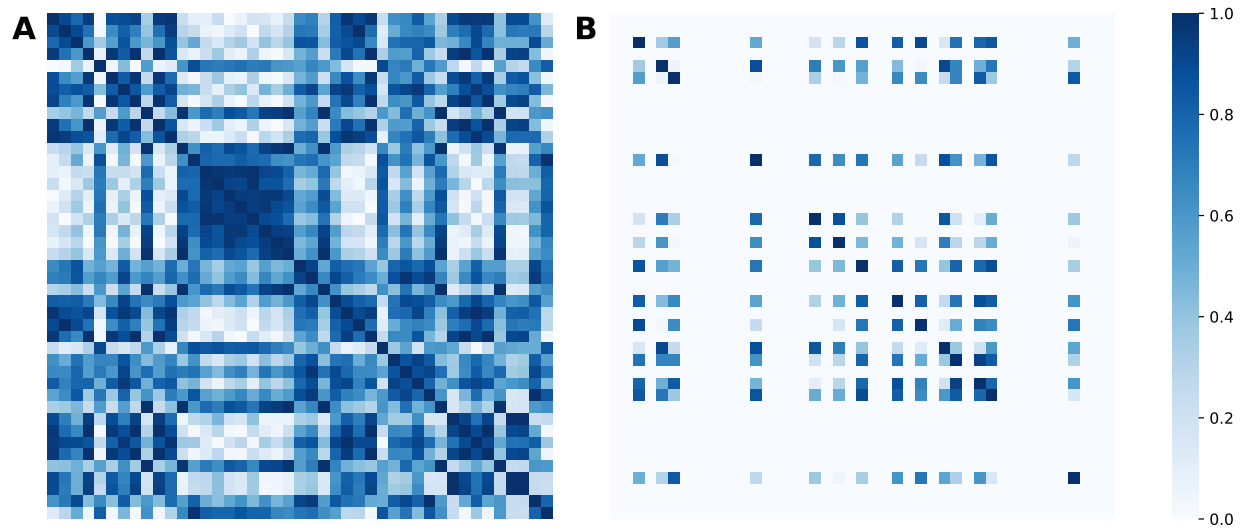


Figure S1: Plots of the absolute value of pair-wise Pearson correlation coefficients for the 43 highly motion robust features from Table 2. A. Before removal of correlated features. B. After removal of correlated features using a threshold of $M < 0.95$.

Table S1: All features with PET_{40%} relative difference < 5%, P-value > 0.05, and CCC > 0.95. Features that have been used in recent radiomics literature for disease state prediction are marked with the corresponding reference. Relative difference is shown with the ± 95% confidence interval.

Feature name	PET _{40%}		CCC	Reference
	Relative difference [%]	P-value		
1 Original Firstorder Energy	1.48 ± 7.85	0.08	0.998	
2 Original Firstorder Entropy	0.69 ± 1.36	0.62	0.991	[1]
3 Original Firstorder InterquartileRange	0.19 ± 7.00	0.69	0.959	
4 Original Firstorder Mean	4.44 ± 5.23	0.14	0.986	[1]
5 Original Firstorder MeanAbsoluteDeviation	2.30 ± 5.90	0.94	0.977	
6 Original Firstorder Median	4.49 ± 5.20	0.12	0.986	
7 Original Firstorder RobustMeanAbsoluteDeviation	1.03 ± 6.30	0.79	0.970	
8 Original Firstorder RootMeanSquared	4.40 ± 5.22	0.12	0.986	
9 Original Firstorder SUV peak	2.24 ± 4.84	0.80	0.993	
10 Original Firstorder TotalEnergy	1.48 ± 7.85	0.08	0.998	
11 Original Firstorder Uniformity	-0.92 ± 4.47	0.59	0.994	[1]
12 Original GLCM Imc1	-2.21 ± 9.63	0.36	0.974	
13 Original GLCM JointAverage	2.95 ± 5.68	0.65	0.976	[6]
14 Original GLCM JointEnergy	-3.83 ± 5.08	0.14	0.989	[5]
15 Original GLCM JointEntropy	0.89 ± 1.18	0.10	0.997	[2, 7-9]
16 Original GLCM MaximumProbability	-2.86 ± 7.85	0.36	0.995	[10]
17 Original GLCM SumEntropy	0.51 ± 1.24	0.83	0.992	
18 Original GLDM DependenceNonUniformity	-1.65 ± 5.06	0.29	0.995	
19 Original GLRLM GrayLevelNonUniformityNormalized	-1.10 ± 4.40	0.62	0.995	
20 Original GLRLM LongRunHighGrayLevelEmphasis	2.88 ± 12.75	0.87	0.979	[5]
21 Original GLRLM RunEntropy	-0.07 ± 1.55	0.65	0.982	
22 Original GLRLM RunLengthNonUniformity	-4.04 ± 4.85	0.06	0.990	
23 Original GLSZM GrayLevelNonUniformity	-3.06 ± 6.82	0.31	0.991	
24 Original GLSZM ZoneEntropy	1.19 ± 3.83	0.27	0.953	
25 Original NGTDM Busyness	-4.57 ± 9.02	0.19	0.992	[8, 10]
26 Original NGTDM Coarseness	-1.37 ± 3.63	0.29	0.991	[8]
27 Original Shape Flatness	0.75 ± 2.73	0.87	0.962	
28 Original Shape MajorAxisLength	-1.39 ± 2.00	0.10	0.999	
29 Original Shape Maximum2DDiameterColumn	-0.46 ± 2.25	0.55	0.998	
30 Original Shape Maximum2DDiameterRow	1.79 ± 4.17	0.43	0.992	
31 Original Shape Maximum2DDiameterSlice	-1.32 ± 2.89	0.33	0.998	
32 Original Shape Maximum3DDiameter	-0.89 ± 1.99	0.12	0.999	
33 Original Shape MinorAxisLength	-0.01 ± 2.42	0.62	0.999	
34 Original Shape SurfaceArea	0.97 ± 4.25	0.83	0.994	
35 Wavelet-HHH GLRLM RunLengthNonUniformity	1.82 ± 5.48	0.98	0.996	
36 Wavelet-HLL GLCM Imc1	2.31 ± 9.09	0.72	0.969	
37 Wavelet-HLL GLDM DependenceEntropy	0.85 ± 1.10	0.27	0.966	
38 Wavelet-HLL GLRLM RunLengthNonUniformity	4.11 ± 5.74	0.87	0.995	
39 Wavelet-HLL NGTDM Coarseness	-0.98 ± 7.82	0.38	0.971	
40 Wavelet-LHH GLDM DependenceEntropy	0.33 ± 1.35	0.69	0.954	
41 Wavelet-LHL GLCM Imc1	-2.31 ± 7.32	0.41	0.977	
42 Wavelet-LHL GLCM Imc2	0.86 ± 2.27	0.79	0.969	
43 Wavelet-LHL GLDM DependenceEntropy	-0.30 ± 0.85	0.12	0.984	
44 Wavelet-LHL GLRLM RunLengthNonUniformity	0.79 ± 5.43	0.52	0.992	
45 Wavelet-LHL GLSZM ZoneEntropy	0.59 ± 1.61	0.72	0.971	
46 Wavelet-LHL NGTDM Coarseness	-0.19 ± 6.93	0.79	0.988	
47 Wavelet-LLH GLCM Imc2	-1.07 ± 2.21	0.18	0.984	
48 Wavelet-LLH GLDM DependenceEntropy	-0.04 ± 1.03	0.49	0.982	
49 Wavelet-LLL Firstorder 90Percentile	4.24 ± 5.10	0.14	0.986	
50 Wavelet-LLL Firstorder Energy	1.08 ± 7.87	0.08	0.998	
51 Wavelet-LLL Firstorder Entropy	0.44 ± 0.97	0.65	0.993	
52 Wavelet-LLL Firstorder InterquartileRange	-1.10 ± 6.60	0.23	0.966	
53 Wavelet-LLL Firstorder Maximum	3.97 ± 5.25	0.19	0.987	
54 Wavelet-LLL Firstorder Mean	4.25 ± 5.19	0.14	0.986	
55 Wavelet-LLL Firstorder MeanAbsoluteDeviation	1.45 ± 5.96	0.69	0.976	
56 Wavelet-LLL Firstorder Median	4.10 ± 5.18	0.15	0.986	
57 Wavelet-LLL Firstorder Minimum	3.37 ± 5.37	0.46	0.984	
58 Wavelet-LLL Firstorder Range	4.42 ± 5.29	0.12	0.987	
59 Wavelet-LLL Firstorder RobustMeanAbsoluteDeviation	0.20 ± 6.46	0.69	0.967	
60 Wavelet-LLL Firstorder RootMeanSquared	4.18 ± 5.19	0.15	0.986	
61 Wavelet-LLL Firstorder TotalEnergy	1.08 ± 7.87	0.08	0.998	
62 Wavelet-LLL Firstorder Uniformity	-1.37 ± 4.14	0.59	0.996	
63 Wavelet-LLL GLCM Imc1	0.17 ± 5.84	0.59	0.985	
64 Wavelet-LLL GLCM JointEnergy	1.87 ± 5.57	0.98	0.977	
65 Wavelet-LLL GLCM JointEntropy	-0.17 ± 1.06	0.41	0.997	
66 Wavelet-LLL GLCM MaximumProbability	-1.92 ± 7.52	0.46	0.983	
67 Wavelet-LLL GLCM SumEntropy	0.11 ± 0.94	0.83	0.995	
68 Wavelet-LLL GLDM DependenceEntropy	-0.65 ± 0.99	0.12	0.983	
69 Wavelet-LLL GLDM DependenceNonUniformity	-2.68 ± 3.26	0.09	0.994	
70 Wavelet-LLL GLDM SmallDependenceEmphasis	4.26 ± 6.06	0.06	0.982	
71 Wavelet-LLL GLRLM GrayLevelNonUniformityNormalized	-1.39 ± 4.13	0.76	0.996	
72 Wavelet-LLL GLRLM RunEntropy	0.13 ± 1.06	0.98	0.991	
73 Wavelet-LLL GLSZM GrayLevelNonUniformity	-4.16 ± 6.10	0.08	0.982	
74 Wavelet-LLL GLSZM GrayLevelNonUniformityNormalized	-2.21 ± 4.05	0.38	0.994	
75 Wavelet-LLL GLSZM GrayLevelVariance	4.29 ± 12.36	0.59	0.974	
76 Wavelet-LLL GLSZM SizeZoneNonUniformity	-1.47 ± 5.65	0.16	0.994	
77 Wavelet-LLL GLSZM SizeZoneNonUniformityNormalized	0.59 ± 2.81	0.38	0.973	
78 Wavelet-LLL GLSZM SmallAreaEmphasis	0.16 ± 1.28	0.46	0.985	
79 Wavelet-LLL GLSZM ZoneEntropy	0.33 ± 0.79	0.69	0.990	
80 Wavelet-LLL NGTDM Coarseness	2.32 ± 3.33	0.49	0.998	

Table S2: All features with PET_{MoCo} relative difference $< 5\%$, P-value > 0.05 and CCC > 0.95 . Features that have been used in recent radiomics literature for disease state prediction are marked with the corresponding reference. Relative difference is shown with the $\pm 95\%$ confidence interval.

	Feature name	PET_{MoCo}			Reference
		Relative difference [%]	CCC	P-value	
1	Original Firstorder Entropy	1.08 ± 1.54	0.14	0.988	[1]
2	Original Firstorder Median	5.00 ± 4.71	0.06	0.984	
3	Original Firstorder SUV peak	4.98 ± 4.52	0.08	0.987	
4	Original Firstorder Uniformity	-3.34 ± 4.97	0.83	0.994	[1]
5	Original GLCM Correlation	-2.58 ± 4.47	0.11	0.967	[8]
6	Original GLCM Id	-2.47 ± 3.54	0.23	0.993	[7–9]
7	Original GLCM Idm	-2.10 ± 6.12	0.36	0.996	
8	Original GLCM Idmn	-0.25 ± 0.36	0.18	0.961	[10]
9	Original GLCM Idn	-0.20 ± 0.41	0.41	0.983	
10	Original GLCM Imc1	4.01 ± 7.09	0.27	0.972	
11	Original GLCM Imc2	0.37 ± 0.85	0.25	0.982	[3]
12	Original GLCM JointEnergy	-4.93 ± 6.40	0.49	0.991	[5]
13	Original GLCM JointEntropy	0.73 ± 1.18	0.15	0.995	[2, 7–9]
14	Original GLCM MaximumProbability	-0.69 ± 7.89	0.72	0.981	[10]
15	Original GLDM DependenceEntropy	0.81 ± 1.07	0.55	0.983	
16	Original GLDM DependenceNonUniformity	-0.46 ± 7.71	0.55	0.994	
17	Original GLDM DependenceVariance	-4.24 ± 6.82	0.27	0.996	
18	Original GLDM LargeDependenceEmphasis	-1.05 ± 7.00	0.38	0.993	
19	Original GLDM SmallDependenceEmphasis	-1.17 ± 6.75	0.19	0.960	
20	Original GLRLM GrayLevelNonUniformityNormalized	-3.27 ± 4.93	0.79	0.993	
21	Original GLRLM LongRunEmphasis	0.38 ± 1.36	0.62	0.991	
22	Original GLRLM RunEntropy	1.15 ± 1.19	0.16	0.983	
23	Original GLRLM RunLengthNonUniformity	-1.01 ± 7.73	0.08	0.992	
24	Original GLRLM RunLengthNonUniformityNormalized	-0.16 ± 0.75	0.62	0.996	
25	Original GLRLM RunPercentage	-0.15 ± 0.49	0.62	0.996	
26	Original GLRLM RunVariance	4.17 ± 14.81	0.52	0.991	
27	Original GLRLM ShortRunEmphasis	-0.10 ± 0.35	0.65	0.996	
28	Original GLSZM SizeZoneNonUniformity	-4.34 ± 13.78	0.83	0.994	[10]
29	Original GLSZM ZonePercentage	-0.98 ± 10.02	0.14	0.980	[9]
30	Original NGTDM Coarseness	-1.98 ± 5.42	0.27	0.980	[8]
31	Original Shape LeastAxisLength	0.55 ± 2.01	0.69	0.997	
32	Original Shape MajorAxisLength	-1.66 ± 3.16	0.15	0.998	
33	Original Shape Maximum2DDiameterColumn	-0.38 ± 4.03	0.57	0.995	
34	Original Shape Maximum2DDiameterRow	0.15 ± 3.68	0.79	0.993	
35	Original Shape Maximum2DDiameterSlice	-1.22 ± 3.66	0.33	0.996	
36	Original Shape Maximum3DDiameter	-0.91 ± 2.82	0.47	0.998	
37	Original Shape MeshVolume	-0.66 ± 7.85	0.07	0.993	
38	Original Shape MinorAxisLength	2.05 ± 3.84	0.41	0.992	
39	Original Shape Sphericity	-0.62 ± 1.46	0.62	0.996	[4, 10]
40	Original Shape SurfaceArea	0.13 ± 5.91	0.23	0.996	
41	Original Shape SurfaceVolumeRatio	1.34 ± 2.25	0.27	0.990	
42	Original Shape VoxelVolume	-0.69 ± 7.74	0.07	0.992	[1, 10]
43	Wavelet-HHH GLDM GrayLevelNonUniformity	-0.47 ± 7.79	0.07	0.993	
44	Wavelet-HHH GLDM SmallDependenceEmphasis	-2.85 ± 5.31	0.08	0.967	
45	Wavelet-HHL Firstorder 90Percentile	0.79 ± 7.69	0.79	0.956	
46	Wavelet-HHL Firstorder Energy	-0.19 ± 13.45	0.21	0.978	
47	Wavelet-HHL Firstorder Entropy	-0.03 ± 3.81	0.62	0.968	
48	Wavelet-HHL Firstorder MeanAbsoluteDeviation	0.30 ± 8.13	0.65	0.957	
49	Wavelet-HHL Firstorder RootMeanSquared	0.23 ± 7.74	0.46	0.964	
50	Wavelet-HHL Firstorder TotalEnergy	-0.19 ± 13.45	0.21	0.978	
51	Wavelet-HHL Firstorder Uniformity	1.54 ± 6.17	0.79	0.968	
52	Wavelet-HHL GLCM JointEntropy	-0.75 ± 3.68	0.41	0.964	
53	Wavelet-HHL GLCM MaximumProbability	4.76 ± 9.95	0.41	0.968	
54	Wavelet-HHL GLCM SumEntropy	0.72 ± 3.07	0.91	0.968	
55	Wavelet-HHL GLDM DependenceNonUniformity	-4.76 ± 7.01	0.06	0.979	
56	Wavelet-HHL GLDM GrayLevelNonUniformity	1.81 ± 13.09	0.59	0.991	
57	Wavelet-HHL GLDM LowGrayLevelEmphasis	2.96 ± 15.65	0.49	0.995	

Continued on next page.

Table S2 – continued from previous page

Feature name	PET _{MoCo}		CCC	Reference
	Relative difference [%]	P-value		
58 Wavelet-HHL GLRLM GrayLevelNonUniformity	-0.53 ± 11.11	0.23	0.989	
59 Wavelet-HHL GLRLM GrayLevelNonUniformityNormalized	1.49 ± 5.85	0.83	0.973	
60 Wavelet-HHL GLRLM LowGrayLevelRunEmphasis	2.66 ± 16.01	0.65	0.994	
61 Wavelet-HHL GLRLM RunEntropy	0.89 ± 1.17	0.08	0.962	
62 Wavelet-HHL GLRLM RunLengthNonUniformityNormalized	-2.56 ± 2.91	0.06	0.959	
63 Wavelet-HHL GLRLM RunPercentage	-1.57 ± 1.89	0.08	0.964	
64 Wavelet-HHL GLRLM ShortRunEmphasis	-1.25 ± 1.57	0.08	0.973	
65 Wavelet-HHL GLRLM ShortRunLowGrayLevelEmphasis	1.23 ± 16.45	0.41	0.985	
66 Wavelet-HHL GLSzM GrayLevelNonUniformityNormalized	0.40 ± 5.83	0.88	0.975	
67 Wavelet-HHL GLSzM GrayLevelVariance	1.12 ± 8.87	0.68	0.968	
68 Wavelet-HHL NGTDM Busyness	-3.07 ± 18.37	0.18	0.963	
69 Wavelet-HLH Firstorder Energy	2.67 ± 18.04	0.69	0.966	
70 Wavelet-HLH Firstorder Maximum	3.68 ± 13.09	0.65	0.954	
71 Wavelet-HLH Firstorder Range	1.42 ± 11.76	0.79	0.957	
72 Wavelet-HLH Firstorder TotalEnergy	2.67 ± 18.04	0.69	0.966	
73 Wavelet-HLH GLDM DependenceNonUniformity	-4.00 ± 7.47	0.06	0.984	
74 Wavelet-HLH GLDM GrayLevelNonUniformity	4.14 ± 13.47	0.41	0.991	
75 Wavelet-HLH GLDM LowGrayLevelEmphasis	4.06 ± 18.10	0.18	0.980	
76 Wavelet-HLH GLRLM GrayLevelNonUniformity	0.52 ± 10.46	0.08	0.991	
77 Wavelet-HLH GLRLM LowGrayLevelRunEmphasis	4.13 ± 18.28	0.31	0.980	
78 Wavelet-HLH GLRLM ShortRunLowGrayLevelEmphasis	1.61 ± 17.69	0.29	0.961	
79 Wavelet-HLL Firstorder Entropy	2.70 ± 3.22	0.08	0.961	
80 Wavelet-HLL GLCM DifferenceEntropy	1.67 ± 3.79	0.43	0.967	
81 Wavelet-HLL GLCM Id	-0.07 ± 4.82	0.83	0.972	
82 Wavelet-HLL GLCM Idm	0.26 ± 7.27	0.94	0.973	
83 Wavelet-HLL GLCM InverseVariance	0.14 ± 7.55	0.46	0.954	
84 Wavelet-HLL GLDM DependenceNonUniformity	0.11 ± 7.33	0.52	0.991	
85 Wavelet-HLL GLDM DependenceVariance	-1.24 ± 10.80	0.49	0.977	
86 Wavelet-HLL GLDM LargeDependenceEmphasis	0.16 ± 10.48	0.83	0.974	
87 Wavelet-HLL GLRLM LongRunEmphasis	0.45 ± 3.95	0.76	0.967	
88 Wavelet-HLL GLRLM RunLengthNonUniformity	-1.09 ± 7.71	0.15	0.991	
89 Wavelet-HLL GLRLM RunLengthNonUniformityNormalized	-0.11 ± 2.10	0.98	0.978	
90 Wavelet-HLL GLRLM RunPercentage	-0.16 ± 1.27	0.91	0.979	
91 Wavelet-HLL GLRLM RunVariance	3.15 ± 12.78	0.76	0.970	
92 Wavelet-HLL GLRLM ShortRunEmphasis	-0.15 ± 1.10	0.94	0.976	
93 Wavelet-HLL GLSzM GrayLevelNonUniformityNormalized	-4.35 ± 7.14	0.18	0.972	
94 Wavelet-HLL GLSzM LargeAreaEmphasis	3.04 ± 26.24	0.49	0.964	
95 Wavelet-HLL GLSzM ZoneEntropy	1.08 ± 1.18	0.06	0.982	
96 Wavelet-LHH Firstorder 10Percentile	-4.93 ± 6.67	0.27	0.974	
97 Wavelet-LHH Firstorder 90Percentile	0.21 ± 7.85	0.94	0.963	
98 Wavelet-LHH Firstorder Energy	-0.42 ± 15.44	0.27	0.952	
99 Wavelet-LHH Firstorder Entropy	0.25 ± 4.40	0.83	0.969	
100 Wavelet-LHH Firstorder InterquartileRange	0.07 ± 7.66	0.94	0.969	
101 Wavelet-LHH Firstorder MeanAbsoluteDeviation	-0.77 ± 7.05	0.87	0.976	
102 Wavelet-LHH Firstorder Minimum	1.15 ± 8.53	0.83	0.954	
103 Wavelet-LHH Firstorder RobustMeanAbsoluteDeviation	-0.56 ± 7.39	0.87	0.968	
104 Wavelet-LHH Firstorder RootMeanSquared	-0.69 ± 7.28	0.91	0.976	
105 Wavelet-LHH Firstorder TotalEnergy	-0.42 ± 15.44	0.27	0.952	
106 Wavelet-LHH Firstorder Uniformity	1.04 ± 4.85	0.62	0.976	
107 Wavelet-LHH Firstorder Variance	0.14 ± 14.61	0.91	0.963	
108 Wavelet-LHH GLCM ClusterTendency	4.53 ± 11.65	0.87	0.950	
109 Wavelet-LHH GLCM Contrast	-3.08 ± 10.95	0.55	0.966	
110 Wavelet-LHH GLCM DifferenceAverage	-3.56 ± 5.97	0.27	0.962	
111 Wavelet-LHH GLCM DifferenceEntropy	-1.66 ± 3.66	0.49	0.967	
112 Wavelet-LHH GLCM DifferenceVariance	-0.26 ± 10.44	0.69	0.958	
113 Wavelet-LHH GLCM Id	1.38 ± 2.45	0.14	0.961	
114 Wavelet-LHH GLCM Idm	1.60 ± 3.28	0.15	0.960	
115 Wavelet-LHH GLCM Idmnn	0.21 ± 0.39	0.21	0.951	
116 Wavelet-LHH GLCM JointEnergy	4.99 ± 10.46	0.21	0.974	
117 Wavelet-LHH GLCM JointEntropy	-0.59 ± 4.39	0.72	0.964	
118 Wavelet-LHH GLCM SumEntropy	0.41 ± 3.53	0.94	0.970	

Continued on next page.

Table S2 – continued from previous page

Feature name	PET _{MoCo}		CCC	Reference
	Relative difference [%]	P-value		
119 Wavelet-LHH GLCM SumSquares	1.65 ± 11.19	0.87	0.957	
120 Wavelet-LHH GLDM DependenceEntropy	0.90 ± 0.84	0.06	0.975	
121 Wavelet-LHH GLDM GrayLevelNonUniformity	0.54 ± 9.76	0.65	0.992	
122 Wavelet-LHH GLDM GrayLevelVariance	1.81 ± 11.08	0.98	0.966	
123 Wavelet-LHH GLDM LowGrayLevelEmphasis	4.19 ± 18.92	0.91	0.960	
124 Wavelet-LHH GLDM SmallDependenceEmphasis	-2.43 ± 11.66	0.46	0.956	
125 Wavelet-LHH GLRLM GrayLevelNonUniformity	-2.79 ± 8.02	0.19	0.993	
126 Wavelet-LHH GLRLM GrayLevelNonUniformityNormalized	0.30 ± 4.92	0.98	0.974	
127 Wavelet-LHH GLRLM GrayLevelVariance	3.15 ± 11.59	0.91	0.964	
128 Wavelet-LHH GLRLM LowGrayLevelRunEmphasis	4.40 ± 18.92	0.87	0.967	
129 Wavelet-LHH GLRLM RunLengthNonUniformityNormalized	-2.79 ± 2.94	0.10	0.966	
130 Wavelet-LHH GLRLM RunPercentage	-2.04 ± 1.97	0.07	0.963	
131 Wavelet-LHH GLRLM ShortRunEmphasis	-1.63 ± 1.74	0.09	0.970	
132 Wavelet-LHH GLSZM ZoneEntropy	-0.84 ± 7.28	0.87	0.953	
133 Wavelet-LHH GLSZM ZonePercentage	-1.41 ± 16.33	0.83	0.955	
134 Wavelet-LHL Firstorder Entropy	1.27 ± 1.88	0.19	0.988	
135 Wavelet-LHL Firstorder Maximum	4.76 ± 8.16	0.29	0.966	
136 Wavelet-LHL Firstorder Uniformity	-2.52 ± 5.24	0.55	0.996	
137 Wavelet-LHL GLCM DifferenceAverage	3.45 ± 5.94	0.19	0.977	
138 Wavelet-LHL GLCM DifferenceEntropy	1.35 ± 2.18	0.15	0.986	
139 Wavelet-LHL GLCM Id	-0.22 ± 3.84	0.87	0.990	
140 Wavelet-LHL GLCM Idm	0.69 ± 6.75	0.98	0.991	
141 Wavelet-LHL GLCM Idmn	0.06 ± 0.21	0.14	0.976	
142 Wavelet-LHL GLCM Inverse Variance	-1.06 ± 6.18	0.29	0.979	
143 Wavelet-LHL GLCM JointEnergy	-3.45 ± 6.69	0.36	0.990	
144 Wavelet-LHL GLCM JointEntropy	0.76 ± 1.43	0.10	0.995	
145 Wavelet-LHL GLCM MaximumProbability	-1.58 ± 8.14	0.76	0.992	
146 Wavelet-LHL GLDM DependenceEntropy	0.67 ± 0.64	0.06	0.990	
147 Wavelet-LHL GLDM DependenceNonUniformity	-0.28 ± 9.55	0.18	0.990	
148 Wavelet-LHL GLDM LargeDependenceEmphasis	1.28 ± 14.85	0.72	0.990	
149 Wavelet-LHL GLDM SmallDependenceEmphasis	3.88 ± 10.96	0.59	0.954	
150 Wavelet-LHL GLRLM GrayLevelNonUniformityNormalized	-2.78 ± 5.20	0.55	0.996	
151 Wavelet-LHL GLRLM LongRunEmphasis	0.33 ± 2.99	0.87	0.980	
152 Wavelet-LHL GLRLM RunEntropy	1.36 ± 1.14	0.07	0.981	
153 Wavelet-LHL GLRLM RunLengthNonUniformity	-0.90 ± 8.10	0.08	0.992	
154 Wavelet-LHL GLRLM RunLengthNonUniformityNormalized	-0.10 ± 1.61	0.79	0.991	
155 Wavelet-LHL GLRLM RunPercentage	-0.13 ± 1.02	0.87	0.992	
156 Wavelet-LHL GLRLM ShortRunEmphasis	-0.19 ± 0.91	0.94	0.991	
157 Wavelet-LHL GLSZM GrayLevelNonUniformity	-2.54 ± 9.23	0.07	0.992	
158 Wavelet-LHL GLSZM GrayLevelNonUniformityNormalized	-3.73 ± 6.52	0.33	0.984	
159 Wavelet-LHL GLSZM ZoneEntropy	0.14 ± 1.32	0.27	0.986	
160 Wavelet-LHL GLSZM ZonePercentage	4.19 ± 11.03	0.59	0.966	
161 Wavelet-LLH GLCM DifferenceAverage	4.42 ± 7.99	0.55	0.975	
162 Wavelet-LLH GLCM DifferenceEntropy	2.64 ± 4.05	0.14	0.984	
163 Wavelet-LLH GLCM Id	-0.60 ± 4.15	0.69	0.987	
164 Wavelet-LLH GLCM Idm	-0.11 ± 6.76	0.65	0.987	
165 Wavelet-LLH GLCM Inverse Variance	-0.25 ± 7.76	0.31	0.951	
166 Wavelet-LLH GLDM DependenceNonUniformity	0.67 ± 7.58	0.59	0.990	
167 Wavelet-LLH GLDM DependenceNonUniformityNormalized	2.10 ± 6.23	0.79	0.958	
168 Wavelet-LLH GLDM DependenceVariance	-0.10 ± 10.66	0.76	0.982	
169 Wavelet-LLH GLDM LargeDependenceEmphasis	-0.92 ± 10.28	0.98	0.988	
170 Wavelet-LLH GLDM SmallDependenceEmphasis	4.29 ± 11.39	0.65	0.964	
171 Wavelet-LLH GLDM SmallDependenceLowGrayLevelEmphasis	-1.88 ± 16.86	0.09	0.976	
172 Wavelet-LLH GLRLM LongRunEmphasis	0.27 ± 4.66	0.98	0.976	
173 Wavelet-LLH GLRLM LongRunLowGrayLevelEmphasis	-4.99 ± 19.94	0.49	0.969	
174 Wavelet-LLH GLRLM RunLengthNonUniformity	-0.34 ± 8.26	0.10	0.993	
175 Wavelet-LLH GLRLM RunLengthNonUniformityNormalized	0.36 ± 2.69	0.65	0.990	
176 Wavelet-LLH GLRLM RunPercentage	-0.02 ± 1.78	0.79	0.990	
177 Wavelet-LLH GLRLM RunVariance	1.85 ± 12.01	0.98	0.977	
178 Wavelet-LLH GLRLM ShortRunEmphasis	-0.17 ± 1.68	0.83	0.989	
179 Wavelet-LLH GLSZM GrayLevelNonUniformity	1.79 ± 10.62	0.55	0.988	

Continued on next page.

Table S2 – continued from previous page

Feature name	PET _{MoCo}		CCC	Reference
	Relative difference [%]	P-value		
180	Wavelet-LLH GLSZM GrayLevelNonUniformityNormalized	-4.59 ± 6.54	0.23	0.983
181	Wavelet-LLH NGTDM Contrast	1.90 ± 8.18	0.41	0.995
182	Wavelet-LLL Firstorder Entropy	0.73 ± 0.85	0.12	0.994
183	Wavelet-LLL Firstorder Median	4.79 ± 4.78	0.06	0.984
184	Wavelet-LLL Firstorder Uniformity	-3.08 ± 3.86	0.55	0.996
185	Wavelet-LLL GLCM Correlation	-2.43 ± 4.24	0.06	0.968
186	Wavelet-LLL GLCM Id	-4.19 ± 3.25	0.08	0.996
187	Wavelet-LLL GLCM Idmn	-0.21 ± 0.20	0.06	0.985
188	Wavelet-LLL GLCM Idn	-0.24 ± 0.32	0.16	0.988
189	Wavelet-LLL GLCM Imc1	2.48 ± 3.13	0.19	0.987
190	Wavelet-LLL GLCM JointEnergy	-0.90 ± 6.06	0.52	0.992
191	Wavelet-LLL GLCM JointEntropy	0.22 ± 1.08	0.52	0.996
192	Wavelet-LLL GLCM MaximumProbability	-4.03 ± 6.08	0.49	0.993
193	Wavelet-LLL GLDM DependenceEntropy	0.40 ± 0.82	0.49	0.989
194	Wavelet-LLL GLDM DependenceNonUniformity	0.67 ± 7.33	0.11	0.992
195	Wavelet-LLL GLDM DependenceNonUniformityNormalized	1.65 ± 2.51	0.41	0.984
196	Wavelet-LLL GLDM LargeDependenceEmphasis	-2.23 ± 3.22	0.15	0.997
197	Wavelet-LLL GLDM SmallDependenceEmphasis	-0.13 ± 1.97	0.55	0.995
198	Wavelet-LLL GLRLM GrayLevelNonUniformityNormalized	-3.06 ± 3.87	0.65	0.996
199	Wavelet-LLL GLRLM LongRunEmphasis	-0.03 ± 0.40	0.23	0.997
200	Wavelet-LLL GLRLM RunEntropy	0.72 ± 0.79	0.14	0.993
201	Wavelet-LLL GLRLM RunLengthNonUniformity	-0.65 ± 7.74	0.08	0.992
202	Wavelet-LLL GLRLM RunLengthNonUniformityNormalized	0.04 ± 0.20	0.46	0.998
203	Wavelet-LLL GLRLM RunPercentage	0.01 ± 0.12	0.41	0.998
204	Wavelet-LLL GLRLM RunVariance	-2.23 ± 6.88	0.12	0.996
205	Wavelet-LLL GLRLM ShortRunEmphasis	0.01 ± 0.08	0.43	0.998
206	Wavelet-LLL GLSZM GrayLevelNonUniformityNormalized	-3.22 ± 3.92	0.52	0.996
207	Wavelet-LLL GLSZM LargeAreaEmphasis	-1.36 ± 3.41	0.41	0.994
208	Wavelet-LLL GLSZM SizeZoneNonUniformity	-1.31 ± 8.47	0.11	0.993
209	Wavelet-LLL GLSZM SizeZoneNonUniformityNormalized	-0.46 ± 2.51	0.79	0.988
210	Wavelet-LLL GLSZM SmallAreaEmphasis	-0.37 ± 1.39	0.87	0.989
211	Wavelet-LLL GLSZM ZoneEntropy	0.82 ± 0.86	0.08	0.986
212	Wavelet-LLL GLSZM ZonePercentage	-0.28 ± 1.89	0.38	0.997
213	Wavelet-LLL NGTDM Coarseness	-2.68 ± 5.22	0.11	0.990

Table S3: All features with $PET_{40\%}$ and PET_{MoCo} relative difference $< 5\%$, P-value > 0.05 , and CCC > 0.95 for the nine most motion-prone patients. Features that are coinciding with Table 3 in the main paper have been marked with a dagger (\dagger). Features that have been used in recent radiomics literature for disease state prediction are marked with the corresponding reference. Relative difference is shown with the $\pm 95\%$ confidence interval.

	Feature name	$PET_{40\%}$			PET_{MoCo}			Reference
		Relative difference [%]	P-value	CCC	Relative difference [%]	P-value	CCC	
1	Original GLCM Idmn	-0.23 \pm 0.19	0.07	0.979	-0.24 \pm 0.23	0.11	0.971	[10]
2	Original GLDM DependenceEntropy	-0.32 \pm 0.71	0.44	0.993	0.68 \pm 1.05	0.44	0.987	
3	\dagger Original GLRLM RunLengthNonUniformity	-2.54 \pm 6.22	0.51	0.991	0.68 \pm 11.03	0.77	0.966	
4	Original GLSZM GrayLevelNonUniformity	-1.47 \pm 10.16	0.86	0.994	-3.07 \pm 9.19	0.59	0.984	
5	Original GLSZM ZoneEntropy	-0.28 \pm 0.81	0.37	0.993	0.96 \pm 1.51	0.26	0.972	
6	\dagger Original NGTDM Coarseness	-1.40 \pm 5.56	0.95	0.975	-4.20 \pm 7.76	0.21	0.966	[8]
7	Original Shape LeastAxisLength	-0.62 \pm 1.42	0.26	0.997	1.03 \pm 2.94	0.26	0.989	
8	\dagger Original Shape MajorAxisLength	-1.81 \pm 3.55	0.31	0.997	-2.58 \pm 4.67	0.21	0.992	
9	\dagger Original Shape Maximum2DDiameterColumn	-0.95 \pm 4.13	0.68	0.984	-0.84 \pm 7.35	0.95	0.961	
10	\dagger Original Shape Maximum2DDiameterRow	0.37 \pm 4.68	0.51	0.976	-1.77 \pm 4.34	0.51	0.971	
11	\dagger Original Shape Maximum2DDiameterSlice	-1.84 \pm 4.87	0.58	0.995	-1.46 \pm 6.63	0.77	0.989	
12	\dagger Original Shape Maximum3DDiameter	-1.67 \pm 3.44	0.44	0.996	-1.77 \pm 4.70	0.33	0.990	
13	Original Shape MeshVolume	-3.70 \pm 6.80	0.51	0.988	-0.22 \pm 11.31	0.77	0.966	
14	\dagger Original Shape MinorAxisLength	-0.37 \pm 2.18	0.68	0.998	2.24 \pm 5.54	0.77	0.971	
15	Original Shape Sphericity	-1.59 \pm 2.02	0.14	0.990	-1.37 \pm 1.76	0.37	0.993	[4, 10]
16	\dagger Original Shape SurfaceArea	-1.03 \pm 4.10	0.68	0.999	1.21 \pm 8.94	0.95	0.965	
17	Original Shape SurfaceVolumeRatio	3.27 \pm 3.87	0.14	0.980	1.97 \pm 2.73	0.17	0.992	
18	Original Shape VoxelVolume	-3.69 \pm 6.68	0.40	0.988	-0.24 \pm 11.23	0.77	0.966	[1, 10]
19	Wavelet-HHH GLRLM GrayLevelNonUniformity	-4.50 \pm 6.36	0.26	0.989	-2.47 \pm 11.01	0.51	0.965	
20	\dagger Wavelet-HLL GLRLM RunLengthNonUniformity	4.22 \pm 4.94	0.37	0.995	2.59 \pm 9.91	0.77	0.965	
21	Wavelet-HLL GLSZM ZoneEntropy	1.55 \pm 2.09	0.21	0.968	1.69 \pm 1.52	0.07	0.980	
22	\dagger Wavelet-LHH GLDM DependenceEntropy	0.02 \pm 1.76	0.95	0.952	1.05 \pm 1.48	0.26	0.954	
23	Wavelet-LHL GLCM Idmn	-0.20 \pm 0.18	0.07	0.969	0.10 \pm 0.24	0.59	0.962	
24	\dagger Wavelet-LHL GLDM DependenceEntropy	-0.26 \pm 1.34	0.26	0.966	0.66 \pm 0.78	0.17	0.983	
25	\dagger Wavelet-LHL GLRLM RunLengthNonUniformity	0.33 \pm 4.83	0.95	0.995	2.28 \pm 10.59	0.77	0.967	
26	\dagger Wavelet-LHL GLSZM ZoneEntropy	-0.44 \pm 0.95	0.21	0.989	0.67 \pm 1.13	0.26	0.984	
27	Wavelet-LLH GLDM DependenceEntropy	-0.28 \pm 0.80	0.59	0.989	1.18 \pm 1.35	0.11	0.965	
28	Wavelet-LLH GLRLM RunLengthNonUniformity	1.75 \pm 5.46	0.95	0.994	4.98 \pm 10.09	0.95	0.967	
29	Wavelet-LLH GLSZM ZoneEntropy	2.93 \pm 4.42	0.26	0.971	3.01 \pm 3.11	0.05	0.984	
30	Wavelet-LLL GLCM Idn	-0.27 \pm 0.29	0.11	0.988	-0.40 \pm 0.46	0.14	0.971	
31	Wavelet-LLL GLCM Imc2	-0.03 \pm 0.10	0.51	0.984	-0.01 \pm 0.06	0.51	0.993	
32	\dagger Wavelet-LLL GLCM JointEnergy	1.67 \pm 5.99	0.77	0.997	-0.99 \pm 8.64	0.44	0.992	
33	\dagger Wavelet-LLL GLCM JointEntropy	-0.35 \pm 0.88	0.68	0.996	0.07 \pm 1.31	0.86	0.992	
34	Wavelet-LLL GLCM SumEntropy	0.48 \pm 0.82	0.44	0.994	1.08 \pm 0.92	0.05	0.988	
35	\dagger Wavelet-LLL GLDM DependenceEntropy	0.40 \pm 0.93	0.51	0.993	1.04 \pm 0.99	0.09	0.986	
36	\dagger Wavelet-LLL GLDM DependenceNonUniformity	-0.45 \pm 4.49	0.59	0.991	3.12 \pm 11.73	0.86	0.956	
37	Wavelet-LLL GLRLM RunLengthNonUniformity	-3.41 \pm 6.46	0.44	0.989	0.07 \pm 11.25	0.77	0.965	
38	\dagger Wavelet-LLL GLSZM SizeZoneNonUniformity	-1.22 \pm 4.36	0.51	0.992	2.72 \pm 11.46	0.77	0.958	

References

- [1] Francesco Bianconi et al. “Texture Analysis on [18F]FDG PET/CT in Non-Small-Cell Lung Cancer: Correlations Between PET Features, CT Features, and Histological Types”. In: *Mol. Imaging Biol.* 21.6 (2019), pp. 1200–1209.
- [2] Marie Charlotte Desseroit et al. “Development of a nomogram combining clinical staging with 18F-FDG PET/CT image features in non-small-cell lung cancer stage I-III”. In: *Eur. J. Nucl. Med. Mol. Imaging* 43.8 (2016), pp. 1477–1485.
- [3] Gurvan Dissaux et al. “Pretreatment 18F-FDG PET/CT Radiomics Predict Local Recurrence in Patients Treated with Stereotactic Body Radiotherapy for Early-Stage Non-Small Cell Lung Cancer: A Multicentric Study”. In: *J. Nucl. Med.* 61.6 (2020), pp. 814–820.
- [4] Mathieu Hatt et al. “Tumour functional sphericity from PET images: prognostic value in NSCLC and impact of delineation method”. In: *Eur. J. Nucl. Med. Mol. Imaging* 45.4 (2018), pp. 630–641.
- [5] Margarita Kirienko et al. “Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery”. In: *Eur. J. Nucl. Med. Mol. Imaging* 45.2 (2018), pp. 207–217.
- [6] Nitin Ohri et al. “Pretreatment 18F-FDG PET textural features in locally advanced non-small cell lung cancer: Secondary analysis of ACRIN 6668/RTOG 0235”. In: *J. Nucl. Med.* 57.6 (2016), pp. 843–848.
- [7] Anastasia Oikonomou et al. “Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy”. In: *Sci. Rep.* 8.1 (2018), pp. 1–11.
- [8] Thomas Pyka et al. “Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy”. In: *Radiat. Oncol.* 10.1 (2015), pp. 1–9.
- [9] Florent Tixier et al. “Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: Prognostic value in non-small cell lung cancer”. In: *J. Nucl. Med.* 55.8 (2014), pp. 1235–1241.
- [10] Stephen S.F. Yip et al. “Associations Between Somatic Mutations and Metabolic Imaging Phenotypes in Non-Small Cell Lung Cancer”. In: *J. Nucl. Med.* 58.4 (2017), pp. 569–576.

10 Paper III

11 Paper IV

Cerebral blood flow measurements with ^{15}O -water PET using a non-invasive machine-learning-derived arterial input function

Journal of Cerebral Blood Flow & Metabolism
0(0) 1–13

© The Author(s) 2021



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0271678X21991393

journals.sagepub.com/home/jcbfm



Samuel Kuttner^{1,2,3} , Kristoffer Knutsen Wickstrøm², Mark Lubberink⁴, Andreas Tolf⁵, Joachim Burman⁵, Rune Sundset^{1,3}, Robert Jenssen², Lieuwe Appel⁴ and Jan Axelsson⁶ 

Abstract

Cerebral blood flow (CBF) can be measured with dynamic positron emission tomography (PET) of ^{15}O -labeled water by using tracer kinetic modelling. However, for quantification of regional CBF, an arterial input function (AIF), obtained from arterial blood sampling, is required. In this work we evaluated a novel, non-invasive approach for input function prediction based on machine learning (MLIF), against AIF for CBF PET measurements in human subjects.

Twenty-five subjects underwent two 10 min dynamic ^{15}O -water brain PET scans with continuous arterial blood sampling, before (baseline) and following acetazolamide medication. Three different image-derived time-activity curves were automatically segmented from the carotid arteries and used as input into a Gaussian process-based AIF prediction model, considering both baseline and acetazolamide scans as training data. The MLIF approach was evaluated by comparing AIF and MLIF curves, as well as whole-brain grey matter CBF values estimated by kinetic modelling derived with either AIF or MLIF.

The results showed that AIF and MLIF curves were similar and that corresponding CBF values were highly correlated and successfully differentiated before and after acetazolamide medication. In conclusion, our non-invasive MLIF method shows potential to replace the AIF obtained from blood sampling for CBF measurements using ^{15}O -water PET and kinetic modelling.

Keywords

Arterial input function, cerebral blood flow, Gaussian processes, kinetic modelling, machine learning, ^{15}O -water PET

Received 23 June 2020; Revised 30 November 2020; Accepted 3 January 2021

Introduction

Measurements of cerebral blood flow (CBF) can be used to separate pathological and healthy brain tissue as well as for functional brain research. Tracer kinetic modelling following dynamic ^{15}O -water positron emission tomography (PET) imaging with arterial blood sampling is considered the reference standard for CBF measurements.^{1–6} However, arterial cannulation is an invasive, laborious and time-consuming procedure, and may, due to induction of pain and risk for complications, discourage patients and volunteers from participating in research studies. Furthermore, a useful arterial input function (AIF) curve cannot be obtained

¹Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway

²Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway

³The PET Imaging Center, University Hospital of North Norway, Tromsø, Norway

⁴Department of Surgical Sciences, Radiology, Uppsala University, Uppsala, Sweden

⁵Department of Neuroscience, Uppsala University, Uppsala, Sweden

⁶Department of Radiation Sciences, Umeå University, Umeå, Sweden

Corresponding author:

Samuel Kuttner, Department of Clinical Medicine, UiT The Arctic University of Norway, 9037 Tromsø, Norway.

Email: samuel.kuttner@uit.no

without careful cross calibration of the blood measurement detector and the PET scanner. In addition, because the blood is most commonly sampled from the radial artery, additional corrections for dispersion and delay of the tracer must be applied, to obtain the true AIF for the brain.^{3,7-9}

The use of an image-derived input function (IDIF) has been proposed as an alternative approach to overcome the challenges with the AIF.¹⁰ In brain PET imaging, an IDIF can be measured inside a suitable intracranial blood vessel directly in the reconstructed PET images, for instance in the intracranial carotid arteries.¹¹⁻¹³ Due to the limited spatial resolution of the PET system, and the need for short time-frames during the first pass of the bolus, image-derived methods suffer from both partial volume effects and image noise. These limitations require complex and standardized methods for partial volume correction and artery delineation, which may be difficult to achieve in practice. Recently, a few clinical studies have suggested the use of integrated PET/magnetic resonance imaging (MRI) for deriving an IDIF in brain, where the latter modality is used for artery delineation or even motion correction.¹⁴⁻¹⁸ However these methods are sensitive to registration errors between the modalities and require detailed knowledge of the scanner resolution. Another recent study overcame potential misregistration problems and formed a corrected IDIF by deriving total number of counts and artery volume from the two modalities separately.¹⁹ However, this method was not yet validated with arterial blood sampling and as hybrid PET/MRI is still an emerging modality, to date, image-derived methods are rarely used in larger clinical or research studies.^{10,20}

Alternatively, a standardized, population-based AIF can be calculated as an average AIF from a group of subjects acquired with the same tracer, injection protocol and population, and scaled to the specific subject.^{21,22} However, this method requires at least one blood sample for curve scaling while individual physiological differences and scan-dependent variations are neglected.

An approach on image data with simultaneous estimation of AIF and kinetic parameters has also been reported.²³⁻²⁵ This method, however, assumes a known mathematical AIF model and requires at least one late blood sample for parameter estimation. Recently, non-invasive simultaneous estimation methods were developed that obviate the need for the single late blood sample by using additional input variables from electronic health records into the models.^{26,27} The limitation of such an approach is that a large set of clinical variables must be collected and handled for each patient. These variables may not necessarily be available in the health records for all patients and may even

complicate inclusion of healthy volunteers in research studies.

In this study we use a machine learning-based approach for AIF estimation. Machine learning-methods are especially useful for function estimation and regression.²⁸ Briefly, one seeks to determine a function, f , that predicts the machine-learning-derived input function (MLIF), based on an input vector, \mathbf{x} , composed of multiple image-derived tissue curves, such that $\text{MLIF} = f(\mathbf{x})$. The function f is determined by optimizing hidden parameters to find the best mapping, $\text{AIF} = f(\mathbf{x})$, for a set of training data, where both the AIF and the tissue curves are known. Once the model has been trained and f is known, the MLIF can be predicted for unseen test data using only the tissue-curves extracted from the image data.²⁸

In our previous work, we developed and validated a machine-learning-based input function for ¹⁸F-fluorodeoxyglucose (FDG) in a mouse PET cohort.²⁹ In short, two learning models were evaluated that predicted an AIF from time-activity curves of up to 7 different tissue regions as input. The main limitation with our previous study was the lack of an AIF, thus the models could only be validated against a reference IDIF. However, in mouse PET scanning, the entire body of the mouse fits in the PET field-of-view, thus, time-activity curves from all organs are readily available as input data for the models. We showed that, for instance, the myocardium and liver were important for AIF prediction, because their time-activity curves closely resembled the reference IDIF. In contrast, these blood-rich organs are outside the field-of-view in clinical brain PET imaging, and thus, alternative input curves had to be derived for the MLIF model in the current study.

In the present study, we have further developed the MLIF approach for human ¹⁵O-water brain PET data and evaluated the models against an AIF obtained from continuous arterial-blood sampling. We aimed to show that an AIF could be accurately predicted by an MLIF model using multiple image-derived input data from the carotid arteries. We hypothesized that there were no significant differences in estimated CBF when using either AIF or MLIF, and investigated similarities and differences between image-derived sampling in the brain versus arterial-sampling from the arm. Further, we investigated whether the MLIF method was capable to predict a clinically relevant CBF difference between scans before and after acetazolamide medication.

Materials and methods

Subjects

Pseudonymized data from 25 subjects were retrospectively collected from a completed clinical research

study at Uppsala University Hospital. The data comprised both patients with multiple sclerosis (MS) and healthy volunteers (mean age (range) in years: 40 (23–56); F:M 15:10). In this methodological study, we did not differentiate between the two groups as we considered that the subject's health status had no impact on our evaluation of the MLIF model. Therefore, all authors were blinded for the health status of each subject and thus, no comparisons were made between healthy subjects and MS-patients. The results of the parent study will be reported elsewhere.

The parent study was approved by the Swedish Ethical Review Authority (reference 2014/453). All subjects signed written informed consent prior to inclusion. Since the present work was purely an image analysis methodology study using pseudonymized data, it was not covered by the Swedish or Norwegian regulations on medical research in humans and as such, no additional ethics approval was necessary.

Image acquisition

All subjects underwent two 10 min dynamic brain PET scans on either an ECAT Exact HR+ stand-alone PET scanner (Siemens, Knoxville, TN; $n = 9$) or a Discovery MI PET-Computed tomography (CT) scanner (GE Healthcare, Waukesha, MI; $n = 16$). The scans started simultaneously with an automated bolus injection of 5 MBq/kg ^{15}O -water (^{15}O -water in 5 ml saline at 1 ml/s followed by 35 ml saline at 2 ml/s). Each subject underwent one scan at baseline and one scan 15–30 min after intravenous administration of acetazolamide (9 mg/kg up to a maximum of 1000 mg; 5 min infusion) such that every subject was its own control. Acetazolamide medication dilates the vascular system and thereby it increases the cerebral arterial blood flow velocity.^{30–32} Attenuation correction was based on a 10 min transmission scan with rotating ^{68}Ge rod sources (ECAT) or an ultra-low-dose CT scan (Discovery MI). Images were reconstructed into 26 time-frames (1×10 , 8×5 , 4×10 , 2×15 , 3×20 , 2×30 and 6×60 s). Image reconstruction algorithms were chosen to result in a matching image resolution for the two scanners: ordered subsets expectation maximization with 6 iterations, 8 subsets and a 4 mm Hanning filter (ECAT) and 3 iterations, 16 subsets and a 5 mm Hanning filter (Discovery MI).

In addition, all subjects underwent MRI on a 3 T MRI scanner (Achieva, Philips Healthcare, Best, The Netherlands) with a 32-channel head coil. A three-dimensional T1-weighted gradient echo sequence was obtained with voxel size $1.0 \times 1.0 \times 1.0 \text{ mm}^3$, repetition time = 8.2 ms and echo time = 3.7 ms.

Blood sampling

Continuous arterial blood sampling was performed during 10 min for each scan (3 ml/min) using either an ABSS V3 (Allogg, Mariefred, Sweden; subjects scanned on ECAT) or PBS-100 (Veenstra-Comecer, Joure, The Netherlands; subjects scanned on Discovery MI). A single arterial blood sample was taken through a three-way-valve on the arterial line 5 min post injection and measured in a cross-calibrated well counter for calibration of the continuous arterial blood data.

The measured blood signal, $g(t)$, was affected by dispersion in the vessels and in the detector system tubes. This could be modelled as a convolution of the true AIF, $C_A(t)$, and a dispersion function, $d(t)$ ⁸

$$g(t) = C_A(t) \otimes d(t) \quad (1)$$

A mono-exponential dispersion model was assumed⁸

$$d(t) = \frac{1}{\tau} e^{-t/\tau} \quad (2)$$

where τ is the dispersion constant. An expression for the true AIF, $C_A(t)$, could be obtained by the Laplace transform³³

$$C_A(t) = g(t) + \tau \frac{dg}{dt} \quad (3)$$

The dispersion constant was fixed to 15 s for all subjects in this study, as suggested in the literature.⁸

The dispersion-corrected AIFs were delay-corrected, as described in the 'Image processing' Section.

A visual assessment of the AIF curves was performed to identify abnormal AIFs due to failures in tracer administration or continuous arterial blood sampling. Three subjects were excluded after the visual assessments.

Image processing

All images were corrected for inter-frame motion using in-house written software in Matlab (Mathworks, MA, USA). A simple and objective multi-VOI thresholding method, that could capture blood information from the carotid arteries, was empirically developed. First, to remove noise close to the edge slices, a PET search-volume was defined by trimming 20 voxels in the x-y-plane periphery, and by removing 5 slices in the z-direction. The algorithm for identifying the time-frame for carotid-VOI thresholding was based on a frame-wise graph of whole-brain gray-matter total intensity. The time-frame used for VOI thresholding was the first frame where the total

intensity was larger than 25% of the maximum value in this graph. Three VOIs were generated, comprising the 10, 100 and 1000 highest-intensity voxels (Figure 1a to c). The median voxel value was derived for each time-frame and VOI, resulting in three IDIF time-activity curves, named IDIF₁₀, IDIF₁₀₀ and IDIF₁₀₀₀ (Figure 1d), which could be interpreted as three different image-derived blood-curves with different amounts of partial volume effect.

Subsequently, to match the AIF sampling, all PET data were interpolated linearly to one second time framing. To correct for delay between the AIF and the PET data, the dispersion corrected AIF was shifted to provide the best overlap with the IDIF₁₀ curve, i.e. where the dot product between the two curves was maximized.

In all following analyses, only PET and AIF data from 0–6 min were used, to minimize noise from late parts of the scans. Calibration of the AIF curves with the single arterial blood sample was only conducted for the training data, and thus no AIF or blood sample was required for the test data.

To extract time-activity curves for whole-brain grey matter, T1-weighted MRI images were co-registered to PET images and segmented using SPM8 (The Wellcome Centre for Human Neuroimaging, UCL Queen Square Institute of Neurology, London, UK).

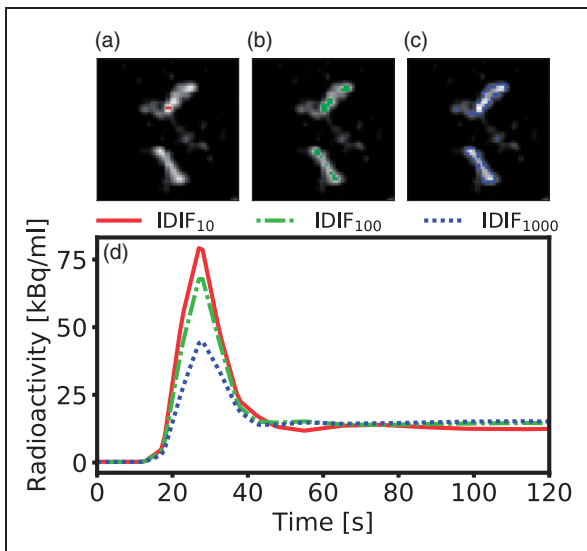


Figure 1. Outline of the VOI thresholding method implemented in this work. (a) to (c) shows an enlarged axial brain PET slice of the optimal time-frame for a representative subject. Highlighted are the parts of the IDIF₁₀ (a), IDIF₁₀₀ (b) and IDIF₁₀₀₀ (c) comprising the 10, 100 and 1000 highest intensity voxels. (d) The resulting time-activity curves during the first two minutes for IDIF₁₀₋₁₀₀₀. The IDIF₁₀ captures the highest activity in the carotid artery, while the IDIF₁₀₀ and IDIF₁₀₀₀ shows a lower activity due to a higher number of voxels included.

All image analysis was performed in the native PET image space.

Function prediction using Gaussian processes

Gaussian processes (GP) is a well-known, non-parametric Bayesian regression method which has been frequently used within machine learning for data-driven function estimation and regression tasks.^{34,35} One advantage with GP is that it predicts not only the mean function, but also its variance, thus providing an uncertainty measure of the model.³⁶ Furthermore, GP, is known to work well with sparse training datasets, as opposed to neural networks.^{37,38}

In GP regression, the output y , is approximated by a probability distribution of functions of the input, \mathbf{x} , such that $f(\mathbf{x}) \sim \mathcal{GP}(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$, which is a generalization of the multivariate Gaussian distribution to infinitely many variables. Here, $m(\mathbf{x})$ is a mean function and $k(\mathbf{x}, \mathbf{x}')$ is a covariance function.³⁶ Assuming N available input-output training samples in a dataset $\{\mathbf{x}_n, y_n\}_{n=1}^N$, each including the three IDIF time-activity curves, \mathbf{x}_n (see Figure 1), with corresponding known AIF, y_n . Then the mean value MLIF of the test sample, $\mathbb{E}[y_*]$, and the variance, $\mathbb{V}[y_*]$, can be predicted by

$$\mathbb{E}[y_*] = \mathbf{k}_*^T (\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{y} \quad (4)$$

$$\mathbb{V}[y_*] = k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^T (\mathbf{K} + \sigma_\epsilon^2 \mathbf{I})^{-1} \mathbf{k}_* \quad (5)$$

Here \mathbf{k}_* is the covariance between the training samples \mathbf{x}_n and the test sample \mathbf{x}_* ; $\mathbf{K} = k(\mathbf{x}_n, \mathbf{x}_m)$ is the covariance between all training samples; $\sigma_\epsilon^2 \mathbf{I}$ is a scalar matrix with diagonal elements equal to the noise level; $k(\mathbf{x}_*, \mathbf{x}_*)$ is the covariance between the test sample and itself.³⁶

Input function prediction

For input function prediction, leave-one-out cross validation was employed, which is a common validation method in machine learning with limited amounts of data.²⁸ In short, one sample was withdrawn from the dataset and assigned as test sample, while the remaining samples were allocated for training. The three time-activity curves (Figure 1) were used as input vectors into the GP framework, and the MLIF and variance of the test sample were predicted using equations (4) and (5), respectively. The process was repeated by assigning a new sample as test sample, until all samples had been tested once.

In all experiments, the Matérn covariance function was chosen, with $\nu = 5/2$, because it has been reported to generate smooth functions.³⁶ Data normalization

was applied on the input-IDIFs, which is a well-known approach to improve convergence of machine learning models.³⁹ Normalization of the IDIF₁₀ curves was performed by identifying the IDIF₁₀ curve with the highest peak value among all subjects. Subsequently, each IDIF₁₀ curve was normalized by dividing with this peak value. Similarly, the IDIF₁₀₀ and IDIF₁₀₀₀ curves were normalized independently with their respectively found highest peak value among the subjects. Thus, the model was trained with three different normalized IDIFs with values ranging between 0 and 1. The normalization was only a scale factor, meaning that relative amplitudes between subjects remained. GP regression was implemented in Python 3.6.8, using Gpflow 1.2.0, in which the matrix inversion of equation (4) was approximated by Cholesky decomposition. The hyperparameters of the covariance function were optimized by maximizing the logarithm of the marginal likelihood of the training data.⁴⁰

Kinetic modelling

Quantification of CBF was performed on the whole-brain grey-matter region. A single-tissue compartment model was used to generate CBF values. This model assumes that water can diffuse freely between vascular and tissue space, with activity concentrations C_A and C_T , respectively, and is described by the following equation^{41,42}

$$\frac{dC_T(t)}{dt} = K_1 \cdot C_A(t) - k_2 \cdot C_T(t) \quad (6)$$

where $C_A(t)$ is the whole-blood arterial time-activity curve, also known as the AIF. The solution to equation (6) is given by⁴³

$$C_T(t) = K_1 \cdot C_A(t) \otimes e^{-k_2 \cdot t} \quad (7)$$

where \otimes denotes mathematical convolution. The activity concentration measured with PET, C_{PET} , is modelled as the sum of the tissue compartment, $C_T(t)$, and the fractional arterial blood volume in the tissue, V_A ,

such that

$$C_{PET}(t) = (1 - V_A) \cdot C_T(t) + V_A \cdot C_A(t) \quad (8)$$

For tracers with high extraction fraction relative to the blood flow, such as ¹⁵O-water, CBF equals K_1 .^{43,44}

Evaluation design

The aim of the current work was to investigate whether an AIF could be accurately predicted using an MLIF model in baseline and acetazolamide scans of the same subject. In a first case, a GP model named MLIF₁ was trained and subsequently tested on baseline scans using leave-one-out cross validation. Similarly, another GP model, named MLIF₂, was trained and subsequently tested on acetazolamide scans using leave-one-out cross validation.

In a routine setting, it is of interest to train a predictive model on normal subjects and apply the same model on a disease, or medicated group. Therefore, in a second case, an additional GP model named MLIF^{*}₁, was trained on all baseline scans, and subsequently tested on all acetazolamide scans. Here, the asterisk (*) in MLIF^{*}₁ emphasises that all subjects from the baseline scan were included in the training data for this model, as opposed to the leave-one-out training for MLIF₁. Essentially, one might see MLIF^{*}₁ and MLIF₁ as two models trained on the same dataset, as the difference is only one subject.

Finally, we hypothesized that MLIF^{*}₁ might be more representative for a local AIF in the brain, compared to an AIF sampled in the wrist. Therefore, in a third case, we aimed to evaluate the CBF increase from baseline to acetazolamide scans obtained by the two different methods in *case 1* and *case 2*, by using the carotid arteries with the ten highest-intensity voxels (IDF₁₀) as input function during kinetic modelling.

The evaluation design is summarized in Table 1.

Evaluation methods

The GP predicted MLIF curves were first compared point-by-point to AIF using orthogonal regression.

Table 1. The evaluation design of the MLIF method.

Case	Training data	Testing data	Input function	Procedure
1	Baseline	Baseline	MLIF ₁	Leave-one-out
	Acetazolamide	Acetazolamide	MLIF ₂	Leave-one-out
2	Baseline	Baseline	MLIF ₁	Leave-one-out
	Baseline	Acetazolamide	MLIF [*] ₁	All baseline scans
3	^a	Baseline	IDIF ₁₀ ^a	Only used to calculated change
	^a	Acetazolamide	IDIF ₁₀ ^a	between baseline and acetazolamide CBF

^aIn *case 3*, no GP prediction was used, but instead, kinetic modelling was based on the 10 highest-intensity carotid artery voxels (IDIF₁₀) as input function.

Subject scans with regression slopes that were more than three scaled-median absolute-deviations away from the median slope, were considered outliers, and removed from further model comparisons.⁴⁵ Time-activity curves averaged over subjects were calculated for: whole-brain grey matter, AIF, the MLIF models and IDIF₁₀. CBF and V_A were estimated for whole-brain grey matter using kinetic modelling with both AIF and the GP-predicted MLIF (*case 1* and *case 2*) as well as with the IDIF₁₀ time-activity curve as input function (*case 3*).

MLIF-based CBF estimates were compared with the one based on AIF by paired t-test ($\alpha=0.05$), ratio calculation, orthogonal regression and Bland-Altman plots.⁴⁶ Normality was assessed using quantile-quantile plots.

The GP variance (equation (5)) was considered as a measure of prediction error. This measure was evaluated by the relationship between the magnitude of the CBF-ratio (CBF_{MLIF}/CBF_{AIF}) and GP variance for *case 1*.

Results

CBF values based on AIF (CBF_{AIF}) and MLIF (CBF_{MLIF}) are shown in Supplementary Table S1. The average CBF_{AIF} and CBF_{MLIF} were similar for both baseline and acetazolamide scans, about 0.45 and 0.60 ml·min⁻¹·g⁻¹, respectively. The mean CBF_{MLIF}/CBF_{AIF} ratio was 1.04 and 1.03 for baseline and acetazolamide, respectively. No significant

differences were found between the average CBF_{AIF} and CBF_{MLIF} for either scan.

Individual data points for MLIF and AIF from *case 1* are shown as a scatter plot in Figure 2(a). Based on the outlier removal criterion, four scans were removed from model comparisons. There is a strong overall linear relationship between AIF and MLIF curves for both baseline (slope: 0.8) and acetazolamide scans (slope: 0.9). Individual r^2 values were high for both baseline (mean: 0.90, SD: 0.1) and acetazolamide (mean: 0.93, SD: 0.07). Histograms of slope values indicate slopes close to unity for most subjects for both baseline (median slope: 0.86, Figure 2b) and acetazolamide (median slope: 0.87, Figure 2c).

Displaying the CBF data from *case 1* as a scatter plot (Figure 3a), a strong linear relationship and high overall correlation ($r^2>0.9$) between CBF_{AIF} and CBF_{MLIF} was obtained. Bland-Altman analysis (Figure 3b) exhibited a prediction bias close to zero.

The fractional arterial blood volume, V_A , from equation (8), was found to be near zero for all scans (0.001 ± 0.003).

As a visual illustration of the effect of prediction errors on estimation of CBF, the AIF and GP-predicted MLIF from the baseline scan for four representative subjects are shown in Figure 4. Figure 4(a) and (b) comprises two examples with less than 3% difference between CBF_{AIF} and CBF_{MLIF} , while Figure 4 (c) and (d) display cases with substantial differences between both methods. Based on the CBF ratios (CBF_{MLIF}/CBF_{AIF}), it can be observed that an over-prediction of the AIF peak (Figure 4c) results in an

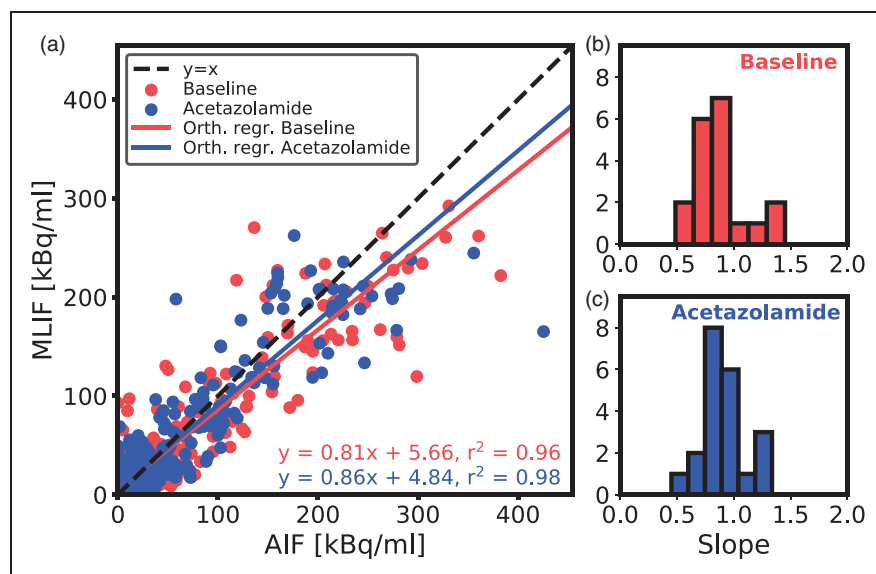


Figure 2. (a) Scatter plot of MLIF and AIF data points for all subjects for baseline (blue) and acetazolamide (red) scans based on *case 1*. The 1 s time frames were interpolated back to the original PET time framing (26 time-frames). The solid lines are the orthogonal regression fits. (b and c) Histogram of orthogonal regression slopes for individual subjects for baseline (b) and acetazolamide (c) scans.

underestimation of CBF, while an underpredicted AIF peak (Figure 4d) ends in an overestimated CBF.

Figure 5 displays box plots of whole-brain grey matter CBF for baseline and acetazolamide scans, when using different prediction models for MLIF (see Table 1).

The GP variance was evaluated as prediction error measure in *case 1* (Supplementary Figure S1). No relationship was found between the magnitude of the $\text{CBF}_{\text{MLIF}}/\text{CBF}_{\text{AIF}}$ ratios and the GP variance values. The predicted variance was not further considered in this work.

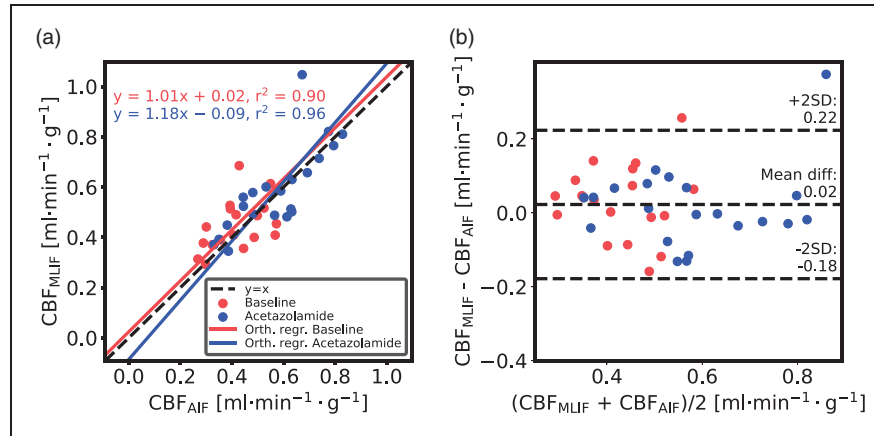


Figure 3. Evaluation of the GP-predicted MLIF for baseline (blue) and acetazolamide (red) scans based on *case 1*. (a) Scatter plot of MLIF-based and AIF-based CBF. The solid lines are the orthogonal regression fits. (b) Bland-Altman plot of *case 1*.

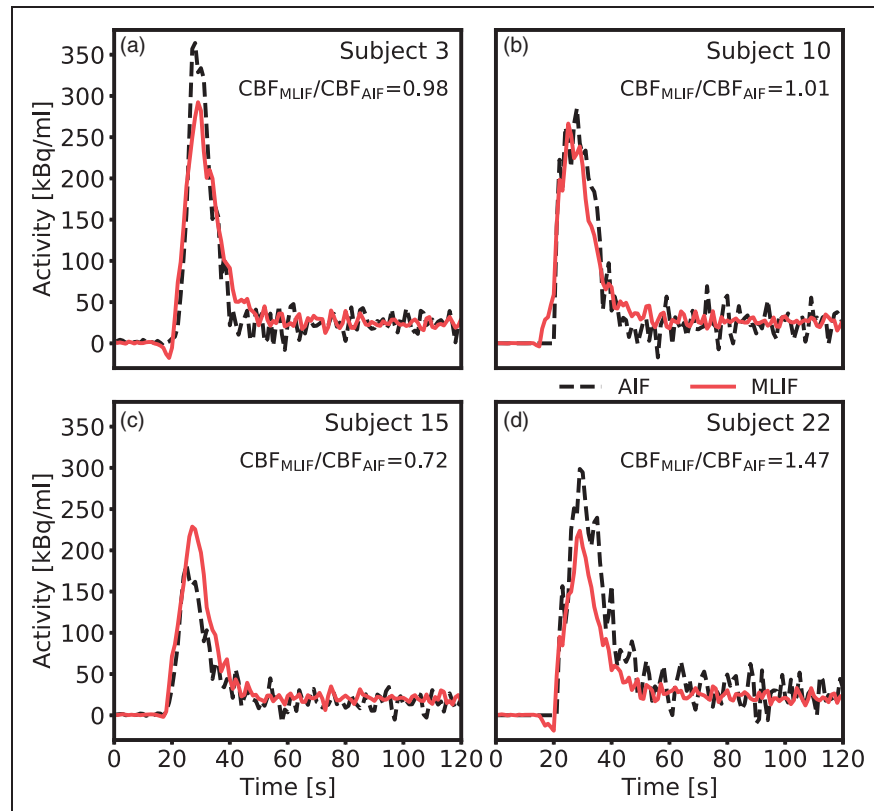


Figure 4. Comparison of AIF (black dashed line) and GP-predicted MLIF (red line) for the baseline scan of four representative subjects. (a and b) Two scans with less than 3% difference between CBF_{AIF} and CBF_{MLIF} . (c) Representative example of a scan where the MLIF overpredicts the AIF peak, resulting in an underestimation of the calculated CBF. (d) Representative example of a scan where the MLIF underpredicts the AIF peak, resulting in an overestimation of the calculated CBF.

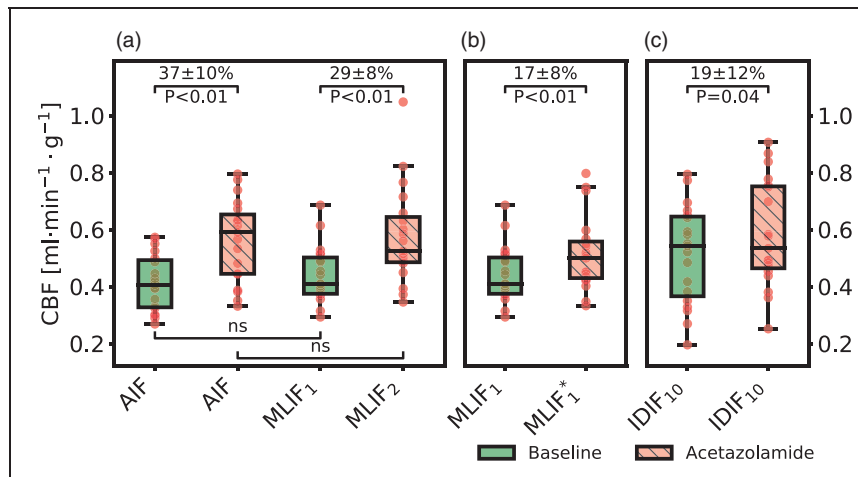


Figure 5. Box plot of estimated whole-brain grey matter CBF for baseline and acetazolamide scans, when using different input functions for *case 1* (a), *case 2* (b) and *case 3* (c). For an explanation of the cases, see Table 1. For visual purposes, the data in panel (c) was scaled to match the range of the AIF-based CBF values in panel (a). Percentage differences are shown as mean \pm 95% confidence interval, and P values are based on paired t-test. In the box plots, red points indicate the data points; the horizontal line and the black box represent median and interquartile range (25th to 75th percentile), respectively; the whiskers indicate the maximum and minimum data point up to $1.5 \times$ interquartile range.

In *case 1* we evaluated the MLIF method by training and applying separate GP models for baseline (MLIF₁) and acetazolamide (MLIF₂) data. There were no significant differences between mean CBF_{AIF} and mean CBF_{MLIF} in neither baseline, nor acetazolamide scans. Furthermore, following acetazolamide medication, we found on average 37% increase ($P < 0.01$) in mean CBF_{AIF} and 29% increase ($P < 0.01$) in mean CBF_{MLIF} (Figure 5a). For *case 1*, both the AIF and the MLIF based methods resulted in similar CBF values, while the CBF increase after acetazolamide medication was lower for the MLIF method. The correlation coefficient between AIF and MLIF-based CBF changes for *case 1* was 0.62.

In *case 2*, a predictive model (MLIF*₁) was trained on baseline scans and then applied on acetazolamide scans. Similar to MLIF₁, the average CBF_{MLIF*1} was non-significantly different from the corresponding CBF_{AIF} after acetazolamide medication. Also, there was still a significant increase (17%, $P < 0.01$) in CBF between baseline and acetazolamide scans (Figure 5b), but notably smaller than the 29% increase observed in *case 1*. The correlation coefficient between AIF and MLIF-based CBF changes for *case 2* was 0.14.

In a final case, the relative CBF increase from baseline to acetazolamide scans was investigated by using the IDIF₁₀ as input function for each scan. A significant CBF increase of 19% ($P = 0.04$) was found after acetazolamide medication (Figure 5c), which was comparable to the change observed in *case 2*.

As CBF is based on underlying time-activity curves, we proceeded to investigate the differences in CBF

after acetazolamide medication in the first two cases by inspecting the mean time-activity curves across subjects in Figure 6. We observed that the local brain input function (IDIF₁₀) (Figure 6e) showed a shape-dependence on acetazolamide medication which was not reflected in the AIF measured in the wrist (Figure 6b). In the AIF, the baseline and acetazolamide curves were similar, while for IDIF₁₀, the acetazolamide curve had a larger area-under-curve compared to baseline. Also, for IDIF₁₀, there was a slight shift in the mean time-activity curves between baseline and acetazolamide, which was not visible for the AIF. Thus, the AIF measured in the wrist does not reflect physiological changes due to acetazolamide, which are apparent in the local brain input (IDIF₁₀).

Discussion

Tracer kinetic modelling of dynamic PET data requires accurate knowledge of an AIF, conventionally measured through arterial blood sampling. Our aim was to investigate whether an AIF could be predicted as accurately by an MLIF model using solely image-derived input data from the carotid arteries.

AIF and predicted MLIF curves were found to be similar, with no significant difference between whole-brain grey matter CBF_{AIF} and CBF_{MLIF} estimates. Furthermore, the correlation between the CBF estimates was $r^2 > 0.9$ and the mean differences were close to zero. The MLIF model was also able to accurately predict an increased CBF after acetazolamide medication, when trained on post-acetazolamide data.

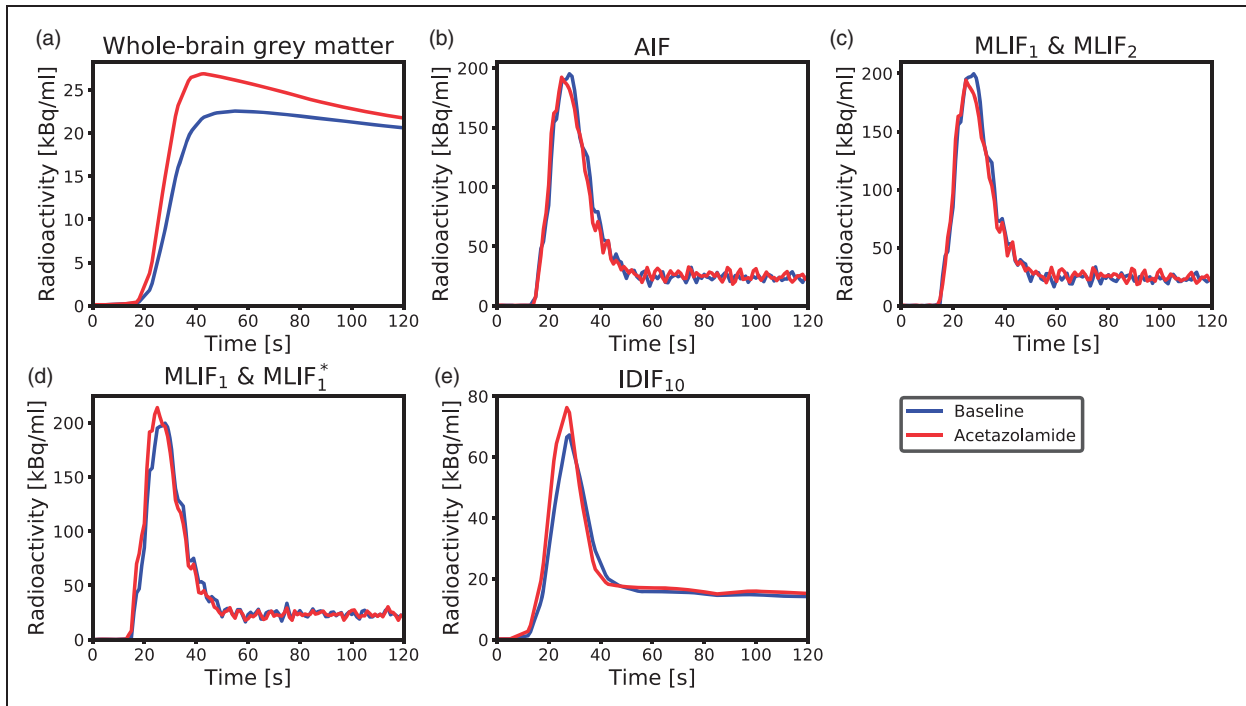


Figure 6. Mean time-activity curves across subjects for baseline (blue) and acetazolamide (red) scans during the first 2 min of PET scanning. (a) Measured radioactivity uptake in whole-brain grey matter. (b) AIF. (c) MLIF₁ model for baseline, and MLIF₂ for acetazolamide scans. (d) MLIF₁ model for baseline, and MLIF₁* for acetazolamide scans. (e) IDIF₁₀ input function based on the 10 highest voxels in the carotid arteries.

Altogether the results indicate that the MLIF method has potential as an alternative AIF for generation of CBF values using ¹⁵O-water PET and kinetic modelling, which in clinical practice implies evading of arterial cannulation.

Initially, we evaluated an MLIF model trained on pooled baseline and acetazolamide data. However, this resulted in inferior generalization to new samples compared to when training was done separately for baseline and acetazolamide scans (data not shown). The reason for this difference between the approaches may be because the input IDIFs under the two conditions vary in amplitude (Figure 6e), while the AIF (Figure 6b) does not. Therefore, in a first case, input functions of baseline scans were predicted using an MLIF model trained on baseline data (MLIF₁), and, similarly, input functions of acetazolamide scans were predicted using an MLIF model trained on acetazolamide data (MLIF₂). There was a strong linear relationship between the data points from AIF and predicted MLIF curves, although the peak values were systematically underestimated (slope <1 in Figure 2). In the time-activity curves, the number of high data values, acquired during the first pass peak, is low compared to the number of low values, acquired during the rest of the scan. Also, the standard deviation around the peaks was observed to be larger than at the

tails (data not shown). We speculate that this imbalance affects the GP models and results in the systematically underestimated peak values. Previous work has shown that when data is limited, the GP model may underestimate the mean function.⁴⁷

For both baseline and acetazolamide, the mean time-activity curves for AIF (Figure 6b), and MLIF₁, MLIF₂ (Figure 6c) appeared similar in shape. Following kinetic modelling, the average CBF values obtained using an AIF are in line with previously published work before^{48,49} and after administration of acetazolamide.⁴⁸ No significant differences were found between AIF-based and MLIF-based CBF estimates in whole-brain grey matter for neither baseline nor acetazolamide scans (Figure 5a). Also, a slope close to unity, an $r^2 > 0.9$ and low bias between CBF_{AIF} and CBF_{MLIF} estimates pointed towards an acceptable agreement of both methods (Figure 3a). Across both baseline and acetazolamide scans, differences between CBF_{AIF} and CBF_{MLIF} estimates were relatively small (Figure 3b), although several subjects had relative CBF errors of > 20% for the baseline scans while there was a somewhat lower spread for the acetazolamide scans. Our hypothesis is that a GP model trained on acetazolamide data (MLIF₂) generalize better to new samples, compared to a model trained on baseline data (MLIF₁). We speculate that the reason for this is that

a larger range of CBF values is a result from a larger range of input values (IDIF₁₀₋₁₀₀₀), for acetazolamide, compared to baseline scans. This may have resulted in MLIF₂ being valid for a wider range of unseen samples, compared to MLIF₁, as mentioned and illustrated in Supplementary material. Nevertheless, the CBF increase between baseline and acetazolamide scans was highly correlated to that of the AIF in *case 1*. Furthermore, it was observed that the shape of the input function had an impact on the accuracy of the MLIF-based CBF estimates (Figure 4). Evidently, an underpredicted AIF peak resulted in an overestimated CBF while an overpredicted AIF peak caused an underestimation of CBF. This can be explained by the inverse relationship between K_1 and C_A in equation (7).

In a second case, we trained an MLIF model on all baseline scans and then applied that model on the acetazolamide scans (MLIF₁^{*}). The idea was to resemble a situation where the baseline scans reflected a database comprising healthy subject with a normal CBF whereas the acetazolamide scans reflected clinical data from patients with an altered CBF. Even in this scenario significant differences were found between the whole-brain grey matter CBF from the baseline and acetazolamide scan. However, the CBF increase between baseline and acetazolamide scans in this case displayed small between-subject variation (data not shown) and, maybe because of that, low correlation to the AIF CBF changes. Furthermore, unexpectedly, the difference in CBF was only 17% (Figure 5b) compared to 29% in the previous case (Figure 5a).

We found this difference between baseline and acetazolamide response striking. The local brain input function (IDIF₁₀) (Figure 6e) showed a shape-dependence on acetazolamide which was not reflected in the AIF measured in the wrist (Figure 6b). Consequently, CBF calculated with a local brain input would by necessity be different from AIF-based CBF. This effect remained also for the MLIF₁ and MLIF₁^{*} models (Figure 6d) and could possibly explain the observed differences in CBF change.

In order to quantify the differences in CBF after acetazolamide medication in the first two cases, we attempted using a carotid artery region (IDIF₁₀) as input function for both baseline and acetazolamide scans (*case 3*). In *case 3* a significant difference of 19% was found between baseline and acetazolamide scans which was similar to the observed difference found in *case 2*. This supports that the difference in acetazolamide provocation results were caused by the different input functions in brain (*case 2*, Figure 6d) and in the wrist (*case 1*, Figure 6c). Note that, *case 3* was used only to investigate the relative increase in CBF between baseline and acetazolamide scans

found in *case 1* and *case 2*. IDIF₁₀ cannot be used as a substitute for AIF, due to the limitations of image derived methods, as described in the Introduction.

We suggest that the above discussed difference in input function curve shape might be explained in part by effects on the vascular system after acetazolamide medication. Acetazolamide dilates the vascular system and increases the cerebral blood flow velocity,³⁰⁻³² which explains the increased mean time-activity curve in whole-brain grey matter after acetazolamide medication (Figure 6a). A lower back-pressure due to dilated vasculature together with increased blood velocity^{30,31} could possibly also explain the observed effect on MLIF₁^{*} (Figure 6d) and on the IDIF (Figure 6e). Figure 6e also indicated that the ¹⁵O-water tracer arrived earlier to the brain after acetazolamide medication compared to the baseline scan, resulting in a slight shift of the mean time-activity curves. This observation might also support that the differences in the IDIF₁₀ peaks between scans were partly due to an enhanced cerebral blood-flow velocity in the acetazolamide scan. An additional contributing effect to the increased amplitude of the IDIF after acetazolamide may be an increased spill-over from tissue due to increased brain uptake.

In summary, when training and testing on the same scans, similar CBF estimates for whole-brain grey matter are obtained when using AIF and MLIF (*case 1*). However, when using baseline scans for training followed by applying the model to the acetazolamide scans (*case 2*) the blood input curves are higher for the MLIF model, possibly due to the increase in blood flow velocity after acetazolamide medication. Consequently, CBF_{MLIF} for whole-brain grey matter was lower for acetazolamide, compared to baseline, and the difference in CBF before and after acetazolamide medication was reduced from 29% to 17%. Although these relative changes were different for *case 1* and *case 2*, both were significant and hence suggesting that the MLIF method has a clinical potential to differentiate baseline from acetazolamide scans.

A prerequisite for the MLIF approach is that representative training data have been collected for the specific tracer and imaging system, including both images and blood AIFs. Once an MLIF model has been trained, it offers several advantages, compared to various other image-derived and population-based methods.^{11-13,20-25} A trained MLIF model is a non-invasive method describing both the shape and the amplitude of an AIF, without any need for calibration blood samples. The MLIF models represents a learned transformation, that directly maps the image-derived input data, to a ready-to-use AIF, by inherently

correcting for partial volume effect, with no predefined assumption of the model function. Furthermore, the input data required by the MLIF approach consists of only three carotid artery regions which can be objectively and automatically segmented in the PET images. This makes the MLIF approach simple and convenient to use, without the need for MRI-based artery segmentation.^{14–19}

One limitation of this study is that the GP models were trained on at most 22 samples (MLIF^{*}₁), which might have resulted in an inferior generalization of the model to new samples which were dissimilar from the training data. To investigate the robustness of the MLIF model to unseen test data, new IDIF curves were created by scaling the existing input IDIF time-activity curves for each subject during the leave-one-out testing (Supplementary material). When the input data was scaled, the MLIF model was stable for input ranges encountered in the training data. For higher and lower scale factors, the model performance was gradually degraded (Supplementary Figure S2A). The drop of performance of machine learning models outside the range of training data, so called *domain shift*, is expected and well known.^{50–52}

In a clinical setting where an existing MLIF model is applied to a patient, it is important that the resulting CBF is reliable. We have reported four outliers based on abnormal regression slopes between the AIF and MLIF data points. However, we cannot know if it was the AIF or the MLIF curves that were abnormal. The corresponding CBF values for these outlier scans all indicate that the AIF was abnormally high as indicated by $CBF_{MLIF}/CBF_{AIF} \ll 1$ (Table S1). This suggests that the outlier was caused by the AIF and not by the MLIF model. Future research should investigate methods for quality control of predicted MLIF curves and CBF values from a trained model.

Another limitation of the study is the combination of healthy and MS patients in the data set, as well as data from two different scanners. This increased the heterogeneity in the input data to the MLIF models, and in combination with the effects of acetazolamide, this might have affected the results. However, the limited number of subjects does not allow to study these effects further in detail. Further, the evaluation of our MLIF model was mainly based on differences in whole-brain gray matter CBF. We expect similar results to be obtained for CBF in smaller brain regions, but this aspect should be investigated in future studies. Also, the test-retest variability of the MLIF method should be investigated and compared to that of the blood AIF,⁵³ as well as the evaluation of the model sensitivity in relation to aspects such as injected dose, time-framing, reconstruction settings and different type of scanners. Finally, the generalizability of the MLIF

method to other diseases should be investigated. For example, it is known that carotid stenosis alters the temporal shape of the AIF,⁵⁴ which might have implications for MLIF models trained on baseline scans and applied to patients with pathological arterial vasculature.

In this study, different MLIF models were evaluated with ¹⁵O-labeled water. In our previous research²⁹ a machine learning approach was also feasible for AIF prediction using FDG, although not yet evaluated in clinical data. Thus, we suggest that the method can be adopted to other tracers by merely training similar MLIF models. With proper validation, it may also be conceivable that tracers requiring metabolite correction of the AIF can be included in the prediction model. As for all data-driven models, the accuracy of the MLIF approach for a particular PET application will depend on the quality and quantity of the available training data. Nevertheless, MLIF opens for simplified and non-invasive input function measurements, and thereby potentially eliminating the need for extensive arterial blood sampling in future PET studies.

In conclusion, we demonstrated that our non-invasive MLIF prediction method may be a viable alternative for CBF measurements using ¹⁵O-water PET and kinetic modelling, which in clinical practice implies evading of arterial cannulation. The MLIF method successfully differentiated CBF values before and after acetazolamide medication.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Northern Norway Regional Health Authority [grant number HNF1349-17]. The parent study was supported by BiogenIdex Sweden AB and The Swedish Foundation for MS Research.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Authors' contributions

All authors contributed substantially to the scientific process leading to this manuscript. JA, KKW, ML, RJ, RS and SK contributed in conception and design. AT, JB, LA and ML were responsible for acquisition of clinical PET data (parent study). KKW, RJ and SK implemented the machine learning models. JA and SK analysed the data. JA, LA, ML and SK interpreted the results and drafted the manuscript. All authors reviewed and approved the final manuscript.

ORCID iDs

Samuel Kuttner  <https://orcid.org/0000-0001-7747-9003>

Jan Axelsson  <https://orcid.org/0000-0002-3731-3612>

Supplementary material

Supplemental material for this article is available online.

References

1. Frackowiak RSJ, Jones T, Lenzi GL, et al. Regional cerebral oxygen utilization and blood flow in normal man using oxygen-15 and positron emission tomography. *Acta Neurol Scand* 1980; 62: 336–344.
2. Lassen NA. Regional cerebral blood flow measurement in man. *Arch Neurol* 1963; 9: 615.
3. Herscovitch P, Markham J and Raichle ME. Brain blood flow measured with intravenous H₂(15)O. I. Theory and error analysis. *J Nucl Med* 1983; 24: 782–789.
4. Raichle ME, Martin WR, Herscovitch P, et al. Brain blood flow measured with intravenous H₂(15)O. II. Implementation and validation. *J Nucl Med* 1983; 24: 790–798.
5. Mintun MA, Raichle ME, Martin WR, et al. Brain oxygen utilization measured with O-15 radiotracers and positron emission tomography. *J Nucl Med* 1984; 25: 177–187.
6. Wintermark M, Sesay M, Barbier E, et al. Comparative overview of brain perfusion imaging techniques. *Stroke* 2005; 36: e83–e99.
7. Dhawan V, Conti J, Mernyk M, et al. Accuracy of PET RCBF measurements: effective of time shift between blood and brain radioactivity curves. *Phys Med Biol* 1986; 31: 507–514.
8. Iida H, Kanno I, Miura S, et al. Error analysis of a quantitative cerebral blood flow measurement using H₂15O autoradiography and positron emission tomography, with respect to the dispersion of the input function. *J Cereb Blood Flow Metab* 1986; 6: 536–545.
9. Koeppe RA, Hutchins GD, Rothley JM, et al. Examination of assumptions for local cerebral blood flow studies in PET. *J Nucl Med* 1987; 28: 1695–1703.
10. Muzi M, O'Sullivan F, Mankoff DA, et al. Quantitative assessment of dynamic PET imaging data in cancer imaging. *Magn Reson Imaging* 2012; 30: 1203–1215.
11. Mourik JEM, van Velden FHP, Lubberink M, et al. Image derived input functions for dynamic high resolution research tomograph PET brain studies. *Neuroimage* 2008; 43: 676–686.
12. Mourik JEM, Lubberink M, Schuitemaker A, et al. Image-derived input functions for PET brain studies. *Eur J Nucl Med Mol Imaging* 2009; 36: 463–471.
13. Sari H, Erlandsson K, Law I, et al. Estimation of an image derived input function with MR-defined carotid arteries in FDG-PET human studies using a novel partial volume correction method. *J Cereb Blood Flow Metab* 2017; 37: 1398–1409.
14. Islam MM, Tsujikawa T, Mori T, et al. Estimation of arterial input by a noninvasive image derived method in brain H₂ 15 O PET study: confirmation of arterial location using MR angiography. *Phys Med Biol* 2017; 62: 4514–4524.
15. Okazawa H, Higashino Y, Tsujikawa T, et al. Noninvasive method for measurement of cerebral blood flow using O-15 water PET/MRI with ASL correlation. *Eur J Radiol* 2018; 105: 102–109.
16. Sundar LKS, Muzik O, Rischka L, et al. Towards quantitative [18F]FDG-PET/MRI of the brain: automated MR-driven calculation of an image-derived input function for the non-invasive determination of cerebral glucose metabolic rates. *J Cereb Blood Flow Metab* 2019; 39: 1516–1530.
17. Su Y, Arbelaez AM, Benzinger TLS, et al. Noninvasive estimation of the arterial input function in positron emission tomography imaging of cerebral blood flow. *J Cereb Blood Flow Metab* 2013; 33: 115–121.
18. Su Y, Vlassenko AG, Couture LE, et al. Quantitative hemodynamic PET imaging using image-derived arterial input function and a PET/MR hybrid scanner. *J Cereb Blood Flow Metab* 2017; 37: 1435–1446.
19. Khalighi MM, Deller TW, Fan AP, et al. Image-derived input function estimation on a TOF-enabled PET/MR for cerebral blood flow mapping. *J Cereb Blood Flow Metab* 2018; 38: 126–135.
20. Zanotti-Fregonara P, Chen K, Liow J-S, et al. Image-derived input function for brain PET studies: many challenges and few opportunities. *J Cereb Blood Flow Metab* 2011; 31: 1986–1998.
21. Takikawa S, Dhawan V, Spetsieris P, et al. Noninvasive quantitative fluorodeoxyglucose PET studies with an estimated input function derived from a population-based arterial blood curve. *Radiology* 1993; 188: 131–136.
22. Eberl S, Anayat AR, Fulton RR, et al. Evaluation of two population-based input functions for quantitative neurological FDG PET studies. *Eur J Nucl Med* 1997; 24: 299–304.
23. Feng D, Wong K-P, Wu C-M, et al. A technique for extracting physiological parameters and the required input function simultaneously from PET image measurements: theory and simulation study. *IEEE Trans Inf Technol Biomed* 1997; 1: 243–254.
24. Wong K-P, Feng D, Meikle SR, et al. Simultaneous estimation of physiological parameters and the input function - in vivo PET data. *IEEE Trans Inf Technol Biomed* 2001; 5: 67–76.
25. Bartlett EA, Ananth M, Rossano S, et al. Quantification of positron emission tomography data using simultaneous estimation of the input function: validation with venous blood and replication of clinical studies. *Mol Imaging Biol* 2019; 21: 926–934.
26. Mikhno A, Zanderigo F, Ogden RT, et al. Toward non-invasive quantification of brain radioligand binding by combining electronic health records and dynamic PET imaging data. *IEEE J Biomed Health Inform* 2015; 19: 1271–1282.
27. Rocca E, Mikhno A, Ogden RT, et al. Quantifying brain [18 F]FDG uptake noninvasively by combining medical

- health records and dynamic PET imaging data. *IEEE J Biomed Health Inform* 2019; 23: 2576–2582.
28. Theodoridis S and Koutroumbas K. *Pattern recognition*. 4th ed. New York: Academic Press, Inc., 2009.
 29. Kuttner S, Wickstrøm KK, Kalda G, et al. Machine learning derived input-function in a dynamic 18 F-FDG PET study of mice. *Biomed Phys Eng Express* 2020; 6: 015020.
 30. Dahl A, Russell D, Rootwelt K, et al. Cerebral vasoreactivity assessed with transcranial Doppler and regional cerebral blood flow measurements. *Stroke* 1995; 26: 2302–2306.
 31. Démolis P, Chalon S and Giudicelli JF. Acetazolamide-induced vasodilation in the carotid vascular bed in healthy volunteers. *J Cardiovasc Pharmacol* 1995; 26: 841–844.
 32. Hubner KF. PET imaging in neurology. *J Nucl Med Technol* 1990; 18: 229–234.
 33. Meyer E. Simultaneous correction for tracer arrival delay and dispersion in CBF measurements by the H215O autoradiographic method and dynamic PET. *J Nucl Med* 1989; 30: 1069–1078.
 34. Roberts S, Osborne M, Ebden M, et al. Gaussian processes for time-series modelling. *Philos Trans R Soc A Math Phys Eng Sci* 2012; 371: 20110550.
 35. Dürichen R, Pimentel MAF, Clifton L, et al. Multitask Gaussian processes for multivariate physiological time-series analysis. *IEEE Trans Biomed Eng* 2015; 62: 314–322.
 36. Rasmussen CE and Williams CKI. *Gaussian processes for machine learning*. Cambridge: MIT Press, 2004.
 37. Cutajar K. Broadening the scope of gaussian processes for large-scale learning, <https://www.eurecom.fr/publication/5852> (2019, accessed 11 February 2020).
 38. McNeish D. On using Bayesian methods to address small sample problems. *Struct Equ Model A Multidiscip J* 2016; 23: 750–773.
 39. Ioffe S and Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. *arXiv*. Epub ahead of print 10 February 2015. DOI: 10.1007/s13398-014-0173-7.2.
 40. Matthews AGDG, van der Wilk M, Nickson T, et al. GPflow: a Gaussian process library using TensorFlow. *J Mach Learn Res* 2016; 18: 1–6.
 41. Watabe H, Ikoma Y, Kimura Y, et al. PET kinetic analysis—compartmental model. *Ann Nucl Med* 2006; 20: 583–588.
 42. Svarer C, Madsen K, Hasselbalch SG, et al. MR-based automatic delineation of volumes of interest in human brain PET images using probability maps. *Neuroimage* 2005; 24: 969–979.
 43. Morris ED, Endres CJ, Schmidt KC, et al. Kinetic modeling in positron emission tomography. In: *Emission tomography*. Amsterdam: Elsevier, 2004, chapter 23.
 44. Quarles RP, Mintun MA, Larson KB, et al. Measurement of regional cerebral blood flow with positron emission tomography: a comparison of [15 O]water to [11 C]butanol with distributed-parameter and compartmental models. *J Cereb Blood Flow Metab* 1993; 13: 733–747.
 45. Hubert M and Van der Veeken S. Outlier detection for skewed data. *J Chemometrics* 2008; 22: 235–246.
 46. Martin Bland J and Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 327: 307–310.
 47. Richardson RR, Osborne MA and Howey DA. Gaussian process regression for forecasting battery state of health. *J Power Sources* 2017; 357: 209–219.
 48. Grandin C, Bol A, Smith A, et al. Absolute CBF and CBV measurements by MRI bolus tracking before and after acetazolamide challenge: repeatability and comparison with PET in humans. *Neuroimage* 2005; 26: 525–535.
 49. Law I, Iida H, Holm S, et al. Quantitation of regional cerebral blood flow corrected for partial volume effect using 0-15 water and PET: II. Normal values and gray matter blood flow response to visual activation. *J Cereb Blood Flow Metab* 2000; 20: 1252–1263.
 50. Kelly CJ, Karthikesalingam A, Suleyman M, et al. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019; 17: 195.
 51. Qayyum A, Qadir J, Bilal M, et al. Secure and robust machine learning for healthcare: a survey. *arXiv* 2020; 1–22.
 52. Thiagarajan JJ, Rajan D and Sattigeri P. Understanding behavior of clinical models under domain shifts. *arXiv* <http://arxiv.org/abs/1809.07806> (2018).
 53. Bremmer JP, van Berckel BNM, Persoon S, et al. Day-to-day test–retest variability of CBF, CMRO₂, and OEF measurements using dynamic 15O PET studies. *Mol Imaging Biol* 2011; 13: 759–768.
 54. Bokkers RP, Bremmer JP, van Berckel BN, et al. Arterial spin labeling perfusion MRI at multiple delay times: a correlative study with H 2 15 O positron emission tomography in patients with symptomatic carotid artery occlusion. *J Cereb Blood Flow Metab* 2010; 30: 222–229.

Supplementary material

Cerebral blood flow measurements with ^{15}O -water PET using a non-invasive machine-learning-derived arterial input function

Samuel Kuttner^{1,2,3*}, Kristoffer Knutsen Wickstrøm², Mark Lubberink⁴, Andreas Tolf⁵, Joachim Burman⁵, Rune Sundset^{1,3}, Robert Jenssen², Lieuwe Appel⁴, Jan Axelsson⁶

¹Nuclear Medicine and Radiation Biology Research Group, Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, Norway.

²UiT Machine Learning Group, Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, Norway.

³The PET Imaging Center, University Hospital of North Norway, Tromsø, Norway.

⁴Department of Surgical Sciences, Radiology, Uppsala University, Uppsala, Sweden.

⁵Department of Neuroscience, Uppsala University, Uppsala, Sweden.

⁶Department of Radiation Sciences, Umeå University, Umeå, Sweden.

*Corresponding and first author:

Samuel Kuttner, Nuclear Medicine and Radiation Biology Research Group,
Department of Clinical Medicine, UiT The Arctic University of Norway, 9037
Tromsø, Norway

Phone: +47-77 66 99 53

e-mail: samuel.kuttner@uit.no

Twitter: @samuel_kuttner

Table S1. Comparison of individual whole-brain grey matter CBF values for baseline and acetazolamide ^{15}O -water brain PET scans estimated by kinetic modelling using AIF and GP-predicted MLIF as input function. Ratios are expressed as $\text{CBF}_{\text{MLIF}}/\text{CBF}_{\text{AIF}}$. For subjects marked with asterisk (*), the AIF and GP-predicted MLIF from the baseline scan are shown in Figure 2. Scans marked with dagger (†) were outliers according to the slope-based criterion defined in the main text.

Subject	CBF [$\text{ml}\cdot\text{min}^{-1}\cdot\text{g}^{-1}$]					
	Baseline			Acetazolamide		
	CBF_{AIF}	$\text{CBF}_{\text{MLIF1}}$	Ratio	CBF_{AIF}	$\text{CBF}_{\text{MLIF2}}$	Ratio
1	0.55	0.61	1.12	0.69	0.66	0.95
2	0.29	0.38	1.30	0.35	0.39	1.12
3*	0.50	0.49	0.98	0.63	0.63	1.00
4	0.33	0.37	1.13	0.48	0.49	1.03
5	0.39	0.51	1.30	0.59	0.59	0.99
6	0.39	0.53	1.34	0.53	0.60	1.13
7	0.30	0.29	0.98	0.33	0.37	1.12
8	0.42	0.49	1.18	0.79	0.77	0.96
9	0.45	0.36	0.80	0.57	0.49	0.86
10*	0.41	0.41	1.01	0.38	0.45	1.17
11	0.36	0.39	1.10	0.45	0.52	1.18
12	0.32	0.37	1.14	0.61	0.48	0.79
13	0.27	0.31	1.17	0.39	0.35	0.89
14	0.43	0.69	1.60	0.67	1.05	1.56
15*	0.57	0.41	0.72	0.63	0.50	0.79
16	0.52	0.52	0.99	0.74	0.71	0.97
17	0.57	0.45	0.79	0.77	0.82	1.06
18	0.76†	0.44†	0.58†	0.83	0.81	0.98
19	0.91†	0.49†	0.54†	0.48	0.58	1.20
20	0.49	0.40	0.82	0.63	0.51	0.82
21	0.57†	0.42†	0.74†	0.69†	0.54†	0.78†
22*	0.30	0.44	1.47	0.44	0.56	1.26
Mean	0.46	0.44	1.04	0.58	0.59	1.03
SD	0.16	0.09	0.28	0.15	0.17	0.19
P		0.67			0.73	

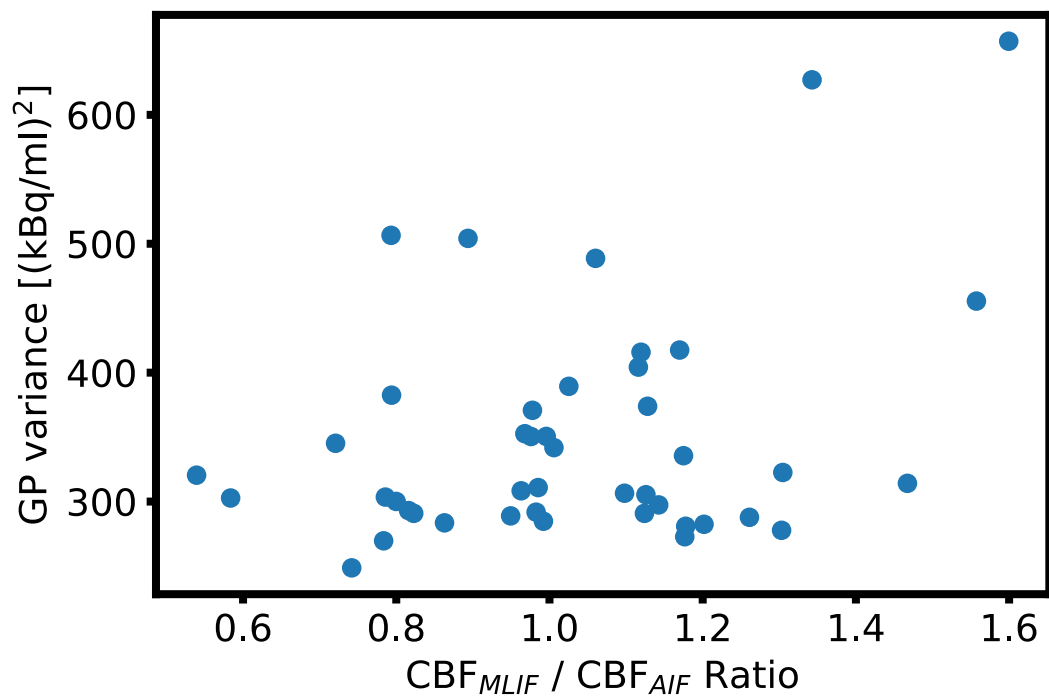


Figure S1. Scatter plot of predicted GP variance versus CBF_{MLIF}/CBF_{AIF} ratio. No clear relation between GP variance and high or low ratio values was observed in the data.

MLIF scale dependency

We have further investigated how the MLIF model responds to input IDIFs outside the range of the training data. This was evaluated by multiplying the input IDIF time-activity curves of the test subject with a scale factor during the leave-one-out testing. Thus, the MLIF model was trained on non-scaled training data, consisting of all subjects except the one currently left out as test patient. Subsequently the input IDIFs of the test subject was scaled by multiplication of the scale factor. The scaled IDIFs were used as input into the trained MLIF model, to generate a predicted scaled MLIF curve. The scale-dependency of the MLIF model was evaluated by comparing the area-under-curve (AUC) of the scaled input IDIFs and AUC of the corresponding scaled output MLIF curves. Scale factors, s , in the range 0.5 to 2 were investigated ($s = 0.5, 0.75, 1.0, 1.25, 1.5, 2$).

Figure S2A displays the generated MLIF AUC as a function of the scaled IDIF AUCs for all subjects. This graph shows that the model-generated MLIFs scales non-linearly with scaled IDIFs. Furthermore, it can be deduced that the model works well within the range of the training data (scale, $s = 1$). This can be expected for any data-driven, machine learning-based method. The model breaks for higher scale factors, at IDIFs with an AUC above 2000 kBq-s/ml, where the AUCs are about 1.3 times the values used for training. There is also a tendency for the lowest scale factor data ($s = 0.5$) to fail at the values furthest from the range of the training data ($s = 1$). Figure S2A suggest that it is safe to predict MLIF data within the training range, but also for data with amplitudes (or AUC) within a range being extended from the training data range by 25% ($s=0.75$ to 1.25)

These results confirm that the normalization used as input for the training does not affect the MLIF model's capability to infer the MLIF amplitude from the IDIF.

The reason for the non-linear dependency of the MLIF AUC on IDIF scaling can be understood from Figure S2B. Also here, in the original arterial input data, there is a non-linear dependency, in this case between the IDIF and the measured AIF AUC values. This non-linearity may be hard to appreciate from Figure S2B but becomes more obvious considering that a line-fit has a slope of 1.5 for the range of training data. Furthermore, the y-intercept of the linear fit is $\gg 0$ kBq-s/ml, which is not describing the reality where zero blood input would give zero IDIF from images.

To summarize, this observation suggests that the non-linear relationship observed at different scaling of the IDIF (Figure S2A) is a feature carried over from the original data. Thus, the non-linearity is not a defect by the MLIF-model but arise from the non-linear scaling of the AIF with the scaled IDIFs (Figure S2B). This is in itself an interesting observation, which may arise from the influence of the neighboring tissue to the pixels defining the IDIF. However, further investigation of this effect is outside the scope of the current work.

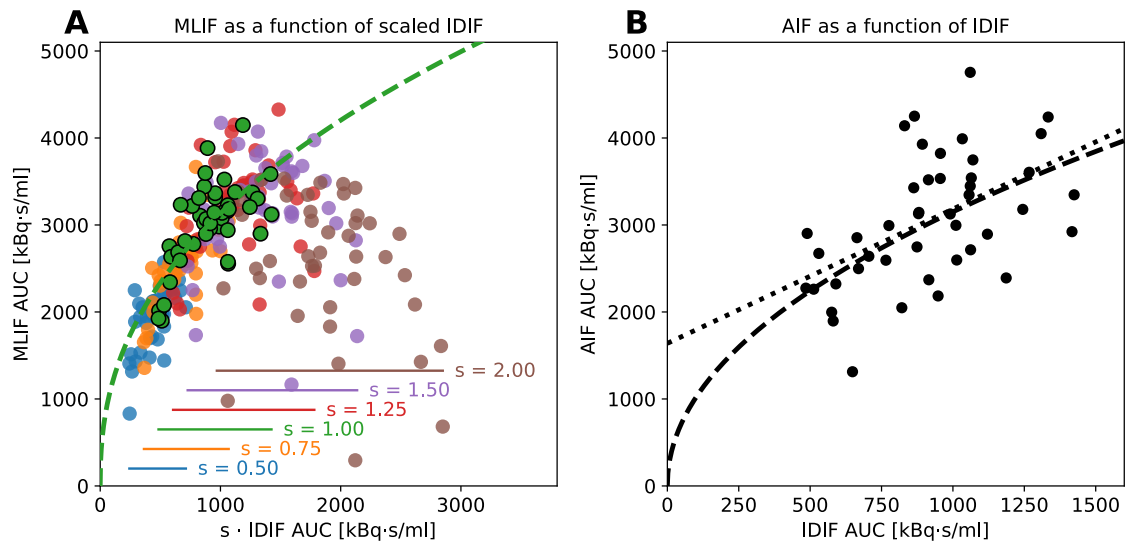


Figure S2. A. Area-under-curve (AUC) of predicted MLIF for different scaled input-IDIF AUCs. IDIF AUC is calculated as the mean AUC of IDIF₁₀, IDIF₁₀₀ and IDIF₁₀₀₀. Colors indicate different scale factors, s . Green circles display non-scaled values ($s = 1$). Horizontal bars indicate the AUC range of predicted MLIF AUC values for each scale factor. The graph indicates that the model explains data well within the trained range but breaks down when extending input data too far outside the trained range ($s = 1$). The curve is a fit to a power function on the non-scaled data, as a visual guide (dashed green line). B. AUC of the measured AIF and IDIF (not scaled) show that the AIF does not scale linearly (linear fit: $y = 1.5x + 1639$) with IDIF, but rather follows a power fit (dashed line).

Bibliography

- [1] Jonathan D Berry and Gary J R Cook. “Positron emission tomography in oncology.” In: *Br. Med. Bull.* 79-80.1 (2006), pp. 171–186.
- [2] Z. Brady et al. “The clinical application of PET/CT: a contemporary review.” In: *Australas. Phys. Eng. Sci. Med.* 31.2 (2008), pp. 90–109.
- [3] Erik Mittra and Andrew Quon. “Positron Emission Tomography/Computed Tomography: The Current Technology and Applications.” In: *Radiol. Clin. North Am.* 47.1 (2009), pp. 147–160.
- [4] Fabian Kiessling and Bernd J Pichler, eds. *Small Animal Imaging*. Springer-Verlag Berlin Heidelberg, 2011.
- [5] Simon R. Cherry and Magnus Dahlbom. “PET: Physics, Instrumentation, and Scanners.” In: *PET: Physics, Instrumentation, and Scanners*. Ed. by Phel. New York, NY: Springer New York, 2004, pp. 1–124.
- [6] S R Cherry, J A Sorenson, and M E Phelps. *Physics in Nuclear Medicine*. 4th ed. Philadelphia, Pennsylvania: Elsevier Inc., 2012.
- [7] Ilja Bezrukov et al. “MR-based PET attenuation correction for PET/MR imaging.” In: *Semin. Nucl. Med.* 43.1 (2013), pp. 45–59.
- [8] A. Martinez-Moller et al. “Tissue Classification as a Potential Approach for Attenuation Correction in Whole-Body PET/MRI: Evaluation with PET/CT Data.” In: *J. Nucl. Med.* 50.4 (2009), pp. 520–526.
- [9] I. Bezrukov et al. “MR-Based Attenuation Correction Methods for Improved PET Quantification in Lesions Within Bone and Susceptibility Artifact Regions.” In: *J. Nucl. Med.* 54.10 (2013), pp. 1768–1774.
- [10] G. Schramm et al. “Influence and Compensation of Truncation Artifacts in MR-Based Attenuation Correction in PET/MR.” In: *IEEE Trans. Med. Imaging* 32.11 (2013), pp. 2056–2063.
- [11] Claes Nøhr Ladefoged et al. “Impact of incorrect tissue classification in Dixon-based MR-AC: fat-water tissue inversion.” In: *EJNMMI Phys.* 1.1 (2014), p. 101.

Bibliography

- [12] Martin Lyngby Lassen et al. “Assessment of attenuation correction for myocardial PET imaging using combined PET/MRI.” In: *J. Nucl. Cardiol.* (2017), pp. 1–12.
- [13] Anders Olin et al. “Reproducibility of MR-Based Attenuation Maps in PET/MRI and the Impact on PET Quantification in Lung Cancer.” In: *J. Nucl. Med.* 59.6 (2018), pp. 999–1004.
- [14] *Global Cancer Observatory: Cancer Today*. Lyon, 2020. URL: <https://gco.iarc.fr/today> (visited on 01/25/2021).
- [15] Arifa Pasic, Pieter E. Postmus, and Thomas G. Sutedja. “What is early lung cancer?” In: *Lung Cancer* 45.3 (2004), pp. 267–277.
- [16] Li Fan et al. “PET/MRI in Lung Cancer.” In: *Semin. Roentgenol.* 49.4 (2014), pp. 291–303.
- [17] Paul Flechsig et al. “PET/MRI and PET/CT in lung lesions and thoracic malignancies.” In: *Semin. Nucl. Med.* 45.4 (2015), pp. 268–281.
- [18] Johannes Czernin, Lisa Ta, and Ken Herrmann. “Does PET/MR Imaging Improve Cancer Assessments? Literature Evidence from More Than 900 Patients.” In: *J. Nucl. Med.* 55.Supplement_2 (2014), 59S–62S.
- [19] Mark Wielpütz and Hans Ulrich Kauczor. “MRI of the lung: State of the art.” In: *Diagnostic Interv. Radiol.* 18.4 (2012), pp. 344–353.
- [20] Yoshiharu Ohno, Kazuro Sugimura, and Hiroto Hatabu. “MR imaging of lung cancer.” In: *Eur. J. Radiol.* 44.3 (2002), pp. 172–181.
- [21] Bingsheng Huang, Martin Wai-ming Law, and Pek-Lan Khong. “Whole-Body PET/CT Scanning: Estimation of Radiation Dose and Cancer Risk.” In: *Radiology* 251.1 (2009), pp. 166–174.
- [22] Wouter van Elmpt et al. “Multiparametric imaging of patient and tumour heterogeneity in non-small-cell lung cancer: quantification of tumour hypoxia, metabolism and perfusion.” In: *Eur. J. Nucl. Med. Mol. Imaging* 43.2 (2016), pp. 240–248.
- [23] Eugene Huo et al. “The Role of PET/MR Imaging in Precision Medicine.” In: *PET Clin.* 12.4 (2017), pp. 489–501.
- [24] Miles N. Wernick et al. “Machine Learning in Medical Imaging.” In: *IEEE Signal Process Mag.* 27.4 (2014), pp. 25–38.
- [25] Yong Xue et al. “Application of Deep Learning in Automated Analysis of Molecular Images in Cancer: A Survey.” In: *Contrast Media Mol. Imaging* 2017 (2017), pp. 1–10.

- [26] Bradley J. Erickson et al. “Machine Learning for Medical Imaging.” In: *RadioGraphics* 37.2 (2017), pp. 505–515.
- [27] Bo He et al. “A biomarker basing on radiomics for the prediction of overall survival in non-small cell lung cancer patients.” In: *Respir. Res.* 19.1 (2018), pp. 1–8.
- [28] Stephen S.F. Yip et al. “Associations Between Somatic Mutations and Metabolic Imaging Phenotypes in Non–Small Cell Lung Cancer.” In: *J. Nucl. Med.* 58.4 (2017), pp. 569–576.
- [29] Weimiao Wu et al. “Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology.” In: *Front. Oncol.* 6.March (2016), pp. 1–11.
- [30] Pierre Lovinfosse et al. “FDG PET/CT texture analysis for predicting the outcome of lung cancer treated by stereotactic body radiation therapy.” In: *Eur. J. Nucl. Med. Mol. Imaging* 43.8 (2016), pp. 1453–1460.
- [31] Hugo J. W. L. Aerts et al. “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach.” In: *Nat. Commun.* 5.1 (2014), pp. 1–8.
- [32] G. J. R. Cook et al. “Are Pretreatment 18F-FDG PET Tumor Textural Features in Non-Small Cell Lung Cancer Associated with Response and Survival After Chemoradiotherapy?” In: *J. Nucl. Med.* 54.1 (2013), pp. 19–26.
- [33] Antonia Dimitrakopoulou-Strauss, Leyun Pan, and Christos Sachpekidis. “Kinetic modeling and parametric imaging with dynamic PET for oncological applications: general considerations, current clinical applications, and future perspectives.” In: *Eur. J. Nucl. Med. Mol. Imaging* (2020).
- [34] Max Wintermark et al. “Comparative Overview of Brain Perfusion Imaging Techniques.” In: *Stroke* 36.9 (2005), pp. 1–29.
- [35] P Herscovitch, J Markham, and M E Raichle. “Brain blood flow measured with intravenous H₂(15)O. I. Theory and error analysis.” In: *J. Nucl. Med.* 24.9 (1983), pp. 782–9.
- [36] M. E. Raichle et al. “Brain blood flow measured with intravenous H₂(15)O. II. Implementation and validation.” In: *J. Nucl. Med.* 24.9 (1983), pp. 790–8.
- [37] Rodney J. Hicks. “PET tracer development - a tale of mice and men.” In: *Cancer Imaging* 6 (2006), S102–S106.
- [38] Wolfgang A. Weber. “Use of PET for monitoring cancer therapy and for predicting outcome.” In: *J. Nucl. Med.* 46.6 (2005), pp. 983–995.
- [39] Habib Zaidi. *Quantitative Analysis in Nuclear Medicine Imaging*. Vol. 52. 11. Springer, 2006.

Bibliography

- [40] Mark Muzi et al. “Quantitative assessment of dynamic PET imaging data in cancer imaging.” In: *Magn. Reson. Imaging* 30.9 (2012), pp. 1203–1215.
- [41] Sune H. Keller et al. “Image artifacts from MR-based attenuation correction in clinical, whole-body PET/MRI.” In: *Magn. Reson. Mater. Physics, Biol. Med.* 26.1 (2013), pp. 173–181.
- [42] Gudrun Wagenknecht et al. “MRI for attenuation correction in PET: methods and challenges.” In: *Magn. Reson. Mater. Physics, Biol. Med.* 26.1 (2013), pp. 99–113.
- [43] Thomas Beyer et al. “Investigating the state-of-the-art in whole-body MR-based attenuation correction: an intra-individual, inter-system, inventory study on three clinical PET/MR systems.” In: *Magn. Reson. Mater. Physics, Biol. Med.* 29.1 (2016), pp. 75–87.
- [44] Jasmine A. Oliver et al. “Variability of image features computed from conventional and respiratory-gated PET/CT images of lung cancer.” In: *Transl. Oncol.* 8.6 (2015), pp. 524–534.
- [45] Stephen Yip et al. “Comparison of texture features derived from static and respiratory-gated PET images in non-small cell lung cancer.” In: *PLoS One* 9.12 (2014), pp. 1–14.
- [46] S H Keller et al. “Image Distortions in Clinical PET/MR Imaging.” In: *PET/MRI*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 21–41.
- [47] V Dhawan et al. “Accuracy of PET RCBF measurements: effective of time shift between blood and brain radioactivity curves.” In: *Phys. Med. Biol.* 31.5 (1986), pp. 507–514.
- [48] H. Iida et al. “Error analysis of a quantitative cerebral blood flow measurement using H²¹⁵O autoradiography and positron emission tomography, with respect to the dispersion of the input function.” In: *J. Cereb. Blood Flow Metab.* 6.5 (1986), pp. 536–545.
- [49] R. A. Koeppe et al. “Examination of assumptions for local cerebral blood flow studies in PET.” In: *J. Nucl. Med.* 28.11 (1987), pp. 1695–1703.
- [50] S Esmail Dorraji et al. “PET imaging of Tertiary Lymphoid Structures (TLS) in development of Lupus nephritis.” In: *8th National Conference in Medical Imaging MedIm*. 2016.
- [51] Frank Herbert Attix. *Introduction to Radiological Physics and Radiation Dosimetry*. Weinheim, Germany: Wiley-VCH Verlag GmbH, 1986.
- [52] G. Delso et al. “Performance Measurements of the Siemens mMR Integrated Whole-Body PET/MR Scanner.” In: *J. Nucl. Med.* 52.12 (2011), pp. 1914–1922.

- [53] Stefaan Vandenberghe and Paul K. Marsden. “PET-MRI: A review of challenges and solutions in the development of integrated multimodality imaging.” In: *Phys. Med. Biol.* 60.4 (2015), R115–R154.
- [54] Andrew L Goertzen et al. “NEMA NU 4-2008 Comparison of Preclinical PET Imaging Systems.” In: *J. Nucl. Med.* 53.8 (2012), pp. 1300–1309.
- [55] Charalampos Tsoumpas, Dimitris Visvikis, and George Loudos. “Innovations in Small-Animal PET/MR Imaging Instrumentation.” In: *PET Clin.* 11.2 (2016), pp. 105–118.
- [56] Srilalan Krishnamoorthy et al. “Performance evaluation of the MOLE-CUBES β -CUBE - A high spatial resolution and high sensitivity small animal PET scanner utilizing monolithic LYSO scintillation detectors.” In: *Phys. Med. Biol.* 63.15 (2018).
- [57] C. C. Watson, D. Newport, and M. E. Casey. “A Single Scatter Simulation Technique for Scatter Correction in 3D PET.” In: *Three-Dimensional Image Reconstruction in Radiation and Nuclear Medicine*. 1996, pp. 255–268.
- [58] John M. Ollinger. “Model-based scatter correction for fully 3D PET.” In: *Phys. Med. Biol.* 41.1 (1996), pp. 153–176.
- [59] Jarmo Teuho et al. “Magnetic Resonance-Based Attenuation Correction and Scatter Correction in Neurological Positron Emission Tomography/Magnetic Resonance Imaging—Current Status With Emerging Applications.” In: *Front. Phys.* 7.January (2020).
- [60] Joseph A. Thie. “Understanding the standardized uptake value, its methods, and implications for usage.” In: *J. Nucl. Med.* 45.9 (2004), pp. 1431–1434.
- [61] Lale Kostakoglu, Harry Agress, and Stanley J Goldsmith. “Clinical Role of FDG PET in Evaluation of Cancer Patients.” In: *RadioGraphics* 23.2 (2003), pp. 315–340.
- [62] Wolfgang A. Weber. “Positron Emission Tomography As an Imaging Biomarker.” In: *J. Clin. Oncol.* 24.20 (2006), pp. 3282–3292.
- [63] Shankar Vallabhajosula, Lilja Solnes, and Brigitte Vallabhajosula. “A broad overview of positron emission tomography radiopharmaceuticals and clinical applications: What is new?” In: *Semin. Nucl. Med.* 41.4 (2011), pp. 246–264.
- [64] Ronald Boellaard et al. “FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0.” In: *Eur. J. Nucl. Med. Mol. Imaging* 42.2 (2015), pp. 328–354.
- [65] Arman Rahmim et al. “Dynamic whole-body PET imaging: principles, potentials and applications.” In: *Eur. J. Nucl. Med. Mol. Imaging* 46.2 (2019), pp. 501–518.

Bibliography

- [66] William W. Moses. “Fundamental limits of spatial resolution in PET.” In: *Nucl. Instrum. Methods. Phys. Res. B* 648.SUPPL. 1 (2011), S236–S240.
- [67] M. Soret, S. L. Bacharach, and I. Buvat. “Partial-Volume Effect in PET Tumor Imaging.” In: *J. Nucl. Med.* 48.6 (2007), pp. 932–945.
- [68] Beverley F. Holman et al. “Improved correction for the tissue fraction effect in lung PET/CT imaging.” In: *Phys. Med. Biol.* 60.18 (2015), pp. 7387–7402.
- [69] T.S. Curry, J.E. Dowdey, and R.C. Murry. *Christensen’s Physics of Diagnostic Radiology*. Lea & Febiger, 1990.
- [70] A.D. Elster and J.H. Burdette. *Questions & Answers in Magnetic Resonance Imaging*. Mosby, 2001.
- [71] J. Valentin. “Annals of the ICRP ICRP PUBLICATION 95 Doses to Infants from Ingestion of Radionuclides in Mothers’ Milk.” In: *Ann. ICRP* 34.3-4 (2004), pp. 1–27.
- [72] Kathy P. Willowson, Elizabeth A. Bailey, and Dale L. Bailey. “A retrospective evaluation of radiation dose associated with low dose FDG protocols in whole-body PET/CT.” In: *Australas. Phys. Eng. Sci. Med.* 35.1 (2012), pp. 49–53.
- [73] Josep M. Martí-Climent et al. “Effective dose estimation for oncological and neurological PET/CT procedures.” In: *EJNMMI Res.* 7.1 (2017), p. 37.
- [74] Sekse Tonje et al. *Yrkeseksponering i Norge. Ioniserende stråling. Ikke-ioniserende stråling. StrålevernRapport 2005:15*. Tech. rep. Østerås: Statens strålevern, 2005.
- [75] H. Zaidi et al. “Design and performance evaluation of a whole-body Ingenuity TF PET–MRI system.” In: *Phys. Med. Biol.* 56.10 (2011), pp. 3091–3106.
- [76] Andrew P Leynes et al. “Zero-Echo-Time and Dixon Deep Pseudo-CT (ZeDD CT): Direct Generation of Pseudo-CT Images for Pelvic PET/MRI Attenuation Correction Using Deep Convolutional Neural Networks with Multiparametric MRI.” In: *J. Nucl. Med.* 59.5 (2018), pp. 852–858.
- [77] Fang Liu et al. “Deep Learning MR Imaging–based Attenuation Correction for PET/MR Imaging.” In: *Radiology* 286.2 (2018), pp. 676–684.
- [78] Waheeda Sureshbabu and Osama Mawlawi. “PET / CT Imaging Artifacts.” In: *J. Nucl. Med. Technol.* 33.3 (2005), pp. 156–162.
- [79] Todd M. Blodgett et al. “PET/CT artifacts.” In: *Clin. Imaging* 35.1 (2011), pp. 49–63.
- [80] Johan Nuyts et al. “Completion of a Truncated Attenuation Image From the Attenuated PET Emission Data.” In: *IEEE Trans. Med. Imaging* 32.2 (2013), pp. 237–246.

- [81] Jan O. Blumhagen et al. “Field of view extension and truncation correction for MR-based human attenuation correction in simultaneous MR/PET imaging.” In: *Med. Phys.* 41.2 (2014), p. 022303.
- [82] D. H. Paulus et al. “Whole-Body PET/MR Imaging: Quantitative Evaluation of a Novel Model-Based MR Attenuation Correction Method Including Bone.” In: *J. Nucl. Med.* 56.7 (2015), pp. 1061–1066.
- [83] Mattijs Elschot et al. “The Effect of Including Bone in Dixon-Based Attenuation Correction for 18 F-Fluciclovine PET/MRI of Prostate Cancer.” In: *J. Nucl. Med.* 59.12 (2018), pp. 1913–1917.
- [84] V. Keereman et al. “MRI-Based Attenuation Correction for PET/MRI Using Ultrashort Echo Time Sequences.” In: *J. Nucl. Med.* 51.5 (2010), pp. 812–818.
- [85] Andoni Toms, Sharief Aboelmagd, and Paul Malcolm. “Magnetic resonance imaging of metal artifact reduction sequences in the assessment of metal-on-metal hip prostheses.” In: *Reports Med. Imaging* 7.1 (2014), p. 65.
- [86] G. Schramm et al. “Evaluation and automatic correction of metal-implant-induced artifacts in MR-based attenuation correction in whole-body PET/MR imaging.” In: *Phys. Med. Biol.* 59.11 (2014), pp. 2713–2726.
- [87] Reto Sutter et al. “Reduction of Metal Artifacts in Patients with Total Hip Arthroplasty with Slice-encoding Metal Artifact Correction and View-Angle Tilting MR Imaging.” In: *Radiology* 265.1 (2012), pp. 204–214.
- [88] I. A. Burger et al. “Hybrid PET/MR Imaging: An Algorithm to Reduce Metal Artifacts from Dental Implants in Dixon-Based Attenuation Map Generation Using a Multiacquisition Variable-Resonance Image Combination Sequence.” In: *J. Nucl. Med.* 56.1 (2015), pp. 93–97.
- [89] M. Reichert et al. “Metal artefact reduction in MRI at both 1.5 and 3.0 T using slice encoding for metal artefact correction and view angle tilting.” In: *Br. J. Radiol.* 88.1048 (2015), p. 20140601.
- [90] Brett S Talbot and Eric P Weinberg. “MR Imaging with Metal-suppression Sequences for Evaluation of Total Joint Arthroplasty.” In: *RadioGraphics* 36.1 (2016), pp. 209–225.
- [91] Claes Nøhr Ladefoged et al. “PET/MR imaging of the pelvis in the presence of endoprostheses: Reducing image artifacts and increasing accuracy through inpainting.” In: *Eur. J. Nucl. Med. Mol. Imaging* 40.4 (2013), pp. 594–601.
- [92] Claes N. Ladefoged et al. “Automatic correction of dental artifacts in PET/MRI.” In: *J. Med. Imaging* 2.2 (2015), p. 024009.

Bibliography

- [93] Niccolo Fuin et al. “PET/MRI in the Presence of Metal Implants: Completion of the Attenuation Map from PET Emission Data.” In: *J. Nucl. Med.* 58.5 (2017), pp. 840–845.
- [94] Norbert Lang et al. “Organ Movement Reduction in PET/CT using Dual-Gated Listmode Acquisition.” In: *Z. Med. Phys.* 16.1 (2006), pp. 93–100.
- [95] Mohammad Dawood et al. “Respiratory gating in positron emission tomography: A quantitative comparison of different gating schemes.” In: *Med. Phys.* 34.7 (2007), pp. 3067–3076.
- [96] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. “Performance of optical flow techniques.” In: *Int. J. Comput. Vis.* 12.1 (1994), pp. 43–77.
- [97] Mohammad Dawood et al. “Lung motion correction on respiratory gated 3-D PET/CT images.” In: *IEEE Trans. Med. Imaging* 25.4 (2006), pp. 476–485.
- [98] S. Furst et al. “Motion Correction Strategies for Integrated PET/MR.” In: *J. Nucl. Med.* 56.2 (2015), pp. 261–269.
- [99] Hiroshi Watabe et al. “PET kinetic analysis—compartmental model.” In: *Ann. Nucl. Med.* 20.9 (2006), pp. 583–588.
- [100] R N Gunn, S R Gunn, and V J Cunningham. “Positron emission tomography compartmental models.” In: *J. Cereb. Blood Flow Metab.* 21.6 (2001), pp. 635–52.
- [101] Evan D Morris et al. “Kinetic Modeling in Positron Emission Tomography.” In: *Emission Tomography*. 2004. Chap. 23.
- [102] E Phelps, J Hoffman, and David E Kuhl. “Noninvasive determination of local cerebral metabolic rate of glucose in man.” In: *Am. J. Physiol. - Endocrinol. Metab.* 238.1 (1980), E69–E82.
- [103] Y.-H. Dean Fang and Raymond F Muzic. “Spillover and Partial-Volume Correction for Image-Derived Input Functions for Small-Animal 18F-FDG PET Studies.” In: *J. Nucl. Med.* 49.4 (2008), pp. 606–614.
- [104] Michael M. Graham et al. “The FDG lumped constant in normal human brain.” In: *J. Nucl. Med.* 43.9 (2002), pp. 1157–1166.
- [105] Kenneth A Krohn, Mark Muzi, and Alexander M Spence. “What Is in a Number ? The FDG Lumped Constant in the Rat Brain.” In: *J. Nucl. Med.* 48 (2007), pp. 5–7.
- [106] Richard E Carson. “Tracer Kinetic Modeling in PET.” In: *Positron Emission Tomography*. London: Springer-Verlag, 2003, pp. 127–159.
- [107] B. A. Everett et al. “Safety of Radial Arterial Catheterization in PET Research Subjects.” In: *J. Nucl. Med.* 50.10 (2009), pp. 1742–1742.

- [108] Andreas Otte et al. *PET and SPECT of Neurobiological Systems*. Ed. by Rudi A.J.O. Dierckx et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.
- [109] Anders N. Christensen et al. “Calibrated image-derived input functions for the determination of the metabolic uptake rate of glucose with [18F]-FDG PET.” In: *Nucl. Med. Commun.* 35.4 (2014), pp. 353–361.
- [110] Richard Laforest et al. “Measurement of input functions in rodents: Challenges and solutions.” In: *Nucl. Med. Biol.* 32.7 (2005), pp. 679–685.
- [111] Laurence Convert et al. “A New Tool for Molecular Imaging: The Microvolumetric Blood Counter.” In: *J. Nucl. Med.* 48.7 (2007), pp. 1197–1206.
- [112] Etienne Croteau et al. “Arterial input function sampling without surgery in rats for positron emission tomography molecular imaging.” In: *Nucl. Med. Commun.* 35.6 (2014), pp. 666–676.
- [113] Malte F Alf et al. “Quantification of brain glucose metabolism by 18F-FDG PET with real-time arterial and image-derived input function in mice.” In: *J. Nucl. Med.* 54.1 (2013), pp. 132–8.
- [114] Laurence Convert et al. “Real-Time Microfluidic Blood-Counting System for PET and SPECT Preclinical Pharmacokinetic Studies.” In: *J. Nucl. Med.* 57.9 (2016), pp. 1460–1466.
- [115] Christian Vanhove et al. “Accurate molecular imaging of small animals taking into account animal models, handling, anaesthesia, quality control and imaging system performance.” In: *EJNMMI Phys.* 2.1 (2015), p. 31.
- [116] Hsiao-Ming Wu et al. “In vivo quantitation of glucose metabolism in mice using small-animal PET and a microfluidic device.” In: *J. Nucl. Med.* 48 (2007), pp. 837–845.
- [117] Karl-Heinz Diehl et al. “A good practice guide to the administration of substances and removal of blood, including routes and volumes.” In: *J. Appl. Toxicol.* 21.1 (2001), pp. 15–23.
- [118] E. Meyer. “Simultaneous correction for tracer arrival delay and dispersion in CBF measurements by the H215O autoradiographic method and dynamic PET.” In: *J. Nucl. Med.* 30.6 (1989), pp. 1069–78.
- [119] Bruno Weber et al. “A femoral arteriovenous shunt facilitates arterial whole blood sampling in animals.” In: *Eur. J. Nucl. Med. Mol. Imaging* 29.3 (2002), pp. 319–323.
- [120] Cyrill Burger and Alfred Buck. “Tracer kinetic modelling of receptor data with mathematical metabolite correction.” In: *Eur. J. Nucl. Med.* 23.5 (1996), pp. 539–545.

Bibliography

- [121] S Takikawa et al. “Noninvasive quantitative fluorodeoxyglucose PET studies with an estimated input function derived from a population-based arterial blood curve.” In: *Radiology* 188.1 (1993), pp. 131–136.
- [122] Stefan Eberl et al. “Evaluation of two population-based input functions for quantitative neurological FDG PET studies.” In: *Eur. J. Nucl. Med.* 24.3 (1997), pp. 299–304.
- [123] Lioe-Fee de Geus-Oei et al. “Comparison of image-derived and arterial input functions for estimating the rate of glucose metabolism in therapy-monitoring 18F-FDG PET studies.” In: *J. Nucl. Med.* 47.6 (2006), pp. 945–9.
- [124] Jurgen E.M. Mourik et al. “Image derived input functions for dynamic High Resolution Research Tomograph PET brain studies.” In: *Neuroimage* 43.4 (2008), pp. 676–686.
- [125] Jurgen E.M. Mourik et al. “Image-derived input functions for PET brain studies.” In: *Eur. J. Nucl. Med. Mol. Imaging* 36.3 (2009), pp. 463–471.
- [126] Hasan Sari et al. “Estimation of an image derived input function with MR-defined carotid arteries in FDG-PET human studies using a novel partial volume correction method.” In: *J. Cereb. Blood Flow Metab.* 37.4 (2017), pp. 1398–1409.
- [127] Bernard Lanz, Carole Poitry-Yamate, and Rolf Gruetter. “Image-Derived Input Function from the Vena Cava for 18F-FDG PET Studies in Rats and Mice.” In: *J Nucl Med* 55.8 (2014), pp. 1380–1388.
- [128] Paolo Zanotti-Fregonara et al. “Image-derived input function for brain PET studies: Many challenges and few opportunities.” In: *J. Cereb. Blood Flow Metab.* 31.10 (2011), pp. 1986–1998.
- [129] Dagan Feng et al. “A technique for extracting physiological parameters and the required input function simultaneously from PET image measurements: theory and simulation study.” In: *IEEE Trans. Inf. Technol. Biomed.* 1.4 (1997), pp. 243–254.
- [130] K P Wong et al. “Simultaneous estimation of physiological parameters and the input function—in vivo PET data.” In: *IEEE Trans. Inf. Technol. Biomed.* 5.1 (2001), pp. 67–76.
- [131] Elizabeth A Bartlett et al. “Quantification of Positron Emission Tomography Data Using Simultaneous Estimation of the Input Function: Validation with Venous Blood and Replication of Clinical Studies.” In: *Mol. Imaging Biol.* 21.5 (2019), pp. 926–934.

- [132] Arthur Mikhno et al. “Toward Noninvasive Quantification of Brain Radioligand Binding by Combining Electronic Health Records and Dynamic PET Imaging Data.” In: *IEEE J. Biomed. Heal. Informatics* 19.4 (2015), pp. 1271–1282.
- [133] Elisa Rocchia et al. “Quantifying Brain [18 F]FDG Uptake Noninvasively by Combining Medical Health Records and Dynamic PET Imaging Data.” In: *IEEE J. Biomed. Heal. Informatics* 23.6 (2019), pp. 2576–2582.
- [134] H. Iida et al. “Evaluation of Regional Differences of Tracer Appearance Time in Cerebral Tissues Using [15 O]Water and Dynamic Positron Emission Tomography.” In: *J. Cereb. Blood Flow Metab.* 8.2 (1988), pp. 285–288.
- [135] I. Kanno et al. “A System for Cerebral Blood Flow Measurement Using an H 2 15 O Autoradiographic Method and Positron Emission Tomography.” In: *J. Cereb. Blood Flow Metab.* 7.2 (1987), pp. 143–153.
- [136] Muhammad M. Islam et al. “Pixel-by-pixel precise delay correction for measurement of cerebral hemodynamic parameters in H2 15O PET study.” In: *Ann. Nucl. Med.* 31.4 (2017), pp. 283–294.
- [137] Maqsood Yaqub et al. “Optimization algorithms and weighting factors for analysis of dynamic PET studies.” In: *Phys. Med. Biol.* 51.17 (2006), pp. 4217–4232.
- [138] William D. Travis et al. *WHO classification of Tumours of the Lung, Pleura, Thymus and Heart*. Tech. rep. Lyon: International Agency for Research on Cancer, 2015.
- [139] J. Lortet-Tieulent et al. “International trends in lung cancer incidence by histological subtype: Adenocarcinoma stabilizing in men but still increasing in women.” In: *Lung Cancer* 84.1 (2014), pp. 13–22.
- [140] Frank C. Detterbeck et al. “The Eighth Edition Lung Cancer Stage Classification.” In: *Chest* 151.1 (2017), pp. 193–203.
- [141] Bennett S. Greenspan. “Role of PET/CT for precision medicine in lung cancer: perspective of the Society of Nuclear Medicine and Molecular Imaging.” In: *Transl. Lung Cancer Res.* 6.6 (2017), pp. 617–620.
- [142] David Ladrón de Guevara Hernández. “The Role of PET/CT Imaging in Lung Cancer.” In: *J. Cancer Ther.* 6.8 (2015), pp. 690–700.
- [143] Sara E. Dahlsgaard-Wallenius et al. “Hybrid PET/MRI in non-small cell lung cancer (NSCLC) and lung nodules—a literature review.” In: *Eur. J. Nucl. Med. Mol. Imaging* 48.2 (2021), pp. 584–591.

Bibliography

- [144] Fernando E. Boada et al. “Improved Detection of Small Pulmonary Nodules Through Simultaneous MR/PET Imaging.” In: *Magn. Reson. Imaging Clin. N. Am.* 25.2 (2017), pp. 273–279.
- [145] Daniel Thomas Ginat, Juan E. Small, and Pamela Whitney Schaefer. *Neuroimaging Pharmacopoeia*. Ed. by Daniel Thomas Ginat, Juan E. Small, and Pamela Whitney Schaefer. Cham: Springer International Publishing, 2015, pp. 1–382.
- [146] A.S. Vagal et al. “The Acetazolamide Challenge: Techniques and Applications in the Evaluation of Chronic Cerebral Ischemia.” In: *Am. J. Neuroradiol.* 30.5 (2009), pp. 876–884.
- [147] C Grandin et al. “Absolute CBF and CBV measurements by MRI bolus tracking before and after acetazolamide challenge: Repeatability and comparison with PET in humans.” In: *Neuroimage* 26.2 (2005), pp. 525–535.
- [148] Julia G. Mannheim et al. “Small Animal PET Cameras — Development, Technology, PET/CT, PET/MRI.” In: *Trends on the Role of PET in Drug Development*. World Scientific, 2012, pp. 289–317.
- [149] Sergios Theodoridis and Konstantinos Koutroumbas. *Pattern Recognition*. 4th. Academic Press, Inc., 2009.
- [150] Frank Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.
- [151] Siddharth Sharma, Simone Sharma, and Anidhya Athaiya. “Activation functions in neural networks.” In: *Int. J. Eng. Appl. Sci. Technol.* 04.12 (2020), pp. 310–316.
- [152] Sebastian Ruder. “An overview of gradient descent optimization algorithms.” In: 55.2 (2017), pp. 591–596.
- [153] Alex Graves. *Supervised Sequence Labelling with Recurrent Neural Networks*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.
- [154] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [155] S. Hochreiter et al. “Gradient flow in recurrent nets: the difficulty of learning long-term dependencies.” In: *A Field Guide to Dynamical Recurrent Neural Networks*. Ed. by S. C. Kremer and J. F. Kolen. IEEE Press, 2001.
- [156] Sepp Hochreiter and Jürgen Schmidhuber. “Long Short-Term Memory.” In: *Neural Comput.* 9.8 (1997), pp. 1735–1780.
- [157] Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian processes for machine learning*. MIT Press, 2004.

- [158] Kurt Cutajar. “Broadening the Scope of Gaussian Processes for Large-Scale Learning.” PhD thesis. 2019.
- [159] Philippe Lambin et al. “Radiomics: Extracting more information from medical images using advanced feature analysis.” In: *Eur. J. Cancer* 48.4 (2012), pp. 441–446.
- [160] Amit Jethanandani et al. “Exploring applications of radiomics in magnetic resonance imaging of head and neck cancer: A systematic review.” In: *Front. Oncol.* 8.5 (2018).
- [161] Ulrike Schick et al. “MRI-derived radiomics: methodology and clinical applications in the field of pelvic oncology.” In: *Br. J. Radiol.* 92.1104 (2019), p. 20190105.
- [162] Gary J.R. Cook et al. “Challenges and Promises of PET Radiomics.” In: *Int. J. Radiat. Oncol. Biol. Phys.* 102.4 (2018), pp. 1083–1089.
- [163] Francesco Bianconi et al. “Texture Analysis on [18F]FDG PET/CT in Non-Small-Cell Lung Cancer: Correlations Between PET Features, CT Features, and Histological Types.” In: *Mol. Imaging Biol.* 21.6 (2019), pp. 1200–1209.
- [164] Hyun Jung Yoon et al. “Decoding Tumor Phenotypes for ALK, ROS1, and RET Fusions in Lung Adenocarcinoma Using a Radiomics Approach.” In: *Medicine.* 94.41 (2015), e1753.
- [165] Florent Tixier et al. “Visual versus quantitative assessment of intratumor 18F-FDG PET uptake heterogeneity: Prognostic value in non-small cell lung cancer.” In: *J. Nucl. Med.* 55.8 (2014), pp. 1235–1241.
- [166] Thomas Pyka et al. “Textural features in pre-treatment [F18]-FDG-PET/CT are correlated with risk of local recurrence and disease-specific survival in early stage NSCLC patients receiving primary stereotactic radiation therapy.” In: *Radiat. Oncol.* 10.1 (2015), pp. 1–9.
- [167] Nitin Ohri et al. “Pretreatment 18F-FDG PET textural features in locally advanced non-small cell lung cancer: Secondary analysis of ACRIN 6668/RTOG 0235.” In: *J. Nucl. Med.* 57.6 (2016), pp. 843–848.
- [168] Marie-Charlotte Desseroit et al. “Development of a nomogram combining clinical staging with 18F-FDG PET/CT image features in non-small-cell lung cancer stage I–III.” In: *Eur. J. Nucl. Med. Mol. Imaging* 43.8 (2016), pp. 1477–1485.
- [169] Mathieu Hatt et al. “Tumour functional sphericity from PET images: prognostic value in NSCLC and impact of delineation method.” In: *Eur. J. Nucl. Med. Mol. Imaging* 45.4 (2018), pp. 630–641.

Bibliography

- [170] Margarita Kirienko et al. “Prediction of disease-free survival by the PET/CT radiomic signature in non-small cell lung cancer patients undergoing surgery.” In: *Eur. J. Nucl. Med. Mol. Imaging* 45.2 (2018), pp. 207–217.
- [171] Anastasia Oikonomou et al. “Radiomics analysis at PET/CT contributes to prognosis of recurrence and survival in lung cancer treated with stereotactic body radiotherapy.” In: *Sci. Rep.* 8.1 (2018), pp. 1–11.
- [172] Gurvan Dissaux et al. “Pretreatment 18F-FDG PET/CT Radiomics Predict Local Recurrence in Patients Treated with Stereotactic Body Radiotherapy for Early-Stage Non-Small Cell Lung Cancer: A Multicentric Study.” In: *J. Nucl. Med.* 61.6 (2020), pp. 814–820.
- [173] H.K. Ahn et al. “Pre-treatment 18F-FDG PET-based radiomics predict survival in resected non-small cell lung cancer.” In: *Clin. Radiol.* 74.6 (2019), pp. 467–473.
- [174] Pierre Lovinfosse et al. “FDG PET radiomics: a review of the methodological aspects.” In: *Clin. Transl. Imaging* 6.5 (2018), pp. 379–391.
- [175] D Fried et al. “MO-DE-207B-07: Assessment of Reproducibility Of FDG-PET-Based Radiomics Features Across Scanners Using Phantom Imaging.” In: *Med. Phys.* 43.6Part30 (2016), pp. 3705–3706.
- [176] Baderaldeen A. Altazi et al. “Reproducibility of F18-FDG PET radiomic features for different cervical tumor segmentation methods, gray-level discretization, and reconstruction algorithms.” In: *J. Appl. Clin. Med. Phys.* 18.6 (2017), pp. 32–48.
- [177] Isaac Shiri et al. “The impact of image reconstruction settings on 18F-FDG PET radiomic features: multi-scanner phantom and patient studies.” In: *Eur. Radiol.* 27.11 (2017), pp. 4498–4509.
- [178] Floris H. P. van Velden et al. “Repeatability of Radiomic Features in Non-Small-Cell Lung Cancer [18F]FDG-PET/CT Studies: Impact of Reconstruction and Delineation.” In: *Mol. Imaging Biol.* 18.5 (2016), pp. 788–795.
- [179] Ralph T.H. Leijenaar et al. “The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis.” In: *Sci. Rep.* 5.1 (2015), pp. 1–10.
- [180] Alex Zwanenburg et al. “The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping.” In: *Radiology* 295.2 (2020), pp. 328–338.

- [181] Constance A. Owens et al. “Lung tumor segmentation methods: Impact on the uncertainty of radiomics features for non-small cell lung cancer.” In: *PLoS One* 13.10 (2018). Ed. by Yong Fan, e0205003.
- [182] Claus Svarer et al. “MR-based automatic delineation of volumes of interest in human brain PET images using probability maps.” In: *Neuroimage* 24.4 (2005), pp. 969–979.
- [183] Stephen S F Yip and Hugo J W L Aerts. “Applications and limitations of radiomics.” In: *Phys. Med. Biol.* 61.13 (2016), R150–R166.
- [184] M Vallières et al. “A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities.” In: *Phys. Med. Biol.* 60.14 (2015), pp. 5471–96.
- [185] Parnian Afshar et al. “From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities.” In: *IEEE Signal Process. Mag.* 36.4 (2019), pp. 132–160.
- [186] Ryan J. Urbanowicz et al. “Relief-based feature selection: Introduction and review.” In: *J. Biomed. Inform.* 85 (2018), pp. 189–203.
- [187] Mari Kleiveland Opsal. “Metal artefacts in PET/MR imaging of lung cancer.” MA thesis. UiT The Arctic University of Norway, 2019.
- [188] Georg Schramm and Claes Nøhr Ladefoged. “Metal artifact correction strategies in MRI-based attenuation correction in PET/MRI.” In: *BJR—Open* 1.1 (2019), p. 20190033.
- [189] Hossein Arabi and Habib Zaidi. “Three-dimensional shape completion using deep convolutional neural networks: Application to truncation compensation and metal artifact reduction in PET/MRI attenuation correction.” In: *2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC)*. IEEE, 2019, pp. 1–3.
- [190] Hossein Arabi and Habib Zaidi. “Truncation compensation and metallic dental implant artefact reduction in PET/MRI attenuation correction using deep learning-based object completion.” In: *Phys. Med. Biol.* 65.19 (2020), p. 195002.
- [191] Karim Armanious et al. “Independent attenuation correction of whole body [18F]FDG-PET using a deep learning approach with Generative Adversarial Networks.” In: *EJNMMI Res.* 10.1 (2020), pp. 1–9.
- [192] Matthias K. Werner et al. “Respiratory gating enhances imaging of pulmonary nodules and measurement of tracer uptake in FDG PET/CT.” In: *Am. J. Roentgenol.* 193.6 (2009), pp. 1640–1645.

Bibliography

- [193] Chi Liu et al. “The impact of respiratory motion on tumor quantification and delineation in static PET/CT imaging.” In: *Phys. Med. Biol.* 54.24 (2009), pp. 7345–7362.
- [194] Ivayla Apostolova et al. “Combined correction of recovery effect and motion blur for SUV quantification of solitary pulmonary nodules in FDG PET/CT.” In: *Eur. Radiol.* 20.8 (2010), pp. 1868–1877.
- [195] Wouter Van Elmpt et al. “Optimal gating compared to 3D and 4D PET reconstruction for characterization of lung tumours.” In: *Eur. J. Nucl. Med. Mol. Imaging* 38.5 (2011), pp. 843–855.
- [196] Robert Grimm et al. “Self-gated MRI motion modeling for respiratory motion compensation in integrated PET/MRI.” In: *Med. Image Anal.* 19.1 (2015), pp. 110–120.
- [197] Avrim L. Blum and Pat Langley. “Selection of relevant features and examples in machine learning.” In: *Artif. Intell.* 97.1-2 (1997), pp. 245–271.
- [198] Egesta Lopci et al. “Imaging with non-FDG PET tracers: outlook for current clinical applications.” In: *Insights Imaging* 1.5-6 (2010), pp. 373–385.
- [199] Francesco Giammarile et al. “Non-FDG PET/CT in Diagnostic Oncology: a pictorial review.” In: *Eur. J. Hybrid Imaging* 3.1 (2019), p. 20.
- [200] F. Lamare et al. “List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations.” In: *Phys. Med. Biol.* 52.17 (2007), pp. 5187–5204.
- [201] Tiantian Li et al. “Motion correction of respiratory-gated PET images using deep learning based image registration framework.” In: *Phys. Med. Biol.* 65.15 (2020), p. 155003.
- [202] Yann Lecun, Yoshua Bengio, and Geoffrey Hinton. “Deep learning.” In: *Nature* 521.7553 (2015), pp. 436–444.
- [203] Pantelis Linardatos, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. “Explainable AI: A Review of Machine Learning Interpretability Methods.” In: *Entropy* 23.1 (2020), pp. 1–45.
- [204] Diogo V. Carvalho, Eduardo M. Pereira, and Jaime S. Cardoso. “Machine learning interpretability: A survey on methods and metrics.” In: *Electron.* 8.8 (2019), pp. 1–34.
- [205] Dario Amodei et al. *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Ed. by Wojciech Samek et al. Vol. 11700. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2019.

