Universidad de la República — Facultad de Ciencias

Instituto de Física

Diagnostic and Prognostic Quantitative Factors in Positron Emission Tomography and Computed Tomography (PET/CT) Images

Doctoral Thesis

presented to Facultad de Ciencias of the Universidad de la República, Montevideo, Uruguay

by

Enrique Cuña

Directors :

Gabriel González

Omar Alonso

Robert Jeraj

2018

Dedicado a

mis padres

mi hermana

y a la memoria de mis abuelitas

Resumen

En esta tesis se abordó una problemática de actualidad vinculada con aspectos cuantitativos de las imágenes médicas, en especial las imágenes funcionales. Estas técnicas constituyen una herramienta poderosa para el diagnóstico y tratamiento médico. El potencial de estas imágenes va más allá de lo que se puede determinar a simple vista. Dicha información, accesible mediante la cuantificación de imágenes, es de mucha utilidad para las predicciones a futuro de la evolución de los pacientes y/o de la efectividad de los tratamientos administrados.

En el presente trabajo se hizo uso de herramientas matemáticas y analíticas para los fines mencionados anteriormente. Los temas abordados se dividen en tres: estudio cuantitativo de imágenes de medicina nuclear de ventilación pulmonar, estudios cuantitativos de imágenes de tomografía por emisión de positrones (PET) en pacientes con cáncer de cabeza y cuello, y estudios cuantitativos de PET de pacientes pediátricos con linfoma.

En el primer estudio hemos realizado una comparación de distintos parámetros cuantitativos de características texturales extraídos de tomografías por emisión de fotón único (SPECT) y PET de pacientes con enfermedad pulmonar obstructiva crónica. También se hizo un análisis de correlaciones entre los valores de estas características y los valores

clínicos relevantes para el diagnóstico de esta enfermedad. En particular, se utilizó el valor del volumen de expiración forzada en un segundo, parámetro que se usa corrientemente para determinar la severidad de la enfermedad pulmonar obstructiva crónica. Se encontraron correlaciones significativas para 18 características texturales de estas tomografías y el volumen de expiración forzado.

En el segundo estudio analizamos marcadores imagenológicos extraídos en base a características texturales cuantificables de las imágenes PET de pacientes con cáncer de cabeza y cuello de células escamosas. Se buscó, de esta manera, una correlación entre estos marcadores y el período de progreso libre de enfermedad. Se lograron identificar 22 posibles marcadores de tiempos de evolución libre de enfermedad para pacientes con esta patología.

Finalmente, en el caso del estudio de pacientes pediátricos con linfoma de Hodgkin, se hizo un estudio cuantitativo también basado en características texturales de las imágenes PET. En este caso, se buscaron marcadores tempranos de respuesta al tratamiento en un intento para predecir anticipadamente la evolución de la enfermedad y permitir así personalizar el tratamiento para cada paciente en específico. Se construyeron modelos de clasificación para estos pacientes usando un número de 8 componentes principales de 51 características texturales. Estos modelos permitieron asignar a los pacientes un grupo de riesgo usando el estudio PET al momento del diagnóstico de la enfermedad. Esta clasificación es crucial ya que dicta el tratamiento a seguir en cada caso.

Abstract

In this thesis we approached a current thematic related to quantitative aspects of medical imaging, especially for functional images. These techniques constitute a powerful tool for medical diagnose and treatment. Potential information contained in these images goes beyond what can be seen with the naked eye. Such information, accessible through image quantification, is very useful to predict patient evolution and/or the treatment effectiveness.

In the present work, we used mathematical and analytical tools in order to achieve the goals mentioned above. Working projects were divided in three: quantitative study of nuclear medicine images of lung ventilation, quantitative studies of positron emission tomography (PET) images from patients with head and neck cancer, and a study of PET images from paediatric patients with lymphoma.

In the first study we did a comparison between quantitative textural features extracted from single photon emission tomography (SPECT) and PET images from patients with chronic obstructive pulmonary disease. An analysis of the correlation of these parameters with the physiologically relevant variable for this disease is also presented. In particular, we used the value of the forced expiratory volume in one second, as a currently used pa-

rameter to evaluate the severity of this disease. We found significant correlations between 18 textural features from these images and the forced volume of expiration.

In the second study we analysed imaging markers extracted from quantifiable textural features from PET images of head and neck squamous cell carcinoma patients. We looked for a correlation between these markers and the disease progression-free survival period. We identified 22 possible markers of progression-free survival for patients carrying this pathology.

Finally, for the study conducted with paediatric patients with Hodgkin lymphoma, we performed an analysis of quantitative parameters. These parameters were also measured from textural features from PET images. In this case we looked for early treatment response markers, in an attempt to predict disease evolution and therefore allow for patient specific treatment tailoring. We constructed classification models for patients using 8 principal components from 51 different textural features. These models allowed for categorizing patients in risk groups using PET studies at the moment of disease diagnose. This classification is key to select the treatment to be administered for each case.

Contents

Acknowledgements

This work and practically every endeavour in my life is motivated by people around me, therefore I would like to express my gratitude to:

Gabriel González, for mentoring this thesis, all the support, useful comments and discussions.

Omar Alonso, for mentoring, for all the support, advices and help at work, from another perspective.

Robert Jeraj, for mentoring, for letting me be part of his group, granting me access to valuable data and advices.

The Image Guided Therapy group for their useful comments and all the help in several aspects related to this thesis.

Carolina Rabin, for taking time at the office to undertake discussions, and advices.

Mónica Rodríguez, for her patience, positivism, friendship and all necessary inputs for this work.

Verónica Martín for her friendship and encouragement throughout this process.

Stephanie Talento, for her friendship and advices.

Alethia de la Fuente, for her friendship and company undertaking some academic chal-

lenges.

Roberto Passera, for his support and all the useful statistics advices.

The CUDIM staff for all the support, and teachings.

My room-mates for their patience and company.

The Physics Institute staff, for all the support.

CAP, for supporting part of these years of work.

My parents Celeste and Washington, for accompanying me through all this process, their advice, love and caring.

My sister Carolina, for her friendship, love and caring.

My girlfriend Maia, for the encouraging and inspiration.

The rest of my family, for being there for me all the time.

All my friends that supported me from the distance: Andreas, Erwin, Alexis, Andrés, Fabián, Marcos.

The program PEDECIBA for taking me in as a PhD student.

The scientific research commission, CSIC, for the economical support and its project number C681.

Last but not least, I would like to thank the big guy for all the strength, motivation and so much more.

Publications and presentations

Research articles

- Cuña E.G.; Gambini J.P.; Servente L.; Savio E.; EnglerH.W.; González G.A.; Alonso O. Quantitative comparison between single positron emission computed tomography and positron emission tomography imaging of lung ventilation with ^{99m}Tc-technegas and ^{68}Ga-gallgas in patients with chronic obstructive pulmonary disease: A pilot study. Accepted for publication at the World Journal of Nuclear Medicine **2018**. http://www.wjnm.org/aheadofprint.asp, number 36.

- Scarpelli M.; Eickhoff J.; Cuña E.; Perlman S.; Jeraj R. Optimal transformations leading to normal distributions of positron emission tomography standardized uptake values. *Physics in Medicine and Biology* **2018**, 63, 1-8.

- Jeraj R.; Harris J.; Rosenthal D. et al. HNSCC Response Assessment Biomarkers: Unplanned Analysis of the NRG Oncology RTOG 0522 PET Sub-Study. To be sent to the *International Journal of Radiation Oncology Biology and Physics*.

- Cuña E.; Rodriguez Taroco M.; Castillo L.; González G.; Engler H.; Alonso O. Evaluation of quantitative imaging markers from ^{18}F-FDG PET/CT for treatment

response evaluation in paediatric patients with Hodgkin Lymphoma. To be sent to the *International Journal of Radiation Oncology Biology and Physics*.

Presentations at meetings

- **Cuña E.**; Rodriguez Taroco M.; Engler H.; Alonso O. Evaluation of quantitative imaging markers from ^{18}F-FDG PET/CT for treatment response evaluation in paediatric patients with Hodgkin Lymphoma. *The journal of Nuclear Medicine* **2018**, 59(1), 23. Presented at the Society of Nuclear Medicine and Molecular Imaging 2018 Annual meeting in Philadelphia, Pennsylvania, USA.

- Jeraj R.; Harris J.; Rosenthal D. et al. Head and Neck Squamous Cell Cancer Response Assessment Biomarkers: Unplanned analysis of the NRG Oncology RTOG 0522 PET sub-study. *International Journal of Radiation Oncology Biology and Physics* **2017**, 99(2), 15. Presented by **Enrique Cuña** at the American Society of Radiation Oncology 59th Annual meeting in San Diego, California, USA.

- **Cuña E.**; Rodriguez Taroco M.; Engler H.; Alonso O. Prognostic value of ^{18}F-FDG PET/CT metabolic quantitative parameters in paediatric Hodgkin Lymphoma. *The Journal of Nuclear Medicine* **2017**, 58(1), 637. Presented at the Society of Nuclear Medicine and Molecular Imaging 2017 Annual meeting in Denver, Colorado, USA.

- **Cuña E.** Comparación entre el SPECT y el PET/CT en la evaluación de volúmenes de ventilación pulmonar en pacientes con EPOC. Presented at the Latin American Association of Societies of Biology and Nuclear Medicine XXV meeting in Punta del Este, Uruguay **2015**.

Acronym list

Acronym	Stands for
AIC	Akaike information coefficient
AUC	Area under the curve
COPD	Chronic obstructive pulmonary disease
CT	Computed tomography
CUDIM	The Uruguayan Center for Molecular Imaging
DFS	Disease-free survival
DIRART	Deformable Image Registration for Adaptive Radiotherapy
DSC	Dice similarity coefficient
EANM	European Association of Nuclear Medicine
FBP	Filtered back projection
FEV1	Forced expiratory volume in 1 second
GLCM	Gray-level co-occurrence matrix
GLRLM	Gray-level run-length matrix
HL	Hodgkin lymphoma

HNSCC	Head and neck squamous cell carcinoma
HR	Hazard ratio
ICC	Intraclass correlation coefficient
IGT	Image Guided Therapy
iPET	^{18}F-FDG PET/CT at end of two cycles of chemotherapy or mid-treatment
IPS	International prognostic score
IVH	Intensity-volume histogram
LG	Lesion glycolysis
LN	Left metastatic nodes
MN	Maximum texture features value from left and right metastatic nodes
MTV	Metabolic tumour volume
MVF	Moving vector field
NGLDM	Neighbouring gray-level dependence matrix
NGTDM	Neighbouring gray-tone dependence matrix
OSEM	Ordered subset expectation maximization
PCA	Principal component analysis
PET	Positron emission tomography
PET/CT	Combination of PET and CT
PFS	Progression-free survival
PPV	Positive predictive value

PT	Primary tumour
QT	Chemotherapy
RN	Right metastatic nodes
RT	Radiotherapy
ROC	Receiver operating curves
ROI	Region of interest
RTOG	Radiation Therapy Oncology Group
SN	Sum of values for texture features from left and right metastatic nodes
SPECT	Single photon emission computed tomography
SUV	Standardized uptake value
TLG	Total lesion glycolysis
TPR	True positive rate
TMTV	Total metabolic tumour volume
TOF	Time of flight
V/Q	Ventilation/Perfusion

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Image quantification in medical images

Imaging in medicine has a wide variety of modalities and applications. They contain a variety of information, ranging from anatomical structures, flow of fluids and function monitoring to name a few. Images constitute a great tool for diagnose by themselves, however, they could also contain useful information for other purposes, information that could be extracted through quantification, for example. In this work we take into account nuclear medicine images, in particular, positron emission tomography (PET) images.

PET is an imaging technique that enables the visualization of functional processes inside the human body. To obtain a PET image, a radiopharmaceutical is injected to the patient, the most common one is 2-deoxy-2-(^{18}F)fluoro-D-glucose, abbreviated ^{18}F-FDG. This radiopharmaceutical is a glucose analogue marked with the β^+-emitter ^{18}F. When

^{18}F-FDG is injected, the molecules are absorbed in the metabolically active tissues in the body. After distribution of the rediopharmaceutical in the body, the emission of a positron from the ^{18}F isotope leads to a nearby anihilation of this particle with a surrounding electron from the environment. This anihilation produces two $512keV$ gamma rays in opposite directions, which are detected in the array of detectors arranged in a ring surrounding the patient's body. Detections of photons along a line connecting a couple of detectors and the time of arrival of these photons allow to predict the site of anihilation. Thus, this process allows for the reconstruction of a tomographic image for visualization (of metabolism in the case of FDG). In particular, cancer cells have high avidity for ^{18}F-FDG [1]. Therefore, ^{18}F-FDG PET allows for tumour and disease characterization as well as treatment efficacy evaluation and quantification [2–4]. Functional imaging is increasingly being taken into account for oncology treatments such as radiation therapy, for which only anatomical images such as computed tomography (CT) were previously used. PET is often complemented by anatomical imaging such as CT, used for anatomical reference, and a fused image is obtained. An example can be seen in figure 1.1.

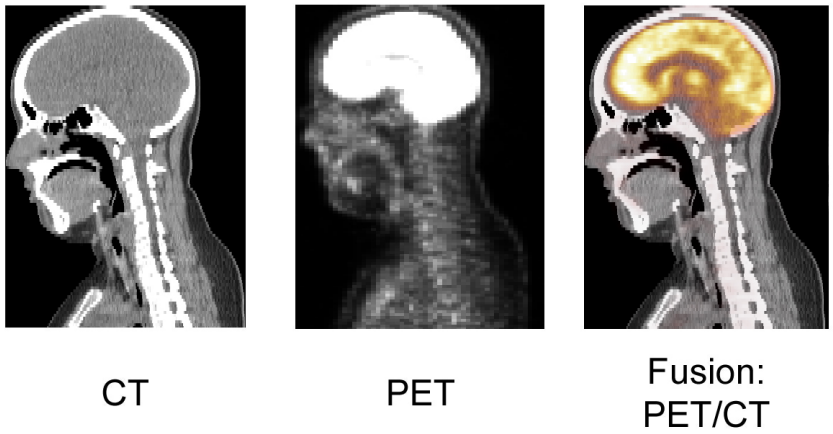

Figure 1.1: Example of a ^{18}F-PET/CT fusion image.

However, ^{18}F is not the only radiopharmaceutical that can be used for PET. A variety of molecules marked with other radioisotopes are used to image different processes as well. Some examples are: ^{15}O in H_2O for oxygen metabolism, ^{13}N in NH_3 for myocardial perfusion, ^{11}C in ^{11}C-acetate for myocardial uptake, oxidative metabolism and studying hepatocellular and prostate carcinomas [5]. All these different molecules are driven to target tissues by the biochemical processed in the body.

As these images bear information about the physiologic processes in the body, quantification of the uptake itself and some other properties of this images are believed to contain much more information about the processes involved in disease manifestation and changes in a patient. In this work we will specially treat PET images with ^{18}F-FDG.

However, to tackle the problem of finding relevant quantitative information in these images, several complementary but important aspects have to be taken into account. These aspects concern image registration, volumes segmentation and quantification itself. The big picture of this process gives birth to a field known nowadays as "radiomics".

1.1.1 Deformable registration

Usually, when using multi-modality imaging, the process of matching two different image acquisitions for the same patient is called "registration". For this purpose, images are acquired consecutively with, ideally, no movement of the patient, in which case, rotations and translations are enough to transform one image to match the other one. However, given that the human body is not rigid, sometimes there is movement and deformation of

the tissues. In this case, registration has to deform one image to match the other one.

1.1.2 Image quantification and texture features

As mentioned before, image quantification is the process of extracting useful measurable magnitudes from medical images, with the objective of extending and/or complementing the visual evaluation done by physicians. In this work this is done by using the so called "textural features". These features could be related to pixel values from an image, their proximity and their spatial distribution. There is a large number of textural features described in the literature [6–12]. These features have generally been devised as a mean to quantify a pattern or a property of an image as seen by the naked eye. In practice, not always the defined features can be related to a visual characteristic of an image.

1.1.3 Radiomics

The field that has been known for a while now as "radiomics" consists in extracting large amounts of data from medical images, in order to be used to make decisions on treatments, or predict outcomes. Basically, all types of quantitative information that can be extracted from images, such as textural features, shape-based features, histogram and intensity level features, etc., are prone to be taken into account in this definition. The founding idea of radiomics is that this features are related to physiologic processess in the body. Threfore, this field is based on the fact that medical images have more much information than what meets the eye. In this thesis, we took into account textural features as described in the following chapters. However, the process of feature extraction in the

case of medical imaging is more complex and involves several consecutive steps: imaging data acquisition, image segmentation, feature extraction and finally data analysis. Each one of these steps has its own inherent challenges, like standarization, reproducibility, ground truths definition, automation and data storage, to name a few. A nice summary of radiomics can be found in Kumar et al. [13]. A conceptual map summarizing the process and challenges is shown in figure 1.2. Some of this challenges are addressed in the chapters of the present work.

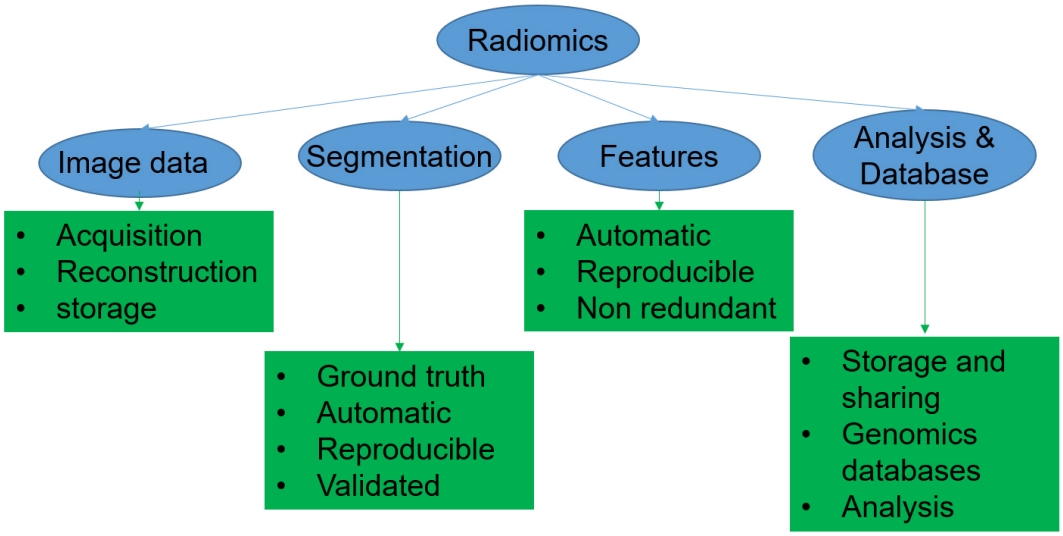

Figure 1.2: Conceptual map summarizing radiomics workflow and challenges.

1.2 Treatment response assessment and early prediction

The perfect scenario for treatment planning would be to know beforehand if the patient is going to respond to a certain treatment. In order to do this, we search for useful indicators

that can be extracted from medical images. Once validated, these indicators would allow for knowing which treatment or drug to administer and eventually, tailor patient specific treatments. This is known as treatment personalization.

To achieve this goal, the methodology consists of evaluating changes in key markers throughout treatment courses, and to correlate this changes to groups of different responses to treatments. Eventually, the best markers would allow for prediction as early as at the moment of disease diagnose. All these goals are very general and in this work we focus specifically on markers extracted from PET images using ^{18}F-FDG as the radiopharmaceutical.

1.3 Thesis layout

As an overview of the structure of this work, we have the following. Chapters 2 and 3 explain the mathematical analytical tools that were used throughout the three problems that were studied in this thesis. Chapter 2 gives an overview of deformable registration, definitions as well as the algorithms used and a brief explanation of how they work. Chapter 3 lists mathematical definitions and methodology for textural features calculation.

All sub-projects presented in this thesis were done using patients data, provided by the Uruguayan Center of Molecular Imaging (CUDIM) and in collaboration with the Image Guide Therapy (IGT) group from the Medical Physics department of the University of Wisconsin. These 3 main frameworks are presented as follows.

Chapter 4 presents a comparison of lung ventilation imaging between PET and single photon emission computed tomography (SPECT). Textural features were used to compare quantitatively both techniques. This constitutes a pilot study as not many patients were involved in the cohort from CUDIM. The interest of this chapter is to present a quantitative analysis that is not much documented between these imaging techniques. We found that quantification of these techniques is in fact comparable, and that a proper standard protocols used for imaging could lead to robust and information-rich texture features quantification.

Chapter 5 presents a study done on head and neck squamous cell carcinoma patients. We had access to this patient cohort from the Radiation Therapy Oncology Group (RTOG) thanks to the collaboration with the University of Wisconsin. In this project, deformable registration was used in order to quantify textural features in scans at disease diagnose (baseline) and after treatment. Fifty textural features were calculated from baseline and post-treatment scans, from which 900 different imaging markers were considered to be predictors of progression-free survival (PFS). Twenty-two of these markers were found to be prognostic when compared to tumour-size prognostics alone.

Finally, in chapter 6 we presents the results of a project carried out with paediatric lymphoma patients from a cohort at CUDIM. Twenty one patients with ages between 6 and 17 diagnosed with Hodgkin lymphoma (HL) were enrolled in the study. Fifty one textural features were calculated from baseline ^{18}F-FDG PET/CT scans. These textural

features were used to train classifier models to separate patients into different groups corresponding to standard clinically used scales to assess treatment response and risk groups. High accuracy classification models were obtained using this method. If further verified and standardized, these findings could have a great clinical impact regarding patient treatment management.

Chapter 2

Deformable Registration

Registration is the process of taking one system of reference, image, structure, to the same space as a similar structure. In the case of medical tomographic images, when two different scans of the same patient are taken at different moments, it is inevitable to have some movement in between. Therefore, in order to account for this movements when comparing the images, registration must be performed.

Registrations belong to two different families: rigid and deformable. The former takes into account basically rotations, translations and isotropic transformations. The latter takes into account deformations of bodies, as is the case in medical images, the body is composed of tissues that not necessarily move rigidly. This is when deformable registration comes into play.

2.1 Demons algorithm

Demons is the name given to a deformable registration algorithm inspired by Maxwell's demon. Maxwell demon is an imaginary entity at the interface between two gases that separates different molecules, allowing some in and taking some out, that would make entropy to decrease. This idea inspired a model in which a moving image, M, is being matched to a fixed image S. The idea is to consider the boundaries of image S as a permeable membrane, where a Maxwell demon acts on every point of the membrane, whereas M is a deformable fluid whose particles are either "attracted" inside the membrane or "repelled" from it. To decide if these moving particles are taken inside or outside the membrane, a definition of similarity between a point P of M and a point P' of S has to be defined, this similarity is noted $K(P, P')$. Knowing the similarity between P and P', and the distance $D(P, P')$, a force $\vec{f}$ on a point P of M can be defined as:

$$\vec{f}(P) = \sum_{P' \in S} \frac{K(P, P')}{D(P, P')} \overrightarrow{PP'}$$

The ensemble of all the forces $\vec{f}$ acting on M defines the motion vector field, and the deformation depends on how rigid M is considered to be. This procedure is described by Thirion [14].

2.2 Optical Flow algorithm

Optical flow method is based on the idea that images are moving frames, therefore each pixel in he image will be at a certain point (x, y) at a certain time t. The deformed image

is then understood as the original image after a certain set of displacement vectors is applied to it. To state the mathematical relationship between the fixed image $I(x, y, t)$ and the moving image $I(x + \Delta x, y + \Delta y, t + \Delta t)$, the condition of constant brightness is imposed:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \tag{2.1}$$

then, a Taylor development to the first order is calculated, and the condition stated in equation 2.1 leads to the *aperture problem* for optical flow:

$$\nabla I^T \cdot \vec{V} = -I_t$$

where $\nabla I = (\partial I/\partial x, \partial I/\partial y)$, the vector $\vec{V}$ is the velocity vector $\vec{V} = (\Delta x/\Delta t, \Delta y/\Delta t)$, and $I_t = \partial I/\partial t$. One frequently used method to find the solutions to this equation is the Horn-Schunck method described by Horn et al. [15].

2.3 Similarity Measurements

Similarity indexes are a way to measure quantitatively the quality of deformable registration. Dice similarity coefficient (DSC) was developed by botanist Lee R. Dice to measure the similarity between two samples [16]. We used this coefficient to judge upon the results of our deformable registrations. To illustrate what DSC is, lets take two ensembles: A and B. The degree of similarity between these two ensembles measured with DSC is defined as:

$$DSC = \frac{2 \mid A \cap B \mid}{\mid A \mid + \mid B \mid}$$

In order to measure DSC from two images, equal gray level values at the same pixels are taken as the elements in the intersection $A \cap B$. Gray levels of an image can be interpolated to a convenient number of intensity levels (256 for example) before calculating DSC.

Lets take ginger-men as an example for fixed and moving images. The moving image is the one that is going to be deformed to match the original image. We used both demons and optical flow methods for testing. We used DIRART (Deformable Image Registration for Adaptive Radio Therapy) toolbox [17, 18] running under MatLab for this purpose. Deformable registrations performed using DIRART do not have a reference metric for stopping iteration, but one can choose a fixed number of iterations and keep the most suitable one, in this case we used up to 5 iterations. Results for deformations using Demons and optical-flow methods can be seen in figures 2.1 and 2.2, respectively.

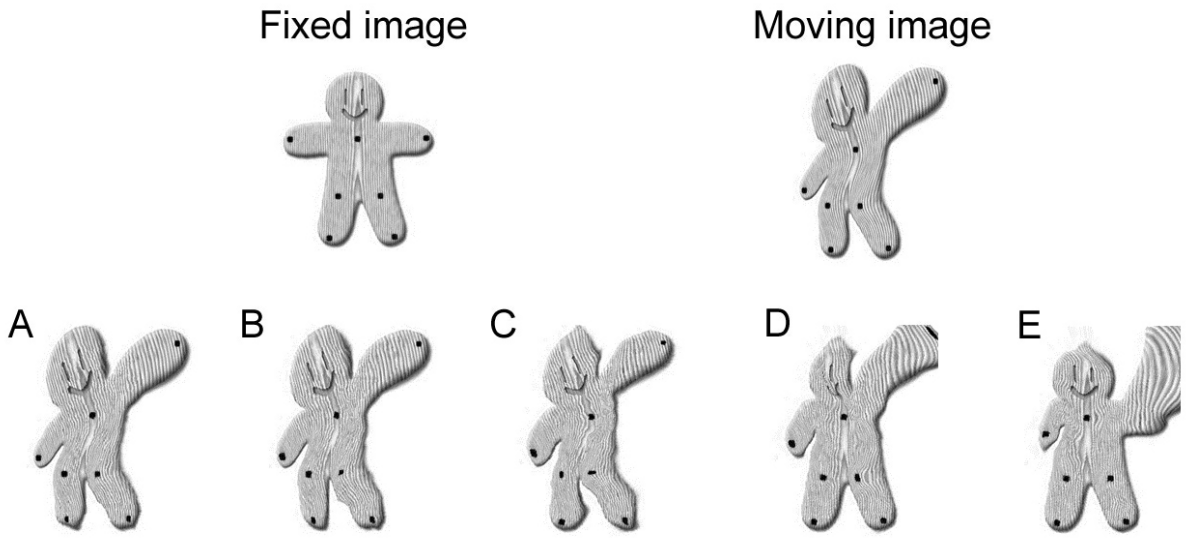

Figure 2.1: Deformable registrations calculated using demons algorithm. Top images correspond to fixed and moving frames, bottom row figures from A) to E) correspond to one to five iterations, respectively. Dice similarity coefficients between deformed image and the original one are the following: A) 0.0059, B) 0.0067, C) 0.0104, D) 0.0126, E) 0.0187; similarity coefficients have been calculated for 256 gray levels.

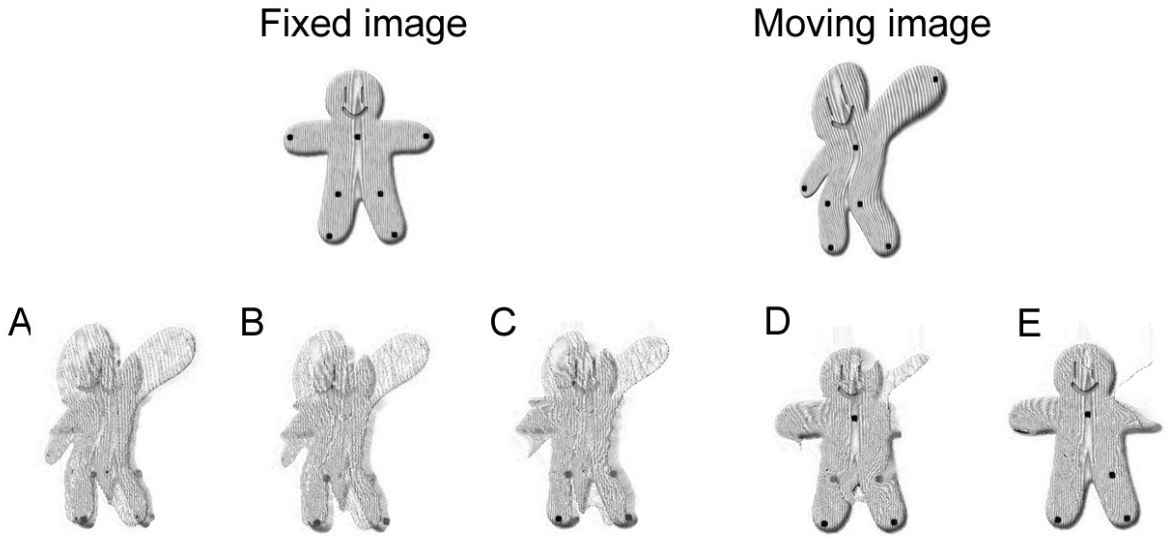

Figure 2.2: Deformable registrations calculated using optical-flow algorithm. Top images correspond to fixed and moving frames, bottom row figures from A) to E) correspond to one to five iterations, respectively. Dice similarity coefficients between deformed image and the original one are the following: A) 0.0079, B) 0.0088, C) 0.0097, D) 0.0155, E) 0.0226; similarity coefficients have been calculated for 256 gray levels.

As can be seen in figures 2.1 and 2.2, both algorithms have strengths and weaknesses depending on the differences between images. Another example of deformable registration results using these both algorithms applied to computed tomography (CT) images can be seen in figure 2.3. In the mentioned example, similarity can be visualized by means of contour lines. These lines run along constant gray-level values borders in an image. In figure 2.3, contour lines corresponding to the deformed image are superposed to the fixed image.

As can be seen on figures 2.1 through 2.3, the measured DSC are small, this is due to the fact that we are taking 256 gray-level values for comparison, which sets a highly demanding threshold for similarity measurements. Therefore, DSC values shown in these figures should be taken from lowest to highest for the sake of comparison only, as we can

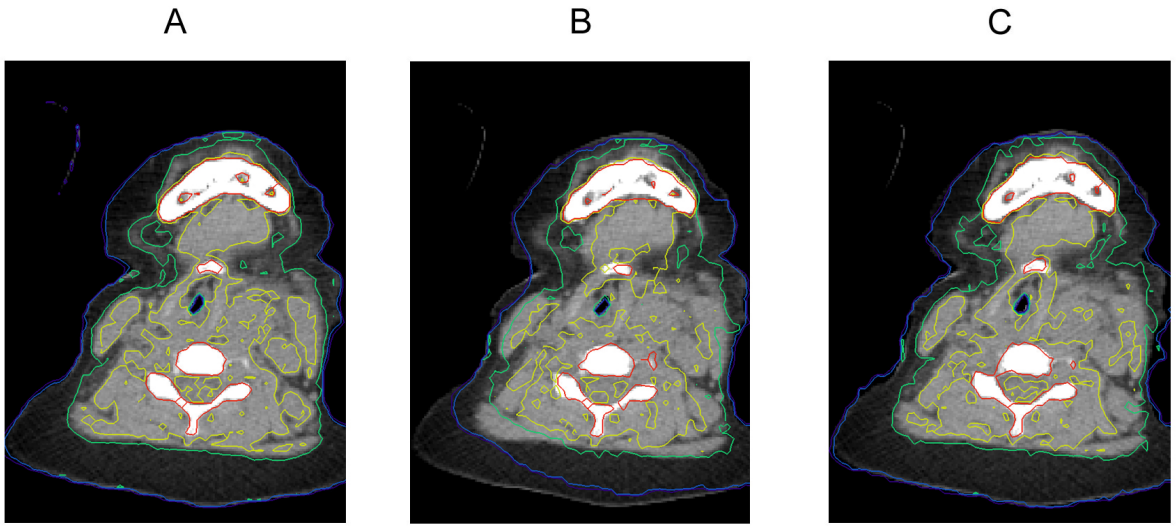

Figure 2.3: Examples of optical flow and demons deformable registration algorithms applied to CT images. A): Fixed image with its own contour lines to illustrate the ideal result of deformable registration; B): Optical flow algorithm result, Dice index 0.0259; C) Optical flow algorithm result, Dice index 0.0308. Dice indexes have been calculated with 256 gray levels.

visually see that deformations are doing well in general.

2.3.1 Examples of application of deformable registration methods

Deformable registration has been used in this work for quantification of response to treatment from medical PET/CT imaging. Lets suppose we have two different ^{18}F-FDG PET/CT scans for one patient, one scan is performed at the moment of disease diagnose and the other one after treatment. There are two different approaches that have been used for deformable registration in this work for response assessment, before applying methods for image quantification.

2.4 Method 1: Deformation of tomographic images

Once we segment the lesions using PET/CT scans, we first perform a local rigid registration around the lesion being considered. After this first registration process corresponding regions should be closer together, then we apply deformable registration. For example, we could take the post-treatment computed tomography as the moving image and the baseline tomography as the fixed image. We use CT images for this purpose since they have much more anatomical detail than PET images. When this process is completed, we obtain a moving vector field (MVF). Finally, we apply the same rigid transformation and MVF to PET images to obtain the corresponding voxel values for the same region before and after treatment.

Using this method, we use the same regions of interest (ROI) as defined in the first scan. Figure 2.4 shows an example of this method applied to a patient with head and neck cancer. This methodology was used for response assessment quantification as described in chapter 5.

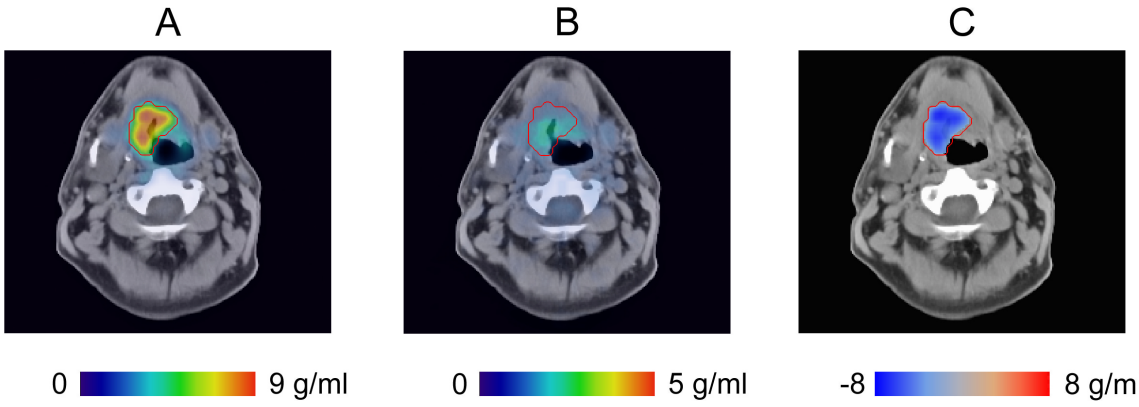

Figure 2.4: Deformable registration applied to PET images. A) Baseline PET/CT with ROI, B) Deformed post-treatment PET/CT with ROI, C) SUV differences inside ROI.

One of the problems with this method is that registration is not always accurate, which is the case with most deformable registration methods. It is also important to notice that when we apply the MVF and interpolate voxel values, we are in fact changing the values of the tomographic image. This is a problem, for example, in the case of PET when comparing standardized uptake values (SUV), where an interpolation could change this quantification. An example of inaccuracies of this method can be seen in figure 2.5.

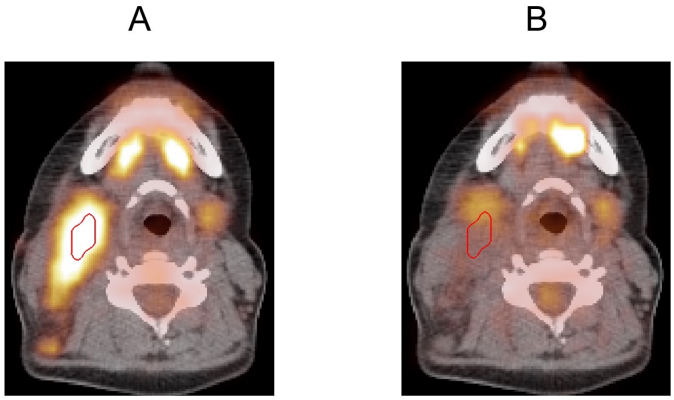

Figure 2.5: Deformable registration applied to PET images. A) baseline PET/CT with ROI as red contour, B) deformed post-treatment PET fused with baseline CT and ROI. Hot glowing zones represent high radiopharmaceutical uptake regions.

An alternative method to deal with the inaccuracies mentioned above is described in the next section.

2.5 Method 2: Deformation of regions of interest

To address some of the cons of the method described in the preceding section, another approach could be considered. Local rigid registration is still used, with the difference that in this case we take post-treatment CT as the fixed image, and we use baseline CT

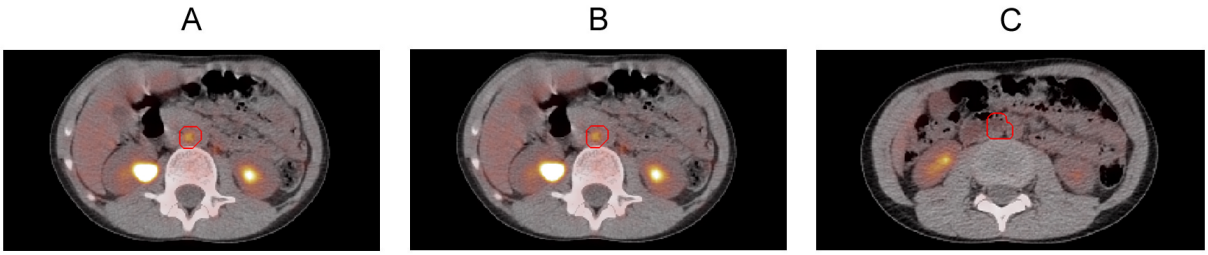

Figure 2.6: Deformable registration (with optical flow algorithm) used to deform a ROI. This ROI is superposed to fused PET/CT images of a patient with Hodgkin lymphoma. A): contoured lesion at baseline scan; B) deformed ROI at mid-treatment scan; C) deformed ROI at end of treatment scan.

scan as the moving image in order to get the MVF. Once the MVF is calculated, we apply both the rigid registration and the MVF deformation to the ROI as defined in the baseline scan. Figure 2.6 shows an example of the application of this method to a patient with Hodgkin lymphoma. The contoured lymph node that can be seen in the baseline scan is followed using this method through mid-treatment and end-treatment scans.

As can be seen in figure 2.6, this method yields visually accurate results. There are two major pros compared to the method 1 in the previous section: when interpolating the deformed ROI, a simple nearest-neighbours interpolation is enough, since a ROI is a binary mask; if the ROI is not accurate enough, it can be modified easily with appropriate software.

Chapter 3

Imaging Texture Features

Information extracted from visual inspection of medical images is of most relevance when used for diagnose purposes. It goes without saying that these images will continue to be used in various medical fields as they are most of the time, non-invasive, and results can be obtained almost immediately. Current advances have made it possible to see anatomical and also functional features with medical imaging. Moreover, medical images contain more information to what can be seen with the naked eye. One form to extract this "hidden" information is through textural features analysis. These features allow us to extract large amounts of information from a single image and use it to our advantage for diagnosis and even prediction of disease/treatment outcomes. This is an emerging field that has been called "Radiomics". There are some reported examples of the usefulness of textural features to predict treatment outcomes, for example, in esophageal cancer [19].

3.1 What are texture features?

Texture features are an attempt to quantify observed patterns, shapes, trends, etc. in images as seen by the naked eye. In this chapter we describe the mathematical definition of the features we use. Not all of these features are easily associated to a visual characteristic of the images they are calculated from.

The motivation for applying texture features analysis to medical images is that different radiotracer uptake patterns could help to recognize distinct pathologies in a tumor. In this regard, possible applications of textural analysis are segmentation of anatomical structures, diagnosis and tissue differentiation [20].

Texture features are image-derived quantitative parameters calculated mathematically from voxel values distributions. Among these we have first, second and higher order features. First order features are taken as statistical moments of the histograms of the voxel values distributions of a given group of voxels. First order features do not depend on the spatial distribution of voxel values, they only depend on the histogram of the image. Second order features are mathematically calculated from the gray-level co-occurrence matrix, which is calculated for any given arrangement of gray-level voxels. High order features are calculated using the gray-level run-length matrix, the neighbouring gray-level dependence matrix and the neighbouring gray-tone dependence matrix. Details for the computation of these matrices will be given in the sections below. Second and higher order features do depend on the spatial distribution of the gray level values of the voxels inside the image. The basic hypothesis to use these features calculated from medical images is that histograms or spatial distribution of gray level values have information about

physiological characteristics captured in a patient's tomographic scans.

Texture features are being increasingly introduced to analyze and quantify medical images, hoping to provide additional information to visual analysis. Some comparisons between visual and textural aspects of imaging have already been reported [21].

In the sections below the mathematical expression used to calculate each one of the before mentioned features is presented.

3.2 First Order Features

First order features are calculated from the histograms of gray values of the images as described in [6]. In our case, we are talking about 3D-images, therefore we will refer to an image as a 3D matrix whose values refer to the gray-level values of the image, and instead of talking about pixels we will refer to each element in the 3D matrix as a voxel. Since first order features do not take into account the spatial distribution of the gray-tones but only the number of voxels in a given range of grey-values, completely different images with the same histograms could have the same first-order features values. Table 3.1 lists definitions of these features in terms of voxel intensities. In this table, n is the length of one side of the image, each voxel is located at coordinates (i, j, k) and has a grey-level $p(i, j, k)$.

Feature	Equation
Mean	$f_1 = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{p(i,j,k)}{n^3} = \mu$
Median	f_2: value separating data in two halves
Variance	$f_3 = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{(p(i,j,k)-\mu)^2}{n^3-1} = \sigma^2$
Coefficient of variation	$f_4 = \frac{\sqrt{\sigma^2}}{\mu}$
Skewness	$f_5 = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{(p(i,j,k)-\mu)^3}{n^3\times\sigma^3}$
Kurtosis	$f_6 = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\frac{(p(i,j,k)-\mu)^4}{n^3\times\sigma^4} - 3$
Energy	$f_7 = \sum_{i}^{L} H(i)^2$
Entropy	$f_8 = -\sum_{i}^{L} H(i) \times \log H(i)$

Table 3.1: **First-order features description.** n is the length of one side of the patch, $p(i,j,k)$ is the voxel intensity value at coordinates (i,j,k) inside the patch, $H(i)$ is the probability histogram value for intensity i and L is the maximum intensity value inside the patch.

3.3 Second order texture features

Second order texture features take into account spatial distributions of voxels in an image. To calculate these features we use run-length matrices, which are described in the following sections.

Before describing these matrices, we show in figure 3.1 what we will consider from now on the directional neighbouring pixels.

3.3.1 Gray-level co-occurence matrix: GLCM

For simplicity, we are going to explain how a GLCM is calculated over a 2D image. This methodology is easily extrapolated to 3 dimensions by adding one more direction in space. Consider a gray scale image as represented by a matrix whose values correspond to a gray

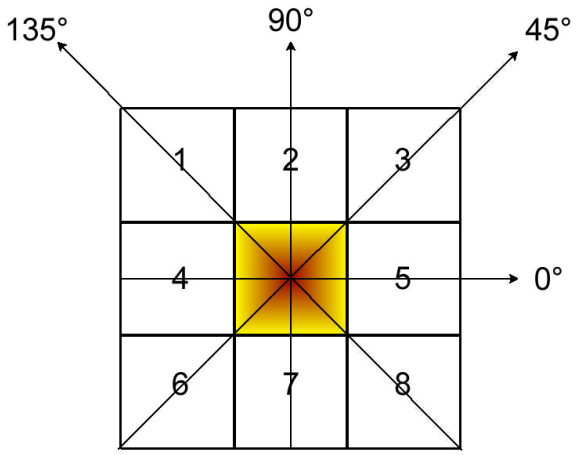

Figure 3.1: Eight neighbouring pixels around a center pixel. There are two specific neighbouring pixels for each given direction: 0, 45, 90 and 135 degrees.

tone. In this sense, let $L_x = \{1, 2, ..., N_x\}$ and $L_y = \{1, 2, ..., N_y\}$ be the horizontal and vertical domains for the image respectively; and $G = \{1, 2, ..., N_g\}$ be the possible gray values. Therefore, the image I can be represented as a function $I : L_x \times L_y \rightarrow G$ that assigns a gray value to every pixel in the image.

We then define four spatially dependent gray-level frequency matrices as follows, where # is the number of elements in the ensemble:

$$p(i, j, d, 0°) = \#\{((k, l), (m, n)) \in (L_x, L_y) \times (L_x, L_y)/|k - m| = d, l = n, I(k, l) = i, I(m, n) = j\}$$

$$p(i, j, d, 45°) = \#\{((k, l), (m, n)) \in (L_x, L_y) \times (L_x, L_y)/k - m = l - n = d \text{ or } k - m = l - n = -d), I(k, l) = i, I(m, n) = j\}$$

$$p(i, j, d, 90°) = \#\{((k, l), (m, n)) \in (L_x, L_y) \times (L_x, L_y)/k = m, |l - n| = d, I(k, l) = i, I(m, n) = j\}$$

$$p(i,j,d,135°) = \#\{((k,l),(m,n)) \in (L_x,L_y) \times (L_x,L_y)/(k-m=n-l=d) \text{ or}$$
$$(m-k=l-n=d), I(k,l)=i, I(m,n)=j\}$$

and then we define the GLCM as the summation of the spatially dependent gray-level frequency matrices:

$$P(i,j) = \sum_{\theta} p(i,j,d,\theta) \tag{3.1}$$

where $\theta = 0°$, $45°$, $90°$ and $135°$.

For example, lets suppose we have the image I defined as a 5×5 matrix with gray values between 0 and 3, and we set $d = 1$. In this particular case, given the matrix I has 4 different values (gray-levels), each one of the $p(i,j,1,\theta)$ will be 4×4 matrices, and the sum $P(i,j) = p(i,j,1,0°) + p(i,j,1,45°) + p(i,j,1,90°) + p(i,j,1,135°)$ is the GLCM for the nearest neighbouring pixels (pixels at distance 1). An example of an image I and its corresponding GLCM can be seen in table 3.2.

Using the GLCM we calculated several texture features as described in [7]. These features are detailed in tables 3.3 and 3.4.

3.4 High-order texture features

High-order texture features are so called because they are more complex than second-order features. The procedures for calculating these texture features are described in [8]. To calculate these high-order features we make use of the gray-level run-length matrix, the neighbouring gray-level dependence matrix and the neighbouring gray-tone dependence

Name	Matrix
Image	$I = \begin{pmatrix} 3 & 0 & 0 & 0 & 2 \\ 3 & 1 & 3 & 1 & 0 \\ 0 & 2 & 3 & 3 & 3 \\ 3 & 3 & 1 & 3 & 3 \\ 2 & 3 & 3 & 3 & 2 \end{pmatrix}$
GLCM	$P(i,j) = \begin{pmatrix} 3 & 1 & 3 & 1 \\ 5 & 0 & 2 & 5 \\ 0 & 1 & 0 & 8 \\ 9 & 10 & 3 & 21 \end{pmatrix}$
GLRLM	$Q(i,j) = \begin{pmatrix} 15 & 1 & 1 & 0 & 0 \\ 12 & 0 & 0 & 0 & 0 \\ 16 & 0 & 0 & 0 & 0 \\ 15 & 11 & 5 & 0 & 0 \end{pmatrix}$
NGLDM	$S(k,l) = \begin{pmatrix} 2 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 2 & 3 & 4 & 1 & 1 & 0 & 0 \end{pmatrix}$

Table 3.2: **Image and spatial dependence matrices examples.** All spatial correlation matrices are related to image I.

matrix.

3.4.1 Gray-level run-length matrix: GLRLM

The GLRLM was first proposed by Galloway in 1974 [11]. This matrix takes into account the number of times a particular gray-level is repeated in straight lines. Given a gray-level i and a run-length j, the matrix $q(i,j,\theta)$ has in each of its elements the number of times this i gray-level is repeated in j consecutive pixels at an angle specified by θ. As in the previous case the GLRLM is the sum of the $q(i,j,\theta)$ matrices for four different angles:

$$Q(i,j) = \sum_{\theta} q(i,j,\theta) \tag{3.2}$$

Feature	Equation
Angular moment	$f_9 = \frac{1}{T}\sum_i^{N_g}\sum_j^{N_g} P(i,j)^2$
Contrast	$f_{10} = \frac{1}{T}\sum_{i=0}^{N_g-1} i^2 P_{x-y}(i)$
Correlation	$f_{11} = \frac{1}{T\sigma_x\sigma_y}\sum_i^{N_g}\sum_j^{N_g} ijP(i,j) - \mu_x\mu_y$
Sum of squares variance	$f_{12} = \frac{1}{T}\sum_i^{N_g}\sum_j^{N_g}(i-\mu)^2 P(i,j)$
Inverse difference moment	$f_{13} = \frac{1}{T}\sum_i^{N_g}\sum_j^{N_g}\frac{P(i,j)}{1+(i-j)^2}$
Sum average	$f_{14} = \frac{1}{T}\sum_{i=2}^{2N_g} iP_{x+y}(i)$
Sum variance	$f_{15} = \frac{1}{T}\sum_{i=2}^{2N_g}(i-f_{16})^2 P_{x+y}(i)$
Sum entropy	$f_{16} = -\frac{1}{T}\sum_{i=2}^{2N_g} P_{x+y}(i)\log P_{x+y}(i)$
Entropy	$f_{17} = -\frac{1}{T}\sum_i^{N_g}\sum_j^{N_g} P(i,j)\log P(i,j)$
Difference variance	$f_{18} = \frac{1}{T}\sigma_{x-y}$
Difference entropy	$f_{19} = -\frac{1}{T}\sum_{i=0}^{N_g-1} P_{x-y}(i)\log P_{x-y}(i)$

Table 3.3: **Second-order features description: first set.** In this table $P(i,j)$ is the GLCM, $T = \sum_{i,j} P(i,j)$, $P_x(i) = \sum_j P(i,j)$, $P_y(j) = \sum_i P(i,j)$, μ_x, μ_y, σ_x and σ_y are the means and variances of P_x and P_y respectively, $P_{x+y}(k) = \sum_{i,j} P(i,j)$, with $k = 2, 3, \cdots, 2N_g$ and $i + j = k$, $P_{x-y} = \sum_{i,j} P(i,j)$ with $k = 0, 1, \cdots, N_g - 1$ and $|i - j| = k$, σ_{x-y} is the variance of P_{x-y}.

where $\theta = 0°$, $45°$, $90°$ and $135°$.

As an example consider again a 5×5 image I in with 4 gray levels for its pixels. The matrix components for the GLRLM following the four different directions will be 4×5 matrices, because there are 4 gray levels and 5 is the maximum run length there could be in a 5×5 matrix. The sum $Q(i,j) = q(i,j,0°) + q(i,j,45°) + q(i,j,90°) + q(i,j,135°)$ is the GLRLM for the nearest neighbouring pixels. An example of such an image I and its corresponding GLRLM can be seen in table 3.2.

Texture features calculated from the GLRLM are described in table 3.5.

Feature	Equation
Info. measure of correlation 1	$f_{20} = \frac{1}{T}\frac{HXY - HXY1}{max\{HX, HY\}}$
Info. measure of correlation 2	$f_{21} = \frac{1}{T}\sqrt{1 - \exp[-2(HXY2 - HXY)]}$
Max. correlation coefficient	$f_{22} =$ 2nd largest eigenvalue $Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$
Maximal probability	$f_{23} = \frac{1}{T}\max p(i,j)$
Diagonal moment	$f_{24} = \frac{1}{T}\sum_i^{N_g}\sum_j^{N_g} \lvert i-j\rvert(i+j-\mu_x-\mu_y)p(i,j)$
Dissimilarity	$f_{25} = \frac{1}{T}\sum_i\sum_j \lvert i-j\rvert p(i,j)$
Difference energy	$f_{26} = \frac{1}{T}\sum_i p_{x-y}(i)^2$
Inertia	$f_{27} = \frac{1}{T}\sum_i\sum_j (i-j)^2 p(i,j)$
Inverse difference moment	$f_{28} = \frac{1}{T}\sum_i^{N_g}\sum_j^{N_g} \frac{p(i,j)}{1+(i-j)}$
Sum energy	$f_{29} = \frac{1}{T}\sum_i p_{x+y}(i)^2$
Cluster shade	$f_{30} = \frac{1}{T}\sum_i^{N_g-1}\sum_j^{N_g-1}(i+j-\mu_x-\mu_y)^3 p(i,j)$
Cluster prominence	$f_{31} = \frac{1}{T}\sum_i^{N_g-1}\sum_j^{N_g-1}(i+j-\mu_x-\mu_y)^4 p(i,j)$

Table 3.4: **Second-order features description: second set.** In this table: $P(i,j)$ is the GLCM, $T = \sum_{i,j} P(i,j)$, $P_x(i) = \sum_j P(i,j)$, $P_y(j) = \sum_i P(i,j)$, μ_x, μ_y, σ_x and σ_y are the means and variances of P_x and P_y respectively, $P_{x+y}(k) = \sum_{i,j} P(i,j)$, with $k = 2, 3, \cdots, 2N_g$ and $i+j = k$, $P_{x-y} = \sum_{i,j} = P(i,j)$ with $k = 0, 1, \cdots, N_g - 1$ and $|i-j| = k$, HX and HY are the entropies of P_x and P_y respectively, $HXY = -\sum_{i,j} P(i,j)\log P(i,j)$, $HXY1 = -\sum_{i,j} P_x(i)P_y(j)\log P(i,j)$, $HXY2 = -\sum_{i,j} P_x(i)P_y(j)\log P_x(i)P_y(j)$.

Feature	Equation
Short run emphasis	$f_{32} = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{Q(i,j)}{j^2}$
Long run emphasis	$f_{33} = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}Q(i,j)j^2$
Gray-level non-uniformity	$f_{34} = \frac{1}{n_r}\sum_{i=1}^{M}(\sum_{j=1}^{N}Q(i,j))^2$
Run-length non-uniformity	$f_{35} = \frac{1}{n_r}\sum_{j=1}^{N}(\sum_{i=1}^{M}Q(i,j))^2$
Run-percentage	$f_{36} = \frac{n_r}{n_p}$
Low gray-level run emphasis	$f_{37} = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{Q(i,j)}{i^2}$
High gray-level run emphasis	$f_{38} = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}Q(i,j)i^2$
Short run low gray-level emphasis	$f_{39} = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{Q(i,j)}{i^2j^2}$
Short run high gray-level emphasis	$f_{40} = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{Q(i,j)i^2}{j^2}$
Long run low gray-level emphasis	$f_{41} = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{Q(i,j)j^2}{i^2}$
Long run high gray-level emphasis	$f_{42} = \frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}Q(i,j)i^2j^2$

Table 3.5: **High-order texture features from GLRLM.** Here, M represents the number of gray levels and N is the maximum run length, n_r is the total number of runs and n_p is the number of pixels in the image.

3.4.2 Neighbouring gray-level dependance matrix: NGLDM)

The NGLDM was introduced by Sun et al. in 1983 [9] which for an image I is defined as follows:

$$\begin{aligned} S(k,l) = \#\{(i,j)|I(i,j) = k \quad \text{and} \quad & \#[(q,r)|\rho((i,j),(q,r)) \leq d \\ \text{and} \quad & |I(i,j) - I(q,r)| \leq a] = l\} \end{aligned} \tag{3.3}$$

where $\rho((i,j),(q,r))$ is the distance between pixel at coordinates (i,j) and (q,r), d is a given distance for neighbouring pixels ($d = 1$ for nearest neighbours), and a is a positive integer. For example, for a 5×5 image I, we set $a = 0$ (equal gray levels) and $d = 1$ (nearest neighbours of a pixel). In this case the element (k,l) of the NGLDM is defined as

the number of pixels with gray level k that exactly have l neighbouring pixels (at distance $d = 1$). Therefore, the NGLDM will be a 4×8 matrix (4 gray levels, 8 possible nearest neighbours for each pixel). An example of an image I and its corresponding NGLDM can be seen in table 3.2.

Texture features calculated from the NGLDM are described in table 3.6.

Feature	Equation
Small number emphasis	$f_{43} = \frac{1}{SS} \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} \frac{S(i,j)}{j^2}$
Large number emphasis	$f_{44} = \frac{1}{SS} \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} j^2 S(i,j)$
Number nonuniformity	$f_{45} = \frac{1}{SS} \sum_{j=1}^{N_d} (\sum_{i=1}^{N_g} S(i,j))^2$
Second moment	$f_{46} = \frac{1}{SS} \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} S(i,j)^2$
Entropy	$f_{47} = \frac{1}{SS} \sum_{i=1}^{N_g} \sum_{j=1}^{N_d} S(i,j) \log S(i,j)$

Table 3.6: **High-order texture features from NGLDM.** $SS = \sum_{i=1}^{N_g} \sum_{j=1}^{S} S(i,j)$, N_g is the number of gray level values in the image and N_d is the number of different values in the NGLDM of the image

3.4.3 Neighbouring gray-tone dependence matrix: NGTDM

The NGTDM was first proposed by Amadasun and King in 1989 [10]. This is a column matrix that takes into account the variability of the surroundings of pixels.

Lets take an image I, with gray-tone value $I(k,l) = i$ for the pixel located at coordinates (k,l). Let d be a measure of the neighbourhood size, the average gray tone value in the neighbourhood centred at coordinates (k,l) is given by:

$$A(k,l) = \frac{1}{w-1} \sum_{m=-d}^{d} \sum_{n=-d}^{d} I(k+m, l+n), \quad (m,n) \neq (0,0)$$

where $w = (2d+1)^2$. The ith element of the NGTDM is defined as:

$$T(i) = \begin{cases} \sum_{(k,l)/I(k,l)=i} \mid i - A(k,l) \mid, & \text{for } i \in N_i \quad \text{if } N_i \neq \oslash \\ 0, & \text{if} \quad N_i = \oslash \end{cases}$$

where N_i is the ensemble of pixels with gray tone i (peripheral regions of size d are excluded from the calculation). For example, lets take the same 5×5 image I from table 3.2. Excluding the peripheral region, there are no pixels of intensity $i = 0$, therefore we will have $T(0) = 0$. For the gray tone $i = 1$ there are three pixels to be taken into account, the calculation for the element $T(1)$ of the matrix will look like this:

$$T(1) = \left|1 - \frac{14}{8}\right| + \left|1 - \frac{14}{8}\right| + \left|1 - \frac{23}{8}\right| = \frac{27}{8} = 3.375$$

Similarly we obtain $T(2) = 0.125$ and $T(3) = 4.75$. These four values constitute the NGTDM.

Texture features calculated from the NGTDM are given in table 3.7.

Feature	Equation
Coarseness	$f_{48} = \left[\epsilon + \sum_{i=0}^{G_h} p(i)T(i)\right]^{-1}$
Contrast	$f_{49} = \left[\frac{1}{N_g(N_g-1)} \sum_{i=0}^{G_h} \sum_{j=0}^{G_h} p(i)p(j)(i-j)^2\right] \left[\frac{1}{n^2} \sum_{i=0}^{G_h} T(i)\right]$
Busynes	$f_{50} = \left[\sum_{i=0}^{G_h} p(i)T(i)\right] \left[\sum_{i=0}^{G_h} \sum_{j=0}^{G_h} ip(i) - jp(j)\right]^{-1}, p(i), p(j) \neq 0$

Table 3.7: **High-order features from the NGTDM.** In this table, $T(i)$ is the ith value of the NGTDM, N_g is the number of gray level values in I, G_h is the highest gray level, p(i) is the probability of appearance of gray level i, ϵ is a small number and n is the number of pixels in I.

3.5 Heterogeneity

We used a way to quantify heterogeneity that has been proven to correspond fairly well to the appreciation of trained radiologists [12]. This quantification uses a three-step method before quantification of heterogeneity itself.

Lets take two voxels inside a 3D matrix, m and n, with corresponding intensities I_m and I_n. The Bresenham algorithm [22] is used to select the voxels that correspond to a discrete approximation to a straight line between voxels m and n. We will refer to this approximation as the Bresenham line. Let's take r_{ml} as a voxel at a distance l from m. Then, for a linear interpolation of gray-level intensities between voxels m and n the intensity of voxel r_{ml} would be:

$$I(r_{ml}) = I_m + \frac{I_n - I_m}{r_{ml}} r_{ml} \tag{3.4}$$

As the second step, let $\mathscr{L}$ be the Bresenham line between m and n, and L be the number of elements of $\mathscr{L}$. Then, the deviation from this linear interpolation is calculated:

$$\overline{\Delta I} = \frac{1}{L} \sum_{l \in \mathscr{L}} \mid I(r_{ml}) - I_l \mid \tag{3.5}$$

where I_l is the actual gray-value of voxel r_{ml}.

The third step is the calculation of the average deviation, defined as the average value of all the $\overline{\Delta I}$ values between all pair of voxels separated by the same Bresenham length.

$$\overline{\overline{\Delta I}} \equiv \langle \overline{\Delta I} \rangle_{ens}, \tag{3.6}$$

where *ens* is the ensemble of all voxels inside the region being considered.

Finally, let $\tilde{L}$ be the longest Bresenham line between two voxels inside this region, then, the heterogeneity is calculated as:

$$h \equiv \int_0^1 \overline{\overline{\Delta I}}(L/\tilde{L})d(L/\tilde{L}) \tag{3.7}$$

As an example, figure 3.2 shows the heterogeneity quantification for four sample images that have all the same pixel mean value. Values for heterogeneity depend on the gray-level values from the image; the values shown in image 3.2 are only indicative and are meant to show how this measure of heterogeneity grows with different image types.

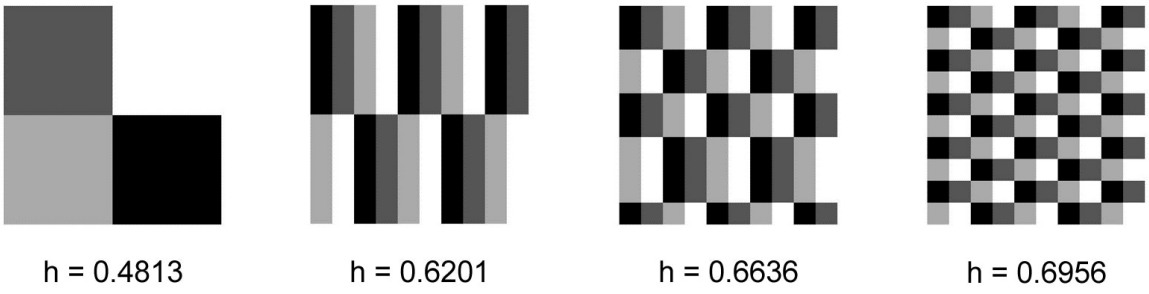

Figure 3.2: Example of quantification of heterogeneity.

3.6 Method for computation of texture features from medical images

In order to calculate texture features from an image, we first define the concept of patch. A patch is defined as a certain number of voxels in a cubic arrangement. For example, a 5-patch is a cubic arrangement of $5 \times 5 \times 5$ voxels. Once a patch is defined, we calculated

the corresponding gray-level histogram, GLCM, GLRLM, NGLDM and NGTDM matrices and, finally, the texture features.

The procedure described above allows for the calculation of a certain texture feature inside a patch (of any given size). These calculations are used to obtain what we call parametric maps. A parametric map is a matrix with the same size as the tomographic image (PET image for example), in which the value of each voxel is replaced by the texture feature value calculated inside the patch surrounding that voxel. Figure 3.3 depicts a schematic for this procedure using an example of PET scan data.

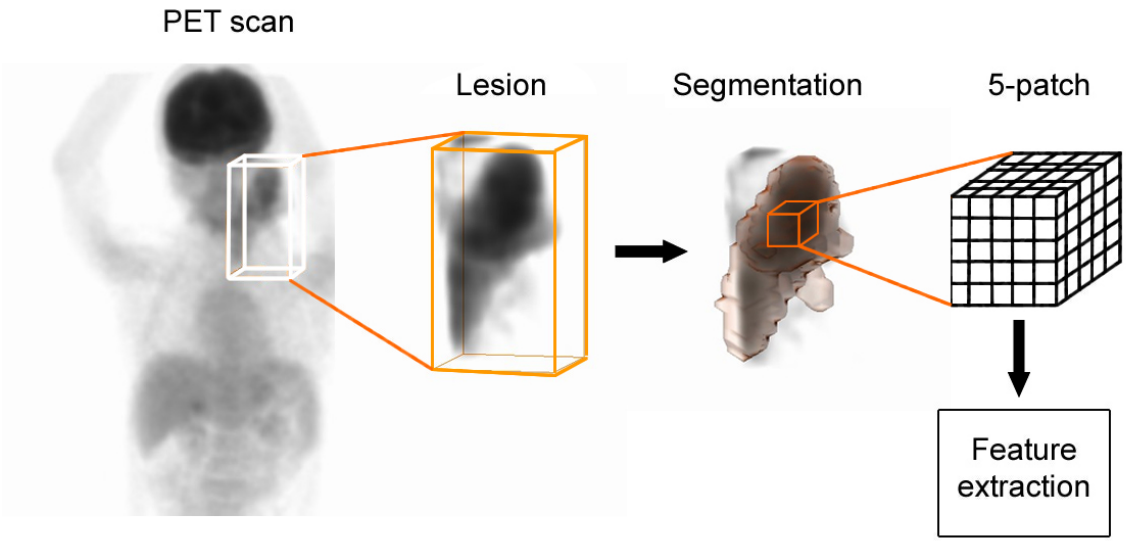

Figure 3.3: Schematic for texture feature extraction from a 5-patch.

Therefore, a parametric map can be obtained for every single texture feature. Finally, we define a ROI in the parametric map, for which the value for the corresponding texture feature is taken as the average of all the voxels inside this region.

3.7 Interpretation of texture features

As mentioned before, texture features are defined in an attempt to assign a value to a certain visual aspect from images. However, it is difficult to associate a certain visual characteristic to every feature described in this chapter. We will explain the interpretation for some of these in the paragraphs below.

First order features take into account grey values from voxels as a whole, it does not matter where they are locate in the image. Therefore, these give an idea about the variation of grey values throughout the image, they do not refer to a texture pattern.

Features taking into account spatial variation s of gray values in the image are second and higher order features. For example, the angular moment (f_9 in table 3.3) could be understood as a measure of homogeneity of the image. This is justified by the number of transitions of gray values of small and large magnitude that are taken into account in the GLCM. As another example, the correlation feature (f_{11} in table 3.3) is a measure of gray tone linear dependencies in the image.

All features from the GLRLM as described in table 3.5 are made to quantify run-length distributions in images, taking into account short and long run-lengths associated to high and low gray values and taking into account all directions of neighbouring voxels in a cube (3D image).

Regarding at features calculated from the NGLDM (see table 3.6), and taking into account how this matrix is constructed using the gray values of voxels, we have the following interpretations. The small number emphasis (f_{43}) is a measure of the fineness, whereas large number emphasis (f_{44}) measures the coarseness of the image. The second moment

(f_{46}) measures the homogeneity of the NGLDM matrix and therefore, of the image itself. Notice at this point that there are several different ways to define the homogeneity of an image. The number nonuniformity (f_{45}) and entropy (f_{47}) are a way to measure homogeneity of the number distribution in the NGLDM. It is difficult to associate the two latter features to a visual texture characteristic of the image, however, they are related to coarseness.

Finally, for textural features calculated from the NGTDM (see table 3.7), coarseness (f_{48}) and contrast (f_{49}) are different ways of measuring these textural characteristics from images. Busyness (f_{50}) is conceived to measure rapid intensity changes between neighbouring voxels.

Chapter 4

PET and SPECT Imaging for lung ventilation

4.1 Introduction

The severity of chronic obstructive pulmonary disease (COPD) is clinically diagnosed with spirometry by measuring forced expiratory volume in 1 s (FEV1). However, this measurement does not give any information about the obstructed regions in the lung. A better assessment of small airway diseases can be obtained with lung scintigraphy [23]. Ventilation/perfusion (V/Q) SPECT has also been incorporated for airway function diagnose [24].

For regional lung obstruction imaging, tomographic 3D images obtained in SPECT have a clear advantage over 2D projection images obtained in scintigraphy. Technegas (Cyclomedica Ltd.) is a technology used in SPECT to assess lung ventilation function [25].

Technegas is a pseudo-gas, consisting of carbon nanoparticles suspended in argon and labeled with ^{99m}Tc. The size of carbon nanoparticles is about 100 - 300 nm. This allows for pulmonary ventilation imaging in nearly physiological conditions, since nanoparticles can reach alveolar space [24].

As an alternative to SPECT, studies have reported the feasibility and implementation of Gallgas-PET scans for lung ventilation [26–29]. *Gallgas* is a radiotracer made with the same carbon nanoparticles as Technegas, but labeled with 68Ga. Gallium-68 is an attractive PET radionuclide due to its short half-life ($\sim$ 68 min), ease of production, availability and compatibility with medical applications [30, 31]. Qualitative comparison studies between Technegas-SPECT and Gallgas-PET have been reported [32], where Gallgas PET shows a better distribution inside the lungs providing a better visualization of ventilation heterogeneities. Other comparative studies showed that diagnoses with Technegas-SPECT and Gallgas-PET are compatible most of the time, however, PET offers more confidence to suspect a different diagnose in some cases [28].

It is well known that PET offers better spatial resolution than SPECT [33]. However, SPECT has the potential to become a more sophisticated technology for image analysis techniques [24]. On the other hand, PET is gaining relevance in diagnose of pulmonary imaging [34], and will most likely continue to be used for this purpose in the future. For these reasons, we considered a quantitative comparison between Gallgas-PET and Technegas-SPECT to be of interest.

Previous authors investigated the feasibility of selecting obstructed lung regions in the lungs using textural features calculated from CT images [35]. They found that there is a

very good agreement with the segmentation obtained through texture features quantification and visual assessment done by experienced physicians.

Moreover, advanced quantitative parameters like texture features, measured from ventilation PET and SPECT scans, are likely to show correlations to FEV1 as the physiologic variable used to assess COPD severity.

4.2 Methods

4.2.1 Patients

This study has been approved by the institutional review board and all subjects signed an informed consent form. For this retrospective study, 5 patients with appropriate non-attenuation-corrected Gallgas-PET scans and available Technegas non-attenuation-corrected SPECT scans were selected. Non-attenuation correction scans were used in order to compare PET to SPECT acquisition mode. All these patients had moderate to severe COPD and underwent Gallgas-PET scans at our center between September 2011 and July 2012. PET and SPECT scans were performed within 4 days of each other. Table 4.1 shows the details for the patients.

Patient	Diagnose	Age (years)	Sex
1	bullous emphysema	48	F
2	panlobular emphysema	55	M
3	chronic bronchitis	74	M
4	emphysema	61	M
5	diffuse panlobular emphysema	37	M

Table 4.1: **Details for the patient cohort.** Images from five patients were retrospectively retrieved for this study.

4.2.2 Radiopharmaceutical production

Technetium 99m was obtained from a $^{99}Mo/^{99m}Tc$ generator. For Gallium-68, an eluate from a $^{68}Ge/^{68}Ga$ generator was purified and concentrated by a QMA cartridge before production of Gallgas nanoparticles. In both cases, carbon nanoparticles were produced with a commercial device from Cyclomedica Ltd. Both radioisotopes were bind to carbon particles in a pure argon atmosphere and a temperature of 2500 °C. For Gallgas, carbon nanoparticles were labelled with ^{68}Ga as $GaCl_3$. For Technegas, nanoparticles were labelled with ^{99m}Tc as $^{99m}TcO_{4-}$.

4.2.3 Imaging

An estimated mean inhaled activity of 70 MBq for Technegas and 27 MBq for Gallgas was administered to the patients, instants prior to scanning. Ventilation SPECT was performed acquiring 120 images around a 360° angle, with 15 seconds per step. Gallgas PET/CT was performed using 2 or 3 bed positions, 3 min each, with a 64-slice multidetector CT component. Each bed position comprises 46 slices. A bed overlap of 11 slices was used (approximately 24%). Both images were reconstructed using an iterative OSEM algorithm. PET images were corrected using time of flight (TOF).

4.2.4 Region of interest definition

The analyzed voxels from PET and SPECT scans were limited to the pulmonary region. This ROI was segmented using the inspiration CT images and extrapolated to PET and SPECT scans. All ROIs underwent visual inspection for further verification. Trachea and

main bronchus regions were removed manually from the ROI.

4.2.5 Image analysis

SPECT data was not corrected for attenuation, since it was acquired using a dedicated SPECT camera. In order to compare both imaging techniques, we used the non attenuation-corrected PET data. SPECT scans were rigidly registered to PET using normalized mutual information maximization criteria. After rigid registration, SPECT scans data grid sizes were resampled to PET scans data grids using nearest-neighbor interpolation method.

4.2.6 Volume segmentation

Threshold based segmentation was performed by taking into account voxels with values within a fixed range from SPECT and PET scans data. Pearson correlation coefficients between segmented volumes and FEV1 were calculated to find optimal segmentation thresholds. Voxel histograms were calculated considering voxels inside the segmented pulmonary region. For these voxels, an estimated value for background activity was calculated using the mean value of voxels from the surrounding area outside the lungs. Voxels with values lower than background were not considered for the histograms.

4.3 Texture features

Texture features were measured from PET and SPECT scans. To quantify these values we calculated parametric maps from scans using 5-patches and parametric maps as described

in section 3.6.

To assign a feature a value from a given ROI, the mean value of the corresponding parametric map voxels inside this ROI is calculated. Voxels having neighbouring patches that fell outside the pulmonary region were not taken into account. Details and enumeration for each of these features is given in chapter 3.

4.3.1 Correlations between texture feature segmentation and FEV1

Using the previously described parametric maps, we used threshold based segmentation to divide voxels inside the lung region. We measured Pearson correlation coefficients between the total volume of the segmented regions and FEV1 volumes, for all 50 texture features, after proper standardization of features quantitative values. The highest Pearson correlation coefficient was used as criteria to find optimal segmentation threshold using texture features.

4.3.2 Statistical comparison between texture features

Using the optimal segmentation thresholds measured from PET and SPECT, we selected the voxels inside the obstruction regions common to both techniques. Thus, we took into consideration voxels corresponding to obstructed regions for both PET and SPECT scans. We compared the mean and variance values for each texture feature from all these voxels taken from all 5 patients to compare SPECT and PET features.

4.4 Results

Correlation coefficients between FEV1 and segmented volumes for several thresholds are shown in figure 4.1. Maximum correlation coefficient values are obtained for a threshold of 27% of maximum uptake for PET and 31% of maximum uptake for SPECT scans, with Pearson correlation coefficients of 0.90 ($p = 0.039$) and 0.98 ($p = 0.002$) respectively.

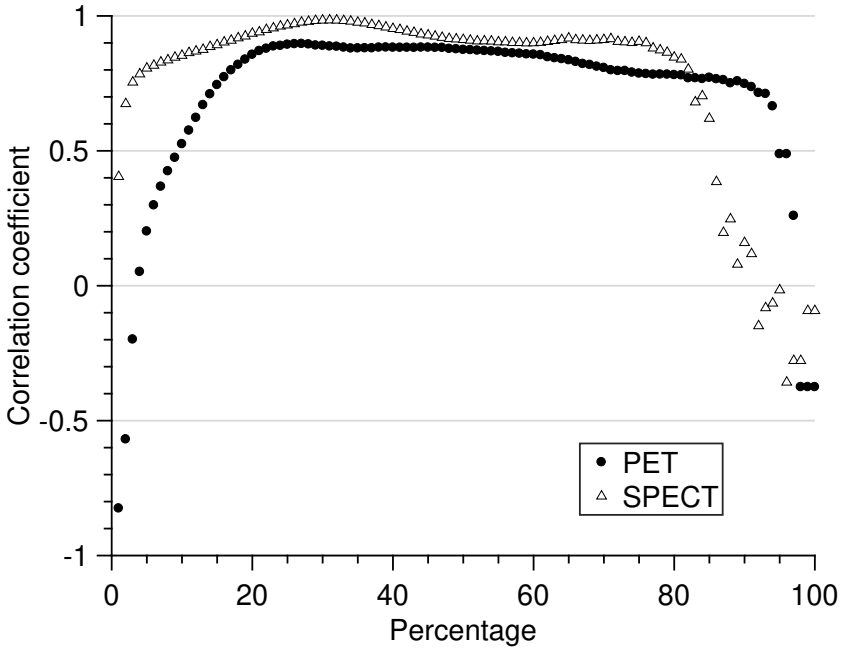

Figure 4.1: Correlation coefficients between threshold-segmented volumes and FEV1.

The histograms measured for each patient, from PET and SPECT scans, and the intensity volume histograms (IVH) are shown in figure 4.2. IVH show the volume in liters taking into account all the voxels with values above a given threshold as percentage of maximum uptake. Figure 4.3 shows maximum intensity projections of PET and SPECT scans limited to the lung regions for all 5 patients in the study.

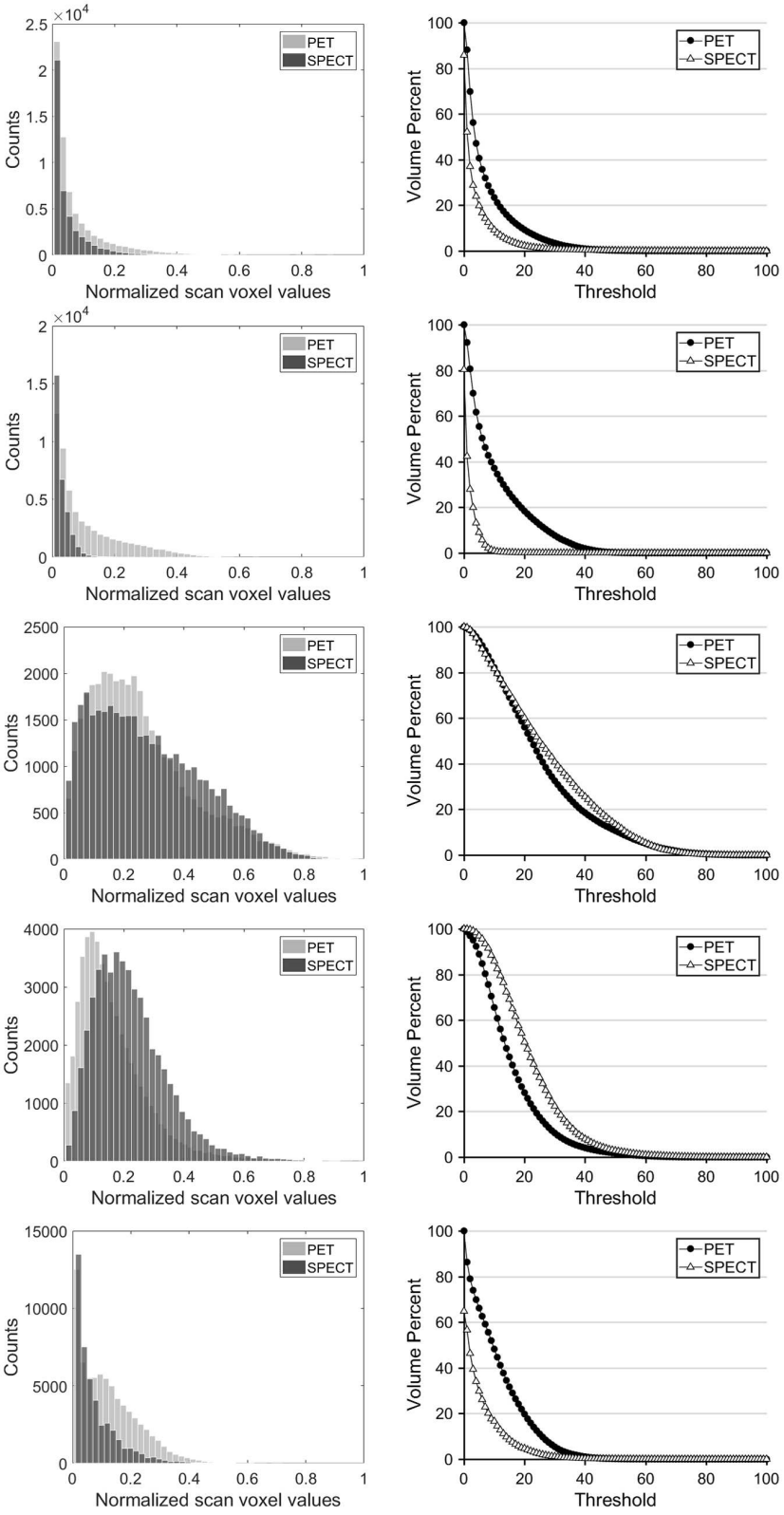

Figure 4.2: Comparison between histograms of PET and SPECT scans (first column) and intensity-volume histograms (second column), for each patient.

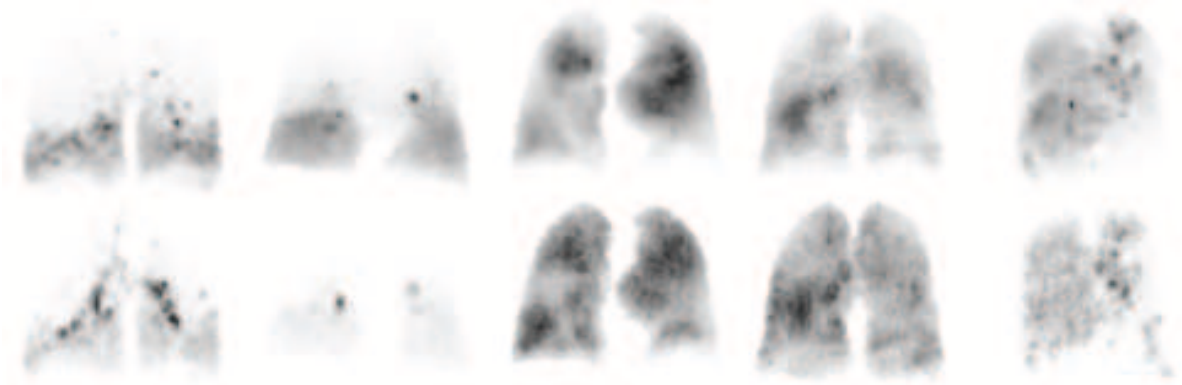

Figure 4.3: Normalized voxel values PET (top row) and SPECT (bottom row) maximum intensity projection scans. Each column corresponds to scans for the same patient.

Segmentation volumes using some of the 50 different texture features were found to show high Pearson correlation coefficients (> 0.9) when compared to FEV1 as the physiologic variable. These features, their corresponding threshold values for segmentation, together with correlation coefficients and their p-values are shown in tables 4.2 and 4.3 for Gallgas-PET and Technegas-SPECT respectively.

Feature (number)	Threshold	Pearson coefficient	P value
3	32	0,96	0,011
5	52	0,91	0,031
10	86	0,94	0,018
23	76	0,93	0,020
27	86	0,94	0,012
31	91	0,97	0,005
35	81	0,98	0,004

Table 4.2: **PET texture features used for segmentation.** Features for which resulting volumes show high correlation with FEV1.

Mean and standard deviation values of voxels inside obstructed lung regions were calculated from parametric maps corresponding to these features, for both Technegas-SPECT and Gallgas-PET scans. Comparison bar graphs for mean and standard deviation

Feature (number)	Threshold	Pearson coefficient	P value
1	43	0,97	0,006
2	39	>0,99	<0,0001
3	18	0,97	0,006
4	71	0,97	0,005
8	92	0,9	0,036
12	51	>0,99	0,0003
14	65	0,98	0,004
16	100	0,95	0,014
17	98	0,95	0,014
19	99	0,96	0,011
29	49	0,97	0,006
33	53	>0,99	0,0001
34	96	0,97	0,008
36	99	0,96	0,01
38	85	0,97	0,006
41	53	>0,99	0,0003
43	96	0,95	0,012
50	49	0,98	0,003

Table 4.3: **SPECT texture features used for segmentation.** Features for which resulting volumes show high correlation with FEV1.

can be seen in figures 4.4 and 4.5 respectively.

4.5 Discussion

FEV1 is the clinical quantitative variable generally used to assess COPD severity, this is why we used these values in order to find a correlation with lung ventilation volumes and thus obtain significant threshold for segmentation in PET and SPECT scans. Optimal thresholds (as percentages of maximum voxel value) used to segment ventilation volumes were similar for PET and SPECT scans (27% and 31% respectively). Previous studies analyzed the relationship between ventilation/perfusion ratio and FEV1 [36, 37]. In this case we analyzed correlations between ventilation scans and FEV1. A correlation with a

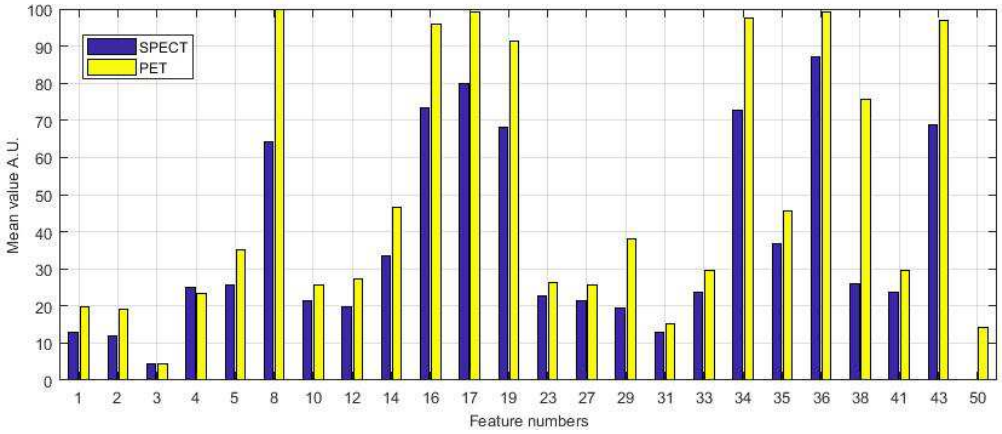

Figure 4.4: Mean values for the features with which segmentation volumes exhibit high correlation with FEV1, inside obstructed lung regions.

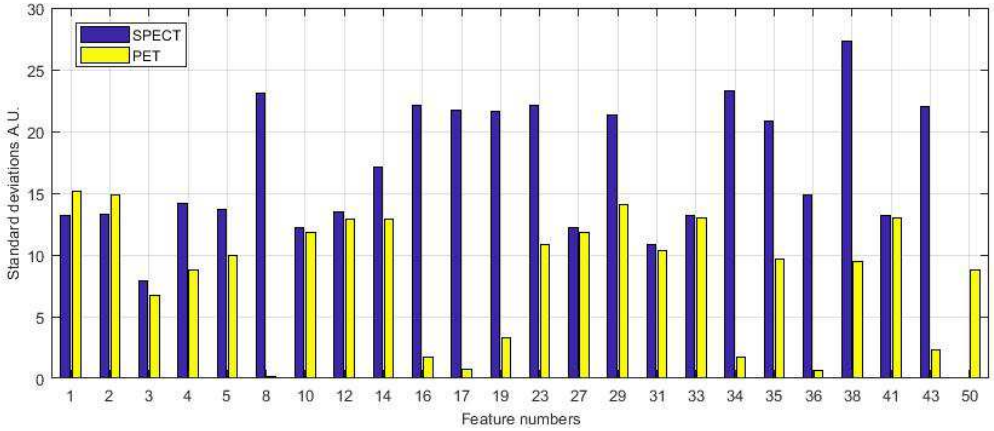

Figure 4.5: Standard deviations for the features with which segmentation volumes exhibit high correlation with FEV1, inside obstructed lung regions.

larger patient cohort between ventilation segmented volumes and FEV1 could reveal a link between tracer distribution and this physiological uptake. Visually, we can see that the uptake regions above this thresholds correspond to what would be visually assigned to the ventilated regions of the lungs. Figure 4.6 shows the ventilated regions calculated using these thresholds in the same slice of the same patient using PET and SPECT (images have been fused with CT for clarity).

In an attempt to characterize tracer distribution, we plotted histograms and IVH from

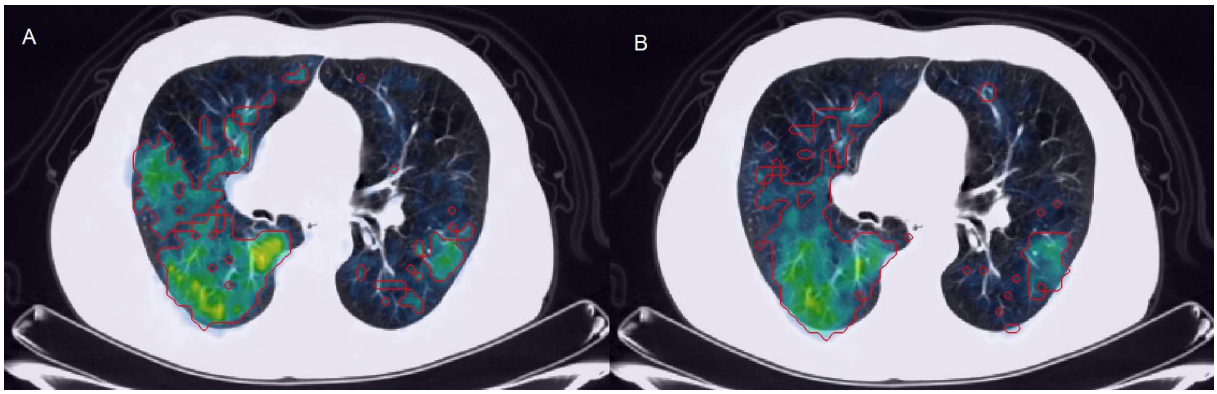

Figure 4.6: Technegas-SPECT segmentation of ventilated region with a threshold of 31% over the SPECT scan fused with the CT (A). Gallgas-PET segmentation of ventilated region with a threshold of 27% over the PET scan fused with the CT (B).

PET and SPECT scans and compared them for each one of our five patients. Comparisons between histograms from PET and SPECT show similarities in some regions, but most of them show differences for a given range of voxel values. These histograms are complemented by the IVHs. Regions where histograms show more voxel counts in PET, together with the higher level of the IVH curves agree with the fact that there is a better peripheral distribution of Gallgas when compared to Technegas. This can also be seen by looking at non attenuation-correction PET and SPECT scans, which allows for a better visual comparison of peripheral distribution. Interestingly, even if three of the cases in this study showed the same tendency, the other two showed an opposite, yet less marked, behaviour. This suggests the radiopharmaceutical distribution could strongly depend on the cause of pulmonary disease.

Texture features are useful when they can be related and used to assess certain aspects of patient clinical data. In this case, we used textural features for threshold-based segmentations using both SPECT and PET data in the lung tissues. Textural features in the lungs have been studied as measured from CT scans [35]. To our knowledge there

are no studies where textural features have been quantified using SPECT or PET scans for lung ventilation. We chose a patch size of $5 \times 5 \times 5$ voxels (5-patch), which by the voxel size accounts for 2.5 cm, as we considered this to be large enough for the features not to be greatly affected by respiratory motion. Nevertheless, it would be interesting to conduct sensitivity studies with a larger patient sample and perform gated acquisitions. The segmented volumes we obtained showed a good correlation to FEV1 volumes for 18 features in SPECT and 7 features in PET. This suggests that there are intrinsic differences between the features measured in the obstructed and ventilated uptake regions of the lungs, in both SPECT and PET, and that textural features could play an important role in lung parenchyma segmentation using these imaging modalities. Also, as we can see in figures 4.4 and 4.5, some features exhibit similar statistical distributions in SPECT and PET scans. We think that a standardized protocol could provide more robust texture features quantification in both techniques, as they seem to be consistent between these two scanning modalities. Although these measurements should be taken with caution as we have only a few patients, these results are encouraging to continue with similar trials.

4.6 Conclusion

To our knowledge, this is the first study analyzing the differences between PET and SPECT using quantitative metrics other than SUV and ventilation/perfusion. Taking this small cohort of patients into consideration, results suggest that tracer distributions could strongly depend on the cause of the disease. Segmentation using some texture features quantified from SPECT and PET were found to have a better correlation to

physiological variable FEV1, motivating to continue the research in this field. Studies including a larger cohort of patients are necessary in order to have statistically significant results.

Chapter 5

Quantitative imaging of head and neck cancer

5.1 Introduction

Head and neck cancer accounts for more than 630,000 new cases and 350,000 deaths per year worldwide. From these, 90% are head and neck squamous cell carcinomas (HNSCC) [38]. PET or PET/CT functional imaging with ^{18}F-FDG is an accurate imaging technique for tumor staging for cervical node metastases in HNSCC [39], solid tumors [40] and various types of cancer [41,42]. ^{18}F-FDG PET has also been shown to be a useful technique for tumor target delineation in head and neck cancer patients [43]. Recent published works report metabolic tumor volume (MTV) as a potentially prognostic biomarker for HNSCC [44]. There have also been reports on the usefulness of standardized uptake values (SUV) as prognostic markers. Maximal SUV value (SUV_{max}) from primary tumor has been found

to be predictor of overall survival, and SUV values for lymph nodes have been proven useful as predictors of extra capsular extension and distant recurrence in patients with head and neck cancer [45]. Textural features have also been found to be useful for segmentation and differentiation between normal and diseased tissues in head and neck cancers using CT and PET imaging [46, 47]. Moreover, imaging features derived from PET scans such as SUV measurements, intensity-volume histograms, texture and shape based features, have been previously studied as prediction markers for treatment outcomes [48]. Therefore, we think that ^{18}F-FDG PET/CT images could be used to extract quantitative information, which would improve the prediction of treatment outcomes and/or personalize treatment to patients. In this context, Wong et al. [49] pointed out the need of a fully automatic processing and analysis scheme, together with a summary of relevant studies in the field of radiomics applied to head and neck cancer. In this work, we develop a framework to get closer to that goal.

We could access images from a clinical trial identified by the number 0522 from the Radiation Therapy Oncology Group (RTOG), which is part of the NRG oncology, a non-profit organization formed to conduct oncology clinical research. (RTOG) protocol 0522 is a randomized clinical trial that was carried out to investigate the benefits of the addition of cetuximab (a drug used to treat colon cancer and other types of squamous cell cancers of the head and neck) to concurrent radiochemotherapy (radiotherapy and chemotherapy at the same time). A particular group from the RTOG 0522 patients had the opportunity to voluntarily participate in a PET sub-study to evaluate treatment outcomes using baseline and post-treatment ^{18}F-FDG PET/CT scans. This is a valuable database to investigate

prediction power of ^{18}F-FDG PET/CT in HNSCC. Access to clinical patient data was granted by our collaboration with the Image Guided Therapy group lead by Prof. Robert Jeraj from the medical physics department of the University of Wisconsin-Madison.

5.2 Methods

5.2.1 Patients

RTOG 0522 is a completed phase III clinical trial investigating the addition of cetuximab to chemoradiation in patients with HNSCC, where patients were randomized to chemoradiation with or without cetuximab. No benefit to progression free survival (PFS) was observed with the addition of cetuximab to this treatment. PFS is defined as time between randomization and local, regional, or distant progression, or death. This database was used retrospectively for this study.

5.2.2 Imaging

Patient database came from different centers, as RTOG 0522 was conceived as a multicenter trial. Patients underwent PET/CT scans on hybrid scanners with CT components of at least 4 slices and full rings for PET imaging. PET scan data was acquired on 2D high-sensitivity mode or 3D mode in scanners without this option.

Patients were asked to fast for 4 to 6 hours before ^{18}F-FDG injection. They were injected a dose ranging from 370 to 740 MBq of ^{18}F-FDG intravenously, and PET images were taken 50-70 min after injection. Patients were imaged with a 120 keV/300 mA, 0.5-seconds

detector rotation time CT scan with intravenous contrast; followed by a 120 keV/80 mA, 0.8-second detector rotation time CT scan for PET attenuation correction, before PET scans. PET scans were reconstructed with the scanner-specific recommended reconstruction algorithm. Post-treatment scans were taken 8-9 weeks after treatment completion (treatments correspond to groups 1 or 2 from the RTOG 0522 protocol) and before any nodal dissection. All post-treatment scans were performed on the same scanner as baseline scans.

5.2.3 Image processing

Only complete datasets (baseline and post-treatment 18F-FDG PET/CT scans) from patients database were taken into account for this study. We drew manually the ROIs enclosing the primary tumor and lymph nodes under the supervision of one experienced nuclear medicine physician at CUDIM. These regions were grown isotropically and used as inputs for a custon semiautomatic segmentation algorithm [50] in order to reduce region definition variability. This algorithm makes use of imaging features and gradients, to define a resulting non-connected ROI, after a starting volume is established as an input. After processing, we visually verified each ROI to avoid undesired regions to be included in the analysis. Three main groups were considered for the ROIs: primary tumor (PT), right metastatic nodes (RN) and left metastatic nodes (LN). We registered post-treatment scans to baseline scans using deformable registration algorithms. As a first step, we performed a local rigid registration between associated CT scans and used as starting configuration to perform deformable registration using DEMONS algorithms from Deformable Image

Registration and Adaptive Radiotherapy (DIRART) toolbox [17] running under MatLab R2017a. We chose the Demons algorithm because it outperformed the Optical Flow. Thus, the obtained motion fields from CT scans were applied to PET data, in order to use the same ROIs as for the pre-treatment scans. To maintain SUV values as close as possible to original scans, all interpolations applying motion vector fields where done using nearest neighbours method. Figure 5.1 depicts the workflow for processing each patient's scans.

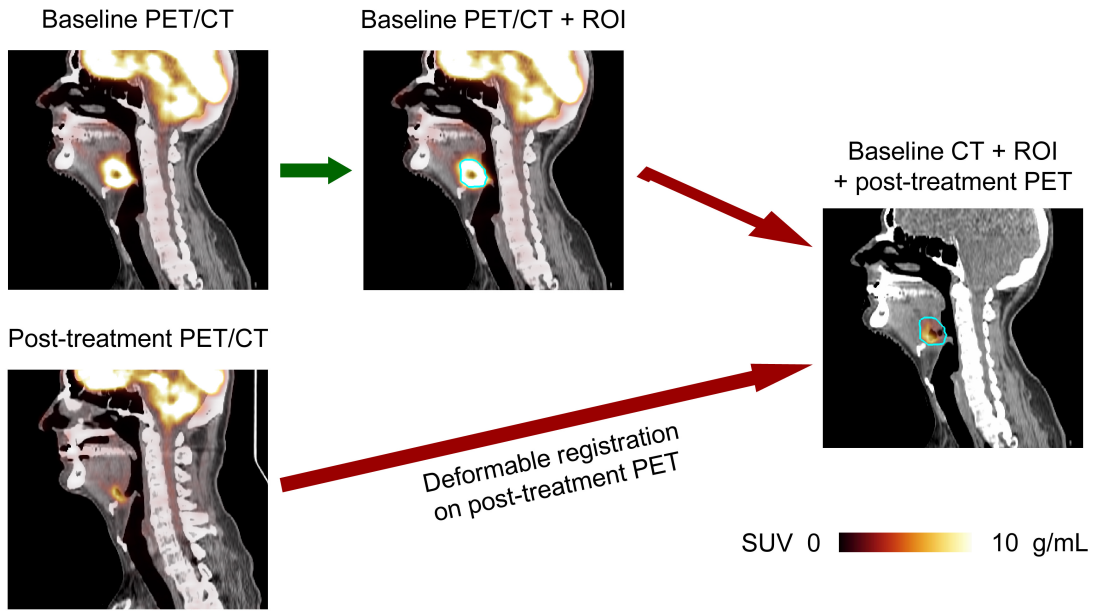

Figure 5.1: Workflow for processing HNSCC patients ^{18}F-FDG PET/CT scans.

A similar method was reported in the literature using rigid registration only, based on expectation maximization algorithms. Authors found this method, used to compare PET scans at different time-points, outperformed threshold based methods when performing voxel-to-voxel comparisons [51]. Baseline, post-treatment and changes in ^{18}F-FDG uptake parameters were evaluated. Figure 5.2 shows an example of SUV changes observed in a

single patient after image processing.

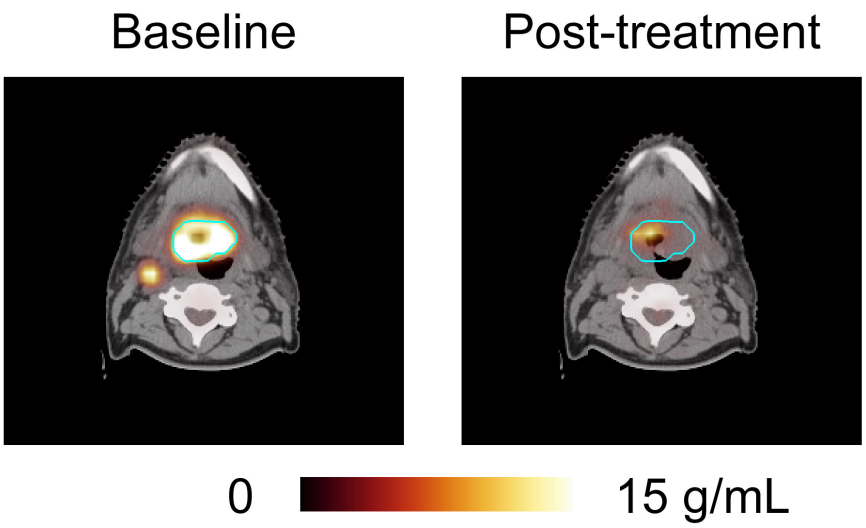

Figure 5.2: SUV changes in a single patient between baseline and post-treatment ^{18}F-FDG PET scans.

5.2.4 Texture features

We computed texture features through parametric maps, using 5-patches as described in chapter 3.6. We decided to use 5-patches for feature computation because this was considered to bee a reasonable size to distinguish feature variations in the lesions. A total of 50 texture features were calculated for each group of lesions, as described and numbered in chapter 3.

5.2.5 Imaging markers

Imaging texture features from metastatic nodes were separated into 2 groups: maximal values among all left and right nodes (MN) and summed values from left and right nodes (SN). Features were treated in 2 ways: continuous and dichotomized by median for PT,

MN and SN. Therefore, a total of 900 variables were assigned to each patient: 50 features, dichotomized or continuous (×2), from T, MN or SN (×3),for baseline, post-treatment or change between baseline and post-treatment (×3).

5.3 Statistical analysis

Markers were correlated to PFS using Cox models. Cox models are also known as "proportional hazards" models, and are used to predict the time of survival using markers as the variables. Using this type of models, hazard ratios (HR) were calculated. HR represent the rate of increased risk of an individual to another when assigned value of a marker (as described in the previous section) above or below he median. In the case of continuous variables, increases in 10% in the total range of values were used to calculate HR. High values of a marker are considered to be protective for $HR < 1$, otherwise, they signal an increased risk.

Statistical Cox models were calculated taking into account markers alone (univariate) or adding markers to either primary tumour or nodal staging (bivariate). Akaike information coefficients (AIC) were calculated for PFS prediction models. AIC is a coefficient used to assess the quality of a statistical model, it takes into account a compromise between fitting and simplicity of the model. The criterion for considering a marker potentially prognostic was reduction in AIC by any amount relative to both primary tumour stage and nodal stage alone.

Finally, Pearson correlation coefficients were calculated between all features meeting this condition.

All statistical analysis were made at the University of Wisconsin-Madison, since clinical data is not available outside the institution.

5.4 Results

One-hundred sixteen patients agreed to participate in the PET sub-study, from which 50 are included in this analysis. Characteristics of the patients are shown in tables 5.1.

Age (years)	
Mean	56.8
Std. Dev.	6.59
Median	56
Min - Max	43 - 71
Q1 - Q3	53 - 61
Gender	
Male	44 (88.0%)
Female	6 (12.0%)
Race	
Asian	1 (2.0%)
Black or African-American	3 (6.0%)
White	46 (92.0%)
Smoking history: pack-years	(n=39)
Mean	26.0
Std. Dev.	31.02
Median	14
Min - Max	0 - 135
Q1 - Q3	0 - 40

Table 5.1: **Patient cohort characteristics (50 patients).** Std. Dev., standard deviation; Q1, first quartile; Q3, third quartile. A pack-year is defined as the equivalent of smoking one pack of cigarettes a day for 1 year.

Tumour and nodes segmentation, as well as deformable registration results were validated by measuring intraclass correlation coefficients (ICC), based on values of SUV and

MTV between our measurements, done under the supervision of an experienced nuclear medicine physician at the Uruguayan center of molecular imaging (CUDIM), and 2 other different groups of nuclear medicine physicians from the University of Wisconsin-Madison. Results for these coefficients can be seen in table 5.2.

	Primary			Right nodes			Left nodes		
	SUV		MTV	SUV		MTV	SUV		MTV
	max	peak		max	peak		max	peak	
Baseline	0,98	0,97	0,93	0,98	0,98	0,83	0,99	0,99	0,95
Post-treatment	0,92	0,83	N/A	0,78	0,77	N/A	0,74	0,79	N/A

Table 5.2: **Intraclass correlation coefficients between observers for MTV and SUV values.** N/A: not applicable due to lack of values for some observers.

Twenty-five PFS events have been reported; median follow-up for the remaining 25 patients is 5.0 years (min-max $4.0 - 6.2$). Models taking into account texture features provided lower AIC values than both primary tumour stage and nodal stage alone for 22 out of 900 features. Descriptive statistics for these 22 features are shown in table 5.3. These values are given only as a reference.

Analysis of these 22 markers for univariate, bivariate adjusted for primary tumour stage and bivariate adjusted for nodal stage is shown in tables 5.4 through 5.6. For reference, a model with only primary tumour stage has an AIC of 176.08 and with only nodal stage an AIC of 179.78.

Marker	Min	Q1	Median	Q3	Max	Mean	Std. dev.
P:#5 (pre-tx)	-0,3	0,3	0,6	0,9	1,6	0,6	0,5
P:#5 (change)	-11,6	-0,4	0,0	0,2	1,2	0,0	0,6
P:#12 (pre-tx)	0	15000	18000	21000	31000	18000	6000
P:#12 (post-tx)	0	12000	15000	17000	22000	14000	4000
P:#14 (pre-tx)	0	200	230	250	330	220	60
P:#33 (pre-tx)	0	14000	17000	20000	30000	17000	6000
P:#33 (post-tx)	0	12000	14000	17000	21000	14000	4000
P:#33 (change)	-17000	-6000	-2000	1000	8000	-3000	5000
P:#41 (pre-tx)	0	14000	17000	20000	30000	16000	6000
P:#41 (post-tx)	0	11000	14000	16000	21000	14000	4000
P:#42 (pre-tx)	0	14000	18000	21000	31000	17000	6000
NM:#6 (post-tx)	0	0	0	0	7	1	1
NM:#6 (change)	-8	-1	0	0	5	-1	2
NM:#14 (post-tx)	0	210	240	260	320	240	50
NM:#23 (pre-tx)	0,000	0,022	0,023	0,024	0,030	0,023	0,004
NM:#32 (post-tx)	0,000	0,004	0,006	0,008	0,015	0,006	0,003
NM:#39 (post-tx)	0,000	0,004	0,006	0,007	0,015	0,006	0,003
NM:#41 (post-tx)	0	14000	17000	19000	26000	17000	5000
NS:#4 (change)	-1,2	-0,6	-0,4	-0,3	0,0	-0,4	0,3
NS:#12 (pre-tx)	0	20000	20000	30000	40000	20000	10000
NS:#20 (change)	-0,00	0,03	0,06	0,09	0,17	0,06	0,04
NS:#24 (change)	-4000	-1000	-1000	0	1000	-1000	1000

Table 5.3: **Statistical variables for markers found potentially prognostic for HNSCC.**Min, minimum; Q1, first quartile; Q3, third quartile; Max, maximum; Std. Dev., standard deviation; pre-tx, pre-treatment; post-tx, post-treatment; change, change from pre-treatment to post-treatment; P, Values for primary tumour; NM, Nodes(max) indicates the the maximum of the values for the right and left side nodes; NS, Nodes(sum) indicates the the sum of the values for the right and left side nodes; # designates the corresponding textural feature number.

For univariate analysis, the lowest AIC corresponded to dichotomized skewness in the primary tumour at pre-treatment ^{18}F-FDG PET/CT scan with a corresponding HR of 0.37 (higher values of skewness are prognostic of a longer PFS time). The p-value, $p = 0.02$ shows there is a significant difference in PFS for values of skewness above the median (within a 95% confidence interval), taking only this variable into the survival model.

Marker	Comparison	AIC	HR (95%CI)	p-value
P:#5 (pre-tx)	by M	173.89	0.37 (0.16-0.84)	0.02
P:#12 (pre-tx)	by M	174.37	2.61 (1.14-5.97)	0.02
P:#33 (pre-tx)	by M	174.37	2.61 (1.14-5.97)	0.02
P:#41 (pre-tx)	by M	174.37	2.61 (1.14-5.97)	0.02
P:#42 (pre-tx)	by M	174.37	2.61 (1.14-5.97)	0.02
P:#14 (pre-tx)	by M	174.53	2.57 (1.12-5.87)	0.03
P:#33 (change)	by M	175.36	0.42 (0.18-0.96)	0.04
NS:#12 (pre-tx)	by M	175.53	2.32 (1.02-5.28)	0.04
P:#5 (change)	by M	175.76	2.26 (1.01-5.08)	0.05
P:#12 (post-tx)	by M	176.18	2.17 (0.96-4.92)	0.06
P:#33 (post-tx)	by M	176.18	2.17 (0.96-4.92)	0.06
P:#41 (post-tx)	by M	176.18	2.17 (0.96-4.92)	0.06
NS:#20 (change)	by M	176.77	0.49 (0.22-1.11)	0.09
NM:#23 (pre-tx)	by M	177.03	1.97 (0.87-4.46)	0.11
NM:#24 (change)	by M	177.34	0.53 (0.23-1.19)	0.12
NM:#6 (post-tx)	continuous	177.38	0.80 (0.56-1.12)	0.19
NM:#41 (post-tx)	continuous	177.52	1.20 (0.94-1.53)	0.14
NM:#39 (post-tx)	continuous	177.57	1.17 (0.95-1.45)	0.13
NM:#14 (post-tx)	continuous	177.58	1.26 (0.91-1.75)	0.17
NM:#32 (post-tx)	continuous	177.80	1.16 (0.95-1.42)	0.15
NM:#6 (change)	continuous	177.81	0.86 (0.70-1.05)	0.13
NS:#4 (change)	by M	178.32	0.61 (0.27-1.37)	0.23

Table 5.4: **Progression-Free Survival Cox Models for markers with Possible Prognostic Value (n=50; 25 events) Univariate Analysis, sorted by AIC (Lowest to Highest).** AIC, Akaike Information Criterion; HR, hazard ratio; CI, confidence interval; pre-tx, pre-treatment; post-tx, post-treatment; change, change from pre-treatment to post-treatment; P: Values for primary tumour; NM, Nodes(max) indicates the the maximum of the values for the right and left side nodes; NS, Nodes(sum) indicates the the sum of the values for the right and left side nodes; by M, $>$ median vs. $\leq$ median; # designates the corresponding textural feature number.

Concerning bivariate analysis adjusted for primary tumour stage, the lowest AIC corresponded to continuous kurtosis for maximum values of nodes in post-treatment ^{18}F-FDG PET/CT scan with a HR of 0.75, higher values of kurtosis are indicative of a longer PFS. The obtained p-value for differences between survival curves according to increasing values

of kurtosis is not significant ($p = 0.10$) within a 95% confidence interval.

Marker	Comparison	AIC	HR (95%CI)	p-value
NM:#6 (post-tx)	continuous	173.86	0.75 (0.54-1.06)	0.10
NM:#6 (change)	continuous	174.39	0.80 (0.65-0.99)	0.04
P:#12 (post-tx)	by M	174.42	2.18 (0.96-4.95)	0.06
P:#5 (pre-tx)	by M	174.43	0.43 (0.18-1.04)	0.06
NS:#24 (change)	by M	174.77	0.47 (0.21-1.07)	0.07
NM:#41 (post-tx)	continuous	174.79	1.26 (0.98-1.63)	0.08
NM:#14 (post-tx)	continuous	175.03	1.34 (0.94-1.92)	0.11
NS:#12 (pre-tx)	by M	175.10	2.06 (0.89-4.77)	0.09
NM:#39 (post-tx)	continuous	175.17	1.20 (0.98-1.46)	0.08
NM:#23 (pre-tx)	by M	175.24	1.99 (0.88-4.51)	0.10
NM:#32 (post-tx)	continuous	175.29	1.19 (0.98-1.45)	0.08
NS:#4 (change)	by M	175.35	0.50 (0.22-1.15)	0.10
P:#33 (post-tx)	by M	175.35	1.98 (0.86-4.52)	0.10
P:#41 (post-tx)	by M	175.35	1.98 (0.86-4.52)	0.10
P:#12 (pre-tx)	by M	175.39	2.17 (0.86-5.46)	0.10
P:#33 (pre-tx)	by M	175.39	2.17 (0.86-5.46)	0.10
P:#41 (pre-tx)	by M	175.39	2.17 (0.86-5.46)	0.10
P:#42 (pre-tx)	by M	175.39	2.17 (0.86-5.46)	0.10
P:#14 (pre-tx)	by M	175.51	2.13 (0.84-5.35)	0.11
P:#33 (change)	by M	175.68	0.50 (0.21-1.21)	0.12
NS:#20 (change)	by M	175.76	0.53 (0.24-1.21)	0.13
P:#5 (change)	by M	175.97	1.89 (0.80-4.48)	0.15

Table 5.5: **Progression-free survival Cox models for markers with possible prognostic value (n=50; 25 events). Bivariate analysis, adjusted for primary tumour stage, sorted by AIC (lowest to highest).** AIC, Akaike Information Criterion; HR, hazard ratio; CI, confidence interval; pre-tx, pre-treatment; post-tx, post-treatment; change, change from pre-treatment to post-treatment; P, values for primary tumour; NM, nodes(max), indicates the the maximum of the values for the right and left side nodes; NS, nodes(sum) indicates the the sum of the values for the right and left side nodes; by M, $>$ median vs. $\leq$ median; # designates the corresponding textural feature number.

However, change of maximum kurtosis of the nodes between pre and post-treatment scans is also considered a prognostic marker with a HR of 0.80 (longer PFS for higher values of this marker). Differences between survival times for patients with increasing

values of this marker are significant (p=0.04) within a 95% confidence interval.

Marker	Comparison	AIC	HR (95%CI)	p-value
NS:#12 (pre-tx)	by M	175.54	3.29 (1.28-8.51)	0.01
P:#5 (pre-tx)	by M	175.73	0.36 (0.16-0.83)	0.02
P:#12 (pre-tx)	by M	176.08	2.71 (1.17-6.31)	0.02
P:#33 (pre-tx)	by M	176.08	2.71 (1.17-6.31)	0.02
P:#41 (pre-tx)	by M	176.08	2.71 (1.17-6.31)	0.02
P:#42 (pre-tx)	by M	176.08	2.71 (1.17-6.31)	0.02
P:#14 (pre-tx)	by M	176.27	2.66 (1.15-6.17)	0.02
P:#33 (change)	by M	177.35	0.42 (0.18-0.96)	0.04
P:#5 (change)	by M	177.74	2.27 (1.01-5.10)	0.05
P:#33 (post-tx)	by M	178.10	2.19 (0.96-4.98)	0.06
P:#41 (post-tx)	by M	178.10	2.19 (0.96-4.98)	0.06
P:#12 (post-tx)	by M	178.17	2.17 (0.96-4.92)	0.06
NS:#20 (change)	by M	178.46	0.46 (0.20-1.07)	0.07
NM:#23 (pre-tx)	by M	178.68	2.11 (0.90-4.91)	0.08
NM:#39 (post-tx)	continuous	179.19	1.20 (0.96-1.51)	0.11
NS:#24 (change)	by M	179.22	0.52 (0.23-1.18)	0.12
NM:#6 (post-tx)	continuous	179.38	0.80 (0.56-1.12)	0.19
NM:#32 (post-tx)	continuous	179.47	1.18 (0.95-1.47)	0.13
NM:#41 (post-tx)	continuous	179.51	1.20 (0.94-1.54)	0.14
NM:#6 (change)	continuous	179.56	0.84 (0.67-1.05)	0.12
NM:#14 (post-tx)	continuous	179.57	1.26 (0.90-1.78)	0.18
NS:#4 (change)	by M	179.74	0.52 (0.21-1.29)	0.16

Table 5.6: **Progression-free survival Cox models for markers with possible prognostic value (n=50; 25 events). Bivariate analysis, adjusted for nodal stage, sorted by AIC (lowest to highest).** AIC, Akaike Information Criterion; HR, hazard ratio; CI, confidence interval; pre-tx, pre-treatment; post-tx, post-treatment; change, change from pre-treatment to post-treatment; P, values for primary tumour; NM, nodes(max), indicates the the maximum of the values for the right and left side nodes; NS, nodes(sum) indicates the the sum of the values for the right and left side nodes; by M, $>$ median vs. $\leq$ median; # designates the corresponding textural feature number.

Finally, for the case of bivariate analysis adjusted for nodal stage, the lowest AIC corresponded to dichotomized sum of square variance for sum of nodes values in pre-treatment ^{18}F-FDG PET/CT scan, with a HR of 3.29, pointing towards a shorter PFS

for high values of this marker. The obtained p-value for differences between survival curves according to values of this marker above and below the median is significant $p = 0.01$ within a 95% confidence interval.

From all these 22 markers, we found 15 pairs of highly correlated variables with Pearson correlation coefficients ≥ 0.96.

5.5 Discussion

As can be seen in table 5.2, we found an excellent agreement between the measurements for SUV and MTV we performed and the measurements of the other two groups of physicians at the University of Wisconsin-Madison. At baseline it is expected to have a very good concordance given that we are looking for the same lesions that were initially diagnosed for the patients. However, for post-treatment values, we relied on the results of our deformable registration procedure. For this case, it is also expected to get a lower agreement, nevertheless, the agreement we obtained for SUV values is remarkable, for values of ICC above 0.75. ICC are higher for primary tumour compared to left/right nodes, this is attributed to the fact that primary tumours are bigger in general, therefore, there is a better chance to obtain a better deformable registration in this case. Based on these results, we decided this processing offered a good degree of confidence in order to measure texture features and perform the analysis of potentially prognostic markers. Unfortunately, the used registration method does not allow for visual inspection of the corresponding regions after application of deformable registration to PET data.

This methodology allowed us to identify potentially prognostic markers from texture fea-

tures for PFS in patients with HNSCC. In contrast to texture features, previous studies found shape-based features as possible predictors for treatment outcomes in head and neck cancer [48]. A similar study conducted with CT images from HNSCC revealed skewness and entropy of lesions were associated with overall survival [52]. In our case, results also point out to skewness as a possible prognostic marker for PFS.

An option worth considering in the future would be to deform ROIs instead of PET data, which would have the advantage of visual confirmation and not modifying PET scans data.

The size of the patient cohort and the small number of events allowed for univariate and bivariate analysis only. Among the markers found as potentially prognostic, some are highly correlated. Therefore, to further test the findings in this work, a larger patient cohort would be necessary to rule out redundant variables, randomness and false positive discoveries. The high correlations between variables allow to reduce the number of meaningful predictors and choose the features that are simpler for computation purposes and/or interpretation, which is often the problem when trying to explain a single feature with a visual characteristic of the images. Thus, prioritizing pre-treatment features in our case, we could further reduce redundant variables to 14. Some candidates worth considering as good predictors measured from pre-treatment scans would be: primary tumour histogram skewness (first order feature), sum of nodes sum of square variance and max maximum probability from nodes (the last two are second order features from GLCM).

For future studies, this type of trials could greatly benefit from a standardized protocol specific for head and neck tumors, such as the one described by Cistaro et al. [53].

5.6 Conclusion

Twenty two texture markers alone and combined with primary tumour and nodal staging, using univariate and bivariate analysis, were found to be possible good predictors of progression-free survival in patients with advanced HNSCC, yielding better model fit when compared to survival models without these markers. The potentially prognostic markers found in this work are very promising, and could be useful for early treatment response assessment and treatment personalization. This motivates to continue in this direction and conduct trials with larger patient cohorts to further verify and standardize these procedures to be included in clinical practice.

Chapter 6

Quantitative imaging of paediatric lymphoma

6.1 Introduction

Since Thomas Hodgkin described the histopathology features of the disease that has his name in 1832, several treatment options have been used. Success rates of treatment for patients with Hodgkin Lymphoma (HL) based in chemotherapy (QT) alone or combined with radiotherapy (RT) have progressively risen. Thanks to these developments, adult and pediatric patients with HL now benefit from PFS survival rates at five years of above 90% [54].

Despite the therapeutic success, associated long term secondary mortality and morbidity has not been reduced sufficiently. Secondary tumors, cardiopulmonary toxicity and endocrine disorders may affect HL survival and quality of life [55–58].

Therefore, there is a need for treatment optimization. The best strategy to achieve this goal is to identify valid, reproducible clinical and imaging prognostic factors that allow more precise patient risk classification. Moccia et al. reported that the international prognostic score (IPS) continues to be useful for stratification in adult patients with advanced HL [59]. However, IPS provides insufficient data to prescribe different therapies [60]. In addition, there is no evidence that supports the use of this score in paediatric patients. As an exemple of the ongoing efforts towards finding early indicators of treatment response, Schwartz et al. identified four clinical poor prognostic factors in a sample of paediatric patients with intermediate risk HL [61]. These factors are, however, not image-derived, they are clinical.

For all ages in HL, ^{18}F-FDG PET/CT is the recommended imaging technique for stratification and treatment response assessment, due to high ^{18}F-FDG uptake in all pathologic subtypes of HL [62]. Moreover, this studies offer also the possibility to distinguish residual metabolically active tissues from post-therapy fibrosis. In addition, ^{18}F-FDG PET/CT offers a better sensitivity, specificity and accuracy compared with anatomical images such as CT alone and bone marrow biopsy [63–65].

In both pediatric and adult patients, ^{18}F-FDG PET/CT at end of two cycles of chemotherapy (called intermediate or iPET) is an excellent prognostic test. A stable or progressive disease found in an iPET scan points towards patients having an increased risk of relapse [66]. Response evaluation criteria from iPET has been simplified to the a visual 5-point scale called the Deauville scale [67–70]. The Deauville scale is used to assign patients to risk groups and therefore, decide upon the therapy they will be prescribed.

Although this scale is currently used in the clinic, it is a purely visual scale, subject to subjective evaluations.

In order to increase the precision of a visual response evaluation scale, such as Deauville, the development of quantitative markers to complement this classification would be of great utility. Hence, this is another case to which radiomics can be applied to. Previous authors have explored the feasibility of response classification in non-Hodgkin lymphoma by exploring texture features changes measured from anatomical magnetic resonance images [71], and texture classification of tissues using texture features measured from stained tissues [72]. In the latter study, authors achieve a high classification accuracy, nonetheless, collecting tissues for staining implies the use of invasive methods.

In this study, we calculate several quantifiable metabolic variables and texture features from baseline and/or interim ^{18}F-FDG PET images from paediatric Hodgkin lymphoma lesions, searching for potentially prognostic imaging markers of disease evolution.

6.2 Methodology

6.2.1 Patient cohort

In this prospective study, twenty-one newly diagnosed patients with confirmed HL were enrolled at the Uruguayan Center of Molecular Imaging (CUDIM) between 2013 and 2017. This trial was approved by the institutional review board and ethics committee. All patients had an informed consent signed by their legal tutors. Patient enrollment criteria were, ages between 0 and 18 years old and confirmed histopathology diagnosis of HL.

Exclusion criteria were previous oncologic disease, previous chemotherpy or radiotherapy, pregnancy and lactation. Patients underwent ^{18}F-FDG PET/CT scans for disease staging (baseline, bPET), at end of two cycles of QT (iPET) and after first line of therapy (end, ePET). For PET/CT staging, we used the Ann Arbor classification system [73]. This staging system takes into account the location of malignant tissues and symptoms related to the disease. Bulky disease was identified when lesions had any diameter larger than 5 cm, this is a clinically relevant factor. Any non-physiologic increase in metabolic activity found in ^{18}F-FDG PET/CT at a lymph-node or extra lymph-node site that could not be explained by another pathologic entity found in clinic or CT was judged as related to lymphoma. Pathological findings in bone marrow were classified as diffuse or focal. Details for patient cohort are summarized in table 6.1.

6.2.2 Patient management

After ^{18}F-FDG PET/CT staging, patients were treated according to the following QT protocol: ABVE (Adriamycin, Bleomycin, Vinblastine, and Etoposide) for patients in the low risk and intermediate risk groups, (14 patients) and BEACOP (Bleomycin, Etoposide, Adriamycin, Cyclophosphamide, Oncovin, and Procarbazine) for patients in the high risk group, (7 patients). This is a standard protocol used in Uruguay. After iPET, patients were classified as responders and non-responders according to the Deauville visual score [68]. If complete response was observed in iPET for low risk and intermediate risk patients, treatment was reduced to only chemotherapy or decreasing radiotherapy doses according to Deauville scale classification. For high risk patients, treatment was not

Characteristics	No. of patients (n=21)
Sex	
Male	8
Female	13
Age	
Mean (years)	11
Range (years)	6 - 17
Stage	
I	2
II	7
III	10
IV	2
Bulky disease	52%
B-symptoms	24%
Extra-nodal disease	5
Histology	
Lymphocyte rich	3
Mixed cellularity	10
Nodular sclerosis	8
Median follow-up (days)	270
Range (days)	92 - 414
Non-complete responders	4
Complete responders	17

Table 6.1: **Paediatric HL patients**. Details for the cohort from CUDIM.

changed after iPET evaluation. Patients were classified by a nuclear medicine specialist at CUDIM as responders or non-responders at end of treatment using ePET. After the prescribed treatment was administered, a clinical and imaging follow up was performed for all patients, in order to find potential relapses.

6.2.3 Image acquisition and processing

For PET, patients were asked to fast for 6 hours before ^{18}F-FDG injection. Injected doses were calculated using the European Association of Nuclear Medicine (EANM) 2008 dosage card. Scans were acquired in 3D mode, 60 minutes after radiopharmaceutical

injection. Reconstruction was done using an ordered subset expectation maximization (OSEM) algorithm with 2 iterations and 28 subsets. PET slices had a resolution of $128x128$ voxels. Overlap between bed positions was set to 50%. For baseline imaging, whole body scans were mandatory, for iPET, scanning regions were decided based on the findings of the baseline scans. If no compromise was detected below femoral diaphysis, scans were taken from skull base to this point. PET/CT scans were acquired on a hybrid PET/CT scanner equipped with a 64 slice CT detector. CT was taken using a kV of 80-120, 22 index noise, automatic tube current modulation (70 - 180 mA), helical scan with 1 second rotation time, slice thickness of 3.75 mm and a pitch of 0.984. CT images were reconstructed using a filtered back projection (FBP) algorithm.

6.2.4 Image segmentation

Lesions were segmented visually using bPET scans, using a semi-automatic threshold-based method under the supervision of two experienced nuclear medicine physicians. For interim PET/CT, we used baseline CT scans and performed at first a local rigid registration to interim CT scans. After rigid registration, a motion vector field was calculated with optical flow method. This motion vector field was applied to ROIs defined on baseline PET/CT to track lesion in iPET scans. According to Deauville criteria, patients were assigned, evaluating iPET, to five different scores (1 to 5). Mediastinum uptake was quantified with a small ROI inside the aortic arc, which represents the lowest physiologic activity uptake. Liver uptake was taken as the SUV_{mean} from a 4 cm^3 ROI in the right lobe and non-vascular region.

6.2.5 Metabolic markers

From bPET we measured several metabolic variables from ^{18}F-FDG PET scans. For SUV we measured SUV_{max} (maximum voxel SUV value) and SUV_{mean} (mean of all SUV values inside a ROI). We also calculated the MTV as the volume of all voxels (in mL) with a SUV value equal or above a certain threshold of the SUV_{max} inside a ROI. Using a fixed threshold value of $SUV = 2$ we calculated a factor called MTV2, and using a variable threshold of 41% of SUV_{max} we calculated MTV41. Lesion glycolysis (LG) was calculated as the product of MTV and the SUV_{mean} inside a given ROI (in grams). As in the previous case, using MTV2 we calculated LG2 and using MTV41 we calculated LG41. Four markers were used as possible predictors to treatment outcome. Total metabolic tumour volume (TMTV) as the sum of MTV of all lesions in a patient, using the 2 different thresholds mentioned above we obtained TMTV2 and TMTV41. Using LG we calculated total lesion glycolysis (TLG) as the sum of LG in all lesions. Similarly, using 2 different thresholds for MTV, we obtained TLG2 (using MTV2) and TLG41 (using MTV41).

Receiver operating curves (ROC) were used to calculate optimal cutoff thresholds to use these four markers as classifiers of responder and non-responder patients after treatment. Quality of this classifiers, as well as sensitivity and specificity, was assessed by the area under the ROC curve (AUC) and the optimal point in the curve respectively. An AUC of 1 is considered to be the best possible classifier, down to an AUC of 0.5 which is attributed to a random classifier.

All quantifications were done using the software MatLab (version 2016b). Figure 6.1

shows an example of these variables calculated in one of the patients of the cohort.

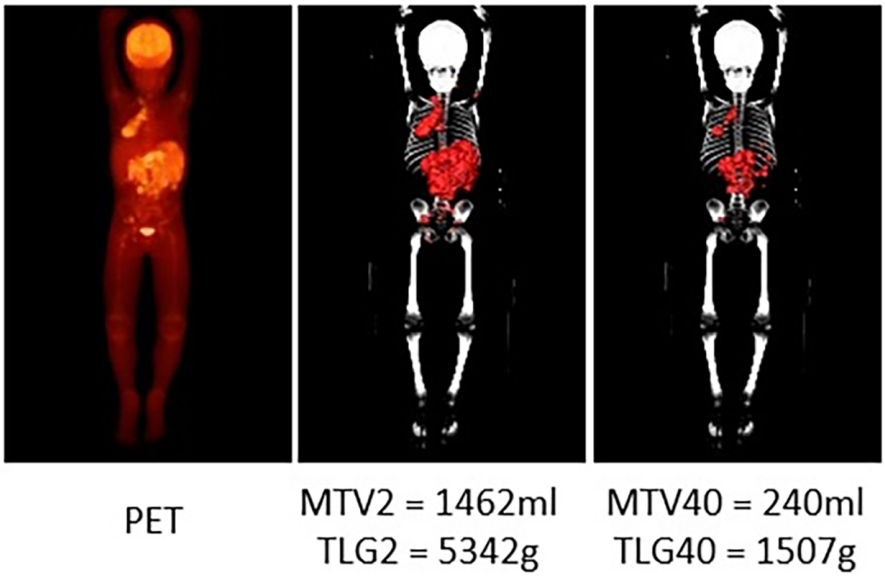

Figure 6.1: Calculation of TMTV and TLG in one of our patients. Middle and right figures show the volumes taken into account for each calculation.

6.2.6 Texture features

To obtain quantitative values for textural features we used 5-patches and parametric maps as described in section 3.6. We calculated 51 different texture features for each lesion, including first, second and higher order features: 8 first order features [6], 23 features from the co-occurrence matrix [7], 11 features from the gray-level-run-length matrix [8], 5 features from the neighboring gray-level matrix [9], 3 features from the neighborhood gray-tone difference matrix [10], and one textural feature designed to quantify heterogeneity as described in [12]. Details for these texture features and methodology for calculations can be found in chapter 3.

6.2.7 Classifier training

Lesions from all patients having an extent of 5 or more voxels in bPET were selected for training supervised classification models. Textural features from lesions measured in bPET were used as classifiers. All features values were standardized for classifier training. Lesions were labelled following two different criteria: patients Deauville score (as obtained from iPET scans) and histopathology. For a single patient, all lesions were assigned the same Deauville score as the overall score of the patient. Accuracy of the trained classifiers was evaluated using a 5-fold cross-validation scheme. This scheme separates data into 5 random groups, one of them is left for testing of the trained classifier, which is trained using the other 4 groups. Accuracy of the classifier is the mean value in percentage, after the before mentioned procedure is repeated five times (once leaving each one of the 5 groups out). ROC curves were used to evaluate discriminator quality using the AUC value.

6.3 Results

For values of all thresholds mentioned below, values above the given thresholds predict a complete responder, otherwise, a non-complete responder corresponds to values below the threshold. For MTV2 we obtained an optimal threshold of 19.0 ml, with a specificity of 50%, sensitivity of 88% and AUC of 0.62. For MTV41 we obtained an optimal threshold of 12.9 ml, with a specificity of 50%, sensitivity of 94% and AUC of 0.66. For TLG2 we obtained an optimal threshold of 54.0 g, with a specificity of 50%, sensitivity of 94% and AUC of 0.63. For TLG41 we obtained an optimal threshold of 44.8 g, with a specificity

of 50%, sensitivity of 94% and AUC of 0.60. ROC curves for these results can be seen in Figure 6.2.

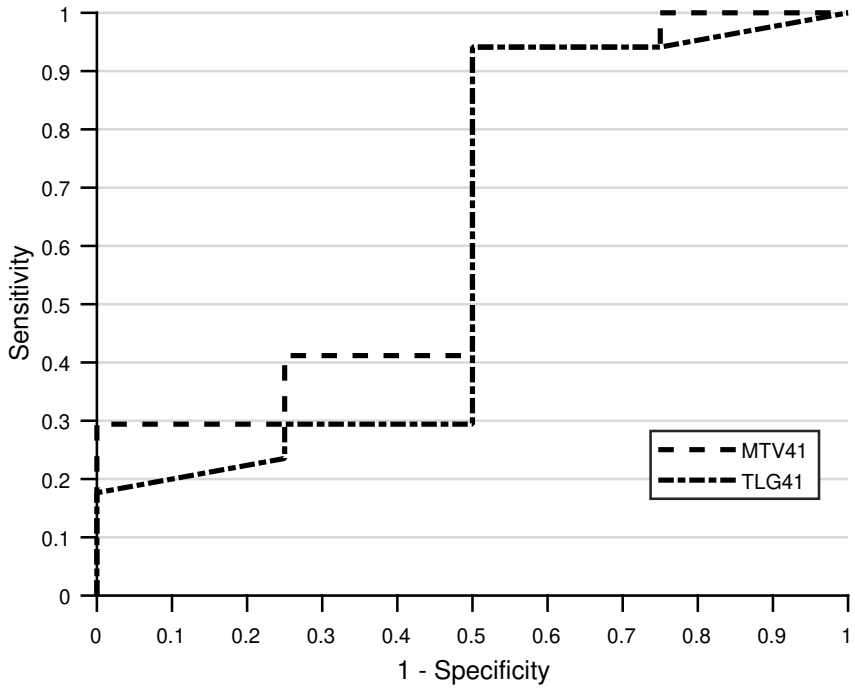

Figure 6.2: ROC curves for MTV41 and TLG41.

At the end of the first-line treatment (QT alone or combined with radiotherapy), 95% of the patients (20 out of 21) were classified as responders.

For histopathology type classification, we selected 120 lesions (57 of mixed cellularity type, 60 of nodular sclerosis type and 3 of lymphocyte rich type) and obtained a 64.2% accuracy using a fine k-nearest neighbours classifier and principal component analysis keeping 8 principal components from the 51 texture features values. True positive rates (TPR) for mixed cellularity, nodular sclerosis and lymphocyte rich are 63%, 68% and 0% respectively. Positive predictive values (PPV) for mixed cellularity and nodular sclerosis are 67% and 65% respectively. The easiest type to discriminate from the others is mixed

cellularity with an AUC of 0.67 for our model. For Deauville scores 1 to 4 (D1-D4) we selected 103 lesions (3 D1, 39 D2, 51 D3 and 10 D4). We obtained a 72.8% accuracy using a linear support vector machine and 8 principal components from the 51 texture features values. TPR for D1-D4 are 0%, 77%, 88% and 0% respectively. PPV for D2 and D3 are 71% and 74% respectively. Using this model, the easiest Deauville score to discriminate from the others was score 3, with an AUC of 0.85.

Results for the trained classifiers performances are shown as confusion matrices in figure 6.3. In our patient sample, after iPET treatment was reduced in 52% of the cases. In

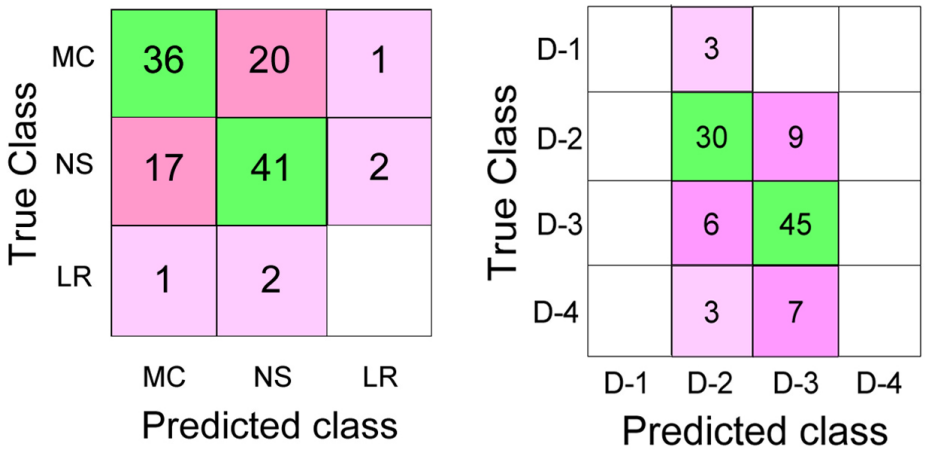

Figure 6.3: Trained classifiers confusions matrices. Left: histopathology classification (MC, mixed cellularity, NS, nodular sclerosis, LR lymphocyte rich). Right: Deauville score classification, 1 through 4.

33% of the cases radiotherapy dose was decreased or not carried out and in 19% of the cases chemotherapy cycles were reduced.

6.4 Discussion

In order to improve response assessment from baseline ^{18}F-FDG PET/CT scans, we measured MTV and TLG using a fixed and an adaptive threshold. Other authors used similar quantitative values such as MTV and TLG with fixed SUV thresholds in pediatric Hodgkin lymphoma [74], and found these do not outperform SUV_{max} in iPET as prognostic factors, from a cohort of 54 patients. Regarding MTV, Kanoun et al. found MTV is a robust PFS prognostic factor in HL adult patients [75]. In our paediatric cohort of HL patients, we found MTV41 and TLG41 to be regular predictors of treatment response. To our knowledge, MTV41 and TLG41 have not been evaluated as predictors of treatment outcome from baseline ^{18}F-FDG FDG-PET scans in paediatric patients with HL. We found that an adaptive threshold of SUV of 41% to measure MTV has a better AUC value compared to the quantification of this parameters using a fixed SUV threshold.

Hoping to further develop a standardized method to characterize paediatric lymphoma lesions using texture features analysis, we tried to train an histopathology type classifiers using 51 different quantifiable texture features. Not all 51 texture features were meaningful for classification. Using principal component analysis (PCA) and keeping 8 principal components was enough to achieve similar precision with classifiers that took all features into account. For histopathology, trained classifiers did not provide optimal accuracy (64,2%), however, we have a small number of lesion samples and a heterogeneous distribution of histopathology types. Moreover, we think PET/CT acquisition parameters could be improved for feature quantifications.

We used the same methodology to train classifiers for lesions which were assigned a Deauville score as described before. In this case, a better accuracy was obtained (72.8%) for all lesions belonging to 4 Deauville rankings. What is more interesting is that the easiest Deauville score to recognize from others is Deauville 3, which is the cutoff point between complete and partial response in the visual scale. We found a PPV of 74% for lesions labeled as corresponding to Deauville 3 with an AUC of 0.85. This is of major relevance because there is some controversy about inter-observer variability when using Deauville score to assess early response in HL. Kluge et al. evaluated Deauville scale variability in a prospective multi-center clinical trial from 1000 pediatric patients iPET scans after two cycles of chemotherapy. They found the biggest variability relies between scores 2 and 3 [76]. There have been previous attempts to quantify the visual Deauville scale to make it less subjective, using a marker called qPET [77]. However, this classification still needs to evaluate iPET, and application of this marker on our patient cohort did not prove to be very accurate. Further validation of these findings, using a larger patient cohort, could constitute a quantitative support for the established Deauville scoring with iPET. Therefore, critical decisions regarding treatment management in paediatric HL patients could benefit from this additional information.

6.5 Conclusion

Total metabolic tumour volume (TMTV) and total lesion glycolysis (TLG) measured in baseline ^{18}F-FDG PET/CT scans, and using a threshold of 41% of SUV_{max}, have been shown to be rather good predictors to treatment outcome in this study. TMTV calculated

using an adaptive SUV_{max} threshold, has a better predictive value than the respective parameter determined using a fixed threshold. Texture features analysis was found to be a promising tool for lesion classification according to visual assessment of images which is prone to inter-observer variabilities.

Conclusions and future directions

Acquiring and analysing medical images for research purposes is a complex process itself. Extracting information from these requires strict protocols for acquisition and standardization of measurements in order to become common practice. What is more important is to find appropriate imaging markers from tomographic molecular imaging that are useful for treatment and prognosis. In the ideal case, we would have a quantity that can be measured from imaging at an early disease stage that could tell us what kind of treatment to give to the patient and/or what is the prognosis, progression free survival in each particular case.

The field of radiomics is advancing in this direction, and textural features have been already proven useful in some particular oncology cases. In this work, we found potentially prognostic markers in the case of head and neck squamous cell carcinoma and Hodgkin lymphoma in paediatric patients. We also did a comparison between textural analysis in two different molecular imaging modalities like PET and SPECT, for the case of lung ventilation imaging.

Three different projects were undertaken during this thesis.

For the lung ventilation project involving patients with COPD, we did a quantitative

comparison between texture features and uptakes in order to find correlations with physiological variables. Several authors have published studies on lung ventilation using nuclear medicine imaging techniques, however, there are no studies concerned with quantification of this images. This image quantification allows to better understand the possibilities of imaging techniques such as PET and SPECT, and allowed to compare the robustness of textural feature quantification across these imaging modalities. Moreover, this quantification tools can be of major relevance as complements of images themselves, therefore, they constitute an aid for diagnosing COPD in this case.

Concerning the retrospective analysis of the HNSCC patient cohort from RTOG protocol 0522, we performed an extraction of imaging markers from ^{18}F-FDG PET images and found several potential prognostic markers for PFS in this type of cancer. These constitute important findings in the field of radiomics applied to head and neck cancer. Imaging markers used as prognostic factors, once standardized, would have great impact in assessing treatment effectiveness and predicting PFS for these patients.

With these same goals on mind, we undertook an analysis of early prognostic imaging markers from ^{18}F-FDG PET images from a patient cohort of paediatric HL. This is a very interesting cohort since there are not so many studies carried out with children. We found out that it is possible to obtain classifier models to differentiate risk groups in this type of paediatric cancer. Since the determination of risk groups is used to decide upon therapy that is going to be administered, standardization of this methods would also have great impact in patient management and healthcare. Also, using these tools, decisions can be made early, as early as at the moment of disease diagnose.

There are a lot of current efforts focusing on texture features analysis and radiomics, as this is a very promising field applying to medical imaging in general. All this work is based on the fact that there is much more information contained in images than what visual inspection is capable of recognizing. This is a growing field, and we hope that ongoing research on radiomics will lead to standardized procedures implemented to clinical practice to improve oncology treatments and diagnose.

In this regard, we think our results are encouraging, and we hope this serves as a motivation to continue the research in the field.

Bibliography

[1] Matthew G. Vander Heiden, Lewis C. Cantley, and Craig B. Thompson. Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation. *Science*, 324(5930):1029–1033, 2009.

[2] F. Fathinul, A. J. Nordin, and W. F. E. Lau. ^{18}[F]FDG-PET/CT is a Useful Molecular Marker in Evaluating Tumor Aggressiveness: A Revised Understanding of an In-Vivo FDG-PET Imaging that Alludes the Alteration of Cancer Biology. *Cell Biochemistry and Biophysics*, 66(1):37–43, 2013.

[3] Martin Fuss. Strategies of assessing and quantifying radiation treatment metabolic tumor response using F18 Positron Emission Tomography (PET). *Acta Oncologica*, 49(7):948–955, 2010.

[4] Xin Gao, Zhong Xue, Jiong Xing, et al. Computer-Assisted Quantitative Evaluation of Therapeutic Response for Lymphoma Using Serial PET/CT Imaging. *Academic Radiology*, 17(4):479–488, 2010.

[5] Paul E. Christian. Technology and Techniques. In Paul E. Christian, Donald Bernier, and James K. Langan, editors, *Nuclear Medicine and PET*, chapter 6. Mosby, St.

Louis, 2004.

[6] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing. 3rd Ed.* Pearson Prentice Hall, 2008.

[7] Robert M. Haralick, K. Shanmugan, and Its'hak Dinstein. Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, 1973.

[8] Xiaoou Tang. Texture Information in Run-Length Matrices. *IEEE Transactions on Image Processing*, 7(11):1602–1609, 1998.

[9] Chengjun Sun and William G. Wee. Neighboring Gray Level Dependence Matrix for Texture Classification. *Computer Vision, Graphics, and Image Processing*, 23:341–352, 1982.

[10] Moses Amadasun and Robert King. Textural Features Corresponding to Textural Properties. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5):1264–1274, 1989.

[11] M. M. Galloway. Texture analysis using grey level run lengths. *NASA STI/Recon Technical Report N*, 75, 1974.

[12] Franck J. Brooks and Perry W. Grigsby. Quantification of heterogeneity observed in medical images. *BMC Medical Imaging*, 13(7):1–12, 2013.

[13] Virendra Kumar, Yuhua Gu, Satrajit Basu, et al. Radiomics: the process and the challenges. *Magnetic Resonance Imaging*, 30(9):1234–1248, 2012.

[14] J.-P. Thirion. Image matching as a diffusion process: an analogy with Maxwell's demons. *Medical Image Analysis*, 2(3):243–260, 1998.

[15] Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. *Artificial Intelligence*, 17(1-3):185–203, 1981.

[16] Lee R. Dice. Measures of the Amount of Ecologic Association Between Species. *Ecology*, 26(3):297–302, 1945.

[17] Deshan Yang, Scott Brame, Issami El Naqa, et al. Technical Note: DIRART - A software suite for deformable image registration and adaptive radiotherapy research. *Medical Physics*, 38(1):67–77, 2011.

[18] Deshan Yand and El Naqa. *DIRART (Deformable Image Registration and Adaptive Radiotherapy) Software Suite User Instruction Manual*. St. Louis, Missouri, 2009.

[19] Florent Tixier, Catherine Cheze Le Rest, Matthiew Hatt, et al. Intratumor Heterogeneity Characterized by Textural Features on Baseline ^{18}F-FDG PET Images Predicts Rresponse to Concomitant Radiochemotherapy in Esophageal Cancer. *The Journal of Nuclear Medicine*, 52:369–378, 2011.

[20] G. Castellano, L. Bonilha, L. M. Li, et al. Texture analysis of medical images. *Clinical Radiology*, 59:1061–1069, 2004.

[21] Fanny Orlhac, Christophe Nioche, Michaël Soussan, et al. Understanding Changes in Tumor Texture Indices in PET: A Comparison Between Visual Assessment and Index

Values in Simulated and Patient Data. *Journal of Nuclear Medicine*, 58(3):387–392, 2017.

[22] J. E. Bresenham. Algorithm for computer control of a digital plotter. *IBM Systems Journal*, 4(1):25–30, 1965.

[23] Marika Baje, Hanna Markstad, Linnea Jarenbäck, et al. Grading obstructive lung disease using tomographic pulmonary scintigraphy in patients with chronic obstructive pulmonary disease (COPD) and long-term smokers. *Annals of Nuclear Medicine*, 29:91–99, 2015.

[24] Geoffrey P. Schembri, Paul J. Roach, Dale L. Bailey, et al. Artifacts and anatomical variants affecting ventilation and perfusion lung imaging. *Seminars in Nuclear Medicine*, 45(5):373–391, 2015.

[25] Jonas Jögi, Björn Jonson, Marie Ekberg, et al. Ventilation-Perfusion SPECT with 99mTc-DTPA Versus Technegas: A Head-to-Head Study in Obstructive and Nonobstructive Disease. *Journal of Nuclear Medicine*, 51:735–741, 2010.

[26] Omar Alonso, Juan Pablo Gambini, Henia Balter, et al. Pet Lung Ventilation Scanning with Gallium-68 Aerosol (Galligas) Versus Technegas SPECT in Patients with Obstructive Lung Disease: A Feasibility Study. *Journal of Nuclear Medicine*, 53(1):2502, 2012.

[27] John Kipritidis, Shankar Siva, Michael S. Hofman, et al. Validating and improving CT ventilation imaging by correlating with ventilation 4D-PET/CT using 68Ga-labeled nanoparticles. *Medical Physics*, 41(1):1–12, 2014.

[28] Michael S. Hofman, Jean Mathiew Beauregard, Thomas W. Barber, et al. 68Ga PET/CT Ventilation perfusion imaging for pulmonary embolism: A pilot study with comparison to conventional scintigraphy. *Journal of Nuclear Medicine*, 52:1513–1519, 2011.

[29] J. Kotzerke, M. Andreeff, G. Wunderlich, et al. Ventilations-Perfusions-Lungen - szintigraphie mit der PET und 68Ga-markierten Radiopharmaka. *Nuklearmedizin*, 6:203–208, 2010.

[30] Clemens Decristoforo. Gallium-68 - A new opportunity for PET available from a long shelf-life generator - Automation and applications. *Current Radiopharmaceuticals*, 5(3):212–220, 2012.

[31] Myrna B. Dolovich and Dal L. Bailey. Positron Emission Tomography (PET) for Assessing Aerosol Deposition of Orally Inhaled Drug Products. *Journal of Aerosol Medicine and Pulmonary Drug Delivery*, 25(1):52–66, 2012.

[32] João Batista Borges, Irina Velikyan, Bengt Langström, et al. Ventilation Distribution Studies Comparing Technegas and "Galligas" Using 68GaCl3 as the Label. *The Journal of nuclear Medicine*, 52(2):206–209, 2011.

[33] Virginia C. Spanoudaki and Sibylle I. Ziegler. *PET and SPECT Instrumentation*. Springer, München, 2008.

[34] Liane Oehme, Klaus Zöphel, Elena Golgor, et al. Quantitative analysis of regional lung ventilation and perfusion PET with 68Ga-labelled tracers. *Nuclear Medicine Communications*, 35(5):501–510, 2014.

[35] Francois Chabat, Yang Guang-Zhong, and David M. Hansell. Obstructive Lung Diseases: Texture Classification for Diffrentiation at CT. *Radiology*, 228:871–877, 2003.

[36] Pierre Yves Le Roux, Shankar Siva, Daniel P. Steinfort, et al. Correlation of 68Gallium ventilation-perfusion PET/CT with pulmonary function test indices for assessing lung function. *Journal of Nuclear Medicine*, 56(11):1718–1723, 2015.

[37] Pierre Yves Le Roux, Shankar Siva, Jason Callahan, et al. Automatic delineation of functional lung volumes with 68Ga-ventilation/perfusion PET/CT. *European Journal of Nuclear Medicine and Molecular Imaging*, 7(82):1–7, 2017.

[38] Ndarajah Vigneswaran and Michelle D. Williams. Epidemiological trends in head and neck cancer and aids in diagnosis. *Oral & Maxillofacial Surgery Clinics of North America*, 26(2):123–141, 2014.

[39] Panayiotis A. Kyzas, Evangelou Evangelos, Despina Denaxa-Kyza, et al. ^{18}f-Fluorodeoxyglucose Positron Emission Tomography to Evaluate Cervical Node Metastases in Patients With Head and Neck Squamous Cell Carcinoma: A Meta-analysis. *Journal of the National Cancer Institute*, 100(10):712–720, 2008.

[40] Gerald Antoch, Nina Saoudi, Hilmar Kuehl, et al. Accuracy of Whole-Body Dual-Modality Fluorine-18-2-Fluoro-2-Deoxy-d-Glucose Positron Emission Tomography and Computed Tomography (FDG-PET/CT) for Tumor Staging in Solid Tumors: Comparison With CT and PET. *Journal of Clinical Oncology*, 22(21):4357–4368, 2004.

[41] Johannes Czernin, Martin Allen-Auerbach, and Heinrich R. Schelbert. Improvements in Cancer Staging with PET/CT: Litterature-Based Evidence as of September 2006. *Journal of Nuclear Medicine*, 48(1):78S–88S, 2007.

[42] K. Facey, I. Brandbury, G. Laking, et al. Overview of the clinical effectiveness of positron emission tomography in selected cancers. *Health Technology Assessment*, 11(44):1–4, 2007.

[43] Christopher Scarfone, William C. Lavely, Anthony J. Cmelak, et al. Prospective Feasibility Trial of Radiotherapy Target Definition for Head and Neck Cancer Using 3-Dimensional PET and CT Imaging. *Journal of Nuclear Medicine*, 45(4):543–552, 2004.

[44] David L. Schwartz, Jonathan Harris, Min Yao, et al. Metabolic Tumor Volume as a Prognostic Image-Based Biomarker for Head-and-Neck Cancer: Pilot results from Radiation Therapy Oncology Group Protocol 0522. *Inernational Journal of Radiation Oncology*, 91(4):721–729, 2015.

[45] Gregory J. Kubicek, Collin Champ, Shannon Fogh, et al. FDG-PET staging and importance of lymph node SUV in head and neck cancer. *Head & Neck Oncology*, 2(19):1–7, 2010.

[46] Huan Yu, Curtis Caldwell, Katherine Mah, et al. Automated Radiation Targeting in Head-and-Neck Cancer Using Region-Based Texture Analysis of PET and CT Images. *International Journal of Radiation Oncology, Biology and Physics*, 75(2):618–625, 2009.

[47] J. V. Raja, M. Khan, V. K. Ramachandra, et al. Texture analysis of CT images in the characterization of oral cancers involving buccal mucosa. *Dentomaxillofacial Radiology*, 41:475–480, 2012.

[48] I. El Naqa, P. W. Grigsby, A. Apte, et al. Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. *Pattern Recognition*, 42(1162-1171), 2009.

[49] Andrew J. Wong, Aasheesh Kanwar, Abdallah S. Mohamed, et al. Radiomics in head and neck cancer: from exploration to application. *Translational Cancer Research*, 5(4):371–382, 2016.

[50] Paulina Galavis. *Robust tumor segmentation for radiotherapy target definition.* PhD thesis, University of Wisconsin-Madison, Madison Wisconsin, 1 2013.

[51] M. David, D. Visvikis, C. Roux, et al. Multi-observation PET image analysis for patient follow-up quantitation and therapy assessment. *Physics in Medicine and Biology*, 56(18):5771–5788, 2011.

[52] Haowei Zhang, Caleb M. Graham, Okan Elci, et al. Locally Advanced Squamous Cell Carcinoma of the Head and Neck: CT Texture and Histogram Analysis Allow Independent Prediction of Overall Survival in Patients Treated with Induction Chemotherapy. *Radiology*, 269(3):801–809, 2013.

[53] Angelina Cistaro, Sergio Palandri, Vito Balsamo, et al. Assessment of a New ^{18}F-FDG PET/CT Protocol in the Staging of Oral Cavity Carcinomas. *Journal of Nuclear Medicine Technology*, 39(1):7–13, 2011.

[54] Christine Mauz-Körholz, Monika L. Metzger, Kara M. Kelly, et al. Pediatric Hodgkin Lymphoma. *Journal of Clinical Oncology*, 33(27):2975–2985, 2015.

[55] Nickhill Bhakta, Ki Liu, Frederick Yeo, et al. Cumulative burden of cardiovascular morbidity in pediatric, adolescent and young adult survivors of Hodgkin's lymphoma: an analysis from the St. Jude Lifetime Cohort Study. *The Lancet Oncology*, 17(9):1325–1334, 2016.

[56] Günther Schellong, Marianne Riepenhausen, Karoline Ehlert, et al. Brest Cancer in Young Women After Treatment for Hodgkin's Disease During Childhood or Adolescence. *Deutsches Ärzteblatt International*, 111(1-2):3–9, 2014.

[57] Smita Bhatia, Yutaka Yasui, Leslie L. Robison, et al. High Risk of Subsequent Neoplasms Continues With Extended Follow-Up of Childhood Hodgkin's Disease: Report From the Late Effects Study Group. *Pediatric Oncology*, 21(23):4386–4394, 2003.

[58] Manuel Gotti, Valeria Fiaccadori, Elisa Bono, et al. Therapy-Related Late Adverse Events in Hodgkin's Lymphoma. *Lymphoma*, 2013(952698):1–7, 2013.

[59] Alden A. Moccia, Jane Donaldson, Mukeshi Chhanabhai, et al. International Prognostic Score in Advanced-Stage Hodgkin's Lymphoma: Altered Utility in the Modern Era. *Journal of Clinical Oncology*, 30(27):3383–3388, 2012.

[60] Andrea Gallamini, Martin Hutchings, Luigi Rigacci, et al. Early Interim 2-[18F]Fluoro-2-Deoxy-D-Glucose Positron Emission Tomography Is Prognostically Superior to International Prognostic Score in Advanced-Stage Hodgkin's Lymphoma: A

Report From Joint Italian-Danish Study. *Journal of Clinical Oncology*, 25(24):3746–3752, 2007.

[61] Cindy L. Schwartz, Lu Chen, Kathleen McCarten, et al. Childhood Hodgkin International Prognostic Score (CHIPS) Predicts event-free survival in Hodgkin Lymphoma: A Report From the Children's Oncology Group. *Pediatric Blood & Cancer*, 64(4):e26278, 2017.

[62] Michal Weiler-Sagie, Olga Bushelev, Ron Epelbaum, et al. 18F-FDG Avidity in Lymphoma Readdressed: A Study of 766 Patients. *The Journal of Nuclear Medicine*, 51(1):25–30, 2010.

[63] Luigi Rigacci, Umberto Vitolo, Luca Nassi, et al. Positron emission tomography in the staging of patients with Hodgkin's lymphoma. A prospective multicentric study by the Intergruppo Italiano Linfomi. *Annals of Hematology*, 86(12):897–903, 2007.

[64] Kevin London, Siobhan Cross, Ella Onikul, et al. ^{18}F-FDG PET/CT in paediatric lymphoma: comparison with conventional imaging. *European Journal of Nuclear Medicine and Molecular Imaging*, 38(2):274–284, 2011.

[65] Sandra Purz, Christine Mauz-Körholz, Dieter Körholz, et al. [^{18}F]Fluorodeoxyglucose positron emission tomography for detection of bone marrow involvement in children and adolescents with Hodgkin's lymphoma. *Journal of Clinical Oncology*, 29(26):3523–3528, 2011.

[66] Christian Furth, Ingo G. Steffen, Holger Amthauer, et al. Early and late therapy response assessment with [18F]fluorodeoxyglucose positron emission tomography in

pediatric Hodgkin's lymphoma: analysis of a prospective multicenter trial. *Journal of Clinical Oncology*, 27(26):4385–4391, 2009.

[67] Bruce D. Cheson, Richard I. Fisher, Sally F. Barrington, et al. Recommendations for Initial Evaluation, Staging, and Response Assessment of Hodgkin and Non-Hodgkin Lymphoma: The Lugano Classification. *Journal of Clinical Oncology*, 32(27):3059–3067, 2014.

[68] Michel Meignan, Andrea Gallamini, and Corinne Haioun. Report on the First International Workshop on interim-PET scan in lymphoma. *Leukemia & Lymphoma*, 50(8):1257–1260, 2009.

[69] Regine Kluge, Lars Kurch, Françoise Montravers, et al. FDG PET/CT in children and adolescents with lymphoma. *Pediatric Radiology*, 43(4):406–417, 2013.

[70] Sameer Bakhshi, Sainath Bhethanabhotla, Rakesh Kumar, et al. Posttreatment PET/CT Rather Than Interim PET/CT Using Deauville Criteria Predicts Outcome in Pediatric Hodgkin Lymphoma: A Prospective Study Comparing PET/CT with Conventional Imaging. *The Journal of Nuclear Medicine*, 58(4):577–583, 2017.

[71] Lara C. V. Harrison, Tiina Luukkaala, Hannu Pertovaara, et al. Non-Hodgkin lymphoma response evaluation with MRI texture classification. *Journal of Experimantal & Clinical Cancer Research*, 28(87):1–13, 2009.

[72] Olcay Sertel, Jun Kong, Umit V. Catalyurek, et al. Histopathological Image Analysis Using Model-Based Intermediate Representation and Color Texture: Follicular Lymphoma Grading. *Journal of Signal Processing Systems*, 55:169–183, 2009.

[73] T. A. Lister, D. Crowther, S. D. Sutcliffe, et al. Report of a Committee Convened To Discuss the Evaluation and Staging of Patients with Hodgkin's Disease: Cotswolds Meeting. *Journal of Clinical Oncology*, 7(11):1630–1636, 1989.

[74] Amr Elsayed M. Hussien, Christian Furth, Stefan Schönberger, et al. FDG-PET Response Prediction in Pediatric Hodgkin´s Lymphoma: Impact of Metabolically Defined Tumor Volumes and Individualized SUV Measurements on the Positive Predictive Value. *Cancers*, 7:287–304, 2015.

[75] Salim Kanoun, Ilan Tal, Alina Berriolo-Riedinger, et al. Methodological Aspects of Total Metabolic Tumor Volume Calculation on Baseline [18F] FDG PET to Predict Survival in Hodgkin Lymphoma. *PLOS One*, 10(10):1–15, 2015.

[76] Regine Kluge, Lidia Chavdarova, Martha Hoffmann, et al. Inter-Reader Reliability of Early FDG-PET/CT Response Assessment Using the Deauville Scale after 2 Cycles of Intensive Chemotherapy (OEPA) in Hodgkin's Lymphoma. *PLOS One*, 11(3):e0149072, 2016.

[77] Dirk Hasenclever, Lars Kurch, Christine Mauz-Körholz, et al. qPET - a quantitative extension of the Deauville scale to assess response in interim FDG-PET scans in lymphoma. *European Journal of Nuclear Medicine and Molecular Imaging*, 41:1301–1308, 2014.