UCL Université catholique de Louvain

École polytechnique de Louvain (EPL)

Mathematical Morphology applied to Radiomics

Dissertation presented by
Nicolas DE VAUCLEROY

for obtaining the Master's degree in
Computer Science and Engineering

Supervisor(s)
Benoit MACQ, Christophe DE VLEESCHOUWER

Reader(s)
Jean LÉGER, Benoit MACQ , Christophe DE VLEESCHOUWER

Academic year 2019

Acknowledgements

Throughout the research and writing of this master I have received a great deal of support and assistance. I would first like to thank Paul Desbordes, whose expertise in Radiomics was invaluable during this research. You provided me with the direction that I needed to successfully complete my thesis.

I would also like to thank my supervisors, Benoit Macq and Christophe De Vleeschouwer for pushing me to my limits and making the submission of this master thesis possible.

I would like to acknowledge the researchers of the imageX lab, working there among these really exceptional people was the best part of this research.

In addition, I would like to thank my friends, who were of great support during the writing.

Finally, thanks to my mother and brother for their wise counsel and comprehensive reading. And special thanks to my mother who took the time to understand the research in order to help me proof-read this master thesis.

Abstract

Personalized medicine is a major and increasing challenge, especially in the diagnostic and treatment of cancer. Traditionally medical imagery was used, but often in a qualitative and subjective way. Radiomics proposes to combine quantitative image features with those extracted from other techniques (clinical, genomic...) to help cancer diagnostic and prediction. The objective of this thesis is to develop new mathematical morphology-based features, and use them for the first time with Radiomics.

We started with listing and analyzing the mathematical foundations of mathematical morphology and the concept of morphological series in order to extract scalar features. These were then compared with classical Radiomic features, namely first order features and gray level matrix based texture descriptors.

These features were implemented and experimented on an open database of 3D Xray-CT segmented non-small cells lung cancer (NSCLC) tumours. A Kaplan-Meier survival analysis was then used in order to select the features that are statistically proven to be relevant. In addition a different machine learning decision tree approach is also tried and the feature selection of the algorithm is compared to the results of the survival analysis.

The comparison of our new mathematical morphology-based features to the classical Radiomics features managed to show that the mathematical morphology based features are at least as good as the classical features extraction techniques.

Our contribution brings to the table new image biomarkers based on mathematical morphology that are statistically proven effective on this database, and could be used for future Radiomics research and personalized medicine.

Contents

Introduction

Personalized medicine for cancer treatment has developed rapidly in recent years thanks to the progress of techniques such as Genomics, which is the analysis of cancerous cell genome. However many of these techniques are invasive, because they necessitate biopsies. Not only are these painful and costly for the patient, but are also an added risk due to eventual complications.

Medical imagery, which gives a spatial representation of the tumour, has existed for a long time, and can be repeated relatively easily. However it is traditionally used qualitatively by physicians, but not with quantitative data. The medical interpretation of the images is usually based on a simple visual appreciation of the contrast, even if it has shown its effectiveness in the management of oncology patients.

Inspired by the Genomics, a similar approach called Radiomics was developed for the analysis of medical imagery, using quantitative data.

Two beliefs underlie the concept of Radiomics. The first considers that characteristics of the medical image (texture features) are strongly correlated with clinical and biological characteristics. The second rational is that the information carried by the image would be complementary to those from other sources of medical information, thus making it possible to enrich the number of characteristics of the tumour.

In order to obtain this quantitative data from imagery, feature extraction image processing techniques are needed and are researched. However, mathematical morphology based feature extraction in radiomics was not yet attempted.

Mathematical morphology is an image processing technique that was developed in the eighties in France and that excels at using the spatial relationship between pixels. It is based on a robust mathematical framework that makes it very successful for various image processing tasks.

The objective of this thesis is to develop mathematical morphology-based algorithms to use in Radiomic image feature extractions. In this master thesis, only texture features are studied.

Our study uses an open medical imagery dataset of 422 patients with non small cell lung cancer (NSCLC). The imagery is done with X-Ray computational tomography with manually segmented regions of interests.

Mathematical morphology was previously used in order to extract information from cancerous cells studied with a microscope[27], however it was never used for texture analysis in radiomics.

Chapter 1 describes in detail the various mathematical morphology techniques, with a focus on mathematics and generalization. During the chapter mathematical morphology is put in perspective with linear signal analysis based techniques. It continues by describing mathematical morphology extended to complete lattice and gray-scale images. Practical usage in image processing are also

described in this chapter, including denoising, Top-hat transform, and morphological gradient. The chapter ends with morphological series which is the main tool for texture feature extraction.

Chapter 2 describes Radionomics, Biometrical imagery (mainly Computational tomography) as well as the classical image biomarkers used in radiomics.

Finally in chapter 3 we combine and compare the morphological based image biomarkers to the classical Radiomics image biomarkers. The comparaison was done using correlation analysis, Kaplan-Meier survival analysis and feature selection through machine-learning.

Chapter 1

Mathematical Morphology

Mathematical Morphology (MM) is an image processing technique based on set and lattice theory. It excels at the exploitation of spatial relationship among pixels. Because it has a strong theoretical mathematical base, it is also used outside of image processing in applications such as *graphs*, *surface meshes* and other domains in which topology and geometry have an importance.

The two biggest uses of mathematical morphology in current research are in segmentation and morphological filtering. Top hat transforms is used to highlight details of images in biomedical applications such as small parasites[34], blood vessels[16]. The watershed in addition to other morphological tools makes mathematical morphology very useful for segmentation[44][15][46][13] is biomedical contexts. Applications of mathematical morphology using morphological filtering and segmentation is also very common in image sensing[55][30][52][53]. However this research is the first to use mathematical morphology for texture analysis in radiomics.

Morphological analysis based on mathematical morphology can be used in any number of dimensions, and is also extended beyond simple binary values into continuous values, allowing the use of MM in grayscale images for example. A grayscale image is an image containing only one channel blabla) imagery).

Compared to other methods classically applied to image processing, mathematical morphology is one of the oldest, as its first implementation dates back to 1964 at the *Ecole des Mines* in Paris by *Georges Matheron* and *Jean Serra* [45]. The original goal was to quantify mineral characteristics (something similar to the goal of this research which is to quantify biomedical characteristics). This led to the development of this theoretical framework that ultimately resulted in the creation of topological concepts such as size, shape, *geodesic distance* (geodesic distance is a graph theory concept that refers to the number of edges in the shortest path between two nodes), etc.

Image processing techniques based on mathematical morphology are *non-linear*, and take into account the morphology of the image, unlike classical linear filtering. The non-linearity and topological aspects of mathematical morphology are its main interest and characteristics.

Finally, among the many use cases of image processing and mathematical morphology, this study will focus more on those deemed useful for the goal of this thesis which is texture analysis, hence why some applications like *skeletonization*, *medial axis transform* (compact representations of objects that keep their shape and topological aspects and can be generated with mathematical morphology) and the *watershed* image segmentation technique will not be studied.

Also, any color related image processing techniques will be ignored as they are not relevant to

biomedical imagery which is purely grayscaled.

1.1. Linear signal/image processing

Signal and system analysis is at the base of many fields in engineering, including image processing.

Definition: A signal is a function of one or multiple variables that conveys information.

$$\textit{Continuous real signal} \quad s : \mathbb{R}^n \to \mathbb{R} \tag{1.1}$$

$$\textit{Discrete real signal} \quad s : \mathbb{Z}^n \to \mathbb{R} \tag{1.2}$$

In image processing, a digital image is defined as a two dimensional discrete signal/system. Imagery itself can be extended beyond two-dimensional images into N-dimensional finite signals, as it is the case for the 3D biomedical imagery used in the third part of this document.

Because an image is a discrete signal, most signal processing techniques are also used in image processing, including *Linear Filters* which are main image processing tools.

Furthermore, we first have to define a filter as a transformation of a signal or image. In other words, a filter is a surjective automorphism

$$F_m : E \to E \tag{1.3}$$

where E is the set of signals, typically $\mathbb{R}^{n \times n}$ for images.

Similarly, a linear filter is a linear application. A filter is linear if it follows the same definition as any linear application:

$$F(A + \lambda B) = F(A) + \lambda F_m(B) \tag{1.4}$$

where F is the filter, A and B are two signals and λ is a coefficient. A filter is thus non linear when this relation breaks.

At the base of image enhancement there is for instance *linear stretch* (linear transform that makes an image use its full dynamic range, widely used as a pre-processing tool), but linear filtering is most useful when using the frequency domain.

1.1.1. Fourier transform

The Fourier transform allows to convert a signal into a sum of different frequencies. Using the intensity and phase of those frequencies (or by using complex numbers that allow to represent both) it is possible with the inverse transform to return to the original signal. As a digital image is discrete, the DFT (discrete Fourier transform) fig. 1.2 is an application that allows to transform an image into an array of frequency intensities and their phase that has the same dimensions as the original image. The intensity spectrum, is one of the main tools of image analysis. It allows to see what frequencies are in the image. This makes it a useful tool for texture analysis, as it allows to see the frequencies of patterns[8]. Frequency spectrum is also much sparser than the image itself, making it a more efficient representation of the image.

There are also other very useful transformations used in image processing, notably the *wavelet transform* and the Discrete Cosine Transform (DCT), both being better alternatives to the Fourier transform in certain applications such as compression.

$$\textit{continuous:} \quad \hat{f}(\xi) = \int_{-\infty}^{\infty} f(x) e^{-2\pi j x \xi} dx \tag{1.5}$$

$$\textit{inverse:} \quad f(x) = \int_{-\infty}^{\infty} \hat{f}(\xi) e^{2\pi j x \xi} d\xi \tag{1.6}$$

$$\textit{m-dimentional:} \quad F(\Omega_1, \Omega_2, \ldots, \Omega_m) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, t_2, \ldots, t_m)\, e^{-j\Omega_1 t_1 - j\Omega_2 t_2 \cdots - j\Omega_m t_m} dt_1 \cdots dt_m \tag{1.7}$$

$$\textit{inverse:} \quad f(n_1, n_2, \ldots, n_m) = \left(\frac{1}{2\pi}\right)^m \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} F(w_1, w_2, \ldots, w_m)\, e^{jw_1 n_1 + jw_2 n_2 + \cdots + jw_m n_m} dw_1 \cdots dw_m \tag{1.8}$$

Figure 1.1.: Fourier transform

$$F(K_1, K_2, \ldots, K_m) = \frac{1}{\sqrt{N_1 \cdots N_m}} \sum_{n_1=0}^{N_1-1} \cdots \sum_{n_m=0}^{N_m-1} f(n_1, n_2, \ldots, n_m)\, e^{-j\frac{2\pi}{N_1} n_1 K_1 - j\frac{2\pi}{N_2} n_2 K_2 \cdots - j\frac{2\pi}{N_m} n_m K_m} \tag{1.9}$$

$$f(n_1, n_2, \ldots, n_m) = \frac{1}{\sqrt{N_1 \cdots N_m}} \sum_{K_1=0}^{N_1-1} \cdots \sum_{K_m=0}^{N_m-1} F(K_1, K_2, \ldots, K_m)\, e^{j\frac{2\pi}{N_1} n_1 K_1 + j\frac{2\pi}{N_2} n_2 K_2 \cdots + j\frac{2\pi}{N_m} n_m} \tag{1.10}$$

Figure 1.2.: discrete Fourier transform (DFT) for discrete multi dimensional arrays (such as images), normalized by $\frac{1}{\sqrt{N_1 \cdots N_m}}$

1.1.2. Convolution theorem

The Fourier transform has useful properties, such as the fact that linear operators have linear equivalents in the reciprocal domain. Linear filtering consists thus of doing operations to the frequency domain, like for instance cutting or enhancing some frequencies.

The **convolution theorem** states that any point-wise multiplication in one domain equates to a convolution in the other domain.

$$\mathcal{F}\{f * g\} = \mathcal{F}\{f\} \cdot \mathcal{F}\{g\} \tag{1.11}$$

$$\mathcal{F}\{f \cdot g\} = \mathcal{F}\{f\} * \mathcal{F}\{g\} \tag{1.12}$$

A convolution is an integral transform defined as follows in the continuous case:

$$(f * g)(t) \int_{-\infty}^{\infty} f(\tau) g(t - \tau) d\tau \tag{1.13}$$

In the discrete case:

$$(f * g)[n] = \sum_{m=-\infty}^{\infty} f[m] g[n - m] \tag{1.14}$$

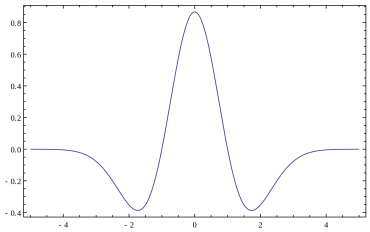


Figure 1.3.: Mexican Hat Filter

A convolution can be seen as the sum/integral of the product of function f and g while the reversed function g slides through it. In practice, this theorem means that any convolution can be implemented efficiently by doing a multiplication in the spectral domain after a fast Fourier transform (FFT, probably the best algorithm of the last century). It means that any frequency based filter can be modelled by a convolution with a matrix that is the impulse response of the filter.

1.1.3. Kernel based image processing

Even if frequency based approaches seem to work on images globally by for instance enhancing or blurring images, convolutions with specific matrices can be used to work on morphological aspects of the image, such as edge, form detection, etc.

In this use case, the matrix is named a *kernel* and it has a striking similarity with the *structuring elements* that are used in mathematical morphology. Their use cases are also similar.

It is worth noting that in computer vision, what is called convolution is in fact *cross-correlation*, a very similar operation:

$$(f(\tau) \star g(\tau))(\tau) = \int_{-\infty}^{\infty} \overline{f(t)} g(t+\tau) dt = [\overline{f(-\tau)} * g(\tau)](\tau) \tag{1.15}$$

$$(f[n] \star g[n])[n] = \sum_{m=-\infty}^{\infty} \overline{f[m]} g[m+n] = [\overline{f[-n]} * g[n]][n] \tag{1.16}$$

in which $\overline{f(t)}$ is the *complex conjugate* (note: this complex conjugate does not mean anything in the context of images usually handled by computer vision, and can thus be ignored in the formula). With cross-correlation, the kernel slides through the image as it is, not as a reversed version, which makes visualisation easier. In most computer vision cases, convolution refers in fact to cross-correlation.

Kernel based convolution is the bread and butter of image processing, allowing to do morphological analysis like edge detection. For texture analysis, convolution with for instance a *Mexican hat kernel* (in reference to the form of the kernel used to convolve), *difference of Gaussians* (tunable filter achieving an effect similar to the L-o-G) or *Laplacian of Gaussian* (derivative operator) allows to analyze details of textures at different scales[41].

1.2. Interest of non-linear image processing methods

$$F_m(A + \lambda B) = F_m(A) + \lambda F_m(B) \tag{1.17}$$

Where F_m is a filter, A and B are two signals and λ a scalar. A filter is said to be non-linear if it doesn't satisfy this relation.

A direct consequence of this is that if we consider A to be an original image and B to be an added noise, some traces of B will inevitably end up in the filtered image . A non-linear filter does not suffer from this limitation.

Moreover, a linear filter cannot be strictly morphological, as the λ will make a linear relationship between the intensity of the input and the intensity of the output. Even kernel based convolutions that have morphological aspects to them are inherently intensity-based due to their linearity.

Finally a succession of convolutions will always be a convolution as $f * g * h = f * (g * h)$ and will thus be always a linear convolution.

There is an infinite amount of non-linear filters, and unlike linear filters they cannot be all characterized by their impulse response.

Basic popular non-linear filters include non-linear stretching (also called levels), thresholding[40] which handle every pixel separately in a non linear way.

Another very popular filter is *median filtering* (which replaces each pixel with the median of its neighbours; the linear equivalent would be to use the mean, that can be simply implemented using a flat kernel), which is a very strong anti-noise non-linear pre-processing filter).

However the main focus of this section will be the **morphological filters** based around *mathematical morphology*.

1.3. Mathematical morphology theory

As *Jean Serra*, the co-inventor of mathematical morphology, explains in his course at *Ecole des Mines Paristech* which serves as the main reference of this section, the concept known today as mathematical morphology was not created in 1981 but rather rediscovered. The concept was first laid out by Hermann Minkowski in 1901, in which he defined *set dilatation* and *Minkowski addition* which is the geometric equivalent. Jean Serra began working on the properties of binary dilatation as a function of geometry in the early 70's.

After being defined for binary sets and binary images, then continuous functions and grayscale images, the subsequent generalization to *lattice theory* is the mathematical morphology theoretical foundation widely accepted today.

In this chapter, the basic operations will first be defined for binary sets and Euclidean geometry, then later extended to complete lattice and grayscale images.

1.4. Mathematical morphology in binary sets and Euclidean geometry

There is both an Euclidean geometrical theory (*Minkowski addition* and *difference*) and a set theory explanation behind binary mathematical morphology.

The Minkowski addition and set dilatation were defined by Hermann Minkowski in 1901.

Let E be a set. We define the following objects.

- The subset $\mathrm{X} \subseteq \mathrm{E}$
- To all $\mathrm{x} \in \mathrm{E}$ are linked two probes $A(x)$ and $B(x)$, those are two automorphisms $\forall z \in E : A(z) \mapsto \mathrm{E}$ and $\forall z \in E : B(z) \mapsto \mathrm{E}$. These are referred to as *structuring elements.*

For the Euclidean geometry aspect:

- set E is equipped with a translation τ

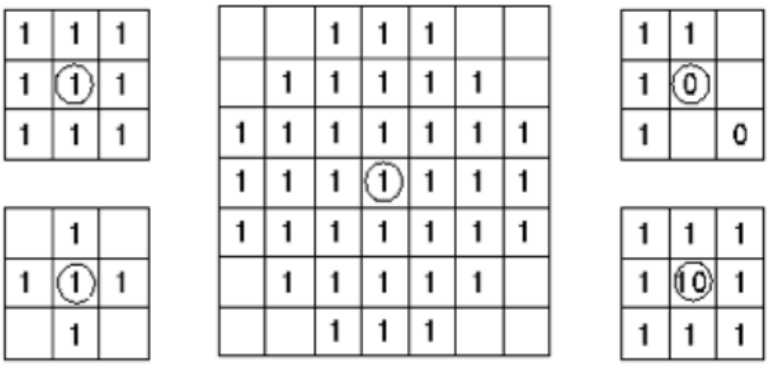


Figure 1.4.: Structuring Element

- set X_b is the translation of X according to the vector b

$$X_b = \{x + b, x \in X\} \tag{1.18}$$

- $\overset{v}{X}$ is the *transposed* or *reflected* of X

$$\overset{v}{X} = \{-x, x \in X\} \tag{1.19}$$

The structuring element can also be transposed

$$x \in B_z \quad \Leftrightarrow \quad z - x \in B \tag{1.20}$$

A structuring element is *symmetrical* when it is equal to its transposed.

It is worth noting that here the transposed is not defined like the transposition of a matrix, but rather as a $180°$ rotation, or a central symmetry.

1.4.1. Structuring element

The set theory definition given previously might not be the easiest to visualize. In the case of a binary image, a structuring element will be similar to a convolution kernel, an array with binary values to it. Like in convolution, the center of this array will be translated to where we are currently processing the image.

In the case of the Euclidean geometry case, the structuring element will be an ensemble of vectors, for instance a geometric form like a square.

1.4.2. Erosion

Unlike the set dilatation, set erosion was not conceptualized by Minkowski, but later by Hadwiger in 1957. Erosion and dilatation are linked and dual to each other.

The *set erosion* is defined as follows:

$$\varepsilon_B(X) = \{z : B(z) \subseteq X\} \tag{1.21}$$

Figure 1.5.: Erosion

For an erosion of the set X using the structuring element B, the result consists of all $z \in X$ such as the whole structuring element for element $z \rightarrow B(z)$ fits the ensemble X.

A geometric analogy would be that the result of the erosion is all the locations in which the structuring element fits inside X.

Because of this definition, the resulting $\varepsilon_{\mathrm{B}}(\mathrm{X})$ is necessarily part of the ensemble X such as $\varepsilon_{\mathrm{B}}(\mathrm{X}) \subseteq X$. The result is thus a **smaller or equal ensemble** to the original ensemble.

Erosion is **commutative** under the *set intersection* operator :

$$\begin{aligned} \varepsilon_{\mathrm{B}}\left(\mathrm{X}_0 \cap \mathrm{X}_1\right) &= \{\mathrm{z} : \mathrm{B}(\mathrm{z}) \subseteq \mathrm{X}_0 \cap \mathrm{X}_1\} && (1.22)\\ &= \{\mathrm{z} : \mathrm{B}(\mathrm{z}) \subseteq \mathrm{X}_0\} \cap \{\mathrm{z} : \mathrm{B}(\mathrm{z}) \subseteq \mathrm{X}_1\} && (1.23)\\ &= \varepsilon_{\mathrm{B}}\left(\mathrm{X}_0\right) \cap \varepsilon_{\mathrm{B}}\left(\mathrm{X}_1\right) && (1.24) \end{aligned}$$

This property is what makes ε_{B} effectively an *erosion* (by definition).

Minkowski subtraction

$$X \ominus B = \{z \in E | B_z \subseteq X\} \tag{1.25}$$

Here B_z is defined as $\mathrm{x} \in \mathrm{B_z} \Leftrightarrow \mathrm{z} - \mathrm{x} \in \mathrm{B}$, as previously stated. This geometric reasoning is more intuitive, especially when we introduce the *Minkowski addition.*

Properties of erosion / Minkowsi subtraction

- **Non-commutativity**

$$A \ominus B \neq B \ominus A \tag{1.26}$$

- **Anti-extensivity**

$$0 \in B \implies A \ominus B \subset A \tag{1.27}$$

- **Increasing**

$$\begin{aligned} A \subseteq C &\implies A \ominus B \subseteq C \ominus B\\ B \supseteq C &\implies A \ominus B \subseteq A \ominus C \end{aligned} \tag{1.28}$$

- **Distributive over set intersection**

$$(A \cap B) \ominus C = (A \ominus C) \cap (B \ominus C) \tag{1.29}$$

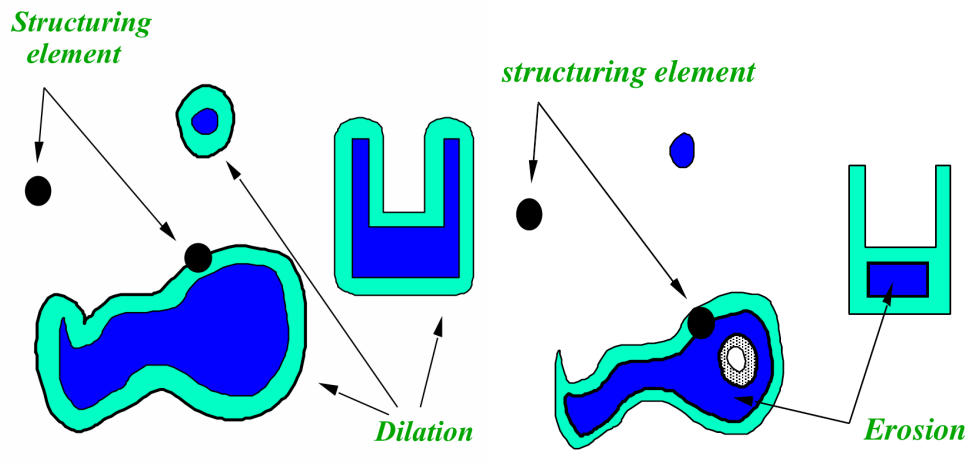


Figure 1.6.: Euclidian geometry: Minkowski subtraction, courtesy of J. Serra's Mathematical Morphology course

1.4.3. Dilatation

From these following relations:

$$\mathrm{X} \subseteq \varepsilon_{\mathrm{B}}(\mathrm{Y}) \Leftrightarrow \{\mathrm{x} \in \mathrm{X} \Rightarrow \mathrm{B}(\mathrm{x}) \subseteq \mathrm{Y}\} \tag{1.30}$$

$$\Leftrightarrow \cup\{\mathrm{B}(\mathrm{x}), \mathrm{x} \in \mathrm{X}\} \subseteq \mathrm{Y} \tag{1.31}$$

$$\Leftrightarrow \delta_{\mathrm{B}}(\mathrm{X}) \subseteq \mathrm{Y} \tag{1.32}$$

Where because X is included in the set erosion of Y ($\mathrm{X} \subseteq \varepsilon_{\mathrm{B}}(\mathrm{Y})$), it is equivalent to saying that x being an element of X as in $\mathrm{x} \in \mathrm{X}$ implies that the structuring element B in x necessarily fits inside Y (as by the definition of set erosion: $\varepsilon_{\mathrm{B}}(\mathrm{Y}) = \{\mathrm{z} : \mathrm{B}(\mathrm{z}) \subseteq \mathrm{Y}\}$) ; which is also equivalent to saying the *union* of all structuring elements in x element of X necessarily is part of Y ($\cup\{\mathrm{B}(\mathrm{x}), \mathrm{x} \in \mathrm{X}\} \subseteq \mathrm{Y}$)

Those relations are very important, as they define the operator of *set dilatation* defined as :

$$\delta_{\mathrm{B}}(\mathrm{X}) = \cup\{\mathrm{B}(\mathrm{x}), \mathrm{x} \in \mathrm{X}\} \subseteq \mathrm{Y} \tag{1.33}$$

Because of the *set union commutativity*, the $\delta_{\mathrm{B}}(\mathrm{X})$ set operator is a dilatation as:

$$\delta_{\mathrm{B}}\left(\mathrm{X}_0 \cup \mathrm{X}_1\right) = \delta_{\mathrm{B}}\left(\mathrm{X}_0\right) \cup \delta_{\mathrm{B}}\left(\mathrm{X}_1\right) \tag{1.34}$$

But they define set erosion and set dilatation to be *ajoint* and make a erosion-dilatation *doublet*, as the previous sequence of relations proves that:

$$\mathrm{X} \subseteq \varepsilon(\mathrm{Y}) \Leftrightarrow \boldsymbol{\delta}(\mathrm{X}) \subseteq \mathrm{Y} \tag{1.35}$$

In the **adjunction theorem** [24], when two operators are linked by the equivalence:

$$\mathrm{X} \subseteq \varepsilon(\mathrm{Y}) \Leftrightarrow \delta(\mathrm{X}) \subseteq \mathrm{Y} \tag{1.36}$$

They are necessarily an *erosion-dilatation doublet*, as per this proof:

If we have $Y_i, i \in I$ and X such as:

$$\delta(\mathrm{X}) \subseteq \cap_{i\in I}\mathrm{Y_i} \quad \Leftrightarrow \quad \delta(\mathrm{X}) \subseteq \mathrm{Y_i} \qquad \forall i \in I \tag{1.37}$$

$$\delta(\mathrm{X}) \subseteq \cap_{i\in I}\mathrm{Y_i} \quad \Leftrightarrow \delta(\mathrm{X}) \subseteq \cap\mathrm{Y_i} \tag{1.38}$$

$$\delta(\mathrm{X}) \subseteq \mathrm{Y_i} \quad \Leftrightarrow \mathrm{X} \subseteq \varepsilon\left(\mathrm{Y_j}\right), \mathrm{i} \in \mathrm{I} \quad \Leftrightarrow \mathrm{X} \subseteq \cap\varepsilon\left(\mathrm{Y_i}\right) \tag{1.39}$$

$$\mathrm{X} \subseteq \varepsilon\left(\cap Y_i\right) \quad \Leftrightarrow \mathrm{X} \subseteq \cap\varepsilon\left(Y_i\right) \tag{1.40}$$

Which implies that $\varepsilon\left(\cap Y_i\right) = \cap\varepsilon\left(Y_i\right)$, meaning that by definition ε is an erosion (analog reasoning for the dilatation)

The **composition product** of two dilatations (respectively for an erosion) is also a dilatation:

$$\delta_{\mathrm{B2}}\delta_{\mathrm{BI}}(\mathrm{X}) = \cup\left\{\mathrm{B_2(y)}, \mathrm{y} \in \cup\left\{\mathrm{B_1(x)}, \mathrm{x} \in \mathrm{X}\right\} = \cup\left\{\delta_{\mathrm{B2}}\left[\mathrm{B_1(x)}\right], \mathrm{x} \in \mathrm{X}\right\}\right. \tag{1.41}$$

$$\delta_{\mathrm{B2}}\delta_{\mathrm{Bl}} = \delta_{\mathrm{A}}, \quad \mathrm{A} = \delta_{\mathrm{B2}}\left(\mathrm{B_1}\right) \tag{1.42}$$

Minkowski addition

The Euclidean geometry *Minkowski addition* is defined as follow:

$$\begin{aligned} \mathrm{X} \oplus \mathrm{B} &= \cup\left\{\mathrm{B_x}, \mathrm{x} \in \mathrm{X}\right\} \\ &= \cup\{\mathrm{x} + \mathrm{b}, \mathrm{x} \in \mathrm{X}, \mathrm{b} \in \mathrm{B}\} \\ &= \cup\{\mathrm{X}, \mathrm{b} \in \mathrm{B}\} = \mathrm{B} \oplus \mathrm{X} \end{aligned} \tag{1.43}$$

Figure 1.7.: Dilatation

The geometrical intuition is quite simple: just place the structuring element on elements of X, and it will make the image grow.

Minkowski addition and difference also have this duality :

$$\mathrm{X} \subseteq \mathrm{Y} \ominus \mathrm{B} \quad \Leftrightarrow \quad \mathrm{X} \oplus \mathrm{B} \subseteq \mathrm{Y} \quad \mathrm{X}, \mathrm{Y} \in \mathrm{E} \tag{1.44}$$

However the most interesting relation between erosion and dilatation comes with the use of the *complement* and *transpose* operator:

$$\varepsilon_B(X) = X \ominus B = \{z : B_z \subseteq X\} \tag{1.45}$$
$$B_z \subseteq X \Leftrightarrow \forall b \in B : b + z \in X \tag{1.46}$$
$$\Leftrightarrow \forall b \in B : z \in X_{-b} \tag{1.47}$$
$$\varepsilon_B(X) = X \ominus B = \cap \left\{ X_b, b \in \overset{v}{B} \right\} \tag{1.48}$$

Using this new expression for the erosion, we can prove that:

$$X \ominus B = \cap \left\{ X_b, b \in \overset{v}{B} \right\} \tag{1.49}$$
$$= \left(\left(\cap \left\{ X_b, b \in \overset{v}{B} \right\} \right)^c \right)^c \tag{1.50}$$
$$= \left(\cup \left\{ X_b^c, b \in \overset{v}{B} \right\} \right)^c \tag{1.51}$$
$$X \ominus B = \left(X^c \oplus \overset{v}{B} \right)^c \tag{1.52}$$

Which proves what seems intuitive and can be seen empirically: that it is possible to do an erosion using only dilatation and a dilatation using only erosion. However it is necessary to use the complement of the ensemble which, in the case of binary and grayscale images, consists of inverting the image and to use the transpose of the structuring element (a rotation of $180°$).

Properties of dilatation/Minkowski addition

- **Commutative**:

$$A \oplus B = B \oplus A \tag{1.53}$$

- **Associativity**

$$A \oplus (B \oplus C) = (A \oplus B) \oplus C \tag{1.54}$$

- **Extensivity**

$$A \subseteq A \oplus B \tag{1.55}$$

- **Increasing**

$$A \subseteq B \implies A \oplus D \subseteq B \oplus D \tag{1.56}$$

1.4.4. Link with kernel based convolution

However it is worth noting that there is a clear analogy between *kernel based convolution* and mathematical morphology. As the analogy was noted in[47]: "The structuring element is to mathematical morphology what the convolutional kernel is to linear filter theory". In this paper *Kevin Locquin Convolutional filtering* proves that kernel based convolutional filtering is an upper-lower bound to the mathematical morphology operations of dilatation and erosion.

More intuitively, adding a non-linear filter (for instance thresholding) to the result of a kernel based convolutional filtering can get results very similar to mathematical morphology.

In fact, binary erosion can be implemented by adding thresholding to a kernel based convolution, which is a technique used in some implementations of mathematical morphology[29][42] and allows to leverage the great speed that convolution operations have on *GPU*s:

$$A \ominus B = \tau_{|B|}(A \star B) \tag{1.57}$$

Different techniques exist to perform mathematical morphology, here is a comparison[38].

Where A and B are binary arrays/images, $|B|$ the *cardinality* (number of 1 in the element) and τ is the thresholding operation.

Furthermore, adding a non-linearity to a kernel convoluted image is exactly what *convolutional neural networks* do, and can be seen as the contemporary evolution of mathematical morphology.

1.4.5. Opening and closing

Many different ensembles can have the same opening or closing. However *there is a single one* that is smaller/bigger than others. These are called *opening/closing*, and they are obtained by applying erosion then dilatation / dilatation then erosion.

- **Adjunction opening**

$$\begin{gathered} \gamma_{\mathrm{B}} = \delta_{\mathrm{B}}\varepsilon_{\mathrm{B}} \\ \mathrm{X} \circ \mathrm{B} = [(\mathrm{X} \ominus \mathrm{B}) \oplus \mathrm{B}] \end{gathered} \tag{1.58}$$

- **Adjunction closing**

$$\begin{gathered} \varphi_{\mathrm{B}} = \varepsilon_{\mathrm{B}}\delta_{\mathrm{B}} \\ \mathrm{X} \bullet \mathrm{B} = [\mathrm{X} \oplus \mathrm{B}) \ominus \mathrm{B}] \end{gathered} \tag{1.59}$$

Idempotence: By adding $\mathrm{Y} = \delta_{\mathrm{B}}(\mathrm{X})$ and $\mathrm{X} = \varepsilon_{\mathrm{B}}(\mathrm{Y})$ to the relation $\delta_{\mathrm{B}}(\mathrm{X}) \subseteq \mathrm{Y} \Leftrightarrow \mathrm{X} \subseteq \varepsilon_{\mathrm{B}}(\mathrm{Y})$, we obtain:

$$\delta_{\mathrm{B}}(\mathrm{X}) \subseteq \mathrm{Y} \Leftrightarrow \mathrm{X} \subseteq \varepsilon_{\mathrm{B}}(\mathrm{Y}) \tag{1.60}$$
$$\delta_{\mathrm{B}}(\varepsilon_{\mathrm{B}}(\mathrm{Y})) \subseteq \mathrm{Y} \Leftrightarrow \mathrm{X} \subseteq \varepsilon_{\mathrm{B}}(\delta_{\mathrm{B}}(\mathrm{X})) \tag{1.61}$$
$$\delta_{\mathrm{B}}(\varepsilon_{\mathrm{B}}(\mathrm{X})) \subseteq \mathrm{X} \Leftrightarrow \mathrm{X} \subseteq \varepsilon_{\mathrm{B}}(\delta_{\mathrm{B}}(\mathrm{X})) \tag{1.62}$$
$$\delta_{\mathrm{B}}\varepsilon_{\mathrm{B}}(\mathrm{X}) \subseteq \mathrm{X} \subseteq \varepsilon_{\mathrm{B}}\delta_{\mathrm{B}}(\mathrm{X}) \tag{1.63}$$
$$\varepsilon_{\mathrm{B}}\left(\delta_{\mathrm{B}}\varepsilon_{\mathrm{B}}(\mathrm{X})\right) \subseteq \varepsilon_{\mathrm{B}}(\mathrm{X}) \subseteq \left(\varepsilon_{\mathrm{B}}\delta_{\mathrm{B}}\right)\varepsilon_{\mathrm{B}}(\mathrm{X}) \tag{1.64}$$
$$\Longrightarrow \ \varepsilon_{\mathrm{B}}\delta_{\mathrm{B}}\varepsilon_{\mathrm{B}} = \varepsilon_{\mathrm{B}} \tag{1.65}$$

Here is how the *idempotence* of the opening (and reciprocally closing) operation is proven. Idempotence means that the result does not change after applying multiple times the operations. So an opening on an opening will not change.

Figure 1.8.: Opening

Figure 1.9.: Closing

Geometrical interpretation

The opening is the *union* of all the structuring elements $B(X)$ that are included in set X, $\gamma_{\mathrm{B}}(\mathrm{X})$ is necessarily smaller that X, and that all parts of X in which the structuring element B does not fit will vanish with the opening operation. The opening will also have an effect on the edges depending on the form of the structuring element(a disk shaped one for instance will round the edges).

Closing can be seen as the complement operation of opening, using the relation between erosion and dilatation demonstrated in the previous sections, we can see that:

$$X \bullet B = \left(X^c \circ \overset{v}{B}\right)^c \tag{1.66}$$

Which means that gaps in X smaller than the structuring element will be filled.

Algebraic opening and closing

Morphological opening and closing are particular cases of *algebraic* opening and closing

Definition :

- An *increasing*, *anti-extensive* and *idempotent* operation is called an (algebraic) **opening**
- An *increasing*, *extensive*, and *idempotent* operation is called an (algebraic) **closing**

Properties

- Any supremum of openings is still an opening
- Any infimum of closing is still a closing
- Idempotence

- Antiextensivity of opening:

$$A \circ B \subseteq A \tag{1.67}$$

- Extensivity of closing:

$$A \subseteq A \cdot B \tag{1.68}$$

1.5. Mathematical morphology extended to complete lattice and grayscale Images

A *complete lattice* is a *partially ordered set* $(L, \leq)$ in which every subset $A \subset L$ has both a greatest lower bound (*infimum*) and a least upper bound (*supremum*).

A *complete lattice* is thus defined from those three elements:

- A partially ordered set $(L, \leq)$
- An infimum $\bigwedge$
- A supremum $\bigvee$

In the complete lattice extension of mathematical morphology:

- A **dilatation** is any *automorphic* operator $\delta : L \to L$ that distributes over the *supremum* and preserves the least element:
  - $\bigvee_i \delta(X_i) = \delta(\bigvee_i X_i)$
  - $\delta(\emptyset) = \emptyset$
- An **erosion** is any *automorphic* operator $\varepsilon : L \to L$ that distributes over the *infimum* and preserves the universe(max element):
  - $\bigwedge_i \varepsilon(X_i) = \varepsilon(\bigwedge_i X_i)$
  - $\varepsilon(U) = U$

  All the properties cited above still hold, including the *Galois connection* that binds to every erosion its doublet dilatation:

$$X \leq \varepsilon(Y) \Leftrightarrow \delta(X) \leq Y \tag{1.69}$$
$$\forall X, Y \in L \tag{1.70}$$

1.5.1. Link with set theory/binary images definition

Even if the definitions look different, definitions for complete lattices also apply to the binary morphology case. L is the *power set* (set of all subsets) of E, $\leq$ is the set inclusion. The infimum then corresponds to the *set intersection* and the supremum to the *set union.*

Complete lattices extend mathematical morphology to a lot of different applications, even functions can benefit from mathematical morphology, transforming them in ways that simple derivatives and filters cannot. Grayscale images are also concerned by this extension.

1.5.2. Grayscale images

Grayscale images are complete lattices, ensembles of pixels that each have a minimum possible value (usually 0, the black color) and a maximum possible value (in the case of 8-bit images, 255 the white color). In this concrete case of 8-bit images, images are functions $f : N^2 \mapsto [0, 255]$ where 0 is the infimum and 255 is the supremum. A flat structuring element $b(x)$ would be defined as follows:

$$b(x) = \begin{cases} 0, & x \in B \\ -255, & \text{otherwise} \end{cases} \tag{1.71}$$

Dilatation

Dilatation swipes the structuring element across the image, and each pixels is replaced with the highest pixel value that fits in the structuring element. It thus returns for each pixel the maximum value within a moving window. Dilatation is thus a particular case of *order statistics*, which makes it very similar to the *median filter*. As it always takes the maximum value within a zone, dilatation will make brighter zones grow bigger, increasing overall the image's luminosity.

$$(f \oplus b)(x) = \bigvee_{y \in E} [f(y) + b(x - y)] \tag{1.72}$$

Erosion

The working is the same as for dilatation and the median filter, except that instead of taking the maximum and the median, it takes the minimum value within the moving window, making the dark zones bigger and the image overall darker.

$$(f \ominus b)(x) = \bigwedge_{y \in B} [f(x + y) - b(y)] \tag{1.73}$$

Opening and closing

In grayscale mathematical morphology, morphological opening and closing are still defined the same way as in the binary images, using a dilatation on an erosion in the case of the opening, and closing of a dilatation in the case of morphological closing.

However other types of algebraic opening and closing can be defined as long as they obey to the properties of algebraic openings and closing defined in a previous section.

1.5.3. Padding

When applying operations like convolution and mathematical morphology in image processing, an important part is to handle border effects. Indeed, near the border both the kernel in the case of convolution, and the structuring element in the case of mathematical morphology will be partly out of the image, and the question is what value to attribute to the pixels that are outside of the image.

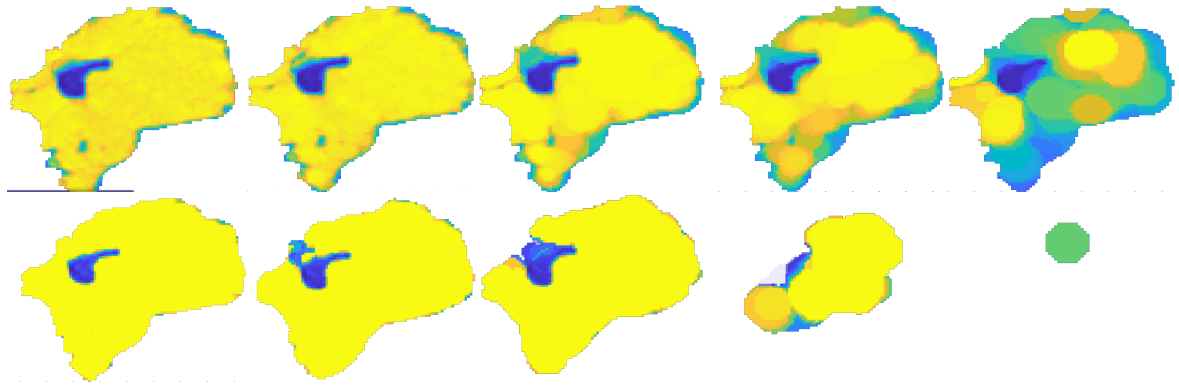


Figure 1.10.: Effect of successive opening with increasing spherical structuring element with and without padding on LUNG1

Convolution

In the case of convolution, different strategies exist when handling edges, and they are particularly relevant in convolutional neural networks(CNN) where convolutions succeed each other, with non-linear functions called activation functions being applied in between them.

Those strategies are:

- **Crop**: Any pixels concerned by those border effects are ignored. The resulting image is thus cropped and smaller than the original one.
- **Kernel-Crop**: The kernel is cropped when it comes close to the edges, the kernel is normalized (divided by the sum of the elements of the matrix) to compensate.
- **Zero padding**: All pixels outside of the image are considered to be equal to zero.
- **Wrap/mirror**: The image is tiled (repeated) after the edges in the case of "Wrapping" and an axial symmetry in the case of "mirror" padding. For texture analysis requiring convolution, using mirror padding usually makes the most sense.
- **Extend**: The pixels on the border are repeated beyond the image.

All of those strategies work for different cases, but in the case of texture analysis, kernel-crop and mirror padding are the most useful as they attenuate the border effects.

Mathematical Morphology

Padding in mathematical Morphology is something that is not addressed in the literature, but it is very important in the case of texture analysis because we want the shape of an object to be totally ignored (in order to focus on the texture itself), which is not the natural behaviour of an operation called mathematical *morphology*.

Padding is somehow implicit in the set theory definitions of mathematical morphology, however practical implementations implement it specifically. The result is the same as the one of the *kernel crop* padding of kernel based convolution. It could in fact be implemented using a "*structuring element crop*", but the properties of mathematical morphology allow using padding values:

- **Erosion**: as erosion $\bigwedge_i \varepsilon(X_i) = \varepsilon(\bigwedge_i X_i)$ preserves the least element, the structuring element will ignore any pixel with the *maximum value*/*supremum*, padding can effectively be done by prolonging the image with maximum value pixels.
- **Dilatation**: as dilatation $\bigvee_i \delta(X_i) = \delta(\bigvee_i X_i)$ preserves the biggest element, putting all the padding values to the *minimum*/*infimum* implements padding.

- **Opening and closing**: padding values used are different for the dilatation and erosion part.

Mathematical morphology implementations do not allow to use a mask to ignore parts of the image, but it can be done by replacing the part of the image to ignore by padding values, effectively suppressing any effect related to the shape of the mask. This is useful when analyzing the texture of a non rectangular shaped object, like it is the case of tumours studied in the third part of this document, which are manually segmented by a surgeon.

Mathematical morphology is very dependent on shape, as its name indicates. In first trials, the morphological series had undesired effects that depended on the shape of their ROI. However, a kind of padding, akin to kernel-crop in convolution is here used to avoid this. This padding makes it so that border effects are totally ignored and the results depend purely on the textures.

A concrete example is showed in 1.10, a snapshot of the experimental part where the problem first appeared, as whithout padding morphological opening would make the size of the tumour have more impact than the texture.

1.6. Practical usage of mathematical morphology in image processing

1.6.1. Denoising

In this case, the use of mathematical morphology is very similar to the median non linear filter. As shown in previous sections, erosion and dilatation take a specific order of statistics of the sample that fits within the structuring element, the minimum (first order statistic of the sample defined by the part of the image that fits in the structuring element) in the case of erosion, the median ($\frac{k}{2}$-order statistic, with k being the size of the sample, the carnality of the structuring element) in the case of the media filter, and the maximum(k-order statistic of the sample) in the case of the dilatation.

Because of that, these three transformations will be effective against different types of noise:

- *Any type of noise* for the median transform, hence its popularity.
- *Pepper noise* (noise consisting of dark dots) for dilatation, as the maximum operator only selects the brightest pixels
- *Salt noise* (noise consisting of light dots) for erosion, as the minimum operator selects the darkest pixels.

However, using dilatation brightens the image by growing the bright spots, and erosion darkens the image by growing the dark spots, leading to an image degradation. This is why *closing* is used to suppress pepper noise, and *opening* to suppress salt noise. With the complementary operation done after the first one, the image is reconstructed, while removing the elements smaller than the structuring element. Opening has the effect of removing bright elements smaller than the structuring element, and closing has the opposite effect of removing dark elements smaller than the structuring element.

When removing salt or pepper noise, the non linear aspect of these transforms is really useful: if salt noise was removed using a linear low pass filter, the DC component would have ended up in the final image making it brighter than it should be. However, if using the non linear opening, salt noise will vanish, and the image reconstructed with little degradation (I personally used it to remove hair of my badly shaven skin in a close up picture).

The useful property of algebraic openings *(resp. closings)* states that any *supremum (resp. infimum)* of openings *(resp. closings)* is still an opening *(reps. closing)*. This allows to use various structuring

elements to filter out elements of different types of shapes, then use a point-wise supremum (maximum) operator to combine those openings into another opening, which is still an algebraic opening of the original image.

This use of openings highlight *slow trends* of an image, namely components in which the structuring elements fit. The usage is thus analogue to a low pass filter.

1.6.2. Opening by reconstruction

In the definition of an algebraic opening in a previous section, a few very useful properties stand out:

- Any operation that has these properties:
  - *Idempotence*
  - *Anti-extensive*
  - *Increasing*

  is an algebraic opening
- Any *supremum* of opening is still an opening.

This allows to create other opening operations for better morphological filtering. The example of using the supremum of multiple morphological openings with different structuring elements, in order to suppress elements of different sizes, was given in the previous section.

The idea of the *opening by reconstruction* (also called *geodesic* opening) is to get the best reconstruction possible of the original image after the initial erosion erases the shapes to be filtered. The way this is achieved is by applying the original image as a mask during the dilatation part, and repeat dilatation until idempotence. Because the reconstruction is constrained / masked by the original image, it is necessarily anti-extensive, and also has the other properties necessary to be an algebraic opening.

$$O_R(F) = R_F^D[(F \ominus B)] \tag{1.74}$$

F is the original image, and B a structuring element. With R_F^D being the reconstruction using dilatation and F as a mask. The reconstruction is a serie of masked openings until idempotence.

$$R_F^D[(F \ominus B)] = D_F^{(k)}[(F \ominus B)] \tag{1.75}$$

where k is the number of necessary iterations until idempotence(an iterator) such that $D_F^{(k)}[(F \ominus nB)] = D_F^{(k-1)}[(F \ominus nB)]$.

$$D_F^{(k)}[(F \ominus B)] = \left(\left(D_F^{(k-1)}[(F \ominus B)]\right) \oplus B\right) \cap F \tag{1.76}$$

$$D_F^{(1)}[(F \ominus B)] = ([(F \ominus B)] \oplus B) \cap F \tag{1.77}$$

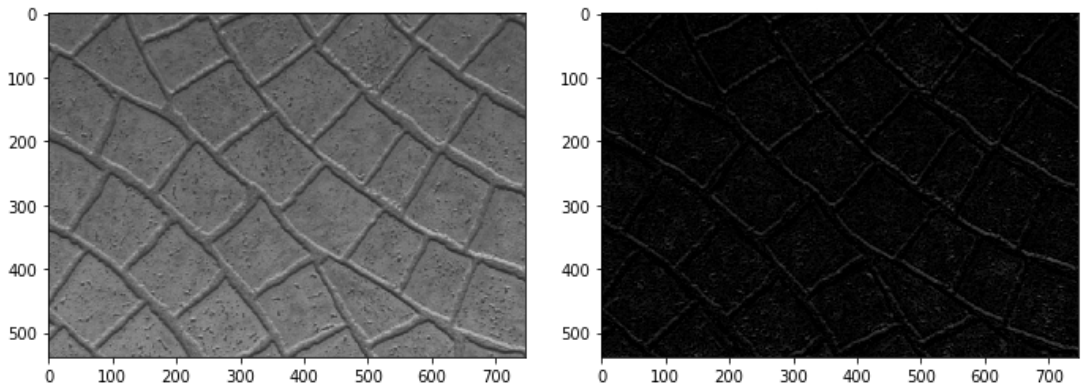


Figure 1.11.: Top hat transform with Disk shaped SE of size $lambda = 9$ on cobblestone texture

Other algebraic openings

This very complete document[36] also mentions other types of closing (can be reciprocally made openings), namely:

- **Parametric closing**: Relaxed structural($\ominus$) closing, also called rank-min closing because usually implemented using a *rank filter*
- **Attribute closing**
- **Flooding**: Based on the watershed transform.

1.6.3. Top-Hat Transform

The top-hat transform extracts the objects that have not been eliminated by an opening, by computing the difference between an image and its opening. The result thus highlights the elements smaller than the structuring element. The transform is just analogue to a high pass filter, highlighting elements smaller than the structuring element. The top-hat transform also highlights parts that are brighter than their surrounding, being then a good pre-processing technique before a threshold operation.

The top-hat transform:

$$T_w(f) = f - f \circ b \tag{1.78}$$

The black-hat transform (the opposite transform that used the closing):

$$T_b(f) = f \bullet b - f \tag{1.79}$$

Using a *geodesic opening* (opening by reconstruction) instead of an adjunction opening by erosion and dilatation, can improve the results of the top-hat transform, as the quality of the reconstruction in the opening is better.

The top-hat transform is one of the most used mathematical morphology applications, as it allows to highlight details normally invisible to the human eye.

1.6.4. Morphological Gradient

The morphological gradient is another residual transform (being the difference between an image and its transform) that uses dilatation and erosion:

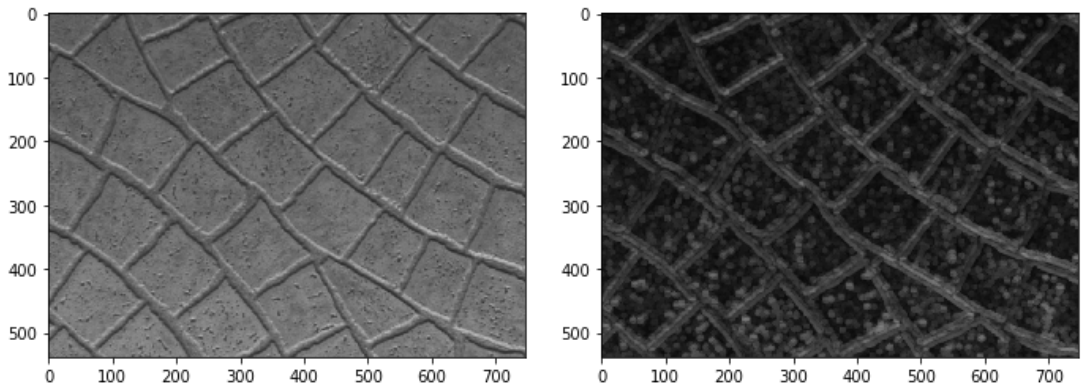


Figure 1.12.: Morphological Gradient transform with Disk shaped SE of size $lambda = 9$

- The *external gradient* is the residual of an image and its dilatation:

$$G_e(f) = f \oplus b - f \tag{1.80}$$

- The *internal gradient* is the residual of an image and its erosion:

$$G_i(f) = f - f \ominus b \tag{1.81}$$

- The *symmetrical gradient* is the sum of the internal and external gradient:

$$G(f) = G_e(f) + G_i(f) \tag{1.82}$$
$$G(f) = f \oplus b - f \ominus b \tag{1.83}$$

- The *morphological Laplacian* is the difference of the external and internal gradient.

$$G(f) = G_e(f) - G_i(f) \tag{1.84}$$

The morphological gradient highlights borders in objects, in this way it acts in a similar way than a derivative operator that can be applied with a linear filter (in this case they are commonly called edge detectors, like the Sobel operator). The morphological nature of the morphological gradient allows it to be very good at edge detection and segmentation.

The external gradient peaks outside of the edges, and the internal gradient peaks inside the edges. The symmetrical gradient peaks on the edges.

However, all of these techniques output images, which are not *Feature Vectors*. The next section will explore how to use these operations as a mean of extracting features.

1.7. Texture Description

Rao (1990) started a taxonomy of textures, according to which textures can be classified into four categories:

- **Strongly Ordered**: repetitive placement of primitive elements according to rules
- **Weakly ordered**: possesses local dominant orientation (that can vary at a global level, as the viewing angle can vary anyway)

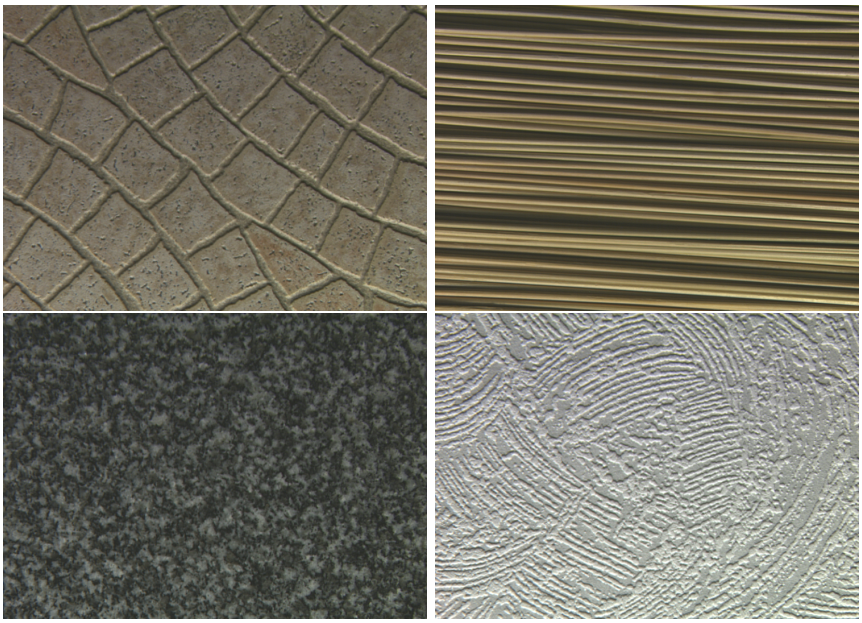


Figure 1.13.: Textures from the OUTEX database

- **Disordered**: lacking the properties above, and thus mostly described according to their roughness.
- **Compositional**: combination of the above.

The most important criteria to extract from a texture are therefore *Granularity*, *Orientation* and *Periodicity*, as specific morphological extraction tools are developed to extract these information that together can define a texture. In order to do so, we will mainly use *Morphological Series*.

1.7.1. Color

Color is also a very important aspect of textures, and some Mathematical Morphology-based approaches to extract this information have been developed, however the final goal of this thesis is to work on biomedical imagery which is grayscale. For this reason color will not be developed in this document.

1.8. Mathematical Morphology for Texture Classification

1.8.1. Image Moments

Most of the math that is presented in this document is inspired by the works of *Sebastien Lefèvre* and *Erchan Aptoula*, who wrote many papers on mathematical morphology. Some of them were written in domains that are very relevant to our subject, such as *Remote Sensing*.

Because moments are scalars, they can even be used as they are in texture descriptors, but the main goal is to use them in order to quantify the difference between an image and its result after a mathematical morphology operation.

Definition of an Image Moment

The unscaled moment $m_i j$ of order (i, j) of an image f of size MxN pixels is given by:

$$m_{ij}(f) = \sum_{x=1}^{M}\sum_{y=1}^{N} x^i y^j f(x,y) \tag{1.85}$$

Further refinement is possible through the use of unscaled central moments:

$$\begin{cases} \mu_{ij}(f) = \sum_{x=1}^{M}\sum_{y=1}^{N}(x-\overline{x})^i(y-\overline{y})^j f(x,y) \\ \overline{x} = m_{10}(f)/m_{00}(f) \\ \overline{y} = m_{01}(f)/m_{00}(f) \end{cases} \tag{1.86}$$

Allows for translation invariant measurements, which is a useful property.

[18]

Volume

The *Volume* of an image f of size $MxN$is defined as follows:

$$V(f) = \sum_{x=1}^{M}\sum_{y=1}^{N} f(x,y) \tag{1.87}$$

Which is simply the sum of all pixels. It is also the first moment $\mu_{0,0}$, and it is the default way of measuring the difference between an image and its result processed by mathematical morphology. However, just using the volume does not represent all spacial characteristics of the transformed image, so using other moments can also add information, even if the information gain of each moment needs to be quantified.

The way we will use the moments is through the *Normalised unscaled moments*, defined by Hu, M.K.: Visual pattern recognition by moment invariants. IRE Transactions on Information Theory 8 (1962) 179–187

$$\begin{cases} \eta_{ij}(f) = \dfrac{\mu_{ij}(f)}{[m_{00}(f)]^\alpha} \\ \alpha = \dfrac{i+j}{2} + 1, \forall (i+j) \geq 2 \end{cases} \tag{1.88}$$

This definition achieves *scale* and *translation invariance* which is important for textures.

1.8.2. Morphological Series

Morphological Series are the main way to make texture descriptors based on *Mathematical Morphology*. The principle is to have a *Morphological Operator* (such as for instance a *Dilatation*) and a *Structuring Element* SE (such as a disk) and compare a *Moment*(for instance Volume) of each successively modified images. Mathematically, it would be defined as follows:

$$S'(f; SE_\lambda) = \mu_{\mathrm{i,j}}\left(\varepsilon_{SE_\lambda}(f)\right) \tag{1.89}$$

Where SE is a *Structuring Element* of increasing size λ, and μ_{ij} is a moment as defined previously. For an increasing λ, a morphological serie gives some characteristic information on the texture.

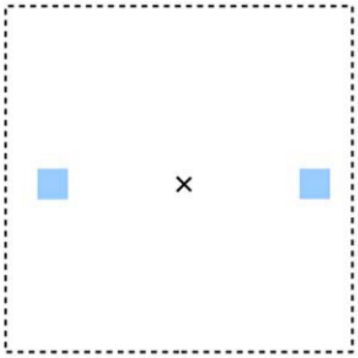


Figure 1.14.: Point pair SE

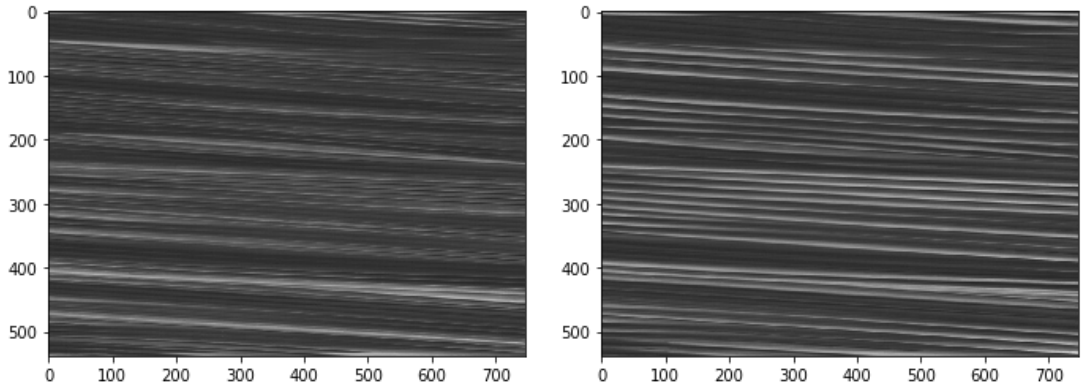


Figure 1.15.: Wood Texture eroded with pair of vertical points *a*) distance λ that gives the least volume *b*) λ that gives high volume (peak)

Morphological Covariance

The morphological covariance K' of an image f, is defined as the volume Vol of the image (the sum of all pixel intensity values), eroded by a pair of points $P_{2,v}$ separated by a vector v:

$$K'\left(f; P_{2,v}\right) = \mathrm{Vol}\left(\varepsilon_{P_{2,v}}(f)\right) \tag{1.90}$$

where ϵ designates the erosion operator. In practice, K' is computed for varying lengths of v, and most often the normalised version K is used for measurements:

$$K(f) = \mathrm{Vol}\left(\varepsilon_{P_{2,v}}(f)\right) / \mathrm{Vol}(f) \tag{1.91}$$

Morphological Covariance gives information about the *periodicity* of the input texture, as the distance between the peaks of the distribution correspond to the periodicity of the texture. The sharpness is also linked to the width of the peaks. The initial drop-off corresponds to the *coarseness*, a sharp drop-off will correspond to a coarse texture.

Because the two points $P_{2,v}$ are oriented, the result is purely *anisotropic*. The *SE* $P_{2,v}$ has both distance between the two points and an orientation. By using multiple directions, such as for instance $(0\,\mathrm{deg}, 45\,\mathrm{deg}, 90\,\mathrm{deg}, 135\,\mathrm{deg})$, it is possible to capture the *orientation* of textures, which is one of the four important criteria we focus on.

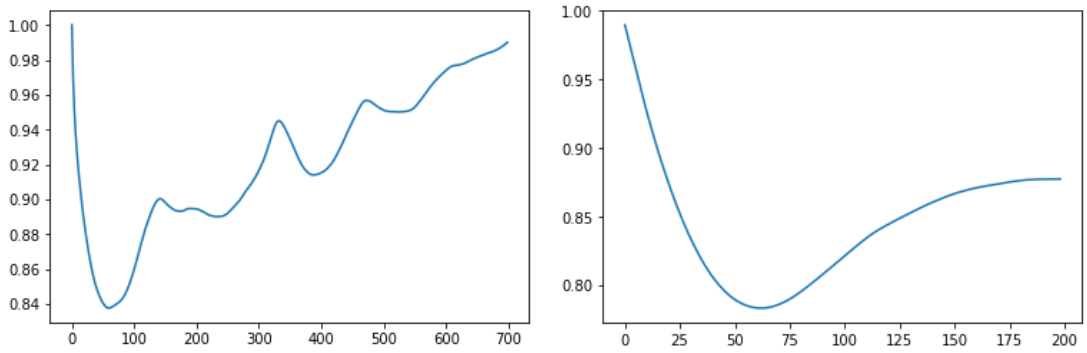


Figure 1.16.: Morphological covariance plot with *a*) horizontal SE *b*) vertical SE

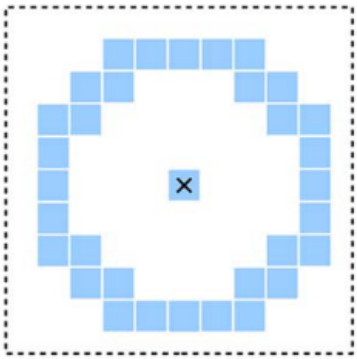


Figure 1.17.: Circle-shaped SE

Circular Covariance Histogram

Circular Covariance Histogram is an attempt to get a more robust *Morphological Covariance* that is *isotropic*.

The principle is to use a growing circular structural element B_λ with a *Morphological Operator*, in order to obtain a series of processed images:

$$\Pi_n^T(f) = \left\{T_{B_i}(f)\right\}_{1\leq i\leq n} \tag{1.92}$$

From there, we create a new double array that has the same size as the original image L, in which each pixel p corresponds to the *SE* size λ that maximizes the absolute difference with the pixel of the original image. In mathematical terms:

$$\forall p, \quad L(p) = \underset{1\leq i\leq n}{\arg\max} \left\{\left|f(p) - \left[T_{B_i}(f)\right](p)\right| / |B_i|\right\} \tag{1.93}$$

with $|B_i|$ being the carnality of B_i

From this we can extract a n dimensional feature vector as follows:

$$\begin{aligned} \mathrm{CCH}_{T,n} &= [v_1, v_2, \ldots, v_n]^T \in [0,1]^n \\ v_i &= P(L(p) = i) \quad i \in [1,n] \end{aligned} \tag{1.94}$$

However, the use of this feature extraction is debatable because of the loss of *orientation* that *Morphological Covariance* has.

Granularity

The *opening* mathematical morphology using a growing *structural element b* of size λ (for instance a disk) has the interesting property of being *Extensive*, meaning that for a series of opening $\Pi^{\lambda}(f)$ defined as follow:

$$\Pi^{\gamma}(f) = \left\{\Pi^{\gamma}_{\lambda}(f) | \Pi^{\gamma}_{\lambda}(f) = \gamma_{\lambda}(f)\right\}_{0 \leq \lambda \leq n} \tag{1.95}$$

Granulometry is defined as follow:

$$\Omega^{\gamma}(f) = \left\{\Omega^{\gamma}_{\lambda}(f) | \Omega^{\gamma}_{\lambda}(f) = \sum_{p \in E} \Pi^{\gamma}_{\lambda}(f)(p)\right\}_{0 \leq \lambda \leq n} \tag{1.96}$$

That is monotonically decreasing, thus worth being normalized:

$$\Gamma^{\gamma}(f) = \left\{\Gamma^{\gamma}_{\lambda}(f) | \Gamma^{\gamma}_{\lambda}(f) = 1 - \frac{\Omega^{\gamma}_{\lambda}(f)}{\Omega^{\gamma}_{0}(f)}\right\}_{0 \leq \lambda \leq n} \tag{1.97}$$

Higher order moments can also be used instead of volume to get some extra information. This is called *granuolmetric moments*[14] and is used in the experimental part of this master thesis, as these granulometric moments are similar to the first order features used in classical radiomics.

If *closing* is used instead of *Opening*, then the result is anti-granulometry.

Rotation Invariant Triplets

- Combines CCH with (rotating) Morphological covariance
- Isotropic
- Loss of directionality

Tests of other morphological series

1.8.3. Texture Classification

With the features extracted, the classification must be done with another machine learning algorithm, such as *K-NN*.

Using Mathematical Covariance and Granularity Using Volume

Arguably the two most simple morphological series to implement are already very powerful as is. *Sebastien Lefèvre* achieved the following results:

Features	Grayscale
Granulometry	67.53
Covariance	73.82
Concatenated	77.75
Combined	**83.53**

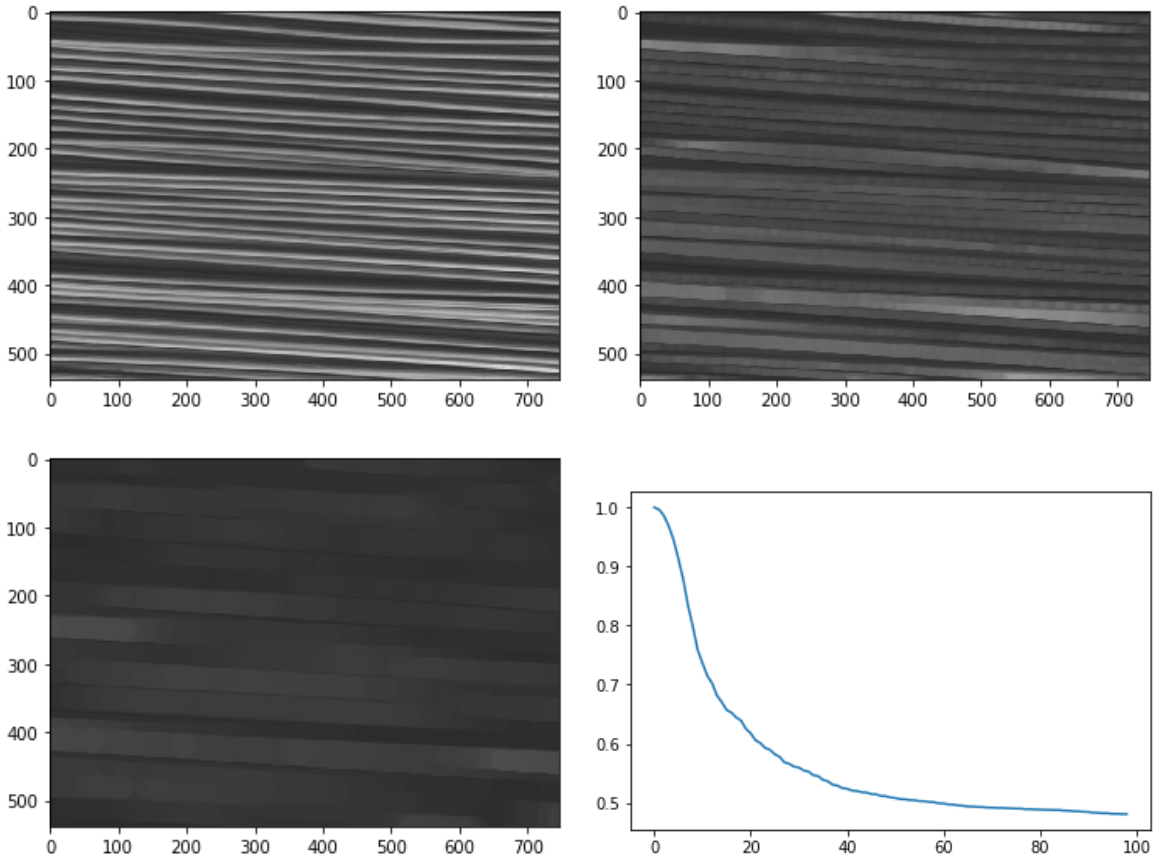


Figure 1.18.: Opened image with *Disk* SE of $\lambda = 0, 10, 30$ and Plot of Normalized Granularity

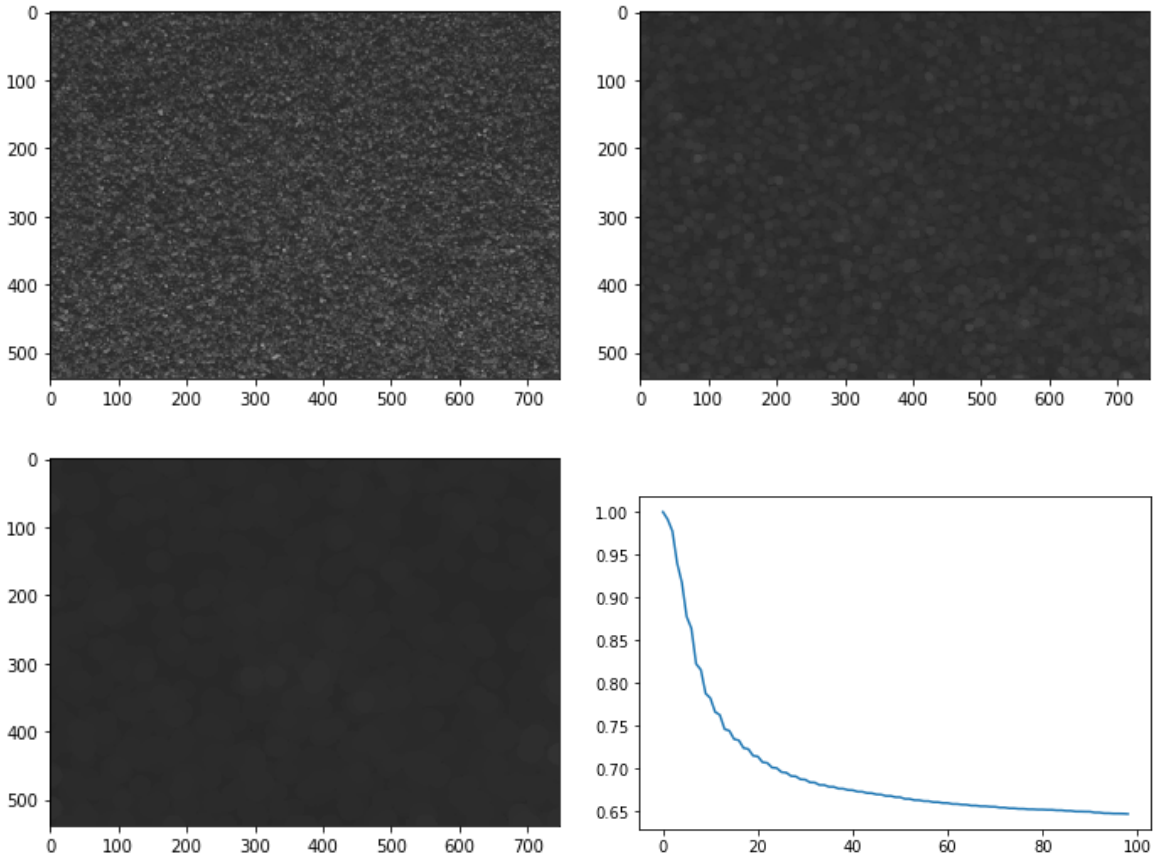


Figure 1.19.: Opened image with *Disk* SE of $\lambda = 0, 10, 30$, Plot of Normalized Granularity

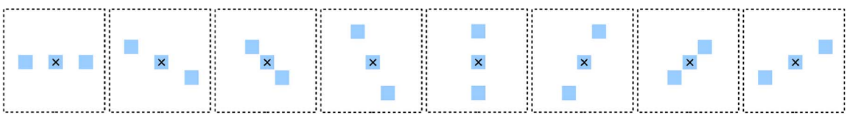
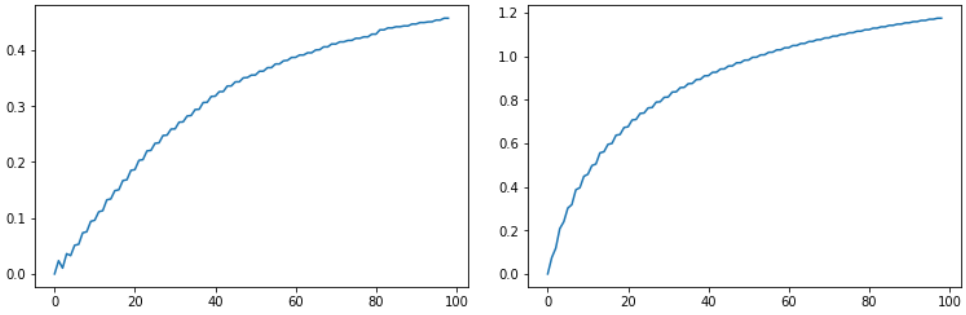
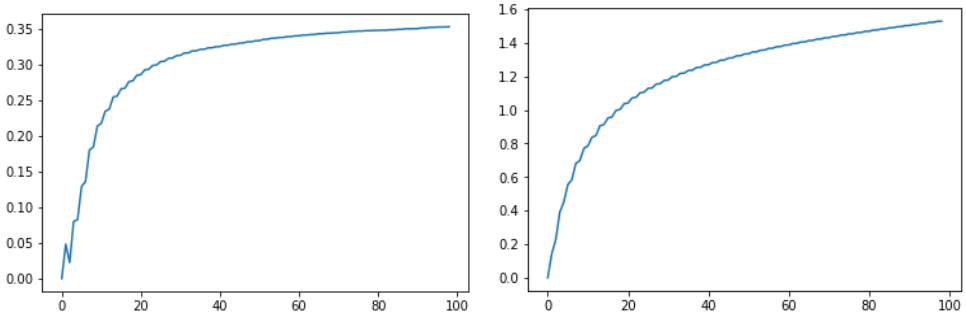


Figure 1.20.: Rotating Point SE



(a) Cobblestone Top Hat

(b) Cobblestone Morphological Gradient



(c) Sand Top Hat

(d) Sand Morphological Gradient

Figure 1.21.: Test of other morphological series with a disk shaped structuring element

Chapter 2

Radiomics

2.1. Introduction

Medical imaging is a non-invasive technique used in clinical routine in the management of patients. It tends to play an increasingly important role in the personalization of treatment based on the individual features of the patient and his/her exams. Personalised medicine has been largely developed with genomics and proteomics, which are invasive techniques (e.g. biopsy).

In order to avoid these invasive operations, the concept of Radiomics was developed [31] [32]. Radiomics, (from "radiology" plus "omics" in comparison with genomics), is a quantitative analysis, consisting of the extraction of a large number of digital data from medical images, in order to obtain predictive and prognostic information concerning the patients treated for a pathology [2]. Initially, this concept was developed in oncology to evaluate the heterogeneity of tumours which was presented as relevant, to predict survival of patients.

Two ideas underlie the concept of Radiomics [32]. The first is that clinical, tissue-level, cellular and/or genomic tumour features would have a phenotypic impact on medical imaging. This amounts to considering that image features are strongly correlated with clinical and biological features. The second rational is that the information carried by the image would be complementary to that from other sources of information, as shown in Fig. 2.1 after Lambin et al. [33], thus enriching the number of tumour features.

The medical interpretation of the images is usually based on a simple visual interpretation of the contrast. Although this type of interpretation has shown its effectiveness in the management of patients in oncology, it remains qualitative and subjective, whereas in Radiomics it becomes necessary to quantify the features for reasons of objectivity, reproducibility, ever increasing number of features studied, and the complexity of modelling some indexes.

Because the concept of Radiomics started in oncology, there are a lot of articles reporting the interest of these features in PET images [17] [21] [22] [23]. More recently, some authors proposed to extend this concept to other type of imaging such as MRI (Magnetic Resonance Imaging) [39] or CT (Computed Tomography) [50]. Radiomics concept is now opening to different pathology than oncology, such as cardiology [39].

The goal of this section is to introduce the Radiomics tools that are used in the final part of this thesis. The process of CT imaging, pre-processing, image feature extraction as well as some level of statistical validation will thus be described.

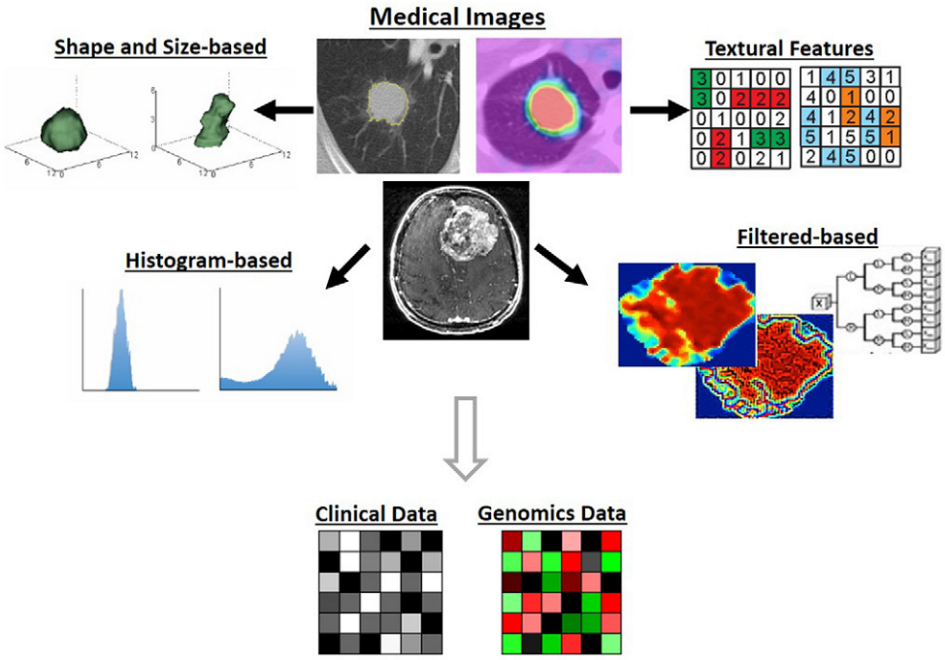


Figure 2.1.: Radiomics principle and features extraction from tumour images [54].

2.2. Cancer

2.2.1. Cancer cells

Cancer cells originate from a mutation of the genetic code. This alteration causes the degeneration of the cell cycle, regulating the life of the cells, as well as their division. The cancer cells then exhibit an exponential and anarchic multiplication creating cell clusters. These groupings are called tumours when they reach about 100,000 cells.

This cell growth is associated with angiogenesis creating new blood vessels providing the energy needed for all these hyperactive cells. Indeed, as cell division is a process requiring a significant amount of energy, these cells have a higher glucose consumption than healthy cells.

2.2.2. Disease management

During the detection of symptoms or a positive screening test, patients perform different exams to confirm or to invalidate the cancer diagnosis. These reviews cover different types of information:

- Clinical data (palpation, stage of the disease, weight loss, ...),
- Biological data (biopsy results, blood test, etc.),
- Genomic and proteomic data.

These exams are usually accompanied by imaging exams such as CT, PET or MRI. All these data make it possible to specify the nature of the tumour and its stage of evolution. They are essential to the implementation of a suitable treatment.

Once the diagnosis is made, it is necessary to choose and plan the best possible treatment. There are three main types: chemotherapy, radiation therapy and surgery. They use different processes and act at different levels. The aim is to amplify the effectiveness of the treatment, sometimes to the detriment of the increase in side effects.

To help the choice of the optimal treatment and to personalise it, the research of features allowing the anticipation of the evolution of the disease is one of the major themes in research. Predictive features that provide information on the efficiency of the treatment are distinguished, with prognostic features providing information on patient survival. These features can come from the various modalities mentioned above: clinical, biological, genetic or proteomic data, as well as from medical images.

Once the treatment has started, the patient is monitored to evaluate the effectiveness of the treatment. Here again, medical imaging plays a fundamental role in monitoring the evolution of the disease during treatment and at a distance, in order to control the appearance of possible recurrences.

2.3. Biomedical Imagery

Eyeballs and cameras have obvious limitations when it comes to observing the inside of a human body, as the visible light-spectrum unfortunately does not pass through human tissues. While it is possible to fit cameras in the human body, a process commonly referred to as endoscopy, or to open the body for visualisation, these techniques are invasive at various levels, and are not applicable to all pathology.

Fortunately, non-invasive imagery methods have been devised, and vary in types such as ultrasound echography, but in Radiomics the most used type of imagery is called tomography as it allows 3D reconstruction into a 3D image, from which countless image features can be extracted.

As sensors can only capture two dimensions, tomography involves image reconstruction, an extremely interesting image processing topic that is unfortunately outside the scope of this thesis which focuses on medical imagery. By nature medical imagery is gray-scaled, meaning that it does not output multiple colour channels like RGB images.

Because Radiomics started in oncology, the most used type of tomography is the PET which is very specific to this domain. Nevertheless, we have shown that different authors studied different type of images (MRI or CT) for Radiomics. In this thesis, we focused on the use of CT images allowing us to extract high resolution bio-markers such as those obtained through morphological operations. This is why this section will introduce as synthetically as possible the way those CT scans are made and what they capture.

2.3.1. Computational Tomography

Computational tomography (CT) refers to X-ray computational tomography, a technique that measures tissue type and density using x-rays. Density differences in tissues below 1% can be measured making CT imagery very precise.

The physical explanation is that X-rays are gradually absorbed by tissues as they go through a body. The rate of absorption is proportional to the density of the tissue, allowing density to be measured by CT.

$$HU = 1000 \times \frac{\mu - \mu_{\text{water}}}{\mu_{\text{water}} - \mu_{\text{air}}} \tag{2.1}$$

where μ_{water} is the linear attenuation coefficient of water, and μ_{air} of air. Because values above 1000 only highlight non relevant elements such as dense bones and foreign metal bodies, the Hounsfield values are clipped in between -1024 and 1024 during pre-processing.

CT images contain information given in a specific scale called the Hounsfield Unit (HU), that describes radio-density, the density of tissues. The scale is defined as follow:

Substance	HU
Air	-1000
Fat	-120 to -90
Water	0
Bone	300-1800
Steel	20000

Table 2.1.: Hounsfield scale values

A CT scan is made by rotation x-ray emitting and x-ray sensor pairs around a body, making what is commonly referred to in tomography as a "slice".

While it is simple to do a 2d projection of a body with a simple radiography, reconstructing a slice from a sinogram (the output of a CT-scanner, where the x-ray intensity values received are mapped on a graph with the angle of the rotation on one axis, and the offset on the sensor on the other), is a complicated task.

Radon transform

The radon transform maps a 2D function (3D extensions have been made and are now used for CT [51]) to another space, for which the coordinates correspond to the sinogram output of a CT-scanner. The radon transform is the perfect mathematical expression of the CT reconstruction problem, however other usage have been made with the radon transform, among which the work of the supervisors of this master thesis of a *soft hash* (hashing techniques in which similar inputs will have similar hashes) based on the radon transform (RaSH [37][43]).

The radon transform is defined as follows for a function g and its radon transform $\mathcal{R}g$:

$$\mathcal{R}\mathrm{g(x,y)} = \int_{\mathrm{L}} \mathrm{g(x,y)d}l \tag{2.2}$$

where $\int_L$ is a line integral, the mathematical equivalent of the physical absorption of an x-ray by tissues, L is given by:

$$p = x \cdot \cos\theta + y.\sin\theta \tag{2.3}$$

with a change of variable it allows to obtain the following integral definition of the radon transform:

$$\mathcal{R}g(p,\theta) = \int_{-\infty}^{\infty} g(p.\cos\theta - q \cdot \sin\theta, p \cdot \sin\theta + q \cdot \cos\theta)dq \tag{2.4}$$

The radon transform indeed has useful properties analogue to those of the Fourier transform (FFT has been used to compute the Radon transform faster[7] [9]), such as for a function g and its radon transform $\mathcal{R}g$:

- for translation by (x_0, y_0):

$$g\,(x - x_0, y - y_0) \longleftrightarrow \mathcal{R}g\,(p - x_0 \cdot \cos\theta - y_0 \cdot \sin\theta, \theta) \tag{2.5}$$

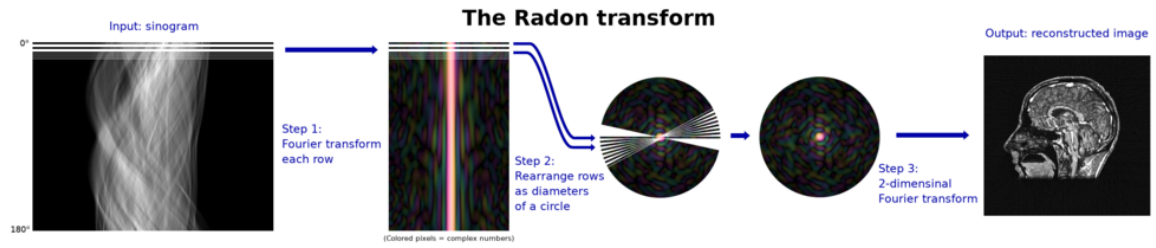


Figure 2.2.: Reconstruction of the original slice using its radon transform sinogram and the Fourier transform. Image courtesy of the wikipedia radon transform article

- for rotation by ϕ:

$$g(x.\cos\phi - y\cdot\sin\phi, x\cdot\sin\phi + y\cdot\cos\phi) \longleftrightarrow \mathcal{R}g(p,\theta+\phi) \tag{2.6}$$

- for scaling by a factor α:

$$g(\alpha.x,\alpha.y) \longleftrightarrow \frac{1}{|\alpha|}\cdot\mathcal{R}g(\alpha.p,\theta) \tag{2.7}$$

- energy conservation:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)dxdy \longleftrightarrow \int_{-\infty}^{\infty}\mathcal{R}g(p,\theta)dp \tag{2.8}$$

These properties make the Radon transform useful beyond tomographic reconstruction, such as in hashes.

Reconstruction from the sinogram/Radon transform

For 2D cases the analytical formulation used to recover is the following [10] :

$$f(\boldsymbol{x}) = \frac{1}{(2\pi)^2}\int_0^{\pi}\left(\mathcal{R}f(\cdot,\theta) * h\right)\left(\langle \boldsymbol{x}, \boldsymbol{n}_\theta\rangle\right)d\theta \tag{2.9}$$

where h is a is a constitutional kernel such as $\hat{h}(\omega) = |\omega|$, oftentimes called a *Ramp filter.*

This formula is a consequence of the **projection slice theorem** that links the Radon transform to the Fourier transform, and is also the practical mean of CT reconstruction.

Theorem : For any $\theta \in [0,\pi]$ the Fourier transform of the projection $\mathcal{R}g(\cdot,\theta)$ satisfies

$$\int \mathcal{R}f(t,\theta)e^{-i\omega t}dt = \hat{f}(\omega\cos\theta, \omega\sin\theta) \tag{2.10}$$

What the projection slice theorem means is that the one dimensional transform of the radon transform along a fixed angle θ corresponds to the 2d Fourier transform of the original image evaluated along the radial line $(\omega\cos\theta, \omega\sin\theta)$.

This theorem also gives out the way CT reconstruction is performed:

- the one dimensional FFT is performed for each angle of sinogram

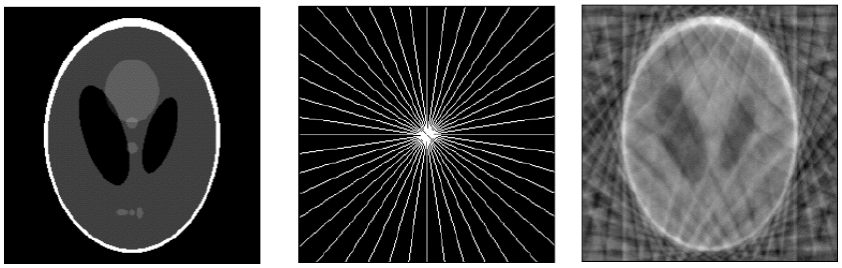


Figure 2.3.: Shep logan phantom with the sinogram sampled for 22 angles (example courtesy of Laurent Jaques computer vision course), however with a sparcity prior it is possible to perfectly recover original image as it is sparce in gradient.

- the FFT's are mapped to a disk, mapping each line to a radial slice
- the inverse iFFT is performed on $\hat{f}$ to reconstruct the image.

However, because the reconstruction process is **ill-posed** in the same way as, for instance, the deconvolution problem, reconstructing the original image perfectly is almost impossible.

Performing better reconstruction with the sampling capability of available scanners, while taking as little time and x-rays as possible is a huge image processing challenge. Good reconstructing algorithms are necessary in all 3d reconstruction tasks.

2.4. Image biomarkers

This chapter will focus on explaining the basics of radiomics necessary for the last part of this thesis. *This section is based on the Image biomarker standardisation initiative[3]* and the documentation of pyradiomics.

The choice of image biomarkers described here is arbitrary, and linked to those that are used in the experimental part of this document. All those image biomarkers are described in more details in the image biomarker standardisation initiative document.

2.4.1. Shape features

Shape features are based on the ROI (region of interest) or VOI (for volume of interest). When the surface of the volume is mentioned, the image biomarker standardisation initiative IBSI states that the volume must be represented by a mesh, in order to not overstate the surface area. The document suggests the use of the *marching cube* algorithm to build these meshes [35].

Volume

There are two types of volume:

- The **mesh volume**

$$\begin{aligned} V_i &= \frac{Oa_i \cdot (Ob_i \times Oc_i)}{6} \\ V &= \sum_{i=1}^{N_f} V_i \end{aligned} \tag{2.11}$$

- The **voxel volume**

Sphericity of the lesion [26] measuring the circular character defined according to the equation: 2.12.

$$\text{sphericity} = \frac{1}{36\pi}\frac{S_{VOI}^3}{V_{VOI}^2} \tag{2.12}$$

Where S_{VOI} and V_{VOI} are respectively the surface and the volume of the VOI studied. The factor $\frac{1}{36\pi}$ ensures that the sphericity is equal to 1 for a sphere.

2.4.2. First Order features

It is possible to summarise the intensity of an image into a histogram. This histogram is a function h that gives for each gray level i, the number of voxel having this intensity in a VOI of size $\omega = X \times Y \times Z$:

$$h(i) = \sum_{x,y,z}^{X,Y,Z} \delta(I(x,y,z),i) \qquad \text{with } i = 0,1,...,G-1 \tag{2.13}$$

Where the image I can be considered as a function of three variables x, y and z with $x \in [1,X]$, $y \in [1,Y]$ and $z \in [1,Y]$. $I(x,y,z)$ can take discrete values such as $i = 0,1,\ldots,G-1$, where G is the total number of gray levels. $\delta(j,i)$ is the Kronecker function such as:

$$\delta(j,i) = \begin{cases} 1 \text{ if } j = i \\ 0 \text{ else} \end{cases} \tag{2.14}$$

The probability density of intensities $p(i)$ is obtained as follow:

$$p(i) = \frac{1}{\omega} \times h(i) \qquad \text{with } i = 0,1,...,G-1 \tag{2.15}$$

The frequency histogram is a simple and concise summary of the first order statistical information contained in the image. The shape of the histogram is indicative of the homogeneity of intensities in VOI. The histogram of a low-contrast image is distributed over a narrow intensity range while a heterogeneous image corresponds to a wide range of intensities.

Several features can be calculated from the histogram. They are defined in the table 2.2 [25]. They are simple to calculate from the moment the ROI has been defined.

2.4.3. Higher Order features

As developed in the previous chapter, the texture of an image corresponds to the mathematical representation of the visual appearance of a surface, which makes it possible to express the frequency of repetition of a pattern or of the homogeneity of the intensities.

There are four main texture matrices proposed in the literature:

- The Gray Level Co-occurrence Matrices (GLCM), introduced by Haralick [20], contain the set of statistics of the second order of the image, characterising the relations of intensity between pairs of neighbouring voxels.
- The Gray Level Difference Matrix (GLDM) has been proposed by Amadasun [5]. It describes the difference in intensity between neighbours, and contains higher order statistics than the previous matrix.
- The Gray Level Runs Length Matrices (GLRLM), introduced by Galloway [19], characterizes the length of runs of the same intensity in a given direction.

Table 2.2.: Main 1st order features.

Features	Formula
Mean Intensity	$\mu = \sum_{i=0}^{G-1} i.p(i)$
Variance	$\sigma^2 = \sum_{i=0}^{G-1}(i-\mu)^2.p(i)$
Skewness	$\mu_3 = \frac{1}{\sigma^3}\sum_{i=0}^{G-1}(i-\mu)^3.p(i)$
Kurtosis	$\mu_4 = \frac{1}{\sigma^4}\sum_{i=0}^{G-1}(i-\mu)^4.p(i)$
Energy	$E = \sum_{i=0}^{G-1}[p(i)]^2$
Entropy	$H = \sum_{i=0}^{G-1} p(i)log_2[p(i)]$
COV	$cov = \frac{\sigma}{\mu}$

- The Gray Level Size Zone matrix (GLSZM) has been proposed by Thibault [49]. It gives the length of the zones having the same intensity in all directions simultaneously.

In the next sections are summarised the different features extracted from these matrices and their mathematical expressions. The construction process for those texture matrices is summarised in the appendix[12].

Grey-Level Co-occurrence Matrix

2.3

Grey-Level Run Length Matrix

2.4

Grey-Level Size Zone Matrix

2.5

Grey-Level Difference Matrix

2.6

Table 2.3.: GLCM features. (μ_i, σ_i) or (μ_j, σ_j) are the mean and the variance for raws or columns for the matrix C [3]

Features	**Formulas**
Variance	$\sum_{i,j}^{G} C(i,j)[(i-\mu_x)^2 + (j-\mu_y)^2]$
Energy	$\sum_{i,j}^{G} C(i,j)^2$
Entropy	$-\sum_{i,j}^{G} C(i,j)\log(C(i,j))$
Correlation	$\sum_{i,j}^{G} C_{ij} \times \frac{(i-\mu_x)(j-\mu_y)}{\sigma_x\sigma_y}$
Dissimilarity	$\sum_{i,j}^{G} C(i,j)\lvert i-j\rvert$
Contrast	$\sum_{i,j}^{G} C(i,j)(i-j)^2$
Homogeneity	$\sum_{i,j}^{G} \frac{C(i,j)}{1+\lvert i-j\rvert}$
IDM	$\sum_{i,j}^{G} \frac{C(i,j)}{1+(i-j)^2}$
Cluster Shade	$\sum_{i,j}^{G} C(i,j) \times (i+j-\mu_x-\mu_y)^3$
Cluster Tendency	$\sum_{i,j}^{G} C(i,j) \times (i+j-\mu_x-\mu_y)^4$

Table 2.4.: Regional features extracted from the GLRLM R. N_r is the total number of homogeneous runs, G is the total number of gray levels and N is the maximum size of homogeneous runs. [3].

Features	**Formulas**
SRE for Small Run Emphasis	$\frac{1}{N_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{R(i,j)}{j^2}$
LRE for Large Run Emphasis	$\frac{1}{N_r}\sum_{i=1}^{M}\sum_{j=1}^{N}R(i,j)\times j^2$
LGRE for Low Intensity Run Emphasis	$\frac{1}{N_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{R(i,j)}{i^2}$
HGRE for High Intensity Run Emphasis	$\frac{1}{N_r}\sum_{i=1}^{G}\sum_{j=1}^{N}R(i,j)\times i^2$
SRLGE (SRE combined with LGRE)	$\frac{1}{N_r}\sum_{i=1}^{G}\sum_{j=1}^{N}\frac{R(i,j)}{i^2\times j^2}$
SRHGE (SRE combined with HGRE)	$\frac{1}{N_r}\sum_{i=1}^{G}\sum_{j=1}^{N}\frac{R(i,j)\times i^2}{j^2}$
LRLGE (LRE combined with LGRE)	$\frac{1}{N_r}\sum_{i=1}^{G}\sum_{j=1}^{N}\frac{R(i,j)\times j^2}{i^2}$
LRHGE (LRE combined with HGRE)	$\frac{1}{N_r}\sum_{i=1}^{G}\sum_{j=1}^{N}R(i,j)\times i^2\times j^2$
GLNUr for Gray Level Non-Uniformity	$\frac{1}{N_r}\sum_{i=1}^{G}\left[\sum_{j=1}^{N}R(i,j)\right]^2$
RLNU for Run Length Non-uniformity	$\frac{1}{N_r}\sum_{j=1}^{N}\left[\sum_{i=1}^{G}R(i,j)\right]^2$
RP for Run Percentage	$N_r/\sum_{i=1}^{G}\sum_{j=1}^{N}(R(i,j)\times j)$

Table 2.5.: Features extracted from GLSZM Z. $N_z$is the total number of homogeneous zones, G is the total number of gray levels and N is the maximum size of homogeneous zones.

Features	Formulas
SZE for Small Zone Emphasis	$\frac{1}{N_z}\sum_{i=1}^{G}\sum_{j=1}^{N}\frac{Z(i,j)}{j^2}$
LZE for Large Zone Emphasis	$\frac{1}{N_z}\sum_{i=1}^{G}\sum_{j=1}^{N}Z(i,j)\times j^2$
LGZE for Low Intensity Zone Emphasis	$\frac{1}{N_z}\sum_{i=1}^{G}\sum_{j=1}^{N}\frac{Z(i,j)}{i^2}$
HGZE for High-Intensity Zone Emphasis	$\frac{1}{N_z}\sum_{i=1}^{G}\sum_{j=1}^{N}Z(i,j)\times i^2$
SZLGE (SZE combined with LGZE)	$\frac{1}{N_z}\sum_{i=1}^{G}\sum_{j=1}^{N}\frac{Z(i,j)}{i^2\times j^2}$
SZHGE (SZE combined with HGZE)	$\frac{1}{N_z}\sum_{i=1}^{G}\sum_{j=1}^{N}\frac{Z(i,j)\times i^2}{j^2}$
LZLGE (LZE combined with LGZE)	$\frac{1}{N_z}\sum_{i=1}^{G}\sum_{j=1}^{N}\frac{Z(i,j)\times j^2}{i^2}$
LZHGE (LZE combined with HGZE)	$\frac{1}{N_z}\sum_{i=1}^{G}\sum_{j=1}^{N}Z(i,j)\times i^2\times j^2$
GLNUz for Gray Level Non-Uniformity	$\frac{1}{N_z}\sum_{i=1}^{G}\left[\sum_{j=1}^{N}Z(i,j)\right]^2$
ZLNU for Zone Length Non-Uniformity	$\frac{1}{N_z}\sum_{j=1}^{N}\left[\sum_{i=1}^{G}Z(i,j)\right]^2$
ZP for Zone Percentage	$N_z/\sum_{i=1}^{G}\sum_{j=1}^{N}(Z(i,j)\times j)$

Table 2.6.: Features extracted from the GLDM N. E is the total number of voxel in the VOI and G is the total number of gray level.

Features	Formulas
Coarseness	$1/\sum_{i=1}(N(i,1)\times N(i,2))$
Contrast	$\frac{1}{E\times G(G-1)}\left[\sum_{i=1}\sum_{j=1}N(i,1)\times N(j,1)\times(i-j)^2\right]\times\left[\sum_{i=1}N(i,2)\right]$
Busyness	$\sum_{i=1}N(i,1)\times N(i,2)/\sum_{i=1}\sum_{j=1}(i\times N(i,1)-j\times N(j,1))$
Complexity	$\sum_{i=1}\sum_{j=1}\frac{\lvert i-j\rvert}{E(N(i,1)+N(j,1))}\times(N(i,1)\times N(i,2)+N(j,1)\times N(j,2))$
Strength	$\sum_{i=1}\sum_{j=1}(N(i,1)+N(j,1))(i-j)^2/\sum_{i=1}N(i,2)$

Chapter 3

Implementation and experimentation of Mathematical Morphology to Radiomics

In the previous chapters, we presented how mathematical morphology tools can be used to extract texture features and what Radiomics brings to medicine. In this chapter, we will study the interest of these features in the context of a Radiomics statistical study on a cohort of lung cancer patients. For this study, we used the methodology published by the IBSI (Image Biomarker Standardization Initiative) consortium [3] for the pre-processing and for the extraction of features as well as the article by Chalkidou et al. [11] for statistical analysis. Finally, after the statistical analysis, a classification step is performed using relevant features and a machine learning algorithm based on decision trees, to predict patient outcomes.

3.1. Dataset

In this thesis we studied the database NSCLC-Radiomics published by Aerts et al. [1]. This cohort contains CT scans from 422 patients suffering from Non-Small Cell Lung Cancer (NSCLC) (see Fig. 3.1). These scans are all acquired before treatment. Several clinical features are also registered and summerized Table 3.1. This dataset contains the manually tumour segmentation of 297 patients. These ROI are the ground truth validated by physicians and thus takes the segmentation problem out of the equation.

3.2. Pre-processing

The goal of the pre-processing step is to turn the data stocked inside the dataset into data structures usable for computing the image bio-markers that will be used for Radiomics. CT images are stored in DICOM format, for *digital imaging and communications in medicine*. This format is the standard one for medical images and is widely used for the storing and transmitting of medical images. Each slice is stored as a *.dcm* file that is in fact a compressed image associated with metadata concerning the patient (age, gender, personal practitioner, device used for the imaging, etc.). In public databases such as the one used here, all personal information is rendered anonymous. The DICOM-RT format, (where the RT stands for Radiation Therapy, a domain in which defining regions of interest is important due to the need to target it with beams), contains one or several ROI.

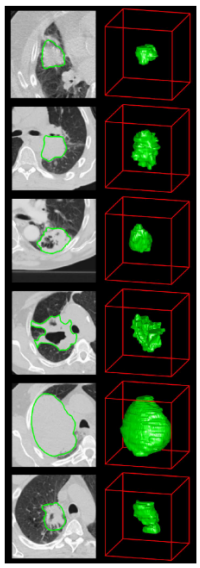


Figure 3.1.: Different 3D lung tumours of the NSCLC-Radiomics dataset

Table 3.1.: List of the clinical features from the medical file of 422 patients suffering from a lung cancer. *TNM stages are developed in appendix.

Features	**Number of patients**
Demographic	
Patient's age (years)	
Median (range)	68.5 (33.7-91.7)
Mean (standard deviation)	68.0 ± 10.1
Patient's gender	
Male	290 (69 %)
Female	132 (31 %)
Clinical	
Histology	
Adenocarcinoma	51 (12 %)
Nos	63 (15 %)
Large cell	114 (27 %)
Squamous Cell Carcinoma	152 (36 %)
Unknown	42 (10 %)
Clinical stage*	
I	93 (22 %)
II	40 (9 %)
III	288 (68.8 %)
Unknown	1 (0.2 %)
Outcomes	
Follow-up (months)	
Median (range)	13.3 (0.3-71.0)
Mean (standard deviation)	17.6 ± 13.6
Last news survival	
Alive	178 (42 %)
Dead	244 (58 %)
Two-years survival	
Alive	208 (49.3 %)
Dead	214 (50.7 %)

A ROI is defined as a series of millimetric coordinates that lay out the contour. These coordinates have to be converted into a binary mask before being used in Radiomics.

CT slices and their ROI made into binary can then be made into 3D arrays. However, the resolution of the voxels can vary according to the devise used for the acquisition, or the parameters set-up by the practitioners. To compare features extracted from a database, it is necessary to homogenise it. The pre-processing step thus makes the resolution of each voxel identical for all patients and in all directions. In this case, the spacing was set to $1 \times 1 \times 1$ mm^3, in order to have a relatively high resolution.

During the pre-processing step it is also necessary to eliminate all patients that do not fit specific criteria, such as corrupted images or ROI, etc. Furthermore, for computing texture features, the IBSI consortium suggests that only tumours with a volume higher than 10 cm^3 should be used. After removing all the patients that did not fit those conditions, the available data shrank from 297 to 168.

3.3. Features extraction

From the pre-processed scans 3D arrays, a whole range of image features were extracted, separated into two types:

- the classical Radiomics features summarised Table 3.2 and introduced in chapter 2.
- the studied mathematical morphology based image features summarised Table 3.3.

Table 3.2.: Classical Radiomics features studied. *Texture matrices are extracted among all the 13 possibles directions (separated by $45°$), with a distance of one. Features are extracted from the resulting mean matrix.

Type of features	**Features**
1st order	Volume, Sum of intensities (I_{sum}), I_{max}, I_{min}, I_{mean}, I_{std}, I_{cov}, Skewness, Kurtosis, Energy, Entropy
Shape based	Sphericity
Texture based*	10 from the Gray-Level Co-occurence Matrix (GLCM), 11 from the Gray-Level Run Length Matrix (GLRLM), 11 from the Gray-Level Size Zone Matrix(GLSZM), 5 from the Gray-Level Difference Matrix (GLDM)

The mathematical morphology based image features are tested for the first time in the context of Radiomics. The classical Radiomics bio-markers are extracted to compare their performances with the proposed features. As per their nature as series, the mathematical morphology features are very high dimensional and non-parse, which makes their practical use very challenging. The classical Radiomics features are in comparison a lot easier to use.

3.3.1. Morphological based image bio-markers

These new image biomakers which are tested for the first time in the context of Radiomics are for the main part based on *Morphological series*. The concept was explained in the first part of this document (Chapter II), but as a reminder *Morphological series* consists of repeating a morphological

Table 3.3.: Studied mathematical morphology based image features studied.

Type of features	Features
Granularity	Volume(Sum of intensities)
Granulometric moments	η_{300} η_{030} η_{003} Maximum, Standard deviation, Covariance, Skewness, Kurtosis, Entropy
Morphological covariance	Volume(Sum of intensities)
Morphological covariance moments	η_{300} η_{030} η_{003} Maximum, Standard deviation, Covariance„ Skewness, Kurtosis, Entropy

operation such as *erosion* or *opening* with a structuring element increasing in size, and quantifying the change in between operations with scalar metrics.

The two morphological series used here are:

- **Granularity**, the most robust morphological series, able to measure texture coarseness and heterogeneity.
- **Morphological covariance**, a morphological series that can measure the periodicity as well as directionality of textures

As per their nature as series, the features given by these series are **very high dimensional** and non-parse, which makes their practical use very challenging. The classical radiomic features are in comparison a lot easier to use.

Granularity

Granularity is computed by performing successive openings with a growing structuring element. In case of particular textures with specific patterns, doing granularity with specific structuring elements can be very effective, but this does not apply to tumour texture that have no repeating motifs.

Moreover, as stated by the IBSI, directionality does not matter in Radiomics. This is why ultimately the structuring element chosen was a sphere with increasing radius. Because the spacing was made to be the same in all directions during the preprocessing step, there was no need to distort the sphere to take into account the spacing, but it would have been needed in case the spacing was not equal in all directions.

The main metric that is measured in between successive openings is *volume*. Volume in the context of granularity is not the same as in the case of radiomics as it refers in fact to the sum of all voxels values in the granularity case, unlike the classical radiomic feature volume that refers to the size of the tumour in voxels. It is also the first image moment, (more details are explained in the first part of this document).

All the metrics applied to granularity are normalized by the value of this metric on the non-opened image. This is to avoid having the morphological features correlated with similar first order features (without normalization, granularity with as metric volume would be correlated with the first order feature *sum of intensities*). Another way to state this goals of having lowest correlation with first order features, is to have the mathematical morphology based features based solely on texture, meaning that the size of the tumour should have no impact (it would have without normalization).

Antigranularity Anti granularity is essentially the same as granularity, except that closing is used instead of opening. Volume is thus increasing instead of decreasing.

Image Moments Image moments are analogue to their equivalent in statistics statistical moments. They are defined in the first part of this document, where we also defined *central unscaled moments*, and their extension to 3D. In granularity we tested moments $\eta_{3,0,0}$, $\eta_{0,3,0}$ and $\eta_{0,0,3}$ as they are the ones used in [6]

Granulometric moment Granulometric moments as defined in [14] are the most interesting metric to use in this context, as they are mostly the same as those used as first order image biomarkers in classical radiomics. Combining them with mathematical morphology makes the most sense as these metrics are proved effective in classical radiomics. The metrics used are shown in 3.3.

Other first order metrics such as volume (as number of voxels), sum of pixels are not used as metrics on granularity for obvious reasons (sum of pixels normalised is just granulometric volume, as for volume as number of voxels it does not change in between openings). And as for the minimum, the first order statistic, is in fact conserved by the opening operation as seen in Chapter. It thus does not change in between openings.

In the code here the number of iterations is *10*, and as such the length of the feature vector combining all the granulometry features is 120, which is already high dimensional, but not as much as with morphological covariance.

Morphological covariance

Morphological covariance, which was defined in the first section of this document, consists of taking a structuring element that is a pair of points in a specific direction, and eroding an image with it. The main challenge with morphological covariance is to achieve rotation invariance, as the SE in only in one direction. The other challenge is the very high dimentionality of the resulting feature vector. The length of the feature vector is :

$$\text{number of iterations} * \text{number of directions} * \text{number of directions} \tag{3.1}$$

Which is very big and not sparse, so some kind of *dimentionality reduction* (reducing the size of the feature vector while loosing the least meaningful information possible) must occur.

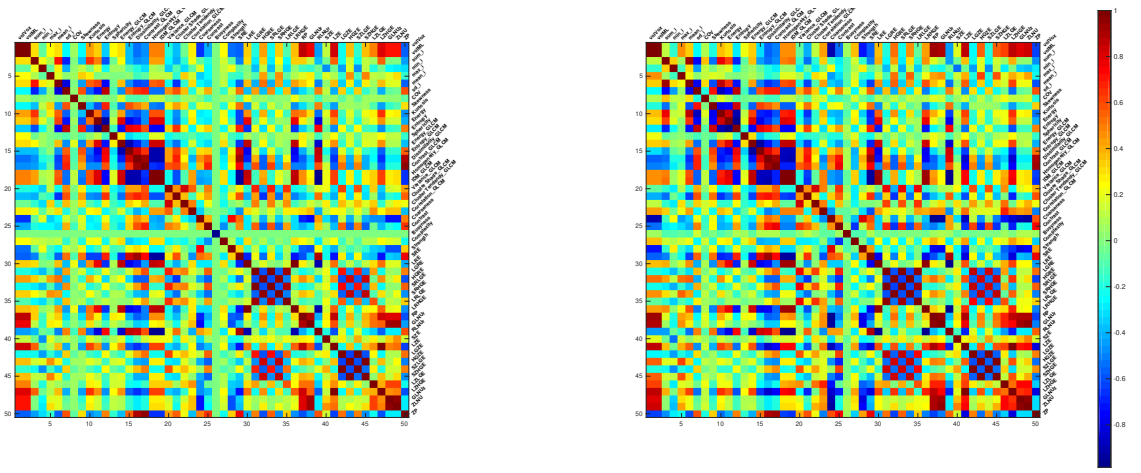
Due to its directionality, morphological covariance is very similar in nature to GLCM and GLRLM that are also directional. These matrices are computed for 13 directions of the 3D space (directions separated by 45°), and morphological covariance has been computed for these same 13 directions.

As the same number of iterations and metrics as granularity were used, the length of the feature vector containing morphological covariance features is 1430, a very long feature vector that necessitates drastic dimentionality reduction.

3.4. Correlation analysis

3.4.1. Methodology

The first step of our study was to do a correlation analysis. Not only can a correlation analysis show the redundancies between features and help as to where to perform dimentionality reduction, it is especially usefull when there are so many features.



(a) First Order features Pearson correlation matrix

(b) First Order features Spearman correlation matrix

Figure 3.2.: Correlation matrices of fist order features

Correlation is defined as follows in the case of Pearson's correlation:

$$\rho_{X,Y} = \frac{1}{\sigma_X \sigma_Y} \mathrm{E}\left[(X - \mu_X)(Y - \mu_Y)\right] \tag{3.2}$$

$$= \frac{1}{\sigma_X \sigma_Y} \sum_{x,y} (x - \mu_X)(y - \mu_Y) \mathrm{P}(X = x, Y = y) \tag{3.3}$$

and is directly linked to covariance:

$$\rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{3.4}$$

$\rho_{X,Y}$ is a value between [-1, 1] where 1 represents the maximum correlation (when X increases so does Y) and -1 is the maximum inverse correlation (when X increases Y decreases). Variables are decorrelated when $\rho_{X,Y} = 0$.

This correlation analysis is purely linear, and cannot capture non linear relations, like other more complicated concepts, such as the Spearman's rank correlation analysis.

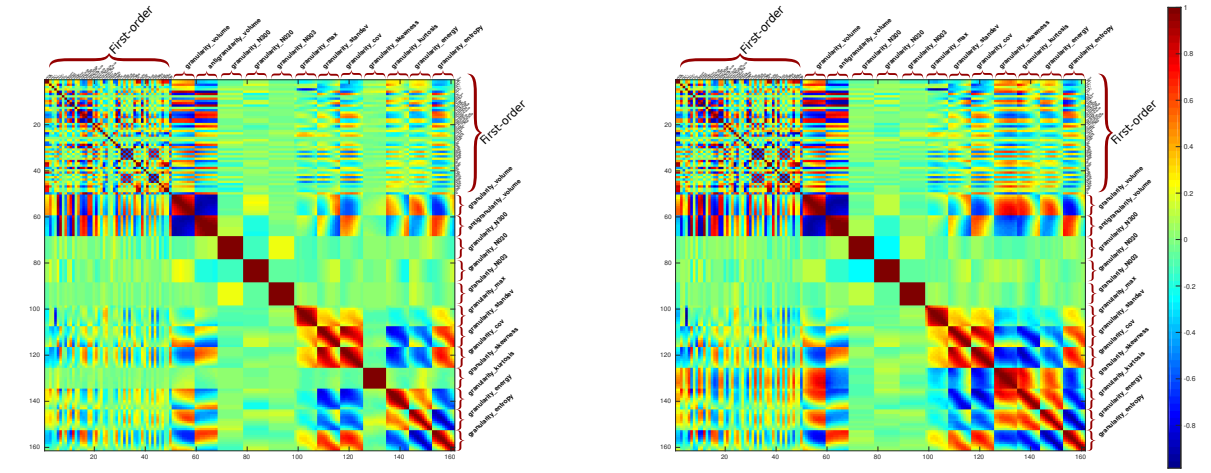
We define X and Y as two features of a list composed of N observations. Observations are sorted according to each feature. Their rank is raised, corresponding to the position they occupy, noted rk_X and rk_Y. The correlation coefficient of Spearman's ranks ρ_s is then defined by the following equation:

$$\rho_{s(X,Y)} = 1 - \sum_{i=1}^{N} (rk(i)_X - rk(i)_Y)^2 \times \frac{6}{N(N^2 - 1)} \tag{3.5}$$

Other approaches exist to seize non-linear correlations, such as mutual information.

3.4.2. Results and discussion

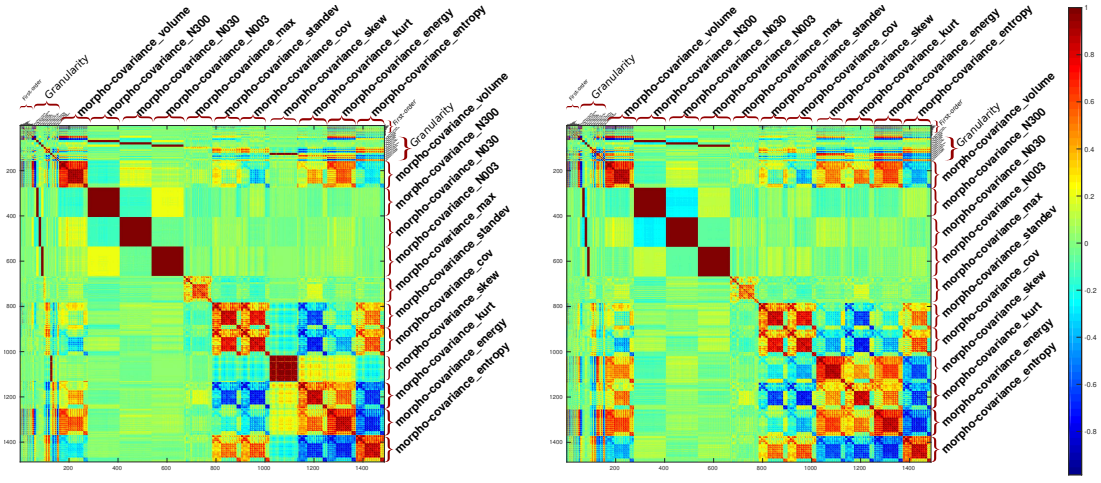
In the case of very high dimensional data such as shown here, the analysis of their correlation is very important, as it shows the redundancy. Most machine-learning algorithms tend to suffer from the curse of dimentionality, and the Kaplan-Meier statistical analysis is uni-variate.



(a) Granulometry features Pearson correlation matrix

(b) Granulometry features Spearman correlation matrix

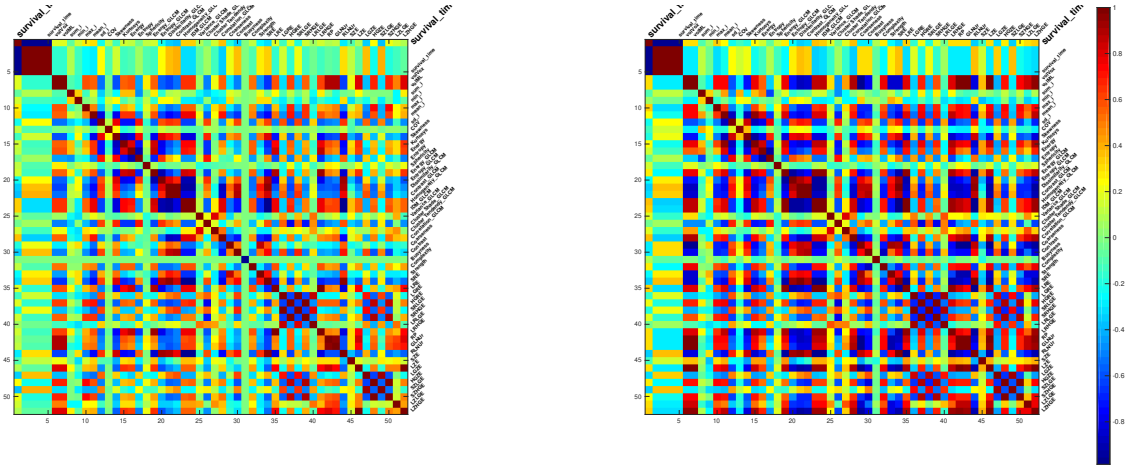
Figure 3.3.: Correlation matrices of granulometry (around 10 features per metric)



(a) Morphological covariance features Pearson correlation matrix

(b) Morphological covariance features Spearman correlation matrix

Figure 3.4.: Correlation matrices of morphological covariance features (around $10 * 13 = 130$ features per metric, 10 iterations per 13 directions)



(a) Pearson correlation matrix

(b) Spearman correlation matrix

Figure 3.5.: Correlation matrices of first order features with survival time

The correlation, as well as anti-correlation (which is a negative relationship, but still a relationship) can show interesting relationships between features. Granularity and anti-granularity are anti-correlated, which is logical due to their duality (granularity is based on opening, anti-granularity is based on closing).

Correlation can also be noticed between granulometry with volume (as in sum of intensities) as metric and other first order features such as standard deviation, skewness, kurtosis, energy, or entropy. This is logical, as granulometry is sensitive to variations of voxel intensities, as well as these other first order metrics that are sensitive to non-homogeneity.

Furthermore another interesting point is that the analysis of classical texture features, such as gray-level co-occurrence matrices, are not only correlated to granularity but also inter-correlated in a way that is similar to mathematical morphology based features: the "correlation squares" are visible on the first order feature correlation matrix in the same way that they are on the mathematical morphology-based features correlation matrices, showing the analogy and similarity between state of the art texture analysis tools and mathematical morphology-based features.

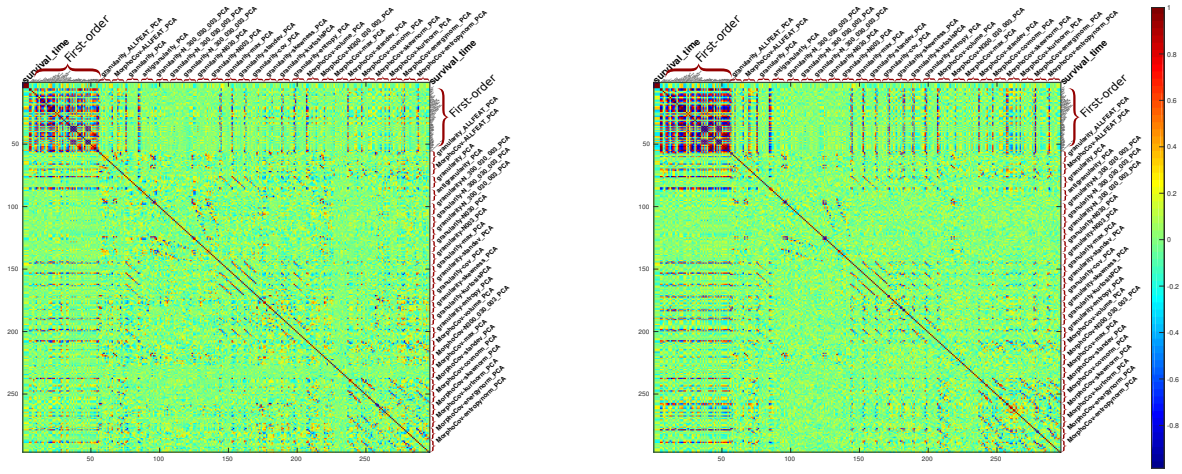
Finally, the most important conclusion of this correlation analysis is the fact that, as expected, values of different iterations of mathematical morphology-based features for a same metric are very inter-correlated, showing a clear redundancy in the data. This is why a method of de-correlating the features is necessary due to the high redundancy and non sparsity of the features. This is why *principal component analysis* (PCA) was used instead of relying on the raw features.

Because PCA is based on linear correlation, in this case the Pearson correlation is more relevant to our analysis.

3.4.3. Dimensionality reduction and survival time correlation

Because the correlation was so strong in between iterations of a morphological series with the same metric, PCA was applied on each metric separately, as well as on all of the granularity metrics and morphological covariance metrics together. As it can be seen on figure 3.6, the transformed features are much less correlated.

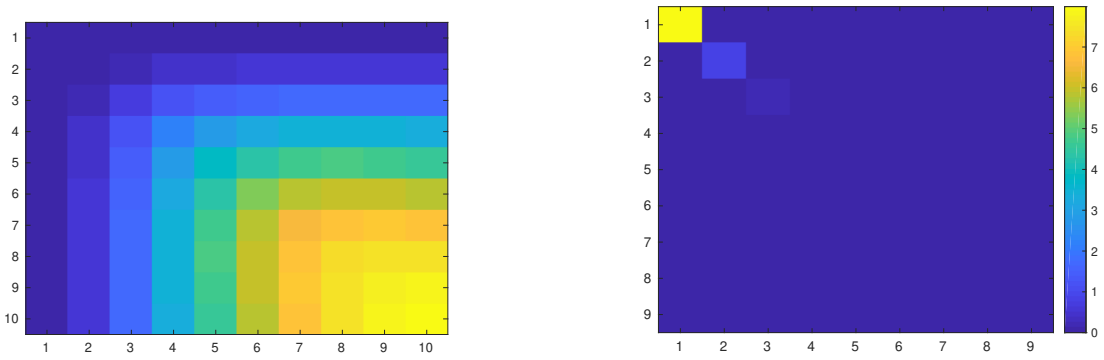
The way that PCA works is that it maps the features to a new set of coordinates, that is a linear



(a) Pearson correlation matrix

(b) Spearman correlation matrix

Figure 3.6.: PCA features correlation, with survival time



(a) Granularity

(b) Granularity-PCA

Figure 3.7.: Covariance matrix before and after PCA comparison (Granularity, volume metric)

combination of the original features. One of the definitions of PCA that is based on covariance

$$\begin{aligned} \operatorname{cov}(PX) &= \mathbb{E}\left[PX(PX)^*\right] && (3.6)\\ &= \mathbb{E}\left[PXX^*P^*\right] && (3.7)\\ &= P\mathbb{E}\left[XX^*\right]P^* && (3.8)\\ &= P\operatorname{cov}(X)P^{-1} && (3.9) \end{aligned}$$

Defines the orthonormal transformation matrix P as being the one that makes $\operatorname{cov}(PX)$ a diagonal matrix, meaning that there is no covariance in between axis in the transformed domain, meaning that there is no correlation, as the Pearson correlation is defined as being the covariance divided by the product of the standard deviations.

As it can be seen in figure 3.7, the result after PCA is not only completely decorrelated, but also more *sparce*.

The final step of the correlation analysis is to see if there is any correlation between the features and an important target such as survival time. This is why for figures 3.5 and 3.6, the surviving patients were filtered out (there is no survival time for surviving patients, what is written in the field instead is the number of days the patient participated to the survey) in order to accurately see any simple correlation between a feature and the survival time.

If some obvious correlation existed, it would be easy, however as it can be seen in figure 3.5, there is no clear correlation between survival time and any of the features. However a certain degree of correlation exists, as it can be seen, for example, in the negative correlation between the volume of the tumour and the survival rate. The size of the tumour is an obvious prognostic factor, however just seeing a correlation between survival and a feature is not enough to demonstrate its relevance. Looking at the correlation with survival time is in no way rigorous and is just a first look at the data. This is why rigorous statistical hypothesis testing is required.

3.5. Prognostic study

In order to analyse the relevance of a feature, simple correlation does not suffice and a true statistical hypothesis testing is necessary. Classical hypothesis testing consists of trying to reject a hypothesis, called H_0 the null hypothesis. Using the probability distribution of the target, a *confidence interval* is made, and H_0 can be rejected if the estimation of the target is outside of the confidence interval.

If the null hypothesis H_0 is rejected, then the feature is deemed relevant. If the H_0 is rejected when it should not have been (the feature was deemed relevant when it was in fact not) it is called a *Type 1 error.* When the null hypothesis H_0 is wrongfully not rejected, it is a *Type 2 error.* The so called *p-value* is such that $1 - p_value$ is the largest possible confidence interval for which there is no type 1 error. The p-value is thus a good indicator of the potential relevance of a feature.

3.5.1. Kaplan-Meier estimator

The Kaplan–Meier estimator[28], also known as the product limit estimator, is a non-parametric statistic used to estimate the survival function from lifetime data. In order to do hypothesis testing on survival data, a simple mean test based on a normal distribution is not enough, due to the nature of survival.

One of the big advantages of Kaplan-Meier is the ability to include "right-censoring". Right-censoring is for a data point that is above a certain value but unknown by how much. It the case of survival analysis, patients who are not dead by the end of the study are known to survive, but it is impossible to know how long they will survive because they are not dead yet. The right-censoring is indicated on Kaplan-Meier survival graphs by vertical tick marks. Dying patients are represented by a step down.

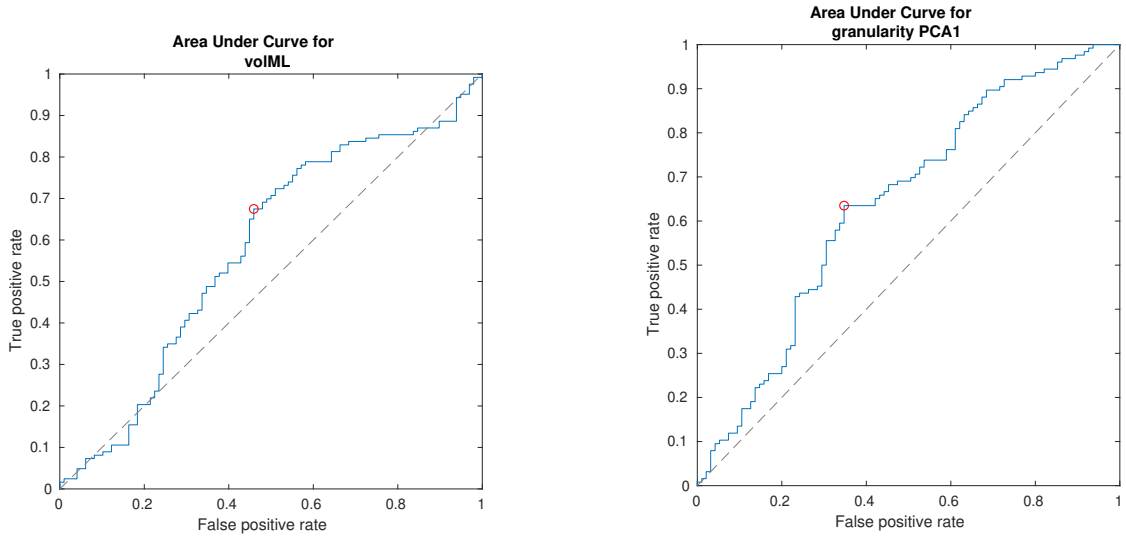
In order to do a Kaplan-Meier survival analysis it is necessary to have:

- Group assignements, which is done in our case by a thresholded feature.
- The status at last observation (dead or alive status)
- The time to event (time until death for dying patients, and time to censoring (time until the patient leaves the study) for surviving patients)

3.5.2. Methodology

A ROC (Receiver Operating Characteristic) curves analysis is performed for each feature, to define the most discriminating cutoff value allowing the differentiation of two groups of patients (alive or dead). This cutoff is defined by the Youden index. A sensitivity (Se), a specificity (Sp) and an AUC (Area Under the Curve) were also obtained through this analysis (see Fig. 3.8).

$$\text{Se} = \frac{\text{True positive}}{\text{True positive} + \text{False Negative}} \tag{3.10}$$



(a) Volume tumour

(b) Granularity PCA 1

Figure 3.8.: Receiver Operating Characteristic (ROC) curves with best cutoff.

$$\text{Sp} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \tag{3.11}$$

To evaluate the prognostic value of the features, a univariate Kaplan-Meier analysis was performed to estimate the survival distribution. The survival was defined as the date of the initial diagnosis to the date of death, or until the end of the follow-up. The association between the survival and each feature was performed after a binarization process using the threshold value previously defined by the analysis of ROC curves. The prognostic value of each feature in terms of survival was assessed using the log-rank test. A p-value less than 5% was considered statistically significant.

To avoid false conclusions, appropriate statistical corrections for type I errors were made according to [11]. A correction of the p-value obtained from the optimal cutoff was performed using the formula of Altman [4]. If P_{min} is the minimum p-value obtained from the optimum cutoff approach for each study, the corrected p-value P_{cor} (for $0.0001 < P_{\text{min}} < 0.1$) is obtained as follows:

$$P_{\text{cor}} = -1.63 \times P_{\text{min}} \times (1 + 2.35 \times \ln(P_{\text{min}})) \quad for\ \epsilon = 10\% \tag{3.12}$$

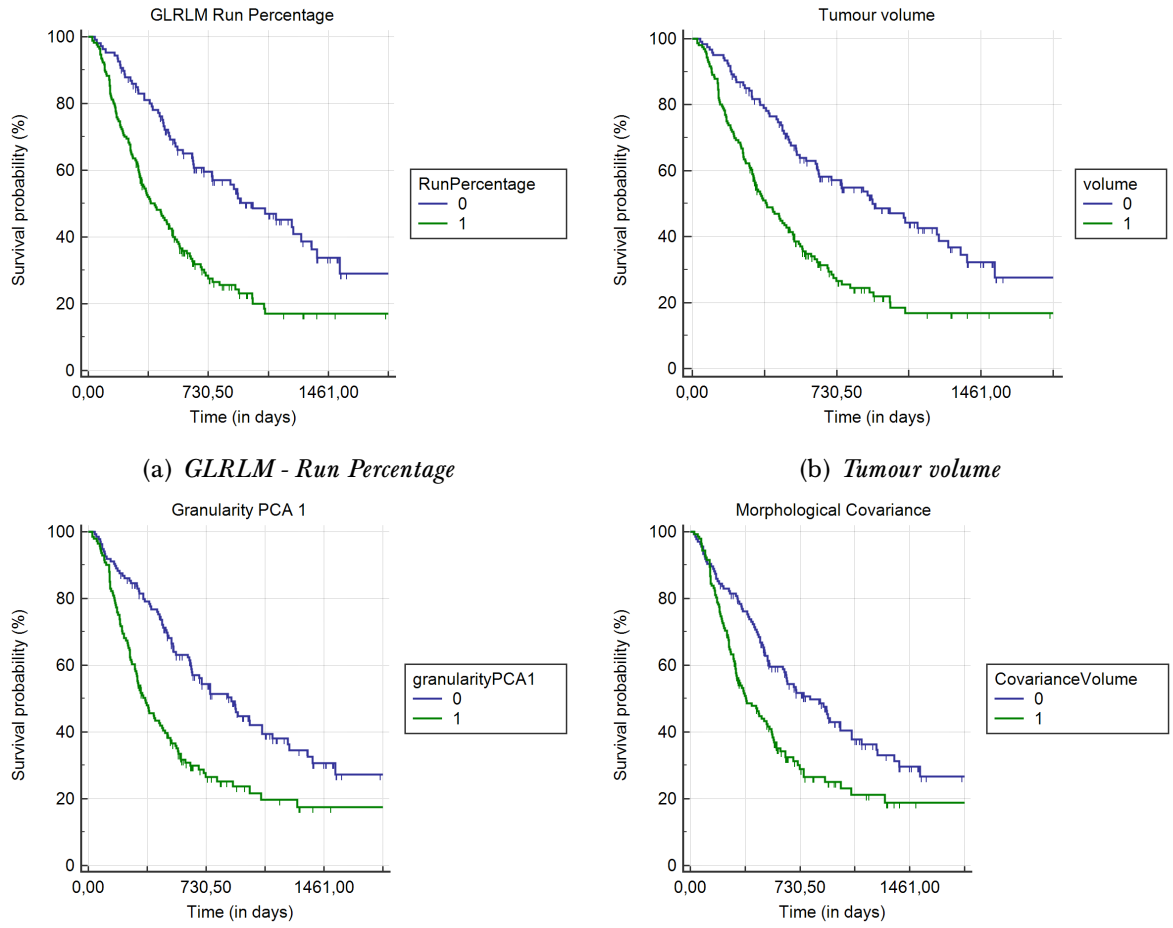
where ϵ is the proportion of values from the tails of continuous variable distribution that is excluded from ROC analysis (from each end of the distribution).

After these corrections, features that pass the threshold are considered relevant.

Validation set

A totally independent validation set was also made for the rest of the study. It is totally independent and was picked after ordering patients in order of:

- tumour volume
- survival time
- final death status



(a) *GLRLM - Run Percentage*

(b) *Tumour volume*

(c) *Granularity PCA 1*

(d) *Morphological Covariance PCA 1*

Figure 3.9.: Kaplan-Meier survival analysis survival curves (survival times in days, horizontal tick is a year). Curves in blue (0) are for the group of patients having the studied features lower than the optimal cutoff while green(1) curves are for the others.

so that the validation set has a similar population to the one used for statistics and training. The validation set was chosen to contain 20 patients.

3.5.3. Results and discussion

A Kaplan-Meier survival analysis needs a horizon, which in this case was tested for 1, 2 and 3 years. The results of the survival analysis are summarised in Table 3.4 for the one year horizon, Table 3.5 for the two years horizon, and in Table 3.6 for the three years horizon.

These results are very interesting, not only because they look very good (a lot of features have a p-value under 0.05 even after the Altman correction) but also because they are consistent (correlated/anti-correlated are selected together). The *granularity PCA 1* feature appears consistently as the best feature is this dataset, and this is thus overall a success for these new mathematical morphology based features.

As shown in Table 3.7, PCA concentrates almost all of the variance in the first few components, and this is illustrated here in the results. Principal components of a higher order (between 5 and 8) also appear to a certain degree down in the scale, however these should be handled with a grain of salt: even if this rigorous hypothesis testing using Kaplan-Meier with type 1 Altman correction can prove the effectiveness of these features, there is a "*confirmation bias*" that is very common in these

Table 3.4.: Results of the prognostic study of patients with a lung cancer using Kaplan-Meier univariate analysis. with horizon at 1 years ($\alpha = 5\%$).

Features	**Se (%)**	**Sp (%)**	**AUC**	**Threshold**	**Logrank** p_{value}	**Altman** p_{value}
GLRLM - Run-Percentage	0.782	0.455	0.599	1.619E+03	1.695E-06	8.351E-05
Granularity PCA 1	0.696	0.606	0.654	8.031E-01	1.865E-06	9.119E-05
GLSZM - GLNUz	0.720	0.555	0.615	2.027E+02	3.412E-06	1.590E-04
GLDM - Complexity	0.603	0.727	0.667	2.628E+06	4.608E-06	2.094E-04
GLDM - Coarseness	0.711	0.538	0.606	2.958E-04	5.634E-06	2.517E-04
GLSZM - ZLNU	0.711	0.503	0.605	8.660E+02	7.314E-06	3.194E-04
GLRLM - LRE	0.375	0.837	0.611	4.788E+00	7.828E-06	3.398E-04
Tumor volume (mL)	0.711	0.517	0.600	1.177E+01	9.396E-06	4.013E-04
MC-volume PCA 10	0.859	0.338	0.537	-3.651E-01	1.196E-05	4.998E-04
1^{st} order - I_{mean}	0.654	0.573	0.615	-6.049E+01	1.798E-05	7.233E-04
GLRLM - HGRE	0.768	0.525	0.668	1.291E+02	2.291E-05	9.004E-04
1^{st} order - Energy	0.446	0.768	0.625	5.791E-03	2.832E-05	1.090E-03
antigranularity PCA 2	0.719	0.568	0.641	-1.040E-01	2.929E-05	1.123E-03
GLRLM - RLNUr	0.395	0.800	0.607	6.320E-01	3.021E-05	1.155E-03
GLCM - Contrast	0.651	0.587	0.626	7.264E+00	3.065E-05	1.170E-03
antigranularity PCA 1	0.704	0.507	0.611	-6.436E-01	3.269E-05	1.240E-03
GLRLM - SRE	0.807	0.399	0.620	8.143E-01	5.601E-05	2.009E-03
GLSZM - LZHGE	0.582	0.718	0.642	1.741E+07	5.680E-05	2.034E-03
1^{st} order - Kurtosis	0.632	0.649	0.657	7.213E+00	5.780E-05	2.066E-03
MC-energynorm PCA 2	0.780	0.396	0.585	-1.002E+00	7.812E-05	2.703E-03
GLCM - Dissimilarity	0.451	0.806	0.652	1.451E+00	7.949E-05	2.745E-03
GLRLM - SRHGE	0.750	0.504	0.660	9.779E+01	8.121E-05	2.798E-03
GLCM - Energy	0.468	0.782	0.643	5.100E-02	8.148E-05	2.806E-03
MC-skewnorm PCA 8	0.175	0.950	0.570	-2.311E-01	8.532E-05	2.923E-03
MC-entropynorm PCA 2	0.329	0.827	0.560	2.417E+00	9.305E-05	3.157E-03
GLDM - Contrast	0.628	0.650	0.644	8.174E-02	9.358E-05	3.173E-03
GLCM - Entropy	0.805	0.431	0.637	2.007E+00	1.553E-04	4.964E-03
1^{st} order - I_{max}	0.697	0.469	0.545	2.400E+02	1.572E-04	5.019E-03
MC-volume PCA 1	0.864	0.400	0.661	-1.953E+00	2.108E-04	6.490E-03
GLRLM- GLNUr	0.623	0.611	0.611	1.966E+04	2.650E-04	7.929E-03
GLSZM - LZE	0.641	0.594	0.611	6.108E+04	2.812E-04	8.350E-03
GLCM - Cluster Shade	0.654	0.643	0.666	2.574E+02	3.572E-04	1.028E-02
GLDM - Strength	0.554	0.701	0.624	8.555E-02	4.065E-04	1.150E-02
MC-covnorm PCA 1	0.513	0.734	0.641	2.759E+00	4.665E-04	1.295E-02
granularity-skewness PCA 5	0.759	0.448	0.595	-2.588E-03	5.208E-04	1.423E-02
MC-skewnorm PCA 4	0.247	0.882	0.520	-5.219E-01	5.315E-04	1.448E-02
1^{st} order - I_{sd}	0.718	0.497	0.601	2.020E+02	5.981E-04	1.603E-02
GLSZM - HGZE	0.690	0.489	0.594	1.189E+02	7.761E-04	2.003E-02
1^{st} order - I_{sum}	0.290	0.888	0.513	-3.979E+04	8.285E-04	2.117E-02
GLRLM - LRHGE	0.566	0.645	0.614	7.297E+02	8.656E-04	2.197E-02
MC-skewnorm PCA 5	0.301	0.851	0.559	-6.813E-01	1.253E-03	3.002E-02
GLSZM - LZLGE	0.875	0.305	0.597	3.548E+02	1.269E-03	3.035E-02
MC-kurtnorm PCA 3	0.750	0.511	0.635	4.206E-01	1.434E-03	3.362E-02
GLRLM - SRLGE	0.519	0.681	0.622	3.537E-02	1.561E-03	3.609E-02
granularity-cov PCA 6	0.690	0.445	0.533	-6.300E-03	1.566E-03	3.620E-02
granularity ALLFEAT PCA 2	0.650	0.589	0.623	-3.142E-01	1.602E-03	3.688E-02
GLCM - Homogeneity	0.654	0.543	0.619	5.816E-01	1.760E-03	3.989E-02
GLCM - IDM	0.654	0.543	0.619	5.378E-01	1.760E-03	3.989E-02
GLRLM - LGZE	0.462	0.643	0.522	3.463E-02	1.854E-03	4.166E-02
1^{st} order - Skewness	0.642	0.629	0.641	-1.766E+00	1.985E-03	4.407E-02

Table 3.5.: Results of the prognostic study of patients with a lung cancer using Kaplan-Meier univariate analysis. with horizon at 2 years ($\alpha = 5\%$).

Features	**Se (%)**	**Sp (%)**	**AUC**	**Threshold**	**Logrank** p_{value}	**Altman** p_{value}
granularity PCA 1	0.635	0.653	0.640	7.899E-01	1.134E-06	5.760E-05
GLSZM - GLNUz	0.664	0.606	0.626	2.019E+02	2.665E-06	1.267E-04
GLDM - Complexity	0.512	0.760	0.619	2.623E+06	3.549E-06	1.648E-04
antigranularity PCA 1	0.571	0.632	0.592	-1.087E+00	4.637E-06	2.106E-04
GLDM - Coarseness	0.748	0.490	0.593	4.456E-04	4.919E-06	2.223E-04
GLRLM - Run-Percentage	0.744	0.490	0.585	1.490E+03	4.919E-06	2.223E-04
GLRLM - GLNUr	0.738	0.515	0.598	1.105E+04	5.170E-06	2.326E-04
MC-volume PCA 10	0.806	0.326	0.525	-3.838E-01	5.576E-06	2.493E-04
GLRLM - LRE	0.305	0.833	0.550	4.774E+00	6.320E-06	2.795E-04
antigranularity PCA 2	0.800	0.516	0.640	2.029E-02	7.562E-06	3.293E-04
MC-volume PCA 1	0.632	0.635	0.646	1.376E+00	9.045E-06	3.876E-04
GLSZM - LZHGE	0.824	0.378	0.606	1.198E+06	9.050E-06	3.878E-04
Tumour Volume	0.675	0.541	0.578	1.053E+01	1.183E-05	4.949E-04
GLCM - IDM	0.379	0.787	0.584	5.782E-01	1.445E-05	5.932E-04
GLCM - Energy	0.442	0.761	0.596	4.714E-02	1.650E-05	6.691E-04
GLCM - Homogeneity	0.364	0.798	0.585	6.170E-01	1.677E-05	6.791E-04
1^{st} order - I_{mean}	0.592	0.604	0.597	-6.049E+01	1.798E-05	7.233E-04
GLRLM - HGRE	0.711	0.505	0.609	1.238E+02	1.857E-05	7.448E-04
GLRLM - SRHGE	0.835	0.362	0.595	8.374E+01	2.933E-05	1.125E-03
GLSZM - LZE	0.769	0.407	0.582	1.423E+04	6.347E-05	2.246E-03
GLDM - Strength	0.705	0.556	0.633	1.396E-01	6.539E-05	2.307E-03
MC-N PCA 4	0.846	0.282	0.522	-3.999E-01	6.959E-05	2.438E-03
GLSZM - ZLNU	0.764	0.439	0.596	5.574E+02	7.076E-05	2.475E-03
GLCM - Dissimilarity	0.366	0.816	0.605	1.455E+00	7.160E-05	2.501E-03
GLRLM - RLNUr	0.832	0.289	0.548	7.453E-01	7.956E-05	2.747E-03
1^{st} order - I_{max}	0.702	0.470	0.564	2.240E+02	1.006E-04	3.384E-03
GLSZM - SZHGE	0.820	0.330	0.571	5.643E+01	1.051E-04	3.517E-03
MC-energy PCA 2	0.815	0.407	0.602	-6.995E-01	1.121E-04	3.723E-03
MC-entropy PCA 2	0.292	0.835	0.554	2.350E+00	1.224E-04	4.024E-03
GLCM - Entropy	0.742	0.452	0.595	2.009E+00	1.261E-04	4.132E-03
GLDM - Contrast	0.675	0.568	0.630	9.265E-02	1.288E-04	4.208E-03
GLRLM - LRHGE	0.580	0.578	0.568	6.255E+02	2.055E-04	6.348E-03
GLCM - Correlation	0.735	0.416	0.550	8.446E-01	2.326E-04	7.074E-03
GLCM - Cluster Shade	0.632	0.635	0.624	3.631E+02	2.643E-04	7.909E-03
1^{st} order - I_{sd}	0.689	0.475	0.560	2.062E+02	2.827E-04	8.388E-03
GLSZM - HGZE	0.511	0.636	0.565	1.309E+02	3.498E-04	1.009E-02
1^{st} order - Kurtosis	0.598	0.652	0.625	6.758E+00	3.620E-04	1.040E-02
granularity-max PCA 2	0.765	0.382	0.566	4.361E-01	4.170E-04	1.175E-02
GLRLM - SRE	0.331	0.812	0.567	7.737E-01	4.415E-04	1.234E-02
GLSZM - SZLGE	0.886	0.214	0.514	2.427E-02	6.648E-04	1.755E-02
GLSZM - LGZE	0.937	0.158	0.504	5.086E-02	7.289E-04	1.898E-02
GLSZM - ZP	0.627	0.540	0.590	8.470E-02	7.641E-04	1.976E-02
GLDM - Contrast	0.376	0.795	0.581	6.236E+00	7.973E-04	2.049E-02
GLRLM - SRLGE	0.928	0.208	0.546	8.013E-02	8.777E-04	2.223E-02
MC-volume PCA 2	0.921	0.213	0.559	-9.537E-01	1.065E-03	2.619E-02
granularity-skewness PCA 5	0.677	0.560	0.614	-2.321E-03	1.211E-03	2.917E-02
granularity-cov PCA 6	0.659	0.489	0.536	-5.808E-03	1.411E-03	3.318E-02
MC-skew PCA 5	0.711	0.490	0.599	4.563E-01	1.488E-03	3.469E-02
MC-kurt PCA 3	0.723	0.582	0.645	4.749E-01	1.549E-03	3.587E-02
granularity-standev PCA 1	0.720	0.458	0.601	-1.667E-01	1.550E-03	3.589E-02
GLRLM - LGRE	0.889	0.242	0.549	9.532E-02	1.692E-03	3.860E-02
granularity ALLFEAT PCA 2	0.615	0.538	0.553	-6.546E-01	1.708E-03	3.890E-02
granularity-skewness PCA 7	0.438	0.714	0.583	-5.748E-04	1.801E-03	4.065E-02
1^{st} order - Skewness	0.579	0.674	0.612	-1.766E+00	1.985E-03	4.407E-02
MC-max PCA 10	0.130	0.980	0.551	-5.901E-01	2.068E-03	4.559E-02
MC-skew PCA 8	0.608	0.563	0.578	-4.138E-03	2.230E-03	4.852E-02

Table 3.6.: Results of the prognostic study of patients with a lung cancer using Kaplan-Meier univariate analysis. with horizon at 3 years ($\alpha = 5\%$).

Features	**Se (%)**	**Sp (%)**	**AUC**	**Threshold**	**Logrank** p_{value}	**Altman** p_{value}
granularity PCA 1	0.596	0.638	0.603	7.899E-01	1.134E-06	5.760E-05
GLSZM - GLNUz	0.642	0.609	0.622	2.019E+02	2.665E-06	1.267E-04
antigranularity PCA 1	0.546	0.625	0.565	-1.087E+00	4.637E-06	2.106E-04
GLDM - Coarseness	0.735	0.506	0.592	4.456E-04	4.919E-06	2.223E-04
GLRLM - GLNUr	0.724	0.529	0.591	1.105E+04	5.170E-06	2.326E-04
GLRLM - LRHGE	0.420	0.731	0.563	8.028E+02	5.284E-06	2.373E-04
GLDM - Complexity	0.507	0.735	0.604	2.584E+06	5.480E-06	2.454E-04
antigranularity PCA 2	0.753	0.493	0.597	2.029E-02	7.562E-06	3.293E-04
GLSZM - LZE	0.701	0.481	0.567	2.434E+04	9.084E-06	3.891E-04
GLRLM - Run-Percentage	0.741	0.488	0.573	1.394E+03	1.039E-05	4.399E-04
Tumour Volume	0.662	0.553	0.568	1.053E+01	1.183E-05	4.949E-04
1^{st} order - I_{mean}	0.551	0.624	0.582	-5.592E+01	1.261E-05	5.245E-04
GLSZM - ZP	0.791	0.370	0.585	9.940E-02	1.891E-05	7.568E-04
GLRLM - LRE	0.848	0.289	0.547	2.495E+00	2.130E-05	8.429E-04
GLCM - Dissimilarity	0.812	0.361	0.590	1.924E+00	3.384E-05	1.279E-03
GLSZM - SZHGE	0.923	0.244	0.582	5.179E+01	3.842E-05	1.434E-03
MC-N PCA 4	0.905	0.192	0.518	-2.839E-01	4.191E-05	1.550E-03
GLCM - Energy	0.396	0.779	0.589	4.905E-02	4.483E-05	1.646E-03
GLCM - Homogeneity	0.816	0.324	0.569	5.367E-01	5.825E-05	2.081E-03
GLSZM - LZHGE	0.759	0.434	0.583	2.186E+06	6.177E-05	2.192E-03
GLRLM - SRE	0.767	0.380	0.563	8.222E-01	6.522E-05	2.302E-03
GLSZM - ZLNU	0.748	0.442	0.587	5.574E+02	7.076E-05	2.475E-03
GLRLM - RLNUr	0.828	0.303	0.542	7.453E-01	7.956E-05	2.747E-03
1^{st} order - I_{max}	0.676	0.459	0.537	2.240E+02	1.006E-04	3.384E-03
GLDM - Strength	0.715	0.536	0.613	1.492E-01	1.007E-04	3.386E-03
GLCM - Correlation	0.786	0.342	0.549	8.367E-01	1.047E-04	3.505E-03
MC-entropy PCA 2	0.276	0.829	0.523	2.350E+00	1.224E-04	4.024E-03
GLCM - Entropy	0.722	0.455	0.574	2.009E+00	1.261E-04	4.132E-03
GLRLM - HGRE	0.843	0.358	0.598	1.060E+02	1.511E-04	4.845E-03
GLCM - IDM	0.789	0.351	0.567	4.895E-01	1.936E-04	6.024E-03
GLCM - Cluster Shade	0.710	0.530	0.606	5.198E+02	2.349E-04	7.135E-03
GLCM - Contrast	0.322	0.806	0.560	6.199E+00	2.357E-04	7.156E-03
GLDM - Contrast	0.627	0.595	0.604	9.151E-02	2.554E-04	7.676E-03
1^{st} order - I_{sd}	0.669	0.476	0.540	2.062E+02	2.827E-04	8.388E-03
1^{st} order - Kurtosis	0.583	0.662	0.612	6.758E+00	3.620E-04	1.040E-02
MC-energy PCA 2	0.755	0.397	0.551	-8.573E-01	3.683E-04	1.055E-02
granularity-max PCA 2	0.750	0.377	0.540	4.361E-01	4.170E-04	1.175E-02
GLSZM - HGZE	0.507	0.636	0.567	1.306E+02	4.222E-04	1.188E-02
MC-cov PCA 1	0.423	0.762	0.586	2.759E+00	4.665E-04	1.295E-02
GLSZM - SRHGE	0.914	0.268	0.590	7.728E+01	7.798E-04	2.011E-02
GLRLM - SRLGE	0.914	0.210	0.539	8.013E-02	8.777E-04	2.223E-02
MC-ALLFEAT PCA 5	0.804	0.333	0.559	-3.882E+00	1.071E-03	2.631E-02
granularity-skewness PCA 5	0.652	0.550	0.592	-2.321E-03	1.211E-03	2.917E-02
MC-volume PCA 1	0.892	0.317	0.612	-4.165E+00	1.282E-03	3.062E-02
MC-kurt PCA 3	0.717	0.539	0.615	6.202E-01	1.410E-03	3.315E-02
granularity-cov PCA 6	0.627	0.456	0.502	-5.808E-03	1.411E-03	3.318E-02
granularity-standev PCA 1	0.714	0.481	0.598	-1.667E-01	1.550E-03	3.589E-02
GLRLM - LGRE	0.879	0.250	0.543	9.532E-02	1.692E-03	3.860E-02
granularity ALLFEAT PCA 2	0.599	0.532	0.549	-6.546E-01	1.708E-03	3.890E-02
1^{st} order - Skewness	0.550	0.667	0.582	-1.766E+00	1.985E-03	4.407E-02
MC-skew PCA 8	0.647	0.506	0.561	2.661E-02	1.991E-03	4.419E-02
MC-skew PCA 5	0.694	0.483	0.580	4.810E-01	2.107E-03	4.630E-02
MC-N300 030 003 PCA 7	0.905	0.243	0.578	-1.480E-01	2.166E-03	4.736E-02

type of analysis that have many features: the more we test variables, the more we are likely to find one that passes the test.

This is why PCA was used to reduce the number of features from 1600 down to 271 which is not a lot by Radiomic standards.

The first year horizon is important because it is where half of the patients die, making the data balanced. Also the features highlighted in the analysis within the 1 year horizon are more likely to highlight more violent tumours that cause a faster death to the patient. In this case, the *homogeneous run percentage* of the gray-level run length matrix appears to be more correlated to faster death (Table 3.4), and has seemingly less importance in longer survival analysis.

The survival curves (Fig. 3.9) show that the selected features with their optimal cutoff are good at separating the population into two groups with a clearly different survival time. The more spread out the curves are, the more effective the feature is. The curves also show that the features separate the patients differently, resulting in different curves.

It is however worth to note that the selected features are overall the same in between the horizons of the survival analysis, and slight disparities might be caused by chance. However, statistical hypothesis testing guarantees that the selected features are relevant. The Kaplan-Meier curves also show the similarity between the survival patterns.

The Kaplan-Meier survival analysis is extremely powerful due to the estimator being conceived for this specific case of handling right-censored data. However, this uni-variate approach is not enough to build the best predictor that could be used across data-sets. This is why the next section will focus on using machine learning classification and regression algorithms.

3.6. Classification using machine learning

Building an accurate predictor for a prognostic is the ultimate goal of Radiomics. It would allow to give an automated prognostic based on purely medical imagery, maybe helped by information obtained from the patient in a non invasive way, such as loss of weight (which is an incredibly effective biomarker).

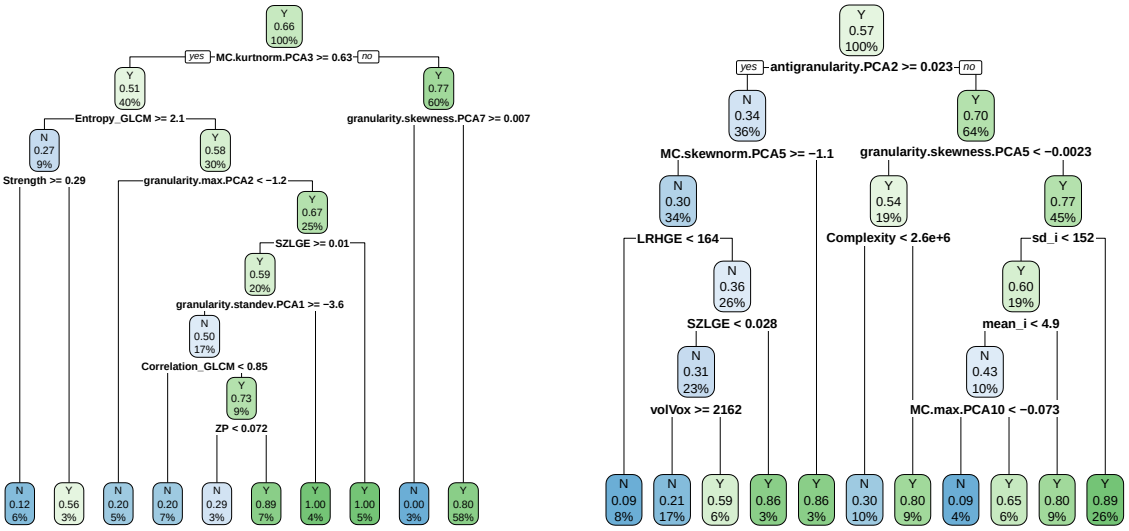
However using machine learning for prognostic suffers from the same problem that Kaplan and Meier managed to solve with their survival analysis method: how to handle right-censored data? Namely, how to regress / predict the survival time when the surviving patients have not died yet?

This is why this section will be more of an experimentation rather than the rigorous statistical analysis that the previous section was. This section will use the validation set that was carefully isolated from any kind of calculations, in order to avoid over-fitting. But as the validation set only has 20 elements (which is the standard for this kind of dataset), it will not be enough to prove anything. This is also why we will not optimize the model using manual cross validation schemes, because doing so on only one data-set is not meaningful and can be some kind of over-fitting (due to the confirmation bias). The goal here is mostly to compare the statistical approach to the machine learning one.

Finally, due to the fact that regression of the survival time without taking into account right-censorship is meaningless, we will concentrate on the classification of survival using decision trees.

3.6.1. Decision trees methodology

Decision trees are extremely interesting in this case of feature analysis, as decision trees on continuous data also compute a threshold on which to binarise the feature, in a similar way that the



(a) Final Death status

(b) Death status after two years

Figure 3.10.: Decision tree for the classification of death status

	Prediction outcome		
	Survival	Death	Total
Survival	87	21	108
Death	33	136	169
Total	120	157	277

(a) *Train accuracy* = 0.805

	Prediction outcome		
	Survival	Death	Total
Survival	7	6	13
Death	1	6	7
Total	8	12	20

(b) *Test accuracy* = 0.65

Figure 3.11.: Classification of death status after two years

optimal threshold was computed for the Kaplan-Meier survival analysis.

Furthermore, decision trees also order features by relevance, putting the features deemed by the algorithm as most relevant on top of the tree. The algorithm also makes feature selection, as the tree is pruned in order to avoid over-fitting.

The result is thus similar in nature as the one of the survival analysis, and comparing the choices of the decision tree algorithm to the ones made in the survival analysis is interesting.

Whilst most of the code in this part was made in MATLAB, this part was implemented in R in order to use the RPART (Recursive Partitioning And Regression Trees [48]) package, which is the best package of this sort available.

The defaults of RPART are very good, including 10 k-fold cross validation, pruning, and a good balance between bias and variance. The testing is done on the independent validation set of 20 patients.

3.6.2. Results

When analyzing decision trees the first thing to look at is the feature selection done by the algorithm. Unlike the Kaplan-Meier analysis in which the best threshold for each variable was chosen from the ROC curves in order to maximize the *precision* - *Recall* for each feature individually, the decision tree algorithm tries to maximize the classification rate for the whole tree. This is why the order of the feature selection is not the same as the one that was achieved after the survival analysis

	Prediction outcome		
	Survival	Death	Total
Survival	55	11	66
Death	38	174	213
Total	93	184	277

(a) *Train accuracy* = 0.823

	Prediction outcome		
	Survival	Death	Total
Survival	4	3	7
Death	2	11	13
Total	6	14	20

(b) *Test accuracy* = 0.75

Figure 3.12.: Classification of final death status

according to the *p*-value.

However, the features chosen by the decision tree were among the best chosen in the survival analysis. The cutoff-threshold values are also only slightly different, despite the difference in their determination.

The biggest strength of a decision tree is also its weakness: the way it prunes (only conserves the first levels of the tree) does not make use of all the available features. Furthermore, the results are decided by the features selected by the tree making algorithm, meaning that a slight change in the training data might lead to a completely different tree and thus classification rates. Finally, in a similar way to the uni-variate Kaplan-Meier analysis, the decision tree binarises the features. The biggest problem is that it becomes then unable to use the continuous data (a decision tree will consider all tumour volumes the same as long as they are above the threshold).

A random forest overcomes some of these limitations by using multiple trees, but the point here is not to build the best classifier but instead to compare the decisions made by the decision trees to those made by Kaplan-Meier, and they interestingly have similar results.

Indeed both K-M and the decision trees value the information about the texture of the tumour above its volume, which shows that even in CT, texture is a very important prognostic factor. Even more interestingly, instead of using the uni-variate threshold of around 1mL, the decision tree uses volume as a final deciding factor, and sentences the patient when the tumor volume is above the threshold of 2.1mL.

Before looking at the classification results it is worth to note that in order to judge a classifier, classification rate is clearly not sufficient, especially when the data is unbalanced. In the case of the final survival status, a simple "All Death" classifier can get 77% classification rate on this dataset, which might look good and would be fairly easy to implement (NSCLC is indeed deadly) but is in fact a very bad classifier. This is why the confusion matrix is necessary in order to judge a classifier.

As for the results of the classification on the training and validation sets, the decision tree computed to classify the death status after two years had average results on both the test set and the validation set (Fig. 3.11).

However, the decision tree made for the classification of the final death status outputs a surprisingly good accuracy, with a good looking confusion matrix (Fig. 3.12). Because the test set is completely independent, it shows that prediction from the computed features is indeed possible.

3.7. Conclusion

The results of the experimental part of this thesis are clearly positive, as the rigorous Kaplan-Meier proved statistically that the computed mathematical morphology based image bio-markers features are relevant and at least as good as classical radiomics image bio-markers.

The initial high dimensional mathematical morphology based image bio-makers were very inter-correlated and too numerous, and this is why PCA was used for dimentionality reduction and was successfully shown to concentrate most of the information in the first principal components.

The correlation analysis also showed the correlation between the classical image bio-markers and the mathematical morphology based ones. These correlations are both useful for the goal of this master thesis that is to prove the relevance of these features: having them correlated to classical proven image bio-markers further attests of their relevance.

However having them correlated also means that they do not bring a lot of additional information, making it harder to prove their true usefulness and superiority when it comes to prognosis. Nevertheless, the correlation between classical image bio-markers and mathematical morphology based ones is relative as it depends on the metric used for the morphological series, and also because PCA is effective at de-correlation.

Finally the feature selection made by the decision trees was not very different to the ones made during the Kaplan-Meier analysis, and it reinforces the relevance of the mathematical morphology based features by selecting them at the root of the trees in most cases. The classification rates on the validation sets were also relatively good for a non-tuned simple decision tree, especially for the classification of the final death status with a good confusion matrix. However the goal of the classification part of this chapter is not to build the best classifier but rather to compare the feature selections.

Building better predictors can be possible with support vector machines(SVM) and other techniques and is a natural follow-up of this research. Making good predictors for multi-factorial time based survival with just a single snapshot image of a tumour is something that is incredibly hard to do no matter how good the image bio-markers are. However combined with other non-radiomics information such as weight loss, age, sex as well as using the evolution of the image bio-markers with time should help complement radiomics information in order to make better predictors for non-invasive personalized medicine.

Conclusion

This thesis aimed to develop mathematical morphology-based texture descriptors in order to apply them to Radiomics. Radiomics aims to provide a personal and non invasive approach to medicine, using quantitative data analysis on features extracted from biomedical images.

By comparing our new mathematical morphology-based features - which are used for the first time in Radiomics - to the classical Radiomics features, we managed to show that the mathematical morphology-based features are at least as good as the classical features extraction techniques. Adding new features to the Radiomics feature "mix" enhances the performance that can be achieved by the subsequent data analysis techniques.

We first took a look the the mathematical foundations of mathematical morphology, and their relation to classical signal analysis techniques. As our analysis was based on N-dimensional signals and complete lattice spaces, it allowed a good generalization of these tools (usually applied to 2D images) to the 3D biomedical images analyzed by Radiomics. We then studied the concept of morphological series in order to extract scalar features from a technique that is designed as a homomorphic transform.

The following step was to analyze the problematic of Radiomics, from image acquisition using biomedical scanners to the computation of classical Radiomic features, namely first order features and gray level matrix based texture descriptors.

Finally, the mathematical morphology-based features were implemented and experimented on an open database of segmented Non-small cells of lung cancer tumours, captured using 3d X-ray computational tomography scans. In order to overcome the very high dimentionality of the mathematical morphology-based features, we did a correlation analysis followed by a reduction of dimentionalty. A rigorous Kaplan-Meier survival analysis was then performed in order to select the features that are statistically proven to be relevant. A different machine learning approach was then tried and the feature selection of the machine learning algorithm was compared to the results of the survival analysis, as well as the accuracy tested on an independent validation set.

After dimentionality reductions, the mathematical morphology based features where proven by the statistical analysis to be at least as good, if not better than the classical Radiomic features. This is, as such, a satisfying result, as the features we used as comparison are state-of-the art features used extensively in Radiomics. In addition, the performances were also carried out to the independent validation set, as the machine learning algorithm used the new mathematical morphology features to a relative success in prediction on the validation test.

However, it is hard to distinguish these new features from the classical ones, as the correlation analysis showed strong similarity between classical and morphology based features. It is a good thing because it validates their effectiveness, but it does not prove them superior to the classical features.

Our contribution has brought to the table new image biomarkers based on mathematical morphology that were statistically proven effective on this database, and could be used for future Radiomics research and increased personalized medicine.

Future Works

Even though these features were proven relevant on this dataset, they would need to be tested on other datasets in order to validate their effectiveness.

In addition, texture analysis is challenging with X-ray computational tomography as it just passively measures the density of tissues, which does not in fact vary that much within a tumour. This is why better results could probably be obtained using MRI and PET. MRI because instead of the density, it shows the types of tissues as well as a lot of active functional information about tissues (for example blood oxygenation). PET is at the genesis of Radiomics, as it allows to explicitly visualize tumour activity.

As such, MRI and PET imagery are seemingly better suited for texture analysis, and thus should bring better results with our mathematical morphology-based image biomarkers.

Furthermore, instead of only using a tumour texture, the shape of the tumour could also be used in order to make new image bio-markers. Indeed the technique of padding that was developed in this research makes total abstraction of the border and shape of the tumour, in order to focus on the texture itself. However the information given by the shape of the tumour could also be relevant, and used as new image bio-markers. Indeed, mathematical *morphology* is after all the mathematical analysis of shapes.

Finally, the biggest current usage of mathematical morphology in biomedical fields is for segmentation. Mathematical morphology is proven very effective in this domain, and can be used for automated tumour ROI (Region of Interest) extraction, a step necessary in this final goal of Radiomics used in automated personalized medicine.

Chapter A

Appendix

A.1. TNM classification

The TNM classification is an international system for classifying cancers according to their anatomical extent. The three letters symbolise the spread of the cancerous disease on the site of the primary tumour (T), in the neighbouring lymph nodes (N) and at a distance for possible metastases (M).

Stage T is evaluated with clinical examination, imaging and/or surgical exploration. This characteristic corresponds to the depth of invasion of the lung wall by the primary tumour and its development in the surrounding tissues:

- T0: no sign of a primary tumour
- Tx: tumour cannot be assessed
- Tis: carcinoma in situ
- T1, T2, T3, T4: size and/or extension of the primary tumour

Stage N is the number of surrounding lymph nodes that contain cancerous cells and their location:

- N0: no sign of regional lymph node involvement
- N1: spread in 1 or 2 neighbouring lymph nodes
- N2: spread in 3 to 6 neighbouring lymph nodes
- N3: spread in more than 7 neighbouring lymph nodes

Stage M (metastasis) is the spread of cancer to other parts of the body:

- M0: absence of distant metastasis
- M1: presence of metastasis, the cancer has spread to another part of the body

Once these characteristics are defined they are combined together forming the TNM stage (Table A.1).

Table A.1.: Grouping of the T (tumour), N (nodes) and M (metastasis) stages into TNM stages.

TNM stage	**T stage**	**N stage**	**M stage**
Stage 0	T *in situ*	N0	M0
Stage IA	T1	N0	M0
Stage IB	T2	N0	M0
Stage IIA	T3	N0	M0
Stage IIB	T1, T2	N1	M0
Stage IIIA	T4a	N0	M0
-	T3	N1	M0
-	T1, T2	N2	M0
Stage IIIB	T3	N2	M0
Stage IIIC	T4a	N1, N2	M0
-	T4b	all N	M0
-	all T	N3	M0
Stage IV	all T	all N	M1

A.2. Texture matrices

[12]

A.2.1. Grey-level Co-occurrence Matrix (GLCM)

The GLCM is built as follows. An element $C(i,j)$ of the GLCM fro an image I of size $\omega = X \times Y \times Z$ is defined as the probability of having 2 voxels of intensity i and j, separated by a vector (d,θ) corresponding to a displacement $(\Delta x, \Delta y, \Delta z)$:

$$C_{d,\theta}(i,j) = \frac{1}{\omega} \sum_{x,y,z}^{X,Y,Z} \begin{cases} 1, \text{ if } I(x,y,z) = i \text{ et } I(x+\Delta x, y+\Delta y, z+\Delta z) = j \\ 0 \text{ else} \end{cases} \tag{A.1}$$

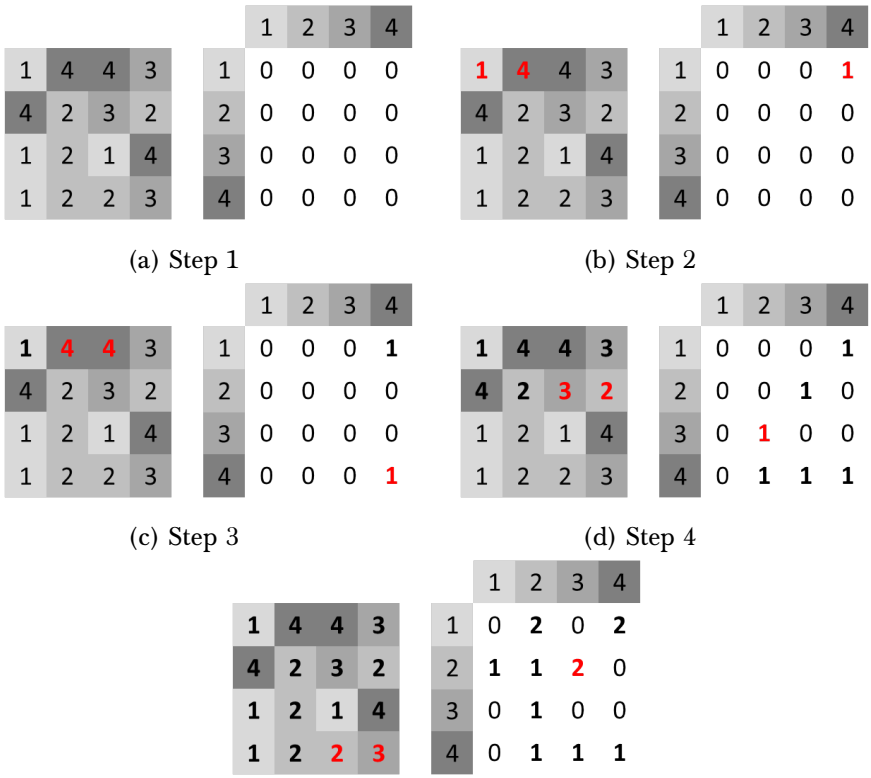
Where the image I can be considered as a function of three variables x, y and z with $x \in [1,X]$, $y \in [1,Y]$ and $z \in [1,Y]$. $I(x,y,z)$ can take discrete values such that $i = 0,1,\ldots,G-1$, where G is the number of grey levels in the image. The construction of a matrix is shown in Figure A.1.

A.2.2. Grey-level Run Length Matrix (GLRLM)

The construction of GLRLM is shown in Figure A.2. The first element of the matrix corresponds to the iteration number that a voxel of intensity equal to one has a neighbour of the same intensity in the 0 degree direction.

A.2.3. Grey-level Size Zone Matrix (GLSZM)

The construction of GLSZM is shown in Figure A.3. The first element of the matrix corresponds to the iteration number that a voxel of intensity equal to one has a succession of neighbours of the same intensity in its neighbourhood.

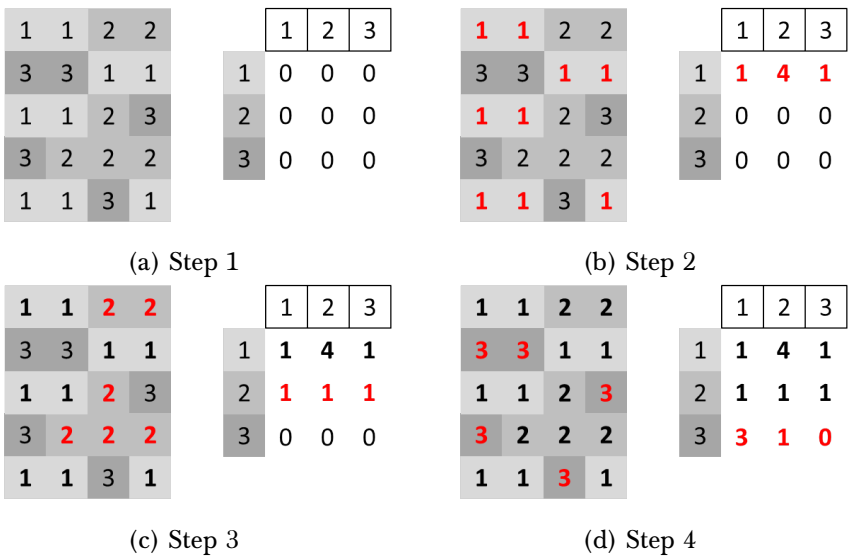


(a) Step 1 (b) Step 2

(c) Step 3 (d) Step 4

(e) Step 5

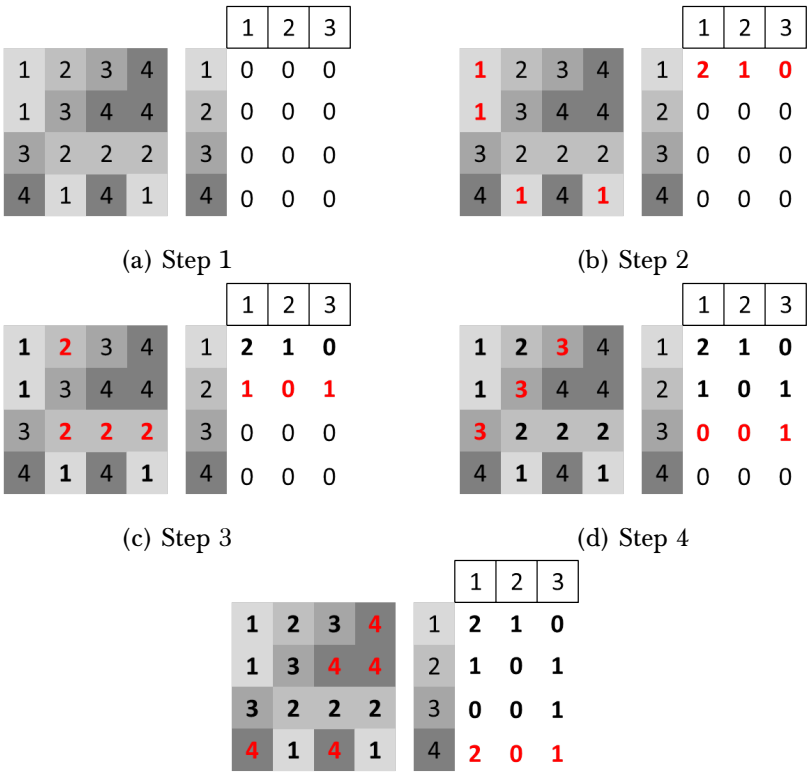
Figure A.1.: Illustration of the stepwise construction of a co-occurrence matrix based on a couple $\theta = 0$ and $d = 1$.



(a) Step 1 (b) Step 2

(c) Step 3 (d) Step 4

Figure A.2.: Illustration of the stepwise construction of a grey-level Run Length matrix based on a couple $\theta = 0$ and $d = 1$.



(a) Step 1 (b) Step 2

(c) Step 3 (d) Step 4

(e) Step 3

Figure A.3.: Illustration of the stepwise construction of a grey-level size zone matrix.

A.2.4. Grey-level Difference Matrix (GLDM)

The construction of GLDM is done in the following way. The element $N(i,1)$ corresponds to the probability of occurrence of the level i and the element $(i,2)$ corresponds to the sum of the difference in absolute value between the voxels of intensity i and the average of their respective neighbourhoods:

$$N(i,2) = \sum_{x,y,z}^{X,Y,Z} \begin{cases} 1, \ |\overline{M}(x,y,z) - i| \text{ if } I(x,y,z) = i \\ 0, \text{ else} \end{cases} \tag{A.2}$$

Where $\overline{M}(x,y,z)$ represents the mean of the neighbours intensity of the coordinates voxel (x,y,z).

Bibliography

[1] HJWL Aerts et al. “Data from NSCLC-radiomics”. In: *Cancer Imaging Archive* (2015).

[2] Hugo J.W.L. Aerts et al. “Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach”. In: *Nature Communications* 5 (2014). ISSN: 20411723. DOI: 10.1038/ncomms5006.

[3] Steffen Löck Alex Zwanenburg, Stefan Leger, Martin Vallières. “Image biomarker standardisation initiative. Reference manual”. In: (2016). DOI: 10.17195/candat.2016.08.1. arXiv: arXiv:1612.07003v6. URL: https://www.cancerdata.org/resource/doi:10.17195/candat.2016.08.1.

[4] Altman D G, Lausen B, Sauerbrei W. “Dangers of using "optimal" cutpoints in the evaluation of prognostic factors”. In: *Journal of the National Cancer Institute* 86.23 (1994), pp. 1798–1799. ISSN: 00278874. DOI: 10.1093/jnci/86.23.1798-a.

[5] Moses Amadasun and Robert King. “Textural features corresponding to textural properties”. In: *IEEE Trans. Syst. Man Cybern.* 19.5 (1989), pp. 1264–1273. ISSN: 00189472. DOI: 10.1109/21.44046.

[6] Erchan Aptoula and Sébastien Lefèvre. “Morphological texture description of grey-scale and color images”. In: *Advances in Imaging and Electron Physics*. Vol. 169. Mm. 2011, pp. 1–74. DOI: 10.1016/B978-0-12-385981-5.00001-X.

[7] Amir Averbuch and Yoel Shkolnisky. “3D Fourier based discrete Radon transform”. In: 15 (2003), pp. 33–69. DOI: 10.1016/S1063-5203(03)00030-7.

[8] Robert Azencott, Jia-Ping Wang, and Laurent Younes. “Texture classification using windowed Fourier filters”. In: *IEEE Transactions on Pattern Analysis & Machine Intelligence* 2 (1997), pp. 148–153.

[9] Thomas Bortfeld. “Image Reconstruction 3 – Fully 3D Reconstruction The 3D Radon transform and its inverse Grangeat ’ s trick Tuy theorem Helical scanning”. In: (2013).

[10] Emmanuel Candes. *Applied Fourier Analysis and Elements of Modern Signal Processing*. 2016.

[11] Anastasia Chalkidou, Michael J. O’Doherty, and Paul K. Marsden. “False discovery rates in PET and CT studies with texture features: A systematic review”. In: *PLoS ONE* 10.5 (2015), pp. 1–18. ISSN: 19326203. DOI: 10.1371/journal.pone.0124165.

[12] Paul Desbordes. “Méthode de sélection de caractéristiques pronostiques et prédictives basée sur les forêts aléatoires pour le suivi thérapeutique des lésions tumorales par imagerie fonctionnelle TEP”. PhD thesis. Normandie Université, 2017.

[13] Belma Dogdas, David W. Shattuck, and Richard M. Leahy. “Segmentation of skull and scalp in 3-D human MRI using mathematical morphology”. In: *Human Brain Mapping* 26.4 (2005), pp. 273–285. ISSN: 10659471. DOI: 10.1002/hbm.20159.

[14] Edward R Dougherty and Jaakko Astola. *An introduction to nonlinear image processing*. Vol. 16. SPIE press, 1994.

[15] Ruoyu Du and Hyo Jong Lee. "An improved region growing method for scalp and skull extraction based on mathematical morphology operations". In: *Proceedings - 4th International Congress on Image and Signal Processing, CISP 2011* 3 (2011), pp. 1201–1204. DOI: 10.1109/CISP.2011.6100498.

[16] Nashwa El-Bendary et al. "Retinal Blood Vessel Segmentation Approach Based on Mathematical Morphology". In: *Procedia Computer Science* 65.Iccmit (2015), pp. 612–622. ISSN: 18770509. DOI: 10.1016/j.procs.2015.09.005. URL: http://dx.doi.org/10.1016/j.procs.2015.09.005.

[17] Issam El Naqa et al. "Exploring feature-based approaches in PET images for predicting cancer treatment outcomes". In: *Pattern Recognit.* 42.6 (2009), pp. 1162–1171. ISSN: 00313203. DOI: 10.1016/j.patcog.2008.08.011. arXiv: NIHMS150003.

[18] Jan Flusser, Barbara Zitova, and Tomas Suk. *Moments and moment invariants in pattern recognition*. John Wiley & Sons, 2009.

[19] Mary M. Galloway. "Texture analysis using gray level run lengths". In: *Comput. Graph. Image Process.* 4.2 (1975), pp. 172–179. ISSN: 0146-664X. DOI: http://dx.doi.org/10.1016/S0146-664X(75)80008-6. URL: http://www.sciencedirect.com/science/article/pii/S0146664X75800086.

[20] Robert M Haralick, K Shanmugam, and Its'hak Dinstein. *Textural features for image classification*. 1973. DOI: 10.1109/TSMC.1973.4309314.

[21] Mathieu Hatt et al. "18F-FDG PET Uptake Characterization Through Texture Analysis: Investigatin the Complementary Nature of Heterogeneity and Functional Tumor Volume in a Multi–Cancer Site Patient Cohort". In: *J. Nucl. Med.* 56.1 (2015), pp. 38–44. ISSN: 1535-5667. DOI: 10.2967/jnumed.114.144055. URL: http://jnm.snmjournals.org/content/56/1/38.abstractN2http://jnm.snmjournals.org/cgi/doi/10.2967/jnumed.114.144055.

[22] Mathieu Hatt et al. "Baseline 18F-FDG PET image-derived parameters for therapy response prediction in oesophageal cancer". In: *Eur. J. Nucl. Med. Mol. Imaging* 38.9 (2011), pp. 1595–1606. ISSN: 16197070. DOI: 10.1007/s00259-011-1834-9.

[23] Mathieu Hatt et al. "Prognostic value of 18F-FDG PET image-based parameters in oesophageal cancer and impact of tumour delineation methodology". In: *Eur. J. Nucl. Med. Mol. Imaging* 38.7 (2011), pp. 1191–1202. ISSN: 16197070. DOI: 10.1007/s00259-011-1755-7.

[24] Henk JAM Heijmans and Christian Ronse. "The algebraic basis of mathematical morphology I. Dilations and erosions". In: *Computer Vision, Graphics, and Image Processing* 50.3 (1990), pp. 245–295.

[25] S. Herlidou et al. "Comparison of automated and visual texture analysis in MRI: Characterization of normal and diseased skeletal muscle". In: *Magn. Reson. Imaging* 17.9 (1999), pp. 1393–1397. ISSN: 0730725X. DOI: 10.1016/S0730-725X(99)00066-1.

[26] Frank Hofheinz et al. "Increased evidence for the prognostic value of primary tumor asphericity in pretherapeutic FDG PET for risk stratification in patients with head and neck cancer". In: *Eur. J. Nucl. Med. Mol. Imaging* 42.3 (2014), pp. 429–437. ISSN: 16197089. DOI: 10.1007/s00259-014-2953-x.

[27] Benoit Macq J-P Thiran. *Morphological Feature Extraction for the Classification of Digital Images of Cancerous Tisssues*.

[28] E . L . Kaplan and MeierPaul. "Nonparametric Estimation from Incomplete Observations Reviewed work (s): Source : Journal of the American Statistical Association , Vol . 53 , No . 282 (Jun ., 1958), pp . 457- Published by : American Statistical Association Stable URL : http://www.js". In: 53.282 (1958), pp. 457–481.

[29] Branislav Kisaäanin and Dan Schonfeld. “A Fast Thresholded Linear Convolution Representation of Morphological Operations”. In: *IEEE Transactions on Image Processing* 3.4 (1994), pp. 455–457. ISSN: 19410042. DOI: 10.1109/83.298399.

[30] Eric W. Koch and Erik W. Rosolowsky. “Filament identification through mathematical morphology”. In: *Monthly Notices of the Royal Astronomical Society* 452.4 (2015), pp. 3435–3450. ISSN: 13652966. DOI: 10.1093/mnras/stv1521. arXiv: arXiv:1507.02289v1.

[31] Virendra Kumar et al. “Radiomics: the process and the challenges”. In: *Magnetic resonance imaging* 30.9 (2012), pp. 1234–1248.

[32] Philippe Lambin et al. “Radiomics: extracting more information from medical images using advanced feature analysis”. In: *European journal of cancer* 48.4 (2012), pp. 441–446.

[33] Philippe Lambin et al. “Radiomics: Extracting more information from medical images using advanced feature analysis”. In: *Eur. J. Cancer* 48.4 (2012), pp. 441–446. ISSN: 09598049. DOI: 10.1016/j.ejca.2011.11.036.

[34] Andrea Loddo, Cecilia Di Ruberto, and Michel Kocher. “Recent advances of malaria parasites detection systems based on mathematical morphology”. In: *Sensors (Switzerland)* 18.2 (2018), pp. 1–21. ISSN: 14248220. DOI: 10.3390/s18020513.

[35] William E Lorensen and Harvey E Cline. “Computer Graphics, Volume 21, Number 4, July 1987”. In: *ACM siggraph computer graphics* 21.4 (1987), pp. 163–169.

[36] CL Luengo Hendriks. “Structure characterization using mathematical morphology”. In: (2004).

[37] Benoit Macq and Jean-didier Legat. “RASH : RAdon Soft Hash algorithm”. In: (), pp. 3–6.

[38] Maxime Moreaud and Frederic Itthirad. “Fast algorithm for dilation and erosion using arbitrary flat structuring element: Improvement of urbach and Wilkinson’s algorithm to GPU computing”. In: *International Conference on Multimedia Computing and Systems -Proceedings*. 2014, pp. 289–294. ISBN: 9781479938247. DOI: 10.1109/ICMCS.2014.6911245.

[39] Edward D Nicol et al. “The Future of Cardiovascular Computed Tomography: Advanced Analytics and Clinical Insights”. In: *JACC: Cardiovascular Imaging* 12.6 (2019), pp. 1058–1072.

[40] Nobuyuki Otsu. “A threshold selection method from gray-level histograms”. In: *IEEE transactions on systems, man, and cybernetics* 9.1 (1979), pp. 62–66.

[41] AG Ramakrishnan, S Kumar Raja, and HV Raghu Ram. “Neural network-based segmentation of textures using Gabor features”. In: *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*. IEEE. 2002, pp. 365–374.

[42] M Razaz and DMP Hagyard. “Efficient convolution based algorithms for erosion and dilation”. In: vol. 1. 1999, pp. 360–363.

[43] Cédric De Roover et al. “Robust Video Hashing Based on Radial Projections of Key Frames”. In: 53.10 (2005), pp. 4020–4037.

[44] Rosniza Roslan, Nursuriati Jamil, and Rozi Mahmud. “Skull Stripping Magnetic Resonance Images Brain Images : Region Growing versus Mathematical Morphology”. In: *International Journal of Computer Information Systems and Industrial Management Applications* 3 (2011), pp. 150–158.

[45] Jean Serra. *Image analysis and mathematical morphology*. Academic Press, Inc., 1983.

[46] P. S. Shijin Kumar and V. S. Dharun. “An efficient skull stripping algorithm using connected regions and morphological operation”. In: *ARPN Journal of Engineering and Applied Sciences* 11.7 (2016), pp. 4305–4309. ISSN: 18196608.

[47] Olivier Strauss and Kevin Loquin. “Linear filtering and mathematical morphology on an image: A bridge”. In: *Proceedings - International Conference on Image Processing, ICIP*. 2009, pp. 3965–3968. ISBN: 9781424456543. DOI: 10.1109/ICIP.2009.5413800.

[48] Terry Therneau et al. "Package 'rpart'". In: *Available online: cran. ma. ic. ac. uk/web/packages/r-part/rpart. pdf (accessed on 20 April 2016)* (2015).

[49] Guillaume Thibault et al. "Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification". In: *Pattern Recognit. Inf. Process.* (2009), pp. 140–145. DOI: Artn1357002\rDoi10.1142/S0218001413570024.

[50] Janna E van Timmeren et al. "Challenges and caveats of a multi-center retrospective radiomics study: an example of early treatment response assessment for NSCLC patients using FDG-PET/CT radiomics". In: *PloS one* 14.6 (2019), e0217536.

[51] M Vassholz et al. "A new x-ray tomography method based on the 3d Radon transform compatible with anisotropic sources, year = 2015". In: (), pp. 1–12.

[52] H Vernieuwe. "A mathematical morphology approach to the identification of drought events in space and time". In: ().

[53] Jiao Wang et al. "An iterative Black Top Hat Transform algorithm for the volume estimation of lunar impact craters". In: *Remote Sensing* 9.9 (2017), pp. 1–10. ISSN: 20724292. DOI: 10.3390/rs9090952.

[54] Stephen SF Yip and Hugo JWL Aerts. "Applications and limitations of radiomics". In: *Physics in Medicine & Biology* 61.13 (2016), R150.

[55] Xia Zhang et al. "Denoising vegetation spectra by combining mathematical-morphology and wavelet-transform-based filters". In: *Journal of Applied Remote Sensing* 13.01 (2019), p. 1. DOI: 10.1117/1.jrs.13.016503.

List of Figures

List of Tables

Rue Archimède, 1 bte L6.11.01, 1348 Louvain-la-Neuve www.uclouvain.be/epl