

Capítulo

1

Precision Radiomic Biomarkers: a brief introduction, some technical development, and several clinical applications

José Raniery Ferreira Junior

Abstract

Computer tools have been part of the clinical routine on a daily basis for radiological image interpretation. However, they are limited to quantifying basic information about the lesions that a medical examination may present, like a nodule size or mass volume. Radiomic biomarkers emerged in this context to address this problem by quantifying the images massively and characterizing them comprehensively to allow the precision phenotyping needed in the current personalized medicine era. Thus, this book chapter introduces some robust radiomic biomarkers identified in the past few years for different pathological imaging patterns of diseases from two critical human systems, i.e., respiratory and musculoskeletal. The text initiates with the primary motivation for radiomic biomarker development, discovery, and validation. The following sections present a quick background with the basic theory for the remainder of the manuscript. Finally, the chapter approaches the state-of-the-art radiomic precision biomarkers for three different diseases and modalities of medical images: covid-19 in chest radiography, lung neoplasms in computed tomography, and spondyloarthritis in magnetic resonance imaging.

1.1. Introduction

Medical imaging is an essential component to evaluate patients with a suspicion of several diseases. Beyond the importance, radiological images are still mostly dependent on the diagnostic level of specialists, and rapid imaging interpretation is not always possible as it relies on the availability of expert physicians [Liang and Zheng 2019, Azevedo-Marques and Ferreira Junior 2021]. Furthermore, the imaging appearance of different diseases can

overlap (Figure 1.1) and mimic other body complications due to the low contrast that the examination may have [Santos et al. 2019, Ferreira Junior et al. 2020a].

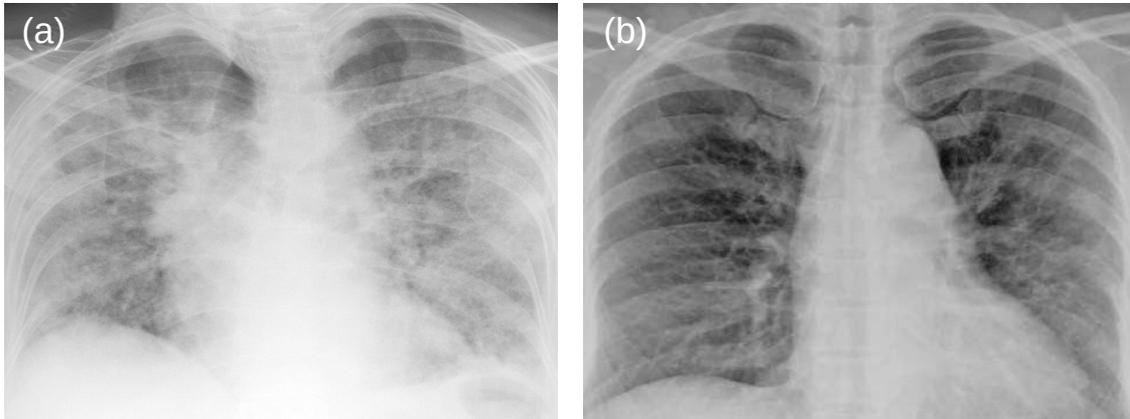


Figura 1.1. Overlapping radiographic characteristics of SARS (a) and covid-19 (b). Both diseases commonly present bilateral ground-glass opacities, as shown in the figures. Source: Author.
SARS, severe acute respiratory syndrome.
covid-19, coronavirus disease 2019.

In the face of these challenges, it is vital to include computer-based tools to aid specialists in evaluating diseases early, as they can improve the accuracy and consistency of medical image interpretation through computational support used as reference [Ferreira Junior et al. 2017]. Radiomics has grown in this context as a quantitative imaging approach that associates computer-extracted image data with clinical endpoints (Figure 1.2). This radiomic association allows a more comprehensive characterization of underlying phenotypes, ultimately increasing the power of decision support models for precision medicine [Tomaszewski and Gillies 2021]. In light of recent advances in target therapies, the need for an inexpensive and easily obtainable imaging approach for phenotyping diseases has become imperative, and radiomics can provide it as is a noninvasive, reproducible tool [Sacconi et al. 2017].

Radiomics emerged as a quantitative approach for personalized medicine to develop medical imaging biomarkers and predictive models for clinical decision support [Santos et al. 2019]. At first, it was based on the extraction of hand-designed features from previously segmented images with the potential to be associated with clinical outcomes (Figure 1.3). During this initial moment, traditional statistical methods, such as test-retest, stability, univariate inference tests, and multivariate regression models, assessed the radiomics associations [Aerts et al. 2014, Sacconi et al. 2017, Ferreira Junior et al. 2020b]. Then, radiomics studies with multivariate analyses started focusing on artificial intelligence and machine learning methods to improve the predictive power of the imaging biomarkers. Some of the machine-learning models studied were standard Bayesian methods, artificial neural networks, and decision trees [Leger et al. 2017, Dawes et al. 2017, Kickingreder et al. 2016].

The main applications of radiomics during those early phases were on the oncological domain for predicting genomic mutations and tumor staging, for instance, [Kickingreder et al. 2016, Aerts et al. 2014, Sacconi et al. 2017, Ferreira Junior et al. 2021d].

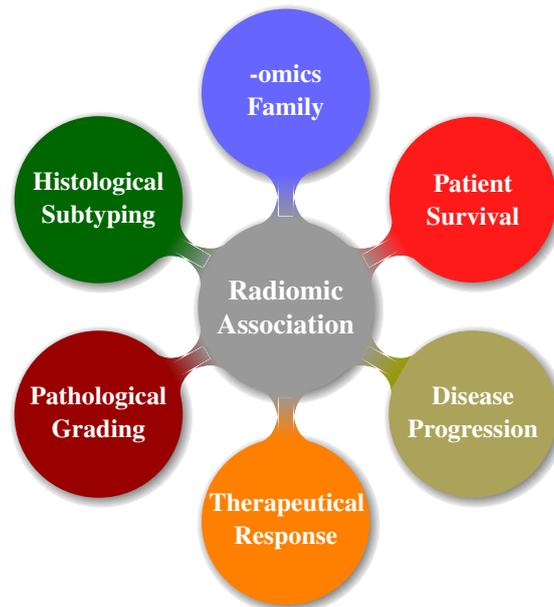


Figura 1.2. Possibilities of clinical endpoints to be investigated through radiomics and medical imaging quantification. Source: Author.

However, literature has shown radiomics could be expanded to analyze other diseases from a wide range of medical specialties. Such as cardiology to predict patient survival, myocardial tissue alterations, and right ventricular failure; and neurology for the characterization of attention deficit hyperactivity disorder [Dawes et al. 2017, Baeßler et al. 2018, Sun et al. 2017].

More recently, research groups have started using deep-learning networks to potentially improve the associative performance [Kermary et al. 2018, Liang and Zheng 2019, Lu et al. 2019]. Those networks are mainly based on deep convolutional neural networks (CNNs) as they are capable of processing high dimensional arrays, like medical images [LeCun et al. 2015, Santos et al. 2019]. However, as those methods characterize the images implicitly (creating the so-called black-box), they will not be further described in this book chapter.

In this manuscript, some theory and practice on robust quantitative biomarker discovery will be introduced to highlight the valuable role that radiomics has in precision

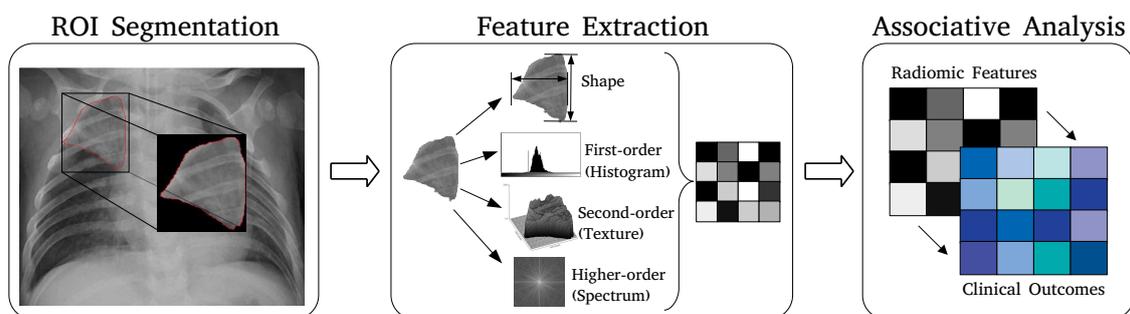


Figura 1.3. Standard investigative pipeline for radiomics. Source: Author.
ROI, region of interest.

medicine. To do that, the state-of-the-art of quantitative precision biomarkers in three different pathological imaging domains will be approached: (I) covid-19 in chest radiography (XR) images, (II) lung neoplasms in computed tomography (CT) scans, and (III) spondyloarthritis in magnetic resonance imaging (MRI).

1.2. Background

1.2.1. Digital radiology images

A standard medical image from the radiology workflow is digitally an n -dimensional discretized matrix where each index identifies a point of interest in the human body. Radiography, for instance, is a two-dimensional matrix with indexes of rows and columns, and the value of each image point (*i.e.*, pixel) corresponds to a gray intensity (Figure 1.4). For example, the grayscale varies 256 levels considering an 8-bit image, from 0 indicating black, 1 to 254 indicating different shades of gray (from dark to light tones), until 255 indicating complete white.

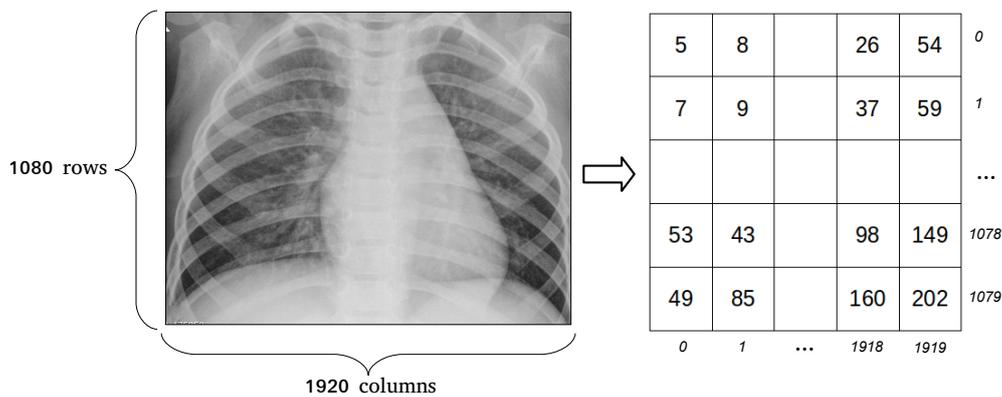


Figura 1.4. Digital XR image representation. Source: Author.

Radiological images as acquired have uniform spacing between pixels and are commonly obtained in the form of a volume of parallel scans, as in the example of CT and MRI. The resulted acquired image is projected in an anatomical plane: coronal, axial, or sagittal (Figure 1.5). The physical distance between the images from a volume is called slice thickness, and each point of the image volume becomes a voxel.

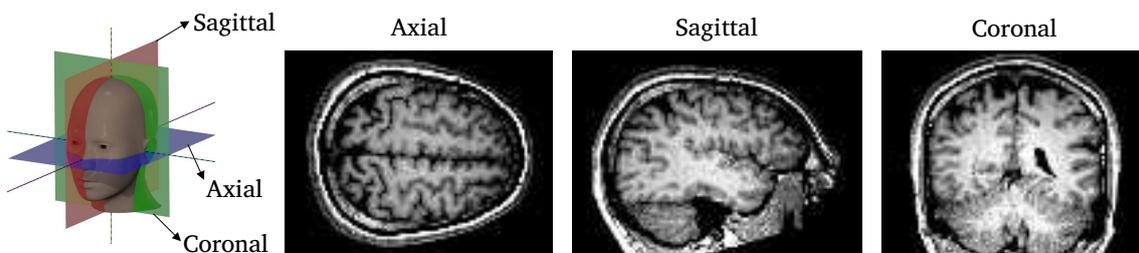


Figura 1.5. Anatomical planes for medical image projection after acquisition. Source: Author licensed under CC BY-SA 4.0. Adapted from Brainscandude / CC BY-SA 3.0 and Slashme / CC BY-SA 4.0.

1.2.2. Image segmentation

Segmentation is a key process for medical image processing, leading to major complications in pattern recognition when presenting low efficiency. Image segmentation aims to identify areas correlated with each other in order to separate a volume of interest (VOI) [Ferreira Junior et al. 2021b].

Methods of medical image segmentation generally use basic properties of gray level discontinuity or similarity [Gonzalez and Woods 2007]. The former approach separates regions according to rough changes on gray intensities, like in the boundaries. The latter strategy is based on comparing gray levels, such as thresholding, the most fundamental technique for segmentation [Ferreira Junior 2019].

Image thresholding is defined as an operation in which an input image is transformed to an output image, as follows:

$$g(i, j) = \begin{cases} f(i, j) & \text{if } T_1 \leq f(i, j) \leq T_2, \\ 0 & \text{if } f(i, j) < T_1 \text{ or } f(i, j) > T_2, \end{cases} \quad (1)$$

where f is the input image, g is the output image, i, j are points of the image matrix, and T_n are threshold values. The threshold is the gray level determined as reference for the comparison between the voxels' intensities (Figure 1.6). In the case of Equation 1.1, the values T_1 and T_2 correspond to a threshold interval, where voxels with intensity out of that gray level range correspond to the image background, and hence, are excluded; while voxels with intensity within the range correspond to the VOI, thus they remained in the segmented output image.

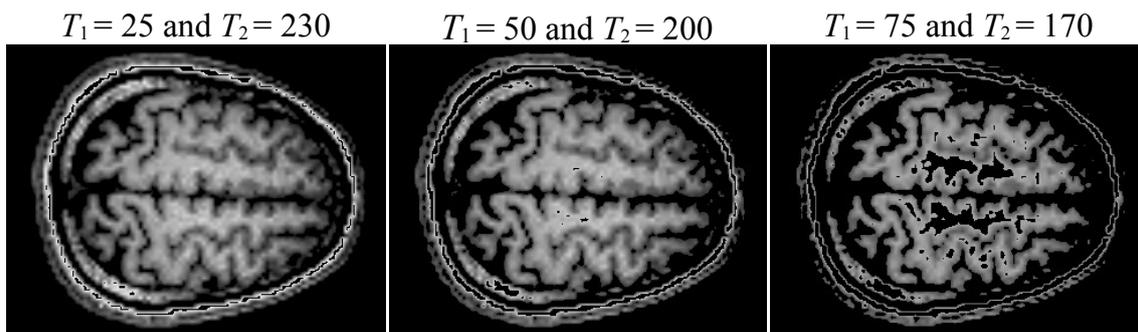


Figura 1.6. Examples of output generated after applying different gray level thresholds for brain image segmentation. Source: Author.

Thresholding is appropriate for very distinct gray intensities between the VOI and background. However, it fails when there is a low contrast in the image, resulting in similar gray levels within the VOI [Ferreira Junior et al. 2020a]. In these particular cases, it is necessary to address this problem using region-oriented approaches, such as region growing methods and convolutional neural networks [Ferreira Junior et al. 2021d, Ferreira Junior et al. 2021c]. But those techniques will not be further described in this book chapter due to space constraints.

1.2.3. Radiomic feature extraction

The process to extract radiomic features is basically the calculation of quantitative measures in a segmented image to represent the visual content. Feature extraction algorithms perform mathematical procedures to characterize the VOIs differently, such as histograms, matrices, transforms, geometrical elements, among others [Tomaszewski and Gillies 2021]. Traditionally, these hand-engineered features describe four primary imaging levels: (I) first-order, characterizing the distribution of intensities in a gray-level histogram; (II) second-order, describing voxel spatial relationships in a gray-level matrix; (III) higher-order, depicting the image spectrum in the frequency domain; and (IV) shape, characterizing geometric and size-related components of a VOI [Ferreira Junior et al. 2020b].

Features from the first imaging order

First-order features are calculated from a gray-level histogram representing a particular VOI (Figure 1.7). These intensity features describe the distribution of voxels' values from the VOI individually without concern for spatial relationships [Santos et al. 2019].

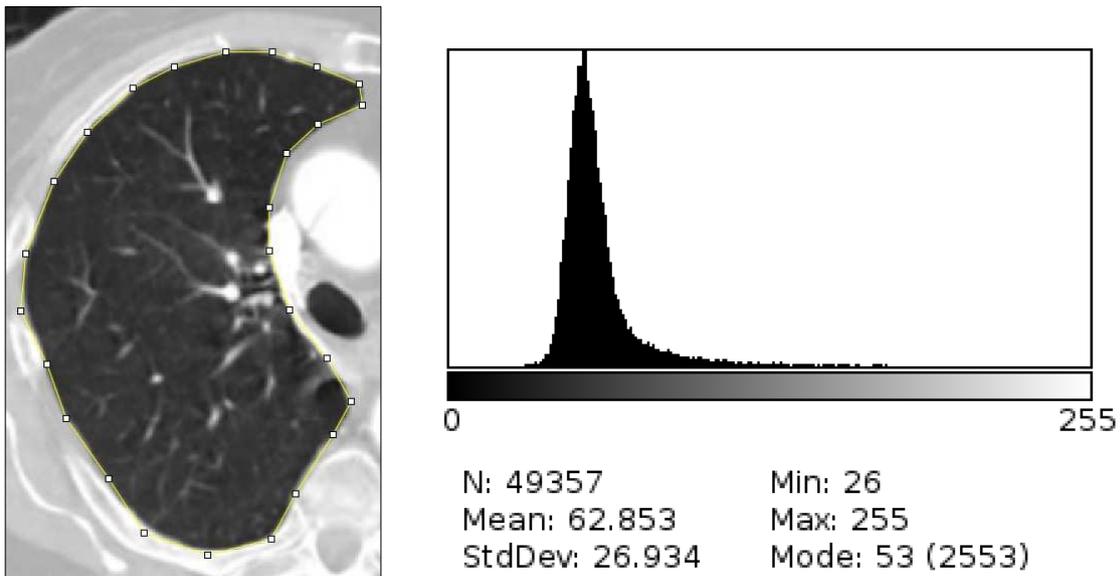


Figura 1.7. Example of an intensity histogram from a cropped CT image of the lung with 256 gray levels. Source: Author.

The first-order features are regular histogram-computed statistical measures, like *mean*, *variance*, *standard deviation*, *coefficient of variation*, *energy*, *entropy*, *mean absolute deviation*, *root mean squared*, *skewness*, and *kurtosis*. Equations 1.2-1.11 list some first-order measures:

$$\text{mean } (\mu) = \frac{1}{n} \sum_{i=1}^n x_i, \quad (2)$$

$$\text{variance } (v) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2, \quad (3)$$

$$\text{standard deviation } (\sigma) = \sqrt{v}, \quad (4)$$

$$\text{coefficient of variation} = \frac{\sigma}{\mu}, \quad (5)$$

$$\text{energy (or uniformity)} = \sum_{i=1}^n x_i^2, \quad (6)$$

$$\text{entropy} = - \sum_{i=1}^n x_i \log_2 x_i, \quad (7)$$

$$\text{mean absolute deviation} = \frac{1}{n} \sum_{i=1}^n |x_i - \mu|, \quad (8)$$

$$\text{root mean squared} = \sqrt{\frac{\sum_{i=1}^n x_i^2}{n}}, \quad (9)$$

$$\text{skewness} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^3}{\sigma^3}, \quad (10)$$

$$\text{kurtosis} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^4}{\sigma^4}, \quad (11)$$

where x is a histogram of n gray levels.

Those statistical measures are important because they can quantify the image density, like the *mean*. The *standard deviation*, *variance*, *coefficient of variation*, and *mean absolute deviation* are dispersion measures and they describe how much the gray levels differ from the mean intensity. While the *energy* and the *root mean squared* are measures of magnitude, the *entropy* can characterize the randomness and variations present in the image, measuring the average amount of information required to encode the image values. *Skewness* and *kurtosis* are the histogram central moments and they mainly quantify the asymmetry and sharpness degrees, respectively, around the mean [Zwanenburg et al. 2020, Aerts et al. 2014].

Features from the second imaging order

Histogram-based features normally are not enough to fully describe the VOI. As previously stated, those first-order measures do not consider spatial relationships, which are crucial for distinguishing similar textural images, such as the ones in Figure 1.1.

On the other hand, second-order features can describe intrinsic characteristics of the image texture using advanced statistical mechanisms. Although there is no consensus about the formal definition of image texture, it can be defined as the repetition of patterns of even minor variations in a VOI [Ferreira Junior 2019]. In medical imaging, that is essential due to the tiniest details at a molecular level that the image may portray.

Several approaches can analyze the images and recognize different texture patterns. Probably the most known in literature is the so-called gray-level co-occurrence matrix (GLCM) [Haralick et al. 1973]. The co-occurrence matrices obtain from the VOI the occurrence probability of pairs of voxel intensities i, j given a distance d and an orientation θ for the x, y dimensions and an orientation ϕ for the z dimension. The GLCM computes in two-dimensional arrays the occurrence of individual intensity pairs in slice by slice manner. The final matrix is the summation of appearances in all image slices. The GLCM tridimensional version directly computes the occurrence of individual intensity pairs in the entire VOI [Ferreira et al. 2017].

The texture features provided by the GLCM are computed by calculating the following measures (Equations 1.12-1.34) on the produced matrix:

$$\text{autocorrelation} = \sum_{i,j=1}^n ijM_{ij}, \quad (12)$$

$$\text{cluster prominence} = \sum_{i,j=1}^n (i+j-\mu_x-\mu_y)^4 M_{ij}, \quad (13)$$

$$\text{cluster shade} = \sum_{i,j=1}^n (i+j-\mu_x-\mu_y)^3 M_{ij}, \quad (14)$$

$$\text{contrast} = \sum_{i,j=1}^n (i-j)^2 M_{ij}, \quad (15)$$

$$\text{correlation} = \frac{1}{\sigma_x \sigma_y} \sum_{i,j=1}^n (i-\mu_x)(j-\mu_y) M_{ij}, \quad (16)$$

$$\text{difference average (or dissimilarity)} = \sum_{k=0}^{n-1} kM_{i-j,k}, \quad (17)$$

$$\text{difference entropy} = - \sum_{k=0}^{n-1} M_{i-j,k} \log_2 M_{i-j,k}, \quad (18)$$

$$\text{difference variance} = \sum_{k=0}^{n-1} (k - \text{dissimilarity})^2 M_{i-j,k}, \quad (19)$$

$$\text{energy (or uniformity or angular second moment)} = \sum_{i,j=1}^n M_{ij}^2, \quad (20)$$

$$\text{IMC}_1 = \frac{- \sum_{i,j=1}^n M_{ij} \log_2 M_{ij} + \sum_{i,j=1}^n M_{ij} \log_2 (M_i M_j)}{- \sum_{i=1}^n M_i \log_2 M_i}, \quad (21)$$

$$\text{IMC}_2 = \sqrt{1 - \exp(-2(- \sum_{i,j=1}^n M_i M_j \log_2 M_i M_j + \sum_{i,j=1}^n M_{ij} \log_2 M_{ij}))}, \quad (22)$$

$$\text{inverse difference} = \sum_{i,j=1}^n \frac{M_{ij}}{1 + |i - j|}, \quad (23)$$

$$\text{inverse difference normalized} = \sum_{i,j=1}^n \frac{M_{ij}}{1 + \frac{|i-j|}{n}}, \quad (24)$$

$$\text{inverse difference moment} = \sum_{i,j=1}^n \frac{M_{ij}}{1 + (i - j)^2}, \quad (25)$$

$$\text{inverse difference moment normalized} = \sum_{i,j=1}^n \frac{M_{ij}}{1 + \frac{(i-j)^2}{n^2}}, \quad (26)$$

$$\text{inverse variance} = 2 \sum_{i,j=1}^n \frac{M_{ij}}{(i - j)^2}, i \neq j, \quad (27)$$

$$\text{joint average} = \sum_{i,j=1}^n iM_{ij}, \quad (28)$$

$$\text{(joint or sum) entropy} = - \sum_{i,j=1}^n M_{ij} \log_2 M_{ij}, \quad (29)$$

$$\text{joint maximum (or max probability)} = \max(M_{ij}), \quad (30)$$

$$\text{(joint) variance (or sum of squares)} = \sum_{i,j=1}^n (i - \text{joint average})^2 M_{ij}, \quad (31)$$

$$\text{sum average} = 2 \times \text{joint average}, \quad (32)$$

$$\text{sum entropy} = - \sum_{k=2}^{2n} M_{i+j,k} \log_2 M_{i+j,k}, \quad (33)$$

$$\text{sum variance (or cluster tendency)} = \sum_{k=2}^{2n} (k - \text{sum average})^2 M_{i+j,k}, \quad (34)$$

where i, j is the gray-level pair, M_{ij} is an element of the GLCM (here computed from a two-dimensional array), n is the number of different gray levels, μ_x, μ_y are mean values in the x, y directions, σ_x, σ_y are standard deviation values in the x, y directions, and IMC is the *informational measure of correlation*.

Inverse difference and *inverse difference moment* were also referred simply as *homogeneity*, but this nomenclature is deprecated. Those two features and their normalized versions measure the local homogeneity of the image. *Energy* analogously is a measure of uniform imaging patterns, in which a greater value implies that there are more intensity pairs in the image that neighbor each other at higher frequencies. On the other hand, features of *entropy* are measures of the randomness/variability in neighborhood intensity values. The *difference variance* measures heterogeneous patterns that place higher weights on differing intensity pairs that deviate more from the mean. *Contrast* is also a measure of the local intensity variation, where a larger value correlates with a greater disparity in intensity values among neighboring voxels. The IMC metrics quantify the texture complexity evaluating the correlation between the probability distributions through mutual information. *Autocorrelation* is a measure of the magnitude of the texture

fineness and coarseness. The features of *cluster prominence*, *shade*, and *tendency* emulate the human perception, and they measure the skewness of the GLCM, where higher values imply greater asymmetry about the mean. The *correlation* shows the linear dependency of gray levels to their respective voxels in a region. The *difference average* and *sum average* measure the relationship between occurrences of pairs. But the former relates similar and differing intensities, and the latter relates lower with higher intensity values [Zwanenburg et al. 2020, Van Griethuysen et al. 2017, Yip et al. 2017, Phillips et al. 2017].

A second alternative for texture analysis is based on the gray-level run-length matrix (GLRLM) [Tang 1998, Galloway 1975]. The run-length matrix tracks the frequencies of sequences with different lengths of the same gray level at a predetermined orientation. The GLRLM is interesting for image characterization in that fine textures have more short sequences with similar gray levels, and rough textures have more long sequences with different intensities [Davnall et al. 2012]. Analogously to the GLCM, the following attributes (Equations 1.35-1.50) can be calculated to form the GLRLM texture features:

$$\text{short run emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i,j|\theta)}{j^2}}{\sum_i^n \sum_j^l p(i,j|\theta)}, \quad (35)$$

$$\text{long run emphasis} = \frac{\sum_i^n \sum_j^l j^2 p(i,j|\theta)}{\sum_i^n \sum_j^l p(i,j|\theta)}, \quad (36)$$

$$\text{gray level non-uniformity} = \frac{\sum_i^n (\sum_j^l p(i,j|\theta))^2}{\sum_i^n \sum_j^l p(i,j|\theta)}, \quad (37)$$

$$\text{gray level non-uniformity normalized} = \frac{\sum_i^n (\sum_j^l p(i,j|\theta))^2}{(\sum_i^n \sum_j^l p(i,j|\theta))^2}, \quad (38)$$

$$\text{run length non-uniformity} = \frac{\sum_j^l (\sum_i^n p(i,j|\theta))^2}{\sum_i^n \sum_j^l p(i,j|\theta)}, \quad (39)$$

$$\text{run length non-uniformity normalized} = \frac{\sum_j^l (\sum_i^n p(i,j|\theta))^2}{(\sum_i^n \sum_j^l p(i,j|\theta))^2}, \quad (40)$$

$$\text{run percentage} = \frac{\sum_i^n \sum_j^l p(i, j|\theta)}{v}, \quad (41)$$

$$\text{gray level variance} = \sum_i^n \sum_j^l q(i, j|\theta) (i - \sum_i^n \sum_j^l q(i, j|\theta) i)^2, \quad (42)$$

$$\text{run variance} = \sum_i^n \sum_j^l q(i, j|\theta) (j - \sum_i^n \sum_j^l q(i, j|\theta) j)^2, \quad (43)$$

$$\text{run entropy} = - \sum_i^n \sum_j^l q(i, j|\theta) \log_2(q(i, j|\theta)), \quad (44)$$

$$\text{low gray level run emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i, j|\theta)}{i^2}}{\sum_i^n \sum_j^l p(i, j|\theta)}, \quad (45)$$

$$\text{high gray level run emphasis} = \frac{\sum_i^n \sum_j^l i^2 p(i, j|\theta)}{\sum_i^n \sum_j^l p(i, j|\theta)}, \quad (46)$$

$$\text{short run low gray level emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i, j|\theta)}{i^2 j^2}}{\sum_i^n \sum_j^l p(i, j|\theta)}, \quad (47)$$

$$\text{short run high gray level emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i, j|\theta) i^2}{j^2}}{\sum_i^n \sum_j^l p(i, j|\theta)}, \quad (48)$$

$$\text{long run low gray level emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i, j|\theta) j^2}{i^2}}{\sum_i^n \sum_j^l p(i, j|\theta)}, \quad (49)$$

$$\text{long run high gray level emphasis} = \frac{\sum_i^n \sum_j^l p(i, j|\theta) i^2 j^2}{\sum_i^n \sum_j^l p(i, j|\theta)}, \quad (50)$$

where $p(i, j|\theta)$ is an element of the GLRLM, $q(i, j|\theta)$ is an element of the normalized GLRLM, i is a gray intensity, j is the frequency of i , θ is the orientation, and n, l, v are the number of gray levels, run lengths, and voxels, respectively, in the image.

Short and *long run emphasis* measure run distributions in the image. The *short run emphasis* is a measure in which a greater value is indicative of shorter run lengths and more fine textural patterns. The *long run emphasis* is a measure in which a greater value is indicative of extended run lengths and more coarse structural textures. *Gray level non-uniformity* and its normalized version measure the similarity of gray-level intensities in the image, where a lower value correlates with a greater similarity in intensities. The *run length non-uniformity* and its normalized version measure the similarity of run lengths throughout the image, with a lower value indicating more homogeneity among run lengths in the image. *Run percentage* measures the coarseness of the texture, and its higher values indicate a larger portion of the ROI consists of short runs and a more fine texture. The features of *gray level* and *run variances* measure the variance in gray level intensity for the runs and the variance in runs for the run lengths, respectively. *Run entropy* is a measure of uncertainty/randomness of run lengths and gray levels, in which a higher value indicates more heterogeneity in the texture patterns. The features of *low* and *high gray level run emphasis* measure the distribution of gray levels with a higher value indicating a greater concentration of low and high gray-level values, respectively, in the image. Finally, the features of *short/long run low/high gray level emphasis* measure the joint distribution of run lengths with the gray-level values [Van Griethuysen et al. 2017, Phillips et al. 2017, Davnall et al. 2012].

A more recent approach for texture characterization, in comparison to the ones in [Haralick et al. 1973, Galloway 1975, Tamura et al. 1978], is the gray level size zone matrix (GLSZM) [Thibault et al. 2013]. A gray level zone is defined as the number of connected voxels (with distance 1 with each other) that share the same gray level intensity. Contrary to GLCM and GLRLM, the GLSZM is rotation independent, with only one matrix calculated for all directions in the ROI [Van Griethuysen et al. 2017]. The following attributes (Equations 1.51-1.66) can be calculated in the GLSZM to form the features:

$$\text{small area emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i,j)}{j^2}}{\sum_i^n \sum_j^l p(i,j)}, \quad (51)$$

$$\text{large area emphasis} = \frac{\sum_i^n \sum_j^l j^2 p(i,j)}{\sum_i^n \sum_j^l p(i,j)}, \quad (52)$$

$$\text{gray level non-uniformity} = \frac{\sum_i^n (\sum_j^l p(i,j))^2}{\sum_i^n \sum_j^l p(i,j)}, \quad (53)$$

$$\text{gray level non-uniformity normalized} = \frac{\sum_i^n (\sum_j^l p(i,j))^2}{(\sum_i^n \sum_j^l p(i,j))^2}, \quad (54)$$

$$\text{size-zone non-uniformity} = \frac{\sum_j^l (\sum_i^n p(i, j))^2}{\sum_i^n \sum_j^l p(i, j)}, \quad (55)$$

$$\text{size-zone non-uniformity normalized} = \frac{\sum_j^l (\sum_i^n p(i, j))^2}{(\sum_i^n \sum_j^l p(i, j))^2}, \quad (56)$$

$$\text{zone percentage} = \frac{\sum_i^n \sum_j^l p(i, j)}{v}, \quad (57)$$

$$\text{gray level variance} = \sum_i^n \sum_j^l q(i, j) (i - \sum_i^n \sum_j^l q(i, j) i)^2, \quad (58)$$

$$\text{zone variance} = \sum_i^n \sum_j^l q(i, j) (j - \sum_i^n \sum_j^l q(i, j) j)^2, \quad (59)$$

$$\text{zone entropy} = - \sum_i^n \sum_j^l q(i, j) \log_2(q(i, j)), \quad (60)$$

$$\text{low gray level zone emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i, j)}{i^2}}{\sum_i^n \sum_j^l p(i, j)}, \quad (61)$$

$$\text{high gray level zone emphasis} = \frac{\sum_i^n \sum_j^l i^2 p(i, j)}{\sum_i^n \sum_j^l p(i, j)}, \quad (62)$$

$$\text{small area low gray level emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i, j)}{i^2 j^2}}{\sum_i^n \sum_j^l p(i, j)}, \quad (63)$$

$$\text{small area high gray level emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i, j) i^2}{j^2}}{\sum_i^n \sum_j^l p(i, j)}, \quad (64)$$

$$\text{large area low gray level emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i, j) j^2}{i^2}}{\sum_i^n \sum_j^l p(i, j)}, \quad (65)$$

$$\text{large area high gray level emphasis} = \frac{\sum_i^n \sum_j^l p(i, j) i^2 j^2}{\sum_i^n \sum_j^l p(i, j)}, \quad (66)$$

where $p(i, j)$ is an element of the GLSZM, $q(i, j)$ is an element of the normalized GLSZM, i is a gray intensity, j is the zone size of i , and n, l, v are the number of gray levels, zone sizes, and voxels, respectively, in the image.

Analogously to the GLRLM, the *small area emphasis* measures the distribution of small size zones, with a greater value indicative of more fine textures. *Large area emphasis* is a measure of the distribution of large size zones, with a greater value indicative of more coarse textures. All GLSZM *non-uniformity, percentage, variance, and entropy* features have equivalent definitions to the GLRLM measures but considering gray-level zones, not runs. The features of *low and high gray level zone emphasis* measure the distribution of gray levels with a higher value indicating a greater proportion of low and high gray-level values, respectively, and size zones in the image. Moreover, the features of *small/large area low/high gray level emphasis* measure the joint distribution of smaller and larger size zones, respectively, with the gray levels [Van Griethuysen et al. 2017].

Another approach vastly used for texture analysis and image characterization is the neighborhood intensity difference matrix (NIDM) [Amadasun and King 1989]. This matrix is actually a one-column array that tracks the average difference between voxel intensities and their neighbors according to a distance. The main advantage of the NIDM is that it examines spatial relationships between three or more voxels at once, not just pairs like the GLCM [Lubner et al. 2017, Yang et al. 2016]. The NIDM attributes are listed in Equations 1.67-1.71:

$$\text{coarseness} = \left[\sum_i^n P(i) S(i) \right]^{-1}, \quad (67)$$

$$\text{contrast} = \left[\frac{1}{N_a(N_a - 1)} \sum_{i_1, i_2}^n P(i_1) P(i_2) (i_1 - i_2)^2 \right] \left[\frac{1}{N} \sum_i^n S(i) \right], P(i_1) \neq 0, P(i_2) \neq 0 \quad (68)$$

$$\text{busyness} = \frac{\sum_i^n P(i) S(i)}{\sum_{i_1, i_2}^n i_1 P(i_1) - i_2 P(i_2)}, P(i_1) \neq 0, P(i_2) \neq 0, \quad (69)$$

$$\text{complexity} = \frac{1}{N_v} \sum_{i_1, i_2}^n |i_1 - i_2| \frac{P(i_1) S(i_1) + P(i_2) S(i_2)}{P(i_1) + P(i_2)}, P(i_1) \neq 0, P(i_2) \neq 0, \quad (70)$$

$$\text{strength} = \frac{\sum_{i_1, i_2}^n (P(i_1) + P(i_2)) (i_1 - i_2)^2}{\sum_i^n S(i)}, P(i_1) \neq 0, P(i_2) \neq 0, \quad (71)$$

where $P(i)$ is the occurrence probability of the gray level i , $S(i)$ is the NIDM element, n is the number of discretized gray levels of the image, N, N_a, N_v are the number of voxels, discretized gray levels with $a > 0$, and voxels with at least 1 neighbor.

The *coarseness* feature is a measure of the border density, and it averages the difference between the center voxel and its neighborhood, where a higher value indicates a lower spatial change rate and a locally more uniform texture. *Contrast* is a measure of local spatial intensity change, yielding a high value when both the dynamic range and the spatial change rate are increased. The *busyness* measures the ratio of spatial intensity change in a region, where higher values indicate rapid changes of intensity between voxels and the neighborhood. *Complexity* and *strength* are measures of primitive texture components in the region, in which high values indicate the primitives are easily defined, visible, and the image is non-uniform [Van Griethuysen et al. 2017, Phillips et al. 2017, Davnall et al. 2012].

One final matrix-based approach worth mention is the gray level dependence matrix (GLDM) [Sun and Wee 1983]. It quantifies intensity dependencies in an image and describe the overall texture coarseness, in which the number of connected voxels within a predetermined distance are dependent on the center voxel [Zwanenburg et al. 2020, Van Griethuysen et al. 2017]. Analogously to the GLSZM, the GLDM is rotation invariant and has the following measures (Equations 1.72-1.85) calculated to form the texture features:

$$\text{small (or low) dependence (or number) emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i,j)}{j^2}}{\sum_i^n \sum_j^l p(i,j)}, \quad (72)$$

$$\text{large (or high) dependence (or number) emphasis} = \frac{\sum_i^n \sum_j^l j^2 p(i,j)}{\sum_i^n \sum_j^l p(i,j)}, \quad (73)$$

$$\text{gray level non-uniformity} = \frac{\sum_i^n (\sum_j^l p(i,j))^2}{\sum_i^n \sum_j^l p(i,j)}, \quad (74)$$

$$\text{dependence non-uniformity} = \frac{\sum_j^l (\sum_i^n p(i,j))^2}{\sum_i^n \sum_j^l p(i,j)}, \quad (75)$$

$$\text{dependence non-uniformity normalized} = \frac{\sum_j^l (\sum_i^n p(i,j))^2}{(\sum_i^n \sum_j^l p(i,j))^2}, \quad (76)$$

$$\text{gray level variance} = \sum_i^n \sum_j^l q(i,j) (i - \sum_i^n \sum_j^l q(i,j) i)^2, \quad (77)$$

$$\text{dependence variance} = \sum_i^n \sum_j^l q(i, j) (j - \sum_i^n \sum_j^l q(i, j) j)^2, \quad (78)$$

$$\text{dependence entropy} = - \sum_i^n \sum_j^l q(i, j) \log_2(q(i, j)), \quad (79)$$

$$\text{low gray level (count) emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i, j)}{i^2}}{\sum_i^n \sum_j^l p(i, j)}, \quad (80)$$

$$\text{high gray level (count) emphasis} = \frac{\sum_i^n \sum_j^l i^2 p(i, j)}{\sum_i^n \sum_j^l p(i, j)}, \quad (81)$$

$$\text{small (or low) dependence low gray level emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i, j)}{i^2 j^2}}{\sum_i^n \sum_j^l p(i, j)}, \quad (82)$$

$$\text{small (or low) dependence high gray level emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i, j) i^2}{j^2}}{\sum_i^n \sum_j^l p(i, j)}, \quad (83)$$

$$\text{large (or high) dependence low gray level emphasis} = \frac{\sum_i^n \sum_j^l \frac{p(i, j) j^2}{i^2}}{\sum_i^n \sum_j^l p(i, j)}, \quad (84)$$

$$\text{large (or high) dependence high gray level emphasis} = \frac{\sum_i^n \sum_j^l p(i, j) i^2 j^2}{\sum_i^n \sum_j^l p(i, j)}, \quad (85)$$

where $p(i, j)$ is an element of the GLDM, $q(i, j)$ is an element of the normalized GLDM, i is the center voxel, j is a neighbouring voxel of i , and n, l are the number of gray levels and dependency sizes, respectively, in the image.

The *small dependence emphasis* is a measure of the distribution of small dependencies with higher values indicative of less homogeneous textures. Opposite to that, the *large dependence emphasis* is a measure of the distribution of large dependencies with a higher value indicative of more homogeneous textures. The GLDM features of *non-uniformity*, *variance*, and *entropy* have equivalent definitions to the GLRLM and GLSZM measures but considering the gray-level dependency, not runs or zones. *Low gray level emphasis* measures the distribution of low intensities with a higher value indicating a greater concentration of low gray-level values in the image. *High gray level*

emphasis measures the distribution of the high intensities with higher values indicating a greater concentration of high gray-level values. *Small/large dependence low/high gray level emphasis* features measure the joint distribution of intensity dependences with lower and higher gray-level values [Van Griethuysen et al. 2017].

Another textural approach was proposed in [Tamura et al. 1978]. Those attributes theoretically correspond to human visual perception features, providing better description of the texture [Faleiros et al. 2020]. Tamura's features are *line-likeness*, *regularity*, *roughness*, *contrast*, *granularity*, and *directionality*, but the last three are known to better describe the image texture. However, *granularity* is the most fundamental feature where higher values indicate greater or less repeated textures. *Directionality* is a global property of the image that considers the textural shape and location, without taking into account the orientation [Tenorio et al. 2020].

A complementary strategy for texture characterization and an alternative to the statistics-based ones previously described uses structural geometry and fractal analysis. This approach implies several regions have a standard statistical pattern of roughness and irregularity in different scales [Faleiros et al. 2020, Ferreira Junior 2019]. The fractal measures represent various aspects of the image and provide essential information about spatial heterogeneity [Kolossvary et al. 2018]. Fractal analyses can result in, for instance, the dimension estimate that quantifies how an object fills spaces, the abundance that measures the volume of space filled, and lacunarity that quantifies structural heterogeneity within an object [Davnall et al. 2012].

Features from the higher imaging order

When first and second-order features are inefficient in image characterization, it is necessary to expand the characteristic spectrum by incorporating descriptors beyond the spatial level. Signal processing methods emerged in this context due to their ability to analyze frequency domain properties of the image [Tomaszewski and Gillies 2021, Ferreira Junior and Cardona Cardenas 2021]. The basics to include higher-order characteristics to the feature multi-dimensional space are applying a transform, filter, or wavelet and then calculating standard mathematical measures on the resulting filtered image.

The most traditional approach to obtain the frequency power spectrum uses the Fourier transform, and polar coordinates of each pixel from the transformed image [Schneider et al. 2012]. The Fourier transform analyzes the frequency content disregarding temporal and spatial locations by converting the image in the spatial domain into a set of sine and cosine components [Davnall et al. 2012]. The identification of frequency peaks, prominence, and location reveals information about the periodicity and directions of the image texture [Ferreira Junior 2019].

However, computing the Fourier transform is a time-demanding process, and thus, several implementations have been proposed in the literature over the decades to reduce the computational cost. The most used strategy is the so-called fast Fourier transform (FFT) [Faleiros et al. 2020, Tenorio et al. 2020]. After applying the FFT in an image, first-order measures could be calculated to compose the Fourier-based features (Figure 1.8). The discrete Fourier transform can be formulated as given by the Equation 1.86 [Parker

2011]:

$$F(w) = \sum_{k=0}^{N-1} f(k) e^{\frac{2\pi jwk}{N}} \quad (86)$$

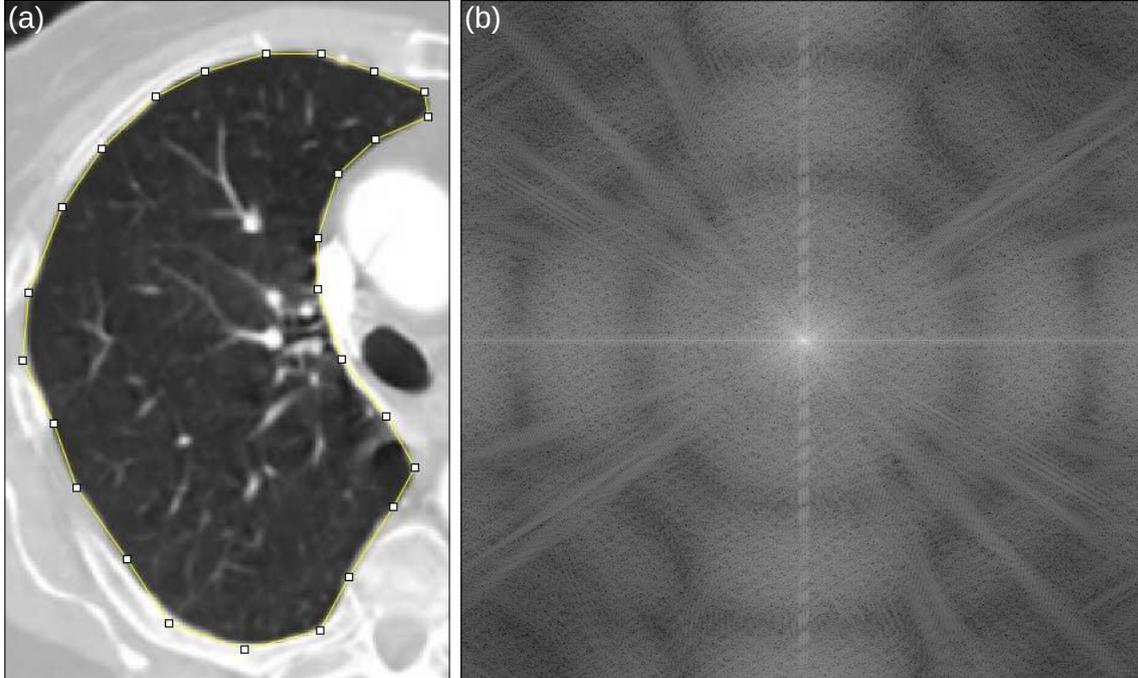


Figura 1.8. Application of the fast Fourier transform in a cropped CT image of the lung: (a) and (b) show the original and resulted images, respectively. Source: Author.

An expanded spectral method for higher-order feature characterization is based on the Gabor transform or filter, which describe textural patterns by sinusoidal functions and allow the spatial, temporal, and frequency representation of the signal [Kolossvary et al. 2018]. A Gabor filter is essentially a windowed Fourier transform after introducing a Gaussian function, resulting in the acquisition of measures in different time-frequency bands according to a determined scale and orientation [Davnall et al. 2012]. Gabor filter bank is a more powerful approach because it can manipulate local texture parameters of frequency, orientation, excentricity, and symmetry [Ferreira Junior 2019]. The formulation of a Gabor filter in the spatial domain is given by the Equation 1.87 [Bianconi and Fernandez 2007]:

$$\psi(x, y) = \frac{F^2}{\pi\gamma\eta} e^{-F^2[(x'/\gamma)^2 + (y'/\eta)^2]} e^{i2\pi Fx'}, \quad (87)$$

where x' is $x\cos\theta + y\sin\theta$, y' is $-x\sin\theta + y\cos\theta$, F is the central filter frequency, θ is the angle between the sinusoidal wave direction and the axis x in the spatial domain, γ and η are Gaussian standard deviations. Analogously to the FFT-filtered image, first-order measures could be calculated to compose the Gabor-based features (Figure 1.9).

Although the Gabor filter is an interesting approach, it is limited by the spatial resolution with single window, which is opposite to the wavelet transforms. These methods

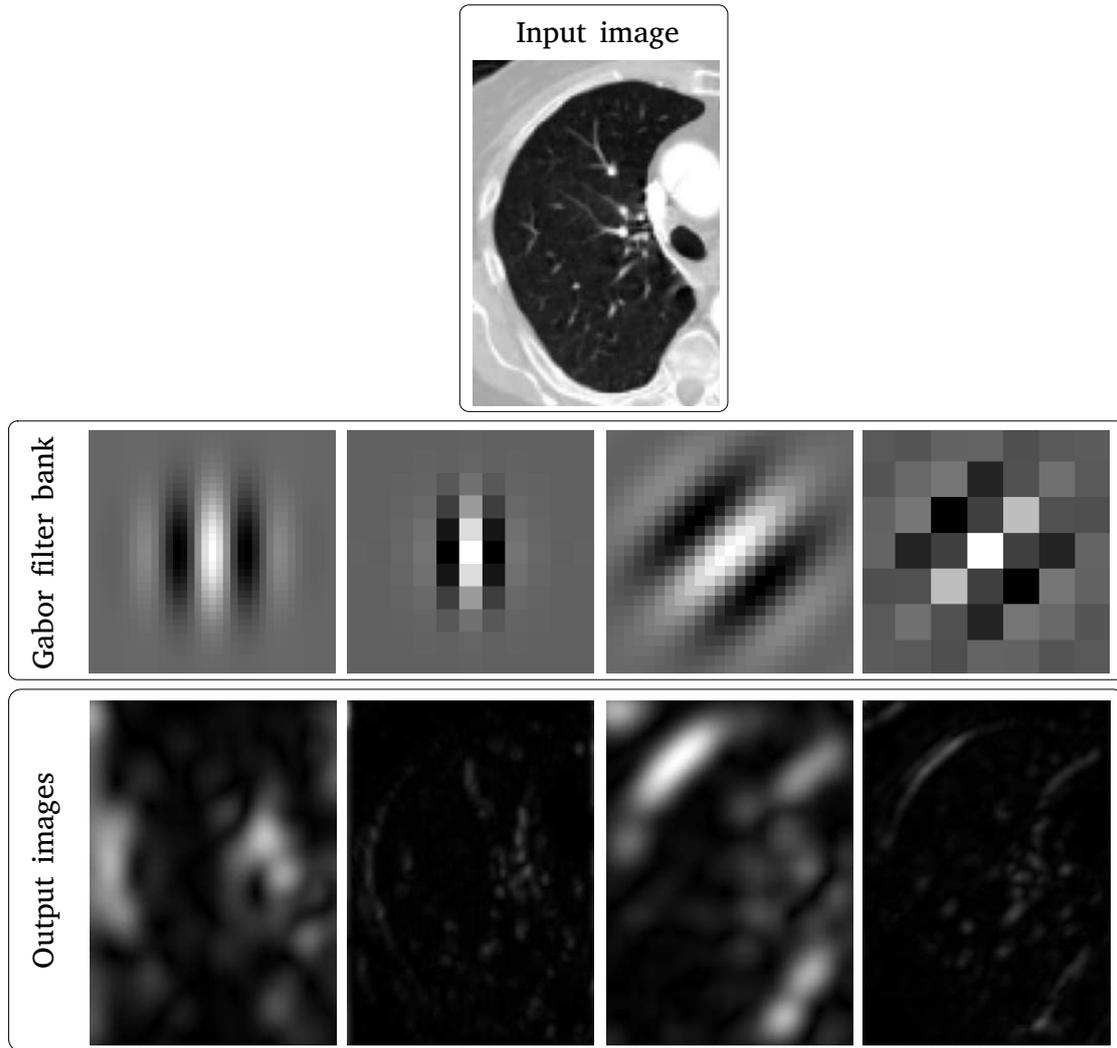


Figure 1.9. Application of a Gabor filter bank in a cropped CT image of the lung.
Source: Author.

use multiple scalable and translatable functions in different frequencies, which could represent the texture more comprehensively than the previous techniques [Kolossvary et al. 2018]. Several wavelets have been proposed in the literature, but they all derived from $\Psi(t)$ in Equation 1.88 [Davnall et al. 2012]:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), \quad a, b \in \mathbb{R}, a \neq 0, \quad (88)$$

where a is a scaling parameter that measures the compression degree and b is the translation parameter that indicates the wavelet time location. Probably the most known wavelet functions are Coiflets, Haar, Daubechies, Symlets, Discrete Meyer, Biorthogonal, and Reverse Biorthogonal (Figure 1.10). They can decompose the image in different frequency domain bands (HH, HL, LH, and LL) and spectrum levels (Figure 1.11) to allow the extensive higher-order feature extraction [Lee et al. 2019].

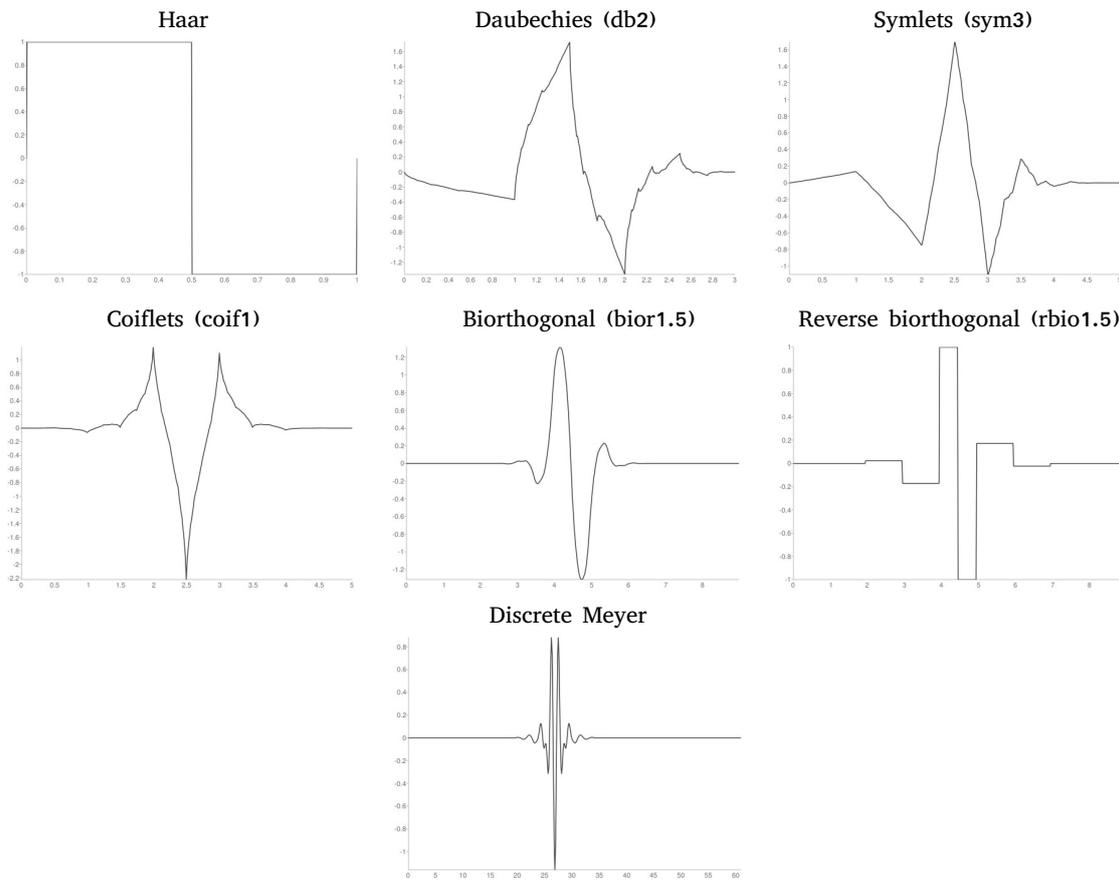


Figura 1.10. Wavelet functions. Source: Adapted from Lee et al. 2019.

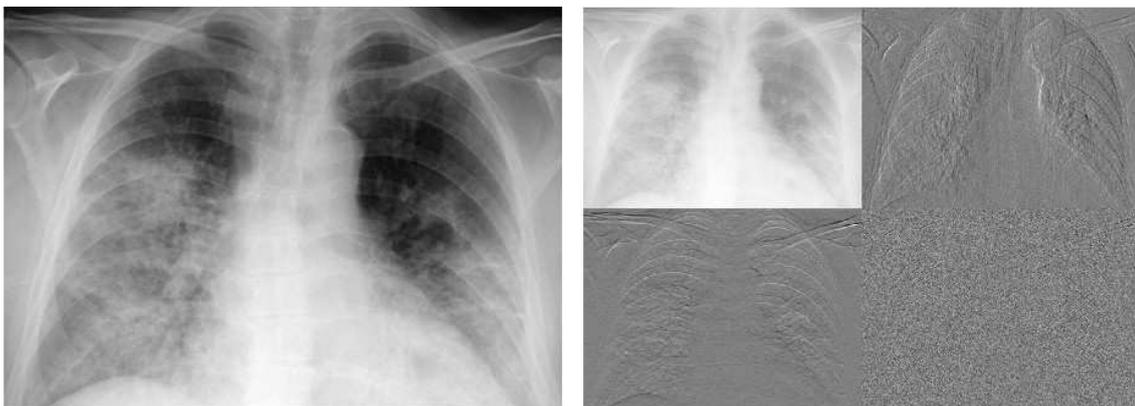


Figura 1.11. Application of a discrete Haar wavelet transform in a chest XR. Source: Author.

Features based on shape

The term *shape* refers to the information inferred from the VOI but could not be represented directly from the intensities, like gray levels and texture. Shape features describe the VOI through geometrical characteristics from the border, contour, curves, among others, which are important in cases where there is a massive difference in the VOI definition of

radiological findings (Figure 1.12) [Echegaray et al. 2015].

Characterizing VOIs quantitatively is a challenging task because it depends directly on the efficiency of image segmentation algorithms [Ferreira Junior et al. 2021b]. Moreover, it is common for image segmentation to have low performances when the VOI has low contrast in the gray levels (Figure 1.13) [Ferreira Junior et al. 2020a], resulting in poor shape-based characterization. Traditionally, shape features are categorized as contour or region-based. The former obtains the features by analyzing the border coordinates from the VOI, and the latter obtains features by considering the region within the VOI. Some of the most used shape features are listed in Equations 1.89-1.95:

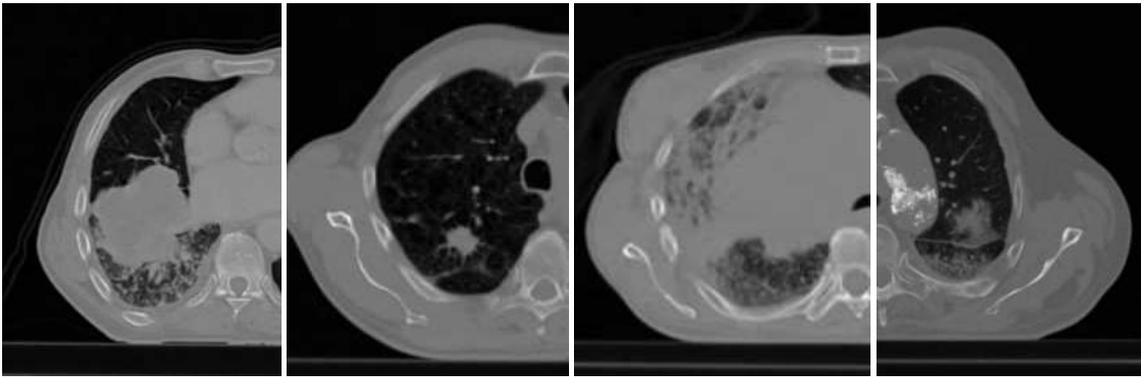


Figura 1.12. Morfológica heterogeneidade de tumores pulmonares apresentados em imagens de TC. Estes são um exemplo em que características de forma poderiam ser úteis em distinguir anomalias radiológicas. Fonte: Autor.

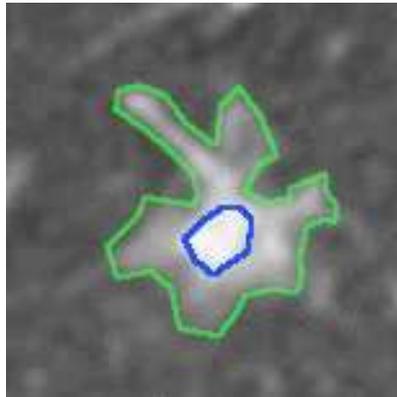


Figura 1.13. Segmentação semi-automática de componentes sólidos (marcação azul) e subsólidos (marcação verde) de um nódulo pulmonar em TC. Este é um exemplo em que a segmentação de imagem poderia impedir a caracterização de forma devido ao baixo contraste dos níveis de cinza. Fonte: Autor.

$$\text{surface area } (A) = \sum_i^N \frac{1}{2} |a_i b_i \times a_i c_i|, \quad (89)$$

$$\text{volume } (V) = \sum_k^v V_k, \quad (90)$$

$$\text{density (or surface area to volume ratio)} = \frac{A}{V}, \quad (91)$$

$$\text{compactness}_1 = \frac{V}{\sqrt{\pi A^{\frac{2}{3}}}}, \quad (92)$$

$$\text{compactness}_2 = 36\pi \frac{V^2}{A^3}, \quad (93)$$

$$\text{spherical disproportion} = \frac{A}{4\pi R^2}, \quad (94)$$

$$\text{sphericity} = \frac{\pi^{\frac{1}{3}}(6V)^{\frac{2}{3}}}{A}, \quad (95)$$

where N is the number of connected triangles covering the surface, v is the number of voxels inside the VOI, V_k is the volume of a single voxel, R is the sphere radius defined as $\sqrt[3]{\frac{3V}{4\pi}}$, and a, b, c are the connected triangle vertices.

Features of *maximum diameter*, *volume*, and *surface area* provide size-related information. Measures of *compactness*, *spherical disproportion*, *sphericity*, and *density* quantify how much the VOI is spherical, compact, and rounded [Zhang et al. 2015].

A special case of shape features consider the specific region in the transition from inside the VOI to the outside. The so-called margin sharpness is essential in cases where the disease grows and invade neighbouring tissues, such as malignant neoplasms [Levman and Martel 2011]. One of the first margin sharpness features was proposed in [Gilhuijs et al. 1998] for the characterization of breast lesions in MR images. The authors based on the spatial gradient of the boundary, as described in Equations 1.96 and 1.97 [Xu et al. 2012]:

$$\text{average of margin gradient} = \max_{i=0, \dots, M-1} \left\{ \frac{\text{mean}_r \|\nabla I_m(r)\|}{\text{mean}_r I_m(r)} \right\}, \quad (96)$$

$$\text{variance of margin gradient} = \frac{\text{var}_r \|\nabla I_m(r)\|}{[\text{mean}_r I_m(r)]^2}, \quad (97)$$

where $I_m(r)$ is a gray-level intensity and the amplitude of vector r in $I_m(\cdot)$ is limited to the region surface.

A second approach developed for margin sharpness quantification also used MR images and breast lesions as application [Levman and Martel 2011]. However, the voxels included in that belonged to both the interior and exterior of the VOI. That method resulted in only one feature defined in Equation 1.98 [Xu et al. 2012]:

$$\text{margin sharpness} = \frac{\overline{I(r_i)} - \overline{I(N(r_i))}}{d}, \quad (98)$$

where $I(r_i)$ is the gray intensity within the VOI, $N(r_i)$ is the three-dimensional operator that provides a set of voxels that neighbor r_i but are outside the VOI, and d is the normalization term.

A different group of researchers developed a two-part feature of margin sharpness, but opposite to the previous ones, it was tested on CT scans [Xu et al. 2012]. The first feature quantifies the intensity difference between gray levels of the surrounding VOI

and itself through normal line segments across the boundary at fixed intervals around its circumference (Figure 1.14). The second feature quantifies the abruptness of the transition in intensity from the VOI to the surrounding region. The authors affirmed that a sharper border has a more abrupt transition and may have a higher intensity difference outside and inside the lesion. In contrast, a blurred border will have a smoother transition and may have a smaller intensity difference. For each normal line segment I perpendicular to the margin, the problem can be formulated as defined in Equation 1.99:

$$\arg \min_{S,W,x_o,L_o} \sum_x \left[L(x) - L_o - \frac{S}{1 + e^{-\frac{x-x_o}{W}}} \right], \quad (99)$$

where x is the distance along the normal, x_o is the intersection of the boundary point with the normal, $L(x)$ is the intensity along the normal at x , and L_o is the intensity offset [Xu et al. 2012].



Figura 1.14. Representation of an ortogonal line over the border to serve as reference for margin sharpness characterization. Source: Author.

1.2.4. Performance evaluation

Given the extraction of a set of n features, a performance evaluation is mandatory to assess the efficiency of all radiomic features to become biomarkers. All levels of evidence for biomarkers (Table 1.1) can use the following validation strategies [Lambin 2021].

At least two image cohorts from different sources should comprise the performance evaluation. One is used for discovery, and the other for validation purposes (Figure 1.15(a)). A widely used strategy in medicine to validate predictive models uses three data sets: one for discovery, one for testing, and one for external validation (Figure 1.15(b)). But it is not the most appropriate approach because not all samples are tested, and thus, not all patient heterogeneity evaluated [Keek et al. 2018, Larue et al. 2017]. Cross-validation, which separates the samples in m folds, from which $m - 1$ is for

Tabela 1.1. Levels of evidence for biomarkers. Source: Adapted from Lambin 2021.

Level	Study Design	Definition
I	Prospective	Marker as primary objective
II	Prospective	Marker as secondary objective
III	Retrospective	Multivariate analysis with outcomes
IV	Retrospective	Univariate analysis with outcomes
V	Retrospective	Correlation with other marker

discovery and 1 is for testing (Figure 1.15(c)), can mitigate this limitation. The cross-validation is performed until all samples of the initial fold are tested. It aims to improve the chances of generalizing the solution and decreasing the risk of overfitting. When m corresponds to the total number of samples, the process is then called leave-one-out cross-validation.

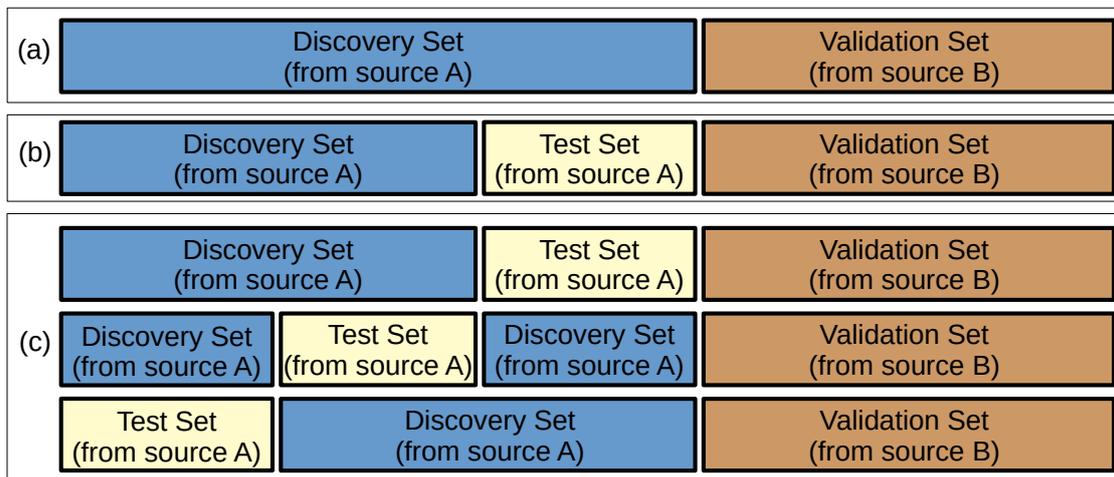


Figura 1.15. Representation of data setting for proper performance evaluation. Source: Author.

The terms *validation* and *testing* are interchangeable, depending on the scientific background. For instance, in medicine it is common to have the *validation* set used independently. But computing and engineering nomenclature considers the *testing* set as the independent one, and the *validation* set is used for fine-tuning.

Two of the most relevant methods used in performance evaluation are the receiver operating characteristic (ROC) curve and the confusion matrix. The ROC curve is defined as a graph of a resulting test where the axis y presents the sensitivity or true positive rate of the test, and the axis x shows the false positive rate defined as $1 - \text{specificity}$ (Figure 1.16). The area under the ROC curve (AUC) measures the final performance from the method. The AUC ranges 0-1 where $0 \leq \text{AUC} \leq 0.50$ indicate bad performance, $0.50 < \text{AUC} \leq 0.70$ indicate low performance, $0.70 < \text{AUC} \leq 0.85$ indicate moderate performance, and $0.85 < \text{AUC} \leq 1$ indicate a high performance [Carvalho et al. 2018, Dou et al. 2018, Yip et al. 2017].

The metrics sensitivity (Equation 1.100) and specificity (Equation 1.101) refer to the proportion of positive and negative cases, respectively, classified correctly as such.

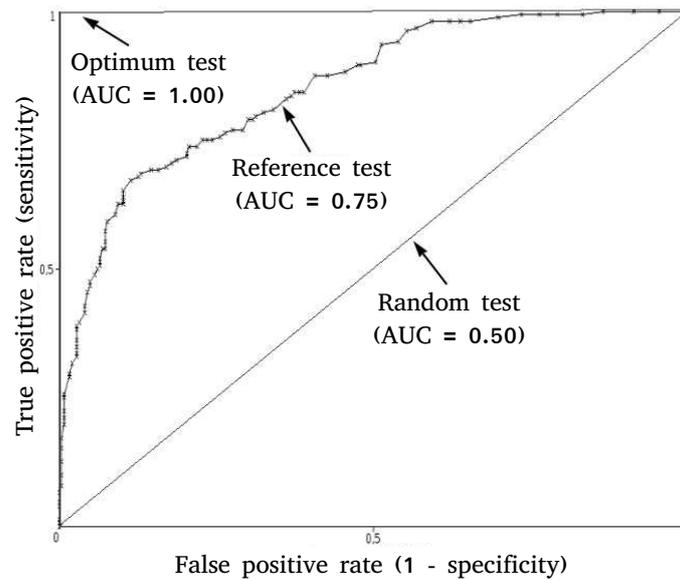


Figura 1.16. ROC curves used for performance evaluation. Source: Author.

Both measures can be calculated from the confusion matrix, which presents the number of samples considered as true-positive (TP), true-negative (TN), false-positive (FP), and false-negative (FN), as in Table 1.2 [Ferreira Junior et al. 2018].

$$\text{sensitivity} = \frac{TP}{TP + FN}, \quad (100)$$

$$\text{specificity} = \frac{TN}{TN + FP}. \quad (101)$$

Tabela 1.2. Confusion matrix construction. TP, TN, FP, and FN are the number of true positives, true negatives, false positives, and false negatives, respectively. Source: Author.

		Test	
		Positive	Negative
Real	Positive	TP	FN
	Negative	FP	TN

A particular method called survival or time-to-event analysis is used to evaluate the performance of a prediction procedure that constitutes a variable relating to the time between the beginning of a study and the occurrence of an event [Ferreira Junior 2019]. The event is not necessarily attached to death, as the term survival indicates, but to any clinical outcome, such as recurrence, hospitalization, among others [George et al. 2014]. Historically, overall survival is considered the temporal outcome of most importance to the clinical practice due to its objectivity and unambiguity [Cheema and Burkes 2013]. In this particular case, the event of interest is the patient's death by any nature. In survival analyses, the patients who did not reach the event or had lost follow-up are censored as the exact survival time is unknown [Ferreira Junior et al. 2021d].

The Kaplan-Meier estimation is a statistical method used for time-to-event evaluation [Kaplan and Meier 1958]. Once defined the event of interest and reasons for censorship, it is possible to build probability curves that describe the event occurrence rate along the time, allowing the comparison of different patient groups (Figure 1.17) [Ferreira Junior 2019]. The survival probability $S(t)$ in a time instant t is computationally given by Equation 1.102:

$$S(t) = S(t - 1) \times \frac{N_s}{N_r}, \tag{102}$$

where $S(t - 1)$ is the likelihood in a previous instant to t , N_r is the number of patients at risk in the study excluding the censored cases in the instant $t - 1$, and N_s is the number of patients that survived until the instant t .

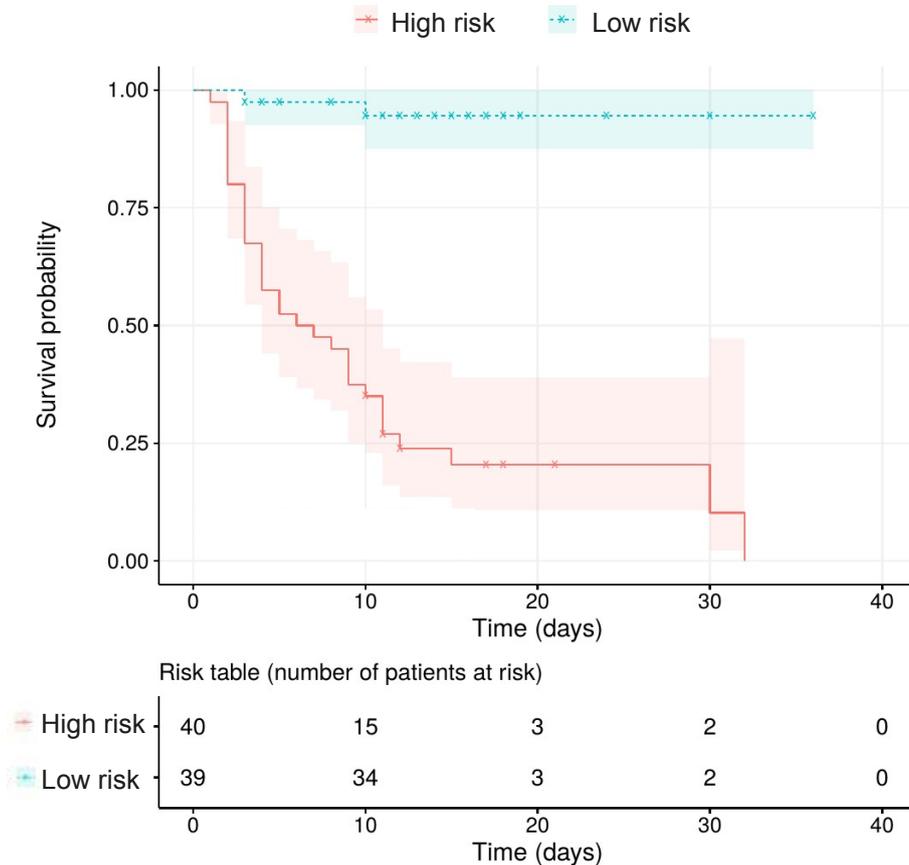


Figura 1.17. Example of Kaplan-Meier curves with confidence intervals and risk table for different patient groups (a high and a low risk). Each marked point in the curves represents a censored patient. Source: Author.

1.3. State-of-the-art

1.3.1. Covid-19 in chest XR

The covid-19 pandemic is the current major public health issue in the world that has caused over 3 million deaths in less than 18 months [European Centre for Disease Prevention and Control 2021]. The disease is mainly characterized by an inflammatory process in

the lungs and airways. The earliest possible diagnosis of covid-19 is imperative for the patient's isolation to prevent virus spread and for rapid treatment decisions to improve the patient's prognosis [Greenhalgh et al. 2020, Osman et al. 2021, Chiu et al. 2020].

A few alternatives exist as a screening tool for appropriate triage of suspected and high-risk patients in several low-incoming healthcare centers with a high demand for suspicious cases. One of those alternatives is chest XR, due to its high availability and portability [Ferreira Junior et al. 2021b, Wehbe et al. 2020, Ferreira Junior et al. 2021a]. However, it is widely known that XR has limited performance in the current clinical environment compared to CT, especially to assess covid-19 pneumonia in early disease stages with very subtle characteristics [Zhang et al. 2021, Wong et al. 2020, Rajaraman et al. 2020]. But radiomics can support the XR assessment of covid-19 and improve the identification of disease-related lesions [Chiu et al. 2020, Zhang et al. 2021, Wehbe et al. 2020, Rajaraman et al. 2020].

Ferreira Junior et al. identified 51 radiomic biomarkers in chest XR; most of them were higher-order features extracted after the Coiflet wavelet, for covid-19 [Ferreira Junior et al. 2021b]. The GLDM feature of *small dependence low gray level emphasis* (Equation 1.82) after the Coiflet transform obtained the highest performance, yielding an AUC of 0.87, sensitivity of 0.85, and specificity of 0.67 ($p < 0.001$). The authors found that higher values of the biomarker correlated with covid-19 patterns, even in cases with XR negative to discrete ground-glass opacities (Figure 1.18).

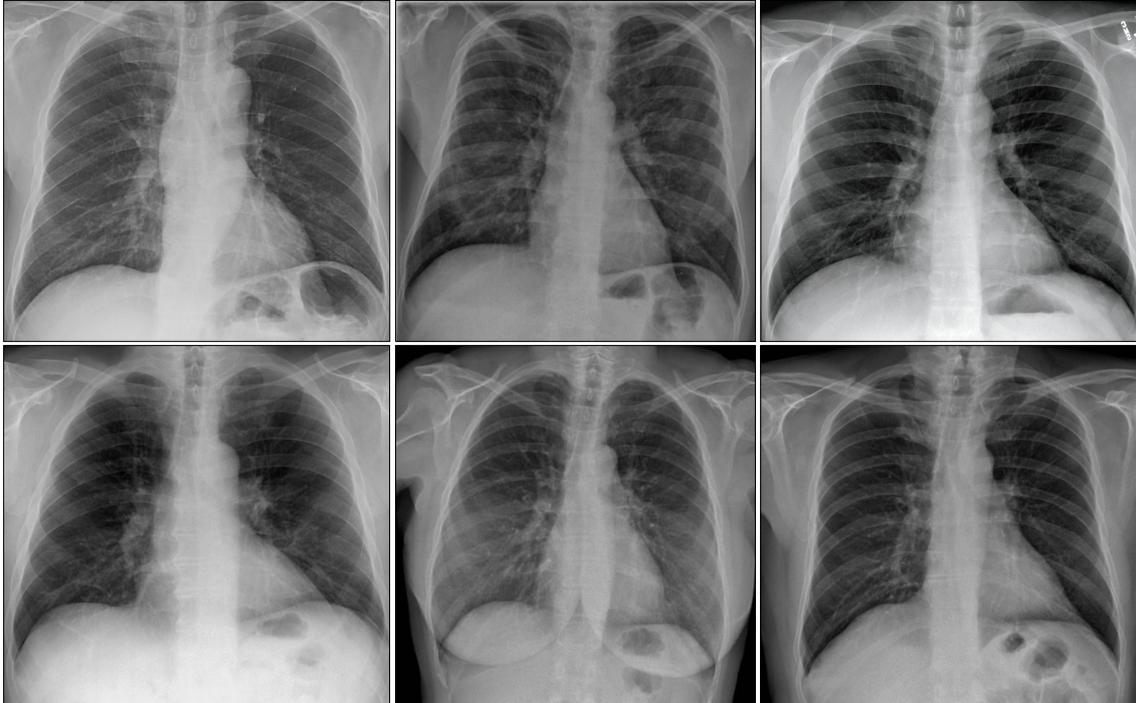


Figura 1.18. Examples of XR images from mild covid-19 patients. Source: Author.

The authors also identified other radiomic biomarkers with prognostic value to predict overall and deterioration-free survival. The first-order feature of *mean absolute deviation* (Equation 1.9) after the Coiflet transform yielded a significant difference in

overall survival rates from the stratified risk groups of covid-19 patients ($p < 0.05$). High values of the biomarker identified low-risk patients with a mean survival time of 25 days. Low values of the feature stratified patients with a higher risk of death, presenting a mean survival time of 13 days.

Furthermore, the biomarker *size zone non-uniformity* (Equation 1.55) of GLSZM after a square filter yielded the highest significant difference in Kaplan-Meier curves to predict short-term deterioration of the patient clinical status ($p < 0.05$). High values of the feature identified lower-risk patients, while low values stratified patients with a higher risk of rapid worsening, presenting a hazard ratio of 3.198 (95% confidence interval – CI: 1.145–8.932). Figure 1.19 presents the survival probabilities of the radiomic biomarkers.

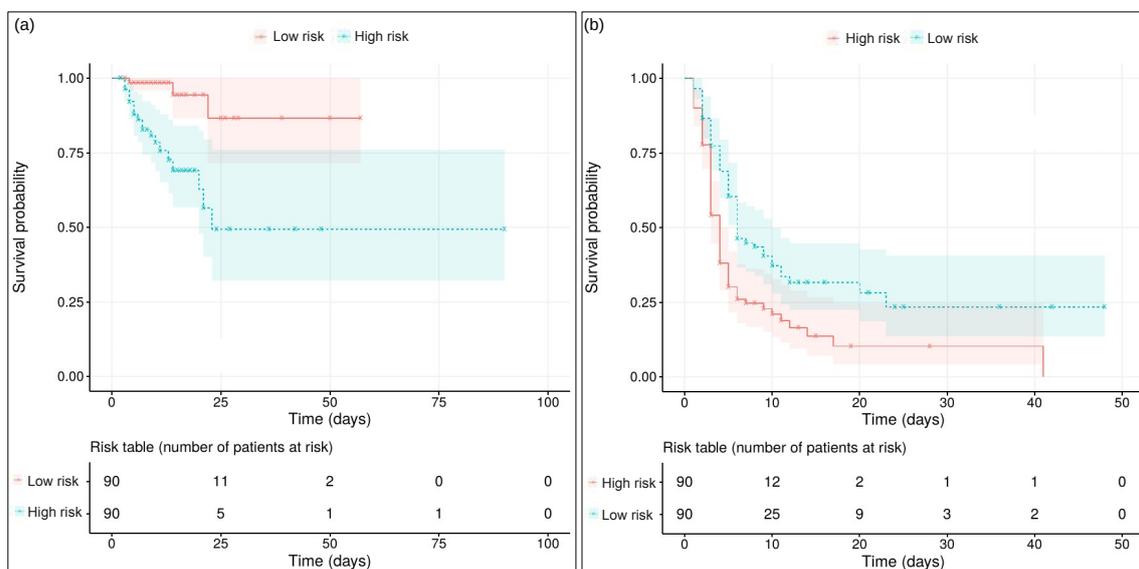


Figura 1.19. Kaplan-Meier curves and risk tables from prognostic radiomic biomarkers to predict death (a) and short-term deterioration (b). Source: Author.

Those XR features significantly associated with covid-19 outcomes could stratify the patient's short-term risk even without comorbidity conditions, at hospitalization, or any early stage of health care. These features could indicate the patient's rapid worsening before the clinical condition deteriorates when intensive therapy is more likely to have greater benefit [Ferreira Junior et al. 2021b]. Moreover, chest XR-based biomarkers may have a significant impact on supporting daily clinical decisions due to the accessibility of radiographic scanners.

The covid-19 biomarkers highlighted the challenge of visually recognizing intricate image characteristics, as they were discovered after algebraic filtering. In the early stages of the disease, small patchy shadows and interstitial changes emerge in the lungs when visible [Li et al. 2020]. But the wavelet transforms enabled to capture higher textural heterogeneity from covid-19 and not from other pneumonia etiologies on radiography [Ferreira Junior et al. 2021b].

Future investigations in this subject include correlating imaging with other clinical outcomes, assessing pathologic aggressiveness, and primarily identifying genetic traits associated with disease progression and therapeutical resistance. Studies have shown that

patients with critical or severe covid-19 have high expression of tyrosine kinase 2 and CC-chemokine receptor 2, respectively [Pairo-Castineira et al. 2021]. Those findings could bring significant clinical benefits with the existing drugs, and radiogenomics could play a key role in decoding covid-19 genotypes [Ferreira Junior and Cardona Cardenas 2021].

1.3.2. Lung neoplasms in computed tomography

Lung cancer is the most lethal malignant neoplasm globally, with a 5-year overall survival rate of about 15%. Moreover, the prognosis of patients with lung neoplasms is still poor and varies markedly according to tumor staging at diagnosis [Ferreira Junior et al. 2020b]. Several clinical aspects may influence therapy decision-making, like staging, histology, genomics, CT imaging, among others. Radiomic biomarkers emerged in this context to increase the clinical applicability of the previous computer-aided detection tools focused on the automated diagnosis of pulmonary nodules [Santos et al. 2019].

Ferreira Junior, Oliveira, and Azevedo-Marques proposed a novel margin sharpness characterization to classify the malignancy likelihood of pulmonary nodules in CT [Ferreira Junior et al. 2018]. The method extracts statistical properties across the nodule boundary in each CT scan (Figure 1.20) [Ferreira Junior and Oliveira 2015]. The authors discovered that the combination of the margin sharpness *amplitude* and the GLCM *inverse difference moment* (Equation 1.25) has potential to identify the malignancy of the lung lesions ($p < 0.05$). To do that, the authors used a standard decision tree to perform the multivariate classification [Ferreira Junior et al. 2018]. Calheiros et al. investigated further margin sharpness features by including perinodular zone characterization [Calheiros et al. 2021]. The developed method increased the performance when integrating parenchyma-originated features of the histogram *skewness* (Equation 1.10), and the GLCM *prominence*, *shade*, *correlation*, *energy*, and *entropy* (Equations 1.13, 1.14, 1.16, 1.20, and 1.29, respectively).

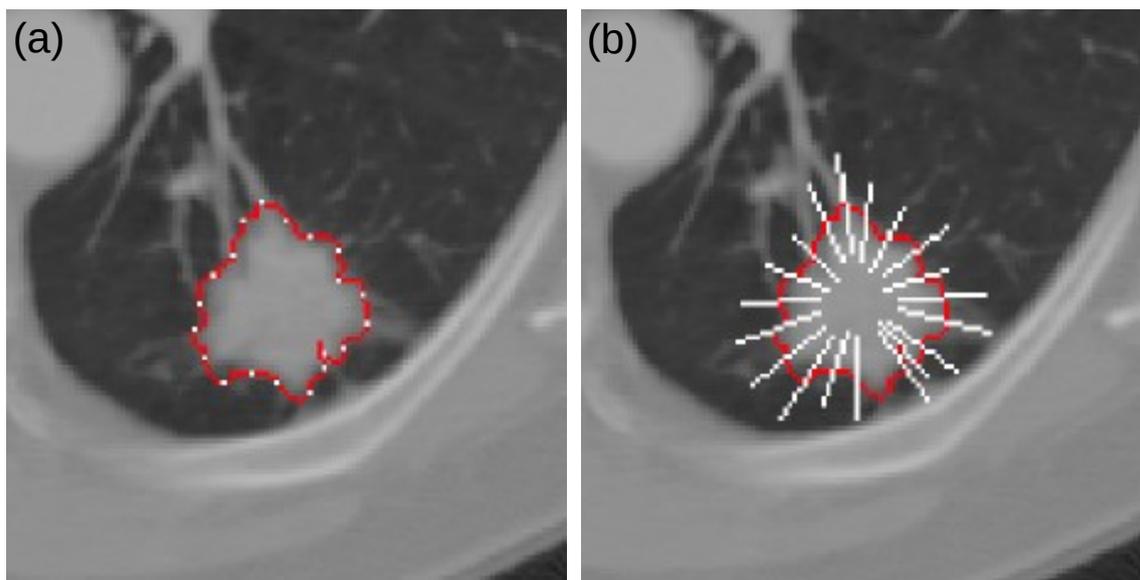


Figura 1.20. Margin sharpness characterization based on boundary control points (a) and normal line segments (b). Source: Author.

In another work, Ferreira Junior et al. found that several features are associated with nodal and distant metastases and could serve as biomarkers for tumor staging ($p < 0.05$) [Ferreira Junior et al. 2020b]. Some of them are the *energy* (Equation 1.6) after the Haar wavelet, GLCM IMCs (Equations 1.21 and 1.22), NIDM *busyness* (Equation 1.69), *directionality* from Tamura, estimation of *fractal dimension*, *surface area* (Equation 1.89), and *volume* (Equation 1.90). Moreover, the shape *diameter* and the GLCM feature of *maximum probability* (Equation 1.30) distinguished types of non-small cell lung cancer (NSCLC) and could serve as biomarkers for histology. The greater are their values, the greater are the chances of tumor to be squamous cell carcinoma, indicating more homogeneous patterns on CT for it and heterogeneous for adenocarcinoma (Figure 1.21). This finding was confirmed in [Zhu et al. 2018, Digumarthy et al. 2019].

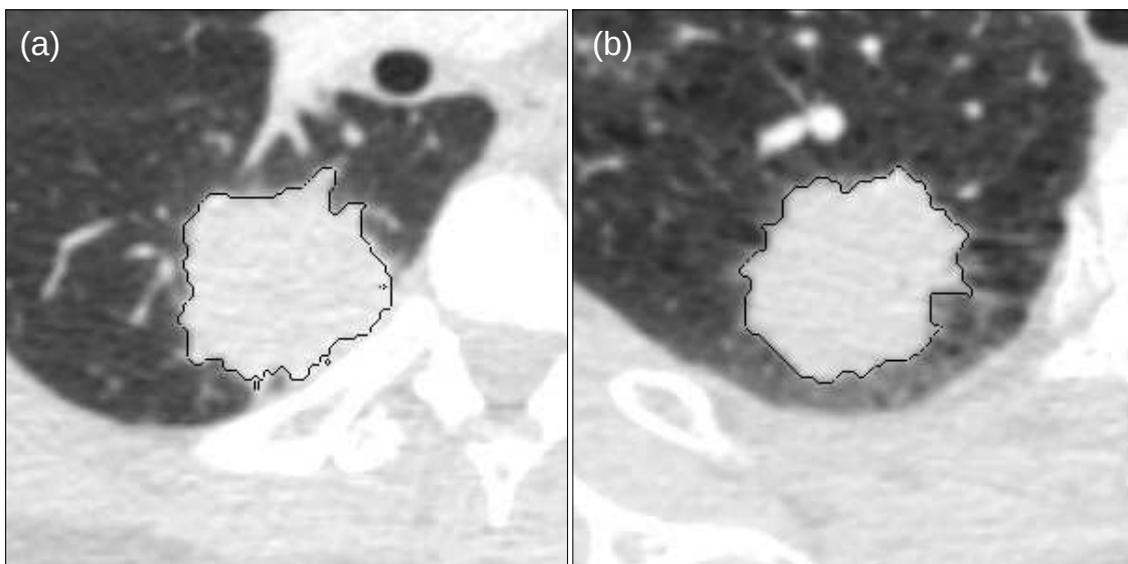


Figura 1.21. Non-small cell lung cancer types of adenocarcinoma (a) and squamous cell carcinoma (b). Source: Author.

van Timmeren et al. discovered three CT features with prognostic value for NSCLC: the mode (most common value) of the image histogram after a Laplacian-of-Gaussian filter, the mean intensity of a VOI centered on the highest gray level, and the GLCM *inverse variance* (Equation 1.27) calculated after a wavelet transform [van Timmeren et al. 2019]. Carvalho et al. discovered the GLRLM feature of *short-run emphasis* (Equation 1.35) on positron emission tomography images combined with CT correlates with the prognosis in patients with lung neoplasms [Carvalho et al. 2018]. Aerts et al. identified a radiomic signature associated with survival in patients with NSCLC composed of four features: first-order *energy* (Equation 1.6), shape *compactness* (Equation 1.93), GLRLM *non-uniformity* (Equation 1.74), and the previous one after a wavelet transform [Aerts et al. 2014]. Ferreira Junior et al. identified another prognostic biomarker for malignant neoplasms of the lung: the *mean* (Equation 1.2) after the Fourier transform ($p < 0.05$) [Ferreira Junior et al. 2021d]. Patients with a high value were identified as being at high risk with a hazard ratio of 2.12 (95% CI: 1.01–4.48). The authors also showed that the lesions from higher-risk patients have greater heterogeneity, and possibly be more aggressive, in comparison to lower-risk lesions, characterized by the presence of

more infiltrating regions (Figure 1.22) [Ferreira Junior et al. 2021c].

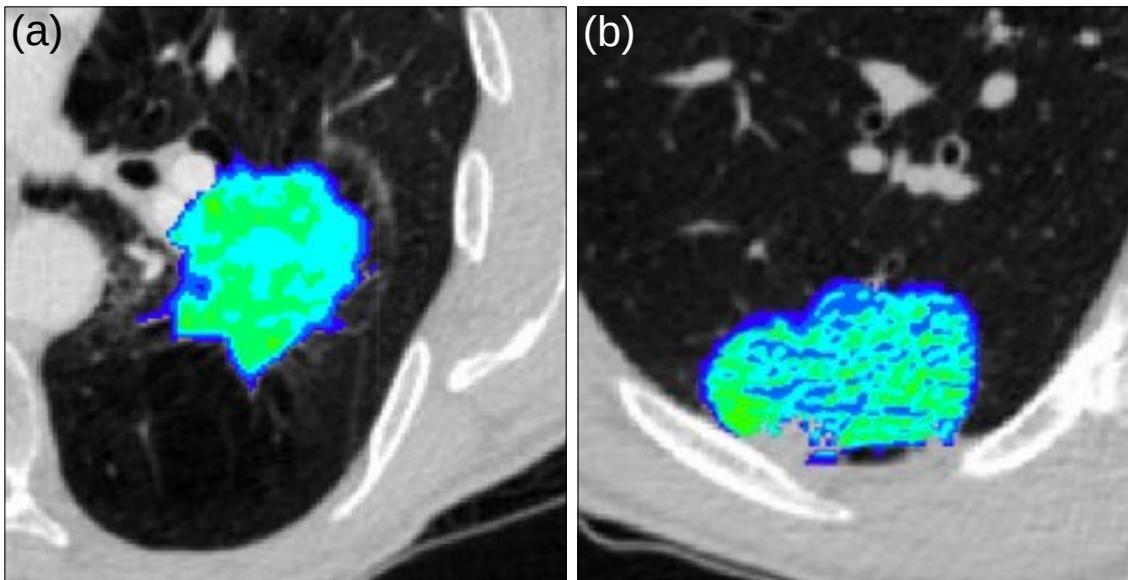


Figura 1.22. Intratumor heterogeneity of lung neoplasms with different staging: (a) stage-I, (b) stage-IV. Source: Author.

These findings highlight the importance of extracting features from all imaging levels. The discovered higher-order biomarkers correlate to smoother or rougher texture variations and thus, could quantify tumor heterogeneity [Ferreira Junior et al. 2021d]. That is important because intratumor heterogeneity can be associated with disease progression and treatment resistance, which ultimately serve as insight to targeted therapies needed for precision medicine [Ferreira Junior et al. 2021c].

Future directions for the area include identifying specific regions of resistance to targeted treatment, which would allow therapies located at a low molecular level.

1.3.3. Spondyloarthritis in magnetic resonance imaging

Spondyloarthritis (SpA) is a set of diseases with common clinical manifestations, such as inflammatory axial pain and peripheral arthritis. The active inflammation in sacroiliac joints, so-called sacroiliitis, is one of the most important criteria to diagnose SpA, and it can be identified in MRI [Rudwaleit et al. 2009]. The major MRI finding of active sacroiliitis is bone marrow edema (Figure 1.23). Therapy decision of SpA consists mainly on the subtype of the disease, *i.e.*, axial or peripheral [Sieper et al. 2009]. In this sense, MRI-based radiomics could play a key role in early diagnosis and therapy decision-making to indicate the presence of SpA and subtyping.

Tenorio et al. identified 63 MRI radiomic biomarkers specific for sacroiliitis, from which most of them were derived from a Gabor bank [Tenorio et al. 2020]. The histogram *skewness* (Equation 1.10) yielded an AUC of 0.86 ($p < 0.001$). High values characterized the active inflammation in MRI, as expected, due to the brightness patterns of the lesion (Figure 1.24). Faleiros et al. confirmed these findings by wrapping six radiomic features into an artificial neural network, yielding an AUC of 0.96 [Faleiros et al. 2020]. Those MRI biomarkers are the mean and standard deviation of the Tamura's *directionality*, *sum*

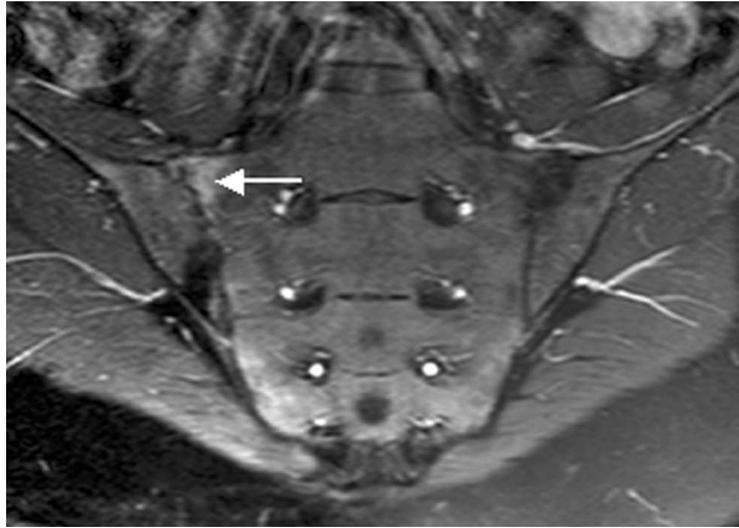


Figura 1.23. MRI of the sacroiliac joints with active inflammation (arrow). Source: Fiona McQueen, Marissa Lassere and Mikkel Ostergaard licensed under CC BY 2.0.

variance of the gray levels, maximum intensity, *mean* after a Gabor filter, and *energy* after the Haar wavelet (LH band on level 2).

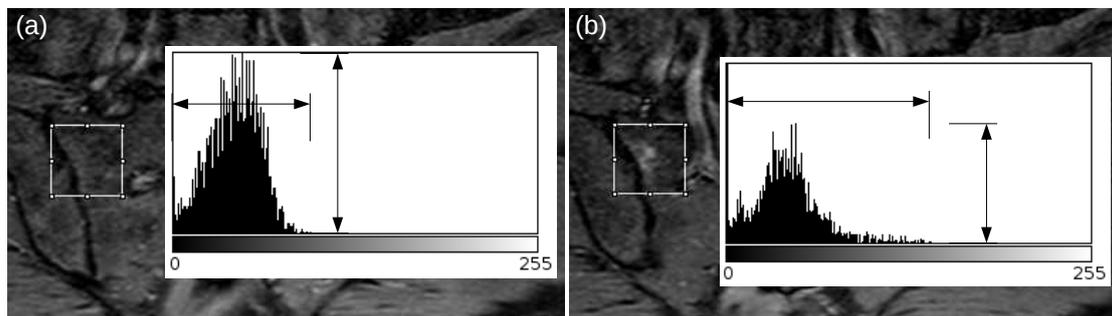


Figura 1.24. Sacroiliitis characterization using histogram and the skewness feature: (a) sacroiliac joint without active inflammation (histogram symmetrical), (b) sacroiliac joint with active inflammation (histogram skewed right). Source: Author.

Tenorio et al. also discovered 27 biomarkers for SpA [Tenorio et al. 2020]. Tamura features were predominant, and the *directionality* yielded the highest performance in identifying SpA with an AUC of 0.80 and distinguishing axial and peripheral with an AUC of 0.97 ($p < 0.001$). High values of the *directionality* standard deviation characterized axial SpA, distinguishing it from the peripheral form and other diseases, like arthrosis, fibromyalgia, and bone injury.

Although this clinical problem does not consider shape features, we highlight the use of all imaging levels for feature extraction. The studies showed that even though the first-order histogram is a simple strategy for characterization, it can quantify the images comprehensively enough to recognize complex clinical patterns.

Future perspectives to improve radiomic biomarker discovery include confirming all findings by testing the features prospectively. It is also advisable the clinical validation of the biomarkers and hence the assessment of whether the biomarkers can impact clinical routine.

1.4. Conclusion

This book chapter introduced theoretical and practical quantitative biomarker development and discovery. We highlighted the valuable role that radiomics can have in precision medicine. Therefore, radiomic biomarkers disclose a vast potential to improve clinical practice.

Referências

- [Aerts et al. 2014] Aerts, H., Velazquez, E., Leijenaar, R., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Monshouwer, R., Kains, B., Rietveld, D., Hoebbers, F., Rietbergen, M., Leemans, C., Dekker, A., Quackenbush, J., Gillies, R., and Lambin, P. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 5(4006).
- [Amadasun and King 1989] Amadasun, M. and King, R. (1989). Textural features corresponding to textural properties. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(5):1264–1274.
- [Azevedo-Marques and Ferreira Junior 2021] Azevedo-Marques, P. M. and Ferreira Junior, J. R. (2021). Medical image analyst: A radiology career focused on comprehensive quantitative imaging analytics to improve healthcare. *Academic Radiology*. DOI:10.1016/j.acra.2020.11.028.
- [Baeßler et al. 2018] Baeßler, B., Mannil, M., Maintz, D., Alkadhi, H., and Manka, R. (2018). Texture analysis and machine learning of non-contrast T1-weighted MR images in patients with hypertrophic cardiomyopathy—preliminary results. *European Journal of Radiology*, 102:61–67.
- [Bianconi and Fernandez 2007] Bianconi, F. and Fernandez, A. (2007). Evaluation of the effects of Gabor filter parameters on texture classification. *Pattern Recognition*, 40(12):3325–3335.
- [Calheiros et al. 2021] Calheiros, J. L. L., Amorim, L. B. V., Lima, L. L., Lima Filho, A. F., Ferreira Junior, J. R., and Oliveira, M. C. (2021). The effects of perinodular features on solid lung nodule classification. *Journal of Digital Imaging*. DOI:10.1007/s10278-021-00453-2.
- [Carvalho et al. 2018] Carvalho, S., Leijenaar, R. T., Troost, E. G., van Timmeren, J. E., Oberije, C., van Elmpt, W., de Geus-Oei, L.-F., Bussink, J., and Lambin, P. (2018). 18F-fluorodeoxyglucose positron-emission tomography (FDG-PET)-Radiomics of metastatic lymph nodes and primary tumor in non-small cell lung cancer (NSCLC)—A prospective externally validated study. *PloS One*, 13(3):e0192859.

- [Cheema and Burkes 2013] Cheema, P. and Burkes, R. (2013). Overall survival should be the primary endpoint in clinical trials for advanced non-small-cell lung cancer. *Current Oncology*, 20(2):e150.
- [Chiu et al. 2020] Chiu, W. H. K., Vardhanabhuti, V., Poplavskiy, D., Yu, P. L. H., Du, R., Yap, A. Y. H., Zhang, S., Fong, A. H.-T., Chin, T. W.-Y., Lee, J. C. Y., et al. (2020). Detection of COVID-19 using deep learning algorithms on chest radiographs. *Journal of Thoracic Imaging*, 35(6):369–376.
- [Davnall et al. 2012] Davnall, F., Yip, C. S., Ljungqvist, G., Selmi, M., Ng, F., Sanghera, B., Ganeshan, B., Miles, K. A., Cook, G. J., and Goh, V. (2012). Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice? *Insights Into Imaging*, 3(6):573–589.
- [Dawes et al. 2017] Dawes, T. J., de Marvao, A., Shi, W., Fletcher, T., Watson, G. M., Wharton, J., Rhodes, C. J., Howard, L. S., Gibbs, J. S. R., Rueckert, D., et al. (2017). Machine learning of three-dimensional right ventricular motion enables outcome prediction in pulmonary hypertension: a cardiac MR imaging study. *Radiology*, 283(2):381–390.
- [Digumarthy et al. 2019] Digumarthy, S. R., Padole, A. M., Gullo, R. L., Sequist, L. V., and Kalra, M. K. (2019). Can CT radiomic analysis in NSCLC predict histology and EGFR mutation status? *Medicine*, 98(1):e13963.
- [Dou et al. 2018] Dou, T. H., Coroller, T. P., van Griethuysen, J. J., Mak, R. H., and Aerts, H. J. (2018). Peritumoral radiomics features predict distant metastasis in locally advanced NSCLC. *PloS One*, 13(11):e0206108.
- [Echegaray et al. 2015] Echegaray, S., Gevaert, O., Shah, R., Kamaya, A., Louie, J., Kothary, N., and Napel, S. (2015). Core samples for radiomics features that are insensitive to tumor segmentation: method and pilot study using CT images of hepatocellular carcinoma. *Journal of Medical Imaging*, 2(4):041011.
- [European Centre for Disease Prevention and Control 2021] European Centre for Disease Prevention and Control (2021). COVID-19 situation update worldwide. Online (last update on Jun 3, 2021); last access on Jun 4, 2021. Available at www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases.
- [Faleiros et al. 2020] Faleiros, M. C., Nogueira-Barbosa, M. H., Dalto, V. F., Ferreira Junior, J. R., Tenorio, A. P. M., Luppino-Assad, R., Louzada-Junior, P., Rangayyan, R. M., and Azevedo-Marques, P. M. (2020). Machine learning techniques for computer-aided classification of active inflammatory sacroiliitis in magnetic resonance imaging. *Advances in Rheumatology*, 60:1–10. DOI:10.1186/s42358-020-00126-8.
- [Ferreira et al. 2017] Ferreira, J. R., Azevedo-Marques, P. M., and Oliveira, M. C. (2017). Selecting relevant 3D image features of margin sharpness and texture for lung nodule retrieval. *International Journal of Computer Assisted Radiology and Surgery*, 12(3):509–517.

- [Ferreira Junior 2019] Ferreira Junior, J. R. (2019). *Framework for Classification, Content-Based Retrieval, and Radiomics of Medical Images: an investigation of quantitative biomarkers for lung cancer*. PhD thesis, Sao Carlos School of Engineering, University of Sao Paulo. DOI:10.11606/T.82.2020.tde-27022020-113956.
- [Ferreira Junior et al. 2021a] Ferreira Junior, J. R., Cardenas, D. A. C., Moreno, R. A., Rebelo, M. d. F. d. S., Krieger, J. E., and Gutierrez, M. A. (2021a). A general fully automated deep-learning method to detect cardiomegaly in chest x-rays. In *SPIE Medical Imaging 2021: Computer-Aided Diagnosis*, volume 11597, page 115972B. DOI:10.1117/12.2581980.
- [Ferreira Junior et al. 2020a] Ferreira Junior, J. R., Cardenas, D. A. C., Moreno, R. A., Rebelo, M. F. S., Krieger, J. E., and Gutierrez, M. A. (2020a). Multi-view ensemble convolutional neural network to improve classification of pneumonia in low contrast chest x-ray images. In *42nd Annual International Conferences of the IEEE Engineering in Medicine and Biology Society*, pages 1238–1241. DOI:10.1109/EMBC44109.2020.9176517.
- [Ferreira Junior et al. 2021b] Ferreira Junior, J. R., Cardenas, D. A. C., Moreno, R. A., Rebelo, M. F. S., Krieger, J. E., and Gutierrez, M. A. (2021b). Novel chest radiographic biomarkers for COVID-19 using radiomic features associated with diagnostics and outcomes. *Journal of Digital Imaging*. DOI:10.1007/s10278-021-00421-w.
- [Ferreira Junior and Cardona Cardenas 2021] Ferreira Junior, J. R. and Cardona Cardenas, D. A. (2021). The potential role of radiogenomics in precision medicine for covid-19. *Journal of Thoracic Imaging*, 36(3):W34.
- [Ferreira Junior et al. 2021c] Ferreira Junior, J. R., Koenigkam-Santos, M., de Vita Graves, C., Correia, N. S. C., Cipriano, F. E. G., Fabro, A. T., and Azevedo-Marques, P. M. (2021c). Quantifying intratumor heterogeneity of lung neoplasms with radiomics. *Clinical Imaging*, 74:27–30.
- [Ferreira Junior et al. 2021d] Ferreira Junior, J. R., Koenigkam-Santos, M., Machado, C. V. B., Faleiros, M. C., Correia, N. S. C., Cipriano, F. E. G., Fabro, A. T., and Azevedo-Marques, P. M. (2021d). Radiomic analysis of lung cancer for the assessment of patient prognosis and intratumor heterogeneity. *Radiologia Brasileira*, 54(2):87–93.
- [Ferreira Junior and Oliveira 2015] Ferreira Junior, J. R. and Oliveira, M. C. (2015). Evaluating margin sharpness analysis on similar pulmonary nodule retrieval. In *Computer-Based Medical Systems, IEEE 28th International Symposium on*, pages 60–65. DOI:10.1109/CBMS.2015.16.
- [Ferreira Junior et al. 2018] Ferreira Junior, J. R., Oliveira, M. C., and Azevedo-Marques, P. M. (2018). Characterization of pulmonary nodules based on features of margin sharpness and texture. *Journal of Digital Imaging*, 31(4):451–463.
- [Ferreira Junior et al. 2017] Ferreira Junior, J. R., Oliveira, M. C., and de Azevedo-Marques, P. M. (2017). Integrating 3D image descriptors of margin sharpness and texture on a GPU-optimized similar pulmonary nodule retrieval engine. *The Journal of Supercomputing*, 73(8):3451–3467.

- [Ferreira Junior et al. 2020b] Ferreira Junior, J. R., Santos, M. K., Tenorio, A. P. M., Faleiros, M. C., Cipriano, F. E. G., Fabro, A. T., Nappi, J., Yoshida, H., and Azevedo Marques, P. M. (2020b). CT-based radiomics for prediction of histologic subtype and metastatic disease in primary malignant lung neoplasms. *International Journal of Computer Assisted Radiology and Surgery*, 15:163–172.
- [Galloway 1975] Galloway, M. (1975). Texture analysis using gray level run lengths. *Computer Graphics and Image Processing*, 4(2):172–179.
- [George et al. 2014] George, B., Seals, S., and Aban, I. (2014). Survival analysis and regression models. *Journal of Nuclear Cardiology*, 21(4):686–694.
- [Gilhuijs et al. 1998] Gilhuijs, K. G., Giger, M. L., and Bick, U. (1998). Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging. *Medical Physics*, 25(9):1647–1654.
- [Gonzalez and Woods 2007] Gonzalez, R. C. and Woods, R. E. (2007). *Digital Image Processing*. Prentice Hall, New Jersey, USA.
- [Greenhalgh et al. 2020] Greenhalgh, T. et al. (2020). Management of post-acute covid-19 in primary care. *BMJ*, 370:m3026.
- [Haralick et al. 1973] Haralick, R., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, SMC-3(6):610–621.
- [Kaplan and Meier 1958] Kaplan, E. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.
- [Keek et al. 2018] Keek, S. A., Leijenaar, R. T., Jochems, A., and Woodruff, H. C. (2018). A review on radiomics and the future of theragnostics for patient selection in precision medicine. *The British Journal of Radiology*, 91(1091):20170926.
- [Kermany et al. 2018] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., et al. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5):1122–1131.
- [Kickingreder et al. 2016] Kickingreder, P., Bonekamp, D., Nowosielski, M., Kratz, A., Sill, M., Burth, S., Wick, A., Eidel, O., Schlemmer, H.-P., Radbruch, A., et al. (2016). Radiogenomics of glioblastoma: machine learning–based classification of molecular characteristics by using multiparametric and multiregional MR imaging features. *Radiology*, 281(3):907–918.
- [Kolossvary et al. 2018] Kolossvary, M., Kellermayer, M., Merkely, B., and Maurovich-Horvat, P. (2018). Cardiac computed tomography radiomics: A comprehensive review on radiomic techniques. *Journal of Thoracic Imaging*, 33(1):26–34.

- [Lambin 2021] Lambin, P. (2021). Radiomics: transforming standard imaging into mineable data for diagnostic and theragnostic applications. In *SPIE Medical Imaging 2021: Physics of Medical Imaging*, volume 11595, page 1159502. DOI:10.1117/12.2585711.
- [Larue et al. 2017] Larue, R. T., Defraene, G., De Ruysscher, D., Lambin, P., and Van Elmpt, W. (2017). Quantitative radiomics studies for tissue characterization: a review of technology and methodological procedures. *The British Journal of Radiology*, 90(1070):20160665.
- [LeCun et al. 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- [Lee et al. 2019] Lee, G. R., Gommers, R., Waselewski, F., Wohlfahrt, K., and O’Leary, A. (2019). Pywavelets: A python package for wavelet analysis. *Journal of Open Source Software*, 4(36):1237.
- [Leger et al. 2017] Leger, S., Zwanenburg, A., Pilz, K., Lohaus, F., Linge, A., Zophel, K., Kotzerke, J., Schreiber, A., Tinhofer, I., Budach, V., et al. (2017). A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. *Scientific Reports*, 7(1):13206.
- [Levman and Martel 2011] Levman, J. E. and Martel, A. L. (2011). A margin sharpness measurement for the diagnosis of breast cancer from magnetic resonance imaging examinations. *Academic Radiology*, 18(12):1577–1581.
- [Li et al. 2020] Li, M., Lei, P., Zeng, B., Li, Z., Yu, P., Fan, B., Wang, C., Li, Z., Zhou, J., Hu, S., et al. (2020). Coronavirus disease (COVID-19): spectrum of CT findings and temporal progression of the disease. *Academic Radiology*, 27(5):603–608.
- [Liang and Zheng 2019] Liang, G. and Zheng, L. (2019). A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Computer Methods and Programs in Biomedicine*. Ahead of print. DOI:10.1016/j.cmpb.2019.06.023.
- [Lu et al. 2019] Lu, M. T., Ivanov, A., Mayrhofer, T., Hosny, A., Aerts, H. J., and Hoffmann, U. (2019). Deep learning to assess long-term mortality from chest radiographs. *JAMA Network Open*, 2(7):e197416–e197416.
- [Lubner et al. 2017] Lubner, M. G., Smith, A. D., Sandrasegaran, K., Sahani, D. V., and Pickhardt, P. J. (2017). CT texture analysis: definitions, applications, biologic correlates, and challenges. *Radiographics*, 37(5):1483–1503.
- [Osman et al. 2021] Osman, A. H., Aljhdali, H. M., Altarrazi, S. M., and Ahmed, A. (2021). SOM-LWL method for identification of COVID-19 on chest x-rays. *PloS One*. DOI:10.1371/journal.pone.0247176.
- [Pairo-Castineira et al. 2021] Pairo-Castineira, E., Clohisey, S., Klaric, L., Bretherick, A. D., Rawlik, K., Pasko, D., Walker, S., Parkinson, N., Fourman, M. H., Russell, C. D., et al. (2021). Genetic mechanisms of critical illness in covid-19. *Nature*, 591(7848):92–98.

- [Parker 2011] Parker, J. R. (2011). *Algorithms for Image Processing and Computer Vision*. Wiley Publishing, Indianapolis, USA.
- [Phillips et al. 2017] Phillips, I., Ajaz, M., Ezhil, V., Prakash, V., Alobaidli, S., McQuaid, S. J., South, C., Scuffham, J., Nisbet, A., and Evans, P. (2017). Clinical applications of textural analysis in non-small cell lung cancer. *The British Journal of Radiology*, 91(1081):20170267.
- [Rajaraman et al. 2020] Rajaraman, S., Sornapudi, S., Alderson, P. O., Folio, L. R., and Antani, S. K. (2020). Analyzing inter-reader variability affecting deep ensemble learning for COVID-19 detection in chest radiographs. *PloS One*, 15(11):e0242301.
- [Rudwaleit et al. 2009] Rudwaleit, M., Jurik, A.-G., Hermann, K. A., Landewé, R., van der Heijde, D., Baraliakos, X., Marzo-Ortega, H., Østergaard, M., Braun, J., and Sieper, J. (2009). Defining active sacroiliitis on magnetic resonance imaging (MRI) for classification of axial spondyloarthritis: a consensual approach by the ASAS/OMERACT MRI group. *Annals of the Rheumatic Diseases*, 68(10):1520–1527.
- [Sacconi et al. 2017] Sacconi, B., Anzidei, M., Leonardi, A., Boni, F., Saba, L., Scipione, R., Anile, M., Rengo, M., Longo, F., Bezzi, M., et al. (2017). Analysis of ct features and quantitative texture analysis in patients with lung adenocarcinoma: a correlation with egfr mutations and survival rates. *Clinical Radiology*, 72(6):443–450.
- [Santos et al. 2019] Santos, M. K., Ferreira Junior, J. R., Wada, D. T., Tenorio, A. P. M., Barbosa, M. H. N., and Azevedo Marques, P. M. (2019). Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine. *Radiologia Brasileira*, 52(6):387–396.
- [Schneider et al. 2012] Schneider, C., Rasband, W., and Eliceiri, K. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7):671–675.
- [Sieper et al. 2009] Sieper, J., Rudwaleit, M., Baraliakos, X., Brandt, J., Braun, J., Burgos-Vargas, R., Dougados, M., Hermann, K., Landewe, R., Maksymowych, W., et al. (2009). The Assessment of SpondyloArthritis international Society (ASAS) handbook: a guide to assess spondyloarthritis. *Annals of the Rheumatic Diseases*, 68(Suppl 2):ii1–ii44.
- [Sun and Wee 1983] Sun, C. and Wee, W. G. (1983). Neighboring gray level dependence matrix for texture classification. *Computer Vision, Graphics, and Image Processing*, 23(3):341–352.
- [Sun et al. 2017] Sun, H., Chen, Y., Huang, Q., Lui, S., Huang, X., Shi, Y., Xu, X., Sweeney, J. A., and Gong, Q. (2017). Psychoradiologic utility of MR imaging for diagnosis of attention deficit hyperactivity disorder: a radiomics analysis. *Radiology*, 287(2):620–630.
- [Tamura et al. 1978] Tamura, H., Mori, S., and Yamawaki, T. (1978). Textural features corresponding to visual perception. *IEEE Transactions on Systems, Man and Cybernetics*, 8(6):460–473.

- [Tang 1998] Tang, X. (1998). Texture information in run-length matrices. *IEEE Transactions on Image Processing*, 7(11):1602–1609.
- [Tenorio et al. 2020] Tenorio, A. P. M., Faleiros, M. C., Junior, J. R. F., Dalto, V. F., Assad, R. L., Louzada-Junior, P., Yoshida, H., Nogueira-Barbosa, M. H., and Azevedo-Marques, P. M. (2020). A study of MRI-based radiomics biomarkers for sacroiliitis and spondyloarthritis. *International Journal of Computer Assisted Radiology and Surgery*, 15(10):1737–1748.
- [Thibault et al. 2013] Thibault, G., Angulo, J., and Meyer, F. (2013). Advanced statistical matrices for texture characterization: application to cell classification. *IEEE Transactions on Biomedical Engineering*, 61(3):630–637.
- [Tomaszewski and Gillies 2021] Tomaszewski, M. R. and Gillies, R. J. (2021). The biological meaning of radiomic features. *Radiology*, 298(3):505–516.
- [Van Griethuysen et al. 2017] Van Griethuysen, J. J., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G., Fillion-Robin, J.-C., Pieper, S., and Aerts, H. J. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer Research*, 77(21):e104–e107.
- [van Timmeren et al. 2019] van Timmeren, J. E., van Elmpt, W., Leijenaar, R. T., Reymen, B., Monshouwer, R., Bussink, J., Paelinck, L., Bogaert, E., De Wagter, C., Elhaseen, E., et al. (2019). Longitudinal radiomics of cone-beam ct images from non-small cell lung cancer patients: evaluation of the added prognostic value for overall survival and locoregional recurrence. *Radiotherapy and Oncology*, 136:78–85.
- [Wehbe et al. 2020] Wehbe, R. M., Sheng, J., Dutta, S., Chai, S., Dravid, A., Barutcu, S., Wu, Y., Cantrell, D. R., Xiao, N., Allen, B. D., et al. (2020). DeepCOVID-XR: An artificial intelligence algorithm to detect COVID-19 on chest radiographs trained and tested on a large US clinical dataset. *Radiology*. DOI:10.1148/radiol.2020203511.
- [Wong et al. 2020] Wong, H. Y. F., Lam, H. Y. S., Fong, A. H.-T., Leung, S. T., Chin, T. W.-Y., Lo, C. S. Y., Lui, M. M.-S., Lee, J. C. Y., Chiu, K. W.-H., Chung, T. W.-H., et al. (2020). Frequency and distribution of chest radiographic findings in patients positive for COVID-19. *Radiology*, 296(2):E72–E78.
- [Xu et al. 2012] Xu, J., Napel, S., Greenspan, H., Beaulieu, C. F., Agrawal, N., and Rubin, D. (2012). Quantifying the margin sharpness of lesions on radiological images for content-based image retrieval. *Medical Physics*, 39:5405–5418.
- [Yang et al. 2016] Yang, J., Zhang, L., Fave, X., Fried, D., Stingo, F., Ng, C., and Court, L. (2016). Uncertainty analysis of quantitative imaging features extracted from contrast-enhanced CT in lung tumors. *Computerized Medical Imaging and Graphics*, 48:1–8.
- [Yip et al. 2017] Yip, S. S., Liu, Y., Parmar, C., Li, Q., Liu, S., Qu, F., Ye, Z., Gillies, R. J., and Aerts, H. J. (2017). Associations between radiologist-defined semantic and automatically computed radiomic features in non-small cell lung cancer. *Scientific Reports*, 7(1):3519.

- [Zhang et al. 2015] Zhang, L., Fried, D., Fave, X., Hunter, L., Yang, J., and Court, L. (2015). IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. *Medical Physics*, 42(3):1341–1353.
- [Zhang et al. 2021] Zhang, R., Tie, X., Qi, Z., Bevins, N. B., Zhang, C., Griner, D., Song, T. K., Nadig, J. D., Schiebler, M. L., Garrett, J. W., et al. (2021). Diagnosis of coronavirus disease 2019 pneumonia by using chest radiography: Value of artificial intelligence. *Radiology*, 298(2):E88–E97.
- [Zhu et al. 2018] Zhu, X., Dong, D., Chen, Z., Fang, M., Zhang, L., Song, J., Yu, D., Zang, Y., Liu, Z., Shi, J., et al. (2018). Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *European radiology*, 28(7):2772–2778.
- [Zwanenburg et al. 2020] Zwanenburg, A., Vallieres, M., Abdalah, M. A., Aerts, H. J., Andrearczyk, V., Apte, A., Ashrafinia, S., Bakas, S., Beukinga, R. J., Boellaard, R., et al. (2020). The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, 295(2):328–338.