Alma Mater Studiorum · University of Bologna

School of Science
Department of Physics and Astronomy
Master Degree in Physics

Implementation of a radiomics pipeline for survival analysis in multiple myeloma patients using ^{18}F FDG PET/CT images: unveiling prognostic markers and predictive models

Supervisor:
Prof. Gastone Castellani

Co-supervisor:
Prof. Federico Álvarez García
Eng. Anaida Fernández García
Prof. Nico Curti

Submitted by:
Alessandro Ceresi

Academic Year 2023/2024

Abstract

Multiple Myeloma (MM) stands out as a cancer with a high mortality rate, presenting a formidable challenge to clinicians. The complexities lie in establishing standardized and reliable diagnostic criteria, determining optimal personalized treatment strategies, and providing a robust prognostic assessment of its progression. Clinical evaluations and medical tests are indispensable for physicians handling this challenging disease. However, medical imaging scans, such as Computed Tomography (CT) and Positron Emission Tomography (PET), play a pivotal role in detecting lesions and abnormalities in patients. Embracing the realm of radiomics, these advanced imaging techniques offer a wealth of information. By extracting quantitative features from medical images, radiomics contribute significantly to refining diagnostic precision, staging lesions, and improving treatment planning. This thesis presents a comprehensive radiomics pipeline involving CT and PET images of MM patients. The image analysis workflow begins with segmentation of the patient's spine performed with MOOSE, in order to focus on the region where the disease is more active. From the segmentation three versions are obtained: original segmentation, one including the bone marrow region, and one considering also the surrounding. Further, the medical scans are shaped following the three types of segmentation and feature extraction is performed. The study employs survival analysis performed with the Cox model, to investigate the Progression-Free Survival (PFS) of MM patients, aiming to identify prognostic biomarkers and develop predictive models. The findings reveal important guidelines on the pipeline. The spine does not seem to be an optimal region for obtaining prognostic insights of this nature, providing poor results compared to other studies, where have been involved all the skeleton regions. However CT images yield higher concordance index (C-index) scores compared to PET images. The study also highlighted the differences in results between the segmentation shapes selected and the image filters employed to extract information. Surprising evidence from the feature selection model suggests that shape features, particularly "flatness", consistently emerge in feature combinations, and the pairing of specific shape and texture features exhibits promising results in Cox model predictions. This study sheds light on valuable insights for MM prognosis, emphasizing the significance of specific image regions, imaging modality, and feature selection strategies in enhancing predictive modeling for patient outcomes.

Index

Index of Figures

Index of Tables

Acronym List

AI	Artificial Intelligence
PET	Positron Emission Tomography
CT	Computed Tomography
UPM	Universidad Politecnica de Madrid
CDSS	clinical-decision support systems
VoI	Volume of Interest
RoI	Region of Interest
DAS	Data Acquisition system
HU	Hounsfield units
LOR	Line Of Response
AC	Attenuation Correction
SUV	Standard Uptake Value
CNN	Convolutional Neural Networks
MOOSE	Multiple-Organ Objective Segmentation
PH	Proportional Hazard
HR	Hazard Ratio
MM	Multiple Myeloma
GCO	Global Cancer Observatory
SEER	Surveillance, Epidemiology, and End Results

MGUS	Monoclonal Gammopathy of Undetermined clinical Significance
IMWG	International Myeloma Working Group
MDE	Myeloma Defining Events
FLC	Free Light Chain
MRI	Magnetic Resonance Imaging
MS	Myelodysplastic Syndromes
SCD	Sickle Cell Disease
ML	Machine Learning
DL	Deep Learning
FL	Federated Learning
BM	Bone Marrow
NDMM	Newly Diagnosed Multiple Myeloma
IRCCS	Istituto di Ricovero e Cura a Carattere Scientific
TCIA	The Cancer Imaging Archive
GE	General Electric
FoV	Field of View
WB	Whole Body
PFS	Progression Free Survival
LASSO	Least Absolute Shrinkage and Selection Operator
C-index	concordance index
ISS	International Staging System
R-ISS	Revised International Staging System
FISH	Fluorescence In-Situ Hybridization
CA	Chromosomal Abnormalities
LDH	Lactate Dehydrogenase Level

IMPeTUs	Italian Myeloma criteria for PET Use
DS	Deauville Score
NIfTI	Neuroimaging Informatics Technology Initiative
DFWG	Data Format Working Group
PFS	Progression-Free Survival
DICOM	Digital Imaging and Communications in Medicine
JSON	JavaScript Object Notation
LoG	Laplacian of Gaussian
CV	Cross-Validation
nCV	nested Cross-Validation
SR	Stepwise Regression
PM	paramedullary
GLDM	Gray Level Dependence Matrix
IQR	InterQuartile Range
DWT	Discrete Wavelet Transform

Chapter 1

Introduction

In this introduction chapter will be outlined the context of the project and the general purposes of this work. The chapter also describes the structure and the processes of this work, providing a comprehensive overview of the research approach.

1.1 Project context

Nowadays, the use of Artificial Intelligence (AI) is experiencing accelerated adoption due to the dramatic growth in large, machine-accessible healthcare data, powerful computing systems, and innovative software technologies [10]. Although it has been a continuous endeavor of scientists and engineers for more than 65 years, today in fields such as industry, transportation, education, and many others, AI has consolidated and penetrated our daily lives [38]. Its effects are evident in today's society; from the automotive field with self-driving cars [18] to the finance sector with AI driven predictive analysis for trading and investment [28] there are many examples of how it is shaping our daily life. Just like an iceberg though, where only a small portion is visible above the waterline, beneath the surface there lies a complex and widespread network of developments and applications in the research environment, contributing to a deeper comprehension and growth of this challenging and powerful tool.

This work is involved in the healthcare field, where the application of AI is becoming more and more crucial over the years. Already in the early 2000s, people recognized the potential of AI defining as promising its application in almost all medical fields to increase the efficiency of treatment and improve patient well-being [68]. Today in fact we can count on many essential AI-driven imaging techniques and treatments that are the results of years of research from the more straightforward pathologies to those that, even in the present day, continue to elude experts complete understanding.

In the realm of extensively researched pathologies, heightened efforts for study and investigation are directed toward cancer diseases. These prevail over many others due to their profound impact on global health, high mortality rates, and the intricate challenges they pose to researchers and clinicians seeking effective treatments and preventive strategies. AI research is directed towards overcoming these obstacles to facilitate the development of suitable technologies for diagnosis, prognosis, and treatment. Numerous AI-driven technologies have been developed and implemented, showcasing their effectiveness. [20] Nevertheless, significant barriers persist in the implementation of AI in oncology:

- Biased and heterogeneous data, issues arising from variations in data sources, potentially impacting the accuracy and generalizability of results.
- Data management and collection burdens, challenges in the process of acquiring and obtaining data which may require significant time, resources, and precision.
- Lack of standardized research reporting, challenges in ensuring consistency and comparability across different studies and publications.
- Insufficient clinical validation, lack of uniform guidelines and opinions from experts.
- Strict regulatory and legal frameworks, issues related to compliance, privacy, and data protection in the context of evolving research practices.

These represent the main challenges and obstacles that must be addressed in the healthcare field [10].

The European project GenoMed4All [23] situates itself within this context and establishes as its goal the creation of a one of the first international implementations of a Federated Learning (FL) platform [13] as shown in Fig. 1.1. Federated learning is a distributed Machine Learning (ML) paradigm where multiple participants collaborate in the training of a unique global predictive model. It is conceived as a scalable and privacy-preserving approach to the joint training of ML models across federated health data repositories in order to overcome issues such as privacy protection, lack of widely adopted sharing solutions, and lack of availability of high-quality datasets comprising large volumes of data [13, 59].

GenoMed4All intends to combine multimodal data in its analysis. Combining medical imaging data, clinical records, genomic information, or sensor data from wearables, a more comprehensive understanding of health and diseases can be achievable. Furthermore, it aims to implement "white box" AI models in 3 real-world pilots, improving diagnosis and prognosis capacity, evaluating treatment options and patient response. The pilots cover common and rare haematological diseases: Myelodysplastic Syndromes (MS),

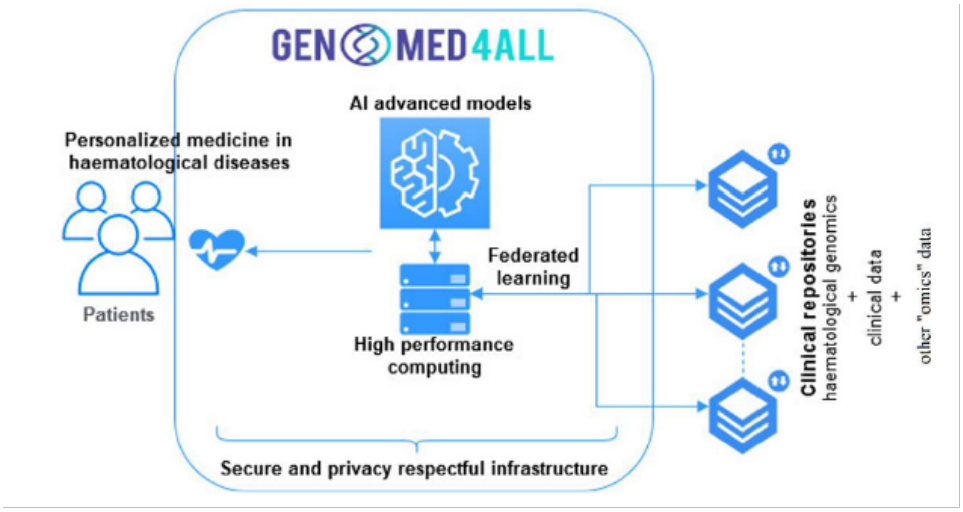


Figure 1.1: Visualization of the GenoMed4All structure. [23, 24]

Multiple Myeloma (MM) and Sickle Cell Disease (SCD). In this work, the attention will be focused on dealing with MM, an oncological disease noted for the short survival (compared to many other haematologic neoplasias) and uniformly fatal outcome of patients [53].

1.2 Objectives

In this document, the focus narrows to a specific facet within the broader project, with a close examination of the multiple myeloma disease by means of the instructions provided by the GenoMed4All project headlines. Given a dataset comprising CT and PET scans from a cohort of patients diagnosed with multiple myeloma, the goal is to extract pertinent insights that can guide clinicians in making informed decisions regarding diagnosis, prognosis, and treatment.

The primary objective of this research is to develop a model that, through survival analysis, can provide valuable information on the progression of MM disease from a prognostic point of view. This requests a comprehensive exploration of mathematical features extracted from medical images, identifying the ones that may be recognized as crucial prognostic markers. The focus extends beyond traditional prognostic factors, delving into details captured by image-derived features. Furthermore, the research emphasizes the significance of specific procedures integrated into the analytical pipeline, revealing their substantial impact on extracting meaningful insights from the dataset. By combining survival analysis with a nuanced examination of mathematical features and pipeline procedures, this study aims to contribute to a more comprehensive under-

standing of MM disease progression and prognosis.

In details the goals that each step wants to accomplish are the following:

- producing a segmentation of the region of the images where the information is relevant, the patient's spine.
- Obtaining an ensemble of compatible medical images, with the relative segmentation of the Region of Interest (RoI), applying to them also operations that enhance the quantity and quality of the information stored.
- Collecting a mathematical values description of the images' properties through feature extraction.
- Analyzing the extracted features to go one step further on the disease understanding, highlighting the most important information contained in the images.

1.3 Structure of the document

In order to make the analysis as complete and linear as possible, this work starts with an overview of the literature and a discussion of the topics treated, with a focus also on the work of a partner university's working team in Bologna. In the following chapters are presented the methods and the development brought in the Universidad Politecnica de Madrid (UPM) working environment and a discussion and comparison of the two different pipelines. Here is described the structure of the document:

- *Introduction*: overview of the motivations and the context of the project involving the work presented, main goals and document structure.
- *State of the art*: in-depth analysis of the literature in the fields of Multiple Myeloma to have a complete knowledge about the illness and the characterizing traits, diagnosis, prognosis, and treatment procedures. Moreover, in the chapter, the radiomics is explored in depth, what its role is in treatment of a disease and which are nowadays the principal State-of-the-Art methods and models in the field together with future further developments and potentiality. Here, the pipeline followed by Bologna's partner team [61] is described to better compare their analysis to the ones explained in this work.
- *Materials and methods*: detailed explanation of the workflow and sequential description of the steps followed, outlining the experimental design with criteria and sample rationale. The description starts from the available data and material and will explain and examine the development of analysis methods with all the models used in the work to extract useful information from the data.

- *Results and analysis*: presentation of key findings and their interpretation. It provides a clear presentation of collected data and employs relevant statistical or analytical methods for comprehensive analysis. This chapter objectively discusses trends, patterns, and relationships within the data, linking back to research objectives. Visual aids such as graphs or tables are used to enhance clarity, supporting a robust and insightful discussion of the results.

- *Conclusions and future lines*: succinct summary of the main results and their alignment with the research objectives. Highlight the contributions of the work to the field, discuss implications of the findings, and address any limitations encountered during the research. Propose avenues for future research, concluding with a reflective statement on the overall significance of your research.

Chapter 2

State of the art

This chapter delves into the current state of the art, presenting the latest and widely adopted workflows in both general radiomics and specific to MM. The techniques and procedures showcased here are backed by years of rigorous scientific research, establishing themselves as the gold standard for addressing specific situations. The section aims to provide a comprehensive overview of the current state of radiomics, particularly as it relates to multiple myeloma, exploring specific methodologies, cutting-edge technologies, and recent advancements that have shaped the landscape of radiomics in medical research. Furthermore, practical applications and case studies will be explored to exemplify the real-world impact of these techniques. By the end of this chapter, readers should gain not only a theoretical understanding of the state of the art, but also practical insights that can inform their approach to similar challenges in the field.

2.1 Radiomics

With the term *radiomics* are addressed all the processes for high-throughput extraction of quantitative features that result in the conversion of medical images into mineable data and the subsequent analysis of these data for decision support . It is a scientific discipline introduced in the last decades, and it was initiated in oncology studies, where it has found fertile ground. While radiomics primarily grew out of basic research, lately it has also elicited interest from those in clinical research and from those in daily clinical practice. The information resulting from a radiomic workflow can be exploited through quantitative image analyses and leveraged via clinical-decision support systems (CDSS) in order to provide a more personalized and efficient healthcare plan for the patient. Thus, radiomics has the potential to help with the diagnosis, prognosis, and treatment planning on both common and rare tumors where the only analysis and opinion of a clinical radiologist may not be enough. [26]
The suffix *-omics* originates from molecular biology disciplines, it addressed the charac-

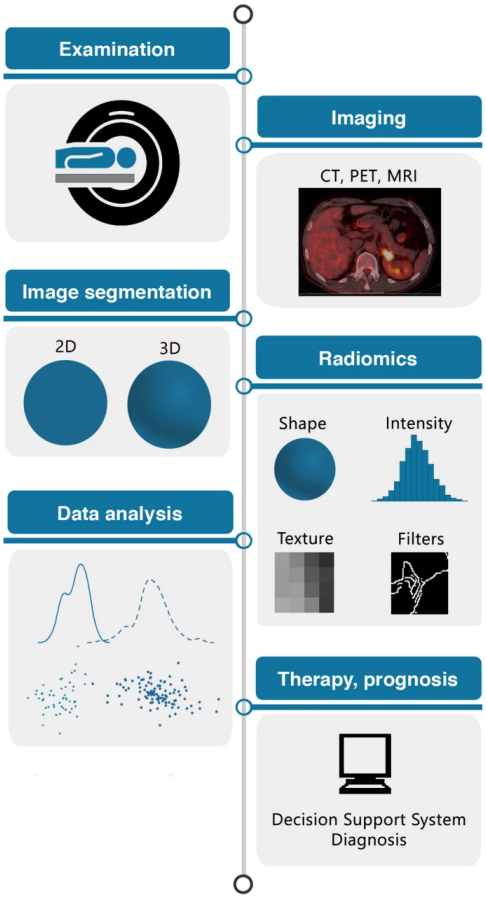


Figure 2.1: The radiomics workflow. Schematic illustration of the patient journey including image acquisition, analysis utilizing radiomics, and derived patient-specific therapy and prognosis [79].

terization of molecules such as DNA (genomics), RNA (transcriptomics), proteins (proteomics), and metabolites (metabolomics), while nowadays it refers to all the disciplines which are capable of producing high-dimensional data from single object or samples, not only biological molecules. Currently, many efforts aim towards radiomics analysis implementations and, as it is recognised as such a promising subject, it is exposed to constant evolution and research. [49]

As represented in Fig. 2.1 the general radiomics workflow follows a specific structure made up of distinct steps. The patients are thoroughly examined and, for further analysis, they are requested to undergo specific imaging techniques. Subsequently, the regions of interest related to the suspected disease are meticulously segmented from the obtained images. After the acquisition and segmentation of images, the next crucial phase involves

the extraction of radiomic features. These features encompass various types, providing a rich and diverse set of information for subsequent analysis. Following this, advanced statistical modelling, integrating ML techniques, comes into play. This sophisticated approach is applied to essential tasks such as disease classification, patient clustering, and individual risk stratification [79, 49].

In order to make the radiomics workflow a standard practice to be implemented diffusely by placing it side by side to a medical expert within the patient treatment, it has to satisfy some further requirements that are not compulsory. It requires standard and adequate implementation of radiomics analyses into the clinical workflow and most of all clinical utility has to be proven in prospective clinical trials [52, 79]. Once these issues are addressed, the radiomics workflow will be able to fully realize the great potential that the research field currently sees in it.

2.1.1 Medical imaging techniques

The initial step in a radiomics workflow involves acquiring images through medical imaging techniques. Among the various imaging methods available, Computed Tomography (CT) and Positron Emission Tomography (PET) are essential for this particular case study.

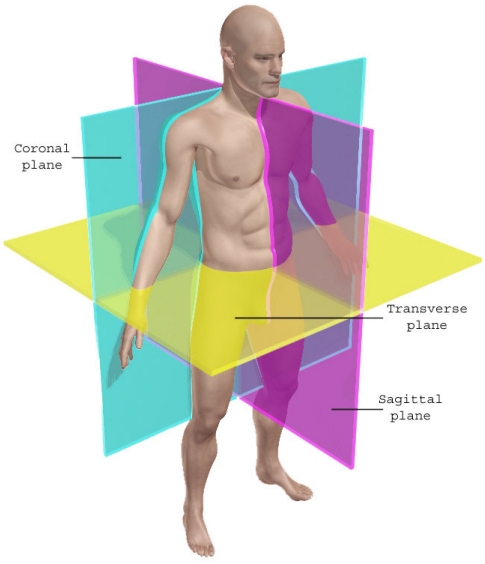


Figure 2.2: The figure represents the three medical imaging planes: transversal plane, coronal plane and sagittal plane. They define the orientation and establish a reference system for the generated image.

These two types of tomography produce a 3D digital image of the patient's spatial domain, encompassing crucial information that medical experts aim to analyse within the designated Volume of Interest (VoI). Both CT and PET technologies capture data distributed across *voxels*, the smallest units capable of storing information within the device. The cumulative count of voxels along each axis determines the image *resolution*. Every tomography image has an orientation which is given by the medical imaging planes, sagittal, coronal and transversal plane. The planes clearly represented in Fig. 2.2 are essential in order to give the image an orientation.

Computed Tomography (CT)

With the development of modern computer technology in the 1960s Computed Tomography (CT) gained popularity in the research field as people recognized its potential and applicability. The first ideas on the cross-sectional imaging, foundational to CT can be traced back to the first half of the century when Johann Radon formulated the "Radon theorem" that states that the image reconstruction is possible from projections: an image distribution function could be obtained from an infinite set of its projections acquired by the rotational scanning. Although the first experiments of this type of reconstructive tomography on medical applications were carried out by the physicist A M Cormack between 1957 and 1963 [42]. In Fig. 2.3 is shown the first CT image of the brain acquired in 1971 alongside a CT brain image acquired with recent devices, indicating the significant advances in CT imaging techniques.

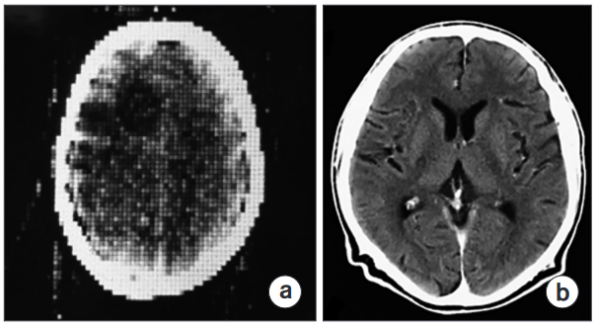


Figure 2.3: First clinical CT image (80×80 image matrix obtained from the Atkinson Morley Hospital, London, UK) of the brain acquired in 1971 (a). Brain CT image acquired recently (512×512 image matrix, Siemens SOMATOM Plus 4 [Siemens medical systems, Erlangen, Germany]) with a more advanced CT scanner (b). [41]

Over the years, the widespread adoption of computed tomography has revolutionized medical diagnostics offering unparalleled insights into the human body, leading also to a continuous development of the technology as we can see, for instance, in Fig 2.4 for what concerns the characteristics of the exploited X-ray beam.

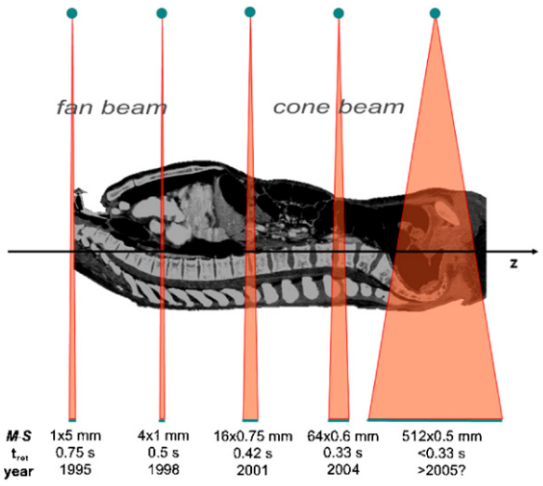


Figure 2.4: The image represents from left to right the development from fan-beam to cone-beam CT termed the 'slice race' in the early 2000s. Currently it seems to come to a halt [42].

CT scanners exploit X-ray radiation beams, and the typical energy used in general CT is in the range of 100 to 150kV. The X-rays propagate through the sample, and some of the photons are absorbed by the medium while the others are transmitted to the detector placed on the other side. The absorption law of the X-ray beam in a medium is a decreasing exponential:

$$I = I_0 e^{-\mu x} \tag{2.1}$$

where I is the intensity of the X-ray beam reaching the detector, I_0 is the beam output intensity, e is the Euler's number, x is the thickness of the sample and μ is the linear attenuation coefficient. If the beam passes through multiple objects with total thickness x as shown in the diagram of Fig 2.5, the resulting intensity acquired by the detector can be expressed by the following equation also called Beer-Lambert law:

$$I = I_0 e^{-\mu_0 x} e^{-\mu_1 x} e^{-\mu_2 x} e^{-\mu_3 x} = I_0 e^{-(\mu_0+\mu_1+\mu_2+\mu_3)x} \tag{2.2}$$

where $\mu = \mu_0 + \mu_1 + \mu_2 + \mu_3$ is the total linear attenuation coefficient, which is the sum of the attenuation coefficients of the single objects. Therefore, the intensity acquired by the detector is proportional to the two-dimensional (2D) transparency of the object, and it is exactly the sum of the linear attenuation coefficient along the path.

The final goal is to have an image representing the complete reconstruction of the sample, and in order to do so the acquisition of projections from many angles are needed.

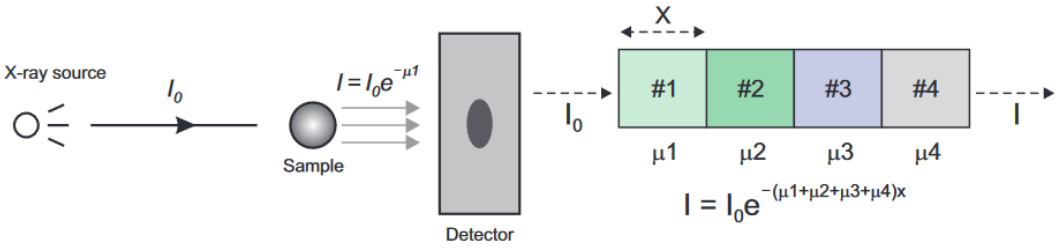


Figure 2.5: X-ray beam attenuation passing through of an object (left) and intensity of an X-ray beam passing through an object with multiple different linear attenuation coefficients (μ_0, μ_1, μ_2, μ_3) (right). [41]

Nowadays, the X-ray source moves along the sample and rotates of 360° (or 180°) around it, and each acquisition is sent to the Data Acquisition system (DAS) in order to form the tomographic image. As described through an historical overview by Willemink [82], over the years many algorithms for the CT image reconstruction have been developed in light of Radon's theorem definition, from filtered back projection to artificial intelligence methods [4, 34].

What a CT scanner device reconstructs is a digital 3D image, divided in a certain number of horizontal square matrices called slices. Each slice is composed of *voxels*, with specific resolution and slice thickness depending on the CT device. These voxel values represent the average linear attenuation coefficients of the internal tissue corresponding to the spatial locations of these pixels. The numerical pixel matrix, associated with a spatial location in the image, is then transformed into an image by assigning/converting the voxel values to a corresponding greyscale value.
In order to have a standard scale of values and allow the comparison between CT scans performed in different conditions, the Hounsfield units (HU) has been selected. With this type of conversion, the CT numbers are defined as the attenuation values of the imaged tissues normalized to that of water [31].

Positron Emission Tomograph (PET)

Positron Emission Tomography (PET) is a functional imaging technique belonging to the field of nuclear medicine. It is a unique imaging modality with proven clinical value in the management of cancer patients, able to show tumor physiology as well as anatomy. The type of PET this work is focusing on is the F-FDG-PET which exploits the tracer F-FDG with F-18 as a positron emitting label, as it is the most frequently selected for tumor imaging. F-FDG stands for Fluorodeoxyglucose, it is similar to glucose, thus it accumulates in high glucose metabolism areas, such as cancer cells [72]. As shown in Tab. 2.1 Fluorine-18 half-life is approximately 110 minutes, therefore the patient undergoes

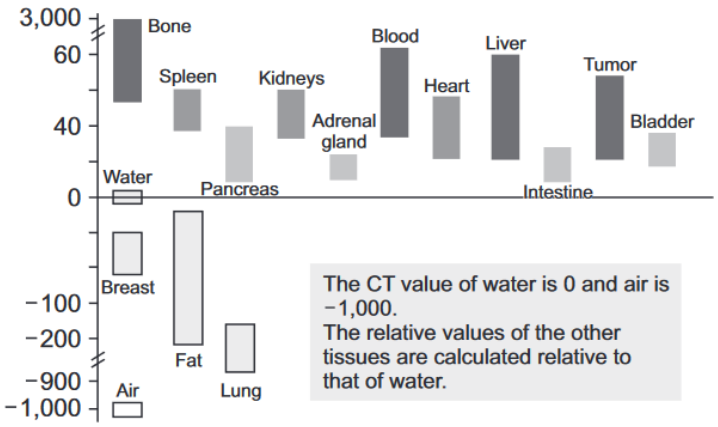


Figure 2.6: Hounsfield scale of computed tomography (CT) numbers for various tissues. [41]

the PET scan when the tracer activity is at its peak.

Radionuclide	Half-life (min)	Average Positron Energy (MeV)
C-11	20.4	0.39
N-13	10	0.50
O-15	2.2	0.72
F-18	110	0.25
Cu-62	9.2	1.3
Ga-68	68.3	0.83
Rb-82	1.25	1.5

Table 2.1: Table reporting half-life and average positron energy for different radionuclides.

When the device is ready and the radiopharmaceutical is at its emission peak, the scan starts. F-FDG emitted positrons, positively charged particles, travel through the surrounding tissues until they encounter an electron. The F-18 positrons emitted have a maximum range (path length between emission and annihilation) in water of 2.4 mm and a mean range in water of 0.6 mm [72]. Consequently, given that the human body has a density similar to that of water, this considered values remain consistent. The positron as soon as it reaches an electron goes through an annihilation, a process where the mass of the positron and the electron are converted into energy also known as positive beta decay [22]. Energy is emitted in the form of two 511 keV photons, moving away from

each other in opposite directions, perfectly aligned at 180 degrees as shown in Fig. 2.7.

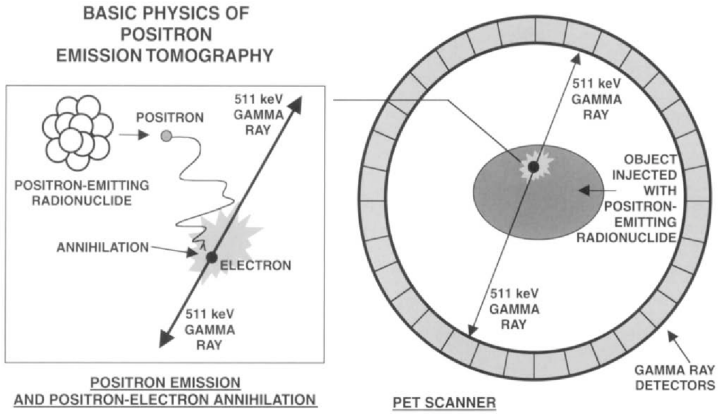


Figure 2.7: The image represents a scheme of the PET scan. On the left is shown the tracer positron emission and the following annihilation resulting in the creation of two gamma photons, on the right the acquisition of the two gamma photons by the circular detector put around the sample.

Special detectors, known as detector rings, surround the patient to capture emitted gamma rays. The distinctive emission direction of the radiation allows the detector to pinpoint the location of the annihilation event. This capability enables the creation of a visual representation of the radiotracer's distribution within the body [56, 72].

Modern detectors can distinguish between two instances occurring extremely close in time. If two gamma photons are detected within a very short period, they are considered as originating from the same event along an imaginary line called the Line Of Response (LOR) between the two detection points. To precisely locate the event along the LOR, the coincidence time window (the time difference between the two detections) is evaluated. Since photons travel at the speed of light, this information allows for the exact determination of the event's position.

In PET scans as well as in CT scans, the photons generated while passing through the sample may be absorbed by the medium, following the Beer Lambert law (Eq. 2.2). To account for this phenomenon, a correction must be applied. Typically, in a CT scan, the linear attenuation coefficient function along the sample is evaluated and utilized to apply a correction to the PET detection. The so called Attenuation Correction (AC) is essential for correctly quantifying the metabolic activity of all the regions of the image.

Although the PET nature leaves a non-negligible variability between patients, which is the reaction of the patient to the radiopharmaceutical. Each individual organism re-

acts differently to the same dose of radiotracer, and it may differentiate the results of the imaging exam. With the aim to standardize the outcome of PET imaging, the Standard Uptake Value (SUV) has been introduced, and the outcomes of every imaging scan are all referenced to that unit of measurement. The SUV depends on three parameters, the radioactivity measured from the PET image, the body weight of the patient and the radiopharmaceutical dose injected into the patient. The equation for its evaluation is as follows:

$$SUV(t) = \frac{A_{img}(t)}{ID/PV} = \frac{A_{img}(t) \cdot PV}{ID} \tag{2.3}$$

with $A_{img}(t)$ being the radioactivity (expressed as volume concentration, e.g. mega-Becquerels per millilitre MBq/mL) measured from the PET image at a time $t > 0$ where $t = 0$ is the time when the radiotracer is injected into the patient. ID is the dose injected at time $t = 0$ and PV is the body volume of the patient evaluated through the patient weight (considering the human body density approximate to the water density 1 g/ml) measured at a time close to the acquisition time. It is clear that SUV depends on time, in fact the radioactivity measure depends strictly on the time interval between the dose administration and the moment when the PET image is acquired, due to the time dependence of the radiotracer activity.

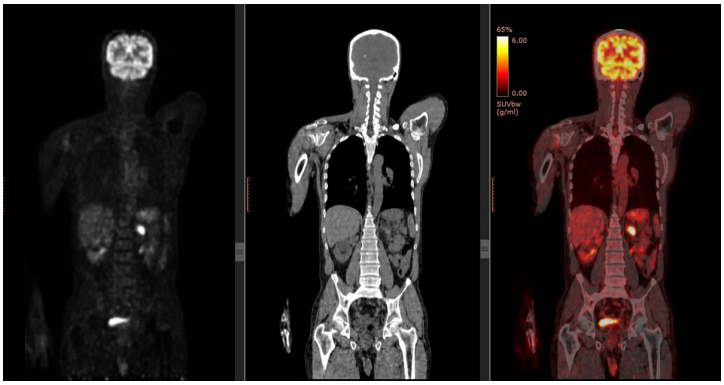


Figure 2.8: The image shows three example scans of a person from the head to half the femur. The first one (left) is a PET scan highlighting metabolic activity along the body, the second one (center) is a CT scan where tissues structure has a great visibility due to the density differences and the last one (right) is the PET/CT combination which combines the features of both. [62]

In conclusion, PET scans can be performed simultaneously to the CT scan, as can be seen in Fig. 2.8. Modern devices are capable of performing the two different imaging techniques on the same patient at the same time, due to the possibility of placing the

two acquisition systems on the same apparatus. This dual imaging approach provides a comprehensive assessment by capturing detailed anatomical information through the CT scan and metabolic activity through the PET scan in a unified procedure, allowing for a more integrated and informative diagnostic evaluation.

2.1.2 Segmentation

Computed Tomography (CT) images play a crucial role in evaluation of lesions and abnormalities in many medical disciplines. It is able to transforms raw biomedical image data into meaningful, spatially structured information, essential ingredients in numerous clinical applications, providing a big contribution for CDSS [1, 21, 33]. While for decades the segmentation of CT images was either manual or semi-automatic, resulting in unpractical procedures and significant time loss, nowadays, there are numerous techniques for automatic segmentation as it is a highly practiced challenge in today's context.

The focus of this work revolves around the segmentation of the skeletal system. Even though it is a much more specific task than general segmentation, it remains a challenge in contemporary applications. In fact, due to the deep variability across datasets, the applicability of the same method is limited and must be supervised and adapted for each case of study. Moreover, the task-specific design and configuration of a method requires high levels of expertise and experience, with small errors leading to large drops in performance [50]. There is a requirement for an out-of-the-box model for automatic segmentation that maintains robustness across diverse datasets, requiring neither expert knowledge nor computing resources beyond standard. Currently, the predominant method to assert itself as the state-of-the-art model for segmentation in this task is the 'nnU-Net'. This approach incorporates a Deep Learning (DL) algorithm founded on the U-Net architecture, complemented by meticulous guidelines for configuring parameters. Despite its prominence, several alternative model proposals are emerging with the aim of surpassing the nnU-Net in performance.

Graph cut algorithm

A graph-based image segmentation approach represents an image as a mathematical graph, with nodes representing image regions or pixels and edges denoting relationships. The segmentation process involves identifying strongly connected groups of nodes, highlighting distinct components in the image. This technique leverages graph theory principles for effective image analysis and segmentation

A graph-based approach to image segmentation, as explained in more complete terms by Peluso et al. [60], involves conceptualizing the image as a mathematical graph. Graph theory, a subset of discrete mathematics, characterizes the relationships between ele-

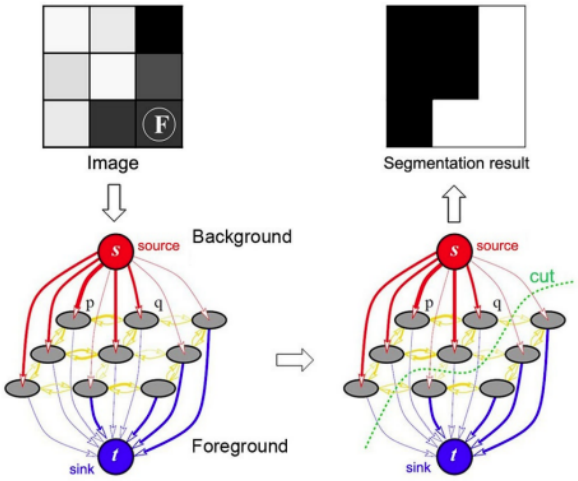


Figure 2.9: Example graph cut segmentation for a simple 3 × 3 image. The image represents in red the source node and in blue the sink node. The weight (cost) of each edge between pixels (yellow links) is reflected by its thickness. The cut is shown by the green dashed line and defines the segmentation of the foreground region from the background image. [84, 60]

ments, termed vertices or nodes, and the connections among them, known as edges or links. In the context of image representation, each pixel becomes a vertex, and edges establish pairwise connections between pixels. These edges carry weights reflecting the affinity or similarity between connected vertices, often defined by pixel features like brightness or position. The segmentation process revolves around identifying optimal cuts in the graph, where a cut partitions vertices into disjoint subsets, guided by a cost function. Image segmentation through this approach entails discovering distinct subgraphs corresponding to different image segments, as illustrated schematically in Fig. 2.9. This technique offers a powerful means of capturing pixel relationships and similarities, contributing to the effective segmentation of diverse images. [60, 7]

nnU-Net

The nnU-Net (short for "no-new-Net") framework is based on a trio of relatively straightforward U-Net models, each incorporating only minor modifications from the original U-Net design. The fundamental contribute of the nnU-Net framework though is brought by the steps around the net architecture: preprocessing (e.g. resampling and normalization), training (e.g. loss, optimizer setting and data augmentation), inference (e.g. patch-based strategy and ensembling across test-time augmentations and models) and a potential post-processing (e.g. enforcing single connected components if applicable).

The automatic or rule-based configuration of these steps allows speeding up the process and provides a standardize framework that leaves little room for inaccuracies [37, 36].

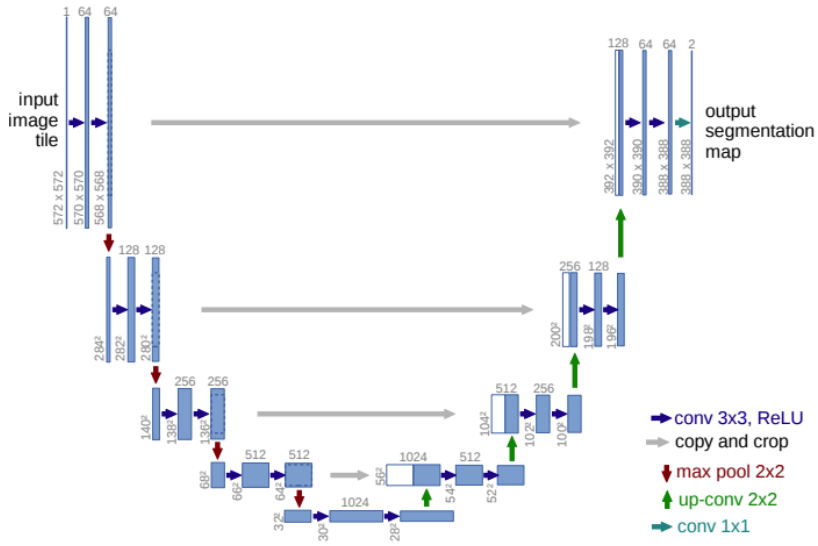


Figure 2.10: U-Net architecture (illustration for a 32×32 pixel representation in the lowest resolution). Each blue box represents a multichannel feature map, with the number of channels indicated on top. The x-y size is provided in the lower-left corner of the box. White boxes denote duplicated feature maps, while arrows indicate different operations. Very relevant are the skip connection (gray arrows) characteristics of this architecture. [70]

The U-Net architecture, as introduced by Ronneberger et al. [70], shares similarities with a Convolutional Neural Networks (CNN) in its encoder-decoder structure. The "U" in U-Net, in fact, is intended to mimic the network's structure, as illustrated in Fig. 2.10. It is made of an encoder that successively aggregates semantic information progressively losing spatial information, and a decoder which receives semantic information from the bottom of the "U" and recombines it in higher resolution feature maps with the spatial information that it gets from the encoder through skip connections. It is crucial to have both kind of information for the success of the network. The nnU-Net advances three fundamental U-Net architectures, 2D U-Net, 3D U-Net, and U-Net Cascade, incorporating slight modifications as detailed by Isensee et al. [37]. The strength of the network, though, lies in designing an automatic training pipeline for these models.

The idea under the automatizing of the configuration process is to have a set of rules for the parameter setting in order to systematize the pipeline and drastically reduce the

search space for empirical design choices. The parameters are divided in three categories: fixed, rule-based and empirical parameters as represented in Fig. 2.11.

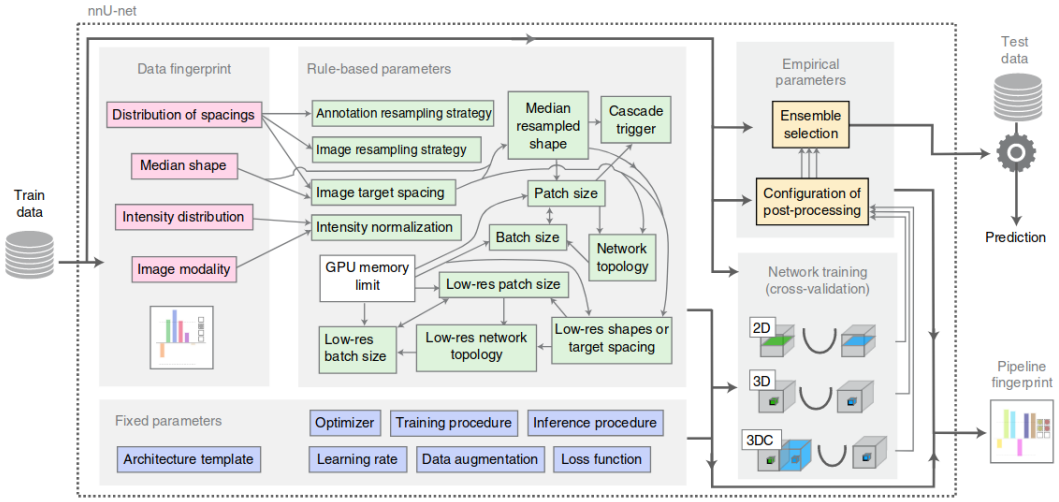


Figure 2.11: The scheme shown in the image represent the pipeline for the parameter setting: the fixed parameters (bottom left in blue), rule-based parameters (top left in green and pink) and empirical parameters (right). [36]

The fixed parameters' selection come from a gather design decisions that remain consistent across datasets, requiring thus no adaptation, and identify a resilient common configuration. The rule-based parameters, whose setting is dictated by heuristic rules allowing an almost-instant adaptation on application, can be of two types: "dataset fingerprint" parameters dependent on dataset properties, and "pipeline fingerprint" dependent on design choices. The lasting parameters to be set are left to an empirical choice dependent on the data available and the case of study [36].

This exposition convincingly establishes the model as the state of the art in medical segmentation, showcasing its proven robustness across various datasets in numerous medical segmentation challenges [37, 36]. Nevertheless, recently the MOOSE model has been proposed as an open source model based on the nnU-Net framework which aims to improvements by introducing a dynamic, data-centric approach, continuously enhancing performance with evolving datasets for improved adaptability in medical segmentation.

MOOSE

With the recent advent of the PET/CT systems with a large axial field of view [44, 74] arose the opportunity to acquire Whole Body (WB) images with the same bed position, facilitating the comparison study between anatomical structure and metabolic informa-

tion. The dual information highly simplifies the multi-organ system analysis, however the large quantity of data generated makes almost impossible the analysis through a non-automated processing pipeline. The MOOSE model represents a cutting-edge approach within the realm of medical image segmentation, building upon the foundation of the nnU-Net framework. It is able to work with both CT images and PET/CT combined images, producing in output a segmentation well differentiated between anatomical structures such as organs, bones and other types of tissues. [69]

Employing a data-centric methodology, MOOSE introduces a novel paradigm where the neural network architecture remains fixed, and model performance is dynamically enhanced through iterative data augmentation. The model demonstrates robust segmentation capabilities across diverse datasets, encompassing 83 cerebral regions and 37 non-cerebral structures, including abdominal organs and bone segments. Continuous monitoring of model performance allows for adaptive retraining when encountering new data deviating from the original training set. [77] As represented in Fig. 2.12, MOOSE

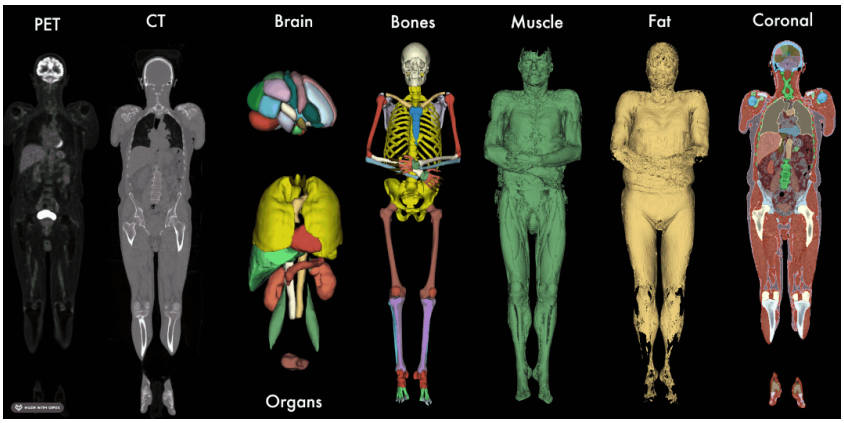


Figure 2.12: In this image is shown an example of MOOSE image segmentation. From left to right can be seen the raw PET and CT scans, segmentation masks with different targets and in the far right a full segmented slice of the coronal projection. [64]

software is able to produce various types of segmentations through a set of different models. Each model has his own presets of parameters suited for a particular kind of segmentation desired. The model already present in the software are:

- *clin-ct-lungs*: model segmenting lungs from CT images,
- *clin-ct-organs*: model segmenting organs in general from CT images,
- *clin-ct-body*: model segmenting the whole body from CT images,
- *clin-ct-ribs*: model segmenting ribs from CT images,

- *clin-ct-muscles*: model segmenting muscles from CT images,
- *clin-ct-peripheral-bones*: model segmenting peripheral bones from CT images,
- *clin-ct-fat*: model segmenting fat from CT images,
- *clin-ct-vertebrae*: model segmenting vertebrae from CT images,
- *clin-ct-cardiac*: model segmenting cardiac tissues from CT images,
- *clin-ct-digestive*: model segmenting digestive organs and tissues from CT images,
- *clin-ct-all-bones-v1*: model segmenting whole skeleton from CT images,
- *clin-pt-fdg-brain-v1*: model segmenting the brain from PET images,
- *preclin-ct-legs*: model segmenting legs from CT images in preclinical stage,
- *preclin-mr-all*: model segmenting the whole body from Magnetic Resonance Imaging (MRI) images in preclinical stage.

Noteworthy is MOOSE's ability to detect and correct segmentation errors in out-of-distribution data, showcasing its adaptability to varying clinical scenarios. The data-centric philosophy, coupled with the nnU-Net architecture, positions MOOSE as a versatile and efficient tool for automated multi-organ segmentation in medical imaging studies.

2.1.3 Feature extraction

Feature extraction is the heart of radiomics pipeline. In this phase, the image and the segmentation are transformed in higher-dimensional data describing attributes of the region of interest. An instrumental contribution in this context was made by Gillies et al. [26] in their comprehensive exploration of the topic. They have argued that features can be divided into two types: *semantic* and *agnostic*. Semantic features are the one describing region of interest, typically they are very informative and easy to interpret by radiologists as they describe lesions and has a very high and direct prognostic value. On the other hand there are agnostic features, they are feature extracted from the image through mathematical procedures on the image numerical properties and thus have a quantitative nature hard to interpret directly. Agnostic features, in turn, can be divided into first-, second-, and higher-order statistical output. The first-order statistical features focus on the single pixel/voxel values, overlooking the relationships with the neighbor ones and the spatial information. They are typically features which tend to describe the region of interest in single numbers representing statistical properties or statistics over

the gray values occurrences. The second-order features are the *texture* features, parameters describing the complex relationship that can be established between pixels/voxels, relationships between neighbors, bigger areas or much more complex structures that an elaborate mathematical algorithm can probe. The higher-order features aim to extract repetitive or non-repetitive patterns from an image after imposing on it filter grids. The nature of this kind of features strictly depends on the type of grid to be applied on the image. For example through fractal analysis it is possible to quantify grid elements with specific voxel values, or evaluating voxel patterns above a threshold Minkowski functional, wavelet transforms extract detailed information across varying frequency scales, and Laplacian transforms of Gaussian band-pass filters for identifying coarser texture patterns in the image.

In this work two of the named filters are employed: wavelet filter and Laplacian of Gaussian (LoG) filter. Wavelet filters operate on the principles of wavelet transformation, a mathematical tool that decomposes signals into different frequency components. Wavelet transform, also called Discrete Wavelet Transform (DWT) [17] is defined by the following equation:

$$X_W(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} \overline{W\left(\frac{t-b}{a}\right)} x(t)\, dt \tag{2.4}$$

where X_W is the transformed function, $x(t)$ the original function, $W(t)$ is the wavelet function shifted in time by a translation parameter b and a dilation parameter a. The discrete form of the wavelet transform is based upon the discretization of parameters (a, b) on the timescale plane [63]. The wavelet filter essentially facilitates this decomposition by convolving the input signal with the chosen wavelet function, capturing variations at different scales. This enables the extraction of relevant features and details from the signal, making it particularly effective for analyzing signals with non-stationary characteristics. [80] The LoG on the other hand is a filter employing morphological operations, involving the mathematical convolution addressed in paragraph 3.4.2. The process involves two operations: convolution with a Gaussian kernel, and then a convolution with a Laplacian filter. The last one is the one extracting interesting information from the image. It highlights regions of rapid intensity change and is therefore often used for edge detection. Being it very sensitive to sudden intensity increase or decrease, it is also particularly sensitive to noise with its characteristic rapid fluctuations. Therefore, the Gaussian kernel comes to use, decreasing noise smoothing a region within the range of a factor σ defining the spread or width of the Gaussian kernel.

$$gaussian = \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \tag{2.5}$$

$$laplacian = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 1 & 0 \\ 1 & 2 & -16 & 2 & 1 \\ 0 & 1 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{2.6}$$

In Eq. 2.5 and Eq. 2.6 are represented respectively an example of Gaussian kernel and an example of Laplacian 5x5 two-dimensional kernels.

2.1.4 Survival analysis

Survival analysis encompasses a range of statistical methodologies designed to examine the duration until a specific event of interest occurs. Widely applied in medical research, it proves particularly useful for assessing the influence of distinct clinical factors on a patient's survival. In diseases such as cancer, where clinical data often include essential *time-to-event* information, survival analysis becomes a pivotal tool. *Time-to-event* data consists of two crucial information: the *event* data, the occurrence of the jump from a status to the other also called *end-point*, and the *time* data, the time from a well-defined *time origin* until the occurrence of the event. Another main concept in this kind of study for what concerns the available data is the *censoring*. The data are subjected to censoring if the total interval length is not known exactly. There are two types of censoring:

- *Right censoring*: given entry at time $t = 0$ and observation at time t, the complete time interval has a length of $T > t$ meaning the event had not occurred at the moment of the observation, and thus the total time interval is unknown.
- *Left censoring*: given entry at time $t = 0$ and observation at time t, the event happens between those time points and thus the information about the exact time-point of the event is not known.

In order to describe this type of data, two functions are used, survivor function $S(t)$ and hazard function $h(t)$, both depending on time. The first one represents the probability over time of an individual to survive from the time origin to a specified future time t, the second one on the other hand is the probability for that individual of having the event at the specific time t.

The first analysis that can be applied on this kind of survival data is the Kaplan-Meier survival estimate [27]. It is a non-parametric method for estimating the survival probability in a given length of time while considering time in many small intervals. For this preliminary analysis we assume that patients who are censored have the same survival prospects as those who continue to be followed, regardless also of the recruitment date, and it is assumed that the event happens at the moment specified by the time interval,

neglecting the left-censoring. Given this conditions, the survival probability $S(t_i)$ at a time t_i knowing that $t_0 = 0$ and $S(0) = 1$ can be evaluated with the following formula:

$$S(t_i) = S(t_{i-1})(1 - \frac{d_i}{n_i}) \tag{2.7}$$

where $S(t_{i-1})$ is the probability of being alive at t_i, n_i is the number of patients alive at t_{i-1}, d_i is the number of events at t_i. The resulting survival curve is a step function characterized by a constant value within the same event time window.

It is possible to compare survival curves between two distinct groups of subjects, such as those receiving standard therapy versus a newer treatment. An example of comparison between survival curves is shown in Fig. 2.13. By examining these curves, gaps

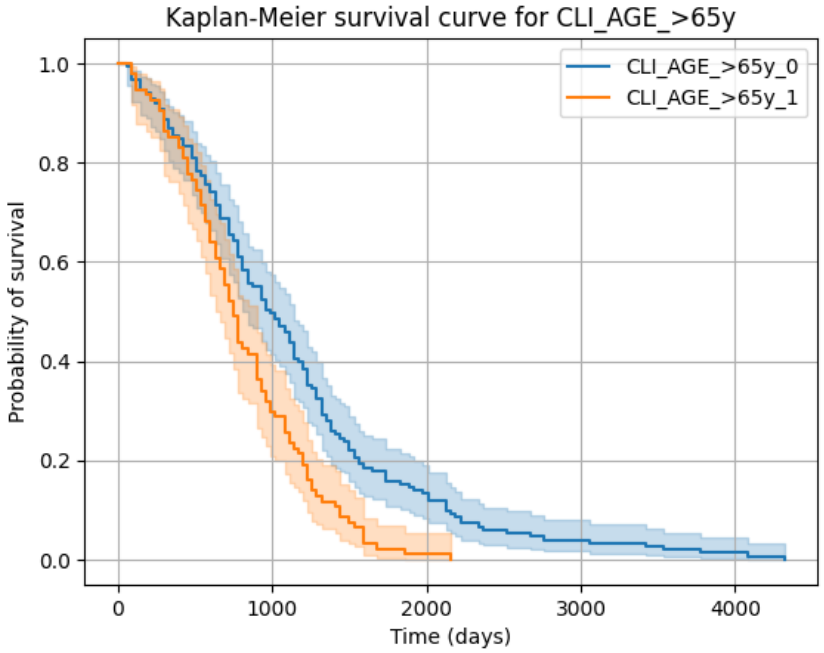


Figure 2.13: This graph represents two Keplan-Meier survival curves. In blue is represented the curve for patients who were diagnosed with MM while they were younger than 65 years old, the orange curve represents people who were diagnosed with MM while being older than 65 years old. The

can be identified, either horizontally or vertically. A vertical gap suggests that, at a specific time point, one group had a higher proportion of surviving subjects, while a horizontal gap indicates that it took one group longer to reach a particular fraction of deaths. The log-rank test emerges as the most widely employed method for comparing

two or more survival curves. Operating under the null hypothesis that there is no discernible difference in survival rates between the compared groups, this non-parametric test avoids assumptions about underlying survival distributions. Essentially, the log-rank test evaluates the observed number of events in each group against the expected count, assuming the null hypothesis holds true, meaning the survival curves are identical. The resulting log-rank statistic is then approximately distributed as a chi-square test statistic.

A single Kaplan-Meier curve exemplifies univariate analysis, portraying survival according to a single factor without considering others. Analysing multiple Kaplan-Meier curves and conducting tests like the log-rank test enables comparisons between factors; however, it is important to note that the predictor variable needs to be a categorical variable. The primary model for handling datasets with numerous, including continuous, variables in survival analysis, allowing for comprehensive multivariate analysis, is the Cox Proportional-Hazards Model [27].

Cox proportional-hazards model

Cox proportional-hazards model, sometimes simply referred to as the *Cox model*, is currently the most popular regression model for the analysis of time-to-event data. It stands out because it is able to estimate the relationship between the hazard rate and explanatory variables, the feature extracted when the model is inserted in a radiomics pipeline, without having to make any assumptions about the shape of the baseline hazard function. Furthermore, the model can compare hazard functions between patients, using their corresponding features to evaluate the curves, to arrange them with an order based on the hazard and eventually elect a priority between different patient conditions.

Working in continuous time and supposing a random number of spells, some are censored and some are complete, the Proportional Hazard (PH) function can be written with the following formula:

$$h(t|\mathbf{X}_i) = h_0(t)e^{\mathbf{X}_i\beta} \tag{2.8}$$

Here, $h_0(t)$ represents the baseline hazard, t signifies the survival time, $\mathbf{X}_i$ is the vector of predictor variables, also known as covariates, and $\beta = \{\beta_1, ...\beta_N\}$ is the parameter vector of the model. Each term $e^{\beta_j X_j}$ is referred to as the Hazard Ratio (HR), indicating the correlation between the respective covariate and the PH. A $\beta_j > 0$ implies that with increasing values of the ith covariate, the event hazard rises, consequently reducing the survival time. The overall exponential of the equation, known as the partial hazard, is a non-negative term independent of time. The Cox model's regression nature stems from this function, which can be expressed as a multiple linear regression by applying a logarithmic function to the hazard function. [25, 46]

The name of the Cox proportional-hazards model comes from an important assumption of the model when evaluating the HR for two subjects with covariate vectors $\mathbf{X}_i$ and $\mathbf{X}_j$:

$$\frac{h_i(t|\mathbf{x_i})}{h_j(t|\mathbf{x_j})} = \frac{h_0(t)e^{\mathbf{X}_i\beta}}{h_0(t)e^{\mathbf{X}_j\beta}} = e^{\beta(\mathbf{X}_i - \mathbf{X}_j)} \tag{2.9}$$

as can be seen in the equation the HR is independent of the time [85].

During the training phase of the Cox model, the algorithm learns from a labelled dataset, adjusting its parameters to optimize its ability to predict the time to an event of interest. This involves utilizing survival data, where each observation is associated with a specific time of event occurrence or censoring. The model undergoes iterative training processes to minimize the loss function, enhancing its predictive capabilities. Once the training is complete, the model enters the testing phase. In this stage, it is evaluated on a separate dataset that it has not encountered during training. The testing dataset helps assess the model's generalization performance, providing insights into how well it can predict survival outcomes on new, unseen data. Evaluation metrics such as C-index or log-rank test may be employed to measure the model's accuracy and effectiveness in predicting survival times. This comprehensive training and testing process ensures the Cox model's robustness and reliability in predicting survival outcomes in various scenarios.

The gold standard metric in survival analysis is the concordance index (C-index). It is a statistical metric employed in survival analysis, specifically for evaluating the predictive accuracy of a survival model. In the context of censored data, where some individuals do not experience the event of interest within the study period, the C-index proves to be a robust measure [9, 30]. Essentially, the C-index assesses the concordance between predicted and observed survival times for pairs of individuals. A higher C-index indicates better discriminatory power, implying that the model is proficient in correctly ranking the survival times of subjects. In the presence of censored data, the C-index considers the pairs in which at least one subject has an event, allowing for an unbiased estimation of concordance while accounting for censoring. The calculation involves counting concordant and discordant pairs, where concordance pertains to correctly ordered survival times. This index, thus, provides a comprehensive and reliable measure of the model's predictive performance in survival analysis, taking into account censored data. Suppose there are n patients, each with covariates and their clinical endpoint time-to-event. A mathematical formulation for the C-index can be explicitly derived. The observed time-to-event (T) is handled as previously explained, accounting for the time until the event occurrence or the last known time for uncensored patients. To encapsulate this information in a variable, a δ variable is introduced, representing event occurrence: $\delta = 1$ if the patient experienced the event and $\delta = 0$ if the patient's time-to-event is censored. The

following formula shows the pairwise evaluation of the C-index:

$$c - index = \frac{\sum_{i \neq j} I(h_i < h_j) I(T_i > T_j) \delta_j}{\sum_{i \neq j} I(T_i > T_j) \delta_j} \tag{2.10}$$

where $I(x)$ is a function yielding 1 if x is true or 0 otherwise, i and j represent the $i-th$ and $j-th$ patient, T_i and T_j respectively the time-to event, and h_i and h_j the associated predicted hazard which is expected to be larger for shorter time-to-event predictions. [30]

Penalization

The model is suited to deal with multivariate survival analysis but in case the covariates set is extremely larger than the input set the regression encounter difficulties in convergence. To address this challenge, the Least Absolute Shrinkage and Selection Operator (LASSO) regularization, or L1 regularization, comes into play. This technique introduces a penalty term based on the absolute values of the coefficients, facilitating variable selection and promoting sparsity in the model. The penalty term's role, called penalizer and represented with λ, is reducing to zero the regression coefficient for the covariates for which contribute is less relevant inside the model. The penalty term, also referred to as L1, is in fact the L1-norm of the regression coefficients β_j as can be seen in the following equation:

$$L1(\beta) = \lambda ||\beta||_1 = \lambda \sum_{j=1}^{n} |\beta_j| \tag{2.11}$$

the variable n denotes the count of covariates, and the penalty term manifests as the sum of absolute coefficients in the regression. Another regularization approach, addressing issues related to the dataset's magnitude, is known as Ridge penalization. This technique is analogous to LASSO penalization, but instead of attempting to force non-contributing covariates to zero, it seeks to reduce their impact on the model. Ridge penalization is characterized as L2 penalization due to its evaluation through the L1-norm of the regression coefficients β_j, expressed as follows:

$$L2(\beta) = \lambda ||\beta||_2^2 = \lambda \sum_{j=1}^{n} |\beta_j|^2 \tag{2.12}$$

2.2 Multiple myeloma

MM is a cancer of the Bone Marrow (BM) plasma cells, and it consists of a cytogenetically heterogeneous clonal plasma cell proliferation. As the underlying causes of this

disease are not yet fully understood, it is known that the transformation of these cells into a malignant neoplasm is lead by multistep genetic and microenvironmental changes. It accounts for about 1% of neoplastic diseases and for 13% of haematological cancers [58]. According to the Global Cancer Observatory (GCO) all over the world there are about 170 000 patients suffering from MM each year, with a raising trend due also to the population increase, as it is shown by the graph in Fig. 2.14. While the diagnosis

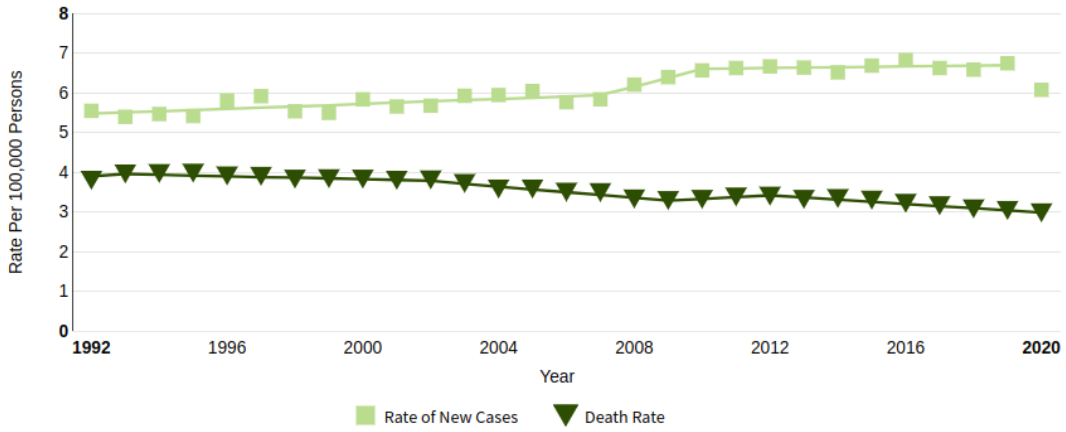
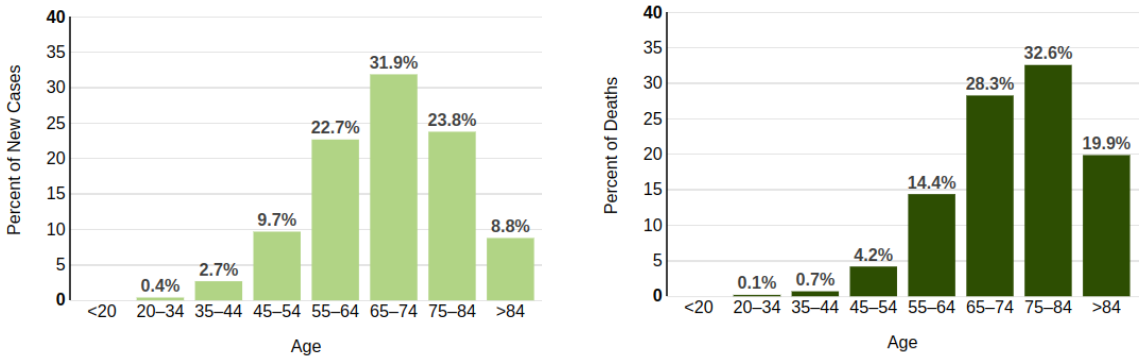


Figure 2.14: This graph provided by the SEER program shows the trend over the years for the incidence of MM and the deaths occurring within the patients suffering from this disease.

occurs at a median age of approximately 70 years, and it is most frequently diagnosed in 65-74-year-olds, there is still a 37% of patients which are younger than 65 years [58]. In Fig. 2.15 can be seen the diagnosis age distribution as well as the death occurrence age distribution according to the Surveillance, Epidemiology, and End Results (SEER) program in the United States. It is clear from the graphs that, even for younger patients, MM is a highly dangerous and life-threatening disease.

MM is a high death rate cancer as it has a 5-Year relative survival of about 59.8%, which is the estimate of the percentage of patients who would be expected to survive the effects of their cancer at least 5 years, according to the SEER. Such a phenomenon is attributed to the nature of the disease, which evolves from a Monoclonal Gammopathy of Undetermined clinical Significance (MGUS) then progresses to smouldering myeloma, all asymptomatic stages, and finally to symptomatic myeloma [67]. MGUS is present in approximately 3–4% of the population over the age of 50 years, making more difficult an early diagnosis of the disease [45, 81].



(a) (b)

Figure 2.15: The image represents two graphs provided by the SEER program. The first one (a) shows the age distribution relative to MM diagnosis, the second one (b) represents the age distribution of the death occurrences within patients suffering from MM.

2.2.1 Diagnosis

The early diagnosis of MM can be crucial for the efficiency of the following treatment, and the difficulties in finding a univocal set of criteria make the process challenging for clinicians still today. The International Myeloma Working Group (IMWG) in a paper published in 2022 [66] suggested a set of revised criteria for the diagnosis, trying to standardize the process. They indicated that patients have to be found with evidence of either 10% or more clonal plasma cells on bone marrow examination or a biopsy-proven plasmacytoma in addition to one or more Myeloma Defining Events (MDE). MDE, a list of symptomatic characteristics that can help the MM diagnostic, comprehend the CRAB features and specific levels of 3 biomarkers. CRAB features are four particular symptoms:

- *Hypercalcemia*: a condition in which the calcium level in blood is above normal.
- *renal failure*: a medical condition in which the kidneys can no longer adequately filter waste products from the blood.
- *Anemia*: a blood disorder in which the blood has a reduced ability to carry oxygen due to a lower than normal number of red blood cells.
- *Lytic bone lesions*: areas where bone has been destroyed, leaving a lesion (hole) in it.

The three specific biomarkers on the other hand are represented by a clonal bone marrow plasma cells level higher by 60% of the normal level, a 100% higher serum Free Light Chain (FLC) ratio and more than one focal lesion captured through MRI. The revised

criteria signify a paradigm shift by enabling early diagnosis and the initiation of therapy prior to the occurrence of end-organ damage.

Stage	Criteria ($\beta 2M$ = Serum $\beta 2$ microglobulin; ALB = serum albumin)
I	$S\beta 2M < 3.5$ mg/L; serum albumin ≥ 3.5 g/dL
II	$S\beta 2M < 3.5$ mg/L; serum albumin < 3.5 g/dL; or $\beta 2M$ 3.5-5.5 mg/L, irrespective of serum albumin)
III	$S\beta 2M > 5.5$ mg/L

Table 2.2: International Staging System (ISS) represented with the three different stages next to the criteria to declare the disease belonging to that stage. [35]

2.2.2 Prognosis

Another important phase for MM disease is the prognostic analysis. It involves the evaluation of the risk stratification to acquire a deeper understanding of the concrete conditions of the patients, the advancing of the disease and be able to propose an effective schedule for organizing the treatment procedure. The IMWG developed in 2005 a criteria to evaluate the stage of the disease called ISS [35] which trough clinical and laboratory data is able to discriminate between three different stages. A combination of $S\beta 2M$ and serum albumin provided the most powerful, simple, and reproducible three-stage classification shown in Tab. 2.2.

In 2015 the IMWG revised the criteria developed with ISS to incorporate two further prognostic factors: genetic risk as assessed by Fluorescence In-Situ Hybridization (FISH), and level of Lactate Dehydrogenase Level (LDH). This new criteria called R-ISS have been updated from years of studies for improving risk stratification and therapeutic decision-making. [66, 29] The summary displayed in Tab 2.3 represents, in comparison with the previous one, the updates developed in the staging system.

2.2.3 Therapeutic approaches

MM poses a significant challenge in the realm of hematology, demanding a multi-faceted therapeutic approach. Traditional approaches, including chemotherapy and immunomodulatory agents, form the foundation of treatment, their mechanisms of action and evolving applications explored in depth. Stem cell transplantation, especially in the

Stage	Criteria ($\beta 2M$ = Serum $\beta 2$ microglobulin; ALB = serum albumin)
I	$S\beta 2M < 3.5$ mg/L; serum albumin ≥ 3.5 g/dL; Standard-risk Chromosomal Abnormalities (CA) by FISH; Normal LDH
II	Not R-ISS stage I or III
III	$S\beta 2M \geq 5.5$ mg/L and either: high-risk CA by FISH OR high LDH

Table 2.3: Revised International Staging System (R-ISS) represented with the three different stages next to the criteria to declare the disease belonging to that stage. [35]

context of autologous stem cell transplantation following high-dose chemotherapy, remains a pivotal consideration. The advent of novel agents, such as proteasome inhibitors and monoclonal antibodies, has significantly expanded the treatment arsenal, warranting comprehensive analysis. [29] As the diagnostic landscape evolves, PET/CT imaging emerges as an indispensable tool, offering precision in disease staging, identification of extramedullary involvement, and assessment of overall disease burden. The novel Italian Myeloma criteria for PET Use (IMPeTUs) criteria, provides a standardized framework for PET/CT response assessment, emphasizing their role in early response evaluation and their implications for personalized treatment strategies.

The IMPeTUs criteria, as elucidated in the work brought by Nanni et al. [55, 54], constitute a set of standardized guidelines for the interpretation of 18F-FDG PET/CT scans in patients diagnosed with MM. Recognized as a valuable imaging tool, PET/CT serves to detect paramedullary (PM) and extramedullary soft tissue masses or solid organ involvement in MM patients, offering a detailed insight into disease presentation. The IMPeTUs criteria are designed to enhance the interpretation of PET/CT images, particularly focusing on the challenging posttreatment phase, where issues like recent fractures, vertebral collapses, or metallic bone implants can pose interpretation challenges. The criteria provide a detailed visual and morphological assessment, incorporating the Deauville Score (DS) [39] for quantifying FDG uptake, evaluation of bone marrow metabolic state, identification of focal lesions, and discerning extramedullary or PM disease. This approach aims to standardize PET readings, addressing the need for consistency, especially in borderline cases.

2.2.4 Radiomics in MM

The recognition and standardization of medical imaging in relation to MM are still evolving. This opens up an exciting potential for gaining valuable insights by applying AI and radiomics pipelines to analyse these images [78]. Each imaging technique is capable of extracting various numerical characteristics and features, but each has its

own drawbacks. The principal imaging techniques used for MM imaging nowadays are CT, MRI, and PET. CT-derived radiomics features may capture structural alterations in the trabecular bone, such as punched-out lytic lesions, expansible lesions with soft tissue masses, diffuse osteopenia, fractures, and, rarely, osteosclerosis. Focal myeloma lesions can be detected on whole-body MRI and they indicate symptomatic multiple myeloma requiring therapy. PET-based radiomics, on the other hand, can evaluate the metabolic activity of MM lesions and assess their response to therapy, although it has a limitation in spatial resolution. PET/CT imaging try to combine the advantages of the two techniques, and it reveals as an efficient method for MM imaging. One of the most significant advantages of PET/CT imaging is its ability to distinguish between active myeloma (FDG positive) and MGUS or smouldering disease [47, 32].

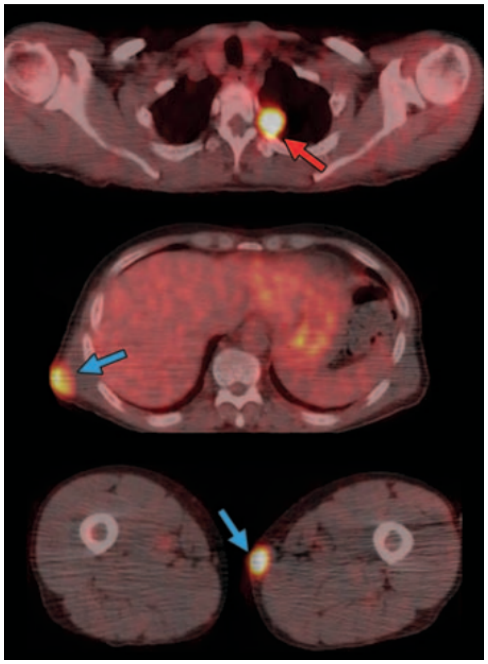


Figure 2.16: MM relapse with cutaneous and pleural plasmacytomas in a 49-year-old man, who initially presented with isolated C5 vertebral involvement and was treated with radiation therapy and systemic chemotherapy. With apparent biochemical and clinical response, the patient was scheduled for allogenic bone marrow transplantation. Restaging axial fused PET/CT images show 18F-FDG avid pleural plasmacytoma (red arrow) and several cutaneous plasmacytomas (blue arrows), compatible with active disease. No FDG-avid lesion was noted at the primary site (C5 vertebra). The patient was subsequently started on systemic chemotherapy. [2]

In the image shown in Fig. 2.16 can be recognised three PET/CT slices of the transversal plane of different anatomical areas: the top image represents the upper thorax, the middle image represents the lower thorax and the bottom image represents the middle part of the femur. Through the arrows are highlighted MM induced lesions: the red arrow points to a PM lesion in proximity of the spine, while the others point to extramedullary lesions located distant from the bone marrow. MM lesions or the region where they are situated can be automatically recognized through image segmentation, and the selected area can be characterized with numerical properties. The contribution of a radiomics pipeline applied to this type of disease is to gain more insights into the progression of the illness. It aims to propose a description through numerical quantities extracted from the medical images, along with other types of information such as clinical data and also physician evaluations using standardized criteria such as R-ISS and IMPeTUs. A statistical analysis with AI methods of all this information can be able to unveil intricate patterns and connections that a clinician may not discern solely by examining the data at their disposal. Therefore, through CDSS the analysis can influence clinicians' decisions over a specific and more adequate procedure for the patient to undergo.

By integrating multi-modal imaging data, radiomics provides a comprehensive perspective on MM lesions, potentially enhancing the accuracy and precision of diagnoses. The future implementation of radiomics with other omics data, including genomics, proteomics, and metabolomics, holds the potential to establish a holistic comprehension of MM. This approach may unveil intricate connections between imaging phenotypes and the progression of molecular mechanisms of the disease.

Chapter 3

Materials and methods

This chapter represents the pipeline of the whole work described in the present document. The first thing described is the data available with the different types of information provided, which make the starting point of the analysis. Then here is pointed out the preprocessing of the data; the preprocessing involves a series of techniques aimed at managing data to prepare it for subsequent analysis procedures. Often, data is provided in various formats, making it heterogeneous and challenging to compare. In conclusion, the chapter focuses on the analysis procedures. All the techniques implemented involving statistical methods, machine learning algorithms and analytical techniques to derive insights, patterns, or trends from the prepared data.

3.1 Data and formats

When working on a radiomics pipeline in medical field, it is important to keep in mind that data could be provided in various formats. The medical images are fundamental, but they are not the only source of information for a radiomics analysis. It is essential to also consider clinical data, as they play a crucial role in a comprehensive analysis. Clinical data provide additional context and valuable insights that complement the information extracted from medical images. They can comprehend blood test or other exams results parameters, such as blood calcium level and creatinine level, and also annotations from the clinicians on significant indicators, based on relevant criteria as the R-ISS and IMPeTUs criteria. The latter are crucial in radiomics pipelines, annotations in general play an important role, as they are the bridge that connects the data analysis and processing to clinician's knowledge.

One cohort of patients is employed in this research, originating from two different sources. The first cohort is composed of a total number of 262 Newly Diagnosed Multiple Myeloma (NDMM) patients collected since 2007 by the Italian Istituto di Ricovero

e Cura a Carattere Scientific (IRCCS) based in Bologna, called Azienda Ospedaliero-Universitaria Sant'Orsola-Malpighi. The patients suffering from MM have been followed by the clinicians in Bologna, where also all the examinations have been performed.

3.1.1 Multiple myeloma cohort

The sample collected by the clinicians of the research institute in Bologna's hospital is composed of a total number of 262 NDMM patients suffering from MM. The privacy of patients is guaranteed through the anonymization of acquired information, this process effectively separates patient identities from the data utilized for analysis. The MM cohort, as usually happens for radiomics studies in the medical field, comprehend two types of data: the medical imaging data and the clinical data. The genomics data, the other type of insights for the cohort, holds significant information, referring to the collection of information about an organism's complete set of genes, known as its genome. It is often useful for the analysis of the disease, although in the present work the clinical data is the only one employed.

The first type of data consists of ^{18}F-fluorodeoxyglucose (^{18}F-FDG) PET/CT digital images acquired with scanners produced by the Italian division of General Electric (GE) Healthcare, GE Medical Systems. An important parameter of the device is the voxel spacing, as it is characteristic of the acquisition system and determines the precision of the image. The dataset available for the present work is distinguished by different voxel spacing values, the distance along different directions of two adjacent voxels centers. The pairs of CT and PET scans from the same patient have consistent dimensions with each other along the z axis, and across the entire dataset this dimension take on two values, 2.78 mm and 3.27 mm. Regarding the other two voxel dimensions along the (x, y) plane, the transverse plane as showed in Fig. 2.2, CT and PET have two different spacings:

- for CT scans the transversal voxel spacing have either a value of 0.98 mm or a value of 1.37 mm in both x and y direction;
- for PET scans the transversal voxel spacing have either a value of 2.73 mm or a value of 5.47 mm in both x and y direction.

As can be observed, the transversal voxel spacing is higher in PET images compared to CT images, which complicates the process of comparing the two images. Moreover, the field of view FoV varies among the devices used, and it is not uniform for all patients. Specifically, a Whole Body (WB) image is obtained only for some patients, while for the remaining individuals, the anatomical region represented in the image differs. Nevertheless, across the entire patient cohort, the scan includes a section of the body extending from the upper part of the skull to the lower part of the femur. A few images with a FoV limited to a small portion of the legs were subsequently excluded in the following

stages of the pipeline for the purposes of this study. The differences in dimensions of the FoV result in a digital images dataset with varying resolution along the z axis. On the other hand, the resolution of each (x, y) plane slice, which is determined by the device specifics and setup, takes on a few discrete values:

- for all the CT scans the slice resolution is 512×512;
- for PET scans the slice resolution have either a value of 256×256 or a smaller value of 128×128.

Despite all the resolution and spacing differences between each pair of CT and PET images they represent the same FoV. In the image format are stored all the information needed to place them in the same spatial domain, as will be discussed later on in this chapter.

While CT images are normalized using the Hounsfield scale, as explained in paragraph 2.2, PET images depend significantly on the type of radiopharmaceutical administered to the patient. In this cohort, all patients received a 3 MBq/kg quantity of ^{18}F-fluorodeoxyglucose (^{18}F-FDG) that is used to evaluate the SUV. The scan was conducted an appropriate time after the administration of the radiopharmaceutical, which, according to the European Association of Nuclear Medicine, is 60 minutes, with an acceptable range of 55–75 minutes [6].

The clinical data on the other hand is categorized into two main components: demographic information and specific clinical data including test results and analyses. In the case of the sample used, clinical data account for 93 of the total number of patients present in the cohort. An overview of the demographic information of this reduced population show that the sample is characterized by 35 female patients and 58 male patients. The mean age at the diagnosis is (64 ± 11) years old, the younger patient to be diagnosed with MM was 23 years old, and the older one was 84. The specific clinical information stored in the clinical data include analysis results such as the creatinine or hemoglobuline value at diagnosis, which can be related to MM as discussed in paragraph 2.2.1. In addition, the provided information encompasses evaluations of the disease by experts capable of assessing lesions and expressing opinions grounded in standardized criteria, akin to the IMPeTUs criteria (Paragraph. 2.2.3), to describe the comprehensive nature of anomalies, whether localized in a specific body region or collectively. Furthermore, the clinical data include information about the progression times of the disease, from the diagnosis and the beginning of the therapy to the end of the monitoring period, or death if it occurs. Among the insights into disease progression, the PFS is of particular interest for this work for its employability in survival analysis statistics. It represents the period of time during which a patient remains without signs of disease progression or deterioration from the initial condition. This information is indicated by whether the

event occurs or not and the time-to-event from the diagnosis.

Lastly, in MM cohort all the CT and PET images are available in the same NIfTI format, which will be presented in the following paragraph (section 3.1.2).

3.1.2 Neuroimaging Informatics Technology Initiative (NIfTI)

The NIfTI file format is a widely utilized standard in medical imaging, particularly well-suited for storing and exchanging neuroimaging data. Developed by the Data Format Working Group (DFWG), the file format architecture of NIfTI files is designed to accommodate both functional and structural imaging data efficiently.

At its core, the NIfTI file typically consists of two parts: a header and an image data array. The header contains essential metadata describing the image, such as voxel dimensions, data type, orientation, and other crucial information for proper interpretation shown in Tab. 5.1. Of the metadata contained in the table the most important for the purposes of this dissertation are the following:

- the *dim[8]*: it is an array that contains the size of the image array. The first element (dim[0]) contains the number of dimensions (1-7). The dimensions 1, 2 and 3 are assumed to refer to space (x, y, z), the 4th dimension is assumed to refer to time, and the remaining dimensions, 5, 6 and 7, can be anything else. The value dim[i] is a positive integer representing the length of the i-th dimension;

- the *pixdim[8]*: in this array is stored the information about the dimensions for each voxel expressed in mm, and each element match its respective in dim[8];

- the *qoffset-x, qoffset-y, qoffset-z*: these three variables represent the displacement in the three spatial directions of the image origin from the standard frame of reference, expressed in mm;

- the *srow-x[4], srow-y[4], srow-z[4]*: they are represented in Eq.. 3.1 as a single matrix with as lines the three arrays, x, y and z respectively. It is the affine matrix which determines the geometric transformations of the images expressed in mm. In Eq. 3.1 are represented the new voxel dimensions in the three axes, dx, dy and dz, and the last column represents the new translation from the origin of the system, tx, ty and tz. The 0s represented in the matrix are 0s for all the images available in the cohorts, however when they are specified with values the matrix.

$$affine = \begin{bmatrix} dx & 0 & 0 & tx \\ 0 & dy & 0 & ty \\ 0 & 0 & dz & tz \end{bmatrix} \tag{3.1}$$

The header is structured as a C structure, ensuring compatibility across different computing environments. The header is followed by the image data array, which stores the actual voxel values. The data array can be stored in a separate file or embedded within the NIfTI file itself. The voxel values are organized in a three-dimensional matrix, corresponding to the spatial dimensions of the image. One notable feature of the NIfTI format is its flexibility in supporting both 3D and 4D (time series) data. This is achieved by allowing multiple 3D datasets to be stored within a single 4D file, facilitating the representation of dynamic processes over time. Furthermore, NIfTI files support different data types, such as integer, floating-point, and complex numbers, ensuring compatibility with various imaging modalities and analysis techniques. The format also accommodates extensions to the standard header to include additional information specific to certain applications.

In conclusion, the NIfTI file architecture is characterized by its simplicity, versatility, and adherence to standards, making it a widely adopted format in the field of medical imaging. Its compatibility with different types of imaging data and support for multidimensional datasets contribute to its effectiveness in facilitating data exchange and analysis within the community. [83, 15]

When a medical image stored in a NIfTI file needs to be modified, it is important to implement changes consistently in both the header and the image array. The changes applied to the image array, such as resolution modifications, image translation, and other types of processing over the grey levels image, must be characterized coherently. It is performed by changing directly the header specifications and the affine matrix which stores the information about geometric transformation of the image, and putting back the three different element into the same NIfTI file.

The cohort made available from TCIA, as discussed previously, comes with DICOM format images. Digital Imaging and Communications in Medicine (DICOM) format [16, 65] is a digital format which stores many more information than a NIfTI file. In fact, it contains also detailed information about the device employed for the image acquisition, the injected doses whether the image is a PET image and other kinds of information. For the purposes of the present work the NIfTI files were easier to handle and manage. Therefore a conversion has been performed with the *dcm2niix* software [71] designed to convert medical imaging data from the from one format to the other. Due to the great amount of information stored in the DICOM file, the conversion produces the desired NIfTI and a JavaScript Object Notation (JSON) file [40]. All the information available in the source file, that the target format cannot contain, are stored in the produced JSON file.

3.2 Architecture of the project

Starting from the data available here will be explained the architecture of the project with all the details about the algorithms and methods employed. The present paragraph aims to describe the structure of the radiomics pipeline adopted.

- *Segmentation*: it is a process where an image is produced, which stores the information about the labeling of the original image.
- *Preprocessing*: it is the ensemble of all the operations that allow to prepare the data for the following analysis and processes. CT and PET images need to be prepared for the following steps of the pipeline, and this involves aligning PET images with the segmentation, producing 3 different segmentation shapes and shaping the original images within those segmentations.
- *Feature extraction*: this process allows to extract, from the images and their segmentation, numerical properties that address among the others shape, grey level distributions and texture image characteristics
- *Survival analysis*: based on the features extracted from the images, survival analysis, performed with the Cox regression model, allows to gain insights about prognostic risk factors and probability against time to experience a relapse in the disease.

The step of the pipeline described here are represented also in the diagram shown in Fig. 3.1.

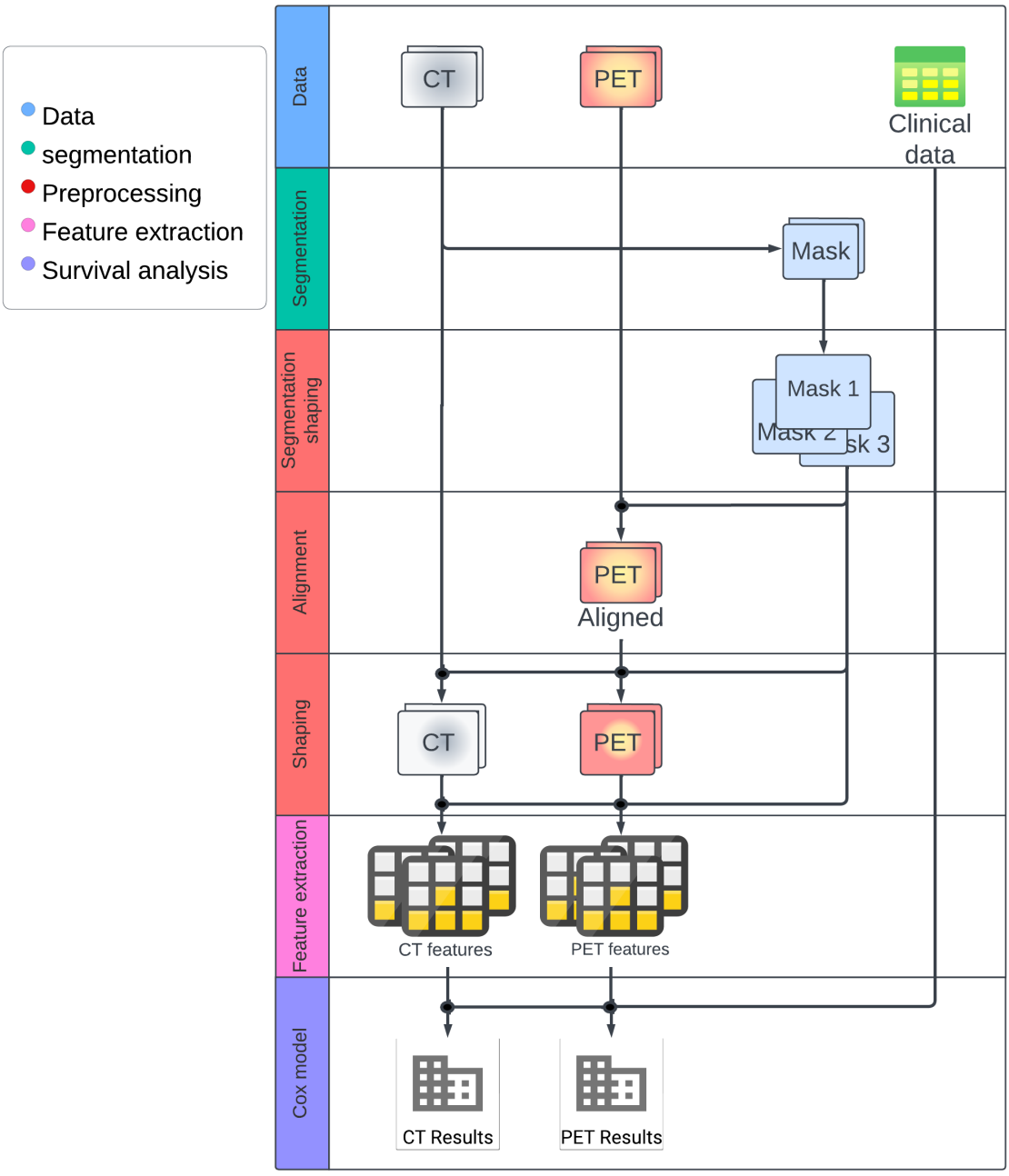


Figure 3.1: Diagram representing the pipeline employed in the present work.

3.3 Segmentation with MOOSE

In the presented pipeline the MOOSE model will be tested, the open source model based on the nnU-Net framework discussed in section 2.1.2. As has been emphasized, it aims to introduce a dynamic, data-centric approach, continuously enhancing performance with evolving datasets for improved adaptability in medical segmentation.

MOOSE software is able to perform accurate multi-organ segmentation both for PET, CT, and MRI images through the different models available in its pipeline. The cohorts used in the present work, as described in details in section 3.1, are composed of couples of CT and PET images covering the same spatial domain but having different resolution. The MOOSE model performing whole skeleton segmentation is the *clin-ct-all-bones-v1* model, and it works with CT images where the body tissues' density is easily distinguishable. The reason of the selection is the high frequency of lesion presented by a MM patients in the spine area and the surroundings, and in general all the bone marrow regions. An example result of a skeleton segmentation is shown in Fig. 3.2.
As can be noted from the figure, the segmentation produced by the model is a segmentation from the CT image of the entire skeleton. To each bone or group of small bones is given a different numerical label which in the figure is represented by the color of the segmentation.

The segmentation mask produced by the model is stored in a NIfTI file. As all the NIfTI files it is made of two parts:

- the *image array* produced is an array with the same resolution as the CT source image. The corresponding pixel that in the CT represents a skeletal tissue, for the segmentation image is filled with the number which represents the label of the bone.
- the *header* produced contains the same information as the CT image, such as the resolution, the dimensions of the voxels and the spatial information in order to place the image in the standard frame of reference.

3.4 Preprocessing

The preprocessing phase is the part of the pipeline where the data available are managed in order to facilitate the following steps. The execution of the operations is highly contingent upon the nature and formats of the data, as well as the subsequent procedures that must be executed.

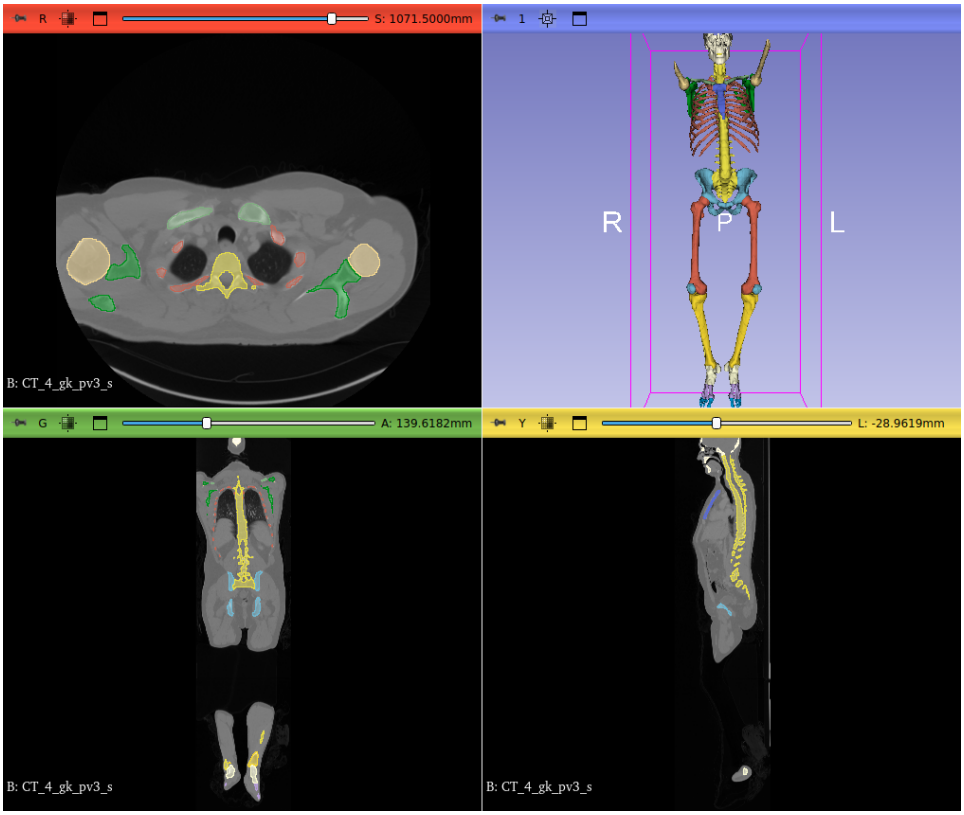


Figure 3.2: Example of MOOSE segmentation. In the image are represented a CT image (black and white) and an example of bone segmentation (colorized), projected on the three medical imaging planes and on top right a 3D segmentation rendering. Each color represents a different bone or small bones group segmented and distinguished by the model. The spine segmentation is well visible in yellow. The original image comes from TCIA public dataset, and they are visualized with 3D Slicer software.

At this point of the pipeline, after performing the segmentation, the data on hand are CT, PET and segmentation mask of each patient from the two cohorts. However, CT and the segmentation have the same resolution and spatial properties, although PET images do not. Moreover, the segmentation performed allows to select the whole skeleton from the CT image, and it may be of interest to make a selection of the most interesting bones to be analyzed.

The following paragraphs describe the operations carried out on the data, how they are performed and the reasons behind the choice that have been made.

3.4.1 Segmentation binarization

The segmentation masks performed on the CT image are, as discussed in paragraph 3.3, segmented the entire skeleton distinguishing it in bones or groups of smaller bones with labels. The images are represented by arrays where the voxel value indicates the label of the segmentation. All the voxels with a value of zero do not depict a region of the skeleton but rather other types of tissues or organs.

MM is a disease which, depending on its stage and progression, generates lesions in different sites of the patient body. As discussed in paragraph 2.2, the lesions in the early development of the disease are usually presented in the bone marrow regions, and while MM progresses the lesions can be originated in the PM zone and extramedullary zone. In order to select the zone with the highest possibility to get information about the disease, the bone selection is crucial. The area where the lesions concentrate the most is the spine, which is also a bone with a high concentration of bone marrow. Therefore, opting for the entire spine for the analysis may be an appropriate selection.

To select the spine from the full skeleton segmentation means generating another segmentation based on the original one, filled with ones in the voxels representing the spine and zeros in the others. It is easily performed dealing directly with the image array values, obtaining a mask as the one represented in Fig. 3.3. In this case there is no need to change the header of the NIfTI file storing the segmentation since none of the metadata is changed in the process, only the image array changes with a pixel-wise operation.

3.4.2 Segmentation shaping

The lesions analyzed in the discussion can be found located in the bone marrow, the PM area and the extramedullary area. The operations of segmenting and binarizing account for the information stored inside the spine, in particular its bone marrow, but PM and extramedullary lesions are excluded from this type of selection. Extramedullary lesions are disconnected from the skeleton tissues and thus difficult to identify and locate without the intervention of a clinician. On the other hand, PM lesions are located in connected or neighbouring regions to the bone marrow inside specific bones, therefore from the bone segmentation it is possible to reach such areas. From the selection of the segmentation of the spine, through morphological operations.

Once selected the spine and binarized the segmentation, it is easily performed the shaping through the application of convolutional kernels to the image. Convolution is a simple mathematical operation, crucial in many image processing applications. It performs a local operation on a digital image, through the application of a kernel, an array of the same dimension of the image also called structuring element, usually much smaller than

Figure 3.3: Example of MOOSE binarization. In the image are represented a CT image (black and white) and an example of bone segmentation binarized (yellow), projected on the three medical imaging planes and on top right a 3D segmentation rendering. It is clear that from the complete skeleton segmentation, just the spine is conserved, specified by a unique label. The original image comes from TCIA public dataset, and they are visualized with 3D Slicer software.

the image. Continuous convolution in three dimensions is generally defined with the following formula:

$$(f * g)(x, y, z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\chi, \xi, \zeta) g(x - \chi, y - \xi, z - \zeta) d\chi d\xi d\zeta \tag{3.2}$$

where f is the original function and g is the kernel function. When working on digital images, the function have a discrete and finite domain, therefore the convolution process needs to be adapted. The following equation represents the operations performed on discrete function:

$$(f * g)[m, n, p] = \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} \sum_{l=0}^{P-1} f[m, n, p] g[n - j, m - k, p - l] \tag{3.3}$$

where f is the discrete function with a finite domain, g is the kernel and N, M, P are the domain's dimensions of the discrete function. To gain an intuitive understanding of its functioning, it involves overlaying one sequence, the kernel, onto another, computing the weighted sum of their overlapping elements at each position. This process efficiently captures the local interactions and transformations within the data. By convolving digital images with appropriate kernels, it is possible to extract meaningful information, enhance features, or apply specific operations to achieve desired outcomes in a wide range of applications.

Therefore, the processes employed are the following:

- *binarydilation* : operation to expand regions of foreground pixels. The main parameters of the function are the structuring element, defining the neighbourhood around each pixel with a default full cross-shaped structuring element, and the number of iteration of the process.

- *binaryerosion* : a morphological operation that involves shrinking the regions of foreground pixels within a binary image. The main parameters of the function are the structuring element, defining the neighborhood around each pixel with a default full cross-shaped structuring element, and the number of iteration of the process.

- $binaryfillholes$: operation for filling holes in binary images, identifying and filling voids, producing a more complete representation of the desired structures. The main parameter in this case is the structuring element, defining the neighbourhood around each pixel with a default full cross-shaped structuring element. This process is a combination of the previous ones. In fact, a preliminary dilation is able to fill the holes present in the figure, in this way they are not anymore part of the border of the image. Then, when applying an erosion the holes only the border parts of the figure are eroded returning back to the original shape.

Three selected shapes could be of particular interest for the analysis: the original spine segmentation, the spine encompassing the entire marrow area within it, and the latter with the additional inclusion of the PM area of the spine. While the first one is the segmentation as it gets out of the MOOSE model, the other two selections had specific reasons to be performed. In the first shaping, the cavity within the spine segmentation is filled to include the bone marrow in the analysis. This addition is crucial, as the bone marrow is the region where lesions are most likely to be present in a patient with MM. The second shaping was selected in order to analyze also the PM region, where the lesions displayed are known to be informative about the progression of the disease. This is underlined by the acR-ISS and acIMPeTUs criteria analyzed in paragraph 2.2.2 and 2.2.3.

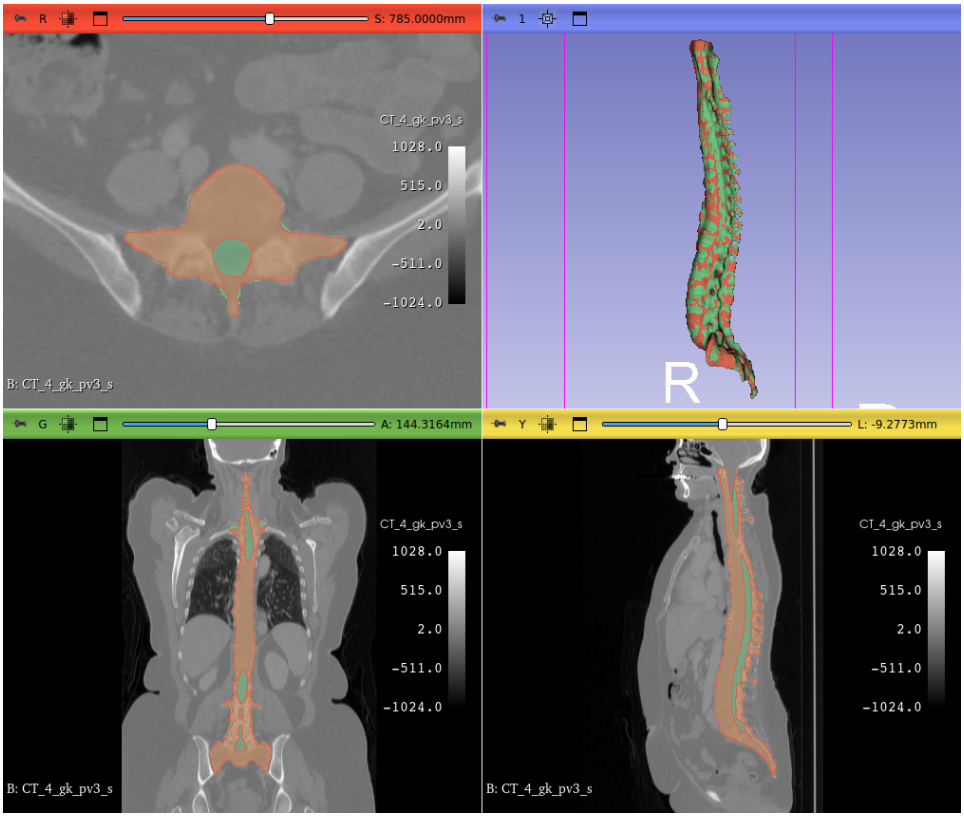


Figure 3.4: Example of *fill_holes* process for segmentation shaping. In the image are represented a CT image (black and white) and two spine segmentation: in red is represented the original spine segmentation, and in green the spine segmentation after the application of the filling process. All the elements are projected on the three medical imaging planes and on top right a 3D segmentation rendering is shown. Looking the 3D rendering it is possible to notice that the borders of the two segmentations do not exactly coincide. The original image comes from TCIA public dataset, and they are visualized with 3D Slicer software.

The first one, referenced throughout the work as *fill_holes*, is performed with a combination of the three morphological operations described below. Firstly, the *binary_dilation* is performed on the original segmentation. Three iterations are performed with the default structuring element of the function, to ensure that small gaps or narrow connections within the regions of interest are closed. Afterward, the function *binary_fill_holes* is executed. It is implemented with a 3x3x3 cube-shaped structuring element. Within this cube, only the middle layer of the structure, which lies horizontally, is filled with 1s. The rest of the layers are entirely filled with 0s. This convolutional kernel is useful to set the function in order to fill all the holes with a closed structure within the transverse plane, such as the ones inside the spine where lays the bone marrow. Finally, three iter-

ations of *binary_erosion* are performed with the default structuring element as employed in the dilation process. This step helps to refine the shapes, eliminating any artifacts introduced during dilation, and restoring the original structure as much as possible. In Fig. 3.4 is represented an example of segmentation which went through the filling process.

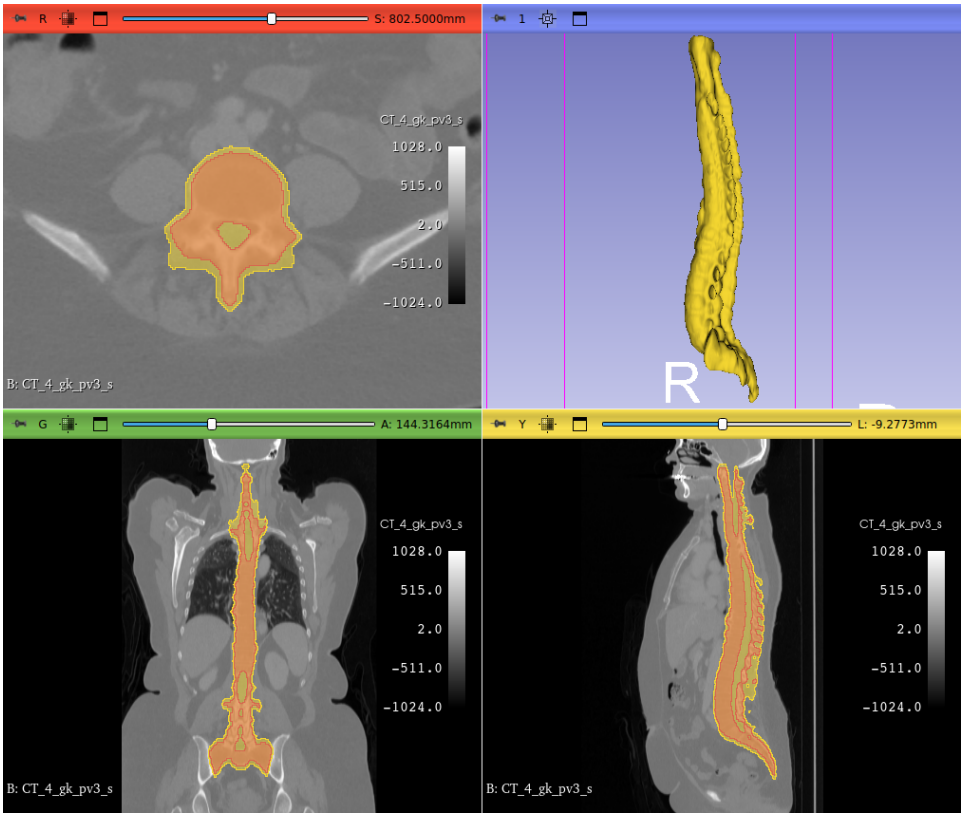


Figure 3.5: Example of *dilation* process for segmentation shaping. In the image are represented a CT image (black and white) and two spine segmentation: in red is represented the original spine segmentation, and in yellow the spine segmentation after the application of the *dilation* process. All the elements are projected on the three medical imaging planes and on top right a 3D segmentation rendering is shown. Looking the 3D rendering it is possible to notice that just the dilated segmentations is visible due to the fact that the original one is enclosed by it, as it is represented in the three projections. The original image comes from TCIA public dataset, and they are visualized with 3D Slicer software.

The second type of spine segmentation shaping, referenced from now on throughout the work simply as *dilation*, is performed with a combination of the previous shaping process and an additional dilation morphological operation. After applying the *fill_holes* process

in order to comprehend the spinal marrow area, a *binary_dilation* is performed on the resulting mask. Three iterations are performed with the default structuring element of the function, to ensure that most part of the PM area and spine surrounding is considered within the resulting mask. In Fig. 3.5 is represented an example of segmentation which went through the *dilation* process.

3.4.3 PET and segmentation alignment

After achieving the segmentation shaping, the next challenge to tackle involves addressing the resolution and voxel spacing disparities between segmentations and PET images. The primary goal in this paragraph is to ensure alignment among PET images and segmentation within their respective image arrays. To achieve this, image offsets must align, and both images need to share the same voxel spacing. This alignment proves crucial for the subsequent step in the pipeline, where the region surrounding the segmentation is set to zero for both CT and PET images.

Firstly, upon analyzing the image cohort, a notable observation is that PET images depict a larger FoV compared to segmentations, despite the fact that they present a greater voxel spacing value in the transversal plane. However, it is important to bear in mind that the two images share the same spatial domain, and that segmentations share the same characteristics as CT images. Given such attributes, the processes performed in this pipeline involves the implementation of a PET image resampling in order to make it share the same voxel spacing with the segmentation. A schematization of the original voxel size, resolution and FoV of PET and segmentation.

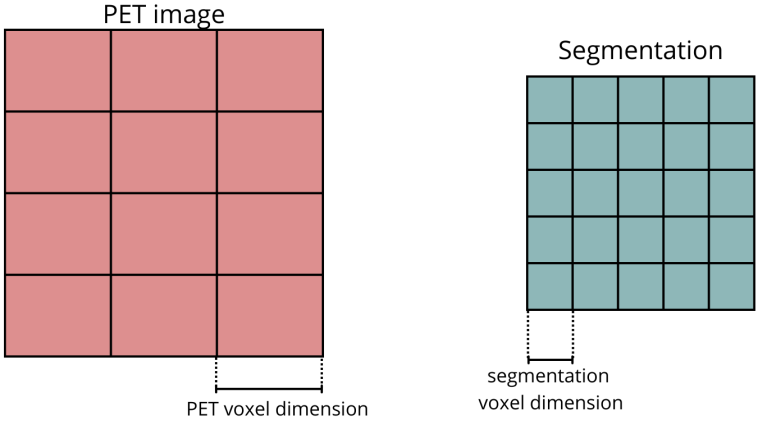


Figure 3.6: Diagram representing the original PET image and the segmentation. In this example is represented a schematization of the voxel dimensions, the resolutions and the FoV of the images.

PET images zooming and segmentation padding

For the PET image resampling is employed an algorithm for manipulating and adjusting the dimensions of arrays or images. When applied to an array, it allows for the resampling or interpolation of the data, effectively altering its shape or size. The core functionality involves specifying a zoom factor, which determines the degree of expansion or contraction in each dimension of the array. This process involves interpolation methods to estimate the values of the new array elements based on the existing data. The one employed for this process is a standard *spline* third order interpolation [48].

Employing this type of zoom for the PET image array, it is possible to manage the resolution of the image keeping the FoV fixed. The objective is to obtain a PET image with the same voxel spacing as the segmentation, and this is performed by using the following zoom factor for each dimension:

$$zoomfactor = \frac{PET\ spacing}{Segmentation\ spacing} \tag{3.4}$$

where $zoomfactor > 1$, being PET spacing greater than segmentation spacing for the cohort employed. Therefore the PET image obtained has the same voxel spacing of the segmentation for each axis. Actually, a small difference remains between voxel spacing of the two images. In fact, in cases where the zoom factor results as an integer PET spacing is perfectly matching the segmentation, but when the zoom factor results a non-integer value this is not true. The resolution of an array must be an integer value and though the process rounds the resolution to the lower integer. Hence, this rounding brings an error when trying to match the two resolutions, with a value of at most the voxel spacing length. It is a quantity, which contribution to the image displacement is distributed among all the voxel in each direction. For the high resolution images employed results as a small error, nevertheless an error to take into consideration that will be addressed as *pixel surplus*.

However, having the PET image a FoV greater than the segmentation, PET resulting image after the zoom operation happens to present a higher resolution than the segmentation. This issue can be solved applying a padding to the segmentation, in order for it to match PET resolution.
The process, represented in a diagram in Fig. 3.7, is performed evaluating first the voxel number difference in each direction between PET and segmentation resolution. The segmentation is placed in the center of an array full of 0s by shifting it of the half number of voxel just evaluated. Therefore, the segmentation is centered in the middle of an image array with the same resolution as the PET image.

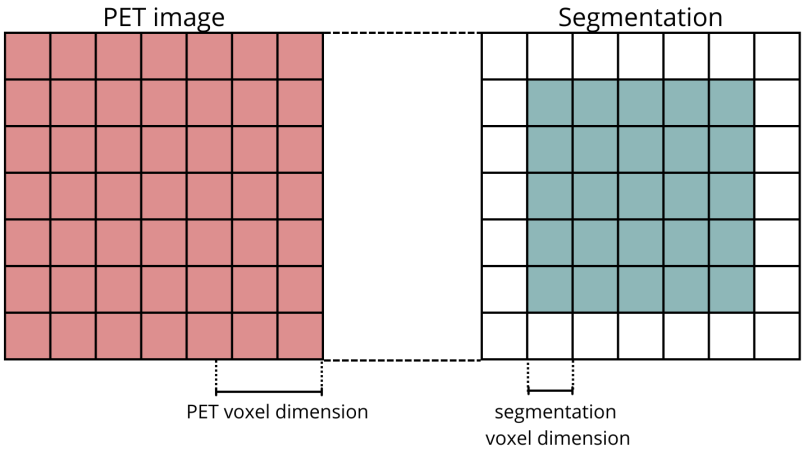


Figure 3.7: Diagram representing the upsampled PET image and the segmentation with applied the padding. In this example is represented a schematization of the voxel dimensions, the resolutions and the FoV of the images.

Every process performed on the image array of a NIfTI file has to be reflected by a change in its header. In this case the changes brought to the images that need to be corrected in the header are the following:

- *PET*: voxel spacing changing in every direction, and *pixel surplus* correction.

$$newPETspacing = \frac{old\ PET\ spacing}{zoom\ factor} - pixel\ surplus \tag{3.5}$$

- *PET*: image offset, due to the voxel spacing variation and being the offset evaluated from the voxel center.

$$newPEToffset = old\ PET\ offset - \frac{old\ PET\ spacing}{2} + \frac{new\ PET\ spacing}{2} \tag{3.6}$$

- *PET*: image resolution, resulting from the zoom operation.

$$newPETresolution = old\ PET\ resolution \cdot zoom\ factor \tag{3.7}$$

  where the result is rounded to the lower value.

- *Segmentation*: image resolution, set equal to the resulting PET resolution.

- *Segmentation*: voxel spacing, with the *pixel surplus* correction. Must result equal to the PET image spacing

$$newSegmspacing = old\ segm\ spacing - pixel\ surplus \tag{3.8}$$

- *Segmentation*: image offset. With the paddling added the offset must be evaluated by measuring the FoV addition in every direction.

$$newSegmoffset = old\ segm\ offset - new\ Segm\ spacing \cdot padding\ number \quad (3.9)$$

where *padding number* is the number of voxel added in the padding process on one side.

At this point of the pipeline PET image and segmentation presents same voxel spacing, resolution and FoV.

PET images shifting

So far, PET images to result aligned have to the segmentation if they are traceable to a standard frame of reference, as the images of the MM cohort are. However, they both have also to be centered, in order to overlap perfectly when the segmentation is put in the center of the frame of 0s. This property, represented schematically in Fig. 3.8 is not valid for every image of the cohort employed in this work.

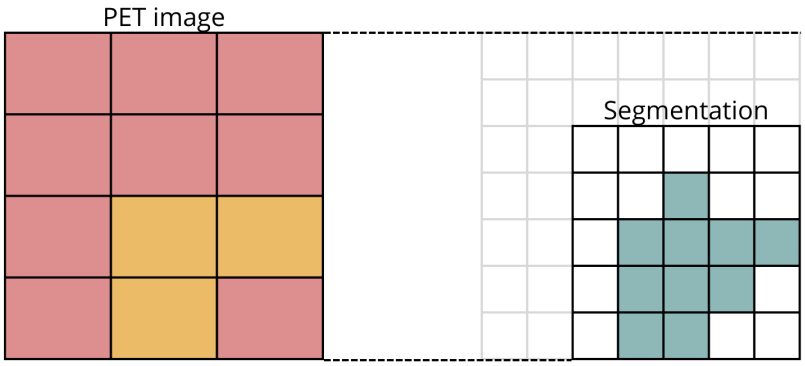


Figure 3.8: Diagram representing the PET image and the segmentation before applying the padding. In this case the images are not centered. Both images represent the same space, in fact, the information stored in the PET image (yellow) corresponds spatially to the information stored in the segmentation (light blue).

Therefore, there is the need to align the spatial information of the two images, and in order to perform such operation the shift between the two image information must be evaluated. It is easy to see that the shift is expressed by the difference between the offset of the two images, which are not the same if the two images are not centered. For every dimansion it is evaluated with the following formula :

$$shift = \frac{new\ segm\ offset - new\ PET\ offset}{newPETspacing} \quad (3.10)$$

which is expressed in voxel unit.

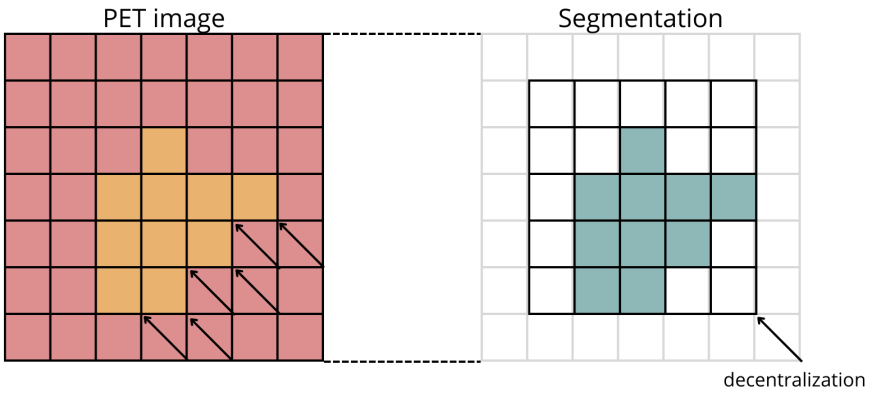


Figure 3.9: Diagram representing the PET image and the segmentation before applying the padding. In this case the images are not centered. Both images represent the same space, in fact, the information stored in the PET image (yellow) corresponds spatially to the information stored in the segmentation (light blue).

Employing the same spline interpolation of the previous step, on the PET image the information stored is shifted in order to produce an image aligned with its segmentation. The information on the borders is negligible, affecting it just the peripheral information of PET images which never comprehends important areas of the patient.

3.4.4 PET and CT shaping

At this stage of the pipeline, there are aligned CT images, PET images and segmentations mask featuring arrays that are also aligned with each other. The last process performed for the preliminary arrangement of the data for the analysis is the shaping of the two types of medical scans according to the three segmentations shapes selected: the original spine segmentation, the *fill_holes* spine segmentation, and the *dilation* spine segmentation.

In the context of a CT image and a set of three spine segmentations, the arrays are extracted from NIfTI files. The image array undergoes a straightforward pixel-wise multiplication with the three binary segmentation masks. As a result, three distinct CT images are generated, each encompassing CT image grey levels within the voxels identified by the segmentation mask labeled as 1, while the remaining voxels contain 0s. PET images undergoes an identical process, however the images employed are the ones obtained from the alignment described in the previous paragraph.

This process is employed in light of the following feature extraction that is performed next in the pipeline. This approach allows a more targeted analysis by concentrating feature extraction solely on the segmented region, aiding in a focused characterization of specific areas of interest. Additionally, it can help mitigate potential boundary effects, providing a more stable feature extraction process, particularly near segmentation edges. Noise reduction is another potential benefit, as excluding irrelevant information from the surrounding regions may result in cleaner radiomic features.

3.5 Feature extraction with pyradiomics

Feature extraction was performed with *PyRadiomics* library [11], providing standardized tools in the field of radiomics. The aim of the library is to establish a reference standard for feature extraction in radiomics. It possesses the capability to extract a diverse array of numerical descriptors from both the original medical image and its corresponding segmentation, facilitating the capture of interesting information embedded in the image's voxel intensities. The integration of segmentation refines this process, allowing for a more focused extraction of features from specific regions of interest. By leveraging various filters, these tools provide a more comprehensive representation of the underlying structures and patterns within the medical images. It is a library used for numerous purposes and applications in medical field [3, 51, 61].

Different filters and features can be selected using *PyRadiomics* for the extraction. For the purposes of this work two type of filters are selected alongside with the original image: *wavelet* filters and *LoG* filters. Wavelet filters are filters where each is one of the 8 combinations of high pass filter and low pass filter applied along the three image axis. LoG filters on the other hand are obtained through the Laplacian transforms of a Gaussian band-pass filter with a specified sigma parameter representing the filter width to use for the Gaussian kernel. For this specific feature extraction in this work's pipeline two LoG filters were employed, with sigma equal to 1.0 mm and 2.0 mm. The sigma selection expresses the coarseness of the smoothing operation in the filter. It was performed keeping in mind the voxel dimensions of about 1 mm of the final images, allowing to obtain two images with a noise reduction without loosing excessive information. Both filters are commonly exploited filters in radiomics for feature extraction due to their ability to extract properties enclosed in the original image, as highlighted in paragraph 2.1.3.

In particular the feature extracted in the present pipeline are, from the original image, the following:

- First Order Statistics (19 features)
- Shape-based (3D) (14 features)

- Gray Level Co-occurrence Matrix (24 features)
- Gray Level Run Length Matrix (16 features)
- Gray Level Size Zone Matrix (16 features)
- Neighbouring Gray Tone Difference Matrix (5 features)
- Gray Level Dependence Matrix (14 features)

with a total number of 102 feature extracted from the original image, with 14 features that only valuate the segmentation shape. The features extracted from an image with applied a wavelet or LoG filter are 88, because of the absence of the 14 shape features. Therefore the total number of features employed in this work, with the original image and 10 different filters employed, amounts to 982 features.
Lastly, it is crucial to scale the feature obtained. Feature scaling aims to standardize or normalize the range of features to ensure that they contribute equally to the model training process. For scaling a specific algorithm is employed which is particularly suited to deal with outliers. The data is scaled by eliminating the median and adjusting the scale based on the InterQuartile Range (IQR), which is the range between the 1st quartile and the 3rd quartile of the data. In simpler terms, it centers the data around its median and scales it in a way that accommodates the variability within the middle 50% of the dataset, making it robust to outliers.

Outliers can significantly impact the performance of standard scaling methods, such as Min-Max scaling or Z-score normalization, by skewing the scaling process toward extreme values. This approach helps maintain the integrity of the feature scaling process, allowing ML models to be more robust and accurate, particularly in scenarios where data may contain extreme values or anomalies.

3.6 Survival Analysis with Cox model

In paragraph 2.1.4, the foundational principles of survival analysis have been established, including the Kaplan-Meier estimator, log-rank test, and Cox proportional hazards model. Building upon this theoretical groundwork, the current section outlines the practical implementation through a step-by-step description of the survival analysis pipeline employed in this study.

The data available at this stage of the pipeline is enclosed in six features datasets and one clinical information dataset:

- *MM_PT_o*: dataset containing the features extracted from 93 patients with *PyRadiomics* from PET images shaped in the original spine segmentation shape.

- *MM_PT_f*: dataset containing the features extracted from 93 patients with *PyRadiomics* from PET images shaped in the *fill_holes* spine segmentation shape.
- *MM_PT_d*: dataset containing the features extracted from 93 patients with *PyRadiomics* from PET images shaped in the *dilation* spine segmentation shape.
- *MM_CT_o*: dataset containing the features extracted from 93 patients with *PyRadiomics* from CT images shaped in the original spine segmentation shape.
- *MM_CT_f*: dataset containing the features extracted from 93 patients with *PyRadiomics* from CT images shaped in the *fill_holes* spine segmentation shape.
- *MM_CT_d*: dataset containing the features extracted from 93 patients with *PyRadiomics* from CT images shaped in the *dilation* spine segmentation shape.
- *Clinical data*: dataset containing all the demographic and clinical information of the 93 patients. Of particular relevance to the survival analysis is the PFS data, which includes 'time-to-event' information. This data is accompanied by a boolean variable indicating whether the event has occurred within the duration of the study.

An important step to bear in mind, for its essential role in the reliability of the model implementation, is to verify the Cox model assumptions for all the features of the datasets employed. It is important, therefore, to exclude from the analysis all the covariates that does not verify Cox model assumptions.

Considering the feature datasets and the associated PFS information, conducting a Cox model analysis poses a challenge due to the substantial number of covariates and the limited size of the patient cohort. To address this issue, strategies for handling the extensive covariate space within the constraints of the available patient data will be explored.

3.6.1 Nested cross-validation Cox model

When dealing with a limited amount of data, employing CV becomes crucial for robust model evaluation and mitigating the risk of overfitting [5]. Overfitting occurs when a model learns the training data too well, capturing noise or specific patterns that may not generalize well to unseen data. In the context of limited data, this is particularly problematic, as the model might end up tailoring itself too closely to the characteristics of the small dataset, failing to generalize accurately to new observations. CV helps address this challenge by partitioning the available data into multiple subsets, typically training the model on a portion and validating it on the rest. This process is repeated multiple times, and the model's performance is averaged over these iterations. By doing so, cross-validation provides a more comprehensive assessment of how well the model

generalizes to different subsets of the data. Moreover, CV is useful for the model hyperparameters tuning, contributing to a better training of the model with the most suited hyperparameter value.

Therefore, to address both the challenges of overfitting in the context of limited data and to ensure effective model hyperparameter tuning, the employment of nCV (also referenced to as double Cross-Validation) [76] emerges as a valuable strategy. nCV involves an outer loop for model evaluation and an inner loop for hyperparameter tuning. The scheme of the model employed is shown in Fig. 3.10. In the outer loop, the data is split into training and test sets, and the model is evaluated. Simultaneously, in the inner loop, the training set is further partitioned for hyperparameter tuning, optimizing the model's performance. This nested approach provides a robust assessment of both generalization to new data and the selection of optimal hyperparameters. By iteratively refining the model in this manner, the risk of overfitting is minimized, ensuring a more reliable and accurate performance evaluation. [75, 5]

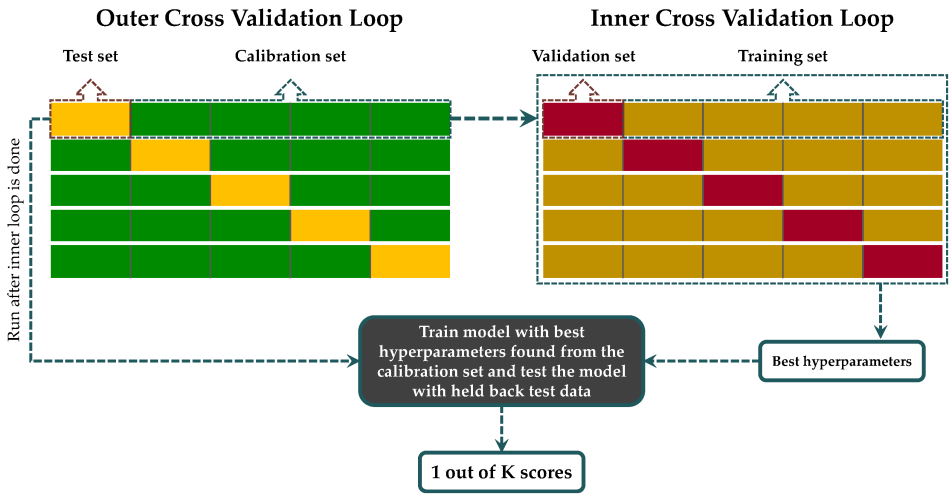


Figure 3.10: Scheme representing the nCV employed in this work. It shows the inner CV for hyperparameter tuning and the outer CV for the training of the model. [43]

In this pipeline a nCV algorithm is employed between a wide range of state-of-the-art ML algorithms for medium-scale supervised and unsupervised problems. The operations employed in this work to develop the algorithm are the following:

- *K-Folding* : slitting the dataset in k sets, then for each iteration providing the indices of one set for testing and the others k-1 for training. The k parameter was set to 5 with a preliminary shuffle to the dataset before the splitting. This

parameter selection allows having more reliable results and a robust training for a dataset of 93 entries.

- *Search for the best hyperparameter*: defining a hyperparameter distribution and a model to be trained, and then employing an algorithm that randomly samples combinations of hyperparameters from that distributions, for a specified number of iterations. A uniform distribution was selected between the values of 0.01 and 3, to have a wide range of possible parameters in the research, with 10 iteration of the process. Afterward it evaluates model performance using simple CV, characterized of another 5-folding algorithm.

- *Cox training, verification and evaluation*: Cox model training is the central part of the process. With a train set and a test set the model can be trained, and once trained the Cox model assumption are verified. Afterward, for evaluating the model performance a score is evaluated featuring the C-index metrics. Inside the model, it is also possible to specify the type of penalization and the penalizer parameters. A LASSO penalization is selected, as it performs a regularization of the covariates dealing with the collinearity issues that could affect such high dimensionality datasets, with such a high number of covariates compared to the samples. LASSO is coarser than the other types of penalization, and for this more appropriate for the present dataset, with a penalizer value selected by the hyperparameter search.

For analyzing the 982 features' datasets of 93 patients the nCV algorithm selected is made of an outer CV loop and an inner CV loop. Both are 5 folds structures, in order to suit a 93 patients dataset increasing test reliability and maximizing the training sets. The hyperparameter tuning happening in the inner CV loop is performed over the "penalizer" for a LASSO penalization of the model, providing a uniform distribution with references 0.01 and 5. For each model training the Cox model assumptions discussed in paragraph 2.1.4 are verified, assessing if the datasets present covariates in concordance with the assumptions for all the datasets' splits. This step is performed after every model training. Finally for each split of the dataset 5 C-index, one for each fold, are evaluated on the best model, and from those the mean value, the standard deviation, and the lower value are obtained.

3.6.2 Most informative features analysis through Cox model

A second analysis on the same dataset was performed. The method employed is called Stepwise Regression (SR) [19, 8] is a method used in statistical modeling to iteratively select the most significant variables for inclusion in the model or remove less significant ones. It involves step-by-step iterations, including adding or removing variables based on certain criteria, such as p-values or other statistical measures, until an optimal set

of variables is determined. The two main variations of SR are "Forward Selection" (adding variables one at a time) and "Backward Elimination" (removing variables one at a time). This technique helps the model focus on the most important variables for predictive accuracy. A simple example of the general idea of the model is represented in Fig. 3.11.

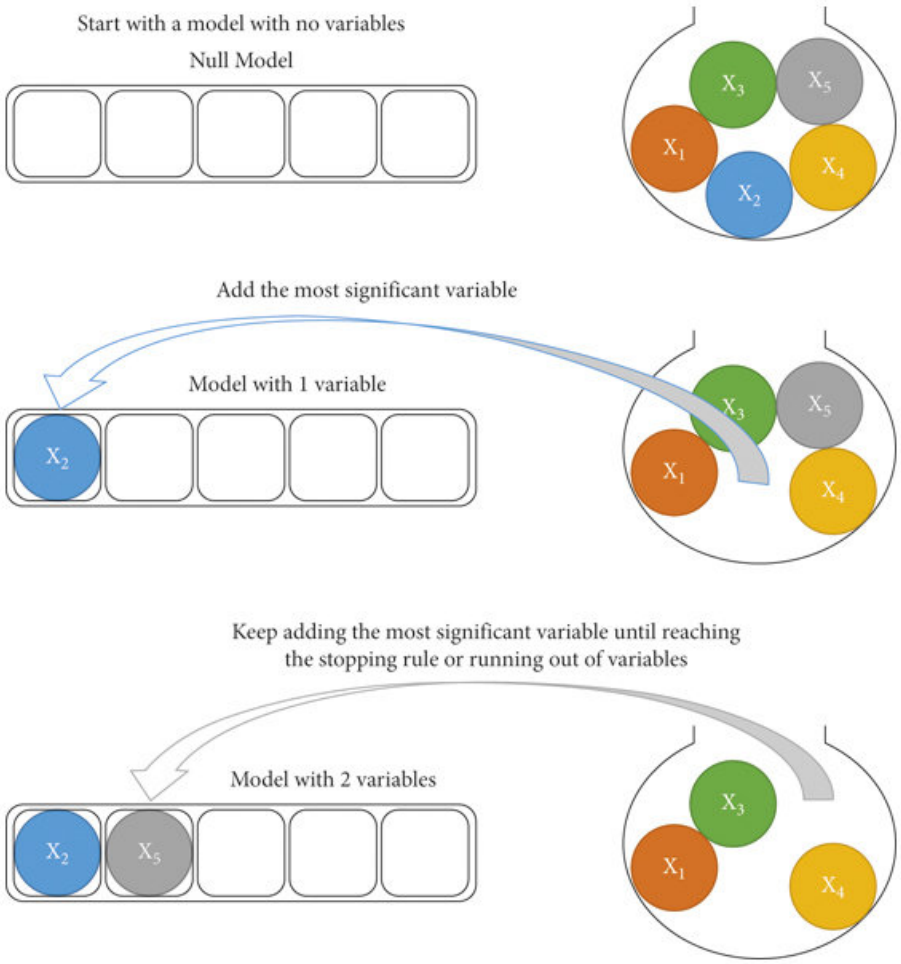


Figure 3.11: Scheme representing an example of SR employed on a dataset of five variables. [73]

In details, the first step involves dividing the dataset into feature groups based on the feature extraction filters used: employing a combination of the original image features and each additional filter. The resulting groups are 10 different datasets, for all the 6 datasets employed (paragraph 3.6), of 190 features, 102 coming from the original image and 88 for each filter. The Cox model penalized with LASSO penalization is trained us-

ing the prepared datasets, with a general penalization parameter. The model evaluates for each covariate their contribution to the regression, associating a p-value assessing the accordance of the covariate with the null hypothesis of non-correlation. The covariates with the lowest p-value are the ones resulting more in contrast with the null hypothesis, and therefore they are the ones contributing more to the regression result showing a higher correlation with the PFS information. However, the training results are affected by the abundance of covariates and the scarcity of data, thus it reveals useful to refine the model. Through the process shown in Fig. 3.11, the most relevant feature is conserved and employed to completely train a new Cox model verifying to get a p-value lower than 0.05. It is a common threshold [57] selected to discriminate the difference between a feature contributing to the regression model and a feature which does not. Furthermore, a second feature is added and the Cox model is evaluated again, verifying again that the p-values are lower than the threshold. This process is repeated until the new feature added results in the model training with a p-value higher than 0.05; at that point the last feature is removed, and the features left are selected as the most informative ones for the final model training. Finally, the model can be trained with a simple 5-fold CV, process to enhance the reliability and robustness, on the only features conserved. After every training, using the *check_assumptions* method provided by *CoxPHFitter*, the assumptions are checked in order to safeguard the reliability of the model. Finally for each split of the dataset 5 C-index, one for each fold, are evaluated on the best model, and from those the mean value, the standard deviation, and the lower value are obtained.

The process employed allows, at the end, to have insights on the most informative features and which is the model that best generalizes the datasets. The selection of the model with the highest performance between the results within the same dataset is made by considering both mean C-index and its standard deviation. In fact, it is important to consider both in the choice of the best model. This allows to obtain the model with a high mean value, meaning that it has a proper performance, and a low standard deviation, in order for it to be more stable and generalizable. For these reasons the function employed to weight the two elements in evaluating the score is the following:

$$Score = \alpha \cdot \overline{\mathbf{C}} - (1 - \alpha)\sigma(\mathbf{C}) \tag{3.11}$$

where $\overline{\mathbf{C}}$ is the C-index mean value, $\sigma(\mathbf{C})$ is its standard deviation and α is the parameter that weights the selection. In this case a value of $\alpha = 0.5$ was selected to give both the variables the same importance in the selection.

Chapter 4

Results and analysis

In this chapter, the analysis results are presented and discussed in comparison with the existing literature and within the context of current research. The main results come from the two analysis approaches explained in the precious chapter: the nCV training of the Cox model on the whole dataset, and the SR analysis for the training of the Cox model with a smaller selection of the most informative features, both on the PFS information.

There are multiple purposes for the analysis. Cox model, as underlined in paragraph 2.1.4, has a great potential of discriminating between hazard probabilities for different patients, and for this it can be helpful for clinical decision support. If the model implementation on the MM cohort datasets reveals satisfying results it can pave the way to a more concrete support to clinicians from radiomics analysis. In fact, from Cox analysis many insights can be obtained that may help to understand the disease and to refine the pipeline for further developments. The first question raised is about the contribution to the analysis of the two types of medical imaging, CT and PET, whether both are informative data or more useful than the other for the analysis sakes. The following question is about the pipeline which allows acquisition of the final information for the analysis. It can be refined from the insights learned from the analysis, through a different selection of segmentation shapes, and a different selection of filters employed and features extracted from the data. The following paragraphs try to answer these questions and others that may emerge discussing the results of the analysis.

4.1 Cox Model results

Here in this paragraph are presented the results of the two Cox model implementations. For both the model is trained with a train set from the 93 inputs dataset and evaluated with the C-index metrics on the test set. Having a total of 6 different datasets,

obtained from CT and PET images forming them in the selection of three shapes, the pipeline is implemented for each of them in order to have a complete set of results. The datasets available for the analysis are the following:

- *MM_PT_o*: dataset containing the entire set of 982 features extracted from 93 patients with *PyRadiomics* from PET images shaped in the *original* spine segmentation shape. From the clinical data is provided the PFS information crucial for Cox model analysis.
- *MM_PT_f*: dataset containing the entire set of 982 features extracted from 93 patients with *PyRadiomics* from PET images shaped in the *fill_holes* spine segmentation shape. From the clinical data the dataset is provided with PFS information crucial for Cox model analysis.
- *MM_PT_d*: dataset containing the entire set of 982 features extracted from 93 patients with *PyRadiomics* from PET images shaped in the *dilation* spine segmentation shape. From the clinical data the dataset is provided with PFS information crucial for Cox model analysis.
- *MM_CT_o*: dataset containing the entire set of 982 features extracted from 93 patients with *PyRadiomics* from CT images shaped in the *original* spine segmentation shape. From the clinical data the dataset is provided with PFS information crucial for Cox model analysis.
- *MM_CT_f*: dataset containing the entire set of 982 features extracted from 93 patients with *PyRadiomics* from CT images shaped in the *fill_holes* spine segmentation shape. From the clinical data the dataset is provided with PFS information crucial for Cox model analysis.
- *MM_CT_d*: dataset containing the entire set of 982 features extracted from 93 patients with *PyRadiomics* from CT images shaped in the *dilation* spine segmentation shape. From the clinical data the dataset is provided with PFS information crucial for Cox model analysis.

4.1.1 Cox model results on the complete feature datasets

Cox model evaluation, as discussed in paragraph 3.6.1, needs a preliminary step involving the verification that all the covariates fulfill the model assumptions. This step is performed after each training process, and it was determined that none of the examined data violated the PH assumption in any cases. One example of the test performed to verify the PH assumptions required by the Cox model is shown in Fig. 4.1. The residuals represented in the image are indexes of the agreement of the covariate with the PH assumption, as underlined in paragraph 2.1.4. Considering the image presented, there is

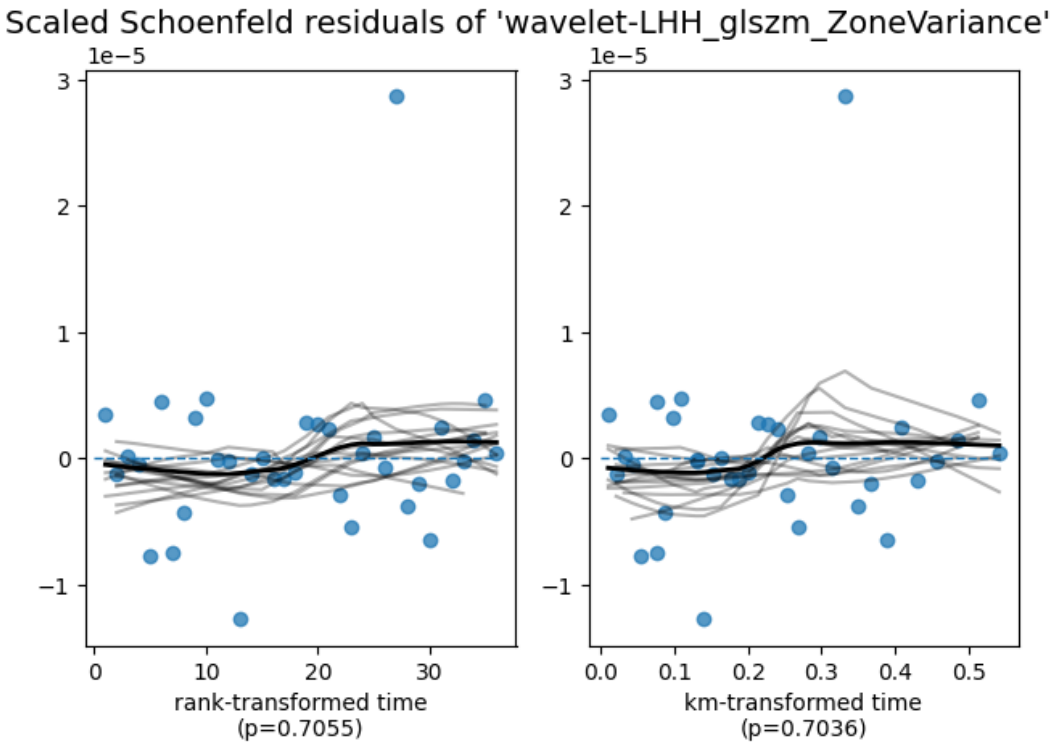


Figure 4.1: Graph produced by *CoxPHFitter* function in the checking assumption process. It represents scaled Schoenfeld residuals against the km-transformed time and rank-transformed time, of the covariate *glszm_ZoneVariance* extracted by *PyRadiomics* from the wavelet filter (LHH) of the original image. Cyan dots represent the evaluated residuals and black lines the fitted lowess along with 10 bootstrapped lowess lines (as an approximation to the confidence interval of the original lowess line) [12].

no clear evidence of violations of the proportional hazards assumption for that covariate. The flatness of the fitted lowess line and the tight clustering of the bootstrapped lines suggest that the effect on the hazard remains relatively constant over the transformed time. The illustrated example offers a visual confirmation, providing valuable insights into the reliability of the proportional hazards assumption for the specified covariate and timescale. The majority of the covariates utilized in the model have been verified for the assumptions. However, the concern becomes of marginal significance in the context of survival prediction. Specifically, when aiming to optimize the model based on the C-index, the non-verification of the assumption by a reduced subset of features becomes a minor consideration, resulting in a slightly diminished predictive capability [14].

Once verified the model assumptions to ensure the reliability and validity of the model's results, it is possible to follow the discussion showing the evaluations of model

training. The process has been realized on the test splits of each dataset for the outer 5 folds of the nCV algorithm, composed of 18/19 samples each, using as evaluation metrics the C-index. The quantity of samples in each test split is insufficient for achieving highly precise results, therefore, it is reasonable to expect that the evaluation will show high variances values. In Tab. 4.1 are shown the final results of the Cox model provided employing the nCV pipeline.

Dataset	C-index
PET spine	0.50 ± 0.03
PET spine + marrow	0.51 ± 0.05
PET spine+marrow+PM area	0.59 ± 0.03
CT spine	0.55 ± 0.07
CT spine+marrow	0.58 ± 0.07
CT spine+marrow+PM area	0.59 ± 0.04

Table 4.1: Results of nCV Cox model training

The obtained C-index from the analysis of the datasets reveal consistently poor results, with all indexes falling within the narrow range of 0.5 to 0.58. A poor result on the C-index metric means that the model struggle at ordering the probability hazard functions of two patients, and in this case the results indicate a near random behavior of the model. As anticipated, the standard deviation accompanying these values is exceptionally high. This collective pattern of low C-index means and elevated standard deviation prompts a critical examination to identify the underlying reasons for such unfavorable outcomes. In the context of training a Cox model with LASSO penalization, the challenge arises when the number of covariates significantly exceeds the number of input samples. This collinearity condition of the dataset, often referred to as a high-dimensional setting, is usually managed by LASSO regularization ensuring that the contribution of a proper number of covariates is precluded. With 982 features and only 93 input samples, the penalization struggles and potential issues are introduced with a high collinearity dataset with covariates that do not manifest either strong relations with the survival information label. Moreover, the dimensionality of the data poses a substantial risk of overfitting, where the model might learn noise in the data rather than capturing meaningful patterns. The challenge intensifies due to the limited number of input samples, making it difficult for the model to discern the true predictive signals from the noise. In such scenarios, the resultant model might lack stability and generalizability, as the estimated coefficients become highly sensitive to small variations in the input data.

The results obtained from the presented pipeline prove themselves to be substantially different from the ones obtained in the work brought by Peluso et al. in her thesis

work [60]. In fact, for the pipeline proposed by Bologna's side, the C-index values evaluating the Cox model, represented in Tab. 4.2, reflect a regression fitting remarkably well the data, with a proper generalization of the data splits. The comparison between

Dataset	C-index mean and std
PET spine+marrow	0.87 ± 0.05
PET PM area	0.90 ± 0.06
CT spine+marrow	0.89 ± 0.06
CT PM area	0.90 ± 0.06

Table 4.2: Results of nCV Cox model training from the work of Peluso et al. [60].

the two approaches seems to highlight a great difference in the model performance. In exploring the reasons behind the considerable discrepancy in results between our current approach and a counterpart with a different pipeline, there is the need to delve into the specific decisions that give rise to such variations. The differences in outcomes underscore the critical importance of understanding the decision consequences for the pipeline construction, examining the medical literature. It is crucial to establish a link between the consequences of a decision tracing back to the root cause based on the medical perspective, in order to gain more awareness on MM disease. In this case, consulting what has been discussed about MM lesions in paragraph 2.2.2, the main reason may be addressed to the selection of the segmentation area. It is known that the regions where MM lesions are more frequently presented is inside the bone marrow, in particular in the region enclosed by the vertebral spine. That made the reason for this work to isolate just the information enclosed inside that region, while Peluso selected the entire skeleton with the exclusion of the patient head. It is therefore reasonable to think that, even if the spine often presents significant lesions, they are not enough informative and predictive for what concerns the PFS analysis. Whereas, other hematopoietic regions, long bones' marrow such as femurs for example, which are excluded from the present pipeline, reveals as more related and informative for the PFS information.

4.1.2 Cox model results on most important features selection

The datasets have been analysed also employing the survival Cox model with SR algorithm, in order to address the excessive amount of variables that may influence the result. This method allows training the model with a subset of few features, as described in paragraph 3.6.2, and the process has advantages and drawbacks. Firstly, training a Cox model with few covariates is convenient to avoid convergence problems contributing to decrease regression reliability. However using a selection of covariates, instead of the entire set, produces a loss of information stored in those features that have been

removed. This may worsen the regression result and thus the prediction power of the model. Furthermore, one of the advantages of this type of feature selection is the insights that can be obtained from the selection itself. It facilitates the comprehension of interesting characteristics of the dataset, such as the type of filter that *PyRadiomics* employs to extract the highest number of informative features, or the type of features which are the most informative in absolute.

In the structure of the procedure employed in the present work, the datasets from where the model selects the best covariates are subsets of features coming from couples of filters: a combination of the features extracted from the original image and from one of the other filters. Therefore for each subset, the model is trained with the best feature selection and the C-index values are evaluated on a test split for each fold of the CV. The complete results of the training, for the 6 datasets and for all the combinations of filters are stored in Tab. 5.2, while here in the following table (Tab. 4.3) are reported the results of the best model for each dataset.

Dataset	**Best Filter**	**C-index**
PET spine	wavelet-LHL	0.65 ± 0.02
PET spine+marrow	wavelet-HHL	0.68 ± 0.03
PET spine+marrow+PM area	wavelet-HHH	0.71 ± 0.04
CT spine	wavelet-LHH	0.71 ± 0.05
CT spine+marrow	wavelet-HHL	0.72 ± 0.03
CT spine+marrow+PM area	wavelet-LLH	0.75 ± 0.04

Table 4.3: Table showing the best results from the Cox model training results for SR algorithm, the best one selected for each dataset. It shows the datasets, the filter alongside the original image for the feature extraction, mean C-index value and C-index standard deviation for the results coming from the five 5 CV.

As discussed in paragraph 4.1.2, the best model was selected by evaluating a score assigning weights to mean C-index and its standard deviation (Eq. 3.11). This type of score was selected for considering the mean metric's score of the model, but also the confidence interval of the model results. From a qualitative point of view, it is worth noting that apparently the models trained with CT datasets show higher C-index values than the ones trained with PET datasets. Moreover, the datasets of features extracted from a region of the image comprehending also bone marrow and PM areas are the ones resulting in models with a higher C-index. This fact is in accordance with the literature about the disease discussed in paragraph 2.2, where the paramedullary regions are discussed as an optimal prognostic factor about the progression of the disease.

In Tab. 4.4 are shown the covariates featuring in the best model selected for each dataset. Based on preliminary examination of the features employed, it is possible to notice that *original_shape_Flatness* results significant for all the best models. Moreover, the optimal models exclusively incorporate features derived from two distinct image types for extraction: the original image and the wavelet image.

Dataset	**Features**
PET spine	original_shape_Flatness, wavelet-LHL_glcm_Idn
PET spine+marrow	original_shape_Flatness, wavelet-HHL_gldm_LargeDependenceHighGrayLevelEmphasis
PET spine+marrow+PM area	original_shape_Flatness, wavelet-HHH_glcm_MCC
CT spine	original_shape_Flatness, wavelet-LHH_glszm_ZoneVariance, original_firstorder_10Percentile
CT spine+marrow	original_shape_Flatness, original_gldm_DependenceVariance
CT spine+marrow+PM area	original_shape_Flatness, original_gldm_DependenceVariance

Table 4.4: In this table are represented the features composing the best model that has been selected for each dataset. Feature names are composed by the filter of the image from which they were extracted, the category and the name of the feature.

4.2 Discussion and analysis

After a detailed exploration of the empirical findings in the preceding chapter, the next phase of the investigation involves a comprehensive discussion and analysis of the results. In this chapter, the observed outcomes are analysed, aiming to extract meaningful insights and implications. Navigating through the results, the focus is not only on summarizing the numerical outcomes but also on interpreting their significance in the context of the research objectives. The main questions raising from the previous chapter, that here will be addressed, are the following:

- *which Cox model training has provided the best regression*: between all the models that have been trained, also with different procedures and techniques, which is the one giving the best results and, at the same time, generalizing the data;

- *between CT and PET scans, which has the highest prognostic value*: each type of imaging technique is useful to store a specific type of information. One of them may be a scan for which the feature extraction evaluate features more descriptive about the disease progression

- *which segmentation shape selection is the most informative*: the shaping of the images with the original segmentation or a derivative shape is the one giving the highest results;

- *are filter essential or not, which filters are the most useful*: the feature extraction process can be carried out either on the original image or on the image resulting from the application of a filter. It is interesting to understand if the process of employing a filter is useful or the original image is already enough informative;

- *which features are the ones describing in a more complete way the information stored in the images*: the features extracted are mathematical properties evaluated on the image grey levels, and an interesting point is to understand which of them describe better the PFS outcome.

4.2.1 Best Cox regression model

The first question to answer is about the best model. In the present work, as thoroughly examined, the Cox Regression is the model employed to analyse the features extracted through the radiomics pipeline. Cox regression is a survival analysis model which aims to discern the relationship between one or more predictor variables and the time until a specific event of interest occurs. This paragraph delves into a comparative analysis of the various models trained, aiming to identify the optimal one based on its ability to provide accurate and reliable predictions in the context of our study. This discussion involves an in-depth exploration of each model's performance, and considering their respective strengths and limitations.

Firstly, the objective is to determine the most suitable method for training the Cox model among the two options, considering both its effectiveness on the available dataset and its ability to generalize to broader scenarios. For both methods the complete set of 6 datasets were used, and the results of the training are stored in Tab. 4.1 for the Cox model complete with all the variables and in Tab. 4.3 for the model featuring few significant covariates. It is evident that the mean C-index values obtained using the first method do not accurately reflect a well-suited model. A C-index value close to 0.5 indicates that the model's discriminatory power is no better than random chance. In other words, the model cannot distinguish between positive and negative cases, and its predictions are equivalent to random guessing. The highest mean C-index achieved through this methodology is (0.59 ± 0.09), attained from the CT dataset featuring PM regions.

However, it remains a value that exhibits only a marginal deviation from the neutral benchmark of 0.5. As already discussed in section 4.1.1, the reasons for the difficulty in achieving good results from Cox regression are multiple:

- The abundance of covariates that add noise to the model, despite the LASSO regularization used, which aims to regularize the variables to ensure that only the most significant ones influence the model.
- The collinearity of variables due to the fact that the *PyRadiomics* features belonging to the same group or family are extracted with mathematical formulas that depend on each other.
- Finally, from the comparison with the optimal results of the analysis conducted on the entire skeleton, it is evident that the method that involves retaining only the information concerning the spine in PET and CT images does not consider regions such as the marrow and PM zones of the long bones, which are more informative than the spine concerning the information on PFS.

Regarding the second method, the C-index values appear quite promising, surpassing those obtained with the previous method where the model predictions were essentially random. This outcome aligns with expectations, as the second method involves feature selection, focusing on the most significant features. The model, that undergoes evaluation through CV, mitigates issues associated with excessive feature abundance and collinearity. The trade-off, understandably, is reflected in the results, which do not reach values indicative of an exceptionally strong model. However, by utilizing a limited set of features, part of the information is sacrificed for the sake of model simplicity. Nevertheless, results from models constructed with 2 or 3 features are promising and sufficiently align with the utilized datasets, suggesting potential for good generalization. The highest C-index value is (0.75 ± 0.09) for the dataset derived from CT images, including the PM zone. The remaining values fall within the range of 0.65-0.75, with notable consistency observed across diverse datasets and imaging modalities. The consistency of these results across different datasets underscores the method's robustness. While the C-index values may not reach the pinnacle of model performance, the focus on essential features and the avoidance of overfitting contribute to the model's reliability. Moreover, the trade-off in predictive power is justified by the increased interpretability and generalization potential of the model.

Finally, being the second method the one giving the most promising results, an interesting thing is highlighting the absolute best model between the ones discovered during the process. The model representing the best performance is the one scoring (0.75 ± 0.09), which, as can be verified in Tab. 4.4, is composed of the subset of two features both coming from the original image without filters: Flatness and DependenceVariance. The

first is a feature of the family "shape" that evaluated on the structure and the geometry of the segmentation. The second one is a texture feature, it evaluates the gray level dependencies where the dependencies are defined as the number of connected voxels within distance δ that are dependent on the center voxel. This pair of features appears to be a suited combination to describe the dataset properly considering the for a survival analysis on the PFS information. It is interesting to analyze the single variables featuring in the resulting best models, therefore it will be discussed in the following sections.

4.2.2 Differences between CT and PET information

Shifting the focus to another pivotal aspect of the investigation, this work delves into the distinctions arising from the utilization of CT and PET images. Illustrated in Tab. 4.3, the C-index metrics stemming from models trained on each imaging modality prompt a nuanced exploration of their respective efficacy in the context of Cox regression. Initial observations hint at potential variations, with mean C-index values displaying fluctuations between the two imaging types. Looking at the image in Fig. 4.2 representing a box-plot graph of the resulting C-index of the two best models for CT and PET datasets, it is possible to notice that the two intervals are not completely overlapping. In fact, for the features extracted from CT images, the interval of the results obtained from the Cox model scoring is placed higher and seems to be considerably different from the PET counterpart, with a minimal overlapping of the box representing the middle quartiles. However, even though the difference between the two medical imaging techniques seems to be significant, it is not statistically meaningful. It is a hypothesis supported by the evidence of the results and also by the medical opinion, asserting that through CT images bone lesions typically provoked by MM are easily recognizable. They can be linked in CT images to discontinuities in bone tissues representing the lytic bone lesions that in paragraph 2.2 are discussed as an important risk factor. However, the available data is insufficient for asserting it with certainty. All the Student t-tests performed on the C-index results for verifying if the difference between PET and CT is consistent were revealed as not significant expressing a p-value higher than 0.05. It suggests that a more detailed and extended dataset may be necessary to verify this trend, and confirm the CT as the most informative image technique for survival analysis with Cox model.

4.2.3 Analysis on the segmentation shape

In the pursuit of refining prognostic models for MM survival analysis, this section delves into another comparative analysis of Cox models, particularly focusing on their performance variance when applied to different segmentation approaches. The primary focus revolves around the original spine-only segmentation and its enhanced counterparts, where additional areas surrounding the spine are incorporated through a meticulous process. As the lens narrows onto the models resulting from the latter method,

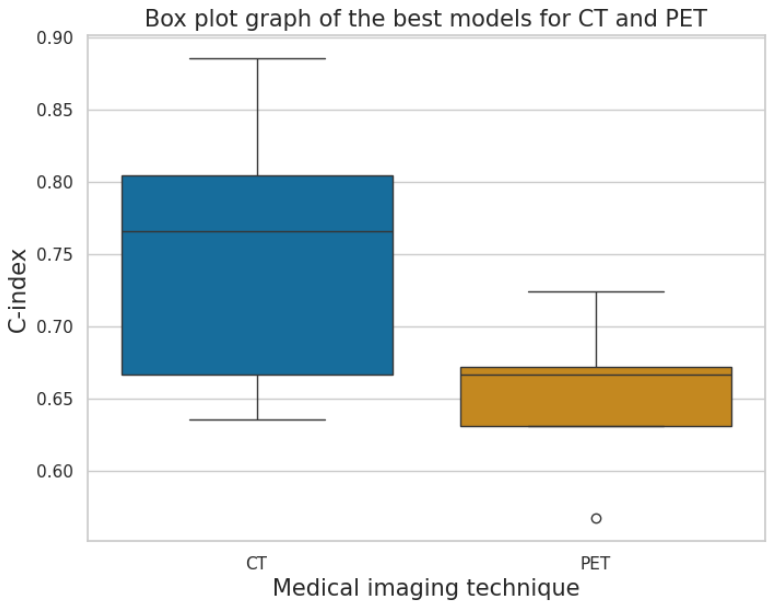
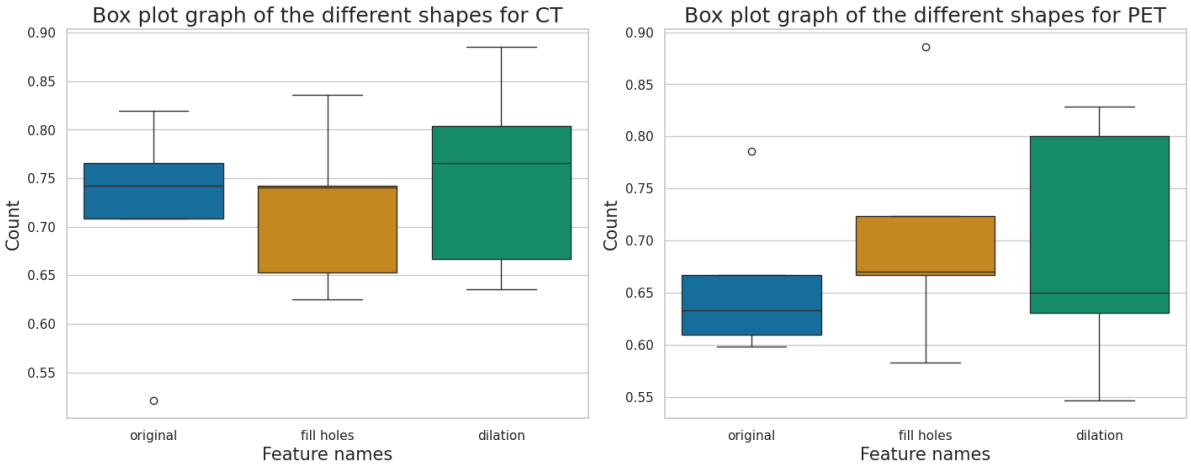


Figure 4.2: Boxplot graph representing the distribution of the C-index values obtained by the best models for CT (blue) and for PET (orange)

specifically the application of SR for the selection of key features, the discussion aims to unveil the relations between segmentation choices and model performance. The pivotal question guiding this exploration is whether the inclusion of supplementary regions beyond the spine, achieved through morphological operations on the original segmentation, contributes significantly to the predictive power of the Cox model.

Considering the C-index metrics resulting from the best models obtained for each of the 6 datasets, as reported in Tab. 4.3, the discussion will focus on determining which segmentation shape is optimal for obtaining informative data and thus achieving favorable outcomes in Cox regression. At first glance, it is apparent that the results obtained using features extracted when the information was solely confined within the spinal cord yield slightly lower mean C-index compared to cases where both the spinal cord and PM areas are considered. The latter appears to yield the best results. From an analysis of the box-plot graphs represented in Fig. 4.3, some divergences in the intervals of the C-index values are highlighted. For the values in the CT case there is a higher overlapping due to the slightly higher results, as discussed in the previous chapter. While for the PET case the higher results in case of feature extracted from an image considering medullary and PM areas. However, none of these differences are statistically significant. The distributions of C-index obtained in various cases with different segmentation shapes, when compared using a Student t-test, return p-values well above 0.05, confirming the null hypothesis that dismisses the possibility that they come from different distributions from



(a) (b)

Figure 4.3: Box-plot graph representing the distribution of the C-index values obtained by the best models for the shapes: original (blue), fill holes (orange), and dilation (green). On the left (a) are shown CT distributions of C-index and on the right (b) PET distributions.

a statistical perspective. However, as confirmed by Tab. 5.2, the mean C-index values of all evaluated models are in line with the assumption made initially. Consequently, with the data obtained, it is not possible to definitively confirm the trend that appears to influence the data given the significant variability in model scores, illustrated by the high standard deviation value. Nevertheless, there is a clear tendency that could be revealed with more precise and accurate results. This is also suggested by the knowledge base provided by the medical literature. Indeed, many of the scores used for MM patients, such as R-ISS or IMPeTUs criteria outlined in paragraph 2.2, to assess the prognostic advancement of the disease include the verification of lesions in the PM area as an important factor for judgment.

4.2.4 Most informative filters

Another insightful analysis concerns the examination of additional elements that characterize the models beyond the results, namely, more significant filters and selected features. It is crucial to note that when dealing with datasets composed of 982 features derived from the same image subjected to various filters before feature extraction, there is a risk of encountering high collinearity and an abundance of features. Moreover, not all features are necessarily significant or contribute essential characteristics from the perspective of the Cox model applied to progression-free survival.

Conducting a study that allows for a focus on specific properties deemed most significant becomes crucial. This approach aims to provide clinicians with potentially valuable insights. Ideally, identifying certain properties that, when corroborated by medical literature on this type of disease, receive confirmation of their importance, opens new avenues for analysis. From the graph in Fig. 4.4, the histogram of all features extracted in the 60 trained models is observable after the selection process through SR, a method explained in Section 4.1.2. The graph encompasses numerous variables; for instance, different colors represent the type of filter: "original", denoting the original image, "wavelet filter", or "laplacian of gaussian". Another aspect to note is the diverse family of features represented by the name or acronym preceding the feature name. It is evident that features extracted from the image without a filter, represented by "original", are the most prevalent, given that each filter pair used for selection is a combination of 'original' and another of the 10 possible filters. The initial features in terms of occurrences appearing in the graph belong to the "shape" family, prominently represented in most of the 60 models, such as Flatness and Sphericity. Another interesting point to note is the significant percentage of wavelet filters, indicating a potential utility in terms of the information they contribute.

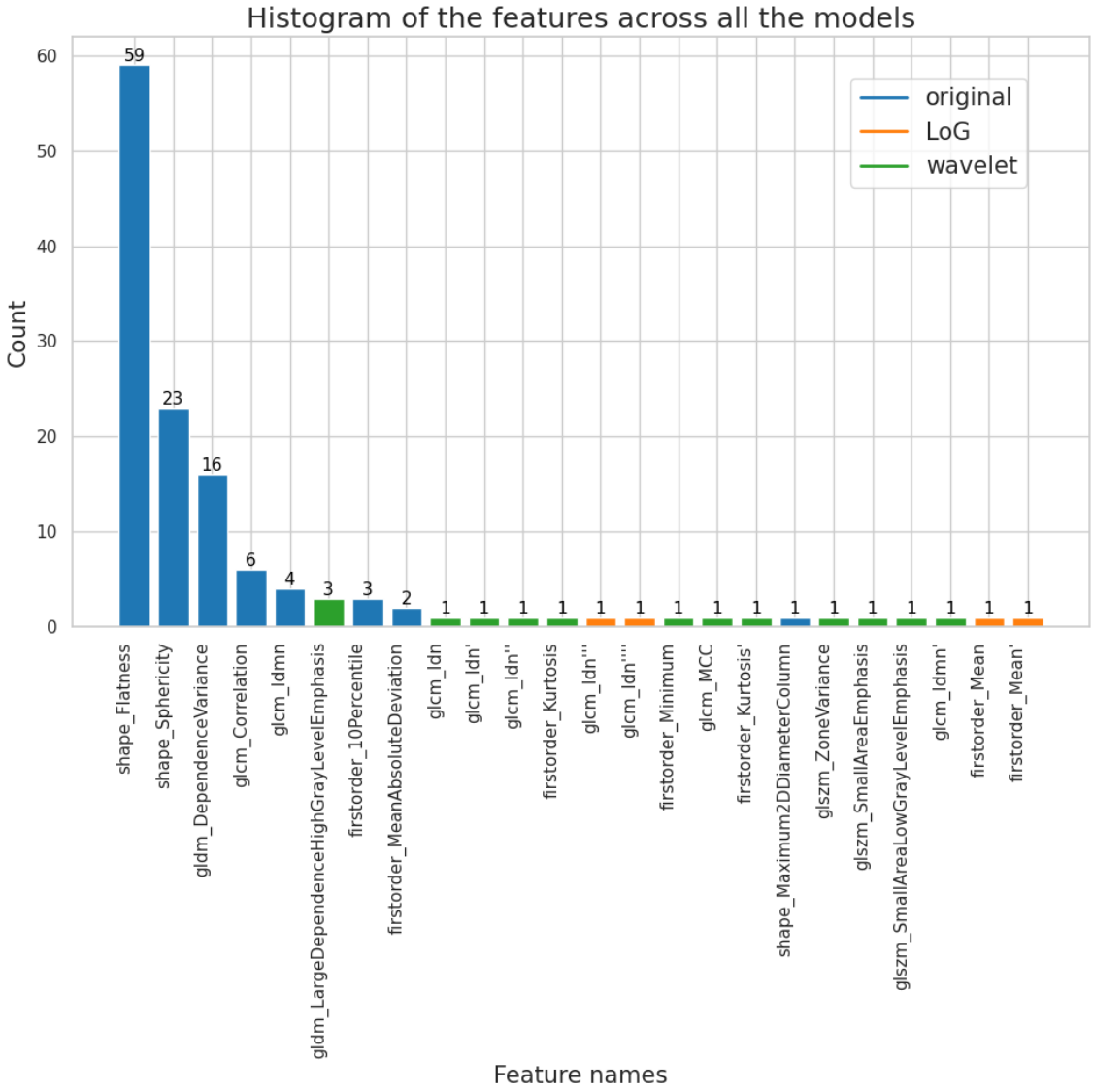


Figure 4.4: Histogram of the features selected for the 60 models. Each color represents the filter employed before the feature extraction on the image and the different colors are represented in the legend.

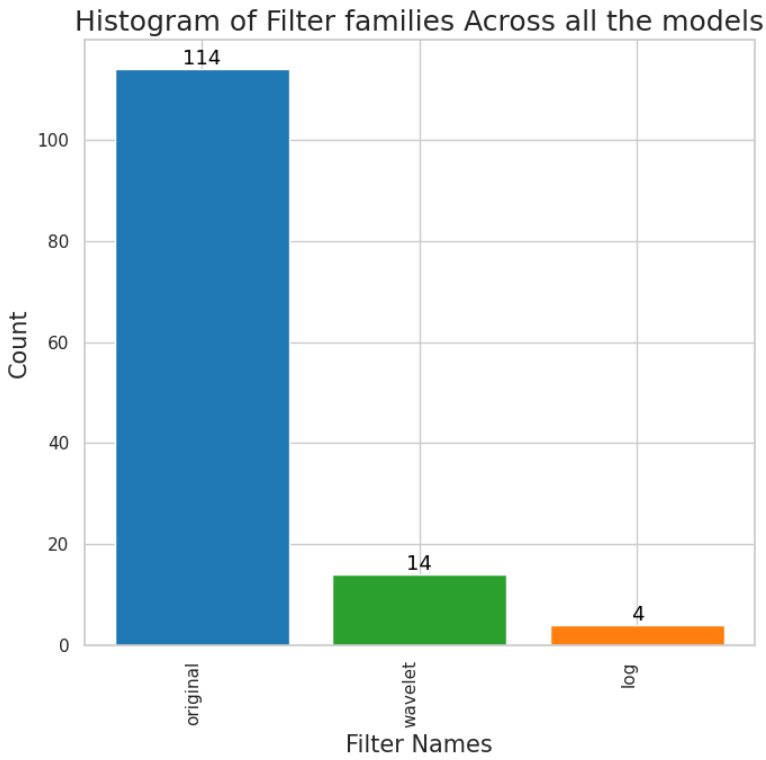


Figure 4.5: Histogram of the features selected for the 60 models. Each color represents the filter employed before the feature extraction on the image and the different colors are represented in the legend.)

The graph in Fig. 4.5 specifically illustrates the frequency with which a filter is used in the model, through the selected feature extracted from it. It is notable that the most utilized filter is the original image, from which, in addition to "shape" features, features from other families are also selected. Features extracted from the original image, although considered in every model, exhibit a high frequency of occurrence, suggesting that they represent particularly informative images. Additionally, it is evident that while the wavelet filter has a significantly lower number of occurrences compared to the previous one, it appears a significant number of times compared to the third, which is selected by the algorithm only twice. In order to verify this trend it is useful to focus on the best model selected for every dataset. With a subset of models, representing the most informative features for every dataset, it is possible to verify if the LoG results as an informative source of features. In Fig. 4.6 is represented this type of histogram, with a total of 13 features selected. The confirmation of the previously stated observations is evident; no LoG filter is utilized in models with higher performance. The reasons for this phenomenon might be found in the explanation of the filter's function and its suitability for specific purposes. Multiple myeloma images can encompass various structures, in-

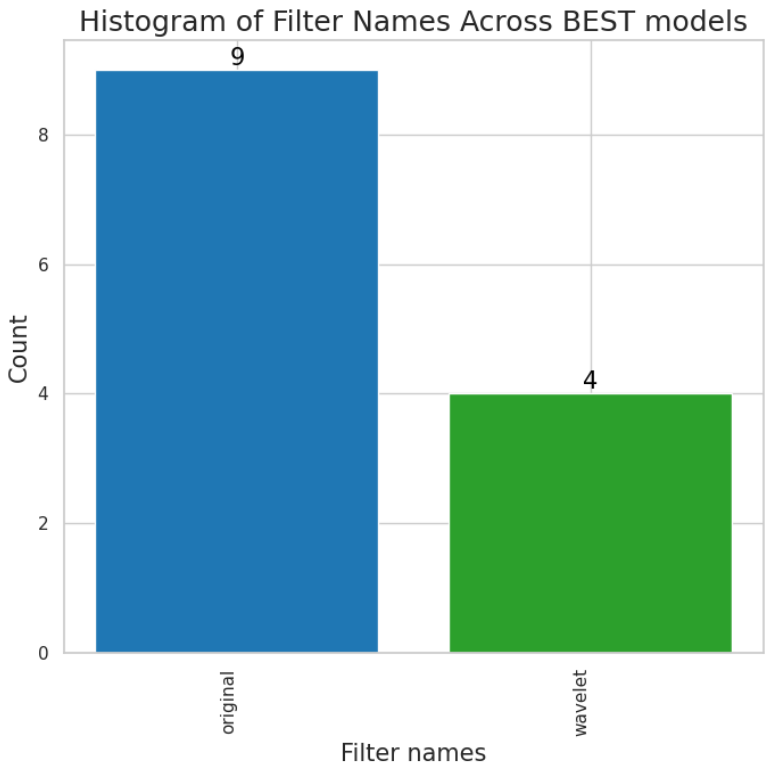


Figure 4.6: Histogram of the features selected for the 60 models. Each color represents the filter employed before the feature extraction on the image and the different colors are represented in the legend.)

cluding bone lesions, soft tissue, and other features. Wavelet filters, with their ability to analyze structures at different frequency scales, could prove useful in capturing details at various resolution levels. Additionally, wavelet filters may exhibit greater robustness to image noise as they segregate information across different scales. On the other hand, the LoG filter might be more sensitive to noise, especially in the presence of high-frequency noise in the images, although it excels in capturing small details or edge structures.

4.2.5 Most descriptive features

In this chapter, we embark on an in-depth exploration of the feature frequency and the pivotal features employed in the models derived from our algorithmic approach. As this work delves into the intricacies of the model training process, the focus shifts towards understanding the prevalence of specific features and their impact on model performance. The graphical representations, as exemplified in Fig. 4.4, provide a comprehensive overview of the feature landscape, illustrating the frequencies and utilization

patterns of various filters and families of features. Our analysis aims to decipher the significance of each feature, shedding light on their individual contributions to the predictive prowess of the models. Through this exploration, this work seeks to unravel the underlying reasons guiding the feature selection process and its implications for enhancing the interpretability and effectiveness of the prognostic models for MM.

As mentioned earlier, the most prominently featured selections in our chosen feature sets are those of Flatness. It belongs to the shape family, representing geometric properties of the segmentation. Flatness, expressing the relationship between the longest and shortest principal components, is calculated using the following formula:

$$\text{flatness} = \sqrt{\frac{\lambda_{\text{least}}}{\lambda_{\text{major}}}} \tag{4.1}$$

where λ_{least} and λ_{major} represent the lengths of the two components. This ratio captures the dimensions of the spine, signifying its length relative to its width, essentially describing how elongated or wide it is. The positive values of β associated with this feature in every Cox model where it appears indicate that it contributes to an increased hazard, favoring cases where the spine exhibits a more elongated characteristic.

As previously discussed, among the prominently featured selections within our chosen feature sets, the DependenceVariance feature stands out. This particular feature, encapsulated within the Gray Level Dependence Matrix (GLDM) family, provides valuable insights into the spatial relationships and dependencies within the image. DependenceVariance calculates the variance of dependence, contrast, in the GLDM and is derived from the formula:

$$DependenceVariance = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)(i-\mu)^2 \tag{4.2}$$

where N_g is the number of gray levels, $p(i,j)$ represents the elements of the GLDM, and μ is the mean of the GLDM. This feature captures the variability in dependence between different gray levels, providing a measure of the contrast variations within the image.

In the context of Cox models, the association of DependenceVariance with a β value signifies its impact on the hazard. Interpretation of this feature involves understanding how changes in the spatial relationships and variations in gray-level dependence contribute to the predictive ability of the model. A $\beta < 0$ value associated with this feature implies a property that decreases the hazard in the Cox model. The feature likely encapsulates information about the spatial relationships and patterns within the image. A smaller DependenceVariance might indicate more homogeneous and predictable spatial

dependencies, which, in turn, may be associated with a more predictable or stable risk profile.

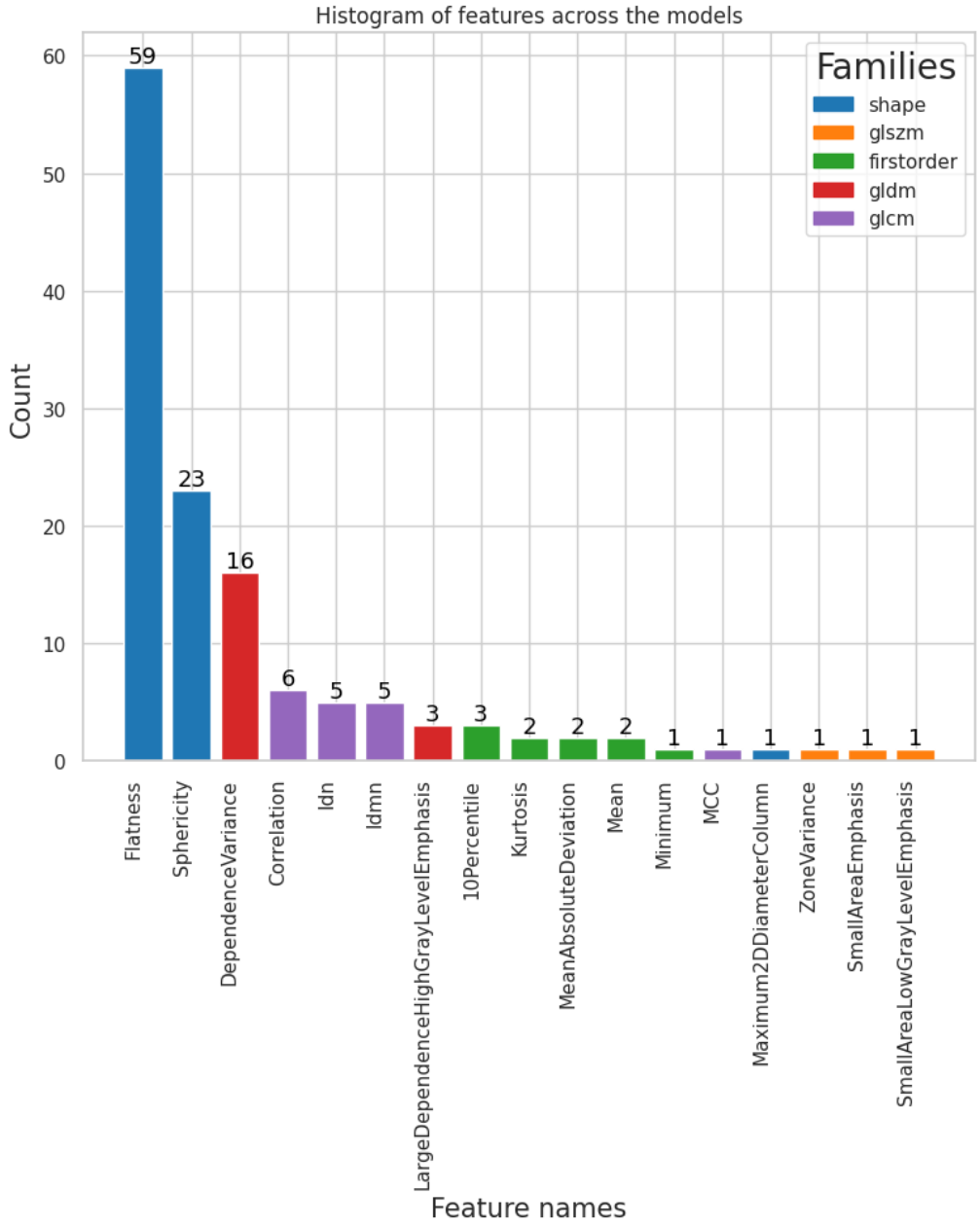


Figure 4.7: Histogram of the features selected for the 60 models. Each color represents the filter employed before the feature extraction on the image and the different colors are represented in the legend.

Both are represented in the graph of Fig. 4.7, as it represents a count of all the features. However for analyzing the other features it may be interesting to take into consideration another graph, an histogram of all the features selected from the resulting best models for each dataset.

In Fig. 4.8 is represented the histogram of all the features from the best model in each dataset. Here we can see the Flatness as the most represented feature, DependenceVariance as the second one and a few of other features, mostly texture features.

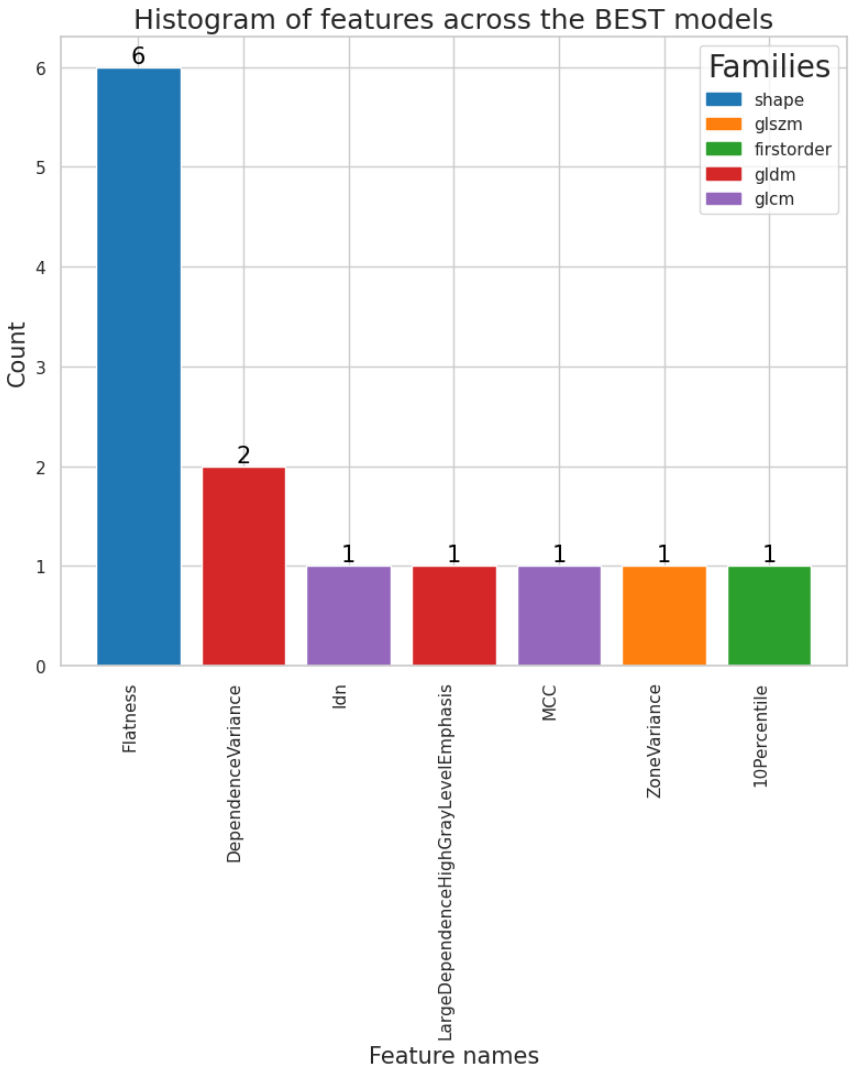


Figure 4.8: Histogram of the features selected for the 60 models. Each color represents the filter employed before the feature extraction on the image and the different colors are represented in the legend.

An intriguing pattern emerges examining models achieving high performance within our dataset. Notably, in every model exhibiting superior predictive accuracy, there is a consistent presence of at least one shape feature, predominantly Flatness, and a texture feature. This recurrent combination of features, often comprised of geometric properties and textural intricacies, proves to be a robust predictor for the PFS in our Cox models. The inclusion of a shape feature like Flatness suggests that the geometric characteristics of the spine play a pivotal role in predicting the PFS. Specifically, the elongation or compression of the spine, as captured by Flatness, appears to hold significant prognostic value. Moreover, the consistent incorporation of a texture feature underscores the importance of textural nuances in the images. These textural features provide insights into the spatial heterogeneity and patterns within the images, contributing crucial information for prognostic assessments.

This recurring pattern prompts the consideration that, for our dataset, the combination of a shape feature and a texture feature is often sufficient for robust predictions of progression-free survival in Cox models. The interplay between geometric attributes and textural intricacies captures nuanced information crucial for understanding disease progression, highlighting the significance of these features in the prognostic landscape of MM.

Chapter 5

Conclusions and future lines

The aim of this study was to unravel the deep relationships between imaging-derived biomarkers and patient outcomes, employing various methods and leveraging the Cox model for the analysis of the Progression-Free Survival (PFS) in Multiple Myeloma (MM) patients. PFS is the duration from treatment start until disease progression or death occurs for the patient, and it is crucial for obtaining insights on the disease progression. The ultimate goal, in fact, was to obtain valuable insights that could help clinicians understand more deeply the nature of this type of hematologic cancer, enhancing the clinical-decision support systems (CDSS).

A cohort of 93 patients suffering from MM, with CT and PET images and clinical information, was employed in this work. A segmentation of the spine was obtained on CT with MOOSE software which was proved to provide accurate bone segmentations, then it was manipulated to provide three different RoI selection: the original spine segmentation, the spine with inclusion of the bone marrow region inside, and a region comprehensive of spine, bone marrow and adjacent regions. The segmentations were used on the images to remove the part outside and keep just the information they enclose. From each image and segmentation, standard numerical features were extracted by PyRadiomics, after employing three types of process on the image: using the original image, applying a set of wavelet filters, and LoG filters. Survival analysis with the PFS information is performed through LASSO penalized Cox regression model with two different methods.

In the first method a nested Cross-Validation (nCV) on the whole 982 features dataset is implemented, where the inner CV is used to set the most suited penalization parameter, and the outer CV for obtaining a more robust and reliable model. From this method poor results were obtained, with a near random regression discriminatory power, in the range [0.50, 0.59]. This proves that limited information can be extracted from the spine region only, in comparison with the substantially higher results obtained considering the complete skeleton obtained by Peluso et al. [60].

For the second method a feature selection was performed implementing a Stepwise Regression (SR) in a "forward selection" fashion where for each combination selected of filter's feature. This process allowed obtaining higher results, in the range [0.65,0.75], with fewer features and also insights on the selections of the most informative ones for the analysis. One of the first result is the trend for which from CT features the model has a higher performance than for PET, as the literature confirms for the presence of distinguishable patterns describing bone lesions. However, it was not possible to confirm the result with a statistically valuable verification for the high variance of the data lead by the few patients available in the research. The same comparison was performed for the different segmentation regions. It highlighted a substantial overlapping in the distributions of the results, leading to the conclusion that it is not possible to verify a significant difference with the high variance of the results with the present cohort. Furthermore, other interesting analysis was performed on the feature selected. It highlighted that between the filters applied on the images the LoG filter was the one from which the lower number of significant features were selected. It was confirmed also considering the best models for the two imaging techniques, and the three shapes, in this case none of the feature selected came from the LoG image. The last analysis was performed on the single features. Features from the "shape" family were the most frequently selected, with the "Flatness" feature indicating the longitudinal nature with respect to the thickness of the spine that was employed in all the resulting best models. Another type of significant feature revealed is the DependenceVariance, a texture feature index of the homogeneity of the spatial grey-level dependencies, featuring on the two best models in the whole analysis. Finally, an interesting pattern highlighted, is the high performance resulting for models where a combination of a shape feature is selected together with a texture feature. With such selection the resulting model appears to have a high regression performance, even though it employs just two features from the complete set.

Numerous developments can be brought to this work. Firstly, a bigger cohort can refine the results. It can allow discriminating between the performance of two different methods employed, in order to have more statistically meaningful results and be able to select with certainty the best pipeline to path. This is a study designed specifically to enhance the CDSS and the support given to clinicians in the MM treatment must have the lowest margin of error. Moreover, the analysis was performed on the PFS information and many others prognostic valuable biomarkers could have been used for highlighting different kind of pattern. A notable enhancement involves incorporating clinicians' annotations on disease progression and integrating genomic data. This comprehensive approach aims to elevate the depth of the analysis, enhance the understanding of MM and most importantly improving the welfare of the patient.

Bibliography

[1] Hugo J. W. L. Aerts et al. "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach". en. In: *Nature Communications* 5.1 (June 2014). Number: 1 Publisher: Nature Publishing Group, p. 4006. ISSN: 2041-1723. DOI: 10.1038/ncomms5006. URL: https://www.nature.com/articles/ncomms5006 (visited on 01/11/2024).

[2] Ankit Agarwal et al. "Evolving role of FDG PET/CT in multiple myeloma imaging and management". eng. In: *AJR. American journal of roentgenology* 200.4 (Apr. 2013), pp. 884–890. ISSN: 1546-3141. DOI: 10.2214/AJR.12.9653.

[3] Pierpaolo Alongi et al. "Radiomics Analysis of Brain [18F]FDG PET/CT to Predict Alzheimer's Disease in Patients with Amyloid PET Positivity: A Preliminary Report on the Application of SPM Cortical Segmentation, Pyradiomics and Machine-Learning Analysis". en. In: *Diagnostics* 12.4 (Apr. 2022). Number: 4 Publisher: Multidisciplinary Digital Publishing Institute, p. 933. ISSN: 2075-4418. DOI: 10.3390/diagnostics12040933. URL: https://www.mdpi.com/2075-4418/12/4/933 (visited on 02/21/2024).

[4] Clemens Arndt et al. "Deep Learning CT Image Reconstruction in Clinical Practice". en. In: *RöFo - Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren* 193.3 (Mar. 2021). Publisher: Georg Thieme Verlag KG, pp. 252–261. ISSN: 1438-9029, 1438-9010. DOI: 10.1055/a-1248-2556. URL: http://www.thieme-connect.de/DOI/DOI?10.1055/a-1248-2556 (visited on 01/08/2024).

[5] Daniel Berrar. "Cross-Validation". In: Jan. 2018. ISBN: 978-0-12-809633-8. DOI: 10.1016/B978-0-12-809633-8.20349-X.

[6] Ronald Boellaard et al. "FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0". eng. In: *European Journal of Nuclear Medicine and Molecular Imaging* 42.2 (Feb. 2015), pp. 328–354. ISSN: 1619-7089. DOI: 10.1007/s00259-014-2961-x.

[7] Yuri Boykov and Gareth Funka-Lea. "Graph Cuts and Efficient N-D Image Segmentation". en. In: *International Journal of Computer Vision* 70.2 (Nov. 2006), pp. 109–131. ISSN: 1573-1405. DOI: 10.1007/s11263-006-7934-5. URL: https://doi.org/10.1007/s11263-006-7934-5 (visited on 02/05/2024).

[8] Harold J. Breaux. "A modification of Efroymson's technique for stepwise regression analysis". In: *Communications of the ACM* 11.8 (1968), pp. 556–558. ISSN: 0001-0782. DOI: 10.1145/363567.363580. URL: https://dl.acm.org/doi/10.1145/363567.363580 (visited on 02/23/2024).

[9] Adam R Brentnall and Jack Cuzick. "Use of the concordance index for predictors of censored survival data". en. In: *Statistical Methods in Medical Research* 27.8 (Aug. 2018). Publisher: SAGE Publications Ltd STM, pp. 2359–2373. ISSN: 0962-2802. DOI: 10.1177/0962280216680245. URL: https://doi.org/10.1177/0962280216680245 (visited on 02/09/2024).

[10] Isaac S. Chua et al. "Artificial intelligence in oncology: Path to implementation". en. In: *Cancer Medicine* 10.12 (2021). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ca pp. 4138–4149. ISSN: 2045-7634. DOI: 10.1002/cam4.3935. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cam4.3935 (visited on 12/14/2023).

[11] *Computational Radiomics System to Decode the Radiographic Phenotype - PubMed*. URL: https://pubmed.ncbi.nlm.nih.gov/29092951/ (visited on 02/21/2024).

[12] *Cox Model Assumptions - Easy Guides - Wiki - STHDA*. en. URL: http://www.sthda.com/english/wiki/cox-model-assumptions (visited on 02/27/2024).

[13] Francesco Cremonesi et al. "The need for multimodal health data modeling: A practical approach for a federated-learning healthcare platform". In: *Journal of Biomedical Informatics* 141 (May 2023), p. 104338. ISSN: 1532-0464. DOI: 10.1016/j.jbi.2023.104338. URL: https://www.sciencedirect.com/science/article/pii/S153204642300059X (visited on 12/15/2023).

[14] Cameron Davidson-Pilon. "lifelines: survival analysis in Python". en. In: *Journal of Open Source Software* 4.40 (Aug. 2019), p. 1317. ISSN: 2475-9066. DOI: 10.21105/joss.01317. URL: https://joss.theoj.org/papers/10.21105/joss.01317 (visited on 02/23/2024).

[15] dfwg. *NIfTI: — Neuroimaging Informatics Technology Initiative*. en. Document. URL: https://nifti.nimh.nih.gov/ (visited on 02/12/2024).

[16] *DICOM*. en. URL: https://www.dicomstandard.org (visited on 02/13/2024).

[17] *Discrete Wavelet Transform - an overview — ScienceDirect Topics*. URL: https://www.sciencedirect.com/topics/mathematics/discrete-wavelet-transform (visited on 03/04/2024).

[18] Vinayak V. Dixit, Sai Chand, and Divya J. Nair. "Autonomous Vehicles: Disengagements, Accidents and Reaction Times". In: *PLOS ONE* 11.12 (Dec. 2016). Publisher: Public Library of Science, pp. 1–14. DOI: 10.1371/journal.pone.0168054. URL: https://doi.org/10.1371/journal.pone.0168054.

[19] M. A. Effroymson. *Multiple regression analysis. Mathematical Methods for Digital Computers, Ed. A. Ralson and HS Wilf*. 1960.

[20] Olivier Elemento et al. "Artificial intelligence in cancer research, diagnosis and therapy". en. In: *Nature Reviews Cancer* 21.12 (Dec. 2021). Number: 12 Publisher: Nature Publishing Group, pp. 747–752. ISSN: 1474-1768. DOI: 10.1038/s41568-021-00399-1. URL: https://www.nature.com/articles/s41568-021-00399-1 (visited on 12/14/2023).

[21] Thorsten Falk et al. "U-Net: deep learning for cell counting, detection, and morphometry". en. In: *Nature Methods* 16.1 (Jan. 2019). Number: 1 Publisher: Nature Publishing Group, pp. 67–70. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0261-2. URL: https://www.nature.com/articles/s41592-018-0261-2 (visited on 01/11/2024).

[22] E. Fermi. "An attempt of a theory of beta radiation. 1." In: *Z. Phys.* 88 (1934), pp. 161–177. DOI: 10.1007/BF01351864.

[23] *Genomed4All*. en-US. URL: https://elixir-italy.org/project/genomed4all/ (visited on 02/05/2024).

[24] *GenoMed4All - Genomics For Next Generation Healthcare*. en-US. URL: http://genomed4all.eu/ (visited on 02/05/2024).

[25] Brandon George, Samantha Seals, and Inmaculada Aban. "Survival analysis and regression models". en. In: *Journal of Nuclear Cardiology* 21.4 (Aug. 2014), pp. 686–694. ISSN: 1532-6551. DOI: 10.1007/s12350-014-9908-2. URL: https://doi.org/10.1007/s12350-014-9908-2 (visited on 02/06/2024).

[26] Robert J. Gillies, Paul E. Kinahan, and Hedvig Hricak. "Radiomics: Images Are More than Pictures, They Are Data". eng. In: *Radiology* 278.2 (Feb. 2016), pp. 563–577. ISSN: 1527-1315. DOI: 10.1148/radiol.2015151169.

[27] Manish Kumar Goel, Pardeep Khanna, and Jugal Kishore. "Understanding survival analysis: Kaplan-Meier estimate". en. In: *International Journal of Ayurveda Research* 1.4 (Dec. 2010). Publisher: Wolters Kluwer – Medknow Publications, p. 274. ISSN: 0974-7788. DOI: 10.4103/0974-7788.76794. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3059453/ (visited on 01/15/2024).

[28] John W. Goodell et al. “Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis”. In: *Journal of Behavioral and Experimental Finance* 32 (Dec. 2021), p. 100577. ISSN: 2214-6350. DOI: 10.1016/j.jbef.2021.100577. URL: https://www.sciencedirect.com/science/article/pii/S2214635021001210 (visited on 12/14/2023).

[29] Sathish Gopalakrishnan et al. “Revised-International Staging System (R-ISS) is Predictive and Prognostic for Early Relapse (¡24 months) after Autologous Transplantation for Newly Diagnosed Multiple Myeloma (MM).” In: *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation* 25.4 (Apr. 2019), pp. 683–688. ISSN: 1083-8791. DOI: 10.1016/j.bbmt.2018.12.141. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6453717/ (visited on 02/09/2024).

[30] F. E. Harrell et al. “Evaluating the yield of medical tests”. eng. In: *JAMA* 247.18 (May 1982), pp. 2543–2546. ISSN: 0098-7484.

[31] Akinori Hata et al. “Effect of Matrix Size on the Image Quality of Ultra-high-resolution CT of the Lung: Comparison of 512×512, 1024×1024, and 2048×2048”. eng. In: *Academic Radiology* 25.7 (July 2018), pp. 869–876. ISSN: 1878-4046. DOI: 10.1016/j.acra.2017.11.017.

[32] C. F. Healy et al. “Multiple Myeloma: A Review of Imaging Features and Radiological Techniques”. en. In: *Bone Marrow Research* 2011 (Aug. 2011). Publisher: Hindawi, e583439. ISSN: 2090-2999. DOI: 10.1155/2011/583439. URL: https://www.hindawi.com/journals/bmr/2011/583439/ (visited on 01/18/2024).

[33] Todd C. Hollon et al. “Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks”. en. In: *Nature Medicine* 26.1 (Jan. 2020). Number: 1 Publisher: Nature Publishing Group, pp. 52–58. ISSN: 1546-170X. DOI: 10.1038/s41591-019-0715-9. URL: https://www.nature.com/articles/s41591-019-0715-9 (visited on 01/11/2024).

[34] Jiang Hsieh et al. “Recent Advances in CT Image Reconstruction”. en. In: *Current Radiology Reports* 1.1 (Mar. 2013), pp. 39–51. ISSN: 2167-4825. DOI: 10.1007/s40134-012-0003-7. URL: https://doi.org/10.1007/s40134-012-0003-7 (visited on 01/08/2024).

[35] *International Staging System (ISS) and Revised ISS (R-ISS)*. en-US. URL: https://www.myeloma.org/international-staging-system-iss-reivised-iss-r-iss (visited on 02/14/2024).

[36] Fabian Isensee et al. “nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation”. en. In: *Nature Methods* 18.2 (Feb. 2021). Number: 2 Publisher: Nature Publishing Group, pp. 203–211. ISSN: 1548-7105. DOI: 10.1038/s41592-020-01008-z. URL: https://www.nature.com/articles/s41592-020-01008-z (visited on 01/10/2024).

[37] Fabian Isensee et al. “nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation”. In: *ArXiv* (Sept. 2018). URL: https://www.semanticscholar.org/paper/nnU-Net%3A-Self-adapting-Framework-for-U-Net-Based-Isensee-Petersen/849e99219259a2cc28209b02d0c646b7183dd40e (visited on 01/10/2024).

[38] Yuchen Jiang et al. “Quo vadis artificial intelligence?” en. In: *Discover Artificial Intelligence* 2.1 (Mar. 2022), p. 4. ISSN: 2731-0809. DOI: 10.1007/s44163-022-00022-8. URL: https://doi.org/10.1007/s44163-022-00022-8 (visited on 12/14/2023).

[39] Junita Rachel John et al. “Validation of Deauville Score for Response Evaluation in Hodgkin’s Lymphoma”. In: *Indian Journal of Nuclear Medicine : IJNM : The Official Journal of the Society of Nuclear Medicine, India* 38.1 (2023), pp. 16–22. ISSN: 0972-3919. DOI: 10.4103/ijnm.ijnm_102_22. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10171757/ (visited on 02/11/2024).

[40] *JSON*. URL: https://www.json.org/json-en.html (visited on 02/13/2024).

[41] Haijo Jung. “Basic Physical Principles and Clinical Applications of Computed Tomography”. In: *Progress in Medical Physics* 32 (Mar. 2021), pp. 1–17. DOI: 10.14316/pmp.2021.32.1.1.

[42] Willi A. Kalender. “X-ray computed tomography”. en. In: *Physics in Medicine & Biology* 51.13 (June 2006), R29. ISSN: 0031-9155. DOI: 10.1088/0031-9155/51/13/R03. URL: https://dx.doi.org/10.1088/0031-9155/51/13/R03 (visited on 01/08/2024).

[43] Eleni Kalopesa et al. “Rapid Determination of Wine Grape Maturity Level from pH, Titratable Acidity, and Sugar Content Using Non-Destructive In Situ Infrared Spectroscopy and Multi-Head Attention Convolutional Neural Networks”. en. In: *Sensors* 23.23 (Jan. 2023). Number: 23 Publisher: Multidisciplinary Digital Publishing Institute, p. 9536. ISSN: 1424-8220. DOI: 10.3390/s23239536. URL: https://www.mdpi.com/1424-8220/23/23/9536 (visited on 02/23/2024).

[44] Joel S. Karp et al. “PennPET Explorer: Design and Preliminary Performance of a Whole-Body Imager”. eng. In: *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine* 61.1 (Jan. 2020), pp. 136–143. ISSN: 1535-5667. DOI: 10.2967/jnumed.119.229997.

[45] Hatem Kaseb, Pavan Annamaraju, and Hani M. Babiker. "Monoclonal Gammopathy of Undetermined Significance". eng. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2023. URL: http://www.ncbi.nlm.nih.gov/books/NBK507880/ (visited on 01/17/2024).

[46] John P. Klein and Prem K. Goel, eds. *Survival Analysis: State of the Art*. Dordrecht: Springer Netherlands, 1992. ISBN: 978-90-481-4133-3 978-94-015-7983-4. DOI: 10.1007/978-94-015-7983-4. URL: http://link.springer.com/10.1007/978-94-015-7983-4 (visited on 02/06/2024).

[47] Michail E. Klontzas et al. "Radiomics Analysis for Multiple Myeloma: A Systematic Review with Radiomics Quality Scoring". en. In: *Diagnostics* 13.12 (Jan. 2023). Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, p. 2021. ISSN: 2075-4418. DOI: 10.3390/diagnostics13122021. URL: https://www.mdpi.com/2075-4418/13/12/2021 (visited on 01/18/2024).

[48] Gary D. Knott. *Interpolating Cubic Splines*. en. Boston, MA: Birkhäuser, 2000. ISBN: 978-1-4612-7092-8 978-1-4612-1320-8. DOI: 10.1007/978-1-4612-1320-8. URL: http://link.springer.com/10.1007/978-1-4612-1320-8 (visited on 03/04/2024).

[49] Philippe Lambin et al. "Radiomics: the bridge between medical imaging and personalized medicine". en. In: *Nature Reviews Clinical Oncology* 14.12 (Dec. 2017). Number: 12 Publisher: Nature Publishing Group, pp. 749–762. ISSN: 1759-4782. DOI: 10.1038/nrclinonc.2017.141. URL: https://www.nature.com/articles/nrclinonc.2017.141 (visited on 11/30/2023).

[50] Geert Litjens et al. "A Survey on Deep Learning in Medical Image Analysis". In: *Medical Image Analysis* 42 (Dec. 2017). arXiv:1702.05747 [cs], pp. 60–88. ISSN: 13618415. DOI: 10.1016/j.media.2017.07.005. URL: http://arxiv.org/abs/1702.05747 (visited on 01/11/2024).

[51] QingLin Liu et al. "Prediction of Aneurysm Stability Using a Machine Learning Model Based on PyRadiomics-Derived Morphological Features". In: *Stroke* 50.9 (Sept. 2019). Publisher: American Heart Association, pp. 2314–2321. DOI: 10.1161/STROKEAHA.119.025777. URL: https://www.ahajournals.org/doi/full/10.1161/STROKEAHA.119.025777 (visited on 02/21/2024).

[52] Marius E. Mayerhoefer et al. "Introduction to Radiomics". In: *Journal of Nuclear Medicine* 61.4 (Apr. 2020), pp. 488–495. ISSN: 0161-5505. DOI: 10.2967/jnumed.118.222893. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9374044/ (visited on 12/18/2023).

[53] Constantine S. Mitsiades et al. "Focus on multiple myeloma". In: *Cancer Cell* 6.5 (Nov. 2004), pp. 439–444. ISSN: 1535-6108. DOI: 10.1016/j.ccr.2004.10.020. URL: https://www.sciencedirect.com/science/article/pii/S1535610804003150 (visited on 02/14/2024).

[54] Cristina Nanni et al. "Image interpretation criteria for FDG PET/CT in multiple myeloma: a new proposal from an Italian expert panel. IMPeTUs (Italian Myeloma criteria for PET USe)". eng. In: *European Journal of Nuclear Medicine and Molecular Imaging* 43.3 (Mar. 2016), pp. 414–421. ISSN: 1619-7089. DOI: 10.1007/s00259-015-3200-9.

[55] Cristina Nanni et al. "Interpretation criteria for FDG PET/CT in multiple myeloma (IMPeTUs): final results. IMPeTUs (Italian myeloma criteria for PET USe)". en. In: *European Journal of Nuclear Medicine and Molecular Imaging* 45.5 (May 2018), pp. 712–719. ISSN: 1619-7089. DOI: 10.1007/s00259-017-3909-8. URL: https://doi.org/10.1007/s00259-017-3909-8 (visited on 02/09/2024).

[56] *Nuclear Medicine*. en. URL: https://www.nibib.nih.gov/science-education/science-topics/nuclear-medicine (visited on 01/08/2024).

[57] D. B. Owen et al. *On the history of statistics and probability: proceedings of a symposium on the American mathematical heritage, to celebrate the bicentennial of the United States of America, held at Southern Methodist University, May 27-29, 1974*. eng. Open Library ID: OL5206547M. New York: M. Dekker, 1976. ISBN: 978-0-8247-6390-9.

[58] Antonio Palumbo and Kenneth Anderson. "Multiple Myeloma". In: *New England Journal of Medicine* 364.11 (Mar. 2011). Publisher: Massachusetts Medical Society _eprint: https://doi.org/10.1056/NEJMra1011442, pp. 1046–1060. ISSN: 0028-4793. DOI: 10.1056/NEJMra1011442. URL: https://doi.org/10.1056/NEJMra1011442 (visited on 01/16/2024).

[59] Willem G. van Panhuis et al. "A systematic review of barriers to data sharing in public health". In: *BMC Public Health* 14.1 (Nov. 2014), p. 1144. ISSN: 1471-2458. DOI: 10.1186/1471-2458-14-1144. URL: https://doi.org/10.1186/1471-2458-14-1144 (visited on 12/15/2023).

[60] Sara Peluso. "Survival analysis in radiomics: a statistical learning approach on 18F-FDG-PET/CT imaging data of patients affected by multiple myeloma". en. Tesi di laurea. Dec. 2023. URL: https://amslaurea.unibo.it/30555/ (visited on 02/20/2024).

[61] Jie Peng et al. "Development and validation of a pyradiomics signature to predict initial treatment response and prognosis during transarterial chemoembolization in hepatocellular carcinoma". In: *Frontiers in Oncology* 12 (2022). ISSN: 2234-943X.

URL: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2022.853254 (visited on 02/21/2024).

[62] *PET Scan vs. CT Scan*. en. June 2021. URL: https://www.neurologica.com/blog/pet-scan-vs-ct-scan (visited on 02/14/2024).

[63] Aleš Procházka, Lucie Grajciarova, and Oldřich Vyšata. *Three-Dimensional Wavelet Transform in Multi-Dimensional Biomedical Volume Processing*. July 2011.

[64] *QIMP-Team/MOOSE*. original-date: 2022-03-24T11:48:24Z. Jan. 2024. URL: https://github.com/QIMP-Team/MOOSE (visited on 01/12/2024).

[65] *Radiopharmaceutical Start Time Attribute – DICOM Standard Browser*. URL: https://dicom.innolitics.com/ciods/pet-image/pet-isotope/00540016/00181072 (visited on 02/13/2024).

[66] S. Vincent Rajkumar. "Multiple myeloma: 2022 update on diagnosis, risk stratification, and management". eng. In: *American Journal of Hematology* 97.8 (Aug. 2022), pp. 1086–1107. ISSN: 1096-8652. DOI: 10.1002/ajh.26590.

[67] S. Vincent Rajkumar et al. "International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma". English. In: *The Lancet Oncology* 15.12 (Nov. 2014). Publisher: Elsevier, e538–e548. ISSN: 1470-2045, 1474-5488. DOI: 10.1016/S1470-2045(14)70442-5. URL: https://www.thelancet.com/article/S1470-2045(14)70442-5/fulltext (visited on 01/16/2024).

[68] A. N. Ramesh et al. "Artificial intelligence in medicine." In: *Annals of the Royal College of Surgeons of England* 86.5 (Sept. 2004), pp. 334–338. ISSN: 0035-8843. DOI: 10.1308/147870804290. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1964229/ (visited on 12/14/2023).

[69] Murray Bruce Reed et al. "Whole-body metabolic connectivity framework with functional PET". In: *NeuroImage* 271 (May 2023), p. 120030. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2023.120030. URL: https://www.sciencedirect.com/science/article/pii/S1053811923001763 (visited on 02/06/2024).

[70] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. arXiv:1505.04597 [cs]. May 2015. DOI: 10.48550/arXiv.1505.04597. URL: http://arxiv.org/abs/1505.04597 (visited on 01/11/2024).

[71] *rordenlab/dcm2niix*. original-date: 2014-10-19T17:20:13Z. Feb. 2024. URL: https://github.com/rordenlab/dcm2niix (visited on 02/13/2024).

[72] A. K. Shukla and Utham Kumar. "Positron emission tomography: An overview". en-US. In: *Journal of Medical Physics* 31.1 (Mar. 2006), p. 13. ISSN: 0971-6203. DOI: 10.4103/0971-6203.25665. URL: https://journals.lww.com/jomp/fulltext/2006/31010/positron_emission_tomography__an_overview.3.aspx (visited on 01/08/2024).

[73] Muhammad Siddiqi et al. "A Precise Medical Imaging Approach for Brain MRI Image Classification". In: *Computational Intelligence and Neuroscience* 2022 (May 2022), pp. 1–15. DOI: 10.1155/2022/6447769.

[74] Benjamin A. Spencer et al. "Performance Evaluation of the uEXPLORER Total-Body PET/CT Scanner Based on NEMA NU 2-2018 with Additional Tests to Characterize PET Scanners with a Long Axial Field of View". eng. In: *Journal of Nuclear Medicine: Official Publication, Society of Nuclear Medicine* 62.6 (June 2021), pp. 861–870. ISSN: 1535-5667. DOI: 10.2967/jnumed.120.250597.

[75] M. Stone. "Cross-validation:a review 2". EN. In: *Statistics: A Journal of Theoretical and Applied Statistics* (Jan. 1978). Publisher: Akademie-Verlag. DOI: 10.1080/02331887808801414. URL: https://www.tandfonline.com/doi/abs/10.1080/02331887808801414 (visited on 02/22/2024).

[76] M. Stone. "Cross-Validatory Choice and Assessment of Statistical Predictions". en. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 36.2 (1974). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1974.tb00994.x, pp. 111–133. ISSN: 2517-6161. DOI: 10.1111/j.2517-6161.1974.tb00994.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1974.tb00994.x (visited on 02/22/2024).

[77] Lalith Kumar Shiyam Sundar et al. "Fully Automated, Semantic Segmentation of Whole-Body ^{18} F-FDG PET/CT Images Based on Data-Centric Artificial Intelligence". en. In: *Journal of Nuclear Medicine* 63.12 (Dec. 2022), pp. 1941–1948. ISSN: 0161-5505, 2159-662X. DOI: 10.2967/jnumed.122.264063. URL: http://jnm.snmjournals.org/lookup/doi/10.2967/jnumed.122.264063 (visited on 01/10/2024).

[78] Alberto Stefano Tagliafico et al. "Quantitative Imaging and Radiomics in Multiple Myeloma: A Potential Opportunity?" In: *Medicina* 57.2 (Jan. 2021), p. 94. ISSN: 1010-660X. DOI: 10.3390/medicina57020094. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7912483/ (visited on 01/18/2024).

[79] Janita E. van Timmeren et al. "Radiomics in medical imaging—"how-to" guide and critical reflection". In: *Insights into Imaging* 11.1 (Aug. 2020), p. 91. ISSN: 1869-4101. DOI: 10.1186/s13244-020-00887-2. URL: https://doi.org/10.1186/s13244-020-00887-2 (visited on 12/19/2023).

[80] Brani Vidakovic and Peter Mueller. "Wavelets for kids". In: *Instituto de Estadística, Universidad de Duke* (1994), p. 161. URL: http://www.isye.gatech.edu/~brani/wp/kidsA.pdf (visited on 03/04/2024).

[81] Rishi K. Wadhera and S. Vincent Rajkumar. "Prevalence of Monoclonal Gammopathy of Undetermined Significance: A Systematic Review". en. In: *Mayo Clinic Proceedings* 85.10 (Oct. 2010). Publisher: Mayo Foundation, p. 933. DOI: 10.4065/mcp.2010.0337. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2947966/ (visited on 01/17/2024).

[82] Martin J. Willemink and Peter B. Noël. "The evolution of image reconstruction for CT—from filtered back projection to artificial intelligence". en. In: *European Radiology* 29.5 (May 2019), pp. 2185–2195. ISSN: 1432-1084. DOI: 10.1007/s00330-018-5810-7. URL: https://doi.org/10.1007/s00330-018-5810-7 (visited on 01/08/2024).

[83] A. M. Winkler. *The NIFTI file format*. en. Sept. 2012. URL: https://brainder.org/2012/09/23/the-nifti-file-format/ (visited on 02/12/2024).

[84] Pengfeng Xiao et al. "Cosegmentation for Object-Based Building Change Detection From High-Resolution Remotely Sensed Images". In: *IEEE Transactions on Geoscience and Remote Sensing* PP (Jan. 2017), pp. 1–17. DOI: 10.1109/TGRS.2016.2627638.

[85] Yishu Xue and Elizabeth D. Schifano. "Diagnostics for the Cox model". en. In: *Communications for Statistical Applications and Methods* 24.6 (Nov. 2017). Publisher: Korean Statistical Society, pp. 583–604. ISSN: 2287-7843. DOI: 10.29220/CSAM.2017.24.6.583. URL: http://www.csam.or.kr/journal/view.html?doi=10.29220/CSAM.2017.24.6.583 (visited on 01/16/2024).

Annexes

A Tables

NIfTI header table

Type	Name	Description
int	sizeof_hdr	Size of the header. Must be 348 (bytes).
char	data_type[10]	Not used; compatibility with analyze.
char	db_name[18]	Not used; compatibility with analyze.
int	extents	Not used; compatibility with analyze.
short	session_error	Not used; compatibility with analyze.
char	regular	Not used; compatibility with analyze.
char	dim_info	Encoding directions (phase, frequency, slice).
short	dim[8]	Data array dimensions.
float	intent_p1	1st intent parameter.
float	intent_p2	2nd intent parameter.
float	intent_p3	3rd intent parameter.
short	intent_code	nifti intent.
short	datatype	Data type.
short	bitpix	Number of bits per voxel.
short	slice_start	First slice index.
float	pixdim[8]	Grid spacings (unit per dimension).
float	vox_offset	Offset into a .nii file.
float	scl_slope	Data scaling, slope.
float	scl_inter	Data scaling, offset.
short	slice_end	Last slice index.
char	slice_code	Slice timing order.
char	xyzt_units	Units of pixdim[1..4].
float	cal_max	Maximum display intensity.
float	cal_min	Minimum display intensity.

Type	Name	Description
float	slice_duration	Time for one slice.
float	toffset	Time axis shift.
int	glmax	Not used; compatibility with analyze.
int	glmin	Not used; compatibility with analyze.
char	descrip[80]	Any text.
char	aux_file[24]	Auxiliary filename.
short	qform_code	Use the quaternion fields.
short	sform_code	Use of the affine fields.
float	quatern_b	Quaternion b parameter.
float	quatern_c	Quaternion c parameter.
float	quatern_d	Quaternion d parameter.
float	qoffset_x	Quaternion x shift.
float	qoffset_y	Quaternion y shift.
float	qoffset_z	Quaternion z shift.
float	srow_x[4]	1st row affine transform.
float	srow_y[4]	2nd row affine transform.
float	srow_z[4]	3rd row affine transform.
char	intent_name[16]	Name or meaning of the data.
char	magic[4]	Magic string.

Table 5.1: Table representing the information stored in a NIfTI file header. [83]

SR Cox model results

Dataset	Filter	C-Index
MM_PT_o	wavelet-HHH	0.61 ± 0.04
MM_PT_o	wavelet-HHL	0.65 ± 0.04
MM_PT_o	wavelet-HLH	0.60 ± 0.04
MM_PT_o	wavelet-HLL	0.66 ± 0.04
MM_PT_o	wavelet-LHH	0.66 ± 0.03
MM_PT_o	wavelet-LHL	0.65 ± 0.02
MM_PT_o	wavelet-LLH	0.62 ± 0.04
MM_PT_o	wavelet-LLL	0.63 ± 0.04

Dataset	Filter	C-Index
MM_PT_o	log-sigma-1-0-mm-3D	0.64 ± 0.03
MM_PT_o	log-sigma-2-0-mm-3D	0.65 ± 0.03
MM_PT_f	wavelet-HHH	0.71 ± 0.04
MM_PT_f	wavelet-HHL	0.68 ± 0.03
MM_PT_f	wavelet-HLH	0.68 ± 0.04
MM_PT_f	wavelet-HLL	0.65 ± 0.04
MM_PT_f	wavelet-LHH	0.68 ± 0.04
MM_PT_f	wavelet-LHL	0.65 ± 0.04
MM_PT_f	wavelet-LLH	0.68 ± 0.04
MM_PT_f	wavelet-LLL	0.68 ± 0.04
MM_PT_f	log-sigma-1-0-mm-3D	0.65 ± 0.04
MM_PT_f	log-sigma-2-0-mm-3D	0.65 ± 0.04
MM_PT_d	wavelet-HHH	0.71 ± 0.04
MM_PT_d	wavelet-HHL	0.69 ± 0.05
MM_PT_d	wavelet-HLH	0.67 ± 0.05
MM_PT_d	wavelet-HLL	0.69 ± 0.04
MM_PT_d	wavelet-LHH	0.69 ± 0.05
MM_PT_d	wavelet-LHL	0.69 ± 0.04
MM_PT_d	wavelet-LLH	0.69 ± 0.05
MM_PT_d	wavelet-LLL	0.69 ± 0.05
MM_PT_d	log-sigma-1-0-mm-3D	0.62 ± 0.05
MM_PT_d	log-sigma-2-0-mm-3D	0.69 ± 0.05
MM_CT_o	wavelet-HHH	0.67 ± 0.07
MM_CT_o	wavelet-HHL	0.65 ± 0.05
MM_CT_o	wavelet-HLH	0.62 ± 0.05
MM_CT_o	wavelet-HLL	0.65 ± 0.06
MM_CT_o	wavelet-LHH	0.71 ± 0.04
MM_CT_o	wavelet-LHL	0.72 ± 0.06
MM_CT_o	wavelet-LLH	0.67 ± 0.07
MM_CT_o	wavelet-LLL	0.66 ± 0.06
MM_CT_o	log-sigma-1-0-mm-3D	0.70 ± 0.06
MM_CT_o	log-sigma-2-0-mm-3D	0.66 ± 0.06
MM_CT_f	wavelet-HHH	0.71 ± 0.06
MM_CT_f	wavelet-HHL	0.72 ± 0.03
MM_CT_f	wavelet-HLH	0.75 ± 0.05
MM_CT_f	wavelet-HLL	0.72 ± 0.03
MM_CT_f	wavelet-LHH	0.72 ± 0.06
MM_CT_f	wavelet-LHL	0.71 ± 0.05

Dataset	Filter	C-Index
MM_CT_f	wavelet-LLH	0.65 ± 0.02
MM_CT_f	wavelet-LLL	0.72 ± 0.05
MM_CT_f	log-sigma-1-0-mm-3D	0.72 ± 0.03
MM_CT_f	log-sigma-2-0-mm-3D	0.66 ± 0.06
MM_CT_d	wavelet-HHH	0.72 ± 0.04
MM_CT_d	wavelet-HHL	0.71 ± 0.06
MM_CT_d	wavelet-HLH	0.75 ± 0.05
MM_CT_d	wavelet-HLL	0.73 ± 0.03
MM_CT_d	wavelet-LHH	0.74 ± 0.05
MM_CT_d	wavelet-LHL	0.72 ± 0.03
MM_CT_d	wavelet-LLH	0.75 ± 0.04
MM_CT_d	wavelet-LLL	0.74 ± 0.04
MM_CT_d	log-sigma-1-0-mm-3D	0.72 ± 0.04
MM_CT_d	log-sigma-2-0-mm-3D	0.71 ± 0.05

Table 5.2: Cox model results for SR algorithm. The results are divided for datasets (first column), filter alongside the original image for the feature extraction (second columns), mean C-index value and C-index standard deviation for the results coming from the five 5 CV.