POLITECNICO
MILANO 1863

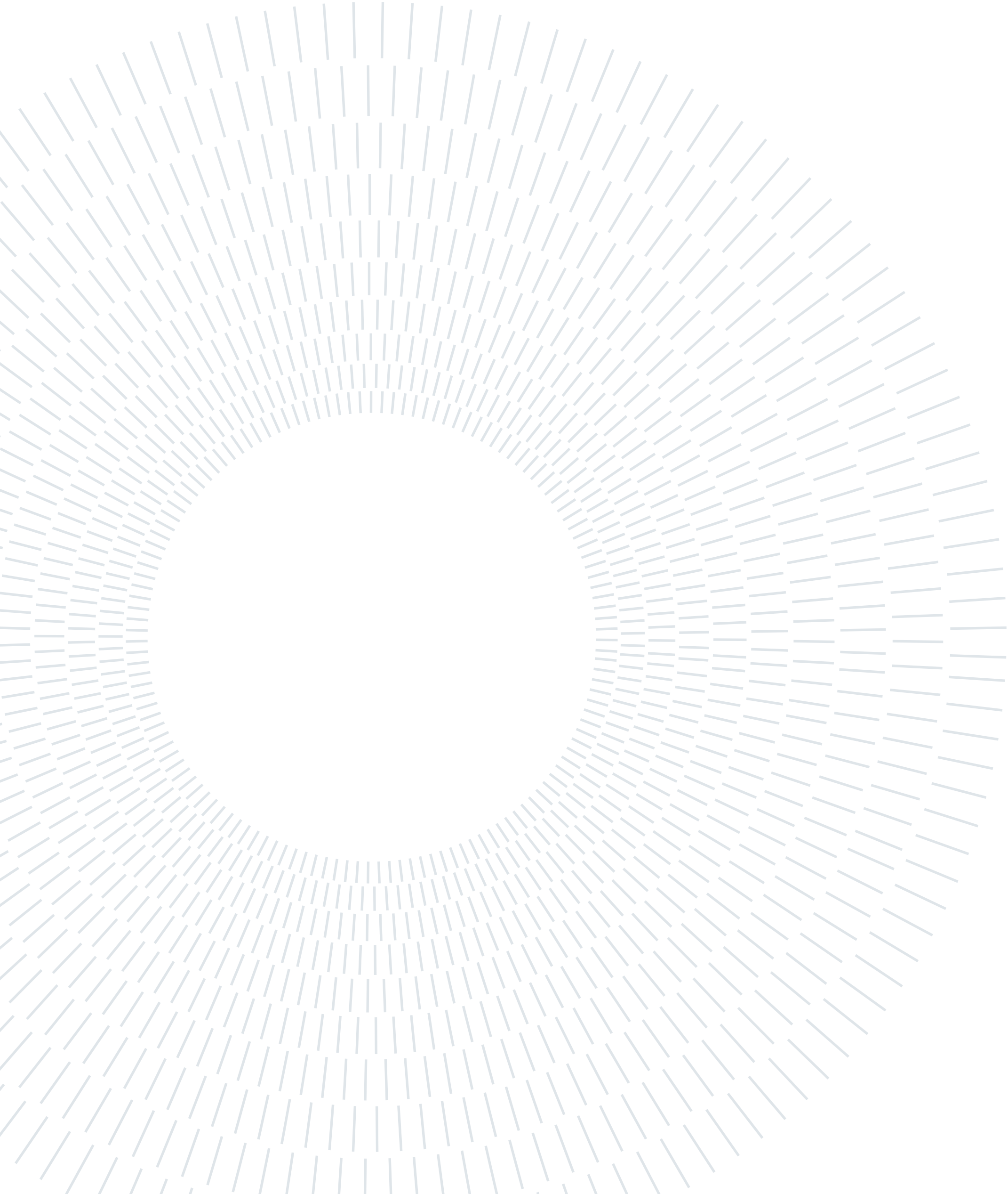
SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Integration of Radiomic Features and Intra-Patient Heterogeneity of Colorectal Cancer Metastases for Prognostic Modeling

Tesi di Laurea Magistrale in
Mathematical Engineering - Ingegneria Matematica

Author: **Gilda Matteucci**

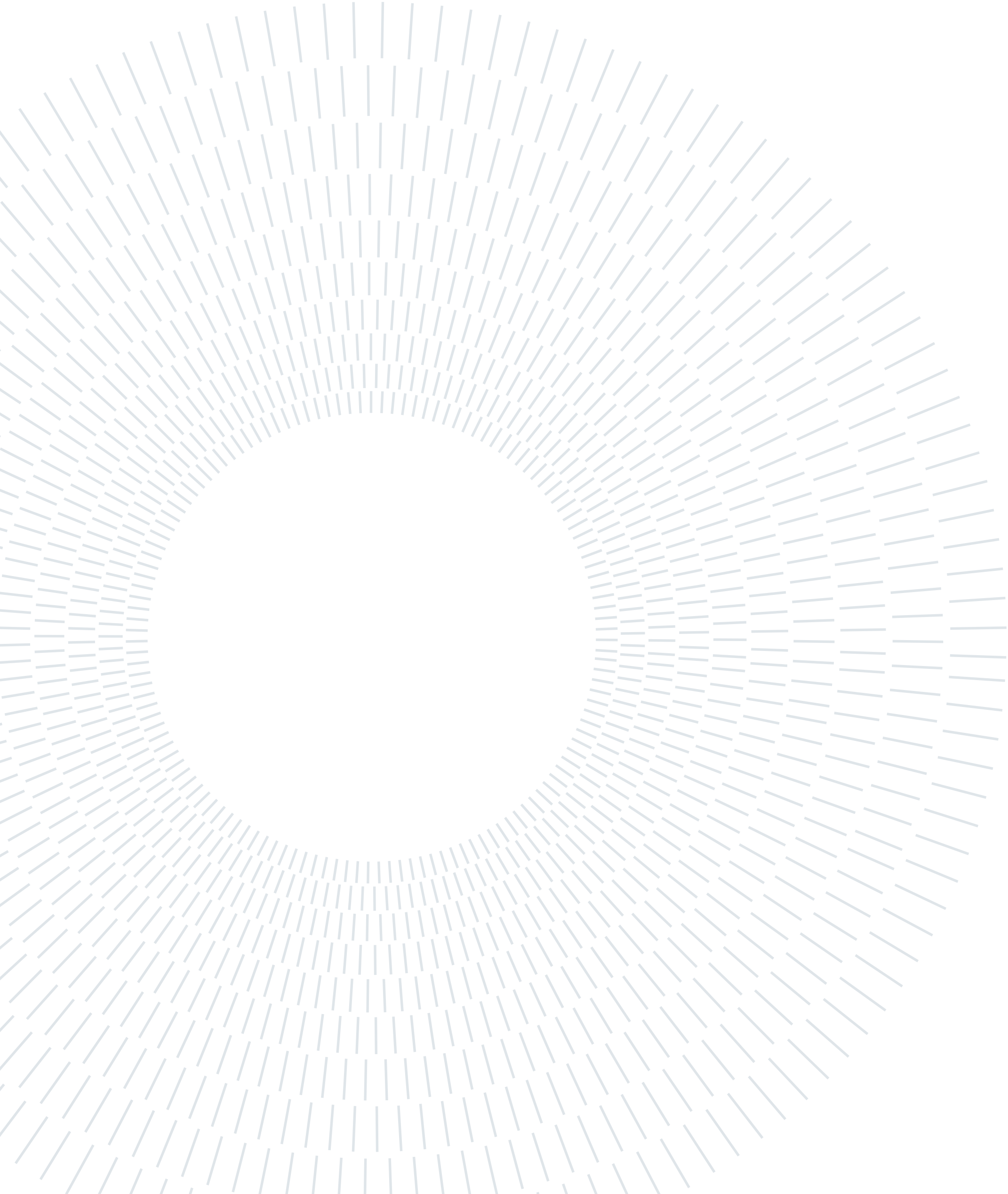
Student ID: 953829
Advisor: Prof. Francesca Ieva
Co-advisors: Chiara Masci, Lara Cavinato
Academic Year: 2023-24



Abstract

Colorectal cancer (CRC) represents one of the leading causes of death worldwide. In recent years, radiomics—a methodology that automatically extracts quantitative information from medical images—has offered new prospects for personalizing oncological treatments by providing non-invasive biomarkers that complement traditional clinical and genetic data. This thesis investigates the integration of clinical, genetic, and radiomic data to improve the prediction of survival in patients with CRC liver metastases. In particular, we apply survival models based on Random Survival Forests, which, through an accurate variable selection process based on the Boruta algorithm, evaluate the predictive power of features extracted from pre-intervention CT scans, both in the neoadjuvant pre-chemotherapy phase and in the pre-operative phase. The results show that the addition of radiomic information from the primary lesion, especially the one related to the central tumor, significantly increases the performance of survival models compared to the isolated use of clinical and genetic data. Furthermore, intra-patient metastasis heterogeneity is explored using tree-clustering techniques based on edit distances between dendrograms, revealing that patients with more homogeneous metastases or with high variability present distinct survival profiles. However, the integration of this stratification into the Random Survival Forest models does not lead to significant improvements in predictive performance, suggesting that the information regarding lesion variability is already partially captured by the radiomic characteristics of the primary lesion. The study supports the use of radiomics as a complementary tool for prognostic assessment in CRC liver metastases and opens new prospects for more personalized treatment strategies.

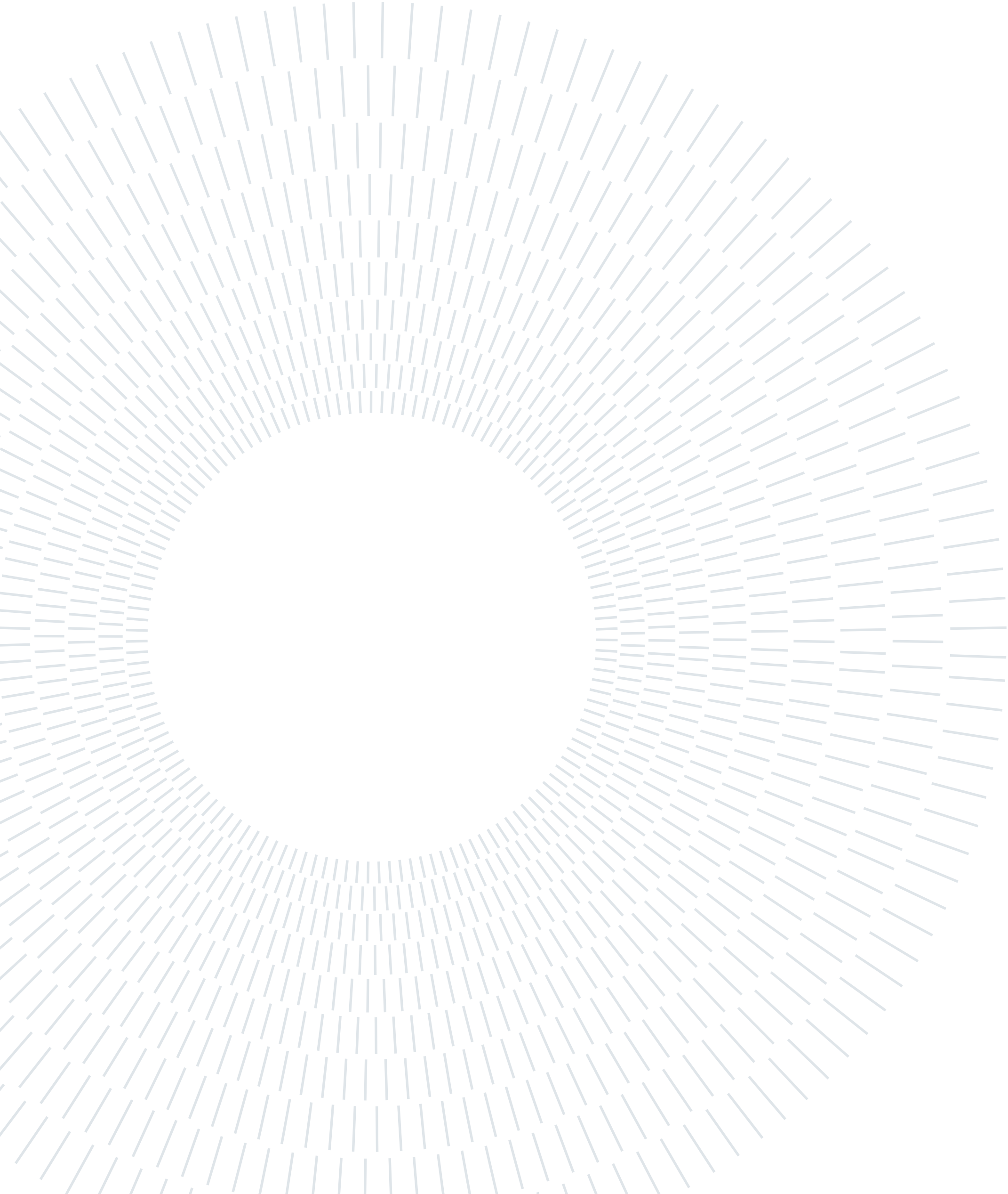
Keywords: Radiomics, Metastatic Colorectal Cancer, Survival analysis, Intra-Patient Heterogeneity Analysis, Tree-clustering



Abstract in lingua italiana

Il carcinoma del colon-retto (CRC) rappresenta una delle principali cause di mortalità a livello mondiale. Negli ultimi anni, la radiomica – una metodologia che consente di estrarre automaticamente informazioni quantitative da immagini mediche – ha offerto nuove prospettive per la personalizzazione dei trattamenti oncologici, fornendo biomarcatori non invasivi e complementari ai tradizionali dati clinici e genetici. Questa tesi indaga l'integrazione di dati clinici, genetici e radiomici per migliorare la previsione della sopravvivenza nei pazienti affetti da metastasi epatiche da CRC. In particolare, abbiamo sviluppato modelli di sopravvivenza basati su Random Survival Forest che, mediante un accurato processo di selezione delle variabili (basato sull'algoritmo Boruta), valutano la potenza predittiva di caratteristiche estratte da scansioni TC pre-intervento, sia nella fase pre-chemioterapia neoadiuvante che pre-operatoria. I risultati mostrano che l'aggiunta delle informazioni radiomiche della lesione primaria, soprattutto quelle relative al tumore centrale, incrementa significativamente le performance dei modelli di sopravvivenza rispetto all'uso isolato di dati clinici e genetici. Inoltre, abbiamo esplorato l'eterogeneità intra-paziente delle metastasi attraverso tecniche di tree-clustering basate su distanze di edit tra dendrogrammi, evidenziando che pazienti con metastasi più omogenee o con elevate variazioni presentano profili di sopravvivenza distinti. Tuttavia, l'integrazione di tale stratificazione nei modelli di Random Survival Forest non porta a miglioramenti significativi nelle prestazioni predittive, suggerendo che le informazioni sulla variabilità delle lesioni siano già in parte catturate dalle caratteristiche radiomiche della lesione primaria. Lo studio supporta l'utilizzo della radiomica come strumento complementare per la valutazione prognostica nelle metastasi epatiche da CRC e apre nuove prospettive per strategie di trattamento più personalizzate.

Parole chiave: Radiomica, Carcinoma Colorettale Metastatico, Analisi di Sopravvivenza, Analisi di Eterogeneità Intra-Paziente, Tree-clustering



Contents

Introduction

Colorectal Cancer (CRC) is a malignant tumor arising in the final segment of the digestive tract, specifically in the colon (right or left) or the rectum, and it often develops from polyps or other precancerous lesions that gradually acquire malignant characteristics over time. It ranks among the most common malignancies worldwide [34]: recent statistics estimate that 1,926,425 new cases of CRC were diagnosed globally in 2022, with 904,019 related deaths [14], and it is estimated that by 2030 there will be more than 2.2 million new cases and 1.1 million deaths [2].

In its early stages, CRC may be asymptomatic or present only mild symptoms that become more apparent as the disease progresses. Common clinical signs include changes in bowel habits, blood in the stool, abdominal pain, unintentional weight loss, fatigue, and anemia. Given the non-specific nature of these symptoms, prevention plays a crucial role in making an early diagnosis and reducing mortality rates. Key preventive measures include polyps removal, Computed Tomography (CT) and fecal occult blood tests [8].

At least 25% of patients affected by CRC will develop Colorectal Liver Metastases (CRLM) at some point during the course of the disease [7, 13, 30]. For this reason, early detection and timely intervention are crucial for reducing the risk of tumor spread and improving treatment outcomes [2]. The liver is one of the most commonly affected organs due to its unique blood supply: blood from the intestines flows through the portal vein directly to the liver, allowing circulating tumor cells to settle in hepatic tissue and form metastatic lesions.

Since CRC patients often exhibit few or no symptoms in the early stages, the disease is frequently discovered only when it is already in an advanced stage and sometimes has already metastasized. Surgery remains the primary and most effective treatment for CRC, especially in the early stages. The procedure may involve the removal of tumors or, in more severe cases, the resection of a portion of the colon or rectum. In certain situations, neoadjuvant chemotherapy can be prescribed prior to surgery to shrink the tumor and facilitate resection, while adjuvant chemotherapy is commonly used after surgery to reduce the risk of recurrence.

Although the morphology of lesions is routinely accessible through surgical resection, identifying non-invasive biomarkers that accurately characterize metastases and predict patient survival remains both challenging and highly desirable. One promising approach is radiomics [11, 18], an emerging field that non-invasively extracts quantitative information from conventional medical imaging, such as CT scans, to provide comprehensive insights into the disease. Many studies have demonstrated its potential as a powerful tool in clinical decision-making and personalized medicine, improving diagnosis and risk stratification across various types of cancer, including rectal cancer [36]. Therefore, integrating radiomics into the diagnostic and prognostic workflows for colorectal liver metastases presents a promising avenue for improving patient care and treatment outcomes.

Although very promising, radiomics faces several notable hurdles [18, 31]. The extraction process yields a high-dimensional set of features, many of which can be highly correlated, thus necessitating rigorous feature selection to avoid overfitting. Additionally, variations in imaging protocols, segmentation methods, and pre-processing techniques can compromise the reproducibility and comparability of radiomic data, especially when collected from different centers. Finally, the computational effort, time, and financial investment required for radiomic extraction further complicate a widespread clinical adoption.

Outline of the study

This study aims to evaluate the prognostic value of radiomics combined with clinical and genetic data to estimate their effects on survival and to assess the impact of intra-patient heterogeneity of metastases on survival outcomes. To achieve this, the primary target variable for the analysis is the time-to-event (e.g., Overall Survival), along with a censoring variable indicating whether the event occurred during the follow-up period or if the data are censored. Specifically, it attempts to:

- Assess the incremental predictive benefit of integrating radiomic data into survival models over traditional clinical and genetic biomarkers by employing Random Survival Forest methods, while also evaluating whether the added cost and complexity of radiomic analysis are justified by its impact on prediction accuracy and clinical decision-making. This was performed separately using information (including radiomics) prior to neoadjuvant chemotherapy and information prior to surgery.
- Analyze intra-patient metastasis heterogeneity using radiomic-based clustering techniques to provide further insights for personalized treatment strategies. In this phase, the study will quantify heterogeneity between multiple metastatic lesions by computing tree-edit distances and applying clustering methods, aiming to stratify

patients based on the variability of their tumor lesions and investigate potential correlations with treatment responses and clinical outcomes.

In terms of survival analysis, numerous studies have investigated patient survival using clinical, radiomic, and genetic biomarkers with methods such as Random Survival Forests (e.g., [20, 22, 35]). However, this work uniquely extracts radiomic features from CT images acquired not only before surgery but also prior to neoadjuvant chemotherapy. In doing so, it examines whether radiomics derived from the pre-chemotherapy phase possess the same prognostic value as those obtained preoperatively, and it evaluates two distinct regions of interest (the tumor core and its margin) to determine if including margin radiomics further enhances survival prediction accuracy.

Regarding intra-patient heterogeneity, most literature focuses on intra-metastasis heterogeneity [37], employs basic analytical approaches [27], or relies on non-radiomic data such as transcriptomics for stratification [29]. In contrast, this study introduces a novel, sophisticated radiomics-based clustering method designed to capture complex intra-patient heterogeneity. Originally applied to prostate cancer [5], this method stratifies patients based on the variability among multiple metastatic lesions, thereby paving the way for more personalized treatment strategies.

To achieve these objectives, the study is organized into the following chapters:

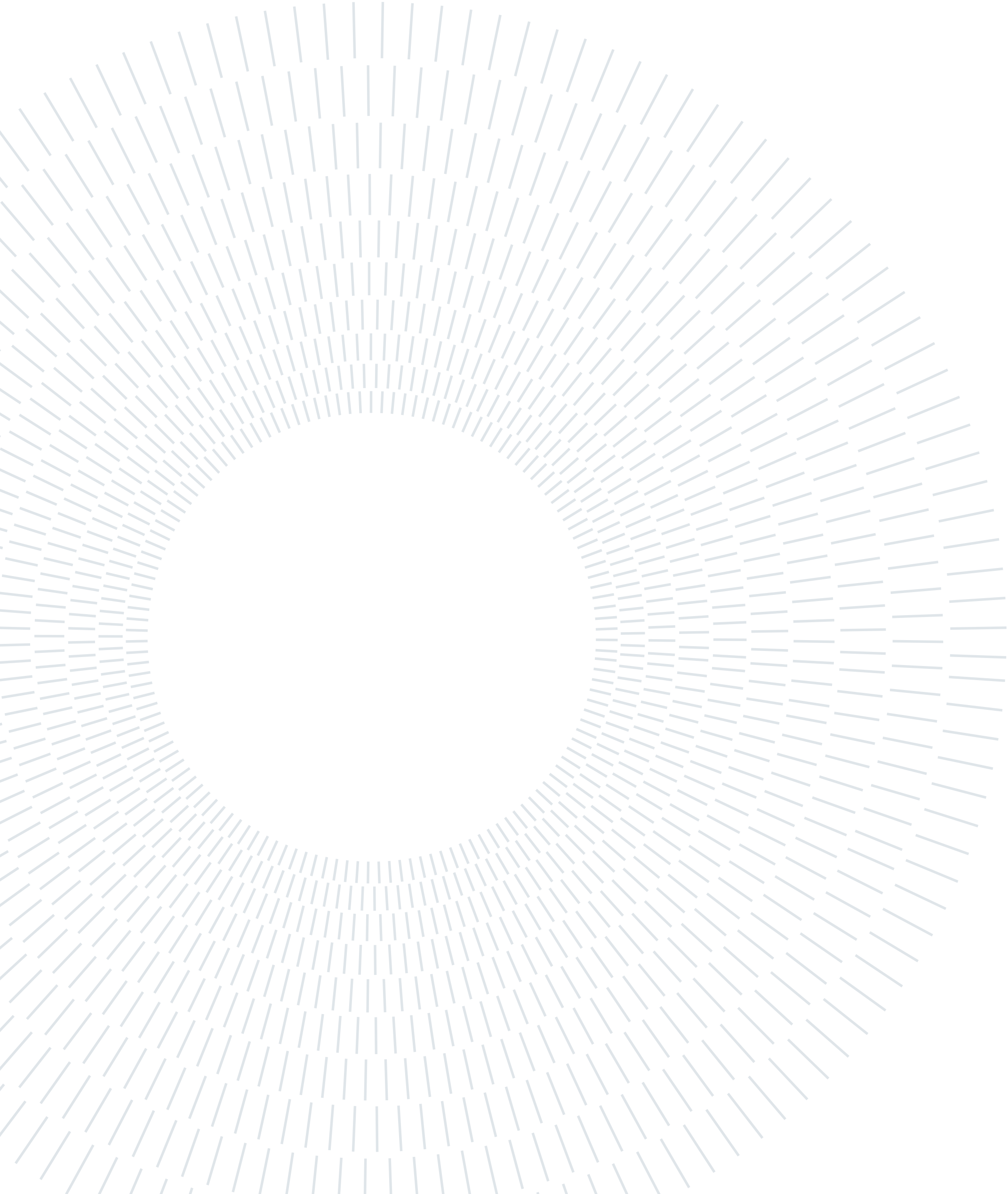
Chapter 1 Data Description and Preprocessing. Description of the dataset and detailed procedures for data imputation and preprocessing techniques, including dimensionality reduction.

Chapter 2 Methodological Pipeline. Overview of the statistical methods employed, such as Random Survival Forest for survival modeling and Tree-Clustering for heterogeneity analysis.

Chapter 3 Results. Presentation of model performance comparisons with and without radiomic data, along with a detailed description of intra-patient heterogeneity results, including patient stratification based on tree clustering and its clinical implications.

Chapter 4 Conclusions. Summary of the main findings, critical evaluation of the challenges associated with radiomic analysis, and suggestions for future research.

All the analysis were conducted using R and Python software.



1 | Dataset Description and Preprocessing

In this chapter, a more in-depth description of the dataset used in the study is provided. Section 1.1 presents an overview of the data, detailing the information available for each patient. The preprocessing procedure for clinical variables is shown in Section 1.2, while Section 1.3 exhibits the method employed to reduce the number of radiomic variables.

1.1. Dataset Description

Data used in this study were provided by the Humanitas Research Hospital and include information on 306 patients affected by CRLM. The selection process of the patients included in the study is summarized in Figure 1.1. For each patient, clinical, genetic and radiomic data of the primary metastasis and some secondary metastases were collected before the tumor removal surgery. Specifically, radiomic features were extracted from the pre-operative CT scan for both the tumor core and the tumor ring. Additionally, among the 216 patients who underwent neoadjuvant chemotherapy, radiomic data of the largest metastasis before therapy are also available for 102 of them.

A complete description of the timeline for the data collection is represented in Figure 1.2.

The primary target variable of the analysis proposed in this study is Overall Survival (`OS`), which indicates, in months, the survival time of the patients during the follow-up period. Another important target variable is the censoring variable (`morto`), a binary variable indicating whether the event occurred during the follow-up period (1 if the patient died) or if the data are censored (0 if the event did not occur or if the data are incomplete). A complete description of these two variables is provided in Table 1.1.

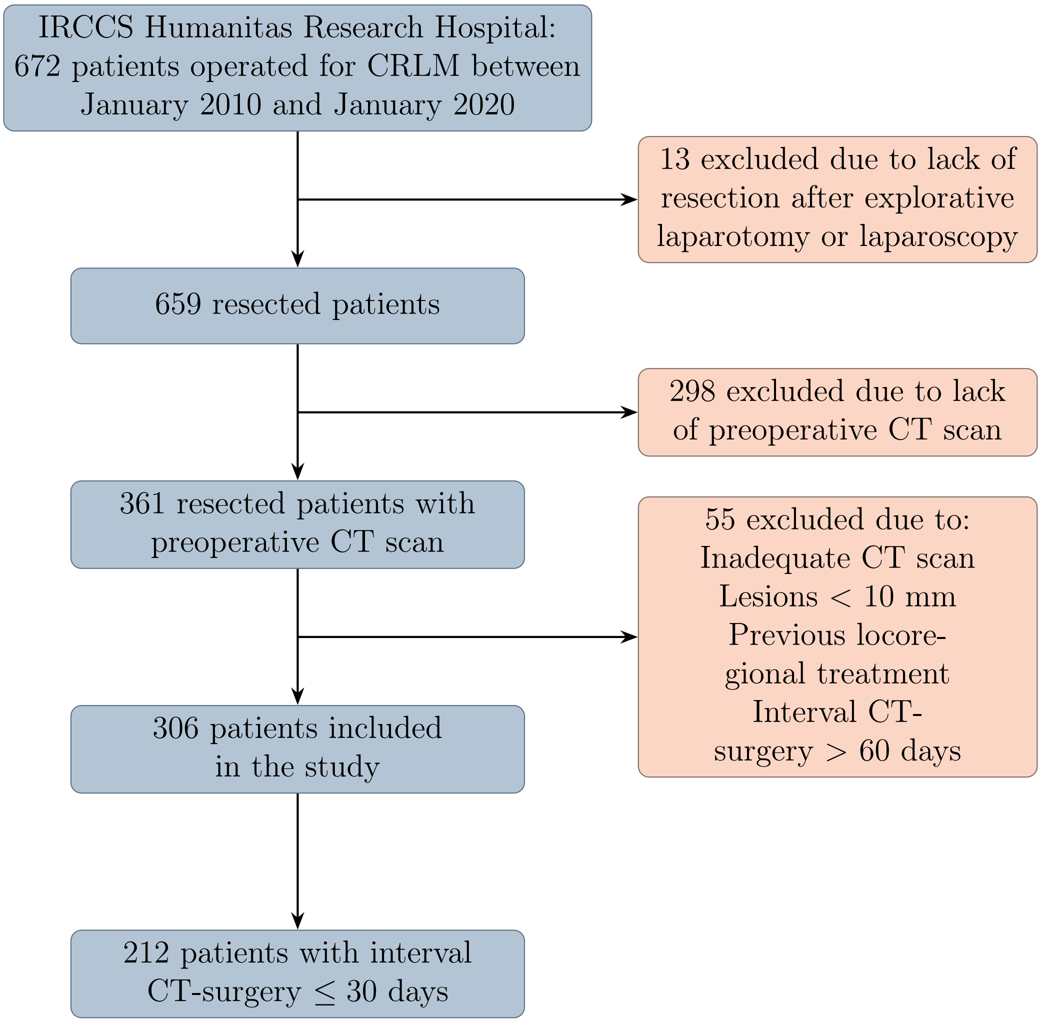


Figure 1.1: Flowchart illustrating the patient selection process for this study. The inclusion and exclusion criteria are detailed, highlighting the number of patients retained at each stage

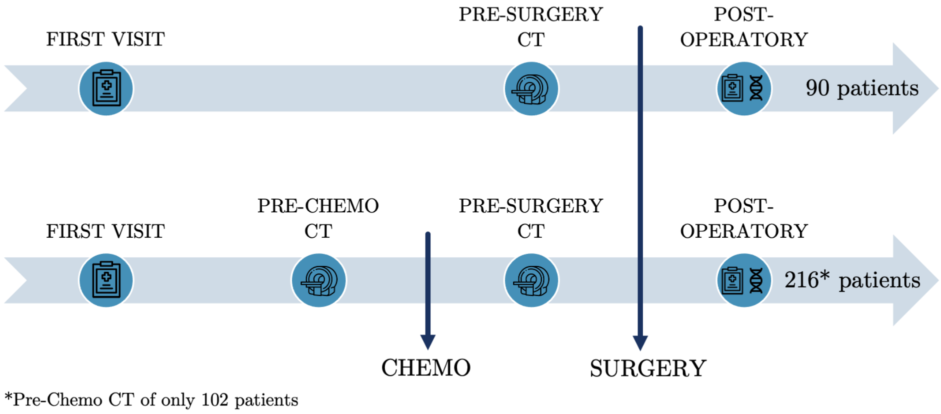


Figure 1.2: Two different timelines illustrating the data collection process for patients, based on whether they underwent neoadjuvant chemotherapy. The upper arrow represents patients who did not receive neoadjuvant chemotherapy, while the lower arrow represents those who did. Each icon represents the type of data collected at that stage (clinical, radiomic, or genetic information). Clinical data are categorized as preoperative if collected during the first visit and as postoperative if collected after surgery

Name	Description	Values

`OS`	Overall Survival (months)	min=0.033;
		mean=34.013;
		max=137.866;

`morto`	Censoring binary variable (1 = event occurred; 0 = censored)	1: 160; 0: 146

Table 1.1: Description of Target Variables

Clinical Variables

Clinical variables include personal information about the patient (such as age and sex), details about the primary tumor and metastases, treatments received, complications, survival and mortality. As mentioned above, clinical variables are classified as preoperative or postoperative, depending on whether information was collected before or after surgery. Postoperative information is further divided into three categories:

Postoperative variables Including presence of lymph node involvement (`stageN`), post-operative complications and the radicality of tumor resection.

Prognostic scores Composite measures that integrate multiple clinical and pathological factors to estimate patient risk and predict outcomes. In this study, widely used scores such as FONG [10], GAME [24], and the modified-clinical score [4] were employed.

Patient outcome variables Overall Survival (`OS`), and mortality status (`morto`).

Tables 1.2, Table 1.3 and Table 1.4 provide a detailed description of the preoperative, postoperative and prognostic variables used.

Note that some of these variables contained missing values. To include them in the analyses, an imputation process was performed, as described in Section 1.2. A descriptive summary of continuous and categorical variables is provided in Tables 1.6, 1.7, 1.8 within the same section.

Radiomic Variables

The radiomic features analyzed in this study were extracted from preoperative CT scans taken before surgery. In addition, for 102 patients who underwent neoadjuvant chemotherapy, radiomic characteristics of the CT scan performed before chemotherapy are also available.

As mentioned above, for each CT scan, two Regions of Interest (ROIs) were segmented: the tumor core and the tumor ring. The tumor ring, which represents the peritumoral liver tissue, is a $5mm$ area automatically generated by the LIFEx software `https://www.lifexsoft.org` around the tumor and subsequently manually adjusted to improve accuracy. The segmentation of the ROIs and the extraction of radiomic features were performed by experienced radiologists using this software to ensure consistency and precision.

From each ROI, 109 radiomic features were extracted, leading to a total of 218 radiomic variables. A detailed description of the extracted radiomic features can be found in the software documentation [9, 41].

Given the large number of radiomic variables extracted in this study, all 109 features are not listed individually here, but a detailed list of all of them can be found in Appendix A.1. The radiomic features are divided into several groups based on their characteristics and the level of analysis. These groups provide a structured way to understand the different types of information they capture.

The features range from those that are more directly interpretable (such as morphological and intensity-based features) to those that are more complex and abstract, making them harder to interpret directly from the images. However, the more complex features

Name	Description	Type
`sex`	sex of the patient (F = female, M = male)	categorical
`age`	age of the patient at the surgery	numeric
`deltaData`	surgery date - CT date	numeric
`stageT34`	T stage of the primitive tumor [a]	binary
`sedePrimitivo`	site of the primitive tumor (colonDx = right colon, colonSx = left colon, retto = rectum)	categorical
`sinc`	metastases timing (1 = synchronous, 0 = metachronous)	binary
`nMets`	number of metastases	numeric
`numClassi`	number of metastases divided into classes (una = 1 met, due_tre = 2-3 mets, quattro_nove = 4-9 mets, 10+ = more than 10 mets)	categorical
`diam_preChemo`	diameter of the major metastasis before neoadjuvant chemo	numeric
`diam_preSurgery`	diameter of the major metastasis before the surgery	numeric
`diam_gt5`	diameter before the surgery divided into classes (1 = diam $\geq$ 5cm, 0 = diam $<$ 5cm)	binary
`CEA_preChemo`	Carcino-Embryonic Antigen value before neoadjuvant chemo	numeric
`CEA_preSurgery`	Carcino-Embryonic Antigen value before surgery	numeric
`malattiaExtrahepSinc`	presence of concomitant extrahepatic disease	binary
`CHT_neoadj`	Neoadjuvant chemotherapy (1 = administred, 0 = not administred)	binary
`TRG`	Tumor Regression Grade (Mandard Scale: 1-3 = responder, 4-5 = non-responder)	ordinal
`TRGsimpl`	Tumor Regression Grade simplified (1 = responder, 0 = non-responder)	categorical
`rispostaUltimaLinea`	response to neoadjuvant chemoterapy (0 = not administred, 1 = partial response, 2 = stable disease, 3 = disease progression)	categorical

Table 1.2: Description of Clinical Preoperative Variables

---

[a] Stage T indicates how much the tumor has grown into the surrounding tissues: Tis (carcinoma in situ), T1 (invasion of the submucosa), T2 (invasion of the muscolaris propria), T3 (invasion into the pericolic or perirectal fat), T4 (direct invasion into adjacent structures)

Name	Description	Type
`stageN`	Lymph node status of the primary tumor (0 = N0, no lymph node involvement; 1 = N+, lymph node metastases present)	binary
`morbSevera`	Severe postoperative complications (1 = Yes, 0 = No)	binary
`morbInfettiva`	Infectious complications (1 = Yes, 0 = No)	binary
`r0_gt1`	Radicality of resection (1 = margin $\geq 1mm$, 0 = margin $< 1mm$)	binary
`CHT_adj`	Adjuvant chemotherapy (1 = Yes, 0 = No)	binary

Table 1.3: Description of Clinical Postoperative Variables

Name	Description	Type
`FONG_score`	Clinical Risk Score [10] for colorectal liver metastases (0-5), computed by summing the following risk factors: `stageN`=1, `nMets`>1, `CEA_preSurgery` > 200, `diam_gt5`=1, and short time to metastasis	ordinal
`FONGsimpl_score`	FONG score simplified (1 = 0-2 low risk, 2 = 3-5 high risk)	ordinal
`GAME_score`	Genetic And Morphological Evaluation score [24] (0-7) for colorectal liver metastases, computed by combining: `RAS`=1, `CEA_preSurgery`>20, `stageN`=1, `malattiaExtrahepSinc`=1 and Tumor Burden Score [b]	ordinal
`GAMEsimpl_score`	GAME score simplified (1 = 0-1 low risk, 2 = 2-3 moderate risk, 3 = 4-5-6-7 high risk)	ordinal
`Mclinical_score`	Modified Clinical score [4] (0-3) for colorectal liver metastases, based on `stageN`, `diam_gt5` and `RAS`.	ordinal

Table 1.4: Description of Clinical Prognostic Scores Variables

---

[b] Tumor Burden Score (TBS) is defined as $TBS^2 = \texttt{diam\_preSurgery}^2 + \texttt{nMets}^2$

provide deeper insights into the tumor's texture and its spatial properties. As the analysis progresses, the features become increasingly specialized and provide more detailed quantitative information about the tumor.

The following is a description of the different groups, outlining the specific characteristics and the type of information each group captures about the tumor. The number of features within each group is indicated in parentheses.

Morphological (12): These features describe the shape and geometry of the tumor. They include measurements such as volume, surface area, compactness, eccentricity, and sphericity, which help characterize the tumor's spatial properties.

Intensity-Based (19): This group includes features based on the statistical properties of voxel intensity values. These features capture various aspects of intensity distribution within ROI, such as mean, variance, skewness and kurtosis.

Intensity Histogram (23): These features represent the distribution of voxel intensities in the ROI, offering a deeper understanding of the intensity patterns. Commonly included are histogram-based metrics such as entropy, energy, and the number of distinct intensity levels.

GLCM (23): Grey Level Co-occurrence Matrix (GLCM) features are derived from the spatial relationship between voxel pairs with specific distances and orientations. They describe the texture of the image and capture patterns such as contrast, correlation, homogeneity, and energy.

GLRLM (11): Grey Level Run Length Matrix (GLRLM) features describe the length of consecutive runs of voxels with the same intensity value within a specified direction. These features provide insight into the image's texture, focusing on the homogeneity and length of intensity runs.

NGTDM (5): Neighbouring Grey Tone Difference Matrix (NGTDM) group measures the difference in intensity values between a voxel and its neighboring voxels. The NGTDM features capture the complexity of the tumor's texture, highlighting patterns such as coarseness and contrast at the voxel level.

GLSZM (16): Grey Level Size Zone Matrix (GLSZM) features quantify zones of uniform intensity within the ROI. These features focus on the size and distribution of such zones, providing information about the spatial heterogeneity of the tumor's texture.

Given the intrinsic nature of radiomic features within each group, many of them naturally exhibit high intercorrelation. Consequently, an initial feature selection step was necessary to mitigate redundancy and retain the most informative variables. This selection was carried out through a correlation analysis, filtering out highly correlated variables. The details of this procedure are discussed in Section 1.3.

Genetic Variables

The genetic variables available for this study represent mutations in KRAS and BRAF genes within the largest metastasis, which have been identified in the literature as relevant prognostic factors for CRLM. These mutations play a crucial role in tumor biology, influencing disease progression, response to targeted therapies, and overall survival. In particular, KRAS and BRAF mutations are widely recognized as predictive biomarkers in colorectal cancer, guiding treatment decisions and patient stratification [23, 28, 39]. Genetic data were obtained through a biopsy of only the largest metastasis, performed after surgical resection of the primary tumor. This approach was chosen because, according to the literature, the genetic profile of the largest metastasis is generally representative of the other metastatic sites [3]. By focusing on this lesion, the objective is to capture the most clinically relevant genetic alterations while minimizing the need for multiple biopsies. A detailed description of these variables can be found in Table 1.5.

Name	Description	Type
`RAS`	Presence of mutations in KRAS genes, which are associated with poorer prognosis and resistance to therapies in CRLM (1 = mutated, 0 = wild-type).	binary
`BRAF`	Presence of mutations in the BRAF gene, which is linked to aggressive tumor behavior and worse survival outcomes in CRLM (1 = mutated, 0 = wild-type).	binary

Table 1.5: Description of Genetic Variables

1.2. Preprocessing of Clinical Variables

As mentioned above, clinical and genetic data required preprocessing due to missing values in some features. To address this, missing values have been imputed using the Hyperimpute method proposed in [16]. Hyperimpute is a deep learning framework that leverages a hypernetwork to dynamically generate parameters for an ensemble of feature-specific conditional models. In an iterative process, each model predicts missing values for a given feature by conditioning on the remaining features, enabling the method to flexibly adapt to various patterns of missingness. Consequently, this approach effectively captures complex inter-feature relationships in high-dimensional and heterogeneous data. The high-level overview of Hyperimpute is shown in Figure 1.3.

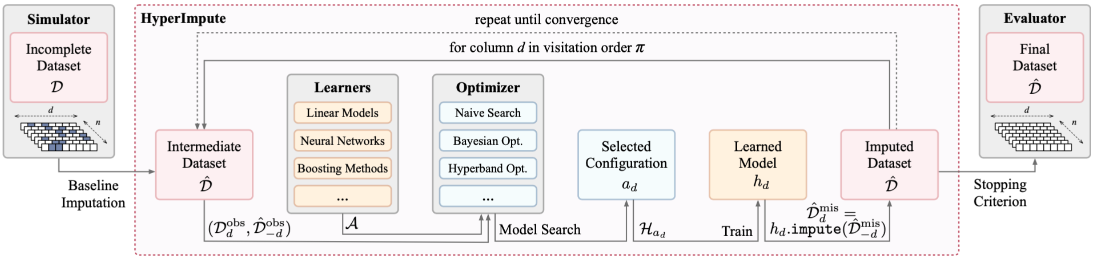


Figure 1.3: High-level overview of HyperImpute. Blue indicates model selection algorithms and their selected output. Orange indicates candidate models and their trained output. Red indicates datasets and imputations. Gray indicates component modules in HyperImpute. Image taken from [16]

For imputation of missing values, all available pre-surgery radiomic and clinical data were employed, excluding only the outcome variables (i.e., mortality status and OS). Given the low percentage of missing data, all values were imputed without excluding any variable from the analysis.

Summary statistics of all variables (before and after the imputation process) are reported in Tables 1.6, 1.7, 1.8.

Variable	n missing (%)	not-imputed mean (std)	imputed mean (std)
`age`	0 (0%)	63.18 (10.28)	-
`deltaData`	0 (0%)	22.90 (20.26)	-
`nMets`	0 (0%)	5.19 (5.08)	-
`diam_preChemo`	0 (0%)	46.44 (29.93)	-
`diam_preSurgery`	0 (0%)	4.13 (2.67)	-
`CEA_preChemo`	0 (0%)	452.56 (1362.54)	-
`CEA_preSurgery`	11 (3.59%)	89.66 (535.48)	89.93 (525.85)

Table 1.6: Summary Statistics of Numerical Clinical Preoperative Variables

Variable	n missing (%)	not-imputed % freq	imputed % freq
`sex`	0 (0%)	M: 61.11% F: 38.89%	-
`stageT34`	19 (6.21%)	1: 86.41% 0: 13.59%	1: 70.59% 0: 29.41%
`sedePrimitivo`	6 (1.96%)	colonSx: 49.33% retto: 27.67% colonDx: 23.00%	colonSx: 50.33% retto: 27.12% colonDx: 22.55%
`numClassi`	0 (0%)	una: 22.22% due_tre: 28.10% quattro_nove: 33.00% 10+: 16.67%	-
`malattiaExtrahepSinc`	0 (0%)	1: 16.01% 0: 83.99%	-
`CHT_neoadj`	0 (0%)	1: 70.59% 0: 29.41%	-
`TRG`	0 (0%)	1: 6.31% 2: 5.34% 3: 16.02% 4: 50.97% 5: 21.36%	-
`TRGsimpl`	0 (0%)	1: 27.67% 0: 72.33%	-
`rispostaUltimaLinea`	0 (0%)	0: 20.41% 1: 54.25% 2: 11.11% 3: 5.23%	-
`RAS`	14 (4.58%)	1: 35.27% 0: 64.73%	1: 33.66% 0: 66.34%
`BRAF`	16 (5.23%)	1: 1.72% 0: 98.28%	1: 1.63% 0: 98.37%

Table 1.7: Summary Statistics of Categorical Clinical Preoperative and Genetic Variables

Variable	n missing (%)	not-imputed % freq	imputed % freq
`stageN`	18 (5.88%)	1: 69.10% 0: 30.90%	1: 69.59% 0: 29.41%
`morbSevera`	0 (0%)	1: 7.19% 0: 92.81%	-
`morbInfettiva`	0 (0%)	1: 10.46% 0: 89.54%	-
`r0_gt1`	0 (0%)	1: 62.09% 0: 37.91%	-
`CHT_adj`	7 (2.29%)	0: 53.18% 1: 46.82%	0: 53.92% 1: 46.08%
`FONG_score`	29 (9.48%)	0: 5.42% 1: 19.86% 2: 23.83% 3: 34.66% 4: 15.16% 5: 1.08%	0: 4.90% 1: 18.30% 2: 25.16% 3: 36.93% 4: 13.72% 5: 0.98%
`FONGsimpl_score`	29 (0.48%)	1: 49.10% 2: 50.90%	1: 48.37% 0: 51.63%
`GAME_score`	41 (13.40%)	0: 3.77% 1: 13.21% 2: 27.55% 3: 29.06% 4: 16.98% 5: 6.79% 6: 2.26% 7: 0.38%	0: 3.27% 1: 11.44% 2: 29.74% 3: 30.39% 4: 16.01% 5: 6.86% 6: 1.96% 7: 0.33%
`GAMEsimpl_score`	41 (13.40%)	1: 16.98% 2: 56.60% 3: 26.41%	1: 14.71% 2: 60.13% 3: 25.16%
`Mclinical_score`	32 (10.46%)	0: 14.60% 1: 46.35% 2: 34.67% 3: 4.38%	0: 13.07% 1: 48.04% 2: 34.97% 3: 3.92%

Table 1.8: Summary Statistics of Categorical Clinical Postoperative and Prognostic Score Variables

1.3. Selection of Radiomics Variables

As mentioned above, radiomic variables are inherently correlated, necessitating variable selection to reduce their high dimensionality. A filtering method based on Spearman correlation analysis was applied to exclude redundant variables.

In particular, the following procedure was applied. Recall that the radiomic features are extracted from both the core and the ring regions and organized into distinct groups. A set of features provided by the clinicians, referred to as the clinician-selected set, is also available and represents features considered interpretable and important in literature. The set of notable features is composed by *Entropy*, *Uniformity*, *Energy*, *Kurtosis*, *Skewness* variables.

For each group (considered separately for core and ring), the absolute Spearman correlation matrix is computed. The features within each group are then processed in their original order. Since features appearing earlier in the order are generally more interpretable, the procedure typically removes the second feature in a pair if their correlation is high ($|\text{correlation}| \geq 0.80$). However, if the second feature belongs to the clinician-selected set while the first does not, then the first feature is removed instead. This approach maintains interpretability by favoring features that appear earlier in the order while preserving those prioritized by clinicians. After processing each group individually, the remaining features from all groups are combined, and the same procedure is applied to the aggregated set to further eliminate any highly correlated features.

The complete procedure is detailed in the Algorithm A.1 described in Appendix A.2.

The threshold of 0.80 was selected as a compromise between maintaining a manageable number of features and keeping inter-feature correlations low. In Appendix A.2, Figure A.1 and Figure A.2 show how the retained feature count varies with the threshold, highlighting that 0.80 represents a stable point, as the number of features remains nearly constant even at slightly higher thresholds. This procedure drastically reduced the total number of features, as reported in the Table 1.9 below.

	n var Core preChemo	**n var Ring preChemo**	**n var Core preSurgery**	**n var Ring preSurgery**
pre-elimination	109	109	109	109
post-elimination	20	25	20	23

Table 1.9: Number of Selected Radiomic Features

In Figure 1.4 and Figure 1.5 the final correlation matrices obtained are reported for core and ring, respectively. The original correlation matrices are reported in Appendix A.2.

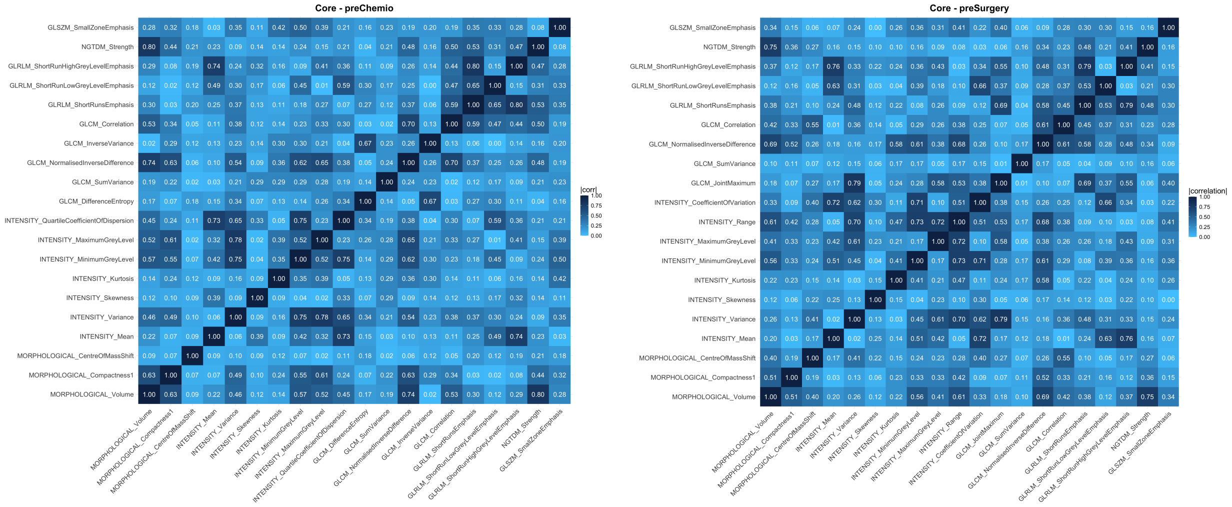


Figure 1.4: Final Correlation Matrix of core radiomics variables for Pre-Chemotherapy (left) and Pre-Surgery (right) cases

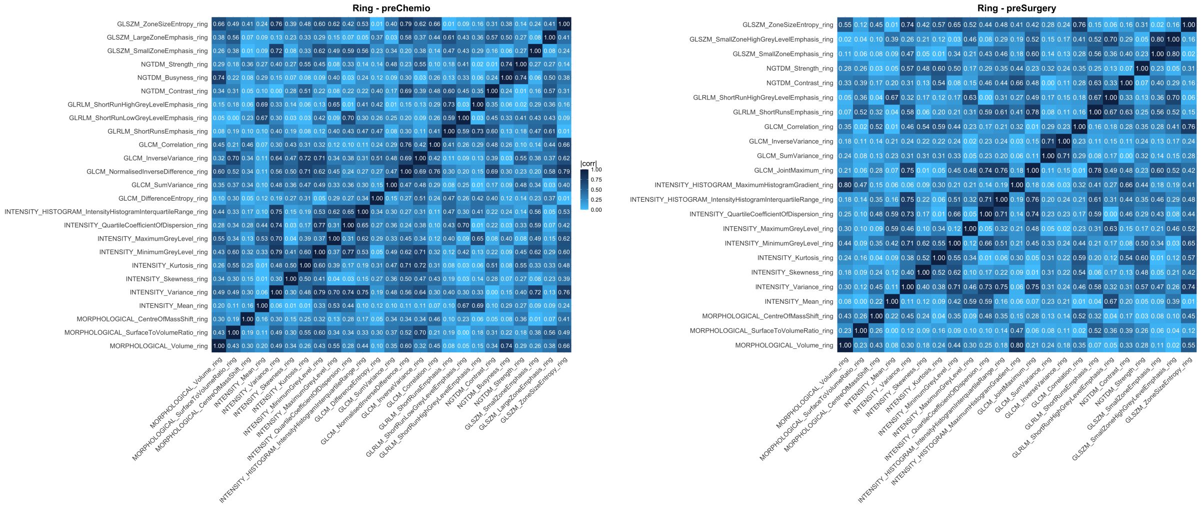
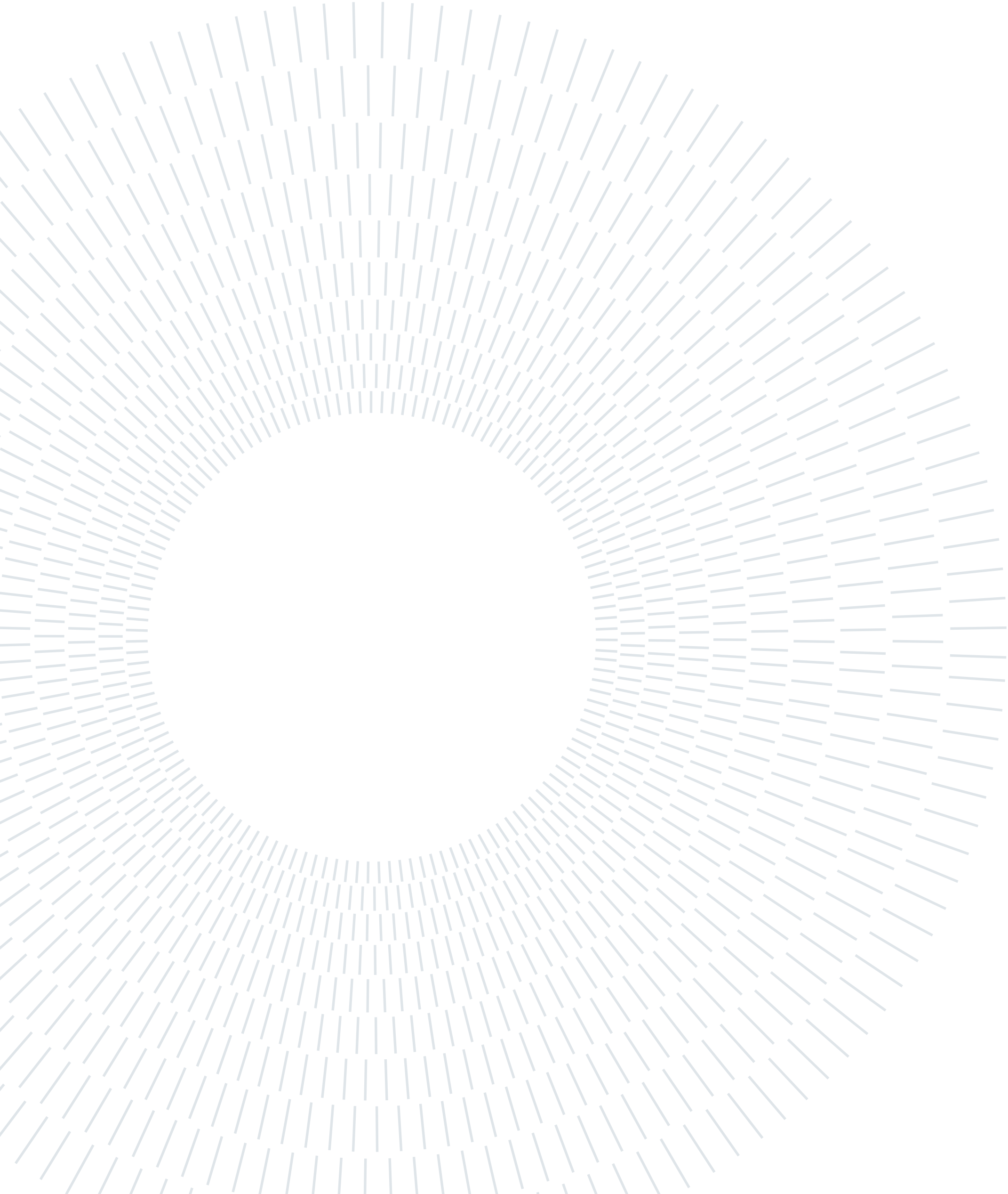


Figure 1.5: Final correlation matrix of ring radiomics variables for Pre-Chemotherapy (left) and Pre-Surgery (right) cases



2 | Methodological Pipeline

In this chapter, the methods used for the analyses are presented. In particular, Section 2.1 focuses on the analysis of primary tumor metastasis using survival analysis, while Section 2.2 describes the examination of intra-patient metastasis heterogeneity, for both pre-Chemo and pre-Surgery CT scans.

Regarding survival analysis, two distinct approaches were used. The first one evaluates the prognostic impact of all available data types (i.e. clinical, genetic, core radiomics and ring radiomics) on patient survival. Given the computational and financial demands of radiomic analyses, one objective is to determine whether the core and ring radiomics of the largest metastasis provide meaningful prognostic information that justifies their use. To achieve this, a series of 12 survival models were developed by progressively incorporating additional variables. Specifically:

pre-chemo models Six models were constructed using the subset of patients who underwent neoadjuvant chemotherapy, with radiomic data acquired before chemotherapy.

pre-surgery models Six models were constructed using the entire cohort, with radiomic data acquired prior to surgery.

The second approach compares the aforementioned models with simpler models that rely solely on the clinical scores described in Chapter 1 for predicting patient survival. The goal is to assess whether clinical scores alone are sufficient for accurate outcome estimation or if the integration of additional patient data yields improved predictions.

All the previously mentioned models were developed using Random Survival Forests [15]. Due to the high number of variables, feature selection was performed with the Boruta [19] algorithm, which is particularly effective for Random Forest-based problems.

Concerning the analysis of intra-patient metastasis heterogeneity, the Tree-Clustering method proposed in [5] was applied to study the heterogeneity of metastases both before neoadjuvant chemotherapy and before resection procedure.

2.1. Survival Models

Overview of Random Survival Forest

Random Survival Forest (RSF) [15] is an extension of the traditional Random Forest algorithm tailored for survival analysis, where the objective is to model the time until an event (e.g. death) occurs. Unlike classical parametric models, such as Cox regression [6] (which assumes proportional hazards and a specific functional form for the relationship between predictors and the hazard rate) RSF is non-parametric and does not impose such restrictions. This makes RSF particularly adept at handling complex and non-linear relationships in the data.

Random Survival Forest operates by constructing an ensemble of decision trees, each built on a bootstrap sample of the original dataset and a randomly selected subset of variables. This approach allows the method to capture intricate interactions and non-linear relationships between predictors without overfitting. Each tree in the forest is designed to handle censored data, i.e. observations where the event of interest has not occurred by the end of the study period.

Predictions are generated by aggregating the survival estimates from all the trees, resulting in a robust and stable estimate of the survival function. In addition, RSF provides valuable insights through Importance Plots, which quantify the contribution of each predictor to the model's performance and help identify the key prognostic factors. Furthermore, RSF produces Partial Effect Plots, which illustrate the marginal effect of individual predictors on the survival outcome, allowing for a more detailed exploration of how variations in a specific variable influence survival, while accounting for other covariates.

Overall, flexibility, ability to handle censored data, and robustness against overfitting make Survival Random Forest a powerful tool in survival analysis, offering both accurate predictions and insightful interpretations of complex biomedical data.

Despite its advantages, Random Survival Forest has certain limitations. The method can be computationally intensive, especially with large datasets and numerous predictors, potentially leading to longer training times compared to simpler models like Cox regression. Moreover, while RSF offers flexibility in modeling complex relationships, it may be less interpretable than traditional regression models, making it challenging to translate the model's insights into clear clinical guidelines.

More in details, the overall survival time for a patient i can be either an observed time O_i (e.g., if the event occurs during the follow-up period) or a censored time C_i. In this setting, let δ_i be the censoring indicator variable, which takes value of 1 when the time is observed, and 0 otherwise. Hence, the time-to-event T_i for the patient i is defined as:

$$T_i = \begin{cases} O_i & \text{if } \delta_i = 1, \\ C_i & \text{if } \delta_i = 0 \end{cases}$$

Finally, RFS aims to estimate the survival function for each patient i, which is defined as the probability that the time-to-event T_i exceeds a given time t, conditional on the covariate profile X_i:

$$S_i(t|X_i) = P(T_i \geq t|X_i)$$

Overview of Boruta Algorithm

Boruta [19] is a feature selection method designed to identify truly important variables for predictive modeling, and it integrates naturally with tree-based approaches like Survival Random Forests. In practice, Boruta works by creating "shadow" features (i.e. randomized copies of the original predictors) and then comparing the importance of each real feature against these shadows. A predictor is considered important if its importance consistently exceeds that of the best-performing shadow feature across multiple iterations.

When applied to a Survival Random Forest framework, Boruta helps in isolating the prognostically significant variables from potentially high-dimensional datasets that include clinical, genetic, and radiomic information. This is particularly useful in survival analysis, where the presence of censored data and complex interactions among variables can obscure the underlying relationships. By filtering out the noise and retaining only the relevant predictors, Boruta not only improves the interpretability of the model but also enhances its predictive accuracy.

Boruta results are commonly shown using the so called Boruta plot: a boxplot-based representation that displays the importance scores of each feature along with those of the randomly created “shadow” features. This visualization facilitates the identification of truly important variables, as it clearly shows which predictors consistently outperform randomized features.

Model Performances of Survival Models

Performances were assessed by computing the Concordance Index (C-index) [12]. This is a widely used measure in survival analysis that quantifies the predictive accuracy of a model by reflecting its ability to correctly rank the survival times of observations, even in the presence of censoring. In other words, a high C-index indicates that the model more reliably predicts which subject will experience the event (e.g., death) earlier than others. A C-index of 1 represents perfect concordance, while a value of 0.5 implies predictions no better than random chance.

In this work, the C-index was computed using 10-fold Cross Validation, repeated 10 times, to provide a more robust assessment of the model's performance in predicting survival times.

Kaplan-Meier Curves

As a final step, Kaplan-Meier [17] were generated to estimate and compare the survival functions of different subgroups within the study cohort. Kaplan-Meier curve provides a nonparametric estimate of the survival function $S(t)$ for a given population, representing the probability that a subject survives beyond time t. The estimator is defined as:

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right)$$

where d_i is the number of events (e.g., deaths) at time t_i and n_i is the number of subjects at risk just prior to t_i (that is, those who have neither experienced the event nor been censored prior to t_i). This method is particularly useful in presence of censored data, as it accurately accounts for subjects whose final event time is unknown. Moreover, Kaplan-Meier analysis is a fundamental tool in survival analysis since it enables the estimation of survival probabilities overtime and provides a clear visualization and statistical comparison of the survival distributions among different groups.

For the key variables identified by RSF, the goal was to examine differences in survival between subgroups. For numeric predictors, each variable was dichotomized standing on its median value, thus splitting the population into two groups (high versus low) to assess whether higher or lower predictor values are associated with different survival outcomes. In the case of categorical predictors, Kaplan–Meier curves were generated for each category, allowing for a direct comparison of survival distributions between the different factor levels.

Pre-Chemo and pre-Surgery Models

In this section, the construction of the pre-chemo and pre-surgery models is described. First, it is important to note that pre-chemo models utilized a cohort of 102 patients who underwent neoadjuvant chemotherapy, for whom radiomic features were available; the second set employed the entire cohort of 306 patients, using radiomic features extracted from PET scans acquired before surgery.

For the pre-chemo models, the objective was to predict patient survival based solely on information available before the beginning of therapy. Consequently, variables collected at a later stage, such as the Tumor Regression Grade (TRG), were excluded from the analysis - except for the genetic variables, which, although collected via biopsy after surgery, were assumed to be obtainable prior to chemotherapy and were therefore included.

Similarly, for the pre-surgery models, all information collected after surgery was excluded, focusing instead on the analysis of patient survival around the time of the intervention (with the exception of genetic variables).

The six models for each set were constructed similarly, with the number of variables gradually increasing, and they include:

model 1 clinical variables

model 2 clinical variables + core radiomics

model 3 clinical variables + core radiomics + ring radiomics

model 4 clinical variables + genetic variables

model 5 clinical variables + core radiomics + ring radiomics + genetic variables

model 6 clinical variables + core radiomics + genetic variables

For clarification, the complete lists of clinical variables used for pre-Chemo and pre-Surgery models is provided in Table 2.1.

To ensure an optimal feature set for the pre-Chemo and pre-Surgery models, a multi-step variable selection process was implemented using the Boruta algorithm. This approach, described in Figure 2.1, systematically refines the dataset by removing irrelevant or redundant features, thereby improving model performance and interpretability.

Initially, Boruta was applied separately to the different types of variables: Clinic, Core Radiomics, Ring Radiomics and Genetic. This step aimed to identify the most relevant features within each category while discarding less informative ones. The selected variables from these independent analyses were then combined to construct the model datasets.

To further refine the feature set and reduce dimensionality, a second Boruta-based selection was performed on the combined dataset. This additional filtering step ensured that only the most significant variables were retained for model training.

Clinical Variables of pre-Chemo
`sex`
`age`
`stageT34`
`sedePrimitivo`
`sinc`
`nMet`
`numClassi`
`diam_preChemo`
`CEA_preChemo`
`malattiaExtrahepSinc`

(a)

Clinical of pre-Surgery
`deltaData`
`sex`
`age`
`stageT34`
`sedePrimitivo`
`sinc`
`nMet`
`numClassi`
`diam_preSurgery`
`diam_preSurgery_gt5`
`CEA_preSurgery`
`malattiaExtrahepSinc`
`CHT_neoadj`
`rispostaUltimaLinea_mod`

(b)

Table 2.1: Clinical Variables of pre-Chemo (a) and pre-Surgery (b)

A key objective of this approach was to maintain the selected variables as similar as possible across different models, allowing for more direct comparisons between them. By ensuring consistency in the feature selection process, this method enhances the interpretability of the results and facilitates meaningful cross-model analysis.

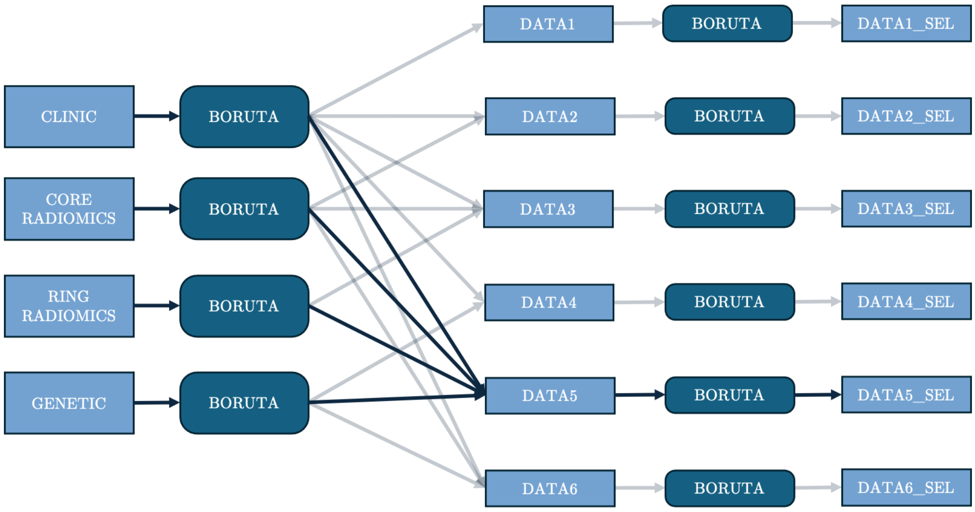


Figure 2.1: Variable selection process using the Boruta algorithm. Initially, Boruta was applied separately to Clinic, Core Radiomics, Ring Radiomics, and Genetic variables. The selected features were then merged, and a second selection step was performed to further reduce dimensionality. As an example, steps specific to model 5 are highlighted in bold

2.2. Heterogeneity Analysis

In this analysis, the focus is on intra-patient metastases heterogeneity, i.e., analyzing for each patient the homogeneity or heterogeneity of their metastases, and stratifying the patients accordingly.

To assess the intra-patient heterogeneity of metastases, the novel tree clustering method proposed in [5] was applied. This non-invasive approach is based on the quantitative analysis of radiomic features extracted from metastatic lesions (separately using both pre-Chemo and pre-Surgery) and is structured in several distinct stages.

Initially, a dimensionality reduction based on correlation analysis was performed, as described in Section 1.3. Subsequently, a tree-based representation was employed to model the heterogeneity within each patient. Specifically, the reduced feature vectors corresponding to all lesions were considered as points in a multidimensional space, and pairwise distances between these points were computed. As for the choice of the distance metric, several attempts were made by testing different distance metrics, including Euclidean, Canberra, Cityblock and Cosine distance. These distance matrices formed the basis for an agglomerative Hierarchical Clustering [32] procedure (implemented with average link-

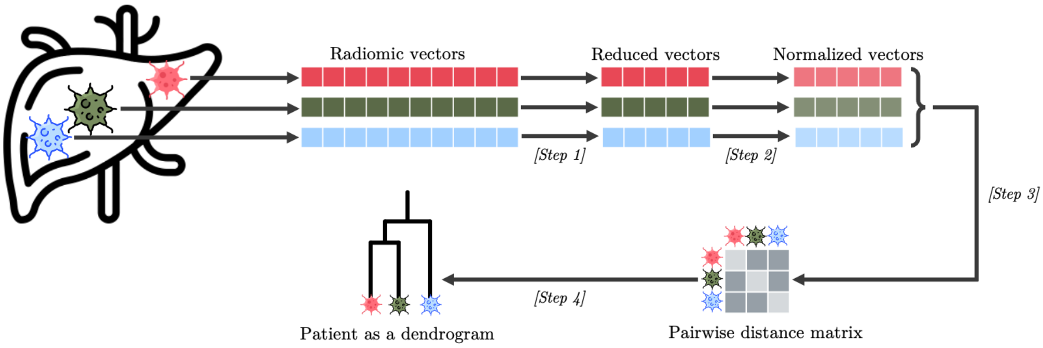


Figure 2.2: Patient representation pipeline: *[Step 1]* lesions' radiomic vectors of each patient are dimensionally reduced according to correlation analysis. *[Step 2]* The resulting features are then standardized using z-score normalization. *[Step 3]* Pairwise distances among patients' lesions are computed, and finally *[Step 4]* Hierarchical Clustering with average linkage is applied to the distance matrices, producing a dendrogram which represents each patient

age) that yielded a dendrogram for each patient. In these dendrograms, branch lengths quantitatively reflect the dissimilarity among lesions, with shorter branches indicating higher similarity (homogeneity) and longer branches signifying greater heterogeneity. A clear visual explanation of this patient representation pipeline is provided in Figure 2.2.

Once the dendrograms are collected, tree-edit distance [21] matrix between these trees is computed. Given the limited number of metastases available per patient in our dataset, the standard tree edit distance was applied rather than the pruned version proposed in [5]. This metric quantifies the minimum cost required to transform one dendrogram into another through a series of edit operations (named edit path) — such as shrinking, insertion, and deletion of edges — with the cost proportional to the differences in branch lengths. Formally, the tree edit distance d_E between two dendrograms T, T' is therefore the total cost of the optimal edit path:

$$d_E(T, T') = \inf_{\gamma \in \Gamma(T,T')} cost(\gamma)$$

where $\Gamma(T, T')$ denotes the set of all possible edit paths that transform T into T'. The computation of this edit distance is carried out using a Linear Integer Programming [40] approach, which requires considerable computational resources.

Through the computation of the tree-edit distance matrix, patients were stratified based on the degree of heterogeneity of their metastases (that is, the similarity between their representative trees). This stratification was performed using two distinct clustering meth-

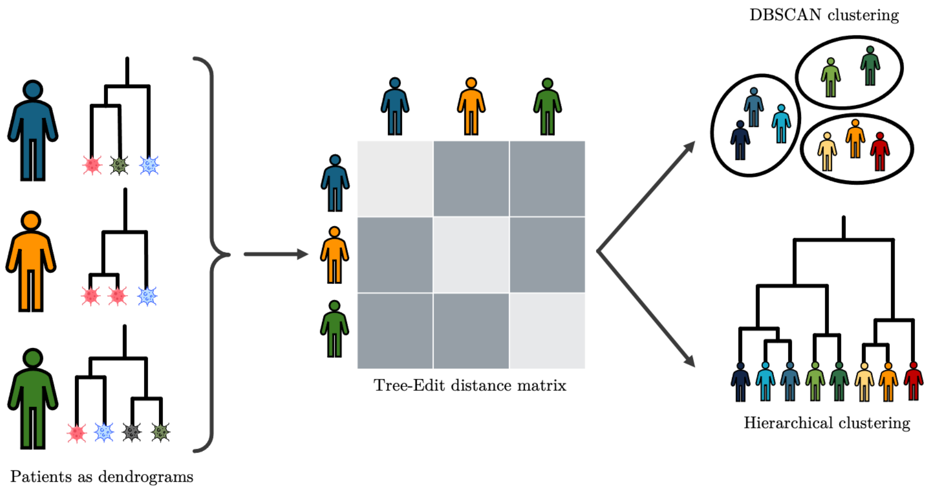


Figure 2.3: Clustering procedure for patient stratification

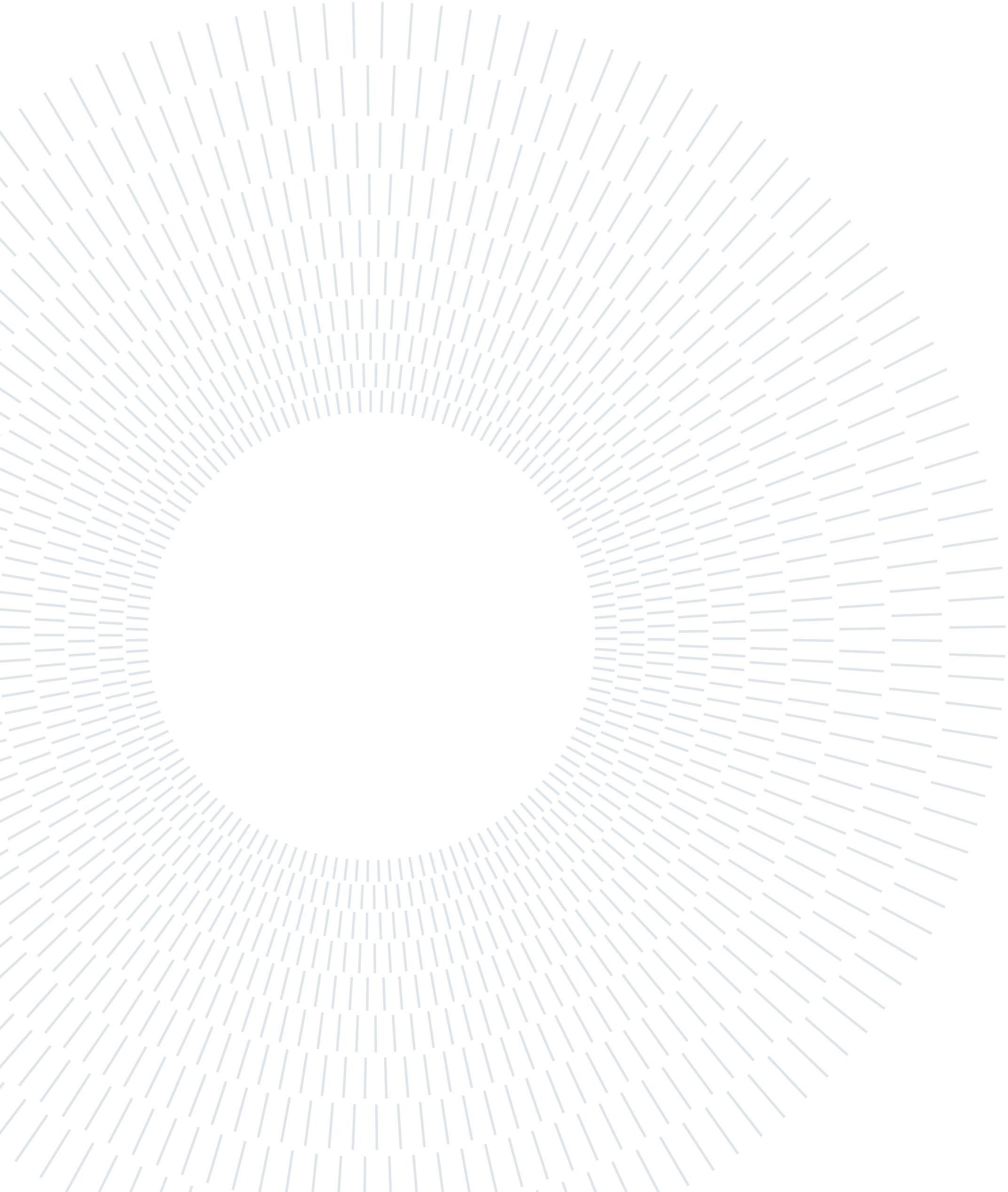
ods on the tree-edit distance matrix: Hierarchical Clustering and Density-Based Spatial Clustering of Applications with Noise [33] (DBSCAN) clustering. This procedure is exemplified in Figure 2.3.

Regarding Hierarchical Clustering, several linkage methods (e.g., single, average, complete, weighted, centroid, median, and Ward's) were explored to determine which one more clearly separates the patients. The cut of the dendrogram was consequently chosen.

On the other hand, DBSCAN requires two key hyperparameters: the minimum number of points (minPts) and the neighborhood radius (eps). In this analysis, minPts was set a priori to 10. Regarding eps, a common approach using a k-distance plot was employed. In this procedure, the k-nearest neighbors for each data point were computed with k=minPts. The eps value is then chosen as the k-distance corresponding to the point in the plot where the slope becomes relatively low (i.e., the "elbow" of the curve).

To further explore the spatial distribution of patients based on the computed distances, a Multidimensional Scaling [38] (MDS) procedure was applied to the tree-edit distance matrix. MDS is a dimensionality reduction technique that projects high-dimensional data into a lower-dimensional space (in this case, two dimensions) while preserving the pairwise distance relationships. This representation provides an intuitive visualization of how patients cluster together according to the heterogeneity of their metastases.

As a final step, Kaplan-Meier [17] curves were compared to assess differences in survival across the identified clusters. The resulting cluster labels were then integrated into the survival models introduced in Section 2.1 to assess the impact of these stratifications on survival prediction.



3 | Results

This chapter presents a comprehensive discussion of the main results obtained from the survival analysis and heterogeneity analysis, as detailed in Section 2.1 and Section 3.2, respectively.

3.1. Survival Models

3.1.1. Pre-Chemotherapy Models

As mentioned above, the pre-chemo models are based on a cohort of 102 patients who underwent neoadjuvant chemotherapy, with available radiomic features from pre-treatment CT scans. In particular, 6 models have been performed integrating increasingly the types of data available:

model 1 clinical variables

model 2 clinical variables + core radiomics

model 3 clinical variables + core radiomics + ring radiomics

model 4 clinical variables + genetic variables

model 5 clinical variables + core radiomics + ring radiomics + genetic variables

model 6 clinical variables + core radiomics + genetic variables

For each model, a feature selection was performed as described in Figure 2.1 and the complete set of Boruta algorithm plots can be found in Section B.1 of Appendix.

Although applying Boruta in the initial selection of clinical and genetic variables did not select any relevant features, all variables from these two categories were nevertheless included in the subsequent selection; the same choice was made in the selection phases for model 1 and model 4.

A complete list of final selected variables for each model is provided in Table 3.1.

model 1	`age, sedePrimitivo, CEA_preChemo, sinc, malattiaExtrahepSinc, diam_preChemo, stageT34, sex, nMet, numClassi`
model 2	`INTENSITY_QuartileRangeOfDispersion, GLRLM_ShortRunHighGreyLevelEmphasis, INTENSITY_Variance`
model 3	`INTENSITY_QuartileRangeOfDispersion, GLRLM_ShortRunHighGreyLevelEmphasis`
model 4	`age, malattiaExtrahepSinc, CEA_preChemo, sedePrimitivo, sinc, BRAF, diam_preChemo, stageT34, sex, RAS, numClassi, nMet`
model 5	`INTENSITY_QuartileRangeOfDispersion, GLRLM_ShortRunHighGreyLevelEmphasis, INTENSITY_Variance, INTENSITY_QuartileRangeOfDispersion_ring`
model 6	`INTENSITY_QuartileRangeOfDispersion, GLRLM_ShortRunHighGreyLevelEmphasis, INTENSITY_Variance`

Table 3.1: Final list of pre-chemotherapy variables for each model obtained after applying the two-step Boruta selection. The features are ranked based on their importance as determined by Boruta

It can be observed that radiomic features appear to be important for survival prediction, as several radiomic variables were selected by Boruta. Conversely, this is not the case for the clinical and genetic variables, which were forcibly included in models 1 and 4 even though Boruta did not identify any of them as relevant.

Once the features were selected, a Random Survival Forest model was used for each set of features to predict the survival time, i.e. the overall survival after the surgery. Performance was assessed by calculating the 95% Confidence Interval of the C-Index for each model using 10-fold cross-validation repeated 10 times to ensure a more robust estimation. In Figure 3.1 a visual comparison of these intervals is provided, while in Table 3.2 each Confidence Interval is explicitly reported.

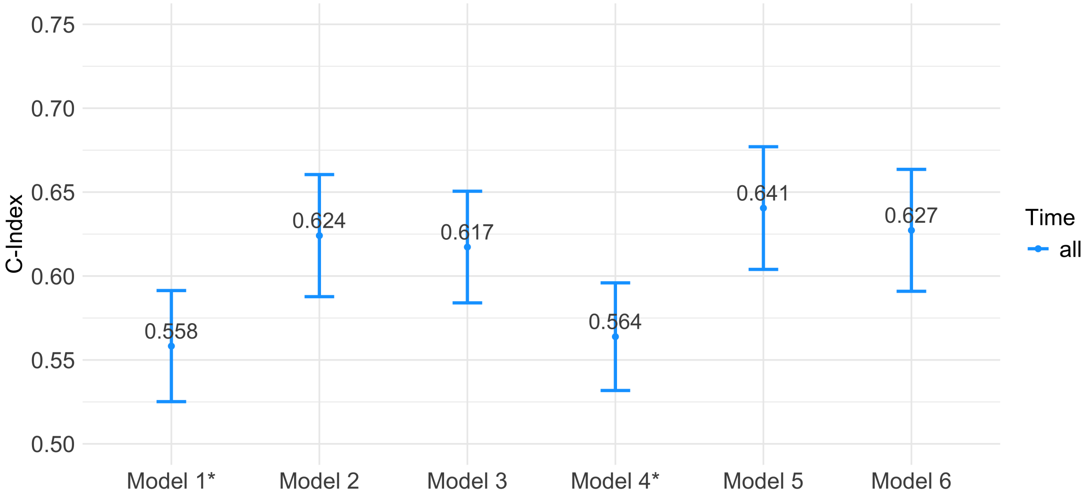


Figure 3.1: Comparison of the 95% CI of C-Index for pre-Chemotherapy models. CI were computed using 10-Fold Cross Validation repeated 10 times. Models marked with an asterisk are those in which the variables were forcefully included, since Boruta did not select any variables

	Lower Bound	Mean	Upper Bound
model 1	0.525	0.558	0.591
model 2	0.588	0.624	0.660
model 3	0.584	0.617	0.651
model 4	0.532	0.564	0.596
model 5	0.604	0.641	0.677
model 6	0.591	0.627	0.664

Table 3.2: 95% Confidence Intervals of C-Index for pre-Chemotherapy models

Comparing the models, it can be observed that clinical and genetic information alone is not sufficient for a reliable survival prediction. In contrast, the addition of radiomic variables improves the estimation, increasing the C-index by 0.1 points. The best-performing model is model 5, which includes all types of variables, although its performance is very similar to model 2, where only core radiomic variables were included. In fact, when comparing model 2 and model 3, it becomes evident that including ring variables does not provide additional information. This suggests that core variables alone are sufficient to enhance survival prediction and that the information contained in ring radiomics is likely already captured by the core radiomic features.

However, it is important to note that the C-index estimates remain relatively low, even in the best-performing models, suggesting that the available information is still insufficient for an accurate prediction of survival.

Subsequently, importance plots generated from random survival forest models were used to identify the variables that had most impact on the prediction of the outcome. Figure 3.2 presents the importance plot for model 5, while the importance plots for the other models are available in the Appendix B.1.

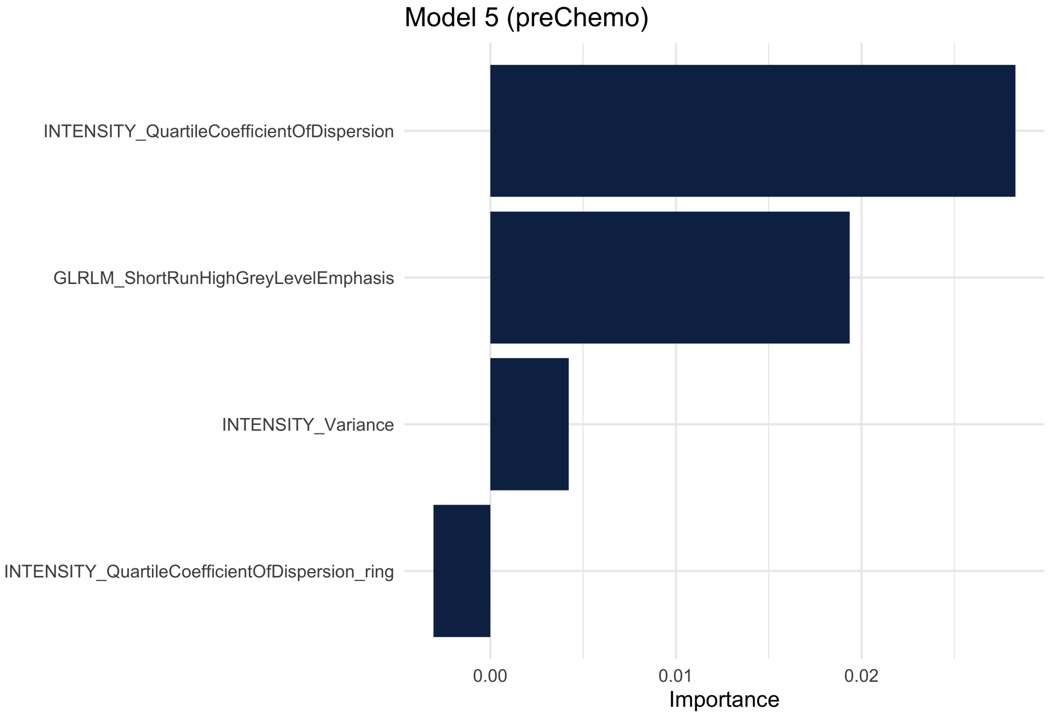
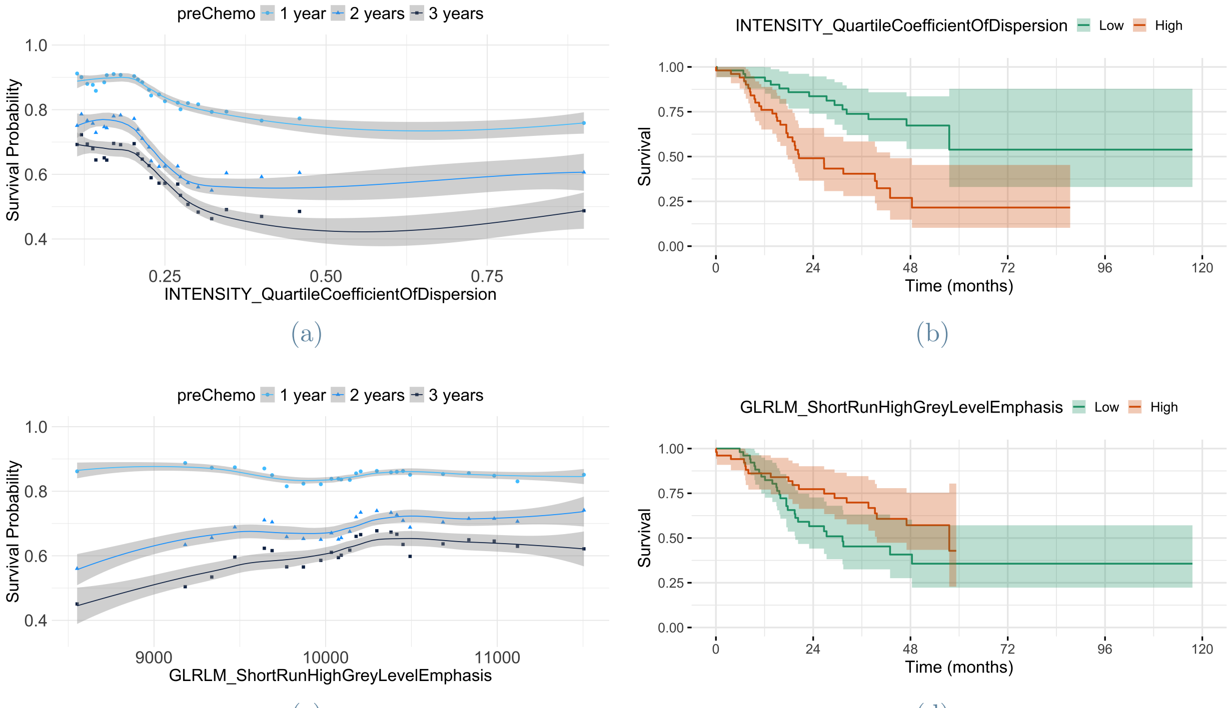


Figure 3.2: Importance Plot of Model 5 (pre-Chemotherapy)

Observing the variable ranking confirms our suspicion about the limited usefulness of the ring's radiomic variables: in model 5, the only radiomic variable from the ring has a negative importance, suggesting that, relative to the other variables, it tends to mislead the model and cause incorrect predictions.

As the final step of the analysis, Partial Effect plots and Kaplan-Meier curves for the most important variables were generated to examine their impact on survival. For model 5, the two most important variables are presented in Figure 3.3. It is worth noting that, to plot the Kaplan–Meier curves, each numerical variable was dichotomized standing on its median, thereby splitting the data into *high* and *low* groups.



(a) (b) (c) (d)

Figure 3.3: Model 5 (pre-Chemo) Partial Effect plot and Kaplan-Meier of the two most important variables according to RSF. Partial Effect plot (a) and Kaplan-Meier curve (b) of variable `INTENSITY_QuartileCoefficientofDispersion` (median split value of 0.22). Partial Effect plot (c) and Kaplan-Meier curve (d) of variable `GLRLM_ShortRunHighGreyLevelEmphasis` (median split value of 10145.30). For each curve, the Confidence Intervals are represented

To have a better understanding of Figure 3.3, it is important to first interpret the two variables, which are used to characterize the tumor's texture and provide insights into its homogeneity or heterogeneity:

`INTENSITY_QuartileCoefficientOfDispersion (QCoD):` this variable quantifies the dispersion of pixel intensities within the tumor, serving as a direct indicator of intra-tumor heterogeneity. Higher `QCoD` values reflect greater variability in intensity, which is interpreted as increased heterogeneity.

`GLRLM_ShortRunHighGreyLevelEmphasis (SRHGE):` this feature is derived from the Grey Level Run Length Matrix (GLRLM) and measures the presence of short runs of high-intensity pixels within the image. A high `SRHGE` value indicates that many small, high-intensity regions are present, suggesting a more uniform (homogeneous) tissue texture.

Hence, the Kaplan-Meier curves and Partial Effect plots presented in Figure 3.3 essentially communicate the same clinical message: tumors that are more homogeneous (as indicated by high `SRHGE` and low `QCoD` values) are associated with higher survival, while greater heterogeneity (low `SRHGE` and high `QCoD`) correlates with a poorer prognosis.

In particular, the Kaplan-Meier curve for *High* group of `SRHGE` exhibits a higher survival probability over time with respect to the *Low* group, while for `QCoD` the *High* group shows lower survival rates. Log-rank tests were performed to assess the differences between the survival curves, and very low p-values were obtained (8.13×10^{-5} for `QCoD` and 0.07 for `SRHGE`), indicating that the differences between the groups are statistically significant (with p-valued below the 0.10 threshold).

On the other hand, Partial Effect plots illustrate how the survival probability changes continuously as the value of a radiomic variable varies. The non-linearity of the estimated functions suggests that the relationships between the two variables and survival are complex. For example, regarding `QCoD`, lower values patients exhibit higher predicted survival probabilities. However, beyond a certain threshold, increases in `QCoD` lead to a steep decline in survival probability. Notably, the negative effect of higher `QCoD` values becomes progressively more pronounced over time. In other words, as follow-up time increases, the worsening prognosis associated with high `QCoD` values is even more evident, underscoring the importance of intra-tumor heterogeneity as a prognostic marker. A similar observation can be applied to `SRHGE`: high values are associated with higher survival probability, whereas lower values predict a poorer outcome, especially as time progresses. For instance, at the 1-year, the survival curve remains nearly constant, indicating no significant difference in survival between patients with high and low `SRHGE` values. However, as follow-up time increases, the divergence becomes more pronounced, with patients exhibiting lower `SRHGE` values showing a steadily worsening prognosis.

3.1.2. Pre-Surgery Models

Pre-Surgery models are based on the entire cohort of 306 patients, using the radiomic features from pre-surgery CT scans. In particular, 6 models have been performed integrating increasingly the types of data available:

model 1 clinical variables

model 2 clinical variables + core radiomics

model 3 clinical variables + core radiomics + ring radiomics

model 4 clinical variables + genetic variables

model 5 clinical variables + core radiomics + ring radiomics + genetic variables

model 6 clinical variables + core radiomics + genetic variables

For each model, a separate analysis was performed including only the 212 patients who underwent a pre-surgery CT scan no more than 30 days prior to surgery: in this study, these models are denoted with the suffix "_30gg" appended to the corresponding original model. The aim was to include in the analysis only those patients whose CT scan was truly representative of the largest metastasis, thus excluding those with outdated CT scans that did not reflect the current stage of tumor progression.

For each model, a feature selection was performed as described in Figure 2.1 and the complete set of Boruta algorithm plots can be found in Section B.2 of Appendix. Although applying Boruta in the initial selection to genetic variables did not produce any significant features, all variables from this category were nevertheless included in the subsequent selection. A complete list of final selected variables for each model (model_30gg) is provided in Table 3.3 (Table 3.4).

model 1	`nMet, diam_preSurgery, malattiaExtrahepSinc`
model 2	`nMet, MORPHOLOGICAL_Volume, INTENSITY_CoefficientOfVariation`
model 3	`nMet, MORPHOLOGICAL_Volume, NGTDM_Strength, NGTDM_Strength_ring, GLSZM_SmallZoneEmphasis_ring, GLRLM_ShortRunsEmphasis_ring, INTENSITY_CoefficientOfVariation, diam_preSurgery, GLCM_InverseVariance_ring`
model 4	`nMet, diam_preSurgery, malattiaExtrahepSinc`
model 5	`nMet, MORPHOLOGICAL_Volume, NGTDM_Strength_ring, NGTDM_Strength, GLSZM_SmallZoneEmphasis_ring, GLRLM_ShortRunsEmphasis_ring, diam_preSurgery, INTENSITY_CoefficientOfVariation, GLCM_InverseVariance_ring`
model 6	`nMet, MORPHOLOGICAL_Volume, INTENSITY_CoefficientOfVariation, NGTDM_Strength`

Table 3.3: Final list of pre-surgery variables for each model obtained after applying the two-step Boruta selection. The features are ranked based on their importance as determined by Boruta

model 1_30gg	`nMet, sedePrimitivo, malattiaExtrahepSinc, diam_preSurgery`
model 2_30gg	`nMet, sedePrimitivo, INTENSITY_Mean, GLRLM_ShortRunHighGreyLevelEmphasis, MORPHOLOGICAL_Volume, GLCM_NormalisedInverseDifference, malattiaExtrahepSinc`
model 3_30gg	`nMet, GLSZM_SmallZoneEmphasis_ring, GLSZM_SmallZoneHighGreyLevelEmphasis_ring, MORPHOLOGICAL_Volume, INTENSITY_Mean, GLRLM_ShortRunHighGreyLevelEmphasis, sedePrimitivo, GLCM_NormalisedInverseDifference, diam_preSurgery, GLRLM_ShortRunsEmphasis_ring, MORPHOLOGICAL_Volume_ring`
model 4_30gg	`nMet, sedePrimitivo, malattiaExtrahepSinc, diam_preSurgery`
model 5_30gg	`nMet, GLSZM_SmallZoneEmphasis_ring, GLSZM_SmallZoneHighGreyLevelEmphasis_ring, MORPHOLOGICAL_Volume, INTENSITY_Mean, GLRLM_ShortRunHighGreyLevelEmphasis, sedePrimitivo, GLCM_NormalisedInverseDifference, GLRLM_ShortRunsEmphasis_ring, diam_preSurgery`
model 6_30gg	`nMet, sedePrimitivo, INTENSITY_Mean, GLRLM_ShortRunHighGreyLevelEmphasis, MORPHOLOGICAL_Volume, GLCM_NormalisedInverseDifference, malattiaExtrahepSinc`

Table 3.4: Final list of pre-surgery variables for each model_30gg obtained after applying the two-step Boruta selection. The features are ranked based on their importance as determined by Boruta

It can be observed that radiomic features seem to be important for survival prediction, as several radiomic variables were selected by Boruta. Conversely, this is not the case for the genetic variables, which were forcibly included in models 4 even though Boruta did not identify any of them as relevant. In addition, the number of metastases `nMet` seems to be a highly relevant predictor, since it ranks as the most important feature in every model.

Once the features were selected, a Random Survival Forest model was used for each set of features to predict the survival time. Performance was assessed by calculating the 95% Confidence Interval of the C-Index for each model using 10-fold cross-validation repeated 10 times to ensure a more robust estimation. In Figure 3.4 a visual comparison of these intervals is provided, while in Table 3.5 each Confidence Interval is explicitly reported.

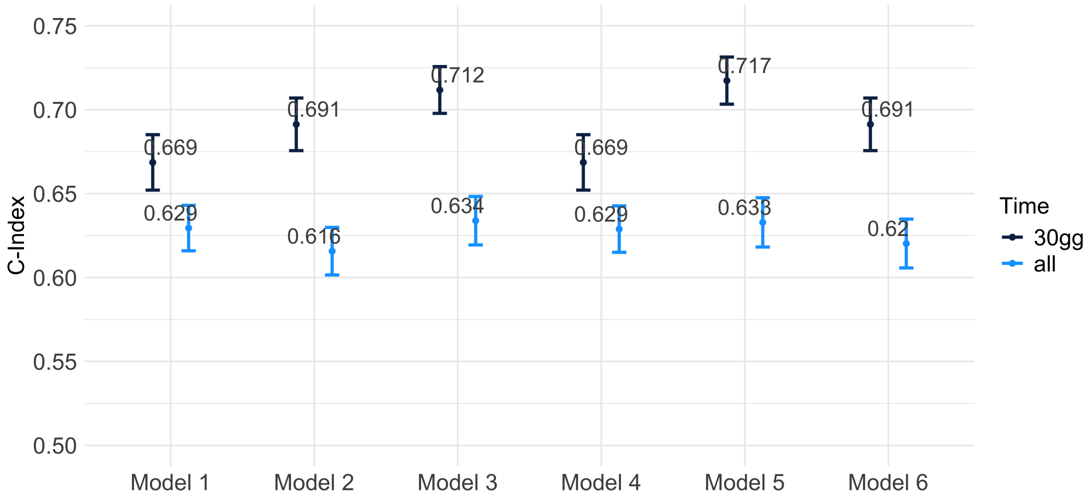


Figure 3.4: Comparison of the 95% CI of C-Index for pre-Surgery models. CI were computed using 10-Fold Cross Validation repeated 10 times

	Lower Bound	**Mean**	**Upper Bound**
model 1	0.616	0.629	0.643
model 1_30gg	0.652	0.669	0.685
model 2	0.602	0.616	0.630
model 2_30gg	0.676	0.691	0.707
model 3	0.619	0.634	0.648
model 3_30gg	0.698	0.711	0.726
model 4	0.615	0.629	0.643
model 4_30gg	0.652	0.669	0.685
model 5	0.618	0.633	0.648
model 5_30gg	0.703	0.717	0.731
model 6	0.606	0.620	0.635
model 6_30gg	0.676	0.691	0.707

Table 3.5: 95% Confidence Intervals of C-Index for pre-Surgery models

Comparing the models, it can be observed that clinical and genetic information alone is not sufficient for a reliable survival prediction. In contrast, the addition of radiomic variables improves the estimation. An outstanding observation from Figure 3.4 is that the 30-day models consistently outperform the base models, which further justifies their development. The best-performing model is model 5_30gg, which includes all types of variables, although its performance is very similar to model 3_30gg. In theory, model 5_30gg was expected to include genetic variables compared to model 3_30gg; however, these variables were never selected by Boruta and thus were not incorporated. As a result, the two models differ only by one radiomic variable `MORPHOLOGICAL_Volume_ring`. Notably in this models, adding ring radiomics information improves considerably model performances.

Subsequently, importance plots generated from random survival forest models were used to identify the variables that had most impact on the prediction of the outcome. Figure 3.5 presents the importance plot for model 5_30gg, while the importance plots for the other models are available in the Appendix B.2.

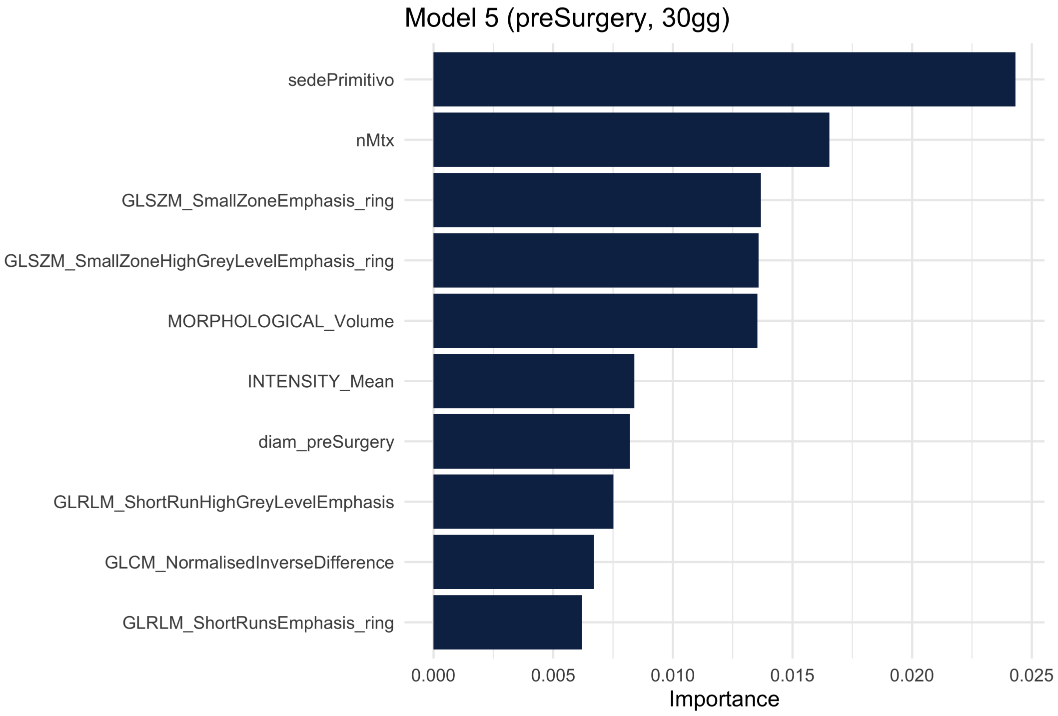
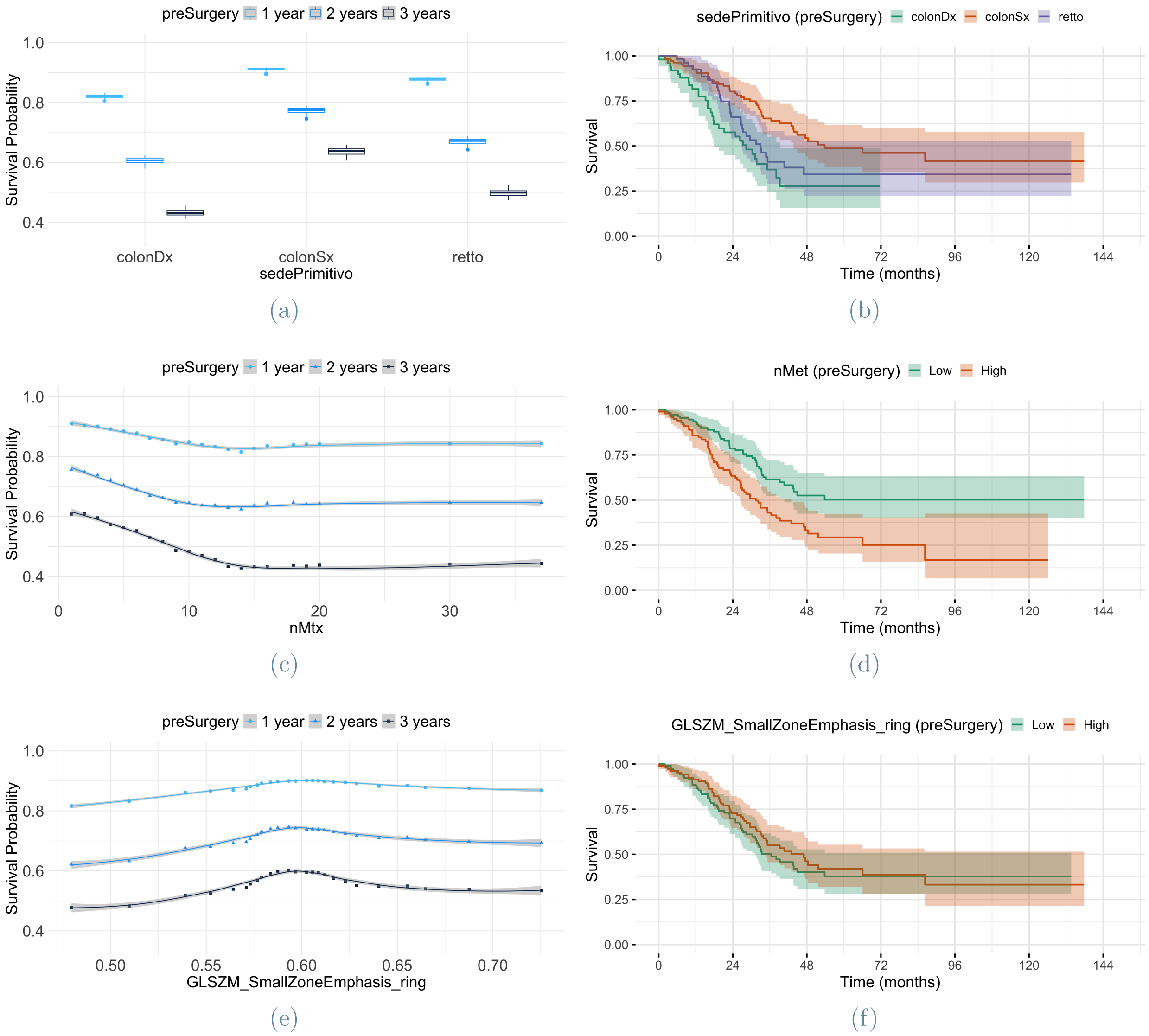


Figure 3.5: Importance Plot of Model 5_30gg

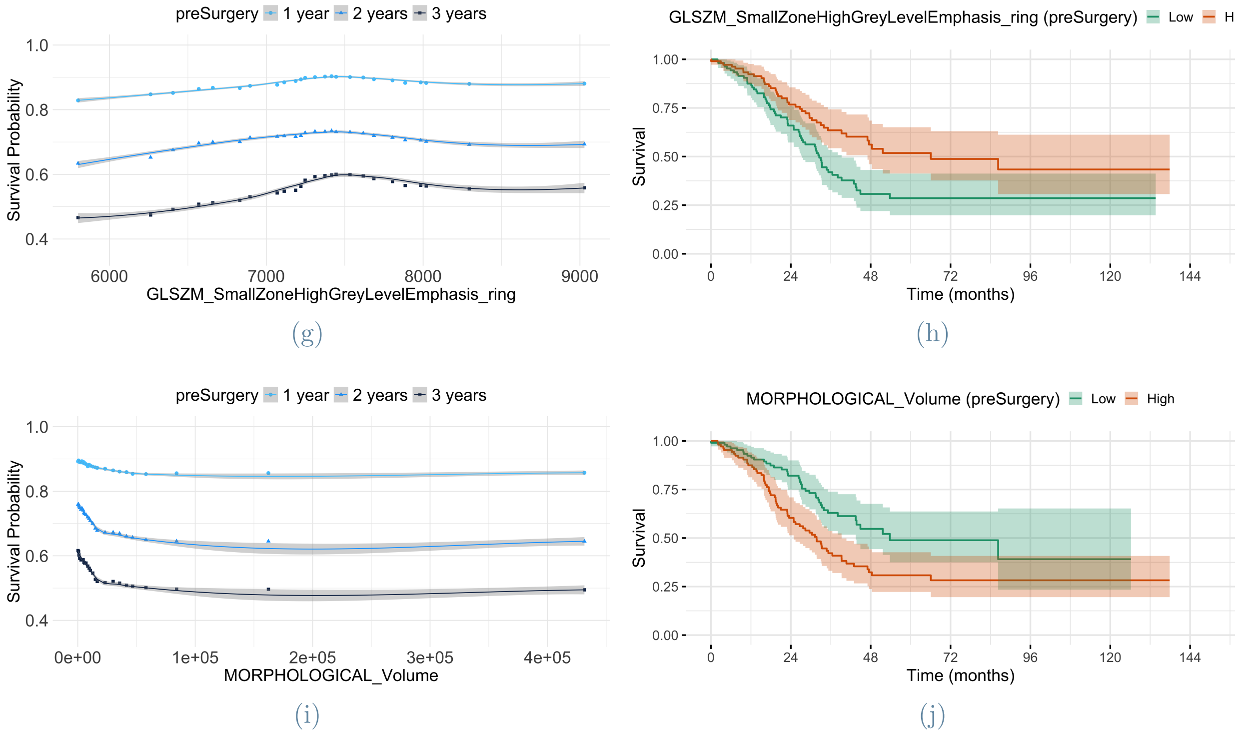
Observing the variable ranking, it is possible to note that among the five most important predictors, there are clinical, core radiomic and ring radiomic variables. Partial Effect plots and Kaplan-Meiercurves for these variables are presented in Figure 3.5 to illustrate their impact on survival. It is important to emphasize that, for the Kaplan-Meier curves of the numerical variables, these were dichotomized at their medians, thereby splitting the data into *High* and *Low* groups.



(a) (b)

(c) (d)

(e) (f)



(g) (h)

(i) (j)

Figure 3.5: Model 5_30gg (pre-Surgery) Partial Effect plot and Kaplan-Meier of the five most important variables according to RSF. Partial effect plots (a, c, e, g, i) and Kaplan–Meier curves (b, d, f, h, j) for the variables in the pre-surgery models. Specifically: (a) and (b) correspond to the variable `sedePrimitivo`; (c) and (d) to `nMet` (median split value of 3); (e) and (f) to `GLSZM_SmallZoneEmphasis_ring` (median split value of 0.597); (g) and (h) to `GLSZM_SmallZoneHighGreyLevelEmphasis_ring` (median split value of 7314.81); (i) and (j) to `MORPHOLOGICAL_Volume` (median split value of 10678.23). In each pair, the left panel shows the estimated partial effect while the right panel displays the survival curve, with confidence intervals indicated

In order to better understand Figure 3.5, it is essential to first interpret the variables. The radiomic variables cited in the figure are described below; for an interpretation of the clinical variables refer to Table 1.2.

`GLSZM_SmallZoneEmphasis_ring (SZE_ring):` this feature is derived from the Grey Level Size Zone Matrix (GLSZM) and measures the prevalence of small zones in the rumor ring. A high value signifies that small zones are very predominant in the tumor's border texture, highlighting a marked fragmentation or heterogeneity in that region.

`GLSZM_SmallZoneHighGreyLevelEmphasis_ring (SZHGE_ring):` this variable is derived from the Grey Level Size Zone Matrix (GLSZM) and quantifies the intensity emphasis within small zones located in the tumor ring (i.e., it measures how high the pixel intensities are specifically within those zones that are defined as "small"). A high value indicates that, among the small zones, the intensities are particularly high, suggesting increased activity or complexity at the tumor's periphery.

`MORPHOLOGICAL_Volume`: this feature captures the overall volume of the tumor core. A larger volume typically reflects a greater tumor burden, which can be associated with more advanced disease and a potentially poorer prognosis.

Examining the Kaplan–Meier curves for `sedePrimitivo` reveals that patients whose primary tumor is located in the left colon exhibit higher survival rates. In contrast, those with tumors in the right colon or rectum show a similar, lower survival, with the rectum representing an intermediate condition. Specifically, on shorter time horizons the rectal subgroup behaves more similarly to the left colon, whereas over the long term its survival curve aligns more closely with that of the right colon. This observation is further supported by the partial effect of the `sedePrimitivo` variable: the boxplots for the right colon consistently indicate a worse prognosis at 1, 2, and 3 years, while the rectum displays an intermediate profile, nearer to the left colon at one year and to the right colon at three years. To assess the differences between the Kaplan-Meier curves, pairwise log-rank tests with Bonferroni correction were performed applied to each pair of tumor sites: a significant difference was observed between the left colon and the right colon ($p = 0.0013$); no statistically significant difference was found between the rectum and the right colon ($p = 0.67$), whereas the comparison between the rectum and the left colon ($p = 0.09$) indicates a notable difference, even if less pronounced than that observed between left and right colon.

Similarly, the number of metastases `nMet` appears to be a critical factor in survival estimation. The Kaplan–Meier curves indicate that patients with a high number of metastases (*High* group) experience markedly lower survival compared to those with fewer metastases (*Low* group), suggesting a more advanced disease stage. Moreover, the partial effect analysis shows that the impact of the number of metastases becomes increasingly significant as the survival horizon extends, implying that having more metastases leads to poorer long-term outcomes. A log-rank test comparing the two Kaplan-Meier curves yielded a highly significant result ($p = 7.45 \times 10^{-4}$), confirming the substantial difference between their survival curves.

Looking at the Kaplan-Meier curves for *High* and *Low* groups of `SZE_ring`, it becomes apparent that the survival curves for these groups remain relatively close, with considerable overlap during the follow-up period- This overlap suggests that the difference in the survival between the two groups is not notably distinct. This is further confirmed by the log-rank, which produced a very high p-value ($p = 0.46$). Turning to the partial effect plot, a slight curvature can be observed, suggesting that both very low and very high `SZE_ring` values may be associated with lower survival probabilities. In particular, this effect becomes more pronounced as the survival horizon extends, especially at the 3-year time point, implying that patients with extreme `SZE_ring` values may experience poorer long-term outcomes. This could imply that an optimal, moderate level of tumor border fragmentation might be favorable, whereas deviations toward either extreme (indicating either overly homogeneous or excessively heterogeneous tumor peripheries) may reflect more aggressive tumor behavior and a worse prognosis. Hence, the simple conventional dichotomy of this variable produced with Kaplan-Meier curves may obscure more subtle relationships, which are more clearly evident in the partial effect plot.

For what concerns `SZHGE_ring`, looking at the Kaplan-Meier curves for *High* and *Low* values, the survival difference between these two groups appears clear, with a p-value of 2.50×10^{-3} from the log-rank test confirming statistical significance. However, while the curves show a clear separation, there is still some overlap in the confidence intervals at certain follow-up time points (in particular for short term survival). In particular, the plot suggests that having higher values of `SZHGE_ring` is associated with an higher survival rate in long term period. This is confirmed when considering the partial effect plot, which evaluates `SZHGE_ring` continuously over time, in which it is possible to notice that lower values of this feature are generally associated with worse survival outcomes. Although higher values also seem to suggest poorer survival, their effect is less pronounced. Moreover, at the 3-year, the impact of very low values becomes more marked. Clinically, these findings indicate that the presence of small, highly intense zones in the tumor periphery (as reflected by high `SZHGE_ring` values) may be a marker of a less aggressive tumor phenotype. In other words, when the tumor border displays numerous small zones with elevated intensity, it could suggest a more organized or well-defined tumor border, leading to better long-term survival. Conversely, lower `SZHGE_ring` values, which indicate fewer or less intense small zones at the tumor edge, may be associated with a more aggressive disease.

Finally, the Kaplan-Meier curves for *Low* versus *High* `MORPHOLOGICAL_Volume` show a clear separation, with the log-rank test yielding a very small p-value ($p = 1.6 \times 10^{-3}$), indicating a statistically significant difference in survival between the two groups. In particular, *High* group exhibits a lower survival, suggesting that having a larger tumor volume is associated with a worse prognosis. The partial effect plot reinforces this observation:

as the tumor volume increases, survival probabilities generally decrease, particularly over longer time horizons. Clinically speaking, a larger tumor core volume often is a symptom of an advanced disease, potentially with greater metastatic potential, and is consequently linked to poorer survival.

3.1.3. Score Models

The last model developed is a survival model that uses clinical scores to predict survival time. The rationale behind this model is that, in this study, it serves as a benchmark for subsequent comparison with the survival models developed so far. In this case, only one model was employed in which `FONGsimpl_score`, `GAMEsimpl_score` and `Mclinic_score` described in Table 1.4 were used as predictors. Given the limited number of predictors, all of them were included in a Random Survival Forest model, without applying any feature selection method.

The model's performance was evaluated using the C-index, computed via 10-fold cross validation repeated 10 times, which allowed to estimate a 95% confidence interval. The estimated confidence interval for the C-index is [0.592, 0.622] (with a mean value of 0.607).

In Figure 3.6 importance plot generated by the RSF is presented.

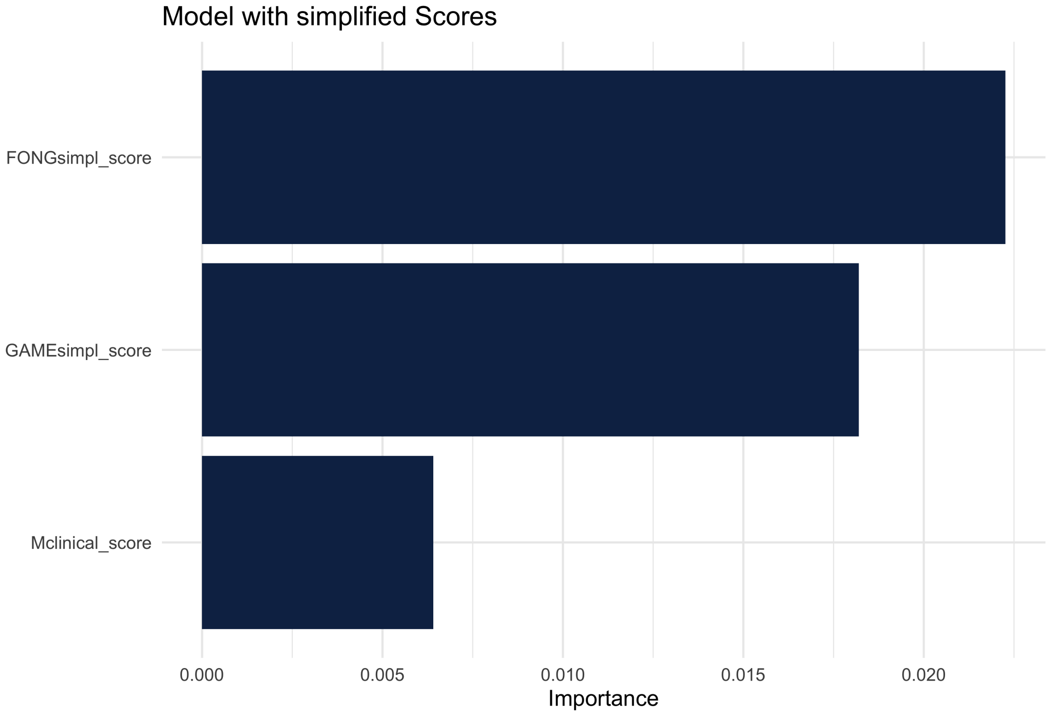
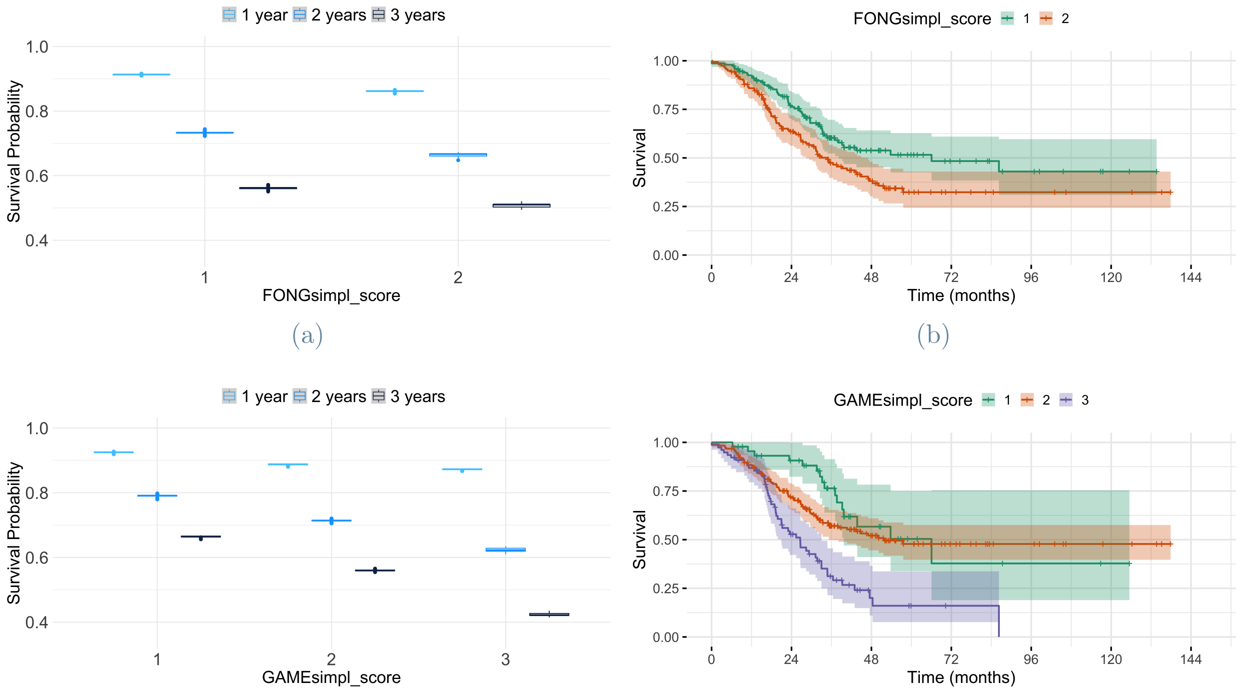


Figure 3.6: Importance Plot of score model

Notably, `FONGsimpl_score`, the most well-known risk score used for colorectal cancer, emerged as the most important feature. Finally, partial effect plots and Kaplan-Meier curve of `FONGsimpl_score` and `GAMEsimpl_score` are provided in Figure 3.7



(a) (b)

(c) (d)

Figure 3.7: Partial Effect plot and Kaplan-Meier for the two most important variables of score model according to RSF. Partial Effect plot (a) and Kaplan-Meier curve (b) of variable `FONGsimpl_score`. Partial Effect plot (c) and Kaplan-Meier curve (d) of variable `GAMEsimpl_score`. For each curve, the Confidence Intervals are represented

Concerning `FONGsimpl_score`, the Kaplan-Meier highlights a clear evidence of differing behavior between the two risk groups (1: low risk vs. 2: high risk). Specifically, the Kaplan–Meier plot shows that lower FONG values are consistently associated with better prognosis. A log-rank test comparing the two curves yielded a highly significant p-value ($p = 0.013$), confirming that the difference in survival between the two groups is statistically significant. Moreover, not only do the two curves remain well separated over the entire follow-up period, but their confidence intervals also rarely overlap. This observation supports the decision to dichotomize the original 0–5 FONG score into two risk categories. Turning to the partial effect plot for `FONGsimpl_score`, the boxplots for the low-risk group always lie above those for the high-risk group. This again confirms that patients in group 1 experience a higher survival probability at each time horizon, and this difference becomes even more pronounced as time passes.

A similar discussion can be made for `GAMEsimpl_score`. From the Kaplan-Meier curves, it appears that the green (`GAMEsimpl_score` = 1, low risk) and the red (`GAMEsimpl_score` = 2, intermediate risk) curves overlap considerably, whereas the purple (`GAMEsimpl_score` = 3, high risk) curve is clearly lower, suggesting worse survival outcomes. This visual inspection is supported by the pairwise log-rank tests: comparison 1 vs 3 and 2 vs 3 yields very small p-values ($p = 3.30 \times 10^{-5}$ and $p = 9.90 \times 10^{-5}$ respectively), while 1 vs 2 yields a high p-value ($p = 0.64$). It is also important to consider that a high censoring rate may contribute to the apparent similarity between the curves for scores 1 and 2 at later time points. In the available data for `GAMEsimpl_score` = 1, a censoring rate of 64% was observed (while for group 2 and 3 censoring rates are 57% and 34%), indicating that 64% of subjects did not experience the event during follow-up. This high censoring rate results in wider confidence intervals and less precise survival estimates at later times, which may make the curves for groups 1 and 2 appear more similar than they are if more data are available. In fact, when focusing on the early part of the curves, where the time horizon is shorter and the confidence intervals (especially for group 1) are narrower, all three curves appear to be clearly distinct. In the partial effect plot, the predicted survival probabilities at 1, 2, and 3 years for `GAMEsimpl_score` = 3 are clearly lower than those for the other two groups. In contrast, the differences between scores 1 and 2 are less pronounced, confirming the log-rank test results. Overall, these findings suggest that `GAMEsimpl_score` effectively identifies a high-risk cohort (`GAMEsimpl_score` = 3), while the survival profiles of patients with scores of 1 or 2 are relatively similar, possibly due in part to the high censoring rate in group 1.

3.2. Heterogeneity Analysis

In this section, the main results of the heterogeneity analysis performed on pre-Chemo and pre-Surgery CT scans are reported. The aim of this study was to assess whether a stratification of patients based on intra-tumor heterogeneity is possible, and whether it is predictive of survival outcomes.

For this purpose, the method outlined in Section 2.2 was applied separately on pre-Chemo and pre-Surgery radiomic datasets, and the main results are presented in Section 3.2.1 for pre-Chemotherapy analysis and in Sections 3.2.2, 3.2.3 for pre-Surgery analysis.

Before discussing the models' results, several clarifications are needed.

Firstly, it should be noted that the study discussed in [5], which proposed the Tree Clustering method, did not employ the conventional edit distance described in Chapter 1. Instead, a novel metric, named the Pruned Edit Distance, was used. This distance represents an extension of the standard edit distance by incorporating a pruning procedure for dendrogram branches. The primary objective of this metric is to reduce very large trees by consolidating multiple metastases, represented by individual leaves, into phenotypic categories. This approach was adopted because the metastases analyzed in [5] originated from prostate cancer and were inherently diverse, involving very different tissues. In that study, the number of metastases per patient was notably high (mean=6.05, median=5), necessitating the aggregation of highly similar metastases into representative phenotypic groups.

In the present study, however, the number of metastases per patient is considerably lower (pre-Chemo: mean=2.70, median=2; pre-Surgery: mean=2.54, median=2. See Table 3.6 for further details) and all the lesions are located within the liver. Consequently, the additional pruning step was retained unnecessary, and the standard tree-edit distance was employed. Preliminary tests with the pruned distance did not reveal any substantial differences compared to the original edit distance.

number of metastases	number of patients (pre-Chemo)	number of patients (pre-Surgery)
1	27	111
2	31	79
3	14	32
4	5	18
5	25	66

Table 3.6: Distribution of the number of available metastases per patients, for pre-Chemotherapy and pre-Surgery radiomics

In addition, due to the very low number of metastases, patients with only one (or, in some cases, two) available radiomics for metastasis were excluded from the heterogeneity analysis. This decision was made to prevent the method from being too influenced by the small number of metastases, ensuring that the stratification occurs based on the heterogeneity of the lesions rather than simply on their count.

Moreover, various metrics were explored to assess the distance between metastases, expanding on the approach used in the previous study (in which only Euclidean distance was explored). In particular, the following distance metrics were evaluated: Euclidean, Canberra, Cityblock, and Cosine.

Furthermore, a notable difference between Colorectal dataset and that of the study [5] is that radiomic features were available for both the tumor core and the peritumoral ring. To extend the analysis, the heterogeneity of the metastases was assessed under three cases: using only core radiomics, only ring radiomics, or a combination of both regions of interest.

In summary, the following hyperparameters were explored to apply the method:

Minimum Number of Metastases (minLes): 2 or 3

Region to Investigate (RoI): Core, Ring, Core+Ring

Distance between Metastases (dist): Euclidean, Canberra, Cityblock, Cosine

In each heterogeneity analysis (pre-Chemo and pre-Surgery), a total of 24 attempts were explored, evaluating every possible combination of the hyperparameters. In this manuscript, only the most relevant (in terms of survival stratification) and interpretable will be presented. Specifically, for the pre-Chemo analysis, the configuration with a minimum number of metastases equal to 2, a RoI set to Core, and the Euclidean distance metric (denoted as `stdCoreEu2`) was selected. For the pre-Surgery analysis, two configurations were highlighted: one with minLes=3, RoI = Core, and Euclidean distance (denoted as `stdCoreEu2`), and another with minLes=3, RoI = Core, and the Canberra distance metric (denoted as `stdCoreCa2`).

In addition, there were few other hyperparameters that needed to be set regarding the final clustering procedures used, such as the linkage method and cut-off of the dendrogram for Hierarchical Clustering, as well as the minimum distance threshold for DBSCAN clustering. The choice of these parameters is discussed with the results.

In the description of the results, it is assumed that the Tree-Edit distance matrix for each attempt has already been computed, and only the final clustering stages through Hierarchical and DBSCAN clustering are discussed. For a complete description of the procedure used to obtain the matrix, please refer to Section 2.2 in the Methodological Chapter.

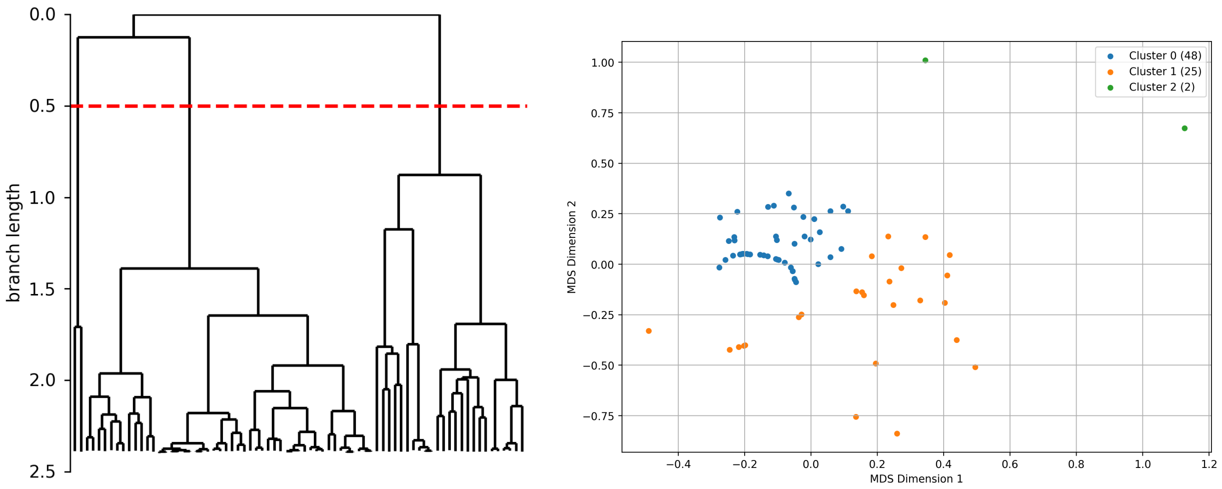
To choose an appropriate linkage method for Hierarchical clustering, a visual inspection of common linkage criteria (e.g., single, average, complete, weighted, centroid, median and Ward's methods) was performed, and the one that best divided the patients was selected. In all the described attempts, Ward's method was chosen, as it always provided the most distinct patient stratification. Furthermore, the cut-off of the dendrogram was also determined by visual inspection; in case where multiple promising values were identified, each one was further explored.

In Section 3.2.1, the main results of the intra-patient heterogeneity analysis for the pre-chemotherapy cohort are presented. This cohort comprises 75 patients who underwent neoadjuvant chemotherapy, each with at least two radiomic lesions on their CT scans. In Sections 3.2.2 and 3.2.3, the analysis is applied to a pre-surgery cohort of 116 patients, each with at least three radiomic lesions available on preoperative CT scans.

3.2.1. Pre-Chemotherapy (`stdCoreEu2`)

As described in Section 2.2, once the tree-edit distance matrix was computed, two clustering algorithms were applied to stratify patients based on this matrix: Hierarchical and DBSCAN clustering.

Concerning Hierarchical clustering, Ward's method was adopted and the dendrogram was cut at `cut=0.5`. To gain further insight into the distribution of patients in the feature space, they were also represented in a two-dimensional space using Multidimensional Scaling (MDS), with the tree-edit distance matrix as input. In this representation, each point was colored according to its cluster membership. Figure 3.8 shows both the dendrogram with the cut-off line and the corresponding MDS plot.



(a) (b)

Figure 3.8: (a) Dendrogram of Hierarchical Clustering for pre-Chemo (`stdCoreEu2`) using Ward linkage, each leaf representing a patient. (b) 2-dimensional visualization of patients through MDS: each point correspond to a patient and it is colored according to its cluster membership

A closer look at Figure 3.8 reveals the presence of three distinct clusters when cutting the dendrogram at `cut=0.5`. The largest cluster (Cluster 0) contains 48 patients, while

Cluster 1 comprises 25 patients, and Cluster 2 includes only 2 patients. Notably, in the MDS representation, the two patients belonging to Cluster 2 appear far from the rest of the cohort, suggesting a substantially different profile. Furthermore, even if Cluster 1 is smaller than Cluster 0, it appears more spread out in the MDS plot, indicating greater internal variability. In contrast, Cluster 0 is relatively more compact, implying a higher degree of similarity among patients. In general, these results underscore the ability of Hierarchical Clustering (with Ward linkage) to capture meaningful patterns of metastatic heterogeneity prior to chemotherapy.

Concerning the DBSCAN algorithm, an epsilon value of 0.2 was set, and the MDS visualization of patients, colored according to the clusters identified by DBSCAN, is provided in Figure 3.9. Notably, DBSCAN aggregated clusters 1 and 2 into a single outlier cluster. Since the stratifications obtained with the two clustering techniques are very similar, only the results from Hierarchical Clustering will be discussed in detail in this study.

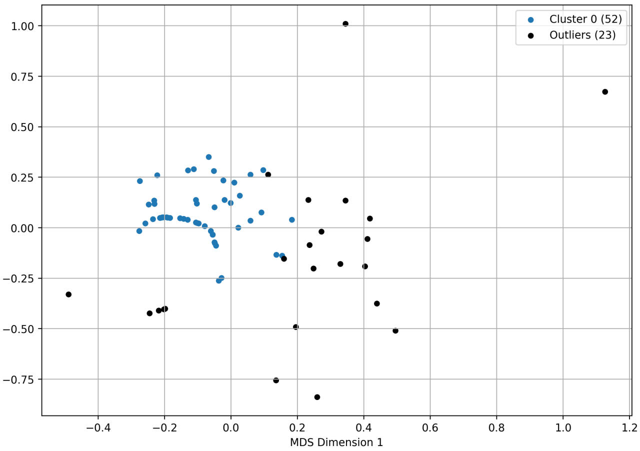


Figure 3.9: MDS visualization of DBSCAN on pre-Chemo (`stdCoreEu2`)

In order to characterize the intra-patient heterogeneity of the metastases within each cluster, the distribution of dendrogram heights for each cluster is provided in Figure A.1 in appendix. Operationally, these curves were constructed by fixing a height value on the x-axis and, for each dendrogram in the selected cluster, counting the number of nodes whose height exceeds that threshold. The y-axis in the plot thus represents the pointwise average (within each cluster) of these counts. By construction, this counting process yields a non-increasing function, offering insights into the heterogeneity of the trees: higher values of the function at larger height threshold indicate the presence of more heterogeneous lesions in the corresponding group.

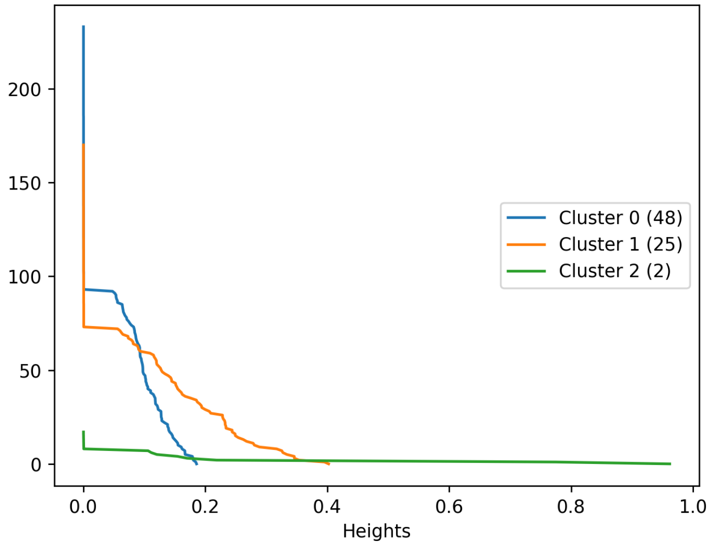


Figure 3.10: Distribution of dendrograms heights for each hierarchical cluster on pre-Chemo (`stdCoreEu2`)

From the figure it can be observed that patients in Cluster 0 exhibit very short branch lengths, suggesting relatively homogeneous metastases. By contrast, patients in Cluster 1 show substantially longer branches, indicating higher heterogeneity. As for Cluster 2, due to its extremely small size (only two patients), it is not possible to draw reliable conclusions about their metastases' heterogeneity.

To further characterize the clusters, all available variables (preoperative and postoperative clinical data, radiomic features, and genetic profiles) were compared across clusters. Unfortunately, this comparison did not further differentiate this stratification, as they appear to have a similar distribution of these variables.

To assess the impact of intra-patient heterogeneity on survival outcomes, Kaplan–Meier curves were generated for each cluster and compared. The resulting survival curves are shown in Figure 3.11.

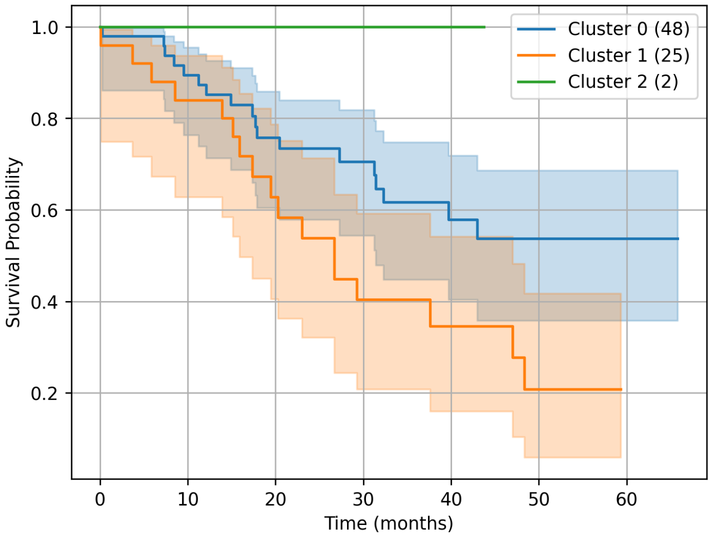


Figure 3.11: Kaplan-Meier curve for each cluster identified by hierarchical clustering on pre-Chemo (`stdCoreEu2`)

Kaplan–Meier analysis reveals a clear difference in survival outcomes between Cluster 0 and Cluster 1. The former (associated with more homogeneous metastases) maintains a higher survival rate, whereas the latter (associated with more heterogeneous metastases) shows a poorer survival profile. In particular, the gap between these two curves becomes more pronounced over time. A log-rank test comparing Clusters 0 and 1 yields a p-value of 0.03, indicating a statistically significant difference in survival. Cluster 2 (with only two patients) remains in 100% survival throughout the observation period, although its very small sample size limits any conclusion. In general, these findings suggest that having more heterogeneous metastases prior to neoadjuvant chemotherapy is associated with a worse prognosis, strengthening the clinical relevance of stratifying patients based on metastasis heterogeneity.

As a final step, the stratification of intra-patient heterogeneity was incorporated into the previous Random Survival Forest models to assess whether it improves the survival predictions. Specifically, for each model described in Section 3.1.1, a new variable containing the cluster labels was added. Patients who were not included in the heterogeneity analysis were assigned a distinct label to maintain the same study cohort across the models. The usual computation of C-Index was performed, and the resulting 95% Confidence Intervals are reported in Figure 3.12, while in Table 3.7 the 95% Confidence Intervals are explicitely reported.

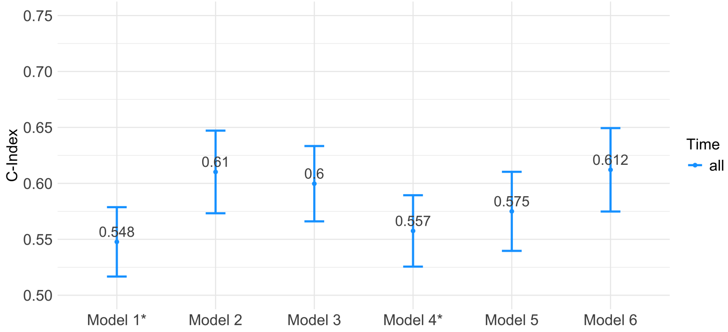


Figure 3.12: 95% Confidence Intervals for pre-Chemotherapy Models (`stdCoreEu2`) including a heterogeneity-based stratification variables

	Lower Bound	Mean	Upper Bound
model 1	0.517	0.548	0.579
model 2	0.573	0.610	0.647
model 3	0.566	0.600	0.633
model 4	0.526	0.557	0.589
model 5	0.540	0.575	0.610
model 6	0.575	0.612	0.649

Table 3.7: 95% Confidence Intervals of C-Index for pre-Chemotherapy models (`stdCoreEu2`)

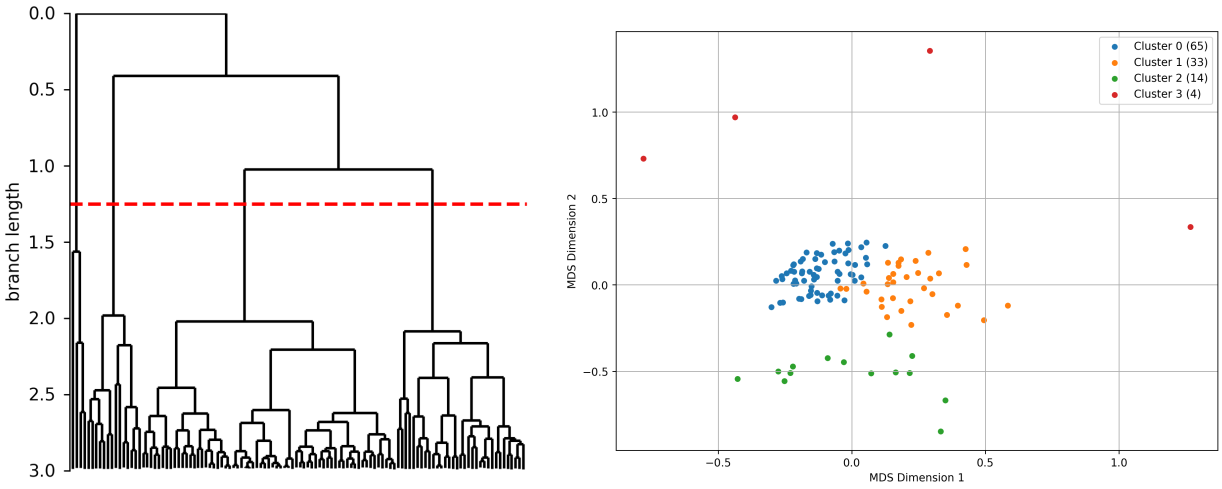
When comparing the C-Index values for models with and without the clustering variable, no improvement in predictive performance is observed. Thus, although the Kaplan–Meier analysis suggests that heterogeneity influences patient survival, this information does not appear to significantly impact the Random Survival Forest model's predictions. This could be due to the possibility that the heterogeneity information is already indirectly captured by the radiomic features, or that it is less influential compared to the other variables in the models.

In addition, the same models (with and without the clustering variable) were re-run using Cox regression to assess whether, in a simpler framework, the additional heterogeneity information might enhance predictive performance—information that might not be fully captured by the other variables. The variable selection process for these Cox models followed a similar two-step approach as for the Random Survival Forest models (described in Figure 2.1). First, a variable selection for each type of variable (i.e., for clinical, core radiomics, ring radiomics and genetic separately) was performed using Lasso, and then a stepwise selection was applied on the aggregated datasets. However, even with the inclusion of the clustering variable, the Cox models did not show any improvement in predictive power; the C-index values remained essentially unchanged (or even slightly worsened, possibly due to an overabundance of categorical variables). For brevity, these Cox model results will not be discussed further.

3.2.2. Pre-Surgery (`stdCoreEu3`)

In this section, results of pre-Surgery analysis are reported with the following set of hyperparameters: RoI=Core, distance=Euclidean, minLes=3. As already mentioned, both Hierarchical and DBSCAN clustering were applied to the tree-edit distance matrix.

Concerning Hierarchical clustering, Ward's linkage was adopted and the dendrogram was cut at `cut=1.25`, dividing the cohort into four clusters. To visualize the distribution of patients in the feature space, a two-dimensional representation was obtained via Multidimensional Scaling (MDS) using the tree-edit distance as input. In this plot, each patient is colored according to its cluster membership. In Figure 3.13 both the pruned dendrogram and the corresponding MDS plot are illustrated.



(a)

(b)

Figure 3.13: (a) Dendrogram of Hierarchical Clustering for pre-Surgery (`stdCoreEu3`) using Ward linkage, each leaf representing a patient. (b) 2-dimensional visualization of patients through MDS: each point correspond to a patient and it is colored according to its cluster membership

Looking at the MDS visualization, Clusters 0 (65 patients) and 1 (33 patients) lie relatively close together, forming the main central cloud. Instead, Cluster 2 (24 patients) is located just below this group yet remains fairly compact, and Cluster 3 consists of only 4 patients positioned much farther from the others.

Concerning DBSCAN clustering, an epsilon value of 0.15 was chosen, and the corresponding MDS visualization of patients, colored according to their DBSCAN stratification, is provided in Figure 3.14. From this visualization, it can be seen that DBSCAN essentially separated one group (Cluster 0) from all the others. However, survival analysis revealed that the two resulting groups did not exhibit significant differences in survival outcomes. Due to these inconclusive survival differences and the challenging interpretation of the heterogeneous point patterns produced by DBSCAN, its results are omitted from further discussion in this study.

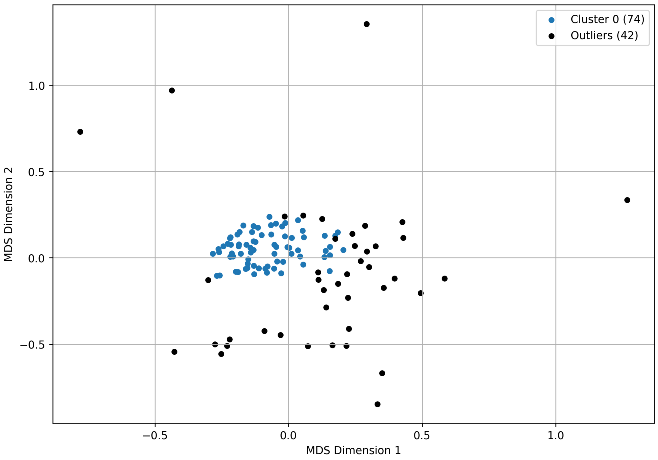


Figure 3.14: MDS visualization of DBSCAN on pre-Surgery (`stdCoreEu3`)

In order to characterize the clusters in terms of intra-patient heterogeneity of metastases, the distribution of dendrogram heights for each cluster was computed, as shown in Figure 3.15. This plot provides insight into the heterogeneity captured by the dendrogram representation: clusters exhibiting higher values at larger heights are indicative of greater heterogeneity.

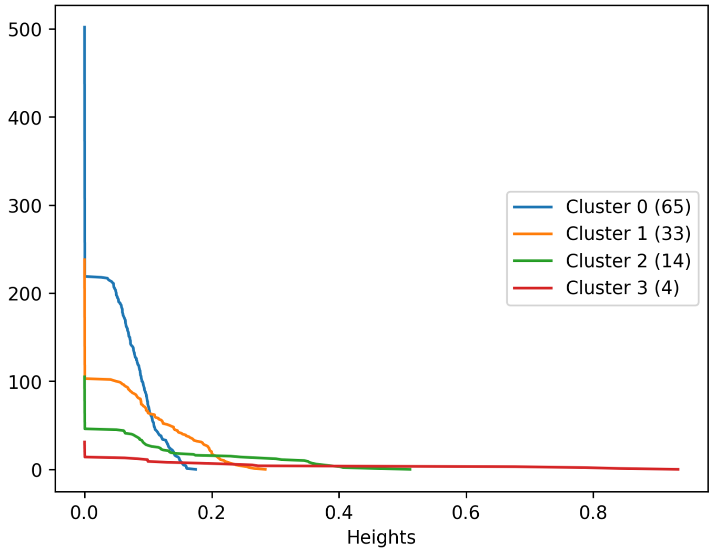


Figure 3.15: Distribution of dendrograms heights for each hierarchical cluster on pre-Surgery (`stdCoreEu3`)

Analyzing the height distributions, it is possible to notice that Cluster 0 is characterized by very short branch lengths, indicating relatively homogeneous metastases. In contrast, Clusters 1, 2, and 3 exhibit longer branches, which suggest higher heterogeneity. Regarding Cluster 3, because of its extremely small size, it is not possible to draw reliable conclusions about the heterogeneity of its metastases.

To further characterize the clusters, all available variables (preoperative and postoperative clinical data, radiomic features, and genetic profiles) were compared across clusters. Unfortunately, this comparison did not further differentiate Clusters 0, 1, and 2, as they appear to have a similar distribution of these variables. The only distinction among these clusters was observed in the intra-patient heterogeneity of metastases, as captured by the dendrogram height distribution.

On the contrary, this comparison allowed for a more in-depth characterization of Cluster 3, that contains patients who appeared very distant from the others in the multidimensional scaling plot. A comparative analysis of clinical variables revealed that all these patients exhibited an advanced T stage and synchronous metastasis timing, suggesting a particularly aggressive and advanced disease. This finding was further supported by the Core radiomic variables: indeed, these patients displayed values that were markedly outside the range of the overall population, at least for one metastasis. In particular, the most divergent variables were `MinimumGreyLevel`, `Kurtosis`, `GLCM_SumVariance`, and `NGTDM_Strength`. Table 3.8 reports, for each of the four patients, the values of the most anomalous metastasis for these variables along with the mean values (and corresponding standard deviations) calculated for the analyzed cohort.

Patient	Minimum Grey Level	Kurtosis	GLCM SumVariance	NGTDM Strength
#50	-551	51.15	6.29	7.38
#56	-369	8.33	0.00	20.89
#244	-643	100.03	3.30	4.84
#256	-819	56.58	9.79	2.04
Mean (std)	-8.52 (75.17)	0.93 (5.92)	0.04 (0.54)	0.84 (1.78)

Table 3.8: Summary statistics of some core radiomics variables for patients in Cluster 3. Only the most anomalous lesion is reported for each patient. Last line refers to the mean (std) of the corresponding variable computed on the entire cohort of patients

In particular, all these patients share the following characteristics: an extremely low `INTENSITY_BASED_MinimumGreyLevel`, along with very high values for `INTENSITY-BASED_Kurtosis`, `GLCM_SumVariance`, and `NGTDM_Strength`. Specifically:

`INTENSITY_BASED_MinimumGreyLevel`: Low values indicate regions with extremely reduced tracer uptake, which may reflect necrosis or less tumor-active tissue.

`INTENSITY-BASED_Kurtosis`: High values indicate that some pixel intensities are very far from the mean, suggesting high internal heterogeneity within the lesion.

`GLCM_SumVariance`: High values denote significant variability in intensity values, again pointing to elevated internal heterogeneity.

`NGTDM_Strength`: High values reflect strong contrast between pixels and a well-defined internal structure, further indicating increased lesion heterogeneity.

Therefore, these patients could be those with tumors that exhibit necrosis or reduced metabolic activity in certain regions combined with pronounced internal heterogeneity, which may be indicative of a more aggressive tumor phenotype.

To assess the impact of intra-patient heterogeneity on survival outcomes, Kaplan–Meier curves were generated for each cluster and compared. The resulting survival curves are presented in Figure 3.16.

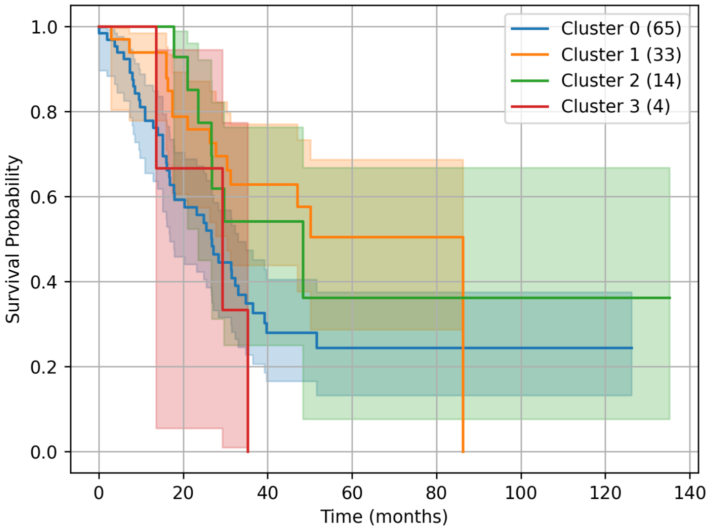


Figure 3.16: Kaplan-Meier curve for each cluster identified by hierarchical clustering on pre-Surgery (`stdCoreEu3`)

The Kaplan–Meier plot shows that patients in Cluster 0 (those with the most homogeneous metastases) have the lowest survival rates, whereas patients in Clusters 1 and 2 (with more heterogeneous metastases) exhibit higher survival. The most distinct separation is observed between Clusters 0 and 1, the largest groups, with Cluster 2 falling in between. Cluster 3, although comprising only four patients and thus should be interpreted with caution, has the worst prognosis as expected from the advanced stage and highly aggressive nature of their metastases. Statistical comparisons using the log-rank test revealed a very low p-value ($p = 0.012$) between Clusters 0 and 1, and a moderately low p-value ($p = 0.125$) between Clusters 0 and 2, while other comparisons produced high p-values. Overall, the data suggest that having homogeneous metastases prior to surgery is associated with a poorer prognosis.

Finally, this stratification was incorporated into the Random Survival Forest models to assess whether it improves the survival predictions. Specifically, a new variable representing the cluster labels was added to each model presented in Section 3.1.2. Patients who were not included in the heterogeneity analysis were assigned a distinct label to maintain the same study cohort across the model for a better comparison. The performances were assessed computing the 95% Confidence Intervals of the C-Index, using 10-Fold Cross Validation repeated 10 times. The results are reported in Figure 3.17 and in Table 3.9

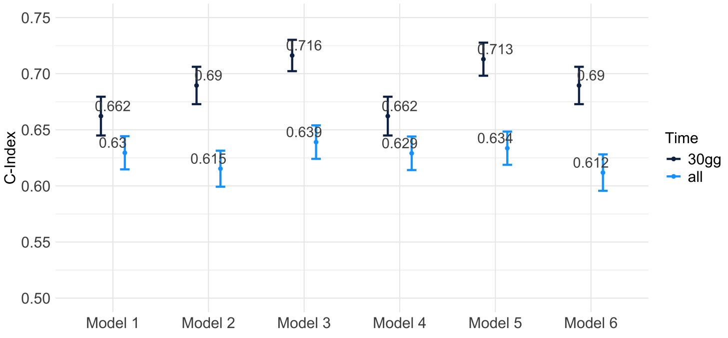


Figure 3.17: 95% Confidence Intervals for pre-Surgery Models including a heterogeneity-based stratification variables for `stdCoreEu3`

	Lower Bound	Mean	Upper Bound
model 1	0.615	0.630	0.644
model 1_30gg	0.599	0.615	0.631
model 2	0.599	0.615	0.631
model 2_30gg	0.672	0.690	0.706
model 3	0.624	0.639	0.654
model 3_30gg	0.702	0.716	0.730
model 4	0.614	0.629	0.644
model 4_30gg	0.645	0.662	0.679
model 5	0.619	0.634	0.648
model 5_30gg	0.698	0.712	0.728
model 6	0.596	0.612	0.628
model 6_30gg	0.673	0.689	0.706

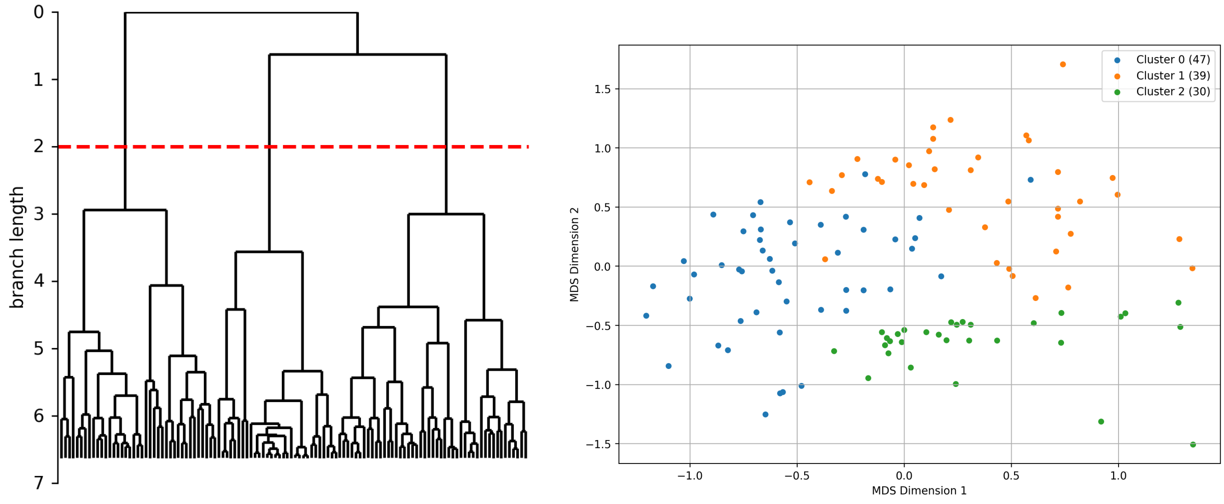
Table 3.9: 95% Confidence Intervals of C-Index for pre-Surgery models for `stdCoreEu3`

Comparing the performance of the models with and without the heterogeneity variable, no noticeable improvement was observed with the addition of this new feature, despite the Kaplan–Meier analyses showing significant survival differences among the clusters. In particular, similar observations as those made for the preChemo models in Section 3.2.1 can be made: either the Random Survival Forest is already capturing the heterogeneity information through the existing radiomic variables, or the stratification based on heterogeneity is not as influential as the other variables.

3.2.3. Pre-Surgery (`stdCoreCa3`)

In this section, results of pre-Surgery analysis are reported with the following set of hyperparameters: RoI=Core, distance=Canberra, minLes=3. As stated before, both Hierarchical and DBSCAN clustering were applied to the tree-edit distance matrix computed.

Regarding Hierarchical Clustering, Ward's linkage was selected, and the resulting dendrogram was cut at 2, separating the cohort into three distinct clusters. Multidimensional Scaling was then applied using the computed tree-edit distances to generate a two-dimensional representation of the patients. Figure 3.18 displays the dendrogram with the cut alongside the corresponding MDS plot.



(a) (b)

Figure 3.18: (a) Dendrogram of Hierarchical Clustering for pre-Surgery (`stdCoreCa3`) using Ward linkage, each leaf representing a patient. (b) 2-dimensional visualization of patients through MDS: each point correspond to a patient and it is colored according to its cluster membership

From the MDS representation, it is possible to observe that the distribution of patients appears more uniform in this case compared to the other analyzed scenarios. However, this distribution should be interpreted with caution, as it is only a two-dimensional projection of a high-dimensional object, and the representation might be misleading.

Concerning DBSCAN clustering, an epsilon value of 0.6 was set, and the resulting MDS visualization of patient colored accordingly is provided in Figure 3.19.

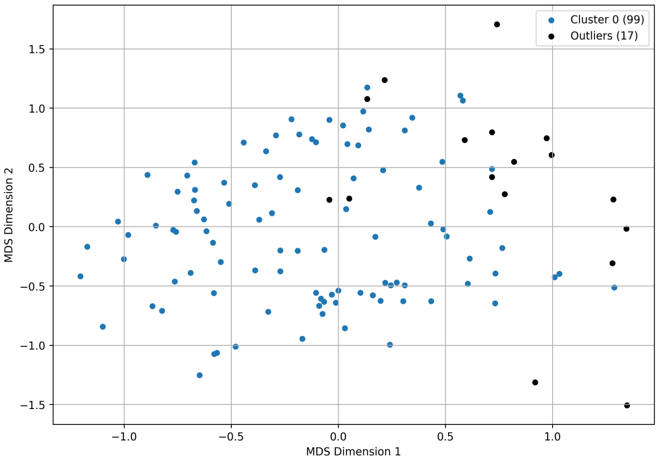


Figure 3.19: MDS visualization of DBSCAN on pre-Surgery (`stdCoreCa3`)

In this case, the MDS representation of patients stratified with DBSCAN shows that DBSCAN fails to capture any interesting patterns. Furthermore, multiple epsilon values were tested, but none resulted in compelling stratifications. For this reason, only the stratification provided by hierarchical clustering will be discussed in this section.

To characterize the clusters in terms of intra-patient heterogeneity of metastases, the distribution of dendrogram heights was computed for each cluster, as illustrated in Figure 3.20. This plot provides insight into the heterogeneity captured by the dendrogram representation: clusters with higher values at larger heights are indicative of greater heterogeneity.

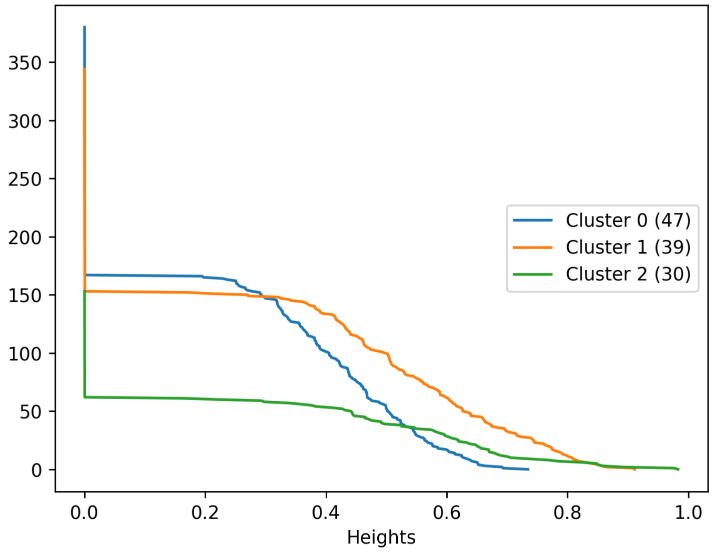


Figure 3.20: Distribution of dendrograms heights for each hierarchical cluster on pre-Surgery (`stdCoreCa3`)

From the plot, it is evident that Cluster 0 represents patients with more homogeneous metastases, while Clusters 1 and 2 include patients with a higher degree of heterogeneity.

In this analysis, all available variables were compared across clusters to achieve a comprehensive characterization of the resulting stratification. Notably, patients in Cluster 2 have a lower metastasis count (mean = 4.53, std = 3.07) compared to those in Cluster 0 (mean = 9.96, std = 6.27) and Cluster 1 (mean = 10.54, std = 5.62). Figure 3.21 illustrates the distribution of metastasis counts for each cluster. This observation is supported by the non-parametric Kruskal–Wallis Test [25], which yielded a very low p-value ($p = 2.0 \times 10^{-08}$), indicating that at least one group exhibits a significantly different number of metastases. Subsequent pairwise comparisons using the Mann-Whitney U Test [26] with Bonferroni correction [1] confirmed that Cluster 2 is statistically different from both Cluster 0 ($p = 2.0 \times 10^{-06}$) and Cluster 1 ($p = 7.3 \times 10^{-08}$), while no significant difference was found between Cluster 0 and Cluster 1 ($p = 1.0$).

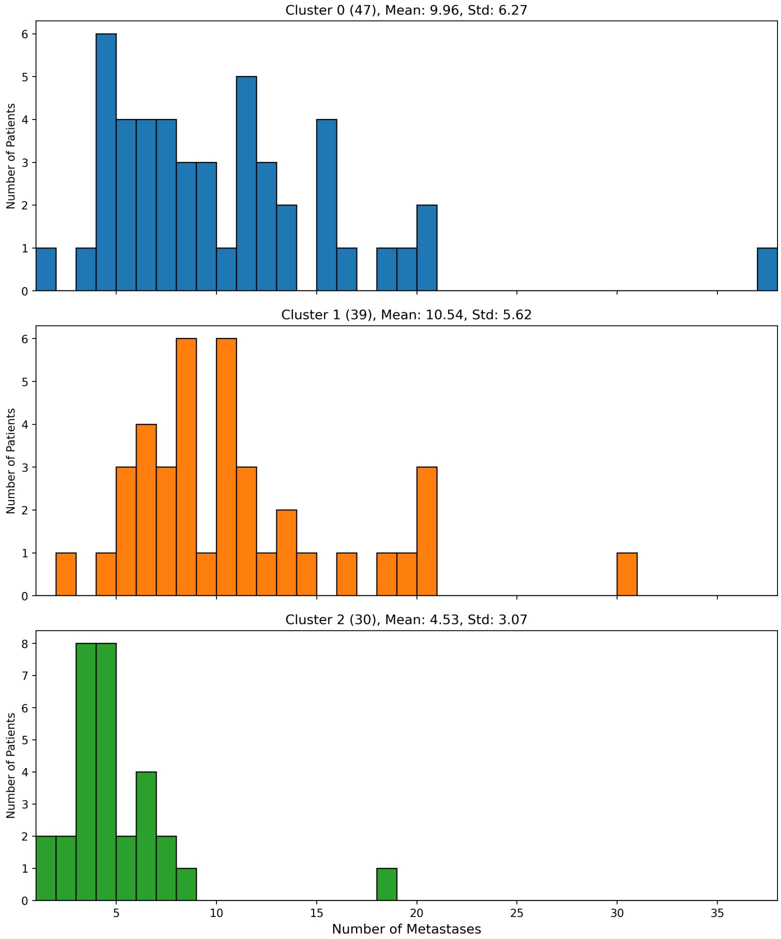


Figure 3.21: Distribution of the number of metastases for each hierarchical cluster on pre-Surgery (`stdCoreCa3`)

To assess the impact of this stratification on survival prediction, Kaplan-Meier curves were plotted for each clustering, as shown in Figure 3.22.

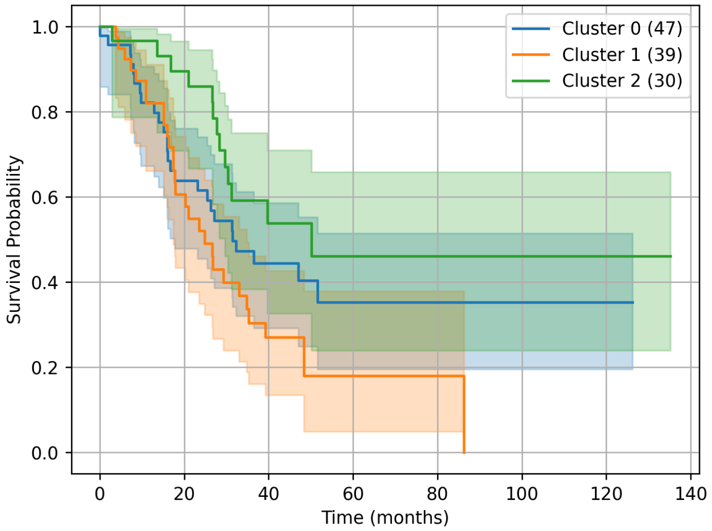


Figure 3.22: Kaplan-Meier curve for each cluster identified by hierarchical clustering on pre-Surgery (`stdCoreCa3`)

This plot shows the survival clusters defined by the heterogeneity and metastasis count. Specifically:

Cluster 0: includes patients with homogeneous metastases and a high metastasis count;

Cluster 1: includes patients with heterogeneous metastases and a high metastasis count;

Cluster 2: includes patients with heterogeneous metastases and a low metastasis count;

Overall, the log-rank test indicates a statistically significant difference among the three survival curves (p=0.031). To better understand where this difference arises, pairwise comparisons were performed. In particular, p-value assesses significant difference between Cluster 1 and Cluster 2 (p=0.006), while higher p-values were obtained when comparing Cluster 0 with Cluster 1 (p=0.186) and with Cluster 2 (p=0.179).

Thus, the main difference is observed between Cluster 1 and Cluster 2. Patients in Cluster 2, who have a lower number of metastatic lesions (even though they are heterogeneous) exhibit a more favorable survival profile compared to those in Cluster 1, who have many heterogeneous metastases. In other words, having few metastases is the most favorable condition overall. Moreover, among patients with a high metastatic burden, those with homogeneous lesions (Cluster 0) fare better than those with heterogeneous lesions (Cluster 1). In this context, the condition of having numerous homogeneous metastases represents an intermediate situation between the two extremes. Overall, these findings emphasize the combined impact of lesion count and heterogeneity on patient survival.

Finally, this stratification included in the Random Survival Forest models to assess its capability of improvement in terms of survival prediction. In particular, a new variable representing this stratification was introduced, where the patients who were not included in the heterogeneity analysis were assigned a new label. The performances were assessed using C-Index, and the results are presented in Figure 3.23 and Table 3.10

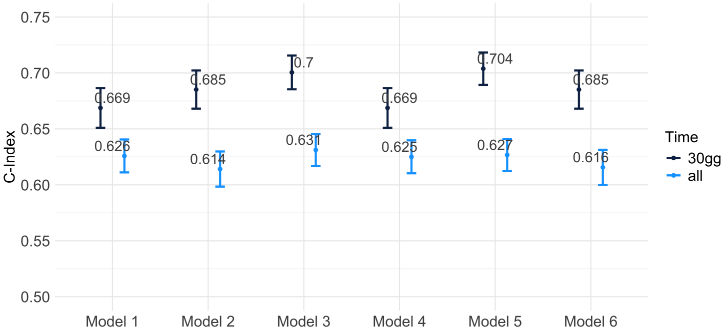


Figure 3.23: 95% Confidence Intervals for pre-Surgery Models including a heterogeneity-based stratification variables for stdCoreCa3

	Lower Bound	Mean	Upper Bound
model 1	0.611	0.626	0.641
model 1_30gg	0.651	0.669	0.687
model 2	0.598	0.614	0.630
model 2_30gg	0.668	0.685	0.702
model 3	0.617	0.631	0.646
model 3_30gg	0.685	0.700	0.716
model 4	0.610	0.625	0.640
model 4_30gg	0.651	0.669	0.687
model 5	0.613	0.627	0.641
model 5_30gg	0.689	0.704	0.718
model 6	0.600	0.616	0.631
model 6_30gg	0.668	0.685	0.702

Table 3.10: 95% Confidence Intervals of C-Index for pre-Surgery models for `stdCoreCa3`

Also in this case, no noticeable improvement was observed when adding the clustering feature compared to a model without the heterogeneity variable, even though the Kaplan–Meier analysis suggested significant differences in survival outcomes among clusters. It is important to note, however, that the most pronounced differences in the KM curves were primarily driven by the number of metastases, a variable already present in the dataset. This finding suggests that the stratification may not be adding substantial new information beyond what is already captured by the metastatic count. Hence, similar observations as those reported in Section 3.2.1 can be made.

4 | Conclusion

This work explored the integration of clinical, genetic, and radiomic data to enhance the prediction of survival outcomes in patients with colorectal liver metastases. The study demonstrated that incorporating radiomic features—especially those extracted from the tumor core—into survival models improves predictive accuracy compared to using clinical and genetic data alone. Models based on Random Survival Forests were developed by gradually integrating additional data types. The observed improvements in the C-index indicate that radiomic features provide useful complementary information (on the other hand, it is noteworthy that the genetic variables did not contribute significantly in any model). For example, in pre-chemotherapy models, adding core radiomic variables to models based solely on clinical variables increased the average C-index from 0.558 (model 1) to 0.624 (model 2). Similarly, in pre-surgery models using CT scans taken within 30 days of the intervention (i.e., CTs that are much more representative of the current tumor state), adding both core and ring radiomic variables to the clinical variables increased the average C-index from 0.669 (model 1_30gg) to 0.711 (model 3_30gg). In contrast, for generic pre-intervention models, the addition of radiomic features did not yield a significant improvement (with model 1 showing 0.629 and model 3 0.634). Furthermore, a comparison between pre-intervention and pre-chemotherapy models reveals that pre-intervention models are more precise in estimating the outcome, with a significantly better performance index (pre-intervention 0.711 versus pre-chemotherapy 0.624). With regard to the score-based models, their performance was found to be comparable to that of models using only clinical information, suggesting that they provide a good summary of the preoperative clinical variables; however, when compared with models that include radiomic features, the score-based models perform considerably worse.

Regarding the heterogeneity analysis, this study proposed a method to stratify patients according to the heterogeneity of their lesions. In particular, patients were stratified according to a tree-clustering method applied to dendrogram representations of metastatic lesions. This approach faces inherent challenges due to the limited number of radiomic profiles available per patient (with at most five radiomic lesions per patient) and the intrinsic homogeneity of the metastases, given that all lesions originate from the liver.

Despite these limitations, the stratification yielded interesting findings. For patients prior to neoadjuvant chemotherapy, those with highly heterogeneous metastases were associated with a poorer prognosis. Conversely, in the pre-surgery analysis, the model using the `stdCoreEu3` configuration indicated that overall, patients with homogeneous metastases had worse outcomes. However, when considering the combined effect of heterogeneity and metastasis count in the `stdCoreCa3` model, the number of metastases emerged as the primary discriminant—with fewer metastases being indicative of better survival—and heterogeneity further discriminated among patients with a high metastatic burden, with many heterogeneous metastases performing worse than many homogeneous ones. Notably, the stratification based on heterogeneity did not translate into a significant improvement in the predictive performance of the Random Survival Forest models, suggesting that the heterogeneity information may already be indirectly captured by the radiomic features or that it is less influential, as also confirmed by alternative Cox regression models.

Limitations

The study faces several limitations. The relatively small sample size, both in terms of the number of patients and the available radiomics of metastases per patient, may affect the robustness of the conclusions. In addition, because all metastases analyzed originated from the liver, the inherent similarity among them may have obscured more subtle variations in heterogeneity analysis. Such uniformity might limit the capacity of the methods to detect meaningful differences that could influence patient outcomes. Another limitation is the high variability in the time interval between the CT scan and the intervention, which compromises the representativeness of the radiomic features for the intervention. In this study, a 30-day cut-off (i.e., analyzing separately patients with a pre-surgery CT scan taken no more than 30 days before the surgical intervention) was applied based on clinical suggestions; however, this criterion significantly reduced the number of patients included in the analysis. Furthermore, the genetic variables available were only two, although the most commonly used in the literature, and there was a significant imbalance between the number of patients with and without mutations, which may have limited their contribution to the models.

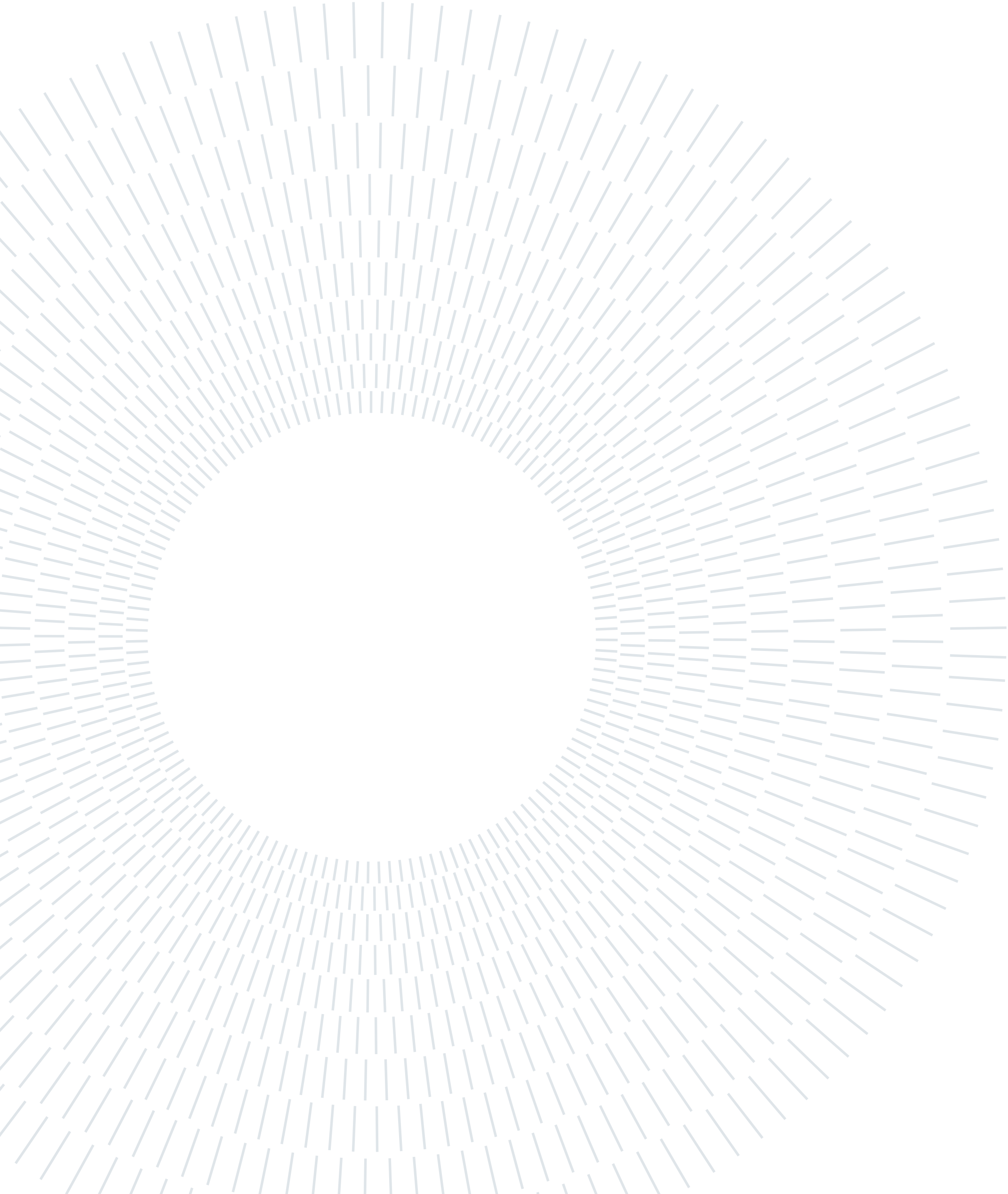
Future Directions

Several future research directions are suggested by the findings. One promising avenue is the implementation of longitudinal analyses to monitor changes in metastatic characteristics over time. Moreover, tracking the evolution of tumor heterogeneity both before and after chemotherapy could provide a dynamic view of treatment response and disease progression.

Another potential direction involves the use of advanced survival models based, for instance, on neural networks. Deep learning approaches may better capture complex, nonlinear interactions among variables, which could further improve the predictive performance of survival models. In addition, pre-treatment heterogeneity measures may serve as valuable inputs in classifiers designed to predict chemotherapy outcomes, thereby contributing to more personalized treatment planning. However, these advanced methods would require a larger dataset for their application and to draw generalizable conclusions.

Finally, future studies could investigate the integration of radiomic features from all lesions of each patient into a comprehensive survival model, rather than relying solely on the radiomic profile of the largest metastasis. This approach could capture additional aspects of tumor heterogeneity and may lead to further improvements in survival prediction.

Finally, future studies should consider re-evaluating pre-surgery models by accounting for the time interval between the CT scan and the intervention. Instead of excluding patients based on a rigid 30-day cut-off, alternative approaches could be developed to assign greater weight to patients with more recent CT scans, or to determine the maximum allowable delta time that maintains the CT's representativeness without unduly reducing the patient cohort.



Bibliography

[1] R. A. Armstrong. When to use the b onferroni correction. *Ophthalmic and physiological optics*, 34(5):502–508, 2014.

[2] M. Arnold, M. S. Sierra, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*, 66(4): 683–691, 2017.

[3] A. R. Brannon, E. Vakiani, B. E. Sylvester, S. N. Scott, G. McDermott, R. H. Shah, K. Kania, A. Viale, D. M. Oschwald, V. Vacic, et al. Comparative sequencing analysis reveals high genomic concordance between matched primary and metastatic colorectal cancer lesions. *Genome biology*, 15:1–10, 2014.

[4] K. W. Brudvik, R. P. Jones, F. Giuliante, J. Shindoh, G. Passot, M. H. Chung, J. Song, L. Li, V. J. Dagenborg, Å. A. Fretland, et al. Ras mutation clinical risk score to predict survival after resection of colorectal liver metastases. *Annals of surgery*, 269(1):120–126, 2019.

[5] L. Cavinato, M. Pegoraro, A. Ragni, M. Sollini, P. A. Erba, and F. Ieva. Imaging-based representation and stratification of intra-tumor heterogeneity via tree-edit distance. *Scientific reports*, 12(1):19607, 2022.

[6] D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

[7] J. Z. Done, A. Papanikolaou, M. Stem, S. N. Radomski, S. Y. Chen, J. R. Maturi, C. Atallah, and B. Safar. Impact of neoadjuvant chemotherapy on perioperative morbidity in combined resection of rectal cancer and liver metastases. *Journal of Surgical Oncology*, 2025.

[8] B. K. Edwards, E. Ward, B. A. Kohler, C. Eheman, A. G. Zauber, R. N. Anderson, A. Jemal, M. J. Schymura, I. Lansdorp-Vogelaar, L. C. Seeff, et al. Annual report to the nation on the status of cancer, 1975-2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates, 2010.

[9] N. et al. *LIFEx User Guide - IBSI-1 Compliant*, 2021. URL `https://www.lifexsoft.org/images/phocagallery/documentation/ProtocolTexture/UserGuide/LIFEx_IBSI1.pdf`.

[10] Y. Fong, J. Fortner, R. L. Sun, M. F. Brennan, and L. H. Blumgart. Clinical score for predicting recurrence after hepatic resection for metastatic colorectal cancer: analysis of 1001 consecutive cases. *Annals of surgery*, 230(3):309, 1999.

[11] R. J. Gillies, P. E. Kinahan, and H. Hricak. Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2016.

[12] F. E. Harrell, R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati. Evaluating the yield of medical tests. *Jama*, 247(18):2543–2546, 1982.

[13] S. R. Horn, K. C. Stoltzfus, E. J. Lehrer, L. A. Dawson, L. Tchelebi, N. J. Gusani, N. K. Sharma, H. Chen, D. M. Trifiletti, and N. G. Zaorsky. Epidemiology of liver metastases. *Cancer epidemiology*, 67:101760, 2020.

[14] International Agency for Research on Cancer (IARC). Global cancer observatory: Cancer today - fact sheets by cancer. `https://gco.iarc.who.int/today/en/fact-sheets-cancers`, 2022. [Accessed 16 January 2025].

[15] H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer. Random survival forests. *The Annals of Applied Statistics*, 2(3):841 – 860, 2008. doi: 10.1214/08-AOAS169. URL `https://doi.org/10.1214/08-AOAS169`.

[16] D. Jarrett, B. Cebere, T. Liu, A. Curth, and M. van der Schaar. Hyperimpute: Generalized iterative imputation with automatic model selection, 2022. URL `https://arxiv.org/abs/2206.07769`.

[17] E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

[18] V. Kumar, Y. Gu, S. Basu, A. Berglund, S. A. Eschrich, M. B. Schabath, K. Forster, H. J. Aerts, A. Dekker, D. Fenstermacher, et al. Radiomics: the process and the challenges. *Magnetic resonance imaging*, 30(9):1234–1248, 2012.

[19] M. B. Kursa and W. R. Rudnicki. Feature selection with the boruta package. *Journal of statistical software*, 36:1–13, 2010.

[20] H. Li, Y. Zhuang, W. Yuan, Y. Gu, X. Dai, M. Li, H. Chen, and H. Zhou. Radiomics in precision medicine for colorectal cancer: a bibliometric analysis (2013–2023). *Frontiers in Oncology*, 14:1464104, 2024.

[21] S.-Y. Lu. A tree-to-tree distance and its application to cluster analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):219–224, 1979. doi: 10.1109/TPAMI.1979.6786615.

[22] L. Lv, B. Xin, Y. Hao, Z. Yang, J. Xu, L. Wang, X. Wang, S. Song, and X. Guo. Radiomic analysis for predicting prognosis of colorectal cancer from preoperative 18f-fdg pet/ct. *Journal of translational medicine*, 20(1):66, 2022.

[23] G. A. Margonis, S. Buettner, N. Andreatos, Y. Kim, D. Wagner, K. Sasaki, A. Beer, C. Schwarz, I. M. Løes, M. Smolle, et al. Association of braf mutations with survival and recurrence in surgically treated patients with metastatic colorectal liver cancer. *JAMA surgery*, 153(7):e180996–e180996, 2018.

[24] G. A. Margonis, K. Sasaki, S. Gholami, Y. Kim, N. Andreatos, N. Rezaee, A. Deshwar, S. Buettner, P. J. Allen, T. P. Kingham, et al. Genetic and morphological evaluation (game) score for patients with colorectal liver metastases. *Journal of British Surgery*, 105(9):1210–1220, 2018.

[25] P. E. McKight and J. Najab. Kruskal-wallis test. *The corsini encyclopedia of psychology*, pages 1–1, 2010.

[26] P. E. McKnight and J. Najab. Mann-whitney u test. *The Corsini encyclopedia of psychology*, pages 1–1, 2010.

[27] K. Menck, D. Wlochowitz, A. Wachter, L.-C. Conradi, A. Wolff, A. H. Scheel, U. Korf, S. Wiemann, H.-U. Schildhaus, H. Bohnenberger, et al. High-throughput profiling of colorectal cancer liver metastases reveals intra-and inter-patient heterogeneity in the egfr and wnt pathways associated with clinical outcome. *Cancers*, 14(9):2084, 2022.

[28] M. Meng, K. Zhong, T. Jiang, Z. Liu, H. Y. Kwan, and T. Su. The current understanding on the impact of kras on colorectal cancer. *Biomedicine & pharmacotherapy*, 140:111717, 2021.

[29] F. L. Pietsch, F. Haag, I. Ayx, F. Grawe, A. K. Vellala, S. O. Schoenberg, M. F. Froelich, and H. Tharmaseelan. Textural heterogeneity of liver lesions in ct imaging-comparison of colorectal and pancreatic metastases. *Abdominal Radiology*, 49(12): 4295–4306, 2024.

[30] N. Reboux, V. Jooste, J. Goungounga, M. Robaszkiewicz, J.-B. Nousbaum, and A.-M. Bouvier. Incidence and survival in synchronous and metachronous liver metastases from colorectal cancer. *JAMA network Open*, 5(10):e2236666–e2236666, 2022.

[31] S. Rizzo, F. Botta, S. Raimondi, D. Origgi, C. Fanciullo, A. G. Morganti, and M. Bel-

lomi. Radiomics: the facts and the challenges of image analysis. *European radiology experimental*, 2:1–8, 2018.

[32] D. Schlee. Numerical taxonomy. the principles and practice of numerical classification, 1975.

[33] E. Schubert, J. Sander, M. Ester, H. P. Kriegel, and X. Xu. Dbscan revisited, revisited: why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, 42(3):1–21, 2017.

[34] R. L. Siegel, A. N. Giaquinto, and A. Jemal. Cancer statistics, 2024. *CA: a cancer journal for clinicians*, 74(1):12–49, 2024.

[35] F. C. Staal, D. J. Van Der Reijd, M. Taghavi, D. M. Lambregts, R. G. Beets-Tan, and M. Maas. Radiomics for the prediction of treatment outcome and survival in patients with colorectal cancer: a systematic review. *Clinical colorectal cancer*, 20 (1):52–71, 2021.

[36] A. Stanzione, F. Verde, V. Romeo, F. Boccadifuoco, P. P. Mainenti, and S. Maurea. Radiomics and machine learning applications in rectal cancer: current update and future perspectives. *World Journal of Gastroenterology*, 27(32):5306, 2021.

[37] H. Tharmaseelan, A. Hertel, F. Tollens, J. Rink, P. Woźnicki, V. Haselmann, I. Ayx, D. Nörenberg, S. O. Schoenberg, and M. F. Froelich. Identification of ct imaging phenotypes of colorectal liver metastases from radiomics signatures—towards assessment of interlesional tumor heterogeneity. *Cancers*, 14(7):1646, 2022.

[38] W. S. Torgerson. Multidimensional scaling: I. theory and method. *Psychometrika*, 17(4):401–419, 1952.

[39] F. Tosi, E. Magni, A. Amatu, G. Mauri, K. Bencardino, M. Truini, S. Veronese, L. De Carlis, G. Ferrari, M. Nichelatti, et al. Effect of kras and braf mutations on survival of metastatic colorectal cancer after liver resection: a systematic review and meta-analysis. *Clinical colorectal cancer*, 16(3):e153–e163, 2017.

[40] L. A. Wolsey and G. L. Nemhauser. *Integer and combinatorial optimization*. John Wiley & Sons, 2014.

[41] A. Zwanenburg, M. Vallières, M. A. Abdalah, H. J. W. L. Aerts, V. Andrearczyk, A. Apte, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, 295 (2):328–338, May 2020. ISSN 1527-1315. doi: 10.1148/radiol.2020191145. URL `http://dx.doi.org/10.1148/radiol.2020191145`.

A | Appendix A

A.1. List of Radiomic Features

Morphological Features

Volume (RNU0) The total volume of the tumor or region of interest (ROI) as measured in cubic units.

ApproximateVolume (YEKZ) An estimated measurement of the tumor volume, often used when precise measurements are difficult.

SurfaceArea (COJK) The total surface area of the tumor, typically calculated from the segmented tumor shape.

SurfaceToVolumeRatio (2PR5) The ratio between the surface area and the volume of the tumor, providing insight into the tumor's geometry.

Compactness1 (SKGS) A measure of how compact the tumor is, typically calculated as a ratio of volume to surface area.

Compactness2 (BQWJ) Another variant of compactness, possibly accounting for different shape or irregularities.

SphericalDisproportion (KRCK) The measure of how far the tumor shape deviates from a perfect sphere.

Sphericity (QCFX) A geometric measure describing how spherical the tumor is, with 1 being perfectly spherical.

Asphericity (25C7) The degree to which the tumor deviates from spherical symmetry.

CentreOfMassShift (KLMA) The shift in the center of mass of the tumor compared to a reference point, indicating irregularity.

Maximum3DDiameter (L0JK) The largest diameter of the tumor, measured in three-dimensional space.

IntegratedIntensity (99N0) A measure that integrates voxel intensities within the tumor ROI, capturing overall intensity variation.

Intensity-Based Features

Mean (Q4LE) The average intensity value across all voxels in the ROI

Variance (ECT3) A measure of how much the intensity values within the ROI differ from the mean intensity.

Skewness (KE2A) The asymmetry of the intensity distribution within the ROI. Positive skew indicates a tail toward higher intensity values.

Kurtosis (IPH6) A measure of the "tailedness" of the intensity distribution. Higher kurtosis indicates more outliers in the intensity values.

Median (Y12H) The middle value of the intensity distribution, with half the voxels having lower intensities and half having higher intensities.

MinimumGreyLevel (1GSF) The lowest intensity value within the ROI.

10thPercentile (QG58) The intensity value below which 10% of the voxels fall, providing insight into the lower range of intensities.

50thPercentile (Y12H) The intensity value at the median, representing the middle value of the intensity distribution.

90thPercentile (8DWT) The intensity value below which 90% of the voxels fall, reflecting the higher end of the intensity distribution.

MaximumGreyLevel (84IY) The highest intensity value within the ROI.

InterquartileRange (SALO) The range between the 25th and 75th percentiles, capturing the spread of the middle 50% of intensity values.

Range (2OJQ) The difference between the maximum and minimum intensity values within the ROI.

MeanAbsoluteDeviation (4FUA) The average of the absolute differences between each voxel intensity and the mean intensity.

RobustMeanAbsoluteDeviation A robust version of the mean absolute deviation, less sensitive to extreme outliers.

MedianAbsoluteDeviation (N72L) The median of the absolute deviations from the median intensity, providing a measure of variability that is less affected by outliers.

CoefficientOfVariation (7TET) The ratio of the standard deviation to the mean intensity, indicating the relative variability within the ROI.

QuartileCoefficientOfDispersion (9S40) The ratio of the interquartile range to the sum of the 25th and 75th percentiles, providing a measure of dispersion.

Energy The sum of squared intensity values, reflecting the overall uniformity of intensity distribution within the ROI.

RootMeanSquare (5ZWQ) The square root of the average of squared intensity values, representing the magnitude of intensity variation in the ROI.

Intensity Histogram Features

IntensityHistogramMean (X6K6) The average intensity of all voxel intensities within the ROI, representing the central tendency of the intensity distribution.

IntensityHistogramVariance (CH89) A measure of how much the intensity values within the ROI vary from the mean intensity, indicating the spread of the intensity values.

IntensityHistogramSkewness (88K1) A measure of the asymmetry of the intensity histogram. Positive skewness indicates a longer tail toward higher intensity values.

IntensityHistogramKurtosis (C3I7) A measure of the "tailedness" of the intensity distribution. High kurtosis indicates a distribution with heavy tails or outliers.

IntensityHistogramMedian (WIFQ) The middle value of the intensity distribution, with half of the voxel intensities falling below it and half above it.

IntensityHistogramMinimumGreyLevel (1PR8) The lowest intensity value present in the ROI.

IntensityHistogram10thPercentile (GPMT) The intensity value below which 10% of the voxel intensities fall, representing the lower end of the intensity distribution.

IntensityHistogram90thPercentile (OZ0C) The intensity value below which 90% of the voxel intensities fall, representing the upper end of the intensity distribution.

IntensityHistogramMaximumGreyLevel (3NCY) The highest intensity value present in the ROI.

IntensityHistogramMode (AMMC) The intensity value that appears most frequently within the ROI, indicating the most common intensity.

IntensityHistogramInterquartileRange (WR0O) The range between the 25th and 75th percentiles, capturing the spread of the middle 50% of intensity values.

IntensityHistogramRange (5Z3W) The difference between the maximum and minimum intensity values, representing the full range of intensities within the ROI.

IntensityHistogramMeanAbsoluteDeviation (D2ZX) The average of the absolute deviations of voxel intensities from the mean, providing a measure of variability within the ROI.

IntensityHistogramRobustMeanAbsoluteDeviation (WRZB) A robust version of the mean absolute deviation, less sensitive to extreme outliers.

IntensityHistogramMedianAbsoluteDeviation (4RNL) The median of the absolute deviations from the median intensity, a measure of dispersion that is more robust than the standard deviation.

IntensityHistogramCoefficientOfVariation (CWYJ) The ratio of the standard deviation to the mean intensity, providing a normalized measure of intensity variability.

IntensityHistogramQuartileCoefficientOfDispersion (SLWD) A measure of dispersion based on the interquartile range, indicating the spread of the central 50% of intensities.

IntensityHistogramEntropyLog (TLU2) A measure of the randomness or disorder in the intensity distribution, with higher values indicating more randomness.

Uniformity (BJ5W) A measure of the uniformity of the intensity distribution, where higher values indicate more uniform intensity values across the ROI.

MaximumHistogramGradient (12CE) The maximum rate of change in intensity values between neighboring voxels, indicating regions of rapid intensity change.

MaximumHistogramGradientGreyLevel (8E6O) The intensity value associated with the maximum gradient in the histogram, identifying the intensity with the steepest gradient.

MinimumHistogramGradient (VQB3) The minimum rate of change in intensity values, highlighting areas with gradual intensity transitions.

MinimumHistogramGradientGreyLevel (RHQZ) The intensity value associated with the minimum gradient, identifying the intensity with the smoothest gradient.

GLCM Features

JointMaximum (GYBY) The maximum value in the joint probability distribution of two voxel intensities at a given distance and orientation.

JointAverage (60VM) The average of all possible pairwise intensity values in the joint probability distribution, representing the average co-occurrence of intensities.

JointVariance (UR99) A measure of the variability or spread of the joint probability distribution, indicating how much the intensity pairs deviate from the average.

JointEntropyLog2 (TU9B) A measure of the unpredictability or randomness of the joint probability distribution, calculated using logarithms to the base 2.

DifferenceAverage (TF7R) The average of the absolute differences between the intensities of neighboring voxels in a given direction, capturing texture contrast.

DifferenceVariance (D3YU) A measure of the variability of the difference in intensity values between neighboring voxels.

DifferenceEntropy (NTRS) A measure of the randomness or unpredictability in the distribution of intensity differences between neighboring voxels.

SumAverage (ZGXS) The average of the summed intensities of pairs of voxels at a given distance and orientation, providing insight into the overall intensity relationships.

SumVariance (OEEB) A measure of the spread or variability of the summed intensities of voxel pairs.

SumEntropy (P6QZ) A measure of the unpredictability in the distribution of summed intensities, indicating texture complexity.

AngularSecondMoment (8ZQL) Also known as the uniformity measure, it quantifies the smoothness or homogeneity of the image texture, where higher values indicate a more uniform image.

Contrast (ACUI) A measure of the intensity contrast between neighboring voxels, where higher values indicate greater difference between neighboring intensities.

Dissimilarity (8S9J) A measure of the spatial dissimilarity of intensities between neighboring voxels, with higher values indicating greater dissimilarity.

InverseDifference (IB1Z) A measure of the local homogeneity in intensity, with smaller values indicating greater similarity between neighboring intensities.

NormalisedInverseDifference (NDRX) A normalized version of the inverse difference measure, providing a scale-invariant texture measure.

InverseDifferenceMoment (WF0Z) A measure that quantifies the smoothness of the intensity difference between neighboring voxels, with higher values indicating more uniformity.

NormalisedInverseDifferenceMoment (1QCO) A normalized version of the inverse difference moment, offering a scale-independent measure of image smoothness.

InverseVariance (E8JP) A measure of the variance of the inverse difference, with higher values indicating greater variability between neighboring voxel intensities.

Correlation (NI2N) A measure of the linear dependency or correlation between pixel pairs in the image, capturing how intensities relate to each other spatially.

Autocorrelation (QWB0) A measure of the correlation between a voxel's intensity and that of its neighboring voxel, indicating local texture regularity.

ClusterTendency (DG8W) A measure of the tendency of voxel intensities to form clusters in the image, with higher values indicating more clustering.

ClusterShade (7NFM) A measure of the skewness of intensity clusters in the image, indicating the direction of intensity distribution.

ClusterProminence (AE86) A measure of the prominence or intensity contrast of the clusters in the image, with higher values indicating more pronounced clusters.

GLRLM Features

ShortRunsEmphasis (22OV) A measure of the emphasis placed on short runs of uniform intensity values. Higher values indicate a greater concentration of short runs.

LongRunsEmphasis (W4KF) A measure of the emphasis placed on long runs of uniform intensity values. Higher values indicate a greater concentration of long runs.

LowGreyLevelRunEmphasis (V3SW) A measure of the emphasis placed on runs of low intensity values. Higher values indicate more frequent runs with lower intensity.

HighGreyLevelRunEmphasis (G3QZ) A measure of the emphasis placed on runs of high intensity values. Higher values indicate more frequent runs with higher intensity.

ShortRunLowGreyLevelEmphasis (HTZT) A measure that combines short runs and low intensity values, highlighting the prevalence of short low-intensity runs.

ShortRunHighGreyLevelEmphasis (GD3A) A measure that combines short runs and high intensity values, highlighting the prevalence of short high-intensity runs.

LongRunLowGreyLevelEmphasis (IVPO) A measure that combines long runs and low intensity values, indicating the prevalence of long low-intensity runs.

LongRunHighGreyLevelEmphasis (3KUM) A measure that combines long runs and high intensity values, indicating the prevalence of long high-intensity runs.

GreyLevelNonUniformity (R5YN) A measure of the non-uniformity in the distribution of grey levels, where higher values indicate a greater variety in intensity values within the runs.

RunLengthNonUniformity (W92Y) A measure of the non-uniformity in the distribution of run lengths, where higher values indicate a greater variation in the length of runs.

RunPercentage (9ZK5) The percentage of the total number of runs of a particular length in the image, providing insight into the proportion of each run length type.

NGTDM Features

Coarseness (QCDE) A measure of the texture's "roughness" or coarseness, where higher values indicate a texture with larger and more distinct intensity differences between neighboring voxels.

Contrast (65HE) A measure of the difference in intensity values between neighboring voxels. High contrast values indicate more pronounced intensity differences.

Busyness (NQ30) A measure of the level of rapid intensity changes in the image, indicating the complexity or "busyness" of the texture. Higher values indicate more rapid intensity variations.

Complexity (HDEZ) A measure of the texture's complexity, where higher values indicate a more intricate texture with a greater number of intensity variations and spatial arrangements.

Strength (1X9X) A measure of the intensity difference between neighboring voxels, with higher values indicating a stronger, more pronounced intensity contrast in the texture.

GLSZM Features

SmallZoneEmphasis (5QRC) A measure of the emphasis placed on small zones with uniform intensity. Higher values indicate that smaller homogeneous areas are more prominent in the texture.

LargeZoneEmphasis (48P8) A measure of the emphasis placed on large zones with uniform intensity. Higher values indicate that larger homogeneous areas are more prominent in the texture.

LowGrayLevelZoneEmphasis (XMSY) A measure of the emphasis on zones with low intensity values. Higher values indicate the presence of more low-intensity homogeneous zones.

HighGrayLevelZoneEmphasis (5GN9) A measure of the emphasis on zones with high intensity values. Higher values indicate the presence of more high-intensity homogeneous zones.

SmallZoneLowGreyLevelEmphasis (5RAI) A measure that combines small zones and low intensity values, highlighting the prevalence of small, low-intensity homogeneous zones.

SmallZoneHighGreyLevelEmphasis (HW1V) A measure that combines small zones and high intensity values, highlighting the prevalence of small, high-intensity homogeneous zones.

LargeZoneLowGreyLevelEmphasis (YH51) A measure that combines large zones and low intensity values, highlighting the prevalence of large, low-intensity homogeneous zones.

LargeZoneHighGreyLevelEmphasis (J17V) A measure that combines large zones and high intensity values, highlighting the prevalence of large, high-intensity homogeneous zones.

GreyLevelNonUniformity (JNSA) A measure of the non-uniformity in the distribution of grey levels within the zones, with higher values indicating a greater variety of intensities across the zones.

NormalisedGreyLevelNonUniformity (Y1RO) A normalized version of grey level non-uniformity, providing a scale-invariant measure of the variation in grey levels across zones.

ZoneSizeNonUniformity (4JP3) A measure of the non-uniformity in the distribution of zone sizes, with higher values indicating a greater variety in the sizes of homogeneous intensity zones.

NormalisedZoneSizeNonUniformity (VB3A) A normalized version of zone size non-uniformity, providing a scale-independent measure of the variability in the sizes of zones.

ZonePercentage (P30P) The percentage of the total number of zones of a particular size in the image, providing insight into the prevalence of different zone sizes.

GreyLevelVariance (BYLV) A measure of the variance in grey levels within the zones, indicating the level of intensity fluctuation within each zone.

ZoneSizeVariance (3NSA) A measure of the variance in zone sizes, with higher values indicating greater variation in the size of homogeneous intensity zones.

ZoneSizeEntropy (GU8N) A measure of the unpredictability in the distribution of zone sizes, with higher values indicating more complex and varied zone sizes.

A.2. Radiomic Feature Selection

Algorithm A.1 Radiomic Feature Selection Based on Spearman Correlation

```
for each group (core and ring separately) do
    Compute the absolute Spearman correlation matrix.
    for each feature $i$ in the group (in order) do
        for each feature $j$ with $j > i$ do
            if $|\text{correlation}(feature_i, feature_j)| \geq 0.8$ then
                if ($feature_j$ is in the clinician-selected set) and (feature $i$ is not) then
                    Remove $feature_i$
                else
                    Remove $feature_j$
                end if
            end if
        end for
    end for
end for
Combine the remaining features from all groups.
Compute the absolute Spearman correlation matrix for the combined set.
for each feature $i$ in the combined set (in order) do
    for each feature $j$ with $j > i$ do
        if $|\text{correlation}(feature_i, feature_j)| \geq 0.8$ then
            if ($feature_j$ is in the clinician-selected set) and (feature $i$ is not) then
                Remove $feature_i$
            else
                Remove $feature_j$
            end if
        end if
    end for
end for
```

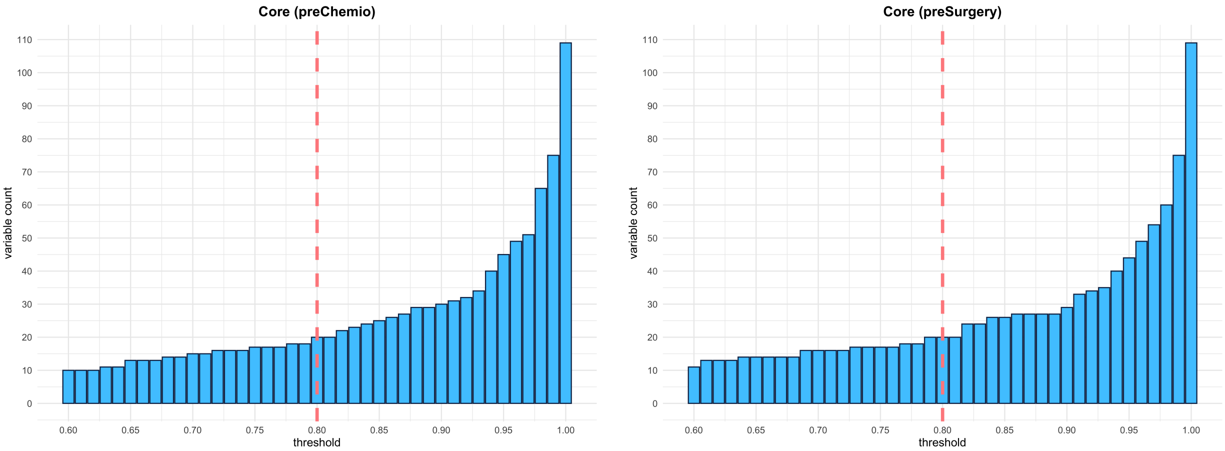


Figure A.1: Number of selected features of core radiomic as a function of the threshold for Pre-Chemotherapy (left) and Pre-Surgery (right) cases. The x-axis represents the threshold values, while the y-axis indicates the number of features retained after applying the filtering procedure

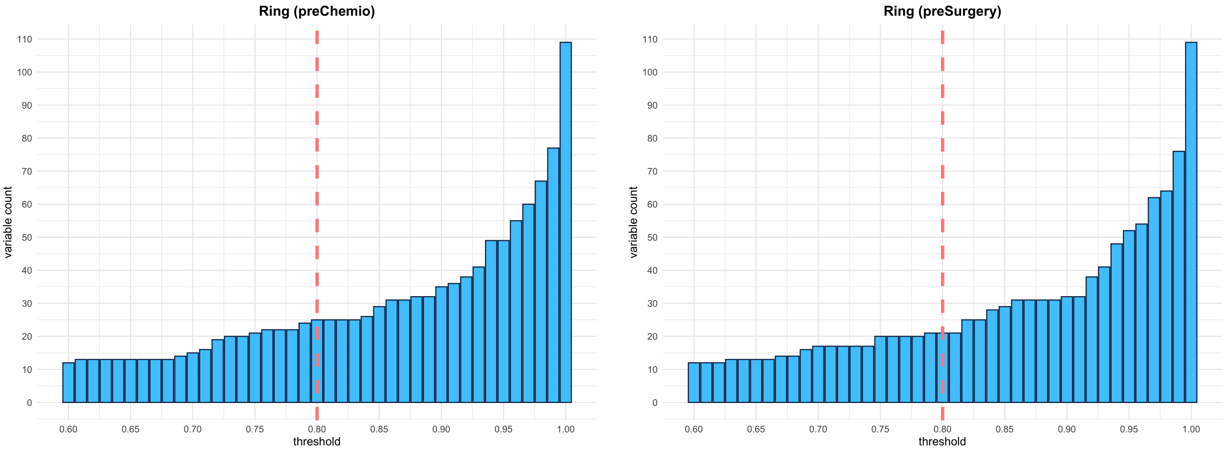
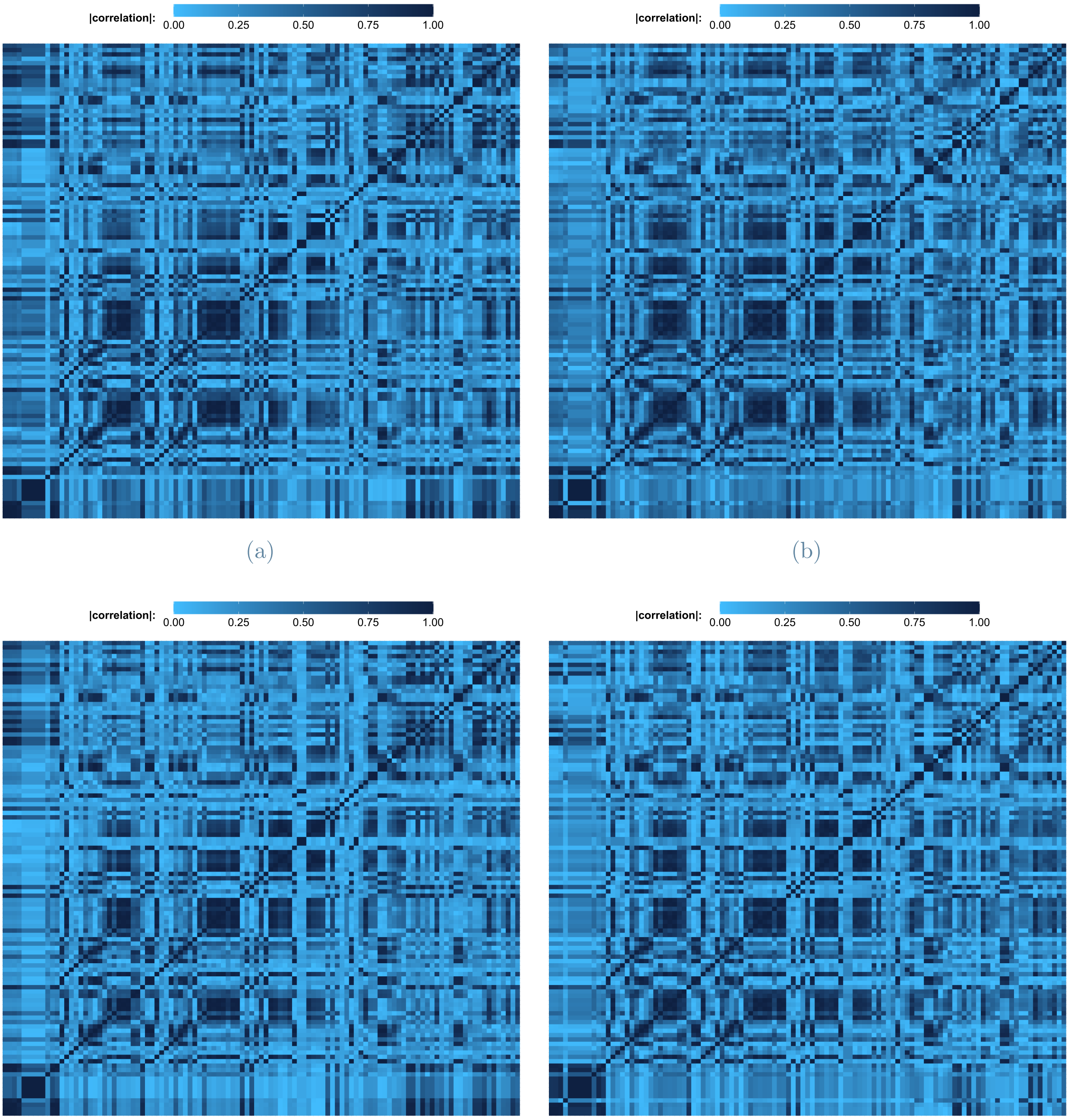


Figure A.2: Number of selected features of ring radiomic as a function of the threshold for Pre-Chemotherapy (left) and Pre-Surgery (right) cases. The x-axis represents the threshold values, while the y-axis indicates the number of features retained after applying the filtering procedure



(a) (b)

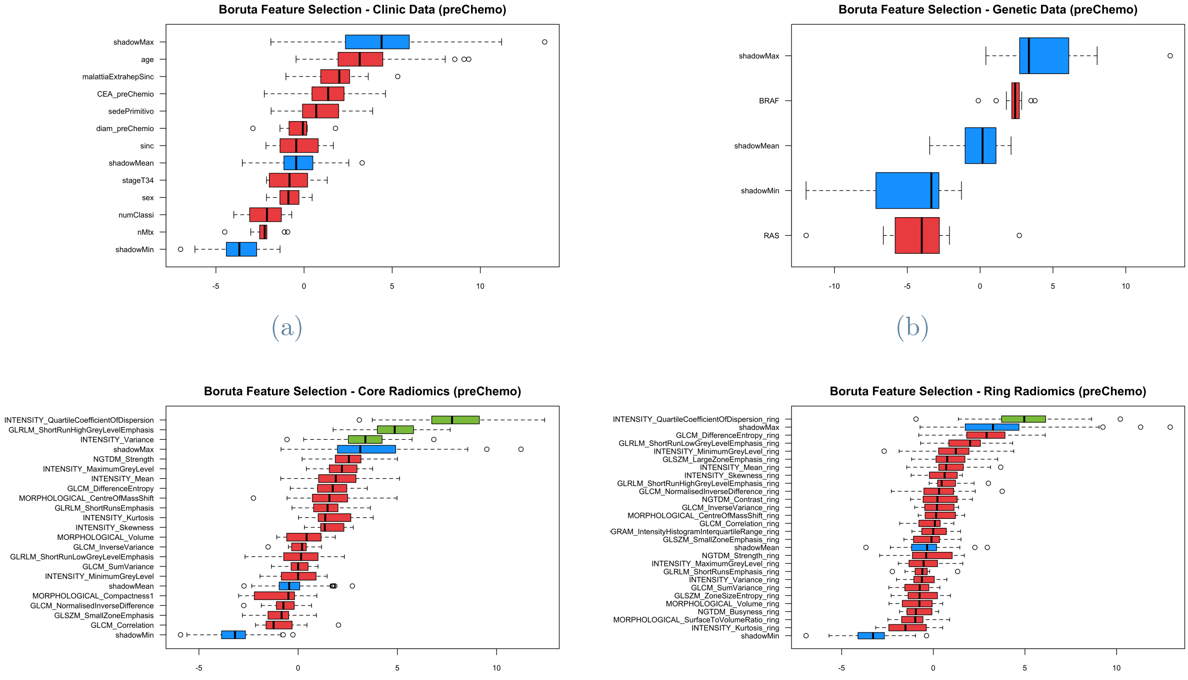
(c) (d)

Figure A.3: Absolute Correlation Matrices of radiomics variables for Pre-Chemotherapy (up) and Pre-Surgery (down) cases, for core (left) and ring (right) RoI. Variable names are omitted for brevity

B | Appendix B

B.1. pre-Chemo Survival Models

Boruta Features Selection



(a)

(b)

(c)

(d)

Figure B.1: Results of the Boruta algorithm applied to pre-Chemo datasets at the first step of feature selection, i.e., applied separately on: Clinical data (a), Genetic data (b), Core radiomics (c), and Ring radiomics (d). Selected variables are colored in green

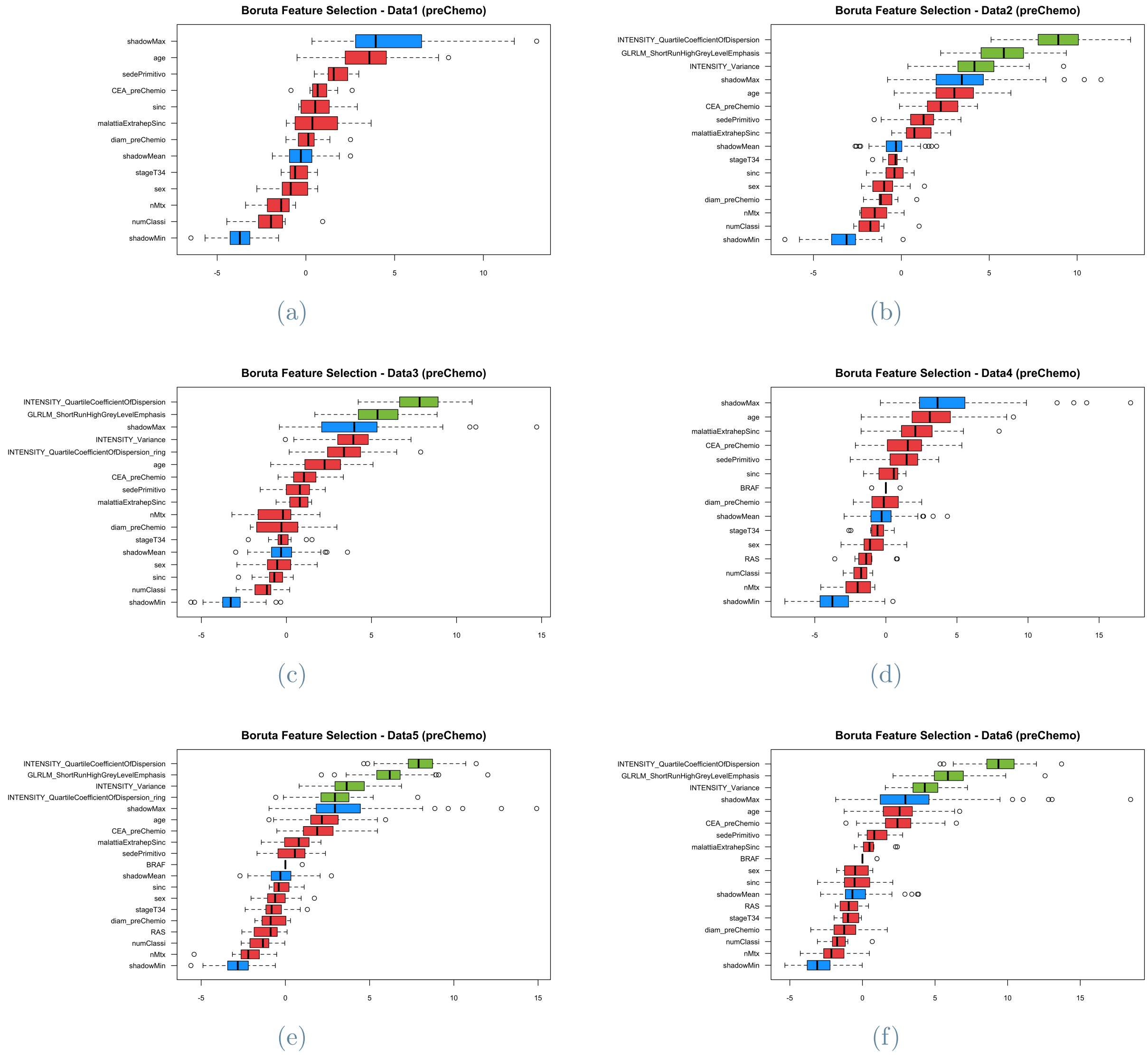
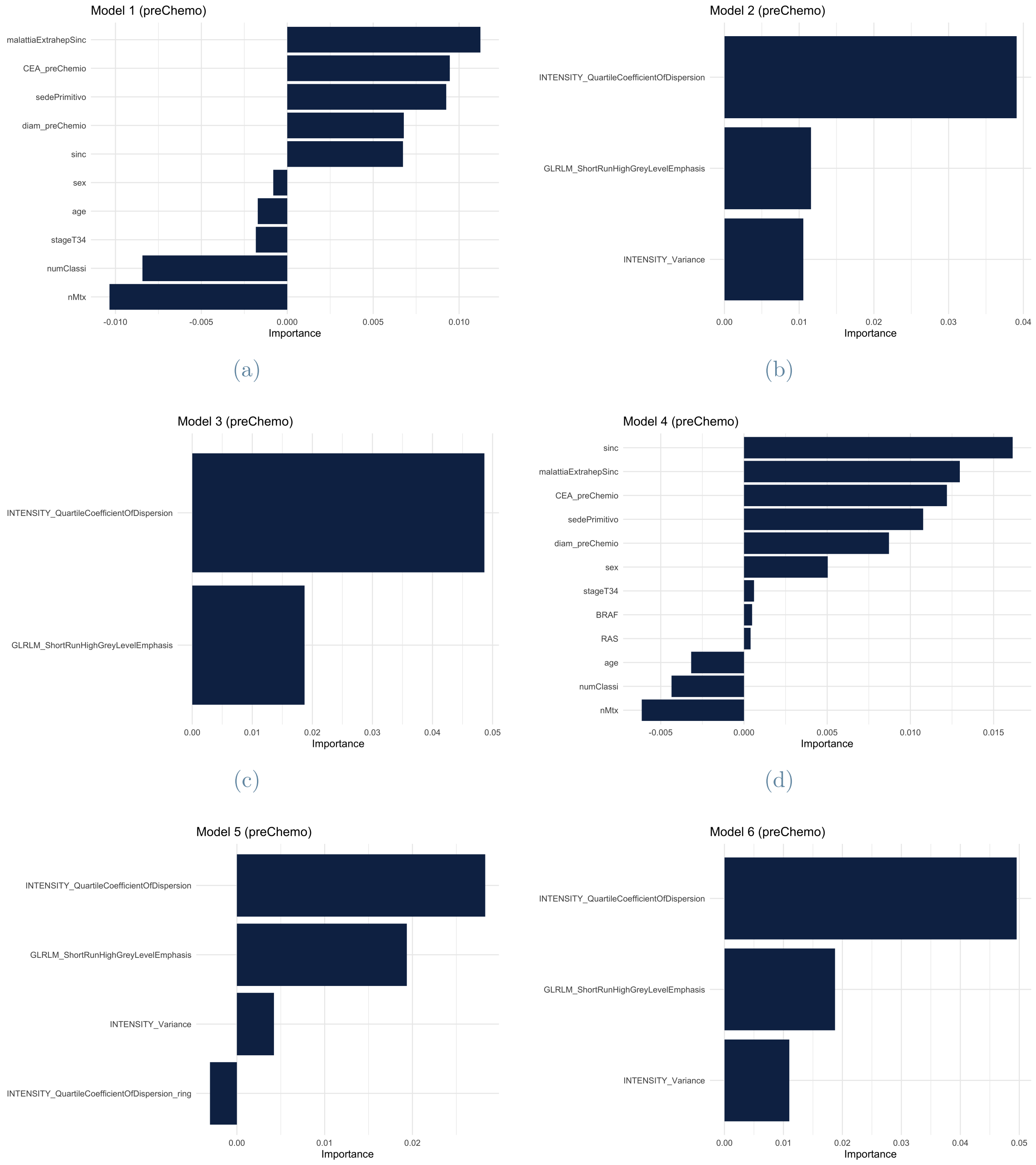


Figure B.2: Results of the Boruta algorithm applied to pre-Chemo datasets at the second step of feature selection, i.e., applied separately on: model 1 (a), model 2 (b), model 3 (c), model 4 (d), model 5 (e), model 6 (f). Selected variables are colored in green

Importance Plots of Random Survival Forest



(a) (b)

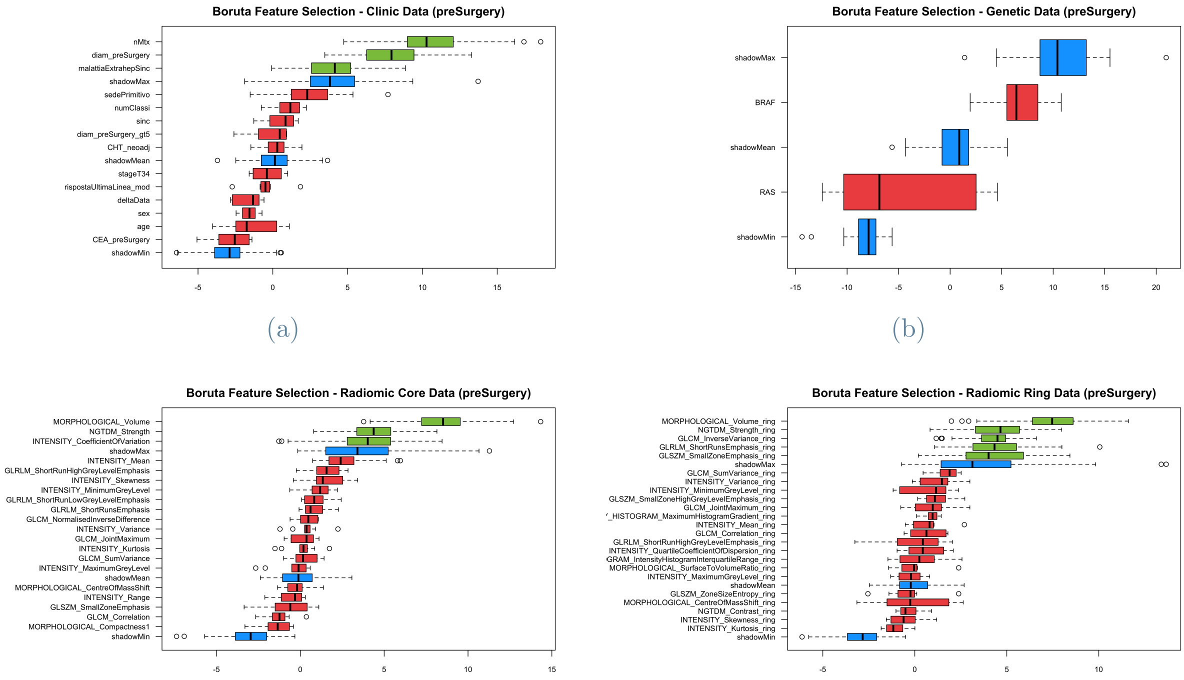
(c) (d)

(e) (f)

Figure B.3: Importance plots of Random Survival Models applied to pre-Chemotherapy models: model 1 (a), model 2 (b), model 3 (c), model 4 (d), model 5 (e), model 6 (f)

B.2. pre-Surgery Survival Models

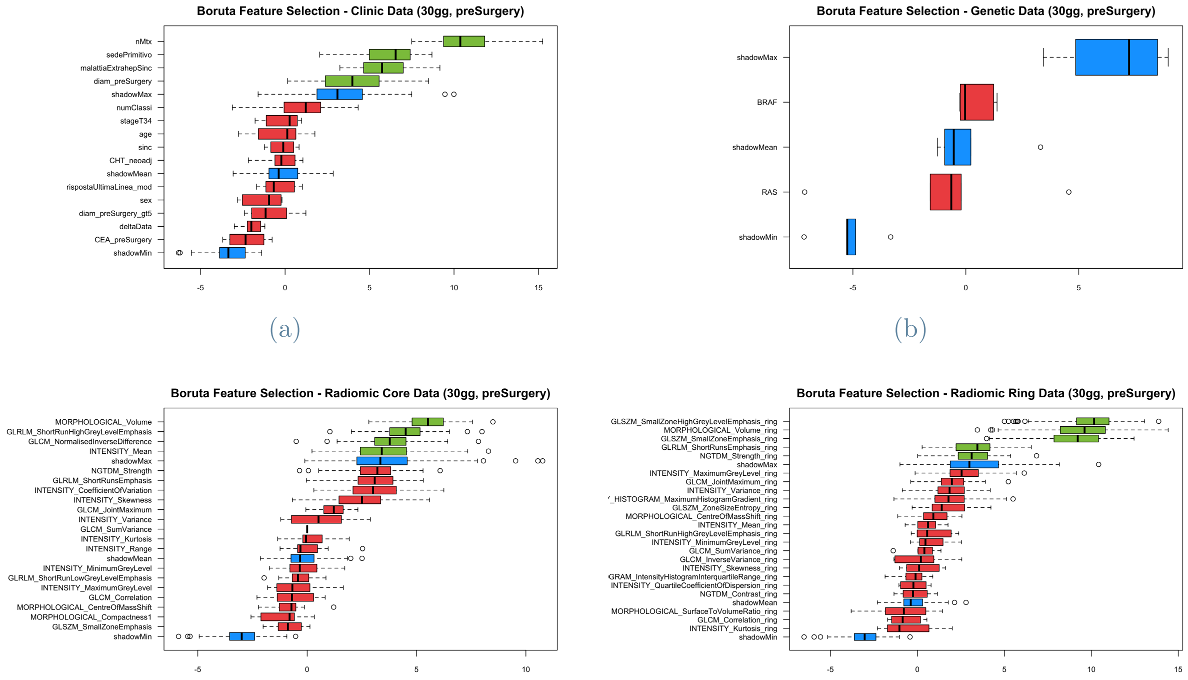
Boruta Features Selection



(a) (b)

(c) (d)

Figure B.4: Results of the Boruta algorithm applied to pre-Surgery datasets at the first step of feature selection, i.e., applied separately on: Clinical data (a), Genetic data (b), Core radiomics (c), and Ring radiomics (d). Selected variables are colored in green



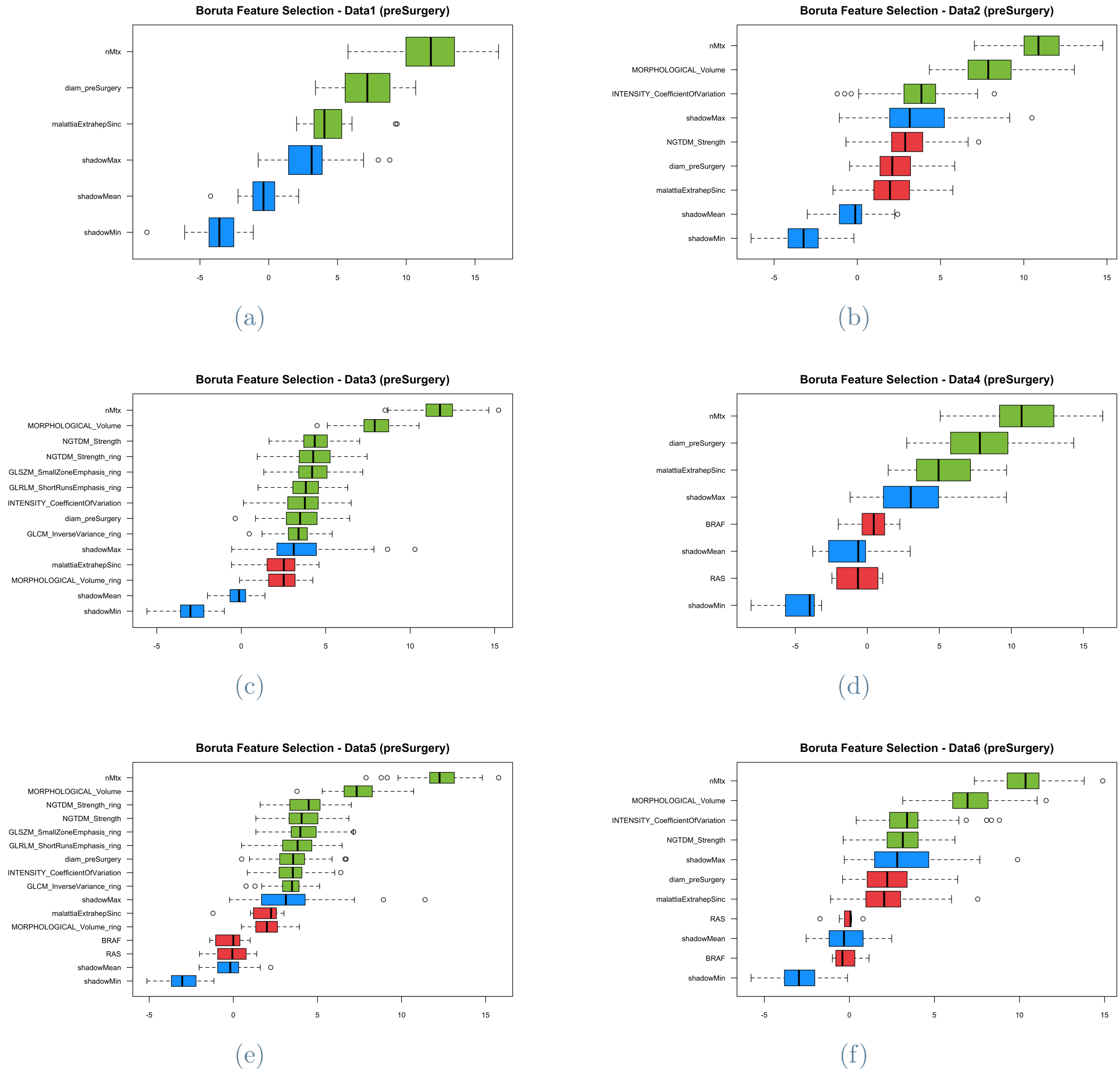
(a)

(b)

(c)

(d)

Figure B.5: Results of the Boruta algorithm applied to pre-Surgery datasets (models at 30gg) at the first step of feature selection, i.e., applied separately on: Clinical data (a), Genetic data (b), Core radiomics (c), and Ring radiomics (d). Selected variables are colored in green


Boruta Feature Selection - Data1 (preSurgery)

Boruta Feature Selection - Data2 (preSurgery)

Boruta Feature Selection - Data3 (preSurgery)

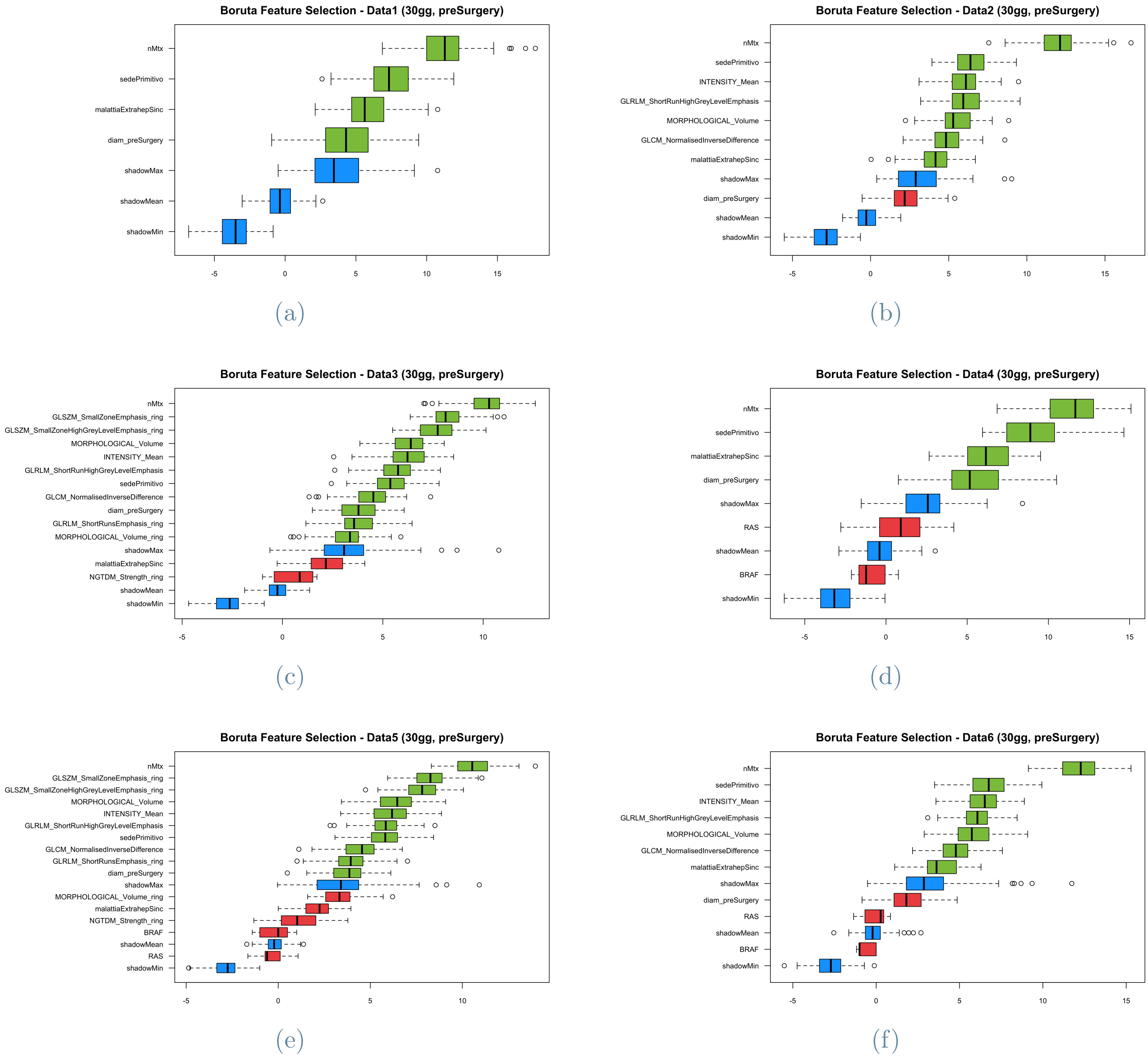
Boruta Feature Selection - Data4 (preSurgery)

Boruta Feature Selection - Data5 (preSurgery)

Boruta Feature Selection - Data6 (preSurgery)

(a) (b) (c) (d) (e) (f)

Figure B.6: Results of the Boruta algorithm applied to pre-Surgery datasets at the second step of feature selection, i.e., applied separately on: model 1 (a), model 2 (b), model 3 (c), model 4 (d), model 5 (e), model 6 (f). Selected variables are colored in green



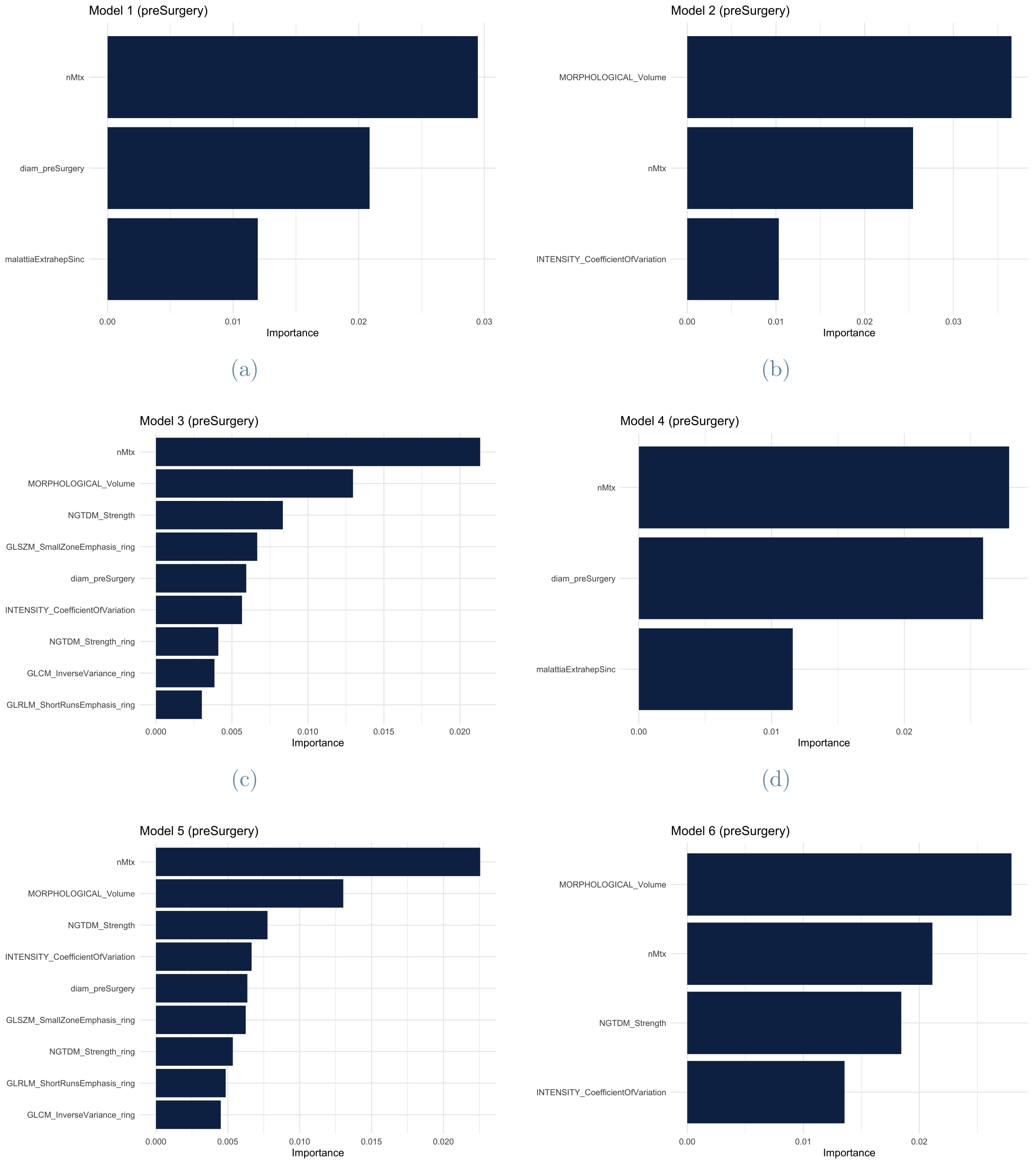
(a) (b)

(c) (d)

(e) (f)

Figure B.7: Results of the Boruta algorithm applied to pre-Surgery datasets (models at 30gg) at the second step of feature selection, i.e., applied separately on: model 1 (a), model 2 (b), model 3 (c), model 4 (d), model 5 (e), model 6 (f). Selected variables are colored in green

Importance Plots of Random Survival Forest

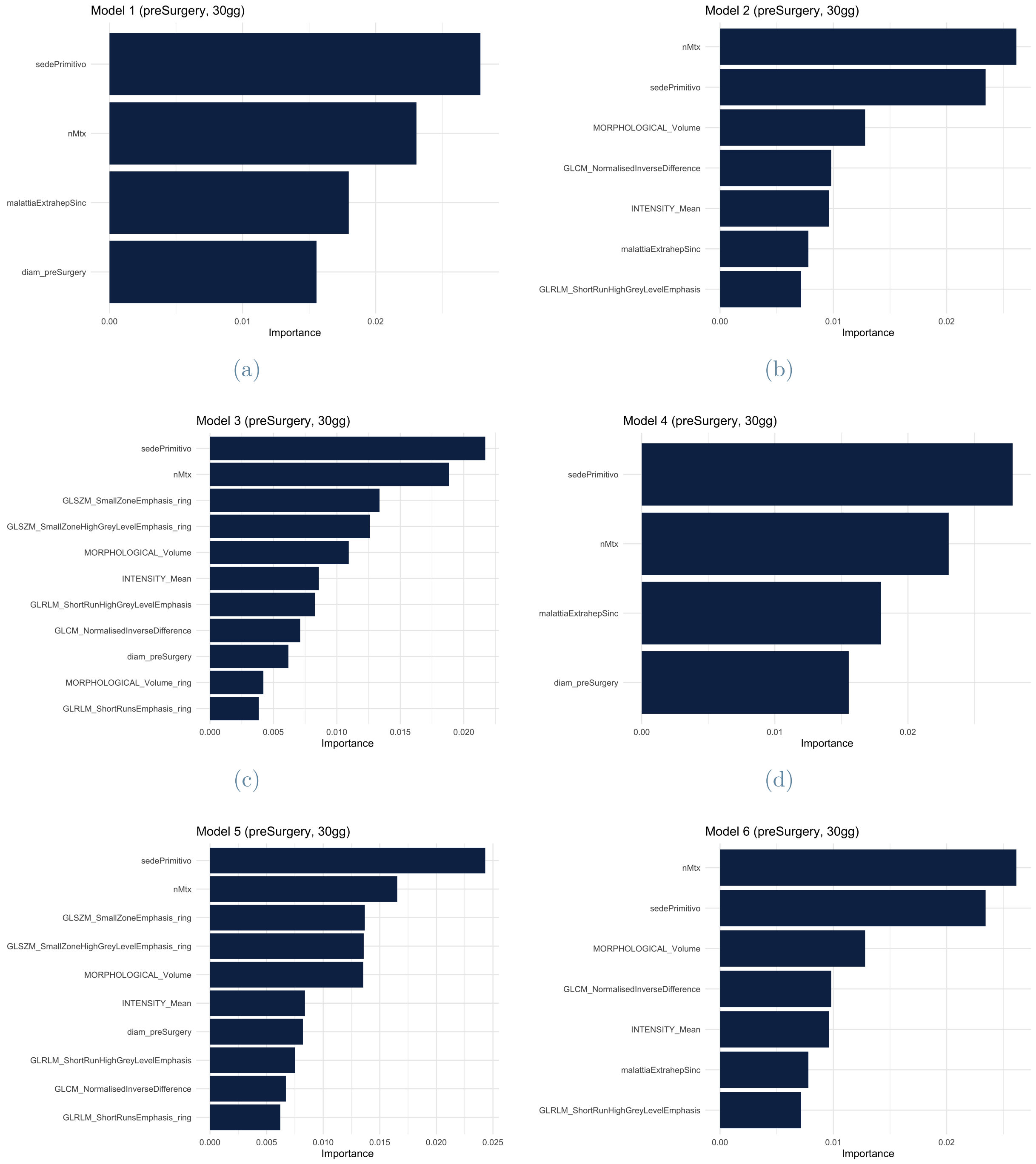


(a) (b)

(c) (d)

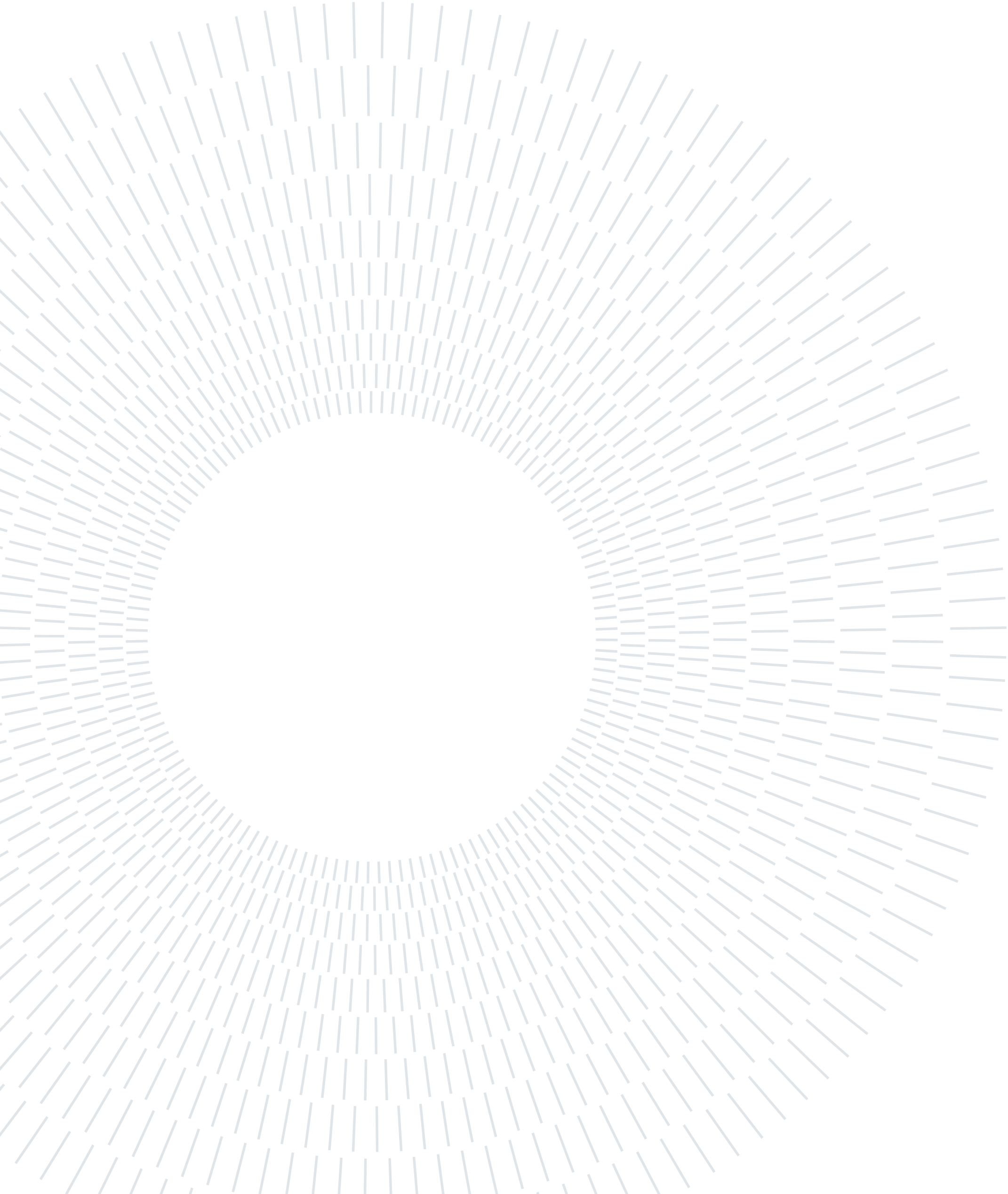
(e) (f)

Figure B.8: Importance plots of Random Survival Models applied to pre-Surgery models: model 1 (a), model 2 (b), model 3 (c), model 4 (d), model 5 (e), model 6 (f)



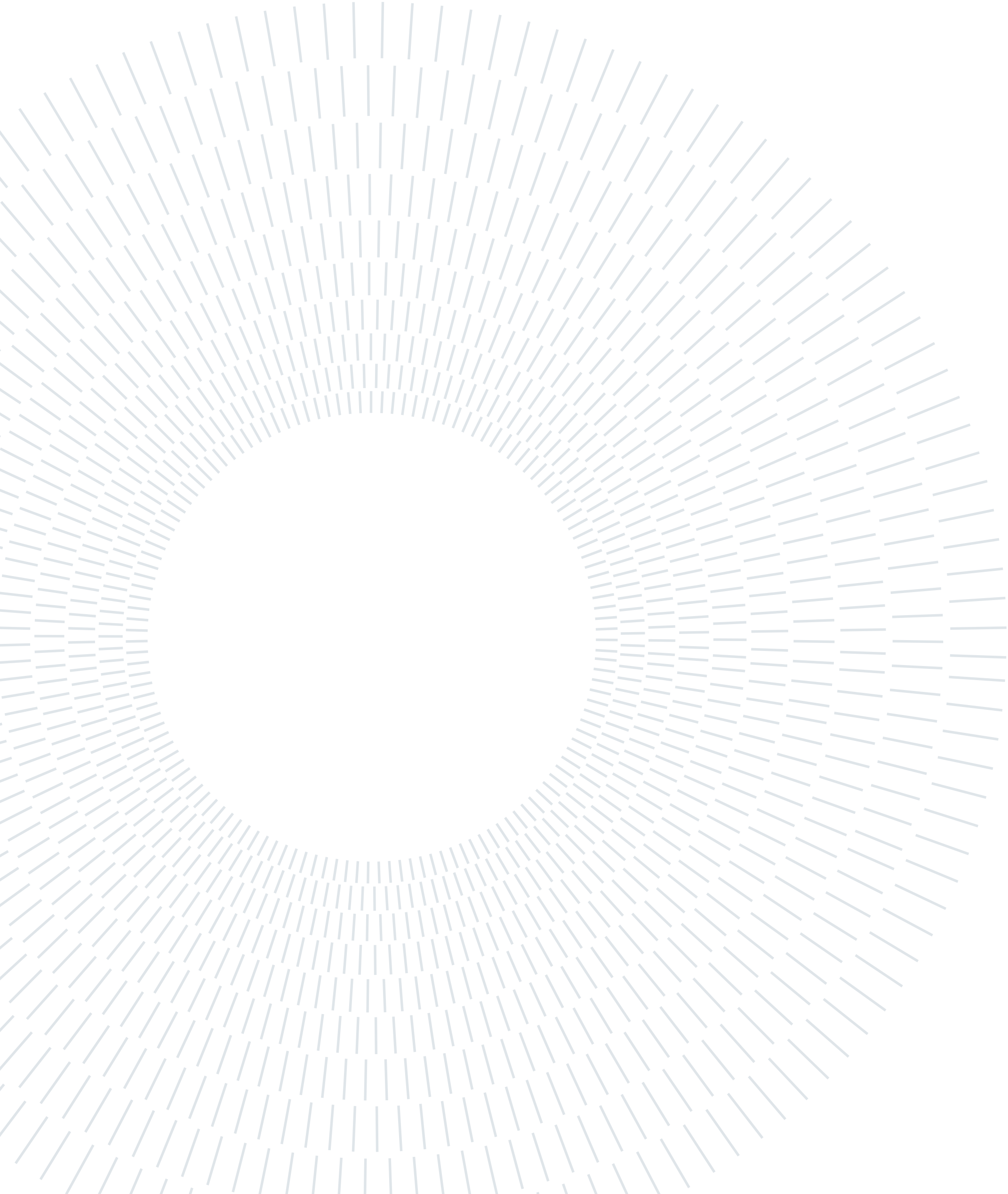
(a) (b) (c) (d) (e) (f)

Figure B.9: Importance plots of Random Survival Models applied to pre-Surgery models (30gg): model 1_30gg (a), model 2_30gg (b), model 3_30gg (c), model 4_30gg (d), model 5_30gg (e), model 6_30gg (f)



List of Figures

List of Tables



Acknowledgements

Data used in this work were provided by Humanitas Research Hospital as part of a monocentric study funded by an AIRC research grant.

I would like to express my gratitude to Prof. Francesca Ieva for her guidance, support, and expert critical insight, which have been fundamental in shaping the direction of this research.

I would also like to thank Prof. Chiara Masci and Prof. Lara Cavinato for their support, contributions, and constructive discussions, which have been essential for refining the methodological framework of this study.

A special acknowledgment goes to Dr. Luca Viganò and Dr. Angela Ammirabile, whose clinical expertise and support have been crucial in integrating the medical perspective into this research.