



UNIVERSITAT DE BARCELONA

Final Degree Project Biomedical Engineering Degree

**“ Development of a preoperative predictive
model of malignancy of indeterminate
ovarian masses “**

Barcelona, 5th of June of 2024

Author: Anna Làzaro Llorens

Director/s: Aida Niñerola i Raúl Tudela

Tutor: Aida Niñerola

Abstract

Technological innovations have transformed healthcare, offering new avenues for disease diagnosis and treatment. Among these advancements, radiomics analysis of MRI images holds great promise for improving the early detection and characterization of ovarian masses. Ovarian cancer, with its high mortality rates attributed to late-stage diagnosis, stands to benefit significantly from these developments.

This study focuses on exploring various pathways before applying radiomics analysis to ovarian masses in MRI images. By investigating different preprocessing methods, such as intensity normalization and registration of MRI images, the project aims to determine their impact on the evaluation of radiomic features.

Through a comparative analysis of the radiomic feature results obtained from these pathways, this research seeks to understand how variations in preprocessing approaches affect the evaluation of radiomic features. Despite challenges such as limited patient samples and time constraints, the study anticipates significant outcomes.

It aims to provide understanding about how various preprocessing techniques impact the assessment of radiomic features. Additionally, it explores the potential for developing a specialized pipeline that incorporates the most effective preprocessing methods for practical use in clinical environments.

By advancing the field of ovarian cancer diagnosis through radiomics and MRI analysis, this research has the potential to improve patient outcomes and healthcare practices. Ultimately, the goal is to contribute to early detection and personalized treatment strategies, ultimately reducing the burden of ovarian cancer on patients and healthcare systems.



Acknowledgements

Estic molt contenta amb la feina que he fet durant aquests mesos i crec que es veu reflectit en el treball. He après un munt del món de les imatges mèdiques i m'ha servit també per plasmar tot el meu aprenentatge d'aquests darrers 4 anys.

Vull donar les gràcies als meus tutors l'Aida i en Raül per haver-me acompanyat durant aquest camí, i per haver-me ajudat cada cop que ho necessitava. Voldria donar les gràcies també a en Marc i la Gemma, ja que aquest camí ha sigut més fàcil amb els nostres esforços units, i estic molt agraïda d'haver compartit aquest semestre amb ells.

D'altra banda, també agrair als amics que m'han recolzat i animat sempre que em feia falta. Especialment a en Beni, que des del primer dia ha sabut escoltar-me i fer el possible per ajudar-me sempre que ho necessitava.

List of figures

Figure 1: 5-year survival rate of ovarian cancer (Based on data from SEER 22 2013-2019)	9
Figure 2: Methodology of the project.	11
Figure 3: Female reproductive system with a tumor in one ovary.	13
Figure 4: Transvaginal ultrasound technique application.	14
Figure 5: T1WI (left) vs T2WI (right) axial MRI pelvic images from a patient.	15
Figure 6: Application of radiomics workflow.	16
Figure 7: Chronological distribution of publications in radiomics for oncology from 2011 to 2020 [23]. .	19
Figure 8: Ovarian cancer diagnostics type market analysis [25].	20
Figure 9: DICOM MRI sequences saved on the computer with hidden names.	31
Figure 10: NifTi MRI sequences after the filtration by name.	31
Figure 11: T2WI (left) vs the registration of T2WI and T1WI (right).	32
Figure 12: Example of a tumor mask segmented with Radiomics Frontiers (Syngo.via).	34
Figure 13: MRI image and tumor mask from 3D Slicer incorrectly overlaid.	34
Figure 14: MRI image and tumor mask from 3D Slicer correctly overlaid.	35
Figure 15: MRI image and tumor mask from Radiomics Frontiers incorrectly overlaid.	36
Figure 16: MRI image and tumor mask from Radiomics Frontiers correctly overlaid.	36
Figure 17: MINT Lesion segmentation interface with delineated tumor.	37
Figure 18: Scheme that indicates the number of each of the methods.	40
Figure 19: Correlation matrix of the features from Method 2.	41
Figure 20: Correlation matrix of the features from Method 2 after the filtration.	41
Figure 21: Skewness average values	43
Figure 22: Skewness standard deviation values.	43
Figure 23: Pearson correlation matrix between methods.	44
Figure 24: WBS of the project.	45
Figure 25: PERT diagram of the project.	47



List of tables

Table 1: Summary table of articles related to radiomics in ovarian masses.	17
Table 2: Proposed solutions for each step of the project.	22
Table 3: Comparison of the MRI sequence proposed options for the project.	23
Table 4: Comparison of the segmentation softwares proposed options for the project.	24
Table 5: Comparison of the preprocessing options for the project.	25
Table 6: Comparison of the implementation of radiomic options for the project.	25
Table 7: Comparison of the groups of radiomic features for the project.	26
Table 8: Proposed solutions of the project in each step.	28
Table 9: Patient's diagnostic and MRI machine acquisition.	30
Table 10: Comparison of conserved features between methods.	42
Table 11: Tasks and timings of the project.	46
Table 12: GANTT diagram of the project.	48
Table 13: SWOT analysis of the project.	49
Table 14: Total cost of the project.	51

Glossary of terms

MRI: Magnetic Resonance Imaging

mRMR: Minimum redundancy maximum relevance

LASSO: Least absolute shrinkage and selection operator

ICC: Intra-class classification

ROI: region of interest

VOI: volume of interest

T2WI: T2 weighted image

T1WI: T1 weighted image

GLCM: Gray Level Co-occurrence Matrix

GLSZM: Gray Level Size Zone Matrix

GLRLM: Gray Level Run Length Matrix

NGTDM: Neighbouring Gray Tone Difference Matrix

GLDM: Gray Level Dependence Matrix

PACS: Picture Archiving and Communication Systems

DICOM: Digital Imaging and Communications in Medicine

NIFTI: Neuroimaging Informatics Technology Initiative

ANTs: Advanced Normalization Tools

STL: Standard Triangle Language

Table of contents

LIST OF FIGURES.....	4
LIST OF TABLES	5
1. INTRODUCTION.....	9
1.1. MOTIVATION.....	9
1.2. OBJECTIVES.....	10
1.3. METHODOLOGY.....	11
1.4. PROJECT LOCATION.....	12
1.5. SCOPE AND LIMITATIONS.....	12
2. BACKGROUND	13
2.1. GENERAL CONCEPTS.....	13
2.1.1. Ovarian cancer	13
2.1.2. Diagnostic methods of ovarian masses	14
2.1.3. Magnetic Resonance Imaging.....	15
2.1.4. Radiomics.....	16
2.2. STATE OF THE ART	16
2.3. STATE OF THE SITUATION.....	18
3. MARKET ANALYSIS	19
3.1. MARKET SECTOR.....	19
3.2. RADIOMICS MARKET EVOLUTION.....	19
3.3. OVARIAN CANCER MARKET ANALYSIS.....	20
4. CONCEPTION ENGINEERING	22
4.1. STUDY OF THE SOLUTIONS.....	22
4.1.1. MRI Sequence.....	23
4.1.2. Segmentation software.....	23
4.1.3. Preprocessing.....	25
4.1.4. Radiomic tool.....	25
4.1.5. Type of radiomic features	26
4.2. PROPOSED SOLUTION.....	28
4.2.1. MRI sequence	28
4.2.2. Segmentation software.....	29
4.2.3. Preprocessing.....	29
4.2.4. Radiomic tool	29
4.2.5. Radiomic features.....	29
5. DETAIL ENGINEERING	30
5.1. PREPARATION OF THE DATA.....	30
5.2. REGISTRATION OF T2 AND T1	32
5.3. INTENSITY NORMALIZATION	33
5.4. PREPARATION OF SEGMENTATIONS	34
5.4.1. 3D Slicer masks.....	34
5.4.2. Syngo.via (Radiomics Frontiers) masks.....	35
5.4.3. Mint Lesion masks.....	36
5.5. APPLICATION OF RADIOMICS	38
5.5.1. Mint lesion.....	38
5.5.2. 3D Slicer and Radiomics Frontiers.....	38
6. RESULTS AND STATISTICAL ANALYSIS	40



7. EXECUTION SCHEDULE	45
7.1. WORK BREAKDOWN STRUCTURE (WBS)	45
7.2. PERT DIAGRAM.....	46
7.3. GANTT DIAGRAM.....	48
8. TECHNICAL VIABILITY	49
8.1. STRENGTHS.....	49
8.2. WEAKNESSES	49
8.3. OPPORTUNITIES	50
8.4. THREATS	50
9. ECONOMIC VIABILITY	51
10. REGULATIONS AND LEGAL ASPECTS	52
11. CONCLUSIONS AND FUTURE LINES	53
12. BIBLIOGRAPHY	55
ANNEX	58

1. INTRODUCTION

1.1. MOTIVATION

Nowadays, cancer is still one of the leading causes of mortality worldwide, laying a significant challenge to humanity's health. The global burden of cancer is immense, with millions of new cases and deaths occurring each year, which is increasing exponentially [1]. This huge problem called cancer highlights the urgent need for improved diagnostic and treatment strategies in all types of cancer.

Among various types of cancer, ovarian masses stand out due to its often-late diagnosis, which leads to a high mortality rate. This explains why it is the 11th most common cancer among women, and the fifth leading cause of cancer-related death among women [2].



Figure 1: 5-year survival rate of ovarian cancer (Based on data from SEER 22 2013-2019)

Figure 1 represents the 5-year survival rate of women with ovarian cancer. Grey people represent those who have died from ovarian cancer, and green figures represent those who have survived 5 years or more. The survival rate for ovarian cancer is notably lower compared to other cancers, primarily because it is often diagnosed at an advanced stage. An early-stage ovarian cancer has a much better prognosis, but unfortunately, only about 20% of cases are detected at this stage [3].

Therefore, one of the primary challenges in combating ovarian cancer is the difficulty in diagnosing the disease at an early stage. Symptoms are often vague and non-specific, such as bloating, pelvic pain, and urinary urgency, which can be easily attributed to benign conditions [4]. Consequently, there is a significant delay in having a diagnosis, leading to an advanced stage of the disease.

Indeterminate ovarian masses cannot be classified as benign or malignant based on initial imaging studies. Current methods, including CA-125 blood tests and transvaginal ultrasonography, are limited for diagnosing which often leads to invasive procedures such as biopsies or surgeries, that have risks and complications [5].

As a woman, I am aware of the health challenges faced by women, particularly in the realm of reproductive health. This project holds special significance, as ovarian cancer exclusively affects women and women's health issues have historically been under-researched and underfunded [6].

Women often encounter biases in healthcare, leading to delayed diagnoses and inadequate treatment. By focusing on ovarian cancer, this project addresses a critical area of women's health, aiming to bridge the gap in diagnostic accuracy and improve the standard of care for women.

My personal motivation stems from a desire to contribute to a field that directly impacts women's lives. Having witnessed the struggles of women battling cancer, I am driven to leverage my skills in biomedical engineering to make a tangible difference. The development of a reliable, non-invasive predictive model has the potential to revolutionize the diagnostic process, offering women a quicker, safer, and more accurate assessment of their health.

1.2. OBJECTIVES

The main goal of this project is to design a pipeline to discriminate malignant and benign ovarian masses from MRI images by using radiomics. To reach this main objective, the project aims to study all the possible paths that can be followed before applying radiomics, to find in the end, which ones give similar radiomic results, which one is more comfortable for the doctors, which one is faster, etc.

Therefore, this project pretends to define the steps needed before the development of a predictive model that will use radiomic features to classify the malignancy of ovarian masses.

In order to achieve the main goal, the project aims to accomplish the following sub-objectives:

- **Stablishment of the best software for segmenting ovarian masses in magnetic resonance images.**

This involves evaluating various segmentation tools to determine which software provides the most accurate, reliable, and user-friendly experience for segmenting ovarian masses from MRI scans. This step is critical as the quality of segmentation directly impacts the subsequent analysis.

- **Study of the preprocessing tools that can be applied to the magnetic resonance images.**

This sub-objective focuses on identifying and evaluating different preprocessing techniques that can be applied to MRI images to enhance their quality and ensure they are suitable for radiomic analysis.

- **Development of a method for extracting texture analysis variables in magnetic resonance images.**

This involves creating or refining methods to extract texture analysis variables from MRI images. Texture analysis is a key component of radiomics, as it helps quantify the heterogeneity of the tissue within the ovarian masses, providing crucial data that can be used to differentiate between malignant and benign masses.

- **Analysis of the radiomic parameters extracted from all the possible paths.**

Once various paths have been explored and data has been collected, this sub-objective involves a comprehensive analysis of the radiomic parameters extracted from these different methods. The goal is to compare and contrast the results to determine which pathways offer the most accurate, reliable, and clinically relevant information.

1.3. METHODOLOGY

As previously mentioned, the primary aim of this project is to systematically compare the different methodologies that can be employed before applying radiomics analysis to the target masks in MRI images. This comparative study is crucial to determine the most efficient, accurate, and clinically practical approach for discriminating between malignant and benign ovarian masses.

The following image provides a visual representation of the various paths explored and how these paths are constructed.

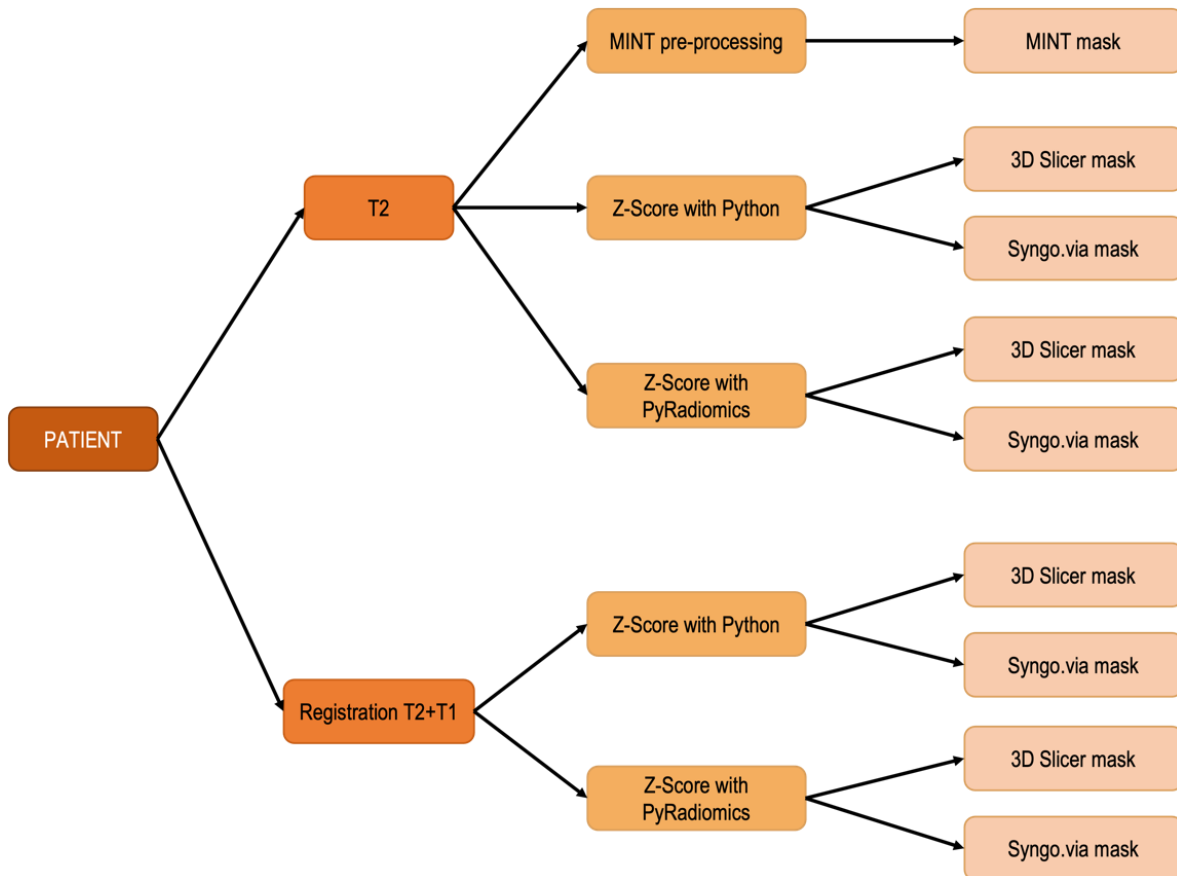


Figure 2: Methodology of the project.

As seen in *Figure 2*, the project involves a comprehensive examination of the preprocessing steps for each patient. Specifically, both T2-weighted imaging (T2WI) and the images resulting from the registration of T2-weighted and T1-weighted imaging (T1WI) are considered for analysis. This dual consideration allows for a robust comparison of the impact of different image types on the final radiomic analysis.

For each type of image (T2WI and registered T2WI-T1WI), two intensity normalization processes are applied from different sources: a python script, and directly from Pyradiomics. Intensity normalization is a critical preprocessing step that adjusts the voxel intensity values of the images to a common scale, thereby reducing variability and enhancing the comparability of the radiomic features extracted from different patients. This normalization is essential for ensuring that the subsequent radiomics analysis is not biased by variations in image acquisition parameters or patient anatomy. Additionally, the project incorporates the MINT preprocessing technique, but it is applied exclusively to the T2WI images due to its closed-software nature.

The final step in the methodology workflow involves the application of segmentation masks. Segmentation is the process of delineating the regions of interest (in this case, the ovarian masses) from the MRI images. The project uses segmentation masks generated by two different segmentation software tools. By superimposing these masks onto the preprocessed images, the radiomic features of the tumors can be calculated. This dual segmentation approach allows for a thorough assessment of the variability and reliability of the segmentation methods, providing insights into which software yields the most consistent and accurate delineations of the ovarian masses. Again, to the MINT preprocessed images, only the segmentations from the same software can be superposed to calculate the radiomics features.

With the preprocessed images and segmentation masks in place, the project proceeds to the extraction of radiomic features. Radiomics involves the quantification of texture, shape, and intensity characteristics of the tumors from the MRI images. These features are probably going to be able to differentiate between malignant and benign ovarian masses, as they capture subtle patterns and variations that are not discernible through visual inspection alone. By analyzing these features across the different preprocessing paths, the project aims to identify the most informative and predictive radiomic parameters, along with the best path for applying this technique.

1.4. PROJECT LOCATION

The project aims to define a pipeline of the steps of another project from a radiologist which works in the same hospital where this work is carried out, the Hospital Clinic in Barcelona, which will be explained in more detail at section 2.3 (State of the situation). Biological images have a really important role in the development of this project, that's why the CDI (Imaging Diagnostic Center) from the hospital and the BIG-UB (BioImaging Group) from the Faculty of Medicine of the University of Barcelona had been involved in it. Weekly online meetings were set up with the tutors, Aida Niñerola and Raúl Tudela, along with Marc Centelles and Gemma Saumell, biomedical engineering students developing their Final Degree Project also about the application of radiomics in Bioimages. In this way they could follow the progress of the project and any doubts that arose were discussed and resolved.

1.5. SCOPE AND LIMITATIONS

The main objective of this research project, as mentioned previously, is to design a pipeline for discriminating between malignant and benign ovarian masses from MRI images using radiomics. The overarching goal is to explore various pathways before applying radiomics analysis, aiming to identify the most effective approach. Additionally, the project seeks to determine if these pathways yield comparable results in terms of accuracy and practicality. In terms of expected outcomes, it is anticipated that the exploration of different pathways will reveal insights into their effectiveness and efficiency. There is a possibility that all paths may demonstrate promising results or produce similar outcomes.

However, despite the project's objectives and anticipated outcomes, certain limitations must be acknowledged. Firstly, the list of patients available for this study is very limited, which may impact the generalizability of the findings. Additionally, the time allocated for conducting the research is constrained, potentially limiting the depth of the analysis. These limitations may influence the scope of the study and the extent to which conclusions can be drawn from the research findings.

2. BACKGROUND

2.1. GENERAL CONCEPTS

2.1.1. OVARIAN CANCER

Ovarian masses or tumors are **abnormal growths** that develop in the ovaries, which are the female reproductive organs responsible for producing ovules and female sex hormones as shown in *Figure 3*. These masses can be formed due to various factors, including genetic predisposition, hormonal factors, environmental factors, and inflammation. The exact cause of ovarian tumors isn't always clear, but certain risk factors, such as family history of ovarian cancer, age, and certain genetic mutations can increase the likelihood of developing ovarian masses.

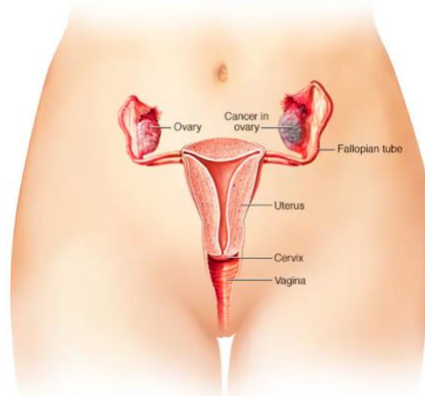


Figure 3: Female reproductive system with a tumor in one ovary [33].

They can be categorized based on their behavior into three main types: benign, malignant, and borderline.

- **Benign Ovarian Tumors:** These are non-cancerous growths that typically do not invade nearby tissues or spread to other parts of the body. They tend to have a well-defined structure and often remain localized within the ovary. Benign ovarian tumors can include various types, such as cystadenomas, fibromas, and mature teratomas. While they are generally not life-threatening, they may still cause symptoms or complications depending on their size and location.
- **Malignant Ovarian Tumors:** These ones are cancerous growths that have the potential to invade nearby tissues and spread (metastasize) to other parts of the body. They are more aggressive and can be challenging to treat, especially if they are diagnosed at an advanced stage. The most common type of malignant ovarian tumor is epithelial ovarian cancer, which includes serous, mucinous, endometrioid, and clear cell carcinomas.
- **Borderline Ovarian Tumors:** These are a type of tumors that exhibit some characteristics of malignancy but lack invasive behavior. They are considered to have a lower potential for spreading compared to malignant tumors. Borderline tumors often have features such as abnormal cellular growth patterns or cellular atypia but do not penetrate the ovarian capsule or invade surrounding tissues. They are typically managed with surgical removal, and the prognosis is generally favorable compared to invasive ovarian cancer.

Ovarian cancer (OC) is the most lethal gynecological cancer and the fifth-leading cause of cancer-related deaths among women. Moreover, this pathology is also related with the **loss of fertility** in a lot of cases. That's why it is important to **detect** those tumors **at an early stage**, allowing a wider range of treatment options, including less invasive surgical procedures and a higher likelihood of successful treatment outcomes [7].

2.1.2. DIAGNOSTIC METHODS OF OVARIAN MASSES

The discrimination of ovarian masses is a common clinical problem, and only a small proportion of these are malignant (30%). **Percutaneous biopsy** of ovarian masses is a **contraindicated technique** due to the **risk of dissemination**, so it is not possible to have a preoperative confirmation of their nature. This failure to determine whether we are dealing with a benign or malignant ovarian mass before surgery causes delays in treatment and inappropriate surgical approaches to malignant ovarian masses that worsen the prognosis and survival of patients. [8]

Transvaginal ultrasound has been a technique of choice for the characterization of ovarian masses during many years. However, up to **25-30%** of ovarian tumors are classified as **indeterminate**, and up to **70%** of those **classified as high risk turn out to be benign**. Therefore, benign masses that could be treated in non-oncological centers occupy oncological surgery spaces in hospitals with the problem of waiting lists and unnecessary healthcare expenses involved [8]. *Figure 4* shows how this technique is done.

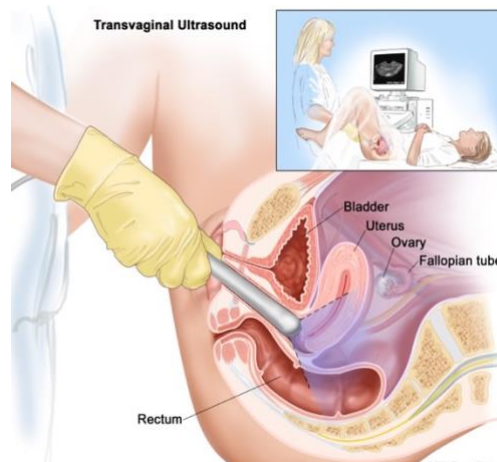


Figure 4: Transvaginal ultrasound technique application [34].

Radiomics from Magnetic Resonance Images (MRI) have recently opened new insights in ovarian cancer tumor detection and discrimination. The most important advantage of this tool is the capability to do an evaluation of the whole tumor which is, in contrary, impossible by doing a biopsy [9].

Moreover, radiomics has recently embraced a more automated pathway with **texture analysis (TA)**. TA includes the **extraction, analysis, and interpretation** of quantitative features from medical images, leading to an exponential amount of data that can be correlated with tumor diagnosis, genomics and/or prognosis [10].

2.1.3. MAGNETIC RESONANCE IMAGING

Magnetic resonance imaging is a medical imaging technique used in medicine as a **diagnostic tool**. It forms anatomical and physiological images from inside the body. To form images of the patients' organs, MRI scanners use strong **magnetic fields, magnetic field gradients, and radiofrequency waves**. These magnetic fields and radiofrequency waves interact with the **hydrogen atoms** in the body, which are abundant in water and fat [11].

When a patient is placed in the MRI scanner, the **strong magnetic field aligns the hydrogen atoms' protons**. A **radiofrequency current** is then pulsed through the patient, **knocking the protons out of alignment**. As the protons realign with the magnetic field, they emit signals that are picked up by the scanner's detectors. These signals are used to generate cross-sectional images of the body. By adjusting the magnetic field gradients and radiofrequency pulses, different tissues can be highlighted, providing detailed images of organs and structures within the body.

MRI images are typically classified based on their weighting: T1-weighted and T2-weighted images. **T1-weighted** images are produced by using **short echo times and repetition times**, making them ideal for visualizing anatomical detail and fat tissue, as fat appears bright, and water appears darker. **T2-weighted** images, on the other hand, are generated using **longer echo and repetition times**, highlighting differences in water content. In T2-weighted images, fluid-filled structures such as edema, cysts, and tumors appear bright, making them useful for identifying pathology.

Figure 5 shows an example of a pelvic MRI, and by looking at the bladder (the big structure in the middle), it can be deduced that the left image corresponds to a T1 because the water inside the bladder appears darker.

By combining T1 and T2-weighted images, radiologists can obtain comprehensive information about tissue composition and pathology, aiding in accurate diagnosis and treatment planning [12].

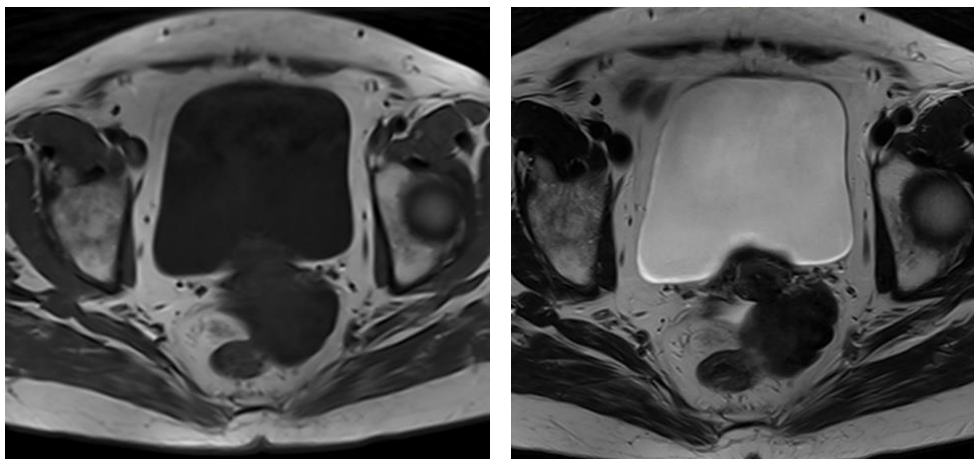


Figure 5: T1WI (left) vs T2WI (right) axial MRI pelvic images from a patient.

2.1.4. RADIOMICS

Radiomics is an advanced field in medical imaging that involves **extracting** many **quantitative features** from medical images using **data-characterization algorithms**. These features, which can include shape, texture, intensity, and spatial relationships, are **not visible to the naked eye** but provide **detailed information** about the **underlying tissue structure and function** [13].

In oncology, radiomics has significant applications. It enhances the ability to **diagnose, characterize, and monitor tumors** through non-invasive imaging techniques such as MRI, CT, and PET scans [14]. By analyzing the extracted features, radiomics can distinguish between benign and malignant tumors, predict tumor behavior, and assess treatment response. This detailed analysis allows for more personalized treatment plans, potentially improving patient's outcome. *Figure 6* shows what are the steps followed in a radiomic study or analysis.

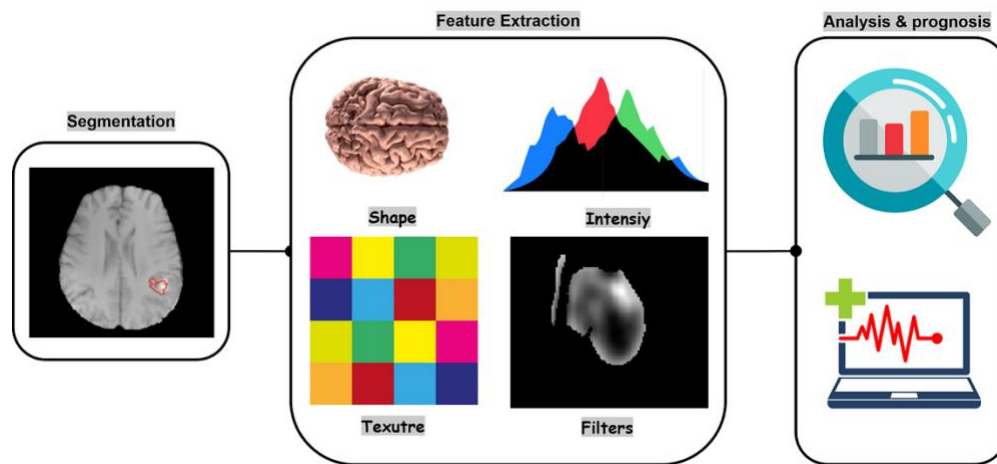


Figure 6: Application of radiomics workflow.

Radiomics also plays a critical role in predicting prognosis. By identifying specific imaging biomarkers, radiomics can help estimate survival rates and the likelihood of disease recurrence. Furthermore, it aids in the development of predictive models that integrate radiomic data with other clinical and genetic information, providing a comprehensive understanding of the tumor biology [15].

Overall, radiomics represents a significant advancement in oncology, offering a deeper insight into tumor characteristics and enhancing the precision and personalization of cancer treatment and management.

2.2. STATE OF THE ART

As explained, radiomics in MR pelvic images has recently been implemented in ovarian cancer tumor detection and discrimination, because of its non-invasiveness, its capacity extracting and analyzing numerous quantitative features and its ability to identify subtle patterns that might indicate early stages of malignancy, among others.

A comprehensive literature review has been conducted to analyze and understand the studies performed to date regarding the use of radiomics in the assessment of ovarian masses in MR pelvic images. This review involved an extensive search of PubMed using the query: "radiomics ovarian masses" and found

a publication that conducts a systematic review of the literature reporting the application of radiomics to imaging techniques in patients with ovarian lesions [16].

I found 6 articles that studied similarly the objective of this project, and I read them. The goal was to gather and synthesize the available evidence on the effectiveness and limitations of radiomics in this specific application. The results of this search, including the objective of the study and how is the feature extraction in radiomics applied, are presented in *Table 1*.

Table 1: Summary table of articles related to radiomics in ovarian masses.

REF.	OBJECTIVE	SEGMENTATION	FEATURE EXTRATION
[17]	Develop a preoperative MRI-based radiomic-clinical nomogram for prediction of residual disease in patients with advanced high-grade serous ovarian carcinoma.	MITK software was used, and the segmentations were manually performed by one gynecological radiologist. The VOI was delineated along the tumor contours slice-by-slice on T2WI and CE-T1WI.	PyRadiomics 841 radiomic features were extracted from each sequence and were subdivided into eight classes.
[18]	Develop and validate an objective MRI-based ML assessment model for differentiating BEOT from MEOT and compare the performance against radiologists' interpretation.	The VOI was delineated manually on the T2WI series in a slice-by-slice manner by an experienced radiologist using the Medical Imaging Interaction Toolkit (MITK) software.	PyRadiomics Multiple MRI features were extracted and were subdivided into eight classes.
[19]	Evaluate the efficiency of 2- and 3-class classification predictive tasks constructed from radiomics features extracted from MRI and pharmacokinetic protocol in discriminating among benign, borderline, and malignant ovarian tumors.	The VOI were manually contoured along the boundary of the tumor in every 2D MRI slice of each ovarian tumor by a radiologist (9 years of experience in MRI) by using ITK-SNAP software.	PyRadiomics Six classes of radiomics features were extracted: 14 shape features, 18 first-order statistical features, 22 GLCM features, 16 GLRLM features, 16 GLSZM features, and 14 GLDM features.
[20]	Investigates whether the texture-based radiomics analysis of magnetic resonance images of the fluid content within ovarian cysts can function as a noninvasive tool in differentiating between benign and malignant lesions.	For image segmentation, a researcher with 15 years of experience in pelvic MRI, reviewed the retrieved sequence and, from every examination, selected a single slice with the largest cross-section area of the previously selected lesion. It doesn't specify the segmentation software.	TexRAD software

[21]	To evaluate the ability of T2-weighted imaging (T2WI)-based radiomics to discriminate ovarian borderline tumors (BOTs) from malignancies based on two-dimensional (2D) and three-dimensional (3D) lesion segmentation methods.	All lesion segmentations were performed by an experienced radiologist. The lesion segmentation on MRI was manually outlined using ITK-SNAP software.	Analysis Kit, version 3.0.0, GE Healthcare
[22]	To evaluate the ability of MRI radiomics to categorize ovarian masses and to determine the association between MRI radiomics and survival among ovarian epithelial cancer (OEC) patients.	All lesion segmentations were performed by an experienced radiologist. They chose one slice with the largest lesion diameter in each protocol as the premium picture for segmenting the whole lesion. The segmentation was manually outlined using MATLAB software.	It doesn't specify the software.

2.3. STATE OF THE SITUATION

As said, this final degree project is the beginning of a project that aims to develop a radiology doctor at the Hospital Clínic in Barcelona. This first approach, as told before, pretends to study all the possible paths that can be followed before applying radiomics to the ovarian masses in the pelvic magnetic resonance images from patients. Then, having these steps defined, the doctor aims to create a database to define a predictive model of malignancy of the ovarian masses.

The need for this methodology is driven by the critical importance of early and accurate detection of ovarian masses. Ovarian cancer is known for its asymptomatic nature in the early stages, often leading to diagnoses at advanced stages when the prognosis is poor. Therefore, having a robust and reliable tool to identify and characterize ovarian masses as benign or malignant as early as possible is crucial. Early detection significantly improves the chances of successful treatment and can greatly enhance patient outcomes.

Moreover, in the future, there's the intention of writing a paper that details the development and validation of the predictive model for ovarian mass malignancy using radiomics. This paper will aim to share the findings with the broader medical and scientific community, highlighting the innovative methodologies used and the clinical implications of the results.

Therefore, it is important to emphasize that this project does not end here, but that it aims to lay the groundwork for ongoing research and development in the field of medical imaging and oncology.

3. MARKET ANALYSIS

3.1. MARKET SECTOR

The market sector for this project lies within the healthcare and medical diagnostics industry, specifically targeting the medical imaging and radiology segment with a focus on oncology and gynecology.

By integrating a predictive model with MR imaging technologies, the project aims to enhance diagnostic accuracy, thereby supporting radiologists in making more informed decisions when dealing with ovarian tumors. This is particularly crucial for women suffering from ovarian cancer, as early and accurate detection significantly improves treatment outcomes.

Furthermore, the project has substantial relevance in gynecological health, aiding gynecologists and oncologists in distinguishing between benign and malignant ovarian tumors preoperatively. Consequently, this enhances operational efficiency and reduces the time and resources required for conclusive diagnoses. By improving diagnostic accuracy and preoperative planning, this work ultimately contributes to better patient outcomes and streamlined healthcare processes.

3.2. RADIOMICS MARKET EVOLUTION

Radiomics is an emerging field that leverages imaging techniques to extract a large number of quantitative features from medical images, which can provide valuable insights into tumor biology, behavior, and pathophysiology. This technology holds significant promise in the field of oncology, as it enables the extraction of data that may not be apparent through traditional radiologic and clinical interpretation.

The radiomics market has experienced exponential growth in recent years. The concepts of image feature extraction have existed for decades, but it is only recently that the application of these techniques has surged, driven by advancements in imaging technology, data analysis, and computational power [23]. As seen in Figure 7, in 2020 there were about 3250 publications referencing radiomics, indicating a rapidly growing interest and research output in the field.

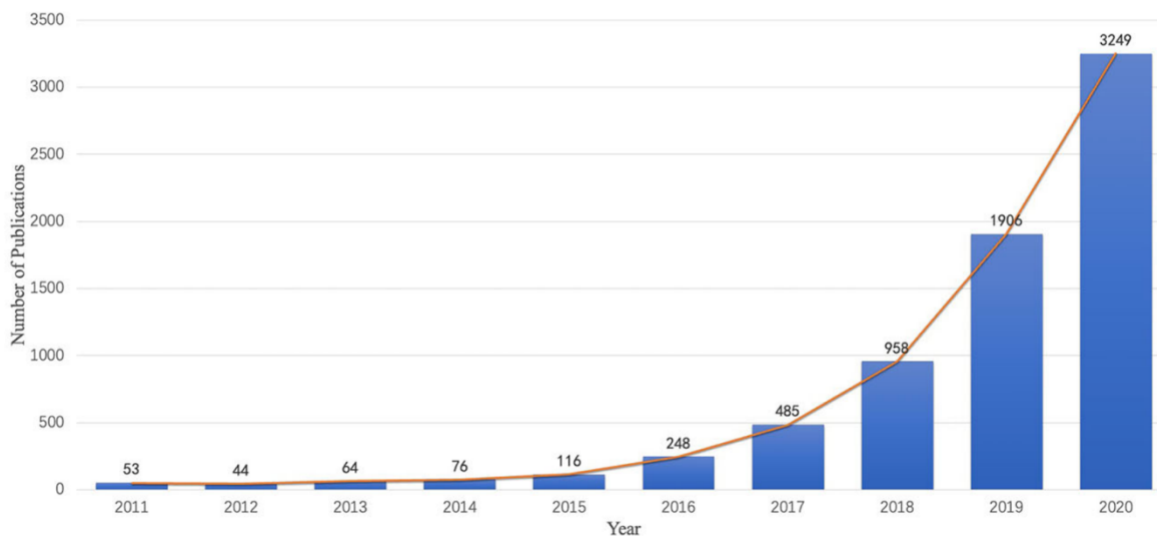


Figure 7: Chronological distribution of publications in radiomics for oncology from 2011 to 2020 [23].

As deduced from the diagram, the number of articles and reviews grew steadily in the first 4 years. From 2015 to 2020, the annual number of publications grew exponentially and peaked in 2020.

According to the results shown in Figure 7, the ten leading countries in the field of radiomics research include a diverse mix of four Asian countries, two American countries, and four European countries. Among these, the United States stands out by contributing a substantial volume of publications, totaling 2,280. Additionally, the USA has been notable for its frequent collaborations with researchers from other nations, enhancing the global exchange of knowledge and expertise. Together, publications from the USA and China account for more than 40% of all research papers in this area, underscoring the significant contributions and influence of these two countries in advancing the field of radiomics [23].

The demand for radiomics in oncology is fueled by its potential to enhance personalized medicine. By analyzing imaging data at various time points, radiomics can track tumor progression and response to treatment, enabling more tailored therapeutic approaches. This has significant implications for improving the accuracy of cancer diagnoses and the efficacy of treatments, ultimately leading to better patient outcomes.

3.3. OVARIAN CANCER MARKET ANALYSIS

The global ovarian cancer diagnostics market was valued at USD 4.41 billion in 2022 and is projected to grow at a compound annual growth rate (CAGR) of 4.93% during the forecast period. This growth is driven by an increase in research and development activities aimed at developing novel diagnostic tests for ovarian cancer.

Innovations in diagnostic technologies are a significant market driver. For instance, scientists from the University of Gothenburg and Uppsala University have developed a blood test that can accurately diagnose ovarian cancer, potentially reducing the need for exploratory surgeries. This advancement can facilitate early detection and treatment, significantly impacting patient outcomes [24].

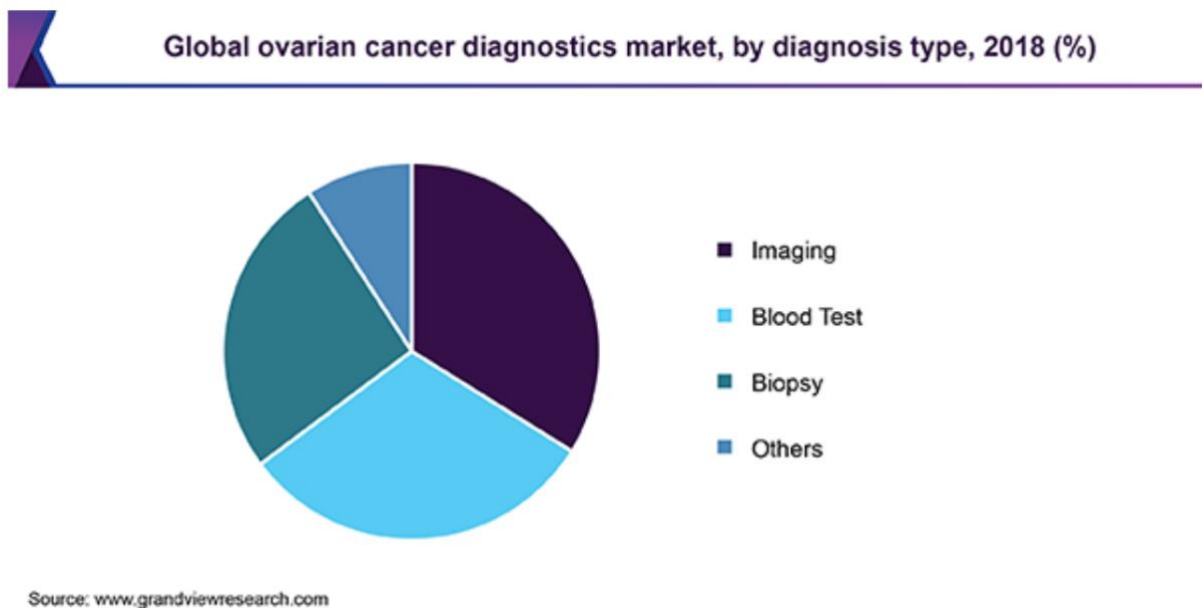


Figure 8: Ovarian cancer diagnostics type market analysis [25].

As seen in Figure 8, imaging techniques accounted for the largest market share in 2018, primarily due to their crucial role in early diagnosis and cancer staging. Transvaginal ultrasound continues to be the preliminary imaging technique for visualizing a tumor internally and measuring its size. Although CT scans are preferred for cancer staging, they have limitations with detecting small tumors. More advanced imaging techniques, such as MRI and PET scans, are increasingly utilized to detect metastasis and assess disease severity.

The growth of the imaging segment can be attributed to several factors. Firstly, there is an increasing patient inflow for cancer diagnosis in hospitals, driven by a growing awareness among consumers about the importance of early detection. Additionally, rising investments in healthcare infrastructure, particularly in developing and underdeveloped countries, are contributing to the segment's expansion. These investments are enhancing the availability and accessibility of advanced imaging technologies, thereby improving diagnostic capabilities and patient outcomes [25].

4. CONCEPTION ENGINEERING

This section conducts a conceptual engineering analysis to evaluate all possible solutions for the project procedure. The evaluation aims to identify and select the best options to achieve the project objectives.

It is important to say that there are options that haven't been discarded in this step of the project. In fact, as this project aims to define the steps to create a pipeline, the study has been carried out using all possible paths to, in the end, carry out a concordance study to analyze if there's similarity between paths to finally decide which one is the most comfortable and simple for those who will use this pipeline in the future.

4.1. STUDY OF THE SOLUTIONS

Creating a pipeline for the classification of ovarian masses means evaluating many tool and method options to obtain an optimal result. In the following table, there are exposed the different solutions proposed for each step of the project.

Table 2: Proposed solutions for each step of the project.

Studied solutions	
MRI sequence	<ul style="list-style-type: none"> - T2WI - T1WI
Segmentation software	<ul style="list-style-type: none"> - 3D Slicer - MINT Lesion - Syngo.via
Preprocessing	<ul style="list-style-type: none"> - Registration of T2 and T1 - Intensity normalization
Radiomic tool	<ul style="list-style-type: none"> - PyRadiomics - MINT Lesion radiomics - Syngo.via (Frontiers) radiomics
Type of radiomic features	<ul style="list-style-type: none"> - First order - Shape - GLCM - GLSZM - GLRLM - NGTDM - GLDM

4.1.1. MRI SEQUENCE

As explained in more detail in the background section, T1WI and T2WI are types of MRI sequences that provide different contrasts and information about tissue properties. T1-weighted images highlight fat and make anatomical structures appear with high contrast, showing clear detail of organs and tissues. T2-weighted images, on the other hand, emphasize fluid content and are particularly useful for detecting edema, inflammation, and other pathological changes. This section pretends to find which MRI sequence is better for segmenting the tumor.

Table 3 outlines the 2 possible solutions proposed for the pipeline.

Table 3: Comparison of the MRI sequence proposed options for the project

Options	Properties
T2WI	<ul style="list-style-type: none"> • Better visualization of the lesion in comparison with T1- weighted images. • Enhanced contrast between pathological tissues and normal tissues. • A lot of studies show a better accuracy using T2-weighted images in comparison with T1 [21]. • Can be more susceptible to motion artifacts, which can affect the quality of the images and the accuracy of radiomic analysis. • The acquisition typically takes longer compared to T1-weighted images.
T1WI	<ul style="list-style-type: none"> • Provide excellent anatomical detail and high spatial resolution. • Less sensitive to fluid content compared to T2-weighted images, which may limit their effectiveness in detecting cystic components within ovarian masses. • Do not provide as high contrast between pathological tissues and normal tissues as T2-weighted images, making it harder to identify certain types of pathologies.

4.1.2. SEGMENTATION SOFTWARE

The evaluation of the segmentation software is a crucial step in this project. Segmentation tools are essential for doctors as they enable the precise delineation of structures of interest, such as ovarian masses in our case.

Each segmentation software comes with its own unique set of features, algorithms, and user interfaces, but they all share the common objective of improving diagnostic accuracy and efficiency. This evaluation involves comparing various software options to determine which one offers the best combination of accuracy, ease of use, and compatibility with existing medical imaging systems. All this information, including a detailed comparison of different software packages, is summarized in *Table 4*.

Table 4: Comparison of the segmentation softwares proposed options for the project.

Options	Properties
<p style="text-align: center;">3D Slicer</p>	<ul style="list-style-type: none"> • Free, open-source software platform for medical image analysis and visualization. • Able to read DICOM (Digital Imaging and Communications in Medicine) files. • Wide range of tools for image segmentation, registration, and visualization, including 3D reconstruction of medical images. • It has semi-automatic segmentation tools, but the best one for the delimitation of ovarian masses is the layer-by-layer contour of the lesion. • Multiple exporting formats, including NifTi (Neuroimaging Informatics Technology Initiative). • Has plugins and extensions available to enhance its functionality. • Doesn't have CE marking. • Not the smoothest.
<p style="text-align: center;">Mint Lesion</p>	<ul style="list-style-type: none"> • Software platform developed by Mint Medical that evaluates and reports imaging studies, particularly in oncology. • Direct access to PACS. • Has a semi-automatic tool that segments the lesion by applying a threshold when it sees a change in intensity in a region delimited in the form of a circle. It is applied in all the layers, so it is quite fast. • Has a radiomic tool that calculates first order and GLCM features and saves them in a CSV. • Has CE marking. • DCM, Planar Figure and NRRD as export format. • Closed software. • Each license and its annual maintenance are quite expensive.
<p style="text-align: center;">Syngo.via</p>	<ul style="list-style-type: none"> • Software platform developed by Siemens Healthineers for medical imaging and healthcare information technology. • Direct access to PACS. • Has some tools for image segmentation, but the best one for the delimitation of ovarian masses is the layer-by-layer contour of the lesion. • Has CE marking. • Only STL (Standard Triangle Language) as export format. • Closed software.

4.1.3. PREPROCESSING

Preprocessing is essential in radiomics to ensure the reliability and comparability of the extracted features, considering that pelvic images come from different MRI machines. *Table 5* shows two preprocessing steps that can be done (one or both) to the images to ensure, as told, reliability.

Table 5: Comparison of the preprocessing options for the project.

Options	Properties
Registration of T2 and T1	<ul style="list-style-type: none"> • Allows the integration of the high-contrast fluid-sensitive information from T2 with the detailed anatomical information from T1. • Registering T1 images to the T2 space can help mitigate the impact of the mentioned artifacts by providing additional anatomical context. • Movement of the patient due to inspiration and expiration between the acquisitions may lead to a bad registration of the sequences. • Requires a script for registration, which takes time to fully execute.
Intensity normalization	<ul style="list-style-type: none"> • Reduces the variability in signal intensity across different MRI scans, which can be caused by differences in acquisition protocols, scanner types, and patient-specific factors. • It makes radiomic features less sensitive to variations in image acquisition parameters, leading to more stable and reliable features. • There are many types of intensity normalization: z-score, min-max, etc. and the choice of method can influence the results.

4.1.4. RADIOMIC TOOL

Choosing the right software/tool for applying radiomics to the MR images can have an influence on the results and, that is why, it is important to check if they lead to the same or similar values. Two of the segmentation softwares provide their own radiomic tool and can lead to a reduction of the overall time. The 3 proposed options are described in more detail in *Table 6*.

Table 6: Comparison of the implementation of radiomic options for the project.

Options	Properties
PyRadiomics	<ul style="list-style-type: none"> • Open-source tool. • A python script and the image with the segmented mass are needed. • Provides a wide array of radiomic features and they can be chosen by interest. • The adjustments/preprocessing applied to the image can also be chosen before the calculation of the features. • It requires well-preprocessed and segmented images.

MINT Lesion radiomics	<ul style="list-style-type: none"> • Only calculates first order and GLCM features. • You don't need a script to calculate them, so it is a fast step. • You can choose the preprocessing applied to the image before calculating the features. • It requires well-segmented images to ensure a good implementation.
Syngo.via (Radiomics Frontiers)	<ul style="list-style-type: none"> • Calculates approximately between 300 and 400 features. • You don't need a script to calculate them, so it is a fast step. • It is based on PyRadiomics. • It requires well-segmented images to ensure a good implementation.

4.1.5. TYPE OF RADIOMIC FEATURES

Radiomics provides us with a lot of information about the ROI, since it allows us to calculate different types of characteristics from the images and their segmentation. *Table 7* describes the properties of each group of radiomic characteristics.

Table 7: Comparison of the groups of radiomic features for the project.

Options	Properties
First order	<ul style="list-style-type: none"> • Describe the distribution of voxel intensities within the image region defined by the mask through commonly used and basic metrics. • Are relatively simple to calculate and interpret. • Its extraction is computationally less demanding compared to higher-order texture features. • Provide a basic quantitative summary of the image. • Only describe the distribution of individual voxel intensities within the ROI and not the spatial relationships between them. • Due to their simplicity, they may have lower discriminative power compared to higher-order texture features.
Shape	<ul style="list-style-type: none"> • They include descriptors of the two and three-dimensional size and shape of the ROI. • They are independent from the gray level intensity distribution in the ROI.

<p>GLCM (Gray Level Co-occurrence Matrix)</p>	<ul style="list-style-type: none"> • Capture the spatial relationships between pixel intensities, providing detailed information about the texture of the ROI. • Have higher discriminative power compared to first-order features. • The calculation of them is more computationally intensive than first-order features. • Are very dependent on the accuracy of the ROI segmentation. • Can be more difficult to interpret than first-order features.
<p>GLSZM (Gray Level Size Zone Matrix)</p>	<ul style="list-style-type: none"> • Capture the size of homogeneous zones of connected pixels with the same gray level. • They often have strong discriminative power for differentiating between different tissue types and pathological conditions. • Their calculation is computationally intensive. • They are also dependent on the accuracy of the ROI segmentation.
<p>GLRLM (Gray Level Run Length Matrix)</p>	<ul style="list-style-type: none"> • Quantify the texture of an image by calculating the length of consecutive pixels that have the same gray level value. • Are useful for distinguishing between homogeneous and heterogeneous tissues. • Their calculation is computationally intensive. • The accuracy and reliability of them can be affected by the resolution and quality of the input images.
<p>NGTDM (Neighbouring Gray Tone Difference Matrix)</p>	<ul style="list-style-type: none"> • Quantifies the texture of an image based on the differences between a pixel's gray level and the average gray level of its neighboring pixels. • Are sensitive to contrast variations, enabling the detection of variations in the tissue density and structure. • It involves analyzing each pixel and its neighborhood's, which can be computationally intensive, particularly for large 3D images. • Its accuracy depends on the quality and resolution of the input images.
<p>GLDM (Gray Level Dependence Matrix)</p>	<ul style="list-style-type: none"> • Evaluates the number of connected voxels within a certain distance that have a similar gray level. • Their calculation can be computationally intensive, particularly for large 3D images. • Its accuracy depends on the quality and resolution of the input images.

4.2. PROPOSED SOLUTION

Considering all the options and the methodology implemented in this project, this section aims to discard those options that are not suitable but, does not necessarily stay with just one solution in each step.

Table 8 shows the proposed solutions of each step.

Table 8: Proposed solutions of the project in each step.

Proposed solutions	
MRI sequence	- T2WI
Segmentation software	- 3D Slicer - MINT Lesion - Syngo.via
Preprocessing	- Registration of T2 and T1 - Intensity normalization
Radiomic tool	- PyRadiomics - MINT Lesion radiomics
Type of radiomic features	- First order - GLCM

4.2.1. MRI SEQUENCE

Although both T2WI and T1WI options exhibit advantageous properties, T2WI images are generally superior to T1WI images when it comes to segmenting ovarian lesions and applying radiomics to the acquired images. This fact is supported not only by the numerous articles referenced in section 2.2, the state of the art, but also by the radiologist involved in the project.

T2WI images provide enhanced contrast between pathological and normal tissues, which is particularly beneficial for visualizing ovarian lesions. In T2 images, fluids appear bright, and most tumors have higher water content than the surrounding normal tissues, making them more conspicuous. This increased visibility facilitates more precise segmentation of the lesions, which is critical for accurate radiomic analysis. By distinguishing the lesion more clearly from the surrounding tissue, T2WI enables a more accurate and reliable extraction of radiomic features, which are essential for the development of predictive models of malignancy.

4.2.2. SEGMENTATION SOFTWARE

The project aims to comprehensively compare the effectiveness, functionality, accuracy, reliability, and user interface of the three proposed segmentation software tools. It is important to note that none of the software tools have been eliminated from the options; all have been rigorously tested and evaluated.

All three segmentation tools—3D Slicer, Radiomics Frontiers from Syngo.via, and Mint Lesion—have been rigorously tested and evaluated during the course of the project. Each tool offers unique features and capabilities that can potentially impact the segmentation process and the subsequent radiomic analysis. By developing different methods using these three tools, the project aims to highlight how each software handles the task of segmenting ovarian lesions and then compare the radiomic features from all the different methods.

4.2.3. PREPROCESSING

Regarding the preprocessing procedures, both approaches have been thoroughly explored and implemented during the development of the project. Each method has been carefully evaluated to ensure the most accurate and reliable outcomes.

The registration of T2WI and T1WI can integrate of the high-contrast fluid-sensitive information from T2 with the detailed anatomical information from T1. This integrated approach may improve the accuracy of tumor segmentation and characterization, leading to more precise diagnostic and prognostic assessments.

Additionally, regarding the normalization of intensities, it has been chosen to make a z-score. This method centers the average intensity at 0 and scales the standard deviation to 1, effectively standardizing the intensity values across all images. By doing so, it ensures that the radiomic features extracted from different images are comparable, reducing the potential for bias and improving the reliability of the subsequent analysis.

4.2.4. RADIOMIC TOOL

For the radiomic tool, the first to be discarded was Syngo.via (Radiomics Frontiers) by Siemens, as it calculates an overwhelming number of features without allowing the user to select specific ones. Consequently, MINT Lesion and PyRadiomics were chosen. MINT Lesion is preferred for calculating a manageable set of 42 features, which can be thoroughly analyzed, while PyRadiomics allows users to select specific features for extraction, offering greater flexibility and control.

4.2.5. RADIOMIC FEATURES

Finally, only first order and GLCM features have been chosen to evaluate the results of all the possible paths in this study. This decision is primarily driven by the capabilities of the MINT Lesion radiomic tool, which is specifically designed to extract these features. To ensure a consistent and fair comparison across different methods, it is crucial that all approaches utilize the same set of radiomic features. By focusing on first order and GLCM features, the study can maintain a standardized framework for analysis, allowing for a more accurate assessment of each method's performance.

5. DETAIL ENGINEERING

In order to carry out the project and meet the objectives, a series of steps had to be taken, which will be described in this section with detail.

5.1. PREPARATION OF THE DATA

The first step was to download from the PACS (Picture Archiving and Communication Systems), that holds patient's imaging records, the pelvic MRI files in DICOM format. The radiologist who intends to continue with the project had previously provided an excel sheet (*Table 9*) with the clinical history numbers of 12 patients with ovarian masses, their diagnosis, and the MRI machine brand that captured the image.

Table 9: Patient's diagnostic and MRI machine acquisition.

Patient	Diagnostic	MRI machine
1	Benign	SIGNA Explorer
2	Benign	Signa HDxt
3	Malign	Signa
4	Benign	Aera
5	Malign	Aera
6	Borderline	MAGNETOM Vida
7	Malign	MAGNETOM Vida
8	Malign	Signa HDxt
9	Benign	Prisma_fit
10	Borderline	MAGNETOM Vida
11	Benign	Aera
12	Malign	SIGNA Explorer

At this point in the work, it was not yet known which MRI series we would use to carry out the study. It is for this reason that all the DICOM files of each patient were downloaded.

When you download a Magnetic Resonance acquisition in DICOM format from a PACS, it may happen that image metadata, such as series names, are hidden or not directly visible. *Figure 9* shows how every patient's data was stored in the computer.

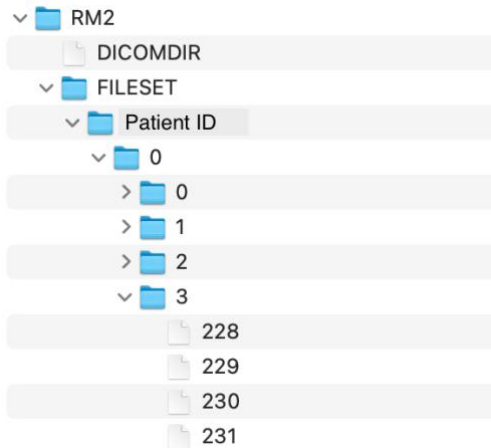


Figure 9: DICOM MRI sequences saved on the computer with hidden names.

Therefore, a Python script was developed to convert these DICOM files to NifTi (Neuroimaging Informatics Technology Initiative), because is a more flexible format and is easier to analyze it.

The series were saved with its corresponding name by accessing at the DICOM metadata stored in the files. The function created in the script, not only converted the images, and saved them with the correct name, but kept only those that met the following conditions:

- T1 or T2 was in their name.
- “Ax” (from axial) or “Tra” (from transversal plane) was also in their name.

The function had as inputs, the directory with the patient folders containing the DICOM images and the directory where the NifTi images were to be saved with the name filtering done.

The transformation of the images was made using DCM2NII, a software tool, that can be easily used in Python, designed to convert medical images from the DICOM format to the NIFTI format. This software is available on GitHub, where users can access the source code, documentation, and updates [26]. The repository provides detailed information on how to use the tool, supported features, and any known issues.

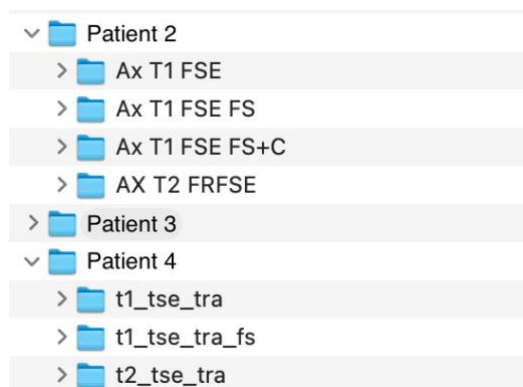


Figure 10: NifTi MRI sequences after the filtration by name.

As seen in *Figure 10*, the filtration of the series' name was not exact, and one or more than 2 images were converted to NifTi format for T2 and T1 series. Depending on the MRI machine, the series had different names, as patient 2 and 4, that show different metadata description of the series.

“FS” stands for fat suppression, and those series were not necessary for this project. So, the next step included another filtration of the name but, the complicity was greater because “fs” could not be a removing condition, because patients like number 2 had in every serie's name this exact character string.

5.2. REGISTRATION OF T2 AND T1

Having in consideration the filtration condition mentioned in the previous section, a registration function was developed in another Python script.

For every patient, two empty lists were created: one to save the paths that contained ‘t1’ and the other for the paths that contained ‘t2’ in their name. As the desired series were the ones that didn't have the “FS” term in their name, the easiest solution was to save only the shortest term of each list. In *Figure 11* there's a screenshot of this part of the explained function.

Finally, with only the desired T2WI and T1WI series, the next step consisted of the registration of these 2, that was applied using the ANTs (Advanced Normalization Tools) module for Python [27]. It is a set of open-source software tools designed for medical image processing, especially for image registration and normalization. This software is also available on GitHub, where users can access the source code, documentation, and updates.

T2 were taken as the fixed images in all cases, and T1 were the moving ones. The type of registration applied was the SyN (Symmetric Normalization), which is a nonlinear deformation method that aligns two images by finding a transformation that minimizes the differences between them. The distinctive feature of SyN is that it performs the registration symmetrically, which means that it treats both images equally during the alignment process.

Figure 11 shows a comparison between a T2 image and the registered T2 with T1 of the same patient.

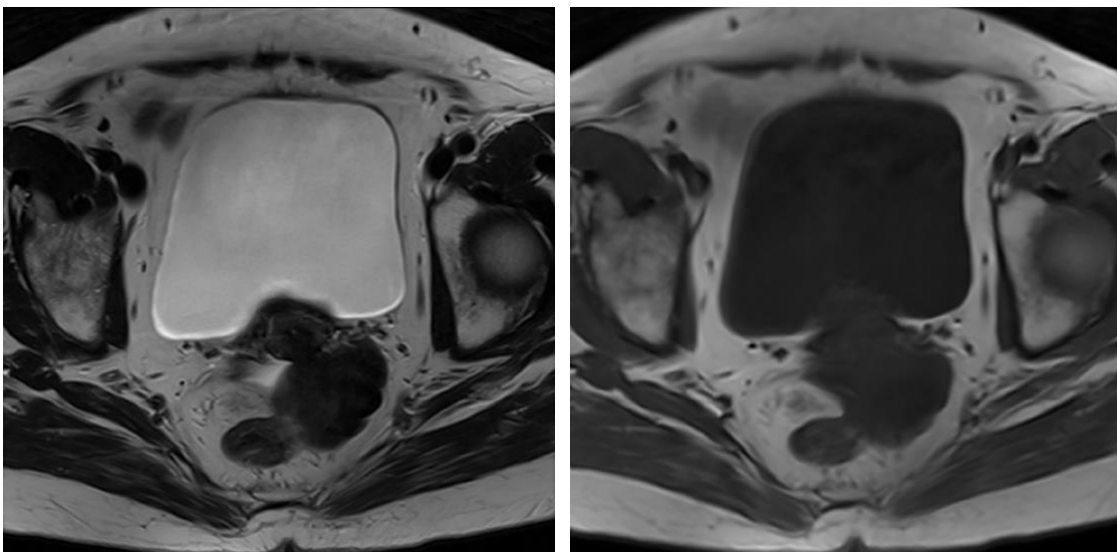


Figure 11: T2WI (left) vs the registration of T2WI and T1WI (right).

It was a necessary step because the registered images (T2 + T1) could yield better results when applying radiomics compared to using only T2 images. By integrating information from both T2 and T1 modalities, a more comprehensive set of features that reflect the underlying tissue characteristics can be captured more accurately.

5.3. INTENSITY NORMALIZATION

The application of intensity normalization to the MRI images was a necessary step due to the variability in intensity levels resulting from differences in acquisition protocols and imaging parameters across the MRI machines variety. As these images were originated from diverse machines, it was probable that they had inconsistencies in intensity values.

A normalization function was created using Python, which aimed to mitigate these discrepancies by standardizing the intensity ranges across all images, thereby ensuring comparability and consistency in the radiomic feature extraction.

A z-score normalization was the one described in the function, which rescales the data so that its mean is 0 and standard deviation is 1, enabling comparison between different datasets by placing them on a common scale. It subtracts the mean from each data point and then divides by the standard deviation, as seen in the following equation:

$$\text{Normalized data} = \frac{X - \mu}{\delta}$$

Being:

- X the image data
- μ the mean value of the image data
- δ the standard deviation of the image data

In the normalizing function created with Python programming software the image path, the output directory and the patient ID were taken as the inputs.

The image was read using the Nibabel package, that can read and write imaging data in various file formats, including NifTi. Through the utilization of the "*get_fdata()*" function, the image data was retrieved as an array, thereby facilitating its subsequent manipulation. Therefore, employing NumPy, a Python library used for working with arrays, calculations were performed to determine the mean and standard deviation of each image to calculate, in the end, using the mentioned formula, the normalized data was calculated.

Finally, the new image was created from the normalized array and saved in the output directory in NifTi format using the affine transformation matrix of the initial image.

5.4. PREPARATION OF SEGMENTATIONS

In this project, a radiologist segmented tumors in 12 patients using the three software tools studied in the previous section. Each software provided unique tools for segmentation and offered different formats for exporting the segmented data. It is important to note that the segmentations were not perfect, primarily due to the depth of each imaging cut, which was approximately 4mm. As illustrated in *Figure 12*, the resulting masks were not as continuous and smooth as anticipated, exhibiting a somewhat scaled appearance.

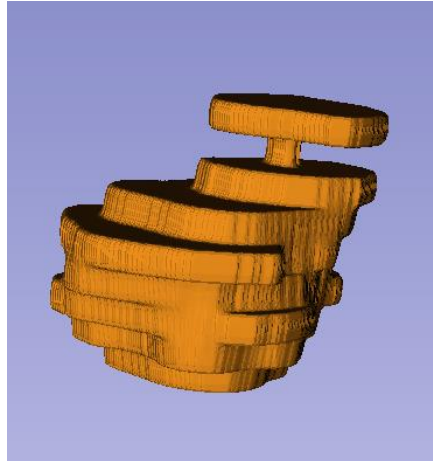


Figure 12: Example of a tumor mask segmented with Radiomics Frontiers (Syngo.via).

5.4.1. 3D SLICER MASKS

The segmentations performed using 3D Slicer involved delineating the tumor layer by layer for each patient. These segmented images were then directly exported into NifTI format.

Before proceeding with the extraction of radiomic features, it was essential to verify that the mask and the corresponding image aligned correctly. This verification was achieved by plotting the overlay of the image with its corresponding segmentation using the matplotlib package in Python, as shown in *Figure 13*.

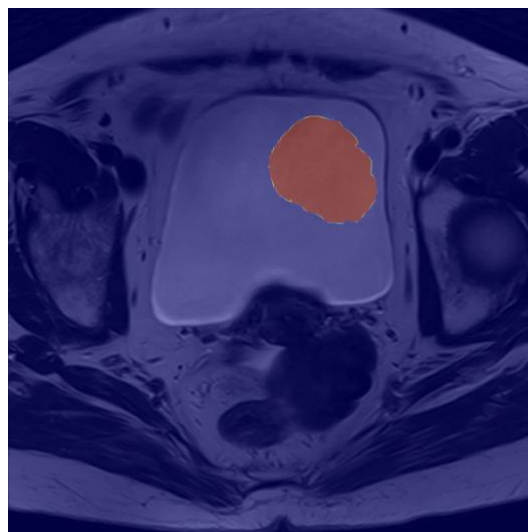


Figure 13: MRI image and tumor mask from 3D Slicer incorrectly overlaid.

As deduced from *Figure 13*, the segmentation was flipped along the X-axis, with the X-axis being the horizontal axis. When dealing with medical imaging data, particularly in formats like NIfTI, it is crucial to understand the spatial orientation conventions used by different software packages and imaging devices.

These conventions can vary, leading to issues such as flipped images when visualizing or processing them together. The common axis conventions include:

- **Radiological Convention:** Images are viewed as if looking from the feet of the patient towards the head. The left side of the image corresponds to the right side of the patient.
- **Neurological Convention:** Images are viewed as if looking from the top of the patient's head down towards the feet. The left side of the image corresponds to the left side of the patient.

To correct this misalignment, a Python script was written to flip the segmentation along the X-axis, ensuring that both the image and the segmentation followed the same orientation. The resulting superposed images, after correction, are shown in *Figure 14*.

This detailed approach highlights the importance of understanding and addressing the orientation conventions in medical imaging. Proper alignment is crucial for accurate radiomic analysis and subsequent data interpretation. Despite the initial challenges with segmentation continuity and orientation, the methodological corrections implemented ensured reliable and consistent results across the different software tools used in this study.

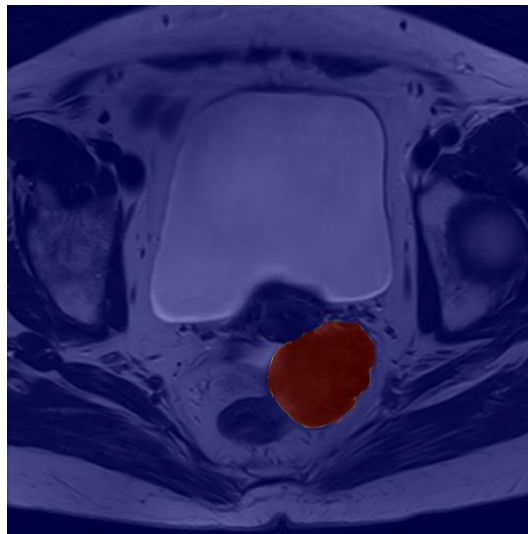


Figure 14: MRI image and tumor mask from 3D Slicer correctly overlaid.

5.4.2. SYNGO.VIA (RADIOMICS FRONTIERS) MASKS

The segmentations performed using Radiomics Frontiers, a module from Syngo.via, were delineated layer by layer for each patient. These segmentations were exported in STL (Standard Triangle Language) format. Unfortunately, a Python script to convert these segmentations to NifTI format could not be developed in time. Consequently, 3D Slicer software was employed to perform this conversion due to its ability to quickly import STL files and export them to NifTI format.

Following the conversion process, it was essential to verify that the mask and the corresponding image were correctly aligned. This verification step was conducted by plotting the overlay of the image with its

corresponding segmentation using the matplotlib package in Python, as described previously. The resulting superposition can be observed in *Figure 15*.

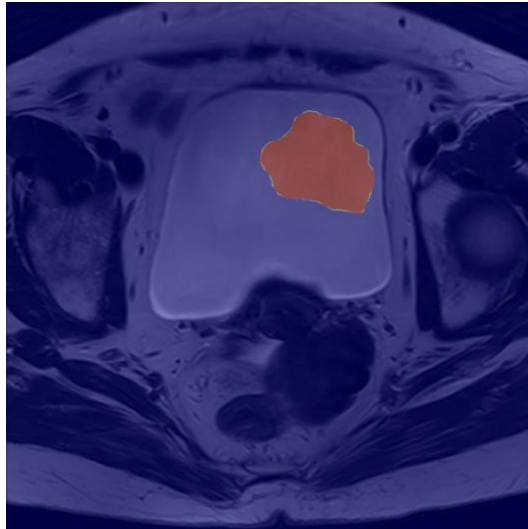


Figure 15: MRI image and tumor mask from Radiomics Frontiers incorrectly overlaid.

This process underscores the importance of ensuring proper alignment between segmented masks and their corresponding images, particularly when different formats and software tools are involved. Despite the initial format incompatibility, the use of 3D Slicer facilitated a swift and efficient conversion, allowing for the continued application of radiomic analysis.

To correct the axis orientation of the mask, the same Python script was used. *Figure 16* shows the result of this last process.

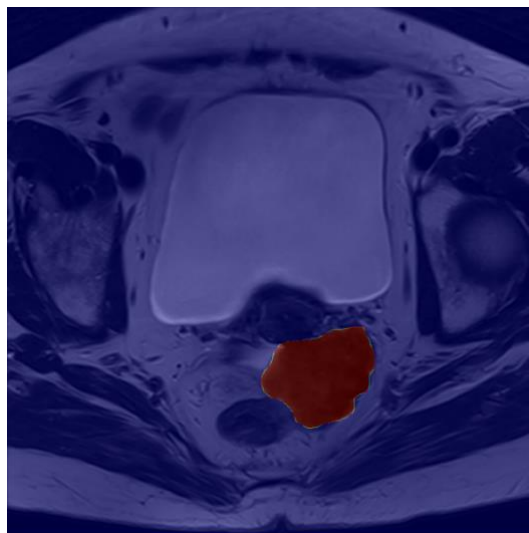


Figure 16: MRI image and tumor mask from Radiomics Frontiers correctly overlaid.

5.4.3. MINT LESION MASKS

Mint Lesion software demonstrated superior performance in segmenting regions of interest (ROIs) when compared to the other two segmentation tools evaluated in this project. This software employed an advanced tool that identified intensity thresholds to delineate the ROIs. Specifically, it segmented the

target region within a predefined circular area, where the object of interest, such as a tumor, was located. The segmentation process is fully automated and applied across all layers of the image, resulting in a highly accurate three-dimensional (3D) segmentation of the entire tumor.

Figure 17 shows the software's interface and the segmented tumor delineated with a very thin red line that is almost not appreciated in the image.

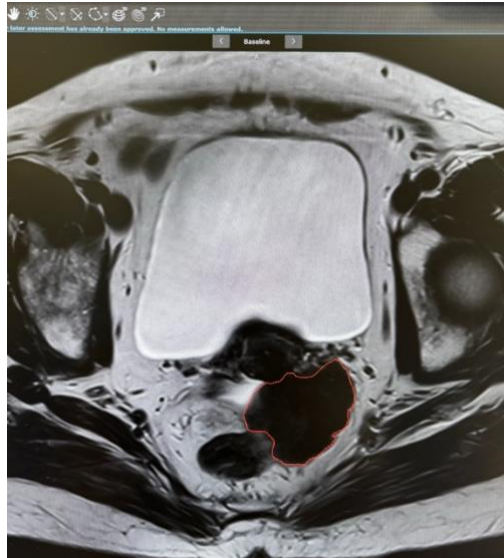


Figure 17: MINT Lesion segmentation interface with delineated tumor.

One of the standout advantages of Mint Lesion software is its integrated radiomics module. This feature facilitates the automatic extraction of radiomic features, including first-order statistics and Gray Level Co-occurrence Matrix (GLCM) features, directly from each segmented region. The software further simplifies the process by providing an option to export the extracted data in CSV (Comma-Separated Value) format. This capability significantly streamlines the workflow for radiomic analysis, making it more efficient and user-friendly. A more comprehensive discussion and analysis of the radiomic features extracted using this software are presented in the subsequent section of this project (Section 5.5: Application of Radiomics).

Despite its many advantages, Mint Lesion software has a notable limitation: it does not support exporting segmentations in the NifTi format, which is widely used in many medical imaging applications. Instead, the segmentations were exported in the DICOM RT-Struct format. Due to this constraint, and the fact that a suitable Python script to convert DICOM RT-Struct to NifTi format was not developed in time, the segmentations could not be automatically converted to this format and, again, 3D Slicer was the solution proposed for this problem.

However, another significant problem emerged during the process. Due to the proprietary nature and closed framework of the software, 3D Slicer was unable to import the masks generated by the MINT Lesion. This incompatibility supposed a big challenge, as it severely limited the flexibility and scope of the radiomics analysis. As a direct consequence of this limitation, only a single radiomics analysis could be performed using the segmentations created in the MINT interface. The inability to integrate masks into 3D Slicer underscores the importance of software compatibility and highlights a critical area for future improvement in radiomics workflows.

5.5. APPLICATION OF RADIOMICS

Once the masks were adequately prepared, the next crucial step in the workflow was the extraction of radiomic features. This step is fundamental in the radiomics pipeline, as it involves quantifying various aspects of the tumor's texture, shape, and intensity from the MRI images. These features are critical for subsequent analyses and model development.

5.5.1. MINT LESION

As previously mentioned, the MINT Lesion software provides an automatic feature extraction tool, which greatly simplifies and speeds up this process. The tool is designed to efficiently extract radiomic features. Specifically, it focuses on first-order statistics and Gray Level Co-occurrence Matrix (GLCM) features.

The automated nature of the MINT Lesion software's extraction tool means that obtaining these features from the MINT segmentations was considerably faster compared to the other segmentation methods. This efficiency in the workflow not only saves time but also reduces the potential for human error during feature extraction.

MINT software offers a variety of preprocessing options that can be applied to MRI images before the extraction of radiomic features. These preprocessing steps are crucial for enhancing the quality of the images and ensuring the accuracy and reliability of the extracted features. Among the available options are the Laplacian of Gaussian (LoG) filter, a threshold filter, and a discretization configuration.

In this project, however, no preprocessing filters were applied to the images. This decision was made to maintain the original image characteristics and avoid introducing any potential biases or alterations that might affect the radiomic analysis. The only preprocessing step applied was the discretization of the image intensities.

Discretization is a process where continuous intensity values of the image are grouped into discrete bins. This step is important for standardizing the intensity values across different images and enhancing the robustness of the radiomic features. In the MINT software, the bin width for discretization was set to 20. This means that the continuous range of intensity values in the MRI images was divided into intervals of 20 units each. As a result, each pixel intensity in the image was assigned to one of these bins based on its value.

5.5.2. 3D SLICER AND RADIOMICS FRONTIERS

To extract features from the 3D Slicer and Radiomics Frontiers masks, a custom Python script was developed utilizing PyRadiomics as the core tool for feature extraction. PyRadiomics is a versatile and powerful library that facilitates the extraction of a wide array of radiomic features from medical imaging data. As detailed in the Conception Engineering section, this tool is capable of calculating up to seven distinct types of radiomic features, including first-order statistics, shape-based features, and texture features such as Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), and others.

The features extracted from the MINT software served as a reference for feature selection due to their suitability and reliability for subsequent analysis. This comparative approach ensured that the feature sets from different preprocessing paths could be consistently evaluated. Consequently, the bin width for discretization in the Python script was set to 20, mirroring the configuration used in the MINT radiomic

extraction. This consistency in the bin width ensured that the discretization process was uniform across different tools, facilitating a fair comparison of the extracted features.

In addition to setting the bin width, the PyRadiomics tool offers numerous settings and preprocessing options to enhance the images before feature extraction. One significant preprocessing step available in PyRadiomics is the normalization of image intensities. Normalization is a critical step that adjusts the intensity values of the images to a common scale, reducing variability due to differences in image acquisition and improving the robustness of the extracted features. Therefore, normalization was applied to the images that did not already have z-score normalization, described in Section 5.3.

In summary, the integration of PyRadiomics for feature extraction from 3D Slicer and Radiomics Frontiers masks, with careful consideration of preprocessing steps such as bin width and intensity normalization, ensured that the radiomic features were robust and comparable across different preprocessing paths.

6. RESULTS AND STATISTICAL ANALYSIS

To facilitate the explanation and understanding of the results a scheme has been created (Figure 18), where a number has been assigned to each of the methods that have been carried out for the extraction of radiomics characteristics.

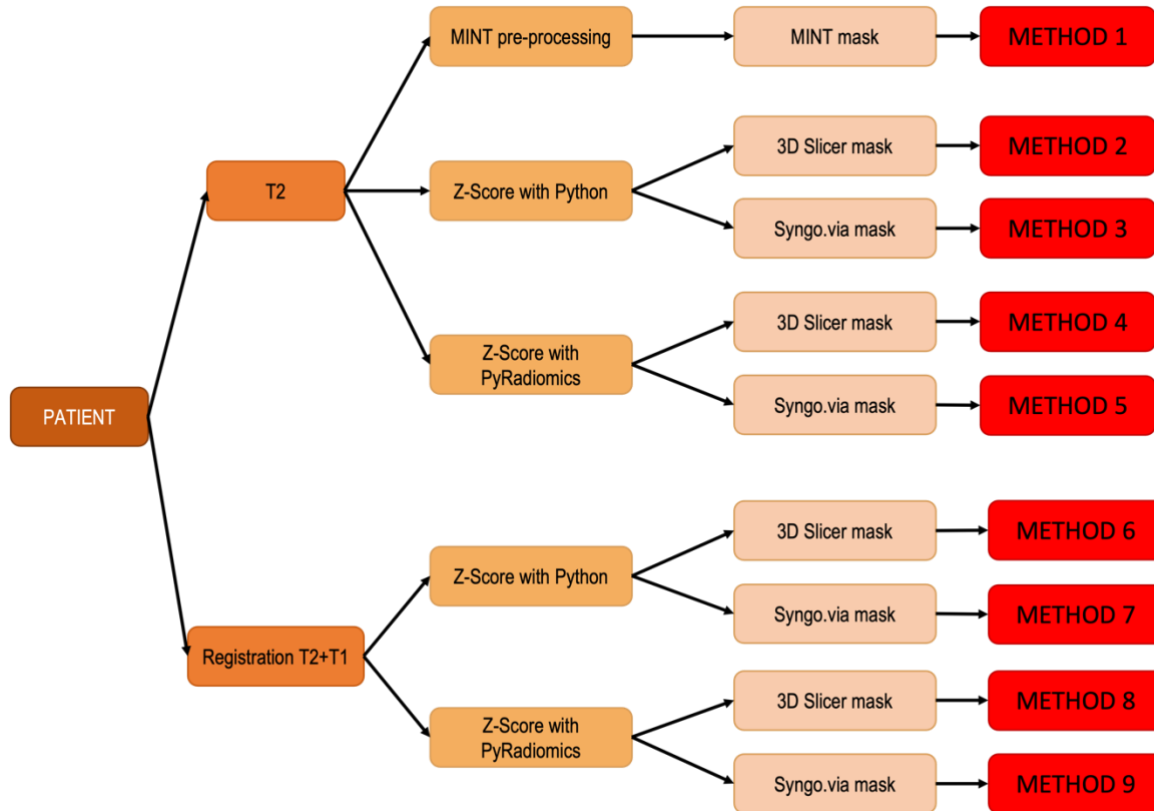


Figure 18: Scheme that indicates the number of each of the methods.

From each of the methods, an excel document with 42 radiomic features in the rows and 12 patients in the columns was created.

If the results obtained from each excel were compared at a glance, it could be seen that there were 5 methods that did not have the normalized intensity values. These were methods 1, 4, 5, 8 and 9.

Method 1, the one created with the MINT software, did not offer any intensities normalization tool and considering that the patients' images came from different MRI acquisition machines, we could not trust the results we got the method.

On the other hand, what methods 4, 5, 8 and 9 had in common was that the intensity values had been normalized with the PyRadiomics package itself. This is most likely the fault of the author of the Python code, as the normalization tool must not have been implemented properly. Since there was not enough time to try to fix the problem, it was decided to remove the affected methods and continue the analysis with the remaining 4.

One clear thing that had to be done was to reduce the dimensionality of the features, as there were too many compared to the number of patients. It was likely that many of the characteristics were correlated with each other regardless of the patients' diagnosis. For this reason, a python code was created that

showed the correlation matrix of the features. *Figure 19* shows the correlation matrix of the features from Method 2. The other three are shown in the Annex section.

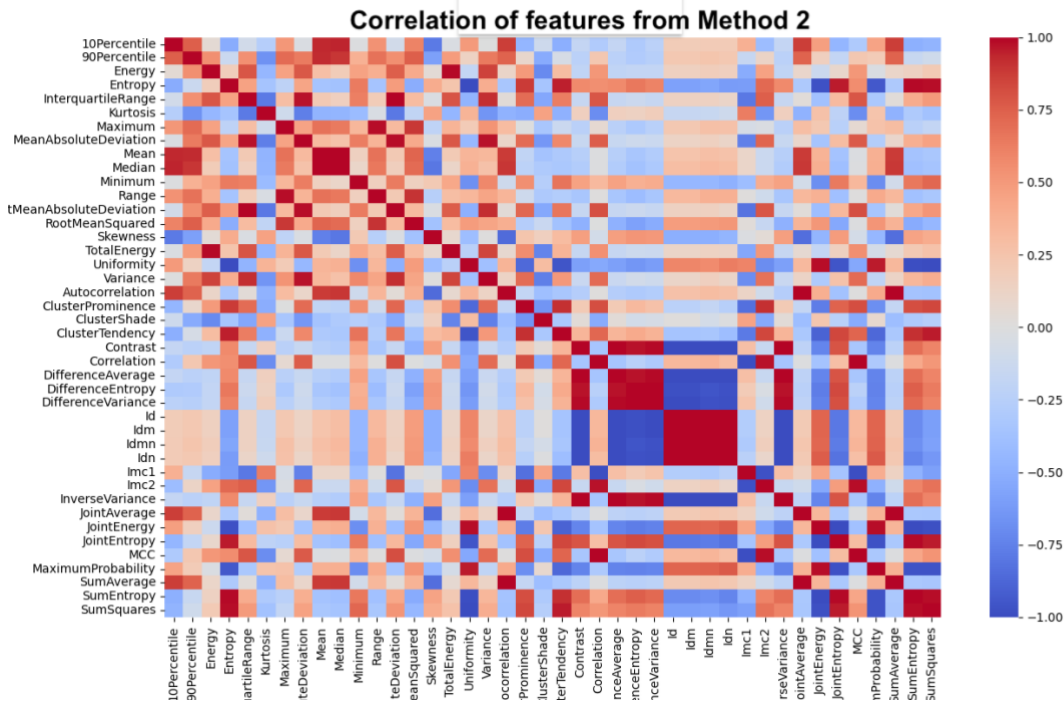


Figure 19: Correlation matrix of the features from Method 2.

As seen in *Figure 19*, lots of features were correlated between them so, another Python function was created to eliminate one of the features when two of them were correlated more than 0.8 in absolute value. Therefore, another correlation matrix was created with the non-eliminated features to see which one were not correlated in each of the methods. In *Figure 20*, there's represented the filtrated correlation matrix of Method 2. The other three are shown in the Annex section.

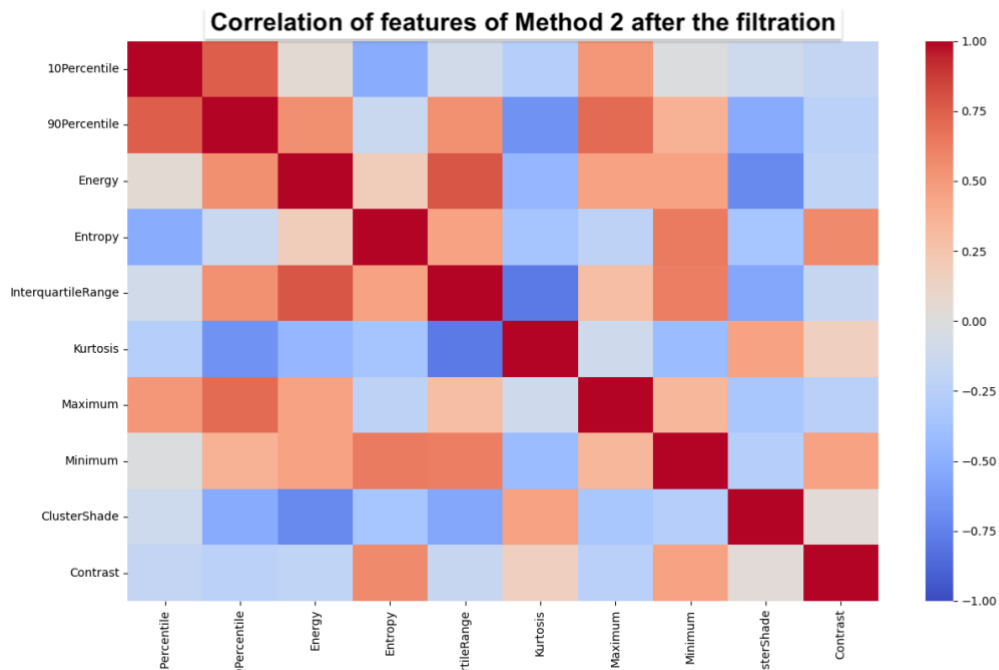


Figure 20: Correlation matrix of the features from Method 2 after the filtration.

Of course, this procedure was also done in the excels corresponding to the methods 3, 6 and 7.

After saving the correlation matrices of the 3 other methods, a table was created to be able to compare between methods, which features had remained in each case. *Table 10* shows this comparison.

Table 10: Comparison of conserved features between methods.

FEATURES	METHOD 2	METHOD 3	METHOD 6	METHOD 7
10th	X	X	X	X
90th	X	X	X	X
Energy	X	X	X	X
Entropy	X	X		
Kurtosis	X	X	X	X
Min	X	X	X	X
Max	X	X	X	
Skewness		X	X	
ClusterShade	X	X	X	X
Correlation			X	
Contrast	X	X		
InterquartileRange	X			
RootMeanSquare			X	

As seen in Table 10, there are many features that are conserved in the four methods: 10th percentile, 90th percentile, Energy, Kurtosis, Minimum intensity, Maximum intensity, and Cluster Shade. In the Annex section there's a table which shows how the features were calculated.

What was done next was to create a new excel for each method with the 13 calculated characteristics, so that they all had the same number of rows. The objective of this new step was to check how different the values of the features of each method were. Therefore, the average and standard deviation of the values of each characteristic were calculated, for example, the average of the Energy values of the 12 patients of each method was calculated, as well as the standard deviation, and so with the remaining characteristics.

These calculations made it possible to make a bar graph for each characteristic and to compare the average and the standard deviation of the 4 methods.

Figure 21 and 22 show the mean and the standard deviation values of the 4 methods for the Skewness feature.

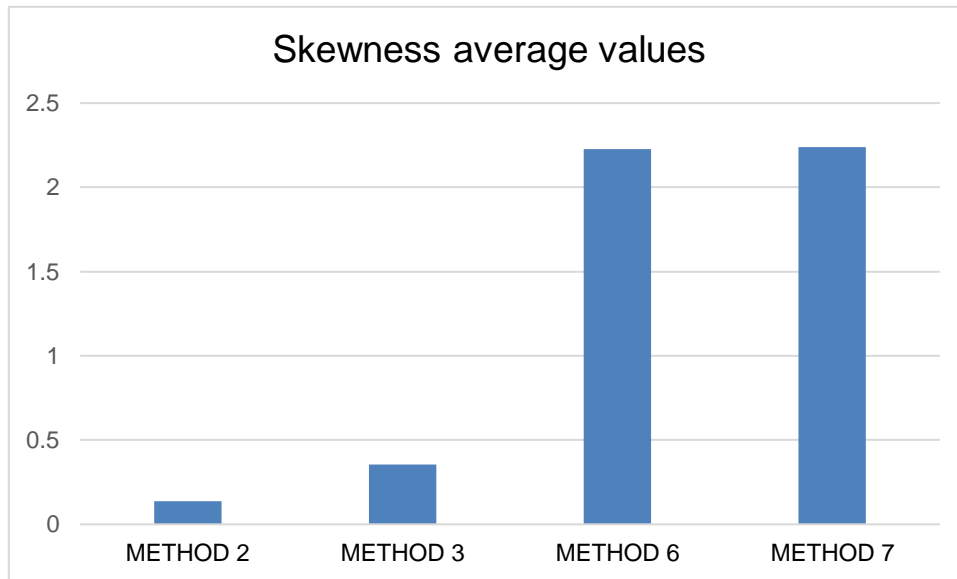


Figure 21: Skewness average values

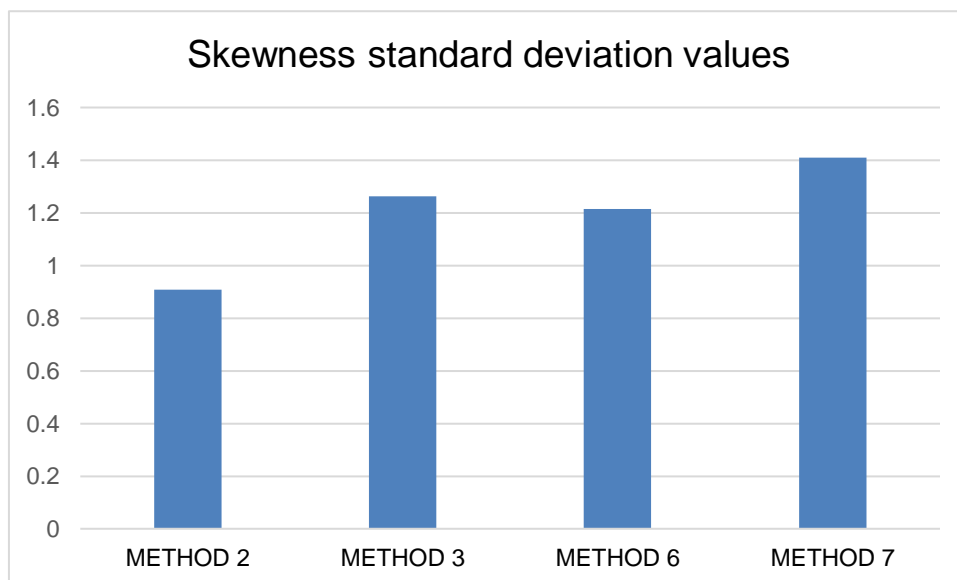


Figure 22: Skewness standard deviation values.

From Figure 21, and the other mean feature's chart it was deduced that maybe Method 2 and Method 3 were correlated between them, as well as Method 6 and Method 7, because they had similar average values in most cases.

To verify this fact, a function was created with python which made a pairwise correlation of the 4 methods. In the context of this project, this pairwise correlation involved comparing each pair of features across the different datasets (Excel files) to determine how similar or dissimilar they are. By calculating the correlation coefficients for each pair, it can be found which features or metrics are consistently related across different methods and which ones vary significantly. This helps in understanding the consistency and reliability of the segmentation and analysis tools the project is comparing. This technique is called Pearson Correlation, and *Figure 23* shows a matrix of correlations between the methods.

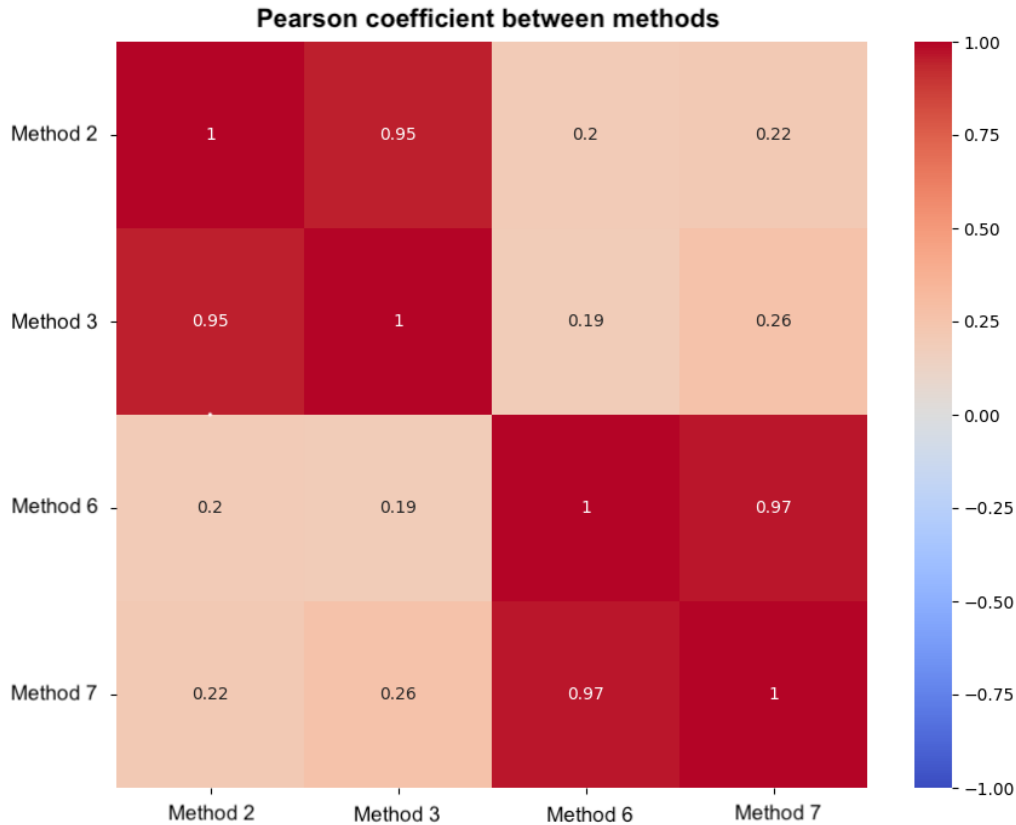


Figure 23: Pearson correlation matrix between methods.

This matrix shows that, indeed, Method 2 and 3, give very similar radiomic values, as well as Method 6 and 7.

Finally, a Scatter Plot Matrix was created using Python to see the relationships between pairs of variables in a dataset. Taking into account the patients diagnose, this tool would help seeing if the relationship of all the pair's combination of a dataset are able to make visible clusters.

Here's a detailed explanation of what a scatterplot matrix does and how it is used:

Each cell in the grid represents a scatterplot of two variables plotted against each other. If you have N variables, the matrix will be $N \times N$ in size.

- **Diagonal Elements:** The diagonal cells typically show the distribution of each variable, often using histograms or density plots. This allows you to quickly see the distribution of each individual variable.
- **Off-Diagonal Elements:** The off-diagonal cells show scatterplots of each pair of variables. This allows you to visually inspect the relationships between every possible pair of variables in the dataset.

The Scatter Plot was tested on Method 2 and Method 6 but there were no interpretable results.

7. EXECUTION SCHEDULE

To minimize the preparation time of the project and not encounter many obstacles during the process, it was necessary to meticulously design and adhere to a comprehensive execution schedule. By strategically allocating resources and delineating clear objectives, this approach ensures an efficient progress and facilitates the achievement of project milestones, ultimately contributing to enhanced diagnostic capabilities in oncology.

Thus, this section presents the Work Breakdown Structure (WBS) and the phases of the project, with the PERT-CPM and the Gantt diagram.

7.1. WORK BREAKDOWN STRUCTURE (WBS)

The Work Breakdown Structure (WBS) is a comprehensive and systematic representation of all the tasks required to develop a project, presented in a hierarchical decomposition. This structured approach facilitates a clear understanding of the project's scope by breaking it down into manageable sections.

At its core, the WBS consists of multiple levels, with the primary two levels being particularly crucial. The first level provides an overarching view of the project, outlining the major components and deliverables. The second level dives deeper into these components, detailing the specific tasks and activities necessary to achieve each principal point.

By organizing tasks in this manner, the WBS not only helps in assigning responsibilities but also in tracking progress and ensuring that all aspects of the project are covered comprehensively.

Figure 24 shows the WBS of this project.

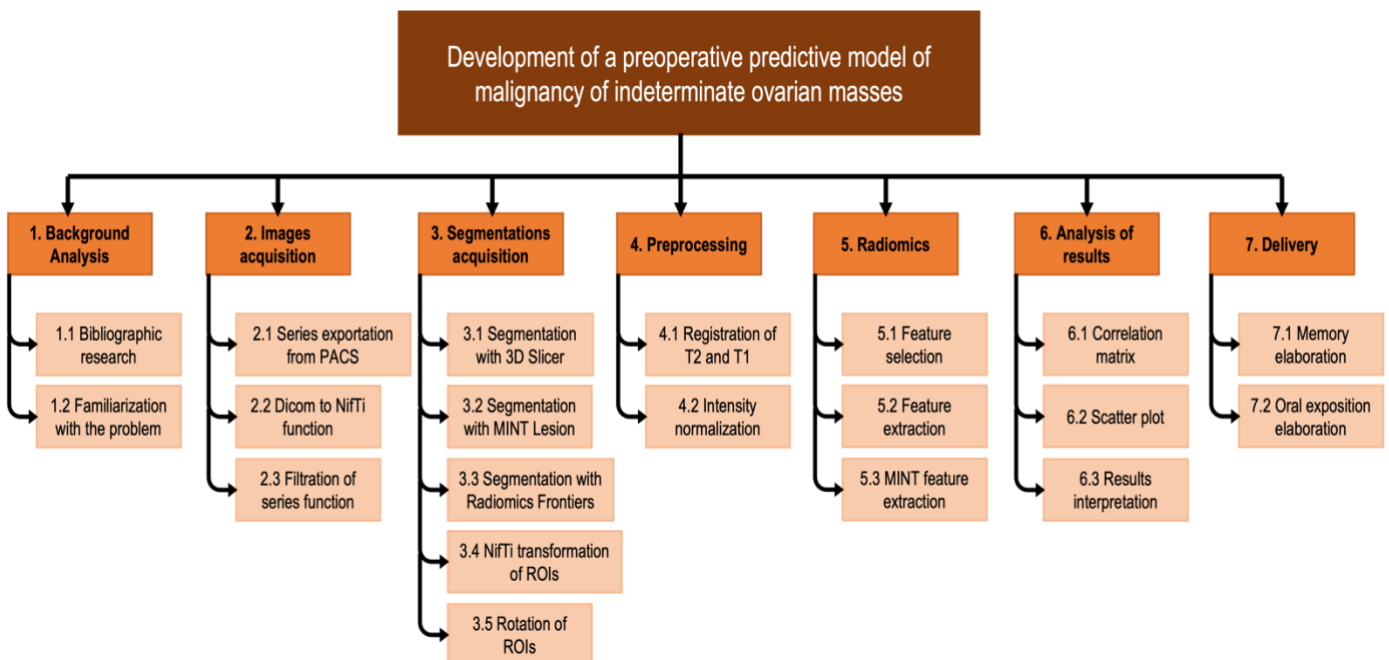


Figure 24: WBS of the project.

7.2. PERT DIAGRAM

As it's seen in Figure x, the project is divided into 7 packages of work. Each of them has several tasks to be accomplished. Therefore, to coordinate them and reach the deadline of the project, the PERT- CPM diagram has been developed. This diagram visually represents the tasks, their timing and their relationship. The chronological dependence of each task and its duration has been defined in *Table 11*.

Table 11: Tasks and timings of the project.

WBS ID	PERT ID	Previous activity	Following activity	Time (days)
1.1 Bibliographic research	A	-	B, M	10
1.2 Familiarization with the problem	B	A	G, H	4
2.1 Series exportation from PACS	C	-	D, F	1
2.2 DICOM to NifTi function development	D	C	E	15
2.3 Filtration of series function development	E	D	K	3
3.1 Segmentation of ROIs with 3D Slicer	F	C	I	5
3.2 Segmentation of ROIs with MINT Lesion	G	B	O	5
3.3 Segmentation of ROIs with Radiomics Frontiers (Syngo.via)	H	B	I	5
3.4 NifTi transformation of the ROIs	I	F, H	J	3
3.5 Rotation of ROIs function development	J	I	N	4
4.1 Registration of T2 + T1 function development	K	E	L	15
4.2 Intensity normalization function development	L	K	N	8
5.1 Feature selection	M	A	N	4
5.2 Feature extraction function development	N	L, J, M, O	P	10
5.3 Feature extraction from MINT Lesion	O	G	N	4
6.1 Correlation matrix's function development	P	N	Q	5
6.2 Scatter plots' function development	Q	P	R	5
6.3 Interpretation of the results	R	Q	S	10
7.1 Elaboration of the memory of the project	S	R	T	15
7.2 Elaboration of the oral exposition	T	S	-	5

From the precedencies in *Table 11*, the chronological order of the tasks has been set, and the PERT-CPM has also been drawn in *Figure 25*. By calculating the early time and last time for each task, it can be determined the critical path of the project. This one represents the sequence of tasks that must be completed for the project to be finished on time. It ensures that a task cannot start until its predecessor has been completed but allows for simultaneous execution of multiple tasks.

To understand better the concept, a proper definition of each term is provided:

- **Early time:** minimum necessary time to perform a task.
- **Last time:** latest time in which a task can be developed. It is the maximum time that an activity can reach.
- **Critical path:** group of activities that, in case of there being a delay, would modify the final timing of the project.

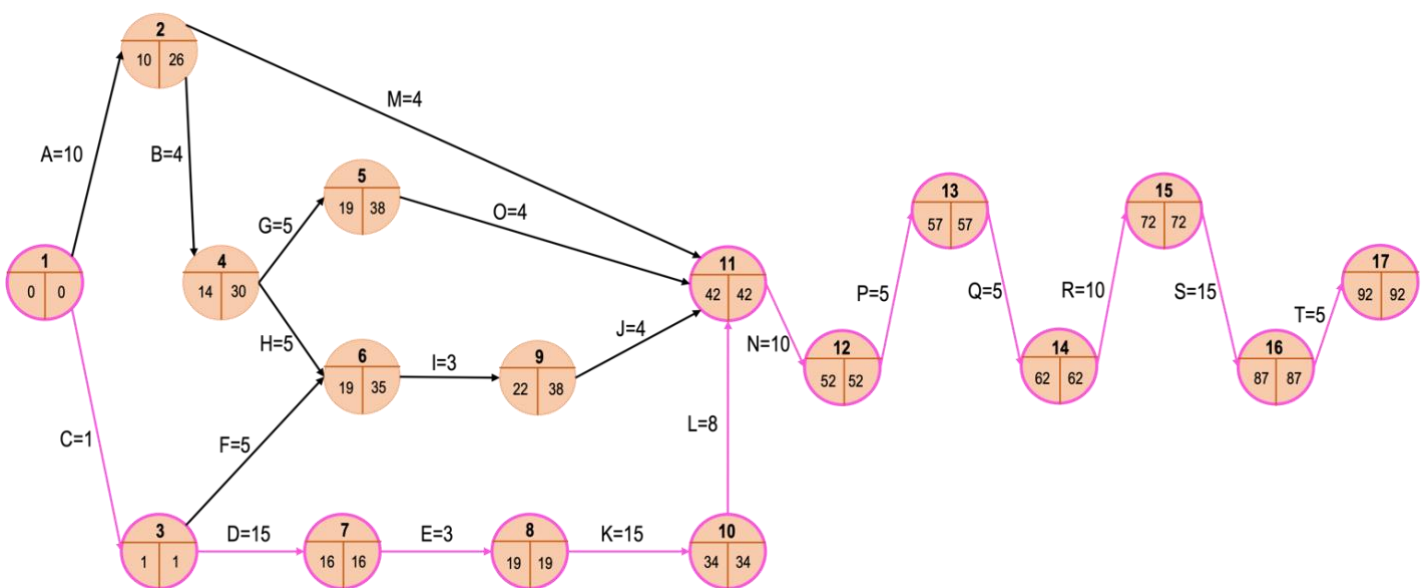


Figure 25: PERT diagram of the project.

As deduced from *Figure 25*, the critical path of this project consists of tasks C, D, E, K, L, N, P, Q, R, S, and T. This sequence is obviously the critical one as it includes the final writing task and the preparation of the presentation. It is clear that the project cannot be completed until the writing portion is finished, which needs a considerable amount of time.

Additionally, the development of the DICOM to NifTi function using Python is another lengthy task. This involves complex coding and validation to ensure the conversion is accurate and reliable. Similarly, the creation of a registration code for T2 and T1 series is a time-consuming process. These tasks, along with the writing task, each require approximately 15 days to complete. The registration code task involves aligning images from different modalities (T2+T1), which is technically challenging and necessitates meticulous attention to detail.

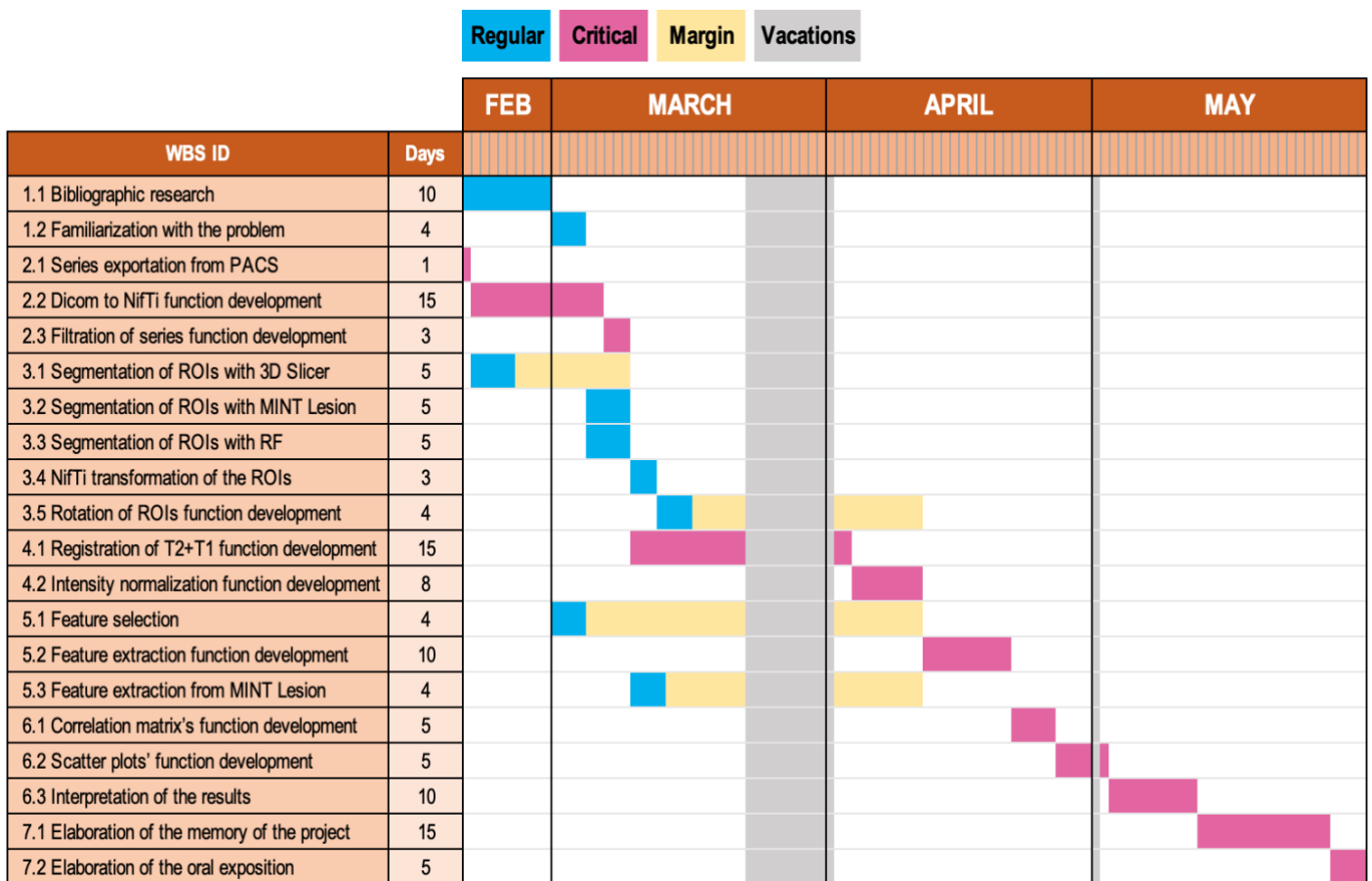
These critical tasks are essential for the project's success and are essential in determining the project's overall timeline. Each task's duration and complexity highlight the importance of careful planning to ensure that deadlines are met and the project is completed in time.

7.3. GANTT DIAGRAM

A gantt diagram is a work planning tool, since it presents all the activities that must be carried out and when they must be carried out. It also allows to have an idea of how the project is progressing and if it is necessary to reschedule the planned actions in order to adapt the project to the new environment or needs.

All the tasks that define the project are illustrated in the GANTT diagram seen in *Table 12*, to provide a clear visualization of the project's timeline. As told, this diagram allows us to see the precise start and end dates for each activity, understand the duration of each task, and identify any overlapping activities. Most importantly, it helps in establishing the overall start and finish dates for the project. This visual representation ensures we have a comprehensive view of the project's progression over time.

Table 12: GANTT diagram of the project.



The grey lines during the final days of March indicate the Easter holidays, ranging from March 23rd to April 1st, while the grey line on May 1st marks Worker's Day, which is also a public holiday. In *Table 12*, the pink lines represent the critical path of the project. This critical path started on February 19th and concluded on May 31st. Given that the final submission of the project report was scheduled for June 5th, with the presentation following on June 10th, the critical path had a maximum flexibility of approximately 10 days. This margin underscores the importance of adhering to the timeline, as any delays in the critical path tasks would significantly impact the overall project completion, leaving little room for unforeseen setbacks.

The careful planning and identification of these key dates ensure that all essential activities are completed on schedule, thereby mitigating the risk of overruns, and ensuring a timely project delivery.

8. TECHNICAL VIABILITY

This section shows the SWOT (Strengths, Weaknesses, Opportunities and Threats) analysis, which provides a comprehensive overview of the project's internal strengths and weaknesses, as well as the external opportunities and threats it faces. By addressing these factors, the project aims to make a significant contribution to the early diagnosis and treatment of ovarian cancer.

The following table shows the schematic SWOT analysis of the project.

Table 13: SWOT analysis of the project.

	STRENGTHS	WEAKNESSES
INTERNAL	<ul style="list-style-type: none"> • High motivation of the author. • Previous development of code/programming knowledge. • Periodic meetings to share the progress and discuss possible improvements. 	<ul style="list-style-type: none"> • Small list of patients. • Lack of reality since not all the radiomics features have been taken into account. • Lack of experience in image processing.
EXTERNAL	OPPORTUNITIES	THREATS
	<ul style="list-style-type: none"> • Increasing interest and investment in radiomics research and applications. • Potential for its integration into clinical workflows. 	<ul style="list-style-type: none"> • Limited time. • Unexpected technical difficulties.

8.1. STRENGTHS

One of the key strengths of this project is the high motivation of the author. This enthusiasm drives the project forward and ensures a committed approach to problem-solving and achieving the project goals.

Additionally, the previous development of code and programming knowledge is a significant advantage. This expertise allows for the efficient handling of technical challenges and the development of sophisticated functions required for radiomics analysis.

Moreover, periodic meetings to share progress and discuss possible improvements provide a structured framework for the project. These meetings facilitate continuous feedback and collaboration, ensuring that the project remains on track and benefits from diverse perspectives.

8.2. WEAKNESSES

Despite these strengths, there are some notable weaknesses. The small list of patients is a limitation, as it can affect the generalizability of the results. A larger dataset would provide more robust and reliable findings.

Furthermore, there is a lack of reality since not all the radiomics features will be taken into account. This omission could result in an incomplete analysis, potentially overlooking important data insights.

Additionally, the lack of experience in image processing is also going to suppose a challenge. Although the author has programming skills, specific expertise in image processing is crucial for accurately analyzing MRI data and extracting meaningful features.

8.3. OPPORTUNITIES

However, there are significant opportunities that can be leveraged. There is an increasing interest and investment in radiomics research and applications. This growing field offers numerous possibilities for innovation and collaboration, enhancing the project's potential impact.

Moreover, the potential for integrating radiomics into clinical workflows presents a substantial opportunity. Successful implementation of this project could lead to improved diagnostic processes and personalized treatment plans, directly benefiting patient care.

8.4. THREATS

Despite the opportunities, there are threats that must be considered. Limited time is a major constraint, as the project must be completed within a set timeframe. This pressure can affect the thoroughness of the research and the quality of the outcomes.

Additionally, unexpected technical difficulties can arise, potentially disrupting progress. These challenges can be unpredictable and may require additional resources and time to resolve, impacting the overall success of the project.

9. ECONOMIC VIABILITY

In this section, information is provided on the financial requirements necessary to be able to adequately achieve the project's objective.

The following table includes various cost elements associated with the development and implementation of the radiomics pipeline for discriminating malignant and benign ovarian masses from MRI images.

Table 14: Total cost of the project.

ITEM	DESCRIPTION	COST PER UNIT	TOTAL COST
IMAGE ACQUISITION			
MRI	Num of patients = 12	200€ [28]	2.400€
SEGMENTATION SOFTWARES			
MINT LESION	Closed software	1 license = 13.000€ Annual revision = 7.000€	20.000€
SYNGO.VIA	Closed software	*	0€
3D SLICER	Open software	Free	0€
DATA PROCESSING			
PYTHON	Python 3.11	Free	0€
ANTs MODULE	Registration module	Free	0€
PYRADIOMICS MODULE	Radiomic module	Free	0€
HUMAN RESOURCES			
BME STUDENT	300 hours	15€/hour	4.500€
DIRECTORS	40 hours/director 2 directors	20€/hour	1.600€
RADIOLOGIST	12 hours	20€/hour	240€
TOTAL			28.740€

*Free if the hospital has bought a machine from Siemens

Therefore, the project's total cost is 28.400€. It is important to note that a huge portion of the budget was allocated to the segmentation software, since one of them is closed and has a considerable cost. Additionally, the human resources budget was also expensive.

10. REGULATIONS AND LEGAL ASPECTS

Ensuring compliance with regulatory and legal requirements is crucial for the successful development and implementation of the radiomics pipeline for discriminating malignant and benign ovarian masses from MRI images. Given that this project is located at the Hospital Clinic in Barcelona, it must adhere to regulations and legislation imposed by both the European Parliament and the Spanish Government.

As this project serves as a final degree project for the Universitat de Barcelona, it must follow the regulations declared in the Royal Decree 1393/2007 [29], which governs the structure and organization of official university education in Spain. This decree outlines the standards and procedures for academic projects, ensuring they meet the educational and scientific requirements set by the institution.

In addition to academic regulations, the project must comply with Spanish laws related to biomedical research. The law 14/2007 [30] regulates biomedical investigation in the country, establishing the principles that researchers must follow and the rights of patients, including non-discrimination, informed consent, and data confidentiality. This law ensures that all biomedical research conducted within Spain adheres to ethical standards that protect patient rights and promote scientific integrity.

The project involves using pelvic MRI scans from patients to develop and validate the radiomics pipeline. That's why it follows GDPR, an European regulation 2016/679 on the protection of natural persons regarding the processing of personal data [31] and a Spanish data protection law (Ley Orgánica 3/2018) also about the protection of patients' data [32], which dictates that all the information should be anonymised and nonpersonal information should be visualized by the researchers like locations, ethnicity, among others. Measures such as encryption, secure data storage, and access controls were implemented to protect patient data.

Therefore, data anonymization techniques were employed to safeguard patient identities. Internally, the Hospital Clínic has strict data protection protocols. The anonymization of patients is handled by creating an independent file that links each patient's anonymous ID to their respective National Health Identifier (NHI). This file is kept secure within the hospital's network and is not accessible outside, adhering to the European Union Regulation 2016/679 [31], mentioned before in the text.

By rigorously following these regulations and implementing robust data protection measures, the project not only meets legal requirements but also ensures the ethical treatment of patient data. This compliance is fundamental to maintaining the trust of patients and the integrity of the research. The commitment to these standards underscores the project's dedication to advancing medical research while safeguarding the rights and privacy of individuals.

11. CONCLUSIONS AND FUTURE LINES

In this study, a comprehensive analysis was conducted to evaluate the effectiveness of different segmentation and radiomic analysis methods for discriminating between benign and malignant ovarian masses from MRI images. Several methods were tested, and various challenges were encountered, particularly with intensity normalization and the limited sample size. The following conclusions summarize the key findings and insights from the project.

Firstly, it is likely that if the intensity normalization had been correctly implemented, the results obtained from the methods that were eliminated would have been similar to those achieved. Specifically, methods 4 and 5, as well as methods 8 and 9, would probably have shown comparable outcomes.

Secondly, the choice of segmentation software appears to be somewhat indifferent, as the radiomic results were very similar across different tools. For instance, methods 6 and 7 produced almost identical results, indicating that the segmentation process is not significantly affected by the software used.

Furthermore, the study had too few patients to definitively determine whether the features were effective in classifying tumors into benign, malignant, and borderline categories. This limitation hindered the ability to make robust conclusions about the classification capabilities of the extracted radiomic features.

Additionally, it is probable that many radiomic features that were discarded during the conceptual engineering phase could have been more effective in classifying the types of tumors. This suggests that a broader inclusion of radiomic features might improve the accuracy and reliability of tumor classification in future studies.

In conclusion, addressing the issues of intensity normalization, increasing the sample size, and re-evaluating the discarded radiomic features could potentially enhance the accuracy and reliability of the radiomics pipeline for discriminating between benign and malignant ovarian masses from MRI images.

The radiologist who intends to further develop this project is already segmenting additional tumors to create a larger database. This expansion is crucial, as it would allow the application of more advanced artificial intelligence algorithms capable of distinguishing between the three types of tumors. With a more extensive patient dataset, the statistical analysis presented in this project's report would become significantly more reliable and robust. The initial sample of 12 patients is too small to draw definitive conclusions, but a larger dataset would provide a more accurate representation and validation of the findings.

Additionally, there is an intention to write and publish an article about the project. The goal is to standardize the methods developed in this project for tumor detection and classification, ensuring that they can be utilized in hospitals. Such standardization would pave the way for wider adoption of these techniques, potentially improving diagnostic accuracy and patient outcomes in clinical settings. Moreover, a published article would contribute to the scientific community by providing detailed insights and methodologies that other researchers and clinicians can build upon.

Furthermore, with a larger dataset, there would be opportunities to explore and validate additional radiomic features that were initially discarded during the conception phase of the project. These features might enhance the classification accuracy and provide deeper insights into the tumor characteristics. This



continuous improvement and validation process is essential for advancing the field of radiomics and integrating these techniques into routine clinical practice.

In summary, the next steps involve not only expanding the database and refining the methods but also disseminating the findings through publication and advocating for the clinical adoption of these advanced radiomic techniques. This comprehensive approach will ensure that the project's outcomes have a lasting and meaningful impact on the diagnosis and treatment of ovarian tumors.

12. BIBLIOGRAPHY

- [1] World Health Organization, "Cancer." Accessed: May 27, 2024. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/cancer>
- [2] OCRA, "Ovarian Cancer by the numbers." Accessed: May 27, 2024. [Online]. Available: <https://ocrahope.org/get-the-facts/statistics/>
- [3] R. T. Greenlee, M. B. Hill-Harmon, T. Murray, and M. Thun, "Cancer statistics, 2001.," *CA Cancer J Clin*, vol. 51, no. 1, pp. 15–36, 2001, doi: 10.3322/canjclin.51.1.15.
- [4] B. A. Goff, L. S. Mandel, C. H. Melancon, and H. G. Muntz, "Frequency of symptoms of ovarian cancer in women presenting to primary care clinics.," *JAMA*, vol. 291, no. 22, pp. 2705–12, Jun. 2004, doi: 10.1001/jama.291.22.2705.
- [5] M. J. Duffy *et al.*, "CA125 in ovarian cancer: European Group on Tumor Markers guidelines for clinical use.," *Int J Gynecol Cancer*, vol. 15, no. 5, pp. 679–91, 2005, doi: 10.1111/j.1525-1438.2005.00130.x.
- [6] G. Baggio, A. Corsini, A. Floreani, S. Giannini, and V. Zagonel, "Gender medicine: a task for the third millennium.," *Clin Chem Lab Med*, vol. 51, no. 4, pp. 713–27, Apr. 2013, doi: 10.1515/cclm-2012-0849.
- [7] T. Arora, S. Mullangi, and M. R. Lekkala, *Ovarian Cancer*. 2024.
- [8] I. Thomassin-Naggara *et al.*, "Ovarian-Adnexal Reporting Data System Magnetic Resonance Imaging (O-RADS MRI) Score for Risk Stratification of Sonographically Indeterminate Adnexal Masses.," *JAMA Netw Open*, vol. 3, no. 1, p. e1919896, Jan. 2020, doi: 10.1001/jamanetworkopen.2019.19896.
- [9] R. Forstner, "Early detection of ovarian cancer.," *Eur Radiol*, vol. 30, no. 10, pp. 5370–5373, Oct. 2020, doi: 10.1007/s00330-020-06937-z.
- [10] S. Nougaret *et al.*, "Ovarian cancer: An update on imaging in the era of radiomics.," *Diagn Interv Imaging*, vol. 100, no. 10, pp. 647–655, Oct. 2019, doi: 10.1016/j.diii.2018.11.007.
- [11] P. A. Rinck, "Magnetic Resonance in Medicine. A critical introduction. e-Textbook (14th ed.).," TRTF – The Round Table Foundation: TwinTree Media. Accessed: May 18, 2024. [Online]. Available: <https://magnetic-resonance.org/>
- [12] L. M. De León-Rodríguez, A. F. Martins, M. C. Pinho, N. M. Rofsky, and A. D. Sherry, "Basic MR relaxation mechanisms and contrast agent design.," *J Magn Reson Imaging*, vol. 42, no. 3, pp. 545–65, Sep. 2015, doi: 10.1002/jmri.24787.
- [13] P. Lambin *et al.*, "Radiomics: extracting more information from medical images using advanced feature analysis.," *Eur J Cancer*, vol. 48, no. 4, pp. 441–6, Mar. 2012, doi: 10.1016/j.ejca.2011.11.036.
- [14] R. J. Gillies, P. E. Kinahan, and H. Hricak, "Radiomics: Images Are More than Pictures, They Are Data.," *Radiology*, vol. 278, no. 2, pp. 563–77, Feb. 2016, doi: 10.1148/radiol.2015151169.
- [15] P. Lambin *et al.*, "Radiomics: the bridge between medical imaging and personalized medicine.," *Nat Rev Clin Oncol*, vol. 14, no. 12, pp. 749–762, Dec. 2017, doi: 10.1038/nrclinonc.2017.141.

- [16] P. Adusumilli, N. Ravikumar, G. Hall, S. Swift, N. Orsi, and A. Scarsbrook, "Radiomics in the evaluation of ovarian masses - a systematic review.," *Insights Imaging*, vol. 14, no. 1, p. 165, Oct. 2023, doi: 10.1186/s13244-023-01500-y.
- [17] H. Li *et al.*, "Noninvasive prediction of residual disease for advanced high-grade serous ovarian carcinoma by MRI-based radiomic-clinical nomogram.," *Eur Radiol*, vol. 31, no. 10, pp. 7855–7864, Oct. 2021, doi: 10.1007/s00330-021-07902-0.
- [18] Yong'ai Li MD, Junming Jian MS, P. Perry J. Pickhardt MD, P. Fenghua Ma MD, and Wei Xia MS, "MRI-Based Machine Learning for Differentiating Borderline From Malignant Epithelial Ovarian Tumors: A Multicenter Study," *Journal of Magnetic Resonance Imaging*, vol. 52, no. 3, Feb. 2020.
- [19] X.-L. Song, J.-L. Ren, D. Zhao, L. Wang, H. Ren, and J. Niu, "Radiomics derived from dynamic contrast-enhanced MRI pharmacokinetic protocol features: the value of precision diagnosis ovarian neoplasms.," *Eur Radiol*, vol. 31, no. 1, pp. 368–378, Jan. 2021, doi: 10.1007/s00330-020-07112-0.
- [20] R.-A. Lupean *et al.*, "Radiomic Analysis of MRI Images is Instrumental to the Stratification of Ovarian Cysts.," *J Pers Med*, vol. 10, no. 3, Sep. 2020, doi: 10.3390/jpm10030127.
- [21] X. Liu *et al.*, "Two-dimensional and three-dimensional T2 weighted imaging-based radiomic signatures for the preoperative discrimination of ovarian borderline tumors and malignant tumors.," *J Ovarian Res*, vol. 15, no. 1, p. 22, Feb. 2022, doi: 10.1186/s13048-022-00943-z.
- [22] H. Zhang *et al.*, "Magnetic resonance imaging radiomics in categorizing ovarian masses and predicting clinical outcome: a preliminary study.," *Eur Radiol*, vol. 29, no. 7, pp. 3358–3371, Jul. 2019, doi: 10.1007/s00330-019-06124-9.
- [23] H. Ding *et al.*, "Radiomics in Oncology: A 10-Year Bibliometric Analysis," *Front Oncol*, vol. 11, Sep. 2021, doi: 10.3389/fonc.2021.689802.
- [24] Grand View Research, "Cancer Diagnostics Industry Data Book - Breast Cancer Diagnostic, Cervical Cancer Diagnostic, Lung Cancer Diagnostic, Liver Cancer Diagnostic, Pancreatic Cancer Diagnostic and Ovarian Cancer Diagnostic Market Size, Share, Trends Analysis, And Segment Forecasts, 2023 - 2030," Jul. 2023. Accessed: Jun. 02, 2024. [Online]. Available: <https://www.grandviewresearch.com/sector-report/cancer-diagnostics-industry-data-book>
- [25] Grand View Research, "Ovarian Cancer Diagnostics Market Size, Share, & Trends Analysis Report By Diagnosis Type (Imaging, Blood Test), By Cancer Type (Epithelial Tumor, Stromal Cell Tumor), By End Use, By Region, And Segment Forecasts, 2019 - 2026," Aug. 2023. Accessed: Jun. 02, 2024. [Online]. Available: <https://www.grandviewresearch.com/industry-analysis/ovarian-cancer-diagnostics-market>
- [26] Rordenlab, "DCM2NIIIX: DICOM to NIFTI converter," GitHub. Accessed: May 28, 2024. [Online]. Available: <https://github.com/rordenlab/dcm2niix>
- [27] ANTsX, "ANTs and antsRegistration," GitHub. Accessed: Jun. 03, 2024. [Online]. Available: <https://github.com/ANTsX/ANTs/wiki/ANTs-and-antsRegistration>

- [28] Blog SaludOnNet., “Cuál es el precio de una resonancia magnética.” Blog SaludOnNet. Accessed: Jun. 02, 2024. [Online]. Available: <https://www.saludonnet.com/blog/cual-es-el-precio-de-una-resonancia-magnetica/>
- [29] Ministerio de Educación y Ciencia, “Real Decreto 1393/2007, por el que se establece la ordenación de las enseñanzas universitarias oficiales.” *BOE*, Oct. 2007.
- [30] Jefatura del Estado, “Ley 14/2007 de Investigación Biomédica,” *BOE*, Jul. 2007.
- [31] European parliament, “REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation),” 2016.
- [32] Jefatura del Estado, “Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.” *BOE*, 2018.
- [33] Mayo Clinic, “Ovarian cancer - Symptoms and causes .” Accessed: May 01, 2024. [Online]. Available: (2023, 9 mayo). Mayo Clinic. <https://www.mayoclinic.org/diseases-conditions/ovarian-cancer/symptoms-causes/syc-20375941>
- [34] NCI, “Transvaginal Ultrasound, NCI Dictionary of Cancer Terms,” National Cancer Institute. Accessed: Jun. 04, 2024. [Online]. Available: <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/transvaginal-ultrasound>

Annex

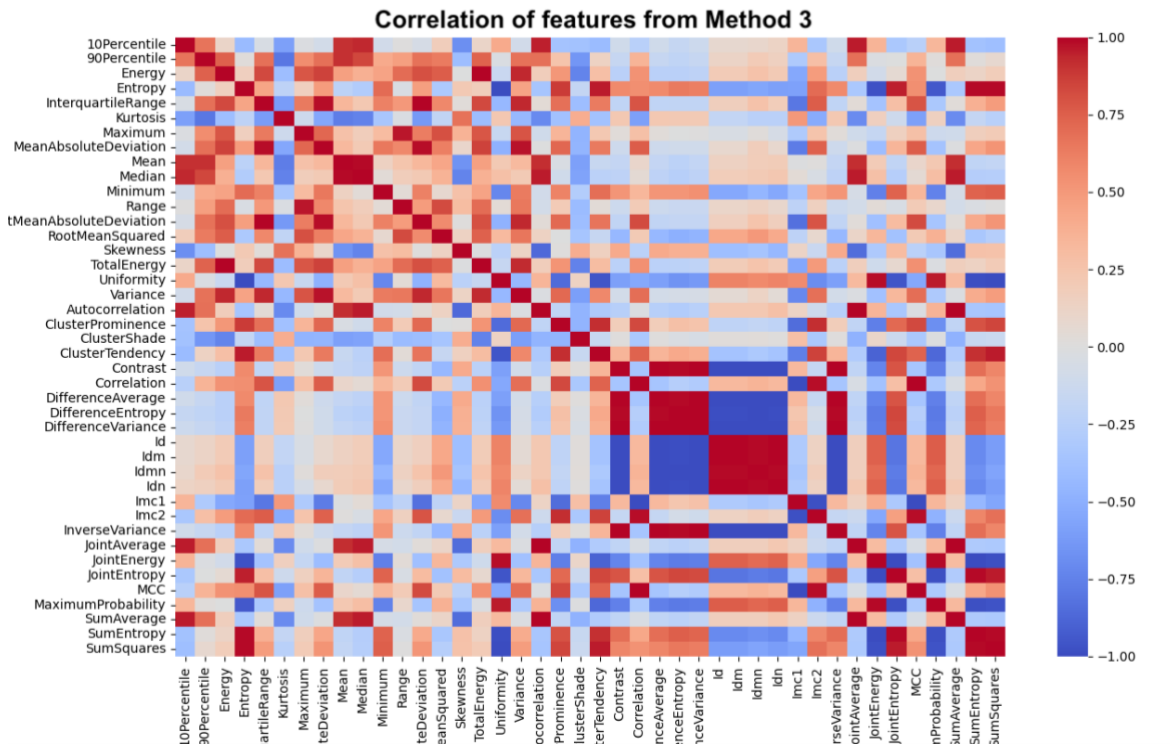


Figure A 1: Correlation matrix of the features from Method 3.

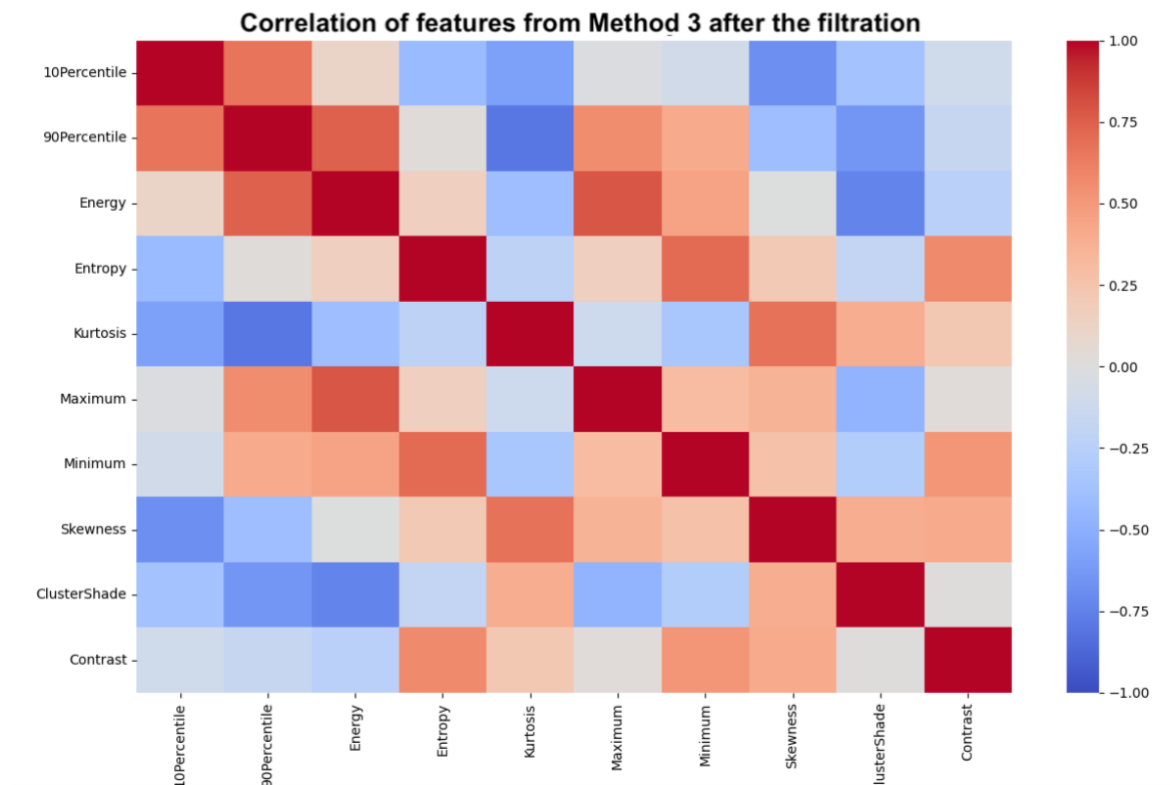


Figure A 2: Correlation matrix of the features from Method 3 after the filtration.

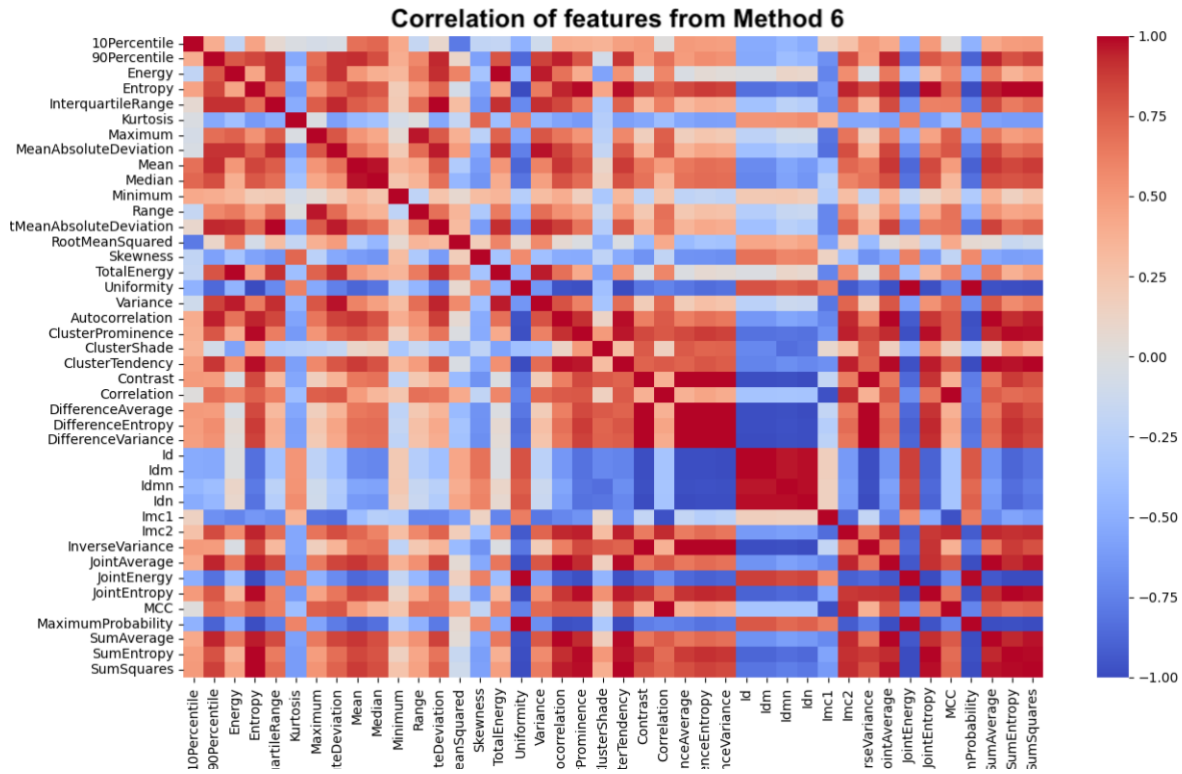


Figure A 3: Correlation matrix of the features from Method 6.

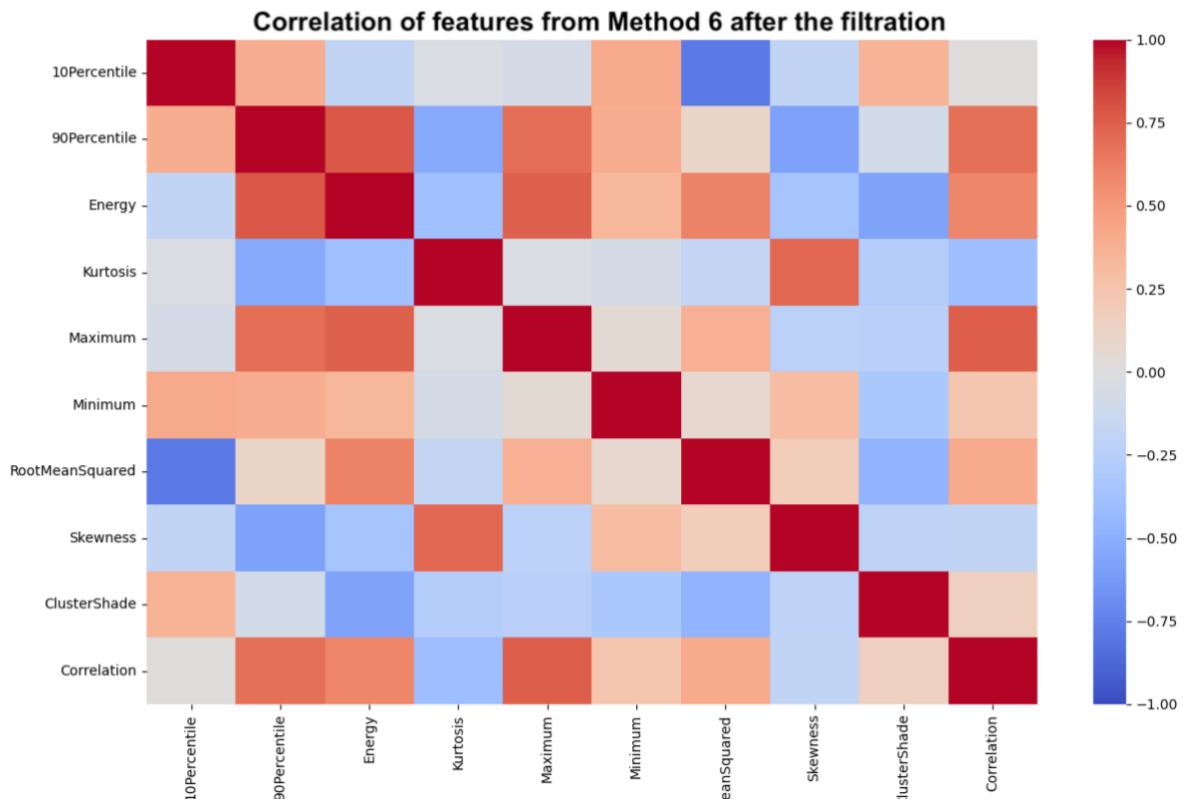


Figure A 4: Correlation matrix of the features from Method 6 after the filtration.

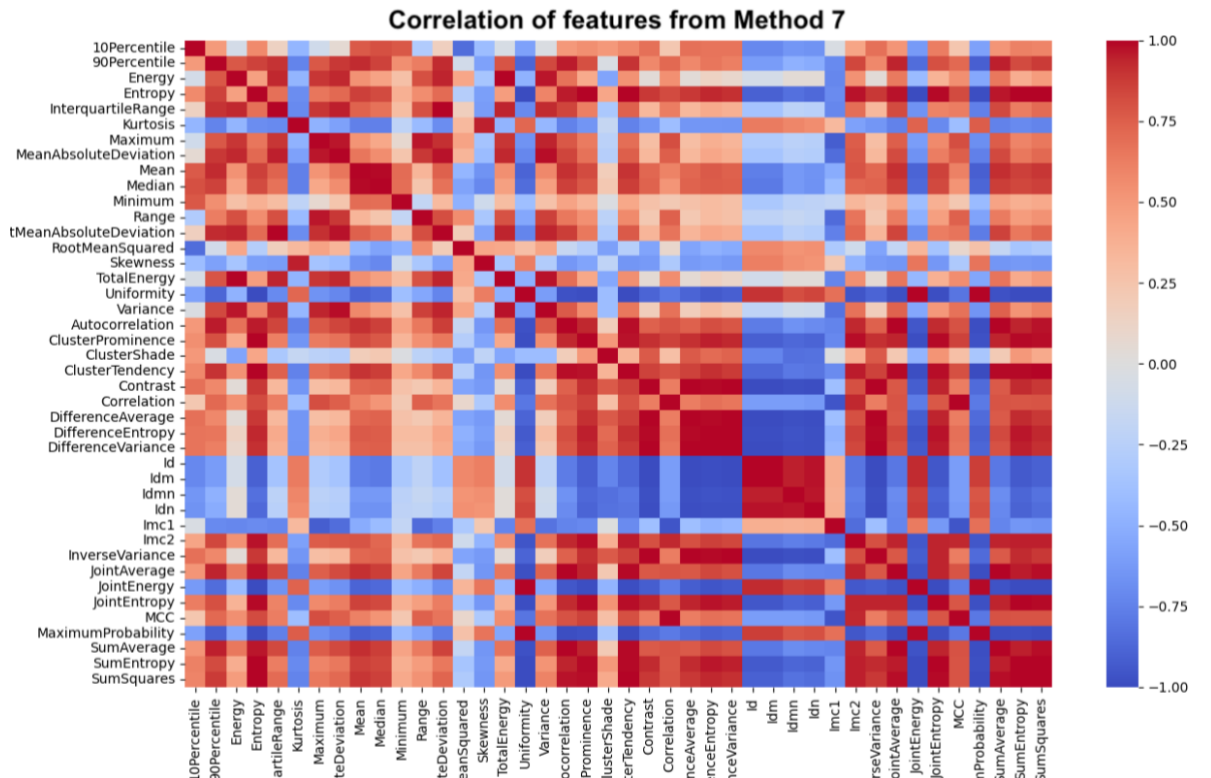


Figure A 5: Correlation matrix of the features from Method 7.

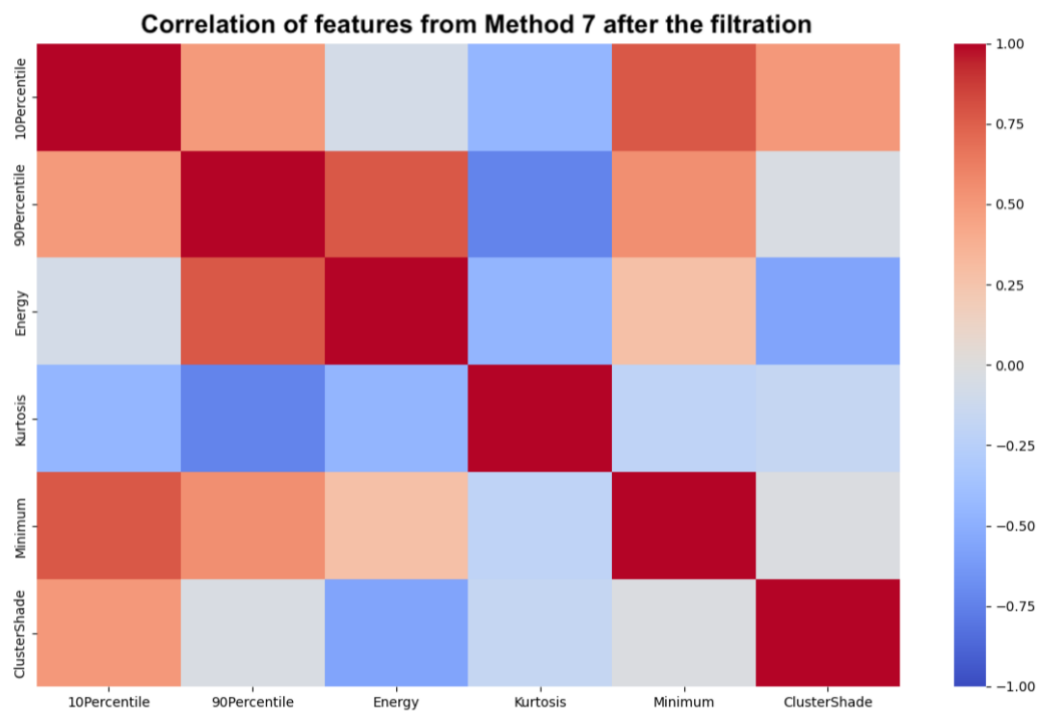


Figure A 6: Correlation matrix of the features from Method 7 after the filtration.

Feature	Mathematical Formula	Pyradiomics function
10 th percentile	-	get10PercentileFeatureValue()
90 th percentile	-	get90PercentileFeatureValue()
Energy	$Energy = \sum_{i=1}^{Np} (X(i) + c)^2$	getEnergyFeatureValue()
Entropy	$Entropy = - \sum_{i=1}^{Np} p(i) \log_2(p(i) + \epsilon)$	getEntropyFeatureValue()
Kurtosis	$K = \frac{\frac{1}{Np} \sum_{i=1}^{Np} (X(i) - \bar{X})^4}{\left(\sqrt{\frac{1}{Np} \sum_{i=1}^{Np} (X(i) - \bar{X})^2} \right)^4}$	getKurtosisFeatureValue()
Min	-	getMinimumFeatureValue()
Max	-	getMaximumFeatureValue()
Skewness	$S = \frac{\frac{1}{Np} \sum_{i=1}^{Np} (X(i) - \bar{X})^3}{\left(\sqrt{\frac{1}{Np} \sum_{i=1}^{Np} (X(i) - \bar{X})^2} \right)^3}$	getSkewnessFeatureValue()
ClusterShade	$CS = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} (i + j - \mu_x - \mu_y)^3 p(i, j)$	getClusterShadeFeatureValue()
Correlation	$C = \frac{\sum_{i=1}^{Ng} \sum_{j=1}^{Ng} p(i, j) ij - \mu_x \mu_y}{\sigma_x(i) \sigma(j)}$	getCorrelationFeatureValue()
Contrast	$Contrast = \sum_{i=1}^{Ng} \sum_{j=1}^{Ng} (i - j)^2 p(i, j)$	getContrastFeatureValue()
InterquartileRange	<i>Interquartile range</i> = $P_{75} - P_{25}$	getInterquartileRangeFeatureValue()
RootMeanSquare	$RMS = \sqrt{\frac{1}{Np} \sum_{i=1}^{Np} (X(i) + c)^2}$	getRootMeanSquareFeatureValue()

Figure A 7: Formulas and functions of the final 13 features.