UNIVERSITAT DE BARCELONA

Final Degree Project

Biomedical Engineernig Degree

“ Development of a Clinical-Radiomic Model for Head and Neck Cancer Outcome Prediction Based on Multi-Scanner PET Imaging “

Barcelona, 11th June 2024

Author: Nuria González Cuesta

Director: Aida Niñerola Baizán

Tutors: Gabriel Reynés Llompart,
Laura Rodríguez Bel

Acknowledgments

First of all, I would like to thank Aida Niñerola for her mentorship and guidance throughout this project. I am especially grateful for her constant willingness to help, clarify doubts, and manage the progress of the project.

I would also like to thank Gabriel Reynés and Laura Rodríguez for their constant support, patience, and commitment. Their dedication has been essential for the development of this project. I also appreciate their generosity in sharing their knowledge and their advice, which has helped me grow professionally. I feel very grateful to have had the opportunity to work with talented professionals and to be part of a team defined by mutual care and passion for this field.

I extend my thanks to all members of the Nuclear Medicine Department at Hospital Universitari de Bellvitge, who warmly welcomed me into their group from the very beginning, providing me with their support and help.

Finally, I am very grateful to my family for their love and support throughout this project. Above all, I want to thank them for instilling in me the importance of culture, critical thinking, and perseverance.

Abstract

Recurrence in patients with head and neck cancer presents a challenge to clinicians due to its poor prognosis. The availability of a single model, user-friendly, and capable of predicting recurrence and exitus across different types of head and neck cancer, is clinically valuable, supporting risk stratification and personalized decision-making.

This study aims to develop a robust and generalizable machine learning model for predicting recurrence and exitus in head and neck cancer patients, with locoregional tumour involvement with no distant lesions suspicious of malignancy, integrating clinical and radiomic data extracted from multi-scanner PET images (Discovery ST, Discovery IQ and Discovery MI).

The process includes tumour lesions and lymph nodes semi-automatic segmentation, followed by spatial standardisation. Subsequently, radiomic features were extracted and harmonized.

To develop the predictive models using clinical and radiomic features, the data was splitted into training and testing sets. For the radiomic data, collinearity was eliminated and univariate feature selection method was applied to select the most relevant variables. Different machine learning algorithms were trained and tested to predict recurrence and exitus outcomes.

This integrative approach underscores the potential of combining radiomics and clinical data to inform clinical decision-making in the management of head and neck cancer.

Keywords: Head and Neck Cancer; Recurrence; Exitus; Outcome Prediction; Radiomics; Multi-scanner; Machine Learning

Resum

La recidiva tumoral en pacients amb càncer de cap i coll representa un repte per als professionals mèdics a causa del seu pronòstic desfavorable. L'existència d'un únic model, de fàcil ús i capaç de predir tant la recidiva com l'èxitus en diferents tipus de càncer de cap i coll, té un gran valor clínic, ja que pot donar suport a l'estratificació del risc i a la presa de decisions personalitzades.

Aquest estudi té com a objectiu desenvolupar un model d'aprenentatge automàtic, robust i generalitzable, per predir la recidiva i l'èxitus en pacients amb càncer de cap i coll, amb afectació tumoral locoregional i sense lesions a distància sospitoses de malignitat, integrant dades clíniques i radiòmiques extretes d'imatges PET, obtingudes mitjançant diferents escàners (Discovery ST, Discovery IQ i Discovery MI).

El projecte inclou la segmentació semiautomàtica de les lesions tumorals i dels ganglis limfàtics, seguida d'una estandardització espacial. Posteriorment, s'extreuen i harmonitzen les característiques radiòmiques.

Per desenvolupar els models predictius utilitzant les característiques clíniques i radiòmiques, les dades es van dividir en conjunts d'entrenament i de prova; es va eliminar la col·linealitat i es va aplicar un mètode de selecció univariable per triar les variables radiòmiques més rellevants. Finalment, es van entrenar i provar diferents algoritmes d'aprenentatge automàtic per predir recidiva i èxitus.

Aquest enfocament posa en relleu el potencial de combinar dades radiòmiques i clíniques per donar suport a la presa de decisions mèdiques en la gestió del càncer de cap i coll.

Paraules clau: Càncer de Cap i Coll; Recidiva; Èxitus; Predicció de l'Evolució; Radiòmica; Multi-escàner; Aprenentatge Automàtic

Figures Content

Tables Content

Index

Glossary

PET	Positron Emission Tomography
CT	Computerized Tomography
HNC	Head and Neck Cancer
HPV	Human Papillomavirus
EBV	Epstein-Barr Virus
DNA	Deoxyribonucleic Acid
18F-FDG	18F-Fluorodeoxyglucose
ML	Machine Learning
2D	Two Dimensional
3D	Three Dimensional
DICOM	Digital Imaging and Communication in Medicine
AI	Artificial Intelligence
SUV	Standarized Uptake Value
MRI	Magnetic Resonance Imaging
IoT	Internet of Things
FDA	Food and Drug Administration
PVE	Partial Volume Effect
PVC	Partial Volume Correction
DL	Deep Learning
CNNs	Convolutional Neural Networks
EANM	European Association of Nuclear Medicine
IBSI	Image Biomarker Standardisation Initiative
GLCM	Gray Level Co-occurrence Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
NGTDM	Neighbouring Gray Tone Difference Matrix
GLDM	Gray Leven Dependence Matrix
ROI	Region of Interest
ANOVA	Analysis of Variance
LASSO	Least Absolute Shrinkage and Selection Operator
SVM	Support Vector Machine
DT	Decision Tree

RF	Random Forest
KNN	K-Nearest Neighbours
LR	Logistic Regression
CSV	Comma-Separated-Values
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve

1. Introduction

This study aims to develop a generalizable machine learning model that integrates clinical and radiomic features from multi-scanner PET imaging to predict recurrence and mortality in head and neck cancer patients. Accurate early prediction of these outcomes can improve clinical decision-making.

1.1. Project Context and Justification

Head and neck cancer (HNC) presents considerable challenges in clinical oncology due to its intrinsic heterogeneity and high mortality rates. This group of cancers accounts for approximately 3% of all cancers globally, with an estimated 660,000 new diagnoses annually. HNC primarily originates from the mucosal linings of the oral cavity, oropharynx, nasopharynx, hypopharynx, or larynx [1]. Despite ongoing advancements in surgical techniques, radiation therapy, and systemic treatments, the 5-year survival rate for HNC has remained around 50% for several decades. For locally advanced disease, this rate can drop to less than 50% [2].

Oropharyngeal cancer is a specific type of HNC that originates in the tissues of the oropharynx. While tobacco and alcohol consumption remain contributing risk factors, the leading cause is now Human Papillomavirus (HPV) infection, particularly HPV-16, HPV-related oropharyngeal cancers generally exhibit a better prognosis and are notably more responsive to treatment [3].On the other hand, nasopharyngeal carcinoma is a type of HNC that originates in the upper throat (nasopharynx), this type of HNC is strongly associated with Epstein-Barr virus (EBV) infection, especially in Southern China, Southeast Asia, and North Africa. High levels of EBV DNA in the blood are linked to poorer treatment response and a higher risk of tumour spread [4].

Outcome prognostication has been marked as one of the approaches to improve management of HNC. While traditional staging systems and established clinical factors are important prognostic indicators, they often do not fully capture the intricate biological characteristics of individual tumours or their spatial complexity. This limitation can lead to variable treatment responses and high rates of recurrence and distant metastasis, which are common causes of treatment failure and death. Radiomics builds upon this foundation by extracting high-dimensional, objective quantitative features from medical images, revealing subtle intratumoral heterogeneity that is invisible to the human eye, thereby offering a more nuanced understanding of tumour behaviour beyond conventional clinical assessment [5].

Positron Emission Tomography (PET) is a nuclear medicine imaging technique that utilizes radioactive tracers, such as 18F-fluorodeoxyglucose ([18F]FDG), to visualize and quantify metabolic activity within organs and tissues. This functional imaging capability allows for the precise localization of disease processes, as malignant cells typically exhibit altered metabolic rates, such as increased glucose uptake. [18F]FDG PET/CT is a highly accurate technique for diagnosing primary HNC lesions, particularly those that cannot be seen by radiological modalities [6].

Machine learning (ML) is increasingly being integrated into clinical oncology to enhance cancer diagnosis, predict patient outcomes, and support treatment planning. Unlike traditional rule-based systems, ML algorithms learn complex patterns directly from data, enabling more adaptive and data-driven decision-making. The growing availability of rich imaging and molecular datasets has further accelerated the application of ML, particularly in linking these multimodal data sources to

cancer progression. By uncovering subtle, often imperceptible correlations within large datasets, ML holds significant potential to improve the accuracy and personalization of oncologic care. [7]

Therefore, this project seeks to leverage radiomics and machine learning applied to [18F]FDG PET/CT imaging to improve risk stratification and prognostic accuracy in HNC. Such integrative approaches are essential for advancing precision oncology and predictive medicine.

1.2. Objectives

There is a critical need to identify more precise prognostic factors that indicate tumour aggressiveness or the risk of recurrence early in the disease course.

This project aims to develop a robust and generalizable machine learning model for predicting recurrence and exitus in head and neck cancer, integrating clinical and radiomic data extracted from multi-scanner PET images.

To support the primary objective, the project also encompasses several secondary objectives that focus on the technical and methodological steps:

- Segmenting tumours and lymphadenopathies from PET/CT images to define regions of interest.
- Transforming DICOM files into numerical matrices and 3D structures corresponding to the segmented regions of interest for computational analysis.
- Develop a software to extract radiomic features from regions of interest.
- Harmonizing radiomic features across different scanners and acquisition protocols to ensure data consistency and model generalizability.
- Developing and training various classification algorithms to predict recurrence and exitus, using both clinical variables and radiomic descriptors as input features, to select the best-performing model based on validation metrics.

1.3. Scope and Limitations

This project explores the availability of a single machine learning model trained with PET radiomic features and clinical data, and capable of predicting recurrence and exitus outcomes across different types of head and neck cancer, specifically focusing on oropharyngeal and nasopharyngeal carcinomas.

While the primary objective focuses on model construction and validation, the development of a graphical user interface for clinical implementation lies beyond the current scope of this project

Given the anatomical and pathological similarities among head and neck cancer subtypes, the approach is well-positioned for extrapolation to other forms such as hypopharyngeal and laryngeal cancers. These subtypes also rely heavily on imaging for diagnosis and treatment planning, making them suitable candidates for radiomics-based predictive modelling. In addition to its relevance within head and neck oncology, the proposed pipeline is adaptable to a wide range of other diseases where medical imaging is central to clinical decision-making.

This project will be developed within the Nuclear Medicine Department of Hospital Universitari de Bellvitge, which provides access to PET scanners from the vendor GE Healthcare, including the

Discovery ST, Discovery IQ, and Discovery MI models, each with distinct intrinsic and reconstruction characteristics

However, the study faces notable limitations that must be acknowledged. First, the relatively modest sample size of 200 patients, consisting of 122 oropharyngeal and 78 nasopharyngeal carcinoma cases. Although sufficient for exploratory analyses, small datasets inherently limit statistical power and increase the risk of model overfitting, where algorithms learn spurious correlations or noise specific to the training data rather than generalizable patterns. This can lead to overly optimistic performance metrics during internal validation, which may not hold when applied to external or prospective datasets. Additionally, the integration of such computational pipelines into clinical practice presents its own set of challenges. These include the need for interpretability and transparency in model outputs, regulatory and ethical considerations surrounding the use of AI in healthcare, and the alignment of predictive tools with existing clinical workflows.

2. Background

To carry out this project, it is essential to understand a series of key concepts and the current state of the art that support its development. The following section outlines the main theoretical and technical foundations.

2.1. Key Concepts

2.1.1. PET/CT Imaging

PET is a nuclear medicine imaging technique that utilizes radioactive tracers to visualize and quantify metabolic activity within organs and tissues. Following the administration of a radiotracer, positrons emitted from the radionuclide undergo annihilation with electrons, resulting in the production of two gamma photons emitted in nearly opposite directions. These photons are detected simultaneously by a ring of scintillation detectors, allowing for the precise reconstruction of three-dimensional images that reflect the tracer's distribution. This functional imaging capability allows for the precise localization of disease processes, as malignant cells typically exhibit altered metabolic rates, such as increased glucose uptake.

PET imaging offers high sensitivity for detecting metabolic activity, making it highly effective for identifying functional abnormalities. However, it has relatively low spatial resolution (typically 4–6 mm), limiting its ability to visualize fine structural details.

The standardized uptake value (SUV) is a semiquantitative measure used in PET imaging to assess radiopharmaceutical uptake, particularly in tumours. It normalizes the activity of a region of interest (ROI) by considering the injected activity and a measure of volume of distribution, like body weight or lean body mass. SUV values are often used to help differentiate between benign and malignant lesions, with higher SUV values generally suggesting increased metabolic activity and potentially malignancy [8].

When PET is combined with Computed Tomography (CT), the resulting PET/CT scans provide detailed three-dimensional images that integrate both metabolic (PET) and anatomical (CT) information. This fusion allows for precise localization and characterization of tumours, offering a comprehensive view of disease extent and biological behaviour. The core principle of PET/CT lies in its ability to fuse functional or metabolic information with anatomical detail. [9]

In most cases, 18F-fluorodeoxyglucose is used as the radiotracer for PET/CT in patients with head and neck cancer. PET involves the injection of a radiopharmaceutical that emits positrons, which, upon interacting with electrons, produce gamma photons. This imaging modality complements conventional imaging methods such as CT and MRI in treatment planning, prognosis, monitoring, and evaluation, due to its ability to provide tumour metabolic information through functional imaging [10].

2.1.2. Head and Neck Cancer

Head and neck carcinoma is a biologically heterogeneous group of cancers arising from the mucosal linings of the oral cavity, oropharynx, hypopharynx, and larynx, accounting for about 3% of all cancers globally with approximately 660,000 new cases annually [11].

Recent epidemiological data indicate a notable shift in HNC incidence patterns. Over the past two decades, there has been a rising incidence of HNC, particularly observed among elderly patients. Projections suggest this trend will continue, with an estimated 1.08 million new cases annually by 2030. This increase is particularly pronounced for oral and oropharyngeal cancers in women [12].

Oropharyngeal cancer is a specific type of HNC that originates in the tissues of the oropharynx, encompassing the base of the tongue, tonsils, soft palate, and posterior pharyngeal wall. Most of these cancers are squamous cell carcinomas. Common signs and symptoms of oropharyngeal cancer include a lump in the neck, a persistent sore throat, a white patch on the tongue or lining of the mouth that does not resolve and coughing up blood.

The epidemiology of oropharyngeal cancer has undergone a significant transformation. While tobacco and alcohol consumption remain contributing risk factors, the leading cause is now Human Papillomavirus infection, particularly HPV-16. HPV-related oropharyngeal cancers are experiencing a drastic increase in incidence, especially among younger populations. HPV-related oropharyngeal cancers generally exhibit a better prognosis and are notably more responsive to treatment [13].

Nasopharyngeal carcinoma is a less common type of HNC that originates in the upper throat, behind the nose. The clinical features of nasopharyngeal carcinoma can be subtle, often leading to advanced-stage diagnosis. Common signs and symptoms at presentation include painless neck lumps, nasal obstruction, epistaxis or headache. It is strongly associated with Epstein-Barr virus infection, especially the non-keratinizing subtype prevalent in endemic regions such as Southern China, Southeast Asia, and North Africa. High levels of EBV DNA in the blood are linked to poorer treatment response and a higher risk of tumour spread. Common symptoms include palpable lymph node metastases in the neck, hearing loss, ear pain, persistent nasal blockages or stuffiness or bloody nasal discharge [14].

In the management of head and neck cancer, therapeutic strategies are meticulously tailored to the primary tumour site, histological subtype, stage, and crucial prognostic indicators. While early-stage disease may often be effectively treated with single-modality therapy, by surgery or radiation, advanced stages frequently necessitate a multimodality approach. This typically involves combinations of surgery, highly conformal radiation to maximize local control while minimizing toxicity to adjacent critical structures, and systemic therapies including chemotherapy and increasingly, immunotherapy as frontline or salvage options.

2.1.3. Radiomics

Radiomics is a field within medical imaging that extracts high-dimensional, quantitative data from medical images. The fundamental aim of radiomics is to identify and quantify complex patterns that may not be discernible by the human eye, thereby enhancing the diagnostic, prognostic, and predictive potential of imaging data. This data-driven approach enables the characterization of lesion properties such as shape, intensity, and heterogeneity. Radiomic

features are mainly categorized into four major groups: statistical, model-based, transform-based, and shape-based features [15].

The radiomics workflow is a multi-stage analytical process encompassing image acquisition, preprocessing (including normalization, standardization, and discretization), lesion segmentation, feature extraction, feature selection, and model training and validation, all aimed at generating robust, reproducible, and clinically meaningful predictive models from medical imaging data [16].

Recent studies have underscored the utility of radiomics in diverse clinical contexts, such as non-invasive tumour grading, forecasting treatment toxicity, and stratifying patient survival outcomes. However, challenges remain, including the standardization of imaging protocols, reproducibility of features, and the need for robust, multi-institutional validation. As it evolves, radiomics is expected to play a pivotal role in precision medicine by integrating imaging biomarkers into personalized treatment strategies [17].

2.2. State of the Art

The growing presence of Artificial Intelligence (AI) in everyday life, along with disruptive advances in the field, poses significant challenges and opportunities both in technological development and in social impact. The integration of AI with other domains, such as the Internet of Things (IoT), points to a shift toward "connected intelligence," where devices are not only interconnected but also intelligent. For example, in the automotive industry, autonomous vehicles rely on AI systems to interpret their environment and make safe decisions while in motion, as well as to prevent manufacturing errors through image analysis. Similarly, in the field of e-commerce, AI-powered recommendation algorithms provide personalized suggestions to consumers, enhancing the online shopping experience. These examples illustrate how the integration of AI into different sectors is transforming the way we interact with technology and the world around us.

In the field of health management and prognostics, AI is used to improve efficiency in various tasks such as error detection, diagnosis, and decision-making. AI techniques such as supervised and unsupervised machine learning, fuzzy systems, and reinforcement learning algorithms are applied from data preparation to decision support. AI in surgical robotics enhances precision, control, and decision-making during procedures. It enables real-time analysis of medical images and patient data to guide robotic movements with high accuracy. AI also assists in preoperative planning, predicting complications, and optimizing surgical strategies. Moreover, AI is also used in the segmentation of medical images, improving accuracy in the detection of tumours and other anomalies [18].

Many studies have been published on the implementation of artificial intelligence models for predicting tumour evolution in PET/CT images. The following tables provide an overview of the most relevant articles, with the models applied according to the type of cancer studied.

Table 1. Title, cancer type studied, and AI model from articles on the implementation of AI for forecasting the progression of different types of cancer [19], [20], [21], [22], [23].

Title	Cancer Type	Model
[18F]FDG PET-CT radiomics and machine learning in the evaluation of prostate incidental uptake	Prostate Cancer	5 different algorithms (unspecified), with and without PCa
Evaluation of survival of the patients with metastatic rectal cancer by staging 18F-FDG PET/CT radiomic and volumetric parameters	Rectal Cancer	Naive Bayes, Logistic Regression, k-Nearest Neighbor, Decision Tree, Support Vector Machine, Random Forest
The application of different machine learning models based on PET/CT images and EGFR in predicting brain metastasis of adenocarcinoma of the lung	Brain Metastasis	Logistic Regression
Prediction of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer by using a deep learning model with 18F-FDG PET/CT	Breast Cancer	Convolutional Neural Network
Prediction of lung malignancy progression and survival with machine learning based on pre-treatment FDG-PET/CT	Lung Cancer	Convolutional Neural Network

The table above presents various studies that use machine learning algorithms to analyse PET/CT images with [18F]FDG in different types of cancer. These studies have applied artificial intelligence for predicting tumour progression in various neoplasms, demonstrating the wide range of machine learning techniques available for cancer evaluation and prediction.

It is also essential to examine how these machine learning methodologies have been specifically applied within the domain of head and neck cancers. Given the anatomical complexity and heterogeneity characteristic of head and neck tumours, the integration of advanced computational techniques with [18F]FDG PET/CT imaging offers significant potential for enhancing diagnostic accuracy, prognostic stratification, and treatment planning.

The following table summarizes key studies applying machine learning models to various subtypes of head and neck cancers, highlighting their clinical endpoints, dataset sizes, model training approaches, and evaluation metrics.

Table 2. Summary of machine learning applications for outcome prediction in head and neck cancer [24].

Author, Year of Publication	Head and Neck or Subsite	Application (End Point)	No. of Patients	Model Evaluation
Haider et al., 2020 (USA, Germany & Canada)	OPSCC (Oropharyngeal Squamous Cell Carcinoma) & MLNs (Metastatic Lymph Nodes)	Outcome prediction (HPV status)	741 (MLNs) / 435 (OPSCC)	AUC: 0.78
M.D. Anderson Cancer Center, Head and Neck Quantitative Imaging Working Group, 2018 (USA)	OPSCC (Oropharyngeal Squamous Cell Carcinoma)	Outcome prediction (local recurrence)	465	Discriminating value: 94%
Vallieres et al., 2017 (Canada)	HNC (Head and Neck Cancer)	Outcome prediction (Locoregional Recurrence, Distant Metastasis, & Overall Survival)	300	Locoregional recurrence AUC = 0.69
Kaźmierska et al., 2022 (Poland & Canada)	LAHNC (Locally Advanced Head and Neck Cancer)	Outcome prediction (Incomplete response and disease progression)	290	AUC: 0.68
Kim et al., 2022 (Republic of Korea)	HNSCC (Head and Neck Squamous Cell Carcinoma)	Outcome prediction (local tumor recurrence)	215	AUC: 0.77

The previous table emphasize the effectiveness of AI-driven models in predicting clinical outcomes across various subtypes of head and neck cancers. The diverse approaches, ranging from HPV status prediction to recurrence and survival estimation, demonstrate the adaptability of machine learning techniques to different diagnostic and prognostic endpoints.

2.3. State of the Situation

This study forms part of a broader research initiative conducted at Nuclear Medicine Department of Hospital Universitari de Bellvitge, aimed at integrating radiomic features with clinical and metabolic data to enhance the characterization and understanding of head and neck cancer, the project seeks to improve diagnostic and prognostic.

The project started after all preliminary processes had been completed, including approval by the ethics committee, patient selection, clinical data extraction, and anonymization of the studies. These preparatory steps ensured the foundations for the development of the study.

3. Market Analysis

3.1. Market Evolution

The market analysis of artificial intelligence algorithms for cancer evolution prediction reveals a rapidly growing and relatively recent field (emerging around the 1990s). This growth is driven by technological advances and the increasing need for accurate and personalized medical solutions. Moreover, the field is shaped by the rising demand for the development of novel biomarkers, which are essential for enabling personalized diagnosis and treatment within the framework of precision medicine [25].

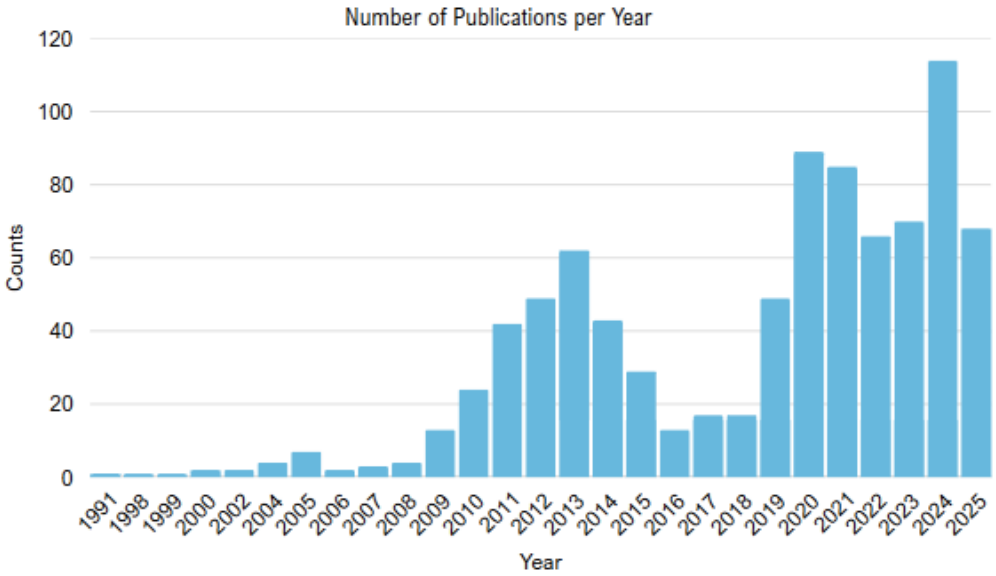

Figure 1. Number of publications about prediction of cancer using AI in PubMed per year [25].

A growing number of companies are investing in the development of such algorithms, motivated by the increasing demand for effective predictive tools and substantial investment in research and development. This has led to intense competition and continuous innovation in the sector. However, no machine learning algorithms specifically designed for prognosis based on PET/CT imaging of head and neck cancer have been identified on the market.

Some of the products and projects currently being developed by companies in this field include:

- **IBM** is developing a range of AI-powered products designed to enhance the medical and healthcare sectors. Having precise, up-to-date, and accessible information is essential for clinical decision-making. *DynaMed* and *Micromedex with Watson* are two of the most advanced and reliable tools available to healthcare professionals, integrating artificial intelligence with extensive databases to improve the quality of care.
- **Arterys** is a company focused on AI-based software solutions aimed at optimizing the interpretation of medical imaging by radiologists, generating reports, and facilitating information sharing with referring physicians. Their products include automated image analysis, with one-click lesion segmentation in CT and MRI images, and configurable AI-based report generation tools that enhance workflow efficiency and consistency. Furthermore, Arterys tools integrate seamlessly with existing systems, including worklists, notification systems, and dictation platforms, thereby optimizing clinical workflows and facilitating the adoption of AI technologies in medical environments.

- **Zebra Medical Vision** is a pioneering company in the use of artificial intelligence to analyze medical images and provide automated diagnostic support. Its platform, *HealthMinds*, offers advanced solutions for the early detection and monitoring of various diseases, including cancer. Zebra's algorithms can analyze X-rays, CT scans, and MRIs to identify anomalies indicative of cancer, as well as other conditions such as cardiovascular, hepatic, and pulmonary diseases. This technology enables faster and more accurate detection, supporting physicians in making informed decisions and improving patient outcomes. Zebra Medical Vision is also working on integrating its solutions with electronic health record systems to facilitate adoption in hospitals and clinics worldwide.
- **Google DeepMind Health**, a division of Google, is dedicated to applying artificial intelligence in medicine. One of its most notable achievements is the development of algorithms for medical image analysis, particularly in radiology and pathology. DeepMind has collaborated with various hospitals and medical institutions to create tools capable of detecting eye diseases from retinal scans and identifying head and neck cancers in CT images. Among its best-known projects is an algorithm for detecting diabetic retinopathy, which has shown accuracy comparable to that of human specialists. Additionally, DeepMind is working on optimizing hospital workflows by predicting complications such as acute kidney injury.

In the public sector, various institutions are also developing projects in this domain. For example, the **Catalan Health Institute** has developed four proprietary AI algorithms that have improved breast cancer diagnosis through the quantification of four biomarkers (HER2, Ki67, estrogen receptors, and progesterone receptors). The project aims to drive the digital transformation of pathology services. Both histological samples and diagnostic-quality images are being digitized, enabling the effective replacement of microscopes with high-resolution screens and allowing all hospitals within the network to share high-precision images [26].

3.2. Market Sector

The main potential customers for this head and neck cancer outcome prediction software (recurrence and mortality) are healthcare facilities, particularly hospitals. Bellvitge Hospital would be the first to implement it, as the algorithm has been specifically designed to integrate with its system. Once validated, the software could be applied in any hospital worldwide that uses similar techniques for patient diagnosis and follow-up.

It also represents an opportunity for software companies specializing in medical technology, such as GE Healthcare, which could incorporate this tool into their portfolio as a distinctive, innovative solution based on artificial intelligence and personalized medicine. Integrating this technology into existing clinical systems could enhance efficiency and improve the quality of medical care.

Finally, the field of biomedical research would also benefit. Academic institutions and research groups could use the software to analyse real-world data and identify key factors in cancer progression, contributing to the development of new therapeutic strategies and advancing knowledge in oncology.

4. Concept Engineering

In this section, the project process is detailed, from the initial conception to the final implementation. Various options for carrying out different tasks are considered to select the best possible solution; a thorough study of the available solutions is done, and the most suitable one is proposed.

4.1. Segmentation

The first step to extract radiomic data from the images is to delineate the regions of interest. To achieve accurate segmentation in medical images, especially in the field of medical imaging, various methods can be used. Below, different segmentation methods are presented and compared, highlighting their advantages and disadvantages, as well as their applicability.

4.1.1. PET VCAR Software

PET VCAR (*PET* Volume Computer Assisted Reading) is a specialized tool for segmenting and analysing lesions in PET/CT images. This software uses advanced algorithms to identify and quantify regions of interest based on PET tracer signal intensity [27].

PET VCAR offers a user-friendly interface that enhances clinical workflow in PET/CT lesion segmentation, and seamless integration with other diagnostic and reporting systems. However, it also presents notable limitations, such as vendor dependency, as it is designed exclusively for GE Healthcare PET/CT systems, restricting its use with other manufacturers' equipment. Additionally, while the software automates much of the segmentation process, manual adjustments by clinicians may still be necessary for optimal results. Another limitation is that it does not allow the export of structures or images. The requirement for specific licenses and proprietary GE software may also pose cost and accessibility challenges, potentially limiting its broader adoption in clinical settings.

4.1.2. MIM Software

MIM (Medical Image Merge) Software is a medical registering and manipulating DICOM medical images. Over time, it has evolved into a comprehensive medical image management and processing system, as recognized by the FDA. MIM Software has segmentation tools, both semiautomated and manual, and multi-modality image integration to enhance accuracy, efficiency, and consistency in medical image segmentation. [28]

MIM Software offers significant advantages, including enhanced accuracy through several tools for manual segmentation or multi-modality image fusion, as well as streamlined workflows via semi-automatic segmentation algorithms. This segmentation software is also well established in Hospital de Bellvitge. However, potential drawbacks include the high costs associated with acquisition and maintenance, a steep learning curve due to its comprehensive features, and the inherent limitations of automated segmentation requiring manual corrections and clinical validation.

4.1.3. 3D Slicer

3D Slicer is a free, open-source, cross-platform software application designed for the visualization, processing, segmentation, registration, and data analysis of medical, biomedical, and other 3D images and meshes. It has established itself as a widely used platform in the field of medical image processing [29].

One of the main advantages of 3D Slicer is its cost and accessibility as free, open-source software. It provides a wide range of tools and extensions for different PET/CT segmentation needs, from manual methods to advanced AI-based techniques. Additionally, 3D Slicer supports multiple medical image formats, including DICOM, the standard for PET/CT imaging. However, despite its benefits, 3D Slicer has some limitations, such as a steep learning curve for new users due to its extensive features and modular interface, as well as potential restrictions in clinical certifications or regulatory compliance for direct use in clinical settings without additional validation.

4.2. Image Preprocessing

Preprocessing standardizes PET/CT scans reducing variability related to acquisition variability or failure. It enhances image quality by noise reduction and artifact correction for reliable radiomic analysis. It also optimizes images for precise feature extraction from defined regions. Various pre-processing steps can be used in nuclear medicine images [30].

4.2.1. Partial Volume Correction

The partial volume effect (PVE) influences the accuracy of PET imaging. Due to the limited spatial resolution of these systems, small structures or regions with high tracer uptake appear blurred, leading to underestimation of activity in small lesions. To address PVE, partial volume correction (PVC) methods are employed, many of which leverage high-resolution anatomical data from MRI or CT scans by integrating structural information from these modalities with the functional PET data, PVC techniques help refine tracer distribution estimates[31]. But current evidence does not support using PVC routinely in PET image analysis [30].

4.2.2. Deep Learning in Image Preprocessing

Deep Learning (DL) techniques are applied to improve the quality of PET images. In PET/CT, DL is applied to three key areas [32]:

- **Denoising**, where neural networks trained on high and low dose images preserve fine details better than classical filters.
- **Deformable image registration**, where convolutional neural networks (CNNs) predict displacement fields to align PET/CT scans, correcting patient motion and reducing artifacts.
- **Super-resolution**, enhancing spatial resolution to improve detection of small lesions. These techniques enable more accurate quantification and improved diagnostics.

DL offers a powerful approach to address PET/CT challenges like noise, attenuation, and limited resolution, demonstrating significant potential for advancing nuclear medicine imaging. Despite significant advances, the clinical use of deep learning for PET/CT preprocessing still faces challenges with regulatory approval [33].

4.2.3. Converting PET Voxel Values to Standard Uptake Value (SUV)

The standardized uptake value (SUV) is a semi-quantitative measure of radiotracer uptake in tissues, normalized to the patient's size and the injected dose. SUV is a widely used metric in clinical PET/CT to assess metabolic activity, such as in tumours. The general formula for calculating SUV is:

$$\text{SUV} = \frac{(\text{Voxel Activity Concentration})}{\left(\frac{\text{Injected Dose}}{\text{Patient Weight}}\right)}$$

Standardization of SUV measurements is crucial for PET radiomics, where quantitative imaging biomarkers require rigorous reproducibility. To ensure reliable radiomic feature extraction guidelines have been established to harmonize all aspects of PET imaging, including SUV normalization [34].

4.2.4. Filters in Image Preprocessing

Image filtering is used to modify pixel values in an image to achieve specific objectives, such as reducing noise, enhancing edges, or smoothing the image [35].

- **Smoothing filters:** are used in PET/CT imaging to reduce noise while preserving important image features. For example, the Gaussian filter applies weighted averaging using a bell-shaped Gaussian function, effectively suppressing high-frequency noise and creating a smoother appearance. However, excessive smoothing can lead to loss of fine details and reduced spatial resolution, which is particularly important for detecting small lesions. Simpler alternatives like the mean filter perform basic neighbourhood averaging, offering more uniform but potentially less refined noise reduction compared to Gaussian approaches.
- **Edge-enhancement filters:** have the opposite purpose of smoothing filters. For example, Laplacian filters detect intensity changes corresponding to anatomical edges and tumour margins. While they are effective for improving tumour delineation, these filters have the drawback of simultaneously amplifying image noise.

A fortiori does not currently provide evidence that pre-processing images with denoising provides significant improvement for radiomics application [30].

4.2.5. Image Resampling

Resampling involves modifying the pixel or voxel dimensions of an image, altering its spatial resolution or alignment. In PET/CT scans, the native PET and CT images often have different resolutions and matrix sizes. Consequently, resampling to a common grid is necessary for tasks like image fusion, coregistration, and to reduce any heterogeneity in acquisition voxel size. Various resampling techniques are employed in clinical and research settings [36]:

- **Nearest-neighbour interpolation:** The simplest method, assigning the value of the closest original pixel to the new location. While computationally efficient, it may introduce block-like artifacts, particularly with significant resolution changes.
- **Linear interpolation:** Bilinear or trilinear interpolation calculates new pixel values using weighted averages of neighbouring pixels. Produces smoother results than nearest-neighbour but may blur sharp edges.

4.3. Radiomics features

Radiomics is a process that extracts quantitative features from medical images to provide insights into underlying pathophysiology. These features are broadly categorized into different types depending on their nature and where they come from.

4.3.1. Radiomics Packages

Numerous open-source radiomics software packages exist, each varying in user interface, adherence to the IBSI standard [37], and capabilities like PET image. While a comprehensive list is constantly evolving, commonly utilized tools include Pyradiomics, SERA, LIFEx, MITK phenotyping, and CERR. These packages offer researchers and clinicians diverse options for extracting quantitative imaging features.

- **PyRadiomics:** Open-source Python package designed for extracting radiomic features from 2D and 3D medical images. It serves as a reference standard for radiomic analysis, offering a robust, tested platform to ensure reproducible feature extraction [38]. Studies confirm PyRadiomics as *IBSI-compliant*, it adheres to the *Image Biomarker Standardisation Initiative,* with high standardization in feature implementation validated by IBSI phantom tests [39].
- **ViSERA:** Radiomics software tools developed at Johns Hopkins University for standardized feature extraction in medical imaging, adhering to the IBSI guidelines. Both tools prioritize reproducibility and interoperability, bridging research and clinical needs in quantitative imaging [40].
- **LIFEx:** It's a free, multiplatform, and user-friendly software designed for the analysis of molecular imaging data, with a particular focus on PET, SPECT, and MRI for quantitative research. Developed for medical imaging professionals, LIFEx doesn't require programming skills. However, while it's IBSI-compliant, its feature calculations may deviate slightly from other platforms due to differences in implementation [41].
- **MITK Phenotyping:** Openly distributed radiomics framework that implements a comprehensive set of radiomic features, adhering to the latest international standards, including IBSI [42]. MITK Phenotyping has been used in studies involving PET/CT images, demonstrating its capability to handle and extract meaningful features from PET data in various research contexts [43].
- **PyCERR:** Platform that facilitates batch calculation and visualization of radiomics features, providing a structured environment for radiomics metadata. PyCERR was developed as a Python-native port of the original MATLAB-based CERR, it maintains the flexible and readable data structure of CERR while providing access to relevant tools within the Python programming environment [44]:

4.3.2. Radiomics features

Radiomics can be categorized into several groups that capture different aspects of the image data, such as texture, intensity, and shape-based features. The primary feature categories are described in the following table: [45]

Table 3. Summary of radiomic feature groups: categories, feature counts, and descriptions.

Feature Group	*Number of Features*	*Description*
First Order Statistics	19	These features describe the distribution of individual voxel intensities within the defined region of interest using basic statistical measures.
Shape-based (3D)	16	This category encompasses features that characterize the three-dimensional size and morphology of the ROI, independent of the intensity values within it.
Shape-based (2D)	10	Like the 3D shape features, this group describes the size and shape of the ROI but in a two-dimensional plane, also independent of intensity.
Gray Level Co-occurrence Matrix (GLCM)	24	These features analyze the spatial relationships between pairs of voxels with specific gray level intensities within the ROI, providing insights into the texture of the image.
Gray Level Run Length Matrix (GLRLM)	16	These features quantify the lengths of consecutive runs of voxels with the same gray level along various directions within the ROI, offering information about the image texture's directionality.
Gray Level Size Zone Matrix (GLSZM)	16	These features determine the size of connected regions of voxels that share the same gray level intensity, it measures the texture homogeneity regardless direction.
Neighbouring Gray Tone Difference Matrix (NGTDM)	5	These features quantify the difference in gray level between each voxel and the average gray level of its neighboring voxels, highlighting local intensity variations and texture uniformity.
Gray Level Dependence Matrix (GLDM)	14	These features quantify the dependencies between a voxel and its neighboring voxels based on their gray level similarity, providing insights into the size and homogeneity of regions with dependent gray levels.

4.4. Harmonization

Harmonization is an important step in our data pre-processing because we are analyzing PET data acquired from three different scanners, variations in equipment can introduce unwanted biases in radiomic features. One possible solution is to discard non-robust features, but it risks losing potentially useful clinical information. Applying harmonization techniques can help mitigate scanner-related variability while preserving predictive biomarkers. Since most radiomic features are sensitive to acquisition differences, harmonization offers a more efficient solution than aggressive feature elimination. The possible harmonization techniques are now described.

4.4.1. EARL

This harmonization technique standardizes PET/CT quantitative metrics across different scanners and institutions to ensure consistency. The EARL accreditation program includes quarterly calibration checks and annual phantom tests to validate SUV accuracy and recovery coefficients. By following EANM guidelines, accredited centers minimize variability in patient preparation, acquisition, and reconstruction. This ensures reliable, comparable image data for multicenter studies and clinical practice [46].

Some studies tested EARL harmonization using a 3D-printed phantom across different PET/CT systems, finding that while EARL harmonization reduced the variability of radiomic features across different scanner models and reconstruction settings, a large percentage of radiomic features still exhibited significant differences. This suggests that EARL harmonization method, though increased the number of comparable features compared to original clinical reconstruction, may be insufficient to make all radiomic features usable in such a setting. The persistence of differences highlights the need for additional harmonization strategies [30].

4.4.2. ComBat

ComBat is a data-driven method that estimates and corrects for site effects without requiring phantom acquisitions. The site effect can be directly derived and adjusted using available image feature values from different sites, without the need for additional image processing or new measurements. ComBat was initially introduced in the field of genomics and has been spread in different fields, like radiomics [47]. ComBat is based on the following formula:

$$y_{ij} = \alpha + \gamma_i + \delta_i \varepsilon_{ij}$$

- j represents a specific measurement of feature y.
- i is the experimental setup or conditions used for that measurement.
- α is the mean value of the feature y.
- γi is an bias caused by batch effects, like differences in scanners.
- δi is a scaling factor on the standard deviation of y, also caused by batch effects.
- εij accounts for random noise or error in the measurement.

This method assumes that any measurement can be affected by two types of errors, additive bias and multiplicative bias. For this technique, it is crucial that all measurements were taken under the same conditions to share the same biases. The site effects γi and δi can be corrected using the following formula:

$$y_{ij}^{ComBat} = \frac{y_{ij} - \hat{\alpha} - \hat{\gamma}_i}{\hat{\delta}_i} + \hat{\alpha}$$

- y_{ij} is the original measurement.
- $\hat{\alpha}$ is the mean of the feature across all data.
- $\hat{\gamma}_i$ is the estimated additive bias for the scanner.
- $\hat{\delta}_i$ is the estimated multiplicative bias for the scanner.

In the following image you can observe the effect of ComBat harmonization:

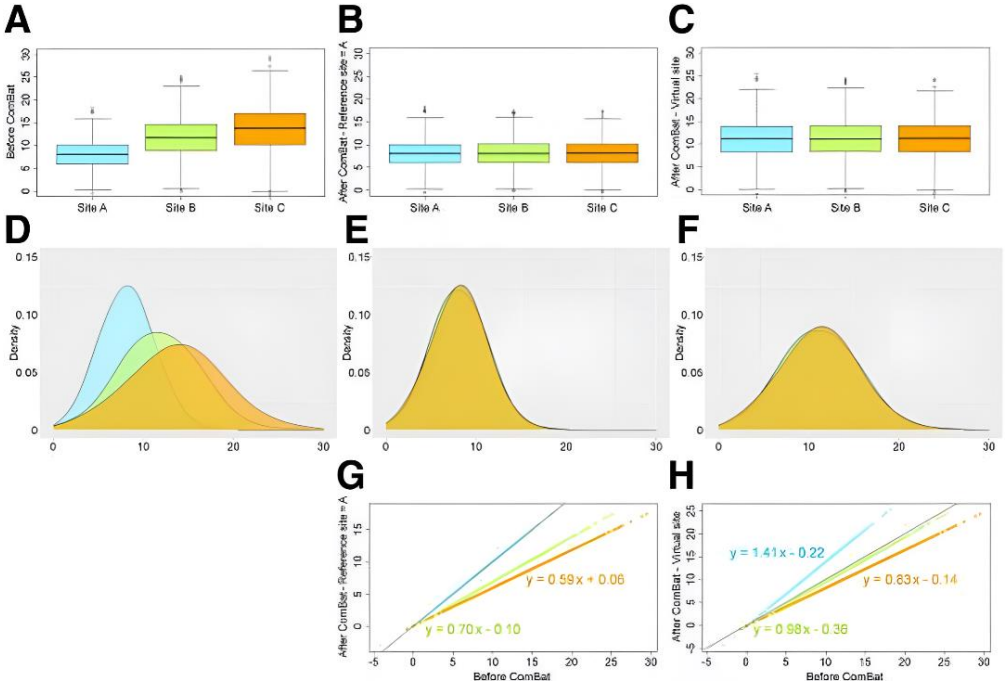

Figure 2. Example of ComBat harmonization with 3 site sources. Box plots and feature value distributions are shown. (A and D) Plots before ComBat. (B, E, and G) Plots after ComBat by aligning data from sites B and C to site A. (C, F, and H) Plots after ComBat by aligning data on virtual site. Bottom graphs show equations of transformations. Source: [47].

ComBat provides an efficient solution for harmonizing radiomic features across imaging sites, though it requires enough data with similar acquisition parameters per batch to reliably estimate corrections. Batch differences reflect technical rather than biological variability, so it must be validated to avoid removing clinically relevant feature variations [30].

4.4.3. Z-Score

Z-score normalization standardizes the intensity distribution of an image by centering it at a mean of 0 and scaling it to a standard deviation of 1. Due to its simplicity and low computational demand, Z-score normalization can be easily applied to either whole images or selected ROIs, making it a practical choice for intensity standardization in medical imaging [48].

$$I(x) = \frac{I(x) - \mu}{\sigma}$$

- $I(x)$ is the original image.
- μ is the mean intensity.
- σ is the standard deviation.

Z-score normalization is simple to implement, as it only requires the voxels within the ROI for standardization. However, it fails to address scanner and protocol-dependent variations, making it insufficient for improving crucial radiomics reproducibility across multi-center studies.

4.5. Feature Selection

In machine learning, feature selection is a critical preprocessing step aimed at identifying and retaining the most relevant subset of features from a dataset. This process is essential for enhancing model performance, reducing computational complexity, and improving model interpretability. Different feature selection methods may identify different features as relevant, and the choice of method can affect the performance of the radiomics model.

4.5.1. Univariate Feature Selection Methods

Univariate feature selection methods evaluate the relevance of each feature independently with respect to the target variable. Among these, Analysis of Variance (ANOVA) is a widely used statistical test.

ANOVA is a statistical method used to determine whether the means of two or more groups are significantly different from each other. It operates under the null hypothesis (H_0) that all group means are equal, while the alternative hypothesis (H_1) states that at least one group mean differs. This method assumes normality, homogeneity of variances, and independent samples. ANOVA can handle categorical independent variables and continuous dependent variables, making it suitable for analyzing medical imaging data and clinical data [49].

4.5.2. Multivariate Feature Selection Methods

Multivariate methods evaluate features collectively, accounting for interactions, dependencies, and redundancies among them to identify the most informative subset with respect to the target variable. Various multivariate feature selection techniques are employed to detect the most relevant group of features:

- **Correlation Matrices:** They are widely used in feature selection to identify relationships between features and reduce redundancy in a dataset. A correlation matrix is a statistical table that measures the relationship between variables, typically calculated using Pearson's correlation for numerical data. When features are highly correlated, they tend to carry similar information, which can lead to redundancy [50].
- **LASSO:** It is an embedded method used with linear regression that adds a special kind of penalty to the model. This penalty helps reduce the size of the coefficients for less important features, often shrinking some of them all the way to zero. When features are similar or strongly correlated, LASSO usually keeps just one and sets the others to zero. After applying LASSO, the features with non-zero coefficients are the selected ones [51].
- **Random Forest:** It works as a bagging method, building multiple unpruned decision trees where, at each split in the tree, the algorithm randomly selects a subset of features to consider. This method can rank features based on their importance, which is measured by how much they reduce impurity across all trees. However, in high-dimensional datasets where only a few features are truly relevant, identifying those key features can be more challenging [51].

4.6. Classification

A classification method is a statistical or computational technique that assigns elements to specific categories or classes based on the observed characteristics of the data.

4.6.1. Traditional Statistical Methods

Traditional statistics, whose origins date back to the 16th century, is based on the use of parametric models and hypothesis testing. This methodology employs predetermined equations, such as linear or logistic regression, to model relationships between variables with deductive reasoning. Classical methods such as logistic regression, linear discriminant analysis, and quadratic discriminant analysis are widely used for classification and prediction tasks.

These statistical approaches require that the data meet certain assumptions, such as normality or independence among predictors, which facilitates the interpretation of results but limits their applicability in complex contexts. However, when faced with highly nonlinear data patterns or multiple interactions, the structural rigidity of traditional methods may become a disadvantage. Their main strength lies in scenarios where phenomena exhibit clear, well-defined relationships that can be modeled through established mathematical formulations [52].

4.6.2. AI Model

Machine learning can be broadly defined as the use of computational methods or models that learn from experience (data) to improve performance or make accurate predictions [53]. To create an accurate Machine Learning model, different types of algorithms can be considered. Various options have been analysed: Support Vector Machine (SVM), Decision Tree, Random Forest, K-Nearest Neighbours (KNN), Naive Bayes, Logistic Regression, and Linear Regression. Below are the pros and cons of each option.

4.6.2.1. Naive Bayes

Naive Bayes is a classification algorithm based on Bayes' theorem, which assumes independence between predictive features. Advantages of Naive Bayes include its simplicity and ease of implementation, the fact that it doesn't require large amounts of training data, its high scalability with the number of predictors and data points, its speed, and its insensitivity to irrelevant features. However, it has disadvantages such as the assumption of feature independence, which rarely holds true in real life, limiting its applicability in practical cases.

4.6.2.2. Support Vector Machines (SVM)

SVM are supervised classification algorithms that aim to find a hyperplane that optimally separates samples from different classes. SVM algorithms are known for being effective in high-dimensional spaces, even when the number of dimensions exceeds the number of samples, and for their efficient memory management, as they use only a subset of points in the decision function. However, they have disadvantages such as being inefficient with large datasets due to slow kernel computations. SVMs also do not provide probability estimates, and their decision boundary directly depends on the closest values—even if they are outliers. Additionally, they are highly dependent on the scale of the data, so proper scaling is required.

4.6.2.3. K-Nearest Neighbours (KNN)

KNN is a very simple machine learning classification algorithm that assigns a class label to a point based on the class labels of its k nearest neighbours, measured by Euclidean distance or another distance metric. The value of k is set beforehand. Advantages of KNN include its simplicity and its ability to perform well with high-dimensional data without a complex training process. It is also easy to implement and understand. However, it has disadvantages such as low efficiency with large datasets—since it must compute the distance to all training points for each new sample—and it is highly sensitive to noise and irrelevant features in the data [54].

4.6.2.4. Decision Tree

A Decision Tree is a classification model that iteratively splits a dataset into more homogeneous subsets, maximizing the purity of the resulting groups. “Purity” refers to how homogeneous a group of data is; a group is purer when it mainly contains data from a single class. Advantages of decision trees include their ease of interpretation and their ability to handle both categorical and numerical data. However, they also have drawbacks such as a tendency to overfit, especially with small datasets, and sensitivity to small changes in the data, which can lead to very different trees.

4.6.2.5. Random Forest

Random Forest is a classification algorithm based on building multiple decision trees from random subsets of the training data. Each tree is trained on a random sample with replacement, and during each node split, only a random subset of predictive features is considered. Each tree votes for the class to be assigned. Advantages of Random Forest include its ability to handle large and complex datasets, its robustness against overfitting thanks to the ensemble of trees, and its strong performance in a wide variety of applications. However, disadvantages include a slight increase in model bias due to the introduction of randomness, and difficulty in interpreting results compared to a single decision tree [55].

4.6.2.6. Logistic Regression

Logistic Regression is a classification algorithm used to determine the probability of success or failure of an event with a binary dependent variable. It learns a linear relationship from the dataset and introduces non-linearity through the sigmoid function. Its advantages include ease of implementation, interpretation, training efficiency, and fast classification of unknown records. Disadvantages include its limitations with non-linear problems, the risk of overfitting with few observations, and difficulty capturing complex relationships in the data.

4.6.2.7. Linear Regression

Linear Regression is an algorithm used for regression tasks. It is primarily used to find the relationship between variables and to make predictions. Advantages of linear regression include its simplicity of implementation and ease of interpreting the output coefficients. It is ideal when the relationship between independent and dependent variables is linear, and it is less complex than other algorithms. However, disadvantages include its susceptibility to outliers, assumptions of linear relationships and independence among attributes, and a

tendency to overfit. Linear regression is useful for analyzing relationships but often oversimplifies real-world problems [56].

4.7. Proposed solution

To predict the progression of head and neck cancer using machine learning techniques, it is crucial to select appropriate data and algorithms. The following table summarizes the chosen option for each step, along with their justifications.

Table 2. Proposed solution table.

Step	Proposed solution	Justification
Segmentation	MIM Software	MIM was chosen because it is the software used in Hospital de Bellvitge, and this segmentation step was carried out during an internship at the hospital.
Image Preprocessing	Converting PET Voxel Values to SUV and Resampling	Standardization to SUV and image resampling are mandatory to ensure data comparability. Other preprocessing techniques did not improve model performance.
Radiomics Features	PyRadiomics package to extract all features	All radiomics features from PyRadiomics were used because they are compliant with the IBSI.
Harmonization	ComBat	Selected as it is a widely established and effective method for harmonizing radiomics data across different scanners and acquisition protocols.
Feature Selection	Univariate and Multivariate Feature Selection	Both univariate and multivariate methods were used to ensure that all relevant and important features are considered, improving model robustness.
Classification	SVM, Random Forest, and Logistic Regression	These models were chosen for their strong performance in classification tasks, interpretability, and ability to handle small datasets well.

5. Detailed Engineering

The pipeline of this project includes several critical stages: image segmentation, preprocessing, radiomic feature extraction, data harmonization, feature selection and final classification. Each step is designed to ensure the accuracy, reproducibility, and robustness of the predictive model.

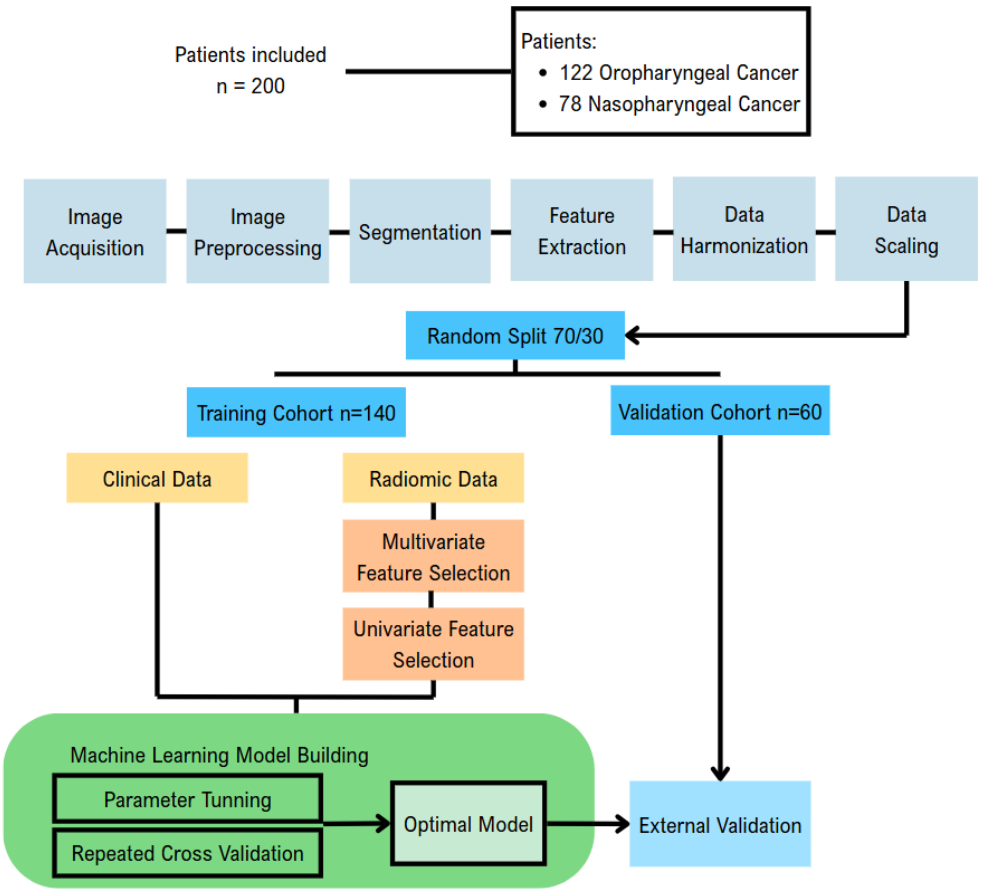

Figure 3. Workflow of the project. (Based on: [57])

5.1. Segmentation

Before to start the project, 200 patient studies had been previously anonymized in accordance with ethical regulations and were readily available for segmentation.

The first step of the project consists of tumour and lymph node segmentation, this process was done during an internship at Hospital Universitari de Bellvitge; it was performed using MIM Software because it is the clinical imaging platform routinely used at Hospital Universitari de Bellvitge.

Among the collected PET/CT scans, the majority comprised a dedicated PET/CT study of the head and neck; with a smaller proportion having only whole-body studies that extended to include the head and neck. When both options were available, head and neck studies were prioritized for segmentation due to their higher spatial resolution and better anatomical detail, which are essential for accurate delineation in this region.

The segmentation process focused on key anatomical targets relevant to prognosis in head and neck cancer. These included the primary tumour, a second tumour if present, the largest metastatic lymph node, and any additional involved lymph nodes when identifiable. The initial step of

segmentation was performed using a fixed threshold approach, applying a Standardized Uptake Value (SUV) cutoff of 3. This method isolates hypermetabolic regions by selecting voxels with SUV values equal to or greater than 3, a commonly used baseline to distinguish malignant from physiologic uptake in oncologic PET imaging.

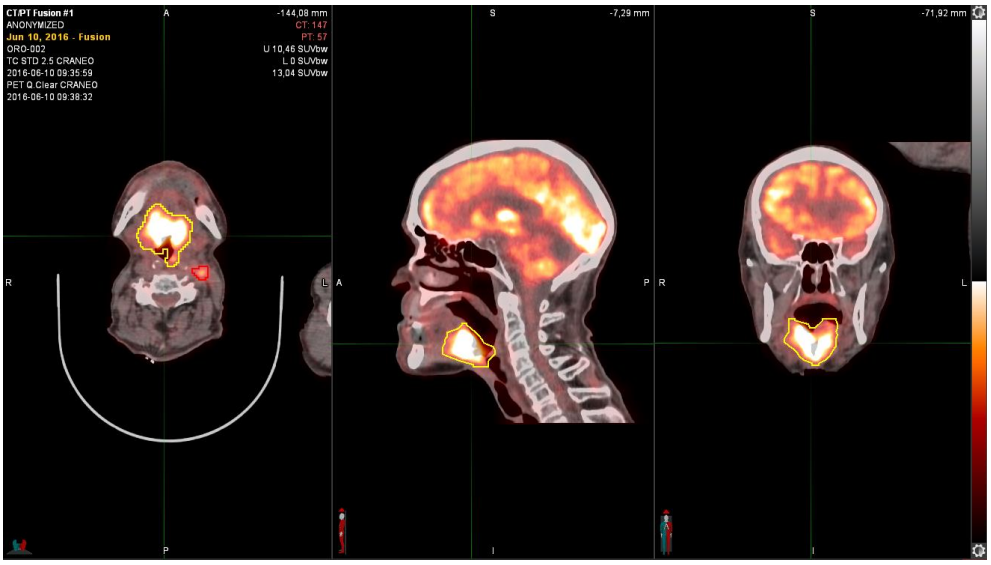

Figure 4. Segmentation of ROIs by SUV threshold of 3.

Following this initial thresholding, only the automatically generated contours corresponding to the anatomical regions of interest were retained, while non-relevant areas were discarded. To improve segmentation accuracy and better define lesion boundaries, a secondary refinement step was applied using an adaptive threshold set at 41% of the maximum SUV (SUVmax) of each lesion. The 41% threshold method is a reliable approximation of the metabolically active tumour volume, helping to mitigate partial volume effects and standardize lesion delineation across patients [58].

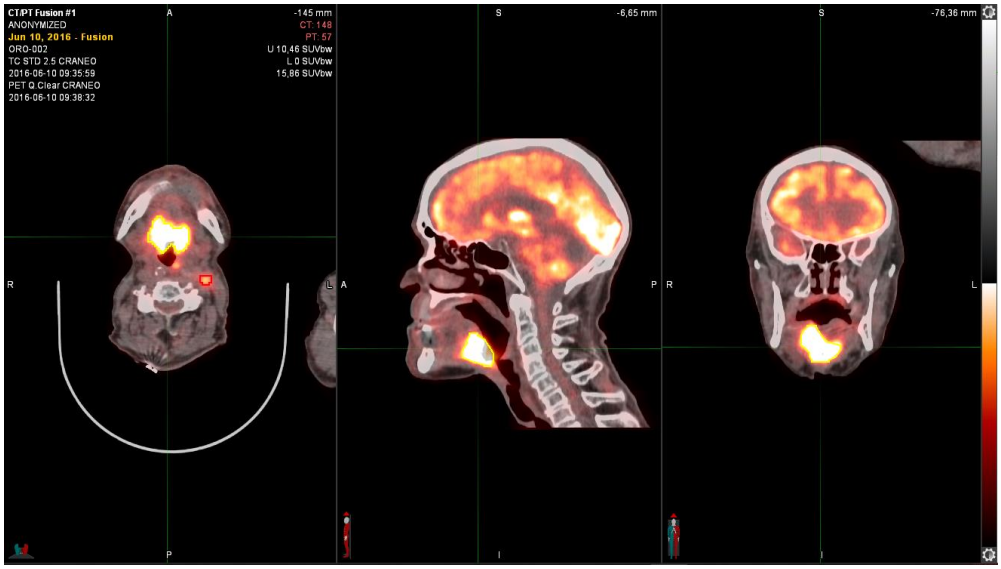

Figure 5. Redefinition with the 41% decay of each lesion.

All contours generated through automated methods were subsequently reviewed and manually adjusted by expert nuclear medicine physicians. Manual corrections were applied when necessary to address inaccuracies such as overextension into adjacent structures or under-segmentation of

lesions with diffuse uptake. Finally, a composite contour encompassing all tumour and nodal regions was created, forming a single total tumour volume.

5.2. Image Preprocessing

All PET/CT images underwent preprocessing to ensure consistency across scans. This included resampling to a uniform voxel size and applying intensity normalization to reduce inter-scanner variability and standardize radiomic feature extraction.

5.2.1. Read DICOM Files

The PET/TC studies and the segmentations performed in MIM Software were saved in DICOM format, the regions of interest specifically using the RT Structure standard. DICOM (Digital Imaging and Communications in Medicine) is the international standard for handling, storing, and transmitting medical imaging information. It allows for the integration of image data and associated metadata, such as patient information, acquisition parameters, and spatial orientation, in a standardized and interoperable format.

The RT Struct file is a particular DICOM object used to encode contours of segmented anatomical structures. It is important to note that RT Struct files do not store the segmentations as volumetric masks or voxel-based data. Instead, they represent the segmentation as a collection of planar closed polygonal contours (also called structure sets) that are drawn over specific slices of the original image. These structures are defined relative to the imaging coordinate system, and each contour corresponds to a specific axial slice of the associated CT or PET scan.

The DICOM series for each study were identified by recursively searching directory paths for DICOM files associated with a specific imaging study. The RT Struct file was identified by scanning for the presence of the *ROIContourSequence* tag within the DICOM metadata. Only one RT Struct file was retained per study, and a warning was issued in cases of multiple matches.

Because RT Struct files store only a geometric representation of the segmented regions, a reconstruction process is required to convert these planar contours into 3D binary masks or volumetric arrays suitable for computational analysis.

5.2.2. Construction of Segmented Structures

Once the appropriate RT Struct files were loaded, the names and indices of all available ROIs were retrieved from the DICOM data. The corresponding contour datasets were extracted, each contour is represented by a set of 3D physical coordinates (in millimeters), which were then converted into 2D pixel coordinates by referencing the appropriate DICOM image slice via its image identificator. This mapping required reading the associated image file and using the pixel spacing and image origin metadata to perform coordinate transformation from physical to image space.

The contours were then rasterized into binary mask, which converted polygonal coordinates into filled 2D Boolean arrays. All masks were stored in a dictionary indexed by image identifier, allowing for alignment with the corresponding DICOM image slices.

To ensure consistent spatial alignment, image slices were sorted in ascending order based on their z-axis positions. Image and mask volumes were constructed by stacking the 2D slices in the sorted order, creating 3D volumes suitable for further analysis.

The entire processing step was developed and executed using Python, leveraging libraries such as Pydicom, Numpy, and PIL for DICOM parsing, numerical operations, and image manipulation. As a result, 3D binary mask matrices corresponding to the segmented regions were successfully generated.

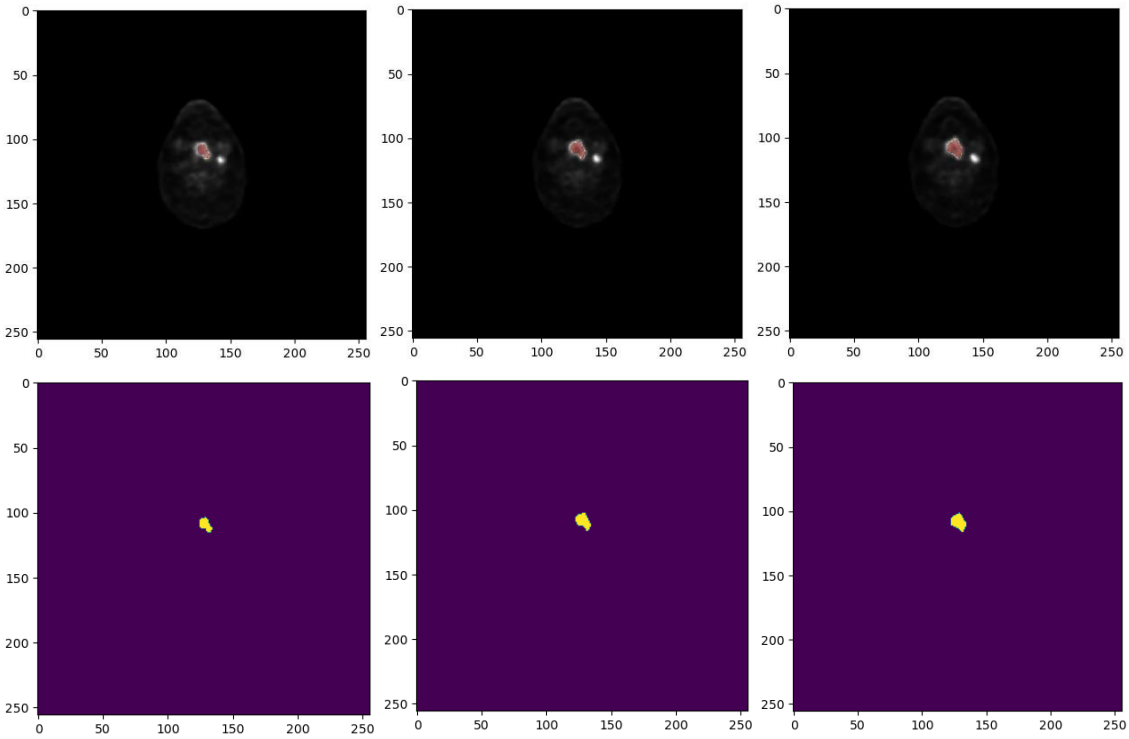

Figure 6. Reconstructed 3D binary mask matrices

5.2.3. Converting PET voxel values to SUV

To convert PET voxel values from arbitrary intensity units to standardized uptake values (SUVs), a Python-based pipeline was implemented. This process ensures the resulting SUVs are quantitatively meaningful and comparable across patients and time points. The approach was implemented using NumPy for array-based operations and follows established clinical standards for SUV computation.

First, patient weight was extracted from the DICOM header and converted from kilograms to grams. If the value was missing or zero, a default value of 70 kg was assumed. The scan start time was obtained from the SeriesTime field and converted in seconds.

To compute the decay-corrected injected dose, information from the DICOM data was retrieved, including the radionuclide half-life and the total injected dose. The algorithm adjusts for radioactive decay from the injection time to the scan time using the standard exponential decay formula:

$$decayed\ dose\ = injeted\ dose\ \times \exp\left[-\ln(2) \times \frac{(tscan - tinjection)}{half - life}\right]$$

Finally, SUV values were computed voxel-wise by multiplying the PET image by the patient's weight and dividing by the decay-corrected dose:

$$SUV = \frac{(PET \times Weight)}{decayed\ dose}$$

This method yields a three-dimensional matrix of SUV values, which provides a normalized metric of radiotracer uptake, essential for quantitative PET analysis.

5.2.4. Resampling

To ensure consistency in spatial resolution across PET studies acquired with different scanners and anatomical protocols (head and neck and whole-body imaging), volumetric data were spatially normalized. Given the heterogeneity of image dimensions, particularly in the axial (x) and coronal (y) planes, a conditional resampling step was implemented. Specifically, a function was used to reshape the 3D by interpolating it to a fixed resolution of 256×256 pixels (corresponds to the transversal resolution of the Discovery IQ scanner), while preserving the original number of slices in the z-direction. This rescaling was performed using cubic interpolation, implemented to preserve anatomical detail and avoid aliasing artifacts.

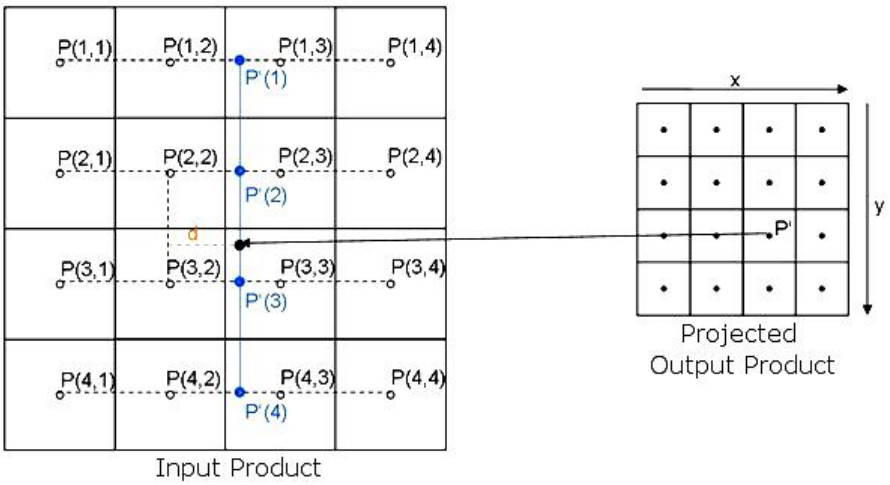

Figure 7. Cubic convolution representation [59].

5.2.5. Morphological Refinement

Given the potential for multiple disconnected components resulting from automatic or semi-automatic segmentation algorithms, a filtering procedure was employed to retain only the largest contiguous segmented region. This was achieved by labelling connected components, computing the voxel count for each labelled region and retaining only the region with the maximum volume. This strategy ensured that spurious or fragmented regions were excluded from further analysis, thereby improving the anatomical fidelity of the segmentation and minimizing downstream errors.

To enhance anatomical precision in the segmentation, non-relevant regions from the masks were removed by superimposing the corresponding CT onto the PET images and applying a threshold of -500 Hounsfield Units to the CT scan. In particular, segmented regions that extended into exterior parts of the body, such as superficial skin-adjacent voxels, were systematically removed. This correction is critical in cases where the tumour

is located near the body's surface, as misclassifications can erroneously include background or air-adjacent voxels. Additionally, given the anatomical context of nasopharynx and oropharynx cancers, regions of segmentation that overlapped with air-filled were eliminated. These areas, though spatially contiguous with tumour regions, do not contain tissue of interest and can introduce noise into quantitative analyses.

5.3. Radiomics Feature Extraction

Radiomic feature extraction was subsequently performed to quantitatively characterize the segmented lesions. This process involves computing a wide range of features that describe the intensity, shape, and texture patterns within the segmented volumes.

5.3.1. Pyradiomics Function

Radiomic features for each lesion type were extracted using the open-source Python package PyRadiomics, which provides a standardized framework for high-throughput feature extraction from medical images.

PyRadiomics input consisted of the pre-processed PET images and the corresponding binary segmentation masks, which delineated the regions of interest. It also provides a wide range of configurable parameters that allow users to tailor the feature extraction process to their specific data and research needs. The table below summarizes the most relevant customizable parameters available in the PyRadiomics extraction configuration:

Table 4. PyRadiomics parameter configuration used for radiomic feature extraction [60].

Parameter	Description	Default Value	Selected Value
normalize	Boolean. Enable normalization of the image prior to any resampling.	False	False
resampledPixelSpacing	List of 3 floats. Defines the spacing to resample the images to isotropic voxels	None	None
minimumROIDimensions	Integer (1–3). Minimum required dimensions of the segmented region.	2	2
minimumROISize	Integer. Minimum number of voxels required in the ROI	1	1
correctMask	Boolean. Attempts to realign the mask to image geometry if mismatch occurs.	False	False
LoG	Laplacian of Gaussian filter. Enhances edges by emphasizing grey level changes.	False	False
sigma	List of floats. Defines scale levels for Laplacian of Gaussian filter.	[1.0, 3.0, 5.0]	[1.0, 3.0, 5.0]
gradientUseSpacing	Boolean. Accounts for image spacing in gradient magnitude computation.	True	True

label	Integer. Label value used to identify the Region of Interest (ROI) in the mask.	1	1
binWidth	Float > 0. Sets the width of histogram bins used for gray level discretization.	25	**0.1**
distances	List of integers. Specifies pixel distances for computing texture matrices like GLCM.	[1]	[1]
symmetricalGLCM	Boolean. Generates symmetrical GLCMs considering both i→j and j→i transitions.	True	True
gldm_a	Float. Alpha cutoff for gray level dependence in GLDM.	0	0
kernelRadius	Integer. Radius used to define the kernel size for local texture features.	1	1
glcmf	Boolean. Enable extraction of Gray Level Co-Occurrence Matrix features.	True	True
glrlmf	Boolean. Enable extraction of Gray Level Run Length Matrix features.	True	True
glszmf	Boolean. Enable extraction of Gray Level Size Zone Matrix features.	True	True
ngtdmf	Boolean. Enable extraction of Neighbouring Gray Tone Difference Matrix features.	True	True
gldmf	Boolean. Enable extraction of Gray Level Dependence Matrix features.	True	True

The parameters selected were optimized to extract radiomic features specifically from PET images of head and neck lesions, where ROIs are typically small. A narrow bin width (0.1) was chosen to preserve subtle texture variations.

5.3.2. Extracted Features

A comprehensive set of radiomic features was computed. From each lesion 350 features were extracted, including first-order statistics, shape descriptors, and higher-order texture features derived from gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), neighboring gray tone difference matrix (NGTDM), and gray level dependence matrix (GLDM). In addition to the radiomic features, several basic lesion characteristics, such as voxel-based and physical size, and the centroid coordinates in x, y, and z, were calculated, along with the distances from the SUVmax location to the lesion centroid and from the SUVmax location to the nearest point on the lesion perimeter. All features were extracted in compliance with IBSI guidelines,

ensuring reproducibility and comparability across studies. The complete feature set was saved in CSV files, with one file generated per lesion type to facilitate organized data analysis.

5.4. Harmonization

The harmonization of radiomic features was conducted using the ComBat algorithm, a well-established method originally developed for genomics data but now widely applied in radiomics to correct for batch effects introduced by different imaging equipment or protocols. In this study, the harmonization process aimed to mitigate the variability arising from the use of different PET scanners (Discovery ST, Discovery IQ and Discovery MI), ensuring comparability of radiomic features across datasets.

The workflow began with the compilation of datasets from two lesion sites: oropharyngeal and nasopharyngeal. Each dataset integrated radiomic features from tumours and lymphadenopathies. After extensive data cleaning, the datasets were merged to create a unified matrix for harmonization.

Prior to applying ComBat, the numeric features were isolated and transposed into a matrix format suitable for the algorithm. Variables with excessive missingness or uniform values were filtered out to preserve statistical robustness. The ComBat model was then fit using a simplified design matrix, treating scanner type as the batch variable.

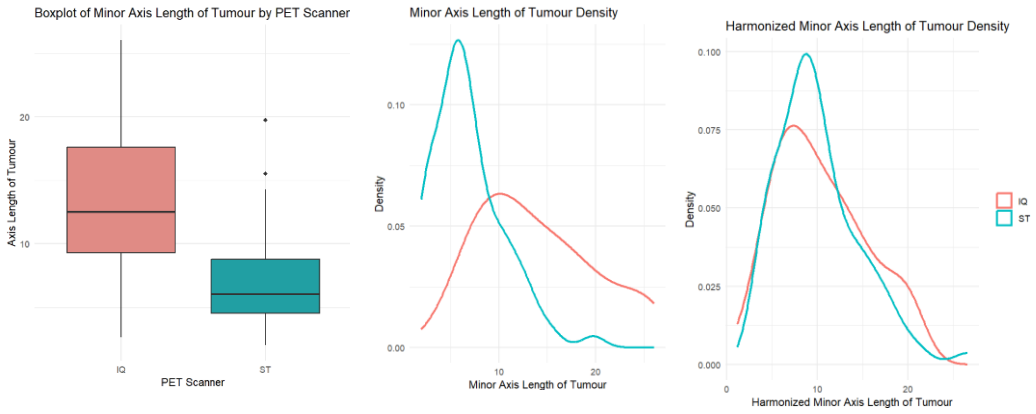

Figure 8. Example of harmonization outcome of Minor Axis Length of Tumour feature.

Finally, the harmonized dataset was saved as a CSV file for downstream analysis. This approach ensured that radiomic features retained their biological relevance while minimizing technical variation.

In accordance with the recommendations of the IBSI, downstream analyses were also performed using the original, non-harmonized radiomic data to assess the impact of batch effects. This evaluation revealed a pronounced batch-related bias, with features exhibiting significant distributional differences across scanner types. Such heterogeneity errors in univariate statistical tests and severely compromised the generalizability of machine learning models. Predictive performance deteriorated markedly when training and testing were conducted across different scanner domains, underscoring the necessity of harmonization.

5.5. Feature selection

To improve model performance and reduce overfitting, a feature selection process was implemented to identify the most informative and non-redundant radiomic features. Prior to this step, a feature scaling process was applied to normalize the range of values across variables, which is essential to ensure that models sensitive to feature magnitudes, such as Logistic Regression and Support Vector Machines. Additionally, the dataset was partitioned into training and testing subsets using a 70:30 split, ensuring that feature selection and model evaluation were performed on independent data.

5.5.1. Data Integration

A data integration process was performed to consolidate three heterogeneous data sources: clinical variables extracted from the electronic health record, including demographic and diagnostic information, as well as other relevant patient characteristics such as age, sex, and history of previous diseases; metabolic features derived from segmented PET images, obtained from MIM Software segmentations, which required custom scripts to merge per-lesion metabolic data into a unified patient-level dataset; and harmonized radiomic features, harmonized to mitigate scanner-related batch effects.

This integrative framework enabled a multidimensional representation of each patient, facilitating robust statistical analysis and machine learning modelling.

5.5.2. Multivariate Feature Selection

Multivariate feature selection was conducted on metabolic and radiomic datasets to optimize the quality and relevance of the variables used in subsequent modelling. The objective was to reduce redundancy, mitigate multicollinearity, and enhance model stability and interpretability. Correlation matrices were computed quantifying linear dependencies between features.

A stringent correlation cutoff threshold of 0.8 was applied; when two or more features were found to be collinear above this threshold, only one representative variable was retained, based on relevance. The correlation matrix after feature selection is presented in the following figure.

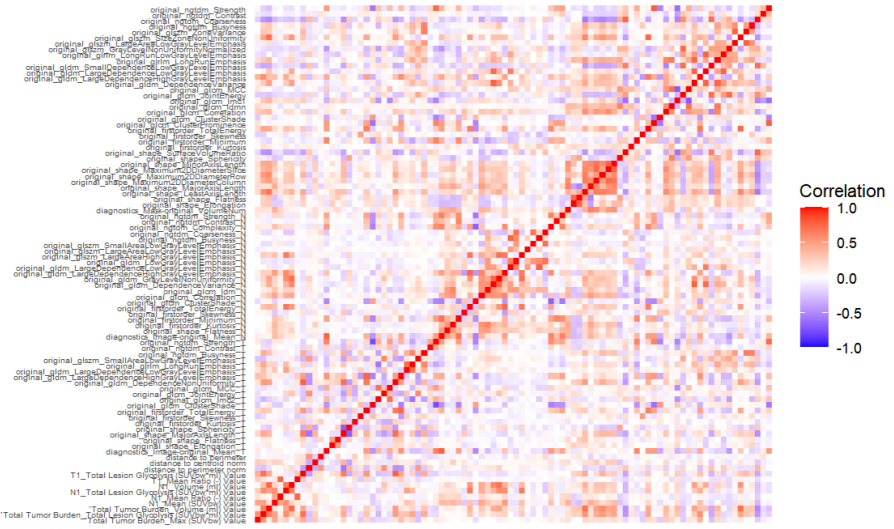

Figure 9. Correlation Matrix after cutoff.

5.5.3. Univariate Feature Selection

Following the multivariate feature selection, ANOVA was conducted to further refine the feature set by evaluating the statistical significance of each variable in relation to clinical outcomes.

Two separate ANOVA tests were performed, each targeting a distinct clinical endpoint: recurrence and overall survival (exitus). For each outcome, the features were assessed individually to determine whether their distributions differed significantly between patient groups defined by the presence or absence of the event. The three variables with the lowest p-values were selected for predicting recurrence, while the variable with the lowest p-value was chosen for predicting exitus. In survival prediction, it is important to select fewer variables to reduce the risk of overfitting due to the limited number of positive cases. The exact p-value is not required; rather, it serves as a criterion for feature reduction.

As a result, two subsets of features were obtained, one optimized for predicting recurrence and another for predicting mortality. In each subset, the most clinically relevant variables for prediction were included alongside the statistically selected. These features served as inputs for the subsequent predictive modelling, ensuring both clinical relevance and statistical rigor with respect to the targeted endpoints.

The selected features for predicting recurrence were: p16 presence, alcoholism, sphericity of the tumour, total lesion volume (ml), smoker status and maximum 2D diameter.

The selected features for predicting mortality were: p16 presence, alcoholism and sphericity of the tumour.

5.6. Classification

For the classification of head and neck cancer recurrence and mortality, supervised machine learning algorithms were applied using the selected radiomic, metabolic, and clinical features. Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR) were used as classifiers. To address class imbalance and ensure adequate learning from both positive and negative outcomes, a weighted method was employed during model training. Additionally, repeated cross-validation (10 repeats of 7-fold cross-validation) was implemented to tune hyperparameters and select the optimal model.

5.6.1. Statistical analysis

Statistical analysis was employed to explore and visualize the relationship between the selected features and clinical outcomes. This included assessing how individual variables correlated with recurrence and survival, providing insight into their potential predictive value. Visual tools such as boxplots and bar plots were used to highlight differences in feature distributions across outcome groups. In the following figures, you can observe the relationship of some related features with the outcome.

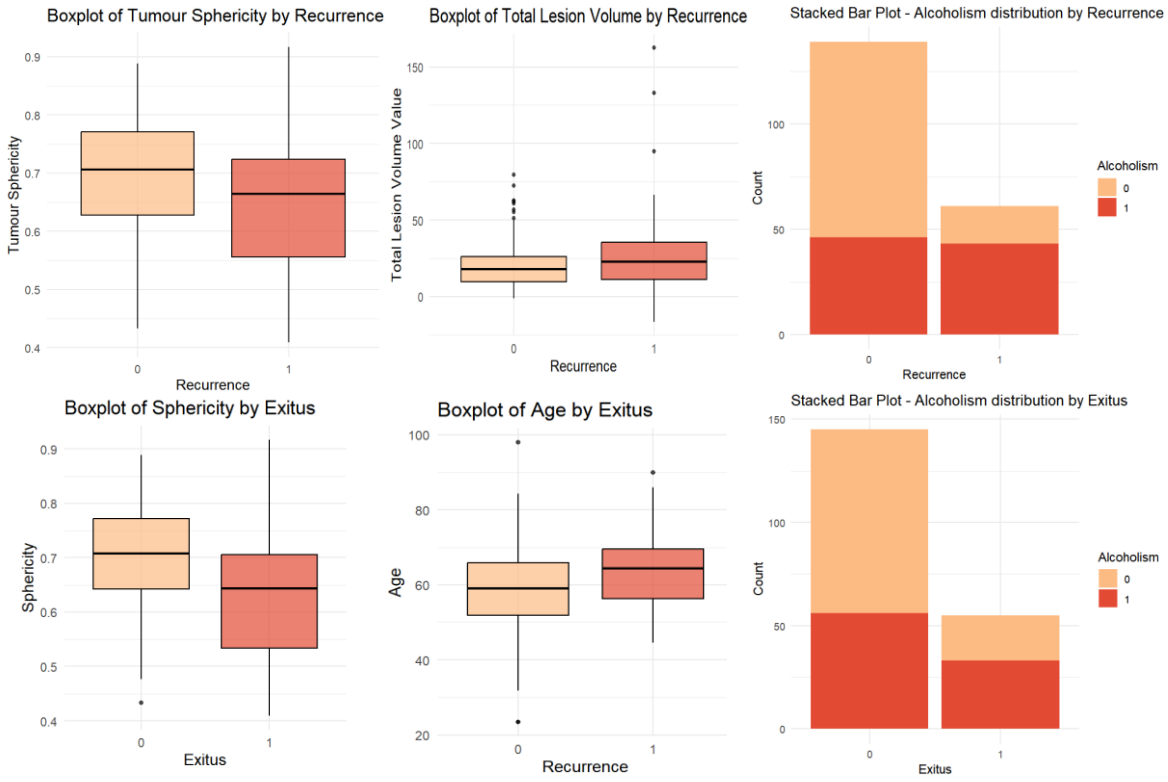

Figure 10. Boxplots of 3 significant features related with recurrence and 3 related with exitus.

5.6.2. Random Forest

Random Forest classifiers were trained to predict both recurrence and exitus outcomes using the selected features. Hyperparameter tuning was conducted by optimizing the number of variables randomly sampled at each split (mtry) by the best AUC value, obtaining the best model with mtry of 1 for predicting recurrence and death.

Following cross-validation and parameter optimization, the best-performing RF models for both outcomes were retrained on the complete training dataset and subsequently evaluated on the held-out test set.

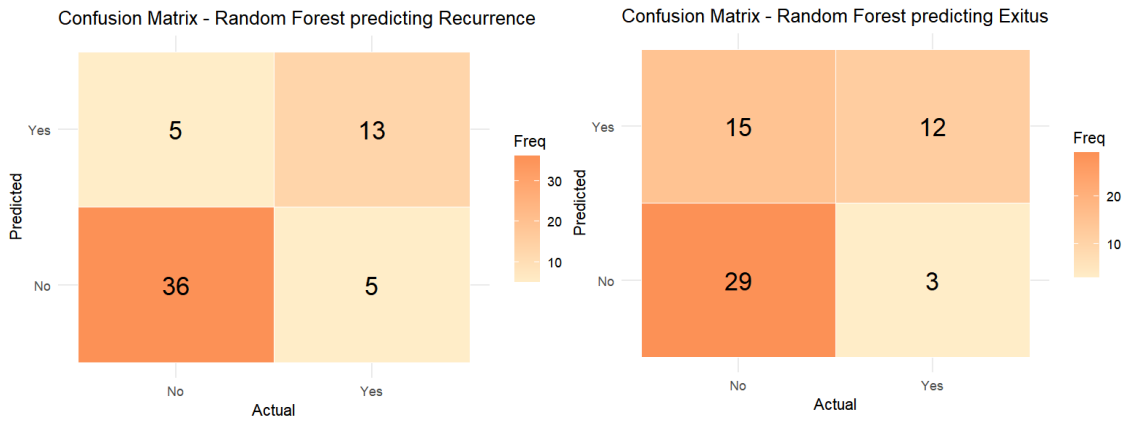

Figure 11. Confusion matrices for Random Forest models predicting recurrence and exitus.

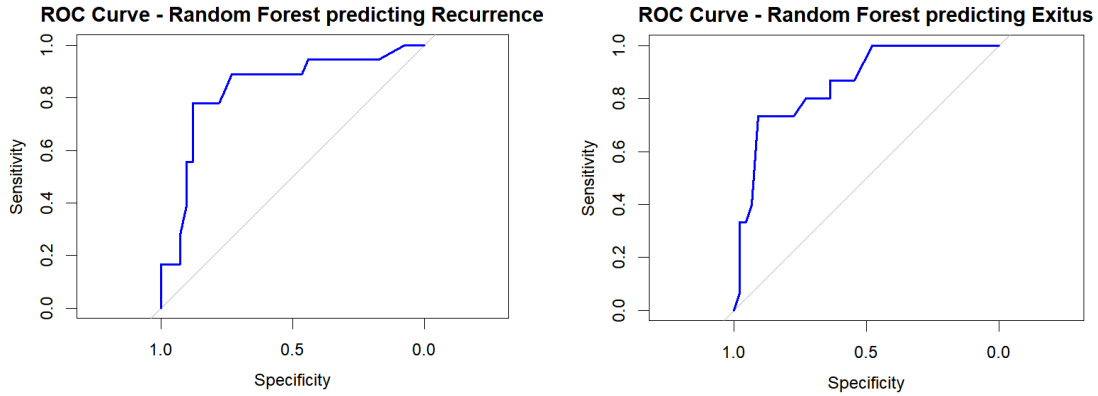

Figure 12. ROC curves for Random Forest models predicting recurrence and exitus.

Performance metrics, including accuracy, sensitivity, specificity, and area under the ROC curve (AUC), were computed to assess model discriminative capacity. The results demonstrated that the optimized models were capable of predict patients with and without recurrence or death with moderate performance.

Table 5. Performance metrics for Random Forest models.

Metric	Recurrence Value	Exitus Value
Sensitivity	0.72	0.66
Specificity	0.88	0.80
Accuracy	0.83	0.70
AUC	0.83	0.86

To enhance model interpretability, feature importance was extracted from the final RF models. This analysis revealed the most influential features contributing to each predictive task. In both models, radiomic shape features emerged among the top predictors, clinical variables such as alcoholism or p16 presence also showed significant relevance, and the variable of the type of tumour was important to differentiate their different behaviour.

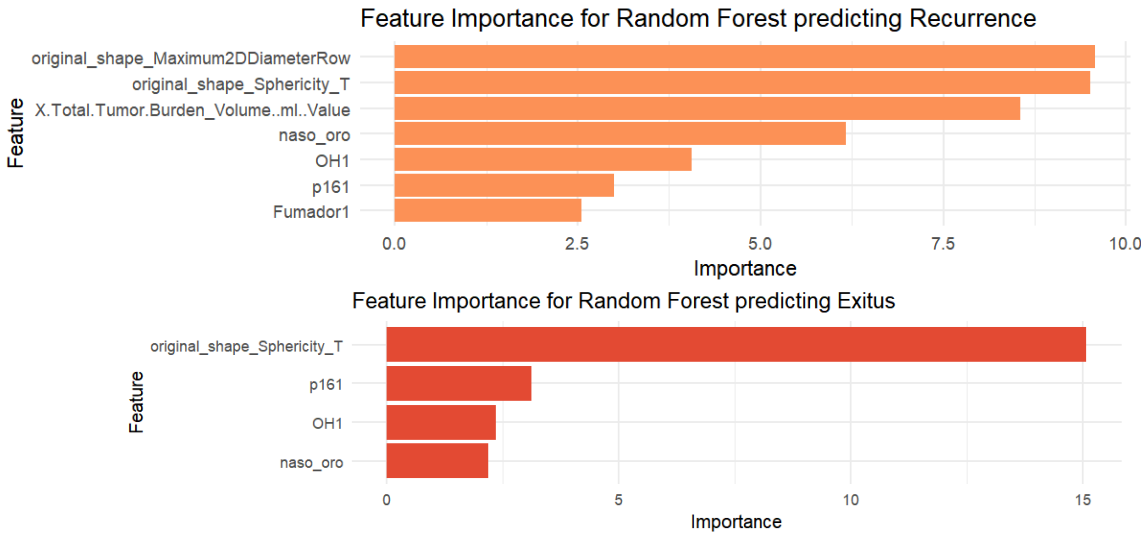

Figure 13. Feature importance plots for random forest models predicting recurrence and exitus,

5.6.3. Support Vector Machine Model

Support Vector Machine (SVM) models were also developed to predict both recurrence and exitus. During this process, hyperparameter tuning was performed for the regularization parameter C and the kernel width parameter sigma, both of which influence the model's ability to delineate complex decision boundaries. A radial basis function kernel was employed to capture potential non-linear relationships among the features.

Optimal parameter configurations (sigma = 0.01 and C = 10 for recurrence, sigma = 0.05 and C = 0.1 for exitus) were identified and the final SVM models were retrained using the full training dataset and subsequently evaluated on the independent testing set.

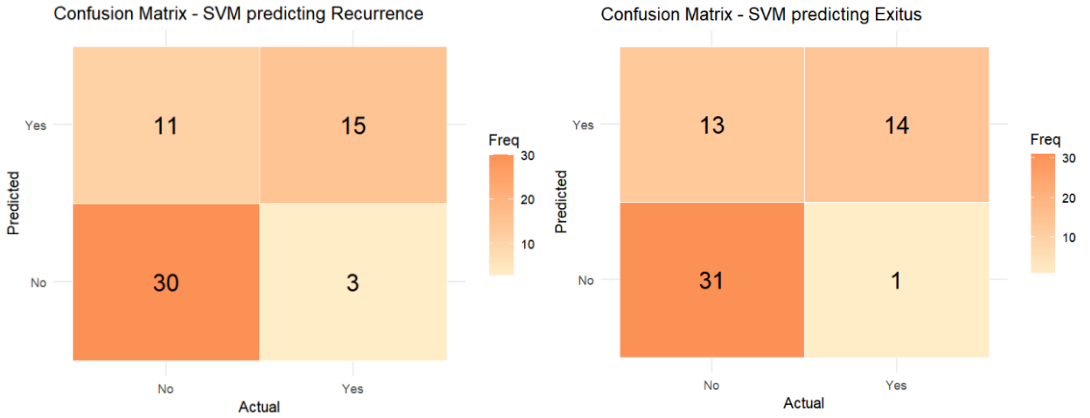

Figure 15. Confusion matrices for SVM models predicting recurrence and exitus.

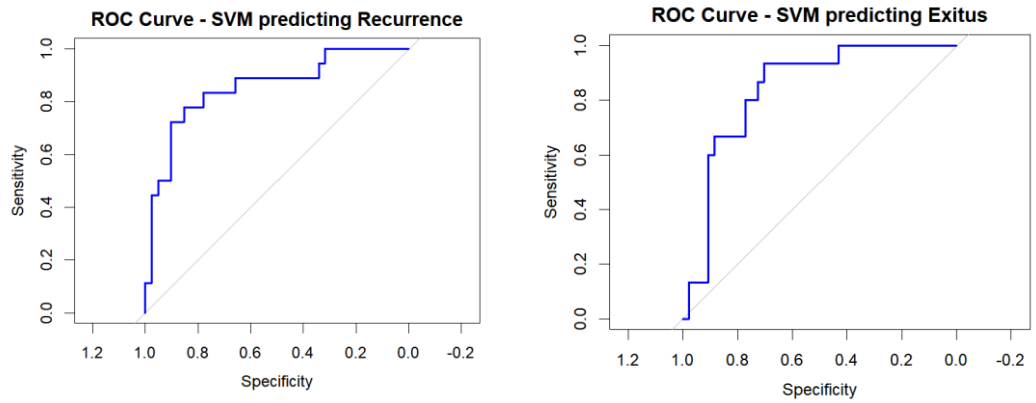

Figure 14. ROC curves for SVM models predicting recurrence and exitus.

The SVM models exhibited moderate classification performance, contingent on the specific outcome variable. For recurrence prediction, the model demonstrated high sensitivity and AUC. In the case of exitus prediction, the model also achieved elevated sensitivity with relatively diminished specificity.

Table 6. Performance metrics for SVM models.

Metric	Recurrence Value	Exitus Value
Sensitivity	0.73	0.93
Specificity	0.83	0.70
Accuracy	0.76	0.76
AUC	0.85	0.84

5.6.4. Logistic Regression Model

Logistic Regression (LR) models were implemented to predict head and neck cancer outcomes. The LR models were trained using the generalized linear model framework ("glm") without hyperparameter tuning. After training, the models were evaluated on the independent test set, derived from a prior 70:30 training-test split.

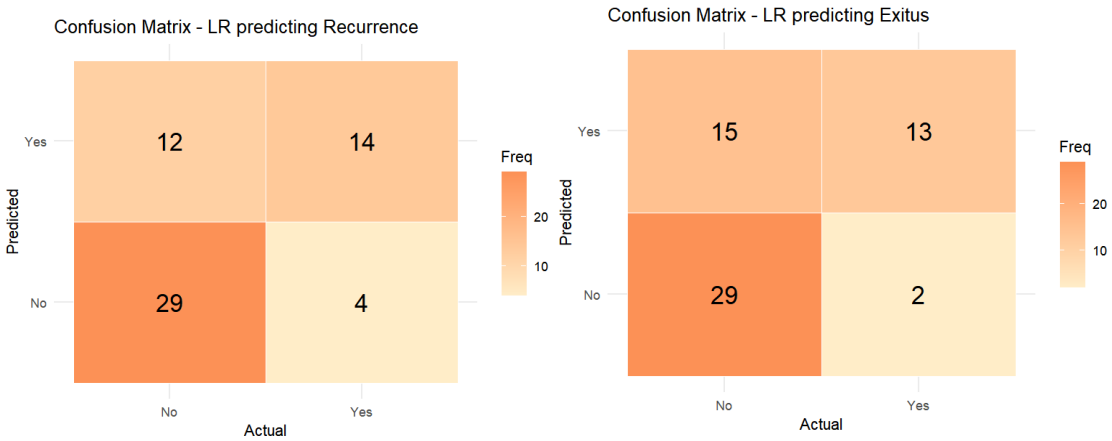

Figure 17. Confusion matrices for Logistic Regression models predicting recurrence and exitus.

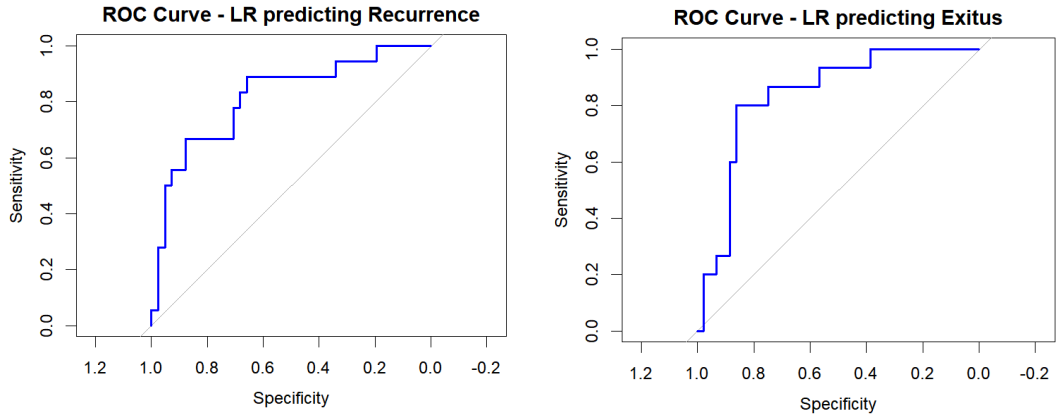

Figure 16. ROC curves for Logistic Regression models predicting recurrence and exitus.

The LR models exhibited distinct predictive performance across the two clinical outcomes. While both models achieved a high area under the ROC curve, the LR model demonstrated greater sensitivity in predicting recurrence and higher specificity in predicting exitus.

Table 7. Performance metrics for Logistic Regression models.

Metric	Recurrence Value	Exitus Value
Sensitivity	0.71	0.66
Specificity	0.78	0.87
Accuracy	0.73	0.71
AUC	0.82	0.84

5.7. Discussion

The present study outlines a comprehensive and reproducible pipeline for radiomic analysis in head and neck cancer, integrating clinical, metabolic, and radiomic features. The process encompasses segmentation, feature extraction, harmonization, feature selection, and classification.

Segmentation represents a key step in the radiomics workflow. Inaccurate or inconsistent delineation of lesions introduces significant variability. An undefined or unstandardized segmentation pipeline can lead to poorly generalizable models, affecting both the reproducibility of features and the robustness of downstream analyses.

Post-processing corrected variability from PET scanners and radiotracer doses, as differences in resolution and acquisition parameters can compromise radiomic feature integrity. Standardizing voxel spacing and intensity normalization helps stabilize features across datasets. Additionally, ComBat harmonization effectively removed scanner-induced batch effects, which otherwise led to biased analyses and reduced model performance.

Feature importance analysis from the Random Forest models revealed shared predictors for recurrence and exitus, reflecting the clinical relationship between these outcomes, since exitus is often associated with tumour recurrence. The prediction of exitus involved fewer features, likely due to the lower number of positive cases in the dataset. Shape features, particularly sphericity, emerged as one of the most influential predictors across both outcomes. Furthermore, the integration of radiomic and clinical data significantly enhanced the models' predictive performance, underscoring the value of a multimodal approach in improving prognostic accuracy.

Clinically recognised variables, such as nationality, known to be relevant in nasopharyngeal cancer, offered limited predictive value in our study, likely due to the low representation of non-local patients in the Barcelona cohort. Although cancer type remained an important variable, likely due to the divergent clinical trajectories and treatment responses associated with different tumour sites.

The performance of the machine learning models revealed important distinctions in their predictive capabilities for recurrence and exitus outcomes. For recurrence, the Random Forest model achieved the highest accuracy (0.83) and specificity (0.88), while the SVM obtained the highest sensitivity (0.73) and AUC (0.85). Although Logistic Regression showed slightly lower performance (AUC of 0.82), it remained consistent across metrics. The AUC, as the most relevant indicator of overall model discrimination, suggests that both RF and SVM offer reliable predictions. Depending on the clinical objective, models can be adjusted to prioritise specificity, reducing unnecessary follow-ups, or sensitivity to minimise missed recurrences.

In contrast, predictive performance for exitus showed lower precision. The SVM model stood out with a notably high sensitivity (0.93) and balanced AUC (0.84), making it suitable for identifying high-risk patients. RF, while slightly less sensitive (0.66), showed a strong AUC (0.86) and higher specificity (0.80), indicating utility in more conservative clinical settings. LR performed similarly, with low sensitivity (0.66), and an AUC of 0.84.

Future improvements to this pipeline could include the use of deep learning for automated segmentation, more advanced harmonization approaches, and external validation on independent multicenter datasets.

6. Execution Schedule

To plan a project adequately, it is necessary to define each of its phases, from preparation or execution to results. The Work Breakdown Structure (WBS) provides a global visualization of the essential task groups that need to be performed to carry out the project, allowing for the development of a timeline.

Table 8. Definition of work blocks and project tasks.

Project Phases	Task ID	Task Name
1. Documentation	1.1. 1.2. 1.3.	Research Written Report Presentation
2. Segmentation	2.1. 2.2.	Patient Upload Segmentation
3. Image Preprocessing	3.1. 3.2. 3.3. 3.4.	Open DICOM Files Construction of Segmented Structures Converting PET Voxel Values to SUV Resampling
4. Radiomics Extraction	4.1. 4.2.	Extract Radiomics Features Save Radiomics Features
5. Harmonization	5.1. 5.2.	Harmonization with ComBat Save Harmonized Features
6. Feature Selection	6.1. 6.2. 6.3.	Multivariate Feature Selection Merge Clinical Data Univariate Feature Selection
7. Classification	7.1. 7.2. 7.3. 7.4.	Statistical Analysis Random Forest Model SVM Model Logistic Regression Model
8. Validation	8.1. 8.2. 8.3.	Segmentation Validation Classification Validation Total Validation

Below is the detailed hierarchical structure of the project's WBS:

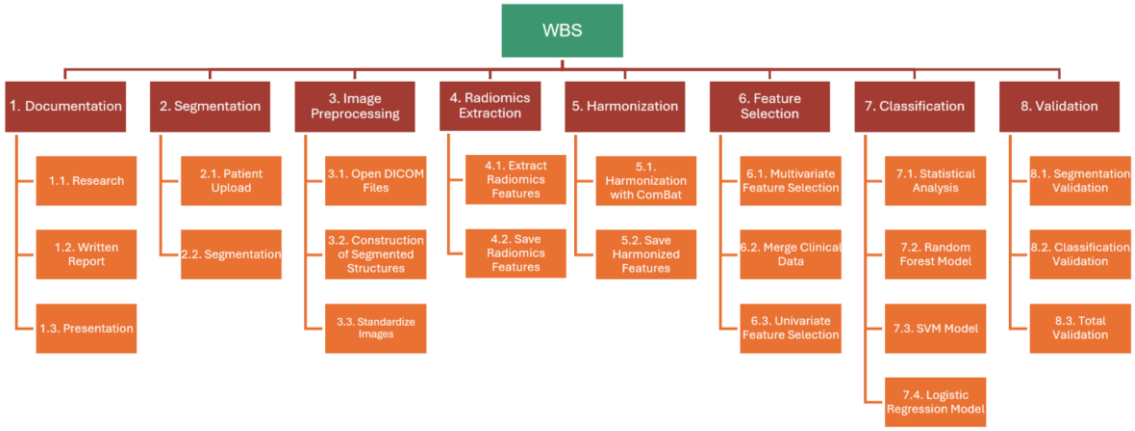

Figure 18. Detailed Work Breakdown Structure (WBS) of the Project.

In total, 22 tasks have been identified. In the annexes, the previously mentioned tasks are defined; a detailed description will be provided, including the required content, deliverables, necessary resources, estimated cost, and duration of each task.

The PERT/CPM diagram shows the time involvement and coordination of the different tasks, indicating their respective durations and interdependencies. This can help identify the critical path, where any delay could impact the final delivery schedule. With this knowledge, it is possible to plan the project workflow more efficiently. As shown in the next figure, even though the report is the largest task, it is written during and after the practical work. This makes documentation part the critical path and needs special attention.

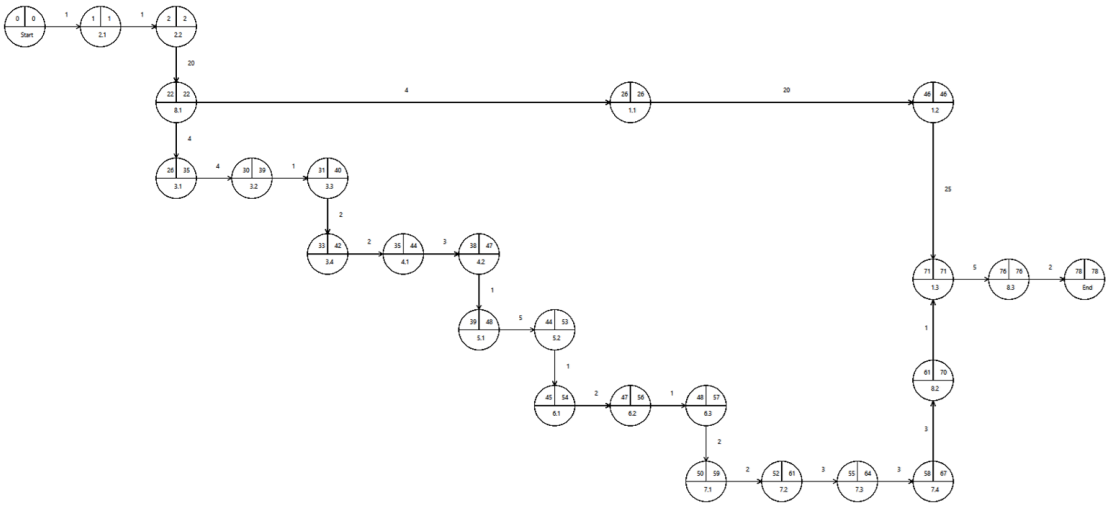

Figure 19. PERT-CPM diagram.

In addition to the PERT/CPM diagram, the last tool used to organize the project was the Gantt chart, which allows for a clearer view of the timeline of each task. In this diagram, each task was assigned to the time specified in the WBS dictionary. This chart shows when tasks should happen and when they should be finished. The relationships between the different tasks were specified using connecting lines. The segmentation process was carried out during practical sessions; because of that, it was the first task to be completed. This information can be seen in in the following diagram, thus providing a complete understanding of the project schedule and task interdependencies.

A color-coding system was used to differentiate tasks based on whether they are critical or not:

1. Critical tasks: These have no slack and cannot be delayed, as this would mean missing the project's final deadline.

2. Non-critical tasks: These have some slack, meaning they can take slightly longer to complete without causing major impact on the overall project.

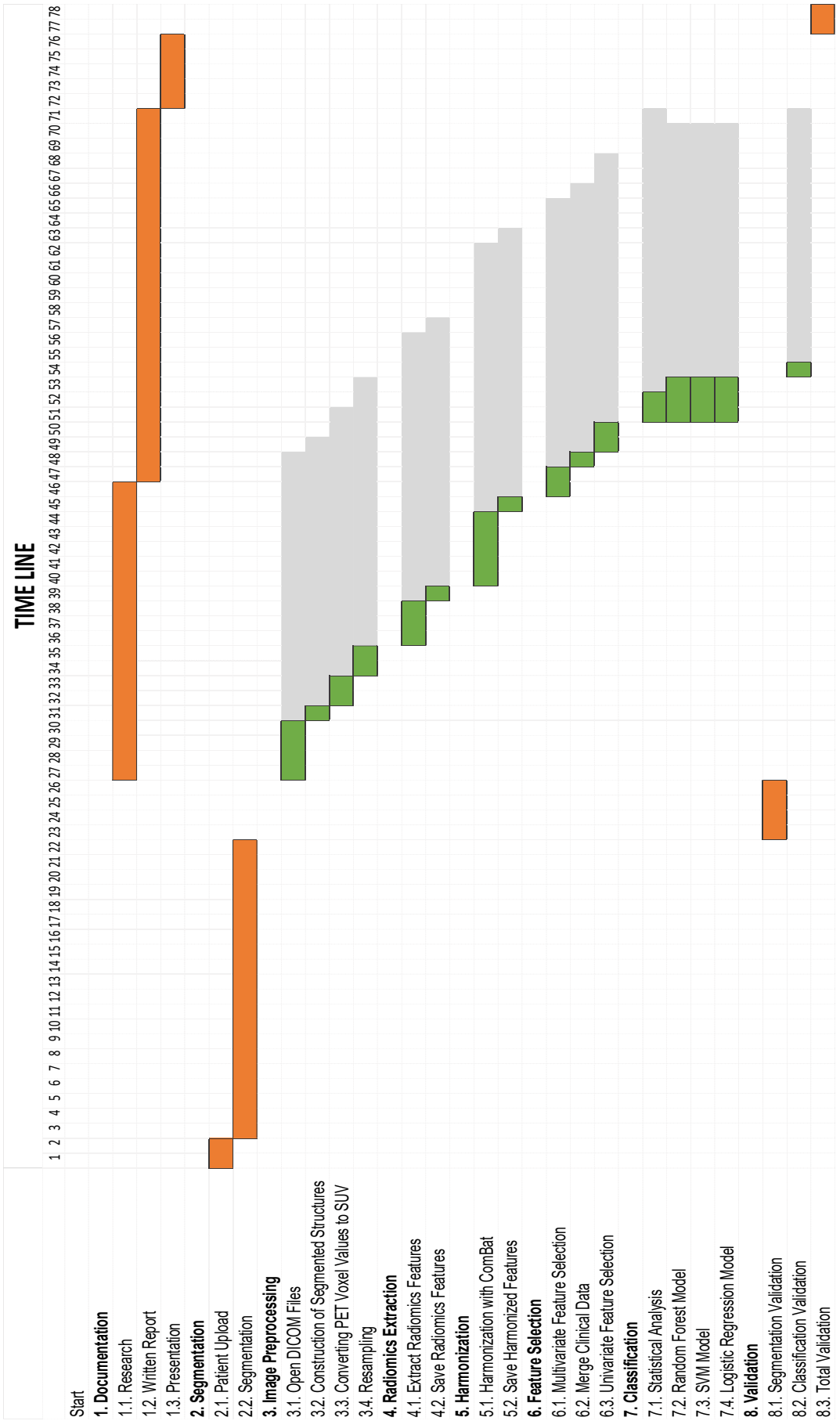

Figure 20. Gantt Diagram of the project.

7. Technical Viability

In the Canvas Model, technical feasibility is a fundamental aspect that assesses whether the project can be carried out with the available technological resources. This evaluation includes reviewing the technical capabilities of the team, the necessary infrastructure, and the specific knowledge required to successfully implement the proposed solutions. Below is the detailed Canvas Model for this project.

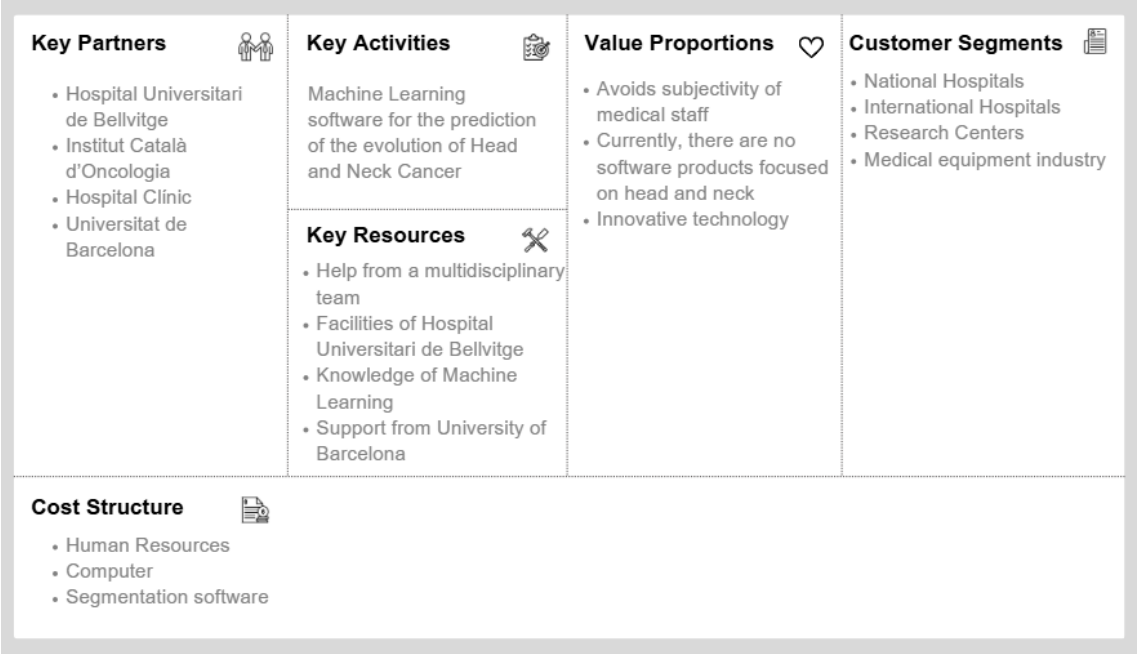

Figure 21. Canvas Model of the project.

In the development and implementation of any engineering project, it is essential to understand both the internal and external factors that can affect its success and its impact on the market. Below, a comprehensive SWOT analysis is carried out to assess the weaknesses, threats, strengths, and opportunities. **Strengths** refer to the internal advantages and resources that give the project a competitive edge, such as innovative technology, skilled personnel, or strong partnerships. **Weaknesses** are internal limitations or deficiencies that may hinder progress, including lack of funding, technical constraints, or insufficient experience. **Opportunities** are external factors in the environment that the project can exploit to its advantage, such as emerging markets, favourable regulations, or advances in related technologies. **Threats** are external challenges or risks that could negatively impact the project's success, including market competition, changing customer demands, or economic instability.

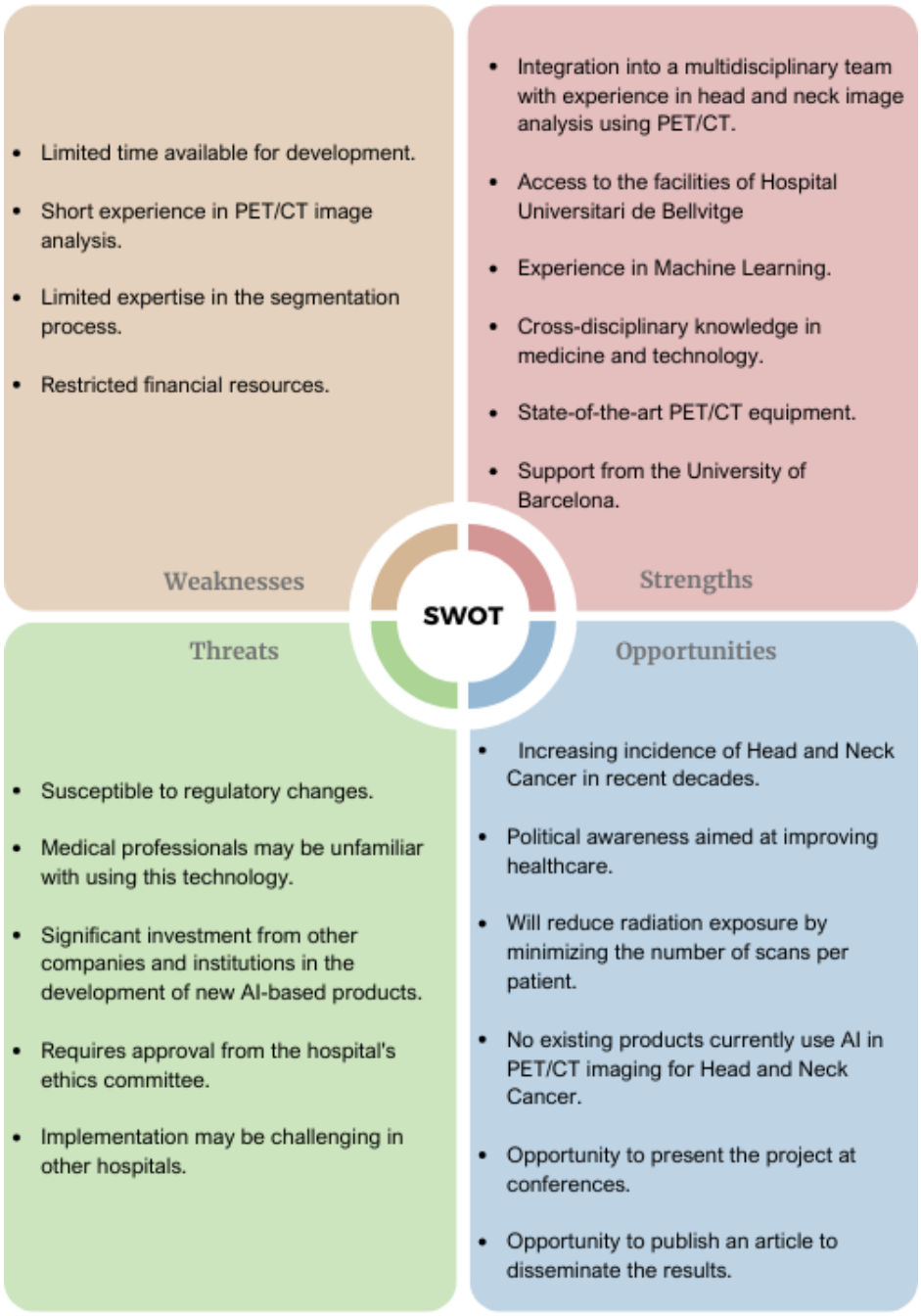

Figure 22. SWOT chart of the project.

The CAME analysis is a strategic tool based on the SWOT analysis to define the actions to be taken depending on the identified weaknesses, threats, strengths, and opportunities. Below is a detailed analysis in different paragraphs, offering a clear action guide focused on improving weaknesses, addressing threats, maintaining strengths, and exploring opportunities.

Correct (Weaknesses):

- **Short development time:** Implement more efficient project management to optimize time, establish clear priorities, and break the project into manageable phases.

- **Short experience in PET/CT image analysis:** Dedicate time to training in this field and collaborate with experts for knowledge transfer.
- **Limited experience in segmentation processes:** Develop specific protocols and guides for segmentation.
- **Restricted financial resources:** Leverage resources from the University of Barcelona and the Bellvitge University Hospital.

Address (Threats):

- **Susceptibility to regulatory changes:** Stay up to date with legal and regulatory updates.
- **Medical professionals unfamiliar with this technology:** Provide training and education programs for medical professionals.
- **Significant investment by other companies and institutions in developing new AI-based products:** Differentiate our product by highlighting its unique features and specific benefits.
- **Approval by the hospital ethics committee:** Prepare robust and detailed documentation for review by the ethics committee and ensure all practices follow ethical standards.
- **Implementation may be challenging in other hospitals:** Develop standardized protocols.

Maintain (Strengths):

- **Integration in a multidisciplinary team with experience in head and neck image analysis using PET/CT:** Maintain and expand joint research activities.
- **Availability of facilities at the Hospital de Bellvitge:** Leverage the infrastructure to carry out the project.
- **Experience in Machine Learning:** Use this experience to develop a complex and precise algorithm.
- **Cross-disciplinary knowledge in medicine and technology:** Leverage this knowledge to efficiently bring technology into the medical field.
- **State-of-the-art PET/TC equipment:** Utilize advanced imaging capabilities to ensure high-quality data.
- **Support from the University of Barcelona:** Maintain strong institutional backing for academic and technical resources.

Explore (Opportunities):

- **Increased incidence of Head and Neck Cancer in recent decades:** Promote the technology so it can be used in other centers.

- **Political awareness towards health improvement:** Take advantage of the situation to gain support.
- **Reducing radiation by decreasing the number of tests per patient:** Use this as an argument to commercialize the technology.
- **Patent possibility:** Investigate intellectual property opportunities and patent newly developed technologies.
- **No existing products using AI in PET/CT images for Head and Neck Cancer:** Conduct market studies to identify specific needs and adapt the product accordingly.
- **Possibility to present at conferences:** Prepare scientific presentations and participate in international conferences to share results.
- **Possibility to publish an article to disseminate the work:** Write a scientific article about the project and submit it to prestigious journals.

8. Economical Viability

The economic forecast is an essential element in the planning of any project, as it allows for estimating the costs and resources needed to carry out the different phases of the project.

In this project, the direct costs are not significant, as the main expenses depend on human resources. It is assumed that an engineering student is working full-time at 12€ per hour, along with the partial cost of the computer used for the work. It should also be noted that the student salary costs are fully covered by the University of Barcelona, while the remaining expenses are assumed by the Hospital Universitari de Bellvitge.

Table 9. Economical viability of the project.

CONCEPT	DESCRIPTION	COST
HUMAN RESOURCES	Includes time dedicated by a student (paid at 12€/hour) for data processing, analysis, and programming tasks; and a clinician (paid at 20€/hour) for supervision of segmentations.	8128€
COMPUTER	A workstation with a computer valued at 850€, amortized over 8 years. This results in an estimated operational cost of approximately 0.012€/hour based on its lifespan.	7,5€
PET/CT IMAGES	Medical imaging data obtained from patient follow-up studies, provided as part of routine clinical care, therefore incurring no additional cost for the project.	0€
MIM SOFTWARE	Commercial software used for medical image segmentation, requiring a 10,000€/year license. This cost can be approximated to about 1€/hour of usage based on typical working hours.	168€
PROGRAMMING SOFTWARE	Software tools like R and Python, which are open-source and free to use, covering all programming and statistical analysis needs for the project.	0€

Thus, based on the detailed cost breakdown, the estimated total cost of the project amounts to approximately 8303.5 €. Taking all the previous factors into account, we can confidently conclude that the project is economically feasible and presents a sound use of available resources.

9. Legal Aspects

The development of this project must be governed by both national and European regulations. This includes adherence to Spanish laws and European Union regulations, particularly those related to biomedical research and data protection. Additionally, it is crucial to follow the ethical principles of medical research, such as those established in the Declaration of Helsinki [60], and to ensure the informed consent of all participants.

This project was formally reviewed and approved by the Bioethical Committee of Hospital Universitari de Bellvitge, which evaluated the research protocols to ensure that they meet ethical standards, safeguard participant rights, and minimize risks.

About data management, it is essential to adhere to the internal policies of the hospital, and to ensure that the handling of medical data follows established protocols. Additionally, a clear policy must be established regarding the duration of data storage, with a retention period that complies with relevant regulations and minimizes the risk of patient privacy breaches.

In compliance with Spanish Law 3/2018 on the Protection of Personal Data and the Guarantee of Digital Rights, all data used in this study must be fully anonymized [61]. This means that before any data is accessed or analysed, all direct and indirect identifiers—such as names, addresses, geographic locations, ethnicity, and other sensitive characteristics—must be removed or masked to prevent re-identification of individual participants. Researchers are strictly prohibited from visualizing or handling any personal data that could compromise participant privacy.

Regulation (EU) 2017/745, which applies to general and implantable devices, defines the term “Medical Device” as any instrument, apparatus, appliance, software, implant, reagent, material, or other article intended by the manufacturer to be used in humans for specific medical purposes, such as diagnosis, prevention, monitoring, prediction, prognosis, treatment, or alleviation of disease, among others [62]. This means that our current prototype, which is designed to predict the progression of head and neck cancer using artificial intelligence, would be considered a Medical Device under this legislation, as it has the specific purpose of predicting a disease. Finally, all model outputs must undergo rigorous human review before being used in clinical practice, thereby ensuring the reliability and safety of the model in the healthcare context.

10. Conclusions and Future Perspectives

This study demonstrates the feasibility of developing predictive models for head and neck cancer outcomes, specifically recurrence and mortality, from PET imaging. The proposed models exhibit applicability across various head and neck tumour subtypes and PET imaging sources, highlighting their potential for broad clinical utility.

All technical and methodological steps were successfully implemented to develop a robust and generalizable machine learning model for predicting recurrence and exitus. The process began with systematic segmentation of tumours and lymph nodes to ensure consistent definition of regions of interest. A dedicated software solution enabled the transformation of DICOM data into analysable 3D numerical matrices, facilitating comprehensive radiomic feature extraction. To reduce scanner-related variability ComBat-based harmonization was applied, resulting in improved consistency and reliability across imaging sources. Finally, a structured feature selection strategy identified a set of interpretable descriptors that integrated both clinical and radiomic variables, each demonstrating relevance to oncological outcomes, and used to create the predictive models.

Machine learning models demonstrated solid predictive performance for both recurrence and exitus, with radiomic features. For recurrence prediction, Random Forest achieved the highest accuracy (0.83) and specificity (0.88), while SVM showed the highest sensitivity (0.73) and AUC (0.85), making both models strong candidates depending on clinical priorities.

For exitus, SVM reached the highest sensitivity (0.93) and a strong AUC (0.84), whereas Random Forest offered better specificity (0.80) and the highest AUC (0.86). The high AUC values indicate robust discriminative ability, and the flexibility in tuning model thresholds supports their integration into personalised decision-making pathways.

Among radiomic features, sphericity has demonstrated significant predictive value in head and neck cancer outcome models and should be considered for clinical application as a predictor variable. Tumours with lower sphericity often indicate more irregular and invasive growth patterns. Incorporating sphericity into clinical practice can improve risk stratification, making it a valuable and non-invasive biomarker for guiding personalized patient management in head and neck oncology. Furthermore, maximum 2D diameter and total lesion volume are also crucial features for prediction, as they provide quantitative measures of tumour size and burden, which are strongly correlated with disease extent and overall patient survival.

In summary, the SVM model demonstrated the highest AUC for recurrence prediction, suggesting its strong potential for clinical application in head and neck cancer management. For mortality prediction, the Random Forest model showed the most reliable performance, based on the AUC. Models with high AUC can be adjusted to prioritise specificity, reducing unnecessary follow-ups, or sensitivity to predict potential recurrences.

The primary limitations of this study include temporal constraints and a limited sample size. Due to the restricted duration of data collection, the number of patients included was relatively small, thereby limiting the scope of analysis. Extending the study period would allow for the inclusion of a larger and more diverse patient cohort, potentially encompassing additional diseases such as hypopharynx cancer. Moreover, an increased sample size would enhance the robustness of the findings, facilitating more comprehensive insights and stronger statistical validity.

Although existing machine learning models for head and neck cancer outcome prediction have shown promising results, this study offers a significant advancement by incorporating data from different PET scanners. This multi-scanner approach enhances the model's robustness and generalizability by accounting for variability across different imaging devices. Consequently, the findings are more applicable to diverse clinical settings.

Future research should continue to explore the clinical relevance of the selected features, many of which reflect key biological and anatomical characteristics of tumour behaviour and treatment response. Indeed, features such as tumour sphericity, standardized uptake values, and lymph node involvement have been demonstrated in the literature to be closely associated with therapeutic outcomes and disease progression.

The integration of deep learning segmentation methods presents another promising avenue. These techniques can significantly reduce inter-observer variability and human error, leading to more consistent and reproducible delineation of tumour regions. It directly contributes to the quality of radiomic feature extraction, thereby strengthening the predictive power of subsequent models.

As already emphasized, expanding the dataset is essential for improving model performance, generalizability, and robustness. Larger and more diverse cohorts would reduce overfitting, enabling more personalized and reliable predictions.

With a larger dataset, becomes feasible the application of more complex artificial intelligence approaches, such as CNNs, as these models can capture complex relationships to enhance prognostic accuracy. Future work should explore the integration of such models for end-to-end predictive pipelines in head and neck cancer prognosis.

References

[1] M. Furukawa, D. R. McGowan, and B. W. Papież, "Prediction of recurrence free survival of head and neck cancer using PET/CT radiomics and clinical information," Feb. 2024, Accessed: May 25, 2025. [Online]. Available: https://arxiv.org/pdf/2402.18417v1

[2] N. Dhariwal and A. Giridharan, "Towards Precision Oncology: Predicting Mortality and Relapse-Free Survival in Head and Neck Cancer Using Clinical Data," Feb. 2025, Accessed: May 25, 2025. [Online]. Available: https://arxiv.org/pdf/2502.11200v1

[3] "Oropharyngeal Cancer Treatment (PDQ®) - NCI." Accessed: May 25, 2025. [Online]. Available: https://www.cancer.gov/types/head-and-neck/hp/adult/oropharyngeal-treatment-pdq

[4] "Nasopharyngeal Carcinoma Treatment (PDQ®) - NCI." Accessed: May 25, 2025. [Online]. Available: https://www.cancer.gov/types/head-and-neck/hp/adult/nasopharyngeal-treatment-pdq

[5] M. Gonçalves *et al.*, "Radiomics in Head and Neck Cancer Outcome Predictions," *Diagnostics*, vol. 12, no. 11, p. 2733, Nov. 2022, doi: 10.3390/DIAGNOSTICS12112733/S1.

[6] E. Eyassu and M. Young, "Nuclear Medicine PET/CT Head and Neck Cancer Assessment, Protocols, and Interpretation," *StatPearls*, Jul. 2023, Accessed: May 25, 2025. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK573059/

[7] K. Swanson, E. Wu, A. Zhang, A. A. Alizadeh, and J. Zou, "From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment," *Cell*, vol. 186, no. 8, pp. 1772–1791, Apr. 2023, doi: 10.1016/J.CELL.2023.01.035.

[8] G. Scott, M. Mahmud, D. R. Owen, and M. R. Johnson, "Microglial positron emission tomography (PET) imaging in epilepsy: Applications, opportunities and pitfalls," *Seizure*, vol. 44, pp. 42–47, Jan. 2017, doi: 10.1016/J.SEIZURE.2016.10.023.

[9] S. S. Anand, H. Singh, and A. K. Dash, "Clinical Applications of PET and PET-CT," *Med J Armed Forces India*, vol. 65, no. 4, pp. 353–358, Oct. 2009, doi: 10.1016/S0377-1237(09)80099-3.

[10] H. Li, Z. Kong, Y. Xiang, R. Zheng, and S. Liu, "The role of PET/CT in radiotherapy for nasopharyngeal carcinoma," *Front Oncol*, vol. 12, Oct. 2022, doi: 10.3389/FONC.2022.1017758,.

[11] N. Dhariwal and A. Giridharan, "Towards Precision Oncology: Predicting Mortality and Relapse-Free Survival in Head and Neck Cancer Using Clinical Data," Feb. 2025, Accessed: May 26, 2025. [Online]. Available: https://arxiv.org/pdf/2502.11200v1

[12] M. Wierzbicka, W. Pietruszewska, A. Maciejczyk, and J. Markowski, "Trends in Incidence and Mortality of Head and Neck Cancer Subsites Among Elderly Patients: A Population-Based Analysis," *Cancers (Basel)*, vol. 17, no. 3, p. 548, Feb. 2025, doi: 10.3390/CANCERS17030548.

[13] "Oropharyngeal Cancer Treatment (PDQ®) - NCI." Accessed: May 26, 2025. [Online]. Available: https://www.cancer.gov/types/head-and-neck/hp/adult/oropharyngeal-treatment-pdq

[14] U. Kontny, C. Rodriguez-Galindo, D. Orbach, and M. Casanova, "Nasopharyngeal Carcinoma," *Pediatric Oncology*, pp. 79–97, Mar. 2024, doi: 10.1007/978-3-030-92071-5_10.

[15] M. E. Mayerhoefer *et al.*, "Introduction to Radiomics," *Journal of Nuclear Medicine*, vol. 61, no. 4, p. 488, Apr. 2020, doi: 10.2967/JNUMED.118.222893.

[16] L. Dercle *et al.*, "Artificial intelligence and radiomics: fundamentals, applications, and challenges in immunotherapy," *J Immunother Cancer*, vol. 10, no. 9, p. e005292, Sep. 2022, doi: 10.1136/JITC-2022-005292.

[17] M. Pallumeera, J. C. Giang, R. Singh, N. S. Pracha, and M. S. Makary, "Evolving and Novel Applications of Artificial Intelligence in Cancer Imaging," *Cancers (Basel)*, vol. 17, no. 9, p. 1510, Apr. 2025, doi: 10.3390/CANCERS17091510.

[18] P. Hamet and J. Tremblay, "Artificial intelligence in medicine," *Metabolism*, vol. 69, pp. S36–S40, Apr. 2017, doi: 10.1016/J.METABOL.2017.01.011.

[19] A. Nieri *et al.*, "[18F]FDG PET-TC radiomics and machine learning in the evaluation of prostate incidental uptake," *Expert Rev Med Devices*, vol. 20, no. 12, pp. 1183–1191, 2023, doi: 10.1080/17434440.2023.2280685,.

[20] N. Agüloğlu and A. Aksu, "Evaluation of survival of the patients with metastatic rectal cancer by staging 18F-FDG PET/CT radiomic and volumetric parameters," *Revista Española de Medicina Nuclear e Imagen Molecular (English Edition)*, vol. 42, no. 2, pp. 122–128, Mar. 2023, doi: 10.1016/J.REMNIE.2022.09.010,.

[21] C. Kong, X. Yin, J. Zou, C. Ma, and K. Liu, "The application of different machine learning models based on PET/CT images and EGFR in predicting brain metastasis of adenocarcinoma of the lung," *BMC Cancer*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/S12885-024-12158-0,.

[22] G. Bulut, H. I. Atilgan, G. Çınarer, K. Kılıç, D. Yıkar, and T. Parlar, "Prediction of pathological complete response to neoadjuvant chemotherapy in locally advanced breast cancer by using a deep learning model with 18F-FDG PET/CT," *PLoS One*, vol. 18, no. 9 September, Sep. 2023, doi: 10.1371/JOURNAL.PONE.0290543,.

[23] B. Huang *et al.*, "Prediction of lung malignancy progression and survival with machine learning based on pre-treatment FDG-PET/CT," *EBioMedicine*, vol. 82, Aug. 2022, doi: 10.1016/j.ebiom.2022.104127.

[24] R. O. Alabi, M. Elmusrati, I. Leivo, A. Almangush, and A. A. Mäkitie, "Artificial Intelligence-Driven Radiomics in Head and Neck Cancer: Current Status and Future Prospects," *Int J Med Inform*, vol. 188, p. 105464, Aug. 2024, doi: 10.1016/J.IJMEDINF.2024.105464.

[25] "((‘Artificial Intelligence’[Mesh]) AND ‘Prognosis’[Mesh]) AND ‘Head and Neck Neoplasms’[Mesh] - Search Results - PubMed." Accessed: May 28, 2025. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/?term=%28%28%22Artificial+Intelligence%22%5BMesh%5D%29+AND+%22Prognosis%22%5BMesh%5D%29+AND+%22Head+and+Neck+Neoplasms%22%5BMesh%5D&sort=date

[26] "Los hospitales del ICS utilizan la IA para mejorar el diagnóstico del cáncer de mama - Institut Català de la Salut." Accessed: May 28, 2025. [Online]. Available: https://icscampdetarragona.cat/es/los-hospitales-del-ics-utilizan-la-ia-para-mejorar-el-diagnostico-del-cancer-de-mama/

[27] W. Xu, S. Yu, Y. Ma, C. Liu, and J. Xin, "Effect of different segmentation algorithms on metabolic tumor volume measured on 18F-FDG PET/CT of cervical primary squamous cell carcinoma," *Nucl Med Commun*, vol. 38, no. 3, p. 259, 2017, doi: 10.1097/MNM.0000000000000641.

[28] F. Raman *et al.*, "Biomarker Localization, Analysis, Visualization, Extraction, and Registration (BLAzER) Methodology for Research and Clinical Brain PET Applications," *J Alzheimers Dis*, vol. 70, no. 4, p. 1241, 2019, doi: 10.3233/JAD-190329.

[29] A. Fedorov *et al.*, "3D Slicer as an Image Computing Platform for the Quantitative Imaging Network," *Magn Reson Imaging*, vol. 30, no. 9, p. 1323, Nov. 2012, doi: 10.1016/J.MRI.2012.05.001.

[30] M. Hatt *et al.*, "Joint EANM/SNMMI guideline on radiomics in nuclear medicine : Jointly supported by the EANM Physics Committee and the SNMMI Physics, Instrumentation and Data Sciences Council," *Eur J Nucl Med Mol Imaging*, vol. 50, no. 2, pp. 352–375, Jan. 2023, doi: 10.1007/S00259-022-06001-6.

[31] K. Erlandsson, I. Buvat, P. H. Pretorius, B. A. Thomas, and B. F. Hutton, "A review of partial volume correction techniques for emission tomography and their applications in neurology, cardiology and oncology," *Phys Med Biol*, vol. 57, no. 21, Nov. 2012, doi: 10.1088/0031-9155/57/21/R119.

[32] C. D. Pain, G. F. Egan, and Z. Chen, "Deep learning-based image reconstruction and post-processing methods in positron emission tomography for low-dose imaging and resolution enhancement," *Eur J Nucl Med Mol Imaging*, vol. 49, no. 9, p. 3098, Jul. 2022, doi: 10.1007/S00259-022-05746-4.

[33] D. Hellwig, N. C. Hellwig, S. Boehner, T. Fuchs, R. Fischer, and D. Schmidt, "Artificial Intelligence and Deep Learning for Advancing PET Image Reconstruction: State-of-the-Art and Future Directions," *Nuklearmedizin*, vol. 62, no. 6, p. 334, Aug. 2023, doi: 10.1055/A-2198-0358.

[34] R. T. H. Leijenaar *et al.*, "The effect of SUV discretization in quantitative FDG-PET Radiomics: the need for standardized methodology in tumor texture analysis," *Scientific Reports 2015 5:1*, vol. 5, no. 1, pp. 1–10, Aug. 2015, doi: 10.1038/srep11075.

[35] A. Demircioğlu, "The effect of preprocessing filters on predictive performance in radiomics," *Eur Radiol Exp*, vol. 6, no. 1, pp. 1–10, Dec. 2022, doi: 10.1186/S41747-022-00294-W/FIGURES/4.

[36] A. Zwanenburg, "Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis," *European Journal of Nuclear Medicine and Molecular Imaging 2019 46:13*, vol. 46, no. 13, pp. 2638–2655, Jun. 2019, doi: 10.1007/S00259-019-04391-8.

[37] A. Zwanenburg *et al.*, "The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping," *Radiology*, vol. 295, no. 2, pp. 328–338, May 2020, doi: 10.1148/radiol.2020191145.

[38] J. J. M. Van Griethuysen *et al.*, "Computational Radiomics System to Decode the Radiographic Phenotype," *Cancer Res*, vol. 77, no. 21, p. e104, Nov. 2017, doi: 10.1158/0008-5472.CAN-17-0339.

[39] I. Fornacon-Wood *et al.*, "Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform," *Eur Radiol*, vol. 30, no. 11, pp. 6241–6250, Nov. 2020, doi: 10.1007/S00330-020-06957-9.

[40] "Quantitative Nuclear Medicine Imaging Using Advanced Image Reconstruction and Radiomics - University of Edinburgh." Accessed: Apr. 04, 2025. [Online]. Available: https://discovered.ed.ac.uk/discovery/fulldisplay?docid=cdi_proquest_journals_2334392614&context=PC&vid=44UOE_INST:44UOE_VU2&lang=en&search_scope=UoE&adaptor=Primo%20Central&tab=Everything&query=null,,xviii,%20214%20pages%20:%20illustrations%20%20%2025%20cm.,AND&mode=advanced&offset=40

[41] C. Nioche *et al.*, "LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity," *Cancer Res*, vol. 78, no. 16, pp. 4786–4789, Aug. 2018, doi: 10.1158/0008-5472.CAN-18-0125.

[42] W. Mu, M. B. Schabath, and R. J. Gillies, "Images Are Data: Challenges and Opportunities in the Clinical Translation of Radiomics," *Cancer Res*, vol. 82, no. 11, pp. 2066–2068, Jun. 2022, doi: 10.1158/0008-5472.CAN-22-1183.

[43] M. Götz, M. Nolden, and K. Maier-Hein, "MITK Phenotyping: An open-source toolchain for image-based personalized medicine with radiomics," *Radiotherapy and Oncology*, vol. 131, pp. 108–111, Feb. 2019, doi: 10.1016/J.RADONC.2018.11.021.

[44] A. P. Apte *et al.*, "pyCERR-A Python-based Computational Environment for Radiological Research," 2024, Accessed: Apr. 05, 2025. [Online]. Available: https://github.com/cerr/pyCERR-

[45] J. J. M. van Griethuysen et al., "Radiomic Features," https://pyradiomics.readthedocs.io/en/latest/features.html.

[46] N. Aide, C. Lasnon, P. Veit-Haibach, T. Sera, B. Sattler, and R. Boellaard, "EANM/EARL harmonization strategies in PET quantification: from daily practice to multicentre oncological studies," *Eur J Nucl Med Mol Imaging*, vol. 44, no. Suppl 1, p. 17, Aug. 2017, doi: 10.1007/S00259-017-3740-2.

[47] F. Orlhac *et al.*, "A Guide to ComBat Harmonization of Imaging Biomarkers in Multicenter Studies," *Journal of Nuclear Medicine*, vol. 63, no. 2, p. 172, Feb. 2022, doi: 10.2967/JNUMED.121.262464.

[48] E. Stamoulou *et al.*, "Harmonization Strategies in Multicenter MRI-Based Radiomics," *J Imaging*, vol. 8, no. 11, p. 303, Nov. 2022, doi: 10.3390/JIMAGING8110303.

[49] A. Chatzi and O. Doody, "The one-way ANOVA test explained," *Nurse Res*, vol. 31, no. 3, pp. 8–14, Sep. 2023, doi: 10.7748/NR.2023.E1885.

[50] A. Farhan AlShammari, "Implementation of Feature Selection using Correlation Matrix in Python," *Int J Comput Appl*, vol. 186, no. 58, pp. 975–8887, 2024.

[51] A. Perniciano, A. Loddo, C. Di Ruberto, and B. Pes, "Insights into radiomics: impact of feature selection and classification," *Multimedia Tools and Applications 2024*, pp. 1–27, Nov. 2024, doi: 10.1007/S11042-024-20388-4.

[52] S. D. Mohaghegh, "Traditional Statistics vs. Artificial Intelligence and Machine Learning," *Journal of Petroleum Technology*, Nov. 2019.

[53] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. W. L. Aerts, "Machine Learning methods for Quantitative Radiomic Biomarkers," *Scientific Reports 2015 5:1*, vol. 5, no. 1, pp. 1–11, Aug. 2015, doi: 10.1038/srep13087.

[54] Theng Mahesh and Theng Dipti, "Machine Learning Algorithms for Predictive Analytics: A Review and New Perspectives," *Science Press Conference*, Jul. 2020.

[55] P. T. R, "A Comparative Study on Decision Tree and Random Forest Using R Tool," *IJARCCE*, pp. 196–199, Jan. 2015, doi: 10.17148/IJARCCE.2015.4142.

[56] H. M. Castro and J. C. Ferreira, "Linear and logistic regression models: when to use and how to interpret them?," *Jornal Brasileiro de Pneumologia*, vol. 48, no. 6, p. e20220439, Jan. 2022, doi: 10.36416/1806-3756/E20220439.

[57] A. Duman, X. Sun, S. Thomas, J. R. Powell, and E. Spezi, "Reproducible and Interpretable Machine Learning-Based Radiomic Analysis for Overall Survival Prediction in Glioblastoma Multiforme," *Cancers (Basel)*, vol. 16, no. 19, Oct. 2024, doi: 10.3390/CANCERS16193351,.

[58] R. Boellaard *et al.*, "FDG PET/CT: EANM procedure guidelines for tumour imaging: version 2.0," *Eur J Nucl Med Mol Imaging*, vol. 42, no. 2, pp. 328–354, Feb. 2015, doi: 10.1007/S00259-014-2961-X,.

[59] "STEP – Science Toolbox Exploitation Platform." Accessed: May 31, 2025. [Online]. Available: https://step.esa.int/main/

[60] "World Medical Association declaration of Helsinki: Ethical principles for medical research involving human subjects," *JAMA*, vol. 310, no. 20, pp. 2191–2194, Nov. 2013, doi: 10.1001/JAMA.2013.281053,.

[61] "BOE-A-2018-16673 Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales." Accessed: May 07, 2025. [Online]. Available: https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673

[62] "Regulation - 2017/745 - EN - Medical Device Regulation - EUR-Lex." Accessed: May 07, 2025. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2017/745/oj/eng

Annexes

1. Documentation	
1.1. Research	
Description	Collection and analysis of relevant information to establish the foundations of the project.
Deliverable	A research report summarizing key findings and data sources.
Resources	- Computer with internet access - Human resources
Estimated cost	481 €
Estimated duration	10 days (20h)
1.2. Written Report	
Description	Writing a detailed written report about all the project.
Deliverable	Report file.
Resources	- Computer with internet access - Human resources
Estimated cost	1658 €
Estimated duration	20 days (138h)
1.3. Presentation	
Description	Design and preparation of a visual presentation summarizing key points from the project.
Deliverable	PowerPoint presentation.
Resources	- Computer with internet access - Human resources
Estimated cost	192 €
Estimated duration	2 days (16h)

Table 10. WBS dictionary for 1. Documentation tasks.

2. Segmentation	
2.1. Patient Upload	
Description	Upload studies to MIM Software.
Deliverable	A folder containing all the studies imported into MIM Software.
Resources	- Computer with internet access - Studies list - MIM Software - Human resources
Estimated cost	48 €
Estimated duration	1 day (8h)
2.2. Segmentation	
Description	Identification and segmentation of the regions of interest (tumour and lymph nodes).
Deliverable	Folder with the structures segmented (RT Structures).
Resources	- Computer with internet access - MIM Software - Human resources
Estimated cost	1922 €
Estimated duration	20 days (160h)

Table 11. WBS dictionary for 2. Segmentation tasks.

3. Image Preprocessing	
3.1. Open DICOM files	
Description	Development of a script to automate the processing from DICOM files.
Deliverable	A functional script capable of batch-processing DICOM files.
Resources	- Computer with internet access - Python - Anaconda - Human resources
Estimated cost	288 €
Estimated duration	4 days (24h)
3.2. Construction of Segmented Structures	
Description	Generation of 3D ROIs from RT-DICOM file vectors and correction of spatial errors based on CT reference alignment.
Deliverable	The code that extracts ROIs from RT-DICOM files, aligns them with the CT space, and exports corrected 3D structures.
Resources	- Computer with internet access - Python - Anaconda - Human resources
Estimated cost	72 €
Estimated duration	1 day (6h)
3.3. Standardize Images	
Description	Conversion of PET image pixel values to Standardized Uptake Values (SUV), followed by spatial rescaling.
Deliverable	Script that reads performs SUV conversion and outputs rescaled images.
Resources	- Computer with internet access - Python - Anaconda - Human resources
Estimated cost	144 €
Estimated duration	2 days (12h)

Table 12. WBS dictionary for 3. Image Preproocessing tasks.

4. Radiomics Extraction	
4.1. Extract Radiomics Features	
Description	Extraction of radiomics features from segmented regions using PyRadiomics with custom parameters.
Deliverable	Script utilizing PyRadiomics to extract predefined radiomics features based on customized parameters, generating a structured dataset.
Resources	- Computer with internet access - Python - Anaconda - Human resources
Estimated cost	216 €
Estimated duration	3 days (18h)
4.2. Save Radiomics Features	
Description	Exporting the extracted radiomics features into a structured CSV file for each type of lesion (tumour and lymph nodes).

Deliverable	A CSV file containing the complete set of extracted radiomics features for each type of lesion.
Resources	- Computer with internet access - Python - Anaconda - Human resources
Estimated cost	72 €
Estimated duration	1 day (6h)

Table 13. WBS dictionary for 4. Radiomics Extraction tasks.

5. Harmonization	
5.1. Harmonization with ComBat	
Description	Application of the ComBat algorithm to harmonize radiomics features across datasets acquired from different scanners.
Deliverable	Script implementing the ComBat harmonization method.
Resources	- Computer with internet access - R - RStudio - Human resources
Estimated cost	360 €
Estimated duration	5 days (30h)
5.2. Save Harmonized Features	
Description	Export of the radiomics features after ComBat harmonization into a structured CSV file.
Deliverable	A CSV file containing the harmonized radiomics features.
Resources	- Computer with internet access - R - RStudio - Human resources
Estimated cost	72 €
Estimated duration	1 day (6h)

Table 14. WBS dictionary for 5. Harmonization tasks.

6. Feature Selection	
6.1. Multivariate Feature Selection	
Description	Application of multivariate statistical methods to select the most relevant radiomics features.
Deliverable	Script that performs multivariate feature selection.
Resources	- Computer with internet access - R - RStudio - Human resources
Estimated cost	144 €
Estimated duration	2 days (12h)
6.2. Merge Clinical Data	
Description	Integration of clinical variables database with the radiomics database.
Deliverable	A unified CSV file containing both clinical and radiomics features.
Resources	- Computer with internet access - R

	- RStudio - Human resources
Estimated cost	72 €
Estimated duration	1 day (6h)
6.3. Univariate Feature Selection	
Description	Application of univariate statistical methods to select the most relevant radiomics features.
Deliverable	Script that performs univariate feature selection.
Resources	- Computer with internet access - R - RStudio - Human resources
Estimated cost	144 €
Estimated duration	2 days (12h)

Table 15. WBS dictionary for 6. Feature Selection tasks.

7. Classification	
7.1. Statistical Analysis	
Description	Perform statistical analysis on the selected radiomics and clinical features to identify relationship with outcome.
Deliverable	Figures of the statistical analysis results.
Resources	- Computer with internet access - R - RStudio - Human resources
Estimated cost	144 €
Estimated duration	2 days (12h)
7.2. Random Forest	
Description	Development and training of a Random Forest model using the selected features, including hyperparameter tuning and performance evaluation.
Deliverable	Script implementing the Random Forest model, showing feature importance and performance metrics.
Resources	- Computer with internet access - R - RStudio - Human resources
Estimated cost	216 €
Estimated duration	3 days (18h)
7.3. SVM Model	
Description	Development and training of a SVM model using the selected features, including kernel selection and performance evaluation.
Deliverable	Script implementing the SVM model, including performance metrics.
Resources	- Computer with internet access - R - RStudio - Human resources

Estimated cost	216 €
Estimated duration	3 days (18h)
7.4. Logistic Regression Model	
Description	Development and training of a Logistic Regression model using the selected features, including regularization and performance evaluation.
Deliverable	Script implementing the Logistic Regression model, including performance metrics.
Resources	- Computer with internet access - R - RStudio - Human resources
Estimated cost	216 €
Estimated duration	3 days (18h)

Table 16. WBS dictionary for 7. Classification tasks.

8. Validation	
8.1. Segmentation Validation	
Description	Consist of a thorough validation of the segmentation in collaboration with an expert nuclear medicine physician.
Deliverable	Folder with the structures segmented (RT Structures).
Resources	- Computer with internet access - Studies list - MIM Software - Expert Clinician
Estimated cost	640 €
Estimated duration	4 days (32h)
8.2. Classification Validation	
Description	Validate the performance of the Machine Learning model by examining the features used, the code, and its accuracy, ensuring there is no overfitting.
Deliverable	Model validation report.
Resources	- Computer with internet access - R - RStudio - Human resources
Estimated cost	72 €
Estimated duration	1 day (6h)
8.3. Total Validation	
Description	Validate all aspects of the project to ensure they meet the requirements established in the planning.
Deliverable	Comprehensive validation report.
Resources	- Computer with internet access - R - RStudio - Human resources
Estimated cost	192 €
Estimated duration	2 days (16h)

Table 17. WBS dictionary for 8. Validation tasks.