UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics and Data Science

Master's thesis

Radiomics-based model for characterization of tissue changes during response to treatment in preclinical glioblastoma.

Sebastián Martin Tanco

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Bioinformatics

SUPERVISOR :
Prof. dr. Ziv SHKEDY

SUPERVISOR :
Ana Paula CANDIOTA

2024
2025

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences

School for Information Technology

Master of Statistics and Data Science

Master's thesis

Radiomics-based model for characterization of tissue changes during response to treatment in preclinical glioblastoma.

Sebastián Martin Tanco

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics and Data Science, specialization Bioinformatics

SUPERVISOR :

Prof. dr. Ziv SHKEDY

SUPERVISOR :

Ana Paula CANDIOTA

Abstract

Glioblastoma (GB) is the most frequent aggressive primary brain tumor in adults, with a 5-year overall survival rate of 5% in GB patients even after aggressive treatment. GB is widely recognized as challenging to diagnose and manage. GB treatment generally comprises surgical excision, followed by radiation and chemotherapy, but relapse is the norm. Decision-making regarding therapy is challenging and uncertain in the early moments, wasting precious time until confirmation on therapy efficacy or the need to switch to second-line approaches, highlighting the need to improve therapy response assessment. Magnetic resonance imaging (MRI) contains a wide range of structural and functional information that can be used by machine learning algorithms to facilitate the management of GB patients. This thesis aims to develop and evaluate machine learning classifiers capable of detecting therapy-induced changes in the tumor microenvironment, using the GL261 murine GB model as a preclinical benchmark. T2-weighted MRI data were analyzed, and a total of 90 quantitative imaging features were extracted per MRI image: 42 texture features derived from various gray-level matrices and 48 morphological features computed from Minkowski functionals. To address the high dimensionality of the dataset, feature selection methods were applied in combination with several classification algorithms, including logistic regression (LR), linear discriminant analysis (LDA), K-nearest neighbors (KNN), and support vector machines (SVM). Our findings suggest that more flexible models, particularly SMV with higher-degree polynomial kernels and fine-grained KNN, achieved the best classification performance. These results highlight the potential of radiomics-based models to non-invasively monitor therapeutic response in GB, with possible future applications in translational and clinical settings by facilitating personalized medicine.

Acknowledgement

First and foremost, I would like to express my sincere gratitude to my external supervisor, Prof. Ana Paula Candiota, for her invaluable guidance and the generous time she dedicated to teaching me about the most relevant characteristics of magnetic resonance images, the GL261 preclinical mouse model, TMZ treatment, key radiomic features and the laboratory's prior experience with machine learning algorithms applied to similar datasets. Her patience, support and inspiration have been instrumental in the successful completion of this thesis, which has undoubtedly been an exceptional learning opportunity for me.

I am also sincerely grateful to the members of the Protein Engineering and Nanomedicine laboratory at the Institute of Biotechnology and Biomedicine, led by Prof. Julia Lorenzo. In particular, I wish to thank Nuria Basanta, Paula Alfonso and Marta Mulero for sharing their expertise in magnetic resonance imaging, GL261 mouse treatment, and image interpretation.

I owe heartfelt thanks to my family (especially to Filip) and friends for their constant support, motivation and patience throughout this journey.

Finally, I wish to acknowledge the Institute of Biotechnology and Biomedicine, the Universitat Autònoma de Barcelona and Hasselt University (with a special thanks to my internal supervisor Ziv Shkedy) for providing the resources, guidance, and opportunities that made this thesis possible.

Abbreviation list

CNS	Central nervous system
CSF	Cerebrospinal fluid
CT	Computerized tomography
EE	Enriched environment
EGFR	Epidermal growth factor receptor
FLAIR	Fluid-attenuated inversion recovery
GABRMN	Group of biomedical applications of nuclear magnetic resonance
GB	Glioblastoma
GLCM	Gray-level co-occurrence matrix
GLRLM	Gray-level run length matrix
GLSZM	Gray-level size zone matrix
IDH	Isocitrate dehydrogenase
IMS	Immune-Enhancing Metronomic Schedule
KNN	K-nearest neighbors
LDA	Linear discriminant analysis
LOOCV	Leave-one-out cross-validation
LR	Logistic regression
MGMT	Methylguanine DNA methyltransferase
NGTDM	Neighborhood gray-tone difference matrix
MRI	Magnetic resonance imaging
MRSI	Magnetic resonance spectroscopic imaging
PET	Positron emission tomography
RANO	Response assessment in neuro-oncology
RARE	Rapid acquisition with relaxation enhancement
RECIST	Response Evaluation Criteria in Solid Tumors
RFE	Recursive feature elimination
ROI	Region of interest
SVM	Support vector machine
SVS	Slice-voting system
TERT	Telomerase reverse transcriptase
TMZ	Temozolomide
WHO	World Health Organization

Table of contents

1. Introduction

1.1 Glioblastoma

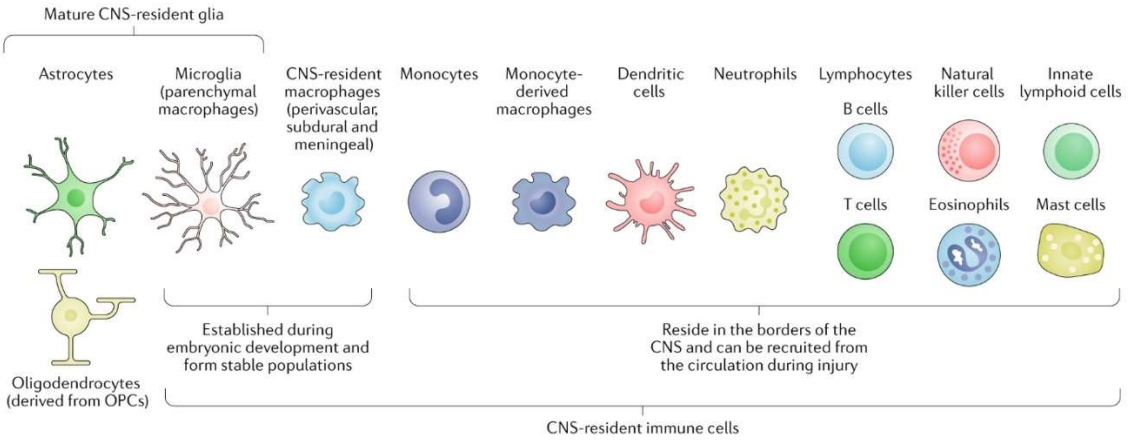
Glioblastoma (GB), the focus of this Master's thesis, is the most frequent of the aggressive primary brain tumor types found in human adults. Tumors of the Central Nervous System (CNS) comprised approximately 2% of all cancer cases diagnosed worldwide in 2022. Because of their poor prognosis, they are responsible for 2.8% of all cancer-related deaths. Each year, CNS cancers lead to over 300,000 new cases and 250,000 deaths globally [1]. Gliomas comprise nearly 30% of all primary brain tumors and around 80% of malignant cases, making them the leading cause of death from primary brain tumors [2]. In 2022, the reported incidence was 5.26 cases per 100,000 population. These tumors originate from mature glial cells (Figure 1) or their precursors and are characterized by their capacity to infiltrate and diffusely invade surrounding brain tissue.

GB is an adult-type (i.e., median age at diagnosis is 62) diffuse glioma, which spreads infiltratively through the brain parenchyma, gradually causing neuronal dysfunction and cell death [3]. GB is the most aggressive and common form of primary astrocytoma, accounting for approximately 57% of all gliomas and 48% of primary malignant CNS tumors [4]. The 2021 World Health Organization (WHO) Classification of Tumors of the CNS grades brain and spinal cord tumors (i.e., with grades going from 1 to 4) based on histological, immunohistochemical, and molecular features, with Grade 4 being the most malignant [5]. GB is classified in most cases as Grade 4 (Figure 1).

Some GB tumors develop *de novo* (i.e., primary tumors), while others (i.e., secondary GBs) progress from lower-grade gliomas [6]. Primary GB is diagnosed in the case of adult diffuse astrocytic tumors that are *IDH*-wildtype (isocitrate dehydrogenase wildtype) when either histological features of grade 4 malignancy (such as microvascular proliferation or necrosis) are present, or when specific molecular alterations are identified. The latter includes telomerase reverse transcriptase (TERT) promoter mutation, epidermal growth factor receptor (EGFR) amplification, or combined whole chromosome 7 gain and whole chromosome 10 loss (+7/−10) (Figure 1). The presence of any one of these molecular changes is sufficient for GB diagnosis, even without classic grade 4 histological features [7]. In contrast, secondary GBs account for only ~10% of GB cases and typically arise from precursor diffuse or anaplastic astrocytomas. These secondary GBs carry *IDH* mutations and are often linked to improved outcomes [8].

GB is a disease generally associated with poor prognosis, especially in older adults, since advanced age or incomplete resection are key negative prognostic factors. Median survival in elderly patients receiving only supportive care is less than 4 months [8]. Unfortunately, despite advances in diagnostic strategies and novel therapies, overall survival has not significantly improved. The typical survival rate is approximately 12-15 months with standard treatment, and the two-year survival rate is below 30% [9].

A)



B)

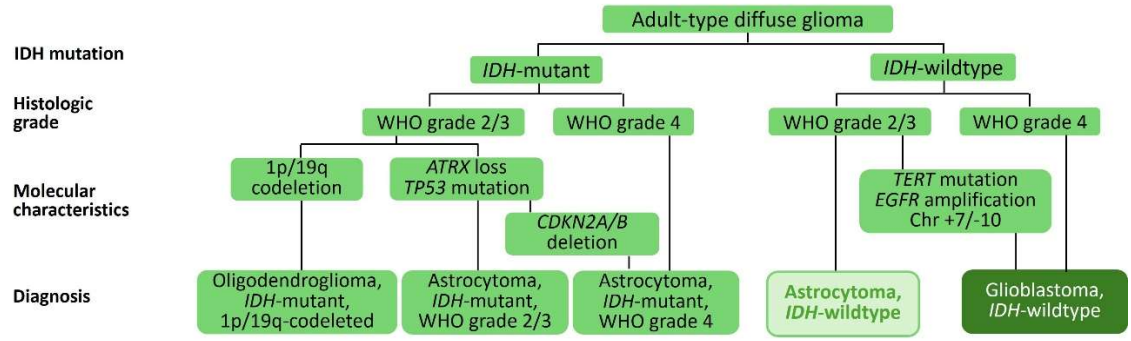


Figure 1 – A) Astrocytes, oligodendrocytes, and microglia are the main glial cell types in the CNS. Microglia also serve as a stable, resident immune cell population within the CNS. Extracted from [10]. B) Classification for adult-type diffuse gliomas. Adapted from [7].

The current standard of care for GB reflects its highly aggressive nature and aims to prolong survival while preserving neurological function. Treatment typically begins with maximal safe surgical resection to reduce tumor burden and alleviate symptoms, followed by a combination of radiotherapy and chemotherapy [11]. Temozolomide (TMZ), an oral alkylating agent, is the chemotherapeutic agent of choice and is administered concurrently with radiotherapy, followed by additional cycles as maintenance therapy. This multimodal regimen, commonly referred to as the Stupp protocol, has become the gold standard for newly diagnosed adult patients up to 70 years of age who are in good general and neurological health. Despite these efforts, GB almost invariably recurs, often locally, highlighting the limitations of current therapies. Overall, treatment response in GB is closely linked to molecular characteristics, with MGMT (methylguanine DNA methyltransferase) promoter methylation being the strongest predictor of benefit from TMZ chemotherapy. For the multimodal regimen, median overall survival is longer in patients with a methylated MGMT promoter (24 months); in unmethylated cases, it averages just 12.6 months [11].

GB are highly heterogeneous tumors, both genetically and phenotypically, which poses a significant challenge to effective treatment. This variability exists not only between patients but also within a single tumor, where diverse subpopulations of cells contribute to treatment resistance [12]. Resistance can be either intrinsic or acquired after therapy and is driven by multiple mechanisms, including the blood-brain barrier (i.e., the natural barrier that separates

the blood and CNS and can hamper the brain delivery of drugs when administered intravenously), evasion of apoptosis [13, 14], the presence of glioblastoma stem cells [15], proneural-mesenchymal transition [16, 17], and immune suppression [18]. Together, these factors reduce the efficacy of conventional and targeted therapies, complicating clinical decision-making and underscoring the urgent need for improved response assessment and personalized treatment strategies.

1.2 Magnetic resonance imaging

Magnetic Resonance Imaging (MRI) has become one of the most widely used imaging techniques today, especially in the medical field. MRI is particularly effective in detecting pathological tissues, making it a key diagnostic tool for healthcare professionals. In particular, in the case of GB, it is the standard modality for diagnosing, characterizing, and clinical management, typically identifying tumors around 4 cm in size at diagnosis [19]. Although GB tumors often develop in the brain's temporal lobe, they can appear anywhere in the CNS. Compared to histopathology, radiological methods like MRI, computerized tomography (CT), and positron emission tomography (PET) offer inherently digital imaging, facilitating analysis and follow-up without the limitations of physical sample processing. MRI advantages include excellent soft tissue contrast, high spatial resolution, unlimited penetration depth, and the absence of ionizing radiation (i.e., unlike CT scans).

MRI is based on the interaction between a magnetic field and the hydrogen nuclei present in water molecules within different tissues. MRI allows the measurement of T1 and T2 relaxation times of these nuclei, which vary depending on their environment, mobility, and concentration. This information is particularly useful for detecting pathological tissues, as cancer cells typically contain more water than healthy ones. The differences in T1 and T2 values contribute to the contrast seen in MRI images.

However, in certain regions of the body, the inherent contrast in MRI is insufficient to clearly distinguish pathological tissues from healthy ones. To enhance image contrast, contrast agents are employed in approximately 40% of MRI examinations. These agents are typically administered intravenously and function by shortening the relaxation times of water protons within tissues, thereby increasing the signal intensity and improving the clarity of the resulting images. Currently, the most widely used contrast agents are based on chelates of the trivalent gadolinium ion (Gd^{3+}), which is paramagnetic [20]. Gadolinium(III) possesses seven unpaired electrons, endowing it with a high longitudinal R_1 relaxivity (i.e., defined as the capacity of a contrast agent to reduce the relaxation times of hydrogen nuclei, thereby enhancing the MRI signal in regions where the agent accumulates).

A standard MRI scan protocol for patients with brain tumors typically includes T1-weighted, T2-weighted, fluid-attenuated inversion recovery (FLAIR), and post-contrast T1-weighted sequences (Figure 2). T1-weighted images are particularly effective for illustrating anatomical detail, with cerebrospinal fluid (CSF) and most tumors appearing as areas of low signal intensity. In contrast, T2-weighted images offer greater sensitivity for lesion detection, presenting most pathological lesions as hyperintense, whereas regions of hemorrhage or chronic hemosiderin deposition often appear hypointense [21]. FLAIR sequences, which are T2-weighted images with CSF suppression, are highly sensitive for detecting pathological changes and enhance the visibility of

lesions such as tumors and edema, which appear with increased signal intensity compared to conventional T2-weighted images. However, tumor margins in FLAIR or T2-weighted images may be poorly differentiated from surrounding edema, gliosis, or ischemic alterations. Post-contrast T1-weighted imaging generally offers superior delineation of the tumor nidus and provides valuable diagnostic information regarding tumor grade, hemorrhage, necrosis, and peritumoral edema [21].

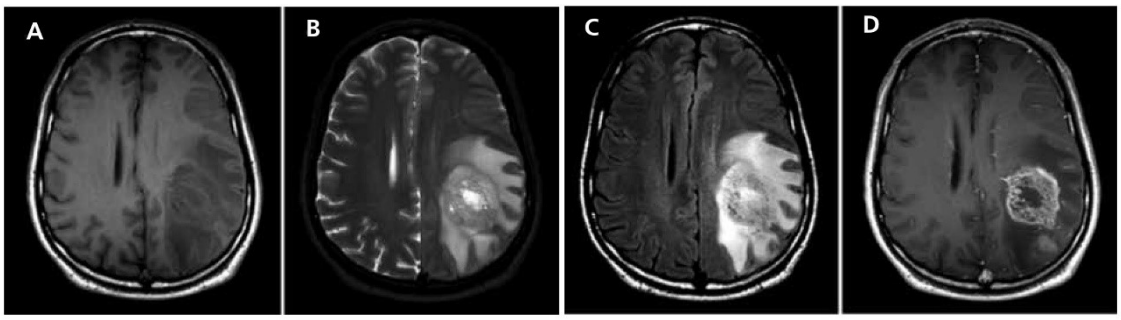


Figure 2 – MRI scans performed for diagnostic purposes revealed a large intracranial mass in a patient presenting right-sided hemiparesis. Subsequent histopathological analysis was consistent with GB. (A) The T1-weighted image shows a hypointense lesion located in the left frontoparietal region. (B, C) T2-weighted and fluid-attenuated inversion recovery (FLAIR) sequences demonstrate a heterogeneous hyperintense lesion accompanied by surrounding edema. (D) The post-contrast T1-weighted image displays a heterogeneous, ring-enhancing lesion, with an additional area of enhancement posterior to the primary lesion, as well as prominent vasogenic edema. Adapted from [21].

There are currently radiological and clinical guidelines, particularly the RANO (Response Assessment in Neuro-Oncology) [22] and Response Evaluation Criteria in Solid Tumors (RECIST) criteria [23], which evaluate GB response to therapy based on changes in enhancing tumor size measured on contrast-enhanced T1-weighted images, as established by the RANO and earlier Macdonald criteria [24].

1.3 Radiomics

Beyond conventional visual assessment, advanced imaging techniques allow for the non-invasive characterization of a tumor's radiographic phenotype across different stages of treatment: before, during, and after therapy. Radiomics builds upon this capability by applying high-throughput computational algorithms to extract and quantitatively analyze a vast array of imaging features [25]. This automated approach enables the transformation of standard radiological scans into high-dimensional, mineable data, capturing subtle patterns in shape, texture, intensity, and spatial relationships that may not be discernible to the human eye. The aim of this approach is to support clinical decision-making by semi-automatically or automatically extracting radiologic features and linking them to clinical outcomes such as disease progression and patient survival. Due to the complex (intra-tumoral and inter-patient) heterogeneity of GB, advanced imaging methods like radiomics offer a valuable alternative to tackle the challenge of tumor heterogeneity, by capturing tumor diversity and guiding treatment based on image-derived features.

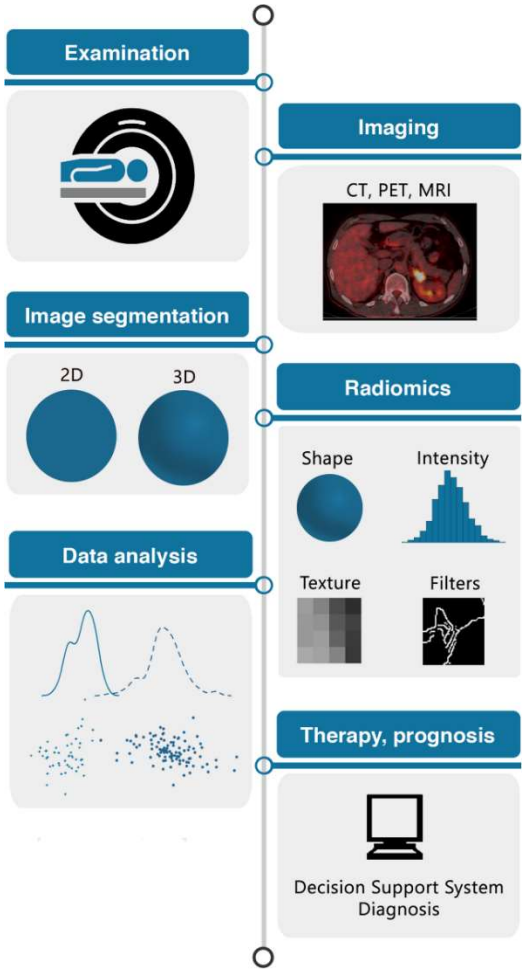


Figure 3 – Overview of the radiomics workflow from image acquisition to feature extraction and model-based clinical decision support. Following image acquisition and segmentation, radiomic features are extracted. Advanced statistical modeling and machine learning techniques are then employed for disease classification, patient subgroup identification, and individualized prognosis assessment. Image extracted from [26].

Because preclinical studies often rely on relatively small datasets (this is usually not the case for clinical trials which may involve hundreds of patients), some radiomics approaches treat each image voxel (i.e., short for "volume element", which associates with a pixel in the case of 2D MRI images) as an individual input, effectively expanding the number of data points and enhancing sensitivity to subtle features [27]. This voxel-wise strategy contrasts with whole-image analysis, which preserves broader contextual and spatial information. Figure 3 displays a typical radiomics workflow. At this point, it is worth noting that although this thesis is focused on radiomics at the preclinical level, the ultimate goal of our research is to develop radiomics tools that serve as support for clinicians' decision-making. In a typical radiomics workflow (Figure 3), the process begins with image segmentation, where regions or volumes of interest are defined. This can be done manually, semi-automatically, or automatically, with deep learning gaining popularity. Manual methods are time-consuming and observer-dependent, affecting reproducibility. In some cases, an image processing step is included, involving the standardization of pixel spacing and grey levels to ensure consistency; tools like 3D Slicer and pyRadiomics support this [25, 26]. Next, a feature extraction step is performed, where quantifiable features (intensity, shape, texture) are computed. Finally, feature selection or dimension reduction eliminates non-robust, redundant, or irrelevant features to avoid model overfitting and improve generalizability.

2. Description of the research questions

This Master's thesis aims to investigate whether machine learning classifiers trained on MRI-derived radiomics and morphological image features can effectively detect therapy-induced changes in the local tumor microenvironment in GB, using TMZ treatments on the GL261 murine GB model as a preclinical benchmark.

The long-term objective of this research is to enable earlier and more accurate assessment of GB therapy response by using quantitative MRI features to detect subtle microenvironmental

changes induced by treatment. By developing and comparing machine learning models trained on a comprehensive set of image features, this thesis aims to support faster, data-driven clinical decision-making in GB management.

3. Description of the dataset

The retrospective dataset used in this thesis consists of MRI images acquired from 63 mice (Table 1) generated as part of the research at Prof. Ana Paula Candiota's (Group of Biomedical Applications of Nuclear Magnetic Resonance, GABRMN) laboratory at Universitat Autònoma de Barcelona (UAB, Spain) [28-32].

Table 1 – Summary of the 63 animals harboring a GL261 GB tumor and analyzed in this thesis. The table also indicates the day post-inoculation when they were last studied (i.e., prior to euthanization), whether they were treated or control mice, and whether they were included in the training set or test set. The Cxxx notation corresponds to the internal unique mouse identifier code in the GABRMN group.

Training set		Test set	
Control	Treated	Control	Treated
C32-Day 16[1]	C414-Day 24[1]	C1109-Day 11[2]	C971-Day 26[2,3]
C69-Day 15[1]	C415-Day 23[1]	C1110-Day 13[2,3]	C975-Day 26[2]
C71-Day 16[1]	C418-Day 22[1]	C1111-Day 16[2,3]	C1023-Day 23[2]
C179-Day 17[1]	C419-Day 24[1]	C1111-Day 13[2]	C1026-Day 23[2]
C233-Day 17[1]	C437-Day 23[1]	C1320-Day 18[4,5]	C1100-Day 26[3]
C234-Day 17[1]	C521-Day 18[1]	C1344-Day 17[4,5]	C1108-Day 29[3]
C255-Day 14[1]	C525-Day 22[1]	C1348-Day 21[4,5]	C1412-Day 23[4,5]
C278-Day 19[1]	C526-Day 18[1]	C1457-Day 23[4,5]	C1445-Day 23[4,5]
C288-Day 18[1]	C527-Day 22[1]	C1459-Day 15[6]	C1447-Day 28[4,5]
C351-Day 13[1]	C572-Day 18[1]	C1461-Day 15[6]	C1450-Day 24[4,5]
C520-Day 18[1]	C574-Day 26[1]	C1462-Day 15[6]	C1451-Day 23[4,5]
C529-Day 18[1]	C575-Day 26[1]	C1465-Day 15[4,5]	C1456-Day 23[4,5]
C583-Day 18[1]	C584-Day 26[1]	C1466-Day 23[4,5]	C1458-Day 23[4,5]
	C586-Day 22[1]	C1471-Day 16[4,5]	C1460-Day 23[4,5]
	C776-Day 34[1]	C1472-Day 13[4,5]	C1463-Day 23[4,5]
	C795-Day 18[1]	C1474-Day 14[4,5]	
	C797-Day 22[1]		
	C808-Day 33[1]		
	C821-Day 34[1]		

1, Mice described in reference [29]. 2, Mice described in reference [28]. 3, Mice described in reference [30]. 4, Mice described in reference [32]. 5, Mice described in reference [33]. 6, Mice described in reference [31].

In all cases, these mice correspond to a GB preclinical model established by stereotactic injection of GL261 cells into the caudate nucleus of mice [29]. Treated mice were, in all cases, administered TMZ following either an Immune-Enhancing Metronomic Schedule (IMS) [28] or three-cycle administration protocol [30]. These animals were followed up with MRI studies and magnetic resonance spectroscopic imaging (MRSI) at chosen time points, although in this thesis we will only focus on the MRI images. However, as a result of this combined MRSI/MRI analysis, the number of MRI brain/tumor slices available for each mouse ranges from 1 to 4 (Figure 4A). Supplementary Table 1 describes additional details of these mice, including the number of slices available for each mouse.

The available MRI images are 256x256 pixels and are accompanied by a mask for each image/slice, determining the region of interest (ROI). The ROI indicates where the tumor is in the image and has been generated by experts from the GABRMN laboratory who are used to seeing MRI images as part of their daily research. A radiomics MATLAB software is used to extract 90 features from each image, which belong to two different types of radiomic features: texture features and Minkowski functions. The texture features are 42 and are described in Supplementary Table 2. These are classified into four different texture matrices: Gray Level Co-Ocurrence Matrix (GLCM), Gray Level Size Zone Matrix (GLSZM), Gray Level Run Length Matrix (GLRLM) and Neighbouring Gray Tone Difference Matrix (NGTDM) [34]. In contrast, the Minkowski functions capture morphological and structural aspects of image heterogeneity. For a chosen number of levels N (set to 16 in this thesis, based on prior experience in the GABRMN lab), N binary images are extracted using equally spaced intensity thresholds within the mask (Figure 4B). From each binary image, three features (i.e., area, perimeter, and Euler characteristics) are computed, yielding 3N (i.e., 48) features.

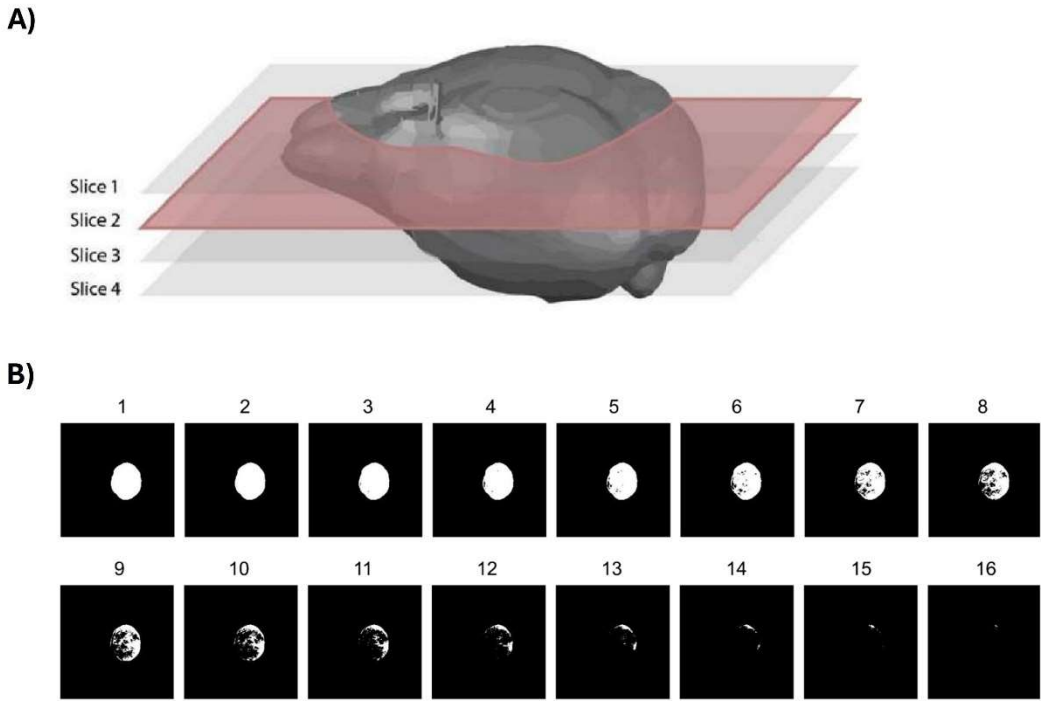


Figure 4 – Dataset and feature characteristics. A) Scheme showing the mouse brain and the positioning of brain slices. B) Illustration of the 16-level Minkowski thresholds for the C526-Day18 tumor mask. Adapted from [31].

4. Methodology

4.1 Generation of the preclinical glioblastoma model for *in vivo* studies

In this thesis, we employed the well-characterized, immunocompetent GL261 orthotopic murine model, which is widely used for GB research [28-30, 35, 36]. The GL261 mouse glioma cells were obtained from the Tumor Bank Repository at the National Cancer Institute (USA) and cultured in RPMI-1640 medium (Sigma-Aldrich, Spain), supplemented with 2.0 g/L sodium bicarbonate, 0.285 g/L L-glutamine, 10% fetal bovine serum (Gibco, Invitrogen, UK), and 1% penicillin/streptomycin (Sigma-Aldrich). Cells were cultured at 37 °C in a humidified incubator with 5% CO_2.

Female C57BL/6 wild-type mice (average weight: 21.1 ± 1.3 g) were obtained from Charles River Laboratories (France) and housed at UAB's animal facility. In particular, the mice described in references [32, 33], were housed in an enriched environment-like (EE-like) setting for three weeks prior to tumor implantation (Figure 5) and maintained there throughout the experiment to promote immune activation and reduce glioma growth, based on evidence that EE exposure enhances survival and brain immunological responses in C57BL/6 mice [37]. For tumor implantation, each animal received preemptive analgesia with subcutaneous Metacam (1 mg/kg; Boehringer Ingelheim, Germany) 15 min prior to anesthesia, and again at 24 and 48 h post-implantation. Anesthesia was induced by intraperitoneal injection of ketamine (80 mg/kg; Parke-Davis SL, Spain) and xylazine (10 mg/kg; Carlier, Spain). Anesthetized mice were placed in a prone position on a stereotaxic frame (Kopf Instruments, USA). Following scalp shaving and disinfection with an iodophor solution, a 1-cm incision was performed along the midline to expose the skull. Using a precision microdrill, a 1-mm burr hole was created at a point 0.1 mm posterior to the bregma and 2.32 mm lateral to the right of the midline (Fine Science Tools, Germany). A 26-gauge Hamilton syringe (Reno, USA), connected to a digital push-pull microinjector (Harvard Apparatus, USA), delivered 4 µL of a suspension containing 100,000 GL261 cells at a depth of 3.35 mm below the skull surface, with an injection speed of 2 µL/min. To minimize backflow, the syringe was maintained in position for an additional 2 min after the injection. It was then slowly and carefully withdrawn, and the incision was closed with 6.0 silk sutures (Braun, Spain). Post-operative care included recovery in a warm environment. Mice were monitored daily, with body weight recorded, and tumor progression was assessed two to three times per week using T2-weighted MRI (Figure 5).

4.2 Treatment of the GB mouse model

The analysis in this thesis was performed using animals that were treated with two distinct treatment protocols. The original protocol used in the laboratory consisted of three cycles of TMZ treatment (Figure 5A). TMZ (Sigma-Aldrich) was dissolved in 10% dimethyl sulfoxide (DMSO) in saline solution (0.9% w/v NaCl). In the three-cycle regimen, mice received a daily dose of 60 mg/kg for 5 consecutive days during the first cycle, followed by a 3-day rest period. The second cycle was administered over 2 consecutive days, followed by another 3-day rest. The third cycle was identical to the second, with a 2-day consecutive dose followed by a 3-day break. The treatment period was carefully structured to ensure that the rest periods were long enough for tumor cells unaffected by previous cycles to regain sufficient proliferative capacity. Control mice were treated with a 10% DMSO vehicle instead of TMZ. This treatment regimen was used in animals with low Cxxx numbers (i.e., animals with numbers lower than C1300). In our study, these mice correspond to the animals described in references [28-30].

Another treatment protocol used was the IMS-TMZ treatment. Tumor-bearing mice were treated at a dose of 60 mg/kg via oral gavage every 6 days, beginning on day 11 post-implantation (Figure 5B). Control mice were treated with a 10% DMSO vehicle. As described for the three-cycle regimen, the IMS-TMZ treatment protocol followed the principle of providing appropriate recovery times between treatments, while monitoring tumor progression using MRI to evaluate treatment efficacy. This treatment regimen was used in animals with high Cxxx numbers (i.e., animals with numbers higher than C1300). In our study, these mice correspond to the animals described in references [32, 33].

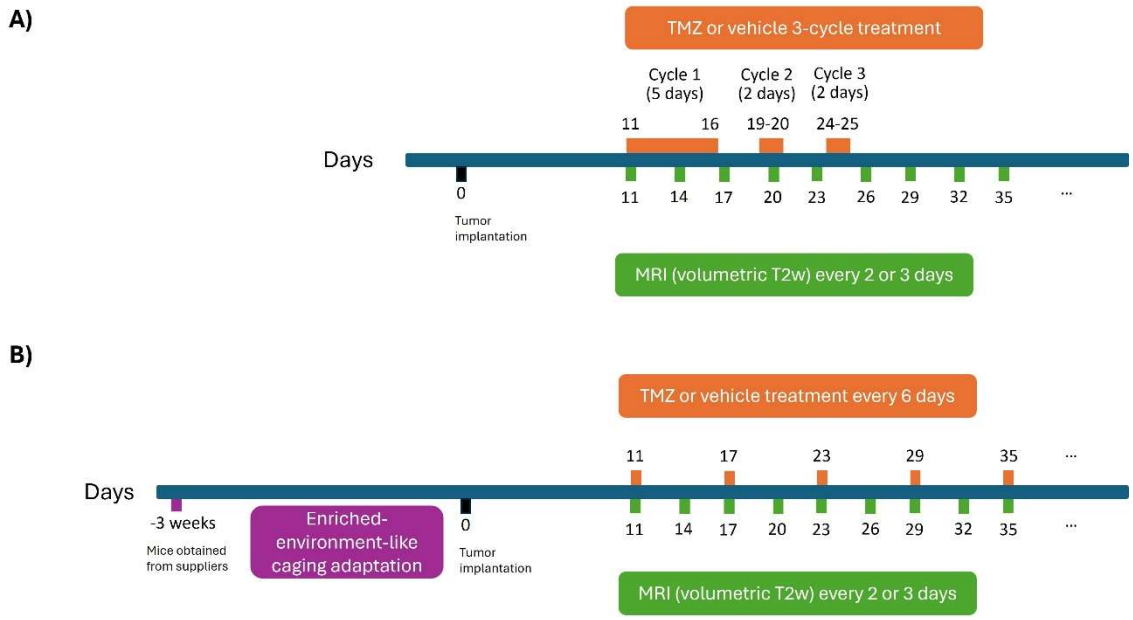


Figure 5 – A) Three-cycle TMZ and B) IMS-TMZ treatment regimens.

In all cases, tumor volumes were assessed through T2-weighted MRI at day 11 after implantation, with randomization ensuring homogeneous tumor sizes and body weights among the experimental groups. MRI scans were performed twice a week to monitor the tumor's response to therapy, and animal health was carefully monitored throughout the study. Upon reaching predetermined endpoint criteria, animals were euthanized by cervical dislocation, and the brain and tumors were resected for analysis. All animal studies were approved by the local ethics committee (https://www.uab.cat/en/research-ethics/animal-experimentation, accessed on 27 July 2025), according to the regional and state legislation.

4.3 MRI analysis

In vivo MRI/MRSI studies were conducted at the joint nuclear magnetic resonance facility of UAB and the Centro de Investigación Biomédica en Red (Ciber)-Bioingeniería, Biomateriales y Nanomedicina (BBN), at Unit 25 of NANBIOSIS (https://www.nanbiosis.es, accessed on 26 July 2025). All experiments were performed on a 7T Bruker BioSpec 70/30 USR spectrometer (Bruker BioSpin GmbH, Germany) equipped with a mini-imaging gradient system of 400 mT/m. For MRI acquisitions, radiofrequency transmission was achieved using a 72 mm inner-diameter linear volume coil, while signal reception was performed with a mouse brain surface coil.

Animals were placed on a dedicated support platform equipped for anesthesia delivery (isoflurane, 1.5-2.0% in oxygen at 1 L/min) and containing a circulating warm-water system to stabilize body temperature. Breathing rate was tracked using a pressure sensor, ensuring a respiratory frequency of 60-80 breaths per minute.

GL261 tumor-bearing mice were first assessed using high-resolution coronal T2-weighted images obtained through a Rapid Acquisition with Relaxation Enhancement (RARE) sequence to identify brain tumor presence and track its stage of development. The MRI acquisition parameters were as follows: repetition time/effective echo time = 4200/36 ms; echo train length= 8; field of view

= 19.2 × 19.2 mm; matrix size = 256 × 256 (75 × 75 µm/pixel); slice thickness = 0.5 mm; inter-slice thickness = 0.1 mm; number of slices = 10; number of averages = 4; and total acquisition time = 6 min and 43 s.

4.4 Radiomics workflow

We performed a radiomics analysis of the MRI images to extract quantitative features from tumor regions in the GL261 GB mouse model. Figure 6 describes the general workflow of radiomics extraction and subsequent analysis. Tumor ROIs 2D masks were delineated manually under the advice and supervision of experts in MRI GB imaging. The purpose of these masks is to avoid the inclusion of surrounding peritumoral and normal tissues in the analysis. This segmentation process ensured that only the tumor tissue was considered for feature extraction, thereby enhancing the precision of the analysis.

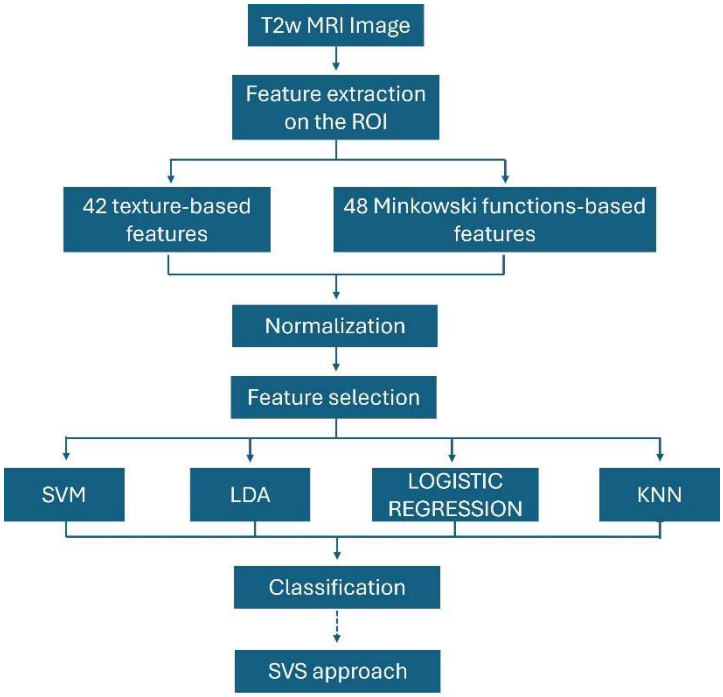


Figure 6 – Image analysis workflow.

All data were processed using MATLAB (version 23.2.0.2859533, R2023b, MathWorks, USA), utilizing the Image Processing Toolbox (version 23.2) and specialized radiomics toolboxes for feature extraction [38, 39]. Radiomic features for the analysis were derived from both texture matrices and Minkowski functions. Specifically, texture features were obtained using the Radiomics MATLAB toolbox [39], which considers four types of texture matrices: Gray Level Co-Occurrence Matrix (GLCM), Gray Level Size Zone Matrix (GLSZM), Gray Level Run Length Matrix (GLRLM), and Neighboring Gray Tone Difference Matrix (NGTDM). These matrices provide various representations of the tumor's texture. As a result, a total of 42 texture features were extracted based on these matrices, as shown in Supplementary Table 2. To calculate these features, a rectangular bounding box was generated around the segmented tumor region, ensuring that the matrix input matched the required dimensions for feature extraction.

In addition to the texture features, Minkowski functions were calculated directly from the tumor mask using the Minkowski MATLAB toolbox [38]. The Minkowski functions are used to describe

the morphological and structural properties of the tumor and its heterogeneity. The process involved the creation of N binary images (where N was set to 16 based on prior experience in the GABRMN lab) by applying equally spaced intensity thresholds within the tumor mask (Figure 4b). For each threshold, three distinct morphological features (i.e., area, perimeter, and Euler characteristics) were computed, providing a comprehensive description of the tumor's structure. This procedure resulted in a total of 48 Minkowski-based features (3 features per threshold level across 16 levels). These features serve as critical descriptors of the tumor's geometric properties, enhancing the overall radiomics analysis by providing insights into the tumor's complexity and heterogeneity.

After feature extraction, features were normalized using the z-score method using the *normalize* MATLAB function. This method centers the data to have a mean of 0 and scales it to have a standard deviation of 1.

To provide a consolidated analysis for multi-slice tumor images, a slice-voting system (SVS) was employed (Figure 7). This method aggregates the outputs from multiple tumor slices by computing a weighted average, with greater emphasis placed on slices that contained a larger proportion of tumor tissue. Weights for each slice were determined based on the number of pixels within the tumor mask to reflect the relative importance of each slice in the overall analysis.

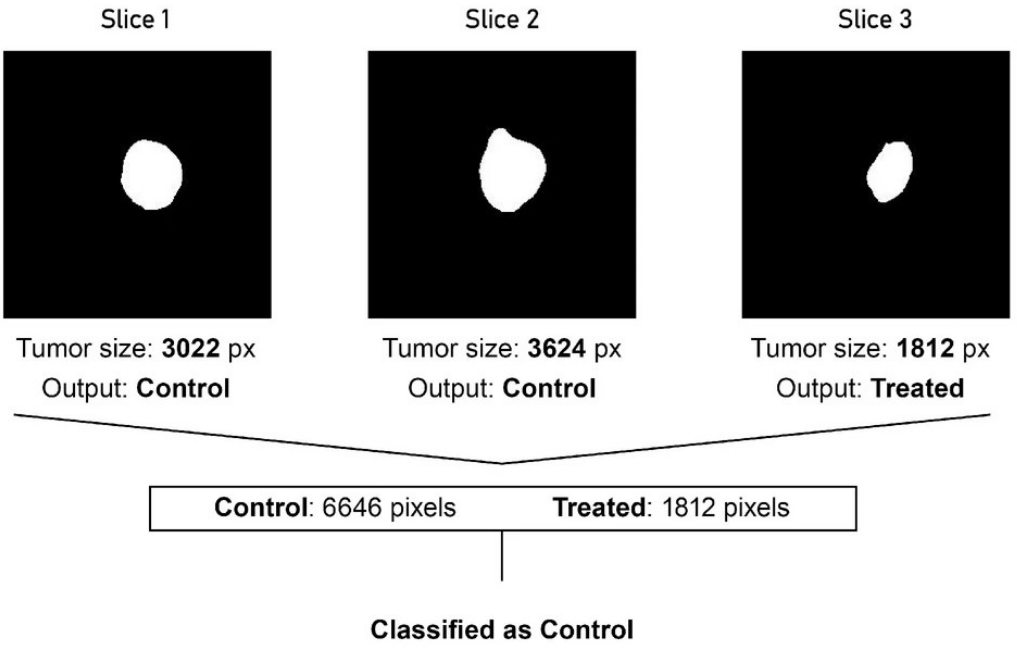


Figure 7 – Example of the SVS system used to aggregate the output of three different slices in the case of a control mouse. Image adapted from [31].

In summary, the radiomics feature set included 90 distinct features (42 texture features and 48 Minkowski features) per slice. Moreover, the outputs of different slices are aggregated to generate a unique output for each mouse.

4.5 Model fitting

To assess the performance of different classifier models, several methods were tested using MATLAB and its Bioinformatics Toolbox (version 23.2), and the Statistics and Machine Learning

Toolbox (version 23.2). Several methods were evaluated, including Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Logistic Regression (LR), and K-Nearest Neighbors (KNN) (Figure 6). Each method was applied to the training data using different configurations (e.g., different kernel types for SVM and distance metrics for KNN). The selected classification models enable evaluating a range of algorithms, each with distinct characteristics and strengths, to determine which is best suited for our dataset.

The LR analysis was performed using the *fitglm* function within MATLAB. The 'Distribution' parameter was set to 'binomial' and the 'link' parameter was set to 'logit' to model the log-odds of the success class.

The SVM analysis was performed with the *fitcsvm* MATLAB function, setting the 'KernelScale' parameter to 'auto' to automatically adjust the kernel's scale to fit the data and optimize the model performance. The 'BoxConstraint' was set to 1 to balance the trade-off between minimizing the classification error and maximizing the margin between classes. The 'Standardize' setting was set to 'true' to standardize the input features to have zero mean and unit variance, ensuring equal contribution of all features during training. The evaluated kernel types include linear, Gaussian, and polynomial (with degrees 2, 3, and 4). The polynomial kernels allow for varying degrees of complexity, with higher degrees enabling more complex decision boundaries.

The LDA analysis was performed using the *fitcdiscr* MATLAB function, which trains an LDA classifier on the input data. The 'DiscrimType' parameter was set to 'linear' to create a linear decision boundary between classes, while the 'Gamma' parameter was set to 0 ensuring that no regularization is applied to the model, while setting FillCoeffs was set to 'off' to ensure that no coefficient filling occurs, allowing the model to remain unaltered by any automatic adjustments that MATLAB might otherwise perform on the coefficients.

The KNN analysis was performed using the *fitcknn* MATLAB function, with the following specified settings. The 'Distance' parameter was set to different distance metrics, including 'Euclidean', 'Cosine', and 'Minkowski', to test different similarity measures between data points. The 'Exponent' was set for the Minkowski distance to 3 for testing the influence of the distance's exponent. The 'NumNeighbors' parameter was adjusted across different values, such as 1 or 10 for the different models, determining how many neighbors influence the classification. The 'DistanceWeight' was set to either 'Equal' or 'SquaredInverse' to explore different weighting schemes for neighbors, where 'SquaredInverse' places more weight on closer neighbors. Additionally, the 'Standardize' option was set to true for all models to ensure that the input features were standardized to have zero mean and unit variance, allowing for standardized comparisons between features with different scales. The evaluated models include fine, medium, cosine, cubic, and weighted KNN configurations, with different combinations of distance metrics and neighbor settings, allowing for exploration of various complexity levels and trade-offs in classification accuracy. A detailed description of the fine, medium, cosine, cubic, and weighted KNN configurations can be found in *Section 5. Results*.

In terms of model performance metrics, after training, the models' predictions were compared to the response labels of the test datasets to assess the best-performing methods. The following performance metrics were used:

- Accuracy: The accuracy of the model was determined by calculating the classification error rate, following the equation Accuracy = (TP+TN)/(TP+TN+FP+FN), where FP refers to false positives; TP refers to true positives; FN refers to false negatives and TN refers to true negatives.
- Sensitivity: This metric was calculated by assessing the proportion of actual positive cases (i.e., treatment-responsive mice) correctly identified by the model. It was calculated using the formula: Sensitivity = TP/(TP+FN)
- Specificity: This metric was calculated by assessing the proportion of actual negative cases (i.e., non-responder mice) that were correctly predicted by the model. This metric was calculated using the formula Specificity = TN/(TN+FP).
- F1-score: The function calculates the F1 score using the confusion matrix to assess the balance between precision and recall. The F1-score was calculated following the equation: F1-score = 2*Prec*Rec/(Prec + Rec) = 2*TP/(2*TP+FP+FN), where Prec = Precision and Rec = Recall.

A leave-one-out cross-validation (LOOCV) strategy was employed to evaluate model performance. In this approach, the models were iteratively trained on all but one subject and tested on the left-out subject, such that each mouse served once as the independent test set.

The MATLAB software used in this thesis can be found at https://gitfront.io/r/stanco78/aJGZ76Z8GDrN/Radiomics/, where additional details can be found. The main scripts used for the analysis are located in the *src* folder.

4.6 Feature selection

Given that our dataset displays a large number of features (i.e., 90) but a limited number of observations (i.e., 32), we decided to reduce the number of features in our models. We expect that many features will not contribute meaningfully to the model learning outcome and, instead, can cause the model to overfit the noise present in the data. Additional expected benefits of feature selection include the reduction of computational resources for model usage and the enhancement of the interpretability of the results.

For comparison purposes, we followed a similar feature selection strategy to the one described in Núñez *et al.* [31]. Thus, we applied two different approaches for feature selection: a filter scheme and an embedded-wrapper recursive feature elimination (RFE) scheme. For the filter scheme, a univariate t-test was performed to rank features based on their individual significance in distinguishing between control and treated groups. For this purpose, the *rankfeatures* MATLAB function was used.

In the embedded-wrapper RFE scheme, the method iteratively removes the least relevant features, starting from the complete set. For the MRI dataset used in this thesis, a pure RFE wrapper method failed to distinguish feature relevance in the early steps, as most accuracies remained unchanged. Consequently, an embedded RFE method was first employed to select features, and a wrapper feature selection was applied to further reduce the number of selected features, optimizing the model for improved performance.

5. Results

This Master's thesis derives from the research published in Scientific Reports by Núñez *et al.* [31], which highlighted the potential of both MRI-based radiomics and MRSI-based source extraction in assessing TMZ therapy response in GB, particularly in distinguishing treated from untreated tumors. The research team led by Prof. Candiota showed that MRI-derived radiomics, based on the use of texture and structural image features, shows adequate accuracy when combined with feature selection techniques like a t-test-based filter method or an RFE method. A key assumption of this analysis was that linear regression models would be the most appropriate for such an analysis, given that they are expected to provide a low degree of over-fitting on our limited number of samples. The dataset used in this research piece also contained MRSI data, which can be used to distinguish between treated and untreated tumors with higher accuracy when compared to the radiomics approach.

In this thesis, we will concentrate exclusively on the MRI dataset, as MRI analysis relies on equipment that is widely available in universities and hospitals, making it a more accessible choice. Moreover, MRI imaging is routinely used as a standard procedure to monitor brain tumor volume in preclinical mouse studies, offering valuable insights into tumor growth and treatment efficacy. Additionally, MRI serves as the primary imaging modality in hospitals for diagnosing and managing GB in patients, due to its ability to provide detailed structural images of the brain. In contrast, while MRSI holds significant potential for enhancing classification methods in GB [28, 30, 31, 33], its use is limited by the availability of specialized, high-end instrumentation. Such equipment is typically found only in select research facilities and highly specialized hospitals, restricting its broader application as a potential classification software in clinical settings.

We will assess whether the previous assumption that linear regression models are the most suitable for this analysis, due to their expected low tendency to overfit, holds true. For this purpose, we will test several alternative classification methods, including Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), and K-Nearest Neighbors (KNN), each with a variety of different configurations, rendering algorithms with diverse degrees of complexity, to assess which is the best-suited for our dataset.

Finally, we will assess different feature selection algorithms to reduce the dimensionality of the dataset by preserving essential information while discarding redundant or irrelevant features, which can improve model performance. High-dimensional data increases the risk of overfitting and complicates the detection of meaningful patterns, as irrelevant or redundant features expand the search space. Feature selection, in particular, is an important process for extracting consistent and relevant insights from the data, ensuring that only the most valuable features are retained.

Ultimately, the selection of the optimal feature selection algorithm is dependent on the characteristics of the data. **Wrapper methods** tend to produce excellent results, albeit at a higher computational cost, as they evaluate subsets of features based on their impact on model accuracy. In contrast**, filter methods** assess feature relevance independently of the classifier, with feature selection occurring before the classifier is trained. A third approach, the **embedded method**, integrates feature selection within the classifier itself, assigning a relevance score to each feature based on the model's internal structure. For comparison purposes, we will follow a

similar approach as the one used in Núñez *et al.* [31]: we will use a filter scheme that uses a univariate t-test to rank features by its ability to distinguish between control and treated groups, and will use a more complex embedded-wrapper RFE scheme that combines both an embedded method followed by a wrapper method (as described in *Section 4.6. Feature selection*)

5.1 Logistic regression

The first step was to reproduce the LR output described by Núñez *et al.* [31] when using the same dataset. For this, we adapted and further developed a MATLAB code that was originally developed for the mentioned research publication, which was not complete. For additional information about the software, check *Section 12. Software code.*

Figure 8 displays the accuracy obtained when the t-test-based filter feature selection method is used. As described in Figure 8 (left), our software can reproduce the published data [31] with small deviations from the published accuracies for the slice classification approach. Thus, we reproduce LR models that use radiomic features to moderately distinguish between TMZ-treated and control cases in the preclinical dataset. The LR models' outcomes exhibit a somewhat irregular pattern with test accuracy generally falling around 60%, with the overall accuracy declining as the number of selected features increases [31]. Table 2 displays the first 30 ranked features based on the t-test filtering method.

In contrast, our models differ markedly from the published results in the case of the SVS approach. By aggregating the results of the different slices available for a given mouse, Núñez *et al.* report that the SVS approach (Figure 8, right) is able to reach a higher test accuracy (i.e., 87%) when 10 features are selected. Our models show a more limited improvement in accuracy with the SVS approach (i.e., 74% with 7 selected features and 77.5% with 18 selected features). After inspection of the MATLAB code, we observed that the data presented in Núñez *et al.* appear to be incorrectly calculated because the MATLAB function used does not aggregate the data as described in Figure 7. Núñez *et al*. appear to have used the *PredictionReport* function shown in *Section 12. Software code*. This function appears to aggregate classifier predictions per mouse, compute individual accuracies per model, and derive a "voting" label based on rounded accuracy per model (i.e., binarizing whether a model is accurate per mouse). When the *PredictionReport* function is used, we can replicate the published data [30] with small deviations (Supplementary Figure 1).

Supplementary Figure 2 displays the specificity and sensitivity of our LR models (i.e., Núñez *et al.* values are not available). While the specificity is high in general (e.g., the specificity of the SVS approach is 100% for models using 1 to 3 selected features), model sensitivity tends to be low, although it reaches 80% for 18 selected features and the SVS approach. Supplementary Figure 17 displays the model performance, evaluated using the F1 score.

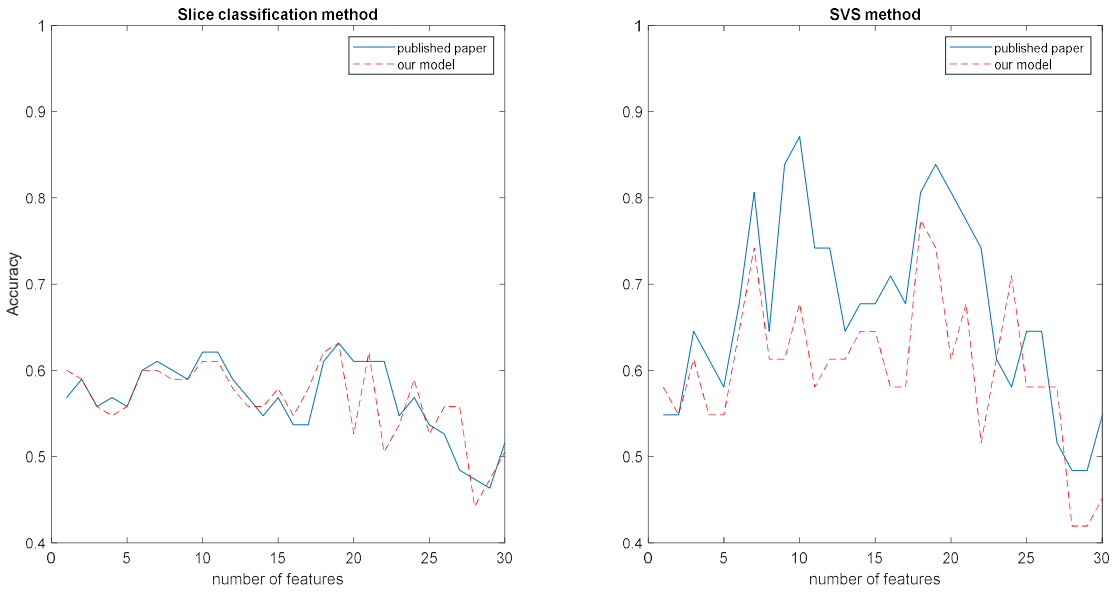


Figure 8 – LR radiomics model performance, in terms of model accuracy over the hold-out set, as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. On the left, the accuracy is evaluated for each slice, and on the right, the SVS approach was applied to aggregate all the slices for a specific mouse.

The embedded-wrapper method was more challenging to reproduce, given the lack of details in the description available for this strategy in the article of Núñez *et al.* [31]. The rationale for adopting a combined embedded-wrapper strategy is that the pure RFE wrapper method struggled to identify feature relevance in the early stages, with most accuracies remaining unchanged. This issue was also evident when we applied our model, further highlighting the need for an integrated approach. A detail not specified in the reference publication is how many features are selected using the embedded method, before changing to the wrapper method. We evaluated different feature numbers at which to transition from the embedded to the wrapper method (i.e., we tested m=50, m=60 and m=70) and found that 60 provides the optimal results in terms of model accuracy. Figure 9 (left) shows that for high feature numbers, we are able to reproduce the accuracy values described in Núñez *et al.* [31], but bigger differences can be observed for low numbers of selected features. Núñez *et al.* described that the embedded-wrapper method reaches 75% test accuracy with two features (i.e., GLCM Entropy and Perimeter9); in our results, 9-11 features are needed to obtain a similar accuracy. When the results are aggregated with the SVS approach, while Núñez *et al.* do not show a significant improvement in model performance as a result of the slice aggregation, our model reaches 87% accuracy with 9 features.

Table 2 displays the first 30 ranked features based on the embedded-wrapper RFE method according to Núñez *et al.* Supplementary Figure 3 displays the specificity and sensitivity of our LR models (i.e., Núñez *et al.* values are not available) constructed using the embedded-wrapper feature selection method. Supplementary Figure 18 displays the performance of these models, evaluated using the F1 score.

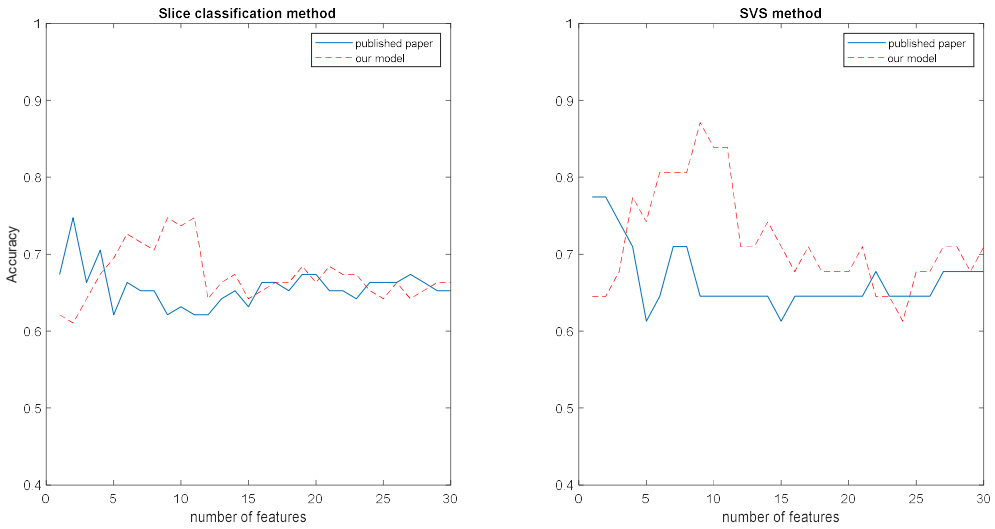


Figure 9 – Embedded-wrapper LR radiomics model performance, in terms of model accuracy over the hold-out set, as a function of the number of selected features (i.e., from 1 to 30 features). On the left, the accuracy is evaluated for each slice, and on the right, the SVS approach was applied to aggregate all the slices for a specific mouse.

In summary, we were able to reproduce with our MATLAB code most findings described for the MRI dataset in Núñez *et al.*, although the embedded-wrapper method is more difficult to reproduce (e.g., in terms of the optimal number of features). As a result, we decided to use our MATLAB code and adapt it to other classification algorithms to determine the optimal approach for our TMZ-treatment MRI dataset.

5.2 Alternative models using a filter-based feature selection method

We started by evaluating how SVM models perform in our MRI dataset. An SVM approach classifies data by finding the optimal hyperplane that separates classes with the maximum margin. It works well in high-dimensional spaces and can handle non-linearly separable data using kernels, which map input data to higher dimensions where separation becomes possible. Common kernels include linear, Gaussian (also known as radial basis function), and polynomial of different degrees. The linear kernel is best for linearly separable data, while the Gaussian kernel handles more complex non-linear patterns by creating smooth decision boundaries. Polynomial kernels can model feature interactions depending on the polynomial degree chosen. Thus, we evaluated SVM models with kernel types that include linear, Gaussian, and polynomial (with degrees 2, 3, and 4).

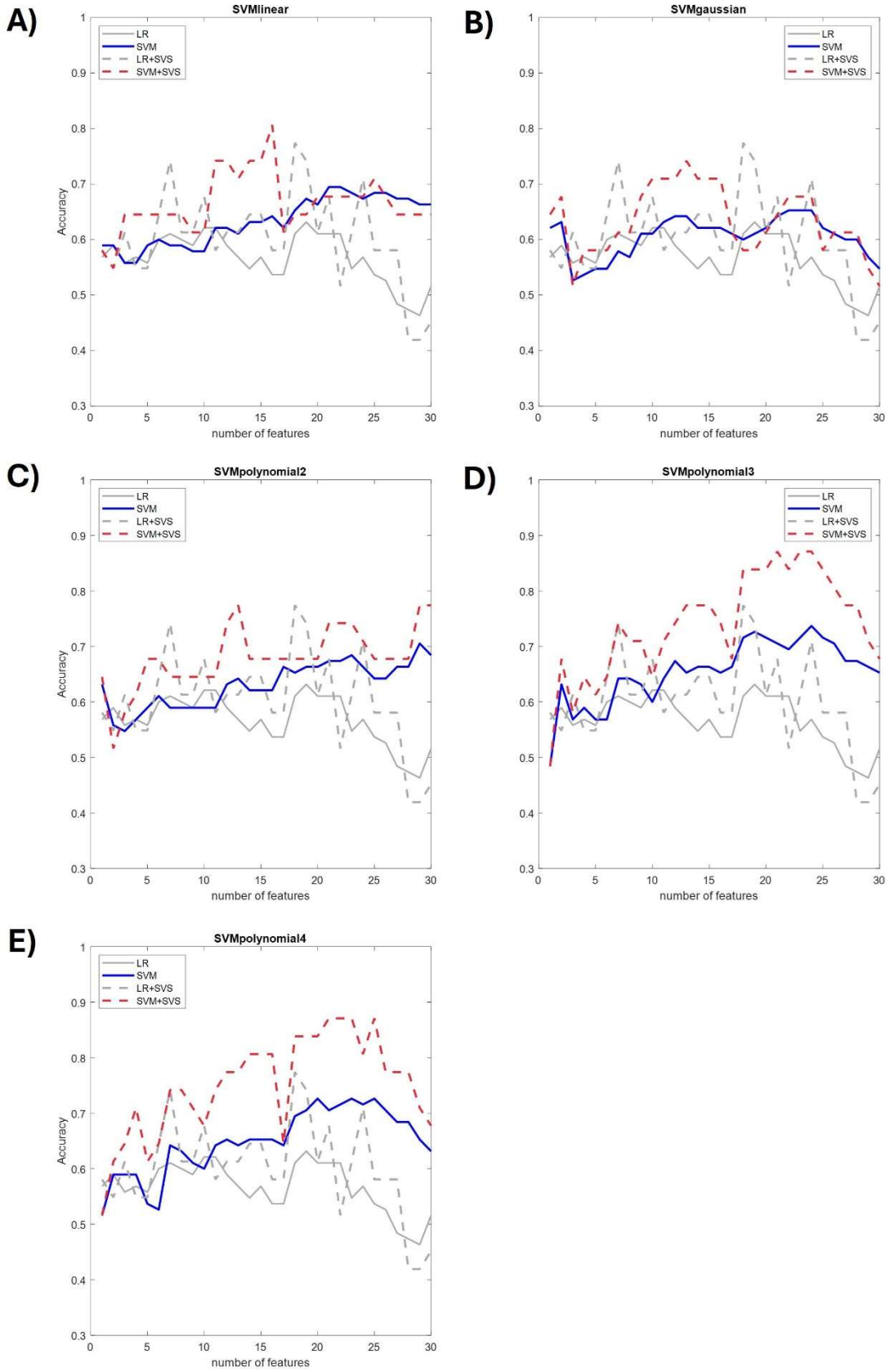


Figure 10 – Performance (i.e., test accuracy) of different SVM radiomics models as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. Different kernel types were evaluated, including A) linear kernel, B) Gaussian kernel, C) polynomial kernel with degree 2, D) polynomial kernel with degree 3, and E) polynomial kernel with degree 4. The performance of the LR models is plotted as a reference.

Figure 10 shows the performance of SVM models both when considering each slice independently (i.e., slice classification approach) and when the SVS approach is used. Figure 10 compares the SVM model performance in terms of classification accuracy with that of the LR models. SVM models display, in general, better performance when compared to LR in the case of the slice classification approach. In particular, the SVM classifiers with a Gaussian kernel underperformed relative to the other SVM variants (i.e., they performed more similarly to the LR models). In contrast, the test accuracy of the SVM models with polynomial kernels of degree 3 and 4 is especially high. While the SVM of polynomial kernel degree 3 reaches 73% accuracy for the model with 25 selected features, when used in the SVS approach, the model with 23-24 features reaches 87% accuracy. Similarly, the SVM model with a polynomial kernel of degree 4 with 21 or 25 selected features achieves 73% accuracy, but the accuracy increases to 87% with 21-23 features when integrated into the SVS approach. Supplementary Figure 4 to Supplementary Figure 8 display the specificity and sensitivity of these models. In general, these models display a better balance between specificity and sensitivity. For example, the SVM models with a polynomial kernel of degree 3 (when used in the SVS approach) reach both 87% specificity and 87% sensitivity with 23-24 selected features. The same occurs with the SVM model with polynomial kernel degree 4, with 21-23 filter-based selected features. Overall, we observe a general trend in which, as the number of selected features increases, model specificity decreases, and model sensitivity increases. Supplementary Figure 19 to Supplementary Figure 21 display the performance of these models in terms of the F1 score.

We also evaluated LDA models, which classify the observations by projecting high-dimensional data onto a lower-dimensional space to maximize class separability. In a binary classification problem like the one studied in this thesis, LDA assumes that both classes are normally distributed with identical covariance matrices but different means. LDA finds a linear combination of features that best separates the two classes by maximizing the ratio of between-class variance to within-class variance. Figure 11, Supplementary Figure 9 and Supplementary Figure 22 show, respectively, the accuracy, specificity/sensitivity and F1 score of the fitted models. The LDA models perform similarly or slightly worse than the published LR models.

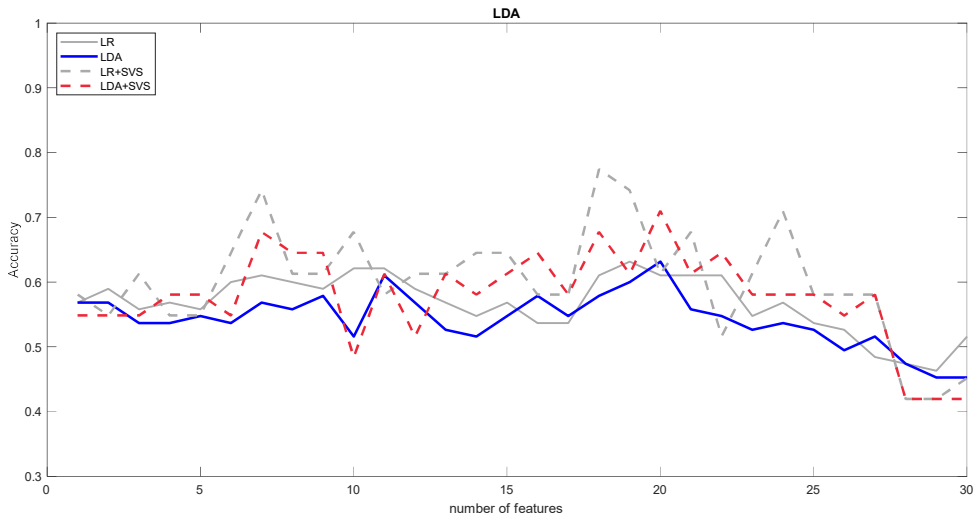


Figure 11 – LDA radiomics model performance, in terms of model accuracy over the hold-out set, as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. The performance of the LR models is plotted as a reference.

Finally, we evaluated different KNN models, which are non-parametric classifiers. In binary classification, KNN predicts the class of a new data point based on the majority class among its *k* closest neighbors in the feature space, typically measured using Euclidean distance. These models do not make assumptions about data distribution and require no training phase. Different KNN model configurations were evaluated (Figure 12): the KNN 'fine' configuration uses Euclidean distance, a *k* value of 1 (also known as nearest neighbor interpolation) and an equal DistanceWeight (i.e., each neighbor gets equal weight); the KNN 'medium' configuration uses also Euclidean distance but using a *k* value of 10 and an equal DistanceWeight; the KNN 'cosine' configuration uses the Cosine distance (i.e., one minus the cosine of the included angle between observations), using a *k* value of 10 and an equal DistanceWeight; the KNN 'cubic' configuration uses a Minkowski distance with exponent 3, with a *k* value of 1 and an equal DistanceWeight; and the KNN 'weighted' configuration is similar to the 'medium' configuration but uses the SquaredInverse distance weight (i.e., each neighbor gets weight equal to $1/d^2$, where d is the distance between this neighbor and the point being classified). Figure 12 compares the KNN performance in terms of classification accuracy with that of the LR models. As in the SVM case, KNN models display generally better performance when compared to LR in the case of the slice classification approach. The KNN fine model with 12 selected features appears to perform best with an accuracy of 71.5% in the case of slice classification (i.e., slightly worse than the best SVM models) and 90% in the case of the SVS approach (i.e., slightly better than the best SVM models). Supplementary Figure 10 to Supplementary Figure 14 display the specificity and sensitivity of these models. Supplementary Figure 23 to Supplementary Figure 25 display the performance of these models in terms of the F1 score.

It is worth noting that the list of selected features in the t-test filter method is the same for all the evaluated models in this section (Table 2), as the list does not depend on the model itself.

5.3 Feature selection using recursive feature elimination

Next, we performed feature selection for some of the models tested in the previous section, but following a similar embedded-wrapper RFE procedure to the one applied by Núñez *et al*. for LR [31]. To this end, we chose to focus on SVM models with polynomial kernels of degrees 3 and 4, as they were among the top-performing models in the previous section and are well-suited for incorporating an embedded-wrapper RFE method.

Figure 13 and Figure 14 display the performance in terms of test accuracy for SVM models with a polynomial kernel of degree 3 and 4, respectively. Similarly, Supplementary Figure 15 and Supplementary Figure 16 represent the test specificity and sensitivity of these models, and Supplementary Figure 26 and Supplementary Figure 27 represent the F1 score of these models. We observe that the SVM models with a polynomial kernel of degree 3 do not display improved performance when compared to the LR models generated based on the embedded-wrapper RFE method (Figure 13). The specificity of these models varies between 50%-60% (Supplementary Figure 15).

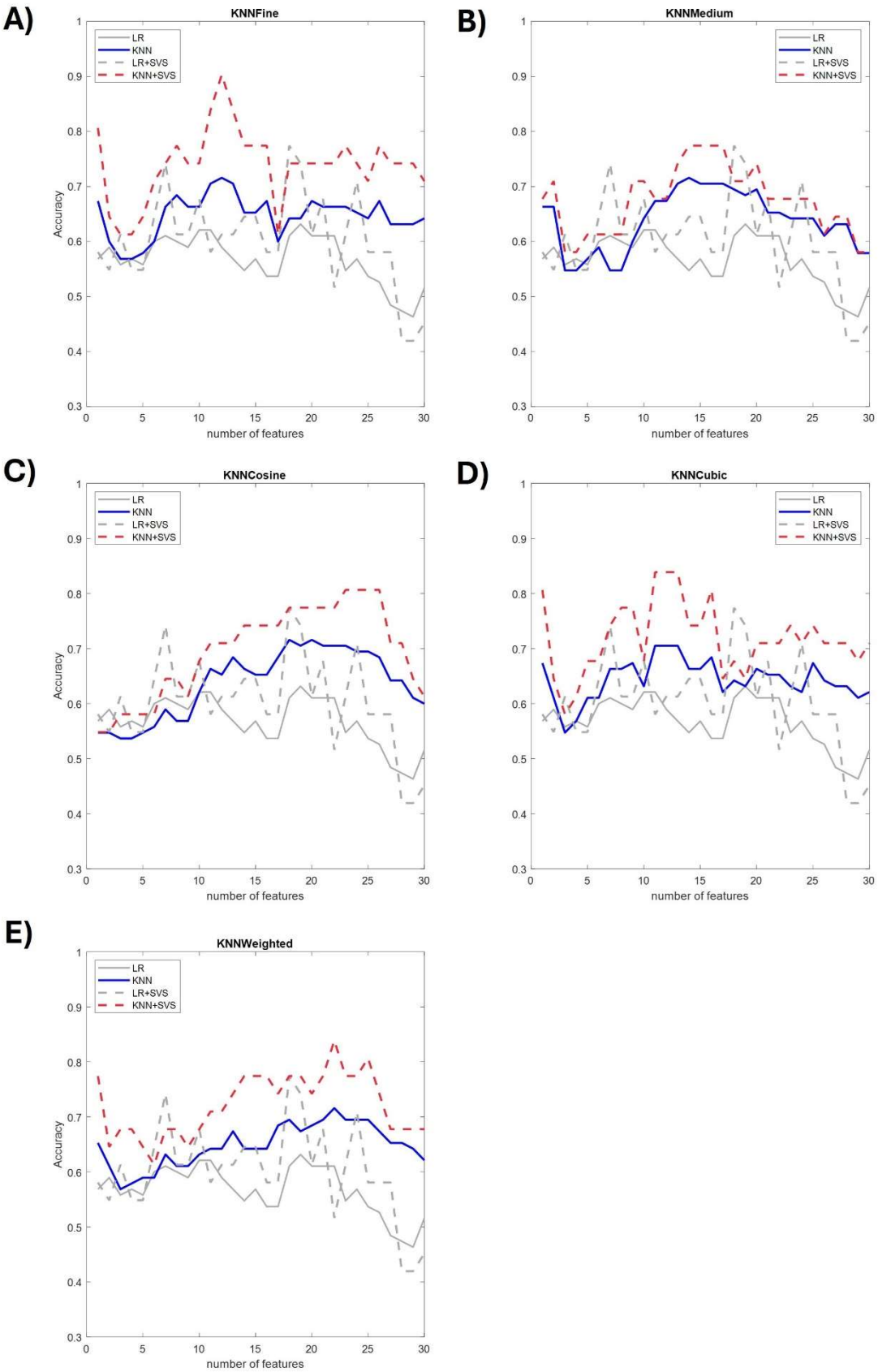


Figure 12 – Performance (i.e., test accuracy) of different KNN radiomics models as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. Different KNN model configurations were evaluated, including A) fine, B) medium, C) cosine, D) cubic, and E) weighted (see main text for a full description of these configurations). The performance of the LR models is plotted as a reference.

In contrast, the SVM model with 14 selected features with a polynomial kernel of degree 4 (Figure 14) performs slightly better than the LR models for the slice classification approach, reaching 77% accuracy. For the SVS approach, the models with 14-15 selected features perform worse than equivalent LR models, reaching 80.5% accuracy. In terms of specificity/sensitivity, the SVM model with 14 selected features and a polynomial kernel of degree 4 provides a 75% specificity and 87% sensitivity (Supplementary Figure 16).

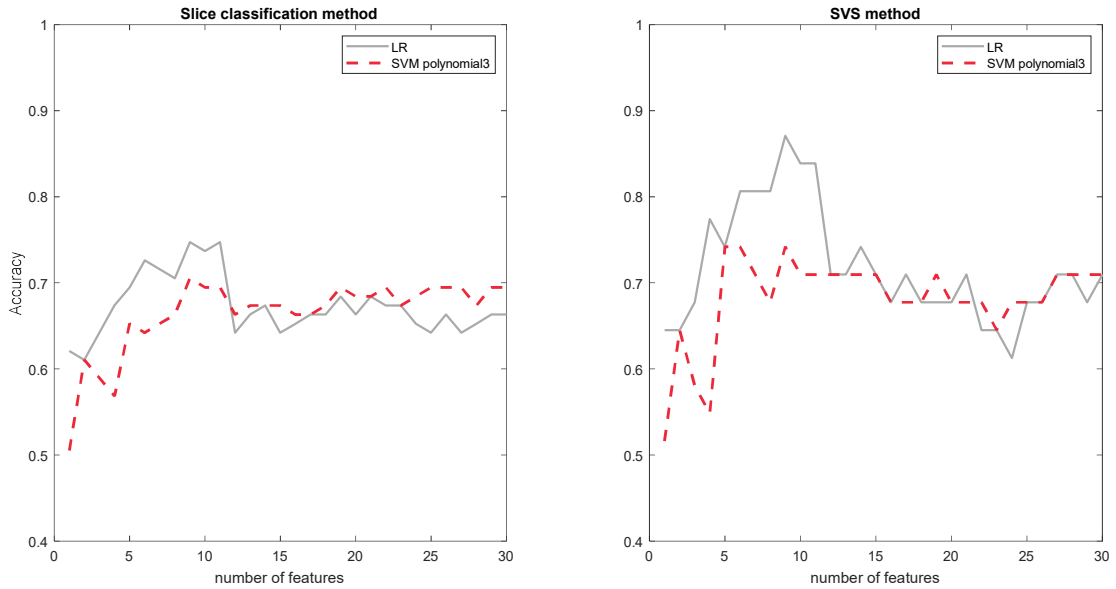


Figure 13 – Test accuracy of SVM radiomics models with a polynomial kernel of degree 3. Model performance is plotted as a function of the number of selected features (i.e., from 1 to 30 features), which were selected using an embedded-wrapper RFE method, in which the embedded RFE feature selection method selects features until 60 features remain in the model, and the wrapper RFE further continues selecting features.

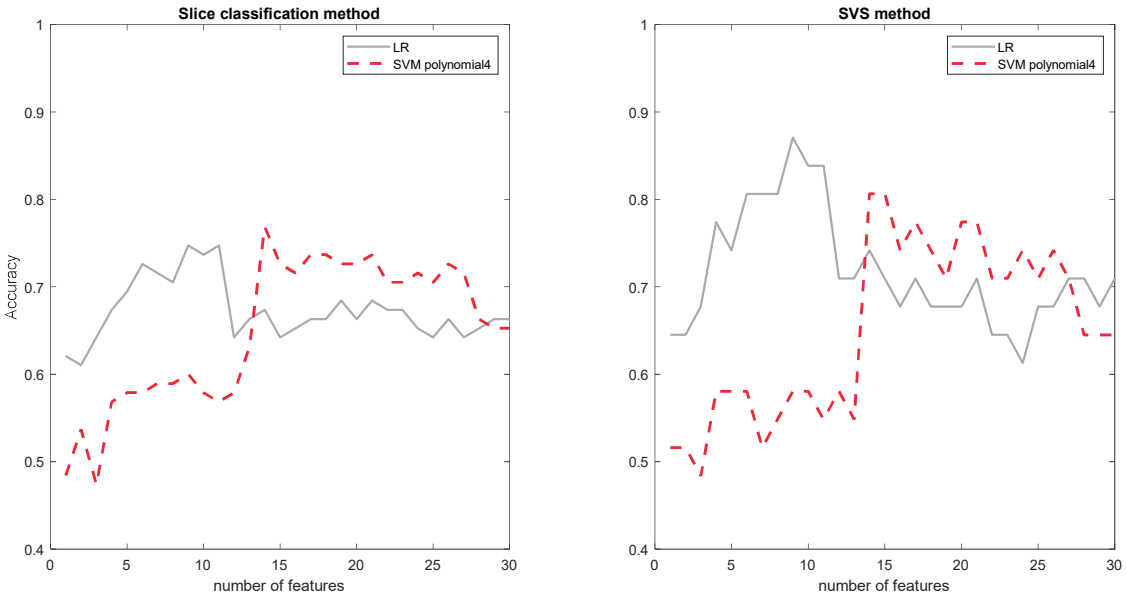


Figure 14 – Test accuracy of SVM radiomics models with a polynomial kernel of degree 4. Model performance is plotted as a function of the number of selected features (i.e., from 1 to 30 features), which were selected using an embedded-wrapper RFE method, in which the embedded RFE feature selection method selects features until 60 features remain in the model, and the wrapper RFE further continues selecting features.

Table 2 displays the 14 selected features in the best-performing model shown in Figure 14. In addition, Table 2 displays the 30 top features in the Figure 14 models and compares them with the t-test-based filter selected features and those published by Núñez *et al*. [31]. Several features selected in the SVM models are also identified by other techniques, including GLRLM-SRE, GLCM-Homogeneity, GLRLM-SRLGE, Euler 15, and various perimeter- and area-related parameters. However, the overlap of the different strategies (Table 2) is quite limited (Figure 15).

Table 2 – Features ranked by the t-test filter method, the embedded-wrapper LR RFE feature selection method [31] and the embedded-wrapper RFE SVM feature selection method.

Ranking	T-test filter method	Embedded-Wrapper RFE LR	Embedded-Wrapper RFE SVM polynomial 4
1	Perimeter9	GLCMEntropy	Perimeter16
2	Perimeter8	Perimeter9	GLRLMSRLGE
3	GLRLMRLV	GTDMComplexity	Perimeter12
4	Perimeter7	GLSZMSZLGE	GLRLMSRE
5	Euler7	Area16	Perimeter2
6	Euler6	Area13	GLRLMHGRE
7	GLRLMGLN	GLRLMRLV	GLRLMLRE
8	Perimeter6	GLRLMRLN	Area6
9	GLCMVariance	Euler1	Perimeter1
10	GLSZMGLN	GLRLMSRE	Area5
11	GLCMEntropy	Euler16	GLCMHomogeneity
12	Perimeter10	GLSZMLGZE	Area15
13	GLCMEnergy	Euler15	Euler15
14	GLRLMGLNN	GLSZMGLV	Euler12
15	Euler5	GLCMCorrelation	GTDMComplexity
16	Perimeter5	Euler9	GLRLMRLNN
17	GLRLMSRHGE	Perimeter16	GLSZMZSN
18	Euler12	Euler2	GLCMAutoCorrelation
19	Euler2	Area12	GLRLMGLNN
20	GTDMContrast	GLSZMZP	GLRLMRLV
21	GLSZMSZLGE	GTDMStrength	GLRLMLRHGE
22	Euler3	Area10	GTDMContrast
23	Perimeter4	GLRLMSRHGE	Perimeter10
24	Euler10	GLCMVariance	GLSZMSZLGE
25	GLCMCorrelation	Euler5	GLRLMLRLGE
26	GLSZMLGZE	GLCMHomogeneity	GLRLMGLN
27	GLSZMSZHGE	Area11	GLSZMLGZE
28	Euler4	GLRLMSRLGE	GLSZMSZHGE
29	GLRLMHGRE	Perimeter6	Euler8
30	GLCMAutoCorrelation	GLRLMRP	GLCMCorrelation

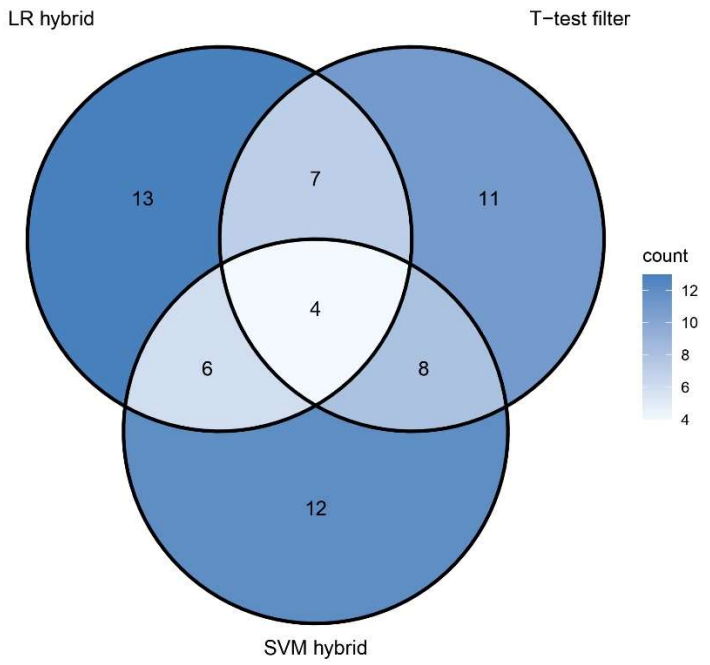


Figure 15 – Venn diagram showing the overlap among the 30 top-selected features using the t-test filter method, the embedded-wrapper LR RFE feature selection method (indicated LR hybrid in the image) and the embedded-wrapper RFE SVM feature selection method (indicated SVM hybrid in the image) according to Table 2. Plot generated using the ggVennDiagram package in RStudio.

6. Discussion and interpretation of the results.

The results presented in this Master's thesis demonstrate the potential of using radiomics and machine learning models to assess GB response to therapy in a preclinical mouse model. We have successfully applied various classifiers and feature selection techniques to MRI-derived features, offering a promising tool for analyzing subtle microenvironmental changes induced by treatment.

The **classification models evaluated included LR, SVM, LDA and KNN**, which exhibit different performance levels, highlighting the complexities of using radiomic features for GB treatment-response classification. While SVM, particularly with polynomial kernels, outperformed other models, it remains important to note that even the best-performing models showed some limitations in their ability to fully distinguish treated and untreated tumors. This may be attributed to the combination of a small sample size and high feature dimensionality, which likely introduces overfitting, limiting the models' overall predictive capacity.

The relatively high accuracy achieved by SVM with polynomial kernels suggests that these models capture important patterns in the data, especially when combined with a slice voting system that aggregates the results of different slices, increasing the confidence of the model prediction for a given mouse. A general trend in the best performing models (i.e., SVM with polynomial kernel of degree 3 or 4, KNN fine) is that these are among the **most flexible and complex models tested** in this thesis. These results suggest that the underlying data distribution is highly nonlinear and benefits from models capable of capturing complex decision boundaries. In fact, SVM has been previously proposed in the literature as the best algorithm to diagnose GB

using histopathological images [40], determine the molecular subtypes of gliomas based on MRI [41] or predict GB treatment survival using magnetic resonance images in a clinical setting [42].

Moreover, we employed different **feature selection methods** such as the t-test filter and embedded-wrapper RFE methods. These approaches aimed to identify the most informative features while reducing the risk of overfitting. Additional expected benefits of feature selection include the reduction of computational resources for model usage and the enhancement of the **interpretability** of the results. In this context, it is worth noting that the classifiers with the best predictive performance (especially SVMs with complex, high-degree kernels) were also the least interpretable. The latter could pose a challenge for applications where interpretability is a desired model feature.

Table 3 – Summary of the top-performing models as a function of the feature selection method.

	t-test filter				**embedded – wrapper**			
	Accuracy	Selected features	SVS accuracy	Selected features	Accuracy	Selected features	SVS accuracy	Selected features
Logistic regression	63%	19	77%	18	75%	9	87%	9
SVM polynomial degree 3	73%	24	87%	21	70%	9	74%	5
SVM polynomial degree 4	73%	20	87%	21	77%	14	80%	14

The effectiveness of these **feature selection techniques varied depending on the classifier used**. Table 3 provides a summary of the model accuracy obtained for the different methods, as a function of the feature selection method, and the use or non-use of the SVS approach. We observe that, in terms of accuracy, while the embedded-wrapper method provides a better model in the case of the LR classifiers, the t-test filter method for feature selection generates better models for the SVM models (i.e., in particular when combined with SVS aggregation per mouse). Another trend that can be observed in Table 3 is that the top-performing models contain fewer selected features in the case of the embedded-wrapper feature method, when compared to the best-performing models generated with the t-test-based filter method. Of particular relevance here is the fact that the t-test filter method offers lower computational cost, as it operates independently of model training and performance evaluation.

In terms of the features selected by each feature selection method, while the t-test filter identified key features such as GLCM Entropy and Perimeter9, the more complex embedded-wrapper RFE method revealed additional features critical for distinguishing treated from untreated tumors. Interestingly, the **overlap between features selected by different methods was limited** (Table 2 and Figure 15). The partial overlap might be linked to the inherent heterogeneity of GB and could suggest that different classification models may capture distinct aspects of tumor behavior.

Among the top selected features in the SVM with a polynomial kernel of degree 4 using the embedded-wrapper RFE method, we found texture **features from the GLRLM (Gray Level Run Length Matrix) matrix**. Developed by Galloway *et al.* in 1975 [43], the GLRLM is a texture descriptor that evaluates sequences of contiguous pixels sharing the same gray level. Among the

top selected GLRLM-derived features we find (Table 2): 1) GLRLM-SRLGLE (Short Run Low Gray-Level Emphasis, which refers to the pattern of short low-intensity areas with homogeneous texture); 2) GLRLM-SRE (Short Run Emphasis, which represents the distribution of short areas of homogeneous texture); 3) GLRLM-HGRE (High Gray-Level Run Emphasis, which refers to the pattern of high grayscale extensions); and 4) GLRLM-LRE (Long Run Emphasis, which describes the pattern of long homogeneous stretches). Moreover, 10 out of the 30 top-selected features in the SVM model (Table 2) belong to the GLRLM matrix. The strong representation of GLRLM-derived features among the top predictors correlated with evidence found in the literature. For instance, prior studies have shown that GLRLM metrics correlate with relevant histopathological markers in gliomas. Specifically, the GLRLM-SRE feature has been linked to proliferation indices like vimentin and CD34 in high-grade glioma cases [44]. Additionally, GLRLM-based features have demonstrated potential in assessing treatment response in GB. As reported by recent work [45], GLRLM features were effective in distinguishing between true progression and pseudoprogression following therapy, a major clinical challenge in the management of GB. Moreover, Cheung *et al*. [40] found several GLRLM features to differ significantly between GB and normal tissue, highlighting their diagnostic utility. Overall, these findings highlight the utility of GLRLM features in capturing tumor heterogeneity relevant to treatment response in GB.

Another feature among the top selected features is **GLCMHomogeneity, from the Gray Level Co-occurrence Matrix** (GLCM). The GLCM matrix [46] quantifies how often pairs of pixel intensities (i and j) occur at specific spatial relationships, typically in vertical, horizontal, or diagonal directions [47]. GB lesions often exhibit irregular shapes and variable nucleoli, leading to greater heterogeneity in pixel intensity. As a result, GLCM features tend to show higher variability and reduced homogeneity in tumor tissue compared to normal tissue, reflecting underlying structural differences. Cheung *et al*. [40] found that GLCMHomogeneity was significantly lower in GB than in normal tissue based on H&E-stained histopathological slides, reinforcing the feature's potential to quantify the loss of textural uniformity characteristic of malignant glioma regions.

Moreover, **Perimeter features from the Minkowski functions** are selected among the top selected features in the SVM based on the embedded-wrapper feature selection method and in the t-test filter method. Minkowski functionals have been identified as valuable tools for quantifying the structural heterogeneity of peritumoral hyperintensity regions in GB. By capturing geometric features such as the boundary complexity in the case of the Perimeter values, these functions have demonstrated prognostic significance, particularly in predicting patient survival outcomes [48].

Finally, one of the notable strategies employed in this research was the **SVS approach**, which aggregates results from multiple MRI slices to produce a more robust classification. We have observed that the SVS approach improved the accuracy of classification for most of the models, although the magnitude of this improvement varied for the different models. In our approach, not only is aggregation a key factor improving accuracy, but also the weighting system used in SVS (i.e., area of the associated image mask). As shown in Figure 16, the use of weights that depend on the area of the tumor outperforms an aggregation strategy in which each tumor slice has the same weight in the aggregation process. All this suggests that a 3D analysis (i.e., rather

than analyzing a few slides) would likely provide a more comprehensive view of the tumor's microenvironment and could enhance the model's predictive power.

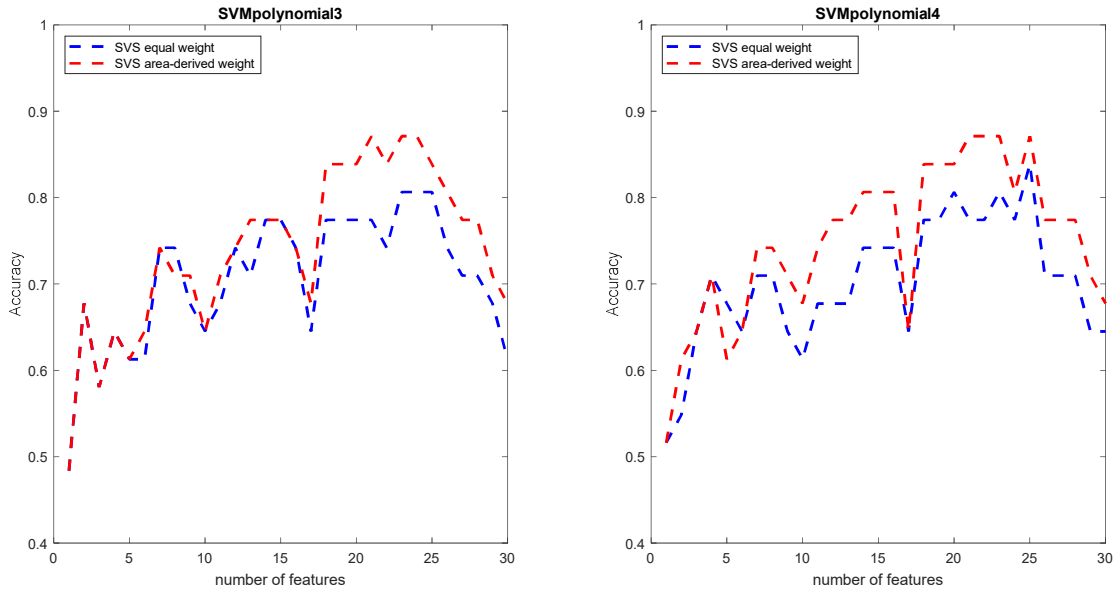


Figure 16 – Comparison of the model accuracy if the SVS approach (Figure 10) uses weights that depend on the area of the masks (i.e., area of the tumors) or if each slice has a similar weight. Model accuracy is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method.

7. Possible drawbacks of the methods used

The methods used in this Master's thesis yielded promising results; however, several limitations and potential drawbacks may be acknowledged.

The study was conducted using a relatively limited number of mice (n = 63), with only 32 samples used for training. This is a small sample size that restricts the models' statistical power and, importantly, can negatively impact the generalizability of the models. The high dimensionality of the feature space (90 features) relative to the number of samples increases the risk of overfitting, despite the use of cross-validation and feature selection techniques. This might be particularly relevant for SVM models with polynomial kernels of degrees 3 and 4, which showed strong performance but might lack robustness on unseen data. These models may still be learning dataset-specific noise and might not generalize well.

Our strategy uses manual ROI segmentation as the image segmentation approach. Here, the segmentation of tumor regions was done manually by expert researchers. While this ensures high-quality segmentation, it also introduces subjectivity and potential inter-individual variability, which might limit the reproducibility of the results described. Automated or semi-automated segmentation methods could provide alternatives to address this potential method limitation. Alternatively, the image segmentation step could be avoided altogether, enabling the model to extract information from the whole MRI images.

Although the SVS approach used to aggregate predictions from multiple slices per mouse improved performance, the weighting scheme may oversimplify spatial heterogeneity within tumors. It is possible that 3D volumetric analysis could provide a more representative picture of

the tumor microenvironment and represent more accurately the intra-tumoral heterogeneity across slices.

For model training and validation, we adopted the same training-test split of the dataset that was proposed by Núñez *et al.* [31], primarily to enable direct comparison with their results. However, this predefined split introduces certain biases that may affect the generalizability of the models. Specifically, the training set consists predominantly of single-slice images per mouse, whereas the test set includes mice with three to four slices each (Supplementary Table 1). Additionally, there is a treatment protocol imbalance: the training set is largely composed of mice treated with the standard 3-cycle TMZ regimen, while the test set primarily includes mice subjected to the IMS-TMZ protocol (Table 1 and Figure 5). Finally, most of the mice in the test set were kept in an EE-like setting before and during TMZ treatment (Figure 5) to boost treatment performance (this was not the case for mice in the training set). These differences could affect model performance in ways unrelated to the features of interest, limiting the reliability of the conclusions. A more balanced approach, which, for example, randomly assigns mice to training and test sets, could potentially mitigate these biases and lead to more robust performance on unseen data.

The analysis presented focuses on images corresponding to the last day these mice were studied (usually prior to euthanization) [31]. This endpoint is informative, as it reflects the full impact of the treatment on the tumor microenvironment. However, it is important to note that the mice were imaged longitudinally, approximately every 3-4 days throughout the study (Figure 5). Despite this, the temporal evolution of the tumor, as captured by these serial MRI scans, was not incorporated into our current modeling approach. Incorporating time-dependent changes in the radiomic features could potentially enhance the predictive performance and clinical relevance of our models, particularly for early response assessment.

The rationale for using a hybrid embedded-wrapper RFE method in this thesis and in Núñez *et al.* [31] originates from the limitations of a pure wrapper-based RFE method, which often fails to distinguish feature relevance in the early steps, as classification accuracies remain largely unchanged. However, we also observed that the embedded RFE method faces similar challenges, encountering many instances in which it cannot distinguish feature relevance. Consequently, the hybrid approach cannot fully resolve the issue it was designed to overcome. Thus, alternative feature selection strategies, such as incorporating a preprocessing step to remove low-variance features [42], could be evaluated in an effort to eliminate variables with minimal discriminatory power that would otherwise hinder both embedded and wrapper RFE methods.

8. Ethical thinking, societal relevance, and stakeholder awareness

From an **ethical point of view**, the research dataset used in this thesis adheres to a responsible and ethical approach to biomedical research, aimed at balancing scientific progress with a strong commitment to animal welfare, societal benefit, and clinical relevance in the context of GB. The dataset used was sourced from previously completed studies, which were performed in line with regional and national ethical guidelines and were approved by UAB's Research Ethics Committee (https://www.uab.cat/en/research-ethics/animal-experimentation, accessed on 27 July 2025), under protocols CEA-OH-3665 and CEEAH-9685.

The research performed at the GABRMN laboratory fully embraces the 3Rs principle—Replacement, Reduction, and Refinement—as a cornerstone of its *in vivo* research practices. Replacement is promoted through the use of *in vitro* systems and existing data from preclinical models, reducing the need for new *in vivo* experiments. Reduction is achieved by minimizing the number of animals to those strictly needed to provide statistically/biologically meaningful conclusions and by applying longitudinal MRI imaging protocols that enable gathering additional information from each mouse. Refinement is ensured through optimal housing conditions, environmental enrichment, and continuous veterinary supervision, all aimed at minimizing stress and discomfort. All imaging procedures were non-invasive and performed under anesthesia, further safeguarding animal welfare.

Additionally, the research is designed to limit animal use to only those stages where therapeutic approaches demonstrated sufficient efficacy in representative *in vitro* models. Only when these preclinical results justified further validation were small cohorts of animals used, with carefully optimized experimental designs that maximize scientific value while minimizing ethical cost.

Crucially, the role of veterinary supervision in this process is paramount. Veterinarians involved in these studies operate under strict animal welfare criteria and are entrusted with the authority to halt any experiment should animal well-being be compromised. Their judgment is sovereign, and animal welfare takes precedence over any scientific goal, including the desire to complete a dataset.

Though the study presented in this thesis did not involve human subjects, future translational steps of these models should adhere to principles of transparency, informed consent, and data protection, particularly if these models are to be integrated into clinical decision-making. The machine learning models developed here are ultimately intended to support, but not to replace, medical expertise. As such, reproducibility, interpretability, and fairness should be some of the principles that guide future deployment.

In terms of **societal relevance**, GB accounts for approximately half of all primary CNS malignancies and remains the most common primary malignant brain tumor in adults, with an incidence of about 3 cases per 100,000 individuals per year [3, 4]. Despite advances in surgical techniques, radiotherapy, and chemotherapy, the prognosis remains extremely poor, with long-term survival being rare. In addition to its high mortality, GB causes significant neurological, cognitive, and emotional impairments, severely diminishing patient independence and imposing a profound burden on families and caregivers. GB also has important economic implications, linked to substantial direct and indirect costs. The median cost of therapy per patient is estimated at approximately $100,000, most of which is concentrated in the first four months after diagnosis (i.e., explained by its aggressive nature and limited survival of the disease). Indirect costs, such as productivity loss and caregiver burden, have been estimated to exceed $1 billion per year in the EU, highlighting the societal urgency of improving GB care.

In this context, the development of predictive models—like those explored in this thesis—has significant potential to enhance clinical decision-making and support precision medicine. By enabling earlier identification of treatment response or failure, these models could help optimize therapeutic strategies, reduce unnecessary toxicity, and improve quality of life. While this work

is based on preclinical (murine) data, it contributes to future translational research aimed at supporting clinicians with quantitative, data-driven tools.

In terms of **stakeholders' awareness**, it is worth pointing out that the long-term impact of this research depends not only on scientific robustness but also on proactive engagement with key stakeholders across the biomedical innovation ecosystem, as a critical step required for future clinical implementation. Thus, establishing and maintaining strong connections with hospitals is central in this respect, enabling early feedback from clinicians and facilitating the integration of new technologies into real-world clinical workflows. In particular, the GABRMN laboratory collaborates closely with Dr. Andreu Gabarrós and Alejandro Fernandez-Coello at Bellvitge Hospital (Bellvitge, Spain), whose expertise in brain tumor treatment and neurological complications provides critical clinical insight. These collaborations ensure alignment with medical needs and provide a realistic view of possible adoption in the complex clinical GB setting. Similarly, interactions with patient associations can ensure critical alignment with patient needs.

Beyond academic, clinical and patient stakeholders, the successful development and deployment of machine learning-based medical technologies would require the involvement of a broader set of stakeholders in the innovation ecosystem. This includes entrepreneurs, early-stage investors, and industry partners capable of supporting regulatory navigation and commercialization of such technologies. Overall, stakeholder engagement is not only critical for accelerating the translation of research into practice but also for ensuring responsible innovation. By involving diverse actors early in the development process, we can better anticipate potential barriers to implementation, such as ethical concerns or data privacy. Stakeholders' involvement enables the development of solutions that are technically robust, realistic, socially acceptable, and ethically sound.

9. Conclusion

This thesis investigated the use of radiomics-based machine learning models to differentiate therapeutic responses in the murine GL261 glioblastoma model. By extracting a comprehensive set of radiomic features from T2-weighted MRI images (i.e., 90 features including texture and morphological features) and applying classification algorithms, we assessed the ability of these models to distinguish between TMZ treatment responses.

We were able to demonstrate the potential of MRI-based radiomics approaches to detect therapy-induced changes in glioblastoma tumors, by capturing subtle microenvironmental changes induced by TMZ. By systematically evaluating a range of diverse classifiers and feature selection strategies, we identified support vector machines with higher-degree polynomial kernels and fine-grained KNN as the top-performing models. Overall, the findings might indicate that the most effective approaches are those relying on flexible and complex models, supporting the notion that the underlying data distribution is highly nonlinear and best addressed by classifiers capable of capturing intricate decision boundaries.

Several texture-based features derived from the GLRLM family were selected, suggesting their potential relevance in characterizing tumor heterogeneity and therapy response. Similarly, GLCM Homogeneity and Minkowski functions' Perimeter features appear to be important in distinguishing treatment groups. However, while certain biological interpretations can be

inferred from existing literature, further validation is necessary to establish their biological relevance and to link them to specific molecular mechanisms.

In terms of future research, this thesis also highlights some possible challenges, particularly regarding model generalizability. Potential imbalances or biases in dataset composition could have limited the robustness of the classifiers. Moreover, the use of a mask generated by expert researchers for image segmentation might introduce subjectivity and potential inter-individual variability, potentially affecting model generalizability. Addressing these limitations through future software developments may help to improve robustness and broader applicability of these models.

Overall, this thesis contributes to highlighting the potential in using radiomics and machine learning methods with the aim of improving treatment response assessment in GB.

10. Ideas for future research

One of the proposed next steps in this project is to apply the developed models to new preclinical datasets, in order to assess how accurately these models can classify other (related) GB treatments. As part of this Master's thesis, I started creating masks for MRI images from different preclinical studies using the GL261 mouse model. These included a dataset in which variations of the IMS-TMZ treatment were evaluated: different TMZ doses (i.e., 140 mg/kg, 200 mg/kg and 240 mg/kg) and different starting days for the TMZ treatment (i.e., 8 and 11 days after tumor implantation) were evaluated. Another available dataset corresponds to the evaluation of a combination of TMZ and the casein kinase 2 inhibitor CX-4945 [49] as a possible combined treatment for GB [50]. Another set of images corresponds to the evaluation of cyclophosphamide (a chemotherapeutic) as GB treatment, evaluated with the metronomic regimen [51]. Moreover, a dataset is available in which immunotherapy using the anti-PD-1 antibody (checkpoint inhibitor) is evaluated as GB treatment alone or in combination with TMZ treatment [52, 53]. Another available dataset evaluated ALDH inhibitors (i.e., a lipidic nanoemulsion of the compound DIMATE) as a possible GB treatment [54]. Finally, there are two datasets that assess biscatechol-based and Fe-based Pt(IV) compounds for GB treatment.

Overall, the available datasets comprise MRI scans from 94 mice, each with three time points on average and 10 slices per scan, resulting in a total of 2,820 slices for analysis. Given the scale and complexity of this data, the development of an automatic or semi-automatic tool is essential to facilitate the generation of segmentation masks and ensure consistency across images and users. Furthermore, the availability of multiple slices spanning the entire brain offers a valuable opportunity to enhance our MATLAB-based pipeline by incorporating multi-slice analysis reflecting the spatial context, enabling a more comprehensive characterization of the tumor microenvironment, even without a full 3D volumetric reconstruction.

A possible interesting direction for future research would be to evaluate classifier ensembles as a way to enhance model robustness and accuracy [55]. Inspired by the brain's modular information processing, ensemble methods apply the divide-and-conquer principle by combining simpler classifiers to address complex tasks [56]. Techniques such as bagging and boosting have shown strong potential in pattern recognition. Bagging trains multiple models on different bootstrapped subsets and combines them via majority voting, while boosting

sequentially adjusts training to focus on previously misclassified instances. Thus, we could evaluate SVM ensembles using bagging or boosting in our future Radiomics studies.

Future research could also explore the integration of multimodal data in preclinical models to enhance the performance of the machine learning models. Recent studies in clinical oncology have demonstrated the benefits of combining imaging, clinical, and molecular data, among others [27]. Incorporating diverse data modalities in addition to MRI, such as histopathology, genomic/transcriptomic profiles (i.e., radiogenomics [57]), and behavioral readouts from animal models, may offer a more integrated understanding of tumor biology. Such multimodal approaches could increase predictive accuracy by overcoming the limitations of MRI and may enable the capture of complementary biological information [41, 42].

Finally, some of the selected features (e.g., GLCM Entropy, Perimeter9, GLRLM-SRE, GLCM Homogeneity) have demonstrated algorithmic importance in distinguishing treatment groups. However, their direct biological relevance in the tumor microenvironment changes remains unclear. Further research could focus on establishing the connections between these radiomic features and biological processes in GB. One possible strategy involves correlating selected radiomic features with histopathological markers, such as cell density, necrosis, vascular proliferation, or immune cell infiltration, obtained from matched tissue sections. This could help determine whether features like texture heterogeneity (e.g., GLCM Homogeneity) reflect cellular disorganization, hypoxia, or immune activity. Similarly, the correlation of selected features with immunohistochemical markers or gene expression could be studied, especially when studied on the same tumors used for MRI acquisition. Overall, such a line of research would highlight the potential of radiomics in shedding light on the molecular mechanisms driving GB disease.

11. References

1. Ferlay, J., et al. *Global Cancer Observatory: Cancer Today. Lyon, France: International Agency for Research on Cancer.* 2024 [cited 2025 12 July]; Available from: https://gco.iarc.who.int/today.
2. Weller, M., et al., *Glioma.* Nat Rev Dis Primers, 2015. **1**: p. 15017.
3. Osswald, M., et al., *Brain tumour cells interconnect to a functional and resistant network.* Nature, 2015. **528**(7580): p. 93-8.
4. Stoyanov, G.S., et al., *Cell biology of glioblastoma multiforme: from basic science to diagnosis and treatment.* Med Oncol, 2018. **35**(3): p. 27.
5. *Central Nervous System Tumours*. WHO Classification of Tumours. Vol. 6. 2021: IARC Publications.
6. Weber, R.G., et al., *Characterization of genomic alterations associated with glioma progression by comparative genomic hybridization.* Oncogene, 1996. **13**(5): p. 983-94.
7. Byun, Y.H. and C.K. Park, *Classification and Diagnosis of Adult Glioma: A Scoping Review.* Brain Neurorehabil, 2022. **15**(3): p. e23.
8. Tan, A.C., et al., *Management of glioblastoma: State of the art and future directions.* CA Cancer J Clin, 2020. **70**(4): p. 299-312.
9. Sipos, D., et al., *Glioblastoma: Clinical Presentation, Multidisciplinary Management, and Long-Term Outcomes.* Cancers (Basel), 2025. **17**(1): p. 146.
10. Greenhalgh, A.D., S. David, and F.C. Bennett, *Immune cell regulation of glia during CNS injury and disease.* Nat Rev Neurosci, 2020. **21**(3): p. 139-52.

11. Stupp, R., et al., *Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial.* Lancet Oncol, 2009. **10**(5): p. 459-66.
12. Parker, N.R., et al., *Molecular heterogeneity in glioblastoma: potential clinical implications.* Front Oncol, 2015. **5**: p. 55.
13. Karpel-Massler, G., et al., *Combined inhibition of Bcl-2/Bcl-xL and Usp9X/Bag3 overcomes apoptotic resistance in glioblastoma in vitro and in vivo.* Oncotarget, 2015. **6**(16): p. 14507-21.
14. Bellail, A.C., et al., *DR5-mediated DISC controls caspase-8 cleavage and initiation of apoptosis in human glioblastomas.* J Cell Mol Med, 2010. **14**(6A): p. 1303-17.
15. Sundar, S.J., et al., *The role of cancer stem cells in glioblastoma.* Neurosurg Focus, 2014. **37**(6): p. E6.
16. Kim, Y., et al., *Perspective of mesenchymal transformation in glioblastoma.* Acta Neuropathol Commun, 2021. **9**(1): p. 50.
17. Segerman, A., et al., *Clonal Variation in Drug and Radiation Response among Glioma-Initiating Cells Is Linked to Proneural-Mesenchymal Transition.* Cell Rep, 2016. **17**(11): p. 2994-3009.
18. Lim, M., et al., *Current state of immunotherapy for glioblastoma.* Nat Rev Clin Oncol, 2018. **15**(7): p. 422-42.
19. Czarnywojtek, A., et al., *Glioblastoma Multiforme: The Latest Diagnostics and Treatment Techniques.* Pharmacology, 2023. **108**(5): p. 423-31.
20. Caravan, P., et al., *Gadolinium(III) Chelates as MRI Contrast Agents: Structure, Dynamics, and Applications.* Chem Rev, 1999. **99**(9): p. 2293-352.
21. Leung, D., et al., *Role of MRI in primary brain tumor evaluation.* J Natl Compr Canc Netw, 2014. **12**(11): p. 1561-8.
22. Vogelbaum, M.A., et al., *Application of novel response/progression measures for surgically delivered therapies for gliomas: Response Assessment in Neuro-Oncology (RANO) Working Group.* Neurosurgery, 2012. **70**(1): p. 234-43; discussion 243-4.
23. Eisenhauer, E.A., et al., *New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1).* Eur J Cancer, 2009. **45**(2): p. 228-47.
24. Macdonald, D.R., et al., *Response criteria for phase II studies of supratentorial malignant glioma.* J Clin Oncol, 1990. **8**(7): p. 1277-80.
25. van Griethuysen, J.J.M., et al., *Computational Radiomics System to Decode the Radiographic Phenotype.* Cancer Res, 2017. **77**(21): p. e104-e107.
26. van Timmeren, J.E., et al., *Radiomics in medical imaging—"how-to" guide and critical reflection.* Insights into Imaging, 2020. **11**(1): p. 91.
27. Perez-Lopez, R., et al., *A guide to artificial intelligence for cancer researchers.* Nat Rev Cancer, 2024. **24**(6): p. 427-41.
28. Arias-Ramos, N., et al., *Metabolomics of Therapy Response in Preclinical Glioblastoma: A Multi-Slice MRSI-Based Volumetric Analysis for Noninvasive Assessment of Temozolomide Treatment.* Metabolites, 2017. **7**(2): p. 20.
29. Delgado-Goni, T., et al., *MRSI-based molecular imaging of therapy response to temozolomide in preclinical glioblastoma using source analysis.* NMR Biomed, 2016. **29**(6): p. 732-43.
30. Wu, S., et al., *Anti-tumour immune response in GL261 glioblastoma generated by Temozolomide Immune-Enhancing Metronomic Schedule monitored with MRSI-based nosological images.* NMR Biomed, 2020. **33**(4): p. e4229.
31. Nunez, L.M., et al., *Unraveling response to temozolomide in preclinical GL261 glioblastoma with MRI/MRSI using radiomics and signal source extraction.* Sci Rep, 2020. **10**(1): p. 19699.

32. Calero-Perez, P., *On the road to improve glioblastoma therapy follow-up. Immune microenvironment: what is behing the MRI-based nosological images?*, in *Department of Biochemistry and Molecular Biology*. 2022, Universitat Autonoma de Barcelona.
33. Calero-Perez, P., et al., *Immune System-Related Changes in Preclinical GL261 Glioblastoma under TMZ Treatment: Explaining MRSI-Based Nosological Imaging Findings with RT-PCR Analyses.* Cancers (Basel), 2021. **13**(11): p. 2663.
34. Vallieres, M., et al., *A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities.* Phys Med Biol, 2015. **60**(14): p. 5471-96.
35. Towner, R.A., et al., *A new anti-glioma therapy, AG119: pre-clinical assessment in a mouse GL261 glioma model.* BMC Cancer, 2015. **15**: p. 522.
36. Kober, C., et al., *Microglia and astrocytes attenuate the replication of the oncolytic vaccinia virus LIVP 1.1.1 in murine GL261 gliomas by acting as vaccinia virus traps.* J Transl Med, 2015. **13**: p. 216.
37. Garofalo, S., et al., *Enriched environment reduces glioma growth through immune and non-immune mechanisms in mice.* Nat Commun, 2015. **6**: p. 6623.
38. Legland, D. *Minkowski toolbox.* 2021 [Last accessed 21th April 2025]; Available from: https://es.mathworks.com/matlabcentral/fileexchange/33690-geometric-measures-in-2d-3d-images.
39. Vallière, M. *Radiomics toolbox.* 2012 [Last accessed 21th April 2025]; Available from: https://github.com/mvallieres/radiomics.
40. Cheung, E.Y.W., et al., *AI Deployment on GBM Diagnosis: A Novel Approach to Analyze Histopathological Images Using Image Feature-Based Analysis.* Cancers, 2023. **15**(20): p. 5063.
41. Lu, C.F., et al., *Machine Learning-Based Radiomics for Molecular Subtyping of Gliomas.* Clin Cancer Res, 2018. **24**(18): p. 4429-4436.
42. Chong, J.K., et al., *Optimizing Glioblastoma, IDH-wildtype Treatment Outcomes : A Radiomics and Support Vector Machine-Based Approach to Overall Survival Estimation.* J Korean Neurosurg Soc, 2025. **68**(1): p. 7-18.
43. Galloway, M.M., *Texture analysis using gray level run lengths.* Computer Graphics and Image Processing, 1975. **4**(2): p. 172-79.
44. Li, J., et al., *High-order radiomics features based on T2 FLAIR MRI predict multiple glioma immunohistochemical features: A more precise and personalized gliomas management.* PLoS One, 2020. **15**(1): p. e0227703.
45. Patel, M., et al. *Radiomic evaluation of treatment response in patients with glioblastoma: a preliminary study*. 2019. European Congress of Radiology-ECR 2019.
46. Haralick, R.M., K. Shanmugam, and I. Dinstein, *Textural Features for Image Classification.* IEEE Transactions on Systems, Man, and Cybernetics, 1973. **SMC-3**(6): p. 610-21.
47. Bhagat, P.K., P. Choudhary, and K.M. Singh, *Chapter 13 - A comparative study for brain tumor detection in MRI images using texture features*, in *Sensors for Health Monitoring*, N. Dey, J. Chaki, and R. Kumar, Editors. 2019, Academic Press. p. 259-87.
48. Choi, Y., et al., *Analysis of peritumoral hyperintensity on pre-operative T2-weighted MR images in glioblastoma: Additive prognostic value of Minkowski functionals.* PLoS One, 2019. **14**(5): p. e0217785.
49. Bova, V., et al., *Casein Kinase 2 Inhibitor, CX-4945, Induces Apoptosis and Restores Blood-Brain Barrier Homeostasis in In Vitro and In Vivo Models of Glioblastoma.* Cancers (Basel), 2024. **16**(23): p. 3936.
50. Ferrer-Font, L., et al., *Targeting Protein Kinase CK2: Evaluating CX-4945 Potential for GL261 Glioblastoma Therapy in Immunocompetent Mice.* Pharmaceuticals (Basel), 2017. **10**(1): p. 24.
51. Ferrer-Font, L., *Tuning response to therapy in preclinical GL261 glioblastoma through CK2 targeting and temozolomide metronomic approaches: non-invasive assessment with*

MRI and MRSI-based molecular imaging strategies., in *Department of Biochemistry and Molecular Biology*. 2017, Universitat Autonoma de Barcelona.
52. Wu, S., *When oncology meets immunology: improving GL261 glioblastoma treatment through cancer-related immunity and MRSI-based non-invasive follow-up of response*, in *Department of Biochemistry and Molecular Biology*. 2020, Universitat Autonoma de Barcelona.
53. Wu, S., et al., *Anti-PD-1 Immunotherapy in Preclinical GL261 Glioblastoma: Influence of Therapeutic Parameters and Non-Invasive Response Biomarker Assessment with MRSI-Based Approaches.* International Journal of Molecular Sciences, 2020. **21**(22): p. 8775.
54. Jimenez, R., *Targeting aldehyde dehydrogenases in combined therapy against glioblastoma*, in *Department of Biochemistry and Molecular Biology*. 2023, Universitat Autonoma de Barcelona.
55. Huang, M.W., et al., *SVM and SVM Ensembles in Breast Cancer Prediction.* PLoS One, 2017. **12**(1): p. e0161501.
56. Kim, H.-C., et al., *Constructing support vector machine ensemble.* Pattern recognition, 2003. **36**(12): p. 2757-67.
57. Monti, S., et al., *Radiomics and Radiogenomics in Preclinical Imaging on Murine Models: A Narrative Review.* J Pers Med, 2023. **13**(8).

12. Software code

The MATLAB software used in this thesis can be found at https://gitfront.io/r/stanco78/aJGZ76Z8GDrN/Radiomics/

- **PredictionReport** function used in Núñez *et al.* [31] to calculate the votation accuracy in the SVS approach

```
function [MiceTable,NMetrics] = PredictionReport (NTable, NMetrics, Methods)
% This function extracts the predictions of the results of classifiers over
% a dataset to extract how the performance in each mouse is individually to
% show which ones are the critical subjects.

MiceTable = TableGenerator(['Mouse';'Samples';'Average';Methods]);

mice = unique([NTable.Mouse],'rows');
VotationLabels = [];
for i = 1: length(mice)
    mouse = mice(i);
    mouse_selection = (NTable.Mouse == mouse);
    NTableMouse = NTable(mouse_selection,:);
    N = sum(mouse_selection);
    right = sum((NTableMouse.Group) == table2array(NTableMouse(:,Methods)),1);
    accmethod = right / N;
    avg = mean(accmethod);
    VotationLabel_mice = logical(round(accmethod));
    VotationLabels = [VotationLabels;VotationLabel_mice];
    MiceTable = [MiceTable;
        num2cell([mouse,N,avg,accmethod])];
end

VotationResults = sum(VotationLabels,1) / size(VotationLabels,1);
NMetrics = addvars(NMetrics,VotationResults','NewVariableNames','VotationAcc');
```

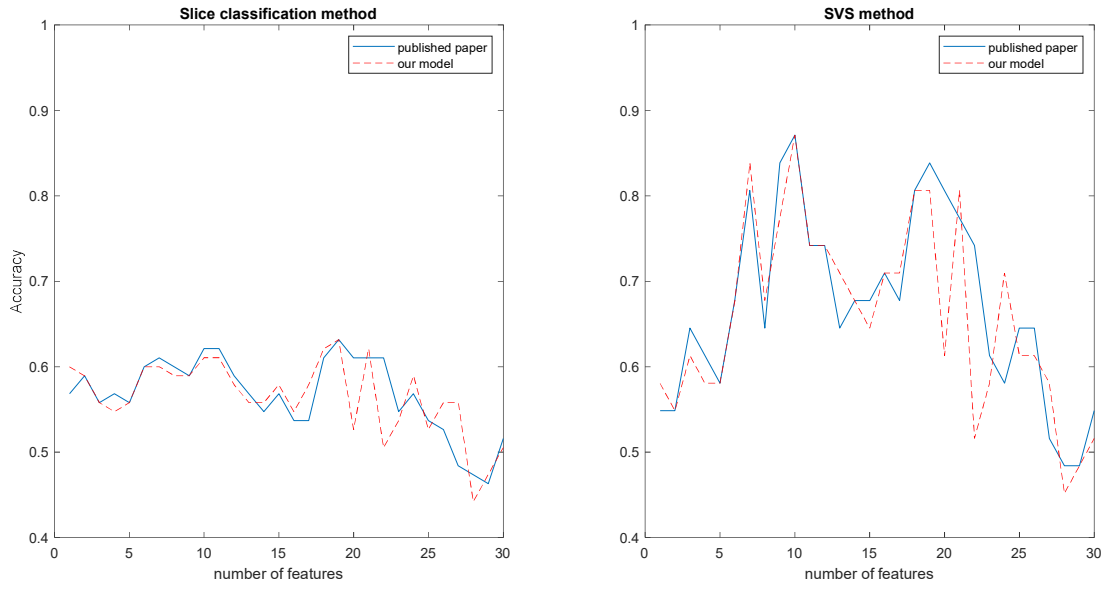
13. Annex

Supplementary Table 1 – Details of the mouse data set, including the day post-treatment at which the image was taken and the number of slices. Adapted from [31].

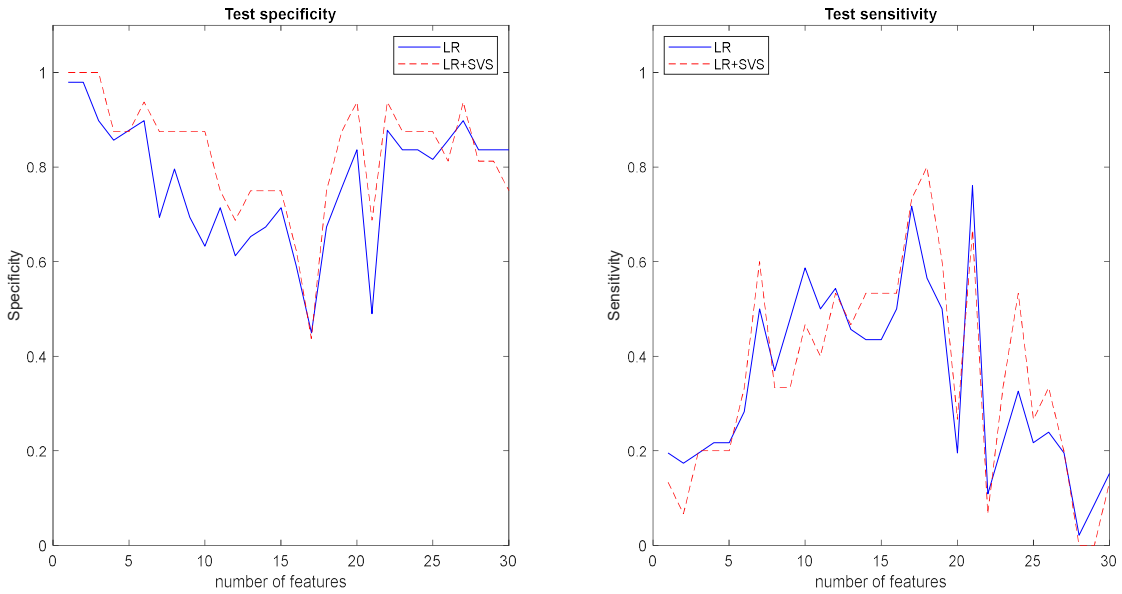
DATASET							
TRAIN				TEST			
MOUSE #	DAY	Number of SLICES	CONTROL/ TREATED	MOUSE #	DAY	Number of SLICES	CONTROL/ TREATED
179	17	1	CONTROL	1109	11	4	CONTROL
233	17	1	CONTROL	1110	13	4	CONTROL
234	17	1	CONTROL	1111	16	4	CONTROL
255	14	1	CONTROL	1112	13	4	CONTROL
278	19	1	CONTROL	1471	16	2	CONTROL
288	18	1	CONTROL	1465	15	2	CONTROL
32	16	1	CONTROL	1472	13	3	CONTROL
351	13	1	CONTROL	1474	14	3	CONTROL
520	18	1	CONTROL	1320	18	3	CONTROL
529	18	1	CONTROL	1344	17	3	CONTROL
583	18	1	CONTROL	1348	21	3	CONTROL
69	15	1	CONTROL	1457	23	3	CONTROL
71	16	1	CONTROL	1459	15	2	CONTROL
526	18	1	TREATED	1461	15	3	CONTROL
572	18	1	TREATED	1462	15	3	CONTROL
574	26	1	TREATED	1466	23	3	CONTROL
776	34	1	TREATED	975	26	4	TREATED
795	18	1	TREATED	1023	23	4	TREATED
797	22	1	TREATED	1026	23	4	TREATED
808	33	1	TREATED	1108	29	3	TREATED
415	22	1	TREATED	1100	26	4	TREATED
418	22	1	TREATED	971	26	4	TREATED
437	23	1	TREATED	1412	23	2	TREATED
525	22	1	TREATED	1445	23	4	TREATED
527	22	1	TREATED	1447	28	3	TREATED
575	26	1	TREATED	1450	24	2	TREATED
584	26	1	TREATED	1451	23	3	TREATED
586	22	1	TREATED	1456	23	2	TREATED
821	34	1	TREATED	1458	23	4	TREATED
414	24	1	TREATED	1460	23	2	TREATED
419	24	1	TREATED	1463	23	1	TREATED
521	18	1	TREATED				

Supplementary Table 2 – Radiomics texture features extracted from the MRI images. Adapted from [31].

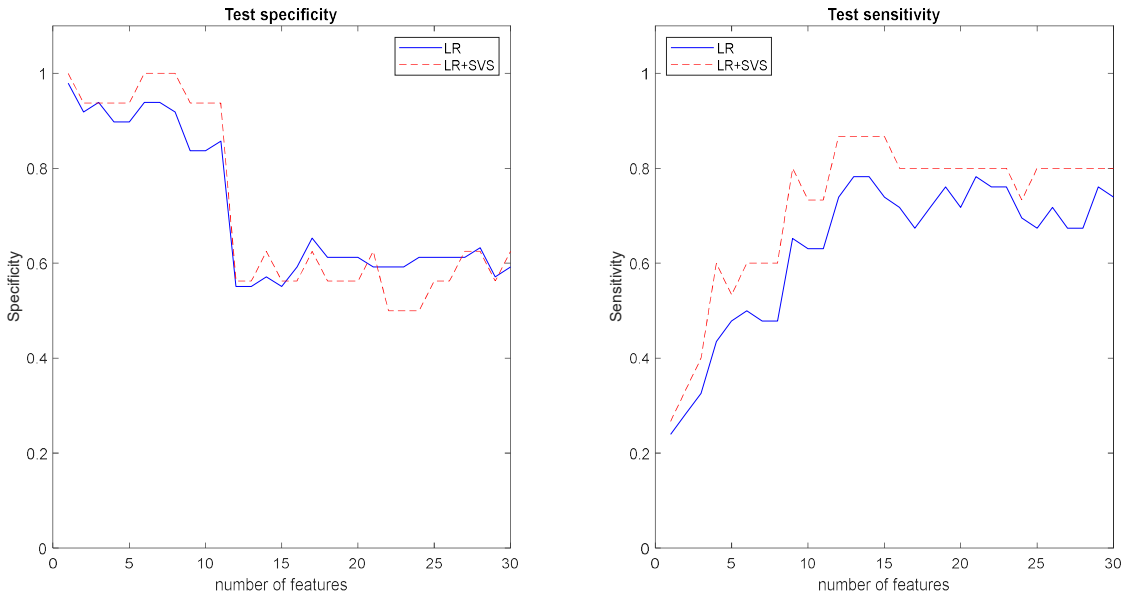
Gray Level Co-occurrence Matrix (GLCM)	
1	Energy
2	Contrast
3	Entropy
4	Homogeneity
5	Correlation
6	SumAverage
7	Variance
8	Dissimilarity
9	Autocorrelation
Gray Level Run Length Matrix (GLRLM)	
1	Short Run Emphasis (SRE)
2	Long Run Emphasis (LRE)
3	Gray Level Non-Uniformity (GLN)
4	Run Length Non-Uniformity (RLN)
5	Run Percentage (RP)
6	Low Gray-Level Run Emphasis (LGRE)
7	High Gray-Level Run Emphasis (HGRE)
8	Short Run Low Gray-Level Emphasis (SRLGE)
9	Short Run High Gray-Level Emphasis (SRHGE)
10	Long Run Low Gray-Level Emphasis (LRLGE)
11	Long Run High Gray-Level Emphasis (LRHGE)
12	Gray-Level Variance (GLV)
Gray-Level Size Zone Matrix (GLSZM):	
1	Small Zone Emphasis (SZE)
2	Large Zone Emphasis (LZE)
3	Gray-Level Nonuniformity (GLN)
4	Zone-Size Nonuniformity (ZSN)
5	Zone Percentage (ZP)
6	Low Gray-Level Zone Emphasis (LGZE)
7	High Gray-Level Zone Emphasis (HGZE)
8	Small Zone Low Gray-Level Emphasis (SZLGE)
9	Small Zone High Gray-Level Emphasis (SZHGE)
10	Large Zone Low Gray-Level Emphasis (LZLGE)
11	Large Zone High Gray-Level Emphasis (LZHGE)
12	Gray-Level Variance (GLV)
Neighborhood Gray-Tone Difference Matrix (NGTDM):	
1	Coarseness
2	Contrast
3	Busyness
4	Complexity
5	Strength



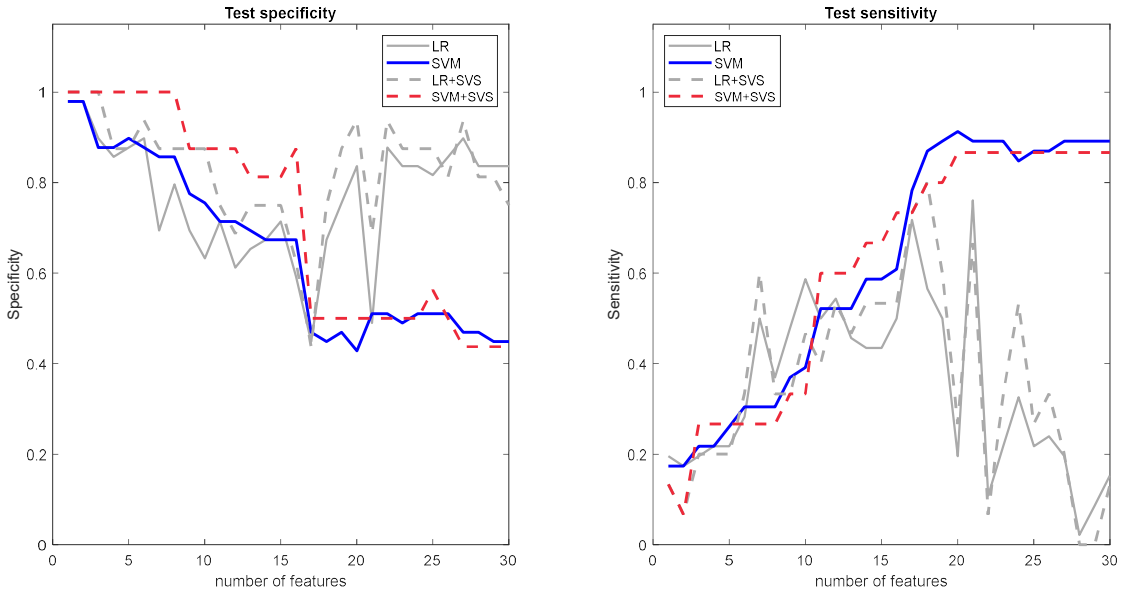
Supplementary Figure 1 – LR radiomics model performance, in terms of model accuracy, when the *PredictionReport* function is used to calculate the accuracy for the SVS approach. Accuracy is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. On the left, the accuracy is evaluated for each slice, and on the right, the SVS approach was applied to aggregate all the slices for a specific mouse.



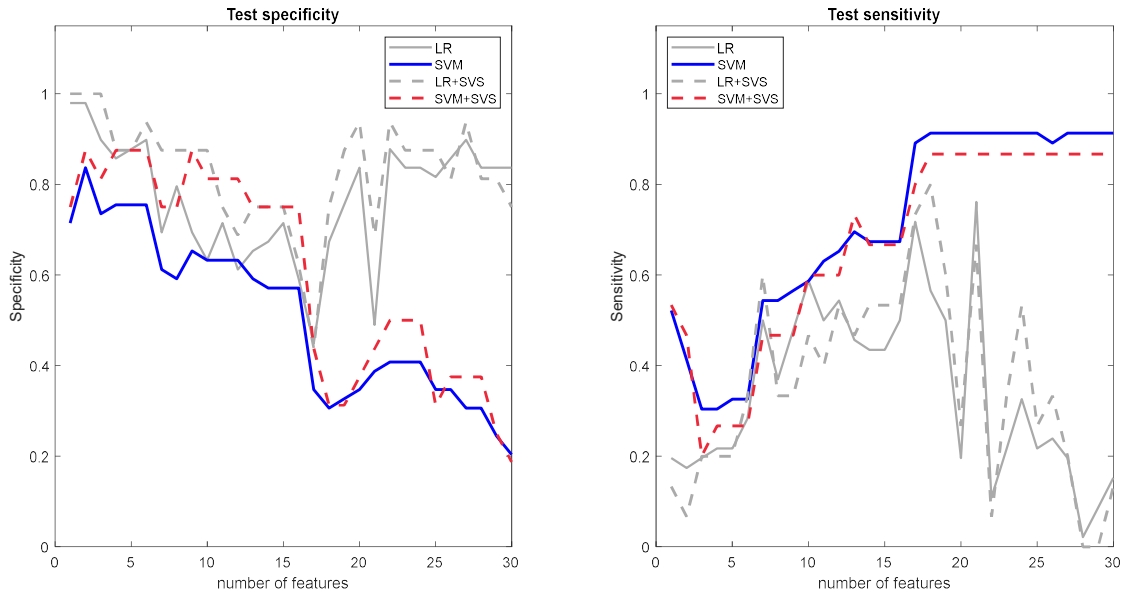
Supplementary Figure 2 – LR model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method.



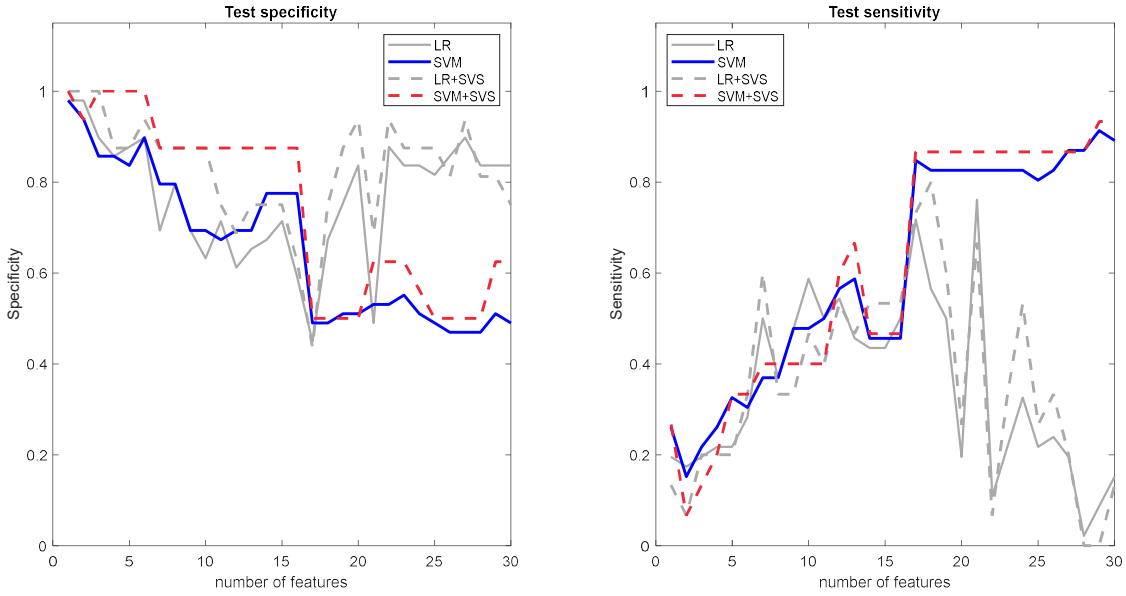
Supplementary Figure 3 – LR model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using an embedded-wrapper method.



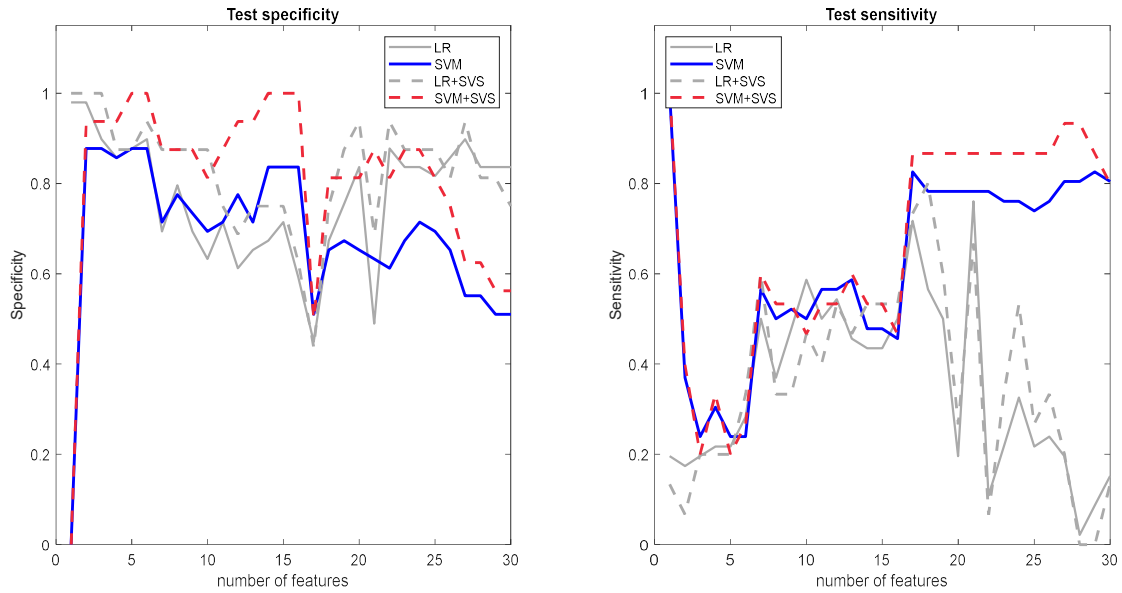
Supplementary Figure 4 – SVM with linear kernel model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. The performance of the LR model is plotted as a reference.



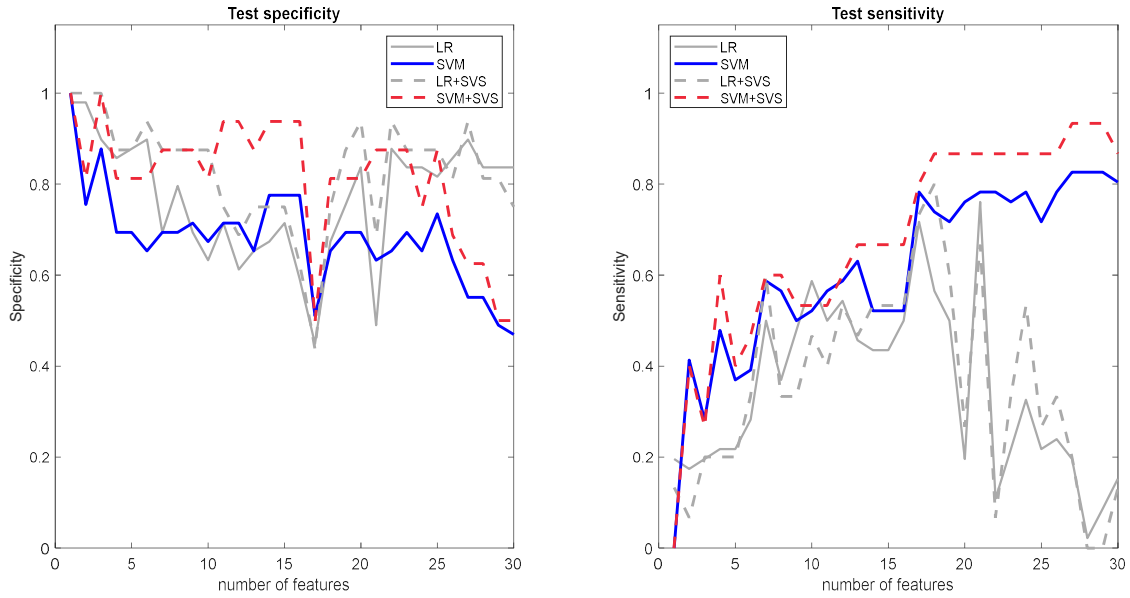
Supplementary Figure 5 – SVM with Gaussian kernel model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. The performance of the LR model is plotted as a reference.



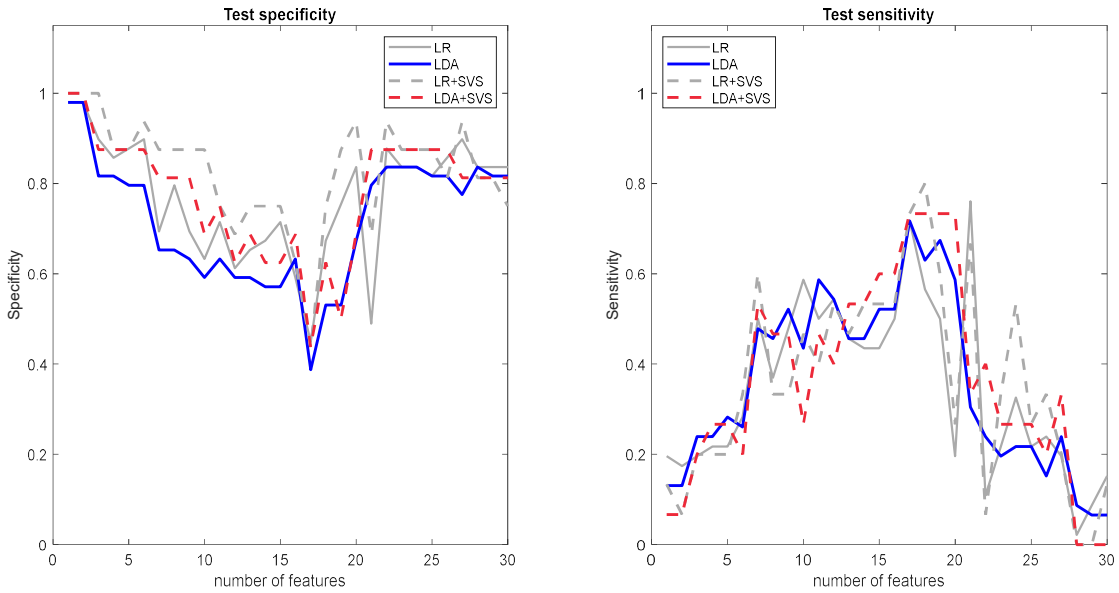
Supplementary Figure 6 – SVM with polynomial kernel of degree 2 model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. The performance of the LR model is plotted as a reference.



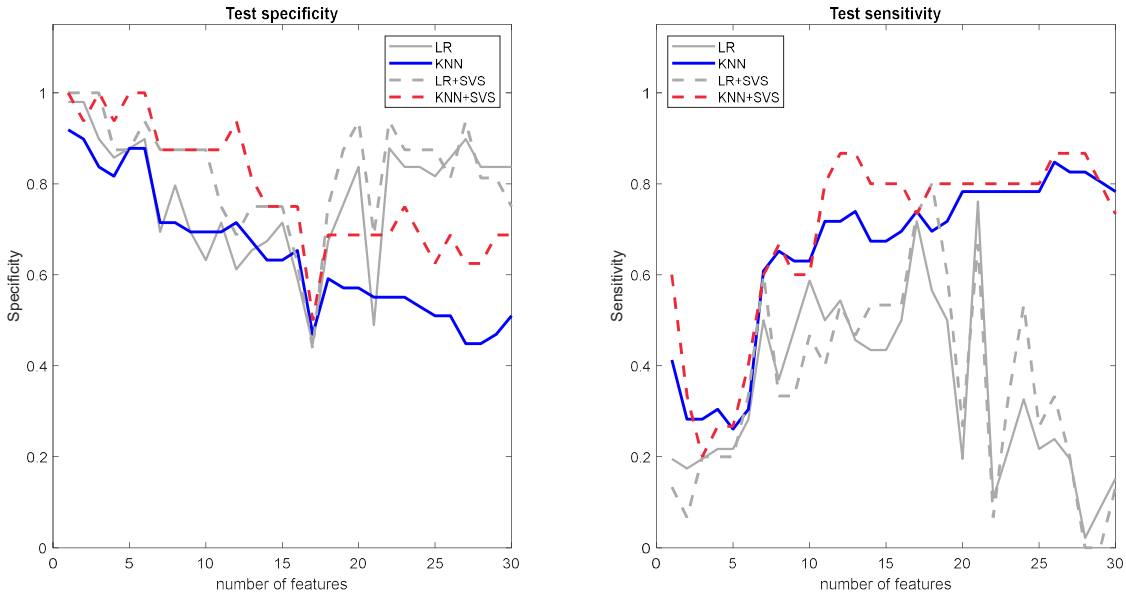
Supplementary Figure 7 – SVM with polynomial kernel of degree 3 model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. The performance of the LR model is plotted as a reference.



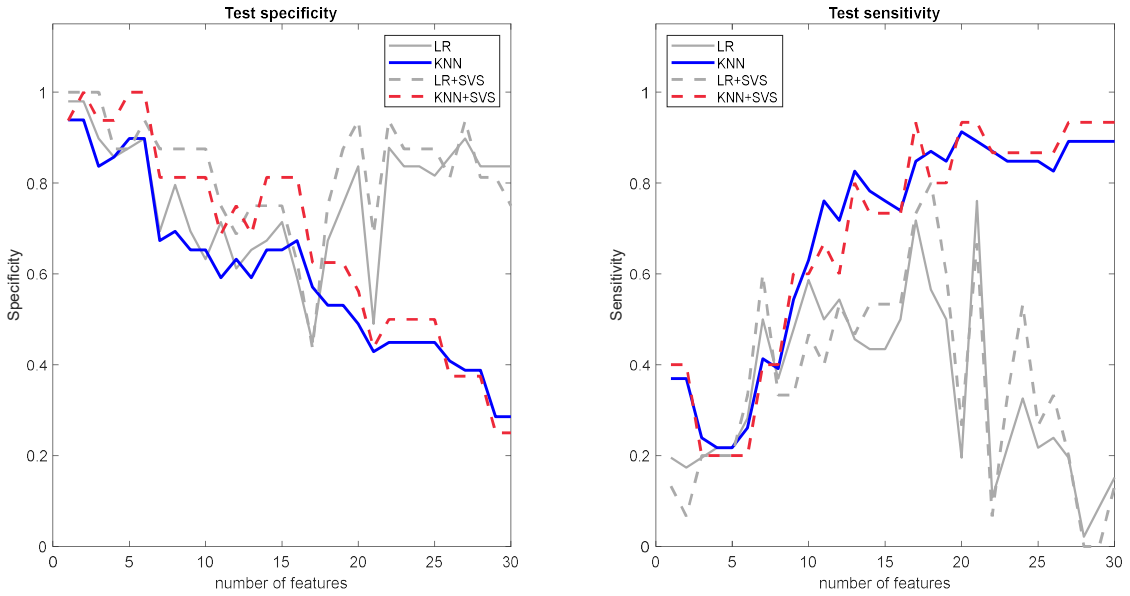
Supplementary Figure 8 – SVM with polynomial kernel of degree 4 model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. The performance of the LR model is plotted as a reference.



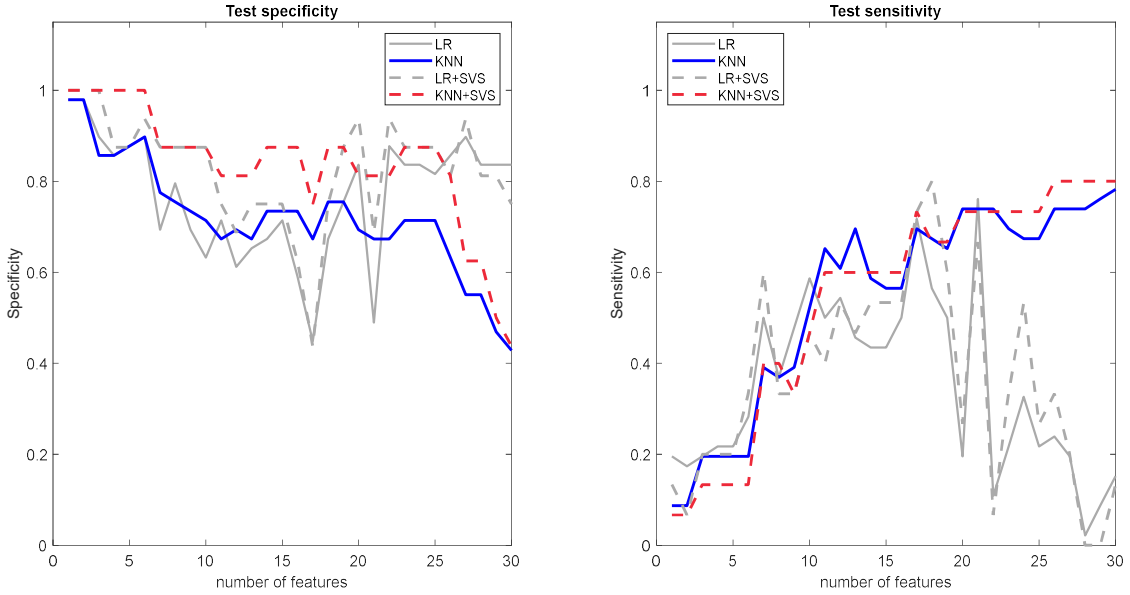
Supplementary Figure 9 – LDA model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. The performance of the LR model is plotted as a reference.



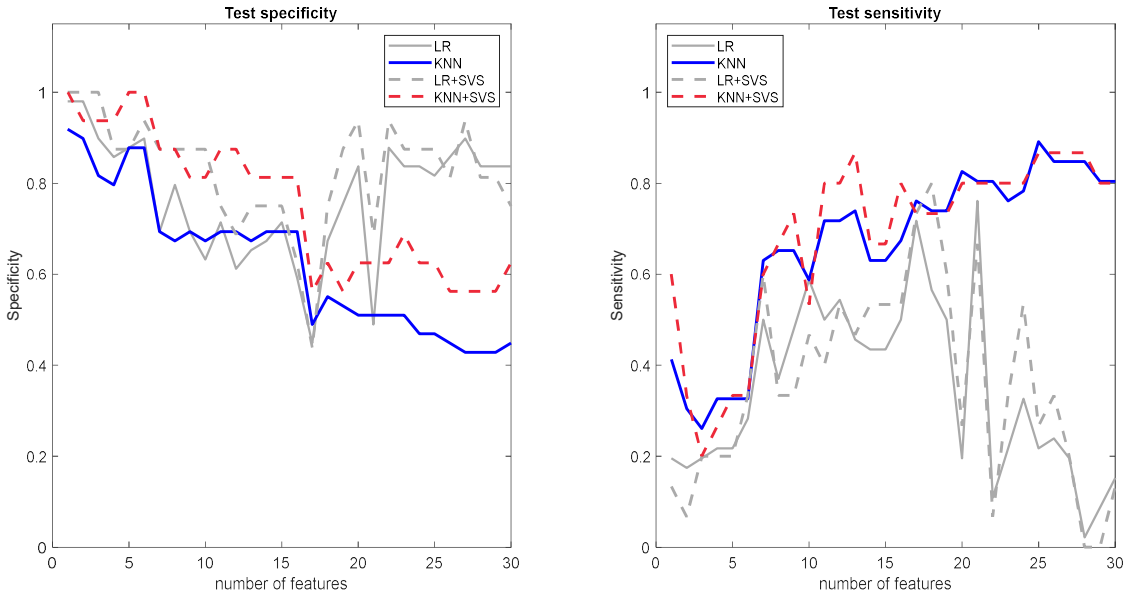
Supplementary Figure 10 – KNN fine model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. The performance of the LR model is plotted as a reference.



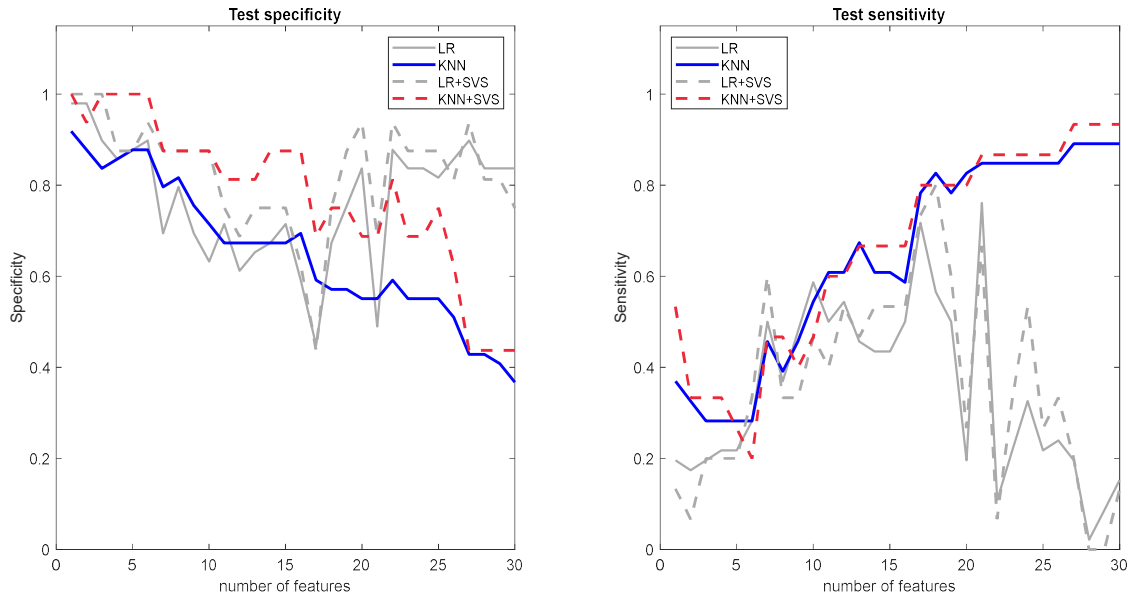
Supplementary Figure 11 – KNN medium model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. The performance of the LR model is plotted as a reference.



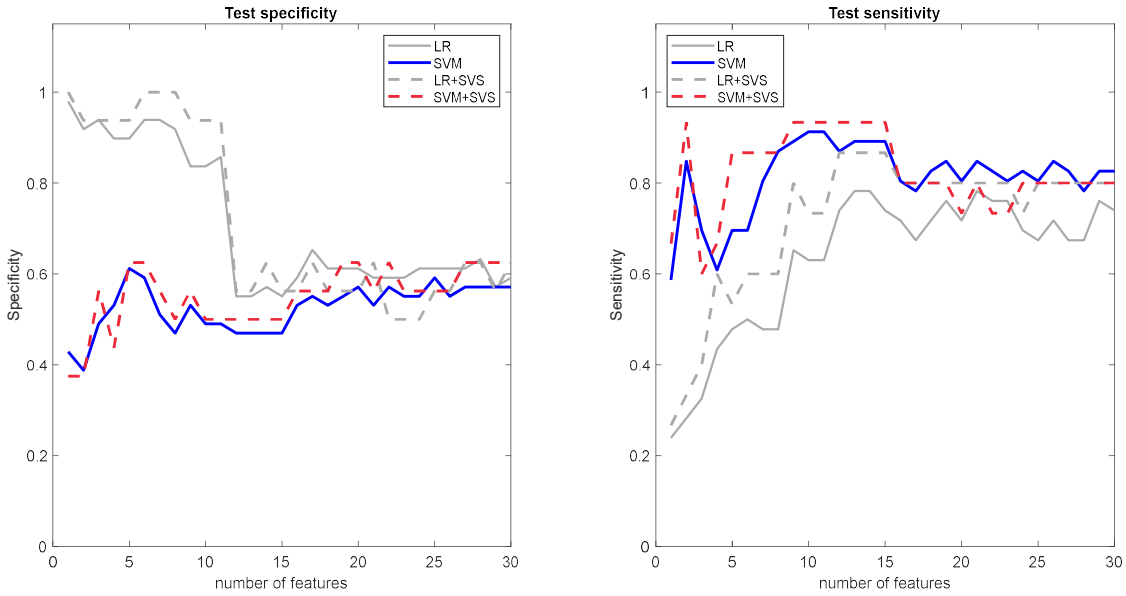
Supplementary Figure 12 – KNN cosine model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. The performance of the LR model is plotted as a reference.



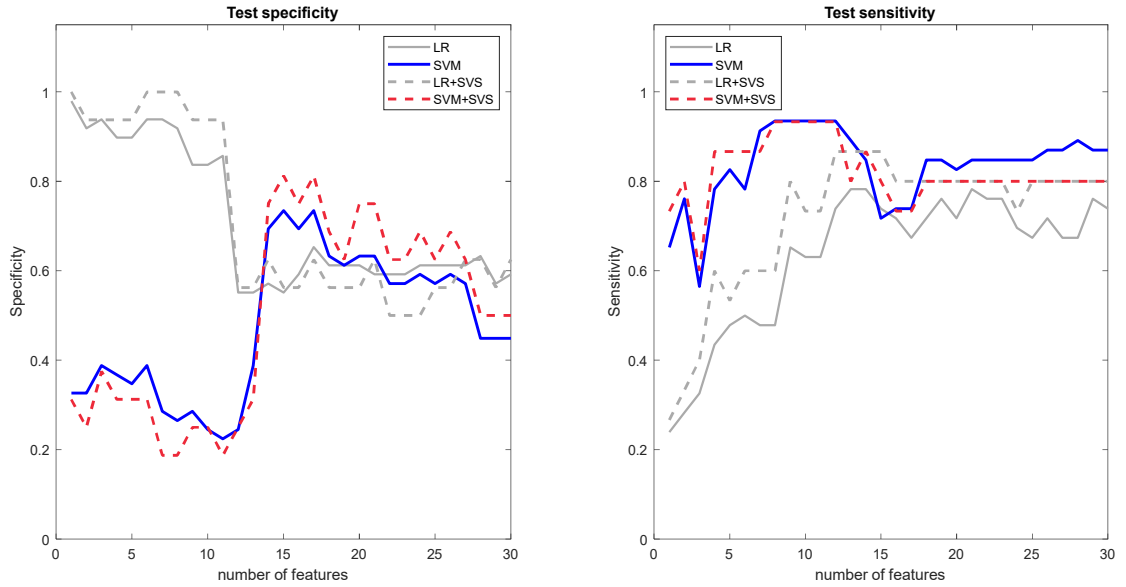
Supplementary Figure 13 – KNN cubic model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. The performance of the LR model is plotted as a reference.



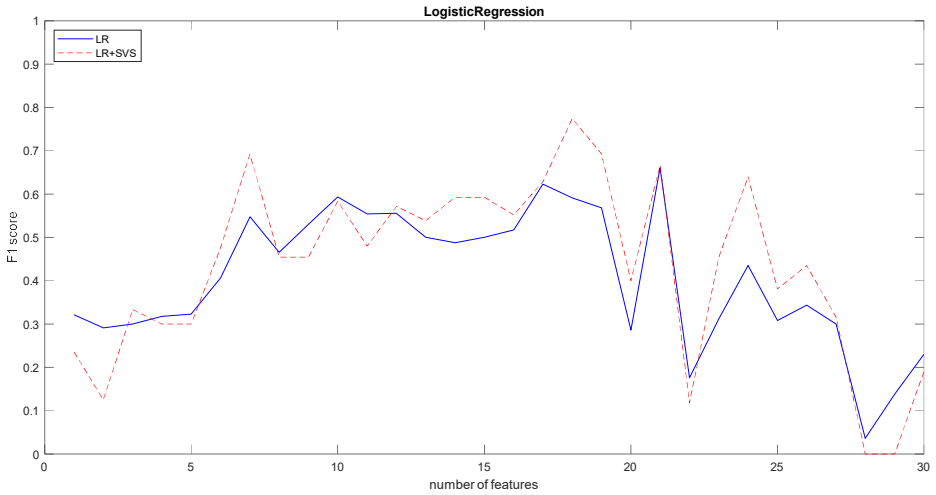
Supplementary Figure 14 – KNN weighted model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using a filter-based method. The performance of the LR model is plotted as a reference.



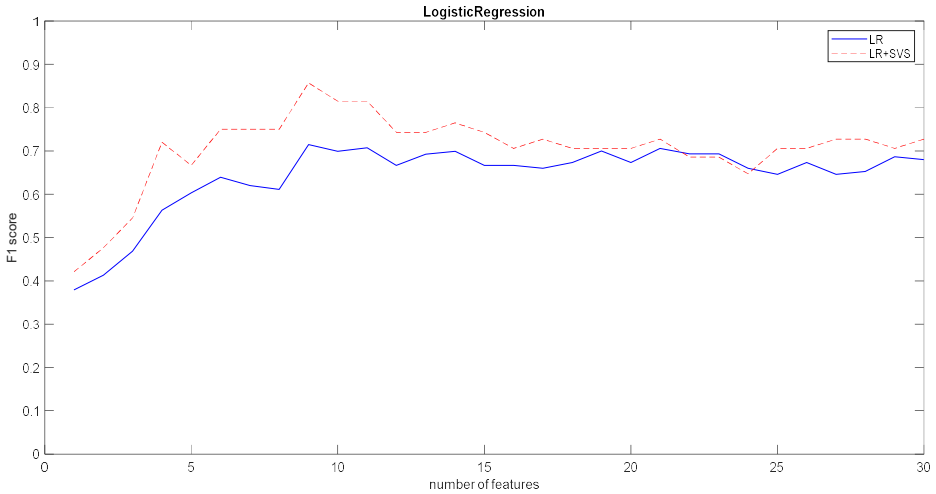
Supplementary Figure 15 – SVM with polynomial kernel of degree 3 model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using an embedded-wrapper method. The performance of the LR model is plotted as a reference.



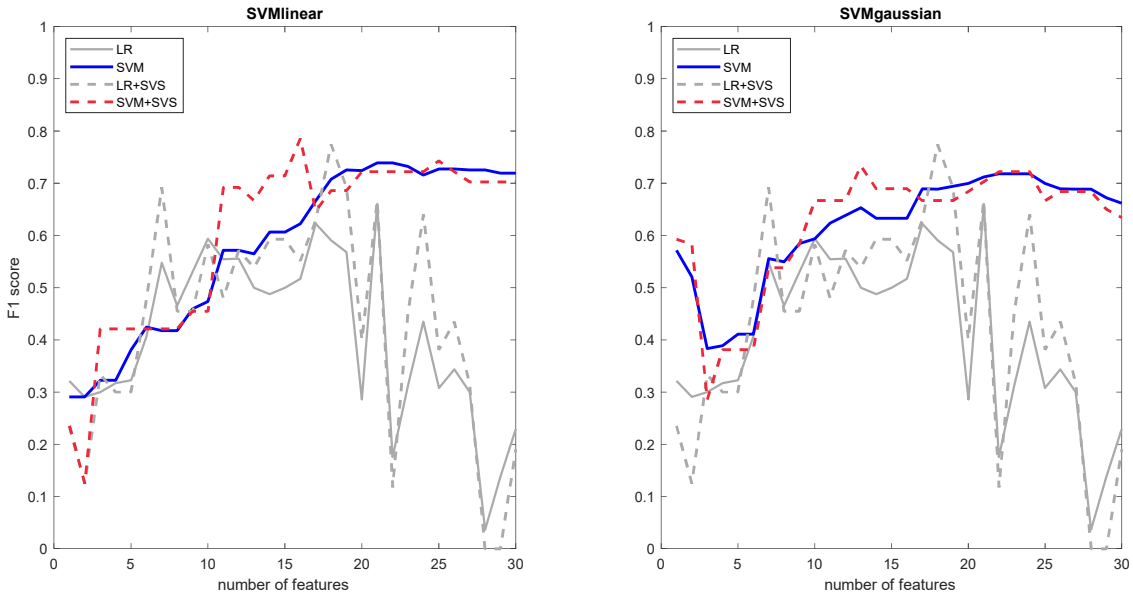
Supplementary Figure 16 – SVM with polynomial kernel of degree 4 model performance, in terms of model test specificity (left) and sensitivity (right). These performance parameters are plotted as a function of the number of selected features (i.e., from 1 to 30 features) using an embedded-wrapper method. The performance of the LR model is plotted as a reference.



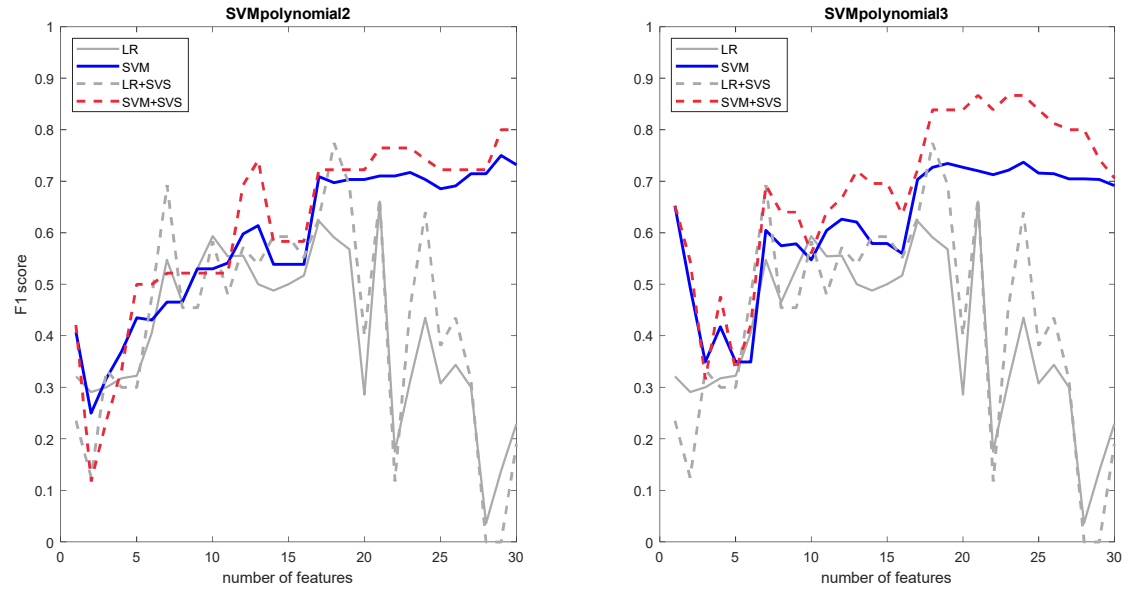
Supplementary Figure 17 – LR model performance, in terms of F1-score. The F1-score is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using the t-test filter method.



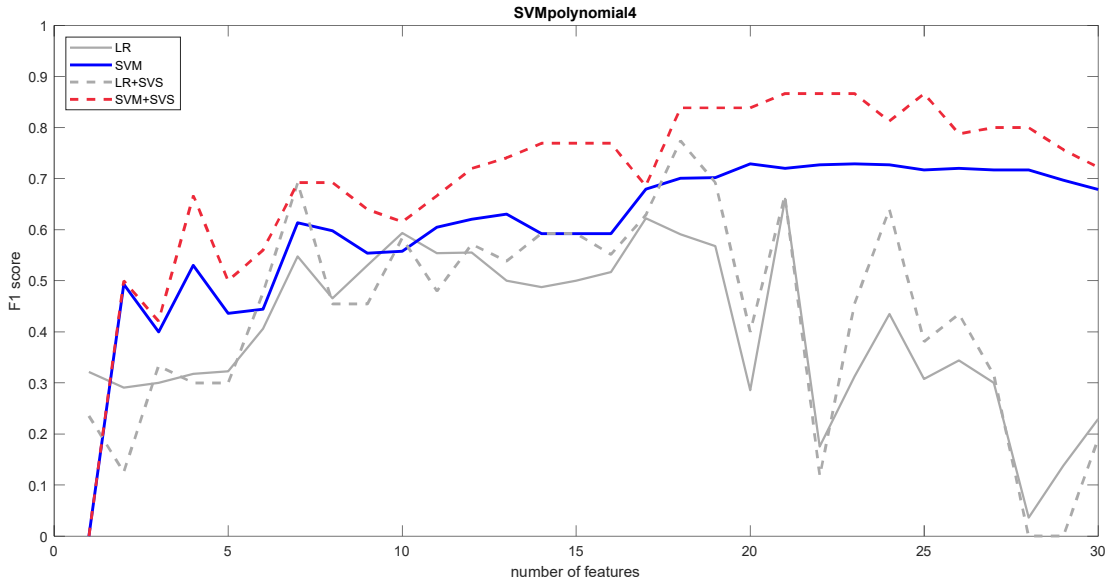
Supplementary Figure 18 – LR model performance, in terms of F1-score. The F1-score is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using the embedded-wrapper method.



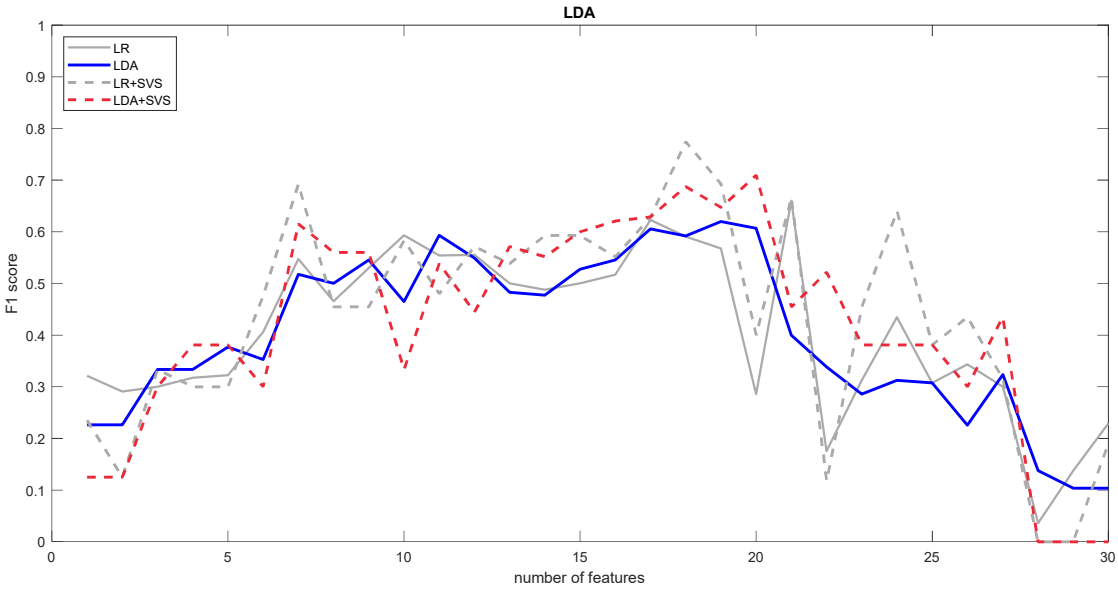
Supplementary Figure 19 – SVM model performance, in terms of F1-score, when a linear kernel (left) or Gaussian kernel (right) is used. The F1-score is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using the t-test filter method.



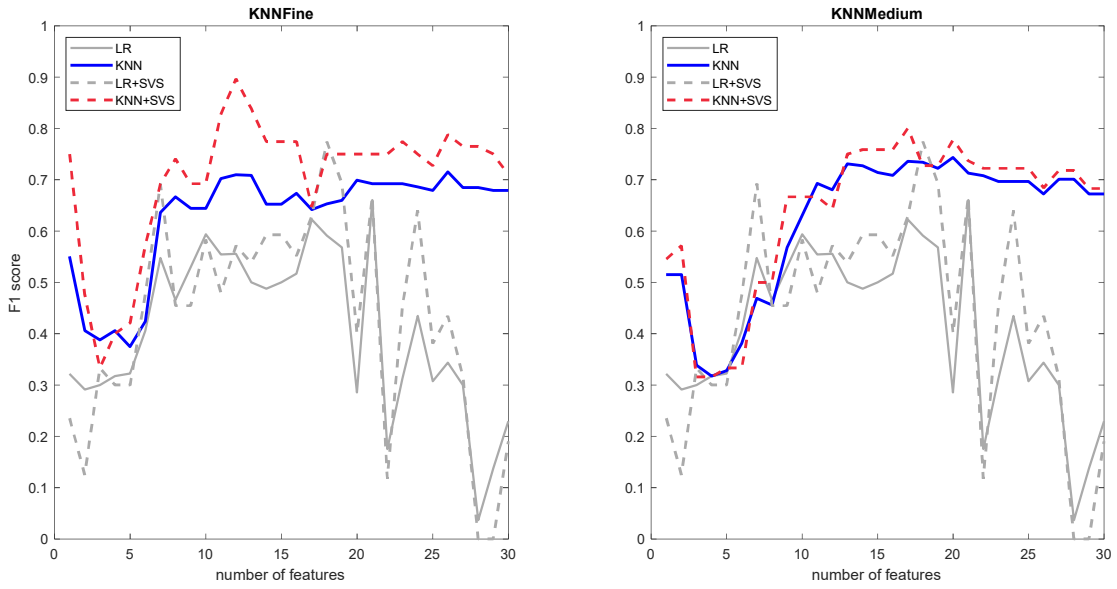
Supplementary Figure 20 – SVM model performance, in terms of F1-score, when a polynomial kernel of degree 2 (left) or 3 (right) is used. The F1-score is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using the t-test filter method.



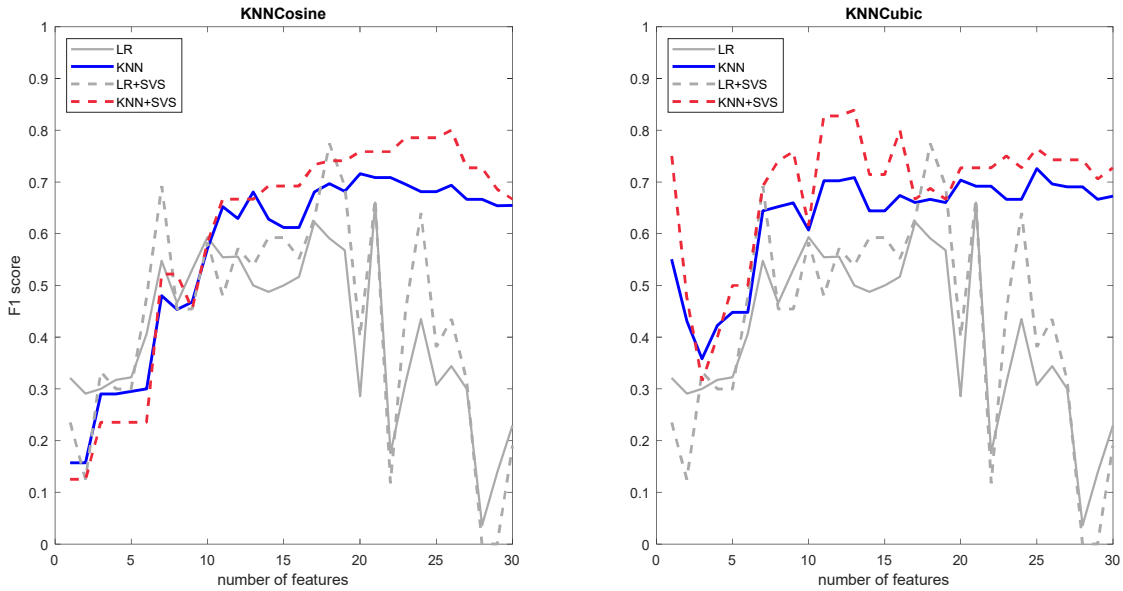
Supplementary Figure 21 – SVM model performance, in terms of F1-score, when a polynomial kernel of degree 4 is used. The F1-score is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using the t-test filter method.



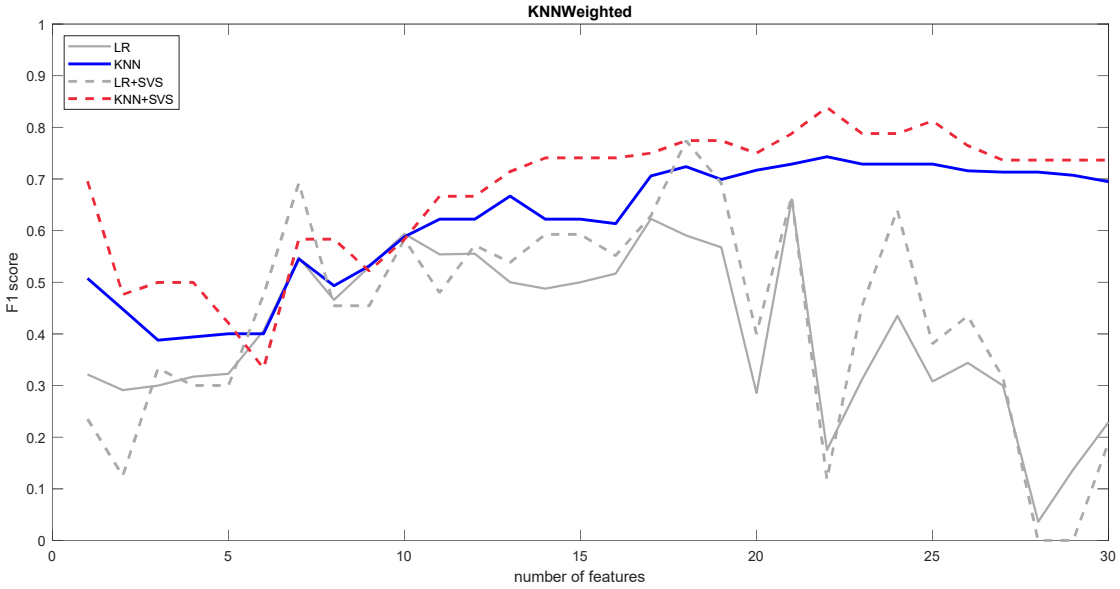
Supplementary Figure 22 – LDA model performance, in terms of F1-score. The F1-score is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using the t-test filter method.



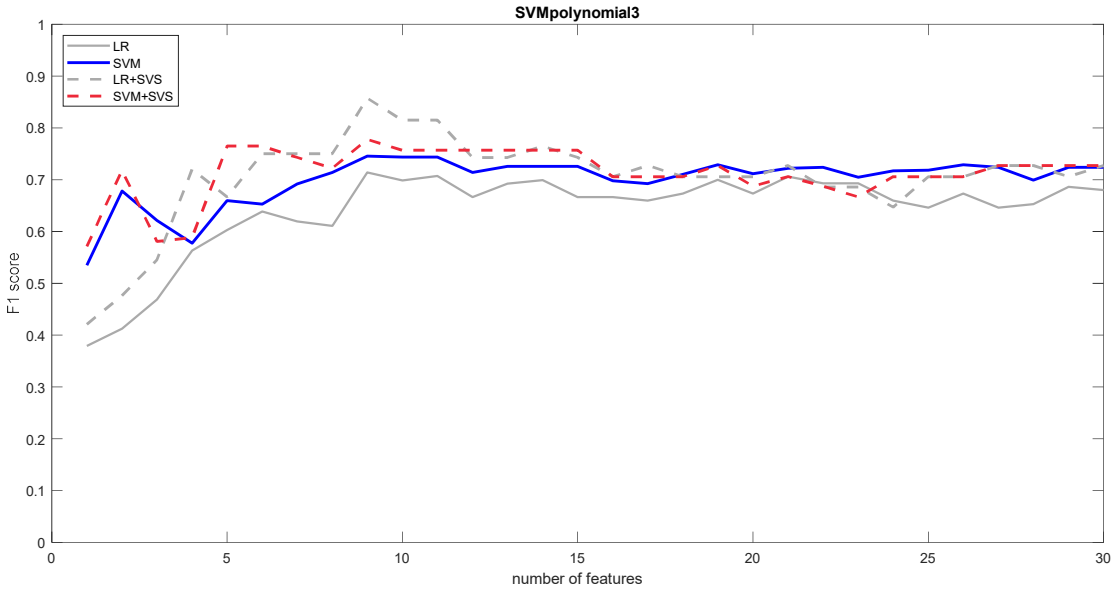
Supplementary Figure 23 – KNN model performance, in terms of F1-score, when a KNN fine (left) or KNN medium (right) model configuration is used. The F1-score is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using the t-test filter method.



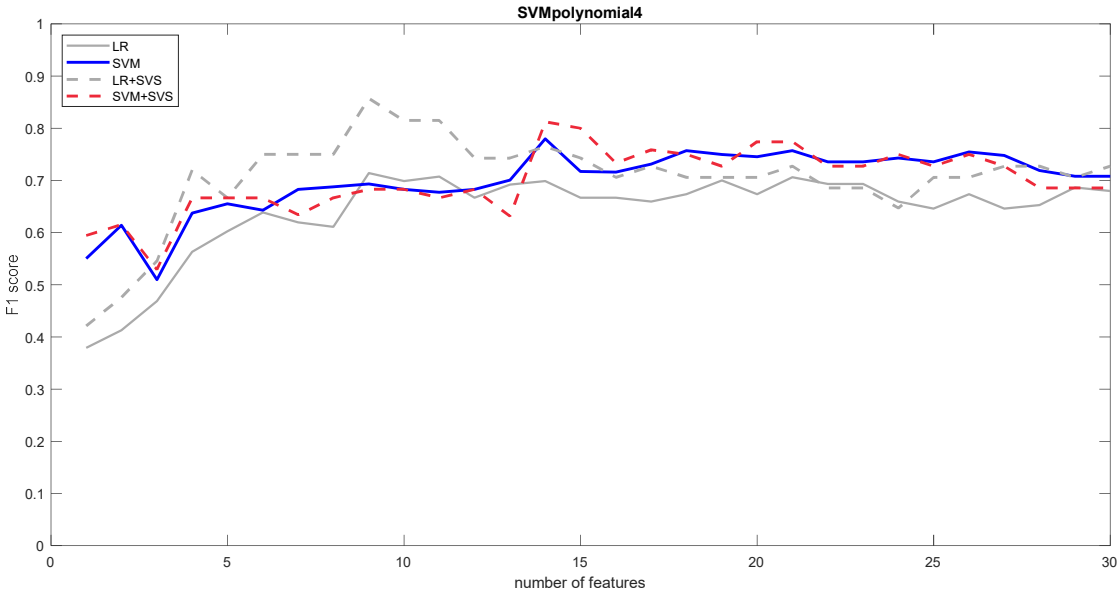
Supplementary Figure 24 – KNN model performance, in terms of F1-score, when a KNN cosine (left) or KNN cubic (right) model configuration is used. The F1-score is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using the t-test filter method.



Supplementary Figure 25 – KNN model performance, in terms of F1-score, when a KNN weighted model configuration is used. The F1-score is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using the t-test filter method.



Supplementary Figure 26 – SVM model performance, in terms of F1-score, when a polynomial kernel of degree 3 is used. The F1-score is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using the embedded-wrapper method.



Supplementary Figure 27 – SVM model performance, in terms of F1-score, when a polynomial kernel of degree 4 is used. The F1-score is plotted as a function of the number of selected features (i.e., from 1 to 30 features) using the embedded-wrapper method.