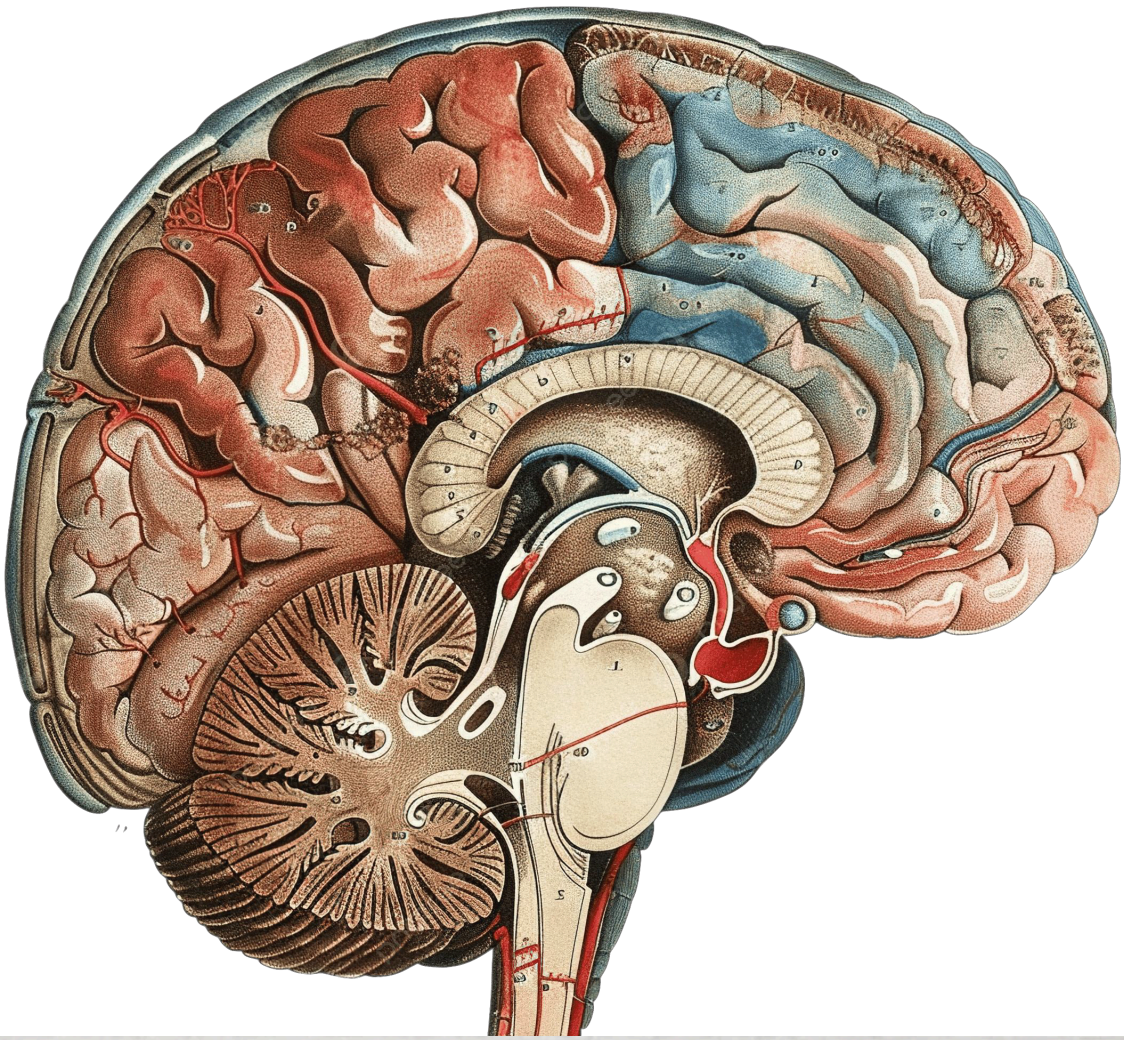
ST. XAVIER'S COLLEGE (AUTONOMOUS), KOLKATA

DISTINGUISHING PSEUDOPROGRESSION FROM TRUE PROGRESSION IN GLIOBLASTOMA

Multiparametric MRI-Based Radiomics and Machine Learning Approach



SUPERVISOR

DR. SOURAV BHADURI

Assistant Professor, TCG Crest

INTERNAL MENTOR

DR. ROMIT BEED

Associate Professor, St.Xavier's College

DECLARATION

I affirm that I have identified all my sources and that no part of this dissertation paper contains unacknowledged materials.

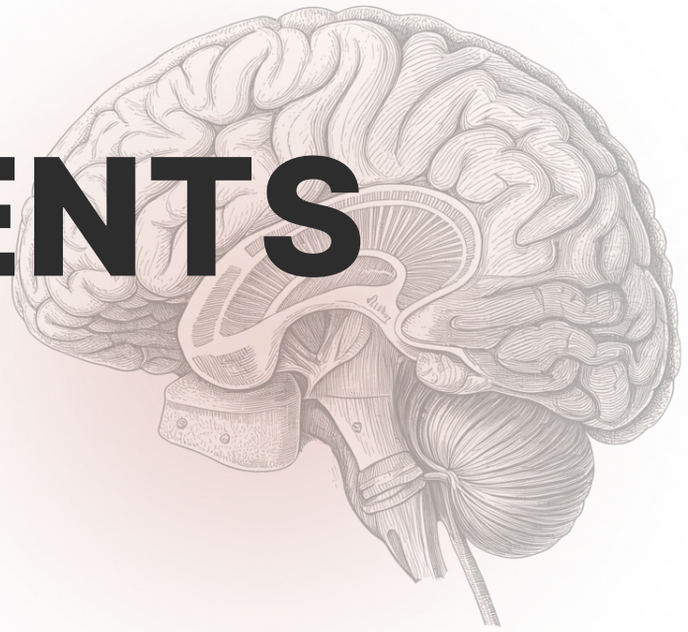
DETAILS

SUCHIBRATA PATRA

CIN **24-500-5-08-0403**

Reg **A01-1112-0867-21**

Academic Year **2026**



CONTENTS

PREFATORY

Declaration & Originality
Acknowledgements
Abstract

Introduction

1 Giloblastoma (GBM)
2 True Progression (TP)
2 Pseudoprogression (PsP)
4 Magnetic Resonance Imaging
5 Dynamic Contrast Enhanced MRI
5 Volume Transfer Constant
6 EES Volume
6 Plasma Volume Fraction
6 Plasma Flow Rate
6 Intracellular Water Exchange Time
6 MGMT Promoter Methylation

Motivation & Objecives

8 Problem Statement
9 Motivation of the Study
9 Research Hypothesis
10 Significance of the Study

Literature Review

11 Detailed Literature Review of the Paper

Methodology

15 Study Cohort
15 Clinical Data Acquisition
17 Extraction of the Parameter Maps
18 Calculation of Tumor Volume
18 Developing Parsimonious Model
21 Feature Selection
21 Class Imbalance & CV
22 Model training
22 Classification Threshold Selection
23 Machine Learning Models
23 Evaluation Metrices

Result & Analysis

24 Extended TOFTS + MGMT
26 T1 Post Contrast + MGMT
28 Parsimonious DCE
29 Parsimonious DCE + MGMT

Discussion & Conclusion

31 Study Discussion
36 Conclusion

Research Limitations

Software & Libraries

References

Supplimentary

47 Two Compartment Model
48 Scanner to Signal Concentrationi
49 Standard TOFTS Model
51 Extended TOFTS Model
52 Two Compartment Exchange Model

Declaration of Originality and Adherence to Academic Ethics

I, Suchibrata Patra, hereby declare that this dissertation is an original work produced through independent inquiry and rigorous academic effort. All arguments, interpretations, and conclusions presented are my own, except where the ideas, words, or findings of other scholars have been duly acknowledged through formal citation and attribution. The conduct of this research has at all times been guided by, and remains fully consistent with, the academic integrity standards and ethical codes of conduct established by St. Xavier's College (Autonomous), Kolkata. All sources - encompassing theoretical frameworks, empirical evidence, quantitative and qualitative analyses, and antecedent scholarship - have been systematically identified, appropriately credited, and referenced in accordance with disciplinary conventions. I affirm that this dissertation, in its entirety or in any substantial portion, has not previously been submitted for any degree, diploma, or academic qualification at this or any other institution, nor has it been submitted for publication or formal assessment, except where such prior submission has been explicitly disclosed. All external sources, theoretical paradigms, research methodologies, datasets, and intellectual contributions of third parties have been recorded with scholarly diligence and precision, in full conformity with accepted academic practice. Where secondary literature or pre-existing datasets have been employed, their scope, provenance, and inherent limitations have been critically recognised and addressed with intellectual honesty. Where this research has engaged with sensitive domains - including human participants, clinical or medical data, or personally identifiable information - such engagement has been conducted in strict accordance with applicable ethical guidelines, institutional protocols, and regulatory frameworks. All requisite approvals from competent ethical review authorities have been formally obtained and documented prior to research initiation. I am fully cognisant that academic misconduct - including plagiarism, fabrication of data, falsification of findings, unauthorised collaboration, or misrepresentation of scholarly contribution - constitutes a grave violation of academic and professional ethics, and is subject to serious disciplinary action with potentially severe academic, institutional, and professional consequences. I accept unconditional and complete responsibility for the integrity, authenticity, originality, and veracity of every dimension of this dissertation as presented. This work is submitted in full awareness of the scholarly obligations it represents, and with the conviction that it upholds the enduring values of intellectual honesty, academic rigour, and ethical scholarship for which this institution is distinguished.



Suchibrata Patra
PG, St.Xavier's College (Autonomous), Kolkata



Dr. Sourav Bhaduri
Institute for Advanced Intelligence, TCG CREST

Dr. Romit Beed
St.Xavier's College (autonomous), Kolkata

Acknowledgement

This dissertation would not exist without the kindness and support of several people whose contributions have been both intellectual and deeply personal.

Dr. Sourav Bhaduri at the Institute for Advancing Intelligence (IAI), TCG CREST, has been instrumental in this work - not just as a mentor, but as someone genuinely invested in opening doors. He went to considerable effort to secure the University of Pennsylvania dataset that sits at the heart of this research, leveraging relationships and advocacy that made what seemed difficult feel possible. More than that, he has engaged with the work itself with real thoughtfulness, pushing me to think more clearly and write more precisely. His combination of generosity and rigor has shaped this entire endeavor.

This Work was carried out under the guidance of Dr. Sourav Bhaduri as part of research activities supported by the Ramalingaswami Re-entry Fellowship funded by the Department of Biotechnology(DBT), Government of INDIA, awarded to him.

I bow with deep gratitude to the Almighty, and with equal reverence to **The Ramakrishna Mission** whose light has quietly guided me through every phase of this journey. There were moments along the way when the road ahead seemed shrouded in uncertainty, and the burden of self-doubt grew heavy enough to slow my steps. It was during those very moments that the timeless teachings and spiritual wisdom of The Ramakrishna Mission became my steadfast companion. The transformation they brought was not sudden or dramatic it was something far more enduring: a gentle, growing clarity; a stillness that settled within and refused to leave. When I felt overwhelmed, their teachings did not merely offer comfort — they offered direction. They reminded me to be patient, to stay rooted, and to trust the process of moving forward, even when the destination was not yet fully visible. Step by step, that inner grounding became the foundation upon which this work was built. Through the difficult seasons and the more hopeful ones alike, that sense of inner stability never wavered. It held me together when I needed it most, and quietly celebrated with me when things began to fall into place. For this , for a strength that came not from the outside, but was awakened from within I am sincerely, humbly, and profoundly grateful.

I am also deeply grateful to **Mr. Soumen Maity**, my mathematics teacher, who built the bedrock upon which my entire intellectual foundation rests. He took time to ensure I grasped the fundamentals deeply - not just memorized, but truly understood. The clarity of thought, the approach to problem-solving, the appreciation for rigor that he instilled in me have become the compass guiding every step of my career. Years later, as I navigate complex research and push myself to achieve more, I find myself returning to lessons he taught me. That solid foundation he gave me is what has allowed me to soar. He shaped not just my knowledge of fundamental of mathematics, but how I think about learning itself. For that, I am profoundly grateful.

I am equally grateful to **Mr. Sourav Basak** at TCG CREST, whose steady presence and collaborative spirit have been a consistent source of support. He has engaged with the technical challenges that arose not as obligations but as genuine problems worth solving together. His encouragement, offered without fanfare, has been quietly sustaining.

I am especially indebted to **Dr. Madhura Dasgupta**, whose belief in me during difficult academic moments became the turning point I needed. When my performance wavered and doubt crept in, she did not retreat but instead offered steady encouragement and real support. She saw beyond the momentary struggle to what might be possible. Her quiet faith that I could recover and eventually flourish made all the difference. Without that timely encouragement and confidence, I would not have found my way to this achievement. To her, I owe a debt of deep gratitude.

I am equally grateful to **Dr. Surupa Chakraborty**, who has been a quiet but steadfast guide in building my understanding of this domain. Her patient guidance on the technical and conceptual foundations of this work, offered without fanfare, has been instrumental in shaping my approach and deepening my expertise. She has the rare quality of helping others develop mastery while remaining quietly in the background. Her contributions to this research, though understated, have been essential.

I owe a special gratitude to **Mr. Rohit Purkait**, my closest friend, who stood by me during the early stages of this project when everything felt uncertain. His belief in what I was attempting to do, his readiness to listen and encourage, and his consistent presence made those foundational moments less daunting. He has been the kind of friend who asks the right questions and offers support without needing to understand every technical detail. I am fortunate to have him. I extend my thanks to **Dr. Romit Beed** and H.O.D **Dr. Durba Bhattacharya** at St. Xavier's College (Autonomous), Kolkata, for her continued support and interest in this research, and to Dr. Durba Bhattacharya, Head of the Department of Data Science, for fostering a departmental culture where serious scholarship is genuinely valued and encouraged. Their grace has been a quiet companion through this entire work.

Finally, my deepest gratitude goes to my parents, **Mr. Kamal Kr. Patra** and **Mrs. Susmita Maity** Their unconditional love, belief, and quiet sacrifices have been the foundation of everything I have accomplished. This dissertation stands as a small reflection of their faith in me and the home they have built. To **Ms. Kathakali** and **Mr. Biswarup**, my fellowmates and collaborators too - I owe them a special gratitude. The meticulous work they poured into the MRI preprocessing pipeline, from skull stripping to careful quality assurance, reflects the kind of conscientiousness that distinguishes good work from excellent work. Beyond the technical contribution, they brought friendship and care to the collaboration. They made this research feel like something genuinely shared.



Warm Regards,
Suchibrata Patra

Abstract

GBM (GBM), the most prevalent and aggressive primary brain malignancy in adults, is designated as WHO Grade IV, IDH-wildtype, adult-type diffuse astrocytic tumor and poses significant therapeutic challenges due to its infiltrative nature and treatment resistance. Distinguishing True Progression (TP) from PsP (PsP) following radiotherapy is paramount for optimizing patient management, yet conventional imaging modalities often fail to provide reliable differentiation. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) emerges as a sophisticated technique that tracks contrast agent kinetics, yielding functional insights into vascular permeability and tissue microenvironment beyond anatomical structure.

Leveraging the PyRadiomics toolbox, this investigation quantifies radiomic features extracted from DCE-MRI-derived pharmacokinetic maps to refine TP versus PsP classification. Five key parameters $K^{trans}, V_e, V_p, F_p, \tau_i$ were modeled and integrated into machine learning frameworks, including Random Forest (RF), Support Vector Classifier (SVC), Logistic Regression (LR), and XGBoost. Feature selection via Lasso, Elastic Net, and Ridge regression was applied alongside full-feature analyses.

The Random Forest classifier integrating parsimonious DCE-derived radiomic features with MGMT promoter methylation achieved the highest performance, with a test AUC of 0.89 ± 0.07, sensitivity of 0.93 ± 0.09, specificity of 0.76 ± 0.19, and F1-score of 0.90 ± 0.07. These findings demonstrate robust discriminative capability and highlight the synergistic value of combining physiologically grounded imaging biomarkers with molecular information.

These findings demonstrate that pharmacokinetic-derived radiomics coupled with machine learning can augment TP-PsP discrimination, thereby refining post-treatment surveillance in GBM and informing precision oncology strategies to enhance clinical outcomes.

Chapter 1
Introduction

Glioblastoma (GBM)

Glioblastoma multiforme (GBM) is the most malignant subtype of glioma and the most common primary brain tumor in adults. It accounts for 45.2% of primary malignant Central Nervous System (CNS) tumors [1], with an estimated incidence of approximately 35 cases per million individuals worldwide. Despite decades of extensive clinical & investigative effort, GBM still remains essentially incurable. The median survival following diagnosis is only 14-16 months , with a 5-year survival rate of just 5.5% [1].

Histologically, GBM is characterised by several features, including but not limited to poorly differentiated pleomorphic cells - irregular and inconsistent in both size and shape that nonetheless retain a degree of astrocytic differentiation, suggesting that even in their highly aberrant state, tumour cells preserve some structural and molecular resemblance to normal astrocytes. This is accomplished by high mitotic activity (rapid and continuous tumor cell division at an abnormally accelerated rate), Microvascular Proliferation (MVP), and necrosis. The majority of GBM cases, around ~90% are classified as Isocitrate Dehydrogenase (IDH) wildtype [1]. These arise de novo, predominantly in older patients, and are driven by a constellation of genetic alterations including EGFR amplification, TERT promoter mutations, and MGMT changes, all of which contribute to their characteristically poor prognosis. On the contrary, a smaller IDH-mutant subset usually tend to develop from the pre-existing (specially in younger patients) lower grade gliomas. This is relatively favourable in terms of clinical outcome. Among the various molecular bio markers studied, O^6- Methylguanine-DNA methyltransferase (MGMT) promoter methylation and IDH mutational status have emerged as the most clinically meaningful, directly informing treatment decisions and serving as key predictors of patient outcomes.

According to the 2021 WHO Classification of CNS Tumors (5th edition), GBM is formally designated as a **WHO Grade IV, IDH-wildtype, adult-type diffuse astrocytic tumor** [2] - the highest tier of malignancy, integrating both histological and molecular criteria to ensure diagnostic precision and accurate prognostic stratification.

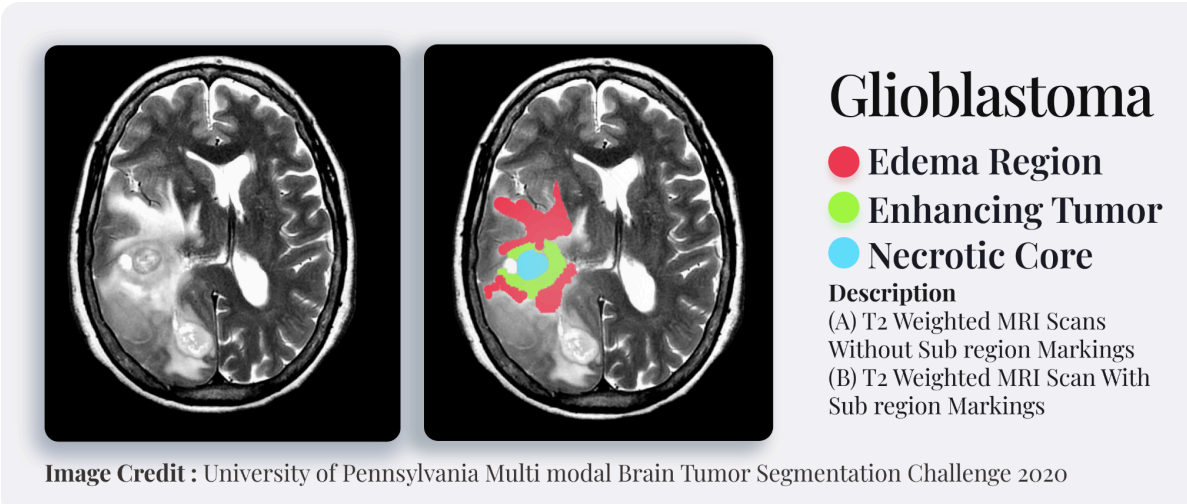


Figure 1: T2 Weighted Scan MRI of GBM patients with and without tumor subregion marking

True Progression (TP)

TP reflects genuine post-operative tumour advancement, driven by active proliferation of residual viable GBM cells and progressive infiltration into the surrounding brain tissue. On Gadolinium based Contrast Agent (GCA) enhanced MRI, $TP\ presents\ as\ a \geq 25\%$ increase in the sum of contrast-enhancing lesion diameter, consistent with mRANO criteria [3]. It represents authentic biological growth rather than treatment-related inflammatory changes. Its identification carries direct clinical consequence, as it definitively signals failure of therapeutic control and correlates with accelerating disease course and poor survival outcomes.

Pseudo Progression (PsP)

PsP is a transient, treatment induced inflammatory and vascular phenomenon, that which closely mimic TP[4]. Following chemoradiotherapy, disruption of the blood-brain barrier (BBB), localised edema, and immune-mediated tissue responses generate new or enlarged contrast-enhancing lesions on MRI[5]. Which subsequently stabilise or regress spontaneously without modification to the oncological regimen. Unlike TP, PsP reflects no underlying neoplastic expansion. It constitutes a recognised, self-limiting biological response to therapy. Nevertheless, its misidentification as genuine recurrence carries significantly negative clinical consequence[6].

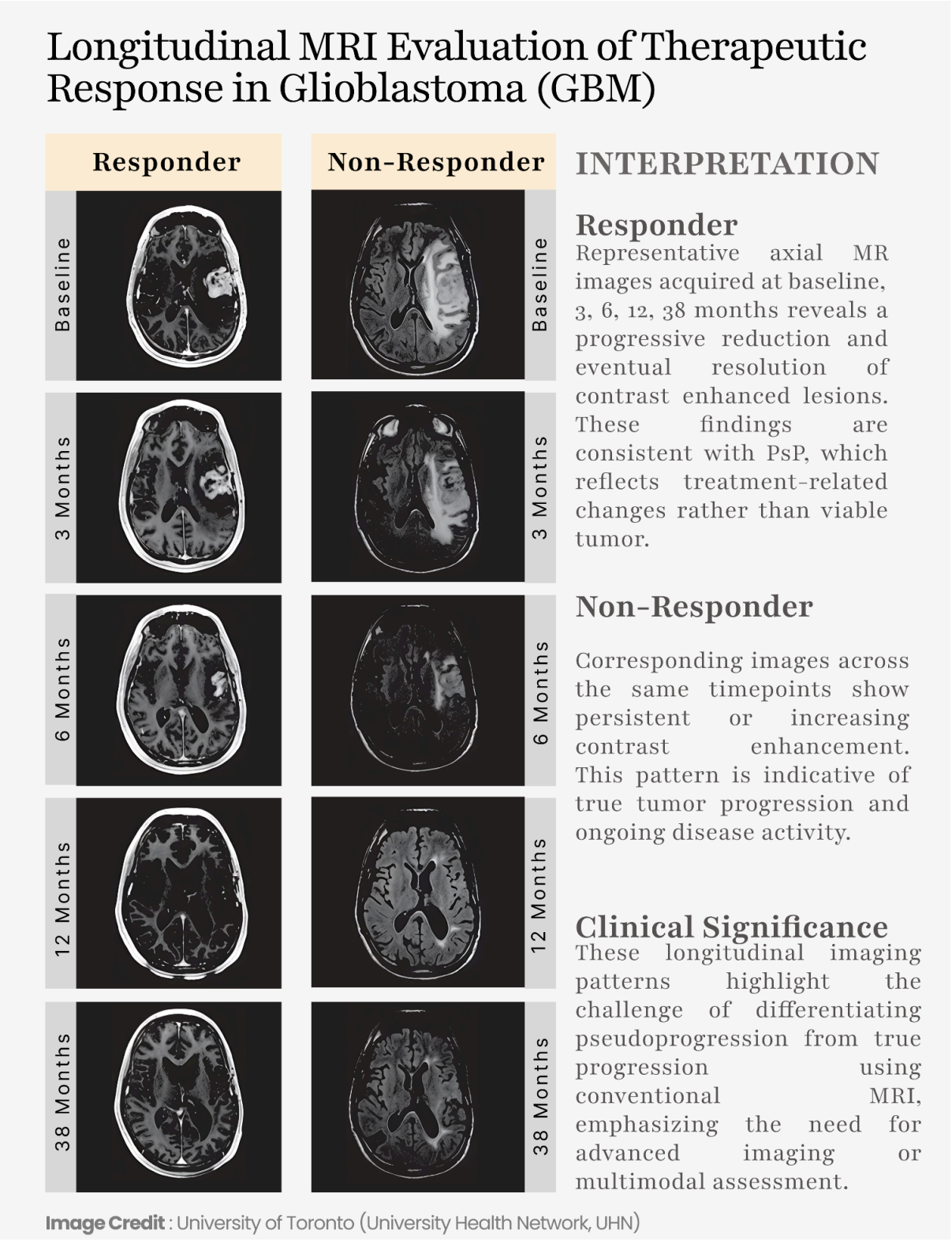


Figure 2: Longitudinal MRI Evaluation of Therapeutic Response in Brain Pathology of GBM

Magnetic Resonance Imaging (MRI)

MRI is a non-invasive diagnostic modality that generates high-resolution structural images by exploiting the magnetic properties of hydrogen protons inherent to biological tissues [7]. Its fundamental advantage over CT and conventional radiography lies in the complete absence of ionising radiation, a clinically significant distinction in GBM management where repeated longitudinal imaging is an unavoidable necessity[8]. The underlying principle draws from nuclear magnetic resonance. Within a strong static $magnetic\ field$ (B_0), hydrogen protons align along the field axis, establishing a net magnetisation [7]. A targeted radio frequency pulse transiently disrupts this equilibrium, exciting the protons into a higher energy state [9]. Upon pulse withdrawal, the protons relax back to stable baseline, releasing tissue-specific signals that are captured by receiver coils and subsequently reconstructed, through complex mathematical transformations, into the cross-sectional images central to clinical diagnosis and disease monitoring [9,10].

Image Formation

Image formation in MRI is a structured, sequential process. An RF pulse first displaces proton magnetisation from its equilibrium state, initiating excitation [9]. Recovery follows through two distinct relaxation mechanisms, $T1\ (longitudinal)\ and\ T2\ (transverse)$ relaxation, each reflecting fundamentally different tissue-specific properties [9]. Spatial localisation of the emitted signals is achieved through magnetic field gradients applied across the imaging volume [2]. The resulting data populates a spatial frequency matrix, k-space, which is subsequently decoded into a diagnostically interpretable anatomical image through an inverse Fourier Transform [2]. In the present study, MRI constitutes the primary imaging modality for brain tumour evaluation. Its soft tissue contrast resolution is, by any clinical measure, unmatched [8]. This allows meaningful visualisation of tumour architecture, internal heterogeneity, and infiltrative margins [8]. Multiparametric protocols, specifically contrast-enhanced $T1, T2/FLAIR$, Diffusion Weighted Imaging (DWI), and Perfusion Weighted Imaging (PWI), together enable characterisation of tumour biology, cellularity, and vascular dynamics that no single sequence can provide alone [9]. Longitudinally, MRI remains the cornerstone of treatment response assessment in GBM. Its capacity to differentiate true progression from post-operative confounders, particularly pseudoprogression and radiation necrosis, is what renders it indispensable in the clinical management of this disease [9].

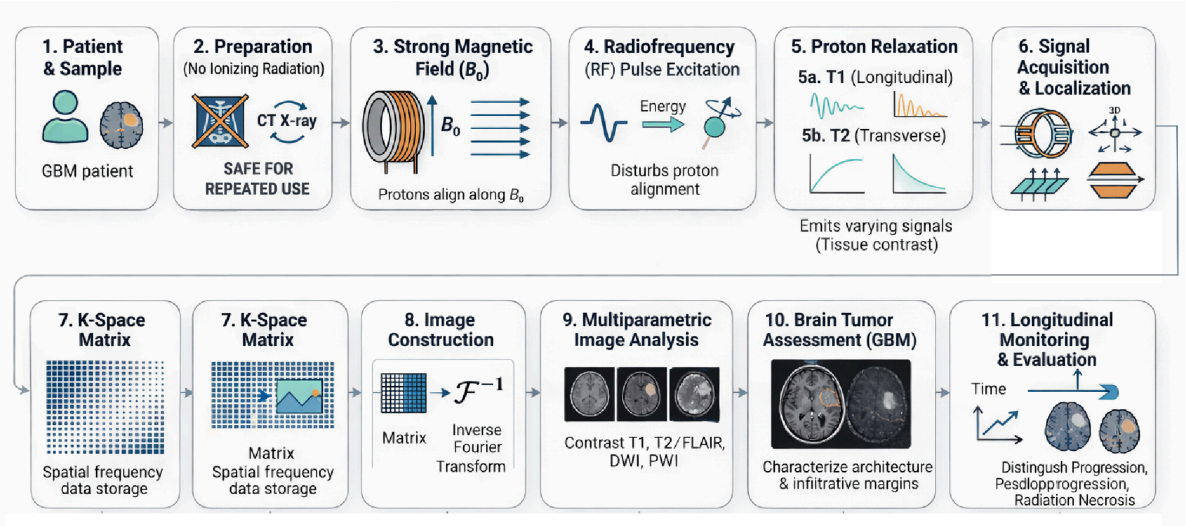


Figure 3: MRI physics steps - excitation relaxation k-space diagram

Dynamic Contrast Enhanced (DCE) MRI

DCE-MRI is an advanced perfusion imaging technique, which evaluates tissue vascularity, capillary permeability, and microenvironmental physiology by tracking the temporal dynamics of a GCA [11]. Serial/sequences of images are captured before, during, and after GCA administration. Which enables us for quantitative assessment of how rapidly and to what extent GCA distributes within *intravascular* and *extravascular* compartments [12]. The signal time curves obtained from DCE-MRI are fitted using pharmacokinetic(PK) models most commonly the Tofts or extended Tofts model to derive quantitative biomarkers that reflect vascular integrity, perfusion, and microstructural properties of tissue [13]. These metrics have become essential in neuro-oncology for differentiating TP from treatment-induced PsP, as GBM profoundly alters vascular permeability and extracellular space [5].

Volume Transfer Constant (K^{trans})

K^{trans} is a measure of capillary permeability obtained using dynamic contrast-enhanced (DCE) MR perfusion [11]. It is calculated by measuring the accumulation of GCA in the Extravascular Extracellular Space (EES).

Increased permeability of vessels in the brain is seen in a variety of conditions [14]. One is the classical example is in GBM, Vascular Endothelial Growth Factor (VEGF) produced by the tumor reduces the effectiveness of gap junction and creates fenestrations in the endothelium of brain capillaries; this has been shown to correlate with tumor grade [1].

K^{trans} is interpreted varied according to permeability and blood flow. Where permeability is very low (near-intact blood brain barrier) then K^{trans} is a useful measure of permeability, whereas in situations where permeability is very high (disrupted blood brain barrier) then K^{trans} will reflect blood flow [13].

Extravascular Extracellular Space Volume (V_e)

Ve represents the fraction of tissue volume occupied by the extracellular space, excluding both blood plasma and intracellular volume [13]. It provides insight into: Cellularity, interstitial space expansion, necrotic or edematous regions and tumor microenvironment looseness. Higher Ve values often correspond to necrosis or edema, whereas lower Ve is seen in densely packed, highly cellular tumor regions [12].

1.8 Plasma Volume Fraction (V_p)

Vp measures the fractional volume of tissue occupied by blood plasma [11]. It serves as a practical indicator of tumour vascular density, microvascular proliferation, and overall capillary supply. In GBM, abnormally elevated Vp values are not incidental. They reflect the chaotic, proliferative vasculature. Which is often recognised biological benchmark of aggressive disease, and are frequently encountered in regions confirmed as TP [14].

Plasma Flow Rate (F_p)

F_p describes the rate of plasma delivery per unit volume of tissue, representing true tissue perfusion [13]. It is shaped by several interacting factors, including blood flow, vessel density, vascular resistance, and capillary transit time. In practice, reduced F_p values, typically falling below $0.1\,ml/min/mL$, tend to reflect hypoperfusion commonly encountered in necrotic cores or treatment-affected tissue [5]. Conversely, elevated F_p, often exceeding $0.3\,ml/min/mL$ in high-grade lesions, is more telling. It points toward active angiogenesis and *aggressive tumour behaviour*, rendering it a physiologically meaningful parameter in distinguishing viable, proliferating tumour from Standard of Care (SOC) related change [11].

Intracellular Water Exchange Time (τ_i)

τ_i quantifies the mean intracellular water lifetime, representing the average duration water molecules spend within cells before exchanging with the extracellular compartment [15]. It reflects cell membrane permeability, cellular viability, metabolic activity, and microstructural integrity. In healthy brain tissue, τ_i typically ranges between $200\ and\ 600\ ms$, with malignant regions demonstrating marked deviations [15]. Prolonged τ_i suggests compromised membrane integrity or reduced cellular turnover. Whereas shortened values point toward heightened metabolic activity and active tumour proliferation. In GBM, τ_i alterations thus serve as a biologically grounded indicator of therapy response and tumour dynamics [11].

O6 Methylguanine DNA Methyltransferase (MGMT)

MGMT promoter methylation is a key molecular biomarker that significantly influences prognosis and therapeutic response in GBMs [2]. MGMT encodes a DNA repair enzyme that reverses alkylation-induced DNA damage, thereby diminishing the cytotoxic efficacy of alkylating agents such as temozolomide (TMZ) and promoting chemoresistance [1]. Methylation of the MGMT promoter leads to reduced gene expression, rendering tumor cells more vulnerable to alkylating chemotherapy [2]. Patients with PsP and MGMT-methylated tumors show better treatment response and longer progression-free and overall better survival than TP patients with an unmethylated MGMT promoter [4].

Chapter 2
Study Motivation and Objectives

Problem Statement

Conventional magnetic resonance imaging (MRI) lacks the specificity required to distinguish TP from PsP in GBM, as both entities manifest nearly identical contrast-enhancing patterns on follow-up scans [5]. Visual interpretation remains inherently subjective, often leading to inter-observer variability, diagnostic discordance, and inappropriate clinical decisions [6]. Multiparametric MRI offers a partial remedy by incorporating physiological imaging and quantitative modelling [11]. Dynamic Contrast-Enhanced (DCE) MRI characterises vascular permeability and microcirculatory dynamics, while pharmacokinetic(PK) modelling yields parameters such as $K^{trans}, V_e, V_p, F_p, \tau_i$ [13].For optimal PK parameters, a parsimonious modelling framework, guided by the Akaike Information Criterion (AIC), is employed [16]. In parallel, complementary imaging sequences, including Diffusion Tensor Imaging (DTI) and Dynamic Susceptibility Contrast (DSC) MRI, characterise tissue microstructural organisation and haemodynamic properties, thereby augmenting the physiological representation of tumour behaviour [11]. Radiomics extends these modalities by extracting high-dimensional quantitative descriptors of intensity, texture, morphology and spatial heterogeneity features inaccessible to visual analysis [17]. When coupled with molecular attributes such as O6-methylguanine-DNA-methyltransferase promoter methylation (MGMT), radiomics generates a biologically informed feature set ideal for ML classifiers [2]. In this investigation, imaging data from 82 patients with GBM underwent radiomic feature extraction from the available multiparametric MRI sequences, producing a large feature space necessitating rigorous reduction [17]. A two-stage selection pipeline was implemented: initial univariate filtering using the Mann Whitney U test, followed by Elastic Net regularisation to manage multicollinearity and isolate the most discriminative predictors [18]. The refined feature subset was subsequently used to train and evaluate multiple machine learning classifiers, including Random Forest (RF), Support Vector Machine (SVM), Logistic Regression, K-Nearest Neighbours (KNN), and XGBoost [19]. While these approaches collectively demonstrate considerable diagnostic promise, a unified and rigorously validated predictive framework that meaningfully integrates PK driven

radiomics, multiparametric MRI, and machine learning based classification remains largely underexplored [11]. This gap directly motivates the present study.

Motivation of the study

Assessing treatment response in GBM remains one of the more demanding problems in neuro-oncology [4]. The central difficulty is distinguishing TP from PsP. Conventional MRI consistently struggles to resolve this with adequate specificity [6]. DSC-MRI offers a physiologically grounded alternative. By tracking GCA transit through brain vasculature, it quantifies cerebral perfusion, with relative cerebral blood volume (rCBV) serving as its principal biomarker [11]. Characteristically elevated in TP and reduced in PsP, rCBV provides a measurable basis for post-treatment differentiation that structural imaging alone cannot reliably offer [5].

The clinical stakes of misclassification are considerable. A false-positive TP diagnosis risks unnecessary escalation of therapy, while misidentifying TP as PsP may lead to premature withdrawal of a regimen that remains effective [4]. In either direction, the error carries direct consequences for patient survival. This diagnostic gap motivates the present study [6].

Radiomics features, derived from the DSC MRI along with the ML classification offers a more objective and reproducible framework for resolving this ambiguity [17]. The primary aim is therefore to extract and systematically analyze radiomic features from DSC-MRI datasets of GBM patients, and to develop ML models capable of discriminating TP from PsP with clinically meaningful reliability [18]. Model performance is evaluated across multiple metrics such as Validation Accuracy, Validation AUC, Test Accuracy, Test AUC, Test Precision, Test Sensitivity, Test Specificity, and Test F1 Score [19]. The study further emphasizes identifying which bio-markers contribute most to the discrimination. This is an interpretability focused objective that reflects a broader intent, which simply is to bring quantitative imaging evidence closer to the point of clinical decision making, where post-treatment ambiguity continues to affect care [20].

Research Hypothesis

The central hypothesis is that quantitative radiomic features derived from DSC-MRI when combined with clinical covariates such as MGMT methylation status will yield significantly greater discriminative performance than perfusion parameters alone [2]. It is further hypothesised that model interpretability analysis will reveal a compact subset of biologically meaningful biomarkers that reliably differentiate TP from PsP, and that these

features will reflect known physiological distinctions between active tumour and treatment-related change rather than emerging as statistical artefacts of the training data [17].

If supported, these findings would suggest that a radiomic-clinical integration framework is not only technically feasible but diagnostically justified with direct implications for post-operative GBM management and individualized treatment planning [19].

Significance of the Study

This work bridges advanced imaging science and clinical decision-making by establishing a reliable, data-driven framework for evaluating GBM progression [11]. By enhancing the ability to distinguish TP from PsP at an earlier and more accurate stage, the study aims to support clinicians in making better-informed therapeutic decisions and improving overall patient management.

Chapter 3

Literature Review

The diagnostic ambiguity between true progression and pseudoprogression in post-treatment GBM is not a recent discovery. It has been an unresolved clinical burden for years, and conventional MRI has consistently failed to resolve it with any meaningful specificity. Both entities produce contrast-enhancing lesions on follow-up scans. They look the same. That fundamental indistinguishability has made this one of the more stubborn problems in neuro-oncology, and it has motivated a growing body of work at the intersection of radiomics, pharmacokinetic modelling, and machine learning.

Early efforts leaned on structural MRI alone. Sun et al. [2021] demonstrated that radiomic features extracted from T1-weighted contrast-enhanced images, fed into a Random Forest classifier, could achieve an accuracy of 72.78% in distinguishing TP from PsP a figure that, notably, surpassed the performance of experienced radiologists on the same task [21]. The result is uncomfortable in the right way. It does not mean radiomics is superior to clinical expertise. It means that quantitative texture features capture something visual inspection routinely misses.

Bani-Sadr et al. [2019] extended this into a cohort of 76 GBM patients, combining MRI-based radiomics with MGMT promoter methylation status. Classification accuracy reached 83% [22]. Specificity, however, remained moderate - a limitation that surfaces repeatedly across the literature and carries real clinical weight. A false-positive TP call changes treatment. That cost is not symmetric.

Ari et al. [2022] approached the problem from a pre-treatment perspective entirely. In 131 high-grade glioma patients, radiomic features were extracted before treatment using semi-automatic segmentation in 3D Slicer, and a Generalized Boosted Regression Model yielded an accuracy of 84.35% with an AUC of 91.49% [23]. The clinical framing here is notably different - predicting pseudoprogression before it develops rather than diagnosing it retrospectively after imaging already shows contrast enhancement. Whether pre-treatment radiomic signatures can reliably anticipate a post-treatment inflammatory phenomenon is a biologically interesting question that the study raises without fully resolving.

Several studies turned specifically to perfusion imaging. Pak et al. [2021] extracted 642 radiomic features from DCE-MRI pharmacokinetic parameter maps K^{trans}, V_e, V_p and derived a radiomic risk score that independently stratified patient prognosis in multivariable Cox regression [24]. The feature count is striking. So is the implicit assumption buried inside it: that a single pharmacokinetic model adequately characterises every voxel across every patient. It likely does not. Voxel environments in GBM are biologically heterogeneous. Forcing a single parametric structure onto that heterogeneity degrades the estimated parameters, and by extension the radiomic features built from them. This is a methodological gap that few studies have explicitly acknowledged, let alone addressed.

Hosseini et al. [2024] combined DTI and DSC-PWI for a related classification task - separating GBM from solitary brain metastases - and reported an accuracy of 87.8% with an AUC of 92.67% [25]. The population and the question are different from TP-PsP discrimination. Still, the consistent benefit of combining diffusion and perfusion-based features across modalities is a finding that generalises. Neither modality alone matched the combined model. Complementarity in multiparametric imaging appears to be a genuine and reproducible phenomenon.

O'Neill et al. [2016] used DCE-MRI alongside 18F-FDG PET to evaluate pharmacodynamic response to VEGF Trap in temozolomide-resistant GBM patients. K^{trans}, V_e fell significantly following treatment [26]. The sample was small and long-term follow-up was absent. Generalizability is limited. But the sensitivity of pharmacokinetic parameters to treatment-induced vascular remodelling is clearly established, and that is the more durable point for the present study.

Kickingereder et al. [2016] operated at considerably larger scale. Across 119 GBM patients, over 12,000 radiomic features were extracted from multiparametric MRI and entered into a Cox proportional hazards framework for survival estimation. Radiomic predictors outperformed conventional clinical and radiological parameters [27]. The dimensionality makes the analysis difficult to interpret mechanistically. But the result reinforces a broader conclusion that the field has slowly converged on: medical images contain prognostic signal that existing staging systems do not extract.

Work outside neuro-oncology has been instructive for methodology if not for direct comparison. Liu et al. [2020] built separate LASSO-regularised logistic regression models - pharmacokinetic, radiomic, and combined - for predicting sentinel lymph node metastasis in breast cancer from DCE-MRI. The combined model consistently outperformed either alone, with AUC values of 0.80 and 0.76 in training and validation respectively [28]. Wang et al. [2022], working in rectal cancer neoadjuvant response prediction, systematically evaluated twenty classifier-feature selection combinations and found meaningful variation in performance across them - a finding that underlines how much pipeline choices matter, independent of the underlying data [29]. Peng et al. [2021] further demonstrated in breast cancer DCE-MRI that the timing of the imaging phase affects radiomic discriminative power substantially, with the third post-contrast phase consistently outperforming earlier acquisitions across multiple model configurations [30].

Aftab et al. [2022], in a broad review of radiomics and radiogenomics applications in GBM, identified dataset scale and acquisition heterogeneity as the primary barriers to clinical translation [31]. The technical pipeline is maturing. The data infrastructure around it has not kept pace. Radiomic features are sensitive to scanner parameters, reconstruction algorithms, and segmentation methodology in ways that remain only partially characterised, and most published cohorts are too small to support the kind of multisite validation that clinical deployment would require [17].

What the literature establishes, taken together, is a coherent but incomplete picture. Multiparametric approaches consistently outperform single-sequence analysis [11]. Molecular biomarkers like MGMT methylation status add discriminative value when incorporated [2]. Pharmacokinetic parameters from DCE-MRI carry genuine signal for the TP-PsP classification problem [13]. And the choice of pharmacokinetic model - a step most pipelines treat as fixed and trivial - shapes the quality of every feature derived downstream. That last point has received almost no direct investigation in the existing literature. The present study addresses it directly [16].

Chapter 4

Methodology

This chapter describes the methodology used in this study, from data collection to model assessment. The objective is to guarantee that the features extracted and machine-learning classifiers used, provide meaningful information to aid in making distinctions between TP and PsP in GBM patients [1, 4, 5, 6]. Due to the clinical importance of this problem, radiomics based feature extraction and clinically validated methods of classification were utilized to produced reliable results [17,20,21]. As shown in the Figure1, the visual abstract of pipeline clearly depicts an overview of analytical components. Such as, MRI scan image preprocessing (such as image coregistration), GBM tissue segmentation among Edema, Necrosis and Enhancing regions, fitting of a pharmacokinetic selection of the parsimonious model, radiomics feature extraction, dimension reduction and development of Machine Learning based models, and evaluation of diagnosis performance matrices.

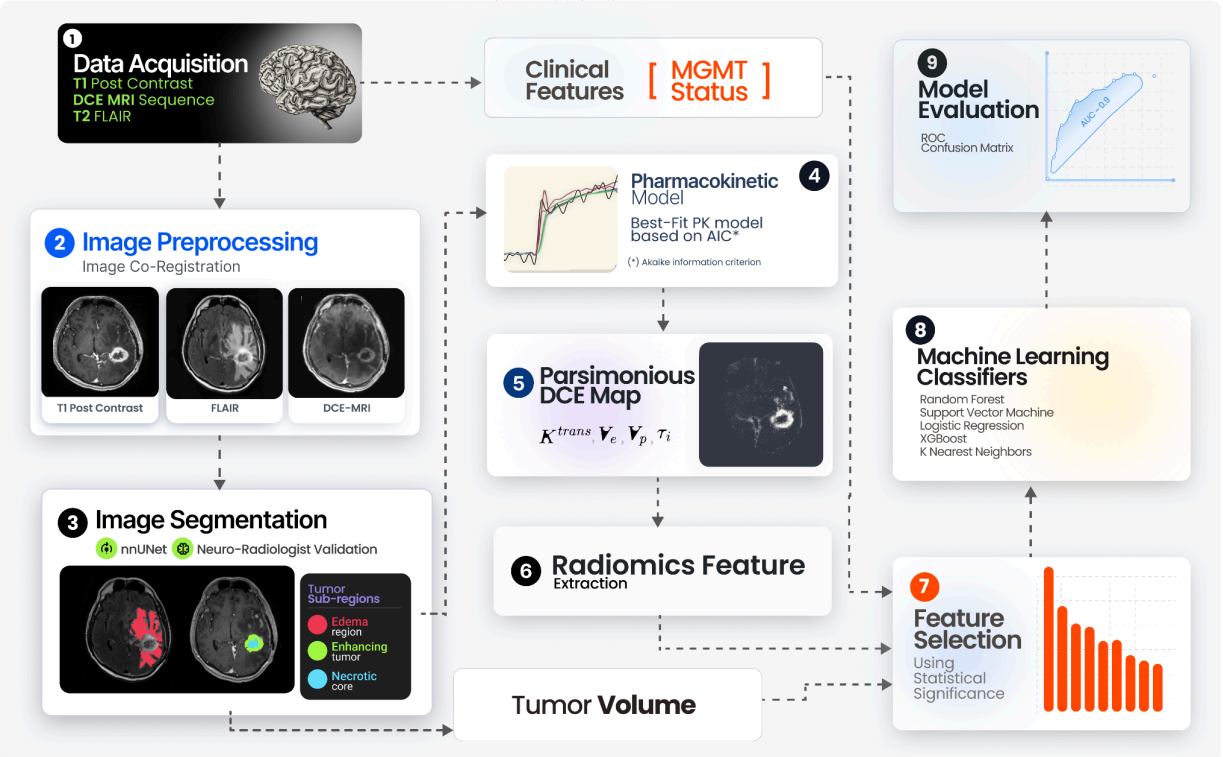


Fig 4: Overview of the proposed radiomics pipeline for MGMT status prediction in GBM using multi-parametric MRI.

Study Cohort and Clinical Data Acquisition

Patients were included if they met the following criteria :

(1) histologically confirmed diagnosis of GBM [1, 2] ;

(2) receipt of standard of care (SOC) therapy, including surgery followed by concurrent chemoradiotherapy (CCRT) [32];

(3) development of a new contrast-enhancing lesion within the radiation field on follow-up MRI within 6 months after completion of SOC therapy [4, 5, 6] ; and

(4) availability of both anatomical and DCE-MRI data [11, 12]. Based on these criteria, a total of 82 patients (49 males, 33 females; mean age 58.4 ± 8.9 years) were included.

Patient data collected at the University of Pennsylvania were shared with TCG CREST. The cohort (n = 82) was categorized into TP(n = 53) and PsP (n = 29). 52 patients with available tumor specimens from repeat surgery or biopsy, histopathological analysis identified TP (>25% malignant features; n = 31) and PsP (<25% malignant features; n = 21). For the remaining 30 patients without available tissue samples, modified Response Assessment in Neuro-Oncology (mRANO) criteria were used [33,34] , resulting in TP (n = 22) and PsP (n = 8). MGMT promoter status was available for all patients and was categorized as unmethylated (n = 41; TP = 27, PsP = 14), methylated (n = 23; TP = 12, PsP = 11), or indeterminate (n = 18; TP = 14, PsP = 4). MGMT status was included as a clinical variable to evaluate its association with progression status (TP vs. PsP) [35, 36]. Following Figure illustrates the patient selection process and subsequent TP/PsP classification.

Data Acquisition :

All patients underwent MRI scans on Siemens Healthineers, Erlangen(Germany), a classic 3T MAGNETOM Trio whole-body MR scanner with a 12 channel phased array head coil. The neuroimaging protocol comprised of T1-weighted magnetization-prepared rapid acquisition of gradient echo (T1-MPRAGE) and axial T2-fluid-attenuated inversion recovery (T2-FLAIR) images [9,10].

The T1-MPRAGE sequence parameters were as follows:

repetition time (TR) of 1820 ms, echo time (TE) of 3.88 ms, flip angle = 15°, slice thickness = 1 mm, bandwidth = 190 Hz/pixel, field of view (FOV) of 256 × 256 mm² with a 75% phase FOV. A total of 192 slices were acquired with one average (NEX).

The T2-FLAIR acquisition parameters were as follows:

Repetition time (TR) of 10,030 ms, Echo time (TE) of 137 ms, flip angle = 150°, slice thickness = 3 mm, bandwidth = 227 Hz/pixel, FOV of 240 × 240 mm², and a 75% FOV

phase. A total of 60 slices were acquired with one average. Gradient recalled spoiled GRE sequence were used for performing the DCE MRI scans [11, 12]. The sequence parameters consisted of the following setup

TR of 3.23 ms, a TE of 1.12 ms, with a flip angle of 23°, a slice thickness of 3.5 mm, FOV was 220 × 220 mm², an acquisition matrix of 256 × 256. A total of 28 slices were obtained with one average, and the pixel bandwidth was 210 Hz/pixel, acquisition time 3min and 33s. A total of 30 sequential measurements were acquired for each slice to assess contrast uptake and washout kinetics, with a temporal resolution of 7 seconds per frame. DCE-MRI data were obtained following intravenous administration of a GCA (MultiHance, Bracco Imaging, Milano, Italy, a widely used stain popular for its dual elimination route (renal and hepatobiliary)) at a dose of 0.07 mmol/kg [38, 39], delivered using a power injector (Medrad, Indianola, PA, USA).



Figure 5 : GBM Patient Inclusion Exclusion Criterion

Extraction of the Parameter Maps :

Signal intensity time curves were converted into contrast concentration time curves using fixed T1 values of 1400 ms for tissue and 1600 ms for blood. The longitudinal relaxivity of the GCA was $4.3\ L.mmol^{-1}s^{-1}$ [38, 39].

DCE-MRI parametric maps, including K_{trans}, V_e, V_p, K_{ep}, F_p, τ_i were generated using an in-house MATLAB toolbox implementing multiple pharmacokinetic (PK) models, including the non-linear Tofts model, extended Tofts model (ETM), shutter-speed model (SSM), two-compartment exchange model (2CXM), and three-site two-exchange model (3S2X) [12, 13, 42, 43, 44].

A vascular input function (VIF) was manually defined for each subject by placing a region of interest over the superior sagittal sinus (SSS). After generating the parametric maps, post-contrast T1 (PC-T1) and T2-FLAIR images were co-registered and resliced to match the DCE-MRI parametric maps using rigid-body affine transformation implemented in the SimpleITK (version 2.0.2) package [46, 47].

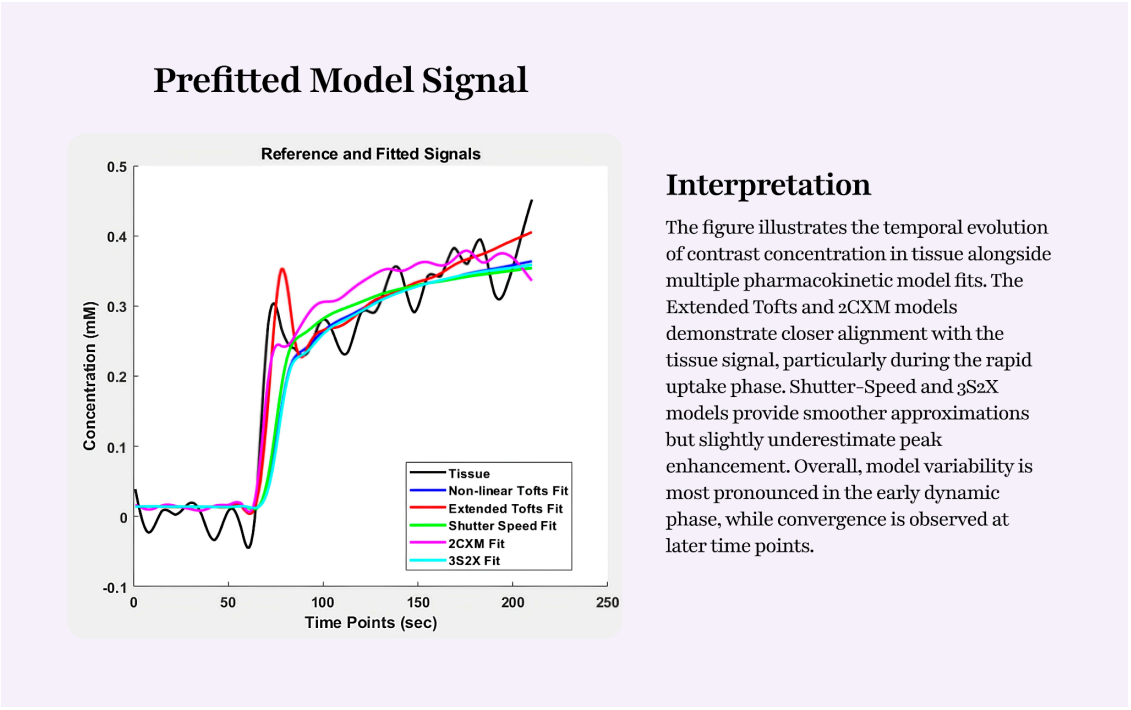


Figure 6: Graph showing the Pre - fitted model signals vs the actual signal

Calculating of Tumor Volumes

Tumors were automatically segmented into contrast-enhancing, non-enhancing/necrotic, and edematous regions using the nnU-Net algorithm [48], based on co-registered post-contrast T1 (PC-T1) and T2-FLAIR images. The model was trained and tested on the BraTS (2018-2021) [48, 50, 51] dataset and achieved a Dice score of 88.95%.

For each patient, total tumor volume was calculated by combining the enhancing and non-enhancing regions across all slices and multiplying the number of voxels in these regions by the voxel size of the co-registered images [48, 49].

Developing the Parsimonious Model

In DCE-MRI, selecting an appropriate PK model is important for characterizing tissue perfusion and vascular permeability [11, 12]. A parsimonious PK model selected based on the AIC value [16], balances model complexity and goodness of fit. DCE-MRI data from each segmented region were was fitted to five PK models: non-linear Tofts, ETM, SSM, 2CXM, and 3S2X [13, 42, 43, 44] to generate the best-fit K^{trans} and V_e values as these are the two parameters that are common across all PK models. A parsimonious model was reconstructed by aggregating best-fit models selected for each voxel, as shown in Figure 2.11. Model specific parameters were also generated: V_p from the ETM, 2CXM, and 3S2X models [15, 43]; and τ_i from the SSM and 3S2X models. Separate parsimonious model selection was performed for these parameters, where V_p was selected as the best-fit value from ETM, 2CXM, and 3S2X models, and τ_i was selected as the best-fit value from the SSM and 3S2X models, based on voxel-wise minimum AIC [16]. Finally, mean AIC and R^2 values were calculated for each model.

Akaike Information Criterion (AIC) was defined as [16]:

$$AIC = N.\ln ln\left(\sum_{i=1}^{N}\left(y_i - C_t(i)\right)^2\right) + \frac{2.p.(p+1)}{N-p-1}$$

Where:

N - number of observations, y_i - observation at time i, $C_t(i)$ - associated fitted value at time i, and p - number of parameters in the model.

The full pharmacokinetic model formulations, ODE derivations, and explicit parameter extraction procedures for all five models are provided in Supplementary Material S1.

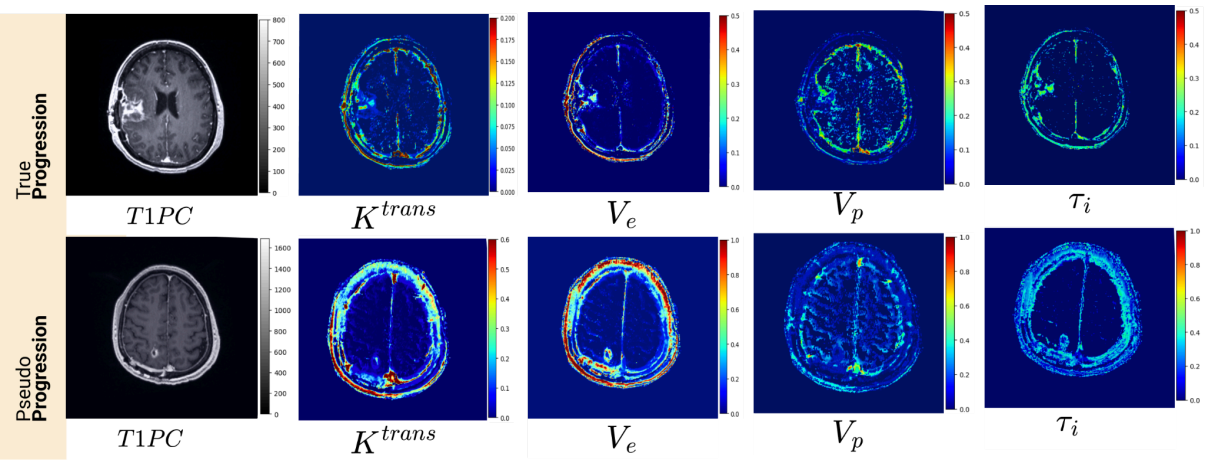


Figure 7 : Pharmacokinetic parameter maps derived from DCE MRI reveal hemodynamic signatures between TP and PsP, including $T1PC$, K^{trans}, V_e, V_p, τ_i.

Radiomics Feature Extraction

Radiomics extracts high-dimensional quantitative features from regions of interest in medical images, including intensity, shape, texture, and spatial heterogeneity [17, 20], a method in differentiating TP from PsP (Ying-Zhi Sun, 2021) [21, 22, 23]. In this study, the tumor region served as the Region of Interest (ROI), subdivided into three anatomically distinct subregions: the enhancing region, non-enhancing core, and peritumoral edema. Radiomic features were extracted from the enhancing region using PyRadiomics (v3.1.0) [52], a validated open-source platform.

The tumour region served as the primary region of interest (ROI), delineated across three anatomically distinct subregions: the contrast-enhancing region, the non-enhancing tumour core, and the peritumoral oedema. This tripartite subdivision reflects the spatial heterogeneity inherent to GBM and captures distinct biological microenvironments - active angiogenesis, necrotic or infiltrative tissue, and oedematous parenchyma - each potentially contributing differential signal to TP-PsP discrimination. Feature extraction was performed exclusively from the enhancing region using PyRadiomics (v3.1.0) [52], a validated open-source platform with documented reproducibility across scanner protocols and segmentation pipelines.

Extracted features spanned four primary categories, each encoding a distinct aspect of tumour phenotype :

First-order features These features characterise the statistical distribution of voxel intensities within the ROI, without regard to spatial relationships between voxels. Formally, for an image with N voxels and intensity values $\{X_i\}\ \forall\ i\ \in\ [1, N]$,the core statistics are :

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \qquad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

$$\gamma_1 = \frac{1}{N\sigma^3}\sum_{i=1}^{N}(x_i - \mu)^3 \qquad \gamma_2 = \frac{1}{N\sigma^4}\sum_{i=1}^{N}(x_i - \mu)^4 - 3$$

Mean and variance capture the overall intensity level and spread. Skewness $\gamma_1$and excess kurtosis γ_2 reflect asymmetry and tail behaviour - subtle distributional shifts that may correspond to heterogeneity in perfusion or contrast retention.

Shape-based features describe tumour geometry independently of intensity. Volume V surface area A, and sphericity Ψ are among the primary descriptors:

$$\Psi\ =\ \frac{\pi^{1/3}(6V)2/3}{A}\ ;\ \Psi\ \in\ (0, 1]$$

A value of unity denotes a perfect sphere. Recurrent tumour tends toward irregular, infiltrative margins. Treatment related change is more often rounded and well demarcated. Shape features attempt to capture precisely this distinction.

Texture features encode spatial relationships between voxels patterns that first-order statistics cannot see. The Gray Level Co-occurrence Matrix (GLCM) $P(i, j \mid \delta, \theta)$ records the joint probability of two voxels, separated by displacement δ orientation θ, having intensity levels i and j. From this matrix, features such as contrast, correlation, energy, and homogeneity are derived. The Gray Level Run Length Matrix (GLRLM) captures consecutive voxel runs of uniform intensity along a direction, with run-length non-uniformity (RLN) and long-run emphasis (LRE) quantifying textural coarseness. The Gray Level Size Zone Matrix (GLSZM) extends this to connected three dimensional zones. The Neighboring Gray Tone Difference Matrix (NGTDM) measures local intensity contrast between a voxel and its spatial neighbourhood. The Gray Level Dependence Matrix (GLDM) quantifies how many neighbouring voxels share intensity dependence with a central voxel, a measure of local consistency [52].

Together, these matrices characterise texture at multiple spatial scales and orientations. Features were extracted across all pharmacokinetic parametric maps from DCE-MRI, spanning both enhancing and non-enhancing subregions. The initial extraction yielded **1,073** features.

Feature Selection

Feature selection was performed using a two-stage reduction pipeline. The dimensionality of the data was high relative to sample size- without reduction, overfitting is essentially guaranteed.

In the first stage, the Mann-Whitney U test [53] assessed each feature independently for discriminatory power between TP and PsP. For samples $X = \{x_1, x_2, \ldots, x_m\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$

$$U = \sum_{i=1}^{m} \sum_{j=1}^{n} S(x_i, y_j) \qquad S(x, y) = \begin{cases} 1 & x > y \\ \frac{1}{2} & x = y \\ 0 & x < y \end{cases}$$

Features with $p < 0.05$ were retained. Radiomic distributions routinely violate normality, making this non-parametric test more appropriate than its parametric alternatives [53]. The set reduced from $1,073\ to$ 86 features.

The second stage applied Elastic Net regression, a regularised linear model that combines $L1$ & $L2$ penalties.

$$\hat{\beta} = \arg\min_{\beta} \left[\frac{1}{2n} \|y - X\beta\|_2^2 + \alpha \left(\rho \|\beta\|_1 + \frac{1-\rho}{2} \|\beta\|_2^2 \right) \right]$$

Here α controls overall regularisation strength and $\rho \in [0, 1]$ balances sparsity against grouping. Elastic Net tends to retain or remove them together, a meaningful distinction in radiomic data, where feature correlations are pervasive. Hyperparameters α and ρ were optimised via grid search with five-fold stratified cross-validation, preserving class balance across folds. All feature selection procedures were implemented in SciPy (v1.17.1) and scikit-learn (v1.8.0).

Class Imbalance and Cross-Validation Strategy

Clinical dataset of GBM is inherently prone to class imbalance. In post SOC GBM cohorts, PsP cases are usually less frequent than TP. Which produces a skewed label distributions. Conventional ML classifiers fail to handle class imbalances data by default. Under such conditions, a naive model achieves deceptively high accuracy by defaulting to the majority class, while offering no genuine discriminative utility. To overcome this, Adaptive Synthetic Sampling (ADASYN) was applied to the training partition to correct for this skew. Unlike

uniform oversampling, ADASYN concentrates synthetic generation near the decision boundary where *minority* samples are most surrounded by *majority* class neighbours. Specifically in the zone, where classification is hardest. For each minority sample i, the augmentation density is governed by:

$$\hat{d} = \frac{\Delta_i}{K}, i \in 1, 2, \ldots, m_s$$

where Δ_i is the number of majority class instances among the K nearest neighbours of sample i. Higher $\hat{d}_i$ implies denser synthetic augmentation. In radiomic features, where TP-PsP distributions substantially overlap, this boundary aware augmentation is more appropriate than uniform minority replication like SMOTE [55]. ***Resampling was applied strictly within the training partition***. The test set remained unmodified throughout the ML pipeline.

Model Training and Hyperparameter Optimisation

Each classifier was trained within a un-stratified five-fold cross-validation framework. Class proportions were preserved across all folds, and each fold served once as a validation set. For every hyperparameter configuration, mean validation AUC was computed :

$$\overline{\mathrm{AUC}}_{\mathrm{val}} = \frac{1}{5}\sum_{k=1}^{5} \mathrm{AUC}^{(k)}$$

The configuration maximising Validation AUC was selected. Validation accuracy was recorded. AUC evaluates discriminative performance across all possible thresholds, a more informative criterion than accuracy when class distributions are unequal.

Classification Threshold Selection

Following hyperparameter selection, an out-of-fold pass under the optimal configuration yielded probability estimates across the full training set. The classification threshold was swept over [0.2, 0.6] in increments of 0.001, selecting the value maximising out-of-fold F1:

$$\hat{t} = arg\,max_t \frac{2\,Precision_t\ Sensitivityy_t}{Percision_t + Sensitivityy_t}$$

A fixed threshold of 0.5 assumes equal class prevalence and equal misclassification cost. Tuning the threshold on OOF predictions, rather than on the test data set, ensures that no information leakage happened during final evaluation.

The final model was trained on the complete resampled training set under the optimal hyperparameter configuration. All experiments were repeated across several random, independent random seeds. Each of the seeds governing the complete pipeline, i.e train-test

split, resampling, fold assignment, and model initialisation. Performance metrics were calculated across runs and reported as $mean\ (\mu)\ \pm\ sd\ (\sigma)$.

Machine Learning Models

Several machine learning models were implemented to compare their effectiveness [19, 56, 57, 58]

SVM [56] : Finds the optimal hyperplane that maximizes the margin between different classes. It is particularly effective in high-dimensional spaces and robust to overfitting with proper regularization.

RF [57]: An ensemble learning method that constructs multiple decision trees and takes a majority vote to improve accuracy and reduce overfitting. It provides feature importance rankings, aiding interpretability.

LR : A simple linear model used for binary classification. It calculates the probability that an input belongs to a particular class using the sigmoid function. 20

XGBoost [58] : A highly effective gradient boosting framework aimed at minimizing bias and variance to enhance the performance of the model. It achieves this through iterative boosting trees to reduce prediction error.

Evaluation Metrics

Performance was assessed at two distinct stages: across validation folds during training, and on the held-out test set following final model fitting. During CV, validation AUC and validation accuracy were recorded. On test data set, multiple metrics was derived from the confusion matrix, using optimal threshold $\hat{t}$.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP} \qquad \text{Sensitivity} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \qquad \text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}$$

Chapter 5

Result and Analysis

Four distinct experimental configurations were systematically evaluated, each representing an increasing level of integration between pharmacokinetic complexity and molecular biomarker information. The progression from ETM-derived features with MGMT, to T1 post-contrast radiomic features with MGMT, to parsimonious DCE parameters without MGMT, and finally to parsimonious DCE parameters with MGMT, provides a structured basis for disentangling the independent and synergistic contributions of imaging and molecular features to classification performance.

Extended Tofts Model [ETM] With MGMT

Method	TA	VA	VAUC	TEAUC	TE	Precision	Sensitivity	Specificity	F1
KNN	0.93 ± 0.12	0.69 ± 0.04	0.73 ± 0.05	0.63 ± 0.07	0.62 ± 0.10	0.67 ± 0.13	0.62 ± 0.16	0.60 ± 0.17	0.64 ± 0.13
LR	0.72 ± 0.05	0.67 ± 0.05	0.72 ± 0.04	0.62 ± 0.13	0.58 ± 0.08	0.64 ± 0.10	0.68 ± 0.17	0.43 ± 0.28	0.64 ± 0.08
RF	0.96 ± 0.05	0.63 ± 0.07	0.70 ± 0.08	0.68 ± 0.13	0.64 ± 0.14	0.63 ± 0.13	0.82 ± 0.19	0.38 ± 0.12	0.70 ± 0.13
SVM	0.72 ± 0.15	0.67 ± 0.07	0.70 ± 0.06	0.62 ± 0.08	0.66 ± 0.17	0.69 ± 0.16	0.74 ± 0.24	0.45 ± 0.32	0.67 ± 0.12
XGB	0.88 ± 0.15	0.62 ± 0.07	0.71 ± 0.07	0.61 ± 0.15	0.58 ± 0.13	0.59 ± 0.12	0.71 ± 0.17	0.35 ± 0.16	0.64 ± 0.13

Table 1 note: No classifier in the ETM + MGMT configuration achieves a TEAUC above 0.68, indicating that single-model pharmacokinetic estimation significantly constrains the discriminative capacity of radiomic features. High training accuracies (RF: 0.96; XGB: 0.88) contrasted against low validation metrics confirm overfitting, particularly in RF and XGBoost. The wide standard deviations in specificity (SVM: ±0.32; LR: ±0.28) indicate extreme instability, clinically unacceptable for deployment.

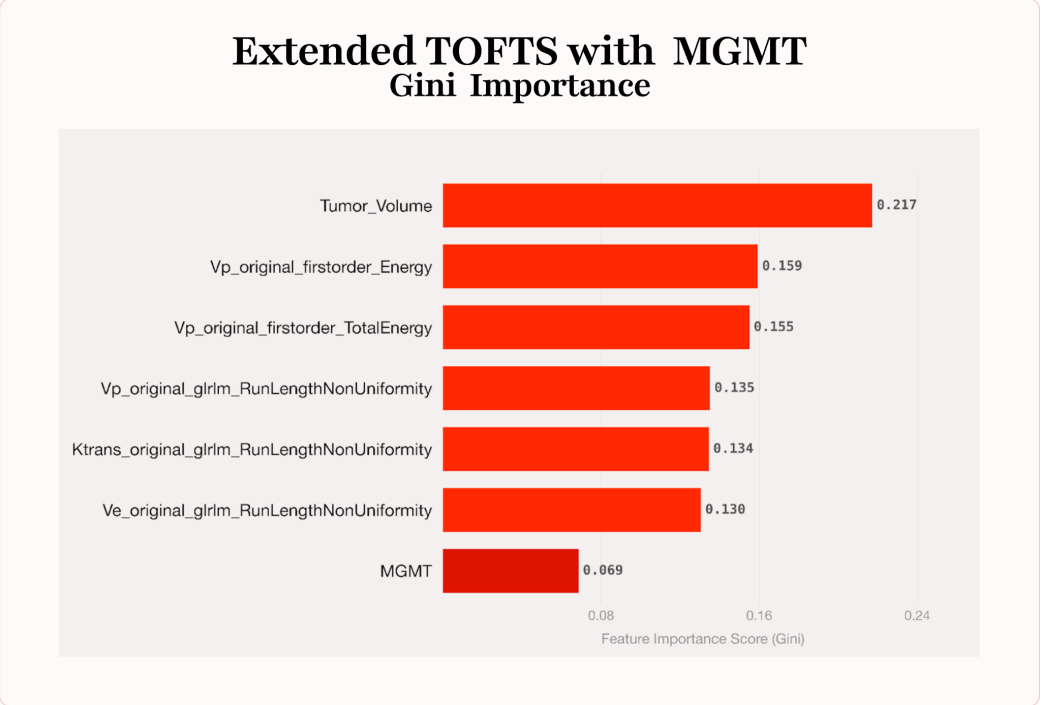


Figure 8: Gini Feature Importance Plot for Extended TOFTS + MGMT

The ETM + MGMT configuration produced the lowest overall performance across all classifiers, with validation AUC values ranging from 0.70 to 0.73 and test AUC (TE AUC) values between 0.61 and 0.68. The KNN classifier achieved the highest TEAUC (0.63 ± 0.07) and specificity (0.60 ± 0.17), while RF demonstrated the highest sensitivity (0.82 ± 0.19) but markedly reduced specificity (0.38 ± 0.12), suggesting pronounced overfitting to the majority class. The wide standard deviations observed across most metrics - particularly specificity (RF: 0.38 ± 0.12; SVM: 0.45 ± 0.32) indicate substantial performance instability that likely reflects the limitations of single-model pharmacokinetic parameter estimation, wherein the ETM may be misspecified for voxel environments better represented by the SSM or 2CXM, resulting in degraded biomarker fidelity. These findings indicate that constraining pharmacokinetic analysis to the ETM, despite its widespread clinical adoption, introduces systematic error that undermines radiomic feature quality and classification consistency.

T1 Post Contrast Model [ETM] With MGMT

Method	TA	VA	VAUC	TEAUC	TE	Precision	Sensitivity	Specificity	F1
KNN	1.00 ± 0.00	0.71 ± 0.03	0.81 ± 0.03	0.75 ± 0.06	0.67 ± 0.06	0.74 ± 0.11	0.76 ± 0.16	0.56 ± 0.23	0.73 ± 0.07
LR	0.75 ± 0.04	0.69 ± 0.04	0.76 ± 0.04	0.75 ± 0.07	0.67 ± 0.05	0.77 ± 0.13	0.71 ± 0.11	0.62 ± 0.25	0.72 ± 0.04
RF	1.00 ± 0.00	0.75 ± 0.04	0.83 ± 0.03	0.72 ± 0.05	0.67 ± 0.05	0.73 ± 0.11	0.79 ± 0.17	0.51 ± 0.23	0.74 ± 0.06
SVM	0.97 ± 0.03	0.78 ± 0.06	0.81 ± 0.06	0.72 ± 0.13	0.69 ± 0.09	0.78 ± 0.11	0.72 ± 0.17	0.66 ± 0.20	0.74 ± 0.08
XGB	0.99 ± 0.04	0.73 ± 0.04	0.82 ± 0.04	0.72 ± 0.09	0.67 ± 0.07	0.74 ± 0.08	0.73 ± 0.11	0.58 ± 0.15	0.73 ± 0.07

Table 2 note: SVM (highlighted) achieves the best balance of VA (0.78), VAUC (0.81), and specificity (0.66), making it the recommended classifier within this configuration. RF achieves the highest VAUC (0.83) but with poor specificity (0.51). LR demonstrates unexpected competitiveness (TEAUC: 0.75; specificity: 0.62), suggesting partial linear separability of T1 post-contrast radiomic features+MGMT.

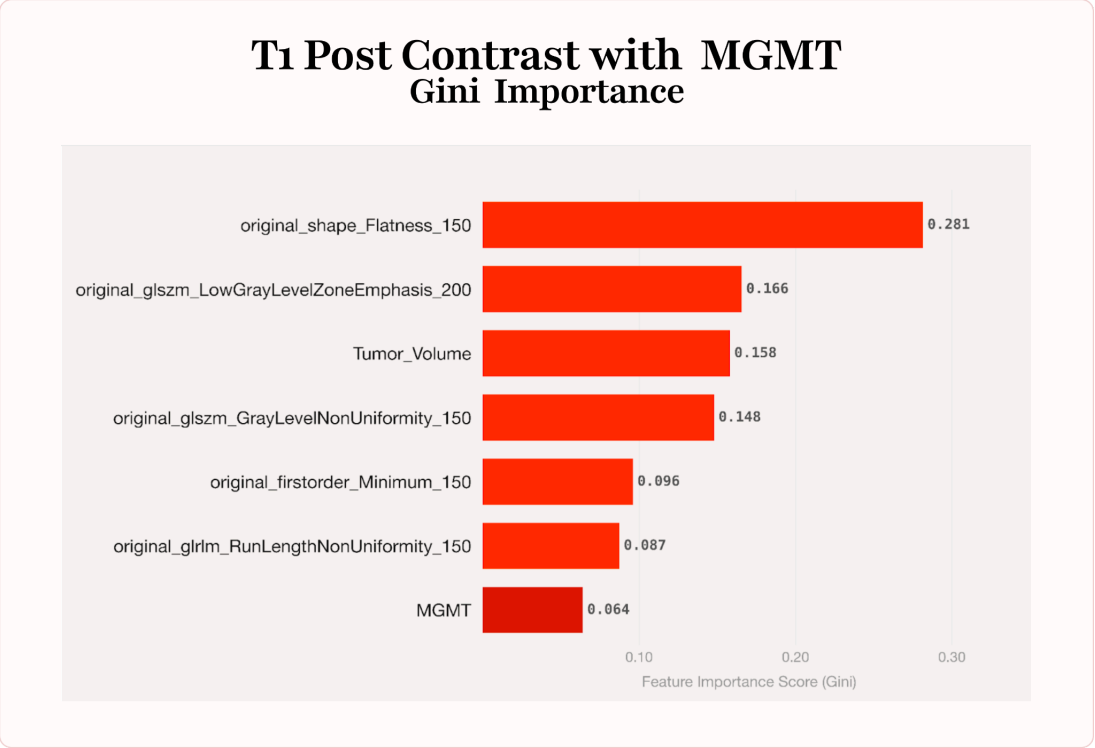


Figure 9: Gini Feature Importance Plot for T1 PosT Contrast + MGMT

The T1 Post Contrast + MGMT configuration yielded a substantially improved performance profile compared to ETM alone, with SVM emerging as the top-performing classifier (VA: 0.78 ± 0.06; VAUC: 0.81 ± 0.06; TEAUC: 0.72 ± 0.13; sensitivity: 0.72 ± 0.17; specificity: 0.66 ± 0.20; F1: 0.74 ± 0.08). RF achieved near-perfect training accuracy (1.00 ± 0.00) with a VAUC of 0.83 ± 0.03, the highest in this configuration, though its specificity remained

suboptimal (0.51 ± 0.23), indicative of a sensitivity-biased decision boundary. KNN and XGBoost demonstrated comparable VAUC (0.81 and 0.82, respectively) and F1 scores (0.73), suggesting relative equivalence among these models in this feature space. Notably, LR achieved competitive performance (VAUC: 0.76; TEAUC: 0.75; specificity: 0.62 ± 0.25) while operating as a linear model, which implies that T1 post-contrast radiomic features with MGMT exhibit sufficient linear separability to support high-performing classification without requiring complex non-linear decision boundaries. This is clinically significant, as T1 post-contrast imaging is universally acquired in GBM follow-up protocols and its radiomic analysis does not necessitate the additional pharmacokinetic modelling overhead of DCE-MRI, potentially enabling broader clinical deployment.

Parsimonious DCE Parameters (Without MGMT)

Model	TA	VA	TEAUC	TE	Precision	Sensitivity	Specificity	F1
RF	1.00 ± 0.00	0.75 ± 0.03	0.84 ± 0.09	0.79 ± 0.09	0.82 ± 0.12	0.91 ± 0.13	0.57 ± 0.31	0.85 ± 0.07
KNN	1.00 ± 0.00	0.72 ± 0.04	0.80 ± 0.09	0.66 ± 0.10	0.74 ± 0.11	0.75 ± 0.14	0.48 ± 0.23	0.74 ± 0.10
LR	0.74 ± 0.04	0.69 ± 0.05	0.74 ± 0.06	0.72 ± 0.05	0.80 ± 0.08	0.78 ± 0.11	0.62 ± 0.14	0.78 ± 0.06
SVM	0.88 ± 0.09	0.77 ± 0.04	0.76 ± 0.12	0.80 ± 0.10	0.75 ± 0.17	0.62 ± 0.17	0.77 ± 0.12	0.73 ± 0.06
XGB	0.99 ± 0.02	0.71 ± 0.03	0.81 ± 0.09	0.74 ± 0.10	0.83 ± 0.12	0.81 ± 0.17	0.63 ± 0.26	0.80 ± 0.10

Table 3 note: RF (highlighted) achieves the best TEAUC (0.84 ± 0.09) and sensitivity (0.91 ± 0.13) in this configuration, confirming that parsimonious model selection substantially improves DCE-derived feature quality. However, RF specificity (0.57 ± 0.31) remains suboptimal and highly variable. SVM provides the best specificity (0.77 ± 0.12) with a lower but acceptable TEAUC (0.76 ± 0.12).

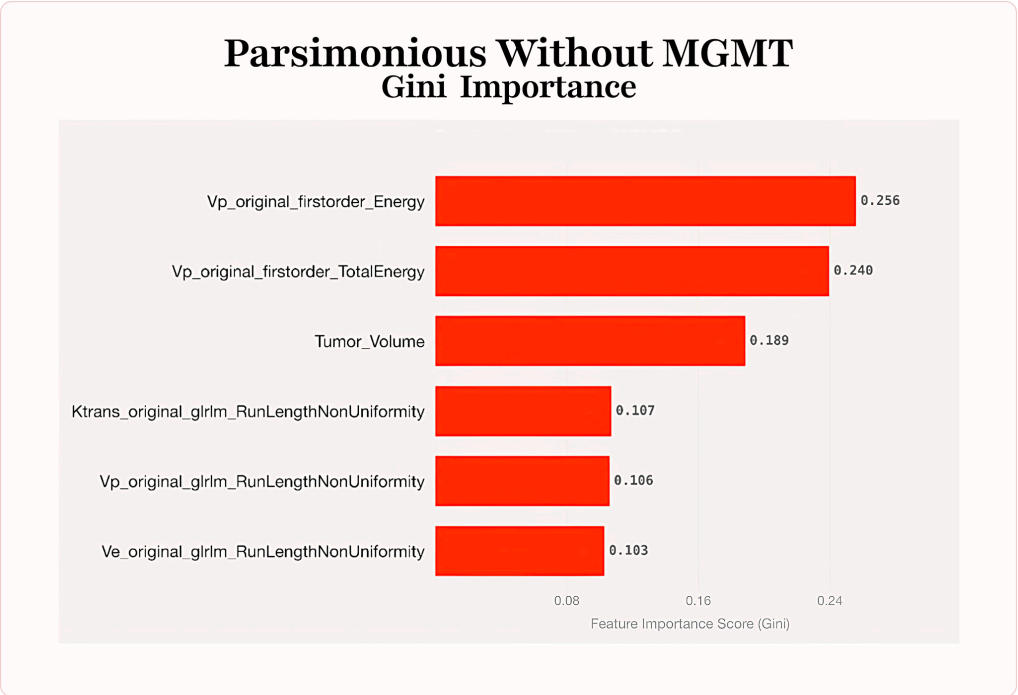


Figure 10: Gini Feature Importance Plot for Parsimonious DCE + MGMT

The most clinically remarkable finding of this configuration is the dramatic improvement in TEAUC observed for the RF classifier (0.84 ± 0.09) compared to all ETM-based results, despite the exclusion of MGMT status. This finding provides compelling evidence that the parsimonious pharmacokinetic modelling framework, by selecting the most biologically appropriate model for each voxel rather than imposing a single parametric structure - extracts substantially more informative and reliable DCE-derived features. The RF sensitivity of 0.91 ± 0.13 represents the highest across all models in this configuration and approaches the sensitivity requirements for clinically useful screening tools for TP. However, the specificity

of 0.57 ± 0.31 remains problematic; the wide confidence interval indicates high across-patient variability, and a specificity below 0.65 may result in an unacceptably high false-positive rate in populations with elevated PsP prevalence (e.g., MGMT-methylated patients receiving full-dose temozolomide). LR showed the most balanced specificity (0.62 ± 0.14) with a TEAUC of 0.74 ± 0.06, suggesting stability under regularisation. SVM demonstrated the highest specificity (0.77 ± 0.12) and test accuracy (TE: 0.80 ± 0.10) among non-RF classifiers, reflecting its aptitude for high-dimensional feature spaces through margin maximisation, but its sensitivity (0.62 ± 0.17) was the lowest in this configuration, limiting clinical utility for identifying TP cases.

Parsimonious DCE [With MGMT Status]

Models	TA	VA	VAUC	TEAUC	TE	Precision	Sensitivity	Specificity	F1
RF	1.0 ± 0	0.74 ± 0.03	0.82 ± 0.03	0.89 ± 0.07	0.86 ± 0.1	0.88 ± 0.1	0.93 ± 0.09	0.76 ± 0.19	0.9 ± 0.07
KNN	1.00 ± 0.00	0.70 ± 0.05	0.78 ± 0.04	0.73 ± 0.11	0.67 ± 0.11	0.74 ± 0.09	0.77 ± 0.17	0.46 ± 0.17	0.75 ± 0.1
LR	0.73 ± 0.04	0.67 ± 0.05	0.72 ± 0.05	0.72 ± 0.05	0.67 ± 0.08	0.76 ± 0.1	0.76 ± 0.13	0.51 ± 0.3	0.76 ± 0.06
SVM	0.89 ± 0.12	0.73 ± 0.07	0.75 ± 0.05	0.71 ± 0.14	0.64 ± 0.12	0.73 ± 0.11	0.73 ± 0.18	0.45 ± 0.22	0.72 ± 0.11
XGB	0.98 ± 0.04	0.69 ± 0.07	0.78 ± 0.04	0.79 ± 0.09	0.72 ± 0.11	0.79 ± 0.12	0.82 ± 0.16	0.56 ± 0.27	0.79 ± 0.08

Table 4 note: RF (highlighted in green) delivers the best overall performance across all metrics in the entire study: TEAUC = 0.89 ± 0.07, Sensitivity = 0.93 ± 0.09, Specificity = 0.76 ± 0.19, F1 = 0.90 ± 0.07. This configuration is the recommended deployment model. Notably, TEAUC improves by 5 percentage points (0.84 → 0.89) and specificity improves by 19 percentage points (0.57 → 0.76) compared to the same RF in the MGMT-free configuration, demonstrating the critical additive contribution of molecular biomarker integration.

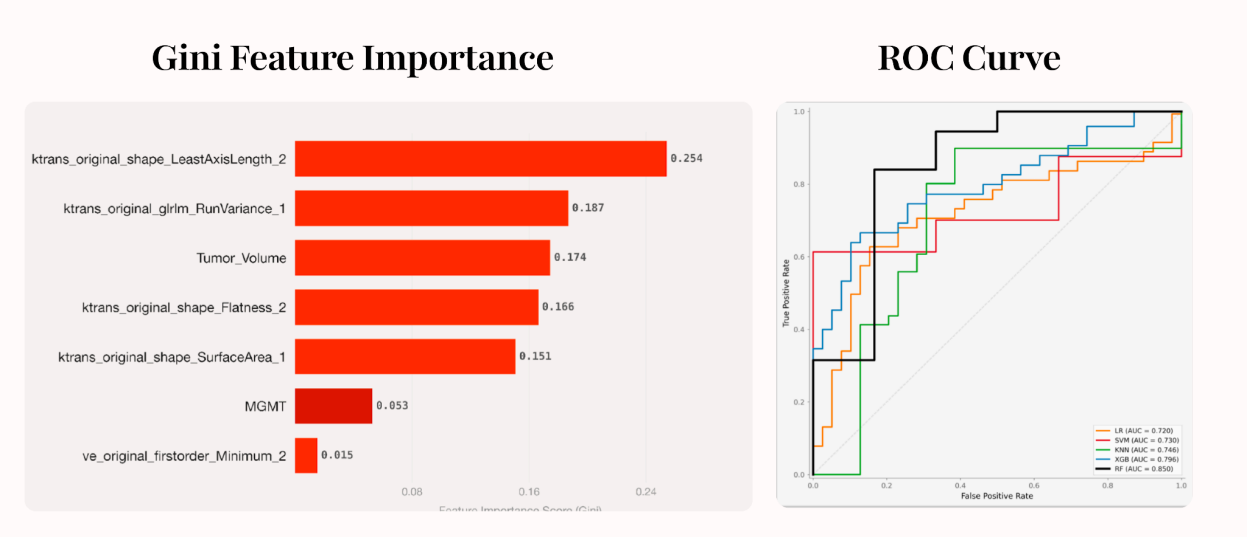


Figure 11: Gini Feature Importance Plot for Parsimonious DCE + MGMT

The integration of MGMT promoter methylation status with parsimonious DCE-derived radiomic features produced the highest performance across all experimental conditions and classifiers. The RF model in this configuration achieved a TEAUC of 0.89 ± 0.07, sensitivity of 0.93 ± 0.09, specificity of 0.76 ± 0.19, precision of 0.88 ± 0.10, and F1-score of 0.90 ± 0.07 metrics that are collectively among the strongest reported for TP/PsP discrimination in the existing radiomics literature. The convergence of high sensitivity, high specificity, and high F1 in the RF model represents a ***clinically meaningful and statistically robust finding***. The improvement in specificity from 0.57 ± 0.31 (parsimonious DCE without MGMT, RF) to 0.76 ± 0.19 (parsimonious DCE with MGMT, RF) is particularly noteworthy, suggesting that MGMT status functions as a biologically informative prior that helps stratify the PsP population and reduce false-positive classification errors.

Chapter 6

Discussion & Conclusion

Across all classifiers in this configuration, TEAUC values ranged from 0.71 (SVM) to 0.89 (RF), with XGBoost performing second (TEAUC: 0.79 ± 0.09; F1: 0.79 ± 0.08) and KNN third (TEAUC: 0.73 ± 0.11; F1: 0.75 ± 0.10). The performance gap between RF and other classifiers suggests that the complex, non-linear, high-order interaction structure of DCE-radiomic features combined with MGMT status is optimally captured by ensemble tree-based methods rather than linear models or distance-based classifiers. The achieved performance of the Parsimonious DCE + MGMT RF model (TEAUC: 0.89; sensitivity: 0.93; specificity: 0.76; F1: 0.90) compares favourably with several key studies in the TP/PsP discrimination literature. Bani-Sadr et al. (2019) reported a diagnostic accuracy of 83% in a cohort of 76 GBM patients using MRI radiomics with MGMT integration, though specificity remained moderate - a limitation similarly observed in lower-performing configurations of the present study. Sun et al. (2021) achieved 72.78% accuracy using T1CE-derived radiomic scores with an RF classifier, a performance level roughly concordant with the T1 post-contrast + MGMT configuration of the present study (VA: 0.75–0.78), thereby providing an important benchmark for validating the incremental value of parsimonious pharmacokinetic modelling.

Across all four experimental configurations and five classifiers evaluated in this study, the Random Forest classifier operating on parsimonious DCE-derived radiomic features supplemented by MGMT promoter methylation status constitutes the unambiguous optimal choice for clinical deployment, based on the following evidence hierarchy :

(i) Highest TEAUC (0.89 ± 0.07): The test ensemble AUC is the most robust performance indicator in this framework, as it evaluates generalisation performance on held-out test data using ensemble voting across multiple cross-validation folds, thereby reducing the variance

of a single train-test split. An AUC of 0.89 indicates excellent discriminative capacity, placing the model well within the clinically useful range (AUC ≥ 0.85) recommended by the TRIPOD reporting framework for clinical prediction tools.

(ii) Highest sensitivity (0.93 ± 0.09): In the clinical context of TP versus PsP discrimination, sensitivity (i.e., the ability to correctly identify TP) is arguably the most critical performance dimension, as failure to detect TP carries the highest clinical harm (continued ineffective treatment, delayed salvage therapy, and accelerated neurological decline). A sensitivity of 0.93 with a relatively narrow confidence interval (±0.09) indicates reliable TP detection across the cross-validation resampling procedure.

(iii) Highest specificity among high-sensitivity models (0.76 ± 0.19): While SVM without MGMT achieves higher specificity in isolation (0.77 ± 0.12), it does so at the cost of substantially reduced sensitivity (0.62 ± 0.17) and TEAUC (0.76 ± 0.12). The RF in the optimal configuration achieves both high sensitivity and high specificity simultaneously, a balance that is far more clinically valuable than optimising either metric in isolation.

(iv) Highest F1-score (0.90 ± 0.07): The F1-score harmonises precision and sensitivity, providing a balanced measure of classifier efficacy in the presence of class imbalance. An F1 of 0.90 is exceptional for a binary medical classification task with n = 82 subjects and is the highest across all 20 classifier-configuration combinations evaluated.

The Random Forest algorithm is particularly well-suited to the feature structure of this problem for several reasons. First, it is inherently robust to overfitting through bagging (bootstrap aggregation), which introduces diversity across base learners while maintaining low bias. The perfect training accuracy (1.00 ± 0.00) observed for RF across multiple configurations reflects the capacity of individual decision trees to memorise training data; however, the ensemble mechanism ensures that generalisation on held-out data is driven by consensus rather than individual tree idiosyncrasies. Second, RF handles high-dimensional input spaces efficiently without requiring a priori feature normalisation or distributional assumptions, making it appropriate for the heterogeneous scale and distribution profiles of radiomic features spanning first-order statistics, shape descriptors, and texture matrices. Third, RF provides built-in feature importance estimates (mean decrease in impurity or permutation importance), which enable post-hoc identification of the most discriminative pharmacokinetic and radiomic features - a critical capability for biological interpretation and future biomarker validation. Fourth, RF is computationally efficient, interpretable at the ensemble level through decision tree visualisation, and does not require hyperparameter

tuning as extensive as SVM kernel selection or XGBoost learning rate and tree depth optimisation, reducing the risk of data-adaptive overfitting in hyperparameter search.

XGBoost achieved a TEAUC of 0.79 ± 0.09 and F1 of 0.79 ± 0.08 in the optimal configuration - competitive but consistently below RF across all key metrics. The gradient boosting framework of XGBoost is more susceptible to overfitting in small-sample, high-dimensional settings when regularisation is not perfectly calibrated. While XGBoost demonstrated strong training accuracy (0.98 ± 0.04) and competitive VAUC (0.78 ± 0.04), the gap between training and validation performance suggests greater sensitivity to sample-specific patterns. For a dataset of n = 82 with SMOTE augmentation, the simpler ensemble architecture of RF appears to generalise more reliably than XGBoost's iterative residual correction approach.

SVM achieved the highest specificity in the parsimonious DCE without MGMT configuration (0.77 ± 0.12), making it attractive for scenarios where false-positive TP diagnoses carry the greatest clinical cost. However, in the optimal (with MGMT) configuration, SVM produced the lowest TEAUC (0.71 ± 0.14) and TEAUC (0.71 ± 0.14) among all classifiers, with moderate sensitivity (0.73 ± 0.18) and wide confidence intervals. The sensitivity-specificity trade-off offered by SVM in this dataset is less favourable than RF across the relevant clinical utility range of the ROC curve.

LR performed competitively in configurations with strong linear signal, particularly in T1 post-contrast + MGMT (TEAUC: 0.75; specificity: 0.62). However, LR produced the lowest TEAUC in the optimal parsimonious DCE + MGMT configuration (0.72 ± 0.05), significantly underperforming RF by 17 percentage points. This performance degradation suggests that the joint feature space defined by parsimonious DCE radiomic features and MGMT status contains important non-linear structure and feature interactions that a linear model cannot adequately exploit. LR may, however, remain the preferred choice in resource-limited settings or where model explainability under regulatory frameworks demands linear coefficients.

KNN showed reasonable performance (TEAUC: 0.73 ± 0.11; sensitivity: 0.77 ± 0.17) in the optimal configuration but was consistently outperformed by RF across all four experimental settings. KNN is inherently sensitive to the curse of dimensionality and the choice of

distance metric, and its reliance on the local neighbourhood structure of the training data makes it vulnerable to SMOTE-induced synthetic sample clustering. The wide standard deviations in sensitivity and specificity (±0.17 and ±0.17, respectively) further indicate instability relative to the RF classifier.

This study presents a comprehensive, methodologically rigorous radiomics and machine learning framework for the non-invasive differentiation of TP from PsP in GBM multiforme following standard-of-care chemoradiotherapy. The principal findings can be summarised as follows.

First, the adoption of a parsimonious pharmacokinetic modelling strategy - selecting the optimal voxel-wise model from among five PK frameworks (non-linear Tofts, ETM, SSM, 2CXM, 3S2X) via AICc minimisation - substantially and consistently improves DCE-MRI radiomic feature quality compared to the conventional single-model (ETM) approach. The 5-percentage-point improvement in TEAUC (0.84 vs. 0.63) for the Random Forest classifier when moving from ETM-derived to parsimonious-model-derived features, even before MGMT integration, establishes parsimonious model selection as a methodologically critical step that the field should more broadly adopt.

Second, the integration of MGMT promoter methylation status as a molecular covariate produces a significant and biologically coherent improvement in classification performance. The synergistic combination of parsimonious DCE-derived radiomic features with MGMT status in the RF model yields a TEAUC of 0.89 ± 0.07, sensitivity of 0.93 ± 0.09, specificity of 0.76 ± 0.19, and F1-score of 0.90 ± 0.07 - collectively representing the strongest discriminative performance in this study and placing the proposed framework among the highest-performing non-invasive TP/PsP classifiers in the current literature.

Third, the two-stage dimensionality reduction strategy combining Mann-Whitney U univariate filtering with Elastic Net regularisation effectively reduces the feature space from 1,073 to a biologically interpretable subset while maintaining discriminative power, supporting the feasibility of high-dimensional DCE radiomic analysis in small clinical cohorts.

Fourth, across all experimental configurations, the Random Forest classifier demonstrates consistent superiority over LR, SVM, KNN, and XGBoost in the parsimonious DCE + MGMT setting, attributable to its ensemble bagging mechanism, robustness to high-dimensional correlated features, and aptitude for non-linear decision boundaries in this biologically complex feature space.

The clinical implications of these findings are substantial. The ability to achieve a sensitivity of 0.93 in distinguishing TP from PsP non-invasively - using advanced but commercially available MRI acquisitions and open-source computational tools - has the potential to fundamentally alter the post-treatment management algorithm for GBM patients. Currently, the standard practice when radiographic progression is detected within 6 months of completing chemoradiotherapy is watchful waiting for 4–8 weeks, followed by repeat imaging, with surgical re-biopsy reserved for ambiguous or deteriorating cases. This approach carries significant delays in treatment modification for TP and unnecessary anxiety, morbidity, and cost for pseudoprogressors who are incorrectly labelled as having failed therapy.

A validated non-invasive classifier with the performance characteristics demonstrated by the proposed framework could enable immediate, data-driven treatment stratification at the first post-treatment MRI scan, directing TP patients to early second-line therapy or clinical trial enrolment while sparing PsP patients from ineffective treatment escalation. The incorporation of MGMT methylation status - a routinely obtained molecular biomarker in standard GBM management - ensures that the model is deployable within existing clinical workflows without requiring additional diagnostic procedures.

The following research directions are identified as priorities for advancing this work towards clinical translation. (i) External validation: The proposed framework must be validated in independent, multicentre cohorts with diverse scanner platforms, acquisition protocols, and patient demographics to assess the generalisability and robustness of the selected radiomic feature signature and classification model. Federated learning approaches may facilitate multicentre validation while preserving data privacy. (ii) Multiparametric integration: Future work should systematically evaluate the incremental value of incorporating DSC-MRI (rCBV) and DTI (FA, MD, CL, CP, CS) parameters within the parsimonious modelling framework, given the complementary biological information they provide regarding haemodynamic and microstructural dimensions of the TP/PsP phenotype. (iii) Longitudinal analysis: The current cross-sectional framework operates on a single post-treatment time point; longitudinal radiomic trajectory modelling, capturing feature dynamics across serial MRI acquisitions, may further enhance discriminative performance and provide earlier detection of TP. (iv) Explainability and clinical acceptance: SHAP-based feature attribution analyses and radiologist-in-the-loop validation studies are essential for

building clinician trust and understanding of the model's decision basis. (v) Prospective trial: A prospective, multicentre trial with pre-specified performance thresholds and decision curve analysis is necessary to formally establish the clinical net benefit of the proposed framework and justify its integration into evidence-based clinical guidelines.

Conclusion

In conclusion, this work advances the state of the art in non-invasive TP/PsP discrimination for GBM by demonstrating that the combination of parsimonious pharmacokinetic model selection, multi-region DCE-MRI radiomic feature extraction, MGMT molecular integration, systematic dimensionality reduction, and Random Forest classification yields a clinically meaningful and technically robust classification framework. The findings underscore the importance of pharmacokinetic model selection as a foundational step in DCE-MRI radiomics analysis, clinical features integration, and the superiority of ensemble tree-based classifiers in the high-dimensional, small-sample setting characteristic of neuro-oncological imaging studies. With appropriate external validation and prospective evaluation, the proposed framework holds significant promise for clinical translation as an objective, reproducible, and non-invasive biomarkers for post-operative GBM assessment.

Research Limitations

Sample Size and Class Imbalance

The most fundamental limitation of this study is its modest cohort size. With a total of 82 patients drawn from a single institution, the dataset is small relative to the dimensionality of the radiomic feature space, even after two-stage reduction. Although Adaptive Synthetic Sampling (ADASYN) was employed to address the inherent class imbalance between true progression (n = 53) and pseudoprogression (n = 29), synthetic oversampling generates plausible rather than real minority samples and cannot fully compensate for an underrepresented PsP cohort. The resulting decision boundaries, particularly for tree-based ensemble classifiers, may reflect characteristics of the synthetic distribution rather than the true biological population. The clinical rarity of pseudoprogression relative to true progression in post-standard-of-care GBM cohorts is a structural challenge that is unlikely to be resolved by data augmentation alone, and future work should seek larger, prospectively collected datasets.

Single-Centre Data Acquisition

All imaging data were acquired on a single 3T Siemens MAGNETOM Trio scanner at the University of Pennsylvania using a fixed neuroimaging protocol. While this ensures internal consistency of acquisition parameters, it substantially limits the generalisability of the trained models. Radiomic features are known to be sensitive to scanner type, field strength, reconstruction algorithms, and acquisition protocols in ways that remain only partially characterised [59,60]. A model trained and validated on data from one scanner at one institution may exhibit non-trivial performance degradation when applied to independently acquired datasets with different vendors, coil configurations, or DCE temporal resolutions. Until multicentre external validation is performed, the clinical deployability of the proposed framework remains uncertain.

Ground Truth Heterogeneity

The ground truth labels used in this study were derived from two distinct sources: histopathological analysis from repeat surgery or biopsy in 52 patients, and modified Response Assessment in Neuro-Oncology (mRANO) criteria [53] in the remaining 30 patients for whom tissue was unavailable. This mixed labelling strategy introduces a degree of label noise and definitional inconsistency. Histopathological classification, while the closest available approximation to biological ground truth, is itself subject to sampling bias, given that biopsies may not be representative of the full spatial heterogeneity of the lesion [61]. mRANO-based labelling, in contrast, relies on longitudinal imaging behaviour and is inherently retrospective, potentially misclassifying cases that stabilise late or progress after an initial plateau. The combination of these two labelling strategies in a single cohort introduces heterogeneity in label reliability that cannot be fully corrected by statistical means.

MGMT Status Categorisation

MGMT promoter methylation status was available for all 82 patients but included a non-trivial proportion classified as indeterminate (n = 18; 21.9% of the cohort). In the present analysis, indeterminate MGMT status was retained as a distinct clinical category rather than excluded or imputed. While this decision preserves sample size, it introduces ambiguity into the molecular covariate that is central to the hypothesis. Indeterminate status may reflect genuine biological uncertainty in promoter methylation levels or technical limitations of the assay [62], and its contribution to classifier performance is difficult to interpret with confidence. Future studies should consider standardised methylation quantification approaches that reduce the proportion of indeterminate calls.

Cross-Sectional Imaging Framework

The present study operates on a single post-treatment MRI acquisition per patient. This cross-sectional design does not capture the longitudinal trajectory of contrast enhancement, which is itself diagnostically informative. Pseudoprogression is characterised by spontaneous stabilisation or regression over subsequent imaging timepoints, and a classifier trained on a single timepoint necessarily discards this temporal signal. A longitudinal radiomic modelling framework - tracking feature dynamics across serial acquisitions - would more closely mirror the clinical diagnostic process and is expected to yield superior discriminative performance [63]. The cross-sectional constraint of the present study is partly a consequence of data availability and partly a deliberate scoping decision, but it represents a meaningful limitation relative to what a clinically deployed system would require.

Absence of External Validation

The classification models were evaluated using an internal held-out test set derived from the same 82-patient institutional cohort. No independent external validation cohort was available. In the absence of external validation, the reported performance metrics - however strong - cannot be interpreted as estimates of true generalisation performance on new, unseen populations. Internal validation tends to overestimate model performance relative to external validation, particularly in small cohorts where random variation in train-test splits can substantially inflate or deflate the apparent metrics [64]. Rigorous external validation on a geographically, demographically, and technically independent cohort is an essential prerequisite before the proposed framework can be considered for clinical translation.

Pharmacokinetic Model Assumptions

Although the parsimonious modelling framework represents a methodological advance over single-model approaches, the five pharmacokinetic models evaluated in this study all share a common assumption of fast water exchange between compartments, with the exception of the Shutter-Speed Model, which relaxes this assumption partially. In regions of GBM where cell membrane permeability is profoundly altered by treatment or necrosis, the fast exchange approximation may be systematically violated, introducing bias into the estimated parameters [42]. Additionally, the vascular input function (VIF) was derived manually from a region of interest placed over the superior sagittal sinus rather than from a population-averaged arterial input function. While this approach is physiologically preferable, manual VIF delineation introduces operator-dependent variability that may affect pharmacokinetic parameter estimates across subjects [42].

Absence of Complementary MRI Imaging Modalities

The current framework does not incorporate Dynamic Susceptibility Contrast (DSC) MRI, Diffusion Tensor Imaging (DTI), or metabolic imaging modalities such as MR spectroscopy or PET. Relative cerebral blood volume derived from DSC-MRI and fractional anisotropy from DTI provide complementary haemodynamic and microstructural information that has demonstrated additive discriminative value in the existing literature [11]. The exclusion of these modalities from the present analysis was a consequence of data availability constraints rather than a principled design decision, and it is expected that a fully multiparametric framework incorporating DCE, DSC, and DTI radiomic features would yield superior discriminative performance to that reported here.

Interpretability and Clinical Trust

While Gini-based feature importance scores were computed for the Random Forest classifier to identify the most discriminative radiomic features, this form of attribution is known to be biased towards features with high cardinality and does not provide instance-level explanations [64]. SHAP-based attribution analysis, which assigns contribution scores to individual features for each patient prediction, was not performed in the present study. Without instance-level interpretability, it is difficult to assess whether the classifier is exploiting biologically meaningful signal or residual acquisition artefacts, and clinical acceptance of the model by radiologists and neuro-oncologists is likely to require a more transparent and individually explicable decision basis than aggregate feature importance alone can provide.

Softwares & LIbraries Used

Tool / Library	Version	Category	Role
MATLAB	R2023a	Pharmacokinetic Modelling	Five-model PK fitting (Tofts, ETM, SSM, 2CXM, 3S2X); voxel-wise AICc parsimonious model selection; parametric map generation (Ktrans, Ve, Vp, Fp, τi)
R	4.3.1	Statistics	Descriptive statistics and clinical variable distribution summaries
RStudio	2023.06.1	Statistics	Integrated development environment for R
Python	3.10.12	Runtime	Primary computational runtime for all preprocessing, segmentation, feature extraction, and machine learning components
nibabel	5.1.0	I/O	NIfTI volume loading, header parsing, voxel geometry extraction, and volume serialisation across all pipeline stages
OS / Pathlib	stdlib 3.10	Utility	File system traversal, directory construction, and path resolution throughout the pipeline
SimpleITK	2.0.2	Preprocessing	Skull-stripping mask application; rigid-body affine co-registration and reslicing of PC-T1 and T2-FLAIR to DCE-MRI space
OpenCV	4.8.0	Preprocessing	Intensity normalisation, spatial filtering, and morphological operations prior to segmentation
nnU-Net	2.0	Segmentation	Automated tumour delineation into enhancing, necrotic, and oedematous subregions; BraTS 2018–2021 trained; Dice = 88.95%
TensorFlow	2.13.0	Deep Learning	Backend framework supporting nnU-Net inference and neural architecture experimentation
Keras	2.13.0	Deep Learning	High-level API over TensorFlow for model definition and training
PyRadiomics	3.1.0	Feature Extraction	Extraction of 1,073 radiomic features (first-order, shape, GLCM, GLRLM, GLSZM, NGTDM, GLDM) from contrast-enhancing ROI
SciPy	1.17.1	Statistics	Mann-Whitney U test for univariate feature filtering at $p < 0.05$
scikit-learn	1.8.0	Machine Learning	Elastic Net regularisation, stratified k-fold CV, grid search optimisation, threshold tuning, SVM, RF, LR, KNN training
imbalanced-learn	0.11.0	Machine Learning	ADASYN oversampling within training partition for class imbalance correction
XGBoost	1.7.6	Machine Learning	Gradient boosted tree classifier evaluated across all experimental configurations
NumPy	1.24.3	Utility	Array operations, numerical computation, and matrix manipulation
Pandas	2.0.3	Utility	Feature matrix construction, tabular data management, and clinical variable organisation
joblib	1.3.2	Utility	Parallelised cross-validation execution and model serialisation
tqdm	4.66.1	Utility	Progress monitoring across CV folds, seed repetitions, and extraction batches
warnings / logging	stdlib 3.10	Utility	Execution logging and suppression of non-critical runtime warnings
Matplotlib	3.7.2	Visualisation	ROC curves, confusion matrices, performance plots, and all result figures

The analytical pipeline was implemented across three environments. Pharmacokinetic modelling was performed in MATLAB (R2023a) via an in-house toolbox implementing five PK models (non-linear Tofts, ETM, SSM, 2CXM, 3S2X) with voxel-wise AICc-based parsimonious model selection, generating Ktrans, Ve, Vp, Fp, and τi parametric maps [16, 42–44]. Descriptive statistical analyses and clinical variable summaries were conducted in RStudio (v2023.06.1) running R (v4.3.1). All remaining components - preprocessing, segmentation, feature extraction, and machine learning - were implemented in Python (v3.10.12), with the library stack detailed below.

Raw MRI volumes were loaded and managed using nibabel (v5.1.0), which handled NIfTI I/O, header parsing, and voxel geometry extraction throughout. Skull stripping, rigid-body affine co-registration, and reslicing of PC-T1 and T2-FLAIR images to DCE-MRI space were performed with SimpleITK (v2.0.2) [46, 47], with OpenCV (v4.8.0) supporting auxiliary intensity normalisation and spatial filtering operations. Automated tumour segmentation into contrast-enhancing, non-enhancing/necrotic, and oedematous subregions was performed using nnU-Net (v2.0), trained on BraTS 2018–2021, achieving a Dice score of 88.95% [48–51], with TensorFlow (v2.13.0) / Keras (v2.13.0) providing the underlying deep learning backend.

Radiomic feature extraction was performed with PyRadiomics (v3.1.0) [52], yielding 1,073 features across first-order, shape, and five texture matrix families (GLCM, GLRLM, GLSZM, NGTDM, GLDM). Univariate filtering via the Mann-Whitney U test was implemented in SciPy (v1.17.1) [53]. The complete machine learning pipeline - Elastic Net regularisation, stratified k-fold cross-validation, grid search optimisation, threshold tuning, and training of SVM, RF, LR, and KNN classifiers - was implemented in scikit-learn (v1.8.0) [56, 57], with imbalanced-learn (v0.11.0) providing ADASYN oversampling [54] and XGBoost (v1.7.6) evaluated as the fifth classifier [58].

Supporting infrastructure comprised NumPy (v1.24.3) and Pandas (v2.0.3) for array operations and data management; joblib (v1.3.2) for parallelised cross-validation and model serialisation; tqdm (v4.66.1) for iterative progress monitoring across folds and seed repetitions; warnings / logging (Python standard library v3.10) for execution logging; and Matplotlib (v3.7.2) / Seaborn (v0.12.2) for all result figures, ROC curves, and feature importance visualisations. File system operations were handled via os / pathlib (Python standard library v3.10).

References

1. Kanderi T, Munakomi S, Gupta V. GBM Multiforme. [Updated 2024 May 6]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK558954/
2. Antonelli M, Poliani PL. Adult type diffuse gliomas in the new 2021 WHO Classification. Pathologica. 2022 Dec;114(6):397-409. doi: 10.32074/1591-951X-823. PMID: 36534419; PMCID: PMC9763975.
3. Huang LL, Xia HH, Zhu SL. Ascitic Fluid Analysis in the Differential Diagnosis of Ascites: Focus on Cirrhotic Ascites. J Clin Transl Hepatol. 2014;2(1):58-64. doi:10.14218/JCTH.2013.00010
4. Young JS, Al-Adli N, Sibih YE, Cha S, Berger MS. Pseudoprogression versus true progression in glioblastoma: what neurosurgeons need to know. J Neurosurg. 2023;139(3):748–759. DOI: 10.3171/2022.12.JNS222173 · PMID: 36774577
5. Strauss SB, Meng A, Ebani EJ, Chiang GC. Imaging Glioblastoma Posttreatment: Progression, Pseudoprogression, Pseudoresponse, Radiation Necrosis. Radiol Clin North Am. 2019;57(6):1199-1216. doi:10.1016/j.rcl.2019.07.003
6. Hygino da Cruz LC Jr, Rodriguez I, Domingues RC, Gasparetto EL, Sorensen AG. Pseudoprogression and pseudoresponse: imaging challenges in the assessment of posttreatment glioma. AJNR Am J Neuroradiol. 2011;32(11):1978-1985. doi:10.3174/ajnr.A2397
7. Pooley RA. AAPM/RSNA physics tutorial for residents: fundamental physics of MR imaging. Radiographics. 2005;25(4):1087-1099. doi:10.1148/rg.254055027
8. Grover VP, Tognarelli JM, Crossey MM, Cox IJ, Taylor-Robinson SD, McPhail MJ. Magnetic Resonance Imaging: Principles and Techniques: Lessons for Clinicians. J Clin Exp Hepatol. 2015;5(3):246-255. doi:10.1016/j.jceh.2015.08.001
9. Currie S, Hoggard N, Craven IJ, Hadjivassiliou M, Wilkinson ID. Understanding MRI: basic MR physics for physicians. Postgrad Med J. 2013;89(1050):209-223. doi:10.1136/postgradmedj-2012-131342
10. MRI From Picture to Proton: 2nd ed. AJNR Am J Neuroradiol. 2008;29(6):e50. doi:10.3174/ajnr.A0980
11. Gonçalves FG, Chawla S, Mohan S. Emerging MRI Techniques to Redefine Treatment Response in Patients With Glioblastoma. J Magn Reson Imaging. 2020;52(4):978-997. doi:10.1002/jmri.27105
12. Sourbron SP, Buckley DL. Classic models for dynamic contrast-enhanced MRI. NMR Biomed. 2013;26(8):1004-1027. doi:10.1002/nbm.2940
13. Tofts PS, Brix G, Buckley DL, et al. Estimating kinetic parameters from dynamic contrast-enhanced T(1)-weighted MRI of a diffusable tracer: standardized quantities and symbols. J Magn Reson Imaging. 1999;10(3):223-232. doi:10.1002/(sici)1522-2586(199909)10:3<223::aid-jmri2>3.0.co;2-s
14. Jain RK. Normalization of tumor vasculature: an emerging concept in antiangiogenic therapy. Science. 2005;307(5706):58-62. doi:10.1126/science.1104819

15. Springer CS Jr, Li X, Tudorica LA, et al. Intratumor mapping of intracellular water lifetime: metabolic images of breast cancer?. NMR Biomed. 2014;27(7):760-773. doi:10.1002/nbm.3111
16. H. Akaike, "A new look at the statistical model identification," in IEEE Transactions on Automatic Control, vol. 19, no. 6, pp. 716-723, December 1974, doi: 10.1109/TAC.1974.1100705.
17. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278(2):563-577. doi:10.1148/radiol.2015151169
18. Zou, H. and Hastie, T. (2005), Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67: 301-320. https://doi.org/10.1111/j.1467-9868.2005.00503.x
19. Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. N Engl J Med. 2016;375(13):1216-1219. doi:10.1056/NEJMp1606181
20. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-762. doi:10.1038/nrclinonc.2017.141
21. Sun YZ, Yan LF, Han Y, et al. Differentiation of Pseudoprogression from True Progressionin Glioblastoma Patients after Standard Treatment: A Machine Learning Strategy Combinedwith Radiomics Features from T1-weighted Contrast-enhanced Imaging. *BMC Med Imaging*. 2021;21(1):17. Published 2021 Feb 3. doi:10.1186/s12880-020-00545-5
22. Bani-Sadr A, Eker OF, Berner LP, et al. Conventional MRI radiomics in patients with suspected early- or pseudo-progression. *Neurooncol Adv*. 2019;1(1):vdz019. Published 2019 Sep 1. doi:10.1093/noajnl/vdz019
23. Ari AP, Akkurt BH, Musigmann M, et al. Pseudoprogression prediction in high grade primary CNS tumors by use of radiomics. *Sci Rep*. 2022;12(1):5915. Published 2022 Apr 8. doi:10.1038/s41598-022-09945-9
24. Pak E, Choi KS, Choi SH, et al. Prediction of Prognosis in Glioblastoma Using Radiomics Features of Dynamic Contrast-Enhanced MRI. *Korean J Radiol*. 2021;22(9):1514-1524. doi:10.3348/kjr.2020.1433
25. Bauer AH, Erly W, Moser FG, Maya M, Nael K. Differentiation of solitary brain metastasis from glioblastoma multiforme: a predictive multiparametric approach using combined MR diffusion and perfusion. *Neuroradiology*. 2015;57(7):697-703. doi:10.1007/s00234-015-1524-6
26. O'Neill AF, Qin L, Wen PY, de Groot JF, Van den Abbeele AD, Yap JT. Demonstration of DCE-MRI as an early pharmacodynamic biomarker of response to VEGF Trap in glioblastoma. *J Neurooncol*. 2016;130(3):495-503. doi:10.1007/s11060-016-2243-5
27. Bani Sadr A, Eker OF, Berner L, et al. P13.02 Conventional MRI radiomics in the diagnosis of early- and pseudo-progression. *Neuro Oncol*. 2019;21(Suppl 3):iii62. doi:10.1093/neuonc/noz126.223

28. Liu C, Ding J, Spuhler K, et al. Preoperative prediction of sentinel lymph node metastasis in breast cancer by radiomic signatures from dynamic contrast-enhanced MRI. *J Magn Reson Imaging*. 2019;49(1):131-140. doi:10.1002/jmri.26224
29. Wang J, Chen J, Zhou R, Gao Y, Li J. Machine learning-based multiparametric MRI radiomics for predicting poor responders after neoadjuvant chemoradiotherapy in rectal Cancer patients. *BMC Cancer*. 2022;22(1):420. Published 2022 Apr 19. doi:10.1186/s12885-022-09518-z
30. Cheng Y, Xu S, Wang H, et al. Intra- and peri-tumoral radiomics for predicting the sentinel lymph node metastasis in breast cancer based on preoperative mammography and MRI. *Front Oncol*. 2022;12:1047572. Published 2022 Dec 12. doi:10.3389/fonc.2022.1047572
31. Singh G, Manjila S, Sakla N, et al. Radiomics and radiogenomics in gliomas: a contemporary update. *Br J Cancer*. 2021;125(5):641-657. doi:10.1038/s41416-021-01387-w
32. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. 2005;352(10):987-996. doi:10.1056/NEJMoa043330
33. Wen PY, Macdonald DR, Reardon DA, et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group. *J Clin Oncol*. 2010;28(11):1963-1972. doi:10.1200/JCO.2009.26.3541
34. Okada H, Weller M, Huang R, et al. Immunotherapy response assessment in neuro-oncology: a report of the RANO working group. *Lancet Oncol*. 2015;16(15):e534-e542. doi:10.1016/S1470-2045(15)00088-1
35. Hegi ME, Diserens AC, Gorlia T, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med*. 2005;352(10):997-1003. doi:10.1056/NEJMoa043331
36. Weller M, van den Bent M, Tonn JC, et al. European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. *Lancet Oncol*. 2017;18(6):e315-e329. doi:10.1016/S1470-2045(17)30194-8
37. Parker GJ, Roberts C, Macdonald A, et al. Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. *Magn Reson Med*. 2006;56(5):993-1000. doi:10.1002/mrm.21066
38. Rohrer M, Bauer H, Mintorovitch J, Requardt M, Weinmann HJ. Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. *Invest Radiol*. 2005;40(11):715-724. doi:10.1097/01.rli.0000184756.66360.d3
39. Pintaske J, Martirosian P, Graf H, et al. Relaxivity of Gadopentetate Dimeglumine (Magnevist), Gadobutrol (Gadovist), and Gadobenate Dimeglumine (MultiHance) in human blood plasma at 0.2, 1.5, and 3 Tesla. *Invest Radiol*. 2006;41(3):213-221. doi:10.1097/01.rli.0000197668.44926.f7
40. Stanisz GJ, Odrobina EE, Pun J, et al. T1, T2 relaxation and magnetization transfer in tissue at 3T. *Magn Reson Med*. 2005;54(3):507-512. doi:10.1002/mrm.20605

41. Shin W, Cashen TA, Horowitz SW, Sawlani R, Carroll TJ. Quantitative CBV measurement from static T1 changes in tissue and correction for intravascular water exchange. *Magn Reson Med*. 2006;56(1):138-145. doi:10.1002/mrm.20937
42. Sourbron SP, Buckley DL. On the scope and interpretation of the Tofts models for DCE-MRI. *Magn Reson Med*. 2011;66(3):735-745. doi:10.1002/mrm.22861
43. Li X, Rooney WD, Springer CS Jr. A unified magnetic resonance imaging pharmacokinetic theory: intravascular and extracellular contrast reagents. *Magn Reson Med*. 2005;54(6):1351-1359. doi:10.1002/mrm.20684
44. Buckley DL, Shurrab AE, Cheung CM, Jones AP, Mamtora H, Kalra PA. Measurement of single kidney function using dynamic contrast-enhanced MRI: comparison of two models in human subjects. *J Magn Reson Imaging*. 2006;24(5):1117-1123. doi:10.1002/jmri.20699
45. Heye AK, Culling RD, Valdés Hernández Mdel C, Thrippleton MJ, Wardlaw JM. Assessment of blood-brain barrier disruption using dynamic contrast-enhanced MRI. A systematic review. *Neuroimage Clin*. 2014;6:262-274. Published 2014 Sep 10. doi:10.1016/j.nicl.2014.09.002
46. Lowekamp BC, Chen DT, Ibáñez L, Blezek D. The Design of SimpleITK. *Front Neuroinform*. 2013;7:45. Published 2013 Dec 30. doi:10.3389/fninf.2013.00045
47. Yaniv Z, Lowekamp BC, Johnson HJ, Beare R. SimpleITK Image-Analysis Notebooks: a Collaborative Environment for Education and Reproducible Research. *J Digit Imaging*. 2018;31(3):290-303. doi:10.1007/s10278-017-0037-8
48. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203-211. doi:10.1038/s41592-020-01008-z
49. Menze BH, Jakab A, Bauer S, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Trans Med Imaging*. 2015;34(10):1993-2024. doi:10.1109/TMI.2014.2377694
50. Bakas S, Akbari H, Sotiras A, et al. Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci Data*. 2017;4:170117. Published 2017 Sep 5. doi:10.1038/sdata.2017.117
51. Baid, U., "The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radiogenomic Classification", arXiv e-prints, Art. no. arXiv:2107.02314, 2021. doi:10.48550/arXiv.2107.02314.
52. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res*. 2017;77(21):e104-e107. doi:10.1158/0008-5472.CAN-17-0339
53. Mann, H. B., & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, *18*(1), 50–60. https://doi.org/10.1214/AOMS/1177730491
54. He, Haibo, Yang Bai, E.A. Garcia, and Shutao Li (2008), "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," *International Joint Conference on Neural Network*, 1322–28.
55. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv*. Published 2014. Accessed April 1, 2026. https://arxiv.org/abs/1106.1813

56. Cortes, C., Vapnik, V. Support-vector networks. *Mach Learn* 20, 273–297 (1995). https://doi.org/10.1007/BF00994018
57. Breiman, L. Random Forests. *Machine Learning* 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324
58. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. arXiv [cs.LG]. 2016:arXiv:1603.02754. Published June 10, 2016. Accessed April 1, 2026. https://arxiv.org/abs/1603.02754
59. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*. 2020;295(2):328-338. doi:10.1148/radiol.2020191145
60. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys*. 2017;44(3):1050-1062. doi:10.1002/mp.12123
61. Albert NL, Weller M, Suchorska B, et al. Response Assessment in Neuro-Oncology working group and European Association for Neuro-Oncology recommendations for the clinical use of PET imaging in gliomas. *Neuro Oncol*. 2016;18(9):1199-1208. doi:10.1093/neuonc/now058
62. Bady P, Delorenzi M, Hegi ME. Sensitivity Analysis of the MGMT-STP27 Model and Impact of Genetic and Epigenetic Context to Predict the MGMT Methylation Status in Gliomas and Other Tumors. *J Mol Diagn*. 2016;18(3):350-361. doi:10.1016/j.jmoldx.2015.11.009
63. Ellingson BM, Wen PY, Cloughesy TF. Modified Criteria for Radiographic Response Assessment in Glioblastoma Clinical Trials. *Neurotherapeutics*. 2017;14(2):307-320. doi:10.1007/s13311-016-0507-6
64. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC Bioinformatics*. 2007;8:25. Published 2007 Jan 25. doi:10.1186/1471-2105-8-25

Supplementary

Section - S1

DCE-MRI is an advanced perfusion imaging technique that evaluates tissue vascularity, capillary permeability, and microenvironmental physiology by tracking the temporal dynamics of a GCA. Serial images are captured before, during, and after contrast administration, allowing quantitative assessment of how rapidly and to what extent the contrast agent distributes within intravascular and extravascular compartments. The signal–time curves obtained from DCE-MRI are fitted using pharmacokinetic (PK) models most commonly the Tofts or Extended Tofts model to derive quantitative biomarkers that reflect vascular integrity, perfusion, and microstructural properties of tissue. These metrics have become essential in neuro-oncology for differentiating true progression (TP) from treatment-induced pseudoprogression (PsP), as GBM profoundly alters vascular permeability and extracellular space.

The Two-Compartment Physiological Model

All pharmacokinetic models employed in this study share a common physiological foundation. The tissue voxel is partitioned into three distinct compartments whose volume fractions must sum to unity:

$$v_e + v_p + v_i = 1 \quad \text{--- (1)}$$

where v_p is the plasma volume fraction (blood plasma within capillaries), ve is the extravascular extracellular space (EES, the interstitial space between cells and outside vessels), and vi is the intracellular volume fraction. GCA distributes between the plasma and EES compartments; the intracellular space is impermeable to GCA in standard models (with the exception of the Shutter-Speed Model, which additionally considers transcytolemmal water exchange).

The GCA concentrations in each compartment are defined as follows:

$C_p(t) = CGA\ Concentration\ in\ blood\ plasma\ [mM]$

$C_t(t) = CGA\ Concentration\ in\ whole\ tissue\ voxel\ [mM]$

$C_e(t) = CGA\ Concentration\ in\ the\ EES\ [mM]$

$K^{trans} = Volume\ transfer\ constant\ [\ min^{-1}\ ]$

$V_e = EES\ Volume\ Fraction\ [\ dimensionless\ ]$

$V_p = Plasma\ Volume\ Fraction\ [\ dimensionless\ ]$

$K^{ep} = Rate\ constant\ from\ EES\ back\ to\ plasma\ [\ min^{-1}\ ] = \frac{K^{trans}}{V_e}$

$F_p = Plasma\ flow\ rate\ [\ mL/min/ML\ tissue\ ]$

The total tissue concentration is therefore:

$$C_t(t) = V_p . C_p(t) + V_e . C_e(t) \quad \text{--- (2)}$$

From Scanner Signal to Contrast Concentration

Before any pharmacokinetic model can be applied, the raw scanner signal S(t) must be converted to a physically meaningful GCA concentration Ct(t) in millimolar. This pre-processing proceeds in two steps.

Step1: Signal to longitudinal relaxation time

The spoiled gradient-echo (GRE) sequence used in this study (TR = 3.23 ms, flip angle $\theta = 23$ θ = 23◦) produces a signal governed by the Ernst equation:

$$S(t) = S_0 \cdot \frac{\sin\theta \left(1 - e^{-\mathrm{TR}/T_1(t)}\right)}{1 - \cos\theta \cdot e^{-\mathrm{TR}/T_1(t)}} \quad \text{--- (3)}$$

where S_0 is a proton-density scaling constant and $T_1(t)$ is the longitudinal relaxation time, which shortens as GCA accumulates. Inverting Equation (3) at each time frame yields $T_1(t)$. The pre-contrast baseline $T_{1,0}$ was fixed at $1400\ ms$ for tissue and $1600\ ms$ for blood, consistent with values reported in the literature at $3T$

Step2: Relaxation rate to GCA concentration

The GCA shortens T1 linearly through its longitudinal relaxivity r_1 (the Solomon Bloembergen Morgan relation under fast exchange) :

$$\frac{1}{T_1(t)} = \frac{1}{T_{1,0}} + r_1 \cdot C_t(t) \quad \text{--- (4)}$$

Solving for Ct(t), with $r_1 = 4.3\ L\ mmol^{-1}s^{-1}$ (Multihande at 3T)

$$C_t(t) = \frac{1}{r_1}\left(\frac{1}{T_1(t)} - \frac{1}{T_{1,0}}\right) \quad [\mathrm{mM}] \qquad \text{--- (5)}$$

The same conversion is applied to a manually defined region of interest (ROI) over the Superior Sagittal Sinus (SSS) to obtain the vascular input function (VIF) $C_p(t)$, using $T_{1,blood} = 1600ms$. This measured $C_p(t)$ serves as the arterial driving term in all pharmacokinetic models. Without a measured $C_p(t)$, every ODE below is mathematically underdetermined.

Pharmacokinetic Model Formulations and ODE Derivations

DCE-MRI data from each segmented tumour region were fitted to five pharmacokinetic models: the non-linear Tofts model, the Extended Tofts Model (ETM), the Shutter-Speed Model (SSM), the Two Compartment Exchange Model (2CXM), and the Three-Site Two-Exchange Model (3S2X). The derivation of each model, together with the explicit mathematical form by which each parameter is extracted, is presented below.

The Standard Tofts Model- Deriving K^{trans} and V_e

Physiological basis:

K^{trans} is a measure of capillary permeability obtained using DCE-MRI perfusion. It is calculated by measuring the accumulation of GCA in the EES. In GBM, VEGF produced by the tumour reduces the effectiveness of gap junctions and creates fenestrations in the endothelium of brain capillaries; this has been shown to correlate with tumour grade. K^{trans} is interpreted variably according to the dominant physiological regime:

Where permeability is very low (near-intact BBB): $K^{trans} \approx PS$ (Permeability Surface area product, the Renkin Crone limit), a pure measure of vascular wall leakiness.

Where permeability is very high (disrupted BBB): $K^{trans} \rightarrow F_p$ (Flow Limited regime), reflecting blood delivery rather than wall permeability.

ODE derivation: The Tofts model (Tofts & Kermode, 1991) assumes $V_p \approx 0$, making it appropriate for tissues with near intact BBB. A mass balance on the EES states that the rate of accumulation of GCA in the EES equals the rate of influx from plasma minus the rate of efflux back to plasma :

$$V_e \frac{dC_e(t)}{dt} = K^{trans}.\, C_p(t) - K^{trans} C_e(t) \text{ --- (6)}$$

Since $C_t(t) \approx V_e.\, C_e(t)$ under the Tofts assumption, dividing through by Ve gives the governing ODE :

$$\frac{dC_t(t)}{dt} = K^{\text{trans}} \cdot C_p(t) - k_{ep} \cdot C_t(t), \quad k_{ep} = \frac{K^{\text{trans}}}{V_e} \text{ --- (7)}$$

Analytical solution:

Equation (7) is a first-order linear ODE. Multiplying both sides by the integrating factor $e^{K_{ep}t}$

$$\frac{d}{dt}\left[C_t(t) \cdot e^{k_{ep}t}\right] = K^{\text{trans}} \cdot C_p(t) \cdot e^{k_{ep}t} \text{ --- (8)}$$

Integrating from 0 to t with initial condition $C_t(t) = 0$

$$C_t(t) = K^{trans} \int_o^t C_p(\tau).\, e^{-K_{ep}(t-\tau)} d\tau$$

This is a convolution of the plasma input function $C_p(t)$ with a mono-exponential impulse response. The wash-in phase corresponds to the rising integral as $C_p(\tau)$ delivers GCA; the washout phase corresponds to the exponential decay term $e^{-K_{ep}(t-\tau)}$ dominating as $C_p(t)$ falls.

Since $C_p(\tau)$ is measured at discrete frames (temporal resolution 7 s, N = 30 frames), the integral is evaluated via the trapezoid rule :

$$C_t(t_n) \approx K^{\text{trans}} \sum_{j=0}^{n-1} \frac{1}{2}\left[C_p(t_j)e^{-k_{ep}(t_n - t_j)} + C_p(t_{j+1})e^{-k_{ep}(t_n - t_{j+1})}\right] \Delta t$$

For trial parameters $\theta = [K^{trans}, K_{ep}]$ this sum produces a predicted curve $C_t(\hat{t_n}, \theta)$. The Levenberg–Marquardt algorithm minimises the residual sum of squares:

$$\mathcal{L}(K^{\text{trans}}, k_{ep}) = \sum_{n=1}^{N} \left[C_t(t_n) - \hat{C}_t(t_n)\right]^2$$

via the iterative update rule,

$$\theta_{\text{new}} = \theta_{\text{old}} - \left(J^\top J + \lambda I\right)^{-1} J^\top r$$

where J is the Jacobian of the residual vector r with respect to θ, and λ is the **Levenberg damping parameter.** The algorithm terminates when $||\frac{\delta L}{\delta \theta}|| < \epsilon_{tol}$. The recovered parameters are K^{trans} and K_{ep} from which:

$$V_e = \frac{K^{trans}}{K_{ep}}$$

The Extended Tofts Model (ETM) - deriving V_p

ODE Derivation:

The ETM (Tofts et al., 1999) adds a plasma compartment term. Total tissue concentration now has two contributions (Equation (2)). The EES mass balance ODE is unchanged:

$$V_e \frac{dC_e(t)}{dt} = K^{\text{trans}} \cdot C_p(t) - K^{\text{trans}} \cdot C_e(t)$$

Differentiating $C_t(t) = V_p C_p(t) + V_e C_e(t)$ and substituting →

$$\frac{dC_t(t)}{dt} = V_p \frac{dC_p(t)}{dt} + K^{\text{trans}} \cdot C_p(t) - k_{ep}\left[C_t(t) - V_p C_p(t)\right]$$

Analytical Solution :

$$C_t(t) = V_p \cdot C_p(t) + K^{\text{trans}} \int_0^t C_p(\tau)\, e^{-k_{ep}(t-\tau)}\, d\tau$$

Parameter Extraction :

The fitting procedure is identical to the Tofts case, now optimising $\theta = [K^{trans}, K_{ep}, V_p]$. The parameter V_p is recovered as the coefficient of the instantaneous $C_p(t)$ term - distinct from the convolution integral because it carries no exponential decay and directly tracks the intravascular bolus. The ETM thus yields the parameter set $\{K^{trans}, V_e, V_p\}$ and the derived rate constant $K_{ep} = \frac{K^{trans}}{V_e}$.

The Two-Compartment Exchange Model (2CXM) - deriving Fp

F_p describes the rate of plasma delivery per unit tissue volume, representing true tissue perfusion. It is influenced by blood flow, vessel density, vascular resistance, and capillary transit time. Reduced F_p may indicate hypoperfusion in necrotic or treatment-affected regions, whereas increased F_p reflects active angiogenesis and aggressive tumour behaviour.

ODE derivation: The 2CXM (Lawrence & Lee, 1998) assigns a full mass balance to the plasma compartment. GCA arrives via arterial inflow, exits via venous drainage, and leaks bidirectionally across the capillary wall.

Compartment 1 – Plasma :

$$V_p \frac{dC_p^{\text{tiss}}(t)}{dt} = F_p \left[C_p^{\text{AIF}}(t) - C_p^{\text{tiss}}(t)\right] - K^{\text{trans}} C_p^{\text{tiss}}(t) + K^{\text{trans}} C_e(t)$$

Applying the ***central volume theorem*** (venous concentration ≈ tissue plasma concentration) :

$$V_p \frac{dC_p(t)}{dt} = F_p C_p^{\text{AIF}}(t) - \left(F_p + K^{\text{trans}}\right) C_p(t) + K^{\text{trans}} C_e(t)$$

Compartment 2 – EES:

$$V_e \frac{dC_e(t)}{dt} = K^{\text{trans}} C_p(t) - K^{\text{trans}} C_e(t)$$

In matrix form, with state vector $X = [C_p, C_e]^T$

$\frac{dX}{dt} = AX + b(t)$

$$A = \begin{pmatrix} -\dfrac{F_p + K^{\text{trans}}}{V_p} & \dfrac{K^{\text{trans}}}{V_p} \\ \dfrac{K^{\text{trans}}}{V_e} & -\dfrac{K^{\text{trans}}}{V_e} \end{pmatrix}, \quad b(t) = \begin{pmatrix} \dfrac{F_p}{V_p} C_p^{\text{AIF}}(t) \\ 0 \end{pmatrix}$$

Since A is diagonalizable with eigenvalues:

$$\lambda_{1,2} = \frac{1}{2}\left[-\left(\frac{F_p + K^{\text{trans}}}{V_p} + \frac{K^{\text{trans}}}{V_e}\right) \pm \sqrt{\Delta}\right]$$

Where the discriminant is

$$\Delta = \left(\frac{F_p + K^{\text{trans}}}{V_p} - \frac{K^{\text{trans}}}{V_e}\right)^2 + \frac{4K^{\text{trans}}F_p}{V_p V_e}$$

The solution for total tissue concentration takes the form of a bi-exponential convolution with the AIF

$$C_t(t) = \left(A_1 e^{\lambda_1 t} + A_2 e^{\lambda_2 t}\right) * C_p^{\text{AIF}}(t)$$

where amplitudes A_1, A_2 are algebraically determined by $\{F_p, V_{p,} K^{trans}, V_e\}$.

Parameter Extraction: At each Levenberg Marquardt iteration, the ODE system *(20)* is integrated forward in time using a fourth-order Runge Kutta scheme. At each time step $h = 7s$

$$\begin{aligned}
k_1 &= Ax(t) + b(t) \\
k_2 &= A\left(x(t) + \frac{h}{2}k_1\right) + b\left(t + \frac{h}{2}\right) \\
k_3 &= A\left(x(t) + \frac{h}{2}k_2\right) + b\left(t + \frac{h}{2}\right) \\
k_4 &= A\left(x(t) + hk_3\right) + b(t + h)
\end{aligned}$$

$$x(t + h) = x(t) + \frac{h}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right)$$

Then $C_t(t) = V_p C_p(t) + V_e C_e(t)$ is computed, compared against measured $C_p(t)$, and the residuals feed back to the optimiser. The parameter Fp is recovered as the free parameter controlling the ratio and temporal separation of λ1 and λ2, yielding the full set $\{K^{trans}, V_{e'}, V_{p'}, F_p\}$.

The Shutter-Speed Model (SSM) - deriving τ_i

τ_i quantifies the average time water molecules remain inside cells before exchanging with the extracellular space. It serves as a sensitive marker for cell membrane permeability, cell viability, metabolic activity, and microstructural integrity. Alterations in τ_i may reflect changes in cell health, therapy response, or tumour proliferation dynamics. Unlike all other

DCE parameters - which describe GCA movement - τ_i describes water movement, exploiting the fact that MRI is fundamentally a water proton signal.

ODE Derivation :

Let $M_i(t)\ and\ M_e(t)$ be the longitudinal magnetisations of intracellular and extracellular water pools, respectively. The Bloch McConnell equations for a two-site exchange system are:

$$\frac{dM_i}{dt} = -\frac{M_i}{T_{1i}} - \frac{M_i - M_i^{eq}}{\tau_i} + \frac{M_e - M_i^{eq}}{\tau_e}$$

$$\frac{dM_e}{dt} = -\frac{M_e}{T_{1e}} - \frac{M_e - M_e^{eq}}{\tau_e} + \frac{M_i - M_i^{eq}}{\tau_i}$$

Where,
τ_{1i} and τ_{1e} are the intracellular and extracellular longitudinal relaxation times.
τ_i is the intracellular water lifetime (time, before a water molecule crosses the membrane)
τ_e is the mean extracellular water lifetime, with the constraint from detailed balance: $\frac{V_i}{\tau_i} = \frac{V_e}{\tau_e}$. Where, V_i, V_e after the intracellular volume fractions. At equilibrium, $M_i^{eq} \propto V_i$, $M_e^{eq} \propto V_e$. The extracellular shortened by the GCA:

$\frac{1}{T_{1e}(t)} = \frac{1}{T_{1e,0}} + r_1 C_e(t)$. Where $r_1$is the GCA longitudinal relaxivity $(4.3\ L.mmol^{-1}s^{-1}\ in\ this\ study)$ and $C_e(t)$ is obtained from the underlying ETM ODE.
In matrix form:

$$\frac{d}{dt}\begin{pmatrix} M_i \\ M_e \end{pmatrix} = \mathbf{K}\begin{pmatrix} M_i \\ M_e \end{pmatrix} + \mathbf{c}$$

$$\mathbf{K} = \begin{pmatrix} -\frac{1}{T_{1i}} - \frac{1}{\tau_i} & \frac{1}{\tau_e} \\ \frac{1}{\tau_i} & -\frac{1}{T_{1e}(t)} - \frac{1}{\tau_e} \end{pmatrix}$$

Because T_{1e} is time-varying (driven by GCA kinetics), K is time-dependent and the system must be solved numerically. The observed MR signal is proportional to $M_i + M_e$ and the apparent tissue T_1 is extracted from the signal recovery curve, from which $\{K^{trans}, V_e, \tau_i\}$ are jointly fitted. Biological significance of τ_i: Viable, metabolically active cells maintain tight membrane integrity, producing longer τ_i(slower exchange). Therapy-damaged or necrotic cells have disrupted membranes, producing shorter τ_i (faster exchange). In the TP/PsP context,τ_i therefore reflects cell viability independently of vascular permeability, providing complementary biological information to K^{trans}.

HyperParameter Grid

Logistic Regression

```
DEFAULT_PARAM_GRID =  [
    # uniform weights + euclidean  (standard distance, equal vote)
    {"knn__n_neighbors": [3],  "knn__weights": ["uniform"],  "knn__metric":
["euclidean"], "knn__algorithm": ["auto"], "knn__leaf_size": [20]},
    {"knn__n_neighbors": [7],  "knn__weights": ["uniform"],  "knn__metric":
["euclidean"], "knn__algorithm": ["auto"], "knn__leaf_size": [30]},
    {"knn__n_neighbors": [15], "knn__weights": ["uniform"],  "knn__metric":
["euclidean"], "knn__algorithm": ["auto"], "knn__leaf_size": [30]},

    # ── distance weights + euclidean  (closer neighbors vote more)
    {"knn__n_neighbors": [3],  "knn__weights": ["distance"], "knn__metric":
["euclidean"], "knn__algorithm": ["auto"], "knn__leaf_size": [20]},
    {"knn__n_neighbors": [7],  "knn__weights": ["distance"], "knn__metric":
["euclidean"], "knn__algorithm": ["auto"], "knn__leaf_size": [30]},
    {"knn__n_neighbors": [15], "knn__weights": ["distance"], "knn__metric":
["euclidean"], "knn__algorithm": ["auto"], "knn__leaf_size": [30]},

    # ── uniform weights + manhattan
    {"knn__n_neighbors": [3],  "knn__weights": ["uniform"],  "knn__metric":
["manhattan"], "knn__algorithm": ["auto"], "knn__leaf_size": [20]},
    {"knn__n_neighbors": [7],  "knn__weights": ["uniform"],  "knn__metric":
["manhattan"], "knn__algorithm": ["auto"], "knn__leaf_size": [30]},
    {"knn__n_neighbors": [15], "knn__weights": ["uniform"],  "knn__metric":
["manhattan"], "knn__algorithm": ["auto"], "knn__leaf_size": [30]},

    # ── distance weights + manhattan  (proximity-sensitive + outlier-robust)
    {"knn__n_neighbors": [3],  "knn__weights": ["distance"], "knn__metric":
["manhattan"], "knn__algorithm": ["auto"], "knn__leaf_size": [20]},
    {"knn__n_neighbors": [7],  "knn__weights": ["distance"], "knn__metric":
["manhattan"], "knn__algorithm": ["auto"], "knn__leaf_size": [30]},
    {"knn__n_neighbors": [15], "knn__weights": ["distance"], "knn__metric":
["manhattan"], "knn__algorithm": ["auto"], "knn__leaf_size": [30]},

]
```

K - Nearhood Neighbours

```
DEFAULT_PARAM_GRID =  {
    "svm__C":            [0.01, 0.1, 1, 10, 100],
    "svm__kernel":       ["rbf", "linear", "poly"],
    "svm__gamma":        ["scale", "auto"],
    "svm__degree":       [2, 3],          # only used when kernel="poly"
    "svm__class_weight": ["balanced"],
}
```

Support Vector Classifier

```
DEFAULT_PARAM_GRID = {
    "rf__n_estimators":      [300, 500, 800],
    "rf__max_depth":         [None, 5, 10, 20],
    "rf__min_samples_split": [2, 5, 10],
    "rf__min_samples_leaf":  [1, 2, 4],
    "rf__max_features":      ["sqrt", "log2"],
    "rf__bootstrap":         [True, False],
    "rf__class_weight":      ["balanced"],
}
```

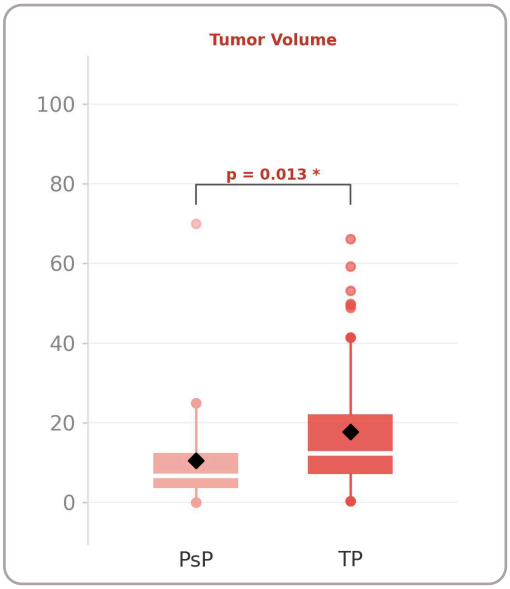
Random Forest

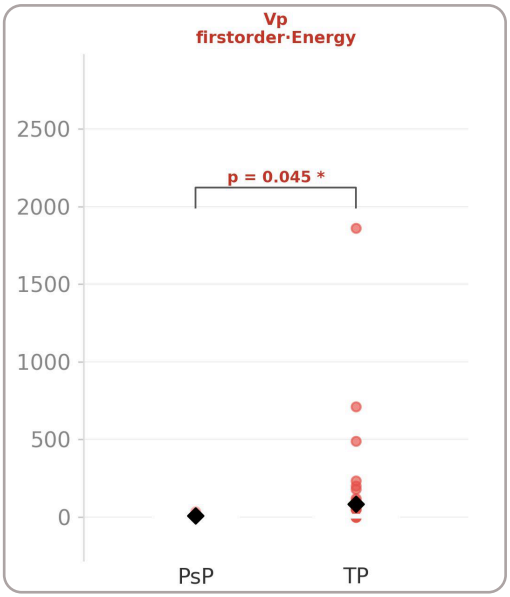
```
DEFAULT_PARAM_GRID = {
    "rf__n_estimators":      [300, 500, 800],
    "rf__max_depth":         [None, 5, 10, 20],
    "rf__min_samples_split": [2, 5, 10],
    "rf__min_samples_leaf":  [1, 2, 4],
    "rf__max_features":      ["sqrt", "log2"],
    "rf__bootstrap":         [True, False],
    "rf__class_weight":      ["balanced"],
}
```

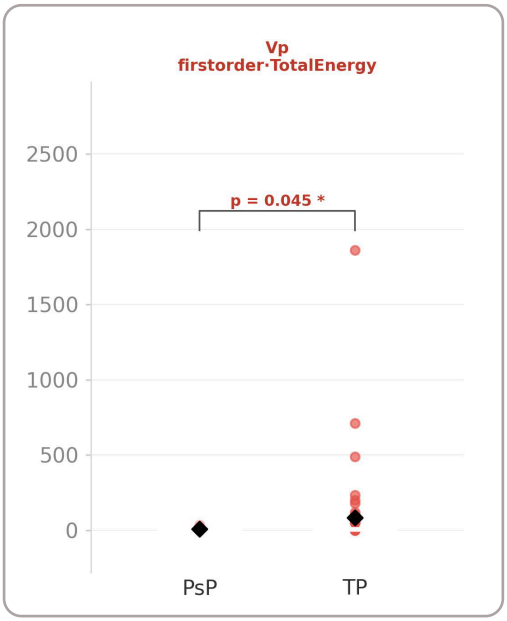
XGBoost

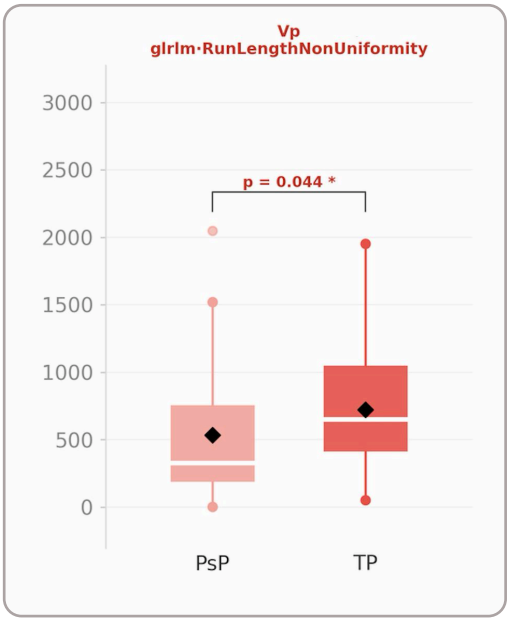
```
DEFAULT_PARAM_GRID = {
    "rf__n_estimators":      [300, 500, 800],
    "rf__max_depth":         [None, 5, 10, 20],
    "rf__min_samples_split": [2, 5, 10],
    "rf__min_samples_leaf":  [1, 2, 4],
    "rf__max_features":      ["sqrt", "log2"],
    "rf__bootstrap":         [True, False],
    "rf__class_weight":      ["balanced"],
}
```

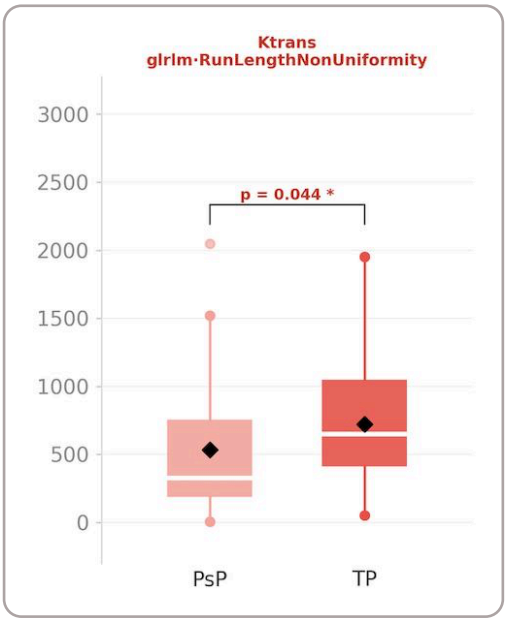
Extended TOFTS + MGMT

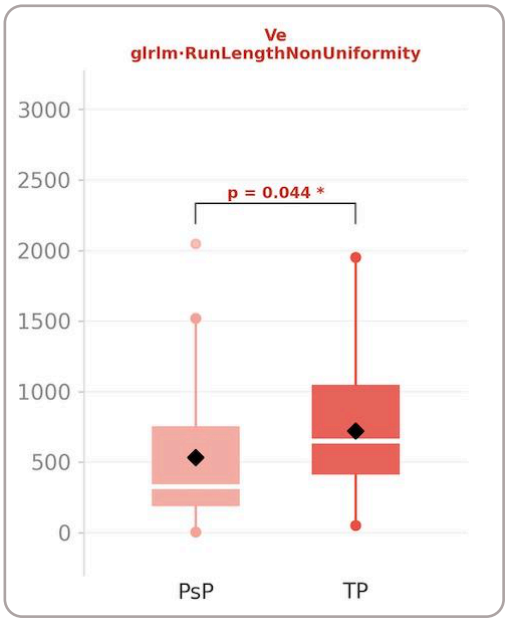




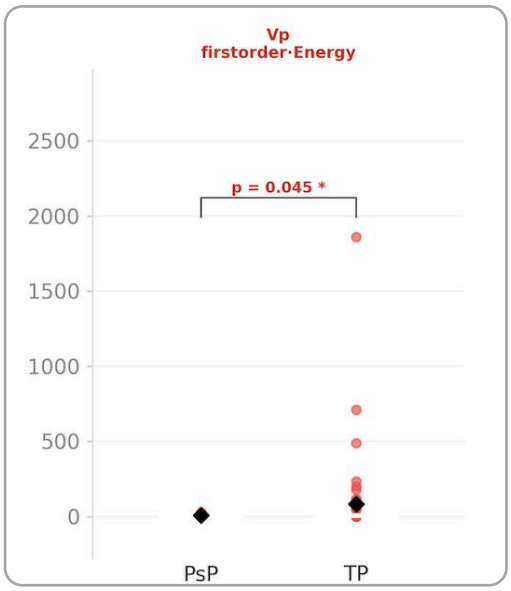


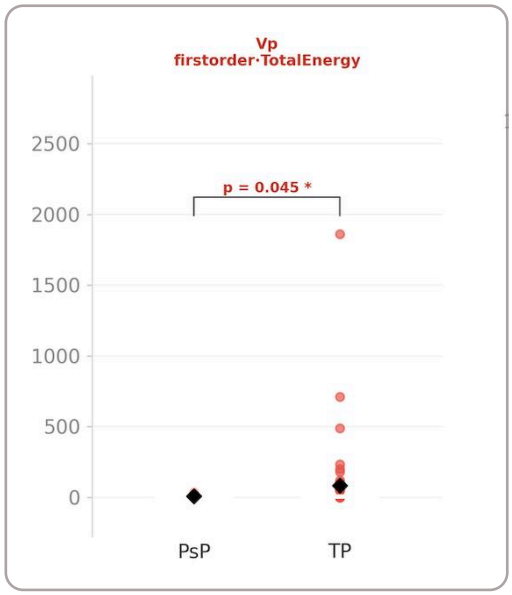


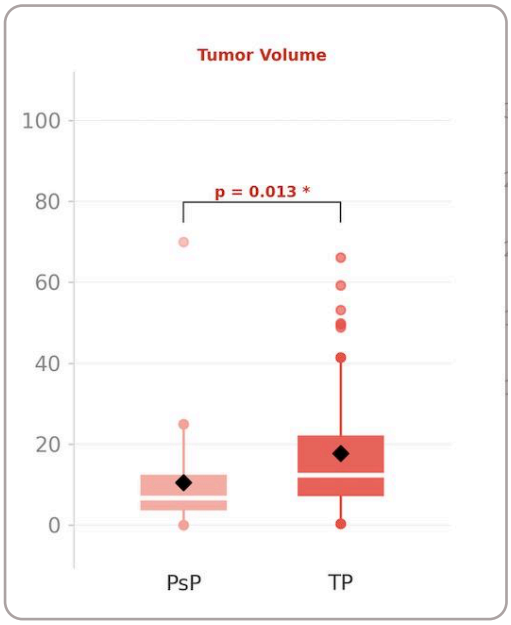


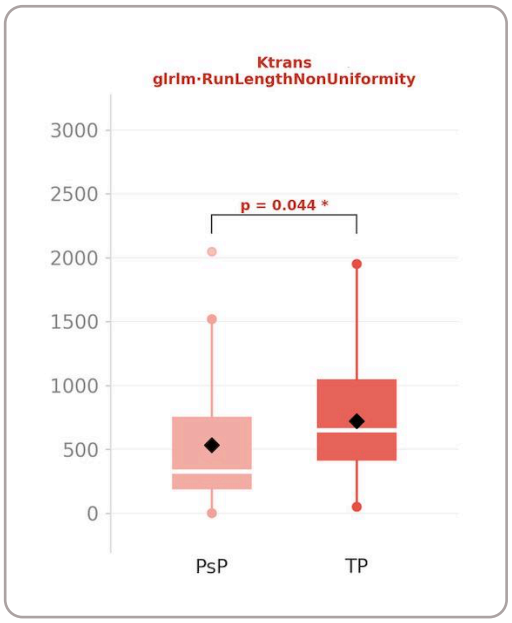


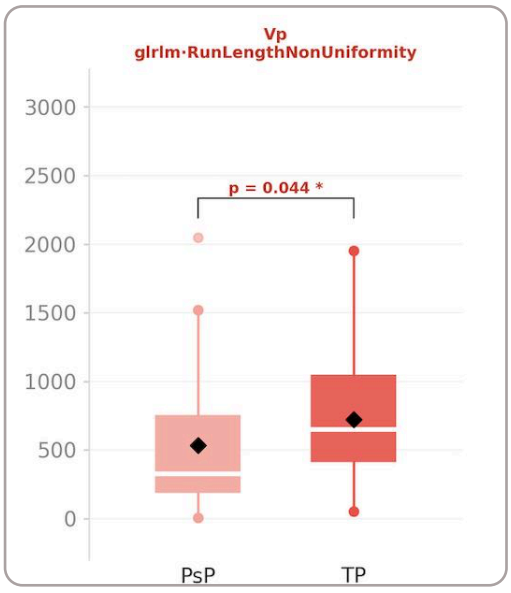
Parsimonious Model

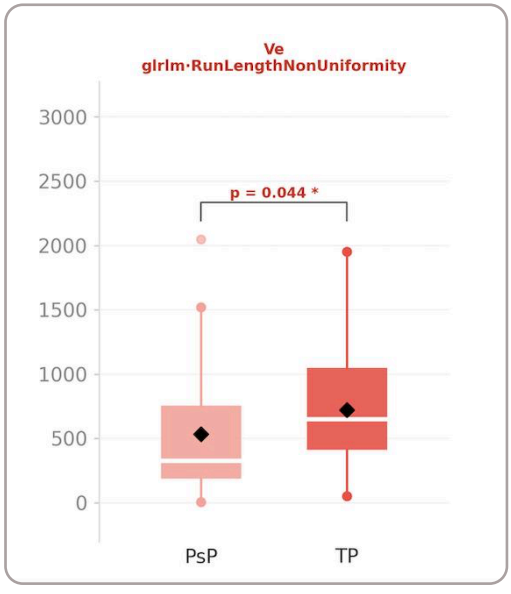




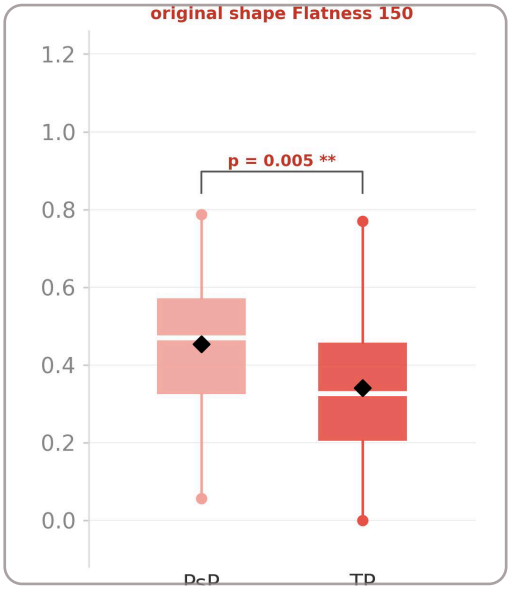


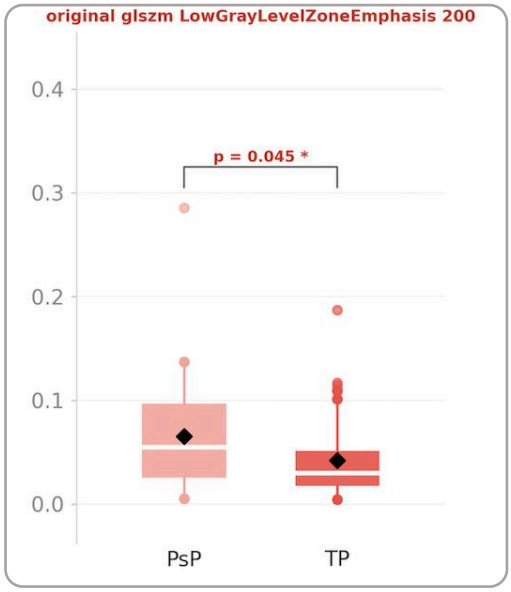


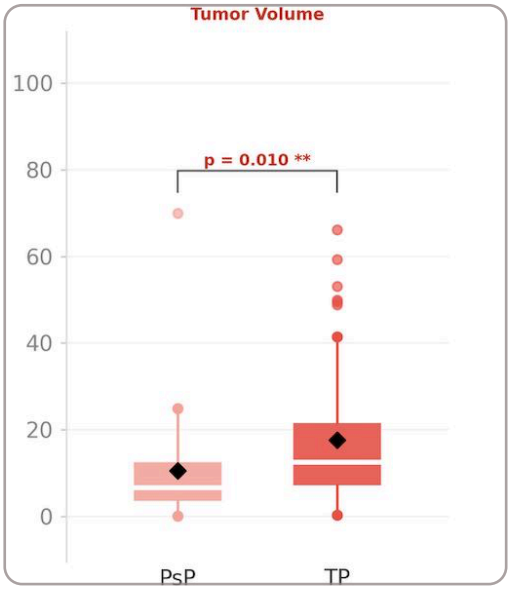


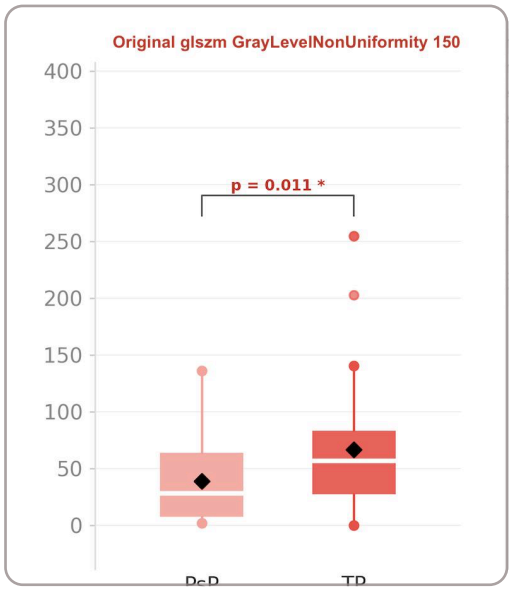


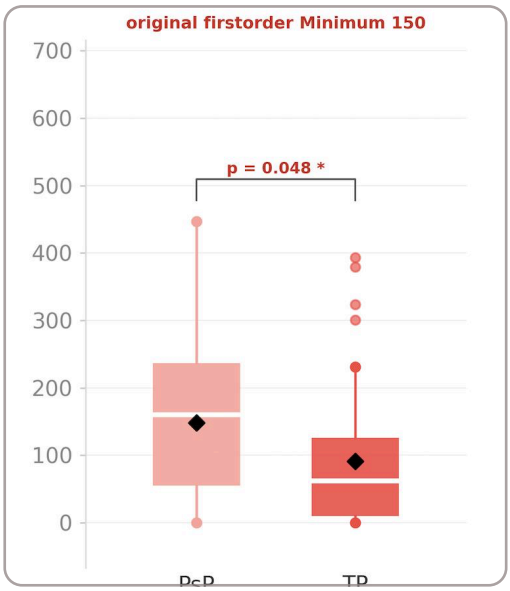
T1 POST Contrast + MGMT

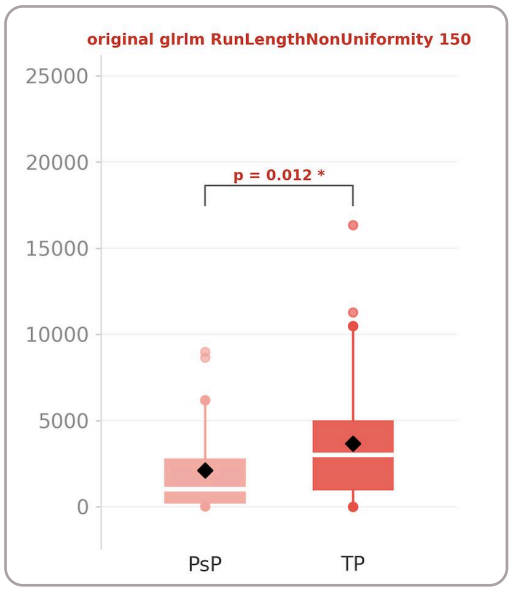












13% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text
- Cited Text

Match Groups

105 Not Cited or Quoted 13%
Matches with neither in-text citation nor quotation marks

0 Missing Quotations 0%
Matches that are still very similar to source material

0 Missing Citation 0%
Matches that have quotation marks, but no in-text citation

0 Cited and Quoted 0%
Matches with in-text citation present, but no quotation marks

Top Sources

9% Internet sources
9% Publications
7% Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.