



HAL
open science

Magnetic Resonance Imaging and Genomic Mutation in Diffuse Intrinsic Pontine Glioma: Machine Learning Approaches for a Comprehensive Analysis

Fahad Khalid

► **To cite this version:**

Fahad Khalid. Magnetic Resonance Imaging and Genomic Mutation in Diffuse Intrinsic Pontine Glioma: Machine Learning Approaches for a Comprehensive Analysis. Cancer. Université Paris-Saclay, 2024. English. NNT: 2024UPAST006 . tel-04573926

HAL Id: tel-04573926

<https://theses.hal.science/tel-04573926>

Submitted on 13 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Magnetic Resonance Imaging and Genomic Mutation in Diffuse Intrinsic Pontine Glioma: Machine Learning Approaches for a Comprehensive Analysis

*Imagerie par résonance magnétique et mutations
génomiques dans le gliome infiltrant du tronc
cérébral: méthodes d'apprentissage automatique pour
une analyse de données détaillée*

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°575 Electrical, Optical, Bio: Physics and Engineering (EOBE)
Spécialité de doctorat: Physique et imagerie médicale
Graduate School: Sciences de l'ingénierie et des systèmes
Réfèrent: Faculté des sciences d'Orsay

Thèse préparée dans le **Laboratoire d'Imagerie Translationnelle en Oncologie** (Institut Curie, Inserm) sous la direction de **Frédérique FROUIN**, chargée de recherche

Thèse soutenue à Paris-Saclay, le 30 Janvier 2024, par

Fahad KHALID

Composition du jury

Membres du jury avec voix délibérative

Claude COMTAT Directeur de recherche, CEA, Université Paris-Saclay	Président
Oscar ACOSTA TAMAYO Maître de conférences (HDR), Université de Rennes	Rapporteur & Examineur
Daniel RACOCEANU Professeur des universités, Sorbonne Université	Rapporteur & Examineur
Alain LALANDE Maître de Conférences - Praticien Hospitalier, Université de Bourgogne	Examineur
Diana MATEUS Professeur des Universités, Centrale Nantes	Examinatrice

Titre : Imagerie par résonance magnétique et mutations génomiques dans le gliome infiltrant du tronc cérébral : méthodes d'apprentissage automatique pour une analyse de données détaillée

Mots clés : radiomique, GIRC, apprentissage automatique, analyse de survie, IRM, traitement d'images médicales

Résumé : Le diagnostic du gliome infiltrant du tronc cérébral (GIRC) chez les enfants est l'un des plus éprouvants en oncologie pédiatrique. Malgré de nombreux essais cliniques explorant divers traitements, le pronostic reste sombre, la plupart des patients succombant entre 9 et 11 mois après le diagnostic. Les mutations génétiques clé associées au GIRC incluent H3K27M, ACVR1 et TP53. Chaque mutation a des caractéristiques distinctes, poussant les médecins à suggérer des thérapies personnalisées, soulignant l'importance d'une détection précise des mutations pour guider le traitement. Situées dans la région cruciale du tronc cérébral, les tumeurs GIRC présentent des risques significatifs liés à la biopsie en raison de potentiels dommages neurologiques. L'IRM est une méthode indispensable pour le diagnostic de ces tumeurs, évaluant leur extension et permettant de mesurer l'évolution de la maladie au cours de la thérapie. Une prédiction des mutations, combinée à l'identification des patients survivant plus de deux ans, pourrait améliorer la thérapie proposée à ces patients. Dans ce contexte, la radiomique transforme les images en vastes sources de données, extrayant des caractéristiques comme la forme et la texture pour aider à la prise de décision. L'objectif de cette thèse est de prédire les principales muta-

tions génétiques et d'identifier les survivants à long terme, en mettant l'accent sur la normalisation des images et l'applicabilité des modèles radiomiques. Notre étude a utilisé une base de données rétrospective de l'Institut Gustave Roussy, comprenant les données IRM de 80 patients et leurs données cliniques respectives. Les données d'IRM ont mis en évidence des problèmes pour les études radiomiques, tels que l'inhomogénéité du champ de biais et l'effet "scanner". Pour répondre à ces défis, un pipeline de normalisation d'images IRM a été mis en place, et les caractéristiques radiomiques ont été harmonisées par la méthode ComBat. Pour faire face au problème de modalités manquantes dans l'ensemble de données, une stratégie multi-modèles a été employée, conduisant à 16 modèles distincts reposant sur diverses combinaisons de caractéristiques radiomiques et cliniques. Cette approche a ensuite été rationalisée en une méthode multimodale, réduisant le nombre de modèles à cinq, après une phase de sélection de caractéristiques indépendantes. Les résultats de l'approche multimodale se sont avérés être prometteurs. Cette stratégie multimodale a été essentielle pour identifier les patients survivant plus de deux ans et a été complétée par l'approche ICARE pour une analyse de survie détaillée.

Title : Magnetic Resonance Imaging and Genomic Mutation in Diffuse Intrinsic Pontine Glioma : Machine Learning Approaches for a Comprehensive Analysis

Keywords : radiomics, DIPG, machine learning, survival analysis, MR, medical image processing

Abstract : The diagnosis of diffuse intrinsic pontine glioma (DIPG) in children stands as one of the most harrowing within pediatric oncology. Despite numerous clinical trials exploring various treatments, the prognosis remains bleak, with most patients succumbing between 9 to 11 months post-diagnosis. Key gene mutations linked to DIPG include H3K27M, ACVR1, and TP53. Each mutation has distinct characteristics, leading physicians to suggest tailored therapies, underscoring the importance of accurate mutation detection in guiding treatment. Located in the crucial region of the brainstem, the pons, DIPG tumors pose significant biopsy risks due to potential neurological damage. Hence, MRI could become a primordial diagnostic tool for these tumors, assessing their spread and gauging therapy responses. Its use to predict accurate gene mutation, and identify long-term survivors, could enhance patient care significantly. Within this framework, radiomics transforms images into vast data sources, extracting features like shape and texture to aid decision-making. The objective of this thesis is to refine mutation prediction and pinpoint long-term survivors, emphasizing image normalization and the applicability of radiomic models. Our study utilized a retrospective database from Gustave Roussy Institute, encompassing 80 patients MRI data and their respective clinical data. These MRI images highlighted issues in radiomic studies, such as bias field inhomogeneity and the "scanner effect". To address these challenges, a dedicated MR image normalization pipeline was implemented, and radiomic features underwent ComBat harmonization. Given the dataset's missing modalities, a multi-model strategy was employed, leading to 16 distinct models based on various radiomic and clinical feature combinations. This approach was then streamlined into a multi-modal method, reducing the number of models to five. The results from the ensemble of these models proved to be the most promising. This multi-modal strategy incorporated a feature selection phase, pinpointing the most pertinent features. Additionally, this method was applied to identify long-term survivors and was complemented by the ICARE framework for a nuanced survival analysis output.

Acknowledgements

First and foremost, I would like to express my deepest appreciation to Dr. Frederique Frouin for her exceptional support and guidance. In an era marked by unprecedented global upheaval and recurring lockdowns, her steadfast commitment and mentorship were invaluable. Dr. Frouin's dedication to sustaining our research endeavors, despite the swirling vortex of global uncertainties, truly showcased her unwavering passion and dedication to our collective scientific pursuit.

I am particularly moved to dedicate this work to the patients whose data formed the foundation of this research. Their contribution, made during challenging times in their own lives, has been invaluable in advancing our understanding and development of predictive models. It is to these individuals that I extend my deepest respect and gratitude. Their willingness to contribute to scientific advancement, despite their own personal battles, is a testament to the human spirit and the collective endeavor to improve healthcare outcomes for future generations. This thesis is not just a reflection of my efforts and those of my academic and support circles; it is, more importantly, a tribute to the patients whose experiences and data have enriched this research, making it both possible and meaningful.

The transition to remote work, characterized by virtual collaborations and the sudden shift from a bustling lab environment to the solitude of a home office, was significantly smoothed thanks to the incredible team at the Institute Curie LITO Unit. The seamless adaptation and continued spirit of cooperation exhibited by my colleagues, even when interfacing through digital screens, were nothing short of remarkable. I extend a special note of gratitude to Irène Buvat, the director of LITO, for her constant support and astute leadership, often extending beyond conventional hours to bridge different time zones.

The challenging period of lockdowns also highlighted the indispensable role of digital research communities. I am profoundly thankful for the wisdom and assistance provided by peers on platforms like Stack Overflow, GitHub, and ResearchGate. Their willingness to share knowledge and resources became particularly pivotal at times when traditional avenues of collaboration were hindered.

I also owe a tremendous debt of gratitude to my friends, from both within and beyond the academic sphere, for their unwavering support through these turbulent times. Our virtual gatherings and in-depth discussions about the world's evolving landscape have been a source of comfort and solace. In addition to the invaluable support from

my mentors and peers, the unwavering encouragement and motivation from my family have been the bedrock of my resilience and perseverance throughout this journey. Their boundless optimism, daily encouragement, and the comfort they provided during moments of doubt and stress have been instrumental in my ability to navigate the complexities of this research. The sanctuary of love and understanding they offered has been my haven, allowing me to recharge and tackle the challenges that came with conducting this thesis. Their belief in the importance of my work and their constant reassurance have been sources of immense strength, for which I am profoundly grateful.

Lastly, to my parents, whose determination to overcome the digital divide for video calls demonstrated their endless support and love, I am eternally thankful. Your resilience and optimism have not only been inspiring but have also infused me with the strength to persevere through these challenging times.

Synthèse en Français

Ce travail a été réalisé au Laboratoire d'Imagerie Translationnelle en Oncologie (LITO) de l'Institut Curie sous la direction de Frédérique Frouin (directrice de thèse) en collaboration avec l'INSERM et les données ont été collectées à l'Institut Gustave Roussy grâce notamment au Dr Jacques Grill, qui a recruté la cohorte de patients et a lancé l'étude. Les tumeurs cérébrales sont aujourd'hui le cancer le plus courant et la première cause de décès lié au cancer chez les enfants de moins de 15 ans. Les cancers infantiles rares représentent environ 185 nouveaux cas diagnostiqués chaque année en France, soit un taux d'incidence annuel moyen de 243 pour un million. Le gliome infiltrant du tronc cérébral (GITC) représente environ 75-80% des tumeurs pédiatriques du tronc cérébral. Le GITC, désormais considéré comme gliome de la ligne médiane, est une tumeur pédiatrique maligne avec une survie moyenne de 11 mois. La localisation de la tumeur rend l'ablation chirurgicale impossible, la radiothérapie est donc le traitement proposé, offrant une amélioration temporaire dans de nombreux cas. Des analyses génomiques basées sur des biopsies tumorales ont révélé que plus de 85% des patients atteints de DIPG présentent des mutations des gènes de l'histone H3, en particulier la substitution de la lysine 27 en méthionine (H3-K27M). Les deux variantes les plus courantes de l'altération H3-K27 sont H3.1 et H3.3. Ces deux mutations et le type non muté H3, se retrouvent associées à différents profils d'âge et de survie. Par exemple, les patients plus jeunes avec la mutation H3.1 montrent souvent une meilleure réponse à la radiothérapie et ont des survies plus longues. Ces mutations H3 K27M coexistent fréquemment avec des mutations somatiques des gènes TP53 et ACVR1. Alors que les mutations TP53 se retrouvent principalement chez les patients H3.3 et que les mutations ACVR1 se produisent principalement chez les patients H3.1, il est essentiel d'identifier séparément ces mutations lors de l'exploration de nouvelles options de chimiothérapie. Des recherches récentes suggèrent que la mutation TP53 pourrait contribuer à la radiorésistance chez les patients atteints de GITC, fournissant des informations précieuses pour adapter les stratégies de ré-irradiation. De plus, la combinaison de vandétanib et d'évérolimus pourrait offrir des bénéfices thérapeutiques pour les patients présentant des mutations ACVR1. Ces nouvelles avancées soulignent la nécessité de prédire la présence de mutations H3.1, ACVR1 et TP53 dans la tumeur à l'aide de données disponibles au moment du diagnostic, telles que les informations de base du patient (âge et sexe) et les résultats de l'IRM multimodale. Cette stratégie de prédiction aiderait à concevoir des traitements personnalisés, en particulier lorsque la biopsie cérébrale ne peut pas être réalisée ou ne donne pas de résultats concluants. Nous

émettons l'hypothèse que l'imagerie médicale pourrait offrir de précieuses informations qui permettraient de prédire la présence de mutations spécifiques chez les patients diagnostiqués avec un DIPG. Cette imagerie non invasive in vivo permet d'évaluer la totalité de la tumeur et de son micro-environnement à divers moments de la chronologie du traitement. Avec l'intégration de l'intelligence artificielle, le potentiel diagnostique de l'imagerie médicale a été considérablement amélioré, la transformant en une riche source de données. Cette nouvelle approche a le potentiel de révolutionner le pronostic pour les DIPG, car elle facilite la détection de mutations cruciales sans nécessiter de biopsie invasive ou d'opération chirurgicale, ouvrant ainsi la voie à des stratégies de traitement plus personnalisées et efficaces. Les prédictions radiogénomiques peuvent être améliorées en utilisant des données d'IRM multimodales régulièrement acquises, à condition que les étapes appropriées de prétraitement soient suivies. La radiomique, un domaine de recherche récent, permet une quantification extensive des phénotypes radiographiques, fournissant ainsi des informations précieuses sur les hétérogénéités tumorales. Cependant, ces caractéristiques radiomiques sont influencées par des facteurs tels que les paramètres d'acquisition et l'implémentation logicielle utilisée pour leur calcul. Pour stabiliser ces valeurs, un prétraitement approfondi est nécessaire. L'IRM, connue pour sa haute résolution spatiale et son contraste, est cruciale pour les tumeurs pédiatriques du système nerveux central. L'extraction de caractéristiques radiomiques cohérentes nécessite des procédures de standardisation, y compris l'utilisation de paramètres de séquence d'impulsion uniformes, de tailles de voxels identiques et de normalisation de l'intensité de l'image. Malgré la standardisation, les caractéristiques radiomiques peuvent varier en raison des différents imageurs, des bobines de gradient et des paramètres d'acquisition. Diverses stratégies, comme la normalisation par le Z-score et l'utilisation d'un tissu de référence pour la normalisation, ont été proposées pour réduire cette variabilité.

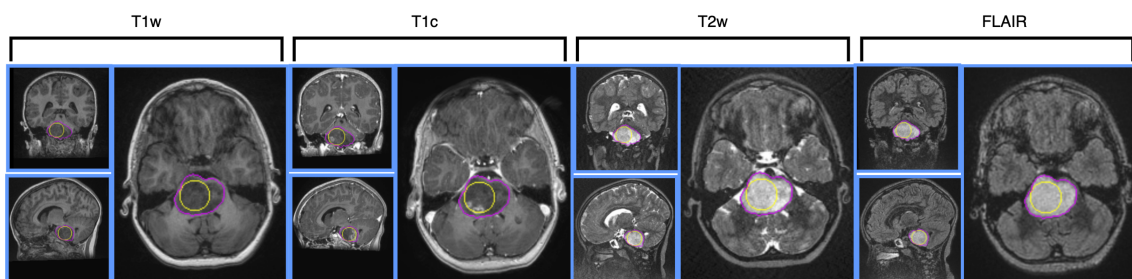


Figure 1: Illustration des données IRM d'un patient de 4 ans, porteur de mutations H3.1 et ACVR1, ne présentant pas de mutation TP53. Les données IRM sont affichées après normalisation de l'intensité à l'aide de la méthode hybride de la bande blanche. Du côté gauche vers le côté droit: images T1, T1c, T2 et FLAIR, en utilisant la disposition $\begin{array}{l} \text{Coronal} \\ \text{Sagittal} \end{array} \left| \text{Axial} \right.$ pour chacune modalité. Les contours de la sphère utilisée pour calculer les caractéristiques radiomiques d'intensité et de texture à l'intérieur de la tumeur sont indiqués en jaune. couleur sur chaque vue. Les contours de la tumeur utilisés pour calculer les caractéristiques de forme sont indiqués en rose.

Pour les patients atteints de DIPG, une procédure spécifique excluant du processus de normalisation les coupes du tronc cérébral porteuses de la tumeur a été suggérée. La méthode ComBat offre également un moyen d'harmoniser

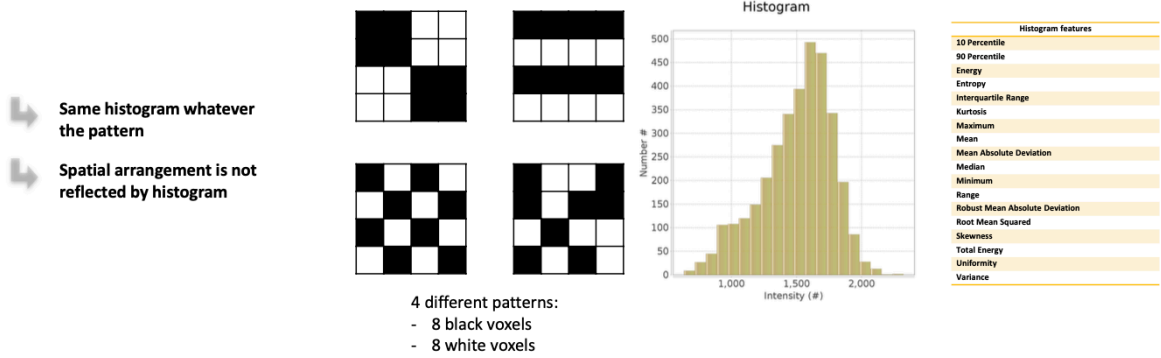
les caractéristiques radiomiques extraites sur différents scanners, réduisant ainsi la variabilité liée au scanner. Dans le Chapitre 1 de cette thèse, nous avons fourni un aperçu du scénario clinique que cette recherche aborde. Nous introduisons les caractéristiques cliniques, radiologiques et génétiques du Gliome Infiltrant du Tronc Cérébral (GITC), une tumeur pédiatrique, qui désormais est considéré comme une sous-catégorie du Gliome Diffus de la Ligne Médiane (GLM) par l'Organisation Mondiale de la Santé. Le chapitre décrit également les options thérapeutiques actuelles et les avancées qui lient les données radiologiques avec le pronostic du patient. De plus, il présente les séquences d'IRM classiquement utilisées (figure 1) pour l'étude du GITC et conclut avec une présentation des bases de données des patients atteints de GITC utilisées pour cette recherche, en introduisant les caractéristiques principales de notre ensemble de données, incluant 80 patients.

Le Chapitre 2 présente l'utilisation de la radiomique pour l'analyse des examens IRM réalisés dans notre cohorte. Ici, nous explorons l'utilisation d'indices prédéfinis tels que les indices de forme, ceux extraits de l'histogramme des intensités, ou ceux liés à la texture à partir des images. Ce processus est encapsulé en huit étapes nécessaires, représentées de manière schématique (la figure 2). Dans ce chapitre, nous passons en revue la standardisation des données IRM, le rôle de la radiomique notamment en imagerie oncologique et les différentes caractéristiques radiomiques. Nous abordons ensuite les principaux éléments d'un modèle d'apprentissage automatique en vue de réaliser des tâches de prédiction, et nous présentons une technique de sélection de caractéristiques basée sur une approche récursive. Finalement, nous introduisons différentes approches pour l'analyse de survie en proposant notamment une nouvelle approche ' Individual Coefficient Approximation for Risk Estimation (ICARE) ' qui nous a permis de remporter le grand challenge HECKTOR lors de la conférence MICCAI en 2022.

Le Chapitre 3 présente l'intérêt des caractéristiques radiomiques pour construire des modèles efficaces de prédiction de la mutation H3K27 de type H3.1. Dans ce chapitre, nous nous concentrons sur l'IRM multimodale (incluant les séquences pondérées T1 sans contraste et après injection de contraste, les séquences pondérées T2 et FLAIR), pour résoudre la question des modalités manquantes. En effet, sur les 80 patients atteints d'eGITC, 56 avaient les quatre modalités d'IRM tandis que les 24 autres avaient une ou plusieurs modalités manquantes. Nous avons proposé une méthode originale sans imputation pour relever ce défi. Seize modèles d'apprentissage automatique sont construits reposant sur les différentes combinaisons de caractéristiques cliniques et radiomiques extraites de chaque modalité d'IRM. En utilisant le cadre de validation croisée "leave one out", nous avons calculé le score F1 pondéré obtenu par chaque modèle pour prédire H3K27M, le modèle ayant le score le plus élevé étant le modèle 2 qui accepte les informations cliniques du patient ainsi que les caractéristiques T1w uniquement pour une valeur de F1 pondéré égale à 0.85. La variabilité présente dans l'ensemble de données en raison de sa nature multicentrique est également abordée dans ce chapitre. L'harmonisation ComBat a été adaptée, et les résultats ont montré un réalignement à travers les caractéristiques extraites des patients de différents centres, comme le montre

la figure 3. Le succès de l'approche à 16 modèles pour la prédiction de la mutation H3K27M a ouvert la voie à l'optimisation de cette approche et à la prédiction de la présence des mutations sur ACVR1 et TP53.

A Histogram or 1st-order features



B Textural or 2nd-order features

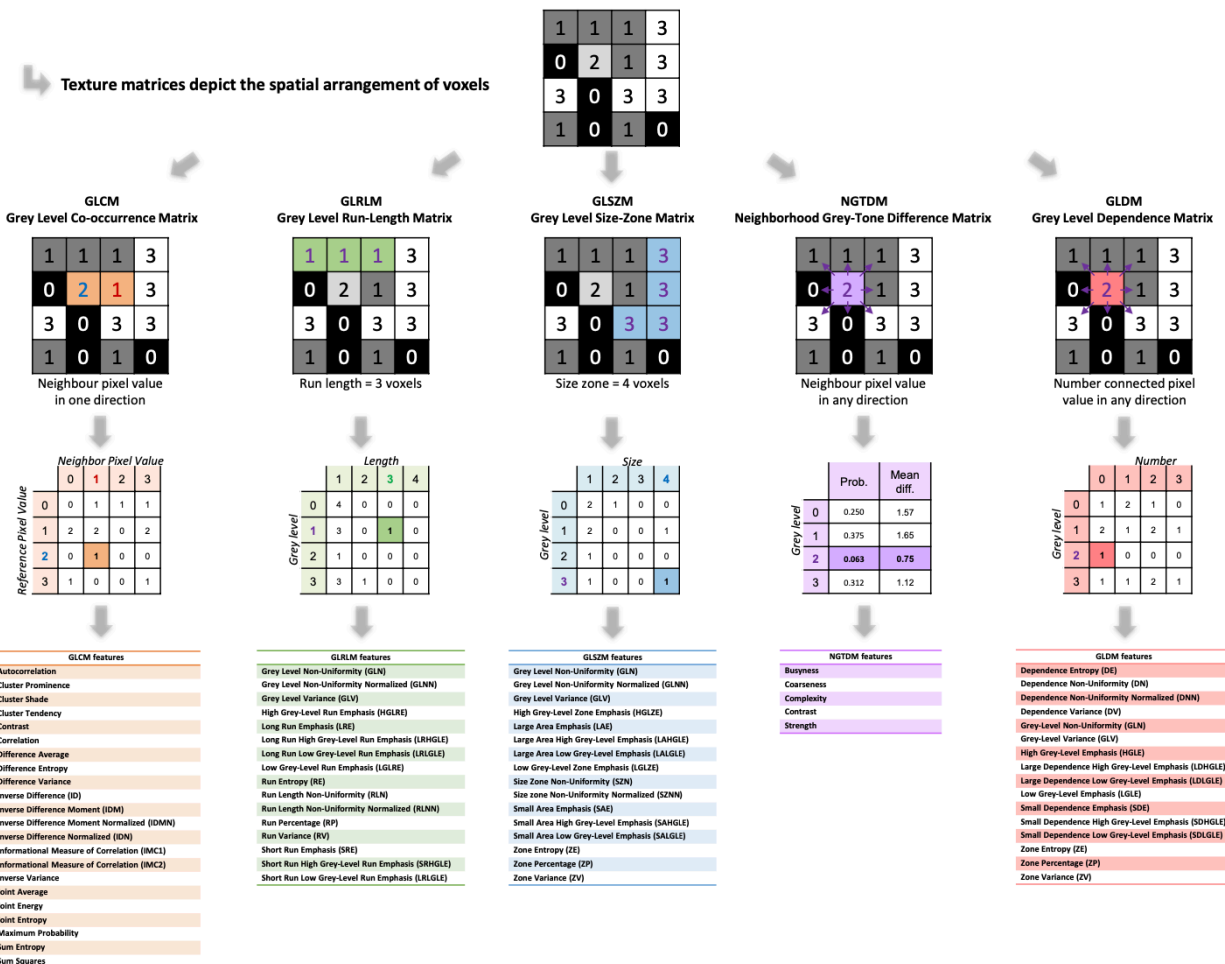


Figure 2: Représentation des types de caractéristiques radiomiques. (A) L'histogramme ou les caractéristiques de premier ordre reflètent uniquement la distribution de l'intensité des voxels. (B) Les caractéristiques texturales ou de second ordre dérivées de matrices de texture (par exemple, co-occurrence, longueur de course, taille-ton, différence, dépendance) reflètent l'arrangement spatial complexe et unique des voxels. Tiré de [1]

Le Chapitre 4 introduit un nouveau pipeline de prédiction qui s'appuie sur la radiomique et l'apprentissage automatique. Trois aspects sont particulièrement étudiés : la sélection des caractéristiques, la performance et l'interprétabilité. La sélection des caractéristiques implique d'identifier les caractéristiques les plus pertinentes pour le modèle de prédiction, tandis que la performance évalue l'exactitude et la robustesse du modèle. L'interprétabilité vise à fournir des éclairages sur la manière dont le modèle est parvenu à ses prédictions. En utilisant une approche "leave one out" pour avoir une prédiction pour chaque patient de l'ensemble de données, nous reconnaissons également les limites de notre ensemble de données, qui comporte des données manquantes, en conservant la validation croisée par leave-one-out de l'approche à 16 modèles. Sur le total combiné de 316 caractéristiques radiomiques, nous sélectionnons un maximum de 16 caractéristiques, quatre par modalité. La sélection des caractéristiques est réalisée par un algorithme d'élimination de caractéristiques récursif modifié pour ne retenir que les quatre caractéristiques les plus importantes si elles sont plus nombreuses. Au lieu de 16 modèles, nous développons cinq modèles monomodaux basés sur la régression logistique, un pour chacune des 4 modalités d'IRM et un modèle clinique + caractéristiques de forme combinant les données cliniques et les caractéristiques de forme extraites de la tumeur délimitée (un exemple est présenté dans la figure 4). Les prédictions des modèles monomodaux sont moyennées pour former le modèle multimodal. Les précisions équilibrées enregistrées pour la prédiction des mutations H3K27M (H3.1 vs H3.3, H3.2, type sauvage), ACVR1 et la mutation TP53 obtenues dans la validation croisée sont respectivement de 87,8 %, 82,1 % et 78,3 %. Le choix du classificateur et de la validation croisée par 'leave-one-out' a été testé en adaptant différents classifieurs tels que les machines à vecteurs de support et les forêts aléatoires dans des validations croisées 'leave-one-out', à 5 et à 10 plis. Enfin, les caractéristiques sélectionnées pour la tâche de prédiction ont également été étudiées. Le cadre final adapté est appelé l'approche multi-modèles avec validation croisée 'leave one out' pour les modalités d'imagerie manquantes. La suite du travail concerne l'analyse de survie globale pour les patients au sein de notre ensemble de données.

Au Chapitre 5, nous fournissons un examen exhaustif et détaillé de la littérature existante la plus pertinente, en nous focalisant sur les études menées sur les patients diagnostiqués avec un G1TC. Dans le G1TC et le gliome de la ligne médiane, l'analyse de survie est généralement effectuée en utilisant des données de temps jusqu'à l'événement, telles que le temps du diagnostic jusqu'à la mort. Les méthodes les plus courantes utilisées dans l'analyse de survie sont la méthode de Kaplan-Meier et le modèle de risques proportionnels de Cox. Sur la base de la littérature, nous effectuons une analyse de Kaplan-Meier sur les patients en utilisant les caractéristiques cliniques et génomiques. De plus, nous avons mis en œuvre l'approche multimodale du chapitre 4 pour prédire les patients survivant plus de 2 ans dans la cohorte. La littérature souligne le fait qu'un patient donné, s'il est prédit comme pouvant vivre longtemps au moment du diagnostic, pourrait bénéficier de stratégies thérapeutiques plus ciblées, qui peuvent procurer un allongement de la durée de vie. La sélection des caractéristiques adaptée pour la prédiction des patients survivant plus de deux ans a abouti à de nouvelles caractéristiques radiomiques et le modèle multi-

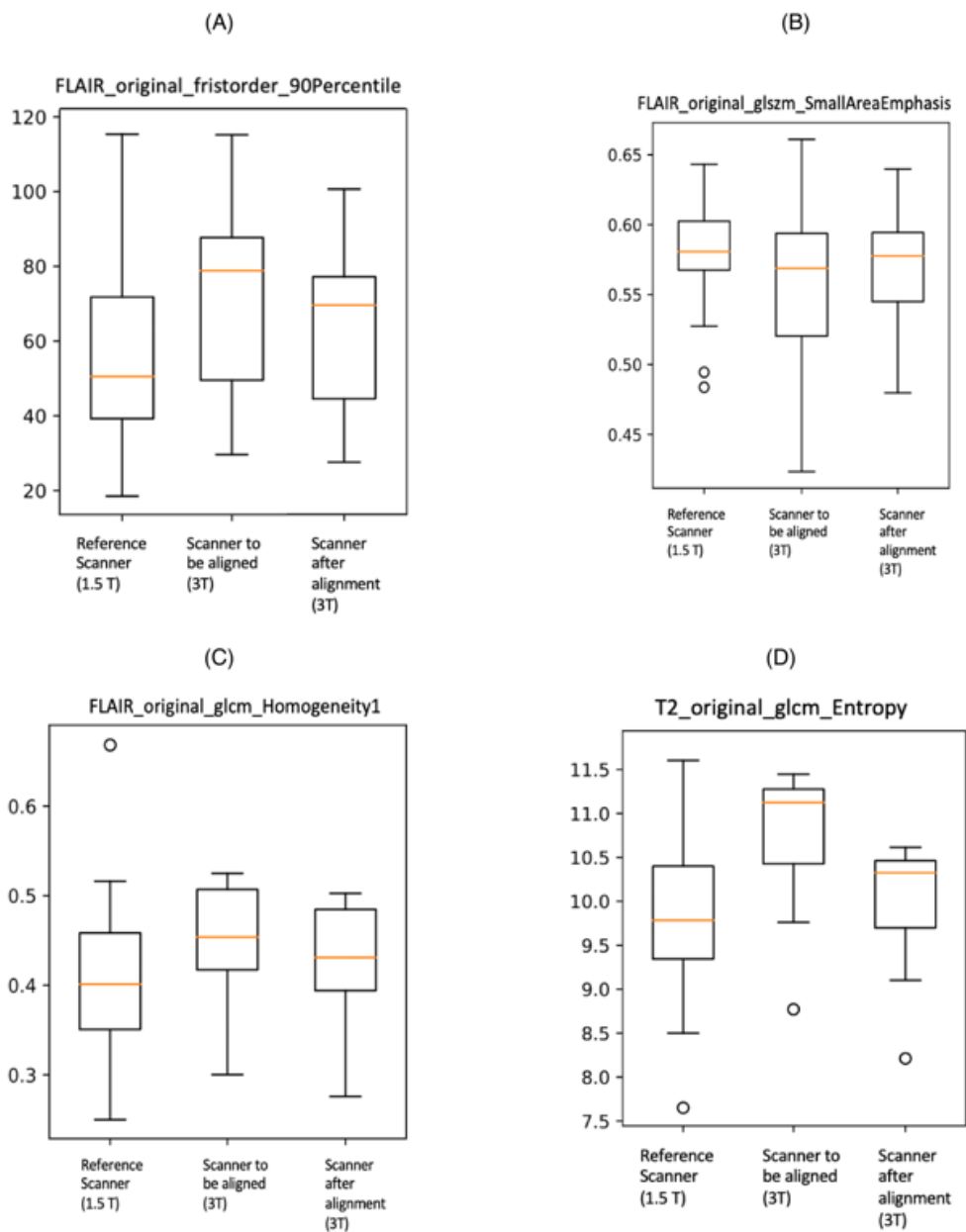


Figure 3: Boîtes à moustaches d'une caractéristique radiomique (FLAIR 90 Percentile, glszm Small Area Emphasis, glcm Homogénéité1 et glcm Entropy) pour les valeurs de caractéristiques acquises à l'aide du scanner 1,5T (Tesla) (à gauche), du scanner 3T à aligner (au milieu) et du 3T scanner après la procédure de réalignement à l'aide de ComBat (à droite).

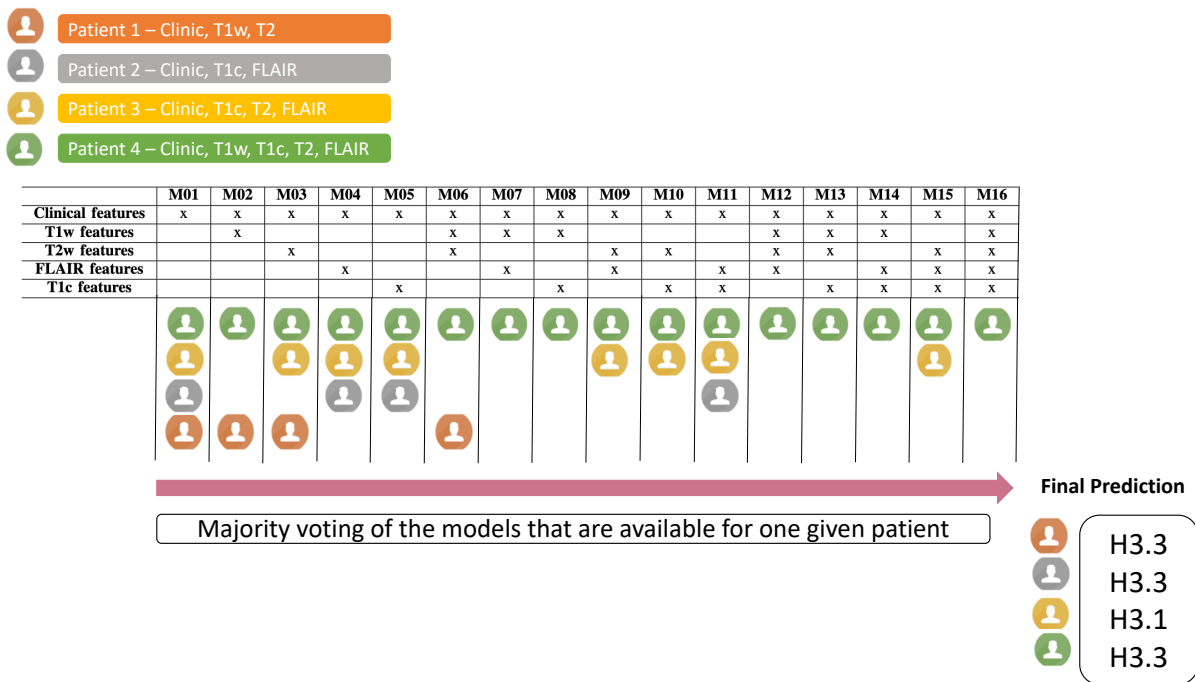


Figure 4: La figure illustre quatre patients hypothétiques évalués par l'approche à 16 modèles.

modal atteint une précision équilibrée de 89% dans le cadre de la validation croisée par "leave one out". L'approche multimodale s'est avérée efficace une fois de plus et supérieure aux approches monomodales considérées individuellement. Dans ce chapitre, nous testons également la prédiction de l'indice de concordance en utilisant le nouvel algorithme ICARE. Cet algorithme a gagné le challenge de prédiction de survie HECKTOR MICCAI 2022. La caractéristique la plus pertinente de cet algorithme est qu'il peut gérer les problèmes de modalités manquantes contrairement au modèle de Cox. Le modèle ICARE a été adapté avec une méthode de validation croisée 'leave pair out' pour rendre l'approche la plus comparable à celle retenue pour l'approche multi-modèles. ICARE nécessite un réglage des hyperparamètres pour des résultats optimaux. En résumé, cette thèse a offert d'importantes avancées méthodologiques dans la formulation d'un pipeline d'analyse complet pour créer des modèles radiomiques. Ces modèles, fondés sur l'imagerie par résonance magnétique, sont conçus pour être résilients dans la réalisation de prédictions à travers des données obtenues de deux institutions différentes, chacune ayant des imageurs IRM différents et des protocoles d'acquisition différents. Nous avons identifié des pistes dans la quête d'informations pertinentes pour prévoir certaines mutations génétiques sans nécessiter de biopsie et avons estimé les caractéristiques des patients survivant plus de deux ans. Les efforts futurs visent à solidifier ces résultats en utilisant des cohortes plus larges, tout en incluant les images de diffusion et/ou de perfusion, pour anticiper la réponse thérapeutique dans les gliomes infiltrant du tronc cérébral au stade le plus précoce possible. L'exploitation des images acquises pendant le traitement devra être réalisée pour mieux caractériser les répondeurs à la radiothérapie d'une part et aux nouvelles chimiothérapies d'autre part.

Contents

1 Clinical context	15
1.1 Neuro-oncology	15
1.2 Grading of brain tumors/ WHO classification	17
1.2.1 Division of diffuse gliomas into adult-type and pediatric- type	17
1.2.2 Pediatric-type low-grade and high-grade diffuse gliomas	17
1.3 Diffuse Midline Gliomas	19
1.3.1 Introduction	19
1.3.2 Genetic alterations associated with DMG/ DIPG	19
1.4 Medical imaging in neuro-oncology	21
1.4.1 Introduction	21
1.4.2 Conventional MRI	22
1.4.3 Advancements in MRI	22
1.5 The Dataset	26
2 Dedicated tools for classification and survival analysis based on radiomic features	29
2.1 Introduction	29
2.2 MRI standardization	30
2.3 Radiomics in medical imaging	33
2.4 Feature Selection	36
2.5 Learning models for medical data	37
2.5.1 Machine Learning	37
2.5.2 Deep learning	37
2.5.3 Loss Functions	43
2.6 Class Balancing	44
2.7 Classifiers	44
2.7.1 Logistic Regression	44

2.7.2 Support Vector Machine	45
2.7.3 Random Forest	46
2.8 Survival Analysis	47
2.8.1 Kaplan Meier curves and Log-rank tests	48
2.8.2 Concordance Index	48
2.8.3 ICARE algorithm	49
2.9 Conclusion	50
3 A first multi-model approach for H3K27M mutation prediction using radiomics and clinical features	51
3.1 Introduction	51
3.1.1 Radiomic studies in how to deal with missing data multiple scanners	52
3.1.2 How to deal with multi-scanner data	53
3.2 Paper published at the 2021 annual international conference of the IEEE in medicine and biology society	55
3.3 Conclusion	60
4 A second Multi-Model approach based on the complementarity between the different sources of data	61
4.1 Introduction	61
4.2 Paper published in Frontiers in Medicine	62
4.3 Additional validation studies	89
4.3.1 Classifier comparison	89
4.3.2 SVM, RF and LR evaluated in the Multi-Model approach	92
4.3.3 Comparing K-Fold cross validation with Leave one out cross validation	95
4.4 Will an Age dependent weight improve results of the multi-model approach?	96
4.5 Discussion	97
4.6 Conclusion	98
5 Radiomic analyses to predict overall survival and long survivors DIPG	99
5.1 Introduction	99
5.2 State of the Art: Survival analysis for DIPG	100
5.2.1 Meta analysis for high grade brain stem glioma	100
5.2.2 Use of MRI to predict survival in DIPG or DMG	100
5.2.3 Survival analysis based on FDG PET and MR	103
5.2.4 New therapeutic protocols: Onc201	104
5.2.5 Radiomics	105

5.3 Methods for assessing Overall Survival	107
5.3.1 Prediction of long survivors using the multi-model framework	107
5.3.2 Predicting Long Term Survivors	108
5.3.3 Survival Prediction continuous output ICARE	108
5.4 Results	109
5.4.1 Kaplan Meier curves and Log-rank tests for overall survival	109
5.4.2 Predicting long surviving patients (2 years) using the Multi-modal approach	113
5.4.3 Predicting long surviving patients (18 months) using the Multi-modal approach	118
5.4.4 Predicting long surviving patients (1 year) using the Multi-modal approach	121
5.4.5 Results for ICARE	123
5.5 Discussion	127
5.6 Conclusion	129
A Patient data	135
B Radiomic Features	139
Glossary	141

List of Figures

1	Illustration des données IRM d'un patient de 4 ans, porteur de mutations H3.1 et ACVR1, ne présentant pas de mutation TP53. Les données IRM sont affichées après normalisation de l'intensité à l'aide de la méthode hybride de la bande blanche. Du côté gauche vers le côté droit: images T1, T1c, T2 et FLAIR, en utilisant la disposition <table border="1" style="display: inline-table; vertical-align: middle;"><tr><td style="padding: 2px;">Coronal</td><td style="padding: 2px;">Axial</td></tr><tr><td style="padding: 2px;">Sagittal</td><td style="padding: 2px;"></td></tr></table> pour chacune modalité. Les contours de la sphère utilisée pour calculer les caractéristiques radiomiques d'intensité et de texture à l'intérieur de la tumeur sont indiqués en jaune. couleur sur chaque vue. Les contours de la tumeur utilisés pour calculer les caractéristiques de forme sont indiqués en rose.	Coronal	Axial	Sagittal		iv
Coronal	Axial					
Sagittal						
2	Représentation des types de caractéristiques radiomiques. (A) L'histogramme ou les caractéristiques de premier ordre reflètent uniquement la distribution de l'intensité des voxels. (B) Les caractéristiques texturales ou de second ordre dérivées de matrices de texture (par exemple, co-occurrence, longueur de course, taille-ton, différence, dépendance) reflètent l'arrangement spatial complexe et unique des voxels. Tiré de [1].	vi				
3	Boîtes à moustaches d'une caractéristique radiomique (FLAIR 90 Percentile, glszm Small Area Emphasis, glcm Homogénéité1 et glcm Entropy) pour les valeurs de caractéristiques acquises à l'aide du scanner 1,5T (Tesla) (à gauche), du scanner 3T à aligner (au milieu) et du 3T scanner après la procédure de réalignement à l'aide de ComBat (à droite).	viii				
4	La figure illustre quatre patients hypothétiques évalués par l'approche à 16 modèles.	ix				
1.1	Recognized tumor types in the 2021 WHO classification of Tumors of the Central Nervous System. Reprinted from (David N. Louis et al 2021). [5].	18				
1.2	Examples of the four most commonly used MR anatomical sequences: T1- weighted (T1), T1-weighted post-gadolinium contrast (T1c), T2- weighted (T2) and FLAIR fluid-attenuated inversion recovery. These images were taken from patient's scan with DIPG. A pink ring is placed with in the tumor region.	22				

1.3	Diffuse midline glioma H3K27Mmut.(A) Conventional MRI: T2 (axial section), contrast-enhanced T1 (axial section), FLAIR (coronal section).(B) PWI: dynamic susceptibility contrast (DSC) curve (red and orange: tumor ROIs; green: normal-appearing white matter (NAWM)), DSC-cerebral blood volume (CBV) (red: tumor ROI), arterial spin labeling (ASL)-cerebral blood flow (CBF). DSC ROI evaluation revealed “aggressive” perfusion features (red ROI: CBVmax, 4 mL/100 g; CBVmax/CBVNAWM, 6.48); ASL-CBF showed highly perfused spots within the tumor tissue.(C) DWI: a clinically feasible single-slice ROI evaluation revealed low diffusion parameters (blue ROI: ADCmean, 0.87 mm ² /s; ADCmean/ADCN AWM ratio, 1.19).(D) Q-ball tractography of corticospinal tracts, exhibiting a mild ventrolateral dislocation of the left tract due to the mass effect of the lesion. Reprinted from [45]	25
1.4	Example of MRS spectra for GBM and metastasis (typical short TE spectra). (A) Glioblastoma multiforme and (B) intracerebral metastases. Reprinted from [50]	26
2.1	Workflow of MR Intensity standardization. Compared to the hWS method (Shinohara et al 2014) [62], the adaptations needed to analyze DIPG scans are shown in purple. Reprinted from [61]	31
2.2	Histograms of intensities in WM and GM before and after standardization for T1, T2, post injection T1 (T1c) and FLAIR scans. Each blue line corresponds to a pre-biopsy scan and each pink line to a post-biopsy scan. For each type of MR scan, the mean Hellinger distance (HD) and its standard deviation between each pair of histograms are shown. Reprinted from [61]	32
2.3	Representation of radiomic feature types. (A) Histogram or first-order features reflect voxel intensity distribution only. (B) Textural or second-order features derived from texture matrices (eg, co-occurrence, run length, size-tone, difference, dependence) reflect the complex and unique spatial arrangement of voxels. Reprinted from [1]	35
2.4	Results on PubMed using the keyword search “Machine Learning” by year up to April 2023 , showing the rapid growth of interest in machine learning.	38
2.5	Examples of underfitting and overfitting using least squares regression. In (a) we have fit a straight line to the points, which slightly underfits the data. Adding another parameter, (b) is a quadratic curve, which fits the data well. (c) is an 8 th degree polynomial fit. Although it fits these data points well (the training data), it will probably not generalise well to new data points and is therefore overfitting the data.	40
2.6	Converge rates of several optimisation algorithms using a multilayer neural network on a handwritten classification challenge. Taken from [151]	42
3.1	The figure illustrates four hypothetical patients evaluated by the 16 model approach.	54

4.1	Top (FLAIR features) and Bottom (T1 features) figures illustrates the performance of four classifiers for predicting the presence of ACVR1 mutation. The y-axis displays the accuracy achieved in (%) the x-axis present the 79 radiomic features extracted from the modality. The blue colored line represents the accuracies achieved by Decision tree classifier and how the addition of features impacts the accuracy achieved. Similarly the orange colored line for Ada-boost, green for Logistic Regression and red for random forest. For visual assistance a dashed line is drawn at 75%.	90
4.2	Top (T1c features) and Bottom (T2 features) figures illustrate the performance of four classifiers for predicting the presence of ACVR1 mutation. The y-axis displays the accuracy achieved in (%) the x-axis present the 79 radiomic features extracted from the modality. The blue colored line represents the accuracies achieved by Decision tree classifier and how the addition of features impacts the accuracy achieved. Similarly the orange colored line for Ada-boost, green for Logistic Regression and red for random forest. For visual assistance a dashed line is drawn at 75%.	91
5.1	Kaplan Meier curves for patients stratified by Sex of patient	110
5.2	Kaplan Meier curves for patients with known H3.1 mutation status	111
5.3	Kaplan Meier curves for patients with known ACVR1 mutation status	112
5.4	Kaplan Meier curves for patients with known TP53 mutation status	113
5.5	The box plots represent the distribution of 20 selected feature for long term survivors. The feature values are scaled between 0 and 100. The blue box plot represent the feature values of patients who are not long survivors marked as '0' and the orange box represents the feature distribution for long term survivors marked as '1'. The x-axis denotes the feature IDs from table 5.2. The y-axis denotes the feature values scaled between 0 and 100.	114
5.6	Pearson correlation heat-maps between the features that have been selected by the five different models to predict long surviving patients	116
5.7	Kaplan Meier estimation of long term surviving patients.	117
5.8	Correlation heatmap between the features that have been selected by the five different models to predict 18 months survival. Feature identifiers (on the right side) of identical features found by the different predictive tasks are shown in color.	119
5.9	Correlation heatmap between the features that have been selected by the five different models to predict 1 year survival. Feature identifiers (on the right side) of identical features found by the different predictive tasks are shown in color.	122
5.10	Importance of the clinical, shape and MR radiomic features. A positive value (red) shows a positive correlation with the risk and a negative value (blue) is a negative correlation.	124
5.11	Correlation heatmaps between the selected features by ICARE models to predict long surviving patients	126

5.12 Histogram distribution displaying overall survival of patients with DIPG 128

List of Tables

1.1	Role of imaging techniques in brain tumors. Reprinted from [26]	23
1.2	Imaging techniques and their major utilities in brain tumor imaging. Reprinted from [27]	24
1.3	Summary of characteristics of patients with DIPG in the combined cohort (PREBIOMEDE and BIOMEDE).	
	¹ Includes 7 patients with CENSORED OS	27
1.4	Patient modalities available (T1, T1c, T2 and FLAIR) with the scanner field strength	28
2.1	List of the main publications based on Radiomics according to their application and the imaging modality studied. Modified from Yep et al., [134] and Desseroit et al., [135]	36
4.1	Prediction results for TP53 for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework using Logistic Regression classifier. For each prediction task and for each model three figures of merit are reported: sensitivity, specificity and balanced accuracy.	93
4.2	Prediction results for TP53 for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework using Support Vector Machine classifier. For each prediction task and for each model three figures of merit are reported: sensitivity, specificity and balanced accuracy.	93
4.3	Prediction results for TP53 for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework using Random Forest classifier. For each prediction task and for each model three figures of merit are reported: sensitivity, specificity and balanced accuracy.	93
4.4	Prediction results for H3.1 for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework using Logistic Regression classifier. For each prediction task for each model three figures of merit are reported: sensitivity, specificity and balanced accuracy.	94
4.5	Prediction results for H3.1 for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework using Support Vector Machine classifier. For each prediction task for each model three figures of merit are reported: sensitivity, specificity and balanced accuracy.	94
4.6	Prediction results for H3.1 for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework using Random Forest classifier. For each prediction task for each model three figures of merit are reported: sensitivity, specificity and balanced accuracy.	94

4.7 Prediction results for TP53 using features from T1w modality only using LOO-CV and 5 and 10 fold CV using Logistic regression. For each prediction frame work three figures of merit are reported: sensitivity, specificity balanced accuracy.	95
4.8 Prediction results for TP53 using features from T1w modality only using LOO-CV and 5 and 10 fold CV using SVM. For each prediction frame work three figures of merit are reported: sensitivity, specificity balanced accuracy.	95
4.9 Prediction results for TP53 using features from T1w modality only using LOO-CV and 5 and 10 fold CV using Random forest. For each prediction frame work three figures of merit are reported: sensitivity, specificity balanced accuracy.	95
4.10 Prediction results for the seven models: M_{Clinic} , M_{T1} , M_{T1c} , M_{T2} , M_{FLAIR} , M_{Multi} and M_{MAge} in a LOO-CV framework. For each prediction task and for each model, four figures of merit are reported: the total number of patients for which the prediction was possible (the number of patients with mutation is between brackets), the sensitivity, the specificity and the balanced accuracy. For each figure of merit, best results are in bold characters and second best results are underlined.	97
5.1 The table lists all papers included in the literature review sorted into different categories based on the type of study conducted in chronological order. The selected papers are categorized into six categories. Twenty-one papers review were based on clinical studies whereas two papers included radiomic studies.	101
5.2 Subsets of features selected by the five different models $M_{Clinic+Shape}$, M_{T1} , M_{T1c} , M_{T2} , and M_{FLAIR} to predict 2 years long survivors. $M_{Clinic+Shape}$ contains the combined features from clinic and shape data.	114
5.3 The prediction result of long term survivors using the five mono modality models and two multi-modal models, combined: $M_{Genomic}$, $M_{ClinicSh}$, M_{T1} , M_{T1c} , M_{T2} , M_{FLAIR} , M_{Multi} and M_{MG} in a LOO-CV framework. For each prediction task and for each model, four figures of merit are reported: the total number of patients for which the long term survival was possible, the sensitivity, the specificity and the balanced accuracy. For each figure of merit, best results are in bold characters and second best results are underlined	117
5.4 Subsets of features selected by the five different models $M_{Clinic+Shape}$, M_{T1} , M_{T1c} , M_{T2} , and M_{FLAIR} to predict 18 months survival. $M_{Clinic+Shape}$ contains the combined features from the model clinic and shape.	118

5.5	The prediction result for the Six models: $M_{ClinicSh}$, M_{T1} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework. For each prediction task and for each model, four figures of merit are reported: the total number of patients for which the long term survival was possible, the sensitivity, the specificity and the balanced accuracy. For each figure of merit, best results are in bold characters and second best results are underlined.	120
5.6	Subsets of features selected by the five different models $M_{Clinic+Shape}$, M_{T1} , M_{T1c} , M_{T2} , and M_{FLAIR} to predict 1 year survival. $M_{Clinic+Shape}$ contains the combined features from the model clinic and shape.	121
5.7	The prediction result for the eight models: $M_{Genomic}$, $M_{ClinicSh}$, M_{T1} , M_{T1c} , M_{T2} , M_{FLAIR} , M_{Multi} and M_{MG} in a LOO-CV framework. For each prediction task and for each model, four figures of merit are reported: the total number of patients for which the long term survival of 1 year was possible, the sensitivity, the specificity and the balanced accuracy. For each figure of merit, best results are in bold characters and second best results are underlined.	123
5.8	The three hyperparameters used to achieve the highest concordance index of 0.578	123
5.9	The table presents the genomic, clinical, shape and MR radiomic features selected by ICARE binary weighted model.	125
A.1	The table presents in detail patients age, sex, volume and genomics mutation for H3, ACVR1 and TP53 along with the modalities available and the OSstatus. Sex: 1 boy, 0 girl, mutation if present is represented by '1' if not present represented by '0' and 'na' if the mutation status is not known to be present or absent. Age: is given in years,	137
B.1	Texture features by feature type Full mathematical details of the features can be found at 238	140

Introduction

Diffuse Intrinsic Pontine Glioma (DIPG) is a devastating pediatric brain tumor, primarily afflicting children. This grade 4 tumor stands out for its aggressive nature and is notorious for its grim prognosis. Despite its lethal nature, biopsies for DIPG are traditionally avoided. This is due to the tumor's location in the pons, a crucial part of the brainstem, where any invasive procedure poses significant risks, including potential neurological damage.

Some therapeutic approaches have shown a longer survival, but the choice of therapy depends on the type of mutation associated with the tumor. Diffuse Intrinsic Pontine Glioma (DIPG) is characterized by several genetic mutations, with the H3K27M mutation being the most prevalent, leading to a specific change in the histone H3 gene. Additionally, ACVR1 mutations are observed in a subset of DIPG cases, especially in conjunction with the H3.1K27M mutation. Furthermore, TP53 mutations frequently co-occur with the H3K27M mutation, adding to the complex genetic landscape of this tumor. Timely detection of these mutations plays a vital role in treatment planning. Given the limitations of biopsies, Magnetic Resonance Imaging (MRI) becomes an indispensable tool in diagnosing and assessing DIPG. Medical imaging serves as a vital supplement to biopsies, providing a non-invasive and repeatable method to evaluate lesions comprehensively, including their surrounding microenvironment. With advancements in artificial intelligence, the capabilities of medical imaging have expanded, viewing images as rich data reservoirs suitable for tasks like predictive analysis. Radiomics, a burgeoning domain, operates on the principle that a tumor's morphology and variance reveal insights about its inherent biological traits. Through the extraction of features, such as shape, intensity, or texture, radiomics taps into the concealed or challenging-to-measure biological data within images.

In the age of technological advancements, the field of MR radiomics emerges as a beacon of hope. Radiomics, in essence, transforms MR images into a treasure trove of data. This data, rich in details, goes beyond mere visual interpretation. It dives deep into the morphological nuances, intensity patterns, and texture features of the tumor. These intricate details are believed to harbor insights about the tumor's biological properties, aggressiveness, and potential response to treatments.

Harnessing the power of radiomics, researchers aim to decode this hidden information in the MR images. By doing so, they aspire to develop predictive models that can aid in treatment decisions, potentially offering a glimmer of hope in the otherwise bleak landscape of DIPG. The ability to accurately predict tumor behavior and response could revolutionize patient care, guiding treatment protocols, and offering more tailored therapeutic interventions for these young patients.

Undertaking quantitative evaluations on MRI presents multiple complexities. Notably, there's the challenge of local inconsistencies in tissues due to bias field amplification and the task of contrasting MR intensities across varied images, given that MR details aren't typically articulated in uniform measures. Additionally, radiomic attributes derived from medical scans are profoundly swayed by data collection parameters, such as the types of scanners or sequences. This phenomenon, often termed as "Scanner effect", complicates the application of radiomic models crafted from images of one diagnostic facility to datasets from different establishments.

In recent years, many MRI-based radiomic studies investigated mutation prediction and overall survival in patients with DIPG, sometimes limiting the models to specific modalities or clinical features, to handle the missing modalities issue.

The focus of this thesis is thus to improve the prediction of DIPG mutations and overall survival of patients while tackling missing modality issues and comparing them with the models found in the literature, with a particular focus on the standardization of images and radiomic features and the exportability of radiomic models.

This work was conducted at the "Laboratoire d'Imagerie Translationnelle en Oncologie" (LITO) under the supervision of Dr. Frédérique Frouin, in collaboration with INSERM and Gustave Roussy with a special thanks to Dr. Jacques Grill, who recruited the cohort of patients and launched the study.

This thesis consists of five chapters.

Chapter 1 introduces DIPG and DMG (Diffuse midline glioma) with respect to neuro-oncology. It provides the 2021 World Health Organization (WHO) brain tumor grading for tumors, highlighting the main changes introduced. Furthermore, the most common mutations of DIPG are presented. Medical imaging and its modalities are presented with a special focus on the advancements in MR technology. Finally, the cohort of 80 patients is introduced, with its main clinical features (age, sex and volume) along with the mutation status of the three mutations under study (H3.1, ACVR1, and TP53). The number of MR modalities present for each patient is also indicated, highlighting the missing modality issue associated with the cohort.

Chapter 2 focuses on MR standardization techniques and the main pipeline adapted for the MR scans used in this thesis work. Radiomics and its main characteristics are presented in detail. Machine learning model-building tools are briefly presented which are used later in this thesis for developing prediction models for mutation and survival.

Chapter 3 introduces a multi-model approach developed for mutation prediction and presents results for the prediction of H3.1 mutation. Radiomic features extracted from MR scanners having different field strengths (1.5 and 3 Tesla) are realigned using ComBat harmonization. The construction of 16 models is explained considering the missing modality issue and a voting mechanism is introduced to have the multi-model prediction result. This chapter presents the paper published at the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) [2].

Chapter 4 presents an updated multi-model approach from Chapter 3 called the Multi-modal approach in leave-one-out cross-validation with missing imaging modalities (LOO-CV-MIM). This approach involves the construction of four radiomic models based on each of the four MR modalities and a fifth model of clinical and shape features. A recursive feature elimination with cross-validation (RFE-CV) approach for feature selection is also introduced aiming to further define the relevant information that can be found in MR images and improve prediction results. This framework is tested to perform the prediction of all three mutations H3.1, ACVR1, and TP53. The Chapter also presents a paper published in Frontier in Medicine [3].

Chapter 5 presents the work done on survival analysis for the patients in the cohort. Kaplan-Meier analysis using clinical and known mutation status of patients to identify patient survival using this data. The prediction of long-term survivors (2 years) using the Multi-modal framework from Chapter 4 for the identification of patients who may be good candidates for testing targeted therapies for prolongation of life. Furthermore, a new algorithm called ICARE for continuous survival output of patients is tested, this algorithm can adapt to missing data. This chapter also presents a comprehensive literature review of studies investigating survival analysis of patients with DIPG in recent years using radiomic and clinical information.

Conclusions and plans for future work, including new methodological developments, are finally exposed.

Chapter 1

Clinical context

This chapter is an introduction to Diffuse Midline glioma (DMG)/ Diffuse intrinsic pontine glioma (DIPG). It presents the characteristics of this pediatric tumor in clinic, radiology and genetics, placing them in the classifications proposed by the World Health Organization (WHO). The chapter gives a brief state of the art of the treatments offered at present, as well as the state of the art of discoveries making it possible to associate radiological information with the prognosis of patients. This chapter also takes stock of the studies which come close to radiomics within the framework of the DIPG. Finally, the databases that we used for this thesis work are presented.

1.1 Neuro-oncology

The discipline of neuro-oncology is a vast and swiftly advancing domain, incorporating elements of neurology, neurosurgery, medical oncology, radiation oncology, neuroradiology, neuropathology, cancer rehabilitation, and palliative care. Its emphasis lies in the identification and treatment of both primary and metastatic malignancies within the central nervous system (CNS), as well as addressing complications arising from systemic cancers or their respective treatments.

Primary tumors of the CNS refer to abnormal growths that occur within the brain or spinal cord. These tumors can be benign (non-cancerous) or malignant (cancerous) and can originate from various cell types within the CNS. Common types of CNS tumors are classified based on their cell of origin and behavior.

1. Gliomas are the most common type of brain tumors. Astrocytic, oligodendroglial, and ependymal gliomas are the most common cell types associated with glioma are primary brain tumors which don't usually spread outside of the brain or spine but range from low-grade to high-grade in terms of aggressiveness. Treatment options for brain gliomas include surgery, radiation therapy, and chemotherapy. The prognosis of individuals with brain gliomas depends on various factors, including the grade of the tumor and the patient's overall health.

2. Meningiomas are a type of brain tumor that originate from the meninges, which are the protective covering of the brain and spinal cord. These tumors are considered to be benign, or non-cancerous, and typically grow slowly. They are more common in women than men and are usually diagnosed in people over the age of 50. Symptoms of meningiomas can vary depending on the location of the tumor, but may include headaches, seizures, visual changes, and weakness or numbness in the limbs. Diagnosis is usually made through imaging studies such as **CT** or **MRI** scans. Treatment options for meningiomas include surgery, radiation therapy, and sometimes a combination of both. Surgery is the most common treatment and is usually the most successful in removing the tumor and relieving symptoms. In some cases, doctors may choose to closely monitor the tumor instead of treating it immediately, especially if it is not causing symptoms or growing rapidly. Overall, meningiomas tend to have a good prognosis, with a high chance of successful treatment and long-term survival. However, it is important to note that the location and size of the tumor can play a significant role in the outcome, and recurrent meningiomas can occur in some cases.
3. Schwannomas, also known as acoustic neuromas, are benign, slow-growing tumors that originate from the Schwann cells of the cranial nerve VIII, which is responsible for balance and hearing. They typically occur in adults between the ages of 30 and 60 and are more common in people with a genetic disorder called neurofibromatosis type 2. Symptoms include hearing loss, tinnitus, balance problems, and facial weakness. Large tumors can also cause headaches, vision problems, and difficulty with speech. Imaging studies such as CT or MRI scans are usually used to diagnose the tumor. Treatment options include surgery, radiation therapy, and a combination of both. Surgery is the most common treatment and is usually successful in removing the tumor and relieving symptoms. The outcome for people with schwannomas is generally good with a high chance of successful treatment and long-term survival, but location and size of the tumor can affect the outcome, and recurrent tumors can occur. In addition, surgical removal of the tumor can result in loss of hearing and facial weakness if the tumor has grown on the nerve that controls the facial movement and hearing.
4. Malignant, rapidly growing brain tumors called medulloblastomas often affect children but can also affect adults. These tumors are categorized as embryonal tumors, indicating their beginning in fetal or embryonic cells, and they develop in the cerebellum, the lower rear region of the brain in charge of balance and coordination. Despite being aggressive in nature, many people with medulloblastomas have a good prognosis if they receive prompt diagnosis and treatment. Chemotherapy, radiation therapy, and surgery are frequently used in conjunction as treatment techniques.
5. Central Nervous System (CNS) lymphomas are a rare subset of non-Hodgkin lymphomas that arise within the brain, spinal cord, or surrounding meninges. Unlike systemic lymphomas, which affect the lymph system, CNS lymphomas are confined to the neurological environment. They can be particularly aggressive, and their

onset is often rapid, presenting symptoms like altered mental functions, focal neurological deficits, or seizures. Treatment typically involves high-dose methotrexate-based chemotherapy, sometimes followed by whole-brain radiation therapy. The prognosis varies based on the patient's age, overall health, and the extent of the disease at the time of diagnosis.

Treatment options for CNS tumors vary depending on the type and stage of the tumor, as well as the patient's overall health. Surgery, radiation therapy, and chemotherapy are all common treatments for CNS tumors. Additionally, targeted therapy and immunotherapy are also being studied as potential treatment options for certain types of brain tumors.

1.2 Grading of brain tumors/ WHO classification

The fifth edition of the **WHO** Classification of Tumors of the Central Nervous System published in 2021 follows the previously published versions from 1979, 1993, 2000, 2007, and 2016 (14). It incorporates numerous advancements in the field that transpired post the 2016 classification, and also takes into account the advice from the Consortium to Inform Molecular and Practical Approaches to CNS Tumor Taxonomy (**CIMPACT-NOW**). The classification emphasizes the significance of molecular diagnostics in classifying CNS tumors, all the while still recognizing and utilizing other traditional tumor characterization methods, such as histology and immunohistochemistry.


1.2.1 Division of diffuse gliomas into adult-type and pediatric- type

Significantly, the WHO CNS5 acknowledges the clinical and molecular variations existing between diffuse gliomas predominantly found in adults (referred to as "adult-type") and those primarily appearing in children (referred to as "pediatric-type"). It is crucial to emphasize the word "primarily" used here, as pediatric-type tumors occasionally manifest in adults, especially those in their youth, while adult-type tumors are sporadically seen in children. However, splitting the classification into adult-type and pediatric-type diffuse gliomas serves as a progressive move in distinctly separating these prognostically and biologically different tumor groups. This separation has been contemplated for an extended period, and the discovery of molecular disparities now makes it achievable. It is anticipated that this distinction will foster enhanced care for children and adults diagnosed with CNS tumors.

1.2.2 Pediatric-type low-grade and high-grade diffuse gliomas

The categorization has been updated to include two novel groupings of pediatric-type of diffuse gliomas low-grade and high-grade. The low-grade category is comprised of four distinct entities, all exhibiting diffusive growth in the brain. However, these entities often present with less distinctive and occasionally overlapping histological traits. In

Newly Recognized Tumor Types
Diffuse astrocytoma, <i>MYB</i> - or <i>MYBL1</i> -altered
Polymorphous low-grade neuroepithelial tumor of the young
Diffuse low-grade glioma, MAPK pathway-altered
Diffuse hemispheric glioma, H3 G34-mutant
Diffuse pediatric-type high-grade glioma, H3-wildtype and IDH-wildtype
Infant-type hemispheric glioma
High-grade astrocytoma with piloid features
<i>Diffuse glioneuronal tumor with oligodendroglioma-like features and nuclear clusters</i> (provisional type)
Myxoid glioneuronal tumor
Multinodular and vacuolating neuronal tumor
Supratentorial ependymoma, <i>YAP1</i> fusion-positive
Posterior fossa ependymoma, group PFA
Posterior fossa ependymoma, group PFB
Spinal ependymoma, <i>MYCN</i> -amplified
<i>Cribriform neuroepithelial tumor</i> (provisional type)
CNS neuroblastoma, <i>FOXR2</i> -activated
CNS tumor with <i>BCOR</i> internal tandem duplication
Desmoplastic myxoid tumor of the pineal region, <i>SMARCB1</i> -mutant
<i>Intracranial mesenchymal tumor, FET-CREB fusion positive</i> (provisional type)
<i>CIC</i> -rearranged sarcoma
Primary intracranial sarcoma, <i>DICER1</i> -mutant
Pituitary blastoma

Figure 1.1: Recognized tumor types in the 2021 WHO classification of Tumors of the Central Nervous System. Reprinted from (David N. Louis et al 2021). 

such cases, molecular investigations are critical in specifying the type of lesion. The four subtypes identified include Diffuse astrocytoma with *MYB* or *MYBL1* alterations; Angiocentric glioma; Polymorphous low-grade neuroepithelial tumor of the young (**PLNTY**); and Diffuse low-grade glioma with MAPK pathway alterations. The last subtype includes tumors displaying either astrocytic or oligodendroglial morphologies. Precise classification of these tumors, like most other glioma types, necessitates molecular characterization and the combination of histopathological and molecular data in a layered diagnostic approach. Uncovering the specific molecular characteristics in turn lays the groundwork for more targeted treatment strategies. The high-grade grouping also includes four types: Diffuse midline glioma, marked by **H3K27M** alterations; Diffuse hemispheric glioma, **H3** G34 mutant; Diffuse pediatric-type high-grade glioma, H3-wildtype and **IDH**-wildtype; and Infant-type hemispheric glioma. The Diffuse midline glioma with H3 K27 alterations was previously classified in 2016, while in the CNS5, the remaining three are newly acknowledged types. The Diffuse pediatric-type high-grade glioma, which is defined as H3-wildtype and IDH-wildtype, is identified as possessing the wildtype for both H3 and IDH gene clusters. Similar to many other CNS tumor variants, this subtype calls for molecular examination and the confluence of histopathological and molecular findings for accurate diagnostics. The Infant-type hemispheric glioma stands out as a unique high-grade glioma, predominantly

affecting newborns and infants and bearing a distinctive molecular footprint

1.3 Diffuse Midline Gliomas

1.3.1 Introduction

Diffuse midline gliomas (DMGs) are a type of brain tumor that typically occur in the midline of the brain, which includes the thalamus, brainstem, and spinal cord. They are considered to be malignant and are often challenging to treat. On imaging, such as CT or MRI scans, diffuse midline gliomas usually appear as a mass that has spread in the middle area of the brain, which most often forms in the pons in the brainstem, thalamus, spinal cord, and cerebellum. It's uncommon for these tumors to occur in other areas of the CNS. The tumors are named, in part, based on the locations where these tumors most often occur. Therefore, tumors in other brain or spine locations thought to be diffuse midline glioma should be reviewed by neuro-oncology providers with experience in treating people with these tumors. Fragments from biopsy-exeresis (open surgery) or fine-needle micro biopsies performed after identification are used to determine the "integrated diagnosis" [6]. Since DMG can be hard to tell them apart, they require molecular testing. Some diffuse midline gliomas have changes in histone-related genes, the most common is H3K27M. Review by a neuropathologist is recommended to confirm this diagnosis. The tumor shows degrees of brightness when contrasted with normal brain tissue. The diagnosis of diffuse midline glioma cannot be made solely based on imaging alone because the tumors can resemble other high-grade brain cancers such as glioblastomas. They can also mimic other tumors that occur in the same area such as lymphomas, metastasis and other midline tumors.

Fragments from biopsy-exeresis (open surgery) or fine-needle micro biopsies performed after identification are used to determine the "integrated diagnosis" [6]. These samples include information that explains the patient's clinical state, including the location of the anomaly physically and its radiological properties, such as contrast changes and border analyses. Therefore, a biopsy is usually required to confirm the diagnosis and to differentiate it from other tumors. This addition to imaging, molecular profiling and genetic testing can be used to confirm a diagnosis of diffuse midline glioma and to help guide treatment decisions. Due to their location and the difficulty of treating these tumors, diffuse midline gliomas often have a poor prognosis. Due to the aggressive nature of these tumors and the challenges in treating them, new treatment options such as targeted therapy and immunotherapy are being researched to improve the outcomes for individuals with diffuse midline gliomas.

1.3.2 Genetic alterations associated with DMG/ DIPG

Brain tumors called diffuse midline gliomas (DMG), which now include diffuse intrinsic pontine gliomas (DIPG), may have distinctive genetic mutations. The H3 K27M mutation in the histone H3 gene is one example of such a change.

DMGs and DIPGs also frequently have mutations in the **TP53**, **ACVR1**, and **PIK3CA** genes. According to Wu et al. (2014) [7], TP53 mutations, which are present in around 30–50% of cases, interfere with this tumor suppressor gene's normal operation and cause uncontrolled cell development. The H3 K27M mutation frequently co-occurs with ACVR1 alterations in DIPG, such as the ACVR1 R206H mutation [8]. Additionally, some DIPG cases have PIK3CA mutations, which impact a crucial protein in a cell signaling pathway [9].

H3 K27M

The H3 K27M mutation, is a mutation in which the lysine (K) at position 27 in the histone H3 protein is changed to a methionine (M). This mutation is linked to the development and spread of DIPG and has a significant effect on the control of gene expression [10]. Interestingly, the WHO classification of brain tumors includes this mutation as a distinguishing criterion for the diagnosis of DIPG because it is present in more than 70% of DIPGs. Furthermore, it has been demonstrated that DIPG patients with the H3 K27M mutation had a worse prognosis than patients without the mutation [11]. The histone H3 gene exists in several variants, and the mutation can be found in different isoforms, namely H3.1, H3.3, and H3.2.

H3K27M Mutations in Different Isoforms:

1. **H3.3K27M:** This is the most common mutation associated with DIPG and other pediatric high-grade gliomas. Tumors with this mutation often show a specific loss of H3K27 trimethylation, leading to altered gene expression patterns and promoting oncogenesis [10].
2. **H3.1K27M:** This variant is less frequent than H3.3K27M, the H3.1K27M mutation has been found in some pediatric gliomas. This mutation occurs in the HIST1H3B gene, which encodes the H3.1 variant [12].
3. **H3.2K27M:** This mutation is relatively rare, with limited reports in the literature. Like the other two mutations, it is associated with a decrease in H3K27 trimethylation, but its specific clinical implications and distribution in tumors require further research [13].

IDH1 and IDH2 gene mutations

Two variants of the enzyme isocitrate dehydrogenase (IDH) are encoded by the genes IDH1 and IDH2, depending on their mutational status [14]. These mutations result in hypermethylation that results in the production of alpha-ketoglutarate. According to present knowledge, this trait that inhibits tumor suppressor genes promotes glioma formation. Because cancers are the result of IDH1 or IDH2 gene mutations, they are collectively referred to as "IDH-mutant" tumors. Nearly all grade II and grade III gliomas (87%) and the glioblastomas that develop from them (85%) have these mutations [15]. The prognosis for grades II, III, and IV gliomas is excellent when IDH mutant status is detected [16].

TP53

The TP53 gene, which typically serves as a tumor suppressor, is frequently discovered to be altered. The regulatory mechanisms that typically stop unchecked cell proliferation are disabled by TP53 mutations, which aid in the development and spread of malignancies. In between 30-50% of patients of DIPG, the TP53 gene has been found to have mutations [7]. The genetic complexity of these tumors is highlighted by the identification of the co-occurrence of TP53 mutations and the histone H3 K27M mutation, a defining trait of DIPG [17]. It is critical to comprehend the function of TP53 mutations in DIPG since doing so could aid in the creation of more focused and efficient treatments.

ACVR1

Pediatric brain tumors with ACVR1 gene mutations are a particularly aggressive variety of tumors. A protein in the bone morphogenetic protein (BMP) pathway, which is crucial for cell development and differentiation, is encoded for by the ACVR1 gene. Mutations in ACVR1 result in hyperactivity, which promotes unchecked cell proliferation and aids in the development of tumors. The histone H3 K27M mutation frequently co-occurs with ACVR1 in DIPG [8]. Contrarily, DIPG cases with TP53 gene mutations often do not have ACVR1 mutations, suggesting that various DIPG subtypes may have unique genetic profiles [18]. In order to create specialized treatment plans for this aggressive disease, research into the role of ACVR1 mutations in DIPG is still ongoing.

1.4 Medical imaging in neuro-oncology

1.4.1 Introduction

Medical imaging stands as a crucial pillar in managing brain tumors, (see table [1.1]). It is used for a variety of tasks, such as diagnosis, predicting illness progression, and directing treatment strategies. For instance this entails finding the exact place for biopsies, figuring out the extent of the surgical resection, and defining and quantifying the radiation dosage used in radiation therapy. Imaging is essential for assessing the effectiveness of the provided therapy and estimating the volume of tumor tissue that hasn't been completely removed post-treatment. It serves as a surveillance tool to monitor tumor evolution during the follow-up phase and to distinguish between resurgent tumor growth and tissue changes brought on by therapies like radiation therapy, such as radiation necrosis. Different types of complementary imaging modalities (CT, PET and MRI) may be used depending on the characterization objectives (see table [1.2]). CT scans are useful in pinpointing the presence of brain cancers, facilitating the planning of radiation therapy, and revealing instances of hemorrhage or edema in the brain. Positron Emission Tomography (PET) provides an avenue for assessing brain tumors, distinguished by a heightened cell proliferation rate compared to normal tissue, through the use of specialized tracers. These tracers signify the absorption of amino

acids by brain cells and include 11C-methionine, 18F-fluorothymidine (FLT), 18F-fluoro-ethyl-tyrosine (FET), and 18F-dihydroxyphenylalanine (DOPA) [19], [20], [21].

1.4.2 Conventional MRI

Conventional MRI is a non-invasive imaging modality that has emerged as an indispensable tool in clinical diagnostics and research, particularly in the realm of neurological, musculoskeletal, and cardiovascular systems [22]. Unlike X-rays, MRI leverages the principles of nuclear magnetic resonance to visualize internal structures, relying on the inherent magnetic properties of atomic nuclei within the human body, predominantly hydrogen protons [23]. The conventional MRI protocols typically encompass T1-weighted (T1w) and T2-weighted (T2w) imaging sequences. T1w images provide excellent anatomical detail, highlighting fat tissues as bright signals. Conversely, T2w images are adept at visualizing fluids, rendering cerebrospinal fluid in the brain as bright, which is crucial for detecting pathologies like edema or tumors [24]. One of the advantages of MRI over other imaging modalities is its ability to produce images in any plane without physical repositioning, a feature that is invaluable in neuroimaging. Moreover, the absence of ionizing radiation in MRI makes it safer for repeated studies, a factor that is critical in longitudinal research or in monitoring disease progression [25]. In the landscape of medical imaging, while newer MRI modalities continue to evolve, conventional MRI remains foundational, providing pivotal anatomical insights that guide clinical decisions and research explorations.

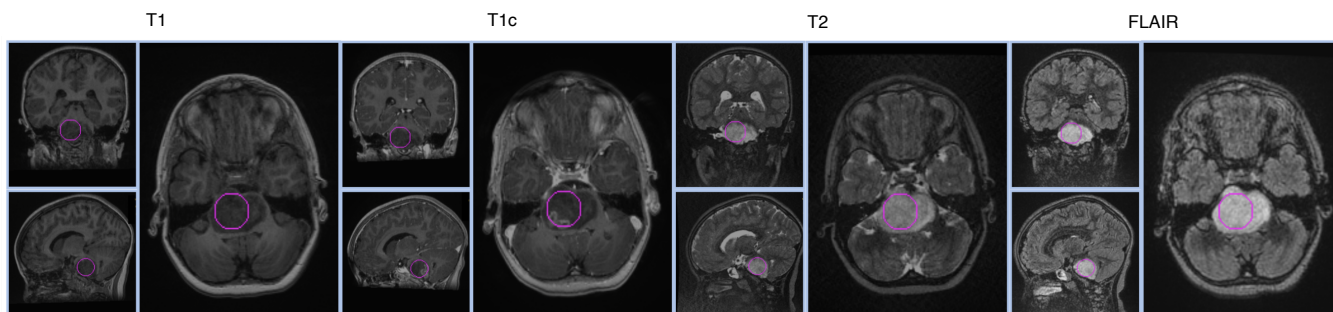


Figure 1.2: Examples of the four most commonly used MR anatomical sequences: T1-weighted (T1), T1-weighted post-gadolinium contrast (T1c), T2-weighted (T2) and FLAIR fluid-attenuated inversion recovery. These images were taken from patient's scan with DIPG. A pink ring is placed within the tumor region.

1.4.3 Advancements in MRI

In order to enhance the capabilities of standard MRI, novel magnetic resonance methods have been developed in neuroradiology [26], [28], [29]. These methods, which include Diffusion-Weighted Imaging (DWI), Perfusion-Weighted Imaging (PWI), Diffusion Tensor Imaging (DTI), and Magnetic Resonance Spectroscopy (MRS). These techniques provide information on the metabolic composition (MRS), white matter infiltration (DTI), tissue perfusion and permeability (PWI), and tumor cellularity (DWI).

Role of Radiology Imaging Techniques		
Preoperatively	Detection	
	Characterization	localization
		size
		margins
		extension
		midline shift
		compression
		contrast enhancement
		vascularity
		supplying vessels
		perifocal oedema
	Differentiation	benign vs malignant
Staging		
tumour embolization		
surgical planing		
Intraoperative	surgical navigation	
Postoperatively	monitoring the effect of treatment	
	exclude recurrence	
	distinguishing recurrent tumour from radiation necrosis	

Table 1.1: Role of imaging techniques in brain tumors. Reprinted from [26]

Perfusion-weighted imaging

An advanced magnetic resonance imaging (MRI) method called perfusion-weighted imaging (PWI) measures the rate at which a contrast agent is administered intravenously passes through the capillaries in the brain to assess tissue blood flow [30]. The dynamics of this process make it possible to extract hemodynamic variables including mean transit time (MTT), cerebral blood flow (CBF), and cerebral blood volume (CBV), which offer substantial insights into tissue microvascular characteristics [30]. It is essential to comprehend these circulatory characteristics in the context of brain tumors. The vascular characteristics of these tumors frequently align with their degree of malignancy and propensity for proliferation. PWI has made a substantial contribution to the ability to differentiate between different tumor grades based on their perfusion profiles and to discriminate tumor recurrence from post-therapeutic changes [31]. The importance of PWI in treatment planning and monitoring is becoming more established as the field of neuro-oncology increasingly embraces customized therapy. Its capacity to provide thorough assessments solidifies its status as a vital tool in the diagnosis and treatment of brain tumors [32].

Arterial Spin Labeling

Arterial Spin Labeling (ASL) is a non-invasive MRI technique used to quantify cerebral blood flow (CBF). ASL uses the water in arterial blood as an endogenous tracer, in contrast to other perfusion imaging techniques that rely on the introduction of exogenous contrast agents. Blood magnetization in the main arteries supplying the brain is changed by ASL using magnetically tagged blood water protons. An picture is taken after a predetermined post-labeling delay, during which time tagged blood water protons enter the imaging zone. To create a perfusion-weighted

Imaging technique	Major utility in brain tumor imaging
CT	Mass effect, herniation, hydrocephalus, hemorrhage, calcifications
Pre and post-contrast T1	Enhancement characteristics, necrosis, extent of the enhancing portion of the tumor
T2/T2 FLAIR	Peri-tumoral edema (vasogenic and infiltrative), non-enhancing tumor
T2* susceptibility sequence (SWI)	Blood products, calcifications, radiation induced chronic micro-hemorrhages
DWI/ADC	Reduced in highly cellular portions of tumor, post-operative injury
DTI	Tractography for surgical planning/navigation
Perfusion (generally DSC)	Tumor/tissue vascularity
MR spectroscopy	Metabolic profile
fMRI	Pre-operative functional mapping, research into treatment effects
PET/MR	Potential new radiotracers

Note. ADC, apparent diffusion coefficient; CT, computed tomography; DSC, dynamic susceptibility contrast-enhanced; DTI, diffusion tensor imaging; DWI, diffusion weighted imaging; FLAIR, fluid attenuated inversion recovery; fMRI, functional magnetic resonance imaging; PET, positron emission tomography; SWI, susceptibility weighted imaging.

Table 1.2: Imaging techniques and their major utilities in brain tumor imaging. Reprinted from [27].

picture, this labeled image is subsequently subtracted from a control image that does not undergo labeling [33]. ASL is especially well suited for repeated measurements, longitudinal investigations, and pediatric or fragile populations where contrast administration may be prohibited since it may quantify cerebral perfusion without using exogenous contrast agents [34]. ASL methods have improved over time, including pulsed, continuous, and pseudo-continuous ASL. This has increased its sensitivity and reliability, allowing it to be used in a variety of clinical and scientific situations. These applications include functional brain mapping and the evaluation of neurodegenerative illnesses, tumor perfusion, and cerebrovascular diseases ([35]).

Diffusion weighted imaging (DWI)

The theory is centered on the spontaneous movement of water molecules across different tissues. This "Brownian motion" hinges on tissue cellularity, cell membrane integrity, and vascularization extent. These factors, when increased, lead to restricted molecular diffusion, assessed by calculating the Apparent Diffusion Coefficient (ADC) [36]. These coefficients can be utilized to generate a voxel-level mapping (ADC map), revealing the water diffusion properties of brain tissue. ADC value variations allow for the delineation of morphological features such as edema, necrosis, and viable tumor tissue [37]. Furthermore, it aids in identifying tumor infiltrated regions that other MRI sequences might not disclose [38]. This technique offers an increased certainty in distinguishing brain abscesses from cystic or necrotic brain tumors compared to traditional MRI [39]. In adults with diffuse gliomas, ADC values show a correlation with the IDH mutation status and overall survival [40]. IDH wild-type gliomas, compared to IDH-mutant gliomas, exhibit lower ADC values and a shorter overall survival period.

Diffusion Tensor imaging (DTI)

Diffusion Tensor Imaging (DTI) is an advanced MRI technique that captures the intricate movement of water molecules within tissues, revealing the structural organization of white matter tracts in the brain [41]. DTI reveals the microstructural traits and coherence of white matter networks by quantifying water diffusion patterns [42]. DTI has become a crucial technique for understanding the effects of brain malignancies, identifying changes or displacements brought on by the tumor in these tracts [43]. This knowledge enables medical professionals to differentiate between different tumor subtypes based on their diffusion characteristics and to plan surgical or therapeutic procedures that protect vital brain pathways [44].

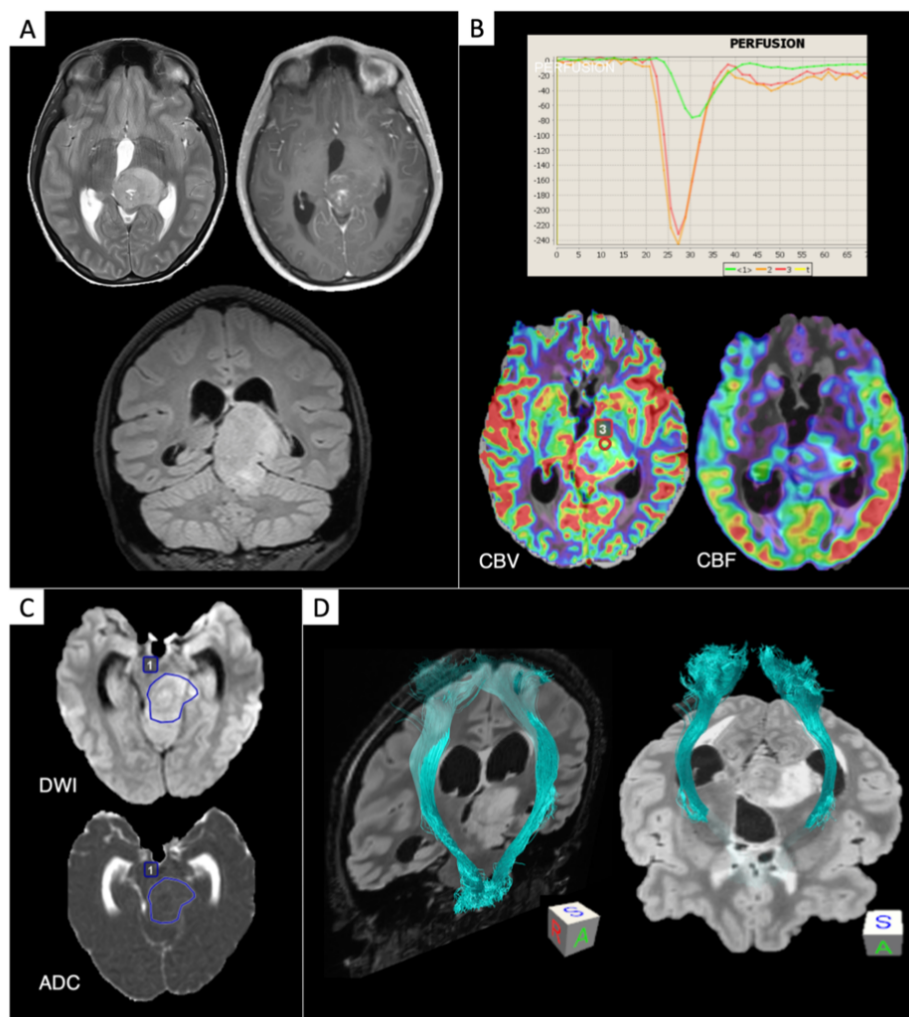


Figure 1.3: Diffuse midline glioma H3K27Mmut. **(A)** Conventional MRI: T2 (axial section), contrast-enhanced T1 (axial section), FLAIR (coronal section). **(B)** PWI: dynamic susceptibility contrast (DSC) curve (red and orange: tumor ROIs; green: normal-appearing white matter (NAWM)), DSC-cerebral blood volume (CBV) (red: tumor ROI), arterial spin labeling (ASL)-cerebral blood flow (CBF). DSC ROI evaluation revealed “aggressive” perfusion features (red ROI: CBVmax, 4 mL/100 g; CBVmax/CBVNAWM, 6.48); ASL-CBF showed highly perfused spots within the tumor tissue. **(C)** DWI: a clinically feasible single-slice ROI evaluation revealed low diffusion parameters (blue ROI: ADCmean, $0.87 \text{ mm}^2/\text{s}$; ADCmean/ADCN AWM ratio, 1.19). **(D)** Q-ball tractography of corticospinal tracts, exhibiting a mild ventrolateral dislocation of the left tract due to the mass effect of the lesion. Reprinted from [45]

Magnetic resonance spectroscopy (MRS)

Magnetic resonance spectroscopy (MRS) is a technique used to analyze the metabolic characteristics of tissues by detecting signals emitted by atomic nuclei within molecules. These signals are identified based on variations in proton resonance frequencies. In the context of tumor pathology, there are six metabolites that hold significant importance [46, 47]. These include N-Acetyl Aspartate (NAA), which serves as a neuronal marker; creatine (Cr), indicating overall metabolism and cellularity; choline (Cho), which increases in cell proliferation scenarios; myoinositol (MI), a sugar specific to glial cells; free lipids (Lip), indicating cellular necrosis; and lactate (Lac), which accumulates due to altered metabolism caused by increased glycolysis. Various metabolic ratios have been identified as relevant for tumor classification and predicting malignancy [48, 49]. Figure 1.4 provides an example of short Time-Echo (TE) spectra from glioblastoma multiforme and intracerebral metastases, illustrating typical patterns observed.

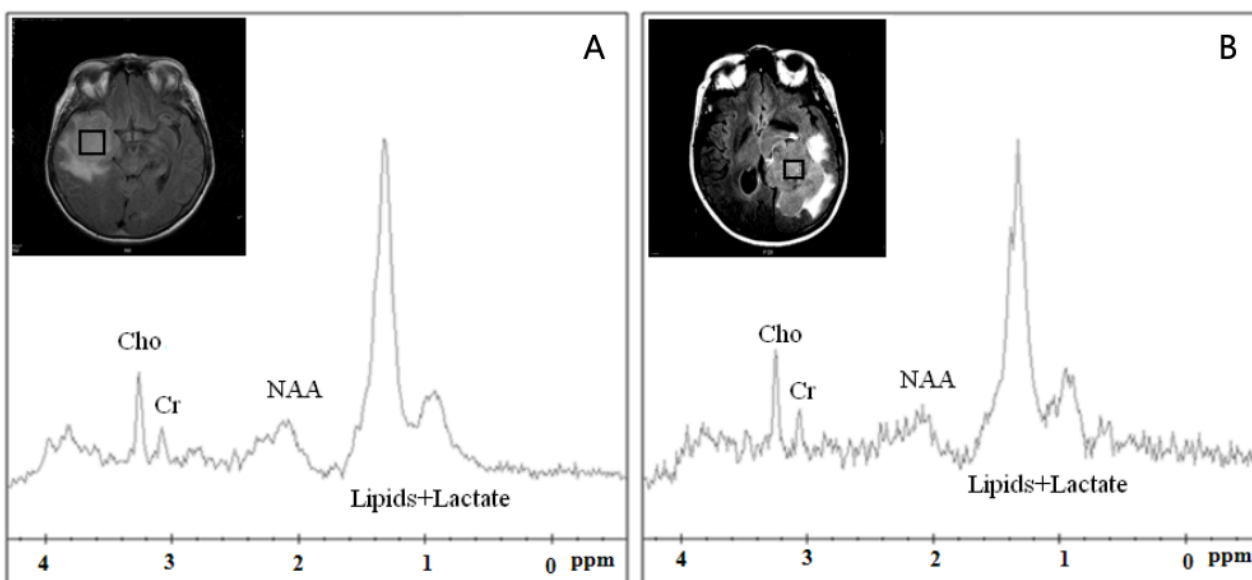


Figure 1.4: Example of MRS spectra for GBM and metastasis (typical short TE spectra). (A) Glioblastoma multiforme and (B) intracerebral metastases. Reprinted from [50]

1.5 The Dataset

In this Ph.D., we focus on two types of data. The first type is imaging data, which was collected using conventional MRI scans and then analyzed using radiomic features. The second type is biological data, which was obtained from biopsy samples and from which we could gather genetic information. In addition clinical data sample were use (age and sex). Our main interest lies in pre-biopsy detection of some somatic mutation and overall survival [OS] through radiomic features. In this thesis, we utilized two cohorts: PREBIOMEDE and a portion of the BIOMEDE study. The patients' families provided consent to use the clinical and radiological data gathered during the treatment process

for research purposes. These consents were obtained following the guidelines of the Institutional Review Board (IRB) at Gustave Roussy.

The images were originally obtained in the Digital Imaging and Communications in Medicine (DICOM) format. For the purpose of this thesis, these images were consistently anonymized and converted into the Neuroimaging Informatics Technology Initiative (NIfTI) format.

Genomic characteristics of the cohort

The final cohort consisting of 80 patients with at least 1 of the 4 MR modalities (T1,T1c,T2 and FLAIR) available along with mutation status of H3K27M , ACVR1 and TP53 for most patients. In the case of Histone (H3), mutation information is available for H3.1, H3.3, H3.2 along with not mutated wild type (WT). For ACVR1 and TP53 mutated or not mutated. In this thesis we worked with the combined sets of patients. The pooled characteristics of the final cohort are presented in table 1.3 and individual patient information is displayed in clinical trial the patient data table present in appendix A.1.

The PREBIOMEDE database concerns patients diagnosed at Necker hospital between the years 2009 and 2014. The database called BIOMEDE [51], [52] contains some of the patients of the National Cancer 2011 BIOMEDE. The full study is a multi-centric study spanning the years 2014 and 2019. The differences between PREBIOMEDE and BIOMEDE are related to therapies. For BIOMEDE, radiotherapy is combined with three different targeted therapies: dasatinib, erlotinib and everolimus. The majority of patients underwent an imaging protocol that included T1w , T2w ,T1c and FLAIR images at the time of diagnosis. A tumor sample was taken by biopsy to determine the genetic mutations associated with the tumor. For some patients, the biopsy was not possible or did not bring conclusive results.

Number of Patients	80
Boys / Girls	45 / 35
Average age at Diagnosis(yr)	8.05 ±4.39
H3.1 / H3.3 / H3.2 / WT / UnK	14/ 44/ 1/ 4/ 17
ACVR1 / WT / UnK	14/ 49 /17
TP53 / WT / UnK	34/ 27 /19
Average Overall Survival (OS) ¹ in days	388.96 ±271.56

Table 1.3: Summary of characteristics of patients with DIPG in the combined cohort (PREBIOMEDE and BIOMEDE).
¹ Includes 7 patients with CENSORED OS

Imaging characteristics of the cohort

The table 1.4 presents the imaging modalities available for the total number of patients on the dataset. The acquisition scanner used for obtaining the scans is also known. A key point to highlight and to take into account several patients scans have missing modalities. Out of the 80 patients 47 patients have all four modalities.

	1.5T	3.0T	Missing
Patients	65	15	
T1	60	13	7
T1c	58	14	8
T2	60	14	6
Flair	56	11	13

Table 1.4: Patient modalities available (T1, T1c, T2 and FLAIR) with the scanner field strength

Chapter 2

Dedicated tools for classification and survival analysis based on radiomic features

2.1 Introduction

In the preceding chapter, we introduced Diffuse Intrinsic Pontine Glioma (DIPG) within the realm of Neuro-oncology, showed its classification under the updated (2021) World Health Organization's grading of pediatric tumors. Based on the characteristics of the data set that was in this Ph.D., we introduce different tools that will be used for testing this dataset for both classification and survival analysis. Binary classification will be used to predict genomic mutations from MRI acquired at diagnosis time. Survival analysis will be based on the combination of MRI data and genomic data. In this chapter, we will focus to the MRI standardization, the role of Radiomics in MR imaging. We will delve into the primary constituents of a machine learning model, specifically adopting a wrapper-based feature selection technique, and ultimately, we introduce a different tool for survival analysis the [HECKTOR](#) [MICCAI](#) [\[53\]](#) challenge-winning classifier ([ICARE](#)).

2.2 MRI standardization

The requirement for standardization is critical in the field of Magnetic Resonance imaging (MRI), particularly in the use of multicentric data. Data from many scanners, each with possibly distinct hardware, software, and scanning methods, ends up with researchers for development of standardization applications as described for instance in [54], [55], [56]. Different scanning techniques can cause inconsistencies in MR images that can result in changes in the image features that are further extracted. Machine learning models, which assume that the training data used is indicative of the data they would encounter in test conditions, are particularly affected by this [57], [58]. The data may become significantly more heterogeneous as a result of the following modification in image capture parameters including field strength, pulse sequence, coil design, slice thickness, and pixel size, among others. The objective of MR image standardization is to harmonize the scanning parameters and protocols across various scanners so that the final pictures are equivalent, no matter where they were acquired. Furthermore, to take into account variations due to retrospective studies. Intensity alignment can reduce discrepancies, enhancing the quality of modeling [59]. Furthermore, standardized MR images can improve data transmission and comparison between various research and institutions as well as allow the creation of imaging biomarkers that are robust and repeatable in a variety of scanning situations [60]. A step by step description of the pre-processing pipeline applied to the dataset used in this thesis is presented in the work of *Goya et al 2018* [61], which is an adaptation of the pipeline proposed by *Shinohara et al* [62] dedicated to brain tumors. A brief description of the main steps is presented below:

Inhomogeneity Correction

The N4 bias field correction is an algorithm created to correct the magnetic field, generating intensity inhomogeneities in MR data. A bias field appears as slow-varying spatial modulations in MRI scans, creating an artifact that may impair further image interpretation. The older N3 (nonparametric nonuniform intensity normalization) technique has been replaced by the N4 algorithm [63]. By changing the B-spline fitting procedure, the N4 method improves upon the N3 approach by making it more resilient in the presence of strong-intensity inhomogeneities. By estimating and reducing the multiplicative bias field, the N4 bias field correction aims at restoring similar intensities of each specific tissue in the images. For hypothesis and mathematical understanding refer to [63]. It's crucial to account for bias field during the preprocessing phase of MRI data analysis. Subsequent analysis, such as segmentation, registration, and any quantitative data obtained from the MRI, may be significantly impacted by it.

Multimodal Co-registration

Registering MRI scans is an essential step in many neuroimaging analyses, and it serves several crucial purposes [64]:

Alignment for comparison: Often, researchers or clinicians need to compare MRI scans taken at different points

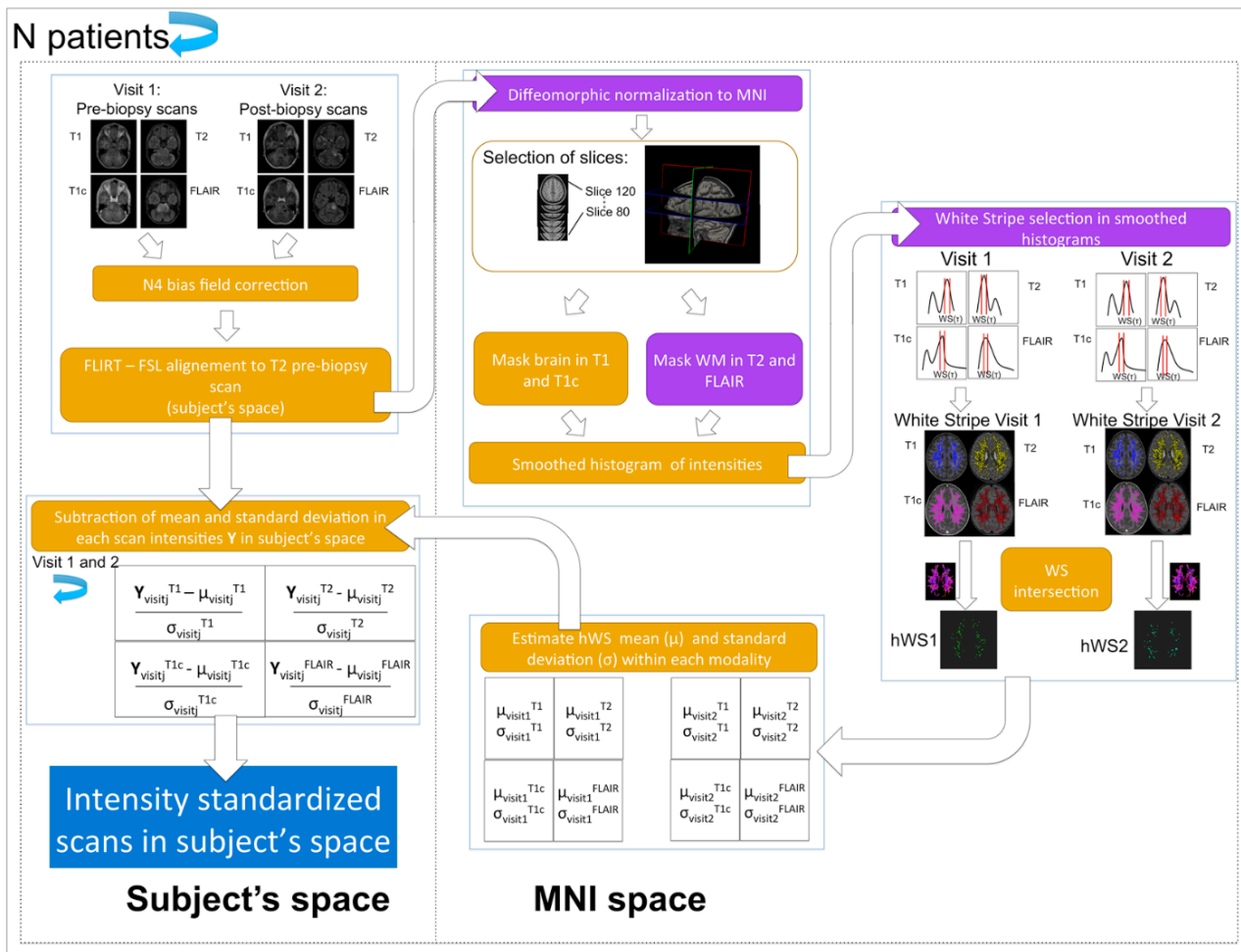


Figure 2.1: Workflow of MR Intensity standardization. Compared to the hWS method (Shinohara et al 2014) [62], the adaptations needed to analyze DIPG scans are shown in purple. Reprinted from [61]

in time (longitudinal studies) or from different individuals (cross-sectional studies). Registration allows these images to be aligned in a common space, making direct comparisons possible.

Multimodal imaging: In many cases, different types of imaging modalities (e.g., T1, T2, T2c) are used to study the same subject. Each modality provides different types of information (structural, functional, metabolic), and registration allows these different types of images to be superimposed or fused together so that their information can be integrated.

Atlas-based analysis: MRI scans are often compared to standard brain templates or atlases, which might contain predefined anatomical regions or functional networks. Registration allows individual scans to be mapped onto these atlases, enabling region-specific analysis.

Preprocessing for machine learning applications: For machine learning applications, registration is an important preprocessing step to ensure that the same anatomical or functional locations align across all input data. In this work, the different modalities associated with one patient were realigned to the T2 diagnostic scan (subject's space) using rigid registration in FSL-FLIRT [65]. Registrations were systematically checked visually and were always found

to be satisfactory.

Hybrid White Stripe

The Hybrid White Stripe (HWS) approach is a brain MR intensity normalization technique developed by Shinohara et al [62]. The original White Stripe technique developed for T1 weighted images has been expanded with the Hybrid White Stripe (HWS) approach. HWS uses a statistical normalizing technique is dedicated to multi-modal MR images account for the entire intensity distribution of the picture, not only the white matter. With the use of histogram distribution of this hybrid strategy, MR images will be more reliably and robustly normalized, enhancing the comparability of images both within and across patients as shown in figure 2.2. The HWS approach is particularly crucial in neuroimaging investigations because changes in image intensities can impair the capacity to identify and measure changes in brain tissue that are caused by diseases. From the R package ‘WhiteStripe: WM Normalization for Magnetic Resonance Images using WhiteStripe, Version 1.1.1’, some modifications were introduced to better take into account the specificities of our data. A detailed description of the modifications can be found in the work of *goya-outi et al.* [61].

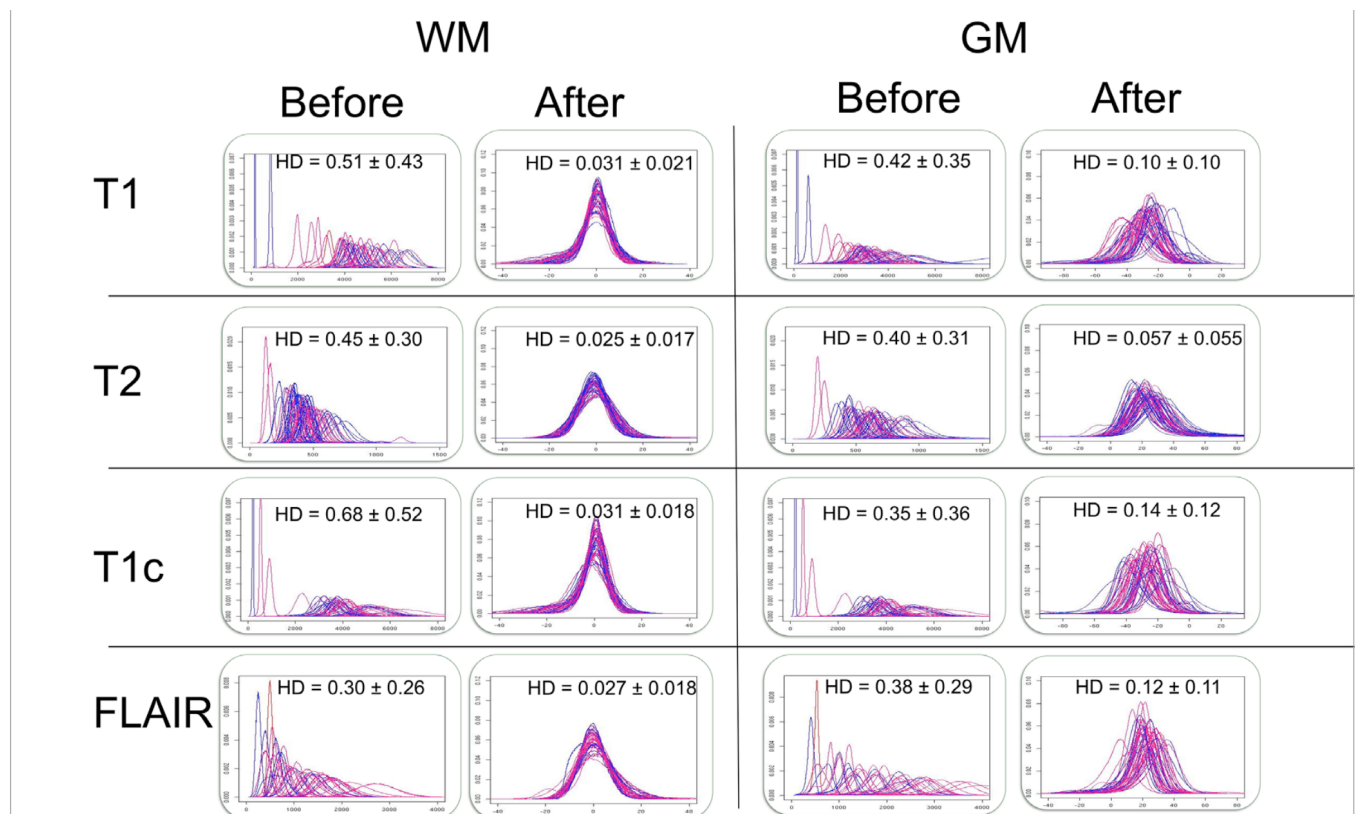


Figure 2.2: Histograms of intensities in WM and GM before and after standardization for T1, T2, post injection T1 (T1c) and FLAIR scans. Each blue line corresponds to a pre-biopsy scan and each pink line to a post-biopsy scan. For each type of MR scan, the mean Hellinger distance (HD) and its standard deviation between each pair of histograms are shown. Reprinted from [61]

2.3 Radiomics in medical imaging

Radiomics is a computational approach that involves extracting a vast number of quantitative image features from radiographic images, such as CT, MR, and PET scans [66]. This process transforms traditionally used simple medical images, which were primarily intended for visual assessment, into a rich and complex dataset that is amenable to high-dimensional analysis and mining [67],[68]. Radiomics aims to create robust decision support tools by integrating radiomic data with other available patient characteristics. This integration enhances the predictive power and effectiveness of decision support models, enabling more accurate and informed clinical decisions [69]. Radiomics encompasses the extraction of morphological and intra-tumoral heterogeneity properties from medical images. These properties are quantified using diverse image processing techniques, which involve the computation of shape parameters, first-, second-, and higher-order intensity. Surface reconstruction is used to extract shape characteristics, and the intensity histogram is used to produce first-order metrics. Using texture analysis techniques on the image, second and higher-order statistics are obtained (Table 2.1). RJ Gillies et al [70] were the first to coin the term 'Radiomics' in 2010, to denote these qualitative image features [68]. Subsequent to its introduction, radiomics has been utilized in diverse fields such as predicting treatment outcomes and survival rates, categorizing and staging tumors, and even establishing correlations between radiomics and genomics. Further details about the application of radiomics in these domains can be found in the referenced materials provided in Table 2.1.

Handcrafted features

Handcrafted features are computed utilizing predefined mathematical algorithms suggested by experts in the field of image processing. These features encompass both semantic and agnostic elements. Semantic features denote the conventional tumor descriptors traditionally used in radiology. Conversely, agnostic features comprise entirely of computational elements based on region of interest in volume. These agnostic features can be categorized further into groups that represent the shape, intensity, and texture inside the region or volume of interest. With numerous alternative methods and formulas available for computing these features, strict adherence to IBSI standards is strongly recommended [71]. In order to accomplish this, several freely available software solutions exist, including stand-alone programs, modules, and libraries like CERR [72], MITK [73], RaCaT [74], S-IBEX [73], LIFEx [75] and Pyradiomics [76] that adhere to the IBSI standard.

Shape-based features

The term "shape" or "morphology" refers to a segmented creation's two-dimensional or three-dimensional geometrical structure. These characteristics, which are determined from the segmentation overlay, are unrelated to the dispersion of grey level intensity contained in a picture. Examining characteristics like the surface area, volume, density, planarity, extensiveness, or roundness of the volume of interest under investigation is particularly benefited

by features based on form. The spiculated characteristics of the lesions suggest a proclivity for expansion and are frequently associated with more progressive stages of the condition [77]. On the other hand, non-threatening or less harmful tumors usually have clear, well-marked edges [78]. A particular research indicated that a significant number of radiomic shape attributes exhibited pronounced direct associations or consistent inverse correlations with the speculated nature of the tumor [79].

Intensity-based features

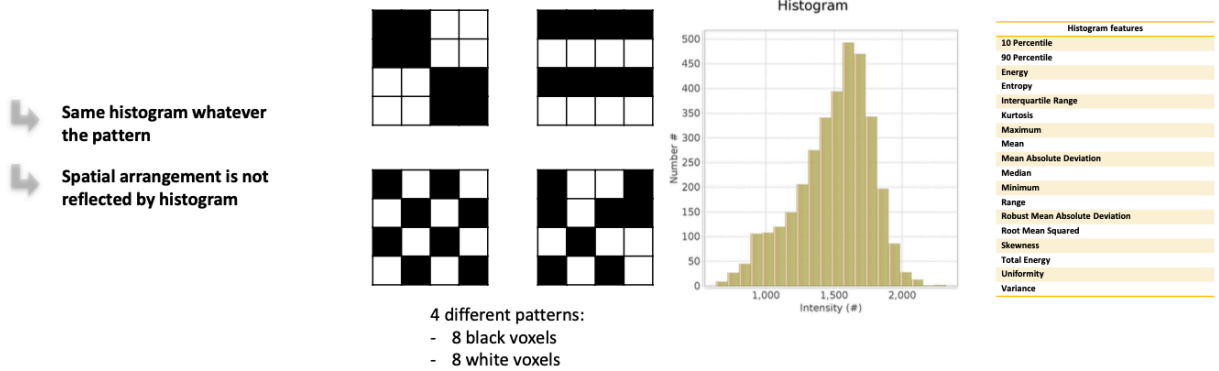
The features based on intensity, often referred to as first-order statistical characteristics, portray the dispersion of voxel intensities within the region or volume of interest. They utilize straightforward descriptors like the upper limit, lower limit, average, range, peakness, and asymmetry of the grey scale intensities as illustrated in Figure 2.3A.

Texture based features

Texture-oriented features, alternatively termed as second-order statistical characteristics, deliver insights on the organization of grey scale values within the region or volume of interest. Texture matrix-derived indices offer a robust approach to capturing tumor heterogeneity. Frequently encountered texture feature classifications encompass Grey Level Cooccurrence Matrix (GLCM), Grey Level Run Length Matrix (GLRLM), Grey Level Size Zone Matrix (GLSZM), Neighbouring Grey Tone Difference Matrix (NGTDM), and Grey Level Dependence Matrix (GLDM). A demonstration of how the various matrix categories are computed is illustrated in the Figure 2.3B. These different matrices thus provide complementary information:

- The Grey Level Co-occurrence Matrix signifies the likelihood of identifying a pair of values in voxels at a specified distance and in a particular direction [78].
- The Grey Level Run-Length Matrix gauges the count of sequential voxels possessing identical value, arranged in a specific direction for each intensity value [79] [80] [81] [82].
- The Grey Level Size-Zone Matrix quantifies the count of adjacent voxels sharing the same intensity, for every intensity value [83].
- The Neighborhood Grey-Tone Difference Matrix calculates the disparity in intensity between adjacent voxels [84].
- The Grey Level Dependency Matrix estimates the interdependencies of the grey scales in an image, that is, the count of neighboring voxels identical to the central voxel [85].

A Histogram or 1st-order features



B Textural or 2nd-order features

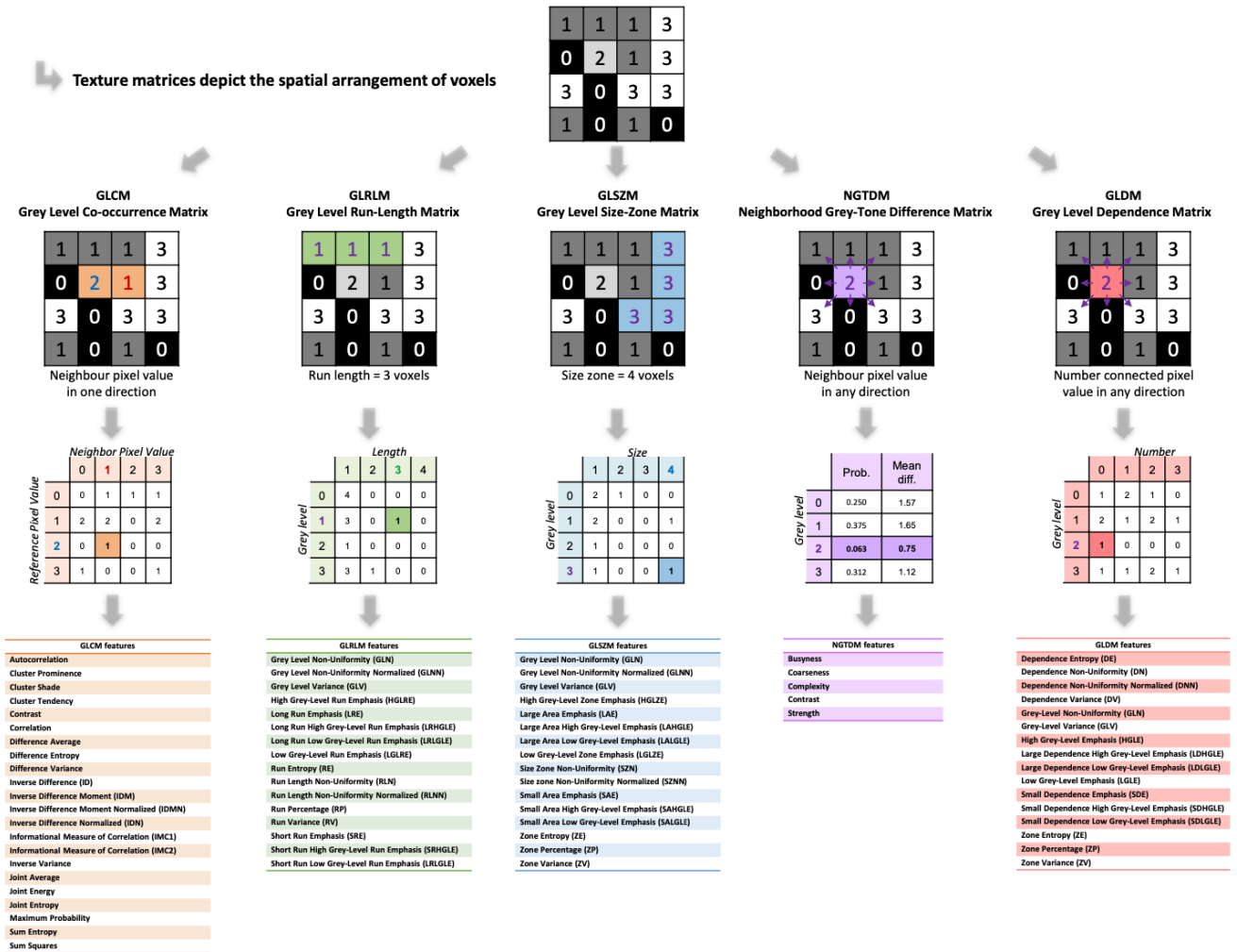


Figure 2.3: Representation of radiomic feature types. (A) Histogram or first-order features reflect voxel intensity distribution only. (B) Textural or second-order features derived from texture matrices (eg, co-occurrence, run length, size-tone, difference, dependence) reflect the complex and unique spatial arrangement of voxels. Reprinted from [1]

Applications	Study	Modality
Radiomic definition	Haralick et al., [78], Galloway et al., [79], Pentland et al., [86], Rahmim et al., [?], Amadasun et al., [84], Davnall et al., [87], Thibault et al., [83]	
Predicting treatment response and survival	Johansen et al., [88], Baek et al., [89], Shukla-Dave et al., [90], Foroutan et al., [91], King et al., [92], Peng et al., [93],	MRI
	Eary et al., [94], El Naqa et al., [95], Yang et al., [96], Cooket al., [97], Tixier et al., [98], Zhang et al., [99]	PET
	Aerts et al., [100], Parmar et al., [101], Tateishi et al., [102], Kim et al., [103], Tixier et al., [104]	CT
	Vallires et al., [105]	MRI + PET
Tumor Staging	Donget al., [106], Muet al., [107]	PET
	Ganeshanet al., [108]	CT
	Zacharaki et al., [109]	MRI
Tumor classification	Lerski et al., [110], Kjaer et al., [111],	MRI
	Mahmoud-Ghoneimet al., [112], Nie et al., [113], McNitt-Gray et al., [114], Kidoet al., [115], Petkovska et al., [116], Wayet al., [117]	CT
	Xu et al., [118], Yuet al., [119],	PET + MRI
	Diehn et al., [120], Ellingson et al., [121], Naeini et al., [122], Gutman et al., [123]	MRI
Radiogenomics	Nair et al., [124], Nair et al., [125]	PET
	Khalid et al., [2][3] Goya-outi et al., [61] [126], Kandemirli., et al [127], Li et al., [128], Chilaca-Rosas et al., [129], Lv et al., [130], Gou et al., [131], Wagner et al., [132], Tam et al., [133]	MRI

Table 2.1: List of the main publications based on Radiomics according to their application and the imaging modality studied. Modified from Yep et al., [134] and Desseroit et al., [135]

2.4 Feature Selection

The goal of feature selection is to optimize the performance of prediction models, comprehend the inherent procedure that produced the data, and, in most instances, provide predictors that are more time-efficient and cost-effective. Assuming a training dataset composed of N instances and P predictive variables or features X_i ($i = 1, \dots, P$) and a class Y in $1, 2, \dots, C$, the goal of feature selection is to choose a concise subset of variables or features, without compromising the predictive information relevant to Y . It's important to understand that feature selection involves choosing a subset from the original set of features, which may offer greater interpretability compared to feature extraction methods, such as principal component analysis, that generate new features through transformations of the initial feature set [136]. The three types of feature selection approaches are filter methods, wrapper methods, and embedding methods. Wrapper approaches, such as forward and backward selection, evaluate subsets of variables using the intended machine learning as a black box to score them based on their predictive power. On the other hand, filter approaches choose subsets of variables prior to the predictor, for example, using criteria based on correlation. Finally, embedded approaches, which are often specialized to machine learning like logistic regression (LR), support vector machines (SVM), and random forests (RF) (Section 2.7), incorporate vari-

able selection throughout the training process [137]. Feature selection is extensively employed in fields such as gene selection, aiming to enhance interpretability and mitigate the dimensionality curse, a challenge presented by high-dimensional data [138] [139]. Sophisticated wrapper or embedded methods improve predictor performance compared to simpler variable ranking methods like correlation methods, so in this work we used an embedded method associated with LR in a cross validation: Logistic Regression Recursive Feature Elimination with cross validation (LR-RFECV) [138].

2.5 Learning models for medical data

2.5.1 Machine Learning

There has been a remarkable surge in the application of machine learning (ML) algorithms in the field of medical imaging over the past two decades, as illustrated in Figure 2.4. Innovative software has been designed to automate a wide array of tasks within clinical oncology, ranging from tumor identification and segmentation to therapeutic decision-making [140] [141] [142]. Increasingly, these sophisticated systems are demonstrating performance that surpasses human specialists. As we accumulate more data, enhance image quality, and harness the capabilities of more potent computing systems, it is projected that machine learning-powered automation will revolutionize the operations within clinical medicine [143] [144].

Machine learning algorithms can essentially be categorized into three groups: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning makes use of data that is labeled, meaning the desired output is already known, to train a model. This model is then applied to make forecasts. Conversely, unsupervised learning uses unlabeled data, constructing models that identify patterns or structures in the data, like data clustering. Reinforcement learning, on the other hand, involves a model that learns through a series of actions guided by a feedback mechanism. A notable example of this is the AlphaGo program which, in 2015, became the first computer program to defeat a professional Go player [145]. In this manuscript, our focus will be on supervised learning.

2.5.2 Deep learning

Deep learning, a specialized subfield of machine learning, employs intricate architectures consisting of multiple layers of neurons known as neural networks [146]. These vast networks undergo optimization during training by minimizing a predefined loss function using gradient descent-based algorithms. Among the various deep learning techniques, Convolutional Neural Networks (CNNs) stand out for their suitability in image analysis. In contrast to handcrafted radiomics, which extracts features from a specific Region of Interest (ROI), deep learning generally processes the entire image. This approach allows CNNs to understand the spatial relationships between voxels through the use of filters and convolutions. In the realm of medical imaging, CNNs are extensively utilized for tasks

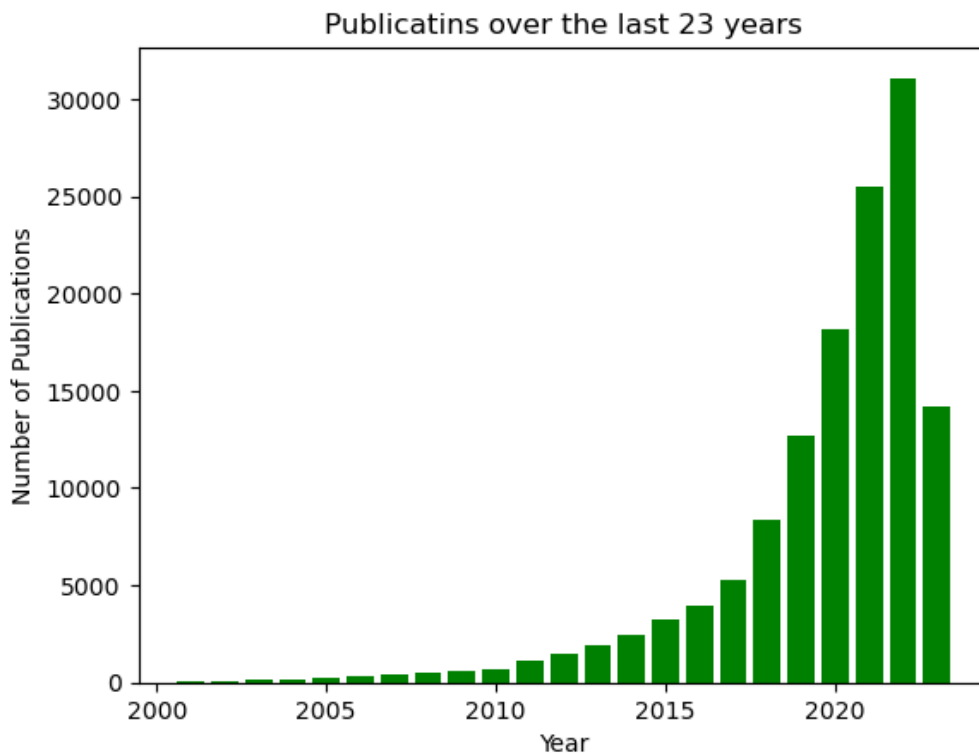


Figure 2.4: Results on PubMed using the keyword search “Machine Learning” by year up to April 2023 , showing the rapid growth of interest in machine learning.

such as lesion and organ segmentation, abnormality detection, and other prediction or classification challenges comparable to what traditional radiomic models aim to achieve [146, 147]. One notable advantage of CNNs is the elimination of the need for tumor delineation, often a limiting factor in radiomic studies. However, the effective training of CNNs necessitates a considerable volume of data to yield meaningful models and mitigate overfitting risks. Given that CNNs often encompass millions of parameters, their optimization demands substantial computational resources.

Supervised Learning

Supervised learning is carried out using a collection of labeled inputs, also known as training data. This training data is leveraged to effectively establish the model. Once this function is appropriately trained, it can be employed to scrutinize an unfamiliar data set and accurately forecast the anticipated result. If the new input data were previously identified, the outcome labels can be evaluated for their exactness and precision.

Unsupervised Learning

Unsupervised learning is realized by deducing the function or algorithm for a labeled data set. In this mode of learning, the outputs are unknown, necessitating that the function ascertain the appropriate response. Algorithms

used in unsupervised learning tackle intricate problems using only the input data.

General principles of model estimation

The field of machine learning lacks a universally accepted standard for mathematical notation. In this context, we'll represent the training set as $(\mathbf{x}^{(i)}, y^{(i)})$ where $1 \leq i \leq n$ represent the n training examples. \mathbf{x} is be the input data and y is the predicted variable/class. The training example $\mathbf{x}^{(i)}$ consists of a number of features m :

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \quad (2.1)$$

Training of a Model

Considering an input (for instance, a collection of radiomic features), our objective is to construct a model capable of generating a prediction (such as determining the presence of cancer in a patient). From a mathematical perspective, we aim to identify a function, f

$$f(x) = y \quad (2.2)$$

If y is not continuous but has distinct categories, we're dealing with a classification problem (like identifying if a patient has cancer). On the other hand, if y is continuous, it's a regression problem (like estimating how long a patient will survive). We find this function (or get as close as we can to it) by using the data and labels in the training set to train a machine learning model. Optimizing the *weights* of the model enables this which we denote as θ_i in the case of linear model.

$$f_{(\theta)}(\mathbf{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 \quad (2.3)$$

The parameter of the function θ_i . Their determination is achieved with the help of a loss function, also known as a cost function, which tells us how well the function $f(x)$ matches the actual value, y . Typically, we start with random values for the weights and then gradually improve them during training using methods like *gradient descent*. Gradient descent works in an iterative manner to minimise the loss function:

$$\theta_i^{t+1} = \theta_i^t - \alpha \frac{\partial J(\theta_i^t)}{\partial \theta} \quad (2.4)$$

J is the loss function and t is the iteration in the gradient decent. α is the learning rate and controls the magnitude of the iteration update. This parameter is crucial to adjust during the training process. If it's too small, the method

of gradient descent will take a long time to reach the optimal point; if it's too large, it might not get there at all. Usually, the learning rate α stays the same, but there are numerous strategies suggested to modify α as the training continues (a deeper insight is presented in [148]).

Evaluating a Model

After training the model enough, we need to check how good it is. A model might do great on the training data but not do well on new data. This issue is known as overfitting. On the other hand, a model might not be strong enough to correctly identify patterns in a dataset, which is referred to as underfitting. Figure 2.5 illustrates these scenarios.

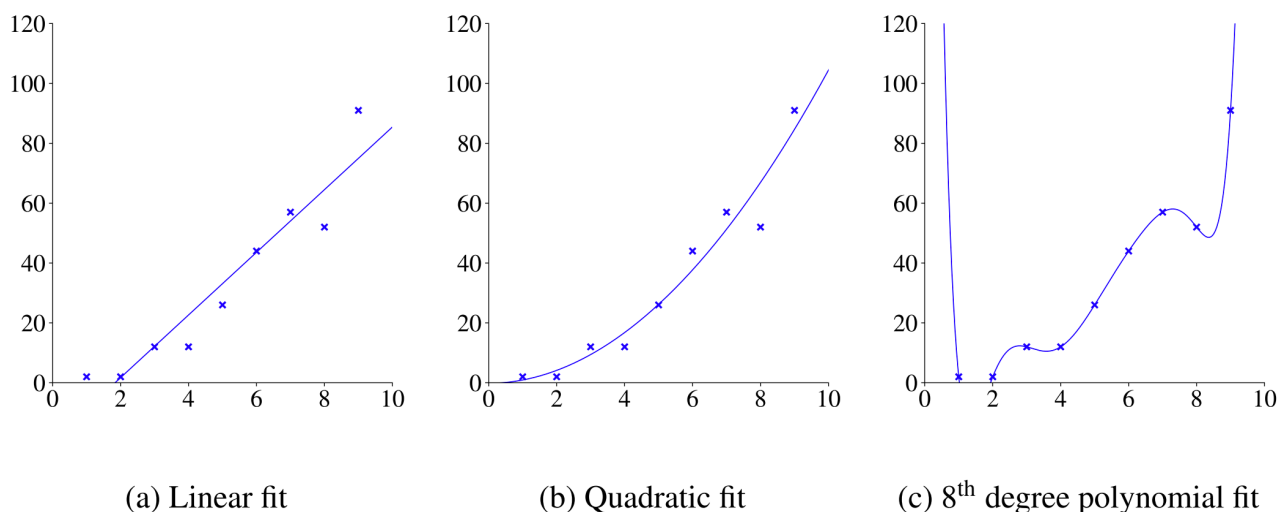


Figure 2.5: Examples of underfitting and overfitting using least squares regression. In (a) we have fit a straight line to the points, which slightly underfits the data. Adding another parameter, (b) is a quadratic curve, which fits the data well. (c) is an 8th degree polynomial fit. Although it fits these data points well (the training data), it will probably not generalise well to new data points and is therefore overfitting the data.

To really know how well a model is doing, the data is divided into three parts: training, validation, and test sets. We use the training data to teach the model. We use the validation data to check how the model is doing as we adjust things like learning rate and loss function. Once we've made the model as good as we can, we use the test set to see how it does on new data. This step is important to make sure the training is reliable. The validation and test sets have to be big enough so we can see clear differences in how well different models or methods work. To evaluate the performance of a model, we usually break our data into three parts: training, validation, and test sets. We use the training data to teach the model. We use the validation data to check how the model is doing as we adjust things like learning rate and loss function. Once we've made the model as good as we can, we use the test set to see how it does on new data. This step is important to make sure results are transferable. The validation and test sets have to be big enough so we can see clear differences in how well different models or methods work. In certain scenarios, especially within the medical domain where available data might be limited (for instance, around 100 images), employing a technique called cross-validation can be more effective than relying on a test

set. In this approach, there isn't a distinct test set. Rather, the data is partitioned into 'k' random subsets, leading to the term k-fold cross-validation. One of these subsets is reserved solely for validation. After model tuning with the initial partition, the model undergoes 'k' training iterations, each time utilizing a different subset for validation. Subsequently, by averaging the results, an approximate performance measure of the model can be obtained. This strategy minimizes potential biases that might arise from relying solely on a single validation set outcome, especially when overfitting is a concern due to excessive fine-tuning. In situations where 'k' equals the dataset size, resulting in a single datapoint for validation, it's termed leave-one-out-cross-validation (LOOCV). However, it's important to note that lacking a distinct test set means that cross-validation may not provide the most robust assessment of model performance. This becomes particularly relevant with complex models, such as deep learning architectures, where numerous training parameters require adjustment. Some research indicates that even with LOOCV, there's potential for overestimating model performance compared to evaluations using an independent test set [149].

The performance evaluation of the final models was done using accuracy, sensitivity and specificity. The definitions for accuracy, sensitivity, and specificity are as follows:

$$Sensitivity = \frac{TP}{TP + FN} \quad (2.5)$$

$$Specificity = \frac{TN}{TN + FP} \quad (2.6)$$

$$Precision = \frac{TP}{(TP + FP)} \quad (2.7)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.8)$$

$$BalancedAccuracy = \frac{(Sensitivity + Specificity)}{2} \quad (2.9)$$

In these expressions, TP denotes the count of true positives, TN represents the count of true negatives, FP is the count of false positives, and FN signifies the count of false negatives [150]. The incorporation of sensitivity and specificity alongside accuracy ensured robust classification for all patient groups.

Optimisation Algorithms Gradient

Gradient descent is the most basic optimization method, which uses the first set of changes (we call these "first-order derivatives") in the loss function. It's simple to understand and put into action. However, newer optimization methods have mostly taken its place. Let's quickly go over the most important ones:

Stochastic and Mini-Batch Gradient Descent

The weights and biases are often changed in the conventional gradient descent approach only after the dataset's loss function has been determined. The optimal point might not be reached for a very long time if the dataset is really large. To get around this issue, the dataset may be broken up into smaller groups, or "batches," and the parameters can be changed after each batch. We get at the best position more quickly since we update the settings more frequently. However, we maintain the loss constant between batches by selecting a reasonable batch size. When optimizing a neural network, the batch size is a crucial aspect to take into account. Stochastic gradient descent (SGD), which is also known as mini-batch gradient descent, is what happens when we employ a batch size of one.

Adaptive Moment Estimation

The Adaptive Moment Estimation (Adam) optimizer uses separate learning rates for each parameter and alters them as the learning process progresses rather than having a single learning rate for all parameters. The published study [151] has all the details regarding this. Adam still uses mini-batches and has a parameter α which sets the overall learning rate of parameters.

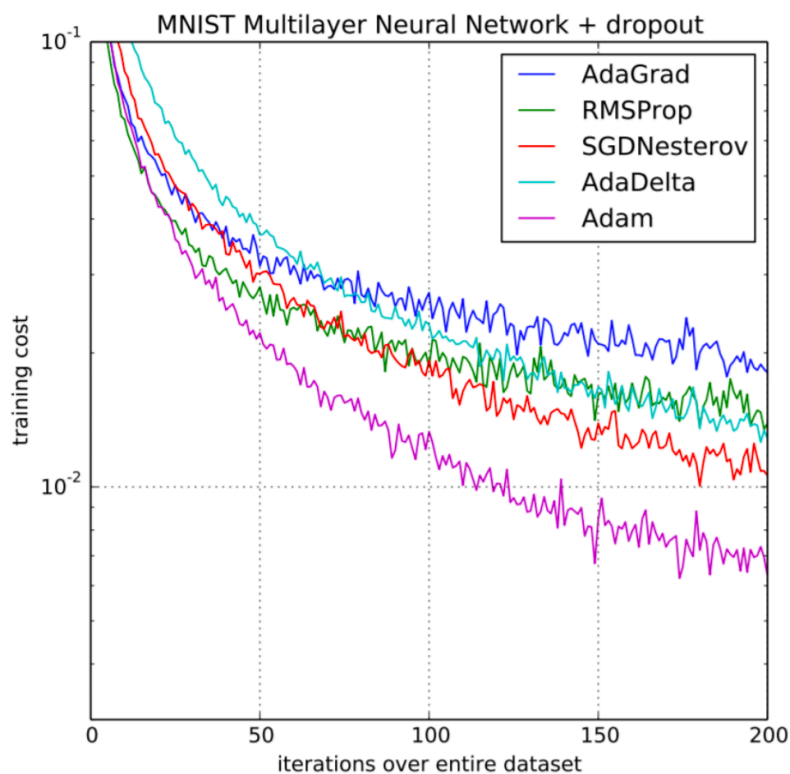


Figure 2.6: Converge rates of several optimisation algorithms using a multilayer neural network on a handwritten classification challenge. Taken from [151]

The Adam optimization technique is perhaps the most popular and is typically suggested as the first choice

[152] although in data science especially machine learning there are exceptions as mentioned in [153]. Adam is the quickest when comparing the converge rates of various optimization techniques, as seen in figure 2.6 from the original Adam article.

2.5.3 Loss Functions

The shape of the loss function dictates the changes during gradient descent, and as a result, sets the final parameters of the tweaked model. So, it's crucial that this loss function accurately mirrors the performance measure we're trying to boost. In other words, reducing the loss should lead to better performance. Let's go through the loss functions that are relevant for this discussion:

Mean Square Error Loss

Mean Square Error (MSE) is a simple quadratic loss, commonly used in regression problems

$$J(\theta) = \frac{\sum_{i=1}^n (\hat{y}^i - y^i)^2}{n} \quad (2.10)$$

Here $\hat{y}_\theta^{(i)} \equiv f_\theta(\mathbf{x}^{(i)})$ are the predictions. Because the loss function is squared, it pays more attention to extreme outliers that are far from the predicted value. When the predicted value (let's call it \hat{y}) is close to the actual value (y), the change in the gradient is small. This means that the gradient descent process is slower and, as a result, more precise.

Cross-Entropy Loss

Cross-entropy is the go-to loss function for problems where we need to classify something, and it needs input values that range from zero to one. For a problem where we need to classify something into two categories, we define the cross-entropy loss as:

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n \left[y^{(i)} \log(\hat{y}^{(i)}) + (1 - y^{(i)}) \log(1 - \hat{y}^{(i)}) \right] \quad (2.11)$$

when simplified to

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=0}^1 y_j^{(i)} \log(\hat{y}_j^{(i)}) \quad (2.12)$$

Here, the sum J considers the classes 0 and 1. For problems where you need to classify something into more than two categories (let's say K categories), we can expand this concept:

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k y_j^{(i)} \log(\hat{y}_j^{(i)}) \quad (2.13)$$

DICE Loss

The **DICE** loss is often used in segmentation problems

$$J(\theta) = 1 - \sum_{i=1}^n \frac{2\hat{y}^{(i)}y^{(i)}}{\hat{y}^{(i)} + y^{(i)} + 1} \quad (2.14)$$

Hinge Loss

$$J(\theta) = \sum_{i=1}^n \max(0, 1 - \hat{y}^{(i)}y^{(i)}) \quad (2.15)$$

2.6 Class Balancing

In many machine learning scenarios, we often encounter a situation where the classes are not balanced, for example, we might have far more negative diagnoses than positive ones. This can be tricky when we use a loss function, because the model could simply predict the majority class all the time to minimize the loss. There are several ways to handle this issue. One is to under-sample the majority class or over-sample the minority class. Another way is to adjust the weight of the classes in the loss function, so that there's a larger penalty when the model wrongly classifies the minority class. Additionally, we can create synthetic data using methods such as the synthetic minority over-sampling technique (also known as **SMOTE**) [154] or data augmentation.

2.7 Classifiers

2.7.1 Logistic Regression

One of the most often used Machine Learning algorithms, within the category of Supervised Learning, is logistic regression. Using a specific collection of independent factors, it is utilized to predict the categorical dependent variable. Logistic Regression, as a classifier, is primarily a statistical approach that uses a logistic function to describe a binary dependent variable, while it may also be used for multiclass classification issues [155]. This method, which is a specific instance of linear regression where the result is a binary classification, predicts the likelihood that an event will occur by fitting data to a logistic function.

The typical logistic regression model is represented as follows:

$$P(Y = 1|X = x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}} \quad (2.16)$$

In this context, $P(Y = 1|X = x)$ indicates the conditional probability of the event $Y = 1$, assuming $X = x$. The element e corresponds to the base of natural logarithms, and β_0 and β_1 are the model's parameters to be determined, with x embodying the independent variable [156].

The logistic regression model, when used with a classifier, computes the odds ratio (a measure of association between an exposure and an outcome) in a multivariate environment and offers a quantitative estimate of the likelihood that a specific event will occur [157]. This offers an in-depth comprehension of the likelihood of events in a manner that takes complicated, real-world data sets with numerous inputs into account [158]. The model is particularly well-suited to applications in the social sciences and in medicine, where it is frequently necessary to comprehend how several interconnected factors affect a binary result.

2.7.2 Support Vector Machine

Support Vector Machines (SVM) are supervised learning models primarily used for classification and regression tasks. They work by figuring out which hyperplane categorizes a dataset into classes the best. The hyperplane that maximizes the margin between the two classes is the one that is selected. When the data cannot be separated linearly, SVM employs a kernel technique to convert the input space into a higher-dimensional space and then locates the best hyperplane there. The equation for the hyperplane can be given as:

$$\mathbf{w} \cdot \mathbf{x} + b = 0$$

Where:

- \mathbf{w} is the weight vector, perpendicular to the hyperplane.
- \mathbf{x} is the input data.
- b is the bias term.

The decision function that assigns a class to a new data point x is:

$$f(x) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b)$$

For a linearly separable dataset, the SVM aims to maximize the margin between the two classes. The margin is defined by the distance between the hyperplane and the closest data points from each class, known as support vectors. The equations for the hyperplanes defined by the support vectors are:

$$\mathbf{w} \cdot \mathbf{x} + b = 1 \quad (\text{for the positive class})$$

$$\mathbf{w} \cdot \mathbf{x} + b = -1 \quad (\text{for the negative class})$$

The margin M can be computed as:

$$M = \frac{2}{\|\mathbf{w}\|}$$

SVMs have been widely employed in a variety of domains, including text categorization, picture classification, and bioinformatics, because of their robustness against overfitting and capacity for handling huge feature spaces. SVMs are a common solution for many classification issues due to their efficiency in high-dimensional spaces and the flexibility provided by the availability of many kernel functions [159, 160].

2.7.3 Random Forest

Random Forest is a versatile and widely-used ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of the classes for classification tasks or the mean prediction for regression tasks [161]. It operates by bootstrapping the data for each tree and selecting a random subset of features at each node of the tree, which ensures the decorrelation of the trees and helps in achieving a lower variance and higher accuracy [162]. One of the primary strengths of Random Forest is its capability to handle large datasets with higher dimensionality. It can manage missing values, and it's relatively unaffected by outliers. It also provides a good indicator of feature importance, making it a valuable tool for feature selection in complex datasets [163]. Due to its robustness and ease of use, Random Forest has been widely adopted in various fields, including medical imaging, bioinformatics, and financial modeling.

The working principle of the Random Forest can be described as follows:

1. **Bootstrapping the data:** Random Forest starts by selecting random samples with replacement from the training dataset. This is known as bootstrapping. Each of these samples is used to train a decision tree.
2. **Feature selection:** Instead of using all features to make a decision at a node, it randomly selects a subset of features and then determines the best split from that subset. This introduces further randomness into the model.
3. **Tree construction:** Trees are grown to the fullest extent without pruning, leading to large fully-grown trees.
4. **Aggregation (majority voting or averaging):**
 - For classification: For a new data point, each tree in the forest predicts a class label. The class label that gets the majority of votes is chosen as the final prediction.

- For regression: For a new data point, each tree in the forest predicts a continuous value. The final prediction is the average of these values.

Mathematically, for classification, the predicted class y is given by:

$$y = \text{mode}\{y_1, y_2, \dots, y_T\}$$

where y_i is the class predicted by the i^{th} tree and T is the total number of trees.

For regression, the predicted value y is:

$$y = \frac{1}{T} \sum_{i=1}^T y_i$$

where y_i is the value predicted by the i^{th} tree.

2.8 Survival Analysis

A crucial field of study in medical imaging is survival analysis, which aims to link image-derived characteristics to time-to-event outcomes like patient survival or disease recurrence. Researchers want to identify prognostic indicators that can forecast clinical outcomes by collecting quantitative information from medical imaging or biopsy. The most often used methods in this field include:

1. **Kaplan-Meier Analysis:** A fundamental method for visualizing and analyzing survival data. Patients are often based on specific feature, and the survival curves in the different groups are compared to discern differences in outcomes.
2. **Cox Proportional Hazards Model:** A regression model that relates several variables to the time of a particular event. In the context of medical imaging, this could involve associating multiple radiomic features with survival time.
3. **Deep Learning Models:** With the advent of artificial intelligence, deep learning models, especially convolutional neural networks (CNNs), are increasingly being used to directly predict survival outcomes from raw image data or in combination with radiomic features.
4. **Time-to-Event Neural Networks:** These are specialized neural network architectures tailored for survival analysis, combining the strengths of deep learning with traditional survival models.

Harnessing these techniques, medical imaging is evolving into a potent tool not just for diagnosis, but also for predicting patient outcomes and personalizing treatment strategies.

2.8.1 Kaplan Meier curves and Log-rank tests

A key statistical tool used to examine and display survival data is the Kaplan-Meier (KM) estimate. It offers a means to display the survival curve and estimate the survival function using lifetime data. Recently, survival analysis, including the Kaplan-Meier approach, has been combined with radiomic characteristics, which are quantitative features generated from medical pictures, to give insights into patient prognosis and disease development.

Kaplan-Meier Survival Analysis:

1. Basic Mechanism: The Kaplan-Meier estimator calculates the probability of an event (e.g., death, relapse) occurring over a defined period. It's particularly suitable for handling "censored" (censored data refers to observations where the event of interest (e.g., death) has not yet occurred or is not observed during the study period.) data, where the event of interest was unknown.
2. Survival Curve: The KM estimator is often visualized as a "survival curve," which shows the proportion of subjects surviving (or free from some event) over time. A drop in the curve indicates an event occurrence.
3. Log-Rank Test: To compare survival curves between two or more groups, the log-rank test is often used. This test determines if there's a statistically significant difference between the groups in terms of survival [164].

In studying DIPG through Kaplan-Meier analysis integrated with radiomics, MRI scans are analyzed to extract radiomic features, which include aspects like shape, texture, and intensity of the tumors. Given the vastness of the radiomic dataset, feature selection techniques are crucial to pinpoint the most clinically relevant attributes. These features then guide the stratification of patients into distinct risk categories, such as "high" or "low" based on tumor characteristics. For each category, Kaplan-Meier survival curves are plotted, shedding light on the survival probabilities over time. Ultimately, to discern the predictive power of specific radiomic features for DIPG prognosis, the survival curves of various groups undergo comparison via the log-rank test [165]. Significant curve disparities highlight the prognostic significance of the corresponding radiomic features.

2.8.2 Concordance Index

The Concordance Index (C-index) in survival analysis is a metric used to evaluate the discriminative ability of a survival model. Specifically, it quantifies the degree of concordance between the model's predicted risk scores and the actual outcomes. A C-index value of 0.5 suggests no discrimination (akin to random guessing), while a value of 1.0 indicates perfect discrimination. For survival data, it measures the pairwise concordance of predicted risks: given two randomly chosen subjects, the one with the higher risk should experience the event (e.g., death) earlier than the other. Mathematically, for each pair of subjects (i, j) , if the actual event time T_i for subject i is less than T_j and the predicted risk for i is also higher than for j , then the pair is considered concordant. The C-index [166] is then computed as the proportion of all pairs that are concordant. The C-index is analogous to the area under the

receiver operating characteristic (ROC) curve used in binary classification, but it is adapted for censored survival data.

2.8.3 ICARE algorithm

During the thesis I took part in an international challenge, the HECKTOR (HEad and neCK TumOR segmentation and outcome prediction in PET/CT images) presented in the MICCAI 2022 conference. Data and challenge description can be found in the link provided ([53]). Working closely with my colleagues we won the challenge with 1st position for the outcome prediction task and 4th position for the segmentation task. The developed algorithm for outcome prediction is a binary weighted model which was later developed by Louis Rebaud into a scikit learn package called ICARE [167]. We hypothesized that it is challenging to precisely quantify the value of a biomarker in outcome prediction based on the literature and our experience. In fact, there is a possibility of biased estimate of the feature weights due to data noise, censoring of the target, such as progression free survival, and very few training samples. We suggested minimizing the learned information in order to counteract this impact and estimating simply the sign that should be given to each feature when estimating the goal. The introduced binary-weighted model's (ICARE) fundamental working principle is this.

Algorithm: There are N samples and M features in the training dataset. Numerous radiomic traits have strong correlations. Only one out of a group of linked features should be preserved in order to uphold the fundamental tenet of the binary-weighted model because using all of them as input would artificially give the information a huge weight. As a result, the p characteristics by determining the absolute Pearson correlation coefficient for each pair of features. The value over which two traits are regarded to be too linked is determined by a threshold. In this scenario, one of the two features is picked at random and removed.

Considering C_{index} to be the Harrell's concordance index [166]. Every given feature x_i is evaluated on its ability to correctly predict the target value y with:

$$c_i = C_{index} \quad (2.17)$$

To lessen the possibility of incorrect sign assessment, the features with $|c_i| < C_{min}$ are dropped, where $|c_i| = \max\{1 - c_i, c_i\}$ and C_{min} is a hyperparameter in $[0.5, 1]$. The remaining features are assigned the following sign:

$$S(i) = \begin{cases} +1, & \text{if } c_i \geq 0.5 \\ -1, & \text{otherwise} \end{cases} \quad (2.18)$$

In order to scale the feature values to the same range, a normalizing step is introduced. Otherwise, the final

prediction would give more weight to traits with high absolute values. The model does this by calculating the z-score of each feature.

$$z_i = \frac{x_i - \mu_i}{\sigma_i} \quad (2.19)$$

where μ_i and σ_i are the mean and standard deviation of x_i in the train set. The estimate \hat{y} of the target y is computed with:

$$\hat{y} = \frac{1}{M} \sum_i^M s_i \times z_i \quad (2.20)$$

The computation of \hat{y} , μ_i and σ_i are done by ignoring the missing values of the dataset. This allows the model to use features with missing values. Here, C_{min} and p are the only two hyperparameters of the model.

An important characteristic of the defined algorithm is that it can handle missing data which is one of the key aspects explored in this thesis. Indeed we focused on the development of models and approaches which can accept missing data/features without introducing any imputation. Since the DIPG cohort with missing modalities and a few censored patients it was interesting to test and analyze results of ICARE on this cohort.

2.9 Conclusion

In this chapter, we delve into the significance of MRI standardization, elaborating on the diverse techniques employed to attain unbiased standardization of MR images acquired from multiple centers, each operating under distinct scanning protocols. We provide an in-depth discussion on radiomics, focusing on its various classes of features and the specific types of information they encapsulate. These include first-order features, second-order texture features, and shape features. Machine and deep learning models are presented and their salient features for optimized output are presented, the frequently used classifiers which we see in later chapters such as logistic regression, support vector machine and random forest are also discussed. Additionally, we introduce some conventional tools used for survival analysis and finally we present the ICARE classifier, a novel tool that contributes significantly to the field. Then, we explore the critical facets of machine learning, with an emphasis on aspects such as feature selection, underscoring their role in constructing effective predictive models.

Chapter 3

A first multi-model approach for H3K27M mutation prediction using radiomics and clinical features

3.1 Introduction

In the preceding chapters, we introduced MRI imaging techniques used in the clinical management of DIPG tumors and impact of radiomic features and machine learning models to build better predictive models. In this chapter, we focus on multimodal framework dealing with classification of MRI (including T1-weighted pre and post-contrast, T2-weighted, and FLAIR sequences), to address the issue of missing modalities. We propose a unique imputation-free method to tackle this challenge. The main objective of this chapter is to develop and validate a methodological framework for predicting H3K27 mutation in DIPG using the aforementioned MRI sequences. We also investigate the impact of ComBat harmonization on the prediction task.

3.1.1 Radiomic studies in how to deal with missing data multiple scanners

A first approach was proposed by Goya-outi et al 2019 [126] selecting patients having the four modalities. In this preliminary study 30 patients were used to predict H3 mutations to better take into account the characteristics of the population (missing data). Several approaches are possible including imputation. Imputation refers to the process of replacing missing data with estimated values based on the information available in the dataset. In the context of medical imaging mutation prediction, missing data may arise due to various reasons, such as technical errors in data acquisition, or incomplete patient data. Imputation can have both positive and negative effects on machine learning algorithms, depending on how it is performed and the characteristics of the dataset. There are several imputation methods commonly used in machine learning, including:

1. Mean imputation: In this method, the missing values are replaced with the mean value of the non-missing values for that feature [168].
2. Median imputation: Similar to mean imputation, but the missing values are replaced with the median value of the non-missing values [169].
3. Mode imputation: For categorical features, the missing values are replaced with the mode (most common value) of the non-missing values [170].
4. Hot deck imputation: In this method, the missing values are replaced with values from other similar observations in the dataset. For example, if a missing value is in a patient record, it can be imputed using the value of the same feature from another patient with similar characteristics [171].
5. K-nearest neighbors imputation: This method imputes missing values by finding the k-nearest neighbors to the observation with missing values and replacing them with the average or median value of those neighbors [171].
6. Regression imputation: In this method, a regression model is trained to predict the missing values based on the other features in the dataset. The predicted values are then used to impute the missing data [172].
7. Multiple imputation: This method involves creating multiple imputed datasets, each with different imputed values, and then combining the results from the analyses on each dataset to obtain a final estimate.

However, the use of imputation methods for a prediction task can affect the accuracy and reliability of the prediction model. An imputation methods can indeed introduce bias in the imputed values, leading to biased estimates of the model parameters and reduced accuracy of the prediction model. A study by Vaida and Xu (2000) [173] showed that imputation methods such as mean imputation, regression imputation, and multiple imputation can lead to biased estimates of the model parameters in linear regression models. A study by Barnard and Rubin (1999) [174] showed that multiple imputation can underestimate the variance of the imputed values, leading to overconfidence

in the model predictions. A study by Jerez et al (2010) [175] showed that imputation methods such as regression imputation and KNN imputation can impute noise into the missing values, leading to increased variability in the prediction model and reduced accuracy. A study by Little and Rubin (2002) [176] showed that different imputation methods can lead to different estimates of the missing values, leading to different model predictions. In conclusion, while imputation methods are a useful tool for handling missing data, they have some drawbacks that can affect the accuracy and reliability of the prediction model. The use of imputation methods should be carefully considered and justified based on the specific characteristics of the dataset and the research question. Generative Adversarial Networks (GANs) have been heralded as a breakthrough in many domains, including the imputation of missing modalities in medical imaging. However, they are not without challenges. One of the primary concerns with GANs is mode collapse, a phenomenon where the network produces a limited variety of outputs rather than a broad representation of the data [177]. This limitation is particularly concerning in medical imaging where diverse pathologies and anatomical variations are essential to capture. Additionally, GANs are notoriously unstable during training, often leading to oscillations or divergence which can result in the creation of unrealistic images [178]. The lack of a ground truth for the produced outputs, which makes it difficult to quantify image quality, presents a more particular issue to medical image synthesis [179]. Additionally, there is a chance of overfitting when GANs are trained on small datasets, which might result in the replication of patient-specific characteristics in created pictures, causing difficulties with accuracy and privacy [180]. Even while GAN-synthesized pictures could look realistic, they could contain anatomical and pathological discrepancies that could result in inaccurate clinical interpretations [181]. Last but not least, the fact that many GAN-based synthesis approaches lack thorough clinical validation is a major source of worry in the medical sector. It is crucial to guarantee that these synthetic pictures are not only aesthetically accurate but also clinically valid and do not add diagnostic errors [182].

Alternative methods that can handle missing data directly without imputation should thus be considered, by defining for instance different models computing 1, 2,.. N types of data as illustrated in Figure 3.1

3.1.2 How to deal with multi-scanner data

Medical imaging data can fluctuate significantly between institutions, locations, devices, and even moments in time. The generalization of machine learning models, which may have been trained on data from one distribution but are anticipated to function well on another, presumably different distribution, is hampered by this variance. Medical imaging models are typically trained on small, single-scanner datasets that do not generalize well to data obtained from different scanners. The quality of the machine, its parameters, the acquisition protocol, and epidemiological variations among patient populations can all impact imaging data. Domain shift refers to the mismatch between the training data distribution, often collected in clinical research settings, and the real-world clinical data distribution. Limited datasets from a single scanner may fail to capture the full range of variability present in imaging data from di-

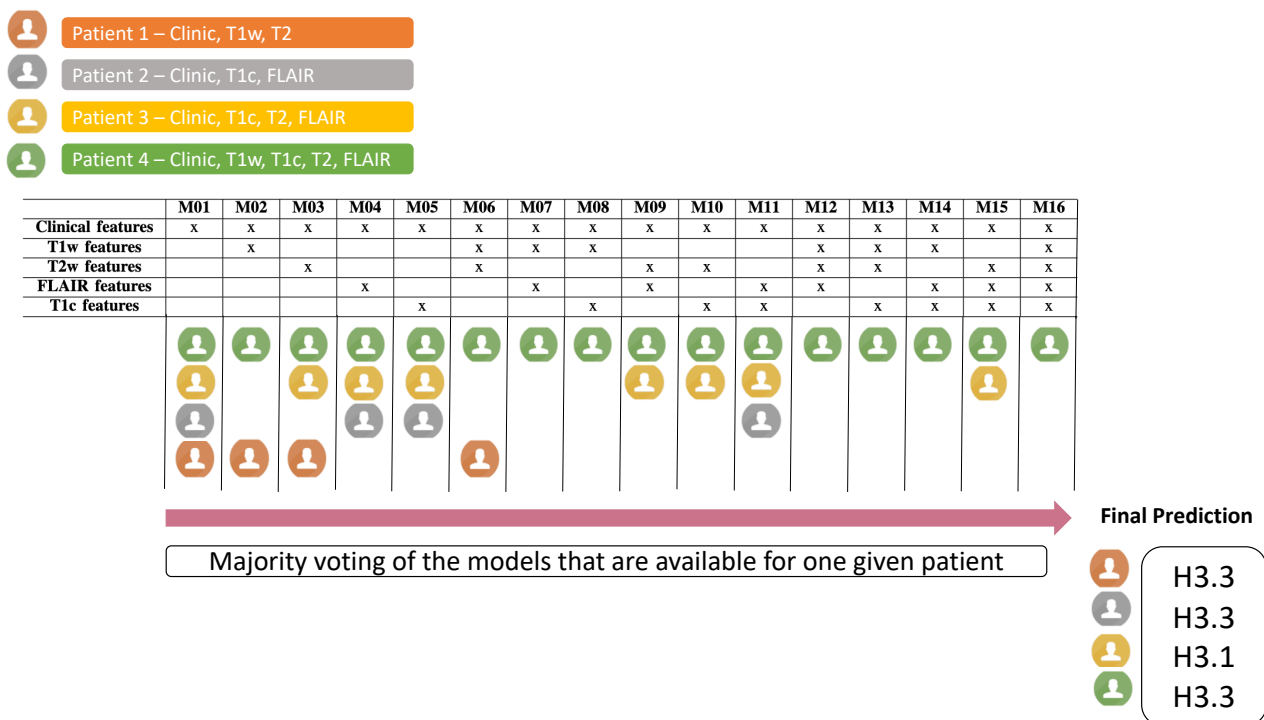


Figure 3.1: The figure illustrates four hypothetical patients evaluated by the 16 model approach.

verse scanners commonly used in clinical practice. This lack of generalizability hinders the successful application of these models to real-world scenarios. Models trained on small datasets predominantly representing a specific population which may struggle to generalize their predictions to populations with different characteristics. To overcome the challenges of domain shift, researchers have proposed various strategies. Augmenting training datasets with a wider range of scanner types and imaging protocols allows models to capture greater variability and improve generalizability. Transfer learning techniques, involving pre-training models on specific datasets and fine-tuning on smaller clinical datasets, enable adaptation to the specific characteristics of the target domain. Addressing domain shift is essential for the successful clinical implementation of automated medical image analysis, especially those based on artificial intelligence. By developing models that can effectively generalize across different scanners, acquisition protocols, and patient populations, we can enhance the reliability and applicability of these automated approaches in real-world healthcare settings. This has been demonstrated in the following studies [183], [184], [185], [186], [187] and [188]. Several solutions have been proposed to solve the problem of domain shift. For instance in [187], [188] and [186] by merging information from various scanners, the heterogeneity of training datasets was increased to produce more generalizable models. This method's drawback is that data must be shared between centers, which is sometimes impractical due to concerns about data privacy. Utilizing transfer learning, which was initially developed using data from a different scanner, is an additional approach. In this case, only the model needs to be shared; the data need not be. This method has been used to adapt models in several studies [189], [190], [191] and [192].

In prospective clinical studies, it is advisable to synchronize imaging protocols at an earlier stage across multiple centers, thereby mitigating the potential influence of disparate imaging protocols on feature values [193], [194]. In order to reduce the effect of multi scanner acquisitions several groups have proposed to reduce the variability by re-sampling the images to a common voxel size or by filtering the images to match spatial resolution [195], [196]. However, this requires accessing the images retrospectively and the filtering procedure reduces the quality of images acquired using the most recent devices. Others apply a z-score transformation [197] to each feature value based on mean and standard deviation measured in each center for that feature, but this assumes that images produced by the different centers have been obtained in similar patient samples (e.g., same proportion of advanced and early-stage tumors), which is sometimes difficult to achieve. In genomics, researchers face a similar problem called batch effect and caused by the handling of scans by different scanners, different technicians, or different days that can obscure individual variations. To deal with that problem in genomics, the ComBat realignment method was introduced [198].

To reduce batch effects and improve the dependability of radiomic characteristics, ComBat harmonization was also applied to radiomic analysis. Radiomic readings may contain unwelcome variability brought on by batch effects, which might result from changes in acquisition techniques or scanner types [183], [199]. The effect of ComBat harmonization on radiomic characteristics in MR imaging has been investigated in a number of research. For instance, in the work of Vallières et al. (2015) [105] improved the prediction accuracy of lung metastases in soft-tissue sarcomas by using ComBat harmonization to coupled **FDG-PET** and MRI texture data. The study published by [200] investigated the application of ComBat harmonization in brain MRIs, highlighting its potential for lowering batch effects and increasing cancer recurrence predicting models. The work of Parmar et al. (2015) [101] highlighted the impact of batch effects on radiomic characteristics and stressed the importance of harmonization strategies in enhancing radiomics-based biomarkers, although not being particular to ComBat. Collectively, these investigations highlight how crucial harmonization is for improving the reliability and consistency of radiomic characteristics.

3.2 Paper published at the 2021 annual international conference of the IEEE in medicine and biology society

Impact of ComBat and a Multi-Model approach to deal with multi-scanner and missing MRI data in a small cohort study. Application to H3K27M mutation prediction in patients with DIPG.

Fahad Khalid¹, Jessica Goya-Outi¹, Vincent Frouin², Nathalie Boddaert³, Jacques Grill⁴ and Frédérique Frouin¹

Abstract—Radiomics was proposed to identify tumor phenotypes non-invasively from quantitative imaging features. Calculating a large amount of information on images, allows the development of reliable classification models. In multi-modal imaging protocols, the question arises of adding an imaging modality to improve model performance. In addition, in the implementation of clinical protocols, some modalities are not acquired or are of insufficient quality and cannot be reliably taken into account. Furthermore, multi-scanner studies generate some variability in the acquisition and data. Some methodological solutions using ComBat and a multi-model approach were tested to take these two issues into account. It was applied to a cohort of 88 patients with Diffuse Intrinsic Pontine Glioma (DIPG). Sixteen models using radiomic features computed using 0, 1, 2, 3 or 4 MRI modalities were proposed. Based on Leave-One-Out Cross-Validation, F1 weighted scores ranged from 0.66 to 0.85. A model of majority voting using the prediction of all the models available for one given patient was finally applied, reducing drastically the number of unclassified patients.

Clinical relevance— In case of patients with DIPG, the prediction of H3 mutation is of prime importance in case of inconclusive biopsy or in the absence of it. It could suggest orientations for new chemotherapy drugs associated with the radiation therapy.

I. INTRODUCTION

Diffuse intrinsic pontine glioma (DIPG) is a highly aggressive pediatric tumor, with a median survival of 9–11 months [1]. Due to its position in the brainstem, surgical intervention is not an option, and conventional chemotherapy has proven to be ineffective. Currently, radiation therapy is the only standard care that temporarily mitigates the symptoms, delays disease progression, and extends median survival by a few months. Recent studies have shown that approximately 80% of DIPG harbor mutations at genes encoding histone H3K27M. Most current mutations are H3.1 (HIST1H3B) and H3.3 (H3F3A). These mutations are currently identified from

*This work was supported by Gustave Roussy grant CAJ 2020.088

¹LITO U1288, Inserm - Institut Curie, 91400 Orsay, France
fahad.khalid@inserm.fr, jessicaouti@gmail.com
frederique.frouin@inserm.fr

²GAlIA, Neurospin, CEA, 91191 Gif-sur-Yvette, France
vincent.frouin@cea.fr

³Pediatric Radiology Department, Hôpital Necker Enfants Malades, AP-HP; IMAGINE Institute, Inserm, Université de Paris, 75015 Paris, France
nathalie.boddaert@aphp.fr

⁴Department of Pediatric and Adolescent Oncology, Gustave Roussy, Inserm, Université Paris-Saclay, 94800 Villejuif, France
jacques.grill@gustaveroussy.fr

biopsy samples and are associated with patient response to therapy [2]. Some clinical trials to assess therapy options according to these mutations are currently under investigation. In a previous work, we have proposed to predict the two types of histone H3K27M mutations non-invasively using MRI-based radiomic features [3]. The ultimate achievement would be to define whether this could avoid biopsy, or at least replace it when it is not feasible or not conclusive, and guide patient care from diagnosis time. The present work aims at optimizing this predictive model [3] in a larger cohort. However, the introduction of new patients (coming from the same center) introduces some variability in the image database due to the use of two different scanners (1.5T and 3T scanners). Furthermore, our first model was based on the joint use of clinical data and four types of structural MR images: T1-weighted (T1w), T2-weighted (T2w), T1-weighted post-contrast injection (T1c) and T2-weighted FLAIR (FLAIR) images. To increase the number of patients, and the predictive power of the prediction models, we consider patients having less than four modalities (at least one among the four).

Radiomics consists in the extraction of quantitative imaging features to identify tumor phenotypes with some predictive values. It faces the critical issue of lack of reproducibility that hampers the successful translation of radiomic model discovery into better diagnosis, patient classification or monitoring. With the introduction of an additional cohort with differences in scanner field and settings, radiomic features were expected to differ between the two scanners, hence ComBat harmonization was introduced [4]. In addition to image intensity standardization [5], ComBat is dedicated to the harmonization of the radiomic features which are associated with one specific tissue, the tumor in the present case. Furthermore, to take advantage of all the MR modalities available for each patient, a multi-model approach is built, using the 16 combinations of the four MR modalities.

II. PATIENTS AND METHODS

A. Clinical and image data

This monocentric retrospective study (2014-2019) included 88 patients with DIPG, scanned at the diagnosis time with one of the two scanners of our center and at least one of the four structural MRI modalities: T1w, T2w, T1c and FLAIR (see Table I).

TABLE I
MRI DATASET PROPERTIES

	1.5T	3.0T
Patients	71	17
T1w	55	10
T1c	52	12
T2w	51	15
FLAIR	50	12

A total of 88 patients were scanned using either a 1.5T (Signa HDxt, GE Medical Systems) or a 3T scanner (MR-750, GE Medical Systems). Among them, 17 presented H3.1 mutation type, 47 H3.3 mutation type and 24 were wild type (WT), or presented another mutation type, or their mutation status was unknown. A total of 17 patients was scanned using the 3T scanner; among them, 9 were H3.3 mutated, 2 were H3.1 mutated, and 6 mutations were unknown or wild type. The clinical feature set consisted of age at the time of diagnosis and sex of patients (see Table II). Patients with H3.3 mutations were older at diagnosis than patients with H3.1 mutation (Wilcoxon test, $p < 0.01$). Table III indicated the number of patients with H3.1 or H3.3 mutation available by considering each combination of 1 to 4 MR modality. Only 47/64 patients (73%) presented the four MR modalities.

TABLE II
CLINICAL CHARACTERISTICS OF PATIENTS INCLUDED IN THE STUDY

	H3.1	H3.3	WT/Unknown
Patients	17	47	24
Age (years)	4.9 ± 1.7	8.7 ± 3.6	8.8 ± 6.1
Girls/boys	10g/7b	22g/25b	9g/15b

B. Image feature extraction

Images were pre-processed by a dedicated pipeline [5] including intensity standardization according to the hybrid white stripe approach, resampling to isotropic voxels (1 mm^3) and multi-modal image registration to each T2w scan (when available, T1 or FLAIR otherwise). For each patient, a large spherical region was drawn inside the tumor on the T2w scans (if available, T1 or FLAIR otherwise) and reported in T1w, T1c and FLAIR scans. Radiomic features were extracted using PyRadiomics [6]. A total of 79 features including first-order and texture features were computed for each MRI modality.

C. Harmonization of features using ComBat

In brain MR, standardization approaches have been proposed to correct for the intensity variability. For instance, the hybrid white stripe method proved to be successful in the context of neurodegenerative diseases and brain cancer [5]. However, we showed that this procedure was not sufficient to explain differences in radiomic features observed for the same patients undergoing 1.5T and 3T scans [4]. The further use of ComBat to harmonize radiomic features has been validated. For each radiomic feature y , computed for one

given modality, measured in patient j with scanner i , the scanner effect on feature y_{ij} can be modelled as (1):

$$y_{ij} \sim \alpha + \gamma_i + \sigma_i \epsilon_{ij}, \quad (1)$$

where α is the overall value of the radiomic feature y , γ_i is an additive scanner effect and σ_i a multiplicative scanner effect associated to an error term ϵ_{ij} . ComBat estimates the $\hat{\alpha}$, $\hat{\gamma}_i$, and $\hat{\sigma}_i$ terms, and corrected values y_{ij}^* are computed by (2):

$$y_{ij}^* = \frac{y_{ij} - \hat{\alpha} - \hat{\gamma}_i}{\hat{\sigma}_i} + \hat{\alpha} \quad (2)$$

Finally, values y_{2j}^{**} obtained for the second scanner (3T) are realigned to the values obtained for the first scanner (1.5T) according to (3), with $\hat{\mu} = \hat{\sigma}_1 / \hat{\sigma}_2$:

$$y_{2j}^{**} = \hat{\mu} y_{2j} + \hat{\alpha}(1 - \hat{\mu}) + (\hat{\gamma}_1 - \hat{\mu} \hat{\gamma}_2) \quad (3)$$

D. Machine Learning Models

Five feature sets were considered as the inputs of the predictive models: one feature set per MRI modality and one clinical feature set. To benefit from all the available patient modalities, 16 models ($M_k, 1 \leq k \leq 16$) were built. Table III represents for each model the type of modality it accepts. This allows the original data set to stay unchanged while addressing the missing data handling problem. For each model, a three steps selection procedure was applied to the imaging features:

- Features were selected according to their robustness to the spherical delineations. Based on features computed in dilated and eroded versions of the tumor region, the absolute agreement intraclass correlation coefficient (ICC) of each feature was computed. Only features with $ICC > 0.9$ were kept.
- Only features presenting an individual Area Under the Receiver Operating Characteristic Curve (AUC) greater than 0.75 for the classification task were kept. This threshold was defined to exclude features that could degrade the model.
- To reduce the total number of features, hierarchical clustering was performed, keeping the minimum absolute Spearman's rank-order correlation between cluster members greater than 0.85. The feature with the greatest AUC of each cluster was finally selected.

Classification task was then achieved using a logistic regression model. Leave-One-Out Cross-Validation (LOOCV) was applied systematically to estimate the performance of each of the 16 models. Feature selection and standardization was performed inside each LOOCV fold, as described in [3]. All experiments have been achieved using radiomic features before the use of ComBat and after the use of ComBat. Finally, an additional prediction model (MMV) was defined as the majority voting of all the models in which each patient can participate.

TABLE III

TABLE SHOWING THE INFORMATION (MARKED BY 'X') USED FOR THE DESIGN OF EACH OF THE 16 MODELS, AND THE NUMBER OF PATIENTS THAT ARE AVAILABLE TO ESTIMATE EACH MODEL

	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12	M13	M14	M15	M16
Clinical features	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
T1w features		x				x	x	x				x	x	x		x
T2w features			x			x			x	x		x	x		x	x
FLAIR features				x			x		x		x	x		x	x	x
T1c features					x			x		x			x	x	x	x
Nb of Patients	64	54	54	51	53	52	51	49	51	52	48	50	49	47	48	47
H3.1	17	13	13	12	14	12	13	12	12	14	12	12	13	12	12	12
H3.3	47	41	41	39	39	40	38	37	39	38	36	38	36	35	36	35

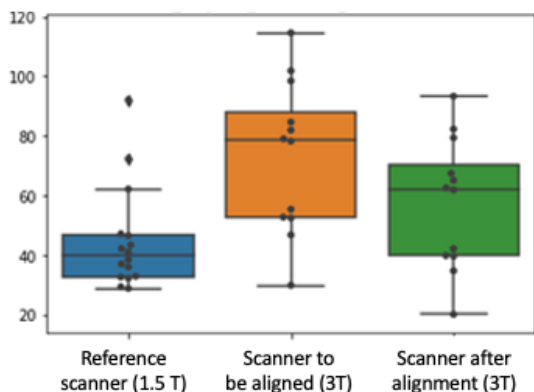


Fig. 1. Box plots of one radiomic feature (FLAIR 90 Percentile) for values coming from the 1.5 T scanner (left), the 3T scanner to be aligned (middle) and the 3T scanner after the realignment procedure using ComBat (right).

III. RESULTS

A. ComBat harmonization

Fig. 1 illustrates the values of one specific radiomic feature (90 Percentile computed in the tumor region of FLAIR image) for the patients acquired with 1.5 T and 3T scanners. After the realignment based on ComBat procedure, we observe a reduction of the values issued from the 3T scanner, which better fit with the values coming from the 1.5T scanner. The ComBat procedure was applied to each radiomic feature independently. Table IV provides the selected features by each of the 16 models without and after the realignment procedure. Following the feature selection, twelve parameters (out of the 318 tested) are involved in the building of the 32 models. Age is selected by all the models. After the realignment procedure based on ComBat, the 15 radiomic models (M02 to M16) are reduced to two features: age and one radiomic feature, this number of features being equal to 2 (7 models), 3 (2 models) and 4 (5 models) when ignoring the realignment procedure. Table V provides the F1 weighted score, obtained by the 16 models following LOOCV without and with realignment. The model showing the lowest performance (M01) is highlighted in red color and the model generating the highest performance (M02) is highlighted in blue color.

B. Prediction of mutation using the multi-model approach

Fig. 2 provides the prediction results obtained by the model showing the highest performance according to LOOCV (M02) and the majority voting process (MMV). Both models used radiomic features after their realignment using ComBat. Of note, for the model MMV, results are quite similar before and after ComBat, one H3.3 case that was miss-classified before ComBat was classified as undecided after ComBat, with an equal number of votes for each class. The number of undetermined cases is drastically reduced when using MMV : two patients are left undecided, with a equal number of votes for both tumor mutation whereas 10 patients could not be classified using M02 approach, due to missing data (lack of T1 modality in that case).

	M02	MMV
True prediction	43 (13 H3.1 ; 30 H3.3)	48 (15 H3.1 ; 33 H3.3)
False prediction	11 (0 H3.1 ; 11 H3.3)	14 (1 H3.1 ; 13 H3.3)
Undetermined cases	10 (4 H3.1 ; 6 H3.3)	2 (1 H3.1 ; 1 H3.3)

Fig. 2. Prediction results provided by the best radiomic model (M02) and by the majority voting approach (MMV).

IV. DISCUSSION

The presented approach makes it possible to adapt prediction of the H3 mutation to the real conditions of an examination and to propose models that adapt to the available data. Indeed, the prediction of the H3 mutation is of prime interest in cases where it is not possible to perform the biopsy or when its results are not conclusive. The different proposed models will be tested on additional data coming from a new clinical trial. As the number of subjects will increase, it will then be possible to refine the models by possibly incorporating more clues. The model M16 that we initiated in [3] incorporates patients with all the 4 MR modalities, in our augmented database, about 25% of the patients could not be analyzed using this model. Model M01 is based on age only and clearly shows worse performance than other models using LOOCV (see Table V). The development of the multi-model approach fills to main objectives, it avoids any imputation based methods for missing data handling along with the benefit of pooling in additional data if a modality is made available for the study.

The MR scanner affects the radiomic feature values extracted from MR images, introducing major confounding factors in

TABLE IV
FEATURES SELECTED BY EACH MODEL BEFORE (IN BLACK COLOR) AND AFTER (IN BLUE COLOR) REALIGNMENT PROCEDURE USING COMBAT FOR THE CLASSIFICATION TASK (H3.1 VERSUS H3.3 MUTATION).

	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12	M13	M14	M15	M16
Clinical	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age
T1w		Id1 ¹				Id6	Id1	Id6				Id1	Id1	Id1		Id1
T2w			Id2													
FLAIR				Id3, Id4			Id3, Id4		Id3	Id3, Id4		Id3	Id3	Id3	Id3	Id3
T1c					Id5					Id5	Id5		Id5	Id5	Id5	Id5
Clinical	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age	Age
T1w		Id6				Id9	Id9	Id9				Id6	Id9	Id6		Id9
T2w			Id2						Id2							
FLAIR				Id7												
T1c					Id8					Id10	Id11					Id11

¹Id1: T1w firstorder median Id2: T2w glm entropy Id3: FLAIR glm homogeneity1 Id4: FLAIR firstorder 90Percentile Id5: T1c firstorder 10Percentile Id6: T1w firstorder root meansquared Id7: FLAIR glszm small area emphasis Id8: T1c glm Idn Id9: T1w firstorder mean Id10: T1c glrlm short run emphasis Id11: T1c glm Entropy

TABLE V
F1 WEIGHTED SCORE OBTAINED BY LOOCV FOR THE 16 MODELS WITHOUT AND WITH REALIGNEMENT BASED ON COMBAT

Model	without ComBat	with ComBat
M01	0.66	0.66
M02	0.88	0.85
M03	0.70	0.69
M04	0.75	0.68
M05	0.71	0.74
M06	0.87	0.78
M07	0.82	0.77
M08	0.84	0.82
M09	0.76	0.74
M010	0.70	0.75
M011	0.77	0.77
M012	0.82	0.75
M013	0.83	0.82
M014	0.81	0.78
M015	0.77	0.77
M016	0.81	0.78

multi-centric studies [4]. Here, we validated a harmonization procedure ComBat realignment for MR radiomic features extracted from different scanners. In patients scanned with 1.5T and 3T scanners, we showed that this harmonization procedure realigns radiomic feature distributions (Fig. 1). In radiomics especially in the light of oncological studies, pooling images acquired using different devices and different acquisition and reconstruction protocols is often needed to increase the size of cohort, or combining different cohorts. In that context, we demonstrated that ComBat could realign feature values so that all data could be analyzed together, even if images had been acquired with different scanners. It is important to highlight that the effects of ComBat on our prediction task are small in with respect to the final decisions. However, changes are seen in feature selection by individual models to decide upon the type of mutation as shown in the Table IV. Fewer features are selected after the feature were subjected to ComBat harmonization. This further reinforces the positive impact of ComBat as it seems to increase the level of robustness. This effect needs to be further investigated, increasing the number of patients with 3T scans.

V. CONCLUSION

The findings of this work demonstrated ComBat harmonization method could efficiently remove the scanner/protocol effect while preserving the individual variations in MR modalities coming from different patients and scanners. Furthermore, it allows the data set to stay unchanged without the need for adding artificially constructed data addressing missing data problem which is commonly used in medical imaging. This approach enables large MR multicentric studies to highlight the added value of radiomic analysis in features acquired from different scans. Furthermore, ComBat harmonization may display visible change in values and rather adds a level of robustness. The multi-model concept utilizes all the available data performs well due to the individual model prediction mechanism. Voting by each model could be associated to a level of confidence for each prediction.

REFERENCES

- [1] Hargrave D, Bartels U, Bouffet E, Diffuse brainstem glioma in children: critical review of clinical trials, *Lancet Oncol*, 7:241–8, 2006. DOI: 10.1016/S1470-2045(06)70615-5
- [2] Castel D, Philippe C, Calmon R et al, Histone H3F3A and HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different prognosis and phenotypes, *Acta neuropathologica*, 130(6):815-827, 2015. DOI: 10.1007/s00401-015-1478-0
- [3] Goya-Outi J, Calmon R, Orhac F et al, Can Structural MRI Radiomics Predict DIPG Histone H3 Mutation and Patient Overall Survival at Diagnosis Time?, 2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI):1-4, 2019. DOI: 10.1109/BHI.2019.8834524
- [4] Orhac F, Lecler A, Savatovski J et al, How can we combat multi center variability in MR radiomics? Validation of a correction procedure, *Eur Radiol*, 31(4):2272-80, 2021. DOI: 10.1007/s00330-020-07284-9
- [5] Goya-Outi J, Orhac F, Calmon R et al, Computation of reliable textural indices from multimodal brain MRI: suggestions based on a study of patients with diffuse intrinsic pontine glioma, *Phys Med Biol*, 63(10):105003, 2018. DOI: 10.1088/1361-6560/aabd21
- [6] van Griethuysen JJM, Fedorov A, Parmar C et al, Computational radiomics system to decode the radiographic phenotype, *Cancer Res*, 77(21):e104–7, 2017. DOI: 10.1158/0008-5472.CAN-17-0339

3.3 Conclusion

In conclusion, the automatic analysis pipeline suggested in this chapter uses a multi-model strategy that includes 16 different models. This design mimics real-world clinical situations closely and strives to adapt to missing modalities. The pipeline's adaptability is demonstrated by its ability to handle patient exams at several locations and at various times. Additionally, each patient's output for predicting the H3K27M mutation is guaranteed by the Leave-One-Out Cross-Validation (LOOCV) approach, offering a vital tool for quantitative comparisons between various patients.

The models' selectivity, with each model often picking just two features, is significant. It's interesting to note that various models occasionally choose the same features. This may cause duplication, but it also highlights how important some features are. For instance, the fact that different models frequently choose the clinical feature "age" shows that it has increased significance. Such knowledge could help determine which traits should be given more weight, thereby improving prediction accuracy.

Although the literature [201] [202] implies that ComBat may not be the optimal choice for every dataset, its application has its merits. The box plot evidence underscores a more aligned representation of features from two distinct scanners post-ComBat application. Additionally, ComBat harmonization affects the choice of features. Notably, after harmonization, fewer traits are selected, suggesting less redundancy. Thankfully, this simplified option does not compromise forecast accuracy. These observations have led us to decide to keep ComBat harmonization in the finished pipeline.

The pipeline does have certain restrictions, though. The 16 models use a majority vote process to determine the final forecasts. This strategy could result in situations where the votes are evenly divided, making a definitive forecast difficult to make. The next chapter's main focus will be on dealing with this uncertainty and the aforementioned issues.

Future research on DIPG stand to benefit greatly from the pipeline architecture, which enables the smooth addition of additional modalities to the database. The pipeline developed in [2], lays the door for an expanded analysis and assessment of imaging data coupled with clinical aspects, giving researchers a powerful tool for further initiatives.

Chapter 4

A second Multi-Model approach based on the complementarity between the different sources of data

4.1 Introduction

This chapter introduces a novel prediction pipeline that builds on the strengths of radiomics and machine learning. Specifically, we focus on three key aspects of the approach: feature selection, performance, and interpretability. Feature selection involves identifying the most relevant features for the prediction model, while performance assesses the accuracy and robustness of the model. Interpretability aims to provide insights into how the model arrived at its predictions. We do not fill in missing data with estimated values. Instead, we have implemented a specialized approach that takes into account the missing data and adjusts the model accordingly. This consideration ensures that our prediction pipeline is robust and can provide accurate predictions even when there is missing data in the input. By addressing this limitation, we are confident that our approach will provide valuable insights for medical imaging analysis and contribute to advancing the field of radiomics and machine learning. This pipeline was initially tested for the prediction of H3.1 versus all (H3.3,H3.2, wild type), then applied to the prediction of ACVR1 and TP53 mutations.

The paper published in *Frontiers in Medicine* and the second part providing additional experiments analysing the approach from different perspectives.

4.2 Paper published in Frontiers in Medicine

Multimodal MRI radiomic models to predict genomic mutations in diffuse intrinsic pontine glioma (DIPG) with missing imaging modalities

Fahad Khalid¹, Jessica Goya-Outi¹, Thibault Escobar^{1,2}, Volodia Dangouloff-Ros^{3,4}, Antoine Grigis⁵, Cathy Philippe⁵, Nathalie Boddaert^{3,4}, Jacques Grill^{6,7}, Vincent Frouin⁵ and Frédérique Frouin^{1,*}

¹Laboratoire d'Imagerie Translationnelle en Oncologie (LITO) - U1288, Institut Curie, Inserm, Université Paris-Saclay, Orsay, France

²DOSIsoft SA, Cachan, France

³Paediatric Radiology Department, Hôpital Universitaire Necker Enfants Malades, AP-HP, Paris, France

⁴Institut IMAGINE, Inserm U1163 and U1299, Université Paris Cité, Paris, France

⁵Neurospin, Institut Joliot, CEA, Gif-sur-Yvette, France

⁶Département Cancérologie de l'enfant et de l'adolescent, Gustave-Roussy, Villejuif, France

⁷Prédicteurs moléculaires et nouvelles cibles en oncologie - U981, Inserm, Université Paris-Saclay, Villejuif, France

Correspondence*:

Frédérique Frouin, LITO-U1288, Inserm, Centre de Recherche de l'Institut Curie, Bâtiment 101B, rue de la Chaufferie, 91400 Orsay, France
frederique.frouin@inserm.fr

2 ABSTRACT

3 **Purpose:** Predicting H3.1, TP53, and ACVR1 mutations in DIPG could aid in the selection
4 of therapeutic options. The contribution of clinical data and multi-modal MRI were studied for
5 these three predictive tasks. To keep the maximum number of subjects, which is essential for a
6 rare disease, missing data were considered. A multi-modal model was proposed, collecting all
7 available data for each patient, without performing any imputation.

8 **Methods:** A retrospective cohort of 80 patients with confirmed DIPG and at least one of the
9 four MR modalities (T1w, T1c, T2w, FLAIR), acquired with two different MR scanners was built. A
10 pipeline including standardization of MR data and extraction of radiomic features within the tumor
11 was applied. The values of radiomic features between the two MR scanners were realigned
12 using the ComBat method. For each prediction task, the most robust features were selected
13 based on a recursive feature elimination with cross-validation. Five different models, one based
14 on clinical data and one per MR modality, were developed using logistic regression classifiers.
15 The prediction of the multi-modal model was defined as the average of all possible prediction
16 results among five for each patient. The performances of the models were compared using a
17 leave-one-out approach.

18 **Results:** The percentage of missing modalities ranged from 6% to 11% across modalities and
19 tasks. The performance of each individual model was dependent on each specific task, with
20 an AUC of the ROC curve ranging from 0.63 to 0.80. The multi-modal model outperformed
21 the clinical model for each prediction tasks, thus demonstrating the added value of MRI.
22 Furthermore, regardless of performance criteria, the multi-modal model came in the first place
23 or second place (very close to first). In the leave-one-out approach, the prediction of H3.1 (resp.
24 ACVR1 and TP53) mutations achieved a balanced accuracy of 87.8% (resp. 82.1%, and 78.3%).

25 **Conclusion:** Compared with a single modality approach, the multi-modal model combining
26 multiple MRI modalities and clinical features was the most powerful to predict H3.1, ACVR1,
27 and TP53 mutations and provided prediction, even in the case of missing modality. It could be
28 proposed in the absence of a conclusive biopsy.

29 **Keywords:** MRI, Radiomics, Prediction, Missing Data, Genomic Mutation, Diffuse Intrinsic Pontine Glioma

INTRODUCTION

30 The diffuse intrinsic pontine glioma (DIPG) is a highly aggressive pediatric tumor, with a median overall
31 survival of 11 months (1, 2). Since this tumor is inoperable, radiotherapy is the standard option that
32 can be proposed systematically, generating in most cases transient improvement (3). Genomic analyses
33 based on tumor biopsies have shown that more than 85% of patients with DIPG harbor mutations (4, 5)
34 at genes encoding histone H3, leading to lysine 27 to methionine substitution (H3-K27M). The new
35 WHO classification of this disease is diffuse midline gliomas, H3 K27-altered (6). Most frequent H3-
36 K27 alterations are H3.1 and H3.3 variants. These two alterations and the H3-wildtype are associated
37 with different age profiles and different overall survivals, patients with H3.1 being younger, having
38 better response to radiotherapy and better overall survivals (1). Furthermore, these H3 K27M mutations
39 are frequently associated with TP53 and ACVR1 somatic mutations (7). If TP53 mutations are mainly
40 encountered in H3.3 patients while ACVR1 mutation mostly occur in H3.1 patients, these mutations need
41 to be separately identified for testing new chemotherapy options. It was recently shown that TP53 mutation
42 can drive radio-resistance in patients with DIPG (8). Thus, the knowledge of this mutation could help to
43 refine re-irradiation strategies. Furthermore, the combination of vandetanib and everolimus was identified
44 as a possible therapeutic option for patient harboring ACVR1 mutations (9). These recent advances in
45 the DIPG patient care, raised the issue of predicting H3.1, ACVR1, and TP53 mutations within tumor
46 independently from each other, using data available at diagnosis time: basic clinical data (age and sex)
47 and multi-modal MRI to help define a personalized treatment strategy when brain biopsy is not possible
48 or is not conclusive.

49 Indeed, multi-modal MRI images are always acquired to confirm diagnosis (10, 11). These data could
50 also be used for radiogenomic prediction tasks, provided that some pre-processing steps are taken.
51 Radiomics is a recent field of research which refers to the comprehensive and automated quantification
52 of this radiographic phenotype (12, 13). This approach aims at enhancing some relevant information
53 contained in the images and made them available to clinicians. It is based on medical image post-
54 processing algorithms and features computation from specific regions of interest (14, 15, 16). Radiomic
55 features belong to different families, including morphological, global image intensity, histogram image
56 intensity distribution and texture families. Texture indices are based on image intensity comparison
57 between neighboring voxels, and potentially reflect biological properties such as tumor heterogeneities
58 (12, 17, 18). The high number of radiomic features and their systematic analysis have accelerated the

59 discovery of potential new biomarkers and has definitively modified the research tools in radiology and
60 nuclear medicine, giving a larger weight to data analysis. However, end-users of radiomic tools should be
61 aware of the pitfalls inherent in these tools (16, 19), including the dependency of radiomic features values
62 to the acquisition parameters and to software implementation, and thus the need of image preprocessing
63 to make these features more reproducible.

64 Magnetic resonance imaging (MRI) with its high spatial resolution and high brain tissue contrast is the
65 imaging modality of choice for children with central nervous system tumors. Current recommendations
66 include the acquisition of T1-weighted images without contrast (T1w) and following the injection of
67 gadoterate meglumine (T1c), T2-weighted images (T2w), and fluid attenuated inversion recovery images
68 (FLAIR) (10, 11). As MRI intensities are non-standardized (20), this prevents the extraction of robust
69 radiomic features, except if specific standardization procedures are defined, including the use of similar
70 pulse sequence parameters and identical size of voxels, and applying image intensity normalization as
71 a preprocessing step (21, 22, 23). Of course, intensity variations depend on the MR scanner and the
72 acquisition parameters, but also on each acquisition. To reduce this variability, many approaches have been
73 proposed (24), including Z-score normalization, and dedicated procedures using a reference tissue, such
74 a white matter for brain studies (25). A refined procedure was proposed for patients with DIPG, removing
75 the slices corresponding to pontine location to avoid the inclusion of the tumor in the normalization
76 process (21). However, despite intensity standardization, some variations in the radiomic features can be
77 due to coils, scanners and/or scanning parameters as it was demonstrated on a breast phantom study (26).
78 To reduce this impact, the ComBat method, providing harmonization of radiomic features across different
79 acquisition scanners (27, 28), has been proposed.

80 In the constitution of our global approach, two specific issues were taken into account: 1) missing data:
81 due to practical constraints some MRI modalities were missing or non-usable; 2) data scarcity for the
82 training of our model: the cohort of patients was small, since DIPG is a rare disease. A compromise
83 was made to incorporate as much relevant information as possible. In a preliminary work of our group, a
84 radiomic model was proposed to distinguish the two types of histone H3-K27M mutations (H3.3 versus
85 H3.1) using a subset of patients having the four MRI modalities (T1w, T1c, T2w, and FLAIR) and clinical
86 data (29). To increase the number of patients (about 20% for each prediction task), all the patients having
87 at least one of the four MRI modalities were included. To have a prediction for each patient, a multi-
88 model strategy was proposed using all the data types among clinical, T1w, T1c, T2w, and FLAIR that
89 were available.

MATERIALS AND METHODS

90 Patient Database

91 This retrospective mono-centric study includes 80 patients having DIPG, who had biopsy and were
92 treated between 2009 and 2018 at Gustave Roussy cancer center (Villejuif, France). Patients were scanned
93 at the time of diagnosis, before biopsy, with either Signa HDx, 1.5T (GE Healthcare) MRI machine or
94 Discovery MR750w, 3T (GE Healthcare) MRI scanner in the pediatric radiology department at Necker
95 Hospital (Paris, France).

96 At least one of the four structural MRI modalities (T1w, T1c, T2w and FLAIR) (see Table 1) was
97 acquired and basic clinical information (age and sex) was also collected. Typical acquisition parameters
98 were described in (21). Figure 1 shows two cases of patients issued from the database. A total of 57 (71%)
99 patients had the four MRI modalities (T1w, T1c, T2w and FLAIR) of sufficient quality, the remaining

100 patients had at least one missing MRI modality. Following the genomic analysis consecutive to biopsy, 63
101 patients have known H3 status, 63 patients (partly different from the H3 subgroup) have known ACVR1
102 status and 61 patients have known TP53 status, as summarized in Table 2. For histone H3, the H3.1
103 mutation was observed in 14 patients, the H3.3 mutation in 44 patients, the H3.2 mutation in 1 patient
104 and histone H3 wild type in 4 patients. Due to the small numbers in the last two classes, the binary task
105 was to predict patients with H3.1 mutation against all other patients grouped together. Three binary tasks
106 of classification were thus defined: prediction of H3.1, ACVR1 and TP53 mutations. Figure 2 gives an
107 overview of the construction of the model which is defined for each prediction task, and the different steps
108 are detailed in the following subsections.

109 MRI Preprocessing and Radiomic Features Extraction

110 All MR Images were first processed through a dedicated pipeline fully described in (21) including bias
111 field correction of MRI using N4 algorithm (30), MRI intensity normalization according to an adaptation
112 of the hybrid white stripe approach (25), resampling to isotropic voxels of 1 mm^3 and multimodal image
113 registration on each T2w scan (when available, T1w or FLAIR otherwise) using FSL FLIRT (31).

114 For each patient, a spherical region was drawn (the largest sphere within the tumor) and transferred to the
115 realigned MR volumes. This region always included the location of the biopsy. For each MRI modality, 79
116 radiomic features were extracted within the spherical region using PyRadiomics (14), including 19 first-
117 order features derived from the distribution of intensity inside the tumor and 60 texture features computed
118 using three different matrices: the grey level co-occurrence matrix (GLCM), the grey level run length
119 matrix (GLRLM), and the grey level size zone matrix (GLSZM). All histogram-based and texture-based
120 features were computed with a fixed bin width equal to 2 (21). As the MRI were acquired using two
121 different scanners, the ComBat harmonization (27, 28) was then applied independently to each radiomic
122 feature to make it more comparable across scanners (32). The spherical region is quick and easy to define
123 and has already shown some promising results (32), but it does not bring any information related to the
124 shape of the tumor. To overcome this drawback, tumor contours were delineated by two skilled operators
125 and 14 additional shape features were extracted. As these features were available for each patient, they
126 were further merged with clinical data. Results of this additional study are provided in supplemental data.

127 Feature Selection

128 A recursive feature elimination cross-validation (RFE-CV) method (33) was used to select the most
129 relevant features. This procedure was repeated for each of the three classification tasks and for each
130 modality m ($1 \leq m \leq 5$). It was implemented using the scikit-learn, a free machine learning library in
131 Python (34). The RFE-CV method iteratively fits a model - a logistic regression model was chosen for our
132 application - and removes progressively the weakest feature. Therefore, the RFE-CV method eliminates
133 dependencies and co-linearity between the different features in the model. To apply the L1 penalty used
134 for the logistic regression model, we used a grid analysis introducing a variation (between 0.1 and 1 with
135 a step size of 0.1) for the inverse of the regularization strength, the C parameter. Feature importance was
136 assessed on the validation set by computing the Brier score loss. The RFE process was repeated 40 times,
137 based on a two-fold cross-validation. The up to four most frequently selected features were kept. The
138 RFE-CV provided a subset of K features f_m^k , $1 \leq k \leq K$, associated with the modality m .

139 Definition of the Mono-Modal and Multi-Modal Models

140 Due to the small number of patients and due to missing imaging modalities, a leave-one-out cross-
 141 validation (LOO-CV) framework, named LOO-CV-MIM, was proposed to compare the different models,
 142 as illustrated in Figure 3. For each training set, a logistic regression model was defined, using a L1 penalty
 143 with $C = 0.5$ (the selected features using the previously described RFE-CV procedure were frequently
 144 selected using this C value), and a balanced mode to automatically adjust weights inversely proportional
 145 to class frequencies of the input data. This process was applied separately to each prediction task.

146 To explain the process more deeply, we have to consider every patient P_i , having m_i modalities such as
 147 $2 \leq m_i \leq 5$, since each patient has one clinical modality and at least one among four MR modalities.
 148 For each patient P_i , having the modality m , a logistic regression model M_m^i is built from the K features
 149 f_m^k selected at the previous step, the feature values inserted in the training set being computed for all the
 150 patients for which the modality m is available, except for the patient P_i . The logistic regression model
 151 M_m^i is then tested on the patient P_i , providing a probability $Pr(P_i, M_m^i)$ that the patient P_i had the
 152 mutation under study, according to the model M_m^i . Using these $Pr(P_i, M_m^i)$ values for all the patients, and
 153 the ground truth classification, receiver operator characteristic (ROC) curve is defined and its associated
 154 area under the curve (AUC) (35) is computed as a first figure of merit. After applying the conventional
 155 threshold of 0.5 to define the final classification: if $Pr(P_i, M_m^i) \geq 0.5$, the patient P_i is classified as
 156 having the mutation under study, else as not having this mutation, confusion matrices are then built. Three
 157 additional figures of merit are then computed: sensitivity, specificity (35), and balanced accuracy (mean
 158 value of sensitivity and specificity). The number of patients for which the prediction is possible is defined
 159 as an additional figure of merit.

160 Finally the multi-modal model approach (M_{Multi}^i) is defined, the probability $Pr(P_i, M_{Multi}^i)$ that the
 161 patient P_i has the mutation under study based on this ensemble model is equal to the mean probability
 162 computed for each model M_m^i (see equation 1):

$$Pr(P_i, M_{Multi}^i) = \frac{1}{m_i} \cdot \sum_{m=1}^{m_i} Pr(P_i, M_m^i) \quad (1)$$

163 Since the number m_i of models for one patient P_i is between 2 and 5, the $Pr(P_i, M_{Multi}^i)$ term can be
 164 defined for each patient. The five figures of merit (AUC, sensitivity, specificity, balanced accuracy and
 165 number of patients for which the prediction can be done) are defined for the multi-modal model M_{Multi}^i ,
 166 too.

RESULTS

167 Feature Selection

168 Two clinical features (age and sex) and 79 radiomic features per imaging modality were initially
 169 considered. The RFE-CV procedure was applied to each modality (Clinical, T1w, T1c, T2w, FLAIR)
 170 independently for the three classification tasks (prediction of H3.1, ACVR1 and TP53 mutations). From
 171 1 to 4 features were selected per modality and resulting features are listed for each task in Tables 3, 4 and
 172 5. From clinical data, age was selected for the three tasks. For imaging modalities, in most cases, both
 173 first-order (between 1 and 2) and texture features were jointly selected. The four feature sets selected
 174 for the four MRI modalities, showed some overlap across the three tasks, but none of these subsets
 175 totally overlapped. Supplemental Figure S1 displays the correlogram between the radiomic features (79

per modality) across the 61 patients selected for prediction of TP53 mutation, showing the potential interest of the four modalities, due to low or moderate correlation between features extracted from two different modalities. Supplemental Table S1 provides the features selected when merging clinical and shape features for each of the three classification tasks.

To further investigate the interest of each MR modality, the correlograms between the selected features are displayed in Figure 4. For the prediction of H3.1 mutation (Figure 4A), four features (h3, h8, h12, and h14) extracted from T1w, T1c, T2w and FLAIR MRI showed high correlation. For the prediction of ACVR1 mutation (Figure 4B), three features (a5, a12, and a15) extracted from T1w, T2w and FLAIR MRI were also highly correlated. For the prediction of TP53 mutation (Figure 4C), two features (t2 and t9) extracted from T1w and T2w MRI were also highly correlated. Interestingly, as shown in Figure 4D, all these nine features had correlation greater than 0.73 with the sphere volume, which could be considered as a surrogate marker of the tumor volume. Except for these nine features, there were no high redundancies between selected features extracted from different modalities, showing the high complementarity between these four MRI modalities. Furthermore, no selected radiomic feature was correlated with age.

Prediction Performance

Table 6 reports the five figures of merit (number of cases, AUC, sensitivity, specificity, and balanced accuracy) obtained by the six models, for the three prediction tasks, using the LOO-CV framework. Supplemental Figures S2, S3, and S4 illustrates for each patient the results of the prediction of H3.1, ACVR1, and TP53 mutations by the six types of models: $M_{Clinical}$, M_{T1w} , M_{T1c} , M_{T2w} , M_{FLAIR} and M_{Multi} . Supplemental Table S2 displays the five figures of merit for two additional models: $M_{ClinicalSh}$ and $M_{MultiSh}$ for which the shape features were merged with the clinical features.

Three points emerge from the analysis of these results.

Radiomics increase the performance of the predictors. Indeed, the simple clinical feature "age" provided alone some pretty good results with a balanced accuracy equal to 71.4% for predicting H3.1 mutation, 70.5% for predicting TP53 mutation and 65.3% for predicting ACVR1 mutation. These values could be considered as baseline. When compared to baseline, adding MR radiomic data through the multi-modal model enabled an increase of 16 percentage points of the balanced accuracy for predicting H3.1 and ACVR1 mutations and of 8 percentage points for predicting TP53 mutation. Finally, the addition of the 14 shape radiomic features slightly improved the prediction of TP53 mutation, with an increase of 1.4 percentage point of the balanced accuracy.

Ensembled multi-modal model outperforms mono-modal predictors. Noticeably the multi-model approach provided the best (or second best) performances for all the figures of merit whatever the predictive tasks. Thanks to its inception, it provided a prediction for each patient, even in case of missing MR data. Following results in Table 6, missing MR data varies between 6% and 11%, according to the MR modality and the task of prediction. The AUC associated with the M_{Multi} model was the highest for predicting ACVR1 (0.91) and TP53 (0.88) mutations, and the second highest for predicting H3.1 mutation (0.91 versus 0.92 for M_{T1c}). Sensitivity was the highest for predicting H3.1 and ACVR1 mutation. It reaches the third position for predicting TP53 mutation (67.6% versus 69.7% for M_{T2w} and 71.9% for M_{FLAIR}), but for that task, it achieves the highest specificity. Taking into account the balanced accuracy as a compromise between sensitivity and specificity, this figure of merit was the highest for predicting H3.1 mutation (87.6%) and ACVR1 mutation (82.1%) and the second highest for predicting TP53 mutation (78.3% versus 78.6% for M_{T2w}), having a prediction for the 61 patients versus 57 for

218 M_{T2w} . The same effects were observed when the clinical features were replaced by the clinical and the
219 shape features, showing the value of the multi-modal model in a slightly different configuration.

220 **Each MR modality brings specific information.** Depending on the task, the ranking of the four models
221 built from each MR modality varied. For instance, the T2w modality appears to be less relevant for
222 predicting H3.1 and ACVR1 mutations, but it proves to have very high figures of merit for the prediction
223 of TP53 mutation. The FLAIR modality appears to be very relevant for predicting ACVR1 mutation but
224 less relevant for predicting TP53 mutation. Furthermore, the shape features which could be extracted
225 without missing values could have an impact for predicting TP53 mutation, too. These results underline
226 the necessity to acquire all the structural modalities to achieve multi-objective classification tasks.

DISCUSSION

227 The proposed approach provides a good prediction of three important mutations (H3, ACVR1 and TP53)
228 encountered in patients with DIPG, within a constrained experimental setting including missing data and
229 small cohort. This result could have a real impact in the coming years to propose a more personalized
230 therapy to patients with DIPG. Our approach is based on clinical and MR data and could be applied in
231 case of absent or not conclusive biopsy. As reported in the literature (1), age was shown to be a relevant
232 predictor of the three mutations, but this study shows that some radiomic models can outperform this
233 baseline predictor, with radiomics originated from T1w, T1c and FLAIR for H3.1 mutation, T1c and
234 FLAIR for ACVR1 mutation, and T2w and the shape features for TP53 mutation (see Tables 6 and S2).
235 With our ensembled multi-modal approach, a prediction can be done for each patient, even if she/he
236 lacks one or more MR modalities, and all the figures of merit were among the highest. In the LOO-CV
237 framework, the number of false positive and false negative cases was reduced to 19% (resp. 24% and
238 23%) for the prediction of H3.1 (resp. ACVR1 and TP53) mutations. This DIPG study illustrates thus the
239 positive impact of radiomic approaches for these three predictive tasks.

240 From a methodological point of view, radiomic studies rely on a succession of steps which have to
241 be optimized. As our database is small, several methods are admissible and can bring some equivalent
242 solutions. Users are recommended to follow best practices (36), some of which depending on MRI. In
243 clinical studies involving MRI, we have demonstrated the interest of MR data preprocessing with image
244 standardization (21, 37) and radiomic feature harmonization (26, 28) to provide more comparable features
245 across scanners, sequences and patients. Furthermore, if automatic tumor segmentation is a major issue
246 to solve and requires additional developments, the precision of segmentation that is required depends on
247 the task to solve. It appears for this study of mutation prediction in DIPG, the definition of a large sphere
248 inside the tumor was sufficient to provide good results and the fine delineation of the tumor in 3D was not
249 absolutely necessary for this discovery step. For feature selection, several approaches are possible. Using
250 a different approach based on feature filtering (and not on RFE-CV) in some preliminary works (29, 32),
251 we found that similar features were found to be predictive of H3.1 mutation. As there are many correlated
252 features for the same MR modality (as shown in Supplemental Figure S1), some equivalent models can
253 be defined using different sets of features.

254 This study shows also a pragmatic but efficient approach to deal with missing (or insufficient quality
255 data) MR modalities, while taking advantage of the complementarity among them. Our objective was
256 to use all the information that was available without data rejection or data imputation. Data rejection,
257 for instance removing patients having less than 4 MR modalities, would have considerably reduced the
258 number of cases (from 80 to 57 patients), and therefore likely decreased the performance of the models

259 (38). In their recent study related to prediction of H3K27M mutation in diffuse midline glioma using
260 multi-modal MRI, more than 50% of patients were excluded due to missing data or insufficient quality
261 (39). Our multi-modal model could remedy such a situation, and enable studies with larger number of
262 patients providing more robust results. Among other conventional approaches used to deal with missing
263 data, MR data imputation appeared to be complex for two main reasons: the low number of cases that
264 were initially available, and the low correlation between the features coming from different modalities
265 (except from those which are highly correlated with the volume or the shape of the region), as underlined
266 by Figure 4. For similar reasons, generative adversarial networks (40) to synthesize missing MR volumes
267 were not retained as a feasible option.

268 In our preliminary work (32), 16 models were defined to deal with missing data for the prediction of
269 H3.1 mutation: one clinical model based on age, four mono-modal models combined with age and 11
270 additional models merging two (providing 6 models), three (providing 4 models) and four (providing 1
271 model) MRI modalities. However, these 11 additional models proved to be redundant with the 4 mono-
272 modal models since they were based on very similar sets of features. Thus the majority voting on all
273 possible models that was applied to each patient could be partially biased.

274 Radiogenomic studies in neuro-oncological studies (41) have shown a small number of studies devoted
275 to DIPG or diffuse midline glioma (DMG). For the specific classification tasks we aimed at solving, we
276 did not find any strictly comparable studies. Indeed, if several studies (39, 42, 43, 44) have proposed some
277 radiomic models to distinguish between H3K7M mutation and Histone H3 Wild-Type groups, all of them
278 included an adult population with DMG, which manifest themselves in several different ways compared to
279 pediatric cancers. Therefore, features and models proposed by those studies could not be compared with
280 ours. Furthermore, we did not find any study aiming at predicting ACVR1 mutations or TP53 mutations
281 in patients with DIPG or DMG.

282 Our study presents several limitations. Despite the selection of a reduced number of features (4 or less
283 features per mono-modal model), some over-fitting could still be present, especially for the prediction
284 of H3.1 and ACVR1 mutations, for which the data sets were strongly imbalanced. However, we are
285 confident in the interest of the multi-modal model, since it proves its superiority for the three different
286 tasks considered here. As the different mono-modal models have the same weight in the definition of
287 the multi-modal approach, optimizing their weight according to their performances could also be tested.
288 However, following this direction, first attempts consisting in removing the 'worst' modality did not
289 show any significant changes. The radiological interpretation of selected features, apart those close to
290 volume or shape, needs also to be refined. For this point, we should test the use of decision maps, as
291 recently introduced in (45). Furthermore, a recent study (46) has shown the superiority of segmenting
292 tumor volume over its ellipsoidal approximation to assess tumor burden in DIPG. The fine delineation of
293 contours will make possible to further test the impact of additional morphological features, including
294 the histogram of oriented gradients (47) for the estimation of the genomic mutations. The manual
295 segmentation is however tedious and its reproducibility still needs to be tested. This task is also difficult
296 to automate due to the particularities of DIPG and the difficulties of obtaining a cohort with numerous
297 data (48). Finally, several works remain to be done. To get rid of the data leakage which was present
298 in our feature selection, the external test of our different models should be done to validate them or to
299 propose some simplified models to travel across the different centers. The model of logistic regression
300 was chosen due to its simplicity and its robustness, and this choice proves to be informative in our context
301 of small number of cases and of imbalanced classes. Regarding prediction performance, our results are
302 certainly overestimated, especially with the LOO-CV process. With a larger database, the performances

303 will be better assessed, and different machine learning models could also be tested, tuned, and compared.
304 Measuring the added value of perfusion and diffusion studies (49), for which the number of missing
305 modalities will be higher, is also a challenge to solve. The interest of MR radiomics to define prognosis
306 (50, 51) should also be further analyzed when compared to simpler models (50, 52, 53).

CONCLUSION

307 The interest of using MRI radiomics in addition to clinical data to predict mutations of H3.1, ACVR1
308 and TP53 was shown on a retrospective cohort of 80 subjects. Each MR modality (T1w, T1c, T2w and
309 FLAIR) demonstrates its interest for at least one of the three prediction tasks. Compared to single-modal
310 models, the multi-modal model combining multiple MRI modalities and clinical features was the most
311 powerful and could provide a prediction for every patient, even in the case of missing MR modalities. It
312 could thus be tested as an alternative in the absence of biopsy or in case of non-conclusive results of the
313 genetic analysis.

CONFLICT OF INTEREST STATEMENT

314 The authors declare that the research was conducted in the absence of any commercial or financial
315 relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

316 FF and FK wrote the manuscript draft, all authors approved it. FF and VF designed the study. VDR and
317 NB provided image data and their radiological expertise. JG provided molecular data and his medical
318 expertise. AG, CP, and VF built the image database. FK, JGO, and TE proposed and implemented MR
319 data processing.

FUNDING

320 All the authors thank Imagine For Margo for funding their research. Fahad Khalid thanks Institut Gustave
321 Roussy for its financial support (grant CAJ 2020-088 IGR).

ACKNOWLEDGMENTS

322 The authors thank Irène Buvat, Raphaël Calmon, Christophe Nioche, and Fanny Orhac for their helpful
323 comments.

SUPPLEMENTAL DATA

324 Supplemental data include one excel file DIPGFeatures.xlsx and one pdf file, including supplemental
325 Figures and Tables.

DATA AVAILABILITY STATEMENT

326 Supplemental file DIPGFeatures.xlsx contains the clinical and biological information, and the radiomic
327 features used for the present study.

REFERENCES

- 328 1 .Cohen KJ, Jabado N, Grill J. Diffuse intrinsic pontine gliomas—current management and new biologic
329 insights. Is there a glimmer of hope? *Neuro Oncol* **19** (2017) 1025–1034. doi:10.1093/neuonc/nox021.
- 330 2 .Hoffman LM, Veldhuijzen van Zanten SE, Colditz N, Baugh J, Chaney B, Hoffmann M, et al. Clinical,
331 Radiologic, Pathologic, and Molecular Characteristics of Long-Term Survivors of Diffuse Intrinsic
332 Pontine Glioma (DIPG): A Collaborative Report From the International and European Society for
333 Pediatric Oncology DIPG Registries. *J Clin Oncol* **36** (2018) 1963–1972. doi:10.1200/JCO.2017.75.
334 9308.
- 335 3 .Vanan MI, Eisenstat DD. DIPG in Children – What Can We Learn from the Past? *Front Oncol* **5**
336 (2015). doi:10.3389/fonc.2015.00237.
- 337 4 .Wu G, Diaz AK, Paugh BS, Rankin SL, Ju B, Li Y, et al. The genomic landscape of diffuse intrinsic
338 pontine glioma and pediatric non-brainstem high-grade glioma. *Nat Genet* **46** (2014) 444–450. doi:10.
339 1038/ng.2938.
- 340 5 .Castel D, Philippe C, Calmon R, Le Dret L, Truffaux N, Boddaert N, et al. Histone H3F3A and
341 HIST1H3B K27M mutations define two subgroups of diffuse intrinsic pontine gliomas with different
342 prognosis and phenotypes. *Acta Neuropathol* **130** (2015) 815–827. doi:10.1007/s00401-015-1478-0.
- 343 6 .Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, et al. The 2021 who
344 classification of tumors of the central nervous system: a summary. *Neuro Oncol* **23** (2021) 1231–1251.
345 doi:10.1093/neuonc/noab106.
- 346 7 .Buczkwicz P, Hoeman C, Rakopoulos P, Pajovic S, Letourneau L, Dzamba M, et al. Genomic
347 analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent
348 activating ACVR1 mutations. *Nat Genet* **46** (2014) 451–456. doi:10.1038/ng.2936.
- 349 8 .Werbrouck C, Evangelista CC, Lobón-Iglesias MJ, Barret E, Le Teuff G, Merlevede J, et al. TP53
350 Pathway Alterations Drive Radioresistance in Diffuse Intrinsic Pontine Gliomas (DIPG). *Clin Cancer*
351 *Res* **25** (2019) 6788–6800. doi:10.1158/1078-0432.CCR-19-0126.
- 352 9 .Carvalho DM, Richardson PJ, Olaciregui N, Stankunaite R, Lavarino C, Molinari V, et al. Repurposing
353 Vandetanib plus Everolimus for the Treatment of ACVR1 -Mutant Diffuse Intrinsic Pontine Glioma.
354 *Cancer Discov* **12** (2022) 416–431. doi:10.1158/2159-8290.CD-20-1201.
- 355 10 .Avula S, Peet A, Morana G, Morgan P, Warmuth-Metz M, Jaspan T, et al. European Society
356 for Paediatric Oncology (SIOPE) MRI guidelines for imaging patients with central nervous system
357 tumours. *Childs Nerv Syst* **37** (2021) 2497–2508. doi:10.1007/s00381-021-05199-4.
- 358 11 .Cooney TM, Cohen KJ, Guimaraes CV, Dhall G, Leach J, Massimino M, et al. Response assessment
359 in diffuse intrinsic pontine glioma: recommendations from the response assessment in pediatric neuro-
360 oncology (rapno) working group. *Lancet Oncol* **21** (2020) E330–E336.
- 361 12 .Gillies R, Anderson A, Gatenby R, Morse D. The biology underlying molecular imaging in oncology:
362 From genome to anatome and back again. *Clin Radiol* **65** (2010) 517–521. doi:10.1016/j.crad.2010.
363 04.005.
- 364 13 .Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data.
365 *Radiology* **278** (2016) 563–577. doi:10.1148/radiol.2015151169.
- 366 14 .van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational
367 Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* **77** (2017) e104–e107. doi:10.
368 1158/0008-5472.CAN-17-0339.
- 369 15 .Nioche C, Orhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: A Freeware
370 for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the

- 371 Characterization of Tumor Heterogeneity. *Cancer Res* **78** (2018) 4786–4789. doi:10.1158/0008-5472.
372 CAN-18-0125.
- 373 **16**. Zwanenburg A, Vallieres M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image
374 Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput
375 Image-based Phenotyping. *Radiology* **295** (2020) 328–338. doi:10.1148/radiol.2020191145.
- 376 **17**. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding
377 tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun* **5**
378 (2014) 4006. doi:10.1038/ncomms5006.
- 379 **18**. Orhac F, Theze B, Soussan M, Boisgard R, Buvat I. Multiscale Texture Analysis: From F-18-FDG
380 PET Images to Histologic Images. *J Nucl Med* **57** (2016) 1823–1828. doi:10.2967/jnumed.116.
381 173708.
- 382 **19**. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A
383 Systematic Review. *Int J Radiat Oncol Biol Phys* **102** (2018) 1143–1158. doi:10.1016/j.ijrobp.2018.
384 05.053.
- 385 **20**. Keenan KE, Delfino JG, Jordanova KV, Poorman ME, Chirra P, Chaudhari AS, et al. Challenges in
386 ensuring the generalizability of image quantitation methods for MRI. *Med Phys* **49** (2022) 2820–2835.
387 doi:10.1002/mp.15195.
- 388 **21**. Goya-Outi J, Orhac F, Calmon R, Alentorn A, Nioche C, Philippe C, et al. Computation of reliable
389 textural indices from multimodal brain MRI: Suggestions based on a study of patients with diffuse
390 intrinsic pontine glioma. *Phys Med Biol* **63** (2018) 105003. doi:10.1088/1361-6560/aabd21.
- 391 **22**. Ford J, Dogan N, Young L, Yang F. Quantitative Radiomics: Impact of Pulse Sequence Parameter
392 Selection on MRI-Based Textural Features of the Brain. *Contrast Media & Molecular Imaging* **2018**
393 (2018) 1–9. doi:10.1155/2018/1729071.
- 394 **23**. Molina D, Pérez-Beteta J, Martínez-González A, Martino J, Velasquez C, Arana E, et al. Lack
395 of robustness of textural measures obtained from 3D brain tumor MRIs impose a need for
396 standardization. *PLoS One* **12** (2017) e0178843. doi:10.1371/journal.pone.0178843.
- 397 **24**. Carré A, Klausner G, Edjlali M, Lerousseau M, Briend-Diop J, Sun R, et al. Standardization of brain
398 MR images across machines and protocols: Bridging the gap for MRI-based radiomics. *Sci Rep* **10**
399 (2020) 12340. doi:10.1038/s41598-020-69298-z.
- 400 **25**. Shinohara RT, Sweeney EM, Goldsmith J, Shiee N, Mateen FJ, Calabresi PA, et al. Statistical
401 normalization techniques for magnetic resonance imaging. *NeuroImage: Clin* **6** (2014) 9–19.
402 doi:10.1016/j.nicl.2014.08.008.
- 403 **26**. Saint Martin MJ, Orhac F, Akl P, Khalid F, Nioche C, Buvat I, et al. A radiomics pipeline dedicated
404 to Breast MRI: Validation on a multi-scanner phantom study. *Magn Reson Mater Phys* **34** (2021)
405 355–366. doi:10.1007/s10334-020-00892-y.
- 406 **27**. Fortin JP, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, et al. Harmonization of multi-site
407 diffusion tensor imaging data. *NeuroImage* **161** (2017) 149–170. doi:10.1016/j.neuroimage.2017.08.
408 047.
- 409 **28**. Orhac F, Lecler A, Savatovski J, Goya-Outi J, Nioche C, Charbonneau F, et al. How can we combat
410 multicenter variability in MR radiomics? Validation of a correction procedure. *Eur Radiol* **31** (2021)
411 2272–2280. doi:10.1007/s00330-020-07284-9.
- 412 **29**. Goya-Outi J, Calmon R, Orhac F, Philippe C, Boddaert N, Puget S, et al. Can Structural MRI
413 Radiomics Predict DIPG Histone H3 Mutation and Patient Overall Survival at Diagnosis Time? *2019*
414 *IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)* (Chicago, IL, USA:
415 IEEE) (2019), 1–4. doi:10.1109/BHI.2019.8834524.

- 416 **30** .Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4itk: Improved n3 bias
417 correction. *IEEE Transactions on Medical Imaging* **29** (2010) 1310–1320. doi:10.1109/TMI.2010.
418 2046908.
- 419 **31** .Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. Fsl. *NeuroImage* **62** (2012)
420 782–790. doi:https://doi.org/10.1016/j.neuroimage.2011.09.015. 20 YEARS OF fMRI.
- 421 **32** .Khalid F, Goya-Outi J, Frouin V, Boddaert N, Grill J, Frouin F. Impact of ComBat and a Multi-
422 Model approach to deal with multi-scanner and missing MRI data in a small cohort study. Application
423 to H3K27M mutation prediction in patients with DIPG. *Annu Int Conf IEEE Eng Med & Biol Soc*
424 (Mexico: IEEE) (2021), 3809–3812. doi:10.1109/EMBC46164.2021.9629704.
- 425 **33** .Guyon I, Weston J, Barnhill S. Gene Selection for Cancer Classification using Support Vector
426 Machines. *Machine Learning* **46** (2002) 389–422. doi:10.1023/A:1012487302797.
- 427 **34** .Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine
428 learning in Python. *Journal of Machine Learning Research* **12** (2011) 2825–2830.
- 429 **35** .Lever J, Krzywinski M, Altman N. Classification evaluation. *Nature Methods* **13** (2016) 603–604.
430 doi:10.1038/nmeth.3945.
- 431 **36** .van Timmeren JE, Cester D, Tanadini-Lang S, Alkadhi H, Baessler B. Radiomics in medical
432 imaging—“how-to” guide and critical reflection. *Insights Imaging* **11** (2020) 91. doi:10.1186/
433 s13244-020-00887-2.
- 434 **37** .Lacroix M, Frouin F, Dirand AS, Nioche C, Orhac F, Bernaudin JF, et al. Correction for Magnetic
435 Field Inhomogeneities and Normalization of Voxel Values Are Needed to Better Reveal the Potential
436 of MR Radiomic Features in Lung Cancer. *Front Oncol* **10** (2020) 43. doi:10.3389/fonc.2020.00043.
- 437 **38** .Dirand AS, Frouin F, Buvat I. A downsampling strategy to assess the predictive value of radiomic
438 features. *Sci Rep* **9** (2019) 17869. doi:10.1038/s41598-019-54190-2.
- 439 **39** .Su X, Liu Y, Wang H, Chen N, Sun H, Yang X, et al. Multimodal MR imaging signatures to identify
440 brain diffuse midline gliomas with H3 K27M mutation. *Cancer Med* **11** (2022) 1048–1058. doi:10.
441 1002/cam4.4500.
- 442 **40** .Conte GM, Weston AD, Vogelsang DC, Philbrick KA, Cai JC, Barbera M, et al. Generative
443 Adversarial Networks to Synthesize Missing T1 and FLAIR MRI Sequences for Use in a
444 Multisequence Brain Tumor Segmentation Model. *Radiology* **299** (2021) 313–323. doi:10.1148/
445 radiol.2021203786.
- 446 **41** .Abdel Razek AAK, Alksas A, Shehata M, AbdelKhalek A, Abdel Baky K, El-Baz A, et al. Clinical
447 applications of artificial intelligence and radiomics in neuro-oncology imaging. *Insights Imaging* **12**
448 (2021) 152. doi:10.1186/s13244-021-01102-6.
- 449 **42** .Pan CC, Liu J, Tang J, Chen X, Chen F, Wu YL, et al. A machine learning-based prediction model
450 of H3K27M mutations in brainstem gliomas using conventional MRI and clinical features. *Radiother*
451 *Oncol* **130** (2019) 172–179. doi:10.1016/j.radonc.2018.07.011.
- 452 **43** .Chen H, Hu W, He H, Yang Y, Wen G, Lv X. Noninvasive assessment of H3 K27M mutational status
453 in diffuse midline gliomas by using apparent diffusion coefficient measurements. *Eur J Radiol* **114**
454 (2019) 152–159. doi:10.1016/j.ejrad.2019.03.006.
- 455 **44** .Raab P, Banan R, Akbarian A, Esmailzadeh M, Samii M, Samii A, et al. Differences in the MRI
456 Signature and ADC Values of Diffuse Midline Gliomas with H3 K27M Mutation Compared to Midline
457 Glioblastomas. *Cancers* **14** (2022) 1397. doi:10.3390/cancers14061397.
- 458 **45** .Escobar T, Vauclin S, Orhac F, Nioche C, Pineau P, Champion L, et al. Voxel-wise supervised
459 analysis of tumors with multimodal engineered features to highlight interpretable biological patterns.
460 *Med Phys* **49** (2022) 3816–3829. doi:10.1002/mp.15603.

- 461 **46** .Lazow MA, Nievelstein MT, Lane A, Bandopadhyay P, DeWire-Schottmiller M, Fouladi M, et al.
462 Volumetric endpoints in diffuse intrinsic pontine glioma: comparison to cross-sectional measures and
463 outcome correlations in the International DIPG/DMG Registry. *Neuro-Oncology* **24** (2022) 1598–
464 1608. doi:10.1093/neuonc/noac037.
- 465 **47** .Alksas A, Shehata M, Atef H, Sherif F, Alghamdi NS, Ghazal M, et al. A Novel System for Precise
466 Grading of Glioma. *Bioengineering* **9** (2022) 532. doi:10.3390/bioengineering9100532.
- 467 **48** .Chegraoui H, Philippe C, Dangouloff-Ros V, Grigis A, Calmon R, Boddaert N, et al. Object Detection
468 Improves Tumour Segmentation in MR Images of Rare Brain Tumours. *Cancers* **13** (2021) 6113.
469 doi:10.3390/cancers13236113.
- 470 **49** .Calmon R, Dangouloff-Ros V, Varlet P, Deroulers C, Philippe C, Debily MA, et al. Radiogenomics
471 of diffuse intrinsic pontine gliomas (DIPGs): Correlation of histological and biological characteristics
472 with multimodal MRI features. *Eur Radiol* **31** (2021) 8913–8924. doi:10.1007/s00330-021-07991-x.
- 473 **50** .Veldhuijzen van Zanten SEM, Lane A, Heymans MW, Baugh J, Chaney B, Hoffman LM, et al.
474 External validation of the diffuse intrinsic pontine glioma survival prediction model: A collaborative
475 report from the International DIPG Registry and the SIOPE DIPG Registry. *J Neurooncol* **134** (2017)
476 231–240. doi:10.1007/s11060-017-2514-9.
- 477 **51** .Tam LT, Yeom KW, Wright JN, Jaju A, Radmanesh A, Han M, et al. MRI-based radiomics for
478 prognosis of pediatric diffuse intrinsic pontine glioma: An international study. *Neuro-Oncology*
479 *Advances* **3** (2021) vdab042. doi:10.1093/noajnl/vdab042.
- 480 **52** .Leach JL, Roebker J, Schafer A, Baugh J, Chaney B, Fuller C, et al. MR imaging features of diffuse
481 intrinsic pontine glioma and relationship to overall survival: Report from the International DIPG
482 Registry. *Neuro-Oncology* (2020) noaa140. doi:10.1093/neuonc/noaa140.
- 483 **53** .Vuong HG, Ngo TNM, Le HT, Dunn IF. The prognostic significance of HIST1H3B/C and H3F3A
484 K27M mutations in diffuse midline gliomas is influenced by patient age. *J Neurooncol* **158** (2022)
485 405–412. doi:10.1007/s11060-022-04027-2.

TABLES

MR Scanner Type	T1w	T1c	T2w	FLAIR
1.5 T	60	58	60	56
3 T	13	14	14	11

Table 1. Number of modalities available according to each type of MR scanner

	Number of patients (F/M)	Mean Age (y)
All patients	80 (35/45)	8.1 ± 4.4
Histone H3 mutation status		
Known	63 (30/33)	7.7 ± 3.7
H3.1	14 (8/6)	5.0 ± 1.6
H3.2	1 (0/1)	4.5
H3.3	44 (20/24)	8.7 ± 3.7
WT	4 (2/2)	6.9 ± 4.6
Others	49 (22/27)	8.5 ± 3.8
Unknown	17 (5/12)	9.3±6.4
ACVR1 mutation status		
Known	63 (28/35)	7.9 ± 3.6
ACVR1 mutation	14 (7/7)	5.9 ± 3.0
WT	49 (21/28)	8.4 ± 3.6
Unknown	17 (7/10)	8.8±6.6
TP53 mutation status		
Known	61 (29/32)	7.8 ± 3.7
TP53 mutation	34 (14/20)	9.0 ± 3.4
WT	27 (15/12)	6.3 ± 3.5
Unknown	19 (6/13)	8.8±6.2

Table 2. Main clinical (age and sex) and molecular features. For Histone H3, "Others" gather H3.2 mutation, H3.3 mutation and Wild-Type (WT).

H3.1 mutation	Features name	Features identifier
M_{Clinic}	Age	h1 (a1, t1)
M_{T1w}	glcm_ClusterShade glrlm_GrayLevelNonUniformity firstorder_90Percentile firstorder_Skewness	h2 h3 (a5) h4 (a2, t4) h5 (a4)
M_{T1c}	glcm_ClusterShade glrlm_ShortRunLowGrayLevelEmphasis glszm_IntensityVariability firstorder_10Percentile	h6 (a6) h7 h8 h9
M_{T2w}	glszm_LowIntensitySmallAreaEmphasis glszm_HighIntensityLargeAreaEmphasis firstorder_TotalEnergy firstorder_Minimum	h10 h11 (a10) h12 (a12, t9) h13
M_{FLAIR}	glrlm_RunLengthNonUniformity firstorder_Skewness glcm_ClusterShade glcm_DifferenceVariance	h14 (a15) h15 (a13) h16 h17

Table 3. Subsets of features selected by the five different models M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2w} , and M_{FLAIR} to predict H3.1 mutation. Inside brackets, features selected by one or two other tasks of mutation prediction.

ACVR1 mutation	Features name	Features identifier
M_{Clinic}	Age	a1 (h1, t1)
M_{T1w}	firstorder_90Percentile gldm_Correlation firstorder_Skewness glrlm_GrayLevelNonUniformity	a2 (h4, t4) a3 a4 (h5) a5 (h3)
M_{T1c}	gldm_ClusterShade glszm_IntensityVariabilityNormalized gldm_ClusterProminence glrlm_LongRunHighGrayLevelEmphasis	a6 (h6) a7 a8 a9
M_{T2w}	glszm_HighIntensityLargeAreaEmphasis firstorder_Skewness firstorder_TotalEnergy	a10 (h11) a11 (t8) a12 (h12, t9)
M_{FLAIR}	firstorder_Skewness gldm_Idmn glrlm_RunLengthNonUniformity glszm_LowIntensityLargeAreaEmphasis	a13 (h15) a14 a15 (h14) a16

Table 4. Subsets of features selected by the five different models M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2w} , and M_{FLAIR} to predict ACVR1 mutation. Inside brackets, features selected by one or two other tasks of mutation prediction.

TP53 mutation	Features name	Features identifier
M_{Clinic}	Age	t1 (h1, a1)
M_{T1w}	glrlm_RunLengthNonUniformity	t2
	glcm_SumAverage	t3
	firstorder_90Percentile	t4 (h4, a2)
	glszm_ZoneEntropy	t5
M_{T1c}	glszm_HighIntensityLargeAreaEmphasis	t6
M_{T2w}	glcm_SumAverage	t7
	firstorder_Skewness	t8 (a11)
	firstorder_TotalEnergy	t9 (h12, a12)
	glcm_AverageIntensity	t10
M_{FLAIR}	glszm_IntensityVariability	t11

Table 5. Subsets of features selected by the five different models M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2w} , and M_{FLAIR} to predict TP53 mutation. Inside brackets, features selected by one or two other tasks of mutation prediction.

Models	M_{Clinic}	M_{T1w}	M_{T1c}	M_{T2w}	M_{FLAIR}	M_{Multi}
H3.1 mutation						
Number of patients	63 (14)	58 (13)	58 (14)	59 (13)	56 (12)	63 (14)
AUC	0.82	0.87	0.92	0.71	0.74	<u>0.91</u>
Sensitivity (%)	85.7	84.6	78.6	<u>92.3</u>	83.3	100
Specificity (%)	57.1	75.6	77.3	52.2	72.7	<u>75.5</u>
Balanced Accuracy (%)	71.4	<u>80.1</u>	77.9	72.2	78.0	87.8
ACVR1 mutation						
Number of patients	63 (14)	58 (13)	58 (13)	59 (13)	56 (13)	63 (14)
AUC	0.74	0.78	0.77	0.72	<u>0.83</u>	0.91
Sensitivity (%)	<u>85.7</u>	76.9	76.9	84.6	76.9	92.9
Specificity (%)	44.9	60.0	71.1	58.7	74.4	<u>71.4</u>
Balanced Accuracy (%)	65.3	68.5	74.0	71.7	<u>75.7</u>	82.1
TP53 mutation						
Number of patients	61 (34)	56 (33)	57 (32)	57 (33)	54 (32)	61 (34)
AUC	0.78	0.83	0.63	<u>0.86</u>	0.75	0.88
Sensitivity (%)	55.9	66.7	28.1	<u>69.7</u>	71.9	67.6
Specificity (%)	85.2	78.3	88.0	87.5	54.5	88.9
Balanced Accuracy (%)	70.5	72.5	58.1	78.6	63.2	<u>78.3</u>

Table 6. Prediction results for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2w} , M_{FLAIR} and M_{Multi} in a LOO-CV framework. For each prediction task and for each model, five figures of merit are reported: the total number of patients for which the prediction was possible (the number of patients with mutation is between brackets), the AUC under the ROC curve, the sensitivity, the specificity and the balanced accuracy. For each figure of merit, best results are in bold characters and second best results are underlined.

FIGURES

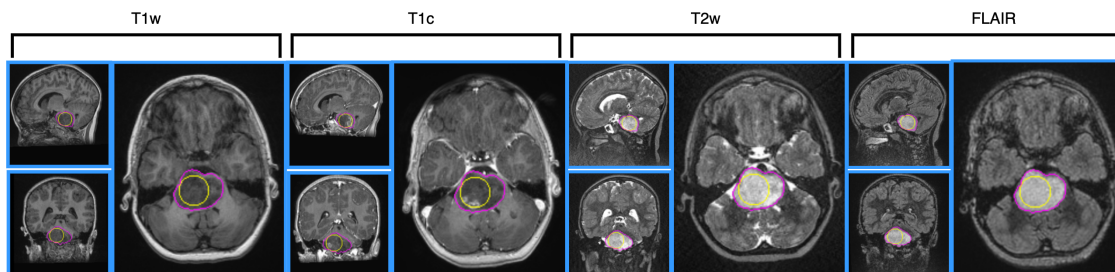


Figure 1. Illustration of MRI data for a 4 year-old patient, having H3.1 and ACVR1 mutations, having no TP53 mutation. MRI data are shown after intensity normalization using the hybrid white stripe method. From the left side to the right side: T1w, T1c, T2w and FLAIR images, using $\begin{matrix} \text{Sagittal} \\ \text{Coronal} \end{matrix}$ Axial layout for each modality. The contours of the sphere used for computing intensity and texture radiomic features inside the tumor are outlined in yellow color on each view. The contours of the tumor used for computing the shape features are outlined in purple color.

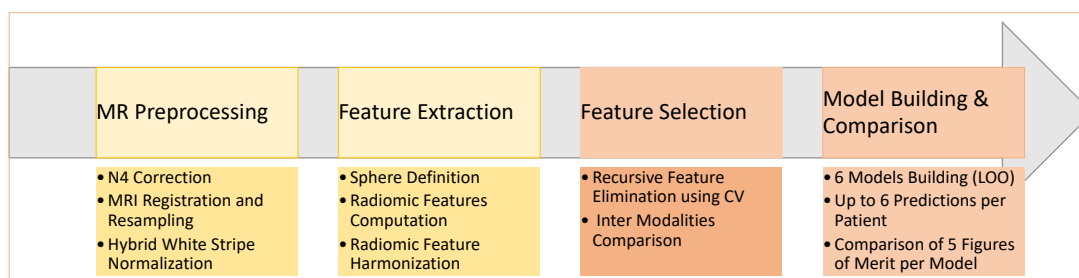


Figure 2. Main steps of the construction of the six machine learning models to predict a molecular mutation.

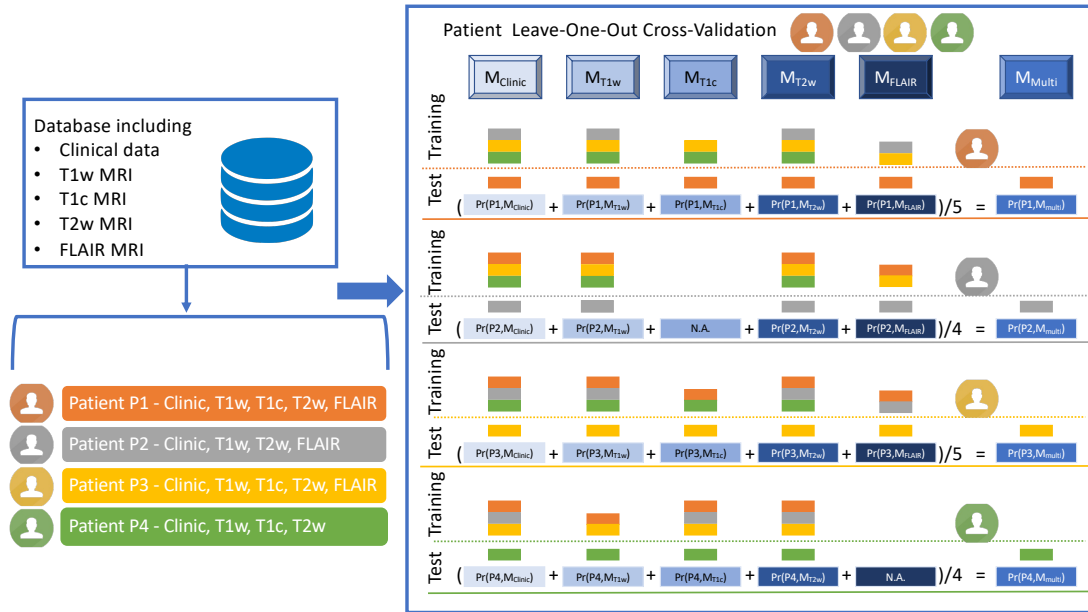


Figure 3. Illustration of the LOO-CV-MIM framework, i.e. the Leave-One-Out Cross-Validation framework dealing with Missing Imaging Modalities. The LOO-CV-MIM framework is applied to a binary classification task (the prediction of a mutation in our current study). The database given as a fictitious example includes four patients (P1, P2, P3, and P4, displayed in orange, grey, yellow, and green colors), P1 and P3 having the five modalities (clinical data, T1w, T1c, T2w and FLAIR MRI), P2 having one missing modality (T1c), and P4 having also one missing modality (FLAIR). For P1 (resp. P3), five models M_j (with $j \in \{Clinic, T1w, T1c, T2w, FLAIR\}$), are defined using as training database all the patients except P1 (resp. P3) for which the modality is present (the training database includes three patients for M_{Clinic} , M_{T1w} , and M_{T2w} , and two patients for M_{T1c} and M_{FLAIR}). These five models are then tested onto the remaining patient P1 (resp. P3), providing five probabilities of mutation $\Pr(P1, M_j)$, with $0 \leq \Pr(P1, M_j) \leq 1$ (resp. $\Pr(P3, M_j)$) and thus five predictions of mutation. A sixth prediction of mutation corresponding to M_{Multi} , is defined as the mean value of the five probabilities $\Pr(P1, M_j)$ (resp. $\Pr(P3, M_j)$). For patients P2 (resp. P4) having one missing modality, a similar process is applied but only four (and not five) models M_j are defined (there is no model M_{T1c} for P2, no model M_{FLAIR} for P4), providing four probabilities $\Pr(P2, M_j)$ (resp. $\Pr(P4, M_j)$). A fifth prediction of mutation corresponding to M_{Multi} , is then defined as the mean value of the four probabilities $\Pr(P2, M_j)$ (resp. $\Pr(P4, M_j)$).

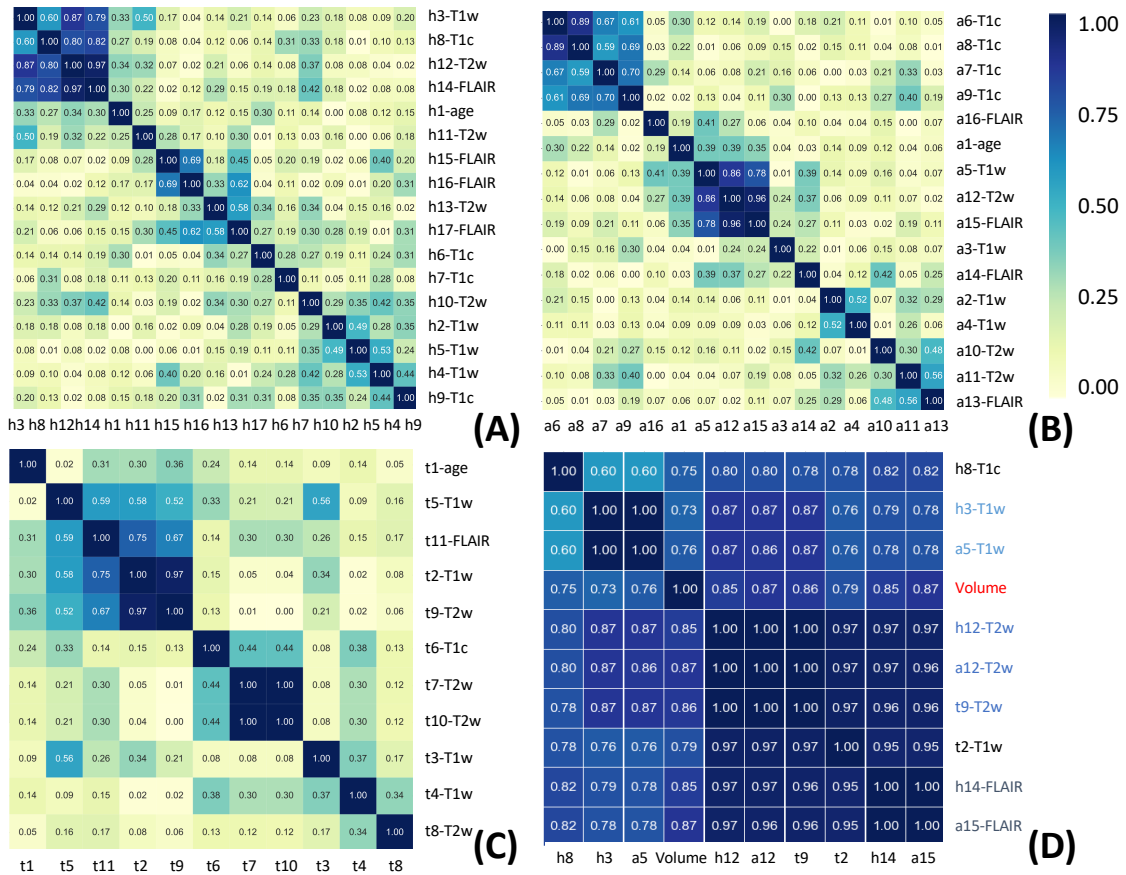


Figure 4. Correlation heatmaps between the features that have been selected by the five different models to predict H3.1 mutation (A), ACVR1 mutation (B), and TP53 mutation (C). Tables 3, 4, and 5 provide correspondence between feature identifiers and the full feature name according to PyRadiomics nomenclature. In (D), correlation matrix heatmap between the previously selected features which are highly correlated with tumor volume. Feature identifiers (on the right side) of identical features found by the different predictive tasks are shown in color.

Supplementary Material

SUPPLEMENTARY DATA

Supplemental file DIPGFeatures.xlsx contains the clinical and biological information, and the radiomic features used for the present study. The first sheet, entitled Classif, provides for each of the 80 subjects, subject code, MR field of the scanner, sex (0 for female, 1 for male), age (in years), volume of tumor (in mm³), H3 mutation (1 for H3.1, 2 for H3.2, 3 for H3.3, 0 for H3 Wild-Type, na when unknown), ACVR1 mutation (1 for ACVR1 mutation, 0 for Wild-Type, na when unknown), TP53 mutation (1 for TP53 mutation, 0 for Wild-Type, na when unknown). Then T1w, T1c, T2w, and FLAIR entitled sheets provide the 79 radiomic features after extraction from T1w, T1c, T2w and FLAIR images (without ComBat transformation), for each subject having the corresponding modality.

SUPPLEMENTARY TABLES AND FIGURES

Classification task	Features name	Features identifier
H3.1 mutation	Age	h1 (a1, t1)
	shape_Elongation	h18 (a17)
	shape_Flatness	h19 (t12)
	shape_SurfaceVolumeRatio	h20 (a19, t13)
ACVR1 mutation	Age	a1 (h1, t1)
	shape_Elongation	a17 (h18)
	shape_Maximum2DDiameterRow	a18
	shape_SurfaceVolumeRatio	a19 (h20)
TP53 mutation	Age	t1 (h1, a1)
	shape_Flatness	t12 (h19)
	shape_SurfaceVolumeRatio	t13 (h20)

Table S1. Subsets of features selected by the model $M^{ClinicSh}$ merging clinical features and shape features to predict H3.1, ACVR1, and TP53 mutations. Inside brackets, features selected by one or two other tasks of mutation prediction.

Models	M_{Clinic}	$M_{ClinicSh}$	M_{Multi}	$M_{MultiSh}$
H3.1 mutation				
Number of patients	63 (14)	63 (14)	63 (14)	63 (14)
AUC	0.82	0.87	0.91	0.95
Sensitivity (%)	85.7	92.9	100	100
Specificity (%)	57.1	61.2	75.5	73.5
Balanced Accuracy (%)	71.4	77.0	87.8	86.7
ACVR1 mutation				
Number of patients	63 (14)	63 (14)	63 (14)	63 (14)
AUC	0.74	0.80	0.91	0.93
Sensitivity (%)	85.7	92.9	92.9	92.9
Specificity (%)	44.9	55.1	71.4	67.3
Balanced Accuracy (%)	65.3	74.0	82.1	80.1
TP53 mutation				
Number of patients	61 (34)	61 (34)	61 (34)	61 (34)
AUC	0.78	0.83	0.88	0.91
Sensitivity (%)	55.9	64.7	67.6	70.6
Specificity (%)	85.2	88.9	88.9	88.9
Balanced Accuracy (%)	70.5	76.8	78.3	79.7

Table S2. Prediction results for four models: M_{Clinic} , as defined in the main text, which is based on age only, $M_{ClinicSh}$, combining clinical features and 14 shape features, M_{Multi} as defined in the main text, including clinic and intensity and texture radiomic features, $M_{MultiSh}$ including clinic, shape, intensity and texture radiomic features, in the LOO-CV-MIM framework. For each prediction task and for each model, the five figures of merit defined in the main text are reported: the number of patients (all of them for the four models), the AUC of the ROC curve, the sensitivity, the specificity and the balanced accuracy. The best performances (in boldface) are provided by the multi-model approaches. There is no clear improvement brought by the introduction of the shape features in the prediction of the H3.1 and ACVR1 mutations. However, there is a slight trend of improvement for the prediction of the TP53 mutation.

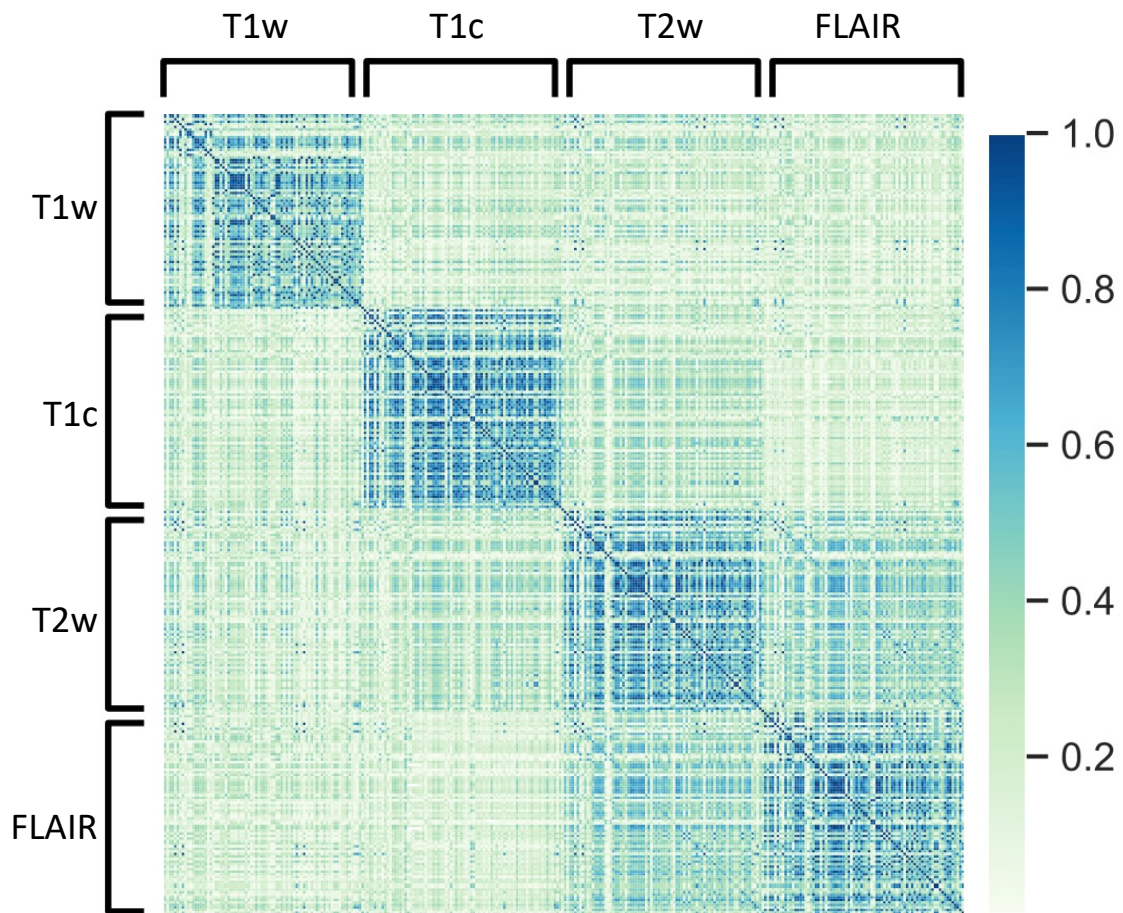


Figure S1. Correlation matrix heatmap between the 4x79 features computed for each modality: T1w, T1c, T2w, and FLAIR. The general trend is that the correlations between two features from two different modalities is less than the correlation between two features from the same modalities and when comparing modalities two by two, the features computed for T2w and FLAIR modalities seem to be the most correlated.

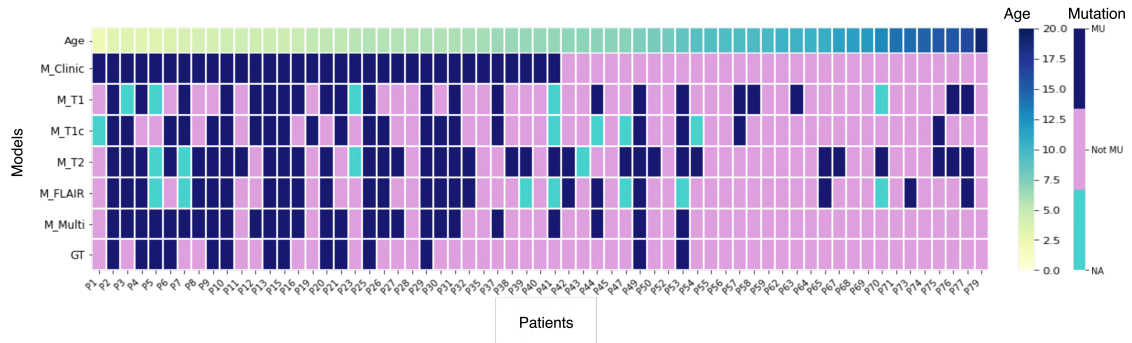


Figure S2. Predictive results of H3.1 mutation by the six models for each patient of the database having a ground truth (n=63). Patients (one per column, being ordered by age) along with the age distributions (yellow/green/blue color). From top row to bottom row: age, binary output of the six predictive models M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2w} , M_{FLAIR} , M_{Multi} and ground truth (GT). MU (dark blue color) stands for mutated cases, Not MU (pink color) stands for not H3.1 mutated cases, NA (cyan color) stands for not available modality.

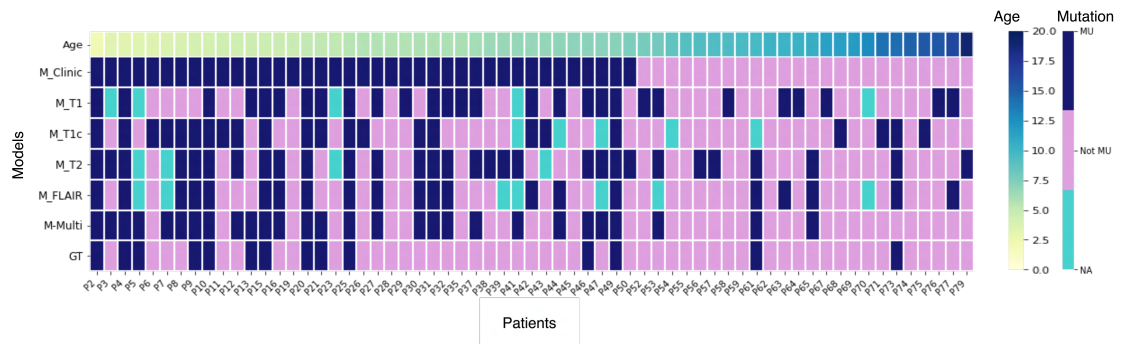


Figure S3. Predictive results of ACVR1 mutation by the six models for each patient of the database, having a ground truth (n=63). Patients (one per column, being ordered by age) along with the age distributions (yellow/green/blue color). From top row to bottom row: age, binary output of the six predictive models M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2w} , M_{FLAIR} , M_{Multi} , and ground truth (GT). MU (dark blue color) stands for mutated cases, Not MU (pink color) stands for not mutated cases, NA (cyan color) stands for not available modality.

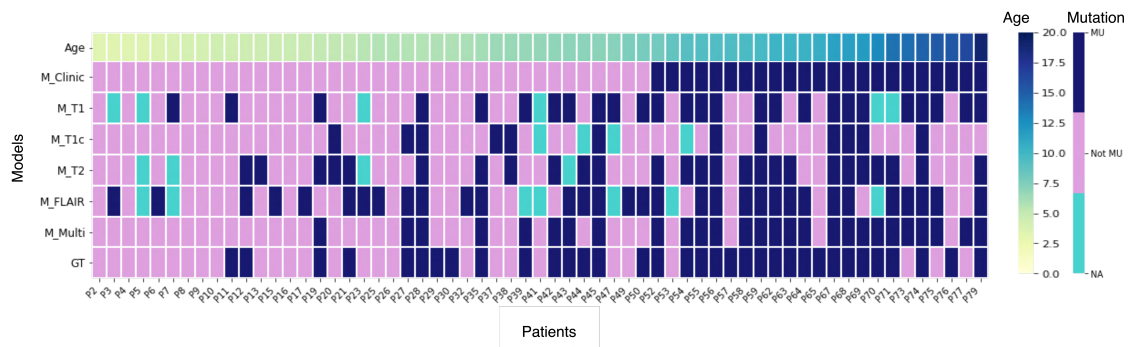


Figure S4. Predictive results of TP53 mutation by the six models for each patient of the database, having a ground truth ($n=61$). Patients (one per column, being ordered by age) along with the age distributions (yellow/green/blue color). From top row to bottom row: age, binary output of the six predictive models M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2w} , M_{FLAIR} , M_{Multi} and ground truth (GT). MU (dark blue color) stands for mutated cases, Not MU (pink color) stands for not mutated cases, NA (cyan color) stands for not available modality.

4.3 Additional validation studies

The following section will present additional experiments that were conducted to estimate the performance of different classifiers for the prediction task with their integration in the multi-modal approach. The stability of results were tested by comparing 5 and 10 folds cross validation with LOOCV. Finally the clinical feature age holding significance in literature with respect to mutation association is utilized for profiling patients and an age weighted model is introduced and evaluated.

4.3.1 Classifier comparison

We have exploited benchmarked machine learning techniques for building predictive models. The four following binary classifiers have been tested:

1. Logistic regression: this linear model forecasts the likelihood that an input will belong to a specific class using a sigmoid function.
2. Decision tree: This model, makes predictions using a series of if-else expressions, based on ones pecific feature for each branch.
3. Random forest: This ensemble model makes predictions by combining various decision trees.
4. AdaBoost : Gives additional weight to incorrectly categorized instances throughout each iteration, it combines several weak classifiers into a powerful classifier.

The following experiment was conducted for each specific modality, comparing performance of classifiers by building classification models. The radiomics features were in addition evaluated by iteratively adding features to the classifier. The four classifiers were tested in a leave one out - cross validation framework (LOO-CV) using their default parameters, no hyper parameter optimization was done. The 79 radiomic features extracted from FLAIR, T1, T1c and T2 were iteratively added from 1st till 79th feature to the four classifier separately in the same order. The MR modality performing the best is also identified. Figures [4.1](#) and [4.2](#) present the accuracy obtained by LOOCV classification for ACVR1 mutation prediction using the four classifiers by considering the full set of features added iteratively.

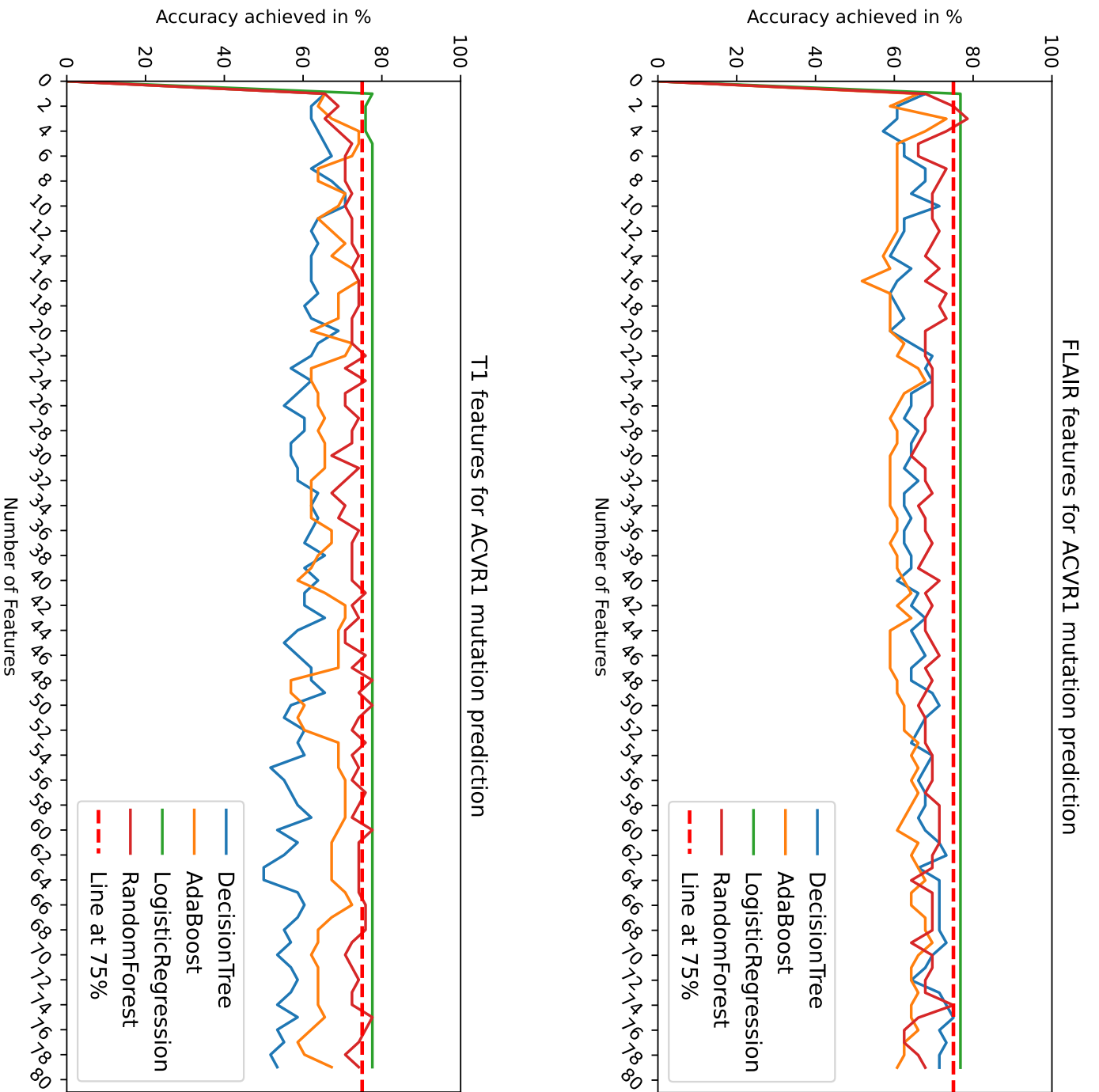


Figure 4.1: Top (FLAIR features) and Bottom (T1 features) figures illustrates the performance of four classifiers for predicting the presence of ACVR1 mutation. The y-axis displays the accuracy achieved in (%) the x-axis present the 79 radiomic features extracted from the modality. The blue colored line represents the accuracies achieved by Decision tree classifier and how the addition of features impacts the accuracy achieved. Similarly the orange colored line for Ada-boost, green for Logistic Regression and red for random forest. For visual assistance a dashed line is drawn at 75%.

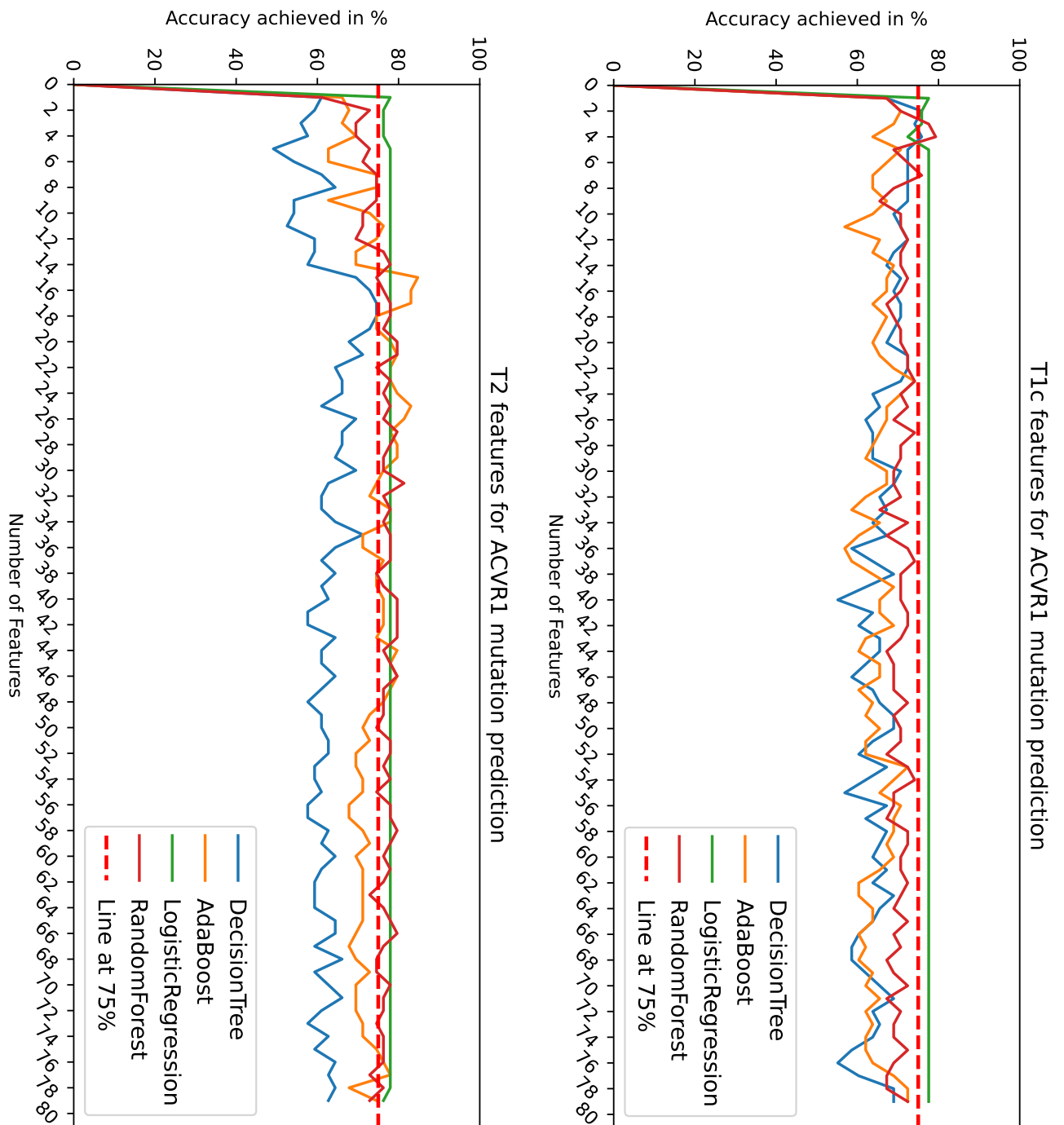


Figure 4.2: Top (T1c features) and Bottom (T2 features) figures illustrate the performance of four classifiers for predicting the presence of ACVR1 mutation. The y-axis displays the accuracy achieved in (%) the x-axis present the 79 radiomic features extracted from the modality. The blue colored line represents the accuracies achieved by Decision tree classifier and how the addition of features impacts the accuracy achieved. Similarly the orange colored line for Ada-boost, green for Logistic Regression and red for random forest. For visual assistance a dashed line is drawn at 75%.

The four plots displayed in figures 4.1 and 4.2 indicate the changes on the balanced accuracy recorded as the features are added. All classifiers except logistic regression (LR) display low and high peaks in accuracy. In the case of LR, the default parameter incorporates a LASSO effect hence after addition of the 5th feature the model drops all added features and maintains the same accuracy. Considering performances of the four classifiers using FLAIR features it seems that only logistic regression achieves accuracies above 75% while the rest of the classifiers fluctuate mostly between 60-70% barely crossing 70% accuracy. Similarly, for classifiers based on T1w features LR achieves the higher accuracy. The iterative addition of T1w features seems to worsen the performance of Ada Boost and Decision Tree classifier. When classifiers are evaluated using T2w features, a global improvement is seen for RF and AdaBoost. Based on the results achieved it was decided to continue with logistic regression for feature selection and mutation prediction, the reason being its simplicity in prediction. These results display the importance of a feature selection step and the individual importance of each MR modality. For instance features extracted from FLAIR and T1c seem to have better predictive power than T1w and T2w as the accuracies curves are better overlapped. In the next section we will see further evaluation of a few classifiers embedded in the multi-modal and missing imaging modality (MM-LOOCV-MIM) framework.

4.3.2 SVM, RF and LR evaluated in the Multi-Model approach

Logistic regression was adapted as the final classifier for the study given the low number of cases and the unbalanced datasets in two out of three mutations being studied (H3.1 and ACVR1). In the previous section evaluating a few classifiers, logistic regression displayed simple and robust classification performance as it does not require tuning a large number of parameters, making it a convenient choice for the multi-modal framework explained in the paper (section Methods and Materials). In addition to LR two additional classifiers SVM and RF were tested. Differing from the previous experiments in last section of this chapter here we will evaluate them in the MM-LOOCV-MIM framework. Only results for TP53 mutation prediction are presented, as the class labels are fairly balanced for this mutation compared to H3.1 and ACVR1. Thus, SVM and Random Forest classifiers were used with their default parameters, (documented with the scikit-learn library, version 1.1.1), with the exception of adding `classweight='balanced'`. Experimental results validated the choice of logistic regression in the final framework. All three classifiers used the same set of features selected by the RFE-CV (Recursive Feature Elimination with Cross-Validation) procedure. To make comparisons easier, the Table using Logistic Regression (LR) given in the Results section of the paper is reported here (Table 4.1). Results displaying performance of SVM and RF classifiers are reported in Tables 4.2 and 4.3 respectively.

TP53(LR)	Sensitivity (%)	Specificity (%)	Balanced accuracy (%)
M_{Clinic}	56	85	70
M_{T1w}	67	78	72
M_{T1c}	28	88	68
M_{T2}	70	87	79
M_{FLAIR}	72	54	63
M_{Multi}	68	89	78

Table 4.1: Prediction results for TP53 for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework using Logistic Regression classifier. For each prediction task and for each model three figures of merit are reported: sensitivity, specificity and balanced accuracy.

TP53(SVM)	Sensitivity (%)	Specificity (%)	Balanced accuracy (%)
M_{Clinic}	73	67	70
M_{T1w}	82	67	76
M_{T1c}	69	44	57
M_{T2}	82	71	76
M_{FLAIR}	84	41	63
M_{Multi}	94	59	77

Table 4.2: Prediction results for TP53 for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework using Support Vector Machine classifier. For each prediction task and for each model three figures of merit are reported: sensitivity, specificity and balanced accuracy.

TP53(RF)	Sensitivity (%)	Specificity (%)	Balanced accuracy (%)
M_{Clinic}	88	85	87
M_{T1w}	94	87	90
M_{T1c}	88	76	82
M_{T2}	91	79	85
M_{FLAIR}	90	71	81
M_{Multi}	100	96	98

Table 4.3: Prediction results for TP53 for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework using Random Forest classifier. For each prediction task and for each model three figures of merit are reported: sensitivity, specificity and balanced accuracy.

In this case of 'approximately balanced labels' task the three tables display similar trend i.e the performances (especially the balanced accuracy) improves when applying the multi-model. Balanced accuracies for logistic regression and SVM are very similar in the multi-model approach, while RF displays very high-performance metrics. LR and SVM are both linear classifiers, RF on the other hand is a non-linear classifier and can capture non-linear relationships between the input features and the output. Furthermore, RF is less sensitive to outliers compared to logistic regression and SVM. Outliers can indeed have a significant impact on the performance of logistic regression and SVM since they try to optimize the decision boundary based on all data points. It was also interesting to test how the three classifiers may or may not differ for a non balanced prediction task, Results for H3.1 prediction in the multi-modal LOO-CV framework using LR, SVM and RF are presented in tables [4.4](#), [4.5](#), and [4.6](#), respectively.

H3.1 (LR)	Sensitivity (%)	Specificity (%)	Balanced accuracy (%)
M_{Clinic}	86	57	71
M_{T1w}	85	76	80
M_{T1c}	79	77	78
M_{T2}	92	52	72
M_{FLAIR}	83	73	78
M_{Multi}	100	76	88

Table 4.4: Prediction results for H3.1 for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework using Logistic Regression classifier. For each prediction task for each model three figures of merit are reported: sensitivity, specificity and balanced accuracy.

H3.1 (SVM)	Sensitivity (%)	Specificity (%)	Balanced accuracy (%)
M_{Clinic}	14	98	56
M_{T1w}	46	98	72
M_{T1c}	29	98	63
M_{T2}	0	100	50
M_{FLAIR}	8	98	53
M_{Multi}	0	100	50

Table 4.5: Prediction results for H3.1 for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework using Support Vector Machine classifier. For each prediction task for each model three figures of merit are reported: sensitivity, specificity and balanced accuracy.

H3.1 (RF)	Sensitivity (%)	Specificity (%)	Balanced accuracy (%)
M_{Clinic}	72	91	82
M_{T1w}	85	89	87
M_{T1c}	71	93	82
M_{T2}	69	98	84
M_{FLAIR}	67	93	80
M_{Multi}	93	100	96

Table 4.6: Prediction results for H3.1 for the six models: M_{Clinic} , M_{T1w} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework using Random Forest classifier. For each prediction task for each model three figures of merit are reported: sensitivity, specificity and balanced accuracy.

In this case of “unbalanced-label” task, two ML classifiers (LR and RF) show similar trends to those already described in the previous experiment i.e., the performances (especially the balanced accuracy) are improved when applying the multi-modal model. However, for the linear SVM, the performance metrics when applying the multi-modal model are not improved. Indeed, too many models have low performances (for instance, sensitivities are always less than 50%) and the mean value of predictions cannot correct for this default.

The different characteristics of classifiers have an impact on how well a particular prediction task is achieved in a LOOCV framework. From the comparison between the three classifiers, the model of logistic regression based on its simplicity and its robustness proves to be informative in our context of small number of cases and of balanced or imbalanced classes. Regarding prediction performance, the results can be overestimated, especially with the LOO-CV process. With a larger database, the performances will be better assessed, and different machine learning models could also be tested, tuned, and compared.

4.3.3 Comparing K-Fold cross validation with Leave one out cross validation

Cross validation is an accepted method in machine learning, used in order to choose and assess models. Cross validation is crucial when working with small sample numbers to ensure model generalization. Given the specific characteristics of our data set (small sample size, unbalanced data for two of the three mutations under study, and some missing imaging modalities), it was challenging to apply the 10-fold or 5-fold cross validation pattern universally on the multi-modal method. This constraint would prevent a direct comparison of the various methods because it is difficult to have comparable folds for the various modalities. But, we put this 10-fold and 5-fold strategy to the model using the T1w modality as an example. The LOO-CV, 10-fold, and 5-fold CV for three classifiers (LR, SVM, and RF) for the prediction of TP53 mutation (roughly balanced-class task) were compared in the Tables 4.7, 4.8 and 4.9. Performances are presented as the mean values and after stratifying and repeating the 10-fold and 5-fold cross-validation procedures 100 times (standard deviation).

LR (TP53) T1w	Sensitivity (%)	Specificity (%)	Balanced accuracy (%)
LOO-CV	67	78	72
5 fold-CV	56(3)	75(3)	65(3)
10 fold-CV	55(3)	74(2)	65(2)

Table 4.7: Prediction results for TP53 using features from T1w modality only using LOO-CV and 5 and 10 fold CV using Logistic regression. For each prediction frame work three figures of merit are reported: sensitivity, specificity balanced accuracy.

SVM (TP53) T1w	Sensitivity (%)	Specificity (%)	Balanced accuracy (%)
LOO-CV	82	67	76
5 fold-CV	64(3)	79(2)	72(2)
10 fold-CV	64(3)	80(2)	73(2)

Table 4.8: Prediction results for TP53 using features from T1w modality only using LOO-CV and 5 and 10 fold CV using SVM. For each prediction frame work three figures of merit are reported: sensitivity, specificity balanced accuracy.

RF (TP53) T1w	Sensitivity (%)	Specificity (%)	Balanced accuracy (%)
LOO-CV	94	87	90
5 fold-CV	48(8)	64(5)	56(7)
10 fold-CV	47(7)	64(6)	55(6)

Table 4.9: Prediction results for TP53 using features from T1w modality only using LOO-CV and 5 and 10 fold CV using Random forest. For each prediction frame work three figures of merit are reported: sensitivity, specificity balanced accuracy.

As anticipated, using a 5-fold CV or 10-fold CV causes a decrease of all predicted metrics as compared to LOO. If the interpretations of these reductions are the same for the LR and SVM models, they reveal a significant decrease for Random Forest models, suggesting a potential overfitting of this model that was not apparent when

merely taking the LOO-CV into account. Studies using 5 and 10 fold cross validation demonstrate the robustness of the logistic regression model. To recap, employing LR instead of other classifiers like SVM and RF allows for the estimated values of performance metrics to remain informative (without fine tuning of their parameters). Because of its simplicity and robustness, the logistic regression model was chosen, and this decision proves to be instructive in our setting of few examples and unbalanced classes. Our results are definitely overstated in terms of prediction performance, especially when using the LOO-CV approach. A larger database will allow for improved performance evaluation and the testing, tuning, and comparison of various machine learning models.

4.4 Will an Age dependent weight improve results of the multi-model approach?

The four radiomic models and the clinic model were given equal weights for the final multi-model computation. From literature review we know the mutation H3.1 is more frequent in younger children (age <10 years) and the mutation H3.3 is more frequent in older people. The ACVR1 mutation is mostly associated with H3.1 mutations. Since we predict for instance H3.1 against H3.3, H3.2 and wildtype, it was interesting to test a clinic (Age) weighted model. This was done by considering the predictions for patients less than 10 years of age. Considering there was no mutation H3.1 for patients older than 10 years, this model was called the Multi-model with Age weights (MMA). This new model is used to evaluate prediction results for the three mutations. Results of the multi-model and multi-model with age weights are displayed in table 4.10. The multi-model with added age weights for H3.1 prediction does not display any change as compared to the non weighted multi-model. Balanced accuracy achieved by both approaches is 87.7%. This is due to the fact all the patient with H3.1 mutation are below 10 years in age. The model clinic and the multi-model predicts these patients as non mutated hence finally the multi-model with age has no effect for H3.1 prediction. Whereas for ACVR1 the balanced accuracy drops by 4%, this is because a patient having age greater than 10 years is miss classified by the age weighted model, initially this patient was accurately classified by the multi-model approach. Multi-modal with age displays increase in balanced accuracy by 3% for TP53 prediction. Two patients which were miss classified by the multi model where corrected by the multi-model with age weights. Due to the low number of positive class for H3.1 and ACVR1 results for multi model with age weights did not seem to work so well. A better evaluation of this idea will be important to test on a larger cohort.

Models	M_{Clinic}	M_{T1}	M_{T1c}	M_{T2}	M_{FLAIR}	M_{Multi}	MM_{Age}
H3.1 mutation							
Number of patients	63 (14)	58 (13)	58 (14)	59 (13)	56 (12)	63 (14)	63 (14)
Sensitivity (%)	85.7	84.6	78.6	<u>92.3</u>	83.3	100	100
Specificity (%)	57.1	75.6	77.3	52.2	72.7	<u>75.5</u>	<u>75.5</u>
Balanced Accuracy (%)	71.4	<u>80.1</u>	77.9	72.2	78.0	87.8	87.8
ACVR1 mutation							
Number of patients	63 (14)	58 (13)	58 (13)	59 (13)	56 (13)	63 (14)	63 (14)
Sensitivity (%)	<u>85.7</u>	76.9	76.9	84.6	76.9	92.9	<u>85.7</u>
Specificity (%)	44.9	60.0	71.1	58.7	74.4	<u>71.4</u>	<u>71.4</u>
Balanced Accuracy (%)	65.3	68.5	74.0	71.7	75.7	82.1	<u>78.5</u>
TP53 mutation							
Number of patients	61 (34)	56 (33)	57 (32)	57 (33)	54 (32)	61 (34)	61 (34)
Sensitivity (%)	55.9	66.7	28.1	69.7	<u>71.9</u>	67.6	73.5
Specificity (%)	85.2	78.3	88.0	<u>87.5</u>	54.5	88.9	88.9
Balanced Accuracy(%)	70.5	72.5	58.1	<u>78.6</u>	63.2	78.2	81.2

Table 4.10: Prediction results for the seven models: M_{Clinic} , M_{T1} , M_{T1c} , M_{T2} , M_{FLAIR} , M_{Multi} and MM_{Age} in a LOO-CV framework. For each prediction task and for each model, four figures of merit are reported: the total number of patients for which the prediction was possible (the number of patients with mutation is between brackets), the sensitivity, the specificity and the balanced accuracy. For each figure of merit, best results are in bold characters and second best results are underlined.

4.5 Discussion

Indeed the selection of logistic regression proved to be an appropriate choice for classifier. When integrating SVM in the multi-modal approach the prediction performance was similar to that of LR in the LOO-CV framework. Balanced accuracies of 78% in the LOO-CV using LR for TP53 mutation prediction (relatively balanced classes) and of 77% when using SVM are recorded. Whereas when using RF we observe a jump in balanced accuracies for each of the mono-modality model and hence so for the multi-modal model, this maybe due to the very high number of trees (100) assigned to the classifier as a default. On the other hand, when predicting H3.1 versus all mutation a relatively unbalanced prediction task, LR maintains its good performance, SVM seems to be not working well and RF display yet again over estimated results. LR was adapted for feature selection integrating a wrapper based feature elimination method. Introducing age profiling of patients in the MM-LOOCV-MIM framework for H3.1 mutation prediction did not seem to enhance or deteriorate performances. Where as, for ACVR1 mutation prediction by Multi-modal model with age weight (MM_{Age}) seemed to have dropped, when investigated this is due to miss-classification of 2 patients by the model. TP53 mutation prediction was seen to have increased its accuracy by 3% by the Age profiling. Published literature highlights most patients with age (>10 years) often do not have H3.1 mutation nor ACVR1 mutation. However, the experiments we conducted setting the age profiling cutoff to 10 years may not be the best for the three mutations. The importance of age alone stands evident for mutation prediction as the mono-model clinic achieves 70% balanced accuracy for two out of the three mutations. When comparing cross validation 5, 10 fold and the leave one out indicates a potential overestimation of performances by the LOO-CV approach

but when working with a relatively small cohort with missing imaging modalities and having an estimation for each patient LOO-CV best suited this purpose.

4.6 Conclusion

Integrating clinical data with radiomic features derived from MR to predict genomic mutations such as H3.1, ACVR1, and TP53 has yielded enhanced outcomes. The evolution from the 16-model strategy shown in chapter 3 to the Multi-modal model within the LOOCV framework has improved mutation prediction accuracy and solved some limitations inherent in the 16-model strategy. The feature selection process efficiently identifies the three or four most pertinent features, which adeptly forecast the presence or absence of mutations. Among various classifiers evaluated in the MM-LOOCV-MIM approach, logistic regression consistently outperformed random forest and support vector machine for the triad mutation prediction task. The inherent LOOCV methodology provides individual patient prediction results, showcasing its capability to manage missing modalities. This approach might offer physicians a viable alternative in situations where biopsies are unfeasible.

Chapter 5

Radiomic analyses to predict overall survival and long survivors DIPG

5.1 Introduction

Survival analysis involves using statistical methods to analyze time until an event such as death occurs. The methods are used to identify factors that are associated with longer survival times. The most common methods used in survival analysis are the Kaplan-Meier curves and Cox proportional hazard model. The Cox proportional hazards model is a parametric method that estimates the hazard ratio (i.e., the probability of an event occurring in a small time interval) for one group of patients to the hazard rate for another group of patients. In the context of DIPG and DMG, survival analysis may take into account imaging features, such as tumor size, shape and location, as well as patient characteristics, such as age, sex, and genomic data. These data can be used to identify prognostic factors and to develop predictive models that can be used to guide treatment decisions and to maybe improve patient outcomes.

5.2 State of the Art: Survival analysis for DIPG

A Pubmed (<https://pubmed.ncbi.nlm.nih.gov/>) search with ((DMG or DIPG AND MRI AND survival)) OR ((DIPG Cox Overall survival)) OR ((DIPG radiomics Overall survival)) (“2008” until [April 2023])) criteria highlighted 34 articles. From these two [132] [133] were about developing radiomic models for predicting overall survival (OS) or PFS while the others focused on radiological; clinical and genomic criteria for addressing survival prediction tasks.

5.2.1 Meta analysis for high grade brain stem glioma

The work of (Hassan et al in 2017) [217] includes sixty-five studies (2336 participants) reports a systematic review and meta-analysis undertaken to determine the survival rates and assess potential prognostic factors including selected interventions. Studies included involved pediatric participants with high grade brain stem gliomas diagnosed by magnetic resonance imaging or biopsy reporting overall survival rates. Meta-analysis was undertaken using a binomial random effects model Meta-analysis showed 1 year overall survival (OS) of 41% (2083 participants), 2 year OS of 15.3% (1329 participants) and 3 year OS of 7.3% (584 participants). Subgroup analysis comparing date of study, classification of tumor, use of temozolomide, non-standard interventions or phase 1/2 versus other studies demonstrated no difference in survival outcomes. There was insufficient data to undertake subgroup meta-analysis of patient age, duration of symptoms, K27M histone mutations and AVCR1 mutations. Survival outcomes of high grade brain stem gliomas have remained very poor, and do not clearly vary according to classification, phase of study or use of different therapeutic interventions. Future studies should harmonize outcome and prognostic variable reporting to enable accurate meta-analysis and better exploration of prognosis. Survival from high-grade brain stem gliomas in childhood remains very poor, with this systematic review estimating that only four in ten young people diagnosed with a DIPG will be alive at one year after diagnosis. The studies do not clearly demonstrate an improvement over time, or show any major impact of chemotherapy or alternative radiotherapy approaches.

5.2.2 Use of MRI to predict survival in DIPG or DMG

The papers encompass all studies reporting on diffuse intrinsic pontine glioma (DIPG) or diffuse midline glioma (DMG), which use magnetic resonance (MR) imaging to investigate tumor localization and characteristics along with overall survival and progression-free survival. Three publications [208], [209] and [210] between 2008 and 2011 that focus on the ability and effectiveness of MR imaging for DIPG prognosis. In one such study, conducted by Darren Hargrave et al. in 2008 [208], 39 patients with confirmed DIPG were included in the analysis. Of these, 37 patients, 2 patients were still alive at the time of follow up, a median survival of 9.5 months was calculated. Overall survival was calculated using Kaplan-Meier methodology, which estimated survival rates of 78% at 6 months, 32% at 1 year, and 5% at 2 years. The authors investigated the clinical and MRI characteristics of patients treated for

Reference	Type of tumor of CNS	Number of patients	Number of tumors	Question	Treatment	Modalities/ Imaging
N. U. Barua et al [203]	DIPG	1M	1	Survival	Radiotherapy / surgery	MRI
K. Muller et al [204]	DIPG	1F	Multiple	Overall survival	Radiotherapy	MRI
Takahiro Ono et al [205]	DMG	1M	1	Therapy	Chemotherapy/ surgery	MRI
V.Ye, et al [206]	DIPG	1M	1	Mutation characteristics	Adjuvant chemoradiotherapy	MRI
T. Picart et al [207]	DMG	17(11M/6F)	Multiple	Overall survival	Radio-chemotherapy / surgery	MR-spectroscopy and perfusion MRI.
D. Hargrave et al [208]	DIPG	39(20M/19F)	Multiple	Overall survival	Radio-chemotherapy	MRI
R.M. Hayward et al [209]	DIPG	16(6M/10F)	Multiple	Overall survival	Radiotherapy/ PEG-Intron	MRI
E.A. Steffen-Smith et al [210]	DIPG	38(13M/25F)	Multiple	Overall survival	Radiation and Chemotherapy	MRI, SVS ^a and MRS ^b
R.M. Lober et al [211]	DIPG	30(14M/16F)	Multiple	Overall survival	Radiation/Chemotherapy/cerebrospinal fluid diversion	MRI (ADC)
M.H. Jansen et al [212]	DIPG	316(160M/156F)	Multiple	12, 25 month OS	Chemo-Radiotherapy and Intravenous chemotherapy	MRI
T.Y. Poussaint et al [213]	DIPG	140	1	PSF and OS	Radiation therapy	MRI (ADC)
Zhou et al [214]	DIPG	32	1	OS	Immunohistochemistry	MRI
Z. Gokce-Samar et al [215]	DIPG	16(9M/7F)	Multiple	OS	Chemo-Radiotherapy and combination therapy	MRI
Z. Zhou et al [216]	DIPG	41(15M/26F)	Multiple	OS	Immunohistochemistry	MRI
H. Hassan et al [217]	DIPG	65	Multiple	OS	Radio-chemotherapy / biopsy	MRI
R. Calmon et al [218]	DIPG	43	Multiple	OS and true prognosis	Radiation therapy	MRI
S. E. M. Veldhuijzen van Zanten et al [219]	DIPG	316(160M/156F)	Multiple	OS	Radiotherapy, oral chemotherapy and Intravenous chemotherapy	MRI
G. S. Colafati et al [220]	DIPG	29(15M/14F)	1	OS	Radiotherapy in combination with Nimotuzumab and Vinorelbine	MRI
F. Garibotto et al [221]	DIPG	18(8M/10F)	Multiple	PFS and OS	Radio-chemotherapy	MRI and FDG PET
J.L. Leach et al [222]	DIPG	357	Multiple	OS	Radio-chemotherapy	MRI
C.L. Tinkle et al [223]	DIPG	105	Multiple	Survival analysis Prognosis	Radiotherapy / fractionated adjuvant chemotherapy	MRI
M. Ianno et al [224]	DIPG	47(22M/25F)	Multiple	PFS and OS	Radiation therapy	MRI
H Ju Kim et al [225]	DIPG	162(89M/73F)	Multiple	prognosis and OS	Radiation therapy	MRI
E. Cantor et al [226]	DIPG	24	1	PFS and long term survivors	None	MRI
J. Liu et al [227]	DMG	38(12M/26F)	Multiple	OS and PFS	Surgery	MRI
S. Venneti et al [228]	DIPG	50	Single	OS and PFS	ONC201	MRI
E. R. Jackson et al [229]	DMG	15 (9M/6F)	Multiple	OS and PFS	ONC201 and Paxalisib	MRI
K.A. Zukiyanski et al [230]	DIPG	33(25F/8M)	Multiple	OS and PFS	Radiation therapy	MRI vs PET
G.Morana [231]	DIPG	19(7M/12F)	1	OS	Chemo-Radiotherapy	MRI vs PET
L. T. Tam [133]	DIPG	177	Multiple	OS	Treatment naive	MRI radiomics
M.W. Wagner [132]	DIPG	89 (41M/48F)	Multiple	PFS	Treatment naive	MRI radiomics

Table 5.1: The table lists all papers included in the literature review sorted into different categories based on the type of study conducted in chronological order. The selected papers are categorized into six categories. Twenty-one papers review were based on clinical studies whereas two papers included radiomic studies.

^aSingle Voxel Spectroscopy.

^bMagnetic Resonance Spectroscopic Imaging

DIPG before and after treatment (biopsy 23%, radiotherapy 97% and chemotherapy 54%) to assess their predictive value for overall survival. Interestingly, no MRI parameter from either the diagnostic or the response scans predicted the prognosis. The study suggests that DIPG clinical trials should focus on overall survival as their primary endpoint and emphasizes the potential of advanced imaging techniques over traditional MRI response criteria. Another study published in 2008 by Robert M. Hayward et al. [209] evaluated inter-observer variability in DIPG measurements using MR imaging. The study used 50 MR scans consisting of fluid-attenuated inversion recovery (FLAIR), T2-weighted, and T1-weighted images from 16 patients with DIPG. The results showed significant variability in DIPG tumor measurements between observers, and FLAIR imaging was the most consistent. Therefore, for patients on clinical trials, DIPG measurements should be performed by a single reader while comparing prior images side-by-side. Lastly, a study published in 2011 by Steffen-Smith et al. [210] evaluated changes in magnetic resonance spectroscopy (MRS) biomarkers in 38 patients with DIPG. The median overall survival for the entire cohort was 14.8 months, with six patients having a survival rate greater than 24 months. Using MRS features in a univariate cox proportional hazards model, the study found that patients with higher single voxel spectroscopy (SVS) Cho:NAA values at their first scan were at greater risk of mortality compared to patients with lower SVS Cho:NAA values. Several studies have explored the use of apparent diffusion coefficient (ADC) as a potential biomarker for predicting survival outcomes in patients with brain tumors. A study by Lober et al. (2014) [211] utilized diffusion-weighted MRI-derived ADC values to categorize tumors into low and high diffusion groups, with distinct median survivals of 3 and 13 months, respectively in 30 patients. The study found that low ADC tumors were only observed in male patients, while high ADC tumors were observed in both male and female patients. Another study by Poussaint et al. (2016) [213] analyzed pre- and post-treatment ADC histograms to identify tumor characteristics associated with shorter or longer progression-free survival (PFS) and overall survival (OS). The study found that tumors with higher ADC was associated with PFS, while enhancing tumors with bimodal enhancement histograms had worse PFS and OS compared to unimodal cases. This study also used the Cox proportional hazards model to analyze the relationship between imaging features and survival outcomes.

In this work Tinkle et al in 2020 [223] used log-rank and Gray's tests and Cox proportional hazard model for identifying survival predictors. Calmon et al [218] examined 43 children with DIPG who had undergone multimodal MRIs at four time points. The patients were divided into two groups based on whether they showed conventional MRI changes that resembled disease progression. The study recorded values for each tumor voxel, avoiding necrotic areas, of the apparent diffusion coefficient, arterial spin labeling cerebral blood flow (ASL-CBF), and dynamic susceptibility contrast perfusion relative cerebral blood volume (DSCrCBV) and flow (DSCrCBF). The study found that 44% of the patients showed radiological signs that mimicked progression after radiotherapy. However, 16 of these patients survived for more than six months, and their pseudo progression lasted for a median of 8.9 months and a maximum of 35.6 months. The study also found that all patients showed an increase in blood volume and flow after radiotherapy, but those with signs of pseudo progression had a greater increase of ASL-CBF. There was no signifi-

cant difference in survival between the two groups. During true progression, only patients who had not experienced pseudo progression showed an increase in DSCrCBF and DSCrCBV values. In the study of Colafati et al in 2019 [220] the authors investigate the involvement of cranial nerve V (CN V) was investigate at diagnosis and its utility as predictor of poor overall survival. Differences in overall survival (OS) and time to progression (TTP) were analyzed for involvement of CN V, sex, age, tumor size, ring enhancement, and treatment regimen. The study claims direct involvement of the cranial nerve V with short surviving patients in a cohort of 29 patients after excluding long survivors. The authors conclude that in DIPG direct involvement of CN V should be routinely evaluated on diagnostic scans.

5.2.3 Survival analysis based on FDG PET and MR

The work of Zukotynski 2017 et al in [230] describes baseline 18F-FDG PET voxel characteristics in pediatric diffuse intrinsic pontine glioma (DIPG) and to correlate these metrics with baseline MRI apparent diffusion coefficient (ADC) histogram metrics, progression-free survival (PFS), and overall survival. The authors indicate In the Pediatric Brain Tumor Consortium's clinical trials for DIPG, 33 children underwent baseline brain scans using both 18F-FDG PET and MRI. The resulting images from 18F-FDG PET, post gadolinium MR, and ADC MR were aligned to the initial FLAIR scan. Three-dimensional regions of interest were generated on FLAIR images and T1c MRI 18F-FDG PET and MRI ADC histograms inside each modality. Metrics evaluated included peak number, skewness, and kurtosis. Correlation between PET and MR ADC histogram metrics was evaluated. PET pixel values within the region of interest for each tumor were plotted against MR ADC values. The association of these imaging markers with survival was described. The work concludes that F-FDG PET and MR ADC histogram metrics in pediatric DIPG demonstrate different characteristics with often a negative correlation between PET and MR ADC pixel values. A higher negative correlation is associated with a worse PFS, which may indicate higher-grade elements within the tumor.

Another similar study exploring the importance of PET and MR imaging published by G. Morana et al in 2020 [231] evaluated the contribution of 18F-dihydroxyphenylalanine (DOPA) PET in association with conventional MRI in predicting treatment response and survival outcome of pediatric patients with DIPGs. Using a cohort of 19 patients, was evaluated the correlation between 18F-DOPA uptake tumor volume at admission and MRI tumor volume following treatment along with statistical analysis of overall survival. Kaplan-Meier OS curves for all the main risk factors were analyzed in this study. Subjects with Telomere to Single-copy gene (T/S) ratios > 1 (markedly increased uptake) had a significantly higher risk of death. Univariate analysis reported that H3K27M-mutant lesions and patients with ring enhancement on post-contrast T1-weighted images had a significantly lower overall survival. Patients with larger MRI tumor volumes at admission did not show lower OS. By contrast, post-treatment MRI tumor volume at maximum response and tumor volume reduction following treatment were significantly associated with

OS. Median age and gender were not correlated with OS. In conclusion, the study reports ring enhancement and MRI tumor volume reduction following treatment resulted to be significant predictors of OS.

5.2.4 New therapeutic protocols: Onc201

Research on therapy measures for DIPG is ongoing and aims to improve the survival rate and quality of life of patients with this disease. Current therapeutic measures being researched for DIPG include radiation therapy, chemotherapy, immunotherapy, and targeted therapies. Radiation therapy is the standard of care for DIPG and can help to relieve symptoms and prolong survival.

ONC201, also known as TIC10 (TRAIL-inducing compound 10), is a small molecule that exhibits anti-cancer properties. Its mechanism of action is distinct; ONC201 induces a potent anti-cancer protein called TRAIL (TNF-related apoptosis-inducing ligand) in various tumor types, leading to selective cancer cell death [232]. Interestingly, ONC201 has been shown to specifically target and kill cancer stem-like cells, which are often resistant to conventional therapies [233]. This molecule works by activating an integrated stress response that results in the up regulation of a transcription factor called ATF4, which in turn leads to the induction of the TRAIL gene [228]. Moreover, ONC201 has demonstrated effectiveness against chemotherapy-resistant colorectal cancer stem-like cells, exerting its effects through an Akt/Foxo3a/TRAIL-dependent mechanism [233]. Clinically, ONC201 has shown promise in early-phase trials. For instance, a first-in-human clinical trial demonstrated its safety and potential efficacy in patients with refractory solid tumors [234]. Another notable aspect of ONC201's mechanism is its ability to induce the ATF4 transcription factor through an atypical integrated stress response, leading to p53-independent apoptosis in hematological malignancies [235]. Given these attributes, ONC201 has garnered attention as a potential novel treatment for various cancers.

Cantor et al in 2022 [226] conducted a study to test the effectiveness of the drug ONC201 on children with H3K27M-mutant glioma, a type of brain cancer. Patients enrolled in the trial underwent serial lumbar puncture for cell-free tumor DNA analysis, and the results were compared to radiographic changes in the tumor. The study found that a decrease in the level of H3K27M variant allele fraction (VAF) over time, as measured in both cerebrospinal fluid (CSF) and plasma samples, was associated with prolonged progression-free survival (PFS) in non recurrent patients. VAF spikes, which are sudden increases in VAF levels, were found to precede tumor progression in a significant percentage of cases. In some individual cases, an early reduction in H3K27M VAF levels predicted a long-term clinical response to ONC201, and did not increase the risk of later-defined pseudo-progression. The multicenter, open-label, single-arm phase II clinical trial reported in [236] aimed to evaluate the effects of ONC201 in pediatric patients diagnosed with H3K27M-mutant DIPG. Within the study, a cohort of 50 pediatric patients, predominantly male (58%) with a median age of 6.3 years, was administered oral ONC201 bi-weekly until either disease progression or the manifestation of unacceptable toxicity. MRI assessments were routinely carried out

every 8 weeks, and available archived tumor samples underwent molecular profiling. In terms of efficacy, the 6-month progression-free survival (PFS6) rate was recorded at 28%, with a median progression-free survival of 4.4 months and an overall survival median of 11.8 months. Notably, ONC201 exhibited commendable tolerability. The predominant treatment-related adverse events reported were increased ALT/AST levels and mild mood alterations. Molecular analysis, available for 24 patients, revealed that 67% had TP53 mutations, 33% manifested ACVR1 mutations, and 29% exhibited both. However, these molecular subgroups did not significantly influence survival outcomes. In conclusion, the study indicates that ONC201 demonstrates moderate single-agent activity in pediatric H3K27M-mutant DIPG patients. Given its well-tolerated nature, there's a compelling case for subsequent studies to explore ONC201 in combination with other therapeutic agents. Another study evaluating the effects of ONC201 was conducted by Jackson et al [229] to assess the safety, tolerability, and preliminary effectiveness of ONC201 when combined with Paxalisib for patients diagnosed with H3K27M-mutant diffuse midline gliomas. The research encompassed a cohort of 15 patients, with an average age of 38.7 years, ranging from 6 to 70 years. These individuals were either new to treatment or had exhibited progression post-standard therapy. Patients in the study were administered ONC201 orally on a weekly basis, in conjunction with a daily dose of Paxalisib. Treatment efficacy was routinely monitored through radiographic evaluations. From the results, 3 out of the 15 participants showed a partial response to the treatment, while 8 maintained stable disease. This translated to an overall response rate of 20% and a disease control rate of 73.3%. In terms of survival metrics, the median progression-free survival stood at 8.1 months, with the median overall survival yet to be determined at the time of reporting. In relation to the treatment's safety profile, 13 out of the 15 patients encountered treatment-related adverse events, predominantly of a low grade, including symptoms like elevated liver enzymes, fatigue, and rash. Conclusively, the combined treatment regimen of ONC201 and Paxalisib appears to hold therapeutic promise for H3K27M-mutant DMG patients, presenting a commendable safety record and encouraging preliminary efficacy results. This is particularly significant considering the aggressive nature of the disease and the limited therapeutic alternatives currently at hand.

5.2.5 Radiomics

Two articles were based on radiomics to assess DIPG's, study conducted by Matthias W. Wagner et al 2022 [132], used a conditional survival forest model to predict progression-free survival (PFS). Having a dataset of 89 patients that they divided (80:20 Training: test). Tumor segmentation was performed by a 4th year radiology resident using 3D Slicer. Semi-automated tumor segmentation on FLAIR and nonenhanced T1-weighted sequences was performed with the Level-Tracing-Effect tool. A final placement of tumor contour was done by an expert of 7-year pediatric neuroradiology research experience. The scans underwent bias field correction along with z-score normalization. Radiomic features extracted included histogram, shape, and texture features with and without wavelet-based filters. The authors provide details of feature extraction in supplementary document. 107 FLAIR and 149 T1 features were

selected after removing features with a high degree of correlation (>90%). The stability of radiomic features across scanners was measured using intraclass correlation coefficient (ICC). The patients were assigned to a training set consisting of 80% of the available cases (66 patients with FLAIR sequences and 54 patients with nonenhanced T1-weighted sequences). The remainder of the patients comprised the test sets (16 patients with FLAIR sequences and 13 patients with nonenhanced T1-weighted sequences). Conditional survival forest-based concordances of the test set were averaged over 100 randomized repeats at 3, 4, 5, 6, and 7 months PFS for both sequences. For the combined prediction of FLAIR and nonenhanced T1-weighted sequences, averaged concordances were .74 at 3 months, .84 at 4 months, .72 at 5 months, .63 at 6 months, and .63 at 7 months. The study concluded that MRI-based radiomic features hold significant potential as non-invasive biomarkers for predicting progression-free survival in pediatric patients with DIPG. This radiomic signature could serve as an essential tool for risk stratification, guiding therapeutic decisions, and tailoring individualized treatment plans for these patients.

The second article presented by Lydia Tam et al 2021 [133] is based on the analysis of an international data set of 177 treatment naive patients with DIPG. Manual delineation of tumor boundaries was performed independently on T2-MRI and over the corresponding tumor boundary on T1-MRI regardless of enhancement using Osirix software. Prior to feature extraction, images were normalized and resampled to isotropic 1 mm voxels. Image features are extracted using T1, T2 images from within the tumor region using open source pyradiomics. A total of 900 features were extracted on each T2-MRI and T1-MRI. Extracted features included size, shape, first-order, and texture-based features computed on original, wavelet, and Laplacian of Gaussian filtered images. The complete dataset was randomly divided into training (60%, n = 106) and test (40%, n = 71) sets. The training set was used to select the optimal features to predict overall survival (OS) and build a Cox regression model. Clinical features alone (age at diagnosis and sex), radiomic features alone, and the combination of clinical and radiomic features were considered. The authors performed 100 repetitions of 10-fold cross-validation to fit a Cox regression model using the least absolute shrinkage and selection operator (LASSO) regularization ($\alpha = 1$). It is important to note not all patients had both T1 and T2 scans available. The authors proceeded by using 95 patients with both T1 and T2 as in the training dataset. The lambda value with the minimum cross-validated error across the 100 repetitions resulted in a total of 5 features with non-zero coefficients. The T1-MRI features included wavelet (LLH) gray-level co-occurrence matrix (GLCM) inverse difference normalized (IDN), wavelet (LHH) GLCM informational measure of correlation 2 (IMC2), and wavelet (HHH) GLCM IMC2. The T2-MRI features were wavelet (LLH) GLCM IDN and wavelet (HHH) first-order mean. The study evaluated the performance of the three T1 features on all patients with gadolinium-enhanced T1-MRI available. Likewise, the two T2 features were evaluated on all patients who had T2-MRI available. The performance of individual sequences was lower compared to the combination of T1 and T2 MRI radiomic features. When clinical features were combined with radiomics, the model performance increased. No significant correlation was found between the selected radiomic features and with clinical features. The developed radiomics signature showcased impressive predictive capabilities, with an Area Under the Curve (AUC) of 0.883

in the training dataset and 0.867 in the validation set, indicating its robustness and reliability. Additionally, the calibration curve of the signature revealed a strong alignment between the predicted and actual outcomes, further emphasizing its accuracy. Through decision curve analysis, the study also highlighted the potential clinical utility of the radiomics signature. In conclusion, the results underscore the potential of MRI-based radiomics as a valuable tool in predicting the prognosis of pediatric patients with DMG carrying the H3K27M mutation. The study advocates for the further exploration and incorporation of radiomics in clinical practices to enhance patient management and treatment decisions.

5.3 Methods for assessing Overall Survival

Our aim is to find potential imaging biomarkers or genomic mutations that may be linked to improved outcomes by predicting the long-term survival of DIPG patients. This information could be used to propose targeted medicines and better understand the disease. In addition to conventional tools used for survival analysis, I implemented the Multi-modal (LOOCV-MIM) approach introduced in the previous chapter for predicting survival of patients surviving at least 2 years, 18 months, and 1 year. The data used for predicting long survival patients is defined in the table [A.1](#). We have survival information for all 80 patients. From these 80 patients, seven are censored, indicating one patient is censored after 1787 days. This patient was included for the long survival (2 years) prediction task. The six censored patients (P36, P46, P47, P61, P64, P69) were removed. For the 74 patients clinical features age, sex and tumor volume are known. Fourteen shape based features extracted from delineated tumor are also known. Hence, a dedicated model ($M_{Clinic+Shape}$) combining clinical and shape features was developed. The mono-modal MR based radiomic models with missing modalities stay the same as described in the previous chapter.

5.3.1 Prediction of long survivors using the multi-model framework

The extraction of radiomic characteristics is a standard procedure in the field of medical imaging analysis. These features are typically hand-crafted or automatically extracted to find distinguishing patterns in the imaging data. When several features are extracted they are connected due to several redundancies. In addition to shape-based, first-, second-, and higher order statistical features were computed, as well as model-based features like fractals and dynamic features they may include a variety of statistical determinants. When dealing with outcome modeling, it's not unusual to deal with hundreds radiomic features, especially when combining features extracted from different modalities in this study. Thus, choosing the best feature subset or feature representation that correlates the most with the endpoint while also having the lowest correlation with other features is a crucial step. Based on the feature subset that has been obtained, various machine learning techniques can be used. The feature selection method described in chapter 4 was used. This approach gradually eliminates the weakest feature while iteratively fitting

a model, a logistic regression model was chosen for this purpose. As a result, the wrapper based RFE approach removes co-linearity and interdependence between the various model features. A grid analysis is employed to introduce variation (between 0.1 and 1 with a step size of 0.1) for the C parameter, which is the inverse of the regularization strength, in order to apply the L1 penalty used for the logistic regression model. On the validation set, feature importance was determined by computing the Brier score loss. A two-fold cross-validation was used to guide the 40 RFE iterations. The top four features that were most commonly chosen were selected. The method is applied to the five models the model based on clinic and shape features and the four models with features extracted from the four MR modalities.

5.3.2 Predicting Long Term Survivors

For predicting long surviving patients we implement the Multi-model approach as it is well adapted to handling the missing data issue. The approach consists of a leave-one-out cross-validation (LOO-CV) framework to compare and combine the independent mono-models to take advantage of all available data with missing MR modalities. We made six separate models using logistic regression for each of the four MRI types, as well as one for combined shape and clinical features, and another that included genomic mutation details. We then combined the predictions from these six models to make our final combined model. We tested how well these models predicted outcomes using a leave-one-out cross-validation method. Suitable criteria, including balanced accuracy, sensitivity and specificity were used to assess each model's performance. The LOO-CV study results gave insights into how well the models predicted the outcomes and made it possible to choose the top-performing model for predicting long surviving patients. The applied methodology offers a mythological strategy for contrasting and choosing models in datasets with few samples and missing data.

5.3.3 Survival Prediction continuous output ICARE

The binary weighted model (ICARE) was employed in survival analysis to find patient prognostic features. Each feature in this model is given a binary weight, indicating whether or not it is a predictor of survival. The risk of mortality is then predicted by the model using these binary weights. A leave-pair-out (LPO) technique and the concordance index (C-index) can be combined to assess ICARE performance for patients in the DIPG cohort. This strategy is not too dissimilar from the leave one out strategy used in the multi-modal with missing modalities strategy. The LPO method entails fitting the model to all patient pairs, except in one pair, then utilizing the results to forecast how the remaining pair will do. Repeated for all pairs the fraction of couples that were correctly predicted is used to determine the C-index. A score of 0.5 for the C-index means that the model's predictions are no better than random chance, the C-index ranges from 0.5 (random chance) to 1.0 (perfect prediction). A C-index of 0.7 or higher is typically regarded as a solid performance, while a score of above 0.8 is regarded as excellent. Overall, the LPO

methodology and the C-index offer a thorough and reliable technique for assessing ICARE's performance in survival analysis.

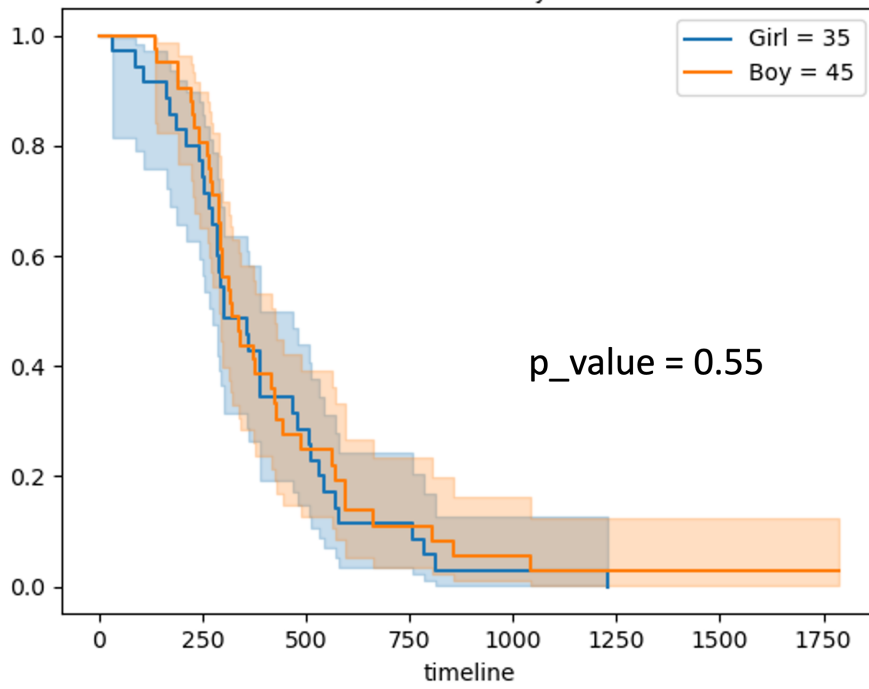
5.4 Results

The following section will display results recorded by the experiments conducted for Kaplan Meier analysis, the prediction of long term survivors and using the MM-LOOCV-MIM and the ICARE algorithm.

5.4.1 Kaplan Meier curves and Log-rank tests for overall survival

The figures [5.1](#), [5.2](#), [5.3](#), and [5.4](#) display Kaplan Meier (KM) curves according to sex, and the three mutations, H3.1 versus others, ACVR1 and TP53. The construction of each curve was based on the Python lifelines library. The p-values of the Logrank test are displayed on each graph. The presence of the TP53 mutation is more aggressive than its absence, tumors with a mutation of H3.1 suggest a slightly higher risk of death, and those with H3.2 or H3.3 mutation and the absence of the ACVR1 mutation is more aggressive than its presence.

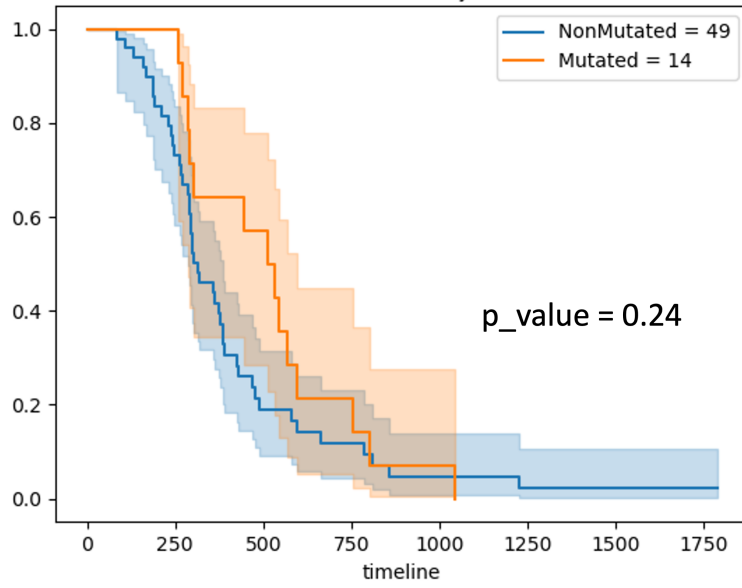
Kaplan Meier curves for OS stratified by Sex of patients
Number of days



Girl = 35								
At risk	35	26	10	4	1	0	0	0
Censored	0	0	0	0	0	0	0	0
Events	0	9	25	31	34	35	35	35
Boy = 45								
At risk	45	33	9	4	2	1	1	1
Censored	0	4	6	6	6	6	6	6
Events	0	8	30	35	37	38	38	38

Figure 5.1: Kaplan Meier curves for patients stratified by Sex of patient

Kaplan Meier curves for OS stratified by H3.1 (versus all others) mutation status
Number of days



NonMutated = 49

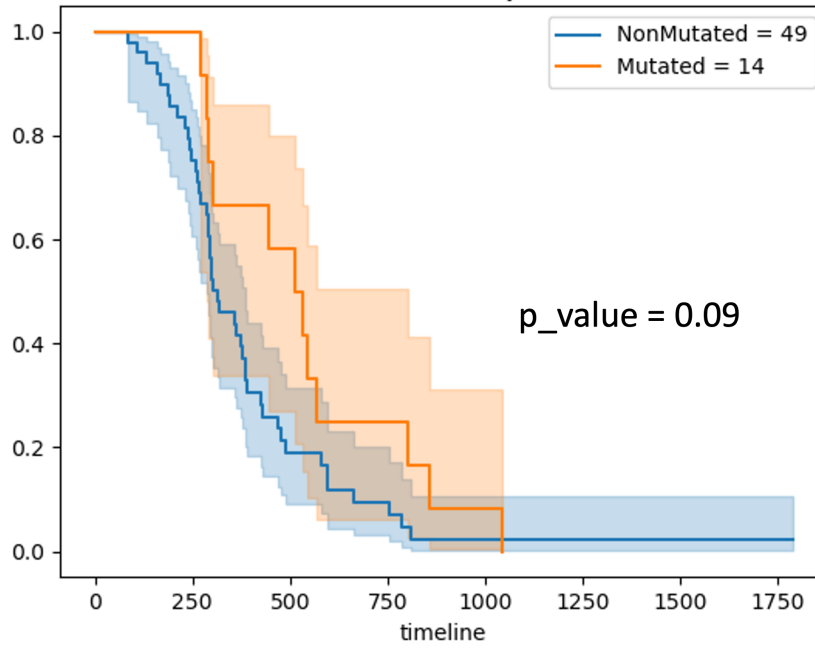
At risk	49	35	8	5	2	1	1	1
Censored	0	1	3	3	3	3	3	3
Events	0	13	38	41	44	45	45	45

Mutated = 14

At risk	14	14	8	3	1	0	0	0
Censored	0	0	0	0	0	0	0	0
Events	0	0	6	11	13	14	14	14

Figure 5.2: Kaplan Meier curves for patients with known H3.1 mutation status

Kaplan Meier curves for OS stratified by ACVR1 mutation status
Number of days



NonMutated = 49								
At risk	49	36	8	4	1	1	1	1
Censored	0	1	3	3	3	3	3	3
Events	0	12	38	42	45	45	45	45
Mutated = 14								
At risk	14	12	7	3	1	0	0	0
Censored	0	2	2	2	2	2	2	2
Events	0	0	5	9	11	12	12	12

Figure 5.3: Kaplan Meier curves for patients with known ACVR1 mutation status

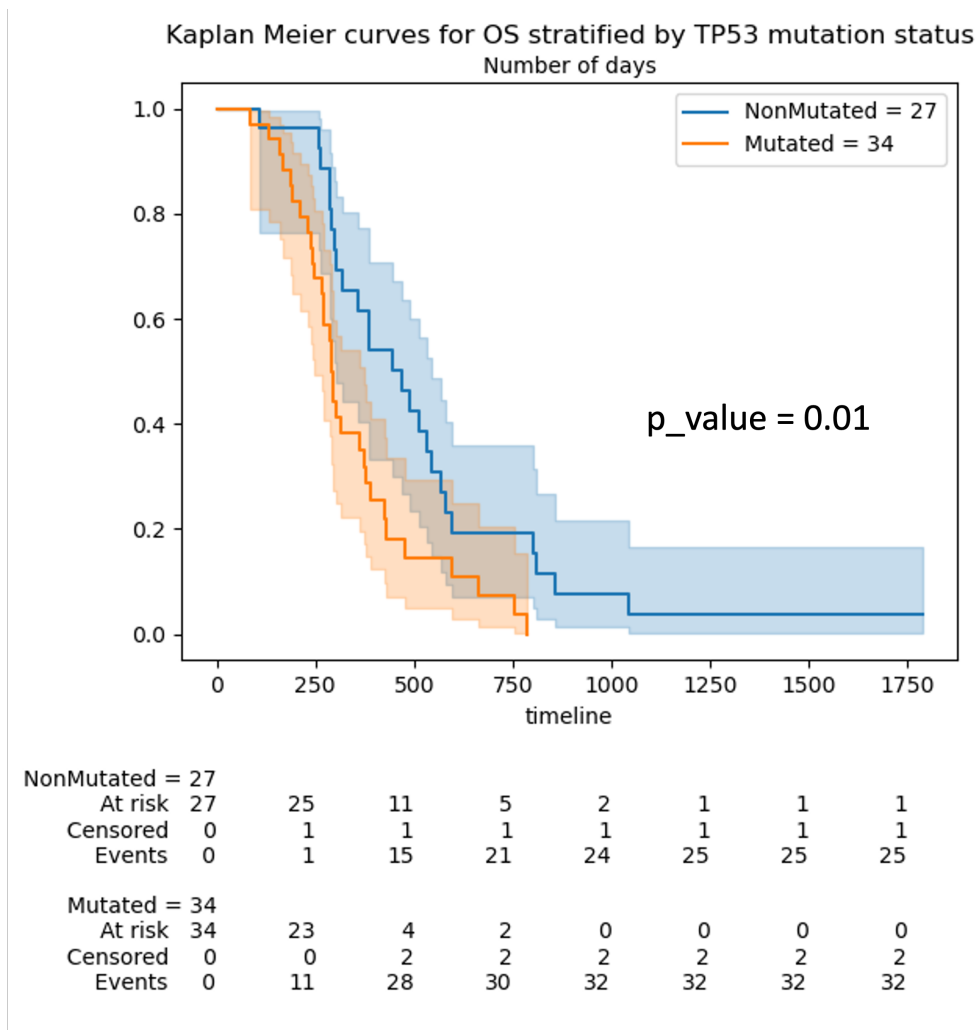


Figure 5.4: Kaplan Meier curves for patients with known TP53 mutation status

5.4.2 Predicting long surviving patients (2 years) using the Multi-modal approach

Feature selection

The final selected features for long survival prediction are presented in the table [5.2](#). These features are obtained using 74 patients for the model combined clinic and shape having a combined set of 17 features, 15 shape features pooled with two clinic features age and sex. Each MR modality had 79 retrieved radiomic characteristics. A total of 67 patients for T1 MR modality, 71 patients for T1c MR modality, 72 patients for T2 MR modality and 66 for FLAIR modality. The final features selected from the MR modalities are mostly first order and a few Gray Level Co-occurrence Matrix (GLCM) features. The final four features selected for the clinical + shape data are volume, MajorAxisLength, Elongation and Flatness. High elongation values are observed for long survivors. Major Axis Length is related to the size of the tumors. The flatness characteristic determine the lesion extent along a specific axis, its partly correlated with elongation.

Mono-Modal	Features Name	Features Identifier
M_{T1}	T1_original_firstorder_InterquartileRange	F1
	T1_original_glcmlmc2	F2
	T1_original_firstorder_Skewness	F3
	T1_original_firstorder_Uniformity	F4
$M_{Clinic+Shape}$	original_shape_MajorAxisLength	F5
	Volume	F6
	original_shape_Elongation	F7
	original_shape_Flatness	F8
M_{T1c}	T1c_original_glrmlm_RunLengthNonUniformity	F9
	T1c_original_firstorder_TotalEnergy	F10
	T1c_original_firstorder_Kurtosis	F11
	T1c_original_glcmlmc2	F12
M_{T2}	T2_original_firstorder_Minimum	F13
	T2_original_firstorder_Skewness	F14
	T2_original_firstorder_Kurtosis	F15
	T2_original_firstorder_10Percentile	F16
M_{FLAIR}	FLAIR_original_glcmlmc2_ClusterShade	F17
	FLAIR_original_glcmlmc2_Idmn	F18
	FLAIR_original_glcmlmc2_InverseVariance	F19
	FLAIR_original_firstorder_Minimum	F20

Table 5.2: Subsets of features selected by the five different models $M_{Clinic+Shape}$, M_{T1} , M_{T1c} , M_{T2} , and M_{FLAIR} to predict 2 years long survivors. $M_{Clinic+Shape}$ contains the combined features from clinic and shape data.

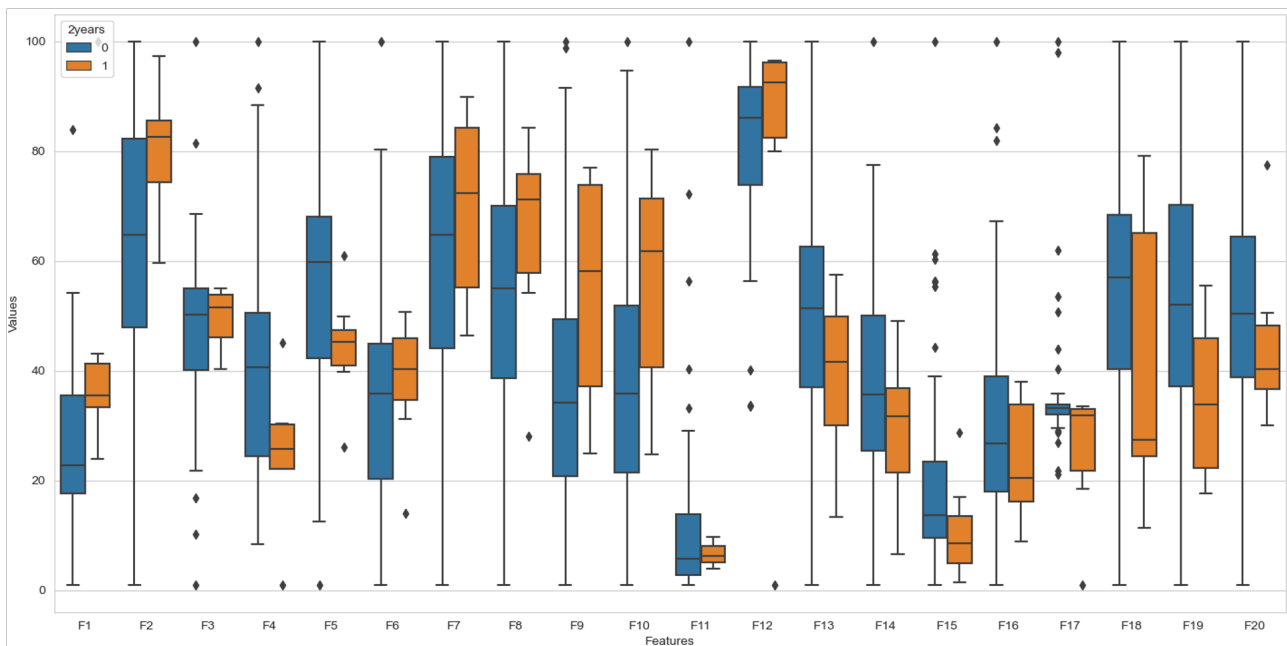


Figure 5.5: The box plots represent the distribution of 20 selected feature for long term survivors. The feature values are scaled between 0 and 100. The blue box plot represent the feature values of patients who are not long survivors marked as '0' and the orange box represents the feature distribution for long term survivors marked as '1'. The x-axis denotes the feature IDs from table 5.2. The y-axis denotes the feature values scaled between 0 and 100.

The figure 5.5 displays a box plot representing the distribution of the twenty selected feature values for long and short term surviving patients. For better visual assessment, the feature values are scaled between 0 and 100. The box plots display how the selected feature values differ between the two groups of patients. For instance features F1, F2, F4, F5, F9, F10, F16, F18 and F19 have large differences in distribution and median values for the two groups of patients. Features F6, F7, F8, F12, F13, F14, F15 and F20 exhibit smaller difference in median values in the two groups. Finally three features, F3, F11 and F17 have very close median values between the two groups.

Feature Correlation

The figure 5.6 illustrates the correlation of the final features selected by the five models ($M_{Clinic+Shape}$, M_{T1} , M_{T1c} , M_{T2} , and M_{FLAIR}) for predicting long surviving patients. Features selected for predicting long surviving patients T1c features `glrlm_RunLenghtNonUniformity` (F9) and `firstorder_TotalEnergy` (F10) show high correlation and again the features from T1 modality `firstorder_InterquartileRange` (F1) and `firstorder_Uniformity` (F4) are highly correlated. For the correlation is moderate to substantially low correlation. Notably, volume (F6) displayed a strong correlation with other features in the context of mutation prediction, whereas in the context of long-term survival prediction, its correlation with features F9 and F10 was only moderate.

Performance of models

The table 5.3 reports the four figures of merit (number of patients, sensitivity, specificity and balanced accuracy) for the six mono-modal models and then for two multi-modal models combining clinical and shape, radiomic (4 MR modalities) and genomic models. As mutation status is known for the three mutation types (H3.1, ACVR1 and TP53) for 57 patients a model is constructed ($M_{Genomic}$) using mutation status as input features for long term survival prediction. The genomic model ($M_{Genomic}$) achieves a balanced accuracy of 69.86% The second model using only the selected clinic and shape $M_{Clinic+Sh}$ features achieves a balanced accuracy of 72.53%. The balanced accuracies achieved by the four (M_{T1} , M_{T1c} , M_{T2} and M_{FLAIR}) radiomic models are (66.94%, 75.51%, 75%, 79.17%) respectively. The model using FLAIR radiomic features achieves the best balanced accuracy for predicting long surviving patients whereas the model based on T1 features the lowest. The ensembled multi-modal M_{Multi} prediction of the five models ($M_{ClinicSh}$, M_{T1} , M_{T1c} , M_{T2} and M_{FLAIR}) achieves the highest balanced accuracy of 89.39% with a sensitivity 100% and a specificity 78.78%. The M_{Multi} performs better than the individual models, improving the balanced accuracy of the best performing radiomic model based on FLAIR features by 10% and provides an estimation for the 74 patient included in this analysis. Introducing the genomic model in the multi-modal approach present in the last column entitled M_{MG} does not improve the sensitivity, specificity and balanced accuracy any further but does not deteriorate it.

1	0.5	0.39	0.41	0.11	0.15	0.18	0.1	0.12	0.28	0.085	0.29	0.26	0.14	0.039	0.0029	0.073	0.067	0.16	0.0039	- F5
0.5	1	0.59	0.57	0.31	0.39	0.14	0.049	0.087	0.2	0.03	0.27	0.38	0.26	0.2	0.13	0.038	0.17	0.16	0.2	- F6
0.39	0.59	1	0.98	0.38	0.28	0.19	0.041	0.26	0.21	0.012	0.2	0.11	0.24	0.2	0.13	0.068	0.011	0.27	0.27	- F9
0.41	0.57	0.98	1	0.38	0.27	0.18	0.041	0.28	0.16	0.0035	0.23	0.093	0.23	0.13	0.1	0.068	0.014	0.21	0.2	- F10
0.11	0.31	0.38	0.38	1	0.7	0.25	0.37	0.15	0.048	0.098	0.11	0.11	0.07	0.035	0.086	0.22	0.14	0.055	0.0091	- F7
0.15	0.39	0.28	0.27	0.7	1	0.14	0.29	0.014	0.077	0.032	0.0045	0.0028	0.012	0.09	0.021	0.13	0.13	0.098	0.011	- F8
0.18	0.14	0.19	0.18	0.25	0.14	1	0.37	0.15	0.24	0.28	0.088	0.041	0.53	0.3	0.22	0.41	0.072	0.084	0.12	- F14
0.1	0.049	0.041	0.041	0.37	0.29	0.37	1	0.075	0.28	0.0051	0.059	0.031	0.21	0.29	0.051	0.079	0.088	0.013	0.05	- F17
0.12	0.087	0.26	0.28	0.15	0.014	0.15	0.075	1	0.4	0.061	0.018	0.1	0.091	0.072	0.083	0.053	0.12	0.14	0.11	- F12
0.28	0.2	0.21	0.16	0.048	0.077	0.24	0.28	0.4	1	0.065	0.2	0.26	0.00024	0.25	0.12	0.004	0.032	0.31	0.34	- F19
0.085	0.03	0.012	0.0035	0.098	0.032	0.28	0.0051	0.061	0.065	1	0.28	0.08	0.31	0.14	0.1	0.11	0.021	0.056	0.085	- F3
0.29	0.27	0.2	0.23	0.11	0.0045	0.088	0.059	0.018	0.2	0.28	1	0.23	0.18	0.081	0.11	0.15	0.035	0.14	0.052	- F11
0.26	0.38	0.11	0.093	0.11	0.0028	0.041	0.031	0.1	0.26	0.08	0.23	1	0.19	0.41	0.62	0.063	0.15	0.056	0.011	- F18
0.14	0.26	0.24	0.23	0.07	0.012	0.53	0.21	0.091	0.00024	0.31	0.18	0.19	1	0.61	0.5	0.056	0.4	0.22	0.14	- F16
0.039	0.2	0.2	0.13	0.035	0.09	0.3	0.29	0.072	0.25	0.14	0.081	0.41	0.61	1	0.68	0.038	0.21	0.16	0.13	- F13
0.0029	0.13	0.13	0.1	0.086	0.021	0.22	0.051	0.083	0.12	0.1	0.11	0.62	0.5	0.68	1	0.024	0.34	0.23	0.11	- F20
0.073	0.038	0.068	0.068	0.22	0.13	0.41	0.079	0.053	0.004	0.11	0.15	0.063	0.056	0.038	0.024	1	0.26	0.34	0.26	- F15
0.067	0.17	0.011	0.014	0.14	0.13	0.072	0.088	0.12	0.032	0.021	0.035	0.15	0.4	0.21	0.34	0.26	1	0.48	0.42	- F2
0.16	0.16	0.27	0.21	0.055	0.098	0.084	0.013	0.14	0.31	0.056	0.14	0.056	0.22	0.16	0.23	0.34	0.48	1	0.85	- F1
0.0039	0.2	0.27	0.2	0.0091	0.011	0.12	0.05	0.11	0.34	0.085	0.052	0.011	0.14	0.13	0.11	0.26	0.42	0.85	1	- F4
F5	F6	F9	F10	F7	F8	F14	F17	F12	F19	F3	F11	F18	F16	F13	F20	F15	F2	F1	F4	

Figure 5.6: Pearson correlation heat-maps between the features that have been selected by the five different models to predict long surviving patients

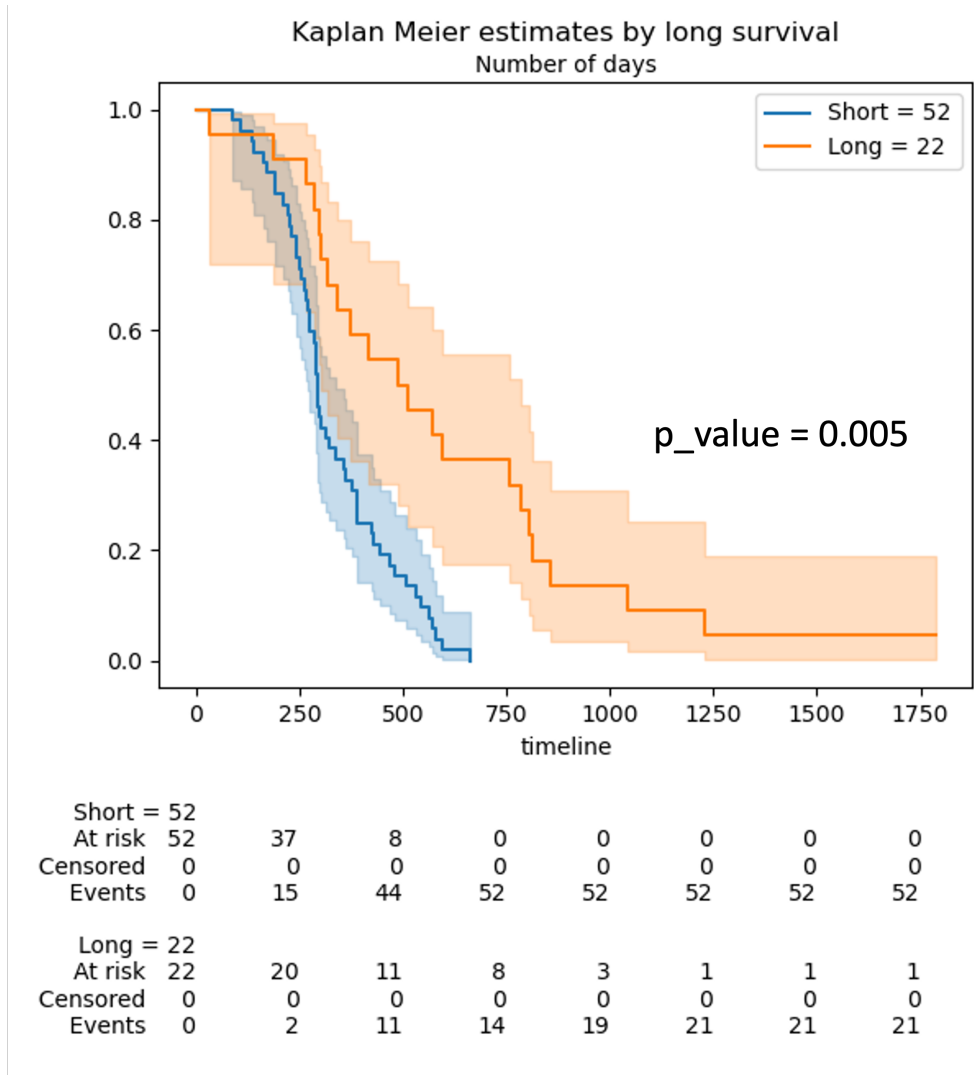


Figure 5.7: Kaplan Meier estimation of long term surviving patients.

Metric	$M_{Genomic}$	$M_{ClinicSh}$	M_{T1}	M_{T1c}	M_{T2}	M_{FLAIR}	M_{Multi}	M_{MG}
Patients	57	74	67	71	<u>72</u>	66	74	74
Sensitivity	85.71	<u>87.5</u>	66.66	83.33	<u>87.5</u>	83.33	100	100
Specificity	54	57.58	67.21	<u>67.69</u>	62.5	75	78.78	78.78
Balanced Accuracy	69.86	72.54	66.94	75.51	75	<u>79.17</u>	89.39	89.39

Table 5.3: The prediction result of long term survivors using the five mono modality models and two multi-modal models, combined: $M_{Genomic}$, $M_{ClinicSh}$, M_{T1} , M_{T1c} , M_{T2} , M_{FLAIR} , M_{Multi} and M_{MG} in a LOO-CV framework. For each prediction task and for each model, four figures of merit are reported: the total number of patients for which the long term survival was possible, the sensitivity, the specificity and the balanced accuracy. For each figure of merit, best results are in bold characters and second best results are underlined

5.4.3 Predicting long surviving patients (18 months) using the Multi-modal approach

For forecasting 18-month survival rates, a group of 74 patients was evaluated using the comprehensive multi-modal strategy. This group included seven individuals classified as censored, notably incorporating one individual censored on the 1787th day into the analysis. A systematic approach was adopted for feature selection, which involved analyzing sets of features from four different MR (Magnetic Resonance) imaging modalities, alongside clinical characteristics of the patients. The selection of pertinent features was carried out through a recursive feature elimination methodology, aimed at refining the pool of features for more accurate analysis.

The results of this feature selection process are detailed in Table 5.4. This table organizes the selected features into categories based on their origins, including mono-modal features and those that combine clinical observations with shape data, labeled as $M_{Clinic+Shape}$.

Mono-Modal	Features Name	Features Identifier
M_{T1}	T1_glcm_Energy	F1
	T1_glcm_SumAverage	F2
	T1_firstorder_Skewness	F3
	T1_glrIm_ShortRunLowGrayLevelEmphasis	F4
$M_{Clinic+Shape}$	shape_Maximum2DDiameterSlice	F5
	shape_VoxelVolume	F6
M_{T1c}	T1c_glcm_Energy	F7
	T1c_glcm_Idn	F8
	T1c_glszm_HighIntensityLargeAreaEmphasis	F9
	T1c_glszm_IntensityVariability	F10
M_{T2}	T2_firstorder_Kurtosis	F11
	T2_glszm_HighIntensityLargeAreaEmphasis	F12
M_{FLAIR}	FLAIR_firstorder_Skewness	F13
	FLAIR_glcm_ClusterShade	F14
	FLAIR_firstorder_Minimum	F15
	FLAIR_firstorder_Kurtosis	F16

Table 5.4: Subsets of features selected by the five different models $M_{Clinic+Shape}$, M_{T1} , M_{T1c} , M_{T2} , and M_{FLAIR} to predict 18 months survival. $M_{Clinic+Shape}$ contains the combined features from the model clinic and shape.

Figure 5.8 presents a correlation heatmap that visualizes the relationships among the chosen features from the five predictive models. Out of an extensive pool of 332 potential features, 16 were meticulously selected for their relevance and potential impact on the study's outcomes. The heatmap reveals predominantly low to moderate correlations among these selected features, indicating a diverse set of data points with minimal overlap in the information they convey.

A noteworthy observation from the heatmap is the highest recorded correlation coefficient of 0.84, which occurs between Feature 5 (shape_Maximum2DDiameterSlice), representing the Maximum 2D Diameter Slice (a morphological attribute), and Feature 6 (shape_VoxelVolume), denoting the Voxel Volume (an original shape characteristic). This particular correlation suggests a significant relationship between these two morphological features, possibly due to their shared basis in the physical dimensions of the observed structures. The heatmap underscores the

distinctiveness of the 16 chosen features, with their low to mild inter-correlations affirming the efficacy of the feature selection process in capturing a wide-ranging and complementary set of variables for analysis. This diversity in the selected features is crucial for constructing a robust and nuanced predictive model that leverages multiple dimensions of the data to forecast 18-month survival rates.

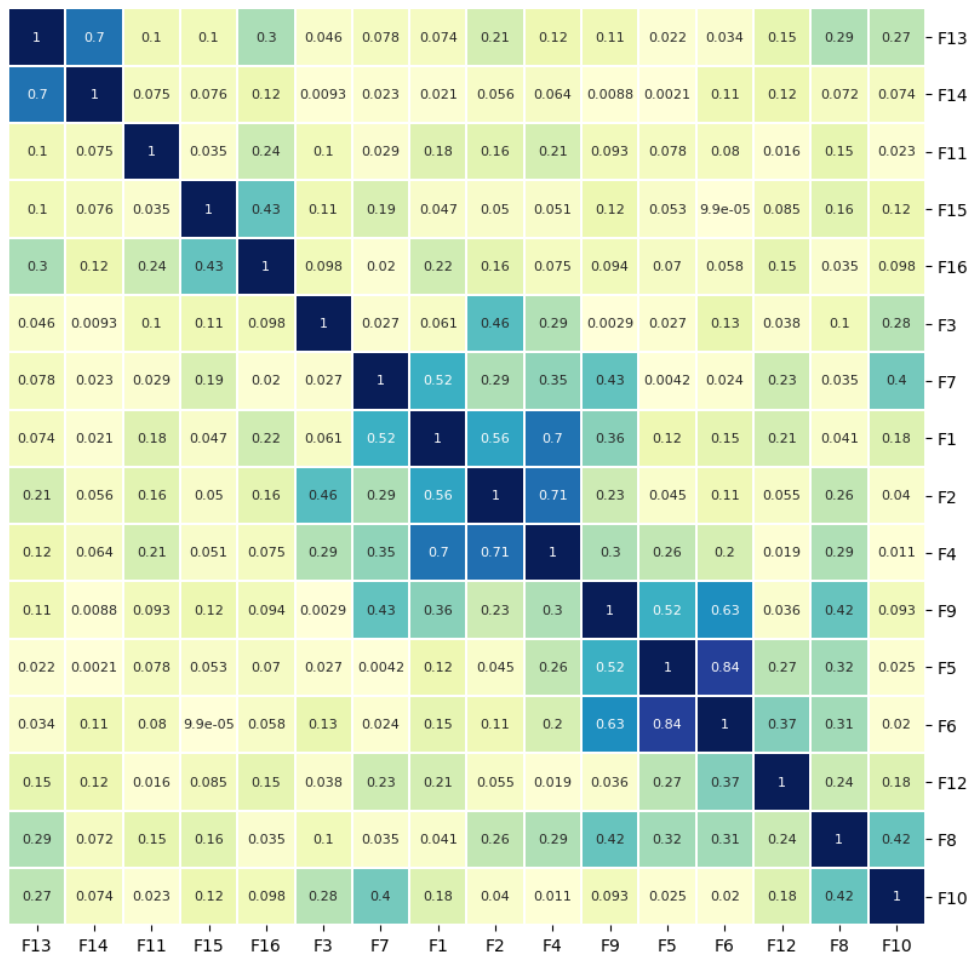


Figure 5.8: Correlation heatmap between the features that have been selected by the five different models to predict 18 months survival. Feature identifiers (on the right side) of identical features found by the different predictive tasks are shown in color.

The predictive performance of the various models deployed in this study is documented in Table 5.5. The analysis encompassed eight distinct predictive models tailored to assess survival rates. Within the subset of single-modality models, the M_{T1} model, which exclusively utilizes features derived from T1-weighted, emerged as the most effective, achieving a balanced accuracy rate of 70.51%. This model outperformed others in its category, notably surpassing the model that integrates genetic mutation status, which itself demonstrated a commendable balanced accuracy of approximately 64.44%.

Of particular note is the performance of the multi-modal model including clinical, genetic, and imaging modalities. This comprehensive approach yielded a superior balanced accuracy rate of 73.79%, underscoring the potential ben-

efits of leveraging a diverse array of data types in predictive modeling. The multi-modal model's superior accuracy highlights the value of integrating multiple data streams to enhance the predictive precision of survival outcomes, suggesting a promising direction for future research in prognostic modeling.

Metric	$M_{Genomic}$	$M_{ClinicSh}$	M_{T1}	M_{T1c}	M_{T2}	M_{FLAIR}	M_{Multi}	M_{MG}
Patients	57	74	64	68	71	63	74	74
Sensitivity	66.67	66.67	83.33	69.23	87.5	83.33	93.33	86.67
Specificity	62.22	47.46	57.69	58.18	38.09	42.11	54.23	52.54
Balanced Accuracy	64.44	57.06	<u>70.51</u>	63.71	62.80	62.72	73.79	69.60

Table 5.5: The prediction result for the Six models: $M_{ClinicSh}$, M_{T1} , M_{T1c} , M_{T2} , M_{FLAIR} and M_{Multi} in a LOO-CV framework. For each prediction task and for each model, four figures of merit are reported: the total number of patients for which the long term survival was possible, the sensitivity, the specificity and the balanced accuracy. For each figure of merit, best results are in bold characters and second best results are underlined.

5.4.4 Predicting long surviving patients (1 year) using the Multi-modal approach

To forecast 1-year survival rates, the multi-modal approach subjected to a cohort of 75 patients. Within this group, seven patients were identified as censored, with two specifically noted for their censorship status post the 365-day mark; these individuals were nevertheless included in the analysis to ensure a comprehensive evaluation. The methodology for selecting pertinent features in this context mirrored the approach previously employed for the 2-year survival prediction analysis, emphasizing consistency in the study's analytical methods.

A recursive feature elimination was employed to select a subset of predictive variables. This method systematically whittled down the feature set by iteratively removing the least significant features, ensuring that only the most impactful variables were retained for the final analysis.

The culmination of this rigorous feature selection process is detailed in Table 5.6, which enumerates the final set of features that were deemed most indicative of 1-year survival outcomes. This table serves as a crucial reference point in the study, highlighting the selected features that combine various data dimensions—ranging from imaging to clinical insights—to enhance the predictive accuracy of the survival models.

Mono-Modal	Features Name	Features Identifier
M_{T1}	T1_glrIm_ShortRunLowGrayLevelEmphasis	F1
	T1_glszm_HighIntensityLargeAreaEmphasis	F2
$M_{Clinic+Shape}$	shape_Sphericity	F3
	shape_Elongation	F4
	Age	F5
	shape_SurfaceVolumeRatio	F6
M_{T1c}	T1c_glszm_ZoneVariance	F7
	T1c_glszm_ZoneEntropy	F8
	T1c_glszm_HighIntensityLargeAreaEmphasis	F9
M_{T2}	T2_firstorder_Kurtosis	F10
	T2_firstorder_Skewness	F11
M_{FLAIR}	FLAIR_firstorder_Skewness	F12
	FLAIR_firstorder_Kurtosis	F13
	FLAIR_glcm_ClusterShade	F14
	FLAIR_firstorder_Uniformity	F15

Table 5.6: Subsets of features selected by the five different models $M_{Clinic+Shape}$, M_{T1} , M_{T1c} , M_{T2} , and M_{FLAIR} to predict 1 year survival. $M_{Clinic+Shape}$ contains the combined features from the model clinic and shape.

The Figure 5.9 illustrated the correlation heatmap between the selected features from from the five models. The 15 selected features from a total of 332 features display very low correlation. The highest correlation recorded is 0.7 between F12 (FLAIR_firstorder_skewness) and F14 (FLAIR_glcM_ClusterShade). Overall the heatmap displays very low to mild correlation between the 15 selected features.

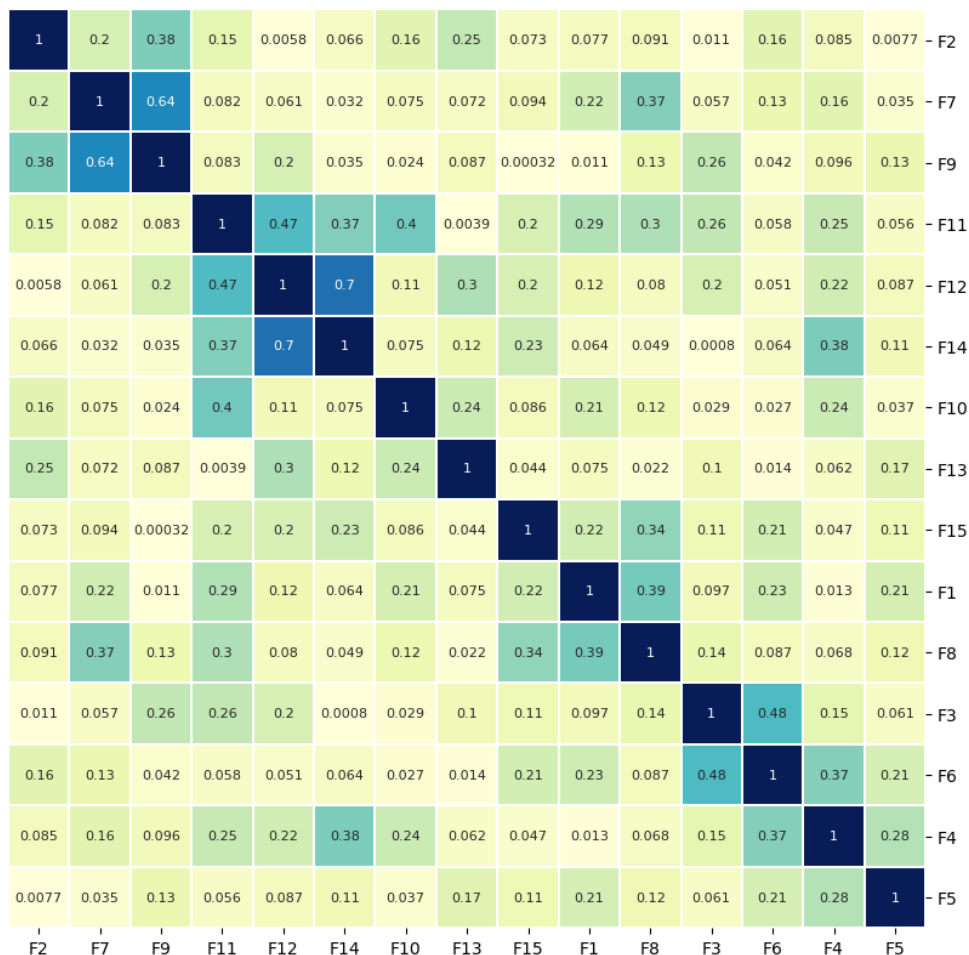


Figure 5.9: Correlation heatmap between the features that have been selected by the five different models to predict 1 year survival. Feature identifiers (on the right side) of identical features found by the different predictive tasks are shown in color.

The prognostic outcomes can be found in Table 5.7. A total of eight predictive models were assessed for this task. Among the mono-models, the model M_{FLAIR} , which leverages FLAIR features, registered the pinnacle of balanced accuracy at 78.34%. This was succeeded by the model incorporating both shape and clinical features, which achieved a balanced accuracy nearing 67%. The multi-modal model recorded a balanced accuracy of 76%. Notably, the integration of genomic data into the multi-modal model led to an enhancement of 4%, elevating the balanced accuracy to 80.84%.

Metric	$M_{Genomic}$	$M_{ClinicSh}$	M_{T1}	M_{T1c}	M_{T2}	M_{FLAIR}	M_{Multi}	M_{MG}
Patients	58	75	64	64	73	64	75	75
Sensitivity	57.1	69.70	51.85	51.85	87.1	96.15	87.87	<u>87.88</u>
Specificity	70	64.29	67.67	67.57	45.23	60.53	64.29	73.81
Balanced Accuracy	63.5	66.99	59.71	59.71	66.17	<u>78.34</u>	76.08	80.84

Table 5.7: The prediction result for the eight models: $M_{Genomic}$, $M_{ClinicSh}$, M_{T1} , M_{T1c} , M_{T2} , M_{FLAIR} , M_{Multi} and M_{MG} in a LOO-CV framework. For each prediction task and for each model, four figures of merit are reported: the total number of patients for which the long term survival of 1 year was possible, the sensitivity, the specificity and the balanced accuracy. For each figure of merit, best results are in bold characters and second best results are underlined.

5.4.5 Results for ICARE

The binary weighted model ICARE was evaluated in a leave pair out approach using the DIPG cohort. As the model is developed for handling missing data and accepts patient survival censoring, all 80 patients were used, presented in the table [A.1](#). The total number of features was equal to 336, 79 radiomic features from each of the four MR modalities, 15 shape features, 2 clinical features age and sex and 3 mutation status of the three known mutations H3, ACVR1 and TP53. The hyper parameter ' ρ ' value determines the correlation of features above which the features are dropped. C-min determines the minimum c-index achieved by each feature and max features controls the total number of features presented to each model created by the ICARE model. The table [5.8](#) below presents the parameters used to achieve the highest concordance index of 0.578 by the leave pair our approach. The final selected features are presented in Figure [5.10](#). A binary weight of +1 or -1 is assigned to selected features. From the twenty clinic, genomics and shape features, seven features are selected age, sex from clinic, H3 and TP53 mutation status and from shape features *SurfaceVolumeRatio*, *Flatness* and *Sphericity* shape features. Eight radiomic features from FLAIR images are selected, seven radiomic features from each of the three (T1 ,T1c and T2) modalities are selected. The table [5.9](#) presents the identifiers of the features selected by ICARE and the correlation between these different features are presented in the figure [5.11](#).

Hyperparameter	Value
ρ	0.65
C-min	0.50
Max Features	0.95

Table 5.8: The three hyperparameters used to achieve the highest concordance index of 0.578

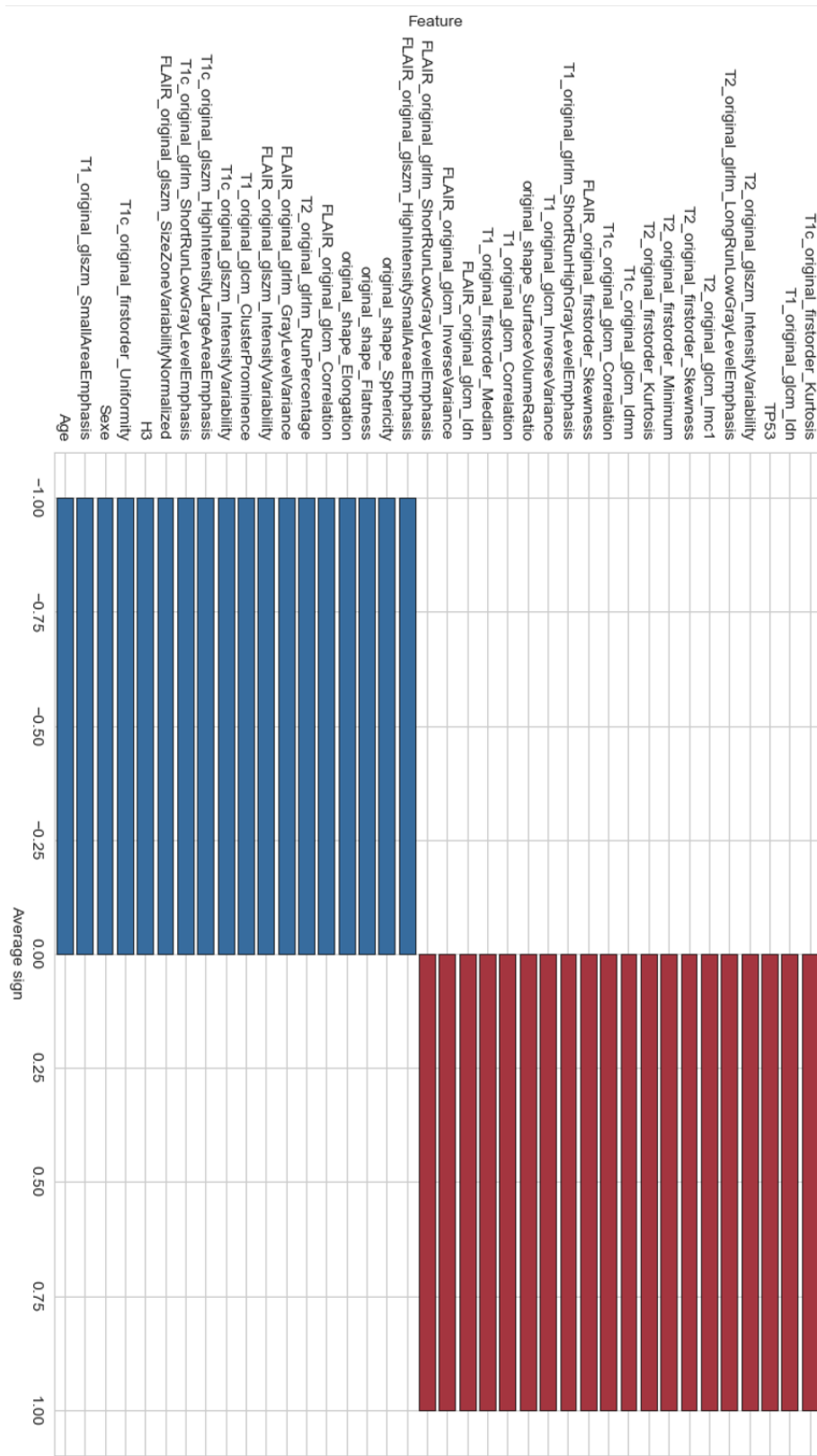
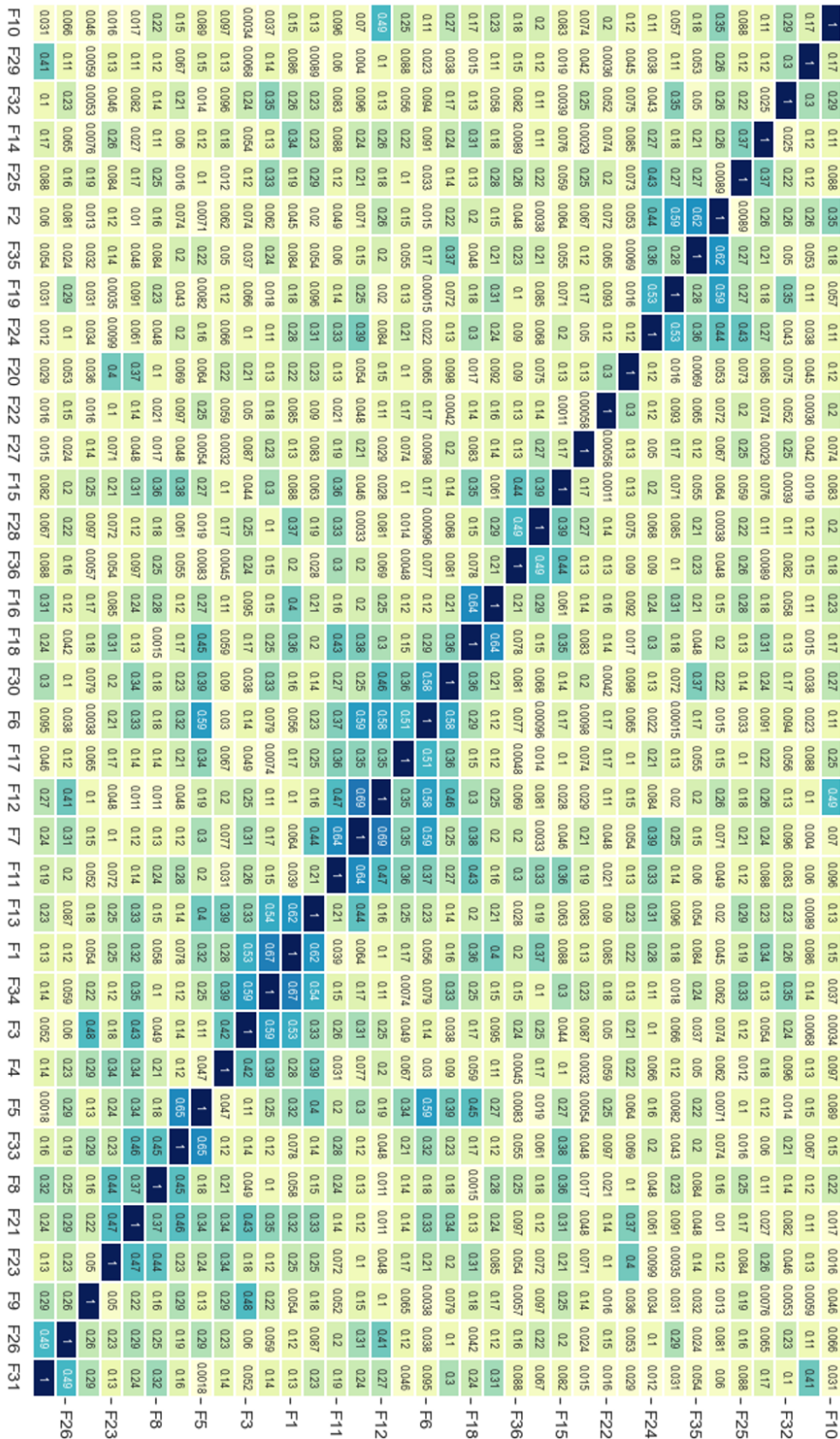


Figure 5.10: Importance of the clinical, shape and MR radiomic features. A positive value (red) shows a positive correlation with the risk and a negative value (blue) is a negative correlation.

ID	Feature Name
F1	T1c_original_glszm_IntensityVariability
F2	T1_original_glcm_Correlation
F3	original_shape_SurfaceVolumeRatio
F4	original_shape_Flatness
F5	FLAIR_original_glrlm_ShortRunLowGrayLevelEmphasis
F6	FLAIR_original_glszm_HighIntensitySmallAreaEmphasis
F7	FLAIR_original_glcm_InverseVariance
F8	T1_original_firstorder_Median
F9	original_shape_Sphericity
F10	FLAIR_original_glcm_Correlation
F11	T2_original_glrlm_RunPercentage
F12	FLAIR_original_glrlm_GrayLevelVariance
F13	FLAIR_original_glszm_IntensityVariability
F14	T1_original_glcm_ClusterProminence
F15	T1c_original_glszm_HighIntensityLargeAreaEmphasis
F16	T1c_original_glrlm_ShortRunLowGrayLevelEmphasis
F17	FLAIR_original_glszm_SizeZoneVariabilityNormalized
F18	T1c_original_firstorder_Uniformity
F19	T1_original_glszm_SmallAreaEmphasis
F20	Age
F21	TP53
F22	Sexe
F23	H3
F24	T1_original_glcm_InverseVariance
F25	T1_original_glrlm_ShortRunHighGrayLevelEmphasis
F26	FLAIR_original_firstorder_Skewness
F27	T1c_original_glcm_Correlation
F28	T1c_original_glcm_Idmn
F29	T2_original_firstorder_Kurtosis
F30	T2_original_firstorder_Minimum
F31	T2_original_firstorder_Skewness
F32	T2_original_glcm_Imc1
F33	T2_original_glrlm_LongRunLowGrayLevelEmphasis
F34	T2_original_glszm_IntensityVariability
F35	T1_original_glcm_Idn
F36	T1c_original_firstorder_Kurtosis

Table 5.9: The table presents the genomic, clinical, shape and MR radiomic features selected by ICARE binary weighted model.



5.5 Discussion

This chapter examines the survival outcome of the cohort of patients with Diffuse Intrinsic Pontine Glioma (DIPG). The interpretation using Kaplan-Meier (KM) curves to estimate the survival probabilities of patients based on different factors such as sex and the presence of specific genetic mutations, the multi-modal approach with missing modalities and the recently developed ICARE binary weighted model were tested. Similarly, KM analysis based on the presence of H3.1 and ACVR1 mutations displays curves for mutated and non-mutated patients are very close together and overlap with a p-value of 0.01 and 0.09 respectively, indicating that these mutations are not strong predictors of survival. In contrast, the curves for patients with TP53 mutations are slightly more separated, suggesting that this mutation may be a better predictor of survival. The multi-modal approach for predicting the long-term survivors with over two years of survival demonstrates strong performance. Notably, this approach selects features from all groups of features, including MR radiomics, shape radiomics, and clinical factors, which were not selected for mutation prediction in the previous chapter (Chapter 4). However, the feature selection process could be further improved, as evidenced by the high correlation observed between two T1c features, namely *RunLengthNonUniformity* and *TotalEnergy*, in the correlation plot presented in Figure 5.6. Similarly, a correlation of 0.85 is observed between the two T1 features *InterquartileRange* and *Uniformity*. Therefore, it may be beneficial to drop highly correlated features and retain only the most relevant features in the model. The study revealed that the FLAIR-based mono-modal model demonstrated the best performance, whereas the T1c and T2 mono-modals exhibited relatively similar performance. However, the multi-modal approach that combined the mono-modals resulted in the most optimal performance. The addition of the genomic mono-modal did not surpass any of the other modalities in performance, but it also did not negatively impact the results of the multi-modal approach. The KM curve in Figure 5.7 was able to separate the the two groups of patients well based on the probabilistic mean achieved by the Multi-modal. The adaptation of leave pair out cross validation for testing ICARE binary models is the closest comparison of the approach with the multi-modal approach. The ability of the binary weighted model to accept missing data without any interpolation methods is coherent with the multi-modal approach with missing data management. Both strategies select the most relevant features and perform their independent survival task. The features selected by both approaches could be similar or different features appear which are highly correlated surrogates of each other. Even though the ρ parameter for ICARE is set to 0.65, two features correlation in figure 5.11 are correlated at 0.67 and 0.69. This could be due to random dropping correlated features. The figure 5.12 illustrates distribution of overall survival of patients, showing a large peak between 9 and 11 months. This cluster of patients may explain the low values of the C-index evaluation especially in a leave pair out approach. It may seem the Multi-model performs well achieving high accuracies it is important to test both approaches independently on a data set. An analysis of the two radiomic studies by L.Tam et al [133] and Wagner et al [132] with the work done during this thesis highlight the transformative potential of MRI-based radiomics in the prognosis and prediction related to pediatric DIPG. A

common thread running through these works is the reliance on MRI as the foundational data acquisition modality and the pivotal role of radiomic features, especially those pertaining to intensity and texture. Notably, the synergistic combination of radiomic and clinical features frequently emerged as a superior predictive tool, underscoring the multifaceted nature of DIPG prognosis. A pivotal clinical factor that stood out, particularly in Wagner’s study, was age, emphasizing its significance in pediatric studies of this nature. However, while these similarities form the bedrock of DIPG research, the nuances in methodologies set each study apart. Lydia Tam’s approach leaned on LASSO Cox regression, Wagner employed a conditional survival forest model, and in this thesis we delved into the intricate realm of radiomic feature realignment across MRI scanners. Additionally, the focus on specific radiomic features and their consequential importance varied, painting a picture of a research landscape that is both rich in diversity and ripe for exploration. The variance in patient numbers among the studies—177 for Lydia Tam, 89 for Wagner, and 80 used in this thesis work highlights the challenges and opportunities in DIPG research. The Multi-models ability to handle missing data is indeed a point which makes it stand out compared to bench mark approaches. In essence, while each study offers unique insights, their collective contribution accentuates the evolving nature of DIPG prognosis research and the promising horizon of MRI-based radiomics.

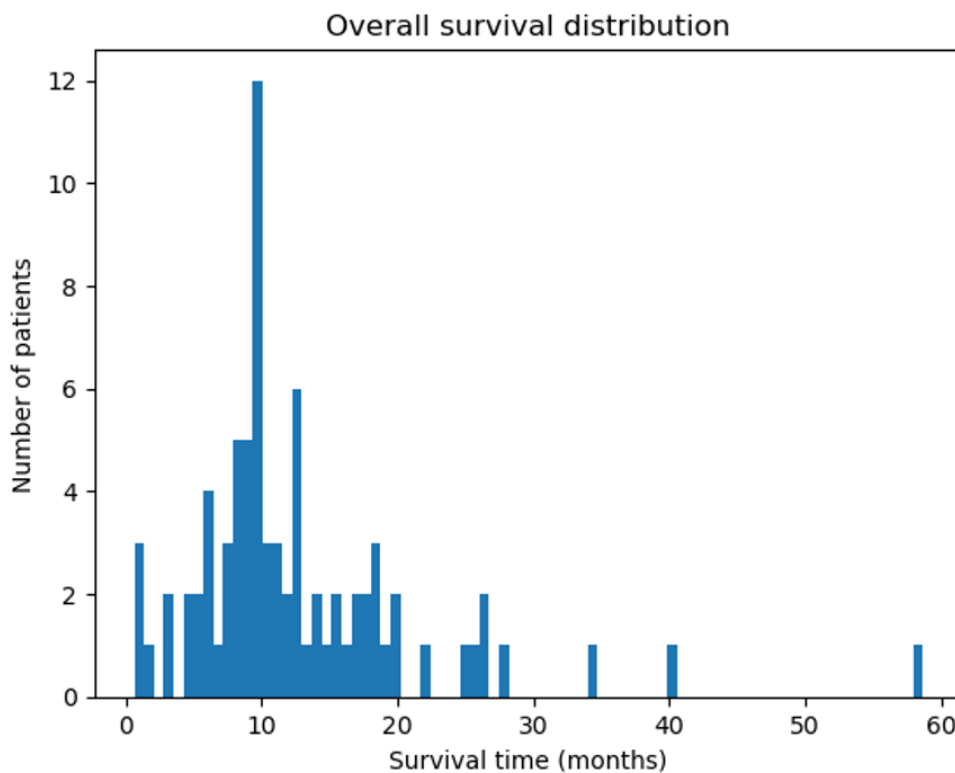


Figure 5.12: Histogram distribution displaying overall survival of patients with DIPG

5.6 Conclusion

This chapter delves deeply into the realm of survival analysis, focusing on our patient cohort with a keen eye on innovation and precision. At the heart of our discussion is the MM-LOOCV-MIM methodology, which has emerged as a beacon of hope in accurately predicting those who will surpass the critical two-year survival threshold. This milestone is of paramount importance in the clinical world, offering a beacon for navigating the complex journey of patient care. The success of the MM-LOOCV-MIM in this context is not just a testament to its robustness but also a clear indication of the untapped potential residing within radiomic features. These features go beyond mere data points; they encapsulate a wealth of insights into the genetic underpinnings of diseases, offering a window into the mutation status of patients and illuminating the path to identifying individuals with a heightened probability of extended survival.

The introduction of novel features into our analytical framework has significantly broadened our understanding, revealing the intricate layers of information that radiomic data can provide. These features, which were meticulously discussed in preceding chapters, serve as a critical bridge to essential insights concerning mutation status. They open up new avenues for predicting patient outcomes, particularly in identifying those with a promising prognosis. Amidst this backdrop, the ICARE algorithm has risen to prominence, distinguished by its adept handling of missing and censored data. This capability sets it apart from traditional models like the Cox proportional hazards model, which struggles with incomplete datasets and often relies on imputation, potentially compromising the integrity of the analysis.

The dual focal points of mutation status prediction and identification of long-term survivors form the cornerstone of our discussion. The ability to predict mutation status before any therapeutic intervention could revolutionize the treatment landscape, offering a pathway to a biopsy-free clinical workflow. This advancement is not merely procedural; it has the potential to transform patient experiences, reducing the invasiveness of diagnostic procedures and allowing for a more targeted therapeutic approach based on specific genetic mutations. In parallel, pinpointing long-term survivors is crucial for optimizing the design and implementation of clinical trials. These individuals represent a unique subset of the patient population, whose resilience and response to treatment make them ideal candidates for exploring novel therapeutic strategies.

Our investigative journey into the predictive value of mutation status for survival outcomes has shed light on its potential significance. Although our findings are preliminary, they hint at a broader narrative where mutation status could play a pivotal role in determining patient prognosis. This area of research beckons for a deeper dive, with a more extensive dataset promising to unveil richer insights.

The application of a multimodal approach to survival prediction has unequivocally demonstrated its worth. Our analyses, spanning one year, eighteen months, and two years, have not only confirmed the predictive capabilities of this approach but also highlighted an intriguing pattern: the consistency of certain features across different time

points and the variability of others. This observation raises compelling questions about the nature of the features that influence long-term survival and suggests that a more granular exploration could unlock the secrets to understanding the determinants of patient longevity.

A critical evaluation of conventional survival analysis tools, such as Kaplan-Meier curves and the log-rank test, revealed their inadequacies in discriminating between patient groups based solely on mutation status. This limitation was starkly contrasted by the significant improvement observed when Kaplan-Meier curves were constructed using the radiomic features we identified. This contrast not only underscores the limitations of traditional methodologies in capturing the nuanced dynamics of patient data but also highlights the transformative potential of radiomic features in redefining the landscape of survival analysis.

In weaving together these threads—advanced analytical techniques, the incorporation of novel features, and the strategic use of sophisticated algorithms like ICARE—this chapter paints a comprehensive picture of the current state and future direction of survival analysis. It underscores the importance of adopting a holistic, data-driven approach to understanding patient data, one that transcends conventional methodologies and embraces the complexity and richness of radiomic information. Through this lens, we aspire to advance the field of personalized medicine, ensuring that each patient's journey is informed by the most accurate, nuanced, and personalized insights available, thereby paving the way for more effective and targeted treatment strategies. This endeavor, while challenging, holds the promise of transforming patient care, ushering in an era of treatment that is not only more effective but also more attuned to the individual needs and circumstances of each patient.

Conclusion and Future work

The exploration of prognostic predictions for genetic mutations in pediatric patients afflicted with Diffuse Intrinsic Pontine Glioma (DIPG) has surged to the forefront of scientific inquiry, propelled by the rapid advancements in the field of precision medicine. The early detection of such mutations harbors the potential to revolutionize the landscape of treatment such as ONC21, paving the way for therapeutic strategies that are meticulously tailored to the individual nuances of each patient's condition. While the burgeoning body of current research, particularly studies focusing on MRI-driven radiomic signatures, demonstrates considerable promise in the realms of mutation and survival outcome prediction, it is universally acknowledged that this is an area ripe for further exploration, refinement, and innovation.

A notable challenge that was progressively refined within the scope of this thesis was the inherent variability introduced by the employment of two disparate scanning devices within our dataset. To counteract the potential discrepancies this variability might introduce, a comprehensive standardization pipeline was meticulously implemented. This was further complemented by the application of ComBat harmonization to the extracted features, ensuring a level of consistency and reliability in the data that forms the backbone of our analysis. This strategic approach stands in contrast to the methodologies employed in previous studies, such as those conducted by Lydia Tam et al. [133], who undertook standardization efforts for datasets emanating from scanners of diverse field strengths but refrained from adjusting the extracted features post-extraction. Similarly, the work of Wagner et al. [132], which focused predominantly on extracting features from FLAIR images to construct a radiomic model, is augmented by our research, which not only emphasizes the predictive capabilities inherent in individual modalities through mono models but also corroborates the enhanced performance that can be achieved through the synergistic integration of clinical and radiomic features in model development—a finding that resonates with the conclusions drawn by both Tam et al. [133] and Wagner et al. [132].

The evolution from an initial array of sixteen distinct model strategies to the refined deployment of multi-modal models has shed light on the vast untapped potential that radiomic features hold, particularly in the context of rare tumors where the absence of certain imaging modalities could potentially pose a significant challenge. The marked success of mono models in achieving precision in prediction underscores the intrinsic value of each distinct imaging modality, reinforcing the notion that each modality brings a unique perspective to the table. This methodological rigor, characterized by a meticulous harnessing of all available data while consciously eschewing data augmentation,

represents a significant forward leap in our approach. The robustness of the multi-modal strategy, particularly evident in its prowess in predicting long-term survivor outcomes, achieving optimal feature selection, and realizing predictive performance, serves as a testament to the approach's efficacy.

As we cast our gaze towards the horizon, future work beckons with the promise of further enhancements. A critical component of this future trajectory involves the expansion of feature extraction to encompass the entire tumor region, leveraging both texture and image-based features. From the work of Fontaine et al [237] it is evident dedicated contours have a greater impact on features extracted from precise tumor segmentation as compared to the implementation of automatic segmentation pipelines it would be interesting to test for DIPG patients. In order to estimate the impact of surrounding tissues on the final features selection this endeavor is anticipated to significantly refine the development of a comprehensive Computer-Aided Diagnosis (CAD) system, bolstering the model's performance with the infusion of new patient data. While the impact of segmentation on the results may not be pronounced within the confines of the current patient cohort, its importance is projected to magnify exponentially when applied to larger datasets.

The prospect of achieving a more balanced dataset, comprising data drawn from a diverse array of scanners, accentuates the critical importance of further validating the ComBat harmonization process. A harmonized dataset is poised to streamline the feature selection process, potentially unmasking relevant features that may currently be obscured by the heterogeneity that is inherently present in data derived from different scanners.

An additional dimension that merits exploration is the testing of the multi-model approach on a validation set. This crucial step is poised to provide a rigorous assessment of the approach's efficacy and the judicious selection of classifiers. It holds paramount importance for the validation of the Leave-One-Out Cross-Validation (LOOCV) framework, ensuring its robustness and applicability to a validation set. Moreover, in light of the heterogeneity characteristic of tumors, segmented tumor masks could serve as a conduit for the extraction of radiomic features at a voxel level. The generation of feature maps from these extractions could unveil regions of varying feature intensity within the tumor, offering a treasure trove of insights when these maps are correlated with clinical data or patient survival metrics.

In the domain of survival analysis, the infusion of additional data is anticipated to facilitate a more comprehensive comparison between the ICARE results and those derived from traditional models such as the Cox model. With a dataset of sufficient breadth and depth, a comparative analysis of ICARE versus Cox could yield invaluable insights into the interplay between different clinical or radiomic features and their impact on survival outcomes. This expanded dataset will not only serve to validate existing methodologies but will also pave the way for more nuanced and detailed analyses, potentially enhancing the predictive accuracy and clinical relevance of our models.

In summation, the trajectory of future research in this domain is laden with potential, holding the promise of evolving into a robust CAD system that could serve as an indispensable tool for clinicians. This system, by offering precise tumor classification, mutation prognosis, and survival predictions from the initial MR scans, could

significantly augment the clinical decision-making process. Realizing this vision will necessitate not only increased collaborative efforts and data sharing, especially given the rarity of conditions like DIPG, but also a steadfast commitment to refining and validating our models against more expansive and diverse datasets. This endeavor, while daunting, is imbued with the potential to fundamentally transform the paradigm of patient care, ushering in an era where treatment is not only more effective but also deeply personalized, reflecting the unique genetic and clinical profile of each patient.

Appendix A

Patient data

The following table presents the clinical and radiomic information available for all the 80 patients present in the cohort. Clinical features include sex, tumor volume and mutation status for the three mutation types H3K27M, ACVR1 and TP53 and the overall survival in days along with patient censoring. In case of H3K27M the sub mutations available are (H3.1K27M denoted by '1', H3.3K27M denoted by '3', H3.2K27M denoted by '2', '0' when no mutation or wild-type and 'na' when the status is not available). The age of patients is presented in years with 1 significant figure. The four MR modalities present for which ever patient are indicated by the modality name for example the patient labeled P2 has all four modalities.

Patient_ID	Sex	ACVR1	TP53	H3	Age	Volume	T1	T2	T1C	FLAIR	OS_status	OS(days)
P1	0	na	na	3	2,2	13997	T1	T2		FLAIR	1	1228
P2	1	1	0	1	3,3	15421	T1	T2	T1C	FLAIR	1	804
P3	0	0	0	0	3,3	22400		T2	T1C	FLAIR	1	469
P4	1	1	0	1	3,4	24303	T1	T2	T1C	FLAIR	1	445
P5	0	1	0	1	3,5	11513			T1C		1	303
P6	1	0	0	1	3,7	6472	T1	T2	T1C	FLAIR	1	260
P7	0	0	0	3	3,9	10144	T1		T1C		1	387
P8	0	0	0	3	4,1	17071	T1	T2	T1C	FLAIR	1	265
P9	0	1	0	1	4,1	11459	T1	T2	T1C	FLAIR	1	543
P10	0	1	0	1	4,4	13424	T1	T2	T1C	FLAIR	1	513
P11	1	0	1	0	4,4	6272	T1	T2	T1C	FLAIR	1	267
P12	1	0	1	2	4,5	12067	T1	T2	T1C	FLAIR	1	272
P13	0	1	0	1	4,6	13424	T1	T2	T1C	FLAIR	1	287
P14	0	na	na	na	4,7	3544	T1	T2	T1C	FLAIR	1	569
P15	1	1	0	1	4,7	18656	T1	T2	T1C	FLAIR	1	569
P16	0	0	0	3	4,7	8144	T1	T2	T1C	FLAIR	1	358
P17	0	na	0	na	4,8	6272	T1	T2	T1C	FLAIR	1	285
P18	1	na	na	na	4,8	14045	T1	T2		FLAIR	1	139
P19	0	0	1	3	5,1	8144	T1	T2	T1C	FLAIR	1	169
P20	0	1	0	1	5,2	12712	T1	T2	T1C	FLAIR	1	531
P21	0	1	1	1	5,2	9093	T1	T2	T1C	FLAIR	1	272
P22	0	na	na	na	5,5	17071	T1	T2	T1C	FLAIR	1	255
P23	0	0	0	3	5,5	10144			T1C	FLAIR	1	107
P24	0	na	na	na	5,6	10905	T1	T2	T1C		1	33
P25	0	1	0	1	5,7	17071	T1	T2	T1C	FLAIR	1	290
P26	1	0	0	3	5,7	18656	T1	T2	T1C	FLAIR	1	318
P27	0	0	1	3	5,7	14136	T1	T2	T1C	FLAIR	1	390
P28	0	0	1	0	5,7	8658	T1	T2	T1C	FLAIR	1	240
P29	0	0	1	1	5,8	13424	T1	T2	T1C	FLAIR	1	755
P30	1	0	1	3	5,8	24303	T1	T2	T1C	FLAIR	1	292
P31	1	0	na	3	5,9	20476	T1	T2	T1C	FLAIR	1	298
P32	1	0	0	3	5,9	10144	T1	T2	T1C	FLAIR	1	297
P33	1	na	na	na	6	17071	T1	T2	T1C	FLAIR	1	337
P34	1	na	na	na	6	15408		T2	T1C		1	321
P35	0	0	1	3	6,2	2512	T1	T2	T1C	FLAIR	1	295

Patient_ID	Sex	ACVR1	TP53	H3	Age	Volume	T1	T2	T1C	FLAIR	OS_status	OS(days)
P36	1	na	na	na	6,3	14045		T2	T1C		CENSORED	50
P37	0	0	0	3	6,5	10144	T1	T2	T1C	FLAIR	1	387
P38	0	0	0	3	6,6	10144	T1	T2	T1C	FLAIR	1	811
P39	1	0	1	3	6,6	15408	T1	T2	T1C		1	376
P40	0	na	na	3	6,7	12067	T1	T2	T1C	FLAIR	1	188
P41	1	0	0	3	6,7	15775		T2			CENSORED	1787
P42	1	0	1	3	6,9	12712	T1	T2	T1C	FLAIR	1	134
P43	0	0	1	3	7	6272	T1		T1C	FLAIR	1	161
P44	1	0	1	3	7,1	15408	T1	T2		FLAIR	1	191
P45	1	0	1	3	7,1	9714	T1	T2	T1C	FLAIR	1	243
P46	1	1	na	na	7,1	12712	T1	T2	T1C	FLAIR	CENSORED	19
P47	1	0	0	3	7,2	29985	T1	T2			CENSORED	193
P48	1	na	na	na	7,4	3544	T1	T2		FLAIR	1	340
P49	1	1	0	1	7,5	12067	T1	T2	T1C	FLAIR	1	1043
P50	0	0	1	3	7,8	5950	T1	T2	T1C	FLAIR	1	212
P51	1	na	na	na	7,9	13997	T1	T2	T1C	FLAIR	1	224
P52	0	0	1	3	7,9	8658	T1	T2	T1C	FLAIR	1	302
P53	1	0	0	1	8,6	12712	T1	T2	T1C		1	595
P54	1	0	1	3	9,1	12712	T1	T2		FLAIR	1	295
P55	1	0	1	3	9,3	2969	T1	T2	T1C	FLAIR	1	373
P56	1	0	1	3	9,3	7679	T1	T2	T1C	FLAIR	1	663
P57	1	0	1	3	9,4	9168	T1	T2	T1C	FLAIR	1	231
P58	1	0	1	3	9,6	14045	T1	T2	T1C	FLAIR	1	288
P59	0	0	1	3	9,7	2332	T1	T2	T1C	FLAIR	1	362
P60	1	na	na	na	9,7	8144	T1	T2	T1C		1	270
P61	1	1	na	na	10	7123	T1	T2		FLAIR	CENSORED	27
P62	1	0	1	3	10	4139	T1	T2	T1C	FLAIR	1	594
P63	1	0	1	3	10,2	17071	T1	T2	T1C	FLAIR	1	314
P64	1	0	1	3	10	6272	T1	T2	T1C	FLAIR	CENSORED	329
P65	0	0	1	3	10,6	5950	T1	T2	T1C	FLAIR	1	249
P66	0	na	na	na	11	1403	T1		T1C	FLAIR	1	508
P67	0	0	1	3	11,4	11459	T1	T2	T1C	FLAIR	1	478
P68	0	0	1	3	11,5	3872	T1	T2	T1C	FLAIR	1	86
P69	1	0	1	3	11,8	8658	T1	T2	T1C	FLAIR	CENSORED	390
P70	0	0	1	3	12,8	11459		T2	T1C		1	786
P71	1	0	1	3	14	4166	T1	T2	T1C	FLAIR	1	291
P72	1	na	na	na	14,2	8144	T1	T2	T1C		1	561
P73	1	1	0	0	14,2	5184	T1	T2	T1C	FLAIR	1	858
P74	1	0	1	3	14,7	11510	T1	T2	T1C	FLAIR	1	424
P75	0	0	0	3	15,2	11510	T1	T2	T1C	FLAIR	1	580
P76	1	0	1	3	15,4	15408	T1	T2	T1C	FLAIR	1	189
P77	1	0	0	3	16,2	11459	T1	T2	T1C	FLAIR	1	488
P78	1	na	na	na	16,5	18656	T1	T2	T1C		1	415
P79	1	0	1	3	19,1	3544	T1	T2	T1C	FLAIR	1	428
P80	1	na	na	na	30,2	6203	T1		T1C	FLAIR	1	221

Table A.1: The table presents in detail patients age, sex, volume and genomics mutation for H3, ACVR1 and TP53 along with the modalities available and the OSstatus. **Sex:**1 boy, 0 girl, mutation if present is represented by '1' if not present represented by '0' and 'na' if the mutation status is not known to be present or absent. **Age:** is given in years,

Appendix B

Radiomic Features

Here we give an overview of the radiomic features used in the thesis. Table [B.1](#) gives a full list of the features used, and full mathematical formulations can be found in the PyRadiomics documentation [\[238\]](#). These are mostly based on IBSI feature definitions [\[239\]](#). Unless otherwise stated we used the default PyRadiomic settings for feature calculation.

Feature Type	Features
Shape (2D)	Mesh Surface, Pixel Surface, Perimeter, Perimeter to Surface Ratio Sphericity, Spherical Disproportion, Maximum 2D Diameter, Major Axis Length, Minor Axis Length, Elongation.
Shape (3D)	Mesh Volume, Voxel Volume, Surface Area, Surface Area to Volume Ratio, Sphericity, Compactness1, Compactness2, Spherical Disproportion, Maximum 3D Diameter, Maximum 2D Diameter (by slice, column, row), Major Axis Length, Minor Axis Length, Least Axis Length, Elongation, Flatness.
First-Order	Energy, Total Energy, Entropy, Minimum, 10th Percentile, 90th Percentile, Maximum, Mean, Median, Interquartile Range, Mean Absolute Deviation, Robust Mean Absolute Deviation, Root Mean Squared, Standard Deviation, Skewness, Kurtosis, Variance, Uniformity.
GLCM	Autocorrelation, Joint Average, Cluster Prominence, Cluster Shade Cluster Tendency, Contrast, Correlation, Difference Average, Difference Entropy, Difference Variance, Joint Energy, Joint Entropy, Informational Measure of Correlation 1, Information Measure of Correlation 2, Inverse Difference Moment, Maximal Correlation Coefficient, Inverse Difference Moment Normalised, Inverse Difference, Inverse Difference Normalised, Inverse Variance, Maximum Probability, Sum Average, Sum Entropy, Sum of Squares.
NGTDM	Coarseness, Contrast, Busyness, Complexity, Strength
GLRLM	Short Run Emphasis, Long Run Emphasis, Grey Level Non-Uniformity, Grey Level Non-Uniformity Normalised, Run Length Non-Uniformity, Run Length Non-Uniformity Normalised, Run Percentage, Grey Level Variance, Run Variance, Run Entropy, Low Grey Level Run Emphasis, High Grey Level Run Emphasis, Short Run Low Grey Level Run Emphasis, Short Run High Grey Level Emphasis, Long Run Low Grey Level Emphasis, Long Run High Grey Level Emphasis
GLSZM	Small Area Emphasis, Large Area Emphasis, Grey Level Non-Uniformity, Grey Level Non-Uniformity Normalised, Size-Zone Non-Uniformity, Size Zone Non-Uniformity Normalised, Zone Percentage, Grey Level Variance, Zone Variance, Zone Entropy, Low Grey Level Zone Emphasis, High Grey Level Zone Emphasis, Small Area Low Grey Level Emphasis, Small Area High Grey Level Emphasis, Large Area Low Grey Level Emphasis, Large Area High Grey Level Emphasis
GLDM	Small Dependence Emphasis, Large Dependence Emphasis, Grey Level Non-Uniformity, Dependence Non-Uniformity, Dependence Non-Uniformity Normalised, Grey Level Variance, Dependence Variance, Dependence Entropy, Low Grey Level Emphasis, High Grey Level Emphasis, Small Dependence Low Grey Level Emphasis, Small Dependence High Grey Level Emphasis, Large Dependence Low Grey Level Emphasis, Large Dependence High Grey Level Emphasis

Table B.1: Texture features by feature type Full mathematical details of the features can be found at [\[238\]](#)

Glossary

ACVR1 Activin A Receptor Type I. [20](#)

ADC Apparent Diffusion Coefficient. [24](#)

ASL Arterial Spin Labeling. [23](#)

CBF cerebral blood flow. [23](#)

CBV cerebral blood volume. [23](#)

cIMPACT-NOW Consortium to Inform Molecular and Practical Approaches to CNS Tumor Taxonomy . [17](#)

CNN Convolutional Neural Network. [37](#)

CNS central nervous system. [15](#)

CT Computed Tomography . [16](#), [21](#)

DICE Sørensen–Dice coefficient. [44](#)

DICOM Digital Imaging and Communications in Medicine. [27](#)

DIPG Diffuse Intrinsic Pontine Glioma. [15](#)

DMG Diffuse Midline glioma. [15](#)

DOPA 18F-dihydroxyphenylalanine. [22](#)

DTI Diffusion Tensor Imaging. [22](#)

DWI Diffusion-Weighted Imaging. [22](#)

FDG-PET Fluorodeoxyglucose Positron Emission Tomography. [55](#)

FET 18F-fluoro-ethyl-tyrosine. [22](#)

FLT 18fluorothymidine. [22](#)

fMRI Functional Magnetic Resonance Imaging. [24](#)

FN False Negative. [41](#)

FP False Positive. [41](#)

GAN Generative Adversarial Networks. [53](#)

GLCM Gray Level Co-occurrence Matrix. [34](#)

GLDM Gray Level Dependence Matrix. [34](#)

GLRLM Gray Level Run Length Matrix. [34](#)

GLSZM Gray Level Size Zone. [34](#)

GM Grey Matter. [5](#) [32](#)

H3 Histone H3. [18](#)

H3K27M Histone H3 Lysine 27 to Methionine mutation. [18](#)

HD Hellinger distance. [5](#) [32](#)

HECKTOR Head And Neck Tumor segmentation and prediction in PET/CT Images . [29](#)

HWS Hybrid White Stripe. [32](#)

ICARE The Medical Image Computing Assisted Intervention Society. [29](#)

IDH Isocitrate Dehydrogenase. [18](#)

IRB Institutional Review Board. [27](#)

KM Kaplan-Meier. [48](#)

KNN K-Nearest Neighbour. [53](#)

LASSO least absolute shrinkage and selection operator . [106](#)

LOOCV Leave One Out Cross Validation . [41](#)

LPO leave-pair-out . [108](#)

LR Logistic Regression. [36](#)

LR-RFECV Logistic Regression Recursive Feature Elimination with cross validation. [37](#)

MI myoinositol. [26](#)

MICCAI The Medical Image Computing Assisted Intervention Society. [29](#)

ML Machine Learning. [37](#)

MM-LOOCV-MIM multi-modal and missing imaging modality. [92](#)

MRI Magnetic Resonance Imaging. [16](#), [21](#)

MRS Magnetic Resonance Spectroscopy. [22](#)

MSE Mean Square Error. [43](#)

MTT mean transit time. [23](#)

N4 Non-Parametric Non-Uniform Intensity Normalization. [30](#)

NGTDM Neighbouring Grey Tone Difference Matrix. [34](#)

NIfTI Neuroimaging Informatics Technology Initiative. [27](#)

OS Overall Survival. [26](#)

PET Positron Emission Tomography . [21](#)

PFS progression-free survival. [102](#)

PIK3CA Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha. [20](#)

PLNTY Polymorphous low-grade neuroepithelial tumor of the young. [18](#)

PWI Perfusion-Weighted Imaging. [22](#), [23](#)

RF Random Forest. [36](#)

ROI Region of Interest. [37](#)

SGD Stochastic Gradient Descent. [42](#)

SMOTE Synthetic Data Augmentation for Tabular Data. [44](#)

SVM Support Vector Machine. [36](#)

SWI Susceptibility Sequence. [24](#)

T1w T1 weighted. [22](#)

T2w T2 weighted. [22](#)

TE Time-Echo. [26](#)

TN True Negative. [41](#)

TP True Positive . [41](#)

TP53 Tumor Protein p53. [20](#)

WHO World Health Organization. [15](#), [17](#)

WM White Matter. [32](#)

WT Wild Type. [27](#)

Bibliography

- [1] S. Reuzé, A. Schernberg, F. Orlhac, R. Sun, C. Chargari, L. Dercle, E. Deutsch, I. Buvat, and C. Robert, “Radiomics in nuclear medicine applied to radiation therapy: methods, pitfalls, and challenges,” *International Journal of Radiation Oncology* Biology* Physics*, vol. 102, no. 4, pp. 1117–1142, 2018.
- [2] F. Khalid, J. Goya-Outi, V. Frouin, N. Boddaert, J. Grill, and F. Frouin, “Impact of combat and a multi-model approach to deal with multi-scanner and missing mri data in a small cohort study. application to h3k27m mutation prediction in patients with dipg,” in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 3809–3812, IEEE, 2021.
- [3] F. Khalid, J. Goya-Outi, T. Escobar, V. Dangouloff-Ros, A. Grigis, C. Philippe, N. Boddaert, J. Grill, V. Frouin, and F. Frouin, “Multimodal mri radiomic models to predict genomic mutations in diffuse intrinsic pontine glioma with missing imaging modalities,” *Frontiers in Medicine*, vol. 10, p. 1071447, 2023.
- [4] P. Kleihues, D. N. Louis, B. W. Scheithauer, L. B. Rorke, G. Reifenberger, P. C. Burger, and W. K. Cavenee, “The WHO classification of tumors of the nervous system,” *J. Neuropathol. Exp. Neurol.*, vol. 61, pp. 215–25; discussion 226–9, Mar. 2002.
- [5] D. N. Louis, A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-Branger, C. Hawkins, H. Ng, S. M. Pfister, G. Reifenberger, *et al.*, “The 2021 who classification of tumors of the central nervous system: a summary,” *Neuro-oncology*, vol. 23, no. 8, pp. 1231–1251, 2021.
- [6] R. M. Young, A. Jamshidi, G. Davis, and J. H. Sherman, “Current trends in the surgical management and treatment of adult glioblastoma,” *Annals of translational medicine*, vol. 3, no. 9, 2015.
- [7] “The genomic landscape of diffuse intrinsic pontine glioma and pediatric non-brainstem high-grade glioma,” *Nature genetics*, vol. 46, no. 5, pp. 444–450, 2014.
- [8] K. R. Taylor, A. Mackay, N. Truffaux, Y. S. Butterfield, O. Morozova, C. Philippe, D. Castel, C. S. Grasso, M. Vinci, D. Carvalho, *et al.*, “Recurrent activating acvr1 mutations in diffuse intrinsic pontine glioma,” *Nature genetics*, vol. 46, no. 5, pp. 457–461, 2014.

- [9] M. Zarghooni, U. Bartels, E. Lee, P. Buczkowicz, A. Morrison, A. Huang, E. Bouffet, and C. Hawkins, "Whole-genome profiling of pediatric diffuse intrinsic pontine gliomas highlights platelet-derived growth factor receptor alpha and poly (adp-ribose) polymerase as potential therapeutic targets," *J clin oncol*, vol. 28, no. 8, pp. 1337–1344, 2010.
- [10] J. Schwartzentruber, A. Korshunov, X.-Y. Liu, D. T. Jones, E. Pfaff, K. Jacob, D. Sturm, A. M. Fontebasso, D.-A. K. Quang, M. Tönjes, *et al.*, "Driver mutations in histone h3. 3 and chromatin remodelling genes in paediatric glioblastoma," *Nature*, vol. 482, no. 7384, pp. 226–231, 2012.
- [11] M. Karremann, G. H. Gielen, M. Hoffmann, M. Wiese, N. Colditz, M. Warmuth-Metz, B. Bison, A. Claviez, D. G. van Vuurden, A. O. von Bueren, *et al.*, "Diffuse high-grade gliomas with h3 k27m mutations carry a dismal prognosis independent of tumor location," *Neuro-oncology*, vol. 20, no. 1, pp. 123–131, 2018.
- [12] "Somatic histone h3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas," *Nature genetics*, vol. 44, no. 3, pp. 251–253, 2012.
- [13] L. Bjerke, A. Mackay, M. Nandhabalan, A. Burford, A. Jury, S. Popov, D. A. Bax, D. Carvalho, K. R. Taylor, M. Vinci, *et al.*, "Histone h3. 3 mutations drive pediatric glioblastoma through upregulation of mycn," *Cancer discovery*, vol. 3, no. 5, pp. 512–519, 2013.
- [14] D. E. Reuss, Y. Mamatjan, D. Schrimpf, D. Capper, V. Hovestadt, A. Kratz, F. Sahm, C. Koelsche, A. Korshunov, A. Olar, *et al.*, "Idh mutant diffuse and anaplastic astrocytomas have similar age at presentation and little difference in survival: a grading problem for who," *Acta neuropathologica*, vol. 129, pp. 867–873, 2015.
- [15] H. Yan, D. W. Parsons, G. Jin, R. McLendon, B. A. Rasheed, W. Yuan, I. Kos, I. Batinic-Haberle, S. Jones, G. J. Riggins, *et al.*, "Idh1 and idh2 mutations in gliomas," *New England journal of medicine*, vol. 360, no. 8, pp. 765–773, 2009.
- [16] M. Sanson, Y. Marie, S. Paris, A. Idbah, J. Laffaire, F. Ducray, S. El Hallani, B. Boisselier, K. Mokhtari, K. Hoang-Xuan, *et al.*, "Isocitrate dehydrogenase 1 codon 132 mutation is an important prognostic biomarker in gliomas," *J Clin Oncol*, vol. 27, no. 25, pp. 4150–4154, 2009.
- [17] B. S. Paugh, C. Qu, C. Jones, Z. Liu, M. Adamowicz-Brice, J. Zhang, D. A. Bax, B. Coyle, J. Barrow, D. Hargrave, *et al.*, "Integrated molecular genetic profiling of pediatric high-grade gliomas reveals key differences with the adult disease," *Journal of clinical oncology*, vol. 28, no. 18, p. 3061, 2010.
- [18] P. Buczkowicz, C. Hoeman, P. Rakopoulos, S. Pajovic, L. Letourneau, M. Dzamba, A. Morrison, P. Lewis, E. Bouffet, U. Bartels, *et al.*, "Genomic analysis of diffuse intrinsic pontine gliomas identifies three molecular subgroups and recurrent activating acvr1 mutations," *Nature genetics*, vol. 46, no. 5, pp. 451–456, 2014.

- [19] N. Galldiks and K.-J. Langen, "Amino acid pet—an imaging option to identify treatment response, posttherapeutic effects, and tumor recurrence?," *Frontiers in neurology*, vol. 7, p. 120, 2016.
- [20] Y. Terakawa, N. Tsuyuguchi, Y. Iwai, K. Yamanaka, S. Higashiyama, T. Takami, and K. Ohata, "Diagnostic accuracy of 11c-methionine pet for differentiation of recurrent brain tumors from radiation necrosis after radiotherapy," *Journal of Nuclear Medicine*, vol. 49, no. 5, pp. 694–699, 2008.
- [21] F. Cicone, G. Minniti, A. Romano, A. Papa, C. Scaringi, F. Tavanti, A. Bozzao, R. Maurizi Enrici, and F. Scopinaro, "Accuracy of f-dopa pet and perfusion-mri for differentiating radionecrotic from progressive brain metastases after radiosurgery," *European journal of nuclear medicine and molecular imaging*, vol. 42, pp. 103–111, 2015.
- [22] S. Avula, A. Peet, G. Morana, P. Morgan, M. Warmuth-Metz, T. Jaspan, and European Society for Paediatric Oncology (SIOPE)-Brain Tumour Imaging Group, "European Society for Paediatric Oncology (SIOPE) MRI guidelines for imaging patients with central nervous system tumours," *Childs Nerv Syst*, vol. 37, pp. 2497–2508, Aug. 2021.
- [23] E. M. Haacke, R. W. Brown, M. R. Thompson, and R. Venkatesan, *Magnetic resonance imaging: Physical principles and sequence design*. John Wiley & Sons, 1999.
- [24] C. Nicholson and E. Syková, "Extracellular space structure revealed by diffusion analysis," *Trends in Neurosciences*, vol. 24, no. 5, pp. 278–283, 2001.
- [25] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince, *MRI from picture to proton*. Cambridge University Press, 2017.
- [26] K. Zeleňák, C. Viera, and P. Hubert, "Radiology imaging techniques of brain tumours," *Clinical Management and Evolving Novel Therapeutic Strategies for Patients with Brain Tumors*, vol. 77, 2013.
- [27] M. C. Mabray, R. F. Barajas, and S. Cha, "Modern brain tumor imaging," *Brain tumor research and treatment*, vol. 3, no. 1, pp. 8–23, 2015.
- [28] R. N. Al-Okaili, J. Krejza, S. Wang, J. H. Woo, and E. R. Melhem, "Advanced mr imaging techniques in the diagnosis of intraaxial brain tumors in adults," *Radiographics*, vol. 26, no. suppl_1, pp. S173–S189, 2006.
- [29] F. G. Dhermain, P. Hau, H. Lanfermann, A. H. Jacobs, and M. J. van den Bent, "Advanced mri and pet imaging for assessment of treatment response in patients with gliomas," *The Lancet Neurology*, vol. 9, no. 9, pp. 906–920, 2010.

- [30] L. Østergaard, R. M. Weisskoff, D. A. Chesler, C. Gyldensted, and B. R. Rosen, "High resolution measurement of cerebral blood flow using intravascular tracer bolus passages. part i: Mathematical approach and statistical analysis," *Magnetic resonance in medicine*, vol. 36, no. 5, pp. 715–725, 1996.
- [31] C. Yue, C. I. Tsien, V. Nagesh, L. Junck, R. t. Haken, B. D. Ross, T. L. Chenevert, and T. S. Lawrence, "Clinical investigation survival prediction in high-grade gliomas by mri perfusion before and during early stage of rt," *International Journal of Radiation Oncology, Biology and Physics*, vol. 64, no. 3, 2006.
- [32] R. Jain, J. Gutierrez, J. Narang, L. Scarpace, L. Schultz, N. Lemke, S. Patel, T. Mikkelsen, and J. Rock, "In vivo correlation of tumor blood volume and permeability with histologic and molecular angiogenic markers in gliomas," *American Journal of Neuroradiology*, vol. 32, no. 2, pp. 388–394, 2011.
- [33] D. C. Alsop, J. A. Detre, X. Golay, M. Günther, J. Hendrikse, L. Hernandez-Garcia, *et al.*, "Recommended implementation of arterial spin-labeled perfusion mri for clinical applications: A consensus of the ismrm perfusion study group and the european consortium for asl in dementia," *Magnetic Resonance in Medicine*, vol. 73, no. 1, pp. 102–116, 2015.
- [34] J. A. Detre, H. Rao, D. J. Wang, Y. F. Chen, and Z. Wang, "Applications of arterial spin labeled mri in the brain," *Journal of Magnetic Resonance Imaging*, vol. 35, no. 5, pp. 1026–1037, 2012.
- [35] J. Wang, D. C. Alsop, H. K. Song, J. A. Maldjian, K. Tang, A. E. Salvucci, and J. A. Detre, "Arterial transit time imaging with flow encoding arterial spin tagging (feast)," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 50, no. 3, pp. 599–607, 2003.
- [36] R. Bammer, "Basic principles of diffusion-weighted imaging," *European journal of radiology*, vol. 45, no. 3, pp. 169–184, 2003.
- [37] A. H. Jacobs, L. W. Kracht, A. Gossman, M. A. Rürger, A. V. Thomas, A. Thiel, and K. Herholz, "Imaging in neurooncology," *NeuroRx*, vol. 2, no. 2, pp. 333–347, 2005.
- [38] J. M. Provenzale, P. McGraw, P. Mhatre, A. C. Guo, and D. DeLong, "Peritumoral brain regions in gliomas and meningiomas: investigation with isotropic diffusion-weighted mr imaging and diffusion-tensor mr imaging," *Radiology*, vol. 232, no. 2, pp. 451–460, 2004.
- [39] S.-C. Chang, P.-H. Lai, W.-L. Chen, H.-H. Weng, J.-T. Ho, J.-S. Wang, C.-Y. Chang, H.-B. Pan, and C.-F. Yang, "Diffusion-weighted mri features of brain abscess and cystic or necrotic brain tumors: comparison with conventional mri," *Clinical imaging*, vol. 26, no. 4, pp. 227–236, 2002.
- [40] C.-C. Wu, R. Jain, A. Radmanesh, L. M. Poisson, W.-Y. Guo, D. Zagzag, M. Snuderl, D. Placantonakis, J. Golfinos, and A. Chi, "Predicting genotype and survival in glioma using standard clinical mr imaging apparent dif-

fusion coefficient images: a pilot study from the cancer genome atlas,” *American Journal of Neuroradiology*, vol. 39, no. 10, pp. 1814–1820, 2018.

- [41] D. K. Jones, T. R. Knösche, and R. Turner, “White matter integrity, fiber count, and other fallacies: the do’s and don’ts of diffusion mri,” *Neuroimage*, vol. 73, pp. 239–254, 2013.
- [42] J.-D. Tournier, R. Smith, D. Raffelt, R. Tabbara, T. Dhollander, M. Pietsch, D. Christiaens, B. Jeurissen, C.-H. Yeh, and A. Connelly, “Mrtrix3: A fast, flexible and open software framework for medical image processing and visualisation,” *Neuroimage*, vol. 202, p. 116137, 2019.
- [43] S. Van Cauter, J. Veraart, J. Sijbers, R. R. Peeters, U. Himmelreich, F. De Keyzer, S. W. Van Gool, F. Van Calenbergh, S. De Vleeschouwer, W. Van Hecke, *et al.*, “Gliomas: diffusion kurtosis mr imaging in grading,” *Radiology*, vol. 263, no. 2, pp. 492–501, 2012.
- [44] M. Jolapara, C. Kesavadas, V. Radhakrishnan, B. Thomas, A. Gupta, N. Bodhey, S. Patro, J. Saini, U. George, and P. Sarma, “Role of diffusion tensor imaging in differentiating subtypes of meningiomas,” *Journal of Neuroradiology*, vol. 37, no. 5, pp. 277–283, 2010.
- [45] F. Sanvito, A. Castellano, and A. Falini, “Advancements in neuroimaging to unravel biological and molecular features of brain tumors,” *Cancers*, vol. 13, no. 3, p. 424, 2021.
- [46] B. Ross and S. Bluml, “Magnetic resonance spectroscopy of the human brain,” *The Anatomical Record: An Official Publication of the American Association of Anatomists*, vol. 265, no. 2, pp. 54–84, 2001.
- [47] D. Soares and M. Law, “Magnetic resonance spectroscopy of the brain: review of metabolites and clinical applications,” *Clinical radiology*, vol. 64, no. 1, pp. 12–21, 2009.
- [48] A. Horská and P. B. Barker, “Imaging of brain tumors: Mr spectroscopy and metabolic imaging,” *Neuroimaging Clinics*, vol. 20, no. 3, pp. 293–310, 2010.
- [49] S. Saraswathy, F. W. Crawford, K. R. Lamborn, A. Pirzkall, S. Chang, S. Cha, and S. J. Nelson, “Evaluation of mr markers that predict survival in patients with newly diagnosed gbm prior to adjuvant therapy,” *Journal of neuro-oncology*, vol. 91, pp. 69–81, 2009.
- [50] E. Kousi, I. Tsougos, and K. Eftychia, “Proton magnetic resonance spectroscopy of the central nervous system,” *Novel Frontiers of Advanced Neuroimaging. InTech*, pp. 19–50, 2013.
- [51] J. Grill, G. Le Teuff, P. Varlet, D. R. Hargrave, K. Nysom, K. Blomgrenn, G. B. McCowage, F. Bautista, D. Van Vuurden, M.-A. Debily, T. Kergrohen, S. Puget, S. Bolle, S. Abbou, P. Leblond, N. Boddaert, G. Vassal, and M.-C. Le Deley, “Biological medicines for diffuse intrinsic pontine glioma (dipg) eradication (biomede):

Final results of an international randomized phase ii platform trial comparing 3 targeted therapies in combination with radiotherapy from itcc, siope-brain and anzchog,” *Journal of Clinical Oncology*, vol. 41, no. 16_suppl, pp. 10003–10003, 2023.

- [52] J. Grill, G. Le Teuff, K. Nysom, K. Blomgren, D. Hargrave, G. McCowage, F. Bautista, D. van Vuurden, V. Dangouloff-Ros, S. Puget, P. Varlet, M. Debily, G. Vassal, and M. Le Deley, “Pdct-01. biological medicine for diffuse intrinsic pontine gliomas eradication (biomede): Results of the three-arm biomarker-driven randomized trial in the first 230 patients from europe and australia,” *Neuro Oncol*, vol. 21, no. Suppl 6, p. vi183, 2019. Epub 2019 Nov 11.
- [53] “Hecktor - a challenge on head and neck tumor segmentation in pet-ct.” <https://hecktor.grand-challenge.org/>. Accessed: yyyy-mm-dd.
- [54] A. Badhwar, Y. Collin-Verreault, P. Orban, S. Urchs, I. Chouinard, J. Vogel, O. Potvin, S. Duchesne, and P. Bellec, “Multivariate consistency of resting-state fmri connectivity maps acquired on a single individual over 2.5 years, 13 sites and 3 vendors,” *NeuroImage*, vol. 205, p. 116210, 2020.
- [55] X. Han, J. Jovicich, D. Salat, A. van der Kouwe, B. Quinn, S. Czanner, E. Busa, J. Pacheco, M. Albert, R. Killiany, *et al.*, “Reliability of mri-derived measurements of human cerebral cortical thickness: the effects of field strength, scanner upgrade and manufacturer,” *Neuroimage*, vol. 32, no. 1, pp. 180–194, 2006.
- [56] A. Zwanenburg, S. Leger, M. Vallières, S. Löck, *et al.*, “Image biomarker standardisation initiative-feature definitions,” *arXiv preprint arXiv:1612.07003*, vol. 10, 2016.
- [57] L. Byrge, D. Kliemann, Y. He, H. Cheng, J. M. Tyszka, R. Adolphs, and D. P. Kennedy, “Video-evoked fmri bold responses are highly consistent across different data acquisition sites,” *Human brain mapping*, vol. 43, no. 9, pp. 2972–2991, 2022.
- [58] L. Y. Cai, Q. Yang, P. Kanakaraj, V. Nath, A. T. Newton, H. A. Edmonson, J. Luci, B. N. Conrad, G. R. Price, C. B. Hansen, *et al.*, “Masivar: multisite, multiscanner, and multisubject acquisitions for studying variability in diffusion weighted magnetic resonance imaging,” *bioRxiv*, pp. 2020–12, 2020.
- [59] L. G. Nyúl, J. K. Udupa, and X. Zhang, “New variants of a method of mri scale standardization,” *IEEE transactions on medical imaging*, vol. 19, no. 2, pp. 143–150, 2000.
- [60] J. P. O’Connor, E. O. Aboagye, J. E. Adams, H. J. Aerts, S. F. Barrington, A. J. Beer, R. Boellaard, S. E. Bohndiek, M. Brady, G. Brown, *et al.*, “Imaging biomarker roadmap for cancer studies,” *Nature reviews Clinical oncology*, vol. 14, no. 3, pp. 169–186, 2017.

- [61] J. Goya-Outi, F. Orlhac, R. Calmon, A. Alentorn, C. Nioche, C. Philippe, S. Puget, N. Boddaert, I. Buvat, J. Grill, *et al.*, “Computation of reliable textural indices from multimodal brain mri: suggestions based on a study of patients with diffuse intrinsic pontine glioma,” *Physics in Medicine & Biology*, vol. 63, no. 10, p. 105003, 2018.
- [62] R. T. Shinohara, E. M. Sweeney, J. Goldsmith, N. Shiee, F. J. Mateen, P. A. Calabresi, S. Jarso, D. L. Pham, D. S. Reich, C. M. Crainiceanu, *et al.*, “Statistical normalization techniques for magnetic resonance imaging,” *NeuroImage: Clinical*, vol. 6, pp. 9–19, 2014.
- [63] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, “N4itk: improved n3 bias correction,” *IEEE transactions on medical imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.
- [64] J. A. Maintz and M. A. Viergever, “A survey of medical image registration,” *Medical image analysis*, vol. 2, no. 1, pp. 1–36, 1998.
- [65] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, “Improved optimization for the robust and accurate linear registration and motion correction of brain images,” *Neuroimage*, vol. 17, no. 2, pp. 825–841, 2002.
- [66] R. Madhogarhia, D. Haldar, S. Bagheri, A. Familiar, H. Anderson, S. Arif, A. Vossough, P. Storm, A. Resnick, C. Davatzikos, *et al.*, “Radiomics and radiogenomics in pediatric neuro-oncology: a review,” *Neuro-Oncology Advances*, vol. 4, no. 1, p. vdac083, 2022.
- [67] M. Hatt, F. Tixier, D. Visvikis, and C. C. Le Rest, “Radiomics in pet/ct: more than meets the eye?,” *Journal of Nuclear Medicine*, vol. 58, no. 3, pp. 365–366, 2017.
- [68] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. Van Stiphout, P. Granton, C. M. Zegers, R. Gillies, R. Boellard, A. Dekker, *et al.*, “Radiomics: extracting more information from medical images using advanced feature analysis,” *European journal of cancer*, vol. 48, no. 4, pp. 441–446, 2012.
- [69] R. J. Gillies, P. E. Kinahan, and H. Hricak, “Radiomics: images are more than pictures, they are data,” *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.
- [70] R. J. Gillies, A. Anderson, R. Gatenby, and D. Morse, “The biology underlying molecular imaging in oncology: from genome to anatome and back again,” *Clinical radiology*, vol. 65, no. 7, pp. 517–521, 2010.
- [71] A. Zwanenburg, M. Vallières, M. A. Abdalah, H. J. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R. J. Beukinga, R. Boellaard, *et al.*, “The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping,” *Radiology*, vol. 295, no. 2, pp. 328–338, 2020.
- [72] J. O. Deasy, A. I. Blanco, and V. H. Clark, “Cerr: a computational environment for radiotherapy research,” *Medical physics*, vol. 30, no. 5, pp. 979–985, 2003.

- [73] A. Bettinelli, M. Branchini, F. De Monte, A. Scaggion, and M. Paiusco, "An ibex adaption toward image biomarker standardization," *Medical physics*, vol. 47, no. 3, pp. 1167–1173, 2020.
- [74] E. Pfaehler, A. Zwanenburg, J. R. de Jong, and R. Boellaard, "Racat: an open source and easy to use radiomics calculator tool," *PLoS One*, vol. 14, no. 2, p. e0212223, 2019.
- [75] C. Nioche, F. Orhac, S. Boughdad, S. Reuzé, J. Goya-Outi, C. Robert, C. Pellot-Barakat, M. Soussan, F. Frouin, and I. Buvat, "Lifex: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity," *Cancer research*, vol. 78, no. 16, pp. 4786–4789, 2018.
- [76] J. J. Van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. Beets-Tan, J.-C. Fillion-Robin, S. Pieper, and H. J. Aerts, "Computational radiomics system to decode the radiographic phenotype," *Cancer research*, vol. 77, no. 21, pp. e104–e107, 2017.
- [77] E. J. Limkin, S. Reuzé, A. Carré, R. Sun, A. Schernberg, A. Alexis, E. Deutsch, C. Ferté, and C. Robert, "The complexity of tumor shape, spiculatedness, correlates with tumor radiomic shape features," *Scientific reports*, vol. 9, no. 1, p. 4329, 2019.
- [78] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Transactions on systems, man, and cybernetics*, no. 6, pp. 610–621, 1973.
- [79] M. M. Galloway, "Texture analysis using gray level run lengths," *Computer graphics and image processing*, vol. 4, no. 2, pp. 172–179, 1975.
- [80] A. Chu, C. M. Sehgal, and J. F. Greenleaf, "Use of gray value distribution of run lengths for texture analysis," *Pattern recognition letters*, vol. 11, no. 6, pp. 415–419, 1990.
- [81] D.-H. Xu, A. S. Kurani, J. D. Furst, and D. S. Raicu, "Run-length encoding for volumetric texture," *Heart*, vol. 27, no. 25, pp. 452–458, 2004.
- [82] X. Tang, "Texture information in run-length matrices," *IEEE transactions on image processing*, vol. 7, no. 11, pp. 1602–1609, 1998.
- [83] G. Thibault, B. Fertil, C. L. Navarro, S. Pereira, P. Cau, N. Lévy, J. Sequeira, and J.-L. Mari, "Texture indexes and gray level size zone matrix. Application to cell nuclei classification," in *10th International Conference on Pattern Recognition and Information Processing, PRIP 2009*, (Minsk, Belarus), pp. 140–145, 2009.
- [84] M. Amadasun and R. King, "Textural features corresponding to textural properties," *IEEE Transactions on systems, man, and Cybernetics*, vol. 19, no. 5, pp. 1264–1274, 1989.

- [85] C. Sun and W. G. Wee, "Neighboring gray level dependence matrix for texture classification," *Computer Vision, Graphics, and Image Processing*, vol. 23, no. 3, pp. 341–352, 1983.
- [86] A. P. Pentland, "Fractal-based description of natural scenes," *IEEE transactions on pattern analysis and machine intelligence*, no. 6, pp. 661–674, 1984.
- [87] F. Davnall, C. S. Yip, G. Ljungqvist, M. Selmi, F. Ng, B. Sanghera, B. Ganeshan, K. A. Miles, G. J. Cook, and V. Goh, "Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice?," *Insights into imaging*, vol. 3, pp. 573–589, 2012.
- [88] R. Johansen, L. R. Jensen, J. Rydland, P. E. Goa, K. A. Kvistad, T. F. Bathen, D. E. Axelson, S. Lundgren, and I. S. Gribbestad, "Predicting survival and early clinical response to primary chemotherapy for patients with locally advanced breast cancer using dce-mri," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 29, no. 6, pp. 1300–1307, 2009.
- [89] H. J. Baek, H. S. Kim, N. Kim, Y. J. Choi, and Y. J. Kim, "Percent change of perfusion skewness and kurtosis: a potential imaging biomarker for early treatment response in patients with newly diagnosed glioblastomas," *Radiology*, vol. 264, no. 3, pp. 834–843, 2012.
- [90] A. Shukla-Dave, N. Y. Lee, J. F. Jansen, H. T. Thaler, H. E. Stambuk, M. G. Fury, S. G. Patel, A. L. Moreira, E. Sherman, S. Karimi, *et al.*, "Dynamic contrast-enhanced magnetic resonance imaging as a predictor of outcome in head-and-neck squamous cell carcinoma patients with nodal metastases," *International Journal of Radiation Oncology* Biology* Physics*, vol. 82, no. 5, pp. 1837–1844, 2012.
- [91] P. Foroutan, J. M. Kreaehling, D. L. Morse, O. Grove, M. C. Lloyd, D. Reed, M. Raghavan, S. Altiok, G. V. Martinez, and R. J. Gillies, "Diffusion mri and novel texture analysis in osteosarcoma xenotransplants predicts response to anti-checkpoint therapy," *PLoS one*, vol. 8, no. 12, p. e82875, 2013.
- [92] A. D. King, K.-K. Chow, K.-H. Yu, F. K. F. Mo, D. K. Yeung, J. Yuan, K. S. Bhatia, A. C. Vlantis, and A. T. Ahuja, "Head and neck squamous cell carcinoma: diagnostic performance of diffusion-weighted mr imaging for the prediction of treatment response," *Radiology*, vol. 266, no. 2, pp. 531–538, 2013.
- [93] S.-L. Peng, C.-F. Chen, H.-L. Liu, C.-C. Lui, Y.-J. Huang, T.-H. Lee, C.-C. Chang, and F.-N. Wang, "Analysis of parametric histogram from dynamic contrast-enhanced mri: application in evaluating brain tumor response to radiotherapy," *NMR in Biomedicine*, vol. 26, no. 4, pp. 443–450, 2013.
- [94] J. F. Eary, F. O'Sullivan, J. O'Sullivan, and E. U. Conrad, "Spatial heterogeneity in sarcoma 18f-fdg uptake as a predictor of patient outcome," *Journal of Nuclear Medicine*, vol. 49, no. 12, pp. 1973–1979, 2008.

- [95] I. El Naqa, P. Grigsby, A. Apte, E. Kidd, E. Donnelly, D. Khullar, S. Chaudhari, D. Yang, M. Schmitt, R. Laforest, *et al.*, "Exploring feature-based approaches in pet images for predicting cancer treatment outcomes," *Pattern recognition*, vol. 42, no. 6, pp. 1162–1171, 2009.
- [96] F. Yang, M. A. Thomas, F. Dehdashti, and P. W. Grigsby, "Temporal analysis of intratumoral metabolic heterogeneity characterized by textural features in cervical cancer," *European journal of nuclear medicine and molecular imaging*, vol. 40, pp. 716–727, 2013.
- [97] G. J. Cook, C. Yip, M. Siddique, V. Goh, S. Chicklore, A. Roy, P. Marsden, S. Ahmad, and D. Landau, "Are pretreatment 18f-fdg pet tumor textural features in non-small cell lung cancer associated with response and survival after chemoradiotherapy?," *Journal of nuclear medicine*, vol. 54, no. 1, pp. 19–26, 2013.
- [98] F. Tixier, C. C. Le Rest, M. Hatt, N. Albarghach, O. Pradier, J.-P. Metges, L. Corcos, and D. Visvikis, "Intra-tumor heterogeneity characterized by textural features on baseline 18f-fdg pet images predicts response to concomitant radiochemotherapy in esophageal cancer," *Journal of Nuclear Medicine*, vol. 52, no. 3, pp. 369–378, 2011.
- [99] H. Zhang, S. Tan, W. Chen, S. Kligerman, G. Kim, W. D. D'Souza, M. Suntharalingam, and W. Lu, "Modeling pathologic response of esophageal cancer to chemoradiation therapy using spatial-temporal 18f-fdg pet features, clinical parameters, and demographics," *International Journal of Radiation Oncology* Biology* Physics*, vol. 88, no. 1, pp. 195–203, 2014.
- [100] H. J. Aerts, E. R. Velazquez, R. T. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, *et al.*, "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature communications*, vol. 5, no. 1, p. 4006, 2014.
- [101] C. Parmar, P. Grossmann, J. Bussink, P. Lambin, and H. J. Aerts, "Machine learning methods for quantitative radiomic biomarkers," *Scientific reports*, vol. 5, no. 1, p. 13087, 2015.
- [102] U. Tateishi, M. Kusumoto, H. Nishihara, K. Nagashima, T. Morikawa, and N. Moriyama, "Contrast-enhanced dynamic computed tomography for the evaluation of tumor angiogenesis in patients with lung carcinoma," *Cancer: Interdisciplinary International Journal of the American Cancer Society*, vol. 95, no. 4, pp. 835–842, 2002.
- [103] C. K. Kim, J. H. Lim, C. K. Park, D. Choi, H. K. Lim, and W. J. Lee, "Neoangiogenesis and sinusoidal capillarization in hepatocellular carcinoma: correlation between dynamic ct and density of tumor microvessels," *Radiology*, vol. 237, no. 2, pp. 529–534, 2005.

- [104] F. Tixier, A. M. Groves, V. Goh, M. Hatt, P. Ingrand, C. C. Le Rest, and D. Visvikis, "Correlation of intra-tumor 18f-fdg uptake heterogeneity indices with perfusion ct derived parameters in colorectal cancer," *PloS one*, vol. 9, no. 6, p. e99567, 2014.
- [105] M. Vallières, C. R. Freeman, S. R. Skamene, and I. El Naqa, "A radiomics model from joint fdg-pet and mri texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities," *Physics in Medicine & Biology*, vol. 60, no. 14, p. 5471, 2015.
- [106] J. Lin, S. Kligerman, R. Goel, P. Sajedi, M. Suntharalingam, and M. D. Chuong, "State-of-the-art molecular imaging in esophageal cancer management: implications for diagnosis, prognosis, and treatment," *Journal of gastrointestinal oncology*, vol. 6, no. 1, p. 3, 2015.
- [107] W. Mu, Z. Chen, Y. Liang, W. Shen, F. Yang, R. Dai, N. Wu, and J. Tian, "Staging of cervical cancer based on tumor heterogeneity characterized by texture features on 18f-fdg pet images," *Physics in Medicine & Biology*, vol. 60, no. 13, p. 5123, 2015.
- [108] B. Ganeshan, S. Abaleke, R. C. Young, C. R. Chatwin, and K. A. Miles, "Texture analysis of non-small cell lung cancer on unenhanced computed tomography: initial evidence for a relationship with tumour glucose metabolism and stage," *Cancer imaging*, vol. 10, no. 1, p. 137, 2010.
- [109] E. I. Zacharaki, S. Wang, S. Chawla, D. Soo Yoo, R. Wolf, E. R. Melhem, and C. Davatzikos, "Classification of brain tumor type and grade using mri texture and shape in a machine learning scheme," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 62, no. 6, pp. 1609–1618, 2009.
- [110] R. A. Lerski, K. Straughan, L. Schad, D. Boyce, S. Blüml, and I. Zuna, "Viii. mr image texture analysis—an approach to tissue characterization," *Magnetic resonance imaging*, vol. 11, no. 6, pp. 873–887, 1993.
- [111] L. Kjaer, P. Ring, C. Thomsen, and O. Henriksen, "Texture analysis in quantitative mr imaging: tissue characterisation of normal brain and intracranial tumours at 1.5 t," *Acta Radiologica*, vol. 36, no. 2, pp. 127–135, 1995.
- [112] D. Mahmoud-Ghoneim, G. Toussaint, J.-M. Constans, and J. D. de Certaines, "Three dimensional texture analysis in mri: a preliminary evaluation in gliomas," *Magnetic resonance imaging*, vol. 21, no. 9, pp. 983–987, 2003.
- [113] K. Nie, J.-H. Chen, J. Y. Hon, Y. Chu, O. Nalcioglu, and M.-Y. Su, "Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast mri," *Academic radiology*, vol. 15, no. 12, pp. 1513–1525, 2008.

- [114] M. F. McNitt-Gray, E. M. Hart, N. Wyckoff, J. W. Sayre, J. G. Goldin, and D. R. Aberle, "A pattern classification approach to characterizing solitary pulmonary nodules imaged on high resolution ct: preliminary results," *Medical physics*, vol. 26, no. 6, pp. 880–888, 1999.
- [115] S. Kido, K. Kuriyama, M. Higashiyama, T. Kasugai, and C. Kuroda, "Fractal analysis of small peripheral pulmonary nodules in thin-section ct: evaluation of the lung-nodule interfaces," *Journal of computer assisted tomography*, vol. 26, no. 4, pp. 573–578, 2002.
- [116] I. Petkovska, S. K. Shah, M. F. McNitt-Gray, J. G. Goldin, M. S. Brown, H. J. Kim, K. Brown, and D. R. Aberle, "Pulmonary nodule characterization: a comparison of conventional with quantitative and visual semi-quantitative analyses using contrast enhancement maps," *European journal of radiology*, vol. 59, no. 2, pp. 244–252, 2006.
- [117] T. W. Way, L. M. Hadjiiski, B. Sahiner, H.-P. Chan, P. N. Cascade, E. A. Kazerooni, N. Bogot, and C. Zhou, "Computer-aided diagnosis of pulmonary nodules on ct scans: Segmentation and classification using 3d active contours," *Medical physics*, vol. 33, no. 7Part1, pp. 2323–2337, 2006.
- [118] R. Xu, S. Kido, K. Suga, Y. Hirano, R. Tachibana, K. Muramatsu, K. Chagawa, and S. Tanaka, "Texture analysis on 18 f-fdg pet/ct images to differentiate malignant and benign bone and soft-tissue lesions," *Annals of nuclear medicine*, vol. 28, pp. 926–935, 2014.
- [119] H. Yu, C. Caldwell, K. Mah, I. Poon, J. Balogh, R. MacKenzie, N. Khaouam, and R. Tirona, "Automated radiation targeting in head-and-neck cancer using region-based texture analysis of pet and ct images," *International Journal of Radiation Oncology* Biology* Physics*, vol. 75, no. 2, pp. 618–625, 2009.
- [120] M. Diehn, C. Nardini, D. S. Wang, S. McGovern, M. Jayaraman, Y. Liang, K. Aldape, S. Cha, and M. D. Kuo, "Identification of noninvasive imaging surrogates for brain tumor gene-expression modules," *Proceedings of the National Academy of Sciences*, vol. 105, no. 13, pp. 5213–5218, 2008.
- [121] B. Ellingson, A. Lai, R. Harris, J. Selfridge, W. Yong, K. Das, W. Pope, P. Nghiemphu, H. Vinters, L. Liau, *et al.*, "Probabilistic radiographic atlas of glioblastoma phenotypes," *American Journal of neuroradiology*, vol. 34, no. 3, pp. 533–540, 2013.
- [122] K. M. Naeini, W. B. Pope, T. F. Cloughesy, R. J. Harris, A. Lai, A. Eskin, R. Chowdhury, H. S. Phillips, P. L. Nghiemphu, Y. Behbahanian, *et al.*, "Identifying the mesenchymal molecular subtype of glioblastoma using quantitative volumetric analysis of anatomic magnetic resonance images," *Neuro-oncology*, vol. 15, no. 5, pp. 626–634, 2013.

- [123] D. A. Gutman, W. D. Dunn, P. Grossmann, L. A. Cooper, C. A. Holder, K. L. Ligon, B. M. Alexander, and H. J. Aerts, "Somatic mutations associated with mri-derived volumetric features in glioblastoma," *Neuroradiology*, vol. 57, pp. 1227–1237, 2015.
- [124] V. S. Nair, O. Gevaert, G. Davidzon, S. Napel, E. E. Graves, C. D. Hoang, J. B. Shrager, A. Quon, D. L. Rubin, and S. K. Plevritis, "Prognostic pet 18f-fdg uptake imaging features are associated with major oncogenomic alterations in patients with resected non-small cell lung cancer," *Cancer research*, vol. 72, no. 15, pp. 3725–3734, 2012.
- [125] V. S. Nair, O. Gevaert, G. Davidzon, S. K. Plevritis, and R. West, "Nf- κ b protein expression associates with 18f-fdg pet tumor uptake in non-small cell lung cancer: a radiogenomics validation study to understand tumor metabolism," *Lung Cancer*, vol. 83, no. 2, pp. 189–196, 2014.
- [126] J. Goya-Outi, R. Calmon, F. Orlhac, C. Philippe, N. Boddart, S. Puget, I. Buvat, V. Frouin, J. Grill, and F. Frouin, "Can structural mri radiomics predict dipg histone h3 mutation and patient overall survival at diagnosis time?," in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pp. 1–4, IEEE, 2019.
- [127] S. G. Kandemirli, B. Kocak, S. Naganawa, K. Ozturk, S. S. Yip, S. Chopra, L. Rivetti, A. S. Aldine, K. Jones, Z. Cayci, *et al.*, "Machine learning-based multiparametric magnetic resonance imaging radiomics for prediction of h3k27m mutation in midline gliomas," *World neurosurgery*, vol. 151, pp. e78–e85, 2021.
- [128] Q. Li, F. Dong, B. Jiang, and M. Zhang, "Exploring mri characteristics of brain diffuse midline gliomas with the h3 k27m mutation using radiomics," *Frontiers in Oncology*, vol. 11, p. 646267, 2021.
- [129] M.-F. Chilaca-Rosas, M. Garcia-Lezama, S. Moreno-Jimenez, and E. Roldan-Valadez, "Diagnostic performance of selected mri-derived radiomics able to discriminate progression-free and overall survival in patients with midline glioma and the h3f3ak27m mutation," *Diagnostics*, vol. 13, no. 5, p. 849, 2023.
- [130] K. Lv, H. Chen, X. Cao, P. Du, J. Chen, X. Liu, L. Zhu, D. Geng, and J. Zhang, "Development and validation of a machine learning algorithm for predicting diffuse midline glioma, h3 k27–altered, h3 k27 wild-type high-grade glioma, and primary cns lymphoma of the brain midline in adults," *Journal of Neurosurgery*, vol. 1, no. aop, pp. 1–9, 2022.
- [131] W. Guo, D. She, Z. Xing, X. Lin, F. Wang, Y. Song, and D. Cao, "Multiparametric mri-based radiomics model for predicting h3 k27m mutant status in diffuse midline glioma: a comparative study across different sequences and machine learning techniques," *Frontiers in Oncology*, vol. 12, 2022.
- [132] M. W. Wagner, K. Namdar, M. Napoleone, N. Hainc, A. Amirabadi, A. Fonseca, S. Laughlin, M. M. Shroff, E. Bouffet, C. Hawkins, *et al.*, "Radiomic features based on mri predict progression-free survival in pediatric

diffuse midline glioma/diffuse intrinsic pontine glioma,” *Canadian Association of Radiologists Journal*, vol. 74, no. 1, pp. 119–126, 2023.

- [133] L. T. Tam, K. W. Yeom, J. N. Wright, A. Jaju, A. Radmanesh, M. Han, S. Toescu, M. Maleki, E. Chen, A. Campion, *et al.*, “Mri-based radiomics for prognosis of pediatric diffuse intrinsic pontine glioma: an international study,” *Neuro-oncology advances*, vol. 3, no. 1, p. vdab042, 2021.
- [134] S. S. Yip and H. J. Aerts, “Applications and limitations of radiomics,” *Physics in Medicine & Biology*, vol. 61, no. 13, p. R150, 2016.
- [135] D. V. Marie-Charlotte Desseroit, Catherine Cheze Le Rest and M. Hatt., “Caractrisation et exploitation de l’htrognit intra-tumorale des images multi-modales tdm et tep,” *Thesis*, 2016.
- [136] I. Jolliffe, “Principal component analysis,” *Encyclopedia of statistics in behavioral science*, 2005.
- [137] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of machine learning research*, vol. 3, no. Mar, pp. 1157–1182, 2003.
- [138] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, “Gene selection for cancer classification using support vector machines,” *Machine learning*, vol. 46, no. 1-3, pp. 389–422, 2002.
- [139] R. Díaz-Uriarte and S. Alvarez de Andrés, “Gene selection and classification of microarray data using random forest,” *BMC bioinformatics*, vol. 7, pp. 1–13, 2006.
- [140] B. A. Skourt, A. El Hassani, and A. Majda, “Lung ct image segmentation using deep neural networks,” *Procedia Computer Science*, vol. 127, pp. 109–113, 2018.
- [141] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, “Dermatologist-level classification of skin cancer with deep neural networks,” *nature*, vol. 542, no. 7639, pp. 115–118, 2017.
- [142] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. Aerts, “Artificial intelligence in radiology,” *Nature Reviews Cancer*, vol. 18, no. 8, pp. 500–510, 2018.
- [143] M. Ghafoorian, N. Karssemeijer, T. Heskes, I. W. van Uden, C. I. Sanchez, G. Litjens, F.-E. de Leeuw, B. van Ginneken, E. Marchiori, and B. Platel, “Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities,” *Scientific Reports*, vol. 7, no. 1, pp. 1–12, 2017.
- [144] K. T. Litjens G, “A survey on deep learning in medical image analysis,” *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.
- [145] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, *et al.*, “Mastering the game of go with deep neural networks and tree search,” *nature*, vol. 529, no. 7587, pp. 484–489, 2016.

- [146] B. Sahiner, A. Pezeshk, L. M. Hadjiiski, X. Wang, K. Drukker, K. H. Cha, R. M. Summers, and M. L. Giger, "Deep learning in medical imaging and radiation therapy," *Medical physics*, vol. 46, no. 1, pp. e1–e36, 2019.
- [147] X. Zhang, Y. Zhang, G. Zhang, X. Qiu, W. Tan, X. Yin, and L. Liao, "Deep learning with radiomics for disease diagnosis and treatment: challenges and potential," *Frontiers in oncology*, vol. 12, p. 773840, 2022.
- [148] J. Konar, P. Khandelwal, and R. Tripathi, "Comparison of various learning rate scheduling techniques on convolutional neural network," in *2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pp. 1–5, IEEE, 2020.
- [149] R. B. Rao, G. Fung, and R. Rosales, "On the dangers of cross-validation. an experimental evaluation," in *Proceedings of the 2008 SIAM international conference on data mining*, pp. 588–596, SIAM, 2008.
- [150] T. H. Jerome Friedman and R. Tibshirani, "The elements of statistical learning," *Neuroimage*, vol. volume 1, no. 1, 2001.
- [151] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization, january 2017."
- [152] F.-F. Li, R. Krishna, and D. Xu, *Cs231n: Convolutional neural networks for visual recognition*. 2020.
- [153] A. C. Wilson, R. Roelofs, M. Stern, N. Srebro, and B. Recht, *The Marginal Value of Adaptive Gradient Methods in Machine Learning*. *arXiv*. 2018.
- [154] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.
- [155] A. Agresti, *Categorical data analysis*, vol. 792. John Wiley & Sons, 2012.
- [156] C.-Y. J. Peng, K. L. Lee, and G. M. Ingersoll, "An introduction to logistic regression analysis and reporting," *The journal of educational research*, vol. 96, no. 1, pp. 3–14, 2002.
- [157] G. D. Garson, "Testing statistical assumptions," 2012.
- [158] E. W. Steyerberg, A. J. Vickers, N. R. Cook, T. Gerds, M. Gonen, N. Obuchowski, M. J. Pencina, and M. W. Kattan, "Assessing the performance of prediction models: a framework for some traditional and novel measures," *Epidemiology (Cambridge, Mass.)*, vol. 21, no. 1, p. 128, 2010.
- [159] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, pp. 273–297, 1995.
- [160] A. C. Chang and A. Limon, "Introduction to artificial intelligence for cardiovascular clinicians," in *Intelligence-Based Cardiology and Cardiac Surgery*, pp. 3–120, Elsevier, 2024.
- [161] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

- [162] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.
- [163] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.
- [164] E. L. Kaplan and P. Meier, "Nonparametric estimation from incomplete observations," *Journal of the American statistical association*, vol. 53, no. 282, pp. 457–481, 1958.
- [165] X. Zhang, Q. Tian, L. Wang, Y. Liu, B. Li, Z. Liang, P. Gao, K. Zheng, B. Zhao, and H. Lu, "Radiomics strategy for molecular subtype stratification of lower-grade glioma: detecting idh and tp53 mutations based on multimodal mri," *Journal of Magnetic Resonance Imaging*, vol. 48, no. 4, pp. 916–926, 2018.
- [166] F. E. J. Harrell, K. L. Lee, and D. B. Mark, "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," *Statistics in medicine*, vol. 15, no. 4, pp. 361–387, 1996.
- [167] L. Rebaud, T. Escobar, F. Khalid, K. Girum, and I. Buvat, "Simplicity is all you need: out-of-the-box nnunet followed by binary-weighted radiomic model for segmentation and outcome prediction in head and neck pet/ct," in *3D Head and Neck Tumor Segmentation in PET/CT Challenge*, pp. 121–134, Springer, 2022.
- [168] D. B. Rubin, *Multiple imputation for nonresponse in surveys*, vol. 81. John Wiley & Sons, 2004.
- [169] A. D. Shah, J. W. Bartlett, J. Carpenter, O. Nicholas, and H. Hemingway, "Comparison of random forest and parametric imputation models for imputing missing data using mice: a caliber study," *American journal of epidemiology*, vol. 179, no. 6, pp. 764–774, 2014.
- [170] . L. D. P. Rahardiantoro, A. D., "Missing data handling with mode imputation in predicting the performance of accounting information system using regression analysis," *Journal of Physics: Conference Series*, vol. 179, no. 6, p. 1339(1), 2019.
- [171] O. Troyanskaya, M. Cantor, G. Sherlock, P. Brown, T. Hastie, R. Tibshirani, D. Botstein, and R. B. Altman, "Missing value estimation methods for dna microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001.
- [172] . R. D. B. Little, R. J., *Statistical analysis with missing data*. John Wiley & Sons, 2002.
- [173] F. Vaida and R. Xu, "Proportional hazards model with random effects," *Statistics in medicine*, vol. 19, no. 24, pp. 3309–3324, 2000.
- [174] J. Barnard and D. B. Rubin, "Miscellanea. small-sample degrees of freedom with multiple imputation," *Biometrika*, vol. 86, no. 4, pp. 948–955, 1999.

- [175] J. M. Jerez, I. Molina, P. J. García-Laencina, E. Alba, N. Ribelles, M. Martín, and L. Franco, "Missing data imputation using statistical and machine learning methods in a real breast cancer problem," *Artificial intelligence in medicine*, vol. 50, no. 2, pp. 105–115, 2010.
- [176] R. J. Little and D. B. Rubin, "Bayes and multiple imputation," *Statistical analysis with missing data*, pp. 200–220, 2002.
- [177] M. Arjovsky and L. Bottou, "Towards principled methods for training generative adversarial networks," *arXiv preprint arXiv:1701.04862*, 2017.
- [178] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," *Advances in neural information processing systems*, vol. 29, 2016.
- [179] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, and I. Išgum, "Deep mr to ct synthesis using unpaired data," in *Simulation and Synthesis in Medical Imaging: Second International Workshop, SASHIMI 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 10, 2017, Proceedings 2*, pp. 14–23, Springer, 2017.
- [180] X. Chen, Y. Duan, R. Houthoofd, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," *Advances in neural information processing systems*, vol. 29, 2016.
- [181] H. Seo, C. Huang, M. Bassenne, R. Xiao, and L. Xing, "Modified u-net (mu-net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in ct images," *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1316–1325, 2019.
- [182] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
- [183] M.-J. Saint Martin, F. Orlhac, P. Akl, F. Khalid, C. Nioche, I. Buvat, C. Malhaire, and F. Frouin, "A radiomics pipeline dedicated to breast mri: validation on a multi-scanner phantom study," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 34, pp. 355–366, 2021.
- [184] B. Glocker, R. Robinson, D. C. Castro, Q. Dou, and E. Konukoglu, "Machine learning with multi-site imaging data: An empirical study on the impact of scanner effects," *arXiv preprint arXiv:1910.04597*, 2019.
- [185] E. H. Pooch, P. Ballester, and R. C. Barros, "Can we trust deep learning based diagnosis? the impact of domain shift in chest radiograph classification," in *Thoracic Image Analysis: Second International Workshop, TIA 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 2*, pp. 74–83, Springer, 2020.

- [186] L. Yao, J. Prosky, B. Covington, and K. Lyman, "A strong baseline for domain adaptation and generalization in medical imaging," *arXiv preprint arXiv:1904.01638*, 2019.
- [187] G. Mårtensson, D. Ferreira, T. Granberg, L. Cavallin, K. Oppedal, A. Padovani, I. Rektorova, L. Bonanni, M. Pardini, M. G. Kramberger, *et al.*, "The reliability of a deep learning model in clinical out-of-distribution mri data: a multicohort study," *Medical Image Analysis*, vol. 66, p. 101714, 2020.
- [188] E. A. AlBadawy, A. Saha, and M. A. Mazurowski, "Deep learning for segmentation of brain tumors: Impact of cross-institutional training and testing," *Medical physics*, vol. 45, no. 3, pp. 1150–1158, 2018.
- [189] K. T. Chen, M. Schürer, J. Ouyang, M. E. I. Koran, G. Davidzon, E. Mormino, S. Tiepolt, K.-T. Hoffmann, O. Sabri, G. Zaharchuk, *et al.*, "Generalization of deep learning models for ultra-low-count amyloid pet/mri using transfer learning," *European journal of nuclear medicine and molecular imaging*, vol. 47, pp. 2998–3007, 2020.
- [190] M. Ghafoorian, A. Mehrtash, T. Kapur, N. Karssemeijer, E. Marchiori, M. Pesteie, C. R. Guttmann, F.-E. de Leeuw, C. M. Tempany, B. Van Ginneken, *et al.*, "Transfer learning for domain adaptation in mri: Application in brain lesion segmentation," in *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*, pp. 516–524, Springer, 2017.
- [191] K. Chang, N. Balachandar, C. Lam, D. Yi, J. Brown, A. Beers, B. Rosen, D. L. Rubin, and J. Kalpathy-Cramer, "Distributed deep learning networks among institutions for medical imaging," *Journal of the American Medical Informatics Association*, vol. 25, no. 8, pp. 945–954, 2018.
- [192] M. J. Sheller, G. A. Reina, B. Edwards, J. Martin, and S. Bakas, "Multi-institutional deep learning modeling without sharing patient data: A feasibility study on brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 4th International Workshop, BrainLes 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Revised Selected Papers, Part I 4*, pp. 92–104, Springer, 2019.
- [193] R. Boellaard, R. Delgado-Bolton, W. J. Oyen, F. Giammarile, K. Tatsch, W. Eschner, F. J. Verzijlbergen, S. F. Barrington, L. C. Pike, W. A. Weber, *et al.*, "Fdg pet/ct: Eanm procedure guidelines for tumour imaging: version 2.0," *European journal of nuclear medicine and molecular imaging*, vol. 42, pp. 328–354, 2015.
- [194] L. P. Clarke, R. J. Nordstrom, H. Zhang, P. Tandon, Y. Zhang, G. Redmond, K. Farahani, G. Kelloff, L. Henderson, L. Shankar, *et al.*, "The quantitative imaging network: Nci's historical perspective and planned goals," *Translational oncology*, vol. 7, no. 1, pp. 1–4, 2014.

- [195] M. Shafiq-ul Hassan, K. Latifi, G. Zhang, G. Ullah, R. Gillies, and E. Moros, "Voxel size and gray level normalization of ct radiomic features in lung cancer," *Scientific reports*, vol. 8, no. 1, p. 10545, 2018.
- [196] D. Mackin, X. Fave, L. Zhang, J. Yang, A. K. Jones, C. S. Ng, and L. Court, "Harmonizing the pixel size in retrospective computed tomography radiomics studies," *PloS one*, vol. 12, no. 9, p. e0178524, 2017.
- [197] A. Chatterjee, M. Vallières, A. Dohan, I. R. Levesque, Y. Ueno, S. Saif, C. Reinhold, and J. Seuntjens, "Creating robust predictive radiomic models for data from independent institutions using normalization," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 3, no. 2, pp. 210–215, 2019.
- [198] W. E. Johnson, C. Li, and A. Rabinovic, "Adjusting batch effects in microarray expression data using empirical bayes methods," *Biostatistics*, vol. 8, no. 1, pp. 118–127, 2007.
- [199] F. Orlhac, F. Frouin, C. Nioche, N. Ayache, and I. Buvat, "Validation of a method to compensate multicenter effects affecting ct radiomics," *Radiology*, vol. 291, no. 1, pp. 53–59, 2019.
- [200] J.-P. Fortin, D. Parker, B. Tunç, T. Watanabe, M. A. Elliott, K. Ruparel, D. R. Roalf, T. D. Satterthwaite, R. C. Gur, R. E. Gur, *et al.*, "Harmonization of multi-site diffusion tensor imaging data," *Neuroimage*, vol. 161, pp. 149–170, 2017.
- [201] F. Orlhac, J. J. Eertink, A.-S. Cottreau, J. M. Zijlstra, C. Thieblemont, M. Meignan, R. Boellaard, and I. Buvat, "A guide to combat harmonization of imaging biomarkers in multicenter studies," *Journal of Nuclear Medicine*, vol. 63, no. 2, pp. 172–179, 2022.
- [202] F. Orlhac, A. Lecler, J. Savatovski, J. Goya-Outi, C. Nioche, F. Charbonneau, N. Ayache, F. Frouin, L. Duron, and I. Buvat, "How can we combat multicenter variability in mr radiomics? validation of a correction procedure," *European radiology*, vol. 31, pp. 2272–2280, 2021.
- [203] N. Barua, S. Lowis, M. Woolley, S. O'sullivan, R. Harrison, and S. Gill, "Robot-guided convection-enhanced delivery of carboplatin for advanced brainstem glioma," *Acta neurochirurgica*, vol. 155, pp. 1459–1465, 2013.
- [204] K. Müller, A. Schlamann, C. Seidel, M. Warmuth-Metz, H. Christiansen, D. Vordermark, R. Kortmann, C. Kramm, and A. von Bueren, "Craniospinal irradiation with concurrent temozolomide and nimotuzumab in a child with primary metastatic diffuse intrinsic pontine glioma. a compassionate use treatment," *Strahlenther Onkol*, vol. 189, no. 8, pp. 693–696, 2013.
- [205] T. Ono, H. Kuwashige, J.-I. Adachi, M. Takahashi, M. Oda, T. Kumabe, and H. Shimizu, "Long-term survival of a patient with diffuse midline glioma in the pineal region: A case report and literature review," *Surgical Neurology International*, vol. 12, 2021.

- [206] C. Y. Vincent, A. P. Landry, T. Purzner, A. Kalyvas, N. Mohan, P. J. O'Halloran, A. Gao, and G. Zadeh, "Adult isocitrate dehydrogenase–mutant brainstem glioma: illustrative case," *Journal of Neurosurgery: Case Lessons*, vol. 1, no. 12, 2021.
- [207] T. Picart, M. Barritault, D. Poncet, L.-P. Berner, C. Izquierdo, E. Tabouret, D. Figarella-Branger, A. Idbaïh, F. Bielle, V. Bourg, *et al.*, "Characteristics of diffuse hemispheric gliomas, h3 g34-mutant in adults," *Neuro-Oncology Advances*, vol. 3, no. 1, p. vdab061, 2021.
- [208] D. Hargrave, N. Chuang, and E. Bouffet, "Conventional mri cannot predict survival in childhood diffuse intrinsic pontine glioma," *Journal of neuro-oncology*, vol. 86, pp. 313–319, 2008.
- [209] R. M. Hayward, N. Patronas, E. H. Baker, G. Vézina, P. S. Albert, and K. E. Warren, "Inter-observer variability in the measurement of diffuse intrinsic pontine gliomas," *Journal of neuro-oncology*, vol. 90, pp. 57–61, 2008.
- [210] E. A. Steffen-Smith, J. H. Shih, S. J. Hipp, R. Bent, and K. E. Warren, "Proton magnetic resonance spectroscopy predicts survival in children with diffuse intrinsic pontine glioma," *Journal of neuro-oncology*, vol. 105, pp. 365–373, 2011.
- [211] R. M. Lober, Y.-J. Cho, Y. Tang, P. D. Barnes, M. S. Edwards, H. Vogel, P. G. Fisher, M. Monje, and K. W. Yeom, "Diffusion-weighted mri derived apparent diffusion coefficient identifies prognostically distinct subgroups of pediatric diffuse intrinsic pontine glioma," *Journal of neuro-oncology*, vol. 117, pp. 175–182, 2014.
- [212] M. H. Jansen, S. E. Veldhuijzen van Zanten, E. Sanchez Aliaga, M. W. Heymans, M. Warmuth-Metz, D. Hargrave, E. J. Van Der Hoeven, C. E. Gidding, E. S. de Bont, O. S. Eshghi, *et al.*, "Survival prediction model of children with diffuse intrinsic pontine glioma based on clinical and radiological criteria," *Neuro-oncology*, vol. 17, no. 1, pp. 160–166, 2015.
- [213] T. Y. Poussaint, S. Vajapeyam, K. I. Ricci, A. Panigrahy, M. Kocak, L. E. Kun, J. M. Boyett, I. F. Pollack, and M. Fouladi, "Apparent diffusion coefficient histogram metrics correlate with survival in diffuse intrinsic pontine glioma: a report from the pediatric brain tumor consortium," *Neuro-oncology*, vol. 18, no. 5, pp. 725–734, 2016.
- [214] Z. Zhou, N. Luther, G. M. Ibrahim, C. Hawkins, R. Vibhakar, M. H. Handler, and M. M. Souweidane, "B7-h3, a potential therapeutic target, is expressed in diffuse intrinsic pontine glioma," *Journal of neuro-oncology*, vol. 111, pp. 257–264, 2013.
- [215] Z. Gokce-Samar, P. Beuriat, C. Faure-Contet, C. Carrie, S. Chabaud, L. Claude, F. Di Rocco, C. Mottolese, A. Szathmari, C. Chabert, *et al.*, "Pre-radiation chemotherapy improves survival in pediatric diffuse intrinsic pontine gliomas," *Child's Nervous System*, vol. 32, pp. 1415–1423, 2016.

- [216] Z. Zhou, N. Luther, R. Singh, J. A. Boockvar, M. M. Souweidane, and J. P. Greenfield, "Glioblastoma spheroids produce infiltrative gliomas in the rat brainstem," *Child's Nervous System*, vol. 33, pp. 437–446, 2017.
- [217] H. Hassan, A. Pinches, S. V. Picton, and R. S. Phillips, "Survival rates and prognostic predictors of high grade brain stem gliomas in childhood: a systematic review and meta-analysis," *Journal of Neuro-oncology*, vol. 135, no. 1, pp. 13–20, 2017.
- [218] R. Calmon, S. Puget, P. Varlet, V. Dangouloff-Ros, T. Blauwblomme, K. Beccaria, D. Grevent, C. Sainte-Rose, D. Castel, M.-A. Debily, *et al.*, "Cerebral blood flow changes after radiation therapy identifies pseudoprogression in diffuse intrinsic pontine gliomas," *Neuro-oncology*, vol. 20, no. 7, pp. 994–1002, 2018.
- [219] S. E. Veldhuijzen van Zanten, A. Lane, M. W. Heymans, J. Baugh, B. Chaney, L. M. Hoffman, R. Doughman, M. H. Jansen, E. Sanchez, W. P. Vandertop, *et al.*, "External validation of the diffuse intrinsic pontine glioma survival prediction model: a collaborative report from the international dipg registry and the siop dipg registry," *Journal of neuro-oncology*, vol. 134, pp. 231–240, 2017.
- [220] G. S. Colafati, I. P. Voicu, C. Carducci, M. Caulo, M. Vinci, F. Diomed-Camassei, P. Merli, A. Carai, E. Miele, A. Cacchione, *et al.*, "Direct involvement of cranial nerve v at diagnosis in patients with diffuse intrinsic pontine glioma: a potential magnetic resonance predictor of short-term survival," *Frontiers in Oncology*, vol. 9, p. 204, 2019.
- [221] F. Garibotto, F. Madia, C. Milanaccio, A. Verrico, A. Piccardo, D. Tortora, G. Piatelli, M. C. Diana, V. Capra, M. L. Garrè, *et al.*, "Pediatric diffuse midline gliomas h3 k27m-mutant and non-histone mutant midline high-grade gliomas in neurofibromatosis type 1 in comparison with non-syndromic children: a single-center pilot study," *Frontiers in Oncology*, vol. 10, p. 795, 2020.
- [222] J. L. Leach, J. Roebker, A. Schafer, J. Baugh, B. Chaney, C. Fuller, M. Fouladi, A. Lane, R. Doughman, R. Drissi, *et al.*, "Mr imaging features of diffuse intrinsic pontine glioma (dipg) and relationship to overall survival: Report from the international dipg registry," *Neuro-Oncology*, 2020.
- [223] C. L. Tinkle, B. Simone, J. Chiang, X. Li, K. Campbell, Y. Han, Y. Li, L. D. Hover, J. K. Molitoris, J. Becksfort, *et al.*, "Defining optimal target volumes of conformal radiation therapy for diffuse intrinsic pontine glioma," *International Journal of Radiation Oncology* Biology* Physics*, vol. 106, no. 4, pp. 838–847, 2020.
- [224] M. F. Iannó, V. Biassoni, E. Schiavello, A. Carengo, L. Boschetti, L. Gandola, B. Diletto, E. Marchesi, C. Vegetti, A. Molla, *et al.*, "A microrna prognostic signature in patients with diffuse intrinsic pontine gliomas through non-invasive liquid biopsy," *Cancers*, vol. 14, no. 17, p. 4307, 2022.

- [225] H. J. Kim, J. H. Lee, Y. Kim, D. H. Lim, S.-H. Park, S. Do Ahn, I. A. Kim, J. H. Im, J. W. Chung, J.-Y. Kim, *et al.*, “Suggestions for escaping the dark ages for pediatric diffuse intrinsic pontine glioma treated with radiotherapy: analysis of prognostic factors from the national multicenter study,” 2021.
- [226] E. Cantor, K. Wierzbicki, R. S. Tarapore, K. Ravi, C. Thomas, R. Cartaxo, V. Nand Yadav, R. Ravindran, A. K. Bruzek, J. Wadden, *et al.*, “Serial h3k27m cell-free tumor dna (cf-tdna) tracking predicts onc201 treatment response and progression in diffuse midline glioma,” *Neuro-oncology*, vol. 24, no. 8, pp. 1366–1374, 2022.
- [227] J. Liu, H. Chen, X. Gao, M. Cui, L. Ma, X. Zheng, B. Guan, and X. Ma, “Surgical treatment of diffuse and multi-lobes involved glioma with the assistance of a multimodal technique,” *Scientific Reports*, vol. 12, no. 1, p. 3343, 2022.
- [228] C. L. B. Kline, A. P. J. Van den Heuvel, J. E. Allen, V. V. Prabhu, D. T. Dicker, and W. S. El-Deiry, “Onc201 kills solid tumor cells by triggering an integrated stress response dependent on atf4 activation by specific eif2 α kinases,” *Science signaling*, vol. 9, no. 415, pp. ra18–ra18, 2016.
- [229] E. R. Jackson, R. J. Duchatel, D. E. Staudt, M. L. Persson, A. Mannan, S. Yadavilli, S. Parackal, S. Game, W. C. Chong, W. S. N. Jayasekara, *et al.*, “Onc201 in combination with paxalisib for the treatment of h3k27-altered diffuse midline glioma,” *Cancer research*, pp. OF1–OF17, 2023.
- [230] K. A. Zukotynski, S. Vajapeyam, F. H. Fahey, M. Kocak, D. Brown, K. I. Ricci, A. Onar-Thomas, M. Fouladi, and T. Y. Poussaint, “Correlation of 18f-fdg pet and mri apparent diffusion coefficient histogram metrics with survival in diffuse intrinsic pontine glioma: a report from the pediatric brain tumor consortium,” *Journal of Nuclear Medicine*, vol. 58, no. 8, pp. 1264–1269, 2017.
- [231] G. Morana, D. Tortora, G. Bottoni, M. Puntoni, G. Piatelli, F. Garibotto, S. Barra, F. Giannelli, A. Cistaro, M. Severino, *et al.*, “Correlation of multimodal 18f-dopa pet and conventional mri with treatment response and survival in children with diffuse intrinsic pontine gliomas,” *Theranostics*, vol. 10, no. 26, p. 11881, 2020.
- [232] J. E. Allen, G. Kringsfeld, P. A. Mayes, L. Patel, D. T. Dicker, A. S. Patel, N. G. Dolloff, E. Messaris, K. A. Scata, W. Wang, *et al.*, “Dual inactivation of akt and erk by tic10 signals foxo3a nuclear translocation, trail gene induction, and potent antitumor effects,” *Science translational medicine*, vol. 5, no. 171, pp. 171ra17–171ra17, 2013.
- [233] V. V. Prabhu, J. E. Allen, D. T. Dicker, and W. S. El-Deiry, “Small-molecule onc201/tic10 targets chemotherapy-resistant colorectal cancer stem-like cells in an akt/foxo3a/trail-dependent manner,” *Cancer research*, vol. 75, no. 7, pp. 1423–1432, 2015.

- [234] M. N. Stein, J. R. Bertino, H. L. Kaufman, T. Mayer, R. Moss, A. Silk, N. Chan, J. Malhotra, L. Rodriguez, J. Aisner, *et al.*, “First-in-human clinical trial of oral onc201 in patients with refractory solid tumors,” *Clinical cancer research*, vol. 23, no. 15, pp. 4163–4169, 2017.
- [235] J. Ishizawa, K. Kojima, D. Chachad, P. Ruvolo, V. Ruvolo, R. O. Jacamo, G. Borthakur, H. Mu, Z. Zeng, Y. Tabe, *et al.*, “Atf4 induction through an atypical integrated stress response to onc201 triggers p53-independent apoptosis in hematological malignancies,” *Science signaling*, vol. 9, no. 415, pp. ra17–ra17, 2016.
- [236] S. Venneti, A. R. Kawakibi, S. Ji, S. M. Waszak, S. R. Sweha, M. Mota, M. Pun, A. Deogharkar, C. Chung, R. S. Tarapore, *et al.*, “Clinical efficacy of onc201 in h3k27m-mutant diffuse midline gliomas is driven by disruption of integrated metabolic and epigenetic pathways,” *Cancer Discovery*, 2023.
- [237] P. Fontaine, V. Andrearczyk, V. Oreiller, D. Ablar, J. Castelli, O. Acosta, and A. Depeursinge, “Cleaning radiotherapy contours for radiomics studies, is it worth it? a head and neck cancer study,” *Clinical and Translational Radiation Oncology*, vol. 33, pp. 153–158, 2022.
- [238] Pyradiomics, “Pyradiomics: Radiomic features,” 2020. [Accessed 2020-02-19].
- [239] A. Zwanenburg, S. Leger, M. Vallières, and S. Löck, “Image biomarker standardisation initiative - feature definitions,” *arXiv*, 2016.