



Article

Deciphering Machine Learning Decisions to Distinguish between Posterior Fossa Tumor Types Using MRI Features: What Do the Data Tell Us?

Toygar Tanyel ¹, Chandran Nadarajan ², Nguyen Minh Duc ³ and Bilgin Keserci ^{4,*}

¹ Department of Computer Engineering, Yildiz Technical University, Istanbul 34349, Türkiye; toygar.tanyel@std.yildiz.edu.tr

² Department of Radiology, Gleneagles Hospital Kota Kinabalu, Kota Kinabalu 88100, Sabah, Malaysia; nadarajan.chandran@gleneaglesdr.my

³ Department of Radiology, Pham Ngoc Thach University of Medicine, Ho Chi Minh City 700000, Vietnam; bsnguyenminhduc@pnt.edu.vn

⁴ Department of Biomedical Engineering, Yildiz Technical University, Istanbul 34349, Türkiye

* Correspondence: bushido.keserci@gmail.com

Simple Summary: This paper focuses on interpreting machine learning (ML) models' decisions in medical diagnoses, specifically for four types of posterior fossa tumors in pediatric patients. The proposed methodology involves using kernel density estimations with Gaussian distributions to analyze individual MRI features, assess their relationships, and comprehensively study ML model behavior. The study demonstrates that employing a simplified approach in the absence of large datasets can lead to more pronounced and explainable outcomes. Furthermore, the pre-analysis results consistently align with the outputs of ML models and existing clinical findings. By bridging the knowledge gap between ML and medical outcomes, this research contributes to a better understanding of ML-based diagnoses for pediatric brain tumors.

Abstract: Machine learning (ML) models have become capable of making critical decisions on our behalf. Nevertheless, due to complexity of these models, interpreting their decisions can be challenging, and humans cannot always control them. This paper provides explanations of decisions made by ML models in diagnosing four types of posterior fossa tumors: medulloblastoma, ependymoma, pilocytic astrocytoma, and brainstem glioma. The proposed methodology involves data analysis using kernel density estimations with Gaussian distributions to examine individual MRI features, conducting an analysis on the relationships between these features, and performing a comprehensive analysis of ML model behavior. This approach offers a simple yet informative and reliable means of identifying and validating distinguishable MRI features for the diagnosis of pediatric brain tumors. By presenting a comprehensive analysis of the responses of the four pediatric tumor types to each other and to ML models in a single source, this study aims to bridge the knowledge gap in the existing literature concerning the relationship between ML and medical outcomes. The results highlight that employing a simplistic approach in the absence of very large datasets leads to significantly more pronounced and explainable outcomes, as expected. Additionally, the study also demonstrates that the pre-analysis results consistently align with the outputs of the ML models and the clinical findings reported in the existing literature.

Keywords: posterior fossa pediatric brain tumors; magnetic resonance imaging; machine learning; exploratory data analysis; kernel density estimation



Citation: Tanyel, T.; Nadarajan, C.; Duc, N.M.; Keserci, B. Deciphering Machine Learning Decisions to Distinguish between Posterior Fossa Tumor Types Using MRI Features: What Do the Data Tell Us? *Cancers* **2023**, *15*, 4015. <https://doi.org/10.3390/cancers15164015>

Academic Editor: Sam Payabvash

Received: 25 June 2023

Revised: 22 July 2023

Accepted: 2 August 2023

Published: 8 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Brain tumors are the most prevalent type of childhood cancer, comprising over a quarter of all cases. Among these tumors, 60–70% arise in the posterior fossa (PF), with

medulloblastoma (MB), ependymoma (EP), pilocytic astrocytoma (PA), and brainstem glioma (BG) being the most common types in children. These tumors can negatively impact mental and physical development.

Clinical information from radiological interpretations and the histopathological features of tumors plays a crucial role in diagnosing, prognosticating, and treating PF tumors in children. Histopathological evaluation, which is necessary for the initial diagnosis, helps to evaluate patient prognosis and direct clinical and therapeutic management. It remains the gold standard in differentiating PF tumors [1,2]. Although biopsies of different PF brain tumors can reveal distinct visual characteristics, they carry significant risks of morbidity and mortality, in addition to being expensive. Recent progress in characterizing tumor subtypes based on cross-sectional diagnostic imaging indicates that it can help to predict differential survival and responses to treatment. This development is particularly promising for future treatment stratification in PF tumors. Hence, developing a novel non-invasive diagnostic tool is essential in classifying tumors based on type and grade and aiding in planning treatment.

Magnetic resonance imaging (MRI) is currently the most preferred non-invasive method. It offers high intrinsic soft-tissue contrast without the risk of ionizing radiation. Conventional MRI protocols, including T1-weighted (T1W), T2-weighted (T2W), and fluid-attenuated inversion recovery (FLAIR) MRI sequences, have shown promising results in differentiating types of PF tumors in children [3–21]. Additionally, diffusion-weighted imaging (DWI) with apparent diffusion coefficient (ADC) maps allows the assessment of physiological features to discriminate between low- and high-grade tumors and their different subtypes [22–37].

While numerous advancements have been made, the diagnosis and prognosis of specific tumor matches still present significant challenges due to the voxel-wise overlap [23,27,38]. The classification process necessitates the inclusion of a tumor's molecular profile as a critical variable to predict the diverse biological behaviors of entities that exhibit histological similarities or even indistinguishability [2]. An extensive exploration of tumor classifications has been conducted using MRI in the literature. Nevertheless, accurately distinguishing between these tumor types remains an active area of research [20,39–42]. The differentiation between MB and EP is of the utmost importance, considering the distinct treatment planning required for each, underscoring the significance of their accurate diagnosis in numerous cases.

Artificial intelligence (AI) applications in pediatric brain tumor research are currently not well documented when compared to the available literature for adults. Challenges arise due to the unique pathology of pediatric cases and limitations in available data, which hinder the development of AI applications specifically tailored to children [43]. While there is growing interest in utilizing AI for pediatric brain tumor classification [44–55], the integration of AI into clinical workflows encounters significant obstacles beyond mere classification. One major challenge is the limited interpretability of many AI methods; creating a “black-box” model raises concerns among clinicians and patients. To address this issue, our research aims to enhance the interpretability of ML models' outcomes, which are frequently either blindly accepted or disregarded due to their black-box nature. To the best of our knowledge, there is a lack of literature specifically focusing on the issue of reasoning and explainability [56].

This study had two main objectives, aiming to bridge the gaps between ML outcomes and medical knowledge. Firstly, it sought to investigate the significance of clinical MRI features in classifying pediatric PF tumors (MB, EP, PA, and BG) through exploratory data analysis (EDA). Secondly, it aimed to offer explanations for the ML outcomes by leveraging the insights gained from the data exploration (Figure 1).

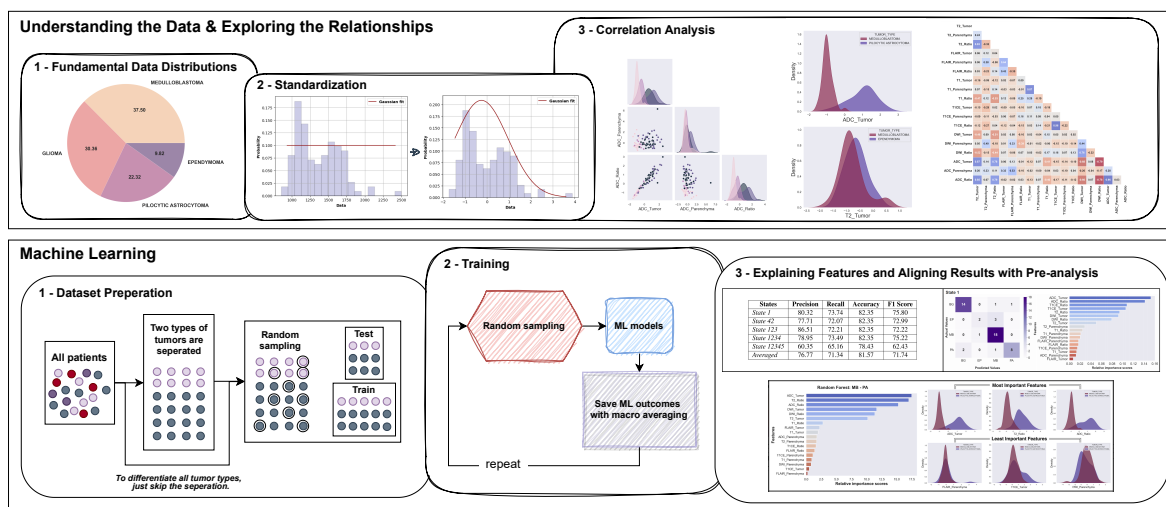


Figure 1. A flowchart depicting the proposed analysis for the classification of pediatric PF tumors: standardization of the dataset, pairwise feature analysis to examine various features of PF tumor types, and aligning interpretations of pre-analysis with ML models’ outcomes.

2. Materials and Methods

2.1. Ethics Statement and Patient Characteristics

This prospective study (Ref: 632 QD-NĐ2 dated 12 May 2019) was conducted in both Radiology and Neurosurgery departments and approved by the Institutional Review Board in accordance with the 1964 Helsinki declaration. Prior to the MRI procedure, written informed consent was obtained from the authorized guardians of the patients. The study included 112 pediatric patients diagnosed with PF tumors, including 42 with MB, 11 with EP, 25 with PA, and 34 with BG. All BG patients were confirmed based on full agreement between neuroradiologists and neurosurgeons, while the remaining MB, EP, and PA patients underwent either surgery or biopsy for histopathological confirmation.

The demographics of the patient population were analyzed to gain insights into their age, gender, and weight distributions. The age statistics revealed a mean age of 6.55 years, with a median age of 6.0 years. The age range varied from a minimum of 0.6 years to a maximum of 15.0 years, reflecting the diversity within our cohort. Regarding gender, we observed a greater representation of males, with a count of 68, compared to females, with a count of 44. The mean weight was calculated to be 22.54 kg, with a median weight of 20.5 kg. The range of weights varied from a minimum of 3 kg to a maximum of 48 kg.

In-depth patient demographics can be found in the accompanying Table 1, which provides a comprehensive overview of the study population. Table 1 includes detailed information on gender, age, and weight for the patients.

Table 1. Patient demographics.

Tumor Type	Gender	Count	Age			Weight		
			Mean ± Std	Min	Max	Mean ± Std	Min	Max
Medulloblastoma	Girl	16	7.16 ± 3.74	0.6	13	20.81 ± 8.53	8	35
	Boy	26	6.77 ± 3.40	1	13	21.19 ± 8.51	9	40
Ependymoma	Girl	3	5.67 ± 0.58	5	6	20.33 ± 4.16	17	25
	Boy	8	4.00 ± 3.42	1	11	18.38 ± 12.35	9	45
Pilocytic Astrocytoma	Girl	11	8.18 ± 3.66	3	14	25.18 ± 12.16	9	48
	Boy	14	5.79 ± 3.24	1	12	24.07 ± 10.22	10	44
Brainstem Glioma	Girl	14	6.43 ± 3.69	1	15	22.86 ± 11.81	3	47
	Boy	20	6.65 ± 2.85	3	14	24.95 ± 7.82	15	48

2.2. Data Acquisition and Assessment of MRI Features

The MRI protocol was performed in the supine position using a 1.5 Tesla MRI scanner (Philips, Best, The Netherlands) and included T1W, T2W, FLAIR, DWI (b values: 0 and 1000) with ADC, and T1 contrast-enhanced (T1CE) sequences with macrocyclic gadolinium-based contrast enhancement (0.1 mL/kg Gadovist, Bayer, Germany or 0.2 mL/kg Dotarem, Guerbet, France).

MR images of all patients were imported into the Medical Imaging Interaction Toolkit, developed by the German Cancer Research Center's Division of Medical Image Computing in Heidelberg, Germany. The radiologists precisely identified the slice in which the largest diameter of the PF tumor was present. For each patient, ROIs corresponding to the posterior fossa tumors and normal-appearing parenchyma were manually delineated on the T1W, T2W, FLAIR, DWI, and ADC images. These delineations were based on the consensus reached by two expert radiologists with over 10 years of experience in interpreting neuro MR images. An example of ROI delineation on a T2W MRI is provided in Figure 2. For additional ROI delineations of other sequences, please refer to Figure S1 in the Supplementary File S1.

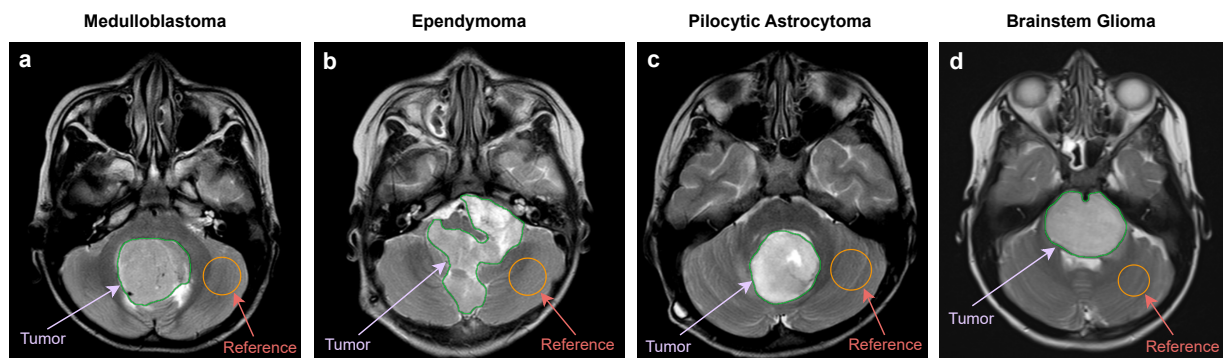


Figure 2. Example of ROI delineation on a T2W MRI. (a) MB: 8 years old, boy. (b) EP: 3 years old, boy. (c) PA: 7 years old, girl. (d) BG: 6 years old, girl.

The following MRI features were evaluated: signal intensities (SIs) of T2, T1, FLAIR, T1CE, DWI, and ADC. The ratio of each MRI feature was calculated as the quotient of the tumor's SI and the SI of the normal-appearing parenchyma ($\text{Ratio} = \frac{\text{Tumor Feature}}{\text{Parenchyma Feature}}$). Additionally, ADC values were quantified for both the PF tumor and parenchyma regions using the MR Diffusion tool available in the Philips Intellispace Portal, version 11 (Philips, Best, The Netherlands).

2.3. Exploratory Data Analysis

The quality of a dataset has a direct impact on the effectiveness of the trained model. Therefore, EDA plays a crucial role in understanding the data by revealing its inherent structure, identifying anomalies and outliers, extracting significant features, and facilitating the appropriate ML models to establish correlations between MRI feature characteristics and the various types of pediatric PF tumors.

In this study, we performed an exploratory analysis using kernel density estimations (KDE) with Gaussian distributions, focusing on the MRI features. The proposed analysis consisted of three parts: standardization, data analysis involving visualization of the distributions of each MRI feature, as well as exploring relationships between different features, and analyzing the ML models' outcomes through the extracted knowledge from EDA. All figures were generated using the Matplotlib package (version 3.5.2) in Python.

2.4. Standardization

The patient dataset underwent a standardization process known as Z-score normalization. This process was carried out using the Python programming language, specifically Python version 3.9.13, along with the Scikit-Learn library version 1.0.2.

To perform the standardization, the StandardScaler function from Scikit-Learn was utilized. This function ensured that numerical attributes within the patient dataset were transformed into a standardized format. It achieved this by subtracting the mean and scaling the values to have a unit variance.

The StandardScaler function normalizes each feature individually, meaning that each column/feature/variable in the input matrix X will have a mean (μ) of 0 and a standard deviation (σ) of 1. The normalization is accomplished using the formula $z = \frac{x_i - \mu}{\sigma}$, where x_i represents the value of a specific feature for a patient.

2.5. Pairwise Feature Analysis

The pairplot function in the Seaborn Python package (version 0.11.2) enables the visualization of the pairwise relationships between variables in a dataset. Numerical variables are split into a single row on the y-axis and a single column on the x-axis by default. The position of one variable on the vertical or horizontal axis indicates its correlation with another variable in the same row of data. The relationship between the MRI features was further examined through Pearson's correlation coefficients, calculated using the default corr() function in the pandas dataframe.

2.6. Revealing Distribution Differences of Patients between Tumor Types

To effectively illustrate the distinctions among the four PF tumor types, we utilized the kdeplot and pairplot functions from Seaborn as necessary. Additionally, we assigned a hue parameter to represent the tumor type, thereby facilitating a semantic mapping. This assignment transforms the default marginal plot into a layered KDE, which helps to address the challenge of reconstructing the density function f using an independent, identically distributed (iid) sample x_1, x_2, \dots, x_n from the respective probability distribution.

The generalized estimate used in plotting can be expressed as follows:

$$\hat{f}(x) = \frac{1}{nh^d} \sum_{i=1}^n k\left(\frac{x - x_i}{h}\right), \quad (1)$$

where h is a bandwidth parameter, and the kernel is commonly a Gaussian,

$$k(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right). \quad (2)$$

2.7. Machine Learning

We employed eight ML models, including support vector machine (SVM), linear support vector machine (LSVM), logistic regression (LR), a random forest classifier (RF), a decision tree classifier (DT), a gradient boosting classifier (GBM), a catboost classifier (CB), and an extreme gradient boosting classifier (XGB), to assess the consistency of our interpretations of the raw data with the outcomes. CB and XGB were obtained from their respective libraries (CatBoost version 1.1.1, XGBoost version 1.5.1), while other models were obtained from the Scikit-Learn library.

To ensure methodological consistency, we utilized the default versions of all ML models, as our primary objective was not to maximize the classification scores. It is important to acknowledge that tuning the model parameters could potentially lead to improved results. However, considering the limited data size, the presence of rare tumor types, and the absence of an external dataset from another hospital, such parameter adjustments carried a significant risk of overfitting on our data. In order to mitigate this risk and uphold the credibility of our findings, we chose to adhere to the default model configurations throughout the analysis. This decision safeguarded the integrity of our research and ensured the validity of our conclusions.

Tree-based models, such as RF, DT, GBM, CB, and XGB, are commonly utilized in ML. However, they can be prone to overfitting when the trees are deep and have a large number of features. To address this issue, RF, which is a bagging model, generates a set of decision trees by training on different data samples or subsets of features. XGB, on the other hand,

is a sequential model that adopts a different approach to building decision trees. To ensure that our models did not have a bias towards certain features and generalized well, we conducted an analysis of the prioritization and proportional distribution of features used by the RF and XGB models during prediction. This analysis strengthened our explanations of the models' high performance and accuracy in predicting outcomes.

To ensure the reliability of our ML models, particularly with a small dataset, we conducted five runs using stratified random sampling based on tumor type with 55% train and 45% test patients. We used random states to obtain samplings and preserve the train/test distributions for the reproducibility of the experiment. Ultimately, we calculated the averaged outcomes with their standard deviations.

The accuracy metric is not employed in the presentation of our results due to significant class imbalance in our dataset. Utilizing the accuracy metric could have led to misleading results. Instead, we relied on the precision, recall, and F1 score, computed based on the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), as fundamental evaluation metrics to assess the performance of our classifiers in both binary and multiclass classification tasks. Precision gauges the proportion of correctly predicted positive instances among all positive predictions, highlighting the accuracy of positive classifications. Conversely, recall assesses the proportion of positive instances that were correctly identified by the classifier, emphasizing the completeness of the positive predictions.

While high precision and high recall are typically desirable, we were aware of the potential trade-off between these two metrics in certain scenarios. To gain a comprehensive understanding of our classifier's effectiveness, we utilized the F1 score, which harmoniously considers both precision and recall.

To ensure a precise interpretation of the ML results, we chose not to equalize the labels. Instead, we utilized the macro precision, macro recall, and macro F1 score metrics to ensure that all labels contributed equally to the results. This approach allowed us to assess the classifier's performance while considering the impact of varying patient counts across different labels.

The validation metrics used in ML are as follows:

$$\text{Macro Precision} = \frac{1}{n} \sum_{i=1}^n \frac{TP_i}{TP_i + FP_i} \quad (3)$$

$$\text{Macro Recall} = \frac{1}{n} \sum_{i=1}^n \frac{TP_i}{TP_i + FN_i} \quad (4)$$

$$\text{Macro F1 Score} = \frac{1}{n} \sum_{i=1}^n \frac{2 \times TP_i}{2 \times TP_i + FP_i + FN_i} \quad (5)$$

where n represents the total number of classes.

2.8. Statistical Analysis

Statistical analysis was performed using the SPSS software (version 25.0, 64-bit edition, IBM Corp., Armonk, NY, USA). A two-sided p -value of <0.05 was considered statistically significant. The statistical summary of the variability of ML outcomes is presented in the mean \pm standard deviation format.

2.9. Hardware Requirements for Machine Learning

Designing an ML pipeline with the current number of patients and their tabular data does not require significant computational power. The entire machine learning system was developed utilizing a system equipped with an Apple M1 chip CPU and a memory capacity of 16 GB, namely the Hynix LPDDR4.

3. Results

3.1. Single MRI Feature Analysis

Our feature analysis, which utilized KDE on the standardized distributions presented in Figure 3a–f, yielded several valuable insights.

T2_Tumor features possess distributions that are expected to differentiate PA from EP and MB but cannot differentiate between MB and EP or PA and BG (Figure 3a). Moreover, the T2_Ratio might aid in distinguishing between MB and EP, as well as PA and BG.

The distributions of FLAIR_Tumor and FLAIR_Ratio generate notably different distributions (Figure 3b), even though the Ratio feature is mathematically dependent on the Tumor feature. The FLAIR distributions might be effective in distinguishing between MB and EP, as demonstrated by FLAIR_Tumor, which exhibits a broad EP and a narrow MB distribution. Furthermore, the FLAIR_Ratio exhibits two distinct and narrow Gaussian distributions, which also might aid in distinguishing between MB and EP. In contrast, the other scenarios do not present any discriminative characteristics.

The DWI characteristics (Figure 3c) demonstrate distributions that allow differentiation between MB and PA. Additionally, although to a lesser extent, discrete distributions can be observed in the differentiation between MB and BG, as well as between EP and PA. On the other hand, despite their high distinctive distributions overall, DWI_Ratio features are not expected to be effective in distinguishing between PA and BG due to significant overlap.

ADC (Figure 3d) demonstrates separate distributions in distinguishing each tumor pair, with the highest distinction observed between MB and PA and the least between PA and BG. When considering tumors as a whole rather than in pairs, ADC and DWI present the most distinct distributions for all tumor types. ADC shows highly distinct distributions for each tumor, with DWI following closely behind.

The T1 features, as shown in Figure 3e, do not demonstrate any distinctive distributions that can effectively differentiate between different tumor scenarios. However, the T1_Ratio appears to be a critical factor in distinguishing PA from other types of tumors. In addition, T1CE presents important distinct distributions for all other tumor matches with BG, as depicted in Figure 3f.

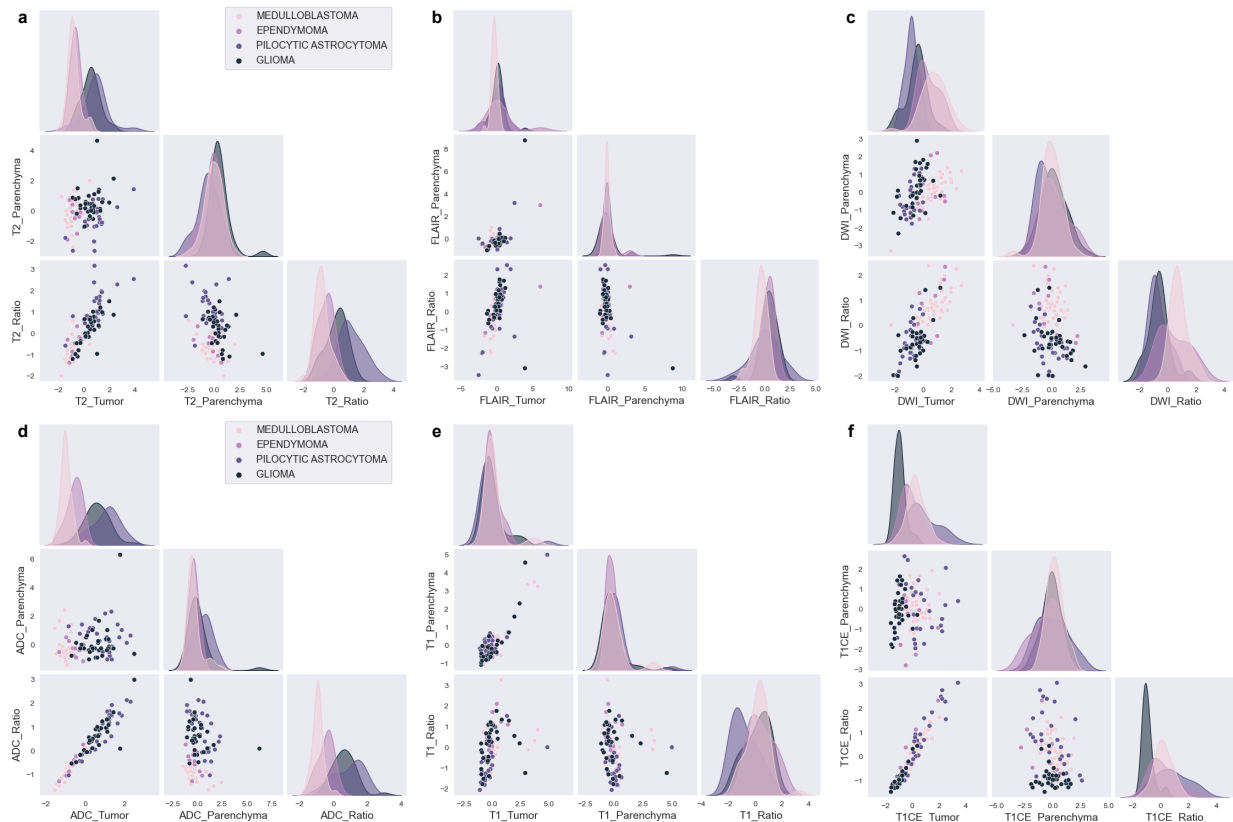


Figure 3. Kernel density estimations with Gaussian distributions of MRI features for PF tumors. (a) T2, (b) FLAIR, (c) DWI, (d) ADC, (e) T1, and (f) T1CE.

3.2. Pairwise Analysis of MRI Features

The scatter correlation plots (Figure S2 in the Supplementary File S1) and Pearson’s correlation coefficients (Figure 4) illustrate varying degrees of correlation between the MRI features and tumor types. Notably, MB exhibits clustered shapes, while PA appears scattered in most cases, and BG and EP show dispersed and uncertain distributions. Outlier patients with correlated features were identified, and some features exhibited no correlations with the tumor types.

The results of the Pearson’s correlation analysis indicated that the T2 and ADC features, with complex distributions compared to other features, exhibited significant positive correlations, particularly T2_Tumor and ADC_Tumor ($r = 0.87, p < 0.0001$), T2_Tumor and ADC_Ratio ($r = 0.85, p < 0.0001$), T2_Ratio and ADC_Tumor ($r = 0.78, p < 0.0001$), and T2_Ratio and ADC_Ratio ($r = 0.79, p < 0.0001$). Conversely, significant negative correlations were observed between the T2 and DWI features, as well as between the DWI and ADC features, namely T2_Tumor and DWI_Tumor ($r = -0.46, p < 0.0001$), T2_Tumor and DWI_Ratio ($r = -0.52, p < 0.0001$), T2_Ratio and DWI_Tumor ($r = -0.51, p < 0.0001$), T2_Ratio and DWI_Ratio ($r = -0.44, p < 0.0001$), ADC_Tumor and DWI_Tumor ($r = -0.68, p < 0.0001$), ADC_Tumor and DWI_Ratio ($r = -0.79, p < 0.0001$), ADC_Ratio and DWI_Tumor ($r = -0.66, p < 0.0001$), and ADC_Ratio and DWI_Ratio ($r = -0.78, p < 0.0001$).

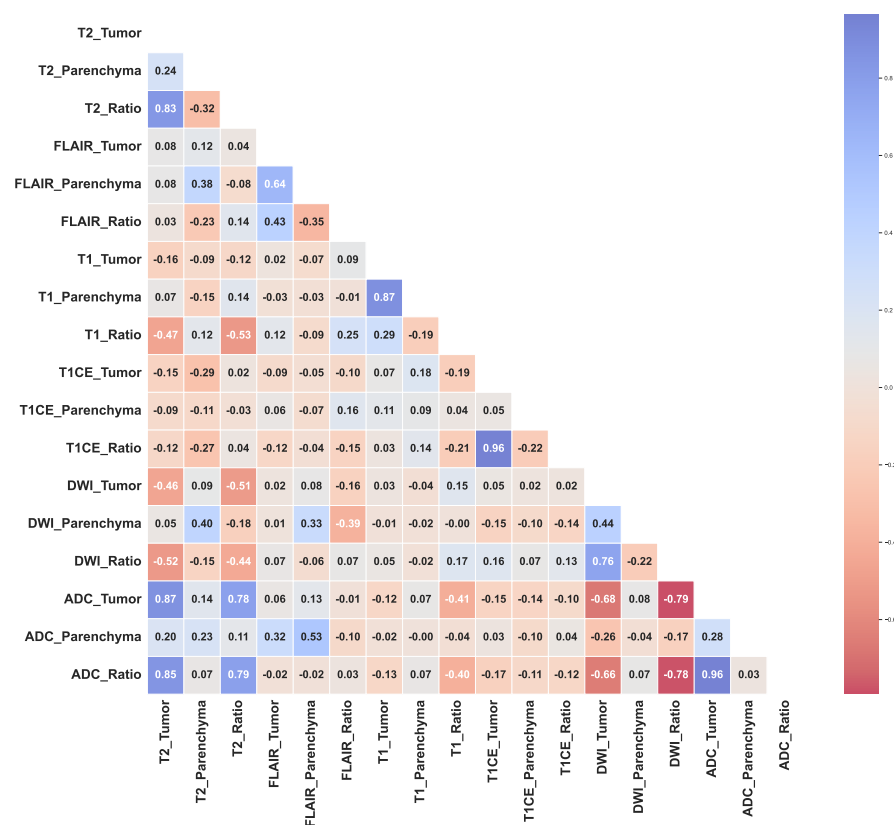


Figure 4. Pearson’s correlation coefficients between each MRI feature.

FLAIR_Tumor did not demonstrate a significant correlation with any other features (T2_Tumor ($r = 0.08, p = 0.39$), T1_Tumor ($r = 0.02, p = 0.84$), T1CE_Tumor ($r = -0.09, p = 0.34$), DWI_Tumor ($r = 0.02, p = 0.85$), and ADC_Tumor ($r = 0.06, p = 0.52$)), while FLAIR_Ratio could exhibit correlations in logarithmic or reduced dimensions (T1_Ratio ($r = 0.25, p = 0.008$)). Similar patterns to FLAIR were observed for T1_Tumor (T2_Tumor ($r = -0.16, p = 0.08$), T1CE_Tumor ($r = 0.07, p = 0.45$), DWI_Tumor ($r = 0.03, p = 0.76$), and ADC_Tumor ($r = -0.12, p = 0.19$)), and T1_Ratio (T2_Tumor ($r = -0.47, p < 0.0001$), T2_Ratio ($r = -0.53, p < 0.0001$), T1CE_Tumor ($r = -0.19, p = 0.045$), T1CE_Ratio ($r = -0.21, p = 0.03$), ADC_Tumor ($r = 0.41, p < 0.0001$) and ADC_Ratio ($r = 0.40, p < 0.0001$)), empha-

sizing the importance of using a ratio computed with reference to parenchyma. In contrast, T1CE_Tumor and T1CE_Ratio showed dispersed distributions, with non-linear patterns that could be observed for certain tumor types.

3.3. Findings from Machine Learning

The ML procedure involved analyzing feature importance scores, the test scores of eight ML models (Tables S1–S7 in the Supplementary File S2), and confusion matrices to assess the accuracy and reliability of the results. We trained the models on all possible tumor pairs to identify unique features for each case, and the most favorable outcomes are summarized in Table 2. Additionally, we conducted a comprehensive analysis of the feature importance scores for all four tumor types, providing further insights into their distinguishing characteristics.

We focused on the RF and XGB models since RF delivered the best scores for the classification of all tumors, while XGB possesses a distinct tree structure compared to RF, allowing us to explore and compare the variations in the ML models' outcomes. Although various ML models could have been employed for this analysis, we specifically chose XGB and RF to illustrate how the methods' structures differ in generating importance scores and to present a clear and concise analysis.

Notably, as shown in Figure 5a, the FLAIR_Ratio was identified as the most discriminating feature in distinguishing between MB and EP in both the RF and XGB models, followed by the ADC_Ratio. However, the two models relied on different features for decision-making. Therefore, relying solely on the analysis presented in Figure 5 may not be sufficient for model comparison, as they prioritize different features. The performance evaluation of both models showed that the RF model, which prioritized diffusion features (4 out of top 5, 65.38%), demonstrated greater accuracy in feature selection compared to the XGB model (2 out of top 5, 60.08%). Therefore, the features highlighted by the RF model should be considered more significant in distinguishing between MB and EP.

Table 2. Best test scores for each case from evaluation of 8 different ML models.

	Best Model	Precision	Recall	F1 Score
MB-EP	LR	70.65 ± 4.55	68.21 ± 7.86	67.70 ± 6.19
MB-PA	LSVM	97.46 ± 1.57	97.28 ± 1.70	97.28 ± 1.52
MB-BG	LR	94.94 ± 2.38	94.77 ± 2.50	94.81 ± 2.42
EP-PA	LR	95.90 ± 4.17	96.33 ± 4.11	95.80 ± 3.85
EP-BG	LR	91.46 ± 7.58	87.50 ± 10.94	87.48 ± 9.73
PA-BG	CB	89.95 ± 5.97	88.69 ± 6.86	89.04 ± 6.50
MB-EP-PA-BG	RF	76.77 ± 9.78	71.34 ± 3.53	71.74 ± 5.41

In differentiating between MB and PA, the RF model showed a more dispersed reliance on various features, whereas the XGB model heavily relied on ADC_Tumor (Figure 5b). The study suggests that DWI features play an important role in distinguishing between these tumors, with T2 features being crucial in the decision-making process of the RF model, leading to a higher F1 score (RF: 93.81%, XGB: 90.15%).

To differentiate between MB and BG, our results indicated that the XGB model heavily relied on the ADC_Tumor feature (~80%), while the RF model utilized a more diverse set of features, such as ADC, DWI, T1CE, and T2 (Figure 5c). Surprisingly, both models performed equally well in this scenario, with an F1 score of 93.62%. Therefore, our results suggest that the ADC_Tumor feature is crucial in distinguishing MB from BG.

In differentiating EP from BG, the ML models were mainly impacted by the ADC_Tumor feature, with a supportive influence of the T1CE_Ratio and T1CE_Tumor features (Figure 5d). The RF model was also influenced by T2_Tumor, which might have led to a slight decrease in the overall F1 score (RF: 69.96%, XGB: 73.34%). As there was significant feature overlap between EP and BG, the performance scores for this classification task were lower compared to other tumor pairs, except MB and EP.

In distinguishing EP from PA, T2 features provided significant discriminative power (Figure 5e). However, the ADC_Tumor and ADC_Ratio features were found to be the leading contributors to the F1 score of 92.18% for the RF model, while the XGB model achieved a score of 81.48% due to its strong dependency on T2.

To differentiate PA from BG, both ML models heavily relied on T1CE features in their decision-making processes, with the T2_Ratio also providing discriminative power (Figure 5f). The XGB model outperformed the RF model, achieving a higher F1 score of 89.01% compared to 87.25%.

Additionally, the LR model achieved the highest F1 score in the MB-EP case and emerged as the dominant model in the MB-BG, EP-PA, and EP-BG cases. Furthermore, the LSVM outperformed other models in distinguishing between MB and PA. For the PA-BG case, the CB model attained high scores.

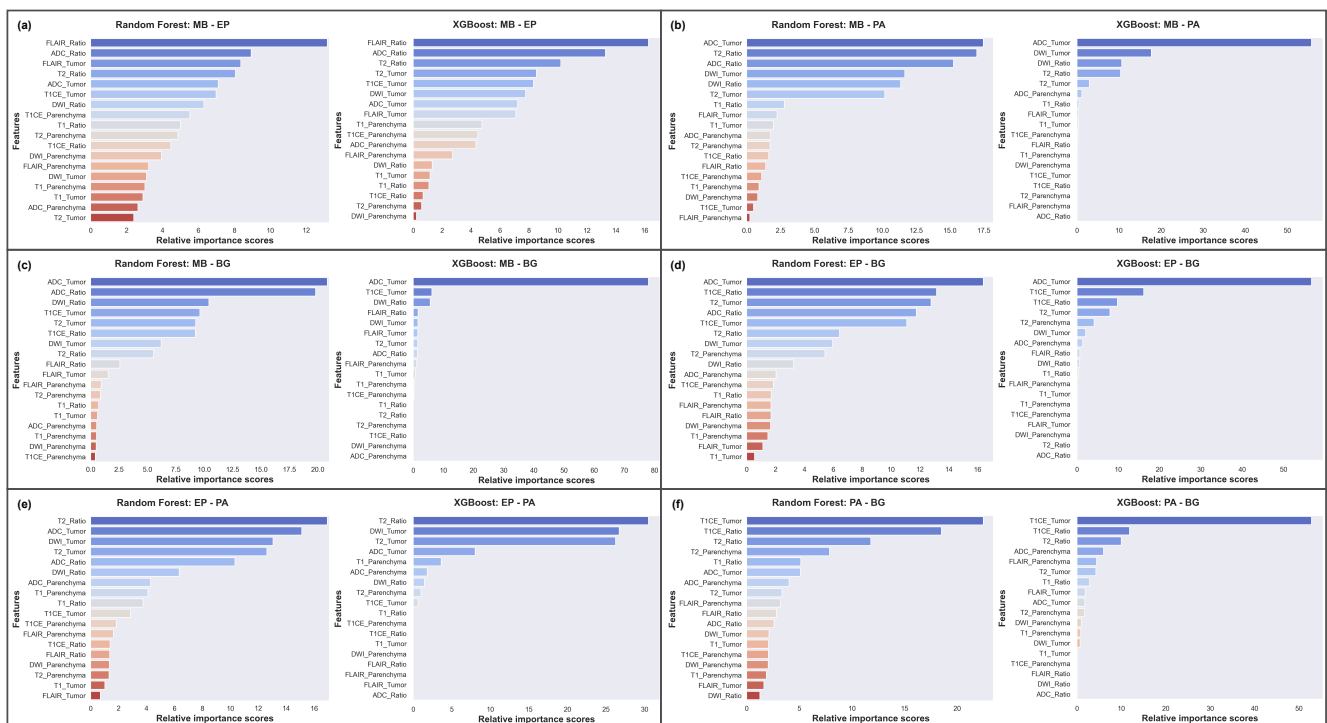


Figure 5. Averaged feature importance scores generated by RF and XGB models for behavior comparison. (a) MB-EP; (b) MB-PA; (c) MB-BG; (d) EP-BG; (e) EP-PA; (f) PA-BG.

The analysis revealed varying degrees of feature importance in differentiating between the four tumor types in the MB, EP, BG, and PA classification task. Among the models, the RF model achieved the highest F1 score of 71.74%, outperforming the other models. Figure 6c illustrates the significant role of ADC features in overall differentiation, followed by the T1CE and T2 features. The DWI and FLAIR features also contributed to the discriminative power, albeit to a lesser extent.

We also identified the most challenging discrimination task, which involved distinguishing between MB and EP (Figure 6a), and the easiest discrimination task, which involved distinguishing between MB and PA (Figure 6b). In the challenging classification problem of MB and EP, the Gaussian distributions of the best distinguishing features were found to overlap significantly, while, in the easiest one, MB and PA, the distributions of the best features did not overlap at all. Conversely, the least important features overlapped completely in every scenario.

The impact of stratified random sampling on the feature selection and performance of the RF model was examined and the findings are as given below (Figure S3 in the Supplementary File S1).

- In State 1, the model misclassified nine patients, including two with a BG tumor, three with an EP tumor, one with an MB tumor, and three with a PA tumor. The most informative features for this classification were ADC_Tumor, ADC_Ratio, and T1CE_Ratio.
- In State 2, the model performed slightly better in predicting BG and could distinguish all MB from other types. However, it misclassified two more PA patients, using ADC_Ratio, ADC_Tumor, and T2_Ratio as the most significant features.
- In State 3, the model could distinguish almost all PA test patients except one. However, it missed one BG, which was previously predicted as EP in State 2.
- In State 4, the model was unable to differentiate three BG, three EP, and three PA test patients from other types.
- In State 5, the model attributed the highest importance to the T1CE_Tumor feature, which led to the misclassification of all EP patients and four BG patients.

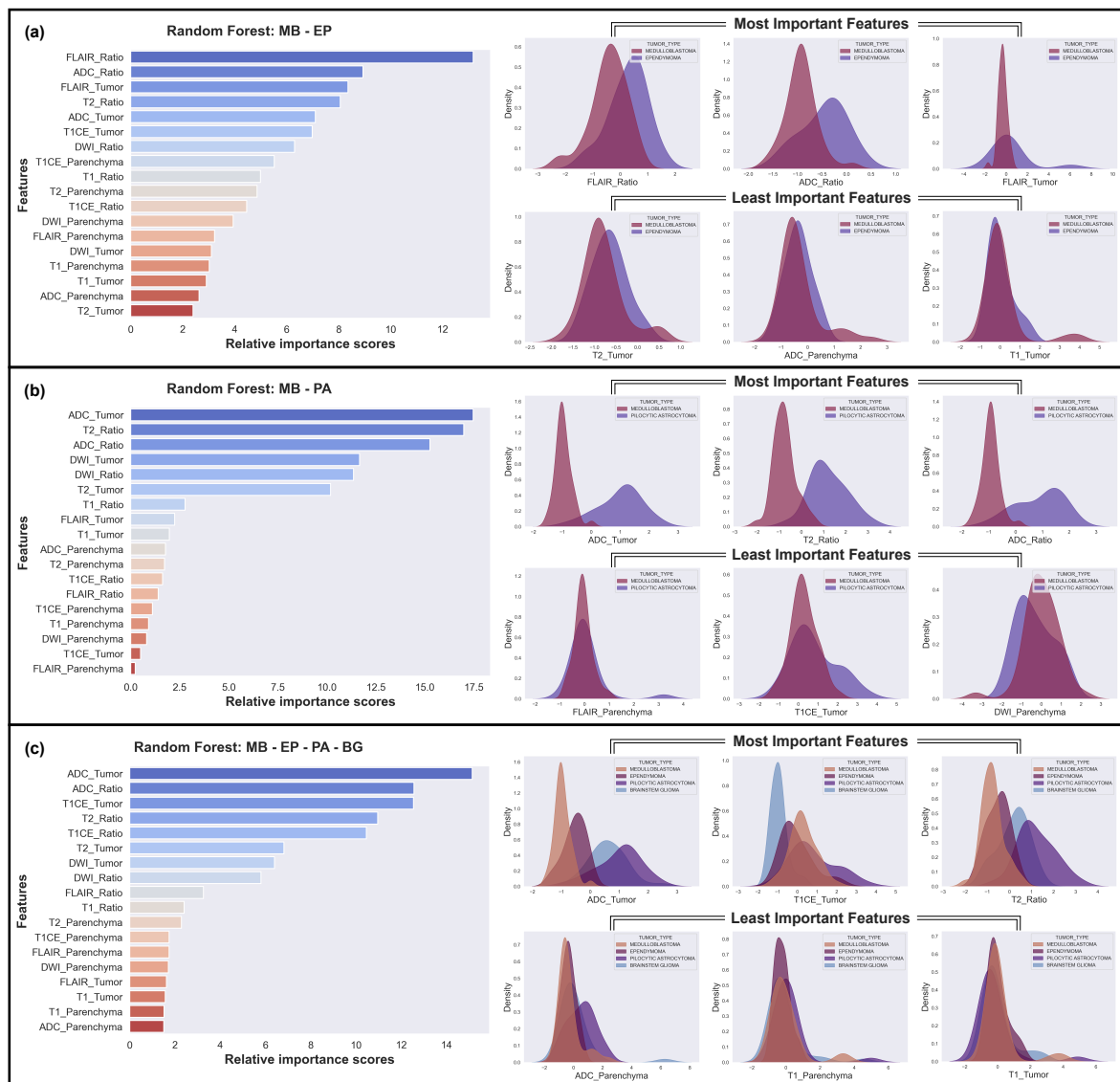


Figure 6. The three most and least effective features for the classifications using the Random Forest (RF) model. (a) The hardest case: MB vs. EP; (b) The easiest case: MB vs. PA; (c) Case of all tumor types.

4. Discussion

Pediatric brain tumors pose a significant clinical challenge due to the substantial degree of spatial heterogeneity in tumor characteristics. Tumors such as those arising

from the posterior fossa have a significant imaging feature overlap, leading to difficulty in differentiation, even among experts. The need to differentiate is important due to the different treatment options available for each of them. Thus, precise diagnosis and treatment are crucial in improving outcomes and enhancing quality of life.

Despite the significant advancements observed in AI and medical imaging, the dependability and accuracy of these approaches are profoundly influenced by the quality of the data, meticulous system design, and the comprehensive dissemination of transparent results. Therefore, we conducted a comprehensive and systematic analysis focusing on four distinct tumor types in pediatric brain tumor research. We employed an approach that integrated EDA to interpret ML outcomes, while the ML models provided additional insights into the underlying patterns and relationships among the MRI features.

This study was motivated by the idea that the feature distribution obtained from KDE can provide reliable estimates of ML results prior to the actual model training. The estimation provides insights into which features are the most effective and which features contribute negligibly to the ML models' decisions. To test this hypothesis, we conducted several pre-training analyses without relying on prior clinical knowledge and analyzed the feature distribution plots. In the present research, a thorough investigation of the diverse characteristics of pediatric PF tumor types was carried out through the utilization of Gaussian distributions, which can be observed in Figure 3. Through this analysis, a number of predictions have been drawn, indicating that certain features are likely to be highly effective in distinguishing particular tumor types, while others are deemed to have a limited impact on classification.

The single-feature analysis using Gaussian distributions, shown in Figure 3, revealed that some MRI features are effective in distinguishing specific pediatric PF tumor types, while others have minimal contributions towards classification. ADC and DWI features are the most effective in differentiating between tumor types, with clear differences in the distributions of these features for different tumors, whereas T1 and T1CE features are less effective in distinguishing between tumor types, although there are some differences in the distributions for different tumors. Moreover, T2 and FLAIR features show some differences in the distributions for different tumors, but these are less pronounced than for ADC and DWI.

Our analyses in the single-feature section are consistent with both the clinical and ML results in almost every instance, and we provide corresponding references in this section to validate our findings. Specifically, our analysis indicates that the T2_Tumor feature can effectively differentiate PA from EP and MB (Figure 5b–d), but cannot differentiate between MB and EP or PA and BG (Figure 5a–f). Remarkably, the incorporation of the hand-crafted feature T2_Ratio further enhances the effectiveness of T2 for tumor classification (Figure 5a,b,e,f). This is particularly evident in the differentiation between MB and EP, as well as PA and BG tumors (Figure 5a–f). Our findings also shed light on the potential of FLAIR features in distinguishing between different tumor types (Figures 3b and 5). The distributions of FLAIR_Tumor and FLAIR_Ratio exhibit notable differences, despite the lack of distributional disparities for parenchyma, which serves as a reference point. Specifically, FLAIR_Tumor shows a broad distribution for EP and a narrow distribution for MB, while FLAIR_Ratio displays two distinct and narrow Gaussian distributions. The ML results highlight that FLAIR features are useful in distinguishing between MB and EP tumors (Figure 5a), although the discriminative characteristics are not evident in the rest of the scenarios (Figure 5b–f). These findings are consistent with those of previous studies [10–12,15,17,19,20].

The results of our study also indicate that the DWI characteristics display distinct distributions that enable the differentiation of MB and PA (Figures 3c and 5b). While the distributions are less clear, there are still noticeable differences in the DWI characteristics when distinguishing between MB and BG, as well as between EP and PA (Figures 3c and 5c,e). However, we found that the DWI_Ratio features, despite having highly distinctive distributions overall, were not likely to be useful in distinguishing between PA and BG due to their significant overlap (Figures 3c and 5f). Moreover, our findings revealed that the ADC

had distinct distributions for each tumor pair, with the most noticeable distinction between MB and PA and the least between PA and BG (Figures 3d and 5a–f). When considering the distributions of all tumors collectively, rather than in pairs, ADC exhibited the most prominent differences among all tumor types (Figure 6c). The difference in diffusivity between various types of PF tumors is due to their cellular characteristics and arrangements, as well as the presence of cystic spaces within the tumor bulk. These results are in line with previous research findings [20,22–24,30–32].

While our study identified several features that can effectively differentiate between different types of PF tumors, not all features are equally informative. Some features may exhibit significant overlap between different tumor types, which can limit their usefulness in certain scenarios. For instance, our study demonstrated that T1_Tumor features did not exhibit any notably distinctive distributions in distinguishing between different tumor types (Figures 3e and 5a–f). However, it could be seen that T1_Ratio is a crucial factor in differentiating PA from other tumor types (Figure 5b,e,f). Additionally, T1CE displays notable distinctive distributions when differentiating all other tumor types from BG (Figures 3f and 5c,d,f).

Based on the pairwise analysis of the dataset, our findings suggest that MB exhibits a more distinct set of MRI features that are strongly correlated with the tumor type. Conversely, PA appears to be more heterogeneous in terms of its MRI features, and the MRI features associated with BG and EP may not be well defined. Furthermore, the positive correlation between both T2 and ADC features may reflect the diverse nature of these tumor types, with different subtypes exhibiting distinct MRI features. The negative correlation observed between DWI and ADC, as well as DWI and T2 features, may reflect differences in tumor cellularity and tissue microstructure. This finding may have important implications for treatment planning, particularly with regard to therapies that target the tumor microenvironment. The findings of this study offer significant insights into the correlations between tumor types and MRI features.

In an ML classification model, a feature's ability to distinguish between different classes, such as different tumor types, is determined by the degree of separation or overlap between the distributions of the feature values for each class. When the distributions are close together and have significant overlap, the feature is unlikely to provide much discriminative value and will have little impact on the classification decision. Conversely, when the distributions are far apart and have minimal overlap, the feature is more likely to provide discriminative value and will significantly impact the classification decision. Examining the distribution of tumor types across various features can help to identify potential biomarkers that may be useful for diagnostic or prognostic purposes.

Our preliminary analysis in this study agrees with the results obtained from the RF model, thus substantiating its decision-making process. However, there is a possibility of the ML model selecting a non-distinctive feature as the most critical factor, which is irrational. Therefore, it is imperative to provide an explanation for the model's decisions. To this end, we have proposed a methodology to elucidate the correlation between the KDE analysis and the averaged feature importance of the ML results, as presented in Figure 6, to bring clarity to this association.

The performance of ML models can be significantly influenced by the distribution of samples in the training and testing sets, even if the samples belong to the same class. To ensure more generalizable results, we utilized stratified random sampling to evaluate the dataset across five different distributions. The feature importance was then computed and averaged over the five distributions, as demonstrated in Figures 5 and 6. However, there was still a considerable degree of variability in the results for each distribution (Figure S3 in the Supplementary File S1). Thus, it is crucial to take into account the sample distribution when assessing ML models and to implement stratified random sampling to ensure robust and generalized outcomes.

We evaluated the performance of eight different ML models and determined that LR is suitable for binary classification tasks. However, in discriminating between all tumor

types simultaneously, RF outperformed all other models. To enhance the interpretability of the RF model's results, we aligned its feature importance values with the KDE predictions. To ensure reproducibility, we used five different random seeds and computed the mean of the resulting outputs. Comprehensive analysis of the ML results side by side revealed that no single model outperformed the others, as demonstrated in Table 2. Furthermore, the results differed depending on the models and data structure. To demonstrate this, we compared the RF and XGB models for PF tumor classification, as shown in Figure 5. This approach provided a clear understanding of how the models' behavior influenced the results, which was crucial for the PF tumor classification task.

Analyzing patient distributions can provide insights into subtypes with unusual patterns that ML models may not detect, leading to errors in calculating tumor characteristics. These outliers can reveal unique features that improve the reliability and accuracy of ML models for medical diagnosis and treatment. Clustering and explaining ML models with larger labeled datasets could enhance our understanding of the heterogeneity within patient subtypes in future studies.

In clinical practice, tumors are assessed based on location, the effect exerted by the tumor on the surrounding tissue, and tumor behavior, including the tendency to invade surrounding tissues or the presence of cystic components or calcification. Tumors with classic imaging features in pathognomonic locations can be identified even by novice radiologists with ease. However, distinguishing between tumors with similar characteristics and locations requires a more in-depth analysis of the imaging features, as demonstrated in this study. When two tumors exhibit near-similar characteristics and locations, the ability to differentiate them based on imaging features such as those studied in this work becomes important. AI models trained on MRI sequences can assist in diagnosing similar lesions and aid in management planning. The transparent use of ML methods with pre-analysis and proper testing procedures is crucial for reliable, reproducible, and accurate findings. Ultimately, the primary goal of any analysis is to produce explainable and reproducible results that can be verified by other researchers, improving the diagnostic features and patient outcomes in medical research.

There are some limitations that need to be considered in the present study. First, the dataset used for analysis was limited in scope and size. Although it contained a sufficient number of samples to train ML models, the dataset may not have been representative of all possible scenarios, and the results may not generalize well to other datasets. To address this, we employed stratified random sampling to ensure that each tumor subtype was represented proportionally in the training and testing datasets. This approach helped to minimize bias in the model training and increase the generalizability of our findings. Second, our dataset only included four types of pediatric PF tumors, which may not fully represent the diversity of pediatric brain tumors. Third, the study was limited to the analysis of a single feature and pairwise interactions between features. Other important features or higher-order interactions that were not considered in this analysis may exist, and their inclusion may change the outcome. Future studies with larger sample sizes and additional advanced MRI protocols, such as semiquantitative and quantitative perfusion MRI and MR spectroscopy, could provide more insights into the diagnostic and prognostic value of MRI features for pediatric PF tumors. Additionally, further research is needed to investigate the potential of ML models and EDA to improve the reliability of pediatric PF tumor diagnosis and treatment.

5. Conclusions

The significance of our study lies in its ability to surpass the constraints of prior research in this field. While previous studies have often focused on only one binary differentiation or incorporated numerous exceptions, leading to reduced transparency, our research stands out by offering a comprehensive analysis of four distinct tumor subtypes within a single source. This paper offers a comprehensive and holistic understanding of the subject matter.

Through our analysis, we have uncovered the effectiveness of specific MRI features, such as ADC and DWI, in accurately distinguishing between tumor types, while also shedding light on the limited impact of features such as T1 and T1CE. The combination of EDA and ML has provided valuable insights into feature distributions and their importance in classification. Additionally, handcrafted features such as T2_Ratio and T1_Ratio enhanced the effectiveness of T2 and T1 features, respectively, in tumor classification. Overall, we identified RF as a suitable model for tumor classification, while LR emerged as the optimal choice for most binary cases.

In our analysis, we focused on MRI features that demonstrated minimal overlap between tumor types within their KDE distributions, as they offered valuable discriminatory information. Our findings have highlighted the potential of specific features, such as ADC_Ratio and ADC_Tumor, in effectively differentiating between tumor types. This effectiveness can be attributed to the distinct cellular characteristics, arrangement, and presence of cystic spaces within the tumor mass.

We have also demonstrated that in situations where patient data are limited, complex systems may not always be necessary to evaluate feature importance; in fact, they could impair both performance and interpretability. By conducting comprehensive analyses using simpler approaches, we can still extract valuable insights into the significance of specific features. This emphasizes the importance of adaptability and resourcefulness in leveraging available data to make informed decisions in clinical settings.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/cancers15164015/s1>, Supplementary File S1: Additional Figures (Figure S1: Examples of ROI delineations on a ADC, DWI, T1 and T1CE MRI; Figure S2: Pairwise plot with no parenchymas; Figure S3: RF model analysis of MB, EP, PA, and BG with classification metrics, confusion matrices, and feature importance distributions for each stratified random sampling state.); Supplementary File S2: Differentiation scores of machine learning models for all tumor types (Table S1: MB-EP; Table S2: MB-PA; Table S3: MB-BG; Table S4: EP-PA; Table S5: EP-BG; Table S6: PA-BG; Table S7: MB-EP-PA-BG.).

Author Contributions: Conceptualization, T.T.; Methodology, T.T.; Software, T.T.; Validation, T.T.; Formal analysis, T.T., C.N. and N.M.D.; Investigation, N.M.D.; Resources, N.M.D.; Data curation, N.M.D.; Writing—original draft, T.T.; Writing—review and editing, C.N., N.M.D. and B.K.; Visualization, T.T.; Supervision, B.K.; Project administration, B.K.; Funding acquisition, B.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Scientific and Technological Research Council of Türkiye (TUBITAK) [2232- 118C221].

Institutional Review Board Statement: After obtaining approval from the Institutional Review Board of Children Hospital of 02 with approval number [Ref: 632 QĐ-NĐ2 dated 12 May 2019], we conducted the study in both Radiology and Neurosurgery departments in accordance with the 1964 Helsinki declaration.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets generated and/or analyzed during the current study are not publicly available due to privacy concerns but are available from the corresponding author upon reasonable request. The source codes, EDA outputs, and entire ML evaluation results of the presented study can be accessed at <https://github.com/toygarr/save-the-kid>.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Louis, D.N.; Ohgaki, H.; Wiestler, O.D.; Cavenee, W.K.; Burger, P.C.; Jouvet, A.; Scheithauer, B.W.; Kleihues, P. The 2007 WHO Classification of Tumours of the Central Nervous System. *Acta Neuropathol.* **2007**, *114*, 97–109. [[CrossRef](#)] [[PubMed](#)]
2. Louis, D.N.; Perry, A.; Reifenberger, G.; Von Deimling, A.; Figarella-Branger, D.; Cavenee, W.K.; Ohgaki, H.; Wiestler, O.D.; Kleihues, P.; Ellison, D.W. The 2016 World Health Organization classification of tumors of the central nervous system: A summary. *Acta Neuropathol.* **2016**, *131*, 803–820. [[CrossRef](#)] [[PubMed](#)]
3. Meyers, S.P.; Kemp, S.S.; Tarr, R.W. MR imaging features of medulloblastomas. *AJR Am. J. Roentgenol.* **1992**, *158*, 859–865. [[CrossRef](#)] [[PubMed](#)]
4. Koeller, K.K.; Rushing, E.J. From the archives of the AFIP: Medulloblastoma: A comprehensive review with radiologic-pathologic correlation. *Radiographics* **2003**, *23*, 1613–1637. [[CrossRef](#)]
5. Koeller, K.K.; Rushing, E.J. From the archives of the AFIP: Pilocytic astrocytoma: Radiologic-pathologic correlation. *Radiographics* **2004**, *24*, 1693–1708. [[CrossRef](#)]
6. Koeller, K.K.; Sandberg, G.D. From the archives of the AFIP: Cerebral intraventricular neoplasms: Radiologic-pathologic correlation. *Radiographics* **2002**, *22*, 1473–1505. [[CrossRef](#)]
7. Meyers, S.; Khademan, Z.; Biegel, J.; Chuang, S.; Korones, D.; Zimmerman, R. Primary intracranial atypical teratoid/rhabdoid tumors of infancy and childhood: MRI features and patient outcomes. *Am. J. Neuroradiol.* **2006**, *27*, 962–971.
8. Arai, K.; Sato, N.; Aoki, J.; Yagi, A.; Taketomi-Takahashi, A.; Morita, H.; Koyama, Y.; Oba, H.; Ishiuchi, S.; Saito, N.; et al. MR signal of the solid portion of pilocytic astrocytoma on T2-weighted images: Is it useful for differentiation from medulloblastoma? *Neuroradiology* **2006**, *48*, 233–237. [[CrossRef](#)]
9. Koral, K.; Gargan, L.; Bowers, D.C.; Gimi, B.; Timmons, C.F.; Weprin, B.; Rollins, N.K. Imaging characteristics of atypical teratoid-rhabdoid tumor in children compared with medulloblastoma. *Am. J. Roentgenol.* **2008**, *190*, 809–814. [[CrossRef](#)]
10. Forbes, J.A.; Chambless, L.B.; Smith, J.G.; Wushensky, C.A.; Lebow, R.L.; Alvarez, J.; Pearson, M.M. Use of T2 signal intensity of cerebellar neoplasms in pediatric patients to guide preoperative staging of the neuraxis. *J. Neurosurg. Pediatr.* **2011**, *7*, 165–174. [[CrossRef](#)]
11. Forbes, J.A.; Reig, A.S.; Smith, J.G.; Jermakowicz, W.; Tomycz, L.; Shay, S.D.; Sun, D.A.; Wushensky, C.A.; Pearson, M.M. Findings on preoperative brain MRI predict histopathology in children with cerebellar neoplasms. *Pediatr. Neurosurg.* **2011**, *47*, 51–59. [[CrossRef](#)] [[PubMed](#)]
12. Poretti, A.; Meoded, A.; Huisman, T.A. Neuroimaging of pediatric posterior fossa tumors including review of the literature. *J. Magn. Reson. Imaging* **2012**, *35*, 32–47. [[CrossRef](#)]
13. Rasalkar, D.D.; Chu, W.C.w.; Paunipagar, B.K.; Cheng, F.W.; Li, C. Paediatric intra-axial posterior fossa tumours: Pictorial review. *Postgraduate Med. J.* **2013**, *89*, 39–46. [[CrossRef](#)] [[PubMed](#)]
14. Plaza, M.J.; Borja, M.J.; Altman, N.; Saigal, G. Conventional and advanced MRI features of pediatric intracranial tumors: Posterior fossa and suprasellar tumors. *Am. J. Roentgenol.* **2013**, *200*, 1115–1124. [[CrossRef](#)] [[PubMed](#)]
15. Porto, L.; Jurcoane, A.; Schwabe, D.; Hattingen, E. Conventional magnetic resonance imaging in the differentiation between high and low-grade brain tumours in paediatric patients. *Eur. J. Paediatr. Neurol.* **2014**, *18*, 25–29. [[CrossRef](#)] [[PubMed](#)]
16. Koob, M.; Girard, N. Cerebral tumors: Specific features in children. *Diagn. Interv. Imaging* **2014**, *95*, 965–983. [[CrossRef](#)] [[PubMed](#)]
17. Orphanidou-Vlachou, E.; Vlachos, N.; Davies, N.P.; Arvanitis, T.N.; Grundy, R.G.; Peet, A.C. Texture analysis of T1-and T2-weighted MR images and use of probabilistic neural network to discriminate posterior fossa tumours in children. *NMR Biomed.* **2014**, *27*, 632–639. [[CrossRef](#)]
18. Moharamzad, Y.; Sanei Taheri, M.; Niaghi, F.; Shobeiri, E. Brainstem glioma: Prediction of histopathologic grade based on conventional MR imaging. *Neuroradiol. J.* **2018**, *31*, 10–17. [[CrossRef](#)]
19. D’Arco, F.; Khan, F.; Mankad, K.; Ganau, M.; Caro-Dominguez, P.; Bisdas, S. Differential diagnosis of posterior fossa tumours in children: New insights. *Pediatr. Radiol.* **2018**, *48*, 1955–1963. [[CrossRef](#)]
20. Duc, N.M.; Huy, H.Q. Magnetic resonance imaging features of common posterior fossa brain tumors in children: A preliminary Vietnamese study. *Open Access Maced. J. Med. Sci.* **2019**, *7*, 2413. [[CrossRef](#)]
21. Duc, N.M.; Huy, H.Q.; Nadarajan, C.; Keserci, B. The role of predictive model based on quantitative basic magnetic resonance imaging in differentiating medulloblastoma from ependymoma. *Anticancer Res.* **2020**, *40*, 2975–2980. [[CrossRef](#)] [[PubMed](#)]
22. Rumboldt, Z.; Camacho, D.; Lake, D.; Welsh, C.; Castillo, M. Apparent diffusion coefficients for differentiation of cerebellar tumors in children. *Am. J. Neuroradiol.* **2006**, *27*, 1362–1369. [[PubMed](#)]
23. Jaremko, J.L.; Jans, L.; Coleman, L.T.; Ditchfield, M.R. Value and limitations of diffusion-weighted imaging in grading and diagnosis of pediatric posterior fossa tumors. *Am. J. Neuroradiol.* **2010**, *31*, 1613–1616. [[CrossRef](#)]
24. Gimi, B.; Cederberg, K.; Derinkuyu, B.; Gargan, L.; Koral, K.M.; Bowers, D.C.; Koral, K. Utility of apparent diffusion coefficient ratios in distinguishing common pediatric cerebellar tumors. *Acad. Radiol.* **2012**, *19*, 794–800. [[CrossRef](#)]
25. Bull, J.G.; Saunders, D.E.; Clark, C.A. Discrimination of paediatric brain tumours using apparent diffusion coefficient histograms. *Eur. Radiol.* **2012**, *22*, 447–457. [[CrossRef](#)] [[PubMed](#)]
26. Pierce, T.; Kranz, P.G.; Roth, C.; Leong, D.; Wei, P.; Provenzale, J.M. Use of apparent diffusion coefficient values for diagnosis of pediatric posterior fossa tumors. *Neuroradiol. J.* **2014**, *27*, 233–244. [[CrossRef](#)] [[PubMed](#)]
27. Porto, L.; Jurcoane, A.; Schwabe, D.; Kieslich, M.; Hattingen, E. Differentiation between high and low grade tumours in paediatric patients by using apparent diffusion coefficients. *Eur. J. Paediatr. Neurol.* **2013**, *17*, 302–307. [[CrossRef](#)]

28. Poretti, A.; Meoded, A.; Cohen, K.J.; Grotzer, M.A.; Boltshauser, E.; Huisman, T.A. Apparent diffusion coefficient of pediatric cerebellar tumors: A biomarker of tumor grade? *Pediatr. Blood Cancer* **2013**, *60*, 2036–2041. [[CrossRef](#)]
29. Gutierrez, D.R.; Awwad, A.; Meijer, L.; Manita, M.; Jaspan, T.; Dineen, R.A.; Grundy, R.G.; Auer, D.P. Metrics and textural features of MRI diffusion to improve classification of pediatric posterior fossa tumors. *Am. J. Neuroradiol.* **2014**, *35*, 1009–1015. [[CrossRef](#)]
30. Zitouni, S.; Koc, G.; Doganay, S.; Saracoglu, S.; Gumus, K.Z.; Ciraci, S.; Coskun, A.; Unal, E.; Per, H.; Kurtsoy, A.; et al. Apparent diffusion coefficient in differentiation of pediatric posterior fossa tumors. *Jpn. J. Radiol.* **2017**, *35*, 448–453. [[CrossRef](#)]
31. Esa, M.M.M.; Mashaly, E.M.; El-Sawaf, Y.F.; Dawoud, M.M. Diagnostic accuracy of apparent diffusion coefficient ratio in distinguishing common pediatric CNS posterior fossa tumors. *Egypt. J. Radiol. Nucl. Med.* **2020**, *51*, 1–11. [[CrossRef](#)]
32. Minh Thong, P.; Minh Duc, N. The role of apparent diffusion coefficient in the differentiation between cerebellar medulloblastoma and brainstem glioma. *Neurol. Int.* **2020**, *12*, 34–40. [[CrossRef](#)] [[PubMed](#)]
33. Dury, R.J.; Lourdasamy, A.; Macarthur, D.C.; Peet, A.C.; Auer, D.P.; Grundy, R.G.; Dineen, R.A. Meta-Analysis of Apparent Diffusion Coefficient in Pediatric Medulloblastoma, Ependymoma, and Pilocytic Astrocytoma. *J. Magn. Reson. Imaging* **2022**, *56*, 147–157. [[CrossRef](#)] [[PubMed](#)]
34. Chen, D.; Lin, S.; She, D.; Chen, Q.; Xing, Z.; Zhang, Y.; Cao, D. Apparent Diffusion Coefficient in the Differentiation of Common Pediatric Brain Tumors in the Posterior Fossa: Different Region-of-Interest Selection Methods for Time Efficiency, Measurement Reproducibility, and Diagnostic Utility. *J. Comp. Assist. Tomogr.* **2023**, *47*, 291. [[CrossRef](#)] [[PubMed](#)]
35. Reddy, N.; Ellison, D.W.; Soares, B.P.; Carson, K.A.; Huisman, T.A.; Patay, Z. Pediatric posterior fossa medulloblastoma: The role of diffusion imaging in identifying molecular groups. *J. Neuroimaging* **2020**, *30*, 503–511. [[CrossRef](#)]
36. Gonçalves, F.G.; Zandifar, A.; Ub Kim, J.D.; Tierradentro-García, L.O.; Ghosh, A.; Khrichenko, D.; Andronikou, S.; Vossough, A. Application of Apparent Diffusion Coefficient Histogram Metrics for Differentiation of Pediatric Posterior Fossa Tumors: A Large Retrospective Study and Brief Review of Literature. *Clin. Neuroradiol.* **2022**, *32*, 1097–1108. [[CrossRef](#)]
37. Phuttharak, W.; Wannasarnmetha, M.; Waraaswapati, S.; Yuthawong, S. Diffusion MRI in evaluation of pediatric posterior fossa tumors. *Asian Pac. J. Cancer Prev. APJCP* **2021**, *22*, 1129. [[CrossRef](#)]
38. Koob, M.; Girard, N.; Ghattas, B.; Fella, S.; Confort-Gouny, S.; Figarella-Branger, D.; Scavarda, D. The diagnostic accuracy of multiparametric MRI to determine pediatric brain tumor grades and types. *J. Neuro-Oncol.* **2016**, *127*, 345–353. [[CrossRef](#)]
39. Deng, J.; Xue, C.; Liu, X.; Li, S.; Zhou, J. Differentiating between adult intracranial medulloblastoma and ependymoma using MRI. *Clin. Radiol.* **2023**, *78*, e288–e293. [[CrossRef](#)]
40. Yamaguchi, J.; Ohka, F.; Motomura, K.; Saito, R. Latest classification of ependymoma in the molecular era and advances in its treatment: A review. *Jpn. J. Clin. Oncol.* **2023**, *53*, hyad056. [[CrossRef](#)] [[PubMed](#)]
41. Wang, S.; Wang, G.; Zhang, W.; He, J.; Sun, W.; Yang, M.; Sun, Y.; Peet, A. MRI-based whole-tumor radiomics to classify the types of pediatric posterior fossa brain tumor. *Neurochirurgie* **2022**, *68*, 601–607. [[CrossRef](#)] [[PubMed](#)]
42. Yearley, A.G.; Blitz, S.E.; Patel, R.V.; Chan, A.; Baird, L.C.; Friedman, G.K.; Arnaout, O.; Smith, T.R.; Bernstock, J.D. Machine Learning in the Classification of Pediatric Posterior Fossa Tumors: A Systematic Review. *Cancers* **2022**, *14*, 5608. [[CrossRef](#)] [[PubMed](#)]
43. Huang, J.; Shlobin, N.A.; Lam, S.K.; DeCuypere, M. Artificial intelligence applications in pediatric brain tumor imaging: A systematic review. *World Neurosurg.* **2022**, *157*, 99–105. [[CrossRef](#)] [[PubMed](#)]
44. Li, M.; Shang, Z.; Yang, Z.; Zhang, Y.; Wan, H. Machine learning methods for MRI biomarkers analysis of pediatric posterior fossa tumors. *Biocybern. Biomed. Eng.* **2019**, *39*, 765–774. [[CrossRef](#)]
45. Li, M.; Wang, H.; Shang, Z.; Yang, Z.; Zhang, Y.; Wan, H. Ependymoma and pilocytic astrocytoma: Differentiation using radiomics approach based on machine learning. *J. Clin. Neurosci.* **2020**, *78*, 175–180. [[CrossRef](#)] [[PubMed](#)]
46. Zhou, H.; Hu, R.; Tang, O.; Hu, C.; Tang, L.; Chang, K.; Shen, Q.; Wu, J.; Zou, B.; Xiao, B.; et al. Automatic machine learning to differentiate pediatric posterior fossa tumors on routine MR imaging. *Am. J. Neuroradiol.* **2020**, *41*, 1279–1285. [[CrossRef](#)]
47. Dong, J.; Li, L.; Liang, S.; Zhao, S.; Zhang, B.; Meng, Y.; Zhang, Y.; Li, S. Differentiation between ependymoma and medulloblastoma in children with radiomics approach. *Acad. Radiol.* **2021**, *28*, 318–327. [[CrossRef](#)]
48. Grist, J.T.; Withey, S.; MacPherson, L.; Oates, A.; Powell, S.; Novak, J.; Abernethy, L.; Pizer, B.; Grundy, R.; Bailey, S.; et al. Distinguishing between paediatric brain tumour types using multi-parametric magnetic resonance imaging and machine learning: A multi-site study. *NeuroImage Clin.* **2020**, *25*, 102172. [[CrossRef](#)]
49. Payabvash, S.; Aboian, M.; Tihan, T.; Cha, S. Machine learning decision tree models for differentiation of posterior fossa tumors using diffusion histogram analysis and structural MRI findings. *Front. Oncol.* **2020**, *10*, 71. [[CrossRef](#)]
50. Novak, J.; Zarinabad, N.; Rose, H.; Arvanitis, T.; MacPherson, L.; Pinkey, B.; Oates, A.; Hales, P.; Grundy, R.; Auer, D.; et al. Classification of paediatric brain tumours by diffusion weighted imaging and machine learning. *Sci. Rep.* **2021**, *11*, 1–8. [[CrossRef](#)]
51. Zhuge, Y.; Ning, H.; Mathen, P.; Cheng, J.Y.; Krauze, A.V.; Camphausen, K.; Miller, R.W. Automated glioma grading on conventional MRI images using deep convolutional neural networks. *Med. Phys.* **2020**, *47*, 3044–3053. [[CrossRef](#)]
52. Quon, J.; Bala, W.; Chen, L.; Wright, J.; Kim, L.; Han, M.; Shpanskaya, K.; Lee, E.; Tong, E.; Iv, M.; et al. Deep learning for pediatric posterior fossa tumor detection and classification: A multi-institutional study. *Am. J. Neuroradiol.* **2020**, *41*, 1718–1725. [[CrossRef](#)] [[PubMed](#)]
53. Saju, A.C.; Chatterjee, A.; Sahu, A.; Gupta, T.; Krishnatry, R.; Mokal, S.; Sahay, A.; Epari, S.; Prasad, M.; Chinnaswamy, G.; et al. Machine-learning approach to predict molecular subgroups of medulloblastoma using multiparametric MRI-based tumor radiomics. *Br. J. Radiol.* **2022**, *95*, 20211359. [[CrossRef](#)] [[PubMed](#)]

54. Fathi Kazerooni, A.; Arif, S.; Madhogarhia, R.; Khalili, N.; Haldar, D.; Bagheri, S.; Familiar, A.M.; Anderson, H.; Haldar, S.; Tu, W.; et al. Automated tumor segmentation and brain tissue extraction from multiparametric MRI of pediatric brain tumors: A multi-institutional study. *Neuro-Oncol. Adv.* **2023**, *5*, vdad027. [[CrossRef](#)] [[PubMed](#)]
55. Kashani, H.R.K.; Azhari, S.; Moradi, E.; Samii, F.; Mirahmadi, M.S.; Towfiqi, A. Predictive Value of Blood Markers in Pediatric Brain Tumors Using Machine Learning. *Pediatr. Neurosurg.* **2022**, *57*, 323–332. [[CrossRef](#)]
56. Tanyel, T.; Ayvaz, S.; Keserci, B. Beyond Known Reality: Exploiting Counterfactual Explanations for Medical Research. *arXiv* **2023**, arXiv:2307.02131.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.