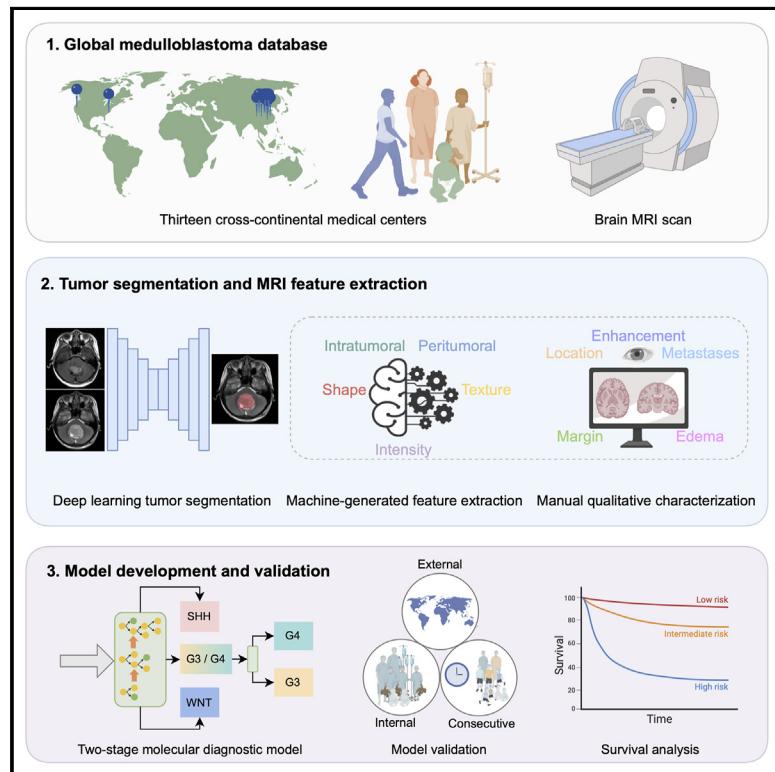


Advancing presurgical non-invasive molecular subgroup prediction in medulloblastoma using artificial intelligence and MRI signatures

Graphical abstract



Authors

Yan-Ran (Joyce) Wang,
Pengcheng Wang, Zihan Yan, ...,
Lu Tian, Feng Wu, Jian Gong

Correspondence

wangyanran100@gmail.com (Y.-R.W.),
gongjian88@vip.163.com (J.G.)

In brief

Wang et al. show that image-based machine learning strategies derived from MRI data have the potential to augment existing molecular testing and create an alternative pathway for noninvasive, presurgical, and low-cost molecular subgroup prediction of medulloblastoma. The open dataset aims to advance global MB investigation and improve patient survival worldwide.

Highlights

- AI-enabled MRI enables presurgical molecular subgroup prediction in medulloblastoma
- Computational and manual MRI signatures are complementary to AI diagnostics in MB
- Internal, external, and consecutive validations highlight model generalizability
- The publicly available dataset from 13 centers advances global medulloblastoma research



Article

Advancing presurgical non-invasive molecular subgroup prediction in medulloblastoma using artificial intelligence and MRI signatures

Yan-Ran (Joyce) Wang,^{1,3,17,19,*} Pengcheng Wang,^{4,17} Zihan Yan,^{2,17} Quan Zhou,^{3,5,18} Fatma Gunturkun,^{3,7,18} Peng Li,^{1,8,18} Yanshen Hu,^{8,18} Wei Emma Wu,^{3,6,18} Kankan Zhao,^{9,18} Michael Zhang,^{3,5} Haoyi Lv,⁸ Lehao Fu,⁸ Jiajie Jin,⁸ Qing Du,¹ Haoyu Wang,⁸ Kun Chen,¹⁰ Liangqiong Qu,¹¹ Keldon Lin,¹² Michael Lv,^{3,5} Hao Wang,^{1,13} Xiaoyan Sun,^{1,8} Hannes Vogel,^{3,14} Summer Han,^{3,7} Lu Tian,^{3,15,16} Feng Wu,^{8,16} and Jian Gong^{2,16,*}

¹Anhui Province Key Laboratory of Biomedical Imaging and Intelligent Processing, Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Hefei 230088, China

²Department of Pediatric Neurosurgery, Beijing Tiantan Hospital, Capital Medicine University, Beijing Neurosurgical Institute, Beijing 100070, China

³School of Medicine, Stanford University, Stanford, CA 94304, USA

⁴Department of Biomedical Engineering, University of Southern California, Los Angeles, CA 90089, USA

⁵Department of Neurosurgery, Stanford School of Medicine, Stanford University, Stanford, CA 94304, USA

⁶Department of Radiology Oncology, Stanford University, Stanford, CA 94305, USA

⁷Quantitative Sciences Unit, Department of Medicine, Stanford University, Stanford, CA 94304, USA

⁸School of Engineering, University of Science and Technology of China, Hefei 230001, China

⁹Paul C. Lauterbur Research Center for Biomedical Imaging, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

¹⁰The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230026, China

¹¹The Department of Statistics and Actuarial Science and the Institute of Data Science, The University of Hong Kong, Hong Kong 999077, China

¹²Mayo Clinic Alix School of Medicine, Scottsdale, AZ 85054, USA

¹³MoE Key Laboratory of Brain-inspired Intelligent Perception and Cognition, School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China

¹⁴Department of Pathology, Stanford School of Medicine, Stanford University, Stanford, CA 94304, USA

¹⁵Department of Statistics, Stanford School of Medicine, Stanford University, Stanford, CA 94304, USA

¹⁶Senior authors

¹⁷These authors contributed equally

¹⁸These authors contributed equally

¹⁹Lead contact

*Correspondence: wangyanran100@gmail.com (Y.-R.W.), gongjian88@vip.163.com (J.G.)

<https://doi.org/10.1016/j.ccel.2024.06.002>

SUMMARY

Global investigation of medulloblastoma has been hindered by the widespread inaccessibility of molecular subgroup testing and paucity of data. To bridge this gap, we established an international molecularly characterized database encompassing 934 medulloblastoma patients from thirteen centers across China and the United States. We demonstrate how image-based machine learning strategies have the potential to create an alternative pathway for non-invasive, presurgical, and low-cost molecular subgroup prediction in the clinical management of medulloblastoma. Our robust validation strategies—including cross-validation, external validation, and consecutive validation—demonstrate the model's efficacy as a generalizable molecular diagnosis classifier. The detailed analysis of MRI characteristics replenishes the understanding of medulloblastoma through a nuanced radiographic lens. Additionally, comparisons between East Asia and North America subsets highlight critical management implications. We made this comprehensive dataset, which includes MRI signatures, clinicopathological features, treatment variables, and survival data, publicly available to advance global medulloblastoma research.

INTRODUCTION

Medulloblastoma (MB) is the most common malignant central nervous system (CNS) tumor and a leading cause of non-accidental

death in children and adolescents.^{1,2} Currently, molecular subgroup testing stands as the only validated means of stratifying risk for MB.^{3–5} The 2021 fifth edition of the WHO Classification of Tumors of the Central Nervous System (WHO CNS5) mandate



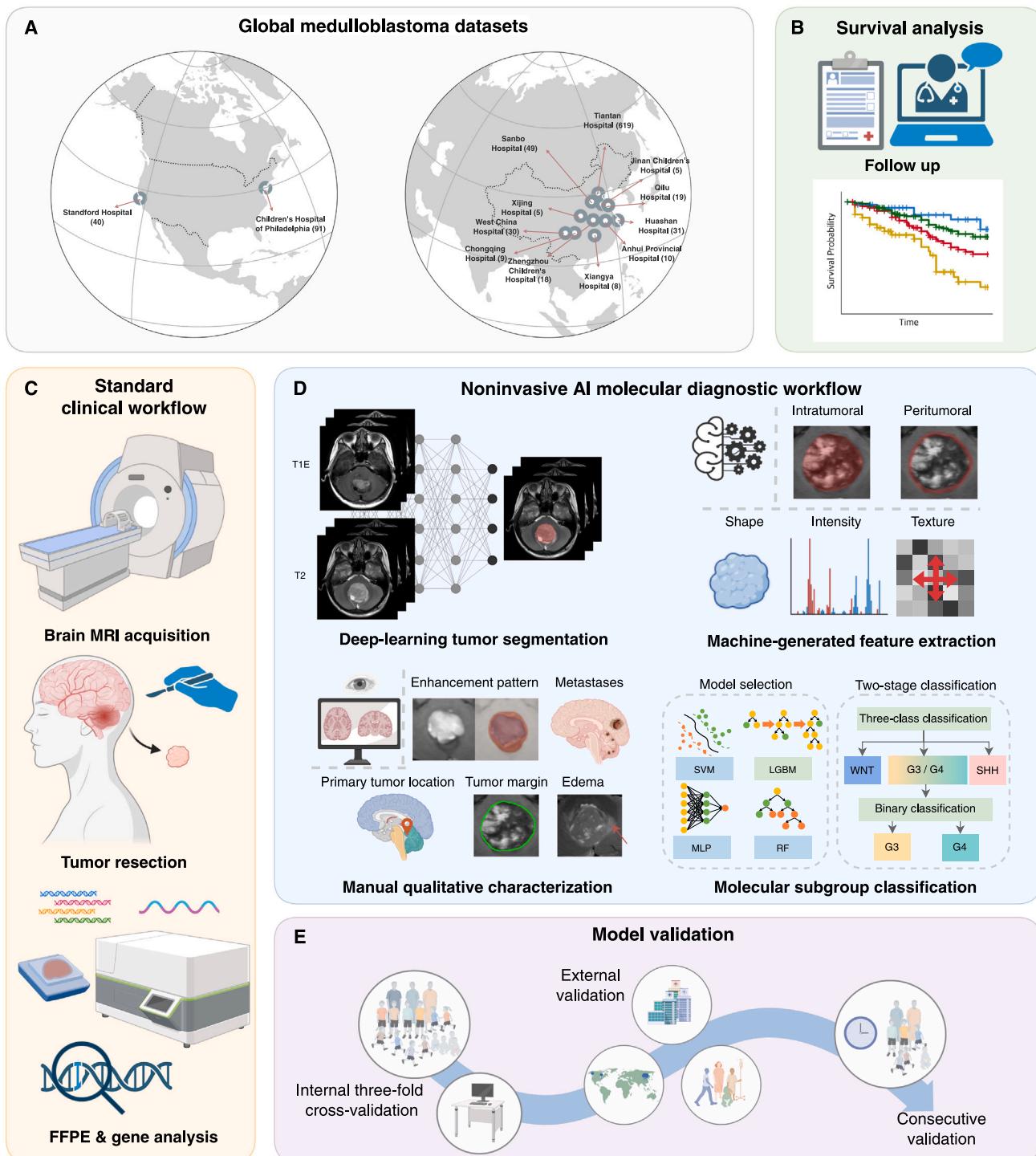
**Figure 1. Methodological overview and global medulloblastoma datasets assembled**

Figure360▷

For a Figure360 author presentation of Figure 1, see <https://doi.org/10.1016/j.ccel.2024.06.002>.

(A) The global medulloblastoma dataset section depicts a visual representation—a map of cities in China and the United States whose medical centers were included in assembling this worldwide medulloblastoma dataset. The cities include Stanford, Philadelphia, Shanghai, Beijing, Jinan, Xi'an, Chongqing, Chengdu, Zhengzhou, Hefei, and Changsha.

(B) Clinical data and prognostic information were collected via phone surveys and served as the primary endpoint for survival analysis in this study.

(C) Depiction of the conventional standard molecular diagnostic workflow for medulloblastoma.

(legend continued on next page)

robust molecular subgrouping in medulloblastoma, underscoring its pivotal role in advancing clinical care.³ However, the complexity and costliness of molecular testing, which involves RNA sequencing or DNA methylation profiling of tumor tissue post-surgery, poses substantial challenges for medical centers globally, particularly in resource-limited regions. This, in turn, drives global health disparities and impedes the exploration of novel subgroup-specific therapeutic opportunities.^{6,7}

Recent evidence reveals that radiographic imaging of patients harbors minable information, including subtle features beyond human visual perception that can be identified by machine learning (ML) to prognosticate clinical outcomes across diverse cancer types.^{8–12} Nevertheless, the very few studies that have piloted image data and machine learning approaches in medulloblastoma were largely inconclusive due to small sample sizes and a restricted identification of MR imaging signatures.^{5,13–16}

In this study, we curated an international molecularly characterized dataset comprising 934 medulloblastoma patients across thirteen medical centers spanning China and the United States. This assembled collection includes MR imaging, demographic data, clinicopathological features, treatment variables, and survival information. Leveraging this extensive dataset, we present an artificial intelligence (AI) system rooted in MR imaging for noninvasive and low-cost presurgical prediction of molecular subgroups.

RESULTS

Largest molecularly characterized medulloblastoma database

This international study enrolled 934 patients with medulloblastoma: 803 patients from China (86%, July 2015–December 2023) and 131 patients from the United States (14%, July 1998–February 2022). The thirteen participating medical centers include two centers in the USA (Lucile Packard Children's Hospital Stanford, Stanford; Children's Brain Tumor Network [CBTN]—mostly assembled from Children's Hospital of Philadelphia [CHOP], Philadelphia) and eleven centers in China (Beijing Tiantan Hospital, Beijing; Sanbo Brain Sciences Institute of Beijing, Beijing; Huashan Hospital, Shanghai; Jinan Children's Hospital, Jinan; Xijing Hospital, Xi'an; Qilu Hospital of Shandong University, Jinan; Children's Hospital of Chongqing Medical University, Chongqing; West China Hospital, Chengdu; Zhengzhou Children's Hospital, Zhengzhou; Anhui Provincial Hospital, Hefei; Xiangya Hospital, Changsha). Figure 1A presents a map of cities whose medical centers were included in assembling this global medulloblastoma dataset. Of the enrolled patients, 689 with complete high-resolution presurgical MRI scans, including both axial contrast-enhanced

T1-weighted and T2-weighted sequences, were utilized for machine learning model development and validation, constituting the largest global presurgical MR imaging dataset for medulloblastoma.

Overview of the presurgical noninvasive AI molecular diagnostic workflow

The machine learning (ML) workflow begins with a patient undergoing a brain MRI scan comprising axial T1-enhanced and T2-weighted MRI images as part of the presurgical assessment (Figure 1C). Our developed tumor segmentation deep learning model automatically delineates the primary tumor area on the MR imaging, followed by manual verification to ensure its accuracy (Figure 1D). A detailed analysis of the variability in segmentation masks leading to the final molecular prediction is outlined in Figure S1. Subsequently, quantitative radiomic features and qualitative tumor characterizations are derived from MR imaging to inform the ML classifiers. This includes computer-generated MR signatures from intra-tumoral and peri-tumoral regions, alongside manual inspection of the primary tumor location, tumor enhancement patterns, intracranial metastases patterns, tumor margins, and peri-tumoral edema on MR imaging (Figure 1D). Utilizing these MR imaging signatures, the three-class classifier distinguishes patients into WNT, SHH, or G3/G4 (non-WNT/non-SHH) classes, and the binary classifier further subclassifies group 3 (G3) and group 4 (G4) within the non-WNT/non-SHH subgroup. Figure 1 illustrates the intricate workflow of the noninvasive AI molecular diagnostic system for medulloblastoma.

Image-based ML strategies enable presurgical classification of MB subgroups

We used the brain MRI data from Beijing Tiantan Hospital¹⁷ as the primary set for model development and data pooled from all the other medical centers across China and the United States as the external validation set. 3-fold cross-validation was conducted within the primary set to further validate performance. Patient characteristics of both the primary set and external validation set for AI-enabled molecular diagnosis are detailed in Table S1.

We first evaluated the three-class classifier which differentiates WNT, SHH, and non-WNT/non-SHH (i.e., G3/G4) simultaneously. Upon 3-fold cross-validation in the primary set, the area under the curve (AUCs) of the three-class classifier were 0.924 (95% confidence interval [CI], 0.885 to 0.954) for WNT subgroup against other two classes, 0.819 (95% CI, 0.780 to 0.857) for SHH subgroup against other two classes, and 0.810 (95% CI, 0.773 to 0.846) for non-WNT/non-SHH subgroup against other two classes. The three-class classifier achieved cross-institution (external validation) AUCs of 0.852 (0.741, 0.938) for WNT, 0.806

(D) Illustration of the noninvasive AI-enabled molecular diagnostic workflow. The workflow initiates with a deep learning model for automatically delineating primary tumor areas in MR images, followed by manual verification for accuracy. Quantitative radiomic features and qualitative human characterization of MR imaging, encompassing peri-tumoral, intra-tumoral, and extra-tumoral areas, are extracted to inform the AI classifiers. The AI classifiers then stratify patients into WNT, SHH, or non-WNT/non-SHH groups, with further subclassification for group 3 and group 4 within the non-WNT/non-SHH category.

(E) Comprehensive validation of the AI model over the primary internal set (3-fold cross validation), the external validation over cross-continental centers, and the consecutive testing set. Key abbreviations: T1E, contrast-enhanced T1; AI, artificial intelligence; MRI, magnetic resonance imaging; FFPE, formalin-fixed paraffin-embedded; LGBM, light gradient-boosting machine; SVM, support vector machine; RF, random forest; MLP, multilayer perceptron neural network; WNT, wingless; SHH, sonic hedgehog; G3, group 3; G4, group 4.

See also Figures S1 and S8, Tables S11–S13.

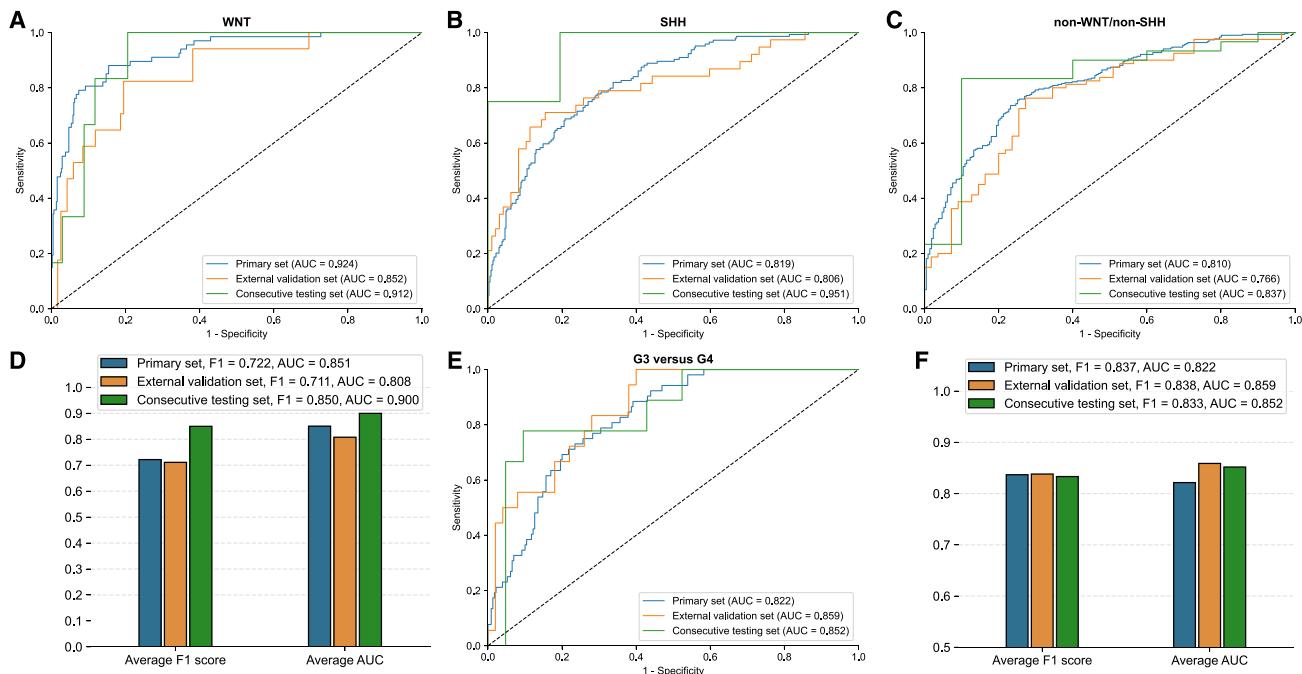


Figure 2. Performance of AI molecular diagnostic models across primary internal, external validation, and consecutive testing sets

(A–C) Receiver-operating characteristic curves illustrating the performance of the three-class classifier (WNT vs. SHH vs. non-WNT/non-SHH, i.e., G3/4) in the primary internal set ($n = 514$), external validation set ($n = 135$), and consecutive testing set ($n = 40$).

(D) Class-averaged performance metrics for the three-class classifier across the three testing datasets.

(E and F) Receiver-operating characteristic curves and corresponding performance metrics for the binary classifier (group 3 vs. group 4) in the primary internal set ($n = 282$), external validation set ($n = 68$), and consecutive testing set ($n = 30$). Key abbreviations: WNT, wingless; SHH, sonic hedgehog; G3, group 3; G4, group 4; AUC, area under the curve.

See also Tables S2 and S3.

(0.712, 0.888) for SHH, and 0.766 (0.686, 0.844) for non-WNT/non-SHH, i.e., G3/G4 (Figure 2). We then examined the binary classifier for discriminating G3 from G4. The binary classifier for G3 versus G4 classification achieved AUCs of 0.822 (95% CI, 0.764 to 0.877) upon 3-fold cross-validation in the primary set and 0.859 (95% CI, 0.760 to 0.939) in the external validation set (Figure 2). It is noteworthy that the external validation set comprises patients with markedly distinct ethical backgrounds compared to those included in the internal set (Table S1, race: $p < 0.0001$). This indicates that the AI model can generalize across diverse patient populations, including medical centers and patient race uninvolved during model training.

To further evaluate the performance of our developed AI models in a real-world clinical setting, we prospectively established a fresh independent testing set comprising 40 medulloblastoma patients consecutively admitted to Beijing Tiantan Hospital¹⁷ in 2023. It's noteworthy that these patients underwent molecular validation via DNA methylation classification, a meticulously calibrated method.¹⁸ Our ML classifiers exhibited exceptional performance on this consecutive testing set, yielding a class-weighted average area under the curve (AUC) of 0.900 (0.797, 0.976) for the three-class classifier and 0.852 (0.667, 1.000) for the binary classifier. The robust performance on this carefully curated testing set, designed to genuinely reflect the clinical prevalence of medulloblastoma molecular subtypes, highlights the model's potential for real-world clinical applications. Comprehensive model performance metrics for both the

three-class classifier and the binary classifier across the three testing sets are presented in Tables S2 and S3, respectively.

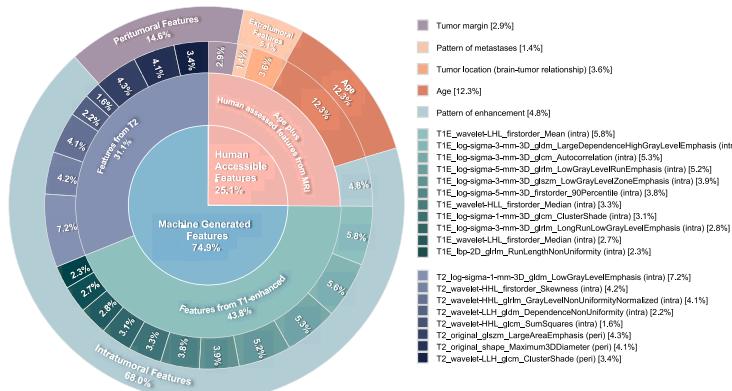
In-depth assessment of feature importance: Regional analysis, modalities, and human versus computer-derived MR signatures

Next, we conducted a thorough analysis of feature importance, gauging the extent to which each input feature contributes to the predictive accuracy of our developed ML model, utilizing Shapley analysis.¹⁹ The significance of each input feature, categorized accordingly, is depicted in sunburst charts within Figures 3A and 3C for the three-class classifier and binary classifier, respectively.

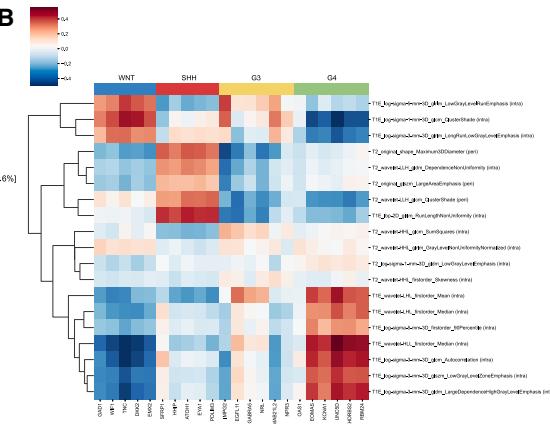
Unsurprisingly, intra-tumoral features emerge as the most influential in accurately predicting the molecular subgroup of a given tumor (relative importance: 68.0% [three-class classifier], 78.7% [binary classifier]). Additionally, in the three-class classifier, peri-tumoral features account for 14.6% of the feature importance, with extra-tumoral features contributing 5.1%. This trend persists in the binary classifier, where peri-tumoral features represent 14.8% of the importance, while extra-tumoral features account for 3.2%. Note that we classify the pattern of metastases and tumor location as extra-tumoral features because their identification requires inference from regions beyond the primary tumor mass.

The contribution of machine-generated features to feature importance was notably significant, comprising 74.9% in the

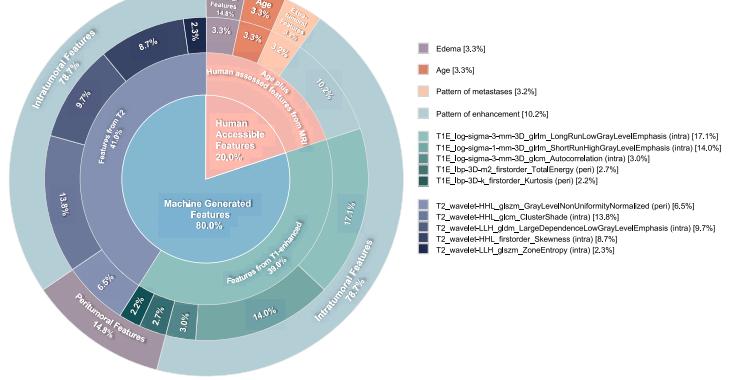
A



B



C



D

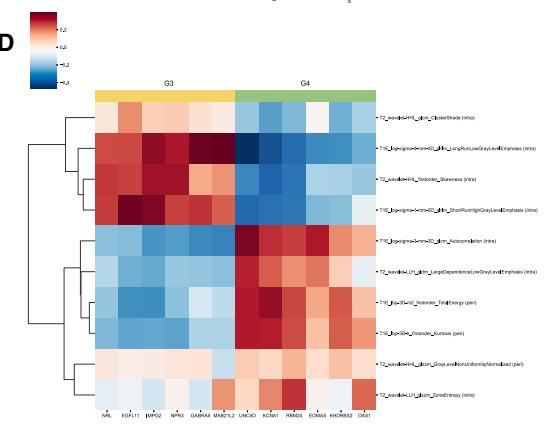


Figure 3. Feature importance in AI molecular diagnosis and radiomics-molecular signature association

(A) Sunburst chart illustrating relative feature importance assessed via Shapley analysis in the three-class classifier (WNT versus SHH versus non-WNT/non-SHH, i.e., G3/4). Features are categorized by region (intra-tumoral, peri-tumoral, and extra-tumoral), modalities (T1E and T2), and extracted source (quantitative radiomics and qualitative characterization by neuroradiologists).

(B) Heatmap depicting the association between radiomics features utilized in the three-class classifier and the expression of NanoString signature genes for medulloblastoma.

(C) Sunburst chart displaying feature importance for the binary classifier.

(D) Heatmap showcasing the association between radiomics features utilized in the binary classifier and NanoString signature genes for distinguishing group 3 and group 4. Key abbreviations: intra, intra-tumoral region; peri, peri-tumoral region.

See also Figures S2–S5 and Tables S4, S5, S7, and S8.

three-class classifier, while human-accessible features represented 25.1%. This trend persisted in the binary classifier, with 80.0% attributed to machine-generated features and 20.0% to human-accessible features. One plausible explanation could be that machine-generated quantitative radiomics on MRI imaging offer multidimensional insights, contrasting with the qualitative manual characterization of tumors.

In the three-class classifier (WNT vs. SHH vs. G3/G4), our analysis revealed that among the 19 machine-generated radiomic features selected, 11 derived from T1E MRI (feature importance: 43.8%) and 8 from T2 MRI scans (feature importance: 31.1%), with 16 arising from intra-tumoral regions and 3 from peri-tumoral regions. Detailed insights into these radiomic features including their qualitative interpretations are provided in Table S4. Additionally, Figure 3B provides heatmap visualizations showcasing the strong correlation between the radiomic features and the NanoString signature genes within each molecular subgroup. Notably, the top five features deemed most

significant for the classifier's efficacy were identified, with age contributing a relative importance of 12.3%. Other influential features included T2-LowGrayLevelEmphasis (region: intra-tumoral; feature class: texture - gray level dependence matrix [gldm]) at 7.2%, T1E-LHL_Mean (intra-tumoral; intensity - first-order) at 5.8%, T1E-LargeDependenceHighGrayLevelEmphasis (intra-tumoral; texture - gldm) at 5.6%, and T1E-Autocorrelation (intra-tumoral; texture - gray level co-occurrence matrix [glcm]) at 5.3%.

In the binary classifier (G3 vs. G4), our examination of ten machine-generated radiomic features revealed a balanced distribution, with 5 derived from T1E MRI (feature importance: 39.0%) and the remaining 5 from T2 MRI scans (41.0%). These features were sourced from both intra-tumoral (7) and peri-tumoral (3) regions. For further insights into the radiomic features utilized to inform the binary classifier, please refer to Figure 3C and Table S5. Additionally, Figure 3D displays heatmap visualizations that highlight the significant association between radiomic

features and NanoString signature genes distinguishing group 3 and group 4. Notably, the top five influential features critical for distinguishing between group 3 and group 4 tumors were identified. These included T1E-LongRunLowGrayLevelEmphasis (intra-tumoral; glrlm; 17.1%), T1E-ShortRunHighGrayLevelEmphasis (intra-tumoral; glrlm; 14.0%), T2-ClusterShade (intra-tumoral; glcm; 13.8%), the enhancement pattern within the primary tumor (10.2%), and T2-LargeDependenceLowGrayLevel Emphasis (intra-tumoral; gldm; 9.7%).

Tumor location is highly associated with molecular subgroups

Statistically significant differences in the primary tumor's location among molecular subgroups were observed ($p < 0.0001$; **Table 1**), with the tumor location feature holding a Shapley value of 3.6% in the three-class classifier (depicted in **Figure 3A**), underscoring its high predictive capacity in distinguishing WNT, SHH, and G3/G4 tumors. No significant statistical difference was found between G3 and G4 in terms of the primary tumor location, and the tumor location feature was not included in the binary classifier designed to differentiate between these two classes.

Among WNT tumors, 24.14% (28 of 116) were centered in the fourth ventricle while extending laterally along one (18.10%; in 21 cases) or both (6.03%; in 7 cases) lateral recesses toward cerebellopontine angles (CPAs). Strikingly, this location pattern (fourth ventricle-uni-CPA and fourth ventricle-bi-CPA), which was initially reported as being pathognomonic for the WNT subgroup,²⁰ was more frequent in group 3 (35.42%, 34 of 96) and group 4 (35.51%, 114 of 321). Nevertheless, WNT tumors primarily exhibited unilateral extension, group 3 and group 4 often presented with bilateral extension along both recesses toward the CPAs (fourth ventricle-bi-CPA; $p < 0.0001$; **Table 1**). Additionally, six tumors located within CPA showed no significant distribution difference among molecular subtypes ($p = 0.308$; **Table 1**). These findings contradict previous studies with limited sample sizes^{16,20} and highlight that the localization of cerebellar peduncle/cerebellopontine angle cistern (CP/CPA) and lateral recess extension are not unique to WNT tumors.

In SHH tumors, 38.83% (80 of 206) exhibited infiltration of the cerebellar hemispheres, with 21.84% (45 of 206) centered within the cerebellar hemisphere and 16.99% (35 of 206) invading both the cerebellar hemisphere and the fourth ventricle. Remarkably, 86.54% (45 of 52) of tumors within the cerebellar hemisphere were classified as the SHH subtype, indicating a strong association between cerebellar hemispheric location and SHH tumors ($p < 0.0001$; **Table 1**), consistent with previous findings.¹⁶ However, unlike previous studies suggesting that the primary location of SHH subtype in the cerebellum,^{16,21} our study revealed that 43.69% of SHH tumors were located in the midline vermis/fourth ventricle, remaining the primary location for the SHH subtype, with the cerebellum serving as the secondary location (occurrence rate: 38.83%).

Across 739 patients analyzed, medulloblastomas, irrespective of molecular subtypes, predominantly resided in the fourth ventricle and were localized within the midline of the fourth ventricle (midline vermis/fourth ventricle: 68.1% for WNT, 43.69% for SHH, 56.25% for group 3, and 57.94% for group 4; **Table 1**). Representative cases illustrating the localization

patterns in medulloblastoma primary tumors are depicted in **Figure S2**.

Intracranial solid metastasis at diagnosis is highly subgroup specific

Among the 733 patients assessed for metastases, 126 individuals (17.2%) displayed visible intracranial solid metastases, a proportion consistent with findings from the Toronto cohort ($p = 0.378$, 21.0%; 25 out of 119).²² To discern the subgroup-specific patterns of metastatic dissemination in medulloblastoma, we delved deeper into factors including location, contrast uptake patterns (enhancement), and diffusion restriction associated with the metastases. Our analysis revealed that the occurrence, anatomical location, and imaging characteristics of metastatic medulloblastoma are markedly subgroup specific (**Table 2**). Furthermore, metastatic patterns contributed to both the binary classifier (3.2%) and the three-class classifier (1.4%) for molecular subgroup diagnosis (**Figures 3A** and **3C**).

The occurrence of intracranial solid metastatic diseases across molecular subgroups was statistically significant ($p = 0.0004$; **Table 2**), with occurrences as follows: 4 out of 115 in WNT (3.48%); 40 out of 204 in SHH (19.61%); 20 out of 95 in group 3 (21.05%); and 62 out of 319 in group 4 (19.44%). Notably, metastatic dissemination at diagnosis is rare in WNT, while it is nearly equally enriched in SHH, group 3, and group 4. Our findings contrast with previous observations based on limited SHH samples (less than 30) that suggested rarity of metastatic dissemination in SHH medulloblastoma.²³

The anatomical distribution of metastases demonstrates subgroup-specific patterns. Among cases with intracranial solid metastases, 56 out of 126 (44.4%) displayed ependymal metastases (including the third ventricular infundibular recess [3rd V.I.R.] metastases), and 79 out of 126 (62.7%) showed leptomeningeal metastases (**Table 2**). Notably, the prevalence of infratentorial leptomeningeal metastases varies significantly across molecular subgroups ($p = 0.017$). In group 3, infratentorial leptomeningeal metastases are most prevalent, accounting for 70% (14 out of 20) of group 3 patients with intracranial solid metastasis, while in WNT, only 25% (1 out of 4 cases) are observed, compared to 52.5% (21 out of 40 cases) in SHH and 33.9% (21 out of 62 cases) in group 4. Leptomeningeal metastases involving both infratentorial and supratentorial regions are notably more frequent in SHH, representing 77.5% (31 out of 40 cases) of instances with intracranial metastasis ($p = 0.04$). In contrast, the incidence of ependymal metastases, including those within the 3rd V.I.R., does not significantly differ across molecular subgroups ($p = 0.205$), with distributions as follows: 50.0% (2 out of 4 cases) in WNT, 32.5% (13 out of 40 cases) in SHH, 40.0% (8 out of 20 cases) in group 3, and 53.2% (33 out of 62 cases) in group 4. These findings contradict a previous study suggesting that suprasellar metastases are specific to group 4 and absent in other molecular subgroups.²³

Next, we delved into the imaging signal patterns of metastases. Notably, the matching C+/D+ pattern, indicating post-contrast enhancement and diffusion restriction, was the prevailing pattern in both ependymal (including 3rd V.I.R.) and leptomeningeal metastases, accounting for 46.4% and 57.0%, respectively (**Table 2**). However, within group 4, among the 33 patients with ependymal metastasis (including those in the 3rd V.I.R.), 18 (54.6%)

Table 1. Neuroradiological assessment of presurgical MR imaging features across the entire medulloblastoma dataset stratified by molecular subtypes

	Molecular Subtypes				Total	p value
	WNT	SHH	Group3	Group4		
Location of the primary tumor						
Midline vermis/fourth ventricle	79 (68.10%)	90 (43.69%)	54 (56.25%)	186 (57.94%)	409 (55.35%)	0.0002***
Fourth ventricle-uni-CPA	21 (18.10%)	17 (8.25%)	14 (14.58%)	53 (16.51%)	105 (14.21%)	0.031*
Fourth ventricle-bi-CPA	7 (6.03%)	17 (8.25%)	20 (20.83%)	61 (19.00%)	105 (14.21%)	<0.0001***
Within CPA	2 (1.72%)	2 (0.97%)	1 (1.04%)	1 (0.31%)	6 (0.81%)	0.308
Cerebellar hemisphere & midline/fourth ventricle	6 (5.17%)	35 (16.99%)	7 (7.29%)	14 (4.36%)	62 (8.39%)	<0.0001***
Cerebellar hemisphere	1 (0.86%)	45 (21.84%)	0 (0.00%)	6 (1.87%)	52 (7.04%)	<0.0001***
Total	116 (100.00%)	206 (100.00%)	96 (100.00%)	321 (100.00%)	739 (100.00%)	
Enhancement of the primary tumor						
The proportion of enhancement ^b	median (Q1, Q3)	34.80% (13.60%,53.85%)	21.85% (10.25%,37.85%)	32.60% (19.45%,47.15%)	14.40% (6.75%,28.30%)	<0.0001***
	mean (std)	34.46% (22.09%)	25.49% (18.92%)	34.87% (19.82%)	19.66% (16.96%)	25.58% (19.74%)
The proportion of enhancement, stratified by three classes	none/minimal ($\leq 10\%$)	21 (18.26%)	50 (24.51%)	12 (12.63%)	125 (39.18%)	<0.0001***
	heterogeneous (10%–50%)	60 (52.17%)	126 (61.76%)	63 (66.32%)	168 (52.66%)	417 (56.9%)
	diffuse (>50%)	34 (29.57%)	28 (13.73%)	20 (21.05%)	26 (8.15%)	108 (14.7%)
	Total	115 (100.0%)	204 (100.0%)	95 (100.00%)	319 (100.00%)	733 (100.0%)
The intensity of enhancement ^c	none enhancement	4 (3.48%)	7 (3.43%)	2 (2.11%)	17 (5.33%)	30 (4.1%)
	lower enhancement	45 (39.13%)	106 (51.96%)	35 (36.84%)	177 (55.49%)	363 (49.5%)
	equal enhancement	66 (57.39%)	91 (44.61%)	58 (61.05%)	125 (39.18%)	340 (46.4%)
	Total	115 (100.0%)	204 (100.0%)	95 (100.0%)	319 (100.0%)	733 (100.0%)
Tumor margin						
Well-defined	70 (60.34%)	88 (43.14%)	62 (64.58%)	208 (64.80%)	428 (58.07%)	<0.0001***
Ill-defined	46 (39.66%)	116 (56.86%)	34 (35.42%)	113 (35.20%)	309 (41.93%)	
Total	116 (100.00%)	204 (100.00%)	96 (100.00%)	321 (100.00%)	737 (100.00%)	
Peri-tumoral edema						
Y	49 (42.24%)	110 (53.92%)	44 (45.83%)	132 (41.12%)	335 (45.45%)	0.032*
N	67 (57.76%)	94 (46.08%)	52 (54.17%)	189 (58.88%)	402 (54.55%)	
Total	116 (100.00%)	204 (100.00%)	96 (100.00%)	321 (100.00%)	737 (100.00%)	
Cystic change/necrosis						
Y	85 (73.28%)	147 (72.06%)	71 (73.96%)	262 (81.62%)	565 (76.66%)	0.047*
N	31 (26.72%)	57 (27.94%)	25 (26.04%)	59 (18.38%)	172 (23.34%)	
Total	116 (100.00%)	204 (100.00%)	96 (100.00%)	321 (100.00%)	737 (100.00%)	

(Continued on next page)

Table 1. *Continued*

		Molecular Subtypes			Group3	Group4	Total	p value
		WNT	SHH	Hydrocephalus before surgery				
Y		90 (68.18%)	194 (76.68%)	84 (73.04%)	304 (78.55%)	672 (75.76%)	0.096	
N		42 (31.82%)	59 (23.32%)	31 (26.96%)	83 (21.45%)	215 (24.24%)		
Total		132 (100.00%)	253 (100.00%)	115 (100.00%)	387 (100.00%)	887 (100.00%)		
Intracranial solid metastases^d								
Y		4 (3.48%)	40 (19.61%)	20 (21.05%)	62 (19.44%)	126 (17.19%)	0.0004***	
N		111 (96.52%)	164 (80.39%)	75 (78.95%)	257 (80.56%)	607 (82.81%)		
Total		115 (100.00%)	204 (100.00%)	95 (100.00%)	319 (100.00%)	733 (100.00%)		

* $p = 0.01$ – 0.05 , ** $p = 0.001$ – 0.01 , *** $p < 0.0001$. p values were obtained using chi-square tests for categorical features, and Fisher's exact test was applied for categorical features with expected counts less than five. For continuous data, we utilized ANOVA and the Kruskal-Wallis rank-sum test to compare mean and median values, respectively. See also Figures S2–S5, and Table S1.

^aThe chi-square test was adopted and combined three CPA classes into one and two cerebellar classes into another when calculating the p value.

^bThe enhancement ratio quantifies the percentage of the primary tumor mass that exhibited enhancement.

^cUsing the superior sagittal sinus enhancement as the reference, the primary tumor mass enhancement was categorized into three groups: avid, mild, or non-enhancing.

^dThe complete assessment of signal patterns of intracranial solid metastases, encompassing ependymal and leptomeningeal metastases, is detailed in Table 2.

exhibited a C-/D+ pattern. Remarkably, the C-/D+ pattern was absent in ependymal metastases within group 3. Specifically, out of the total 22 patients displaying the C-/D+ ependymal metastasis (including 3rd V.I.R.) pattern, 18 belonged to group 4, and the remaining four patients were classified under the SHH subgroup, with no subjects in WNT and group 3. Therefore, ependymal (including 3rd V.I.R.) C-/D+ metastasis emerged as highly specific to group 4 ($p = 0.005$; Table 2) and was subsequently utilized as a characteristic feature in the machine learning binary classifier designed to differentiate between group 3 and group 4. This finding corroborates earlier research.²² It is noteworthy that although group 4 predominantly exhibited a C-/D+ pattern in ependymal locations, a matching C+/D+ pattern was primarily observed in leptomeningeal locations (62.5%; 20 out of 32). Conversely, in group 3, metastases predominantly demonstrated a matching C+/D+ pattern, regardless of their ependymal (75%; 6 out of 8) or leptomeningeal (57.1%; 8 out of 14) locations. Figure S3 showcases representative cases displaying diverse MR imaging patterns in leptomeningeal and ependymal metastases, respectively.

Primary tumor enhancement patterns are distinctive across molecular subgroups

The vast majority of medulloblastomas exhibited enhancement on post-contrast T1-weighted sequences (95.9%; 703 out of 733; Table 1). However, this enhancement was predominantly partial, with 56.9% (417 out of 733) of cases displaying enhancement covering between 10% and 50% of the primary tumor mass. Additionally, 28.4% (208 out of 733) showed none or minimal enhancement, i.e., less than 10% of the primary tumor mass, while 14.7% (108 out of 733) exhibited enhancement across more than 50% of the mass. The intensity of enhancement was almost evenly distributed between mild (49.5%; 363 out of 733) and avid (46.4%; 340 out of 733) categories (Table 1).

Significant differences were observed in both the proportion and intensity of enhancing primary tumors among molecular subgroups (proportion: $p < 0.0001$; intensity: $p = 0.0035$). Notably, group 4 displayed the lowest proportion of primary tumor mass exhibiting enhancement, with a median proportion of 14.40% (Q1: 6.75%, Q3: 28.30%). Conversely, both WNT and group 3 exhibited relatively high median proportions of primary tumor enhancement, with 34.80% (Q1: 13.60%, Q3: 53.85%) and 32.60% (19.45%, 47.15%), respectively. Corresponding with the low proportion of enhancement, the intensity of enhancement was also lowest in group 4, with only 39.18% (125 out of 319) of cases exhibiting strong avid enhancement, compared to 57.39% (66 out of 115) in WNT and 61.05% (58 out of 95) in group 3. SHH showed an intermediate enhancement pattern among the molecular subgroups, with 44.61% (91 out of 204) exhibiting strong avid enhancement and a median enhancement ratio of 21.85% (Q1: 10.25%, Q3: 37.85%). While previous studies have highlighted the minimal enhancement observed in group 4 in terms of proportion,^{15,16,22} our study further reveals that the intensity strength of enhancement in group 4 is also the lowest among molecular subtypes. This distinct enhancement pattern, encompassing both intensity and proportion, demonstrated high subgroup specificity and was leveraged in the machine learning three-class classifier (feature importance: 4.8%) and binary classifier (feature

Table 2. Patterns of intracranial solid metastases across the entire medulloblastoma dataset stratified by molecular subgroups

	Molecular subgroups				Total	<i>p</i> value
	WNT	SHH	Group3	Group4		
Intracranial solid metastases						
Y	4 (3.48%)	40 (19.61%)	20 (21.05%)	62 (19.44%)	126 (17.19%)	0.0004***
N	111 (96.52%)	164 (80.39%)	75 (78.95%)	257 (80.56%)	607 (82.81%)	
Total	115 (100.00%)	204 (100.00%)	95 (100.00%)	319 (100.00%)	733 (100.00%)	
Site of metastases^a						
Ependymal metastases	1 (25.0%; 1/4)	6 (15.0%; 6/40)	2 (10.0%; 2/20)	21 (33.9%; 21/62)	30 (23.8%; 30/126)	0.051
3rd V.I.R. metastases	2 (50.0%; 2/4)	9 (22.5%; 9/40)	8 (40.0%; 8/20)	22 (35.5%; 22/62)	41 (32.54%; 41/126)	0.327
Ependymal +3rd. V.I.R. metastases	2 (50.0%; 2/4)	13 (32.5%; 13/40)	8 (40.0%; 8/20)	33 (53.2%; 33/62)	56 (44.4%; 56/126)	0.205
Supratentorial leptomeningeal metastases	1 (25.0%; 1/4)	13 (32.5%; 13/40)	3 (15.0%; 3/20)	12 (19.4%; 12/62)	29 (23.0%; 29/126)	0.324
Infratentorial leptomeningeal metastases	1 (25.0%; 1/4)	21 (52.5%; 21/40)	14 (70.0%; 14/20)	21 (33.9%; 21/62)	57 (45.2%; 57/126)	0.017*
Supratentorial + Infratentorial leptomeningeal metastases	2 (50.0%; 2/4)	31 (77.5%; 31/40)	14 (70.0%; 14/20)	32 (51.6%; 32/62)	79 (62.7%; 79/126)	0.04*
Signal patterns of ependymal metastases						
Ependymal metastases ^b	C-/D+	0 (0.0%)	3 (50.0%)	0 (0.0%)	15 (71.4%)	18 (60%)
	C+/D+	1 (100.0%)	3 (50.0%)	2 (100.0%)	5 (23.8%)	11 (36.7%)
	C+/NA	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (4.8%)	1 (3.3%)
	Total	1 (100.0%)	6 (100.0%)	2 (100.0%)	21 (100.0%)	30 (100.0%)
3rd V.I.R. metastases	C-/D+	0 (0.0%)	1 (11.1%)	0 (0.0%)	7 (31.8%)	8 (19.51%)
	C+/D+	1 (50.0%)	4 (44.4%)	6 (75.0%)	12 (54.5%)	23 (56.10%)
	C+/D-	0 (0.0%)	3 (33.3%)	2 (25.0%)	1 (4.5%)	6 (14.63%)
	C+/NA	1 (50.0%)	1 (11.1%)	0 (0.0%)	2 (9.1%)	4 (9.76%)
	Total	2 (100.0%)	9 (100.0%)	8 (100.0%)	22 (100.0%)	41 (100.0%)
Ependymal and 3rd. V.I.R. metastases	C-/D+	0 (0.00%)	4 (30.77%)	0 (0.00%)	18 (54.55%)	22 (39.29%)
	C+/D+	1 (50.00%)	6 (46.15%)	6 (75.00%)	13 (39.39%)	26 (46.43%)
	C+/D-	0 (0.00%)	3 (23.08%)	2 (25.00%)	1 (3.03%)	6 (10.71%)
	C+/NA	1 (50.00%)	1 (7.69%)	0 (0.00%)	3 (9.09%)	5 (8.93%)
	Total	2 (100.00%)	13 (100.00%)	8 (100.00%)	33 (100.00%)	56 (100.00%)
C-/D+ pattern in ependymal and 3rd. V.I.R. metastases ^c		0 (0.0%; 0/4)	4 (10.0%; 4/40)	0 (0.0%; 0/20)	18 (29.03%; 18/62)	22 (17.46%; 22/126)
Signal patterns of leptomeningeal metastases						
Supratentorial leptomeningeal metastases	C-/D+	0 (0.0%)	1 (7.7%)	0 (0.0%)	0 (0.0%)	1 (3.45%)
	C+/D+	1 (100.0%)	5 (38.5%)	1 (33.3%)	5 (41.7%)	12 (41.38%)
	C+/D-	0 (0.0%)	5 (38.5%)	1 (33.3%)	3 (25.0%)	9 (31.03%)
	C+/NA	0 (0.0%)	2 (15.4%)	1 (33.3%)	4 (33.3%)	7 (24.14%)
	Total	1 (100.0%)	13 (100.0%)	3 (100.0%)	12 (100.0%)	29 (100.00%)
Infratentorial leptomeningeal metastases	C-/D+	0 (0.0%)	1 (4.8%)	0 (0.0%)	0 (0.0%)	1 (1.75%)
	C+/D+	0 (0.0%)	12 (57.1%)	8 (57.1%)	15 (71.4%)	35 (61.40%)
	C+/D-	0 (0.0%)	3 (14.3%)	4 (28.6%)	1 (4.8%)	8 (14.04%)
	C+/NA	1 (100.0%)	5 (23.8%)	2 (14.3%)	5 (23.8%)	13 (22.81%)
	Total	1 (100.0%)	21 (100.0%)	14 (100.0%)	21 (100.0%)	57 (100.00%)

(Continued on next page)

Table 2. Continued

	Molecular subgroups					Total	<i>p</i> value
	WNT	SHH	Group3	Group4			
Supratentorial and infratentorial leptomeningeal metastases	C-/D+	0 (0.0%)	2 (6.5%)	0 (0.0%)	0 (0.0%)	2 (2.5%)	0.517
	C+/D+	1 (50.0%)	16 (51.6%)	8 (57.1%)	20 (62.5%)	45 (57.0%)	0.547
	C+/D-	0 (0.0%)	6 (19.4%)	4 (28.6%)	4 (12.5%)	14 (17.7%)	0.637
	C+/NA	1 (50.0%)	7 (22.6%)	2 (14.3%)	8 (25.0%)	18 (22.8%)	0.654
	Total	2 (100.0%)	31 (100.0%)	14 (100.0%)	32 (100.0%)	79 (100%)	

Percentages may exceed 100% due to certain subjects exhibiting two metastases with different signal patterns.

"C-/D+" denotes cases with diffusion restriction but minimal or no postcontrast enhancement, while "C+/NA" indicates postcontrast enhancement with unavailable diffusion information due to the absence of a DWI sequence.

p* = 0.01–0.05, *p* = 0.001–0.01, ****p* = 0.0001–0.001. *p* values were obtained using chi-square tests for categorical features, and Fisher's exact test was applied for categorical features with expected counts less than five.

Key abbreviation: 3rd V.I.R. = the third ventricular infundibular recess.

See also Figure S3.

^aThe percentage and corresponding *p* value for each metastatic site (categorized into 6 classes as listed below) were computed specifically among subjects exhibiting intracranial solid metastases.

^bThe percentage and corresponding *p* value for each signal pattern of ependymal metastases were computed among subjects exhibiting ependymal metastases.

^cThe percentage and corresponding *p* value were computed within subjects exhibiting intracranial solid metastases.

importance: 10.2%) to distinguish between molecular subgroups (Figures 3A and 3C). Figure S4 illustrates representative cases exhibiting diverse MR imaging patterns of enhancement within primary tumors, encompassing variations in both enhancement proportion and intensity.

Correlative image phenotypes and vascular density analysis in medulloblastoma subgroups

To unravel potential correlative image phenotypes associated with medulloblastoma, we conducted 3D immunostaining on 16 random formalin-fixed specimens. We utilized anti-CD31 antibody staining to specifically target endothelial cells and assess vascular density. Each subgroup was represented by four specimens. Figure 4 and Table S6 present the qualitative and quantitative results of blood vessel density via immunostaining across the four molecular subgroups. The mean vascular density exhibited the highest values in WNT (10.3% ± 1.55%), followed by group 3 (5.5% ± 1.23%), group 4 (4.7% ± 0.47%), and SHH (4.2% ± 0.75%). This observation aligns with prior studies,²⁴ reinforcing the notion of notably elevated vascular density within the WNT subgroup in comparison to others. Correspondingly, WNT exhibits the most pronounced enhancement in T1-contrast enhanced MRIs, as evidenced by Table 1. No significant differences were observed in vascular density between SHH, group 3, and group 4 subtypes (Figure 4K). However, group 4 and SHH exhibited significantly lower enhancement on T1E MRI compared to group 3, with median proportions of enhancement for group 4 at 14.40% and for SHH at 21.85%, in contrast to 32.60% in group 3 (Table 1). A plausible explanation is that SHH and group 4 may uphold blood-brain barrier (BBB) integrity, restricting contrast agent passage into the primary tumor and thus limiting signal enhancement on T1E MRI. The preservation of BBB integrity in a highly malignant embryonal tumor like medulloblastoma, although unexpected, is consistent with prior studies observed in group 4 and SHH subtypes.^{5,22,24,25} Conversely, group 3 may exhibit a disrupted BBB, facilitating contrast agent leakage into the primary tumor and

thereby enhancing the signal on T1E MRI.^{5,26,27} Video S1 showcases four representative tissue samples with 3D immunostaining (CD31) of endothelial cells for WNT, SHH, group 3, and group 4 molecular subgroups of medulloblastoma, respectively.

Other differential MR imaging signatures across molecular subgroups

Tumor margin

Significant difference was observed in tumor margin characteristics among molecular subgroups (*p* < 0.0001; Table 1). Ill-defined tumor margins were notably present in 56.86% (116 out of 204) of SHH patients, serving as a distinctive imaging feature for the SHH subgroup. In contrast, only 39.66% (46 of 116) of WNT, 35.42% (34 of 96) of group 3, and 35.20% (113 of 321) of group 4 patients exhibited ill-defined tumor margins. Additionally, machine learning analysis identified the definition of tumor margins as a predictor of medulloblastoma subgroups, with a relative importance of 2.9% as indicated by the Shapley value (out of a total of 100%) in the three-class classifier designed to distinguish WNT, SHH, and G3/G4 subgroups (Figure 3A). Figure S5 depicts representative cases showcasing well-defined and ill-defined tumor margins.

Peri-tumoral edema and cystic change/necrosis

A notable proportion of medulloblastomas exhibited peri-tumoral edema (45.45%, 335 out of 737) and cystic change/necrosis within their primary tumors (76.66%, 565 out of 737). Significant statistical differences were observed in both peri-tumoral edema (*p* = 0.032) and cystic change/necrosis (*p* = 0.047) among molecular subgroups, with SHH displaying the highest rate of peritumoral edema at 53.92%, while group 4 exhibited the highest rate of cystic change/necrosis at 81.62% (Table 1). Interestingly, machine learning analysis identified peri-tumoral edema as a predictive feature, with an importance of 3.3% in the binary classifier aimed at distinguishing between group 3 and group 4, despite the lack of a statistical difference observed for this feature between the two groups. Figure S5 showcases the MR imaging

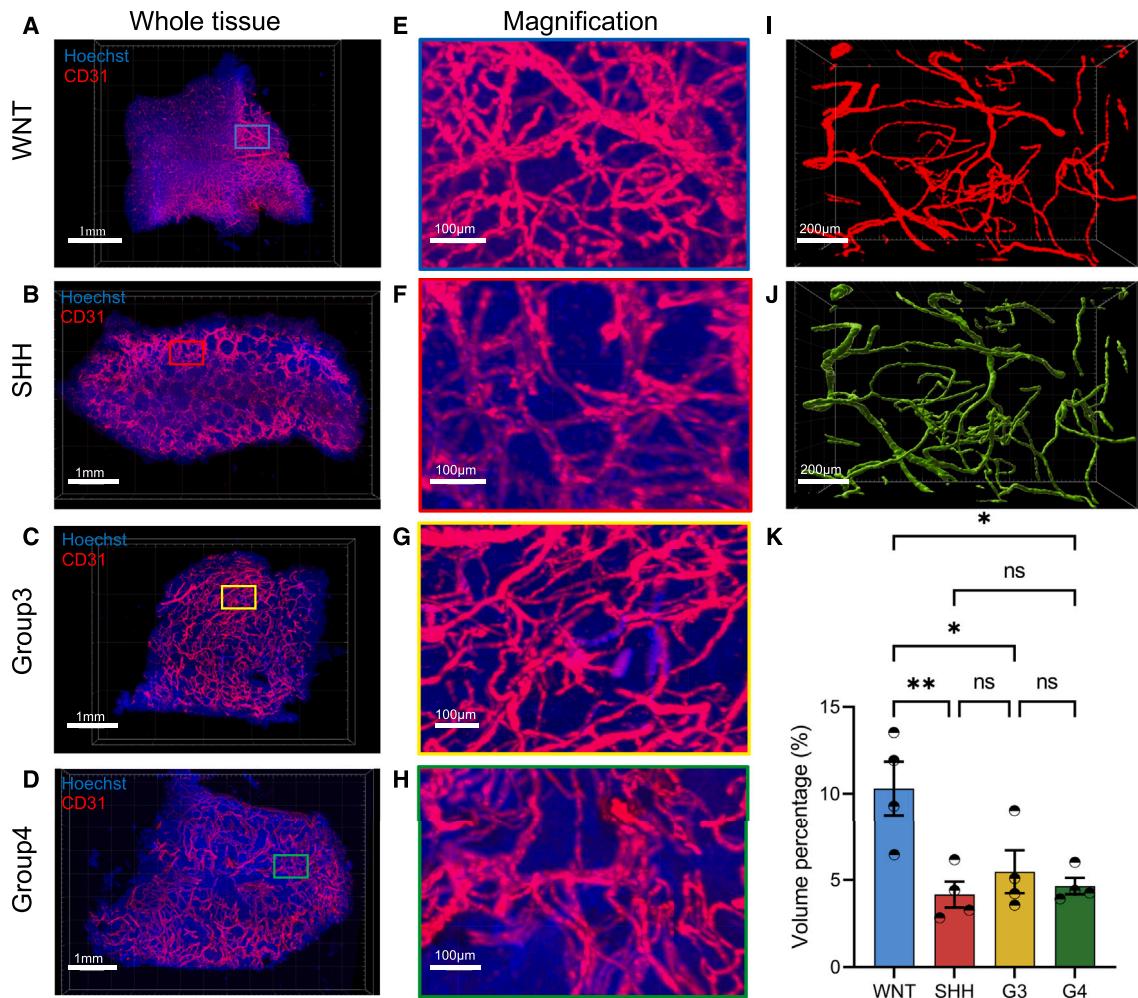


Figure 4. Analysis of blood vessel density via immunostaining in medulloblastoma surgical specimens

(A–D) Representative samples displaying immunostaining (CD31) of endothelial cells across WNT, SHH, group 3, and group 4 molecular subgroups of medulloblastoma.

(E–H) Enlarged crops of the regions highlighted in rectangles in (A–D), respectively.

(I) Volumetric rendering illustrating the distribution of vessels in a representative medulloblastoma sample.

(J) Surface rendering of the segmented vessels depicted in (I).

(K) Comparative analysis of vascular density across the four molecular subgroups in medulloblastoma. Sample size (n) equals 4 in each group. Statistical significance is denoted by asterisks: * indicates $p = 0.01\text{--}0.05$, ** indicates $p = 0.001\text{--}0.01$, and "ns" denotes not significant (p value >0.05). One-way ANOVA. Error bars represent the mean \pm SEM (the standard error of the mean).

See also Table S6.

characteristics of peri-tumoral edema and cystic change/necrosis in the primary tumor.

No significant statistical difference was found among the molecular subgroups regarding the presence of hydrocephalus before surgery ($p = 0.096$) (Table 1).

Integrating computational and manual MR imaging signatures enhances diagnostic performance

Next, we conducted extensive comparative analyses encompassing both computational and manual MR imaging signatures, elucidating their complementary value in molecular subgroup prediction. It is noteworthy that previous endeavors in machine learning applications for medulloblastomas have solely focused on either manual qualitative characterization^{16,20,22,28–30} or

limited computational radiomics,^{5,15} constraining their broader applicability.

We explored three input schemes: (1) radiomics, (2) manual MR imaging features, and (3) the combination of radiomics and MR imaging features for the three-class AI model. The pure radiomics scenario achieved an average AUC of 0.821 across the three testing sets, serving as the benchmark for comparison (Table S7). Human-accessible MR imaging features, combining all features, yielded an AUC of 0.771; for combining selected characteristic MR imaging features demonstrating statistical significance (tumor location, enhancement pattern, tumor margin, and pattern of metastases; Table S7), it reached a slightly higher AUC of 0.774. The pure radiomics scheme outperformed pure MR imaging features. This finding is consistent with the feature

importance comparison depicted in **Figure 3A**, highlighting the efficacy of computer-generated radiomics in providing multidimensional and distinctive information for molecular diagnosis. Notably, the combination of radiomics and MR imaging features achieved the highest average AUC of 0.853 across all three testing sets, surpassing the capabilities of either radiomics or human-accessible MR imaging features alone. This demonstrated a 7.9% point improvement for the human-accessible MR imaging features scheme and a 3.2% point improvement for the radiomics scheme in the averaged AUC metric. Specifically, each single characteristic MR imaging feature contributed incrementally to performance improvement (**Table S7**), highlighting the highly complementary values of manual characterization of MR imaging to the machine learning model.

The trend persists in the binary classifier analysis. The pure radiomics scenario, serving as the benchmark, achieved an average AUC of 0.836 across the three testing sets (**Table S8**). Combining all human-accessible MR imaging features resulted in an average AUC of 0.698. Notably, improved performance was observed when integrating benchmark radiomics with characteristic MR imaging features—specifically, enhancement pattern, metastases pattern, and peri-tumoral edema. This combined approach yielded an average AUC of 0.844 over the three testing sets, representing a substantial 14.6% point improvement compared to the human-accessible MR imaging features alone and a 0.8% point improvement over the pure radiomics scheme (**Table S8**). This underscores the efficacy of integrating both human characterization and radiomics in molecular diagnosis.

Patient characteristics in the entire medulloblastoma cohort

Among 934 patients (65.20% male; 85.97% treated in East Asia) with overall survival, median follow-up was 22 months (range 0–258.4), during which time there were 179 (19.16%) deaths. The one-year overall survival probability stood at 90.9% (95% CI, 88.9%–92.9%), followed by a three-year overall survival probability of 75.4% (95% CI, 71.9%–79.0%), and a five-year overall survival probability of 68.2% (95% CI, 63.8%–72.9%). The median age at diagnosis was 8 years (range 0.3–54), and 76.02% had hydrocephalus before primary tumor resection (**Table S9**).

Of the 934 patients, 792 had histological subgrouping available with the following distribution (**Table S9**): medulloblastoma with extensive nodularity (MBEN)—34 (4.29%), large cell/anaplastic (LCA)—35 (4.42%), desmoplastic/nodular (DN)—164 (20.71%), and classic—559 (70.58%). Additionally, 798 patients had a tumor texture description with the following distribution: soft—501 (62.78%), soft-hard-mixed—165 (20.68%), and hard—132 (16.54%). Molecular subgroup distribution showed WNT—132 (14.13%), SHH—253 (27.09%), and non-WNT/non-SHH—549 (58.78%). Further subclassification of non-WNT/non-SHH showed G3—115 (20.95%) and G4—387 (70.49%); Subclassification of SHH showed TP53-mutant—60 (23.72%) and TP53-wildtype—150 (59.29%). 47 patients within non-WNT/non-SHH and 43 patients within SHH were not further subclassified (**Table S9**).

We assessed survival in a Kaplan-Meier analysis of the medulloblastoma subgroup classification. The log rank test revealed highly significant differences in survival across molecular subgroups according to the classic 2016 WHO Classification, i.e.,

WNT, SHH, group 3, and group 4 ($p = 0.00014$; **Figure 5E**). Favorable survival outcomes for WNT subtype were confirmed in Kaplan-Meier analysis. Meanwhile, the histopathological subgroups also demonstrated noteworthy differences in survival, albeit less pronounced than those observed among the molecular subgroups ($p = 0.005$; **Figure 5K**).

Association of clinicopathological features and the collective molecular subgroups

In this extensive medulloblastoma cohort, with a particular focus on the Asian demographic, we delved deeper into the correlation between clinicopathological features and the molecular subgroups (WNT, SHH, G3, and G4; where SHH is additionally categorized into SHH TP53+ and SHH TP53-). **Tables 3** and **S10**, and **Figure 5** show demographic, clinicopathological, and prognostic features of molecular subgroups and their association. Clinicopathological features were non-randomly distributed in subgroups.

WNT subgroup occurs primarily in children to early adulthood (median age of diagnosis [IQR], 9 [7, 12]) and exhibits a balanced male versus female ratio (male percentage, 43.94%). Classic histology characterizes the majority of WNT tumors (91.38%), consistent with previous investigations in EU/USA cohorts (**Figure 5D**; **Table 3**).³¹ WNT tumors exhibit the lowest proportion of hard tumor texture (9.02%) and the highest proportion of soft tumor texture (77.05%) among the molecular subgroups (**Figure 5C**; **Table 3**). Preoperative hydrocephalus is least prevalent in the WNT subgroup (68.18%; **Table 1**). The outcomes for patients with WNT are the most favorable among molecular subgroups, with 3-year overall survival of 91.7% and 5-year overall survival of 88.3% (**Figure 5E**; **Table 3**).

SHH represents the most common molecular subgroup across both infant and adult populations (**Figure 5B**), displaying a higher prevalence among males, with a ratio of approximately 2:1. Notably, classic and DN histologies occur at comparable frequencies (37.50% and 45.98%; **Figure 5D**; **Table 3**), consistent with findings from EU/USA cohorts.³¹ Following the classification outlined in the 2021 WHO guidelines, which delineate SHH-activated TP53 mutant tumors as a distinct high-risk category, we conducted TP53 sequencing across 210 SHH samples. Subgroup SHH TP53+ constitutes 28.57% of SHH patients, with a notably heightened prevalence of large cell/anaplastic (LCA) histology compared to SHH TP53- counterparts ($p = 0.0065$; **Table S10**). Kaplan-Meier analysis of survival probability reveals no statistically significant disparity between SHH TP53- and SHH TP53+ across all ages ($p = 0.068$; **Figure 5F**). However, within the 4–16 age range, SHH TP53+ exhibits a significantly poorer prognosis compared to SHH TP53- ($p = 0.022$; **Figure 5G**), aligning with prior research indicating TP53 mutational status as a prognostic marker specifically within the alpha SHH subgroup aged 4–16 years.³² Notably, no significant differences are observed in age, gender, or tumor texture between SHH TP53+ and SHH TP53-. **Table S10** provides a comprehensive overview of the clinicopathological features associated with the SHH subgroup, stratified by TP53 mutational status.

The male-to-female ratio in group 3 and group 4 medulloblastoma is consistently higher than 2:1, aligning with findings from other EU/USA cohorts.³¹ Among the four molecular subgroups, group 4 exhibits the highest prevalence (43.63%; **Table 3**),

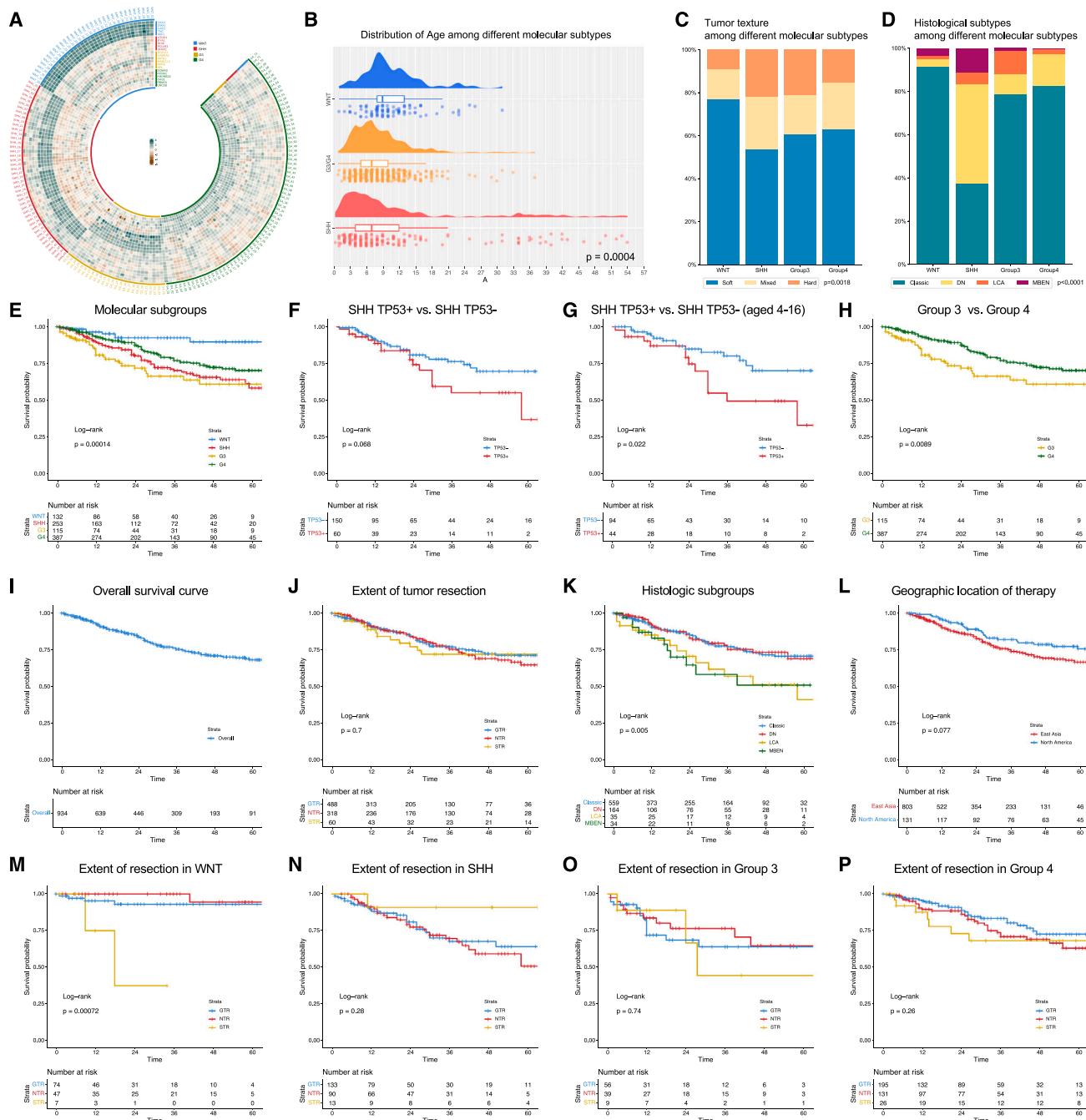


Figure 5. Clinicopathological characteristics and survival analysis of subgroups in medulloblastoma

(A) An expression heatmap based on 22 signature genes specific to medulloblastoma (NanoString CodeSet), derived from 174 primary samples at Beijing Tiantan Hospital.

(B-D) Distribution of age, tumor texture, and histologic subgroups across molecular subtypes. The chi-square test was adopted to calculate the *p*-value.

(E-H) Kaplan-Meier curves illustrate overall survival by molecular classification (WNT, SHH, G3, and G4). TP53 SHH subgroup analysis is limited to patients aged 4–16 in (G).

(I) Five-year overall survival curve for all medulloblastoma patients (*n* = 934) in the study.

(J-L) Kaplan-Meier analysis for extent of tumor resection, histologic subgroups, and geographic location of therapy.

(M-P) Kaplan-Meier analysis for extent of resection by molecular subgroups. *p* values denote log rank test across different strata in Kaplan-Meier analysis. Key abbreviations: MB, medulloblastoma; WNT, wingless; SHH, sonic hedgehog; G3, group 3; G4, group 4; LCA, large cell/anaplastic; DN, desmoplastic/nodular; MBEN, medulloblastoma with extensive nodularity; GTR, gross total resection; NTR, near-total resection; STR, sub-total resection.

See also Tables S9 and S10.

Table 3. Association of clinicopathological features and the collective molecular subgroups with a focus on the Asian population

	Molecular Subtypes				Total	p value
	WNT	SHH	Group3	Group4		
Frequency N (%)	132 (14.88%)	253 (28.52%)	115 (12.97%)	387 (43.63%)	887 (100.00%)	
Age						
mean (SD)	10.077 (4.812)	10.370 (10.911)	7.359 (5.058)	8.713 (4.801)	9.214 (7.190)	0.0004***
median (Q1, Q3)	9 (7, 12)	7 (4, 12)	6.2 (4, 9.5)	8 (6, 11)	8 (5, 11)	
Sex						
M	58 (43.94%)	165 (65.22%)	78 (67.83%)	277 (71.58%)	578 (65.16%)	<0.0001****
F	74 (56.06%)	88 (34.78%)	37 (32.17%)	110 (28.42%)	309 (34.84%)	
Total	132 (100.00%)	253 (100.00%)	115 (100.00%)	387 (100.00%)	887 (100.00%)	
Race						
Asian	124 (93.94%)	220 (86.96%)	94 (81.74%)	336 (86.82%)	774 (87.26%)	0.0656
Black or African American	1 (0.76%)	0 (0.00%)	0 (0.00%)	4 (1.03%)	5 (0.56%)	
White	5 (3.79%)	21 (8.30%)	16 (13.91%)	37 (9.56%)	79 (8.91%)	
Other or Not Reported	2 (1.52%)	12 (4.74%)	5 (4.35%)	10 (2.58%)	29 (3.27%)	
Total	132 (100.00%)	253 (100.00%)	115 (100.00%)	387 (100.00%)	887 (100.00%)	
Tumor texture						
Hard	11 (9.02%)	47 (22.07%)	20 (21.28%)	51 (15.18%)	129 (16.86%)	0.0018**
Mixed	17 (13.93%)	52 (24.41%)	17 (18.09%)	73 (21.73%)	159 (20.78%)	
Soft	94 (77.05%)	114 (53.52%)	57 (60.64%)	212 (63.10%)	477 (62.35%)	
Total	122 (100.00%)	213 (100.00%)	94 (100.00%)	336 (100.00%)	765 (100.00%)	
Histological Subtypes						
Classic	106 (91.38%)	84 (37.50%)	67 (78.82%)	269 (82.52%)	526 (70.04%)	<0.0001****
DN	4 (3.45%)	103 (45.98%)	8 (9.41%)	48 (14.72%)	163 (21.70%)	
LCA	2 (1.72%)	12 (5.36%)	9 (10.59%)	8 (2.45%)	31 (4.13%)	
MBEN	4 (3.45%)	25 (11.16%)	1 (1.18%)	1 (0.31%)	31 (4.13%)	
Total	116 (100.00%)	224 (100.00%)	85 (100.00%)	326 (100.00%)	751 (100.00%)	
LCA-enrichment (%)	1.72%	5.36%	10.59%	2.45%	4.03%	0.0055**
Extent of tumor resection						
GTR	74 (57.81%)	133 (56.36%)	56 (53.85%)	195 (55.40%)	458 (55.85%)	0.9314
NTR	47 (36.72%)	90 (38.14%)	39 (37.50%)	131 (37.22%)	307 (37.44%)	
STR	7 (5.47%)	13 (5.51%)	9 (8.65%)	26 (7.39%)	55 (6.71%)	
Total	128 (100.00%)	236 (100.00%)	104 (100.00%)	352 (100.00%)	820 (100.00%)	
Radiotherapy						
Y	114 (95.00%)	180 (82.19%)	71 (80.68%)	295 (93.35%)	660 (88.83%)	<0.0001****
N	6 (5.00%)	39 (17.81%)	17 (19.32%)	21 (6.65%)	83 (11.17%)	
Total	120 (100.00%)	219 (100.00%)	88 (100.00%)	316 (100.00%)	743 (100.00%)	
Chemotherapy						
Y	98 (85.96%)	186 (85.71%)	65 (78.31%)	275 (89.00%)	624 (86.31%)	0.001***
N	16 (14.04%)	31 (14.29%)	18 (21.69%)	34 (11.00%)	99 (13.69%)	
Total	114 (100.00%)	217 (100.00%)	83 (100.00%)	309 (100.00%)	723 (100.00%)	
Survival probability						
1-year OS (%)	96.1%	90.3%	76.6%	91.7%	90.9%	0.00014***
2-year OS (%)	91.7%	79.3%	70.6%	86.1%	83.9%	
3-year OS (%)	91.7%	68.0%	67.9%	74.7%	75.6%	
4-year OS (%)	88.3%	63.5%	64.2%	68.7%	71.0%	
5-year OS (%)	88.3%	59.8%	64.2%	65.2%	68.0%	

(Continued on next page)

Table 3. Continued

	Molecular Subtypes		Group3	Group4	Total	<i>p</i> value
	WNT	SHH				
Survival status						
Alive	124 (93.94%)	197 (77.87%)	84 (73.04%)	315 (81.40%)	720 (81.17%)	0.0001***
Decease	8 (6.06%)	56 (22.13%)	31 (26.96%)	72 (18.60%)	167 (18.83%)	
Total	132 (100.00%)	253 (100.00%)	115 (100.00%)	387 (100.00%)	887 (100.00%)	

p* = 0.01–0.05, *p* = 0.001–0.01, ****p* = 0.001–0.0001, *****P* = <0.0001. *p* values were obtained using chi-square tests for categorical features, and Fisher's exact test was applied for categorical features with expected counts less than five. For continuous data, we utilized ANOVA test to compare mean values.

See also Table S10.

whereas group 3 has the lowest occurrence (12.97%). Notably, group 3 tumors show a significantly higher incidence of LCA pathology compared to group 4 (group 3: 10.59%, group 4: 2.45%, *p* = 0.0055; Table 3). Furthermore, patients within group 3 experience significantly worse prognostic outcomes compared to those in group 4 (*p* = 0.0089, Figure 5H; Table 3).

Comparison of geographic treatment location and tumor resection extent

No significant difference in survival rates was observed between cohorts from East Asia and North America (*p* > 0.05, Figure 5L; Table S9). Furthermore, there were no notable disparities between the receipt of radiotherapy and chemotherapy across East Asia and North America (radiotherapy: *p* = 1, chemotherapy: *p* = 0.3351, Table S9). However, a distinct contrast emerged in the extent of tumor resection between patients treated in East Asia and those in North America (*p* < 0.0001, Table S9). Notably, 23.46% of patients in North America underwent subtotal resection (STR), contrasting with only 5.22% in East Asia. In East Asia, 56.82% and 37.96% of patients underwent gross-total resection (GTR) and near-total resection (NTR), respectively, compared to 51.85% and 24.69% in North America. This suggests a more aggressive tumor resection approach in East Asia as opposed to North America.

While no statistically significant difference was observed in the occurrence of intracranial solid metastases or tumor texture between cohorts from East Asia and North America, there was a notable discrepancy in the proportion of patients presenting with hydrocephalus before surgery. Specifically, the incidence of hydrocephalus before surgery was nearly double in East Asia (81.32%) compared to North America (43.51%) (*p* < 0.0001, Table S9). The comparatively reduced incidence of hydrocephalus before surgery in the North America cohort may reflect patients in North America receive timelier diagnoses compared to those in the East Asia cohort, potentially before the tumor grows too large that block the normal circulation of cerebrospinal fluid (CSF) in the fourth ventricle of the brain and causes obstructive hydrocephalus.

Next, we investigated the prognostic value of the extent of tumor resection. Across 866 patients, no significant disparity in survival was observed among those undergoing GTR, NTR, or STR (*p* = 0.7; Figure 5J). We further examined the prognostic impact of resection extent within specific molecular subgroups. Notably, WNT subgroup patients exhibited significant differences, with STR indicating the poorest prognostic effect (*p* = 0.00072; Fig-

ure 5M). Conversely, no substantial survival advantage was evident with greater resection extent for patients with SHH, group 3, and group 4 tumors (Figures 5N–5P). It's important to acknowledge the limited number of patients undergoing STR in our study (60 patients, ratio: 6.93%), which may influence the conclusion regarding its prognostic significance compared to GTR/NTR due to the small sample size. Overall, our findings, derived from the largest cohort to date evaluating the extent of resection's value for medulloblastoma (866 patients; primarily managed in East Asia), corroborate previous research suggesting that overly aggressive surgical resection, especially in cases of heightened neurological morbidity risk, may not be recommended as there is no definitive benefit of GTR over NTR.³³

DISCUSSION

Molecular testing forms the cornerstone of contemporary precision oncology, yet its accessibility is hindered by the intricate and expensive techniques involved, including clustering, gene expression analysis, and DNA methylation profiling.⁷ This challenge exacerbates the lack of access to molecular testing for most cancer patients worldwide,⁶ a particularly pressing issue in the context of medulloblastoma, the most prevalent malignant brain tumor in children. In this study, we developed an artificial intelligence-driven diagnostic workflow that integrates machine learning with presurgical MR imaging, offering a noninvasive and cost-effective approach to predict medulloblastoma molecular subgroups. Through comprehensive validation on an international cohort of 689 patients from 11 medical centers, our AI-enabled diagnostic models demonstrated accurate presurgical prediction of molecular subgroups with high fidelity (AUCs ≥ 0.8 and F1 scores ≥ 0.7 ; internally, externally, and consecutively). These findings highlight the transformative potential of imaging-based AI-enabled molecular diagnosis in medulloblastoma, suggesting its capability to reduce health disparities, enhance risk stratification, and accelerate the personalization of therapy, especially in resource-constrained regions.

Presurgical molecular classification enabled by the developed AI strategy has a critical impact on the surgical planning and care of MB patients. Although standard treatment of MB requires maximal safe tumor resection, larger resections are associated with increased surgical complications that can lead to neurological morbidity, cognitive deficits, speech and auditory deficits, and ataxia.^{34–37} Recent studies have suggested that surgical goals should be tailored based on molecular subgroups.^{5,33,38,39}

Furthermore, the presurgical prediction of molecular subgroup allows the consideration of clinical trials that introduce subgroup-specific neoadjuvant chemotherapy and create future opportunities for an individualized approach to MB treatment.^{40–42} Neoadjuvant chemotherapy aims to decrease tumor size prior to safer surgical resection and could potentially yield substantial clinical benefits, particularly for non-WNT/non-SHH tumors, which are typically associated with a higher incidence (31–35%) of cerebellar mutism syndrome.⁴³

ML presents an unparalleled opportunity to mine high-dimensional image data and uncover subtle signatures that enhance precision analytics. However, prior studies exploring radiogenomics approaches in medulloblastoma have been inconclusive due to limited dataset validation.^{5,13–16} Most importantly, existing machine learning methods in this context have predominantly depended on features extracted solely from isolated tumor volumes within the intra-tumoral region, overlooking the crucial tumor-brain relationships and distinct characteristics of peri-tumoral and extra-tumoral regions.^{5,15} In contrast, our approach adopts a comprehensive strategy by integrating computational radiomics extracted from both intra-tumoral and peri-tumoral regions, complemented by human qualitative characterization of MR imaging. This inclusive methodology significantly boosted the accuracy of AI-enabled molecular subgroup diagnosis for medulloblastoma. Meanwhile, the feature importance analysis fills the gap, offering a holistic perspective of the tumor landscape in molecular subgrouping.

Leveraging the largest molecularly characterized medulloblastoma (MB) dataset with MR imaging, our study provides an in-depth analysis of MR imaging signatures across molecular subgroups, substantially extending previous investigations constrained by limited samples.^{16,20,22,28–30} Specifically, the location of CP/CPA and lateral recess extension are present across all subgroups, contrary to earlier studies that suggested their uniqueness to WNT tumors.^{16,20} In line with earlier studies, the cerebellar hemisphere remains highly specific to the SHH subgroup; however, we delineate that the primary tumor location for SHH remains the midline vermis/fourth ventricle, with the cerebellum serving as a secondary site.^{16,21} Intriguingly, our study shows that intracranial solid metastasis at diagnosis, while rare in WNT, is nearly equally prevalent in SHH, group 3, and group 4, reversing earlier assumptions about the rarity of metastatic dissemination in SHH.²³ In our dataset, leptomeningeal metastases involving both infratentorial and supratentorial regions are most prevalent in the SHH subgroup. We corroborated a notable finding that ependymal (including 3rd V.I.R.) C-/D+ metastasis is highly specific to group 4, while absent in group 3.²² Our analysis also reveals that group 4 tumors exhibit a minimal degree of enhancement in both intensity and proportion, suggesting a preserved BBB.^{5,22,24} Additionally, we identified that SHH tumors display significantly ill-defined tumor margins, a distinctive imaging feature not emphasized in prior studies. This comprehensive investigation of MRI characteristics replenishes the understanding of MB from a nuanced radiographic imaging perspective, which complements existing genetic knowledge of molecular subgroups.

Notably, MB embodies a worldwide malady, yet research pertaining to it has been predominantly focused on European and American (EU/USA) populations.⁴⁴ The substantial Asian population, constituting about 60% of the global populace, has

been significantly underrepresented in published MB literature.^{2,45} Utilizing our international MB database, we conducted a detailed comparison of treatment strategies and clinicopathological variables between East Asia and North America. Our findings reveal that East Asian medical practices tend to adopt a more aggressive tumor resection strategy, a contrast to more conservative approaches observed in North America. However, in this largest cohort examined to date analyzing the impact of resection extent on survival, our results reveal no significant survival advantage of gross total resection over near-total resection, corroborating previous findings.³³ This suggests the potential for re-evaluating surgical strategies, particularly in East Asia, where the risk of neurological morbidity from attempting to remove small residual tumor portions is high.

Limitations of the study

Our study has several limitations. Extensive evaluation through prospective studies, larger populations, and clinical trials is indispensable before the AI model's clinical implementation. Generalizability across geographic regions beyond East Asia and North America remains to be tested to confirm its broad utility. Moreover, manual inspection revealed that the model yielded less promising performance in MR imaging with pronounced artifacts, particularly for patients with implanted ventriculoperitoneal (VP) shunt valves.⁴⁶ These metal pumps and shunt tubes, designed to drain excess fluid and alleviate intracranial pressure, cause substantial artifacts on MRI scans, some exceeding 30 mm, which obstruct tumor visualization and impact feature extraction (Figure S6). The presence of these artifacts disrupts MRI preprocessing including intensity normalization, essential for accurate feature representation, thus compromising the AI model's diagnostic guidance. The ability of the AI diagnostic model to accurately diagnose cases with pronounced MRI artifacts represents a critical area for future research. Additionally, we explored the use of deep learning in modeling MRI imaging directly for molecular subgroup prediction, it demonstrated limited efficacy, likely due to the data-intensive demands of patient-level classification. Future studies with accumulated larger sample sizes could consider combining deep learning and MR imaging signatures, given that the continuous learning capabilities of deep learning could enhance test performance over time as the training dataset expands.⁴⁷ In addition, future studies could explore the clinical utility of integrating image-based biomarkers for predicting outcomes in patients with medulloblastoma.

In conclusion, our study demonstrates how ML-based methods derived from MRI data have the potential to augment existing molecular testing and create an alternative pathway for presurgical, noninvasive molecular diagnosis of medulloblastoma. If confirmed in clinical settings, this can significantly advance the clinical management of MB patients. Additionally, the underlying open dataset (Table S16) can be used to advance global investigation of MB, and to better understand, recognize, and enhance MB survival for patients treated across the world.

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE

• RESOURCE AVAILABILITY

- Lead contact
- Materials availability
- Data and code availability

• EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

- MRI acquisition

• METHOD DETAILS

- Dataset curation
- Study design
- Molecular subgroup analysis
- NanoString-based subgrouping
- Next generation sequencing
- Molecular characterization through DNA methylation profiling in the consecutive testing set
- Extent of tumor resection
- Qualitative MRI imaging features
- MRI pre-processing
- ML-enabled molecular subgroup prediction
- Deep learning automated tumor segmentation
- Deep learning in molecular subgroup prediction
- Manual annotation procedure
- Immunostaining in medulloblastoma surgical specimens
- Missing data imputation

• QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.ccel.2024.06.002>.

ACKNOWLEDGMENTS

The authors extend their gratitude to the Children's Brain Tumor Network (CBTN, <https://cbtn.org/>) and the Children's Hospital of Philadelphia (CHOP). Furthermore, the authors wish to acknowledge the China Pediatric Neurosurgery Federation (CPNF), National Center for Neurological Disorders, for their collaboration in granting access to the medulloblastoma database scrutinized within this study. Grateful recognition is directed toward the following doctors: Yunwei OU, Yongji TIAN, Wei LIU, Chunde LI, Zhenyu MA, from Beijing Tiantan Hospital, Capital Medical University; Rong ZHANG, from Huashan Hospital, Fudan University; Qing CHANG, from the Department of Neuropathology, Beijing Neurosurgical Institute, Beijing Tiantan Hospital, Capital Medical University; Junping ZHANG, from Sanbo Brain Hospital, Capital Medical University; Guangyu WANG, from Qilu Children's Hospital of Shandong University; Jie ZHAO, from Xiangya Hospital, Central South University; Jie GONG, from Qilu Hospital of Shandong University and Institute of Brain and Brain-Inspired Science; Ping LIANG, from the Children's Hospital of Chongqing Medical University; Xiaosheng HE, from Xijing Hospital.

We express our sincere appreciation to the anonymous reviewers for their valuable and constructive feedback, which greatly enhanced the quality of this paper. The molecular subgroup analysis and dataset curation at Beijing Tiantan Hospital was partially supported by a grant from Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, grant number 21KT012, and the National Natural Science Foundation of China (grant no. 62276027). The statistical analysis was partially supported by the Biostatistics Shared Resource (B-SR) of the NCI-sponsored Stanford Cancer Institute: P30CA124435.

Graphic Abstract and Figure 1 were created in part using BioRender.com.

AUTHOR CONTRIBUTIONS

Y.-R.J.W. and J.G. conceived the research project. Y.-R.J.W., P.W., P.L., and H.L. carried out data processing, developed the machine learning experiments, and performed the model and data analysis. J.G., Z.Y., Q.Z., Y.-R.J.W., P.L., and P.W. conducted data acquisition and quality controlled the data. Y.-R.J.W., F.G., Y.H., L.F., K.Z., P.W., and J.J. conducted the statistical analysis. Y.-R.J.W., P.W., Z.Y., F.G., Y.H., K.C., H.W., Q.D., W.E.W., K.Z., and L.Q. contributed to the tables and figures of the paper. Y.-R.J.W. designed

the study and drafted the paper. J.G. and Y.-R.J.W. jointly supervised the work. All authors provided valuable feedback and ensured the integrity of the work.

DECLARATION OF INTERESTS

The authors declare no competing interests.

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

During the preparation of this work the author used ChatGPT in order to reframe certain sentences. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

Received: August 18, 2023

Revised: April 25, 2024

Accepted: June 5, 2024

Published: June 27, 2024

REFERENCES

1. Ostrom, Q.T., Cioffi, G., Gittleman, H., Patil, N., Waite, K., Kruchko, C., and Barnholtz-Sloan, J.S. (2019). CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2012–2016. *Neuro Oncol.* 21, v1–v100.
2. Amayiri, N., Swaidan, M., Ibrahim, A., Hirmas, N., Musharbash, A., Bouffet, E., Al-Hussaini, M., and Ramaswamy, V. (2021). Molecular subgroup is the strongest predictor of medulloblastoma outcome in a resource-limited country. *JCO Glob. Oncol.* 7, 1442–1453.
3. Louis, D.N., Perry, A., Wesseling, P., Brat, D.J., Cree, I.A., Figarella-Branger, D., Hawkins, C., Ng, H.K., Pfister, S.M., Reifenberger, G., et al. (2021). The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro Oncol.* 23, 1231–1251.
4. Taylor, M.D., Northcott, P.A., Korshunov, A., Remke, M., Cho, Y.-J., Clifford, S.C., Eberhart, C.G., Parsons, D.W., Rutkowski, S., Gajjar, A., et al. (2012). Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol.* 123, 465–472.
5. Zhang, M., Wong, S.W., Wright, J.N., Wagner, M.W., Toescu, S., Han, M., Tam, L.T., Zhou, Q., Ahmadian, S.S., and Shpanskaya, K. (2022). MRI radiogenomics of pediatric medulloblastoma: a multicenter study. *Radiology* 304, 406–416.
6. Casolino, R., Johns, A.L., Courtot, M., Lawlor, R.T., De Lorenzo, F., Horgan, D., Mateo, J., Normanno, N., Rubin, M., Stein, L., et al. (2023). Accelerating cancer omics and precision oncology in health care and research: a Lancet Oncology Commission. *Lancet Oncol.* 24, 123–125.
7. Ramaswamy, V., Remke, M., Bouffet, E., Bailey, S., Clifford, S.C., Doz, F., Kool, M., Dufour, C., Vassal, G., Milde, T., et al. (2016). Risk stratification of childhood medulloblastoma in the molecular era: the current consensus. *Acta Neuropathol.* 131, 821–831.
8. Hiremath, A., Shiradkar, R., Fu, P., Mahran, A., Rastinehad, A.R., Tewari, A., Tirumani, S.H., Purysko, A., Ponsky, L., and Madabhushi, A. (2021). An integrated nomogram combining deep learning, Prostate Imaging-Reporting and Data System (PI-RADS) scoring, and clinical variables for identification of clinically significant prostate cancer on biparametric MRI: a retrospective multicentre study. *Lancet. Digit. Health* 3, e445–e454.
9. Kesler, S.R., Sleurs, C., McDonald, B.C., Deprez, S., van der Plas, E., and Nieman, B.J. (2021). Brain imaging in pediatric cancer survivors: correlates of cognitive impairment. *J. Clin. Oncol.* 39, 1775–1785.
10. Dercle, L., Zhao, B., Gönen, M., Moskowitz, C.S., Firas, A., Beylergil, V., Connors, D.E., Yang, H., Lu, L., Fojo, T., et al. (2022). Early readout on overall survival of patients with melanoma treated with immunotherapy using a novel imaging analysis. *JAMA Oncol.* 8, 385–392.
11. Sun, R., Limkin, E.J., Vakalopoulou, M., Dercle, L., Champiat, S., Han, S.R., Verlingue, L., Brandao, D., Lancia, A., Ammari, S., et al. (2018). A radiomics

- approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol.* 19, 1180–1191.
12. Jiang, Y., Wang, H., Wu, J., Chen, C., Yuan, Q., Huang, W., Li, T., Xi, S., Hu, Y., Zhou, Z., et al. (2020). Noninvasive imaging evaluation of tumor immune microenvironment to predict outcomes in gastric cancer. *Ann. Oncol.* 31, 760–768.
 13. Chen, X., Fan, Z., Li, K.K.-W., Wu, G., Yang, Z., Gao, X., Liu, Y., Wu, H., Chen, H., Tang, Q., et al. (2020). Molecular subgrouping of medulloblastoma based on few-shot learning of multitasking using conventional MR images: a retrospective multicenter study. *Neurooncol. Adv.* 2, vdaa079.
 14. Yan, J., Liu, L., Wang, W., Zhao, Y., Li, K.K.-W., Li, K., Wang, L., Yuan, B., Geng, H., Zhang, S., et al. (2020). Radiomic features from multi-parameter MRI combined with clinical parameters predict molecular subgroups in patients with medulloblastoma. *Front. Oncol.* 10, 558162.
 15. Iv, M., Zhou, M., Shpanskaya, K., Perreault, S., Wang, Z., Tranvinh, E., Lanzman, B., Vajapeyam, S., Vitanza, N.A., Fisher, P.G., et al. (2019). MR imaging-based radiomic signatures of distinct molecular subgroups of medulloblastoma. *AJNR. Am. J. Neuroradiol.* 40, 154–161.
 16. Perreault, S., Ramaswamy, V., Achrol, A.S., Chao, K., Liu, T.T., Shih, D., Remke, M., Schubert, S., Bouffet, E., Fisher, P.G., et al. (2014). MRI surrogates for molecular subgroups of medulloblastoma. *AJNR. Am. J. Neuroradiol.* 35, 1263–1269.
 17. Liu, W., Tang, J., Van Halm-Lutterodt, N., Luo, S., and Li, C. (2018). History and current state of pediatric neurosurgery at Beijing Tiantan Hospital Neurosurgery Center. *Childs Nerv. Syst.* 34, 797–803.
 18. Capper, D., Jones, D.T.W., Sill, M., Hovestadt, V., Schrimpf, D., Sturm, D., Koelsche, C., Sahm, F., Chavez, L., Reuss, D.E., et al. (2018). DNA methylation-based classification of central nervous system tumours. *Nature* 555, 469–474.
 19. Lundberg, S.M., and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 30.
 20. Patay, Z., DeSain, L.A., Hwang, S.N., Coan, A., Li, Y., and Ellison, D.W. (2015). MR imaging characteristics of wingless-type–subgroup pediatric medulloblastoma. *AJNR. Am. J. Neuroradiol.* 36, 2386–2393.
 21. Orr, B.A. (2020). Pathology, diagnostics, and classification of medulloblastoma. *Brain Pathol.* 30, 664–678.
 22. Mata-Mbemba, D., Zapotocky, M., Laughlin, S., Taylor, M.D., Ramaswamy, V., and Raybaud, C. (2018). MRI characteristics of primary tumors and metastatic lesions in molecular subgroups of pediatric medulloblastoma: a single-center study. *AJNR. Am. J. Neuroradiol.* 39, 949–955.
 23. Zapotocky, M., Mata-Mbemba, D., Sumerauer, D., Liby, P., Lassaletta, A., Zamecnik, J., Krskova, L., Kyncl, M., Stary, J., Laughlin, S., et al. (2018). Differential patterns of metastatic dissemination across medulloblastoma subgroups. *J. Neurosurg. Pediatr.* 21, 145–152.
 24. Phoenix, T.N., Patmore, D.M., Boop, S., Boulos, N., Jacus, M.O., Patel, Y.T., Roussel, M.F., Finkelstein, D., Goumnerova, L., Perreault, S., et al. (2016). Medulloblastoma genotype dictates blood brain barrier phenotype. *Cancer Cell* 29, 508–522.
 25. Tylawsky, D.E., Kiguchi, H., Vaynshteyn, J., Gerwin, J., Shah, J., Islam, T., Boyer, J.A., Boué, D.R., Snuderl, M., Greenblatt, M.B., et al. (2023). P-selectin-targeted nanocarriers induce active crossing of the blood–brain barrier via caveolin-1-dependent transcytosis. *Nat. Mater.* 22, 391–399.
 26. Varatharaj, A., Liljeroth, M., Darekar, A., Larsson, H.B.W., Galea, I., and Cramer, S.P. (2019). Blood–brain barrier permeability measured using dynamic contrast-enhanced magnetic resonance imaging: a validation study. *J. Physiol.* 597, 699–709.
 27. Zhou, Q., Vega Leonel, J.C.M., Santoso, M.R., Wilson, C., van den Berg, N.S., Chan, C.T., Aryal, M., Vogel, H., Cayrol, R., Mandella, M.J., et al. (2021). Molecular imaging of a fluorescent antibody against epidermal growth factor receptor detects high-grade glioma. *Sci. Rep.* 11, 5710.
 28. Teo, W.Y., Shen, J., Su, J.M.F., Yu, A., Wang, J., Chow, W.Y., Li, X., Jones, J., Dauser, R., Whitehead, W., et al. (2013). Implications of tumor location on subtypes of medulloblastoma. *Pediatr. Blood Cancer* 60, 1408–1410.
 29. Yeom, K.W., Mobley, B.C., Lober, R.M., Andre, J.B., Partap, S., Vogel, H., and Barnes, P.D. (2013). Distinctive MRI features of pediatric medulloblastoma subtypes. *AJR Am. J. Roentgenol.* 200, 895–903.
 30. Zhao, F., Li, C., Zhou, Q., Qu, P., Wang, B., Wang, X., Zhang, S., Wang, X., Zhao, C., Zhang, J., et al. (2017). Distinctive localization and MRI features correlate of molecular subgroups in adult medulloblastoma. *J. Neuro Oncol.* 135, 353–360.
 31. Hovestadt, V., Ayraut, O., Swartling, F.J., Robinson, G.W., Pfister, S.M., and Northcott, P.A. (2020). Medulloblastomics revisited: biological and clinical insights from thousands of patients. *Nat. Rev. Cancer* 20, 42–56.
 32. Cavalli, F.M.G., Remke, M., Rampasek, L., Peacock, J., Shih, D.J.H., Luu, B., Garzia, L., Torchia, J., Nor, C., Morrissey, A.S., et al. (2017). Intertumoral heterogeneity within medulloblastoma subgroups. *Cancer Cell* 31, 737–754.e6.
 33. Thompson, E.M., Hielscher, T., Bouffet, E., Remke, M., Luu, B., Gururangan, S., McLendon, R.E., Bigner, D.D., Lipp, E.S., Perreault, S., et al. (2016). Prognostic value of medulloblastoma extent of resection after accounting for molecular subgroup: a retrospective integrated clinical and molecular analysis. *Lancet Oncol.* 17, 484–495.
 34. Cochrane, D.D., Gustavsson, B., Poskitt, K.P., Steinbok, P., and Kestle, J.R. (1994). The surgical and natural morbidity of aggressive resection for posterior fossa tumors in childhood. *Pediatr. Neurosurg.* 20, 19–29.
 35. Korah, M.P., Esiashvili, N., Mazewski, C.M., Hudgins, R.J., Tighiouart, M., Janss, A.J., Schwaibold, F.P., Crocker, I.R., Curran, W.J., Jr., and Marcus, R.B., Jr. (2010). Incidence, risks, and sequelae of posterior fossa syndrome in pediatric medulloblastoma. *Int. J. Radiat. Oncol. Biol. Phys.* 77, 106–112.
 36. Robertson, P.L., Muraszko, K.M., Holmes, E.J., Sposto, R., Packer, R.J., Gajjar, A., Dias, M.S., and Allen, J.C.; Children's Oncology Group (2006). Incidence and severity of postoperative cerebellar mutism syndrome in children with medulloblastoma: a prospective study by the Children's Oncology Group. *J. Neurosurg.* 105, 444–451.
 37. Levisohn, L., Cronin-Golomb, A., and Schmahmann, J.D. (2000). Neuropsychological consequences of cerebellar tumour resection in children: cerebellar cognitive affective syndrome in a paediatric population. *Brain* 123, 1041–1050.
 38. Menyhárt, O., and Györfy, B. (2020). Molecular stratifications, biomarker candidates and new therapeutic options in current medulloblastoma treatment approaches. *Cancer Metastasis Rev.* 39, 211–233.
 39. Thompson, E.M., Bramall, A., Herndon, J.E., Taylor, M.D., and Ramaswamy, V. (2018). The clinical importance of medulloblastoma extent of resection: a systematic review. *J. Neuro Oncol.* 139, 523–539.
 40. Guerrini-Rousseau, L., Abbas, R., Huybrechts, S., Kieffer-Renaux, V., Puget, S., Andreuolo, F., Beccaria, K., Blauwblomme, T., Bolle, S., Dhermain, F., et al. (2020). Role of neoadjuvant chemotherapy in metastatic medulloblastoma: a comparative study in 92 children. *Neuro Oncol.* 22, 1686–1695.
 41. Grill, J., Lellouch-Tubiana, A., Elouahdani, S., Pierre-Kahn, A., Zerah, M., Renier, D., Valteau-Couanet, D., Hartmann, O., Kalifa, C., and Sainte-Rose, C. (2005). Preoperative chemotherapy in children with high-risk medulloblastomas: a feasibility study. *J. Neurosurg.* 103, 312–318.
 42. Iwama, J., Ogihara, H., Kiyotani, C., Terashima, K., Matsuoka, K., Iwafuchi, H., and Morota, N. (2015). Neoadjuvant chemotherapy for brain tumors in infants and young children. *J. Neurosurg. Pediatr.* 15, 488–492.
 43. Jabarkheel, R., Amayiri, N., Yecies, D., Huang, Y., Toescu, S., Nobre, L., Mabbott, D.J., Sudhakar, S.V., Malik, P., Laughlin, S., et al. (2020). Molecular correlates of cerebellar mutism syndrome in medulloblastoma. *Neuro Oncol.* 22, 290–297.
 44. Alcantara, J.H., Ornos, E.D.B., and Tantengco, O.A.G. (2023). Global trends, gaps, and future agenda in medulloblastoma research: a bibliometric analysis. *Childs Nerv. Syst.* 39, 3185–3194.
 45. Cooney, T., Lindsay, H., Leary, S., and Wechsler-Reya, R. (2023). Current studies and future directions for medulloblastoma: A review from the

- pacific pediatric neuro-oncology consortium (PNOC) disease working group. *Neoplasia* 35, 100861.
46. Chen, B., Dammann, P., Jabbari, R., Sure, U., Quick, H.H., Kraft, O., and Wrede, K.H. (2023). Safety and function of programmable ventriculo-peritoneal shunt valves: An *in vitro* 7 Tesla magnetic resonance imaging study. *PLoS One* 18, e0292666.
 47. Esteva, A., Feng, J., van der Wal, D., Huang, S.-C., Simko, J.P., DeVries, S., Chen, E., Schaeffer, E.M., Morgan, T.M., Sun, Y., et al. (2022). Prostate cancer therapy personalization via multi-modal deep learning on randomized phase III clinical trials. *NPJ Digit. Med.* 5, 71.
 48. Northcott, P.A., Shih, D.J.H., Remke, M., Cho, Y.-J., Kool, M., Hawkins, C., Eberhart, C.G., Dubuc, A., Guettouche, T., Cardentey, Y., et al. (2012). Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. *Acta Neuropathol.* 123, 615–626.
 49. Leal, L.F., Evangelista, A.F., de Paula, F.E., Caravina Almeida, G., Carloni, A.C., Saggiorno, F., Stavale, J.N., Malheiros, S.M.F., Mançano, B., de Oliveira, M.A., et al. (2018). Reproducibility of the NanoString 22-gene molecular subgroup assay for improved prognostic prediction of medulloblastoma. *Neuropathology* 38, 475–483.
 50. Kim, J.W., Park, S.-H., Choi, S.A., Kim, S.-K., Koh, E.J., Won, J.-K., Nam, S.M., and Phi, J.H. (2022). Molecular subgrouping of medulloblastoma in pediatric population using the NanoString assay and comparison with immunohistochemistry methods. *BMC Cancer* 22, 1221.
 51. D'Arcy, C.E., Nobre, L.F., Arnaldo, A., Ramaswamy, V., Taylor, M.D., Naz-Hazrati, L., and Hawkins, C.E. (2020). Immunohistochemical and nanoString-based subgrouping of clinical medulloblastoma samples. *J. Neuropathol. Exp. Neurol.* 79, 437–447.
 52. George, S.L., Izquierdo, E., Campbell, J., Koutroumanidou, E., Proszek, P., Jamal, S., Hughes, D., Yuan, L., Marshall, L.V., Carceller, F., et al. (2019). A tailored molecular profiling programme for children with cancer to identify clinically actionable genetic alterations. *Eur. J. Cancer* 121, 224–235.
 53. Tibshirani, R., Hastie, T., Narasimhan, B., and Chu, G. (2002). *Proc. Natl. Acad. Sci. USA* 99, 6567–6572.
 54. Li, Y., Jiang, T., Shao, L., Liu, Y., Zheng, C., Zhong, Y., Zhang, J., and Chang, Q. (2016). Mir-449a, a potential diagnostic biomarker for WNT group of medulloblastoma. *J. Neuro Oncol.* 129, 423–431.
 55. Pugh, T.J., Weeraratne, S.D., Archer, T.C., Pomeranz Krummel, D.A., Auclair, D., Bochicchio, J., Carneiro, M.O., Carter, S.L., Cibulskis, K., Erlich, R.L., et al. (2012). Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations. *Nature* 488, 106–110.
 56. Robinson, G., Parker, M., Kranenburg, T.A., Lu, C., Chen, X., Ding, L., Phoenix, T.N., Hedlund, E., Wei, L., Zhu, X., et al. (2012). Novel mutations target distinct subgroups of medulloblastoma. *Nature* 488, 43–48.
 57. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.
 58. Albright, A.L., Wisoff, J.H., Zeltzer, P.M., Boyett, J.M., Rorke, L.B., and Stanley, P. (1996). Effects of medulloblastoma resections on outcome in children: a report from the Children's Cancer Group. *Neurosurgery* 38, 265–271.
 59. Lowekamp, B.C., Chen, D.T., Ibáñez, L., and Blezek, D. (2013). The design of SimpleITK. *Front. Neuroinf.* 7, 45.
 60. Avants, B.B., Tustison, N., and Song, G. (2009). Advanced normalization tools (ANTS). *Insight j* 2, 1–35.
 61. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., and Gee, J.C. (2010). N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imag.* 29, 1310–1320.
 62. Wang, Y.-R., Wang, H., Chen, S., Katsaggelos, A.K., Martersteck, A., Higgins, J., Hill, V.B., and Parrish, T.B. (2019). A 3D cross-hemisphere neighborhood difference Convnet for chronic stroke lesion segmentation. In *2019 IEEE International Conference on Image Processing (ICIP)* (IEEE), pp. 1545–1549.
 63. Isensee, F., Jaeger, P.F., Kohl, S.A.A., Petersen, J., and Maier-Hein, K.H. (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* 18, 203–211.
 64. Ibanez, L., Schroeder, W., Ng, L., and Cates, J. (2015). The insight segmentation and registration toolkit. Software Guide.
 65. Zwanenburg, A., Vallières, M., Abdalah, M.A., Aerts, H.J.W.L., Andarczyk, V., Apte, A., Ashrafinia, S., Bakas, S., Beukinga, R.J., Boellaard, R., et al. (2020). The image biomarker standardization initiative: standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology* 295, 328–338.
 66. van Griethuysen, J.J.M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R.G.H., Fillion-Robin, J.-C., Pieper, S., and Aerts, H.J.W.L. (2017). Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* 77, e104–e107.
 67. Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., and Deng, S.-H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *J. Electron. Sci. Technol.* 17, 26–40.
 68. Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5–9, 2015, proceedings, part III* 18, pp. 234–241.
 69. Ulyanov, D., Vedaldi, A., and Lempitsky, V. (2016). Instance normalization: The missing ingredient for fast stylization. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1607.08022>.
 70. Maas, A.L., Hannun, A.Y., and Ng, A.Y. (2013). Rectifier nonlinearities improve neural network acoustic models. *Proc. icml* 30, 3. (Citeseer).
 71. Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A.L. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* 40, 834–848.
 72. Drozdzal, M., Vorontsov, E., Chartrand, G., Kadoury, S., and Pal, C. (2016). The importance of skip connections in biomedical image segmentation. In *Deep learning and data labeling for medical applications* (Springer), pp. 179–187.
 73. He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778.
 74. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin transformer: Hierarchical Vision Transformer Using Shifted Windows, pp. 10012–10022.
 75. Wang, H., Zhu, Q., Ding, L., Shen, Y., Yang, C.Y., Xu, F., Shu, C., Guo, Y., Xiong, Z., Shan, Q., et al. (2019). Scalable volumetric imaging for ultrahigh-speed brain mapping at synaptic resolution. *Natl. Sci. Rev.* 6, 982–992. <https://doi.org/10.1093/nsr/nwz053>.
 76. Zheng, W., Mu, H., Chen, Z., Liu, J., Xia, D., Cheng, Y., Jing, Q., Lau, P.-M., Tang, J., Bi, G.-Q., et al. (2024). NEATmap: a high-efficiency deep learning approach for whole mouse brain neuronal activity trace mapping. *Natl. Sci. Rev.* 11, nwae109. <https://doi.org/10.1093/nsr/nwae109>.
 77. Azur, M.J., Stuart, E.A., Frangakis, C., and Leaf, P.J. (2011). Multiple imputation by chained equations: what is it and how does it work? *Int. J. Methods Psychiatr. Res.* 20, 40–49.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
An open-source version of the code base in this study	https://github.com/MedAI-Vision/MedAI-Brain-MB	
SnapITK	http://www.itksnap.org/pmwiki/pmwiki.php	
Slicer3D	https://www.slicer.org/	
Colab notebook for deep learning tumor segmentation	This paper; https://colab.research.google.com/drive/1FoBDfPAeU_PH22VQyn-vrUED_bHM5Vpc?usp=sharing	
Imaris v9.8.2	Oxford Instruments	
Antibodies		
Anti-CD31 primary antibody	Ab28364, Abcam	RRID: AB_726362
Alexa Fluor 647 affiniPure donkey anti-rabbit secondary antibody	711-605-152, Jackson ImmunoResearch	RRID: AB_2492288
Chemicals, Peptides, and Recombinant Proteins		
Quadrol	122262, Sigma	
H2O2	10011218, Sinopharm Chemical Reagent	
Ethanol	G73537B, Greagent	
Hoechst	H3570, Invitrogen	
PBS	10010023, Thermo Fisher	
Triton-X100	T9284, Sigma	
Deposited data		
Medulloblastoma database	This paper	

RESOURCE AVAILABILITY

Lead contact

Further information and requests regarding this manuscript should be sent to and will be fulfilled by the lead contact, Yan-Ran (Joyce) Wang (wangyanran100@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Participating medical centers obtained approval from their institutional review boards (IRB) with a waiver of informed consent for the use of retrospective deidentified data. The database with demographic, molecular subgroup, clinical information, survival data, and MRI imaging features is openly provided in [supplemental information](#). A portion of the de-identified presurgical MRI data is accessible via application to the Children's Brain Tumor Network (CBTN) at <https://cbtn.org/>. The remaining de-identified imaging data can be made available by making a reasonable request to the corresponding authors. All requests will be evaluated by the ethics committee of the hospitals based on institutional policies to determine whether the data requested are subject to intellectual property or patient privacy obligations. Data can be shared only for non-commercial academic purposes and will require a formal material transfer agreement. Generally, all such requests for access to imaging data will be responded to within 1 month.

The entire code was developed using Python with the PyTorch deep learning framework. For accessibility and transparency, an open-source version of the codebase has been made accessible on GitHub at <https://github.com/MedAI-Vision/MedAI-Brain-MB>. In a bid to offer a practical illustration of our approach, we have uploaded presurgical MRI scans of five de-identified testing subjects representing WNT, SHH (both TP53- and TP53+), G3 and G4 molecular subtypes, respectively. These scans can be accessed via our GitHub repository.

To further improve the clarity and accessibility of our research, we have developed an interactive Colab notebook for visualization and testing demonstrations, encompassing the front-end deep learning segmentation model. Please refer to https://colab.research.google.com/drive/1FoBDfPAeU_PH22VQyn-vrUED_bHM5Vpc?usp=sharing. We have made the model weights openly available, and they are released under an open-source license.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

MRI acquisition

All patients had 1.5 or 3.0-Tesla multiparametric brain MRI scans from vendors with the following distribution: Siemens Healthineers (45%), GE Healthcare (37%), Philips Healthcare (17%), and other vendors (1%); 3.0T (87%) and 1.5T (13%). The T1-weighted and T2-weighted MRIs comprised either three-dimensional isovolumetric or two-dimensional spin-echo imaging, with a section thickness of 0.8–7.0 millimeter (mm) and a pixel spacing of 0.21–1.0 mm. Contrast-enhanced T1 (T1E) was acquired 2 minutes after intravenous administration of contrast agents (0.1 mmol/kg).

METHOD DETAILS

Dataset curation

Inclusion criteria comprised pathologically confirmed medulloblastoma (initial diagnosis), molecular subgroup analysis, and known survival status, with cases labeled as Not Otherwise Specified (NOS) and Not Elsewhere Classified (NEC) diagnoses excluded. Age and comorbidities were not exclusion criteria.

The molecular subgroups of medulloblastoma patients in this investigation were identified through specialized post-surgery specimen analysis, including DNA methylation arrays, NanoString-based mRNA expression analysis, and next-generation sequencing panels (Figure 1C).^{18,48–52} Clinical data, encompassing demographic information, extent of tumor resection, tumor texture, and histological subgroup, were gathered from pathology and surgeons' reports. Prognostic information was collected by phone surveys if not reported in medical records and overall survival was applied as the primary endpoint for survival analysis in this study.

Study design

We aim to address four key issues inherent in the current standard of care and understanding of medulloblastoma. Firstly, our developed AI system facilitates presurgical noninvasive molecular subgroup prediction, offering significant value in preoperative guidance. Studies have shown that the prognostic value of extent of surgical resection for medulloblastoma is highly molecular subgroup dependent.^{5,33,38,39} Therefore, preoperative subgroup prediction has the potential to optimize surgical planning by customizing resection goals according to molecular subgroup profiles.

Secondly, we combine machine-generated radiomic features with manual characterization by neuroradiologists, covering intra-tumoral, peri-tumoral, and extra-tumoral regions on MR imaging, all within a unified framework. This holistic integration not only enhances the accuracy of molecular prediction but also allows us to gauge the predictive power of distinct features and regions in molecular subgrouping.

Thirdly, our AI models underwent rigorous validation, including cross-validation on the primary set (n=514), external validation across ten cross-continental centers (n=135), and prospective validation on a consecutive testing set (n=40) designed to reflect real-world prevalence of molecular subtypes based on DNA methylation classification.¹⁸ This diverse validation approach underscores the broad applicability of the developed system as a generalizable molecular subgroup classifier.

Last but not least, as underscored in the 2023 review by the Pacific Neuro Oncology Consortium (PNOC),⁴⁵ the existing literature on MB predominantly stems from European and North American cohorts. By harnessing this international dataset as the foremost and extensive resource of MB including the Asian populace, we examined the association of clinicopathological features and molecular subgroups, and discerned disparities between East Asia and North America subsets. In an earnest answer to the calling for an inclusive exploration of MB on a worldwide scale, we have made available the data of 934 MB patients, inclusive of demographic, molecular subgroup, clinicopathological, and survival information, along with comprehensive MR imaging signatures, as a valuable resource for the neuro-oncology community in advancing the global investigation and comprehension of medulloblastoma.

Molecular subgroup analysis

The molecular subgroups, namely WNT, SHH, Group 3, and Group 4, of medulloblastoma patients enrolled in this study, were delineated through fluorescence *in situ* hybridization, specialized testing including NanoString-based mRNA expression analysis, next-generation sequencing panels, along with DNA methylation arrays.^{18,48–52} The molecular subgroup analyses were generated by third-party commercial companies or research hospitals. The testing outcomes are documented in clinical records, but the individual patients retain ownership of their raw data. Table S11 shows the listing of contributing medical centers by pathology.

NanoString-based subgrouping

The molecular subgroups of medulloblastoma from Beijing Tiantan Hospital¹⁷ and other medical centers were partially identified through the NanoString-based subgrouping method.^{48–51}

NanoString nCounter analysis

RNA was extracted from formalin-fixed paraffin-embedded (FFPE) tissue specimens from medulloblastoma samples based on the NanoString nCounter Analysis System (Nanostring Technologies), and gene expression analysis was conducted using the classic medulloblastoma subtyping gene CodeSet, including 22 genes with 5–6 signature genes for each subgroup: WNT (WIF1, TNC, GAD1, DKK2 and EMX2), SHH (PDLIM3, EYA1, HHIP, ATOH1 and SFRP1), Group 3 (IMPG2, GABRA5, EGFL11, NRL, MAB21L2 and NPR3) and Group 4 (KCNA1, EOMES, KHDRBS2, RBM24, UNC5D and OAS1). Each reaction contained 200 ng of total RNA

in a 4 μ l aliquot, together with the reporter and capture probes, and six pairs of positive controls and six pairs of negative control probes. Analysis and normalization of the raw NanoString data were conducted using nSolver Analysis Software v2.5 (Nanostring Technologies). Raw counts were normalized to internal levels for three reference genes including CLTC, GAPDH and TUBB. A background count level was estimated using the average count of the six negative control probes in every reaction plus two standard deviations (SDs). All procedures related to mRNA quantification including sample preparation, hybridization, detection, and scanning were carried out as recommended by NanoString Technologies.

Prediction analysis for microarrays (PAM) scores

PAM is a statistical method designed for class prediction using gene expression data, specifically employing nearest shrunken centroids.⁵³ In our study, we utilized the PAM algorithm implemented in the *pamr* R package to perform class prediction based on normalized NanoString gene expression data. The analytical process involved two main phases: (1) Training Phase: We utilized a training series comprising 63 samples from the Department of Pathology, Peking University, School of Basic Medical Science.⁵⁴ These samples had previously been classified for medulloblastoma molecular subgroups. Raw data from the training series were normalized, and gene expression values were defined. The training series, along with paired medulloblastoma molecular subgroup labels, was employed to establish a reference dataset. The PAM algorithm was applied during the training phase, utilizing a 10-fold cross-validation scheme to train our model. (2) Prediction Phase: The model trained during the training phase was subsequently applied to the studied cohort for class prediction. PAM scores were obtained during the prediction phase, providing a measure of the probability that each case belonged to a specific medulloblastoma molecular subgroup. Patients with posterior probabilities reported by the *pamr* model no less than 0.9 (threshold) and predicted as Group 3 or Group 4 are labeled accordingly. We provided two case examples for Group 3 and Group 4 (Figure S7) with PAM scores along with heatmap visualization to clarify the criteria used for exclusion and inclusion of cases within the primary cohort in the development of our machine learning model.

Figure 5A presents an expression heatmap illustrating the distinct stratification of four molecular subgroups (WNT, SHH, Group 3, and Group 4) using the 22 NanoString signature genes specific to medulloblastoma. These 22 signature genes constitute the NanoString CodeSet employed in our study. The heatmap is derived from a series of 174 primary medulloblastoma samples admitted to Beijing Tiantan Hospital in the primary set, all of which underwent profiling using the NanoString nCounter System. Note that the NanoString signature was applied to further diagnose the molecular subgroup within medulloblastoma patients, rather than to diagnose medulloblastoma itself. All medulloblastoma patients enrolled in this study were histologically confirmed to have medulloblastoma.

Next generation sequencing

The molecular subgroups of medulloblastoma from Beijing Tiantan Hospital¹⁷ and other medical centers were identified in part through next-generation sequencing techniques.

Whole genome sequencing (WGS)

For DNA extraction, medulloblastoma tissues were dissected from 5 μ m tissue sections. The microdissected tissues were mixed with the ATL buffer (Qiagen, Hilden, Germany), heat-treated for 15 minutes at 98°C, and then underwent proteinase K digestion for 3 days at 56°C. Genomic DNA was extracted using the DNeasy Tissue and Blood kit (Qiagen) according to the manufacturer's protocol. Extracted DNA was quantified using a NanoDrop ND-1000 Fluorospectrometer (NanoDrop, Wilmington, DE). Genomic DNA was fragmented by sonication, and then DNA fragments were endpolished, A-tailed, and ligated with the full-length adapter for Illumina sequencing with further PCR amplification. PCR products were purified (AMPure XP) and libraries were analyzed for size distribution by the 2100 Bioanalyzer (Agilent) and quantified by Qbit3 and realtime PCR. Clustering of the index-coded samples was performed on a cBot Cluster Generation System using the HiSeq X HD PE Cluster Kit (Illumina), per manufacturer's instructions. Libraries were then sequenced on the HiSeq X Ten (Illumina, San Diego, CA) and 150 bp paired-end reads were generated.

Targeted sequencing

The library of gDNA was constructed according to the manufacturer's protocols by a KAPA Hyper Prep: kit (Kapa Biosystems). The quantities of the library were all measured with Qbit 3 (Thermo fisher). A panel of 39 gene were used in our research (APC, ARID1B, BCOR, CDH1, CDK6, CHD7, CREBBP, CSNK2B, CTDNEP1, CTNNB1, DDX31, DDX3X, GABRG1, GFI1, GFI1B, GLI2, GPS2, KDM4C, KDM6A, KMT2B, LDB1, MLL3, MYC, MYCN, NCOR2, OTX2, PIK3CA, PTCH1, PTCH2, PTEN, SMARCA4, SMARCC2, SMARCD2, SMO, SNCAIP, SUFU, TERT, TP53, ZMYM3).^{55–57} gDNA libraries were enriched for regions of this custom designed captured probe manufactured by Agilent. 750ng prepared libraries were hybridized with two different hybridization reagents and blocking agents in SureSelectXT Target Enrichment System (Agilent Technologies). The enriched libraries were amplified with P5/P7 primer. After qualified by the 2200 Bioanalyzer and quantified with Qbit3 and a QPCR NGS library quantification kit (Agilent Technologies), then the libraries were sequenced on a Hiseq X10 platform (Illumina, San Diego, CA). Note that the TP53 status within SHH tumors in this study was discerned employing the targeted sequencing method.

Bio-informatics analysis

Sequencing reads were mapped to the human reference genome (hg19) using the Burrows-Wheeler Aligner (BWA). Duplicate removal, local realignment, and base quality recalibration were performed using PICARD (<http://broadinstitute.github.io/picard/>) and the Genome Analysis Toolkit (GATK).⁵⁷ Somatic single nucleotide variations (SNVs) were called using Mutect, small indels were called using strelka and somatic CNV were called by ExomeCNV. Each variant was confirmed using IGV. The WGS data

were used for testing amplification and deletion in chromosome by software control freeC. By using the variation data of the known classification results in the previous research, the random forest algorithm is used to construct the classification model. And this model will be used to calculate the molecular classification.

Molecular characterization through DNA methylation profiling in the consecutive testing set

The fresh consecutive testing set underwent DNA methylation classification for molecular diagnosis, and the specific procedure for DNA methylation classification is outlined in the following. The genomic DNA was meticulously extracted from formalin-fixed, paraffin-embedded (FFPE) specimens and subjected to bisulfite conversion employing the EZ DNA Methylation kit, in strict adherence to the recommended protocol provided by Zymo Research (Irvine, CA, USA). Subsequent to the conversion process, the DNA underwent amplification, fragmentation, and was meticulously hybridized onto the Illumina Infinium Human Methylation450 Beadchip array as per the specified guidelines of Illumina Inc (San Diego, CA). Data derived from methylation analyses were preprocessed utilizing the ChAMP package (version 1.28.4) within R Bioconductor (version 3.5.3). We adhered to the ChAMP standard pipeline for stringent quality control measures and normalization of the data. For the prediction of Medulloblastoma (MB) subgroups, we accessed a dedicated web-based platform that specializes in DNA methylation-based classification of central nervous system tumors¹⁸ available at www.molecularneuropathology.org (version 11b4).

Extent of tumor resection

Extent of resection was primarily determined through surgeons' reports and validated using postoperative gadolinium-enhanced T1-weighted MRI images, where available. Consistent with prior studies,^{33,58} we categorized gross total resection (GTR) as the absence of residual tumor, near-total resection (NTR) as less than 1.5 cm² residual tumor based on postoperative imaging, and subtotal resection as 1.5 cm² or more of residual tumor.

Qualitative MRI imaging features

We conducted a comprehensive assessment encompassing a range of MR imaging features, including tumor location, tumor margin, enhancement pattern, the existence of peritumoral edema, the presence of cystic change/necrosis within the primary tumor, the occurrence of hydrocephalus before surgery, and the identification and pattern of intracranial solid metastases. Diagnostic MR images obtained before surgery were independently evaluated by three experienced neuroradiologists who were blinded to molecular subgroup and pathologic data. Features of primary tumor as well as intracranial solid metastases were evaluated. For primary tumors, characteristics including location, tumor margin, peri-tumoral edema, cystic change/necrosis, and contrast enhancement were reviewed. Patients with T1-enhanced MRI were included for metastasis analysis. Intracranial metastatic lesions were evaluated for location, enhancement, and diffusion restriction. Additionally, the occurrence of hydrocephalus pre-surgery was reviewed. Note that neuroradiologists based their assessments on a combination of the radiology report for each patient and a reassessment of the presurgical MRI images for qualitative MR imaging feature evaluation.

Tumor location

The tumor location was categorized as follows:

- 1) Midline-vermis/4th ventricle, denoting the presence of the tumor within the fourth ventricle.
- 2) 4th ventricle-uniCPA, characterizing the tumor situated in the fourth ventricle with unilateral extension along the lateral recess toward the cerebellopontine angle (CPA). We followed the definition in a previous article.²²
- 3) 4th ventricle-biCPA, specifying the tumor within the fourth ventricle with bilateral extension along bilateral recesses toward the CPAs.
- 4) Cerebellar hemisphere and midline/fourth ventricle, describing tumors that infiltrate both the cerebellar hemisphere and the fourth ventricle.
- 5) Cerebellar hemisphere, indicative of tumors located specifically within the cerebellar hemisphere.
- 6) Within CPA, referring to tumors strictly localized to the cerebellopontine angle (CPA).

Enhancement pattern of the primary tumor

Primary tumor enhancement patterns were systematically classified based on proportion and strength observed on T1-weighted contrast-enhanced images. The proportion of tumor enhancement was delineated into distinct categories: 1) none or minimal (characterized by enhancement in less than 10% of the primary tumor mass), 2) heterogeneous (encompassing enhancement between 10% and 50% of the mass), and 3) diffuse (manifesting when more than 50% of the mass exhibited enhancement).

In contrast to previous studies that derived enhancement information from human estimations based on a single tumor slide in the axial view (typically showing the largest tumor area),^{16,22} we developed an additional deep learning model. This model automatically delineates the 3D enhancement area within the T1-weighted contrast-enhanced images. The predicted 3D enhancement mask on T1-enhanced MRI underwent further refinement through consensus review by neuroradiologists to ensure accuracy. The proportion of tumor enhancement was calculated volumetrically using the 3D enhancement mask and the 3D tumor mask.

Additionally, using the superior sagittal sinus enhancement as the reference, the intensity of enhancement within the primary tumor mass was categorized into three groups: avid enhancement (comparable to the superior sagittal sinus), mild enhancement (less than that of the superior sagittal sinus), and non-enhancing tumors.

Intracranial solid metastases

Besides features derived from the primary tumor, intracranial solid metastases for each patient were evaluated in MRI imaging sequences. Patients with metastatic disease underwent further characterization into ependymal or leptomeningeal metastases. Ependymal metastases were searched specifically along the ependyma of the lateral and third ventricles, especially in the third ventricular infundibular recess (3rd V.I.R.). We reported ependymal dissemination (excluding 3rd V.I.R.) and 3rd V.I.R. dissemination separately. Leptomeningeal dissemination (LMD) was intricately classified into supratentorial and infratentorial dissemination.

Furthermore, signal patterns of all intracranial metastatic disease were evaluated and characterized as diffusion restricting or not (denoted as D+/D-, assessed on DWI sequence) and as postcontrast enhancing or not (denoted as C+/C-, evaluated on contrast-enhanced T1 sequence). The imaging features of metastases were termed as a "mis-matching pattern" when the metastatic lesion exhibited diffusion restriction but minimal or no postcontrast enhancement (C-/D+), and a "matching pattern" when the lesion demonstrated both diffusion restriction and postcontrast enhancement (C+/D+).

Tumor margin

The tumor margin was characterized as ill-defined if more than 50% of the margin could not be distinguished from the surrounding cerebellar parenchyma on the basis of all imaging sequences; otherwise, it was considered well-defined.^{16,22}

MRI pre-processing

All pre-surgical MRIs were preprocessed to: (1) NIFTI format; (2) resample MRI images to the same spatial resolution; (3) bias field correction; and (4) tumor segmentation. Below we detailed the preprocessing step for pre-surgical MRI scans.

The original DICOM format of MRIs was converted to NIFTI format using "dcm2niix" command. "dcm2niix" (<http://manpages.ubuntu.com/manpages/bionic/man1/dcm2niix.1.html>) is designed to convert neuroimaging data from the DICOM format to the NIFTI format and can be performed using a simple command-line interface from Ubuntu system. Then, we adopted the three-spline interpolation provided by SimpleITK⁵⁹ library (<https://simpleitk.org/>) for resampling, which ensures all of the scans are in the same image dimension – 0.430×0.430×6.5 mm (the most common spatial resolution across all subjects investigated in the study). Rigid registration was performed with ANTs⁶⁰ software on the multi-parameter MR images for each patient using axial T2 as a reference. This ensures T1E MRI and T2 MRI for the same patient are aligned. The N4 Bias Field Correction⁶¹ from ANTs software was adopted to correct bias field of MRI imaging, which was created by magnetic field inhomogeneities, coil loading, RF transmit errors, and head coil receive characteristics.⁶² The top and bottom 0.5% of the pixels in brain MRI images are clipped to avoid pixels that are outliers of the distribution. The MRI images are then normalized using min-max normalization. Finally, we developed a tumor ROI (region of interest) segmentation model (nnU-Net)⁶³ and used it to delineate tumor area for each pre-surgical MRI scan. Predicted tumor volume on MRI was further revised with consensus review by board-certified neuroradiologists with >15 years of experience.

ML-enabled molecular subgroup prediction

We developed presurgical MRI-based machine learning strategies for non-invasive molecular subgroup prediction. The machine learning pipeline encompassed three pivotal phases, elucidated as follows.

Semi-automatic tumor segmentation

As the initial pre-processing step, we proposed a tumor segmentation deep neural network (DNN) for medulloblastoma based on MRI scans (Figure S8). The backbone of the segmentation model is nnU-Net⁶³ as it has previously been shown to perform well on a broad range of medical segmentation tasks. Predicted tumor volume on MRI was further revised with consensus review by board-certified neuroradiologists with >15 years of experience. Meanwhile, we used N4 Bias Field Correction^{61,62,64} to correct bias field of MRI imaging. The preprocessing steps improve robustness of MRI-derived radiomic features and remove the additional burden of the machine learning methods to find patterns between images for molecular subgroup prediction.

Radiomics feature extraction and selection

We extracted Image Biomarker Standardization Initiative (IBSI)-based quantitative features using PyRadiomics.^{65,66} To capture information at the invasive margin besides the tumor core, we defined the peri-tumoral region by explicitly extracting a peripheral ring structure surrounding the primary MB tumor. The peri-tumoral ring was generated with dilation of the tumor boundaries by two mm on the outside and shrinkage of the tumor boundaries by two mm from the inside (MATLAB; imdilate and imerode functions), resulting in a ring structure with a thickness of five mm. For each patient, we have radiomics features (6897 features) extracted from T1E-intratumoral region, T1E-peritumoral region, T2-intratumoral region, and T2-peritumoral region. Random forest (RF) feature selection was performed and repeated for 100 cycles. Features ranked in the top ten and representing greater than 80% of the cycles were retained. In addition, radiomic features with the top 60 highest individual AUCs were assessed and incorporated if they demonstrated supplementary value in addition to the RF-selected features. The detailed mathematical definitions of selected imaging features are described in Tables S4 and S5.

Human-accessible MR imaging feature selection

For each human-accessible MR imaging feature, we expanded the feature set with benchmark radiomics features and evaluated its impact on performance in the testing set. To mitigate randomness, we conducted each experiment, specifically the single-imaging-feature-plus-radiomics, ten times and computed the average AUC as the comparison metric. From the top-performing features, we selectively incorporated MR imaging features into the benchmark ML model if they enhanced performance on the primary set. It's crucial to note that an MR imaging feature was promptly excluded if it did not contribute to a fusion experiment, even if it showed improvement in the initial single-feature-plus-radiomics experiments. Notably, the selected characteristic MR imaging features for

the three-class classifier (WNT vs SHH vs G3/4)—namely tumor location, tumor margin, enhancement, and intracranial solid metastases—consistently exhibit statistically significant differences in distribution across molecular subgroups, all with p-values < 0.001, as outlined in [Table 1](#).

Machine learning model development

We developed two classifiers to predict molecular subgroups: (1) a three-class classifier to classify WNT, SHH, and non-WNT/non-SHH (i.e., G3/4) classes, and (2) a binary classifier to further classify G3 and G4 within non-WNT/non-SHH. Light gradient-boosting machine (LightGBM) emerged as the final model backbone among four candidate classifiers (support vector machine-SVM, random forest-RF, and neural network-NN). Performance comparisons are outlined in [Table S12](#), with optimal classifier parameters estimated through Bayesian optimization or Grid search ([Table S13](#)).⁶⁷ Three-fold cross validation was employed in the primary set. Training and test sets were assigned from the primary set in a 2:1 ratio, with the training cohort undergoing resampling to address sample imbalance. The relative influences of input features were determined using Shapley analysis, which is a principled attribution method used in ML to quantify the contribution of individual input features.¹⁹

Deep learning automated tumor segmentation

We developed an automatic 3D deep neural network (DNN) segmentation model to identify the primary tumor for patients of medulloblastoma. The DNN model was derived from T1E and T2 MRI scans (the most common MRI sequences). The framework of the tumor segmentation model is shown in [Figure S8](#).

The segmentation model shares the classic U-net⁶⁸ backbone with three small adjustments: 1). batch normalization is replaced with instance normalization;⁶⁹ 2). ReLU is replaced with leaky ReLU⁷⁰ as the activation function; 3). additional auxiliary losses are added in the decoder to all but the two lowest resolutions. The segmentation model takes T1E and T2 MRI scans from the axial view as combined inputs and outputs the binary pixel-wise accurate 3D segmentation mask for the primary tumor of medulloblastoma.

We adopted stochastic gradient descent (SGD) with Nesterov momentum ($\mu=0.99$). The initial learning rate is set to be 0.01, and the decay of the learning rate follows the ‘Poly’ learning rate policy.⁷¹ The batch size is set to 2. Data augmentation includes rotations, scaling, gamma correction, and mirroring. The loss function is the sum of cross-entropy and Dice loss.⁷²

The segmentation model was trained upon 250 brain MRI records and validated on 50 records. We used a five-fold cross-validation approach. The average Dice score is 0.876. The high Dice value demonstrates the effectiveness of the DNN model in tumor segmentation for patients of medulloblastoma.

Deep learning in molecular subgroup prediction

We opted for the computational and manual MRI signatures approach in this study due to its superior performance over end-to-end deep learning methods. Despite testing advanced models like ResNet⁷³ and Swin Transformer⁷⁴ for direct brain MRI analysis, our proposed approach consistently outperformed deep learning in end-to-end molecular subgroup classification from presurgical MRIs.

Our methodology followed the 2.5D input scheme, utilizing five consecutive axial slices from both the contrast-enhanced (T1E) and T2 modalities, with the mid-slice displaying the largest tumor area. This resulted in ten input slices per subject. We designed our deep learning model as a three-class classifier to distinguish among WNT, SHH, and G3/G4 (i.e., non-WNT/non-SHH) molecular classes. Training and validation were conducted using the primary Beijing Tiantan dataset, comprising 514 MB subjects. We employed cross-entropy loss for training and performed three-fold cross-validation within the Beijing Tiantan cohort, with each fold involving 343 patients for training and 171 for testing, followed by evaluation on an external test set ($n=135$). Detailed results are presented in [Table S15](#).

The limited size of our training dataset (343 subjects) may explain the less promising performance of deep learning models in end-to-end molecular subgroup prediction. Conversely, our application of deep learning for medulloblastoma tumor segmentation yielded highly promising results, achieving a Dice score approaching 0.9 for tumor segmentation. This notable contrast in performance can be attributed to the inherent differences in the complexity of the tasks. Pixel-wise segmentation, such as tumor segmentation, falls under middle-level vision tasks and requires significantly less training data compared to the high-level molecular diagnostic tasks.

Manual annotation procedure

300 subjects were manually labeled for tumor segmentation by board-certified neuroradiologists (with >15 years of experience). The Scissor tool in 3D Slicer (<https://www.slicer.org/>) and the Polygon Inspector in ITK-SNAP (<http://www.itksnap.org/pmwiki/pmwiki.php>) were used for tumor delineation. The delineated tumor region included the contrast-enhancing area, the non-enhancing area, and the necrotic and cystic portions of the tumor. T1E sequence was used as the main reference, and T2 was used for finetuning and final confirmation. A binary segmentation mask is saved for each brain MRI record. This dataset with ground-truth tumor masks served as the training set for developing deep learning automated tumor segmentation.

Immunostaining in medulloblastoma surgical specimens

Tissue preparation

Surgical specimens from 16 medulloblastoma patients (4 WNT, 4 SHH, 4 Group3, and 4 Group4) underwent fixation in 10% formalin at 4°C for 24 hours. Then, samples were sectioned into 300 μ m slices using a vibratome. These slices underwent photobleaching

under a UV lamp for 16 hours, followed by chemical bleaching in a solution containing 10% Quadrol (Merck), 5% H₂O₂, and 5% ethanol overnight. Afterward, the slices underwent three washes with 1x phosphate-buffered saline (PBS), each lasting one hour. Subsequent clearing was achieved by incubating the slices in 10% Triton X-100 in PBS (PBST) for 24 hours at 37°C with gentle agitation, followed by three one-hour washes with 0.3% PBST.

Immunostaining and imaging

Following tissue clearing, the thick slices were blocked with donkey serum at 4°C for 8 hours. Subsequently, the slices were incubated with an anti-CD31 primary antibody (1:200, ab28364, Abcam) in 0.3% PBST for 24 hours at 37°C with gentle agitation. After three one-hour washes with 0.3% PBST, the slices were incubated with a secondary antibody (1:1000, 711-605-152, Jackson ImmunoResearch) in 0.3% PBST for 18 hours at room temperature with gentle shaking. Following another set of three one-hour washes with 0.3% PBST, the nuclei were stained with Hoechst (1:2000, H3570, Invitrogen) for 4 hours at room temperature. All slices were imaged at a voxel resolution of 1*1*3.5 μm³ for two channels using a custom-built light sheet microscope as described in previous study.⁷⁵

Microscopic image reconstruction and analysis

Microscopic data were reconstructed into 4*4*4 μm³ voxel-sized 3D images for subsequent analysis.⁷⁶ The blood vessels were segmented, and their density was quantified utilizing Imaris (v9.8.2, Oxford Instruments). Statistical analysis between different molecular subgroups was conducted using one-way ANOVA.

Missing data imputation

As the primary endpoint for survival analysis, the overall survival (OS) was defined as the time from date of surgery to death for deceased patients, which was right censored at the date of last follow-up for alive patients. For a small proportion of patients (40 subjects), the precise death time was not available. Instead, a time window containing the death time was observed for those patients inducing interval censoring. Interval-censored death times (missing rate, 4.28%) were accounted for by use of multivariate imputation from chained equation⁷⁷ (MICE package in R). To retain data fidelity, four candidate algorithms (predictive mean matching [PMM], classification and regression trees [CART], random forest [RF], and sample) were evaluated on the complete data (random 20%) using metrics of mean absolute error and root mean square error (Table S14). CART was chosen to perform the imputation, and the imputed OS was bounded by the last follow-up for deceased patients. In multivariable Cox regression analysis, we also accounted for the missing data in clinicopathological features by imputation. More details are provided in the following.

QUANTIFICATION AND STATISTICAL ANALYSIS

P-values were obtained using chi-square tests for categorical features, and Fisher's exact test was applied for categorical features with expected counts less than five. For continuous data, we utilized ANOVA and the Kruskal-Wallis Rank Sum Test to compare mean and median values, respectively. The significance threshold was set at P < 0.05 for all statistical tests in this study. Kaplan-Meier curves with log-rank tests were employed to compare patient subgroups.

For the prediction of molecular subgroups, we assessed the performance of the machine learning (ML) models using the area under the curve (AUC) of the receiver operating characteristic (ROC) and the F1 score, which represents the harmonic mean of the predictive positive value and sensitivity, accompanied by two-sided 95% confidence intervals (CIs). To ensure reliability and reduce variability, each molecular prediction experiment underwent ten replications, ensuring the robustness of the reported performance metrics in our study. Furthermore, we employed Shapley values to gauge the influence of each input modality, including radiomics and human-accessible MRI imaging signatures. The Shapley value, a principled attribution method in artificial intelligence, assigns an importance value to each input feature, thereby quantifying its contribution to a particular prediction.¹⁹