Radiomics Can Distinguish Pediatric Supratentorial Embryonal Tumors, High-Grade Gliomas, and Ependymomas


AJNR Am J Neuroradiol 2022, 43 (4) 603-610
doi: https://doi.org/10.3174/ajnr.A7481
http://www.ajnr.org/content/43/4/603
Radiomics Can Distinguish Pediatric Supratentorial Embryonal Tumors, High-Grade Gliomas, and Ependymomas


ABSTRACT

BACKGROUND AND PURPOSE: Pediatric supratentorial tumors such as embryonal tumors, high-grade gliomas, and ependymomas are difficult to distinguish by histopathology and imaging because of overlapping features. We applied machine learning to uncover MR imaging–based radiomics phenotypes that can differentiate these tumor types.

MATERIALS AND METHODS: Our retrospective cohort of 231 patients from 7 participating institutions had 50 embryonal tumors, 127 high-grade gliomas, and 54 ependymomas. For each tumor volume, we extracted 900 Image Biomarker Standardization Initiative–based PyRadiomics features from T2-weighted and gadolinium-enhanced T1-weighted images. A reduced feature set was obtained by sparse regression analysis and was used as input for 6 candidate classifier models. Training and test sets were randomly allocated from the total cohort in a 75:25 ratio.

RESULTS: The final classifier model for embryonal tumor-versus-high-grade gliomas identified 23 features with an area under the curve of 0.98; the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 0.85, 0.91, 0.79, 0.94, and 0.89, respectively. The classifier for embryonal tumor-versus-ependymomas identified 4 features with an area under the curve of 0.82; the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 0.93, 0.69, 0.76, 0.90, and 0.81, respectively. The classifier for high-grade gliomas-versus-ependymomas identified 35 features with an area under the curve of 0.96; the sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were 0.82, 0.94, 0.82, 0.94, and 0.91, respectively.

CONCLUSIONS: In this multi-institutional study, we identified distinct radiomic phenotypes that distinguish pediatric supratentorial tumors, high-grade gliomas, and ependymomas with high accuracy. Incorporation of this technique in diagnostic algorithms can improve diagnosis, risk stratification, and treatment planning.

ABBREVIATIONS: AUC = area under the curve; EP = ependymoma; GLCM = gray-level co-occurrence matrix; HGG = high-grade glioma; LR = logistic regression; NPV = negative predictive value; PNET = primitive neuroectodermal tumor; PPV = positive predictive value; WHO = World Health Organization; XGB = extreme gradient boosting; LASSO = least absolute shrinkage and selection operator

Pediatric supratentorial embryonal tumors, high-grade gliomas (HGGs), and ependymomas (EPs) can be difficult to differentiate by both imaging and histopathology because of overlapping features.1,2 Given the vastly different treatment approaches and prognoses, accurate diagnosis of these entities is extremely important; however, it requires advanced immunohistochemistry and molecular analyses, which have substantial practical barriers of availability, timeliness, and cost.3,5

Embryonal tumors of the CNS are highly malignant, undifferentiated, or poorly differentiated tumors of neuroepithelial origin, a category that has continuously evolved during the past few decades, reflecting an improving understanding of tumor biology.1,6 The nomenclature of supratentorial HGG has also
changed across the years, including major updates in the 2021 World Health Organization Classification of CNS Tumors (WHO CNS5), with separation of “adult-type” and “pediatric-type” gliomas and further subgrouping based on specific genetic mutations. The term “anaplastic astrocytoma” has been discontinued, and “glioblastoma” is no longer used in the pediatric context.7 Supratentorial EPs have been shown to be biologically distinct from the more common infratentorial counterparts, with different cells of origin and specific genetic mutations.8,9 Superficial embryonal tumors, HGGs, and EPs all demonstrate aggressive behavior, and routine histopathology may be unreliable in accurately differentiating these tumor types.

Recent advances in machine learning and computer vision in medicine offer a new potential for precision in oncology whether it is classification of the tumor subgroup or prognosis. For example, feature extraction, such as radiomics, enables mining of high-dimensional, quantitative image features that facilitate data-driven, predictive modeling. With such approaches, computational algorithms assign probabilities for diagnoses based on quantitative analyses of tumor voxels on imaging.10,11 Prior studies have used various machine learning approaches to separate the different posterior fossa tumors in children, to predict the molecular subtypes for pediatric medulloblastomas and adult high-grade gliomas, and for development of prognostic biomarkers for various tumors.12-19 Here, we present a large multi-institutional cohort of pediatric supratentorial tumors for MR imaging–based radiomics analysis, in an attempt to identify quantitative imaging features and radiomic profiles that can help distinguish these tumors types.

MATERIALS AND METHODS

Study Population
We performed a multi-institutional, retrospective study after institutional review board approval (No. 51059) at participating institutions (Online Supplemental Data) with a waiver of consent. Stanford served as the host institution and executed site-specific data-use agreements. The inclusion criteria were consecutive patients with pathologically confirmed supratentorial embryonal tumors, HGGs, and EPs spanning 2003–2021, nineteen years of age or younger, and with preoperative MR imaging that included both axial T2-weighted and gadolinium-enhanced T1-weighted sequences. For this retrospective study, the original tumor type assignments were based on the older WHO classifications. The HGG group included anaplastic astrocytomas (grade III) and glioblastomas (grade IV); both terms have been discontinued in the 2021 WHO Classification. All supratentorial EPs, regardless of the pathologic grade (grade II or III), were included in the study. We excluded patients if the MR imaging was nondiagnostic or had artifacts.

Imaging Techniques

MR imaging brain scans were performed on 1.5 or 3T MR imaging scanners across centers using the following vendors: GE Healthcare, Siemens, Philips Healthcare, and Toshiba Canon Medical Systems. The T2-MR imaging sequence parameters were the following: T2 TSE clear/sensitivity encoding, T2 FSE, T2 PROPELLER, T2 BLADE (Siemens), T2 DRIVE sensitivity encoding (TR/TE = 2475.6–9622.24/80–146.048 ms); section thickness = 1–5 mm; 0.5- or 1-mm skip; matrix ranges = 224–1024 × 256–1024. T1-MR imaging sequences comprised T1 MPRAGE, T1 axial MRI 3D brain volume, T1 fast-spoiled gradient recalled, T1 echo-spoiled gradient echo, and T1 spin-echo (section thickness = 0.8–1.2 mm; matrix = 256–512 × 256–512).

Feature Extraction and Reduction
One blinded neuroradiologist attending physician (reader 1, K.W.Y.) independently segmented the volumetric whole-tumor boundary on both T2-MR imaging and T1-MR imaging, inclusive of solid, cystic, and hemorrhagic components, excluding perilesional edema. The T2-MR imaging was used as the baseline for tumor segmentation, and the ROI was manually overlaid onto the T1-MR imaging. A second blinded neuroradiologist attending physician (reader 2, A. J.) confirmed tumor boundary delineation. Normalization was performed by normalizing the intensities by centering at the mean (SD), with a scaling factor of 100. Isotropic voxel resampling was performed to 1 × 1 × 1 mm³. A bin width of 10 was used for gray-level discretization in both normalized MR images. Both the normalization and resampling elements are further detailed in the Online Supplemental Data. From each tumor volume, we extracted 1800 (900 each from T2-MR imaging and T1-MR imaging) Image Biomarker Standardization Initiative–based20,21 PyRadiomics features (2.2.0.post7+ gac7458e; https://aim.hms.harvard.edu/pyradiomics) using the Quantitative Image Feature Pipeline (Online Supplemental Data).22 Extracted features underwent sparse regression analysis by a least absolute shrinkage and selection operator (LASSO) on RStudio 1.2.5033 (https://www.rstudio.com/products/rstudio/download/; Online Supplemental Data). We conducted feature selection from the entire cohort given our relatively small data set size and addressed this potential limitation by performing internal cross-validated LASSO (glmnet package; https://glmnet.stanford.edu/articles/glmnet.html) to obviate overfitting.

Binary Classifier Training and Testing

For each binary classifier model, we first conducted feature reduction using the extracted feature set and clinical variables (age at diagnosis and sex) as input. The corresponding reduced feature set was then submitted to train 6 candidate classifiers to identify the best-performing algorithm. The 6 candidate classifiers included support vector machine, logistic regression (LR), k-nearest neighbor, random forest, extreme gradient boosting (XGB), and neural net. Training and test sets were randomly allocated from the total cohort in a 75:25 ratio. The training cohort underwent resampling to correct for sample imbalance. Embryonal tumors were designated as the positive class in classifiers containing such pathologies. For the classifier between EP and HGG, EP was designated as the positive class. Optimal classifier parameters were estimated by a grid search (Online Supplemental Data). The relative influences of imaging features were calculated for the optimal classifiers, namely, feature coefficients for LR and percentage gain for tree-based classifiers.

Single-Stage Multiclass Classifier Model

To compare the performance of multiple individual binary primary models (embryonal tumor versus HGG; embryonal tumor versus EP; EP versus HGG) with that of a single multiclass model, we used...
the same 6 candidate classifiers to perform a multiclass classification across the 3 tumor groups: embryonal tumor, HGG, and EP.

Statistical Analysis
A P value < .05 was considered statistically significant for all analyses. We calculated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and area under the curve (AUC) for each classifier. The accuracy confidence interval was compared with the no-information rate, which was calculated from the prevalence of the more populous class within a binary pairing (Wald statistic). Confidence intervals were obtained by bootstrapping the test sets for 2000 random samples. Classifier development was performed using Python 3.8.5 (https://www.python.org/downloads/release/python-385/). Feature reduction and statistics were calculated with RStudio 1.2.503.

RESULTS

Patient Cohort
Of the 271 patients who were shared by participating sites, 231 met the final inclusion criteria. Reasons for exclusion were lack of either axial plane T2-MR imaging or T1-MR imaging or artifacts. A few patients were excluded due to infratentorial tumor location. There were 50 (21.6%) embryonal tumors, 127 (55.0%) HGGs, and 54 (23.4%) EPs, with pathologic subtypes as detailed in the Online Supplemental Data. The mean ages at diagnosis were 69.3, 138.1, and 87.3 months, respectively.

Embryonal Tumor and High-Grade Glioma Classifier
The subsequent classifier for embryonal tumor and HGG identified 23 features (Online Supplemental Data). These features entailed 1 clinical feature (age), 9 from T1-MR imaging, and 13 from T2-MR imaging, including 6 first-order, 2 shape, and 14 textural features (8 gray-level co-occurrence matrix [GLCM], 5 gray-level size zone, 1 gray-level run length matrix). Among the 6 classifier models, LR showed highest performance (AUC = 0.98) (Online Supplemental Data) with a sensitivity, specificity, PPV, NPV, and accuracy of 0.85, 0.91, 0.79, 0.94, and 0.89, respectively. The top 3 relevant features included age, T2-cluster shade (GLCM), and T1-mean (first-order intensity, Fig 1, and Online Supplemental Data). Accuracy was significantly higher than the no-information rate (P < .001). Metrics from all 6 classifier models are provided in the Online Supplemental Data.

Embryonal Tumor and Ependymoma Classifier
In the binary classifier for embryonal tumor and EP, LASSO regression identified 4 relevant Image Biomarker Standardization Initiative features, with 2 from T1-MR imaging and 2 from T2-MR imaging (Online Supplemental Data), including 3 first-order features and 1 textural feature (1 GLCM). Among the 6 classifier models,
XGB had the best performance (AUC = 0.82) (Online Supplemental Data). The top 3 relevant features included T2-kurtosis (first-order), T1-informational measure of correlation (GLCM), and T1-skewness (first-order, Fig 2 and Online Supplemental Data). Sensitivity, specificity, PPV, NPV, and accuracy were 0.93, 0.69, 0.76, 0.90, 0.81, respectively. Accuracy was statistically greater than the no-information rate ($P = .001$). Metrics from all 6 classifier models are provided in the Online Supplemental Data.

**Ependymoma and High-Grade Glioma Classifier**

Finally, a classifier performed for HGG and EP identified 35 features (Online Supplemental Data), including 1 clinical feature, 16 from T1-MR imaging, and 18 from T2-MR imaging, including 8 first-order, 1 shape, and 25 textural features (11 GLCM, 10 gray-level size zones, 4 gray-level run length matrix). Among the 6 classifier models, neural net showed the highest performance (AUC = 0.96) (Online Supplemental Data) with a sensitivity, specificity, PPV, NPV, and accuracy of 0.82, 0.94, 0.82, 0.94, and 0.91, respectively. The top 3 relevant features included T1-mean (first-order intensity), T1-cluster shade (GLCM), and T2-maximal correlation coefficient (GLCM, Fig 3, and Online Supplemental Data). Accuracy was statistically higher than the no-information rate ($P < .001$). Metrics from all 6 classifier models are provided in the Online Supplemental Data.

**Single-Stage Embryonal Tumor, High-Grade Glioma, Ependymoma Classifier**

The performance of this multiclass classifier was inferior to the above-described binary classifiers, and the metrics stemming from this model are included in the Online Supplemental Data.

**DISCUSSION**

In this multi-institutional study, we constructed machine learning classifiers to identify MR imaging–based radiomics phenotypes that distinguish supratentorial embryonal tumors, HGG, and EP. Our study represents the largest study to date on imaging of pediatric supratentorial tumors and the first one to apply radiomics.

Histopathologic features of embryonal tumors, HGG, and EP can overlap and require immunohistochemistry and/or molecular profiling for accurate diagnosis. Also, recent clinical trials have reported that rates of discordance between central and site pathologic review range between 28% and 38%, further highlighting the difficulties in accurate pathologic diagnosis.\(^1,4,23\) The diagnosis of embryonal tumors from other entities is particularly challenging. In the past, the histologically diagnosed category of primitive neuroectodermal tumors (CNS-PNET) was considered synonymous with embryonal tumors. However, molecular profiling using genome-
wide DNA methylation of CNS-PNETs has revealed that this group comprises disparate entities including embryonal tumors as well as nonembryonal tumors such as HGG and EP, thereby leading to discontinuation of the term CNS-PNET in the WHO Classification.1,2 The supratentorial embryonal tumors include a broad group termed "CNS embryonal tumors (not otherwise specified)" and some more specific entities like embryonal tumors with multilayered rosettes.6 In addition to these, supratentorial embryonal tumors have traditionally included atypical teratoid/rhabdoid tumors and pineoblastomas.6,8 Supratentorial embryonal tumors represent approximately 15% of CNS neoplasms in children and are biologically distinct from medulloblastomas.24

High-grade gliomas constitute 8%–12% of all pediatric CNS neoplasms, and one-third of these are supratentorial.3,25 The 2021 WHO CNS5 places adult and pediatric HGG in separate categories, which are further subdivided on the basis of a complex spectrum of genomic abnormalities.7 In contrast to adult-type HGGs, pediatric HGGs are typically IDH wild-type and demonstrate histone mutations in more than half the cases.26 Ependymomas constitute 10% of all primary CNS neoplasms in children, and 40% are supratentorial with most in a parenchymal location.24,27,28 Supratentorial EPs are now identified as genetically distinct from infratentorial and spinal EPs; WHO CNS5 has introduced genetically defined subgroups of ZFTA fusion-positive and YAP1 fusion-positive for supratentorial EPs, with the former demonstrating more aggressive clinical behavior.7,8

Histopathologic grading of EPs has been controversial with regard to its reproducibility and clinical significance. Although EPs can be either grade II or III, the clinical outcome is poorly correlated with tumor grade; therefore, all EPs regardless of the grade were included in this study.29

There are only a few published studies on the imaging appearance of pediatric supratentorial high-grade tumors.27,30-34 The only study comparing the imaging features of supratentorial embryonal tumors with other high-grade tumors (HGG and EP) concluded that it is not possible to distinguish these entities by conventional MR imaging.30 A prior report compared the MR imaging findings of CNS-PNET not otherwise specified with ependymoblastomas and ependymomas, and although the authors found some differences on imaging, their conclusion was that precise distinction is not feasible.35 All of these high-grade tumors have overlapping imaging appearances and typically present as large, heterogeneous, diffusion-restricting, hemispheric, or ventricular masses with variable cystic and necrotic changes. Enhancement is usually present but can vary in extent and intensity.24

Our radiomic models demonstrated high predictive accuracy for each of the embryonal tumor versus HGG, embryonal tumor versus EP, and HGG versus EP classifiers. The final model for embryonal tumor versus HGG selected age as one of the dominant contributors, which is congruent with the reported propensity of embryonal tumors to occur in younger children, and HGG, in the adolescent age group.24 The other 2 models selected...
purely MR imaging–based radiomic features. One of the advantages of the radiomics technique is that it allows identification of specific computational features that drive model prediction, thus offering some transparency compared with the “black box” nature of deep learning. For the embryonal tumor-versus-HGG classifier, the embryonal tumors demonstrated more balanced T2 voxel intensities around the mean intensity and were overall brighter on T1 postcontrast imaging (Fig 1). For the embryonal tumor-versus-EP classifier, the embryonal tumors demonstrated overall darker voxel intensities on T2, while EPs had more homogeneous texture on T1 postcontrast images (Fig 2). The performance of the embryonal tumor-versus-HGG model was stronger compared with the embryonal tumor-versus-EP model. For the HGG-versus-EP classifier, EPs were overall brighter with more balanced signal intensities around the mean on T1 postcontrast images and had a more “complex” texture involving a greater proportion of brighter intensities on T2-weighted images (Fig 3).

Examples of model-derived probability output are shown on test cohorts of supratentorial embryonal tumors (EMB), EP, and HGG that did not participate in training. Due to overlap in macroscopic features of these malignant supratentorial tumors (eg, a wide range in size, morphology, and enhancement/intensity features), independent binary classifiers that specifically targeted feature separation for EMB versus EP [A, XGB], EMB versus HGG [B, LR], and HGG versus EP [C, Neural network [NN]] were found predictive over a single multiclass classifier. Examples of the same EMB tumors that were separately submitted into XGB and LR models are shown (asterisk) and show strong EMB discrimination against EP and HGG, respectively. In 1 example, the same EP tumor could be distinguished from EMB (yellow arrow) but was not predictive against HGG (gray arrow). ATRT indicates atypical teratoid/rhabdoid tumors; ETMR, embryonal tumor with multi-layered rosettes; NB, CNS neuroblastoma; NOS, not otherwise specified.

A common limitation of radiomics lies in replicability when obscure algorithms are used for feature extraction. Thus, we used the publicly available PyRadiomics package to compute features, as defined by the Imaging Biomarker Standardization Initiative, for future reproducibility.20

We note several limitations, including the small cohort size of each tumor type related to its relative rarity. Nevertheless, our cohort represents the largest imaging study of supratentorial tumors to date with data pooled from multiple institutions. There were institutional differences in MR imaging acquisition techniques, sequence availability, and image quality; however, we identified discriminating features that are retained despite diverse imaging protocols and vendors that may facilitate future generalizability and usability across centers. While the use of an independent institution outside of training would be desirable to show model generalization, this was not feasible due to uneven distribution of the tumor types across institutions. A future larger cohort study could build on our pilot results and further examine the robustness of radiomics-based separation of these supratentorial tumors. Additional imaging sequences such ADC and DWI, which may have predictive information, were excluded to preserve a robust sample size. We extracted radiomics features from isolated tumors and thus did not incorporate spatial relationship. Future design could consider combining radiomics and deep learning approaches that can intake whole-brain MR imaging for feature extraction and thereby assimilate tumor spatial features. While we performed intensity normalization and isotropic voxel resample, incorporation of other preprocessing steps would be desirable to further enhance the reproducibility and generalization of MR imaging–based radiomics classification.
CONCLUSIONS

Accurate pathologic diagnosis of supratentorial tumors often requires advanced immunohistochemistry and molecular analyses. These techniques are not readily available outside a handful of brain tumor centers and can be prohibitively expensive. Also, final diagnosis may take multiple weeks and is often not available for initial surgical and treatment planning. Conventional MR imaging is also of limited utility in distinguishing these tumors. Our MR imaging–based radiomic phenotypes demonstrated high accuracy and provided a rapid, readily available tool that can help provide a more accurate imaging diagnosis or a narrower differential diagnosis. This result in conjunction with initial histopathology can be more effective in guiding the surgery, treatment planning, and prognostication and can improve the overall outcomes of these patients. In recent years, standardization of quantitative image features by the radiology and bioinformatics community now enables potential deployment of such image-derived variables with fidelity in the clinical and bioinformatics community now enables potential deployment of such image-derived variables with fidelity in the clinical environment across centers. Pediatric embryonal tumors, HGGs, and EPs also have a wide and complex spectrum of genomic features involving several oncogenic pathways that can further affect the therapeutic strategies, and noninvasive distinction among these would be the next frontier for machine learning–based imaging techniques.

Disclosure forms provided by the authors are available with the full text and PDF of this article at www.ajnr.org.

REFERENCES