Radiomic Phenotypes Distinguish Atypical Teratoid/Rhabdoid Tumors from Medulloblastoma


AJNR Am J Neuroradiol 2021, 42 (9) 1702-1708
doi: https://doi.org/10.3174/ajnr.A7200
http://www.ajnr.org/content/42/9/1702
ABSTRACT

BACKGROUND AND PURPOSE: Atypical teratoid/rhabdoid tumors and medulloblastomas have similar imaging and histologic features but distinctly different outcomes. We hypothesized that they could be distinguished by MR imaging–based radiomic phenotypes.

MATERIALS AND METHODS: We retrospectively assembled T2-weighted and gadolinium-enhanced T1-weighted images of 48 posterior fossa atypical teratoid/rhabdoid tumors and 96 match-paired medulloblastomas from 7 institutions. Using a holdout test set, we measured the performance of 6 candidate classifier models using 6 imaging features derived by sparse regression of 900 T2WI and 900 TIWI Imaging Biomarker Standardization Initiative–based radiomics features.

RESULTS: From the originally extracted 1800 total Imaging Biomarker Standardization Initiative–based features, sparse regression consistently reduced the feature set to 1 from TIWI and 5 from T2WI. Among classifier models, logistic regression performed with the highest AUC of 0.86, with sensitivity, specificity, accuracy, and F1 scores of 0.80, 0.82, 0.81, and 0.85, respectively. The top 3 important Imaging Biomarker Standardization Initiative features, by decreasing order of relative contribution, included voxel intensity at the 90th percentile, inverse difference moment normalized, and kurtosis—all from T2WI.

CONCLUSIONS: Six quantitative signatures of image intensity, texture, and morphology distinguish atypical teratoid/rhabdoid tumors from medulloblastomas with high prediction performance across different machine learning strategies. Use of this technique for preoperative diagnosis of atypical teratoid/rhabdoid tumors could significantly inform therapeutic strategies and patient care discussions.

ABBREVIATIONS: ATRT = atypical teratoid/rhabdoid tumor; AUC = area under the curve; GLCM = gray level co-occurrence matrix; MB = medulloblastoma

A typical teratoid/rhabdoid tumors (ATRTs) are rare, but-aggressive neoplasms that often affect very young children.1-3 They are classically characterized by rhabdoid cells and divergent differentiation along neuroectodermal, mesenchymal, and epithelial lines. However, many ATRTs often lack rhabdoid cells and are simply dense, small, round, blue cell–rich lesions that mimic medulloblastomas (MBs, Online Supplemental Data).3,4 Whereas most ATRTs may be distinguished from MBs by immunohistochemical confirmation of SMARCB1 (INI1/BAF47/SNF5) loss (Online Supplemental Data),1-7 up to 22% of ATRTs retain the protein marker.3,8,9

Presurgical distinction of ATRT from MB is not possible by human interpretation of MR imaging; both primarily occupy the posterior fossa, share low T1- and T2-weighted intensities and variable enhancement, and have a reduced diffusion characteristic of
dendly packed cellular tumors (Online Supplemental Data). However, if it were possible, this distinction could add value because their different behaviors demand different treatment strategies. Median survival for patients with ATRTs is approximately 1 year, while the 5-year survival rate for pediatric MB is approximately 70%. Thus, an anticipated diagnosis of ATRT may prompt discussion of maximal surgical resection and aggressive adjuvant therapy.

Recent advances in machine learning and computer vision in medicine offer new potentials for precision in oncology, whether it is for tumor subgroup classification or prognosis. For example, feature extraction, such as in radiomics, enables mining of high-dimensional, quantitative image features that facilitate data-driven, predictive modeling. The resulting computational algorithm assigns probabilities for diagnoses and outcomes on the basis of its quantitative analysis of tumor voxels on imaging. While studies have reported various machine learning approaches to MR imaging—based radiomic phenotypes that distinguish ATRT from MB, in part, due to the rarity of ATRT.

Radiomics has the potential to not only uncover quantitative image features that may otherwise be imperceptible to the human eye but also offers interpretability of computational features that drive model prediction—a potential advantage over deep learning, in which learned features remain opaque. In this multicenter study, we applied machine learning to uncover MR imaging—based radiomic phenotypes that distinguish ATRT from MB.

**MATERIALS AND METHODS**

**Study Population**

We conducted a retrospective study after obtaining institutional review board approval (No. 51059) and data-sharing agreements with 7 participating institutions (Online Supplemental Data): Stanford Children’s (ST-Palo Alto, California), Lurie Children’s Hospital of Chicago (CG-Chicago, Illinois), Primary Children’s Hospital (UT-Salt Lake City, Utah), New York University Langone Medical Center (NY-New York, New York), Children’s Hospital Orange County (CH-Irvine, California), Indiana University Riley Hospital for Children (IN-Indianapolis, Indiana), and Tepecik Health Sciences (TK-Izmir, Turkey). We performed a chart review to identify patients with ATRTs and MBs. Inclusion criteria were the following: 1) Patients underwent preoperative MR imaging with gadolinium-enhanced T1WI and T2WI, and 2) surgical specimens of the tumor served as ground truth for pathology, including loss of INI-1 staining to confirm ATRT. Patients were excluded if MR imaging was degraded by motion or other artifacts or was considered nondiagnostic. When available, tumor molecular subgroup information was recorded. To increase the available training information and given the availability of additional MB data, we included twice the number of patients with MB relative to ATRT in the study. The initial MB cohort was randomly match-paired by institution, sex, and age with the ATRT cohort. To avoid overfitting from class imbalance, the ATRT cohort was oversampled to match the number of MBs in the training cohort.

**MR Imaging Acquisition**

MR imaging brain scans were acquired at either 1.5 and 3T using the following vendors: GE Healthcare (Signa Artist, Discovery 750, Optima 360, Signa Excite, Signa HDxt, Signa Explorer, Optima 450w), Siemens (Aera, Skyra, Avantofit, Espree, Symphony, Symphony Vision, Trio), Philips Healthcare (Ingenia, Intera, Achieva), and Toshiba Canon Medical Systems USA. The T2WI scans were the following: T2 TSE constant level appearance/sensitivity encoding, T2 fast-spin-echo, T2 PROPELLER, T2 BLADE (Siemens), T2 drive sense (TR/TE = 2745.6–9622.24/80–146.048; section thickness = 1–5 mm with a 0.5- or 1-mm skip; matrix ranges = 224–1024 × 256–1024). T1WI postgadolinium MR imaging scans included T1 MPRAGE, T1 BRAVO (GE Healthcare), T1 fast-spooled gradient recalled, T1 spoiled gradient recalled, and T1 spin-echo (section thickness = 0.8–1.2 mm, matrix ranges = 256–512 × 256–512). All image data were obtained in DICOM format.

**Image Preprocessing and Feature Extraction**

The volumetric whole-tumor boundary, inclusive of solid and cystic components, was delineated (K.W.Y.) and confirmed (A.J.) by board-certified attending neuroradiologists with Certificates of Added Qualification (K.W.Y., A.J., with >10 years’ experience) using OsiriX Imaging Software (http://www.osirix-viewer.com). We used PyRadiomics software (Version 2.2.0, post7+gac7458e; https://github.com/AIM-Harvard/pyradiomics) for feature extraction with implementation in the Quantitative Image Feature Pipeline (http://qifp.stanford.edu). The configuration files for radiomic feature extraction are included in the Online Supplemental Data.

A total of 1800 features (900 each from T2WI and T1WI) was automatically extracted on tumor volume including the following: first order statistics, 2D/3D shape, gray level co-occurrence matrix (GLCM), gray level run length matrix, gray level size zone matrix, neighboring gray-tone difference matrix, and gray level dependence matrix, as defined by the Imaging Biomarker Standardization Initiative. MR imaging studies were normalized for voxel size (1 × 1 × 1 mm) and intensity (scale factor of 100). A fixed bin width (10) was used for gray-value discretization. Preprocessing filters included wavelet (8 coefficients) and Laplacian of Gaussian (3 σ). Feature extraction was calculated for classes including first order statistics, shape descriptors, and gray level derivatives.

**Feature Reduction**

Training and test sets were randomly allocated from the total cohort in a 70:30 ratio. Feature selection for the allocated training set was performed using sparse regression analysis by a Least Absolute Shrinkage and Selection Operator, performed with 10-fold cross-validation and repeated for 1000 cycles. The mean squared error was calculated for 100 lambdas in each cycle or until a minimum was achieved. The optimal L was identified as the lowest mean squared error value and used for feature reduction and coefficient calculations. Both radiologic and clinical variables were incorporated at this stage into the primary model. Selected features represented in ≥80% of the cycles were retained for subsequent classifier optimization.

Feature reduction was performed using R Studio, Version 1.2.5033 (http://rstudio.org/download/desktop).

**Classifier Model Building and Analysis**

The retained features were submitted to 6 training models, including support vector machine, logistic regression, k-nearest neighbors, random forest, eXtreme Gradient Boosting, and neural net. The cohort underwent resampling to correct for sample imbalance. Training and test sets were randomly allocated from the total cohort in a 75:25 ratio. MB tumor was designated the positive class. Optimal classifier parameters were performed by grid search (Online Supplemental Data). The optimal radiomics classifier was selected by maximizing the area under the curve (AUC). Confidence intervals for each metric were obtained by bootstrap-ping of the test sets for 2000 random samples. Relative influence of the radiologic features was calculated for logistic regression and tree-based models, random forest, and eXtreme Gradient Boosting. Model training was performed using Python, Version 3.8.5.

**Qualitative Evaluation by Human Reader**

Two human experts (K.W.Y., A.J.) performed consensus review of T1WI and T2WI on the ATRT and MB cohorts, blinded to pathologic diagnosis or any clinical variables. The readers scored the degree of enhancement (0, no enhancement; 1, < 50% tumor volume with enhancement; 2, ≥ 50% tumor volume with enhancement) and the presence or absence of a cyst. Categoric variables were compared using the Fisher exact test, as appropriate. A P value < .05 was considered statistically significant for all analyses.

**RESULTS**

**Demographics and Clinical Information**

A total of 48 ATRTs (28 males [58.3%]; median age, 13.7 months; range, 1.0–114.6 months at diagnosis) and 96 patients with MB (61 males [63.5%]; median age, 83.0 months; range, 3.0–231.9 months at diagnosis) met the study criteria (Online Supplemental Data). MB molecular subgroup distribution is shown in the Online Supplemental Data. MB tumor was designated the positive class. Training and test sets were randomly allocated from the total cohort in a 75:25 ratio. MB tumor was designated the positive class.

**Feature Reduction and Model Performance**

Following feature reduction with sparse regression, 6 textural features were consistently selected in >80% of regression cycles, including 3 shape features, 2 first order features, and 1 GLCM feature (Online Supplemental Data), with 1 feature derived from T1WI, and 5, from T2WI. The single T1WI feature, elongation, was also represented among the T2WI features.

The performances of 6 models were evaluated on the holdout test, with logistic regression demonstrating the highest AUC of 0.8582 (Online Supplemental Data). Sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and F1 score were 0.80, 0.82, 0.91, 0.64, 0.81, and 0.85, respectively. The least effective classifier was neural net with an AUC of 0.73, closely followed by eXtreme Gradient Boosting with an AUC of 0.74. Among other models, k-nearest neighbors was notable, with the highest metrics other than AUC (0.84). Sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and F1 score were 0.80, 0.91, 0.95, 0.67, 0.83, and 0.87, respectively.

**Relative Influence of Variables**

Relative influence was assessed by logistic regression, random forest, and eXtreme Gradient Boosting (Fig 1, Fig 2 and Online Supplemental Data). In all classifiers, the voxel intensity at the 90th percentile was the most contributory, ranging from 24% to 40%. In the logistic regression, voxel intensity at the 90th percentile was also the only parameter that positively predicted ATRT. This was consistently followed by 2 other textural features, GLCM inverse difference moment normalized and kurtosis. The last 3 features (by relative importance) included T1WI and T2WI measurements for elongation and flatness within the segmented ROI. T1WI elongation was consistently the lowest contributing feature, ranging from 5.2% to 7.8% of classifiers.

**DISCUSSION**

In this multi-institutional study, we constructed machine learning classifiers to identify MR imaging–based radiomic phenotypes to distinguish ATRT from MB. This is the largest imaging dataset and first radiomics study of ATRT, a rare-but-aggressive neoplasm.

While loss of INI-1 immunohistochemical staining can confirm the diagnosis in most ATRTs, up to 22% of ATRTs may show no alteration. Other CNS tumors, such as oligodendroglioma or anaplastic oligoastrocytoma, may also have INI-1 inactivation. Complex immunophenotypes as well as overlapping histologic features can confound the pathologic diagnosis, particularly with extensive embryonal morphologic components. Thus, reports of pathologic misdiagnoses have included MB, various embryonal tumors, glioblastoma, and, occasionally, choroid plexus carcinoma, in which inactivation of INI-1 may be present.

Here, we identify 6 radiomic features, 1 derived from T1WI and 5 from T2WI, that together distinguish ATRT from MB by logistic regression with AUC = 0.86. Of these radiomic features, 3 describe T2WI-based voxel intensities and texture, and 3 describe tumor morphology. On the basis of blinded human expert review, we found overlap in visually determined, qualitative image features such as the presence of cysts, suggesting morphologic heterogeneity (eg, cysts/cavities) inherent in both ATRT and MB, as previously described. Most interesting, despite variable MB enhancement, human experts scored MB as enhancing over a larger tumor volume (≥50%) in contrast to ATRT, regardless of how brightly or faintly a tumor enhanced (Online Supplemental Data). However, at a quantitative level, tumor brightness that is calculated by first order radiomics features (eg, average intensity/brightness) on T1WI was not selected by our model; suggesting how brightly (or faintly) a tumor enhanced was not a distinguishing feature. Radiomic features of tumor volume and
diameter were also not selected, indicating that tumor size did not contribute.

Overall, T2WI-based voxel intensities were most relevant. For example, 90th percentile voxel intensity emerged as the most important variable, with a higher value associated with ATRT. More heterogeneous texture, as described by the GLCM-based feature inverse difference moment normalized, calculated by larger gradient changes in intensity between neighboring voxels, also predicted ATRT. Lower kurtosis or a wider distribution of voxel intensities was more characteristic of ATRT and similarly suggested a wider range in tissue composition.

The more heterogeneous texture of ATRT might reflect multiple histologic components of rhabdoid cells juxtaposed to embryonal cells and, sometimes, glial, mesenchymal, and/or epithelial differentiation, compared with more homogeneous and, classically, dense cellular sheet growth of MB.19,40,41 In combination, the myxoid background of gelatinous mucopolysaccharide-rich water content that ATRT is known to produce likely contributes to the high T2-voxel intensity value of ATRT.40,41

Prior studies have suggested that ATRT and MB both qualitatively display nondiscriminating, T2-heterogeneous signal.11,12,37,42-44 Applying a filter to an image before calculating radiomic features can capture patterns or highlight additional details within the image that might otherwise be imperceptible to the human eye. Here, we show that features derived from wavelet-filtered images (GLCM inverse difference moment normalized and kurtosis) can uncover textural differences that reside within tumor voxels. Furthermore, radiomics interrogates the

![Barplot of the reduced feature set and its relative influence as calculated by logistic regression, trained to distinguish ATRT and medulloblastoma. IDMN indicates inverse difference moment normalized; HLL, High/Low/Low; LLL, Low/Low/Low.](image1)

![Density plots. A. T2-weighted 90th percentile voxel intensity. B. T2-weighted inverse difference moment normalized. C. T2-weighted kurtosis. D. T2-weighted flatness. E. T2-weighted elongation. F. T1-weighted elongation among patients with ATRT and medulloblastoma. LLL indicates Low/Low/Low.](image2)
entire tumor phenotype before surgical disturbance, a distinct advantage over histology that probes tumor slices. Thus, heterogeneous texture might also reflect focal cysts, necrosis, and CSF clefts/spaces interspersed between tumor clusters unique to ATRT macro- or microenvironment, which may be difficult to identify either by histology or, qualitatively, on gross visual inspection (Fig 3).13,26,36,44

Most interesting, linear and planar morphology suggested ATRT, whereas more circular and spheric morphology suggested MB (Fig 3). The distribution of the elongation feature showed that low values, ie, those that were more linear, were very specific for ATRT. Conversely, the distribution of the flatness feature showed that the most extreme values, ie, those that were more spheric, were specific to MB. Both elongation and flatness derive from the ellipsoid axes underlying the ROI but mathematically differ on the basis of which secondary axis is used in its calculation (\( \sqrt{A_{\text{minor}}/A_{\text{major}}} \) versus \( \sqrt{A_{\text{major}}/A_{\text{least}}} \), respectively). While there may be some redundancy among these 3 features, their selection internally validates the use of ellipsoid dimensions as predictive features. These morphology features may reflect anatomic origins. Both tumors can occupy the cerebellum and vermis with involvement of the fourth ventricle.26,36,45 However, from a histogenic perspective, MBs are derived from the roof of the external granular layer of the fourth ventricle and expand radially in a spheric manner.10,41 Meanwhile, ATRTs are thought to have choroid plexus derivation, commonly lateralizing to the cerebellopontine angle, and may, thus, deform and flatten along its growth trajectory.35,46

The radiomics signatures had consistent performance across different machine learning models, with substantial overlaps in the AUC-confidence intervals of the support vector machine, logistic regression, and k-nearest neighbors models. The k-nearest neighbors, in particular, had high sensitivity and specificity scores, albeit a slightly lower AUC than logistic regression. This feature likely relates to the intrinsic model design of k-nearest neighbors, in which extreme scores are penalized when the parameter for number of neighbors is small. The tree-based classifiers (random forest, eXtreme Gradient Boosting, and neural net) had higher false-negative rates, implying misclassification of a number of MBs. We suspect overfitting during the training phase with these tree-based approaches, given the smaller difference between training error and testing error for the nontree models. A larger ATRT sample size could augment the training pool for better tree-based models.

We note several limitations, including the small cohort size of ATRT due to its rarity. Nevertheless, this is the largest ATRT imaging study to date, with data pooled from multiple institutions. While we describe features derived from T2WI and gadolinium-enhanced T1WI, it is possible that the use of additional MR imaging sequences, such as FLAIR, T2*, or DWI could further optimize the classifier and add new insight into significant radiomic signatures. Although desirable, we did not conduct

FIG 3. MR imaging correlates of radiomics phenotypes. Despite overlap in gross image features of MB and ATRT, unique quantitative radiomics features associated with shape and texture emerged as predictive features of ATRT and MB. For example, more heterogeneous features derived from GLCM-based texture or kurtosis-based wider distribution of voxel intensities were indicative of ATRT. Furthermore, more spheric morphology characterized MBs, compared with the more elongated or planar configuration of ATRT. Gross examples of the heterogeneous texture of ATRT are shown, including areas of mixed low and high T2-signal that might be seen with blood products, variations in tissue components, as well as cystic areas. While some ATRT tumors were round, many were quantitatively more elongated compared with the more spheric contour of many MB tumors. Despite the presence of cysts or T2-dark foci that might stem from blood products or vascularity, quantitatively, MB showed more even distribution of voxel intensities.
radiogenomics analysis of ATRTs because the molecular subgroup information was not available. Our radiomics analysis is contingent on a voxel-based analysis of tumor segmentations. Therefore, it does not identify other potentially useful semantic images features such as anatomic location, perilesional edema, or other features of the brain environment external to the tumor.\textsuperscript{11,13,47} Finally, our model was trained on infratentorial ATRTs and may not infer features of the supratentorial ATRT.

CONCLUSIONS

In this multi-institutional study, we constructed discovery-driven approaches to uncover distinctive MR imaging–based radiomic phenotypes of ATRT and MB. Image intensity, texture, and morphology had high predictive performance across different machine learning strategies. Despite several limitations, including lack of radiogenomics analysis of ATRT tumors, our results suggest potential future roles for machine-enabled classifiers to refine preoperative planning and patient family counseling. Future iterations may additionally incorporate tumor genomics to uncover the biologic significance of quantitative image phenotypes.

ACKNOWLEDGMENTS

We would like to thank Sara Norris of the Intermountain Health Care Imaging Department (Salt Lake City, Utah) for archival image acquisition.


REFERENCES


