



A comprehensive machine-learning model applied to MRI to classify germinomas of the pineal region



Ningrong Ye ^{a,b,1}, Qi Yang ^{a,b,1}, Peikun Liu ^{a,b}, Ziyan Chen ^{a,b}, Xuejun Li ^{a,b,*}

^a Department of Neurosurgery, Xiangya Hospital, Central South University, No. 87, Xiangya Road, Changsha, Hunan, 410008, PR China

^b Hunan International Scientific and Technological Cooperation Base of Brain Tumor Research, Xiangya Hospital, Central South University, No. 87, Xiangya Road, Changsha, Hunan, 410008, PR China

ARTICLE INFO

Keywords:

Radiomics
Pineal region tumor
Germinoma
Machine learning
Medical image

ABSTRACT

Background: Pineal region tumors (PRTs) are highly histologically heterogeneous. Germinoma is the most common PRT and is treatable with radiotherapy and chemotherapy. A non-invasive system that helps identify germinoma in the pineal region could reduce lab exams and traumatic therapies.

Methods: In this retrospective study, 122 patients with histologically confirmed PRTs and pre-operative multi-modal MR images were included. Radiomics features were extracted from different ROIs and image sequences separately. A computational framework that combines a few classification and feature selection algorithms were used to predict histology with radiomics features and demographics. We systematically benchmarked performance of models with feature matrices from all possible combinations of ROIs and image sequences. The Area under the ROC Curve (AUC) was then used to evaluate model performance.

Results: Models with demographics and radiomics features outperform radiomics-only or demographics-only models. The best demographical-radiomics model reached the highest AUC of 0.88 (CI95%: 0.81–0.96). Through the comprehensive evaluation of possible sequence combinations in the differential diagnosis of pineal tumor, T1 and T2 emerged as the most informative sequences for the task. There is imbalanced usage of feature classes as we analyze their proportion in all models.

Conclusions: The demographical-radiomics model can accurately and efficiently identify germinomas in the pineal region. The preference for MRI sequences, radiomics feature classes, features selection and classification algorithms provide a valuable reference for future attempts at developing classifiers on medical images.

1. Introduction

Pineal region tumors (PRTs) are rare midline deep intracranial lesions originating from pineal or para-pineal structures. PRTs comprised 0.2% of all US central nervous system (CNS) lesions [1]. However, it is more common in Asian countries like Japan (3% of CNS lesions) and China [2,3]. PRTs are classified into germinomas, non-germinoma germ cell tumors (NGGCTs), mixed germ cell tumors (mixed GCTs), gliomas, meningiomas, pineocytomas, pineal parenchymal tumors (PPTs), pineoblastomas, and other highly uncommon tumors by histology [4]. Patients with PRTs commonly develop symptoms related to increased intracranial pressure (ICP) due to the compression of the Sylvian aqueduct. Primary resection is the preferred therapy for most PRTs but not for germinomas. As highly radiosensitive, intracranial germinomas

can be treated with radiotherapy and chemotherapy [5,6]. Elevating level of Alpha-fetoprotein (AFP) and Beta human chorionic gonadotropin (β -hCG) in serum or CSF could help identify specific subtypes of NGGCTs. However, germinomas do not have accurate and consistent serum/CSF markers [5,6]. Moreover, the rarity of PRTs and their highly heterogeneous histology make it challenging to accurately diagnose without traumatic procedures like biopsy or surgery. A non-invasive system that helps identify germinoma in the pineal region could reduce lab exams and traumatic therapies.

The advances in medical images based computational assist diagnostic systems have shown potential in different types of tumors [7–10]. The previous study has demonstrated that MRI-based machine learning is capable of distinguishing germinoma from other tumors of the pineal region [11], the basal ganglia [12], and the anterior skull base [13].

* Corresponding author. Department of Neurosurgery, Xiangya Hospital, Central South University, No. 87, Xiangya Road, Changsha, Hunan, 410008, PR China.
E-mail address: lxjneuro@csu.edu.cn (X. Li).

¹ These authors have contributed equally to this work and share first authorship.

However, most studies were based on a selected cohort which would cripple the ability of clinical translation. Therefore, we sought to develop and validate a model with an unselected cohort of PRTs that could differentiate germinoma from other lesions. Moreover, little is known about the significance of each sequence in pre-operative multi-modal MR images. No study has explored which combination of MRI modalities is best for developing a discriminator of pineal lesions. Many previous investigations of radiomics-based discriminators used a single feature selection and/or a single classification algorithm, in this study, we also explored the impact of the commonly used algorithms on

model performance. A total of 122 patients with histologically confirmed PRTs and pre-operative multi-modal MR images were included in this retrospective analysis.

This study aims to determine which MRI sequence is most informative about the differentiation of germinoma and to develop a discriminator that could aid future diagnostic imaging of PRTs.

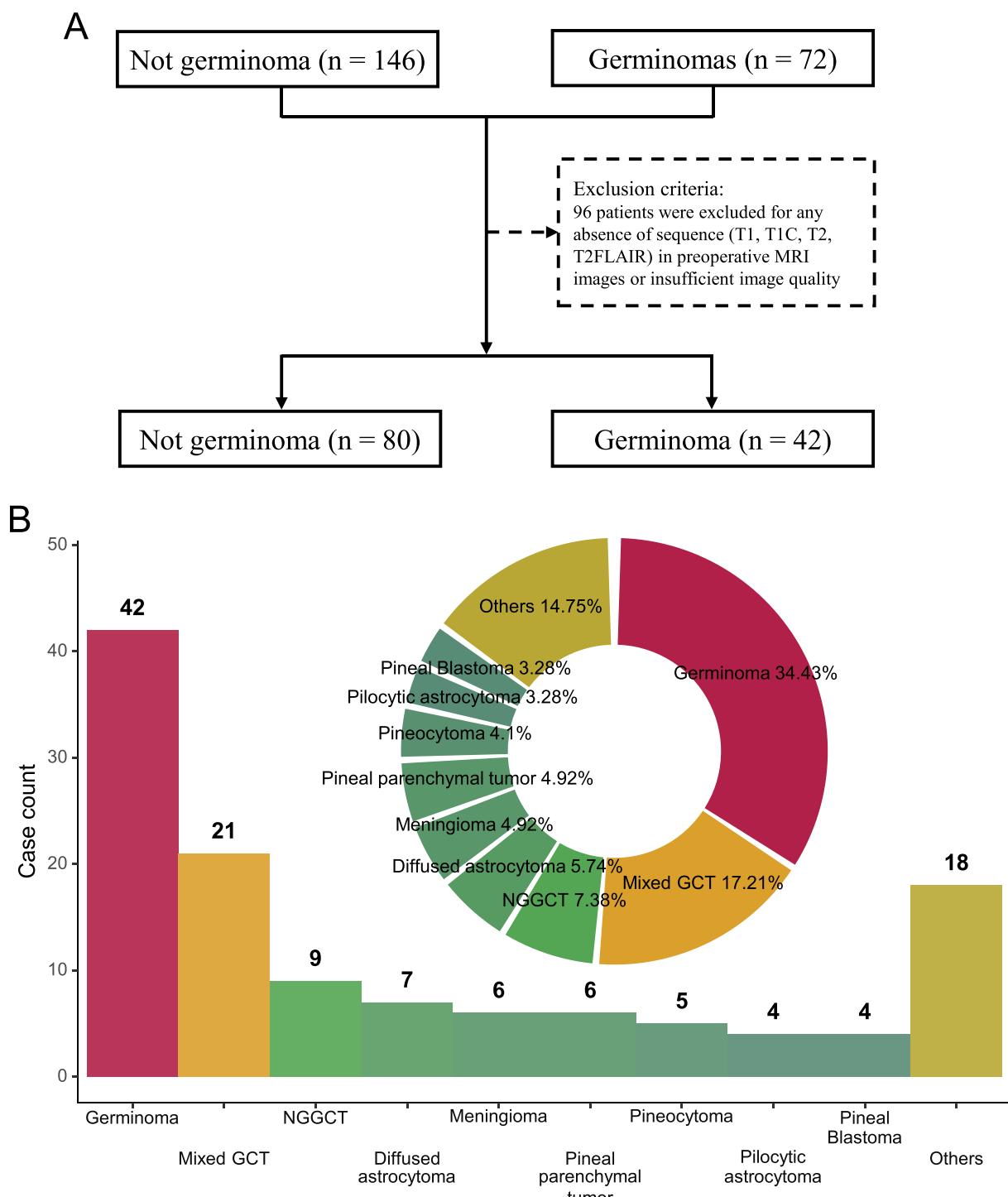


Fig. 1. A, flowchart for patient inclusion. B, disease spectrum of pineal region tumors. The histogram showed the frequency of each disease, and the donut chart showed the proportion of each disease.

2. Materials and methods

2.1. Patient cohort

This study was approved by the Human Investigation Committee at our hospital. Informed consent was waived because of the retrospective nature of the study and the analysis used anonymous clinical data. The study was conducted in accordance with the Declaration of Helsinki.

The inclusion criteria were:

- 1) Cases with pre-operative MRI images;
- 2) Cases with histology confirmed diagnosis;
- 3) No prior history of radiotherapy or chemotherapy.

A total of 218 patients from the Xiangya hospital from 2010 to 2020 passed preliminary selection. The exclusion criteria were:

- 1) Absence of any sequence in T1-weighted (T1), T2-weighted (T2), T1-weighted contrast-enhanced (T1C), and T2-weighted fluid-attenuated inversion recovery (T2 FLAIR);
- 2) Images with noise, blurring, motion artifact, and other interference information.

There were 96 patients excluded for reasons mentioned above. The cohort for the later analysis contains a total of 122 patients (Fig. 1A and B).

2.2. Image preprocessing and lesion labeling

Image preprocessing follows the pipeline of image transformation and re-slicing, N4 bias correction, image gray-level correction, multi-modal affine alignment, and image registration to a template. Delineation of tumor boundaries was performed semi-automated on a slice-by-slice basis using the ITK-SNAP software, an open-source 3D image analysis kit [14]. In addition, two neurosurgeons (NY and PL with over 7 and 4 years of experience, respectively) delineated the enhancing tumor regions (of T1C) and peritumoral edema regions (if there is, of T2 FLAIR) independently for further analysis (Fig. 2). A total of 122 lesions from 122 patients were segmented and labeled. Voxels of all images were resized into isotropic $1 \times 1 \times 1 \text{ mm}^3$ cubes. A bin width of 20 was set for the discretization of the image gray level.

2.3. Radiomics features extraction and harmonization

Radiomics features were extracted for each region of interest (ROI) using the Pyradiomics package [15] (<http://www.radiomics.io/pyradiomics.html>) in Python 3.8.0. Filters, including Original or Wavelet, were applied to the images before radiomics feature extraction. The minimum required ROI size was set to 5 voxels. The Distance parameter that specified the distance between the center and the neighbor voxels, for which angles should be generated was set to 1, 2, or 3. Radiomics features from the following seven feature classes were used in this study: first-order statistics (18 features), shape-based features (14 features), Gray Level Cooccurrence Matrix (GLCM, 24 features), Gray Level Run Length Matrix (GLRLM, 16 features), Gray Level Size Zone Matrix (GLSVM, 16 features), Neighbouring Gray Tone Difference Matrix (NGTDM, 5 features), Gray Level Dependence Matrix (GLDM, 14 features). Detailed names of each feature are provided in Supplementary Table 1. Feature extraction is performed separately for the tumor bulk and peritumoral edema. The results are two feature matrices for a given MRI sequence of a specific lesion (Fig. 2).

2.4. Radiomics features preprocessing

To evaluate the robustness of the radiomics features, we randomly choose 20% of the patients to do segmentation twice. The intra-class correlation coefficient (ICC) was calculated, and the unstable features ($\text{ICC} < 0.8$) were eliminated. ICC values of all the assessed features are reported in Supplementary Table 2 and Table 3 for ROI edema and tumor, respectively. In addition, a robust scaler transform was applied to minimize the effect of the outlier value. The median and interquartile range (IQR) between the 25th and 75th quantile of each radiomics feature were calculated from the training set and then applied to the test set. This can keep the testing set unseen during the feature preprocessing and model-building process.

2.5. Feature selection and classification

This study used four feature selection algorithms: elastic net, stochastic gradient descent classifier (SGD Classifier), Random Forest, and Lasso. Feature preprocessing and selection were performed on training sets and then applied to test sets. The features were selected based on the estimator assigned importance weights. For the Lasso estimator, features whose absolute importance value is greater or equal to $1e-5$ were kept. Otherwise, “mean” values were used as the feature selection threshold. The feature section process runs separately for each model. For the

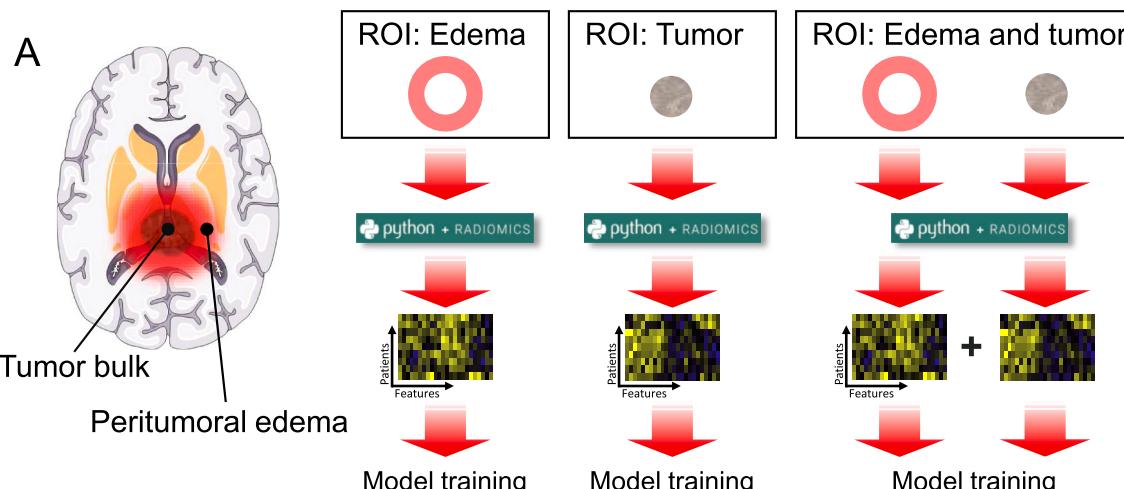


Fig. 2. Diagrams of pineal lesions and the segmentation process. Radiomics features are extracted from the two ROIs separately: tumor and peritumoral edema. Three models are benchmarked with radiomics features from tumors, edema, and both for all combinations of MRI sequences.

classification task, we used ten classifiers: ridge regression, least absolute shrinkage and selection operator (Lasso), Random Forest, Gaussian Naïve Bayes, Support Vector Machine Classifier (SVC), Multilayer Perceptron (MLP), and AutoML based XGBoost, CatBoost, LightGBM, Random Forest, and Extra Trees. All 40 combinations of feature selection algorithms and classifiers were tested, and the model with the highest average AUC on the test sets was then selected.

2.6. Strategy for the construction of input feature matrices

To determine which MRI sequence or sequence combination is most suitable for pineal tumor classification. We sought to exhaust all the combinations of MRI sequences and ROIs. Fifteen combinations for four MRI sequences (T1, T1C, T2, T2 FLAIR) and three combinations for two ROI groups (Tumor, edema) were tested. In total, forty-five (15×3) combinations of radiomics features serve as input for each training session. In addition, radiomics features combined with age and gender were also used to train demographical-radiomics models with the same strategy. Finally, a Random Forest model with only demographic information was trained with a 5-fold CV to provide a performance benchmark.

2.7. Training and evaluation of models

All the model training and evaluation were done in the Python 3.8.0 environment with scikit-learn 0.21.3 [16], AutoML 1.0.7 [17]. Model performance visualization was performed with the R programming language (v 4.1.2). Models were assessed using the average receiver-operating characteristic (ROC) curve with area under the curve (AUC), accuracy, positive predictive value (PPV), negative predictive value (NPV), sensitivity, specificity, and area under the precision-recall curves (AUPRC) of 5-fold, with R package *precrec* (v 0.12.7) [18].

2.8. Statistical analysis

For demographic information and imaging parameters, we used χ^2 and Wilcoxon rank-sum to evaluate differences in categorical as well as continuous variables between two types of lesions, respectively (Table 1). Statistical analysis was performed with the R programming language (v 4.1.2). For ROI groups, MRI sequence combinations, feature selectors and classifiers, mean values of AUCs of radiomics-only models were compared with the Wilcoxon signed-rank test, the Bonferroni adjusted *p* values were calculated. A *p* value < 0.05 was considered statistically significant.

3. Results

3.1. Cohort information and workflow

The most common PRTs were germinomas and mixed GCTs (Fig. 1B). The rest contain NGGCTs, diffused astrocytomas, meningiomas, PPTs, pineocytomas, pilocytic astrocytomas, pineoblastomas, and other tumors. The average age of diagnosis is 22.5 for all patients in our cohort, germinoma patients are slightly younger than the others, but the difference is not significant. Expectedly, germinoma is more common in males than females. For all imaging parameters and scanner manufacturers, there are no significant differences between germinomas and other tumors (Table 1).

Fig. 3 summarizes the workflow of the current study. A total of 122 cases of various pineal lesions are included for further analysis. A single lesion was segmented into two ROIs: tumor bulk and peritumoral edema. Radiomics features were then extracted from these ROIs separately. Feature matrices from two ROIs and four MRI sequences were cross-combined as the input of models. For each combination of ROIs and MRI sequences, 40 models were trained with four feature selection and ten classification algorithms, the model with the highest AUC was

Table 1
Demographic information and imaging parameters.

	Histology		Overall	<i>p.value</i>
	Germinoma	Others		
Age (year)				0.72
Count	42	80	122	Wilcoxon rank-sum
Mean (SD)	17.09 (7.47)	25.34 (20.19)	22.50 (17.33)	
Median (IQR)	15.84 (4.96)	14.25 (32.94)	15.59 (19.73)	
Q1, Q3	13.54, 18.50	9.66, 42.60	11.16, 30.89	
Min, Max	5.84, 45.59	0.67, 67.95	0.67, 67.95	
Missing	0	0	0	
Gender				0.008
Count (%)	42 (34.43%)	80 (65.57%)	122	Chi-square
(Col %)				
Female	3 (7.14%)	24 (30.00%)	27 (22.13%)	
Male	39 (92.86%)	56 (70.00%)	95 (77.87%)	
Missing	0	0	0	
Scanner manufacturer				0.784
Count (%)	42 (34.43%)	80 (65.57%)	122	Chi-square
(Col %)				
Alltech	3 (7.14%)	8 (10.00%)	11 (9.02%)	
GE MEDICAL SYSTEMS	15 (35.71%)	22 (27.50%)	37 (30.33%)	
SIEMENS	17 (40.48%)	37 (46.25%)	54 (44.26%)	
TOSHIBA MEC	7 (16.67%)	13 (16.25%)	20 (16.39%)	
Missing	0	0	0	
Slice thickness (mm)				0.4
Count	42	80	122	Wilcoxon rank-sum
Mean (SD)	4.81 (0.84)	4.59 (1.19)	4.66 (1.08)	
Median (IQR)	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	
Q1, Q3	5.00, 5.00	5.00, 5.00	5.00, 5.00	
Min, Max	1.00, 5.00	1.00, 5.60	1.00, 5.60	
Missing	0	0	0	
Magnetic field strength (T)				0.466
Count	42	80	122	Wilcoxon rank-sum
Mean (SD)	2.11 (0.75)	2.01 (0.71)	2.04 (0.72)	
Median (IQR)	1.50 (1.50)	1.50 (1.50)	1.50 (1.50)	
Q1, Q3	1.50, 3.00	1.50, 3.00	1.50, 3.00	
Min, Max	1.50, 3.00	1.50, 3.00	1.50, 3.00	
Missing	0	0	0	

then reported for evaluation.

3.2. Performance evaluation of radiomics-only models and comparison of the algorithms

As described in the method, radiomics features from two ROIs and four MRI sequences generate 45 combinations of input feature matrices. Four feature selection and ten classification algorithms generate a total of 40 combinations. Fig. 4A provides the mean AUCs of all 1800 models

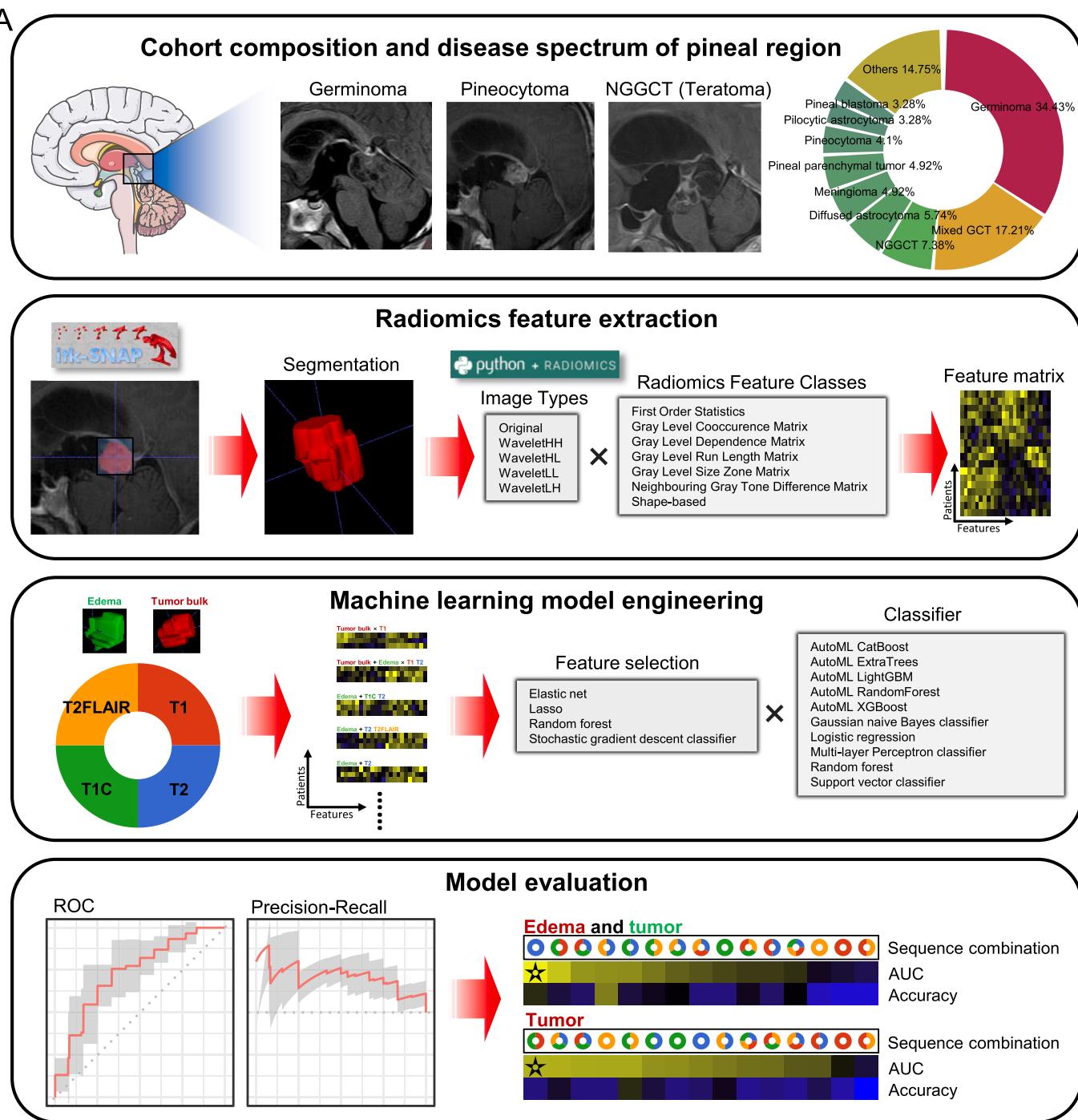


Fig. 3. Workflow of this study. MRI images of 122 patients with lesions of the pineal region were collected from Xiangya hospital. Tumor bulk and peritumoral edema were manually segmented. Radiomics features were extracted for tumor and edema separately. The models were benchmarked using fifteen combinations of MRI sequences, three combinations of ROIs, four feature selection methods and ten classification methods. AUC, accuracy and AUPRC are used for model evaluation.

trained on radiomics features alone, with columns and rows arranged by the mean value. The highest mean AUC for each combination of ROIs and MRI sequences in the heatmap is marked with an asterisk. Interestingly, these “best” models in each row are dominantly constructed with the feature selection methods of Random Forest and SGD classifiers. So, we compared the frequency of the feature selector + classifier combo among all the “best” and “Top 5” models. Random Forest + MLP classifier has overwhelming popularity compared to all other combos (Fig. 5A and B). Then we compared the mean AUCs of models with different feature selectors and classifiers. In the four feature selection methods, models with the Random Forest algorithm as the feature

selection method have significantly better mean AUCs than all other algorithms (Fig. 5C, left). Though logistic regression has the highest mean AUCs for the ten classification methods, it is not significantly better than the others (Fig. 5C, right). All these analyses indicate that the choice of feature selection and classification methods could significantly impact model performance. Detailed statistical comparisons were presented in [Supplementary Table 4](#).

3.3. Comparison of MRI sequences and ROIs

PRTs have distinctly different appearances between T1 and T2 MRI

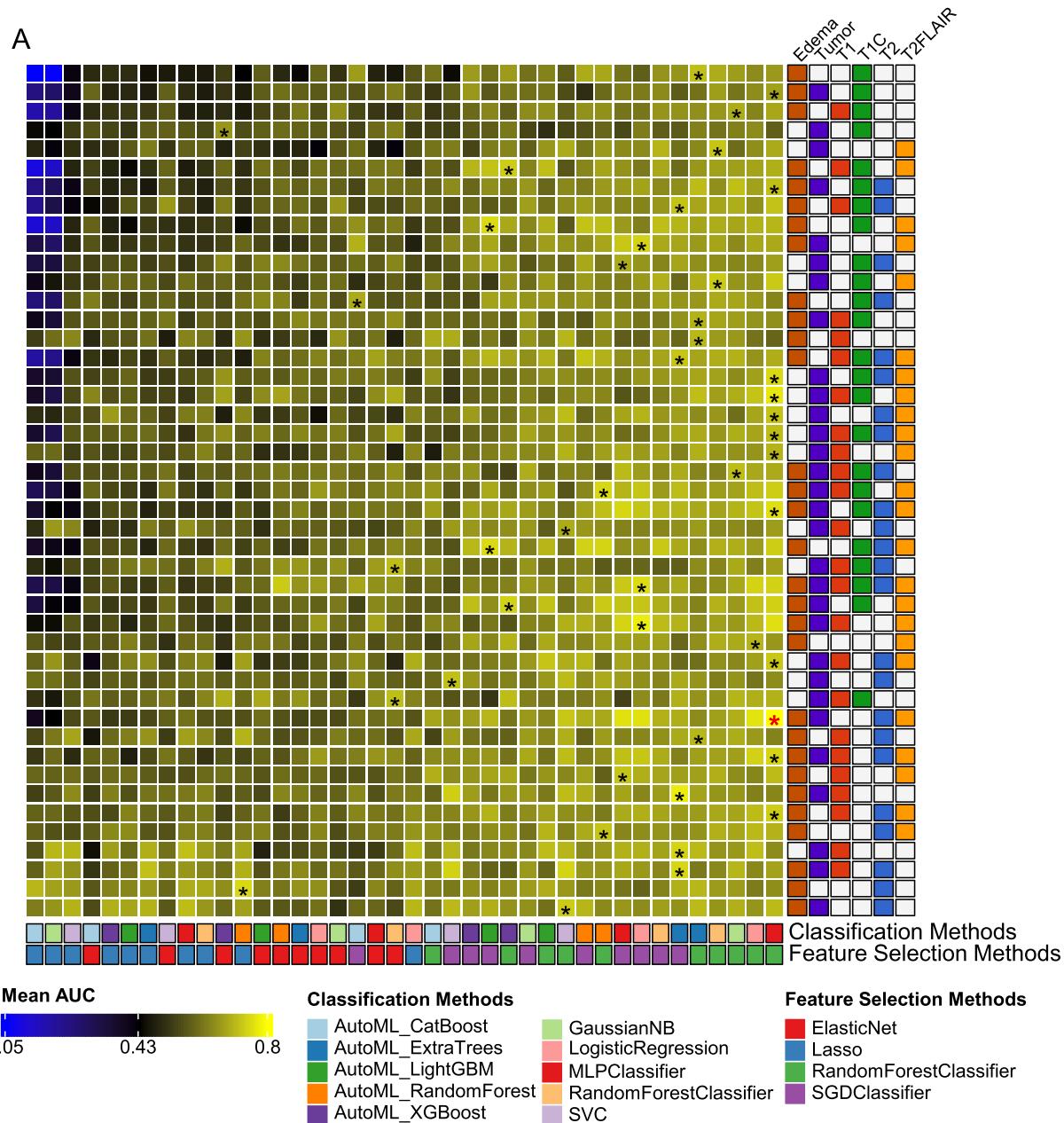


Fig. 4. A, Mean AUC matrix of all models trained only on radiomics features, the heatmap rows represents a specific combination of ROIs and MRI sequences where the radiomics features are extracted from, the heatmap columns depict a specific combination of classification and feature selection methods. The mean value of the rows/columns is used to arrange the heatmap. The black asterisks denote the highest AUC of each row, and the red asterisk denotes the highest AUC in this graph. Sequences: T1, T1-weighted scan. T1C, T1-weighted contrast-enhanced scan. T2, T2-weighted scan. T2F, T2 FLAIR. Classification methods: GaussianNB, Gaussian Naive Bayes classifier. MLPClassifier, Multi-layer Perceptron classifier. SVC, Support Vector Classifier. Feature selection methods: SGDClassifier, Stochastic Gradient Descent (SGD) classifier.

images. So, we compared the performance of radiomics-only models with or without a specific MRI sequence (Fig. 6A). Surprisingly, not all MRI sequences have a positive effect. For example, models with T1C images have significantly inferior performance than their counterparts (p -value: <0.0001). While T1 and T2 MRI images significantly increased the performance of models (T1 p -value: 0.015, T2 p -value: <0.0001). The same comparison was also performed for ROIs (Fig. 6B), the performance of models with the three ROI groups showed no significant difference.

Among all the radiomics-only models, features from Edema-Tumor (ROI) + T2-T2 FLAIR (MRI sequence) are the best for this classification task, according to the AUC and AUPRC values of all the radiomics-only models. This model reached the highest mean AUC of 0.80 (CI95%:

0.74–0.86) (Fig. 6C). Detailed ROC, PRC plots, AUC, AUPRC, and ACC results were presented in [Supplementary Figs. 1A and B](#), [Supplementary Figs. 2A and B](#), [Supplementary Figs. 3A and B](#), and [Supplementary Table 5](#).

3.4. Simple demographic information significantly improves model performance

Demographic information, including age and gender, was collected when a brain MRI was prescribed. As germinoma is more common in males than females, this additional information might improve the classification model's performance when provided with radiomics features. Therefore, we retrained all the previous models with demographic

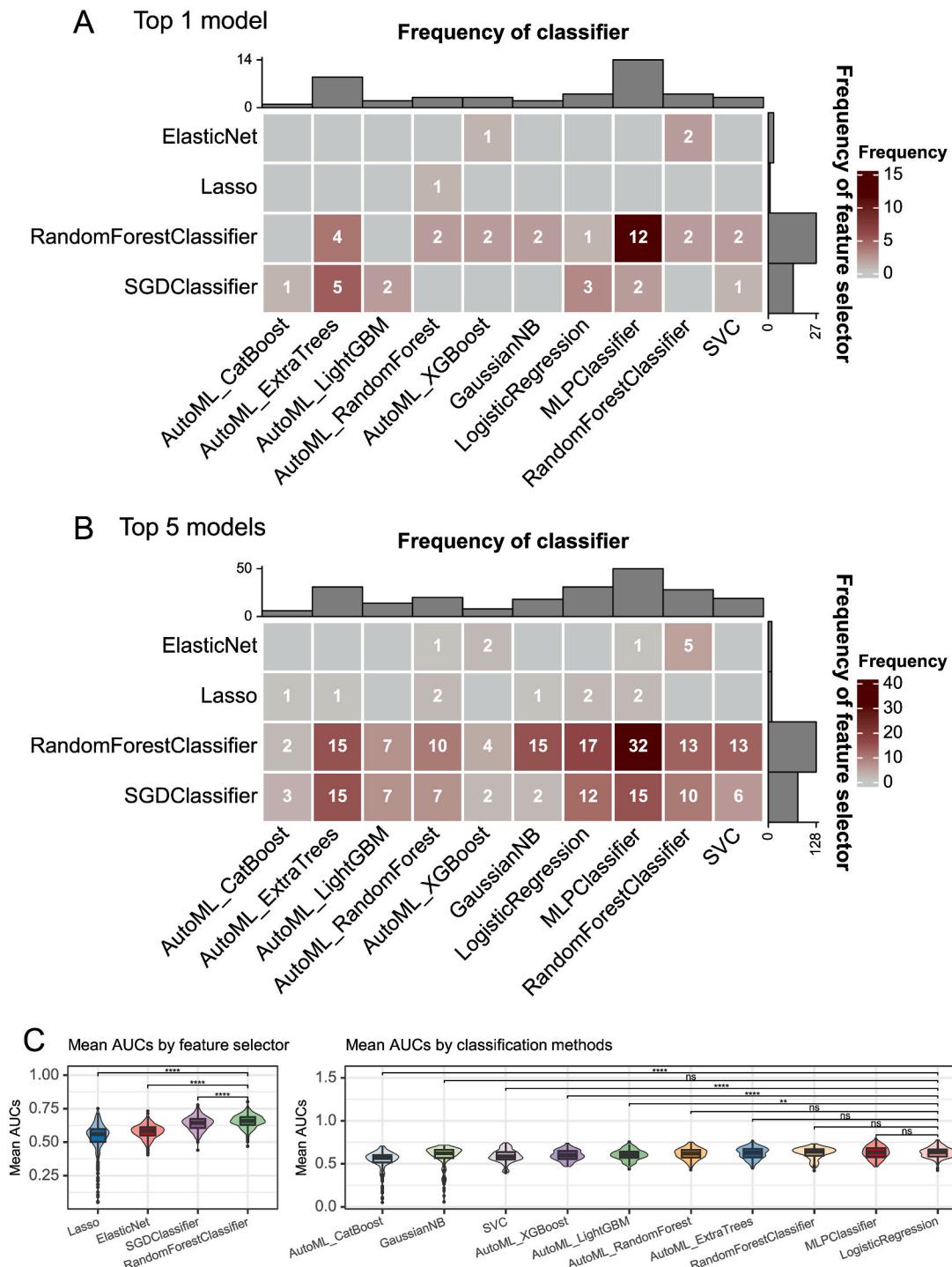


Fig. 5. A, the frequency of classification and feature selection methods of all top 1 model. The definition of the top 1 model: for each combination of ROI and MRI sequences, we trained 40 models. The one with the highest mean AUC was reported as the top 1. B, the frequency of classification and feature selection methods of all top 5 models. The definition of the top 5 model: for each combination of ROI and MRI sequences, we trained 40 models. The top 5 models with the highest mean AUC were reported. C, the mean AUCs of radiomics-only models grouped by feature selection and classification methods. Mean values were compared with the Wilcoxon signed-rank test, Bonferroni adjusted p values were calculated, **** adjusted $p < 0.0001$, *** adjusted $p < 0.001$, ** adjusted $p < 0.01$, * adjusted $p < 0.05$, ns not significant.

information + radiomics features. In addition, a model with only demographic information was also trained to provide a performance benchmark (Fig. 6D), which reached a mean AUC of 0.82 ± 0.04 (CI95%: 0.78–0.86 mean \pm SD), and a mean AUPRC of 0.73.

Demographics-radiomics (R + D) models significantly outperform their radiomics-only (R) counterparts with the same ROI and MRI sequence as we evaluate the mean AUC and accuracy values (Fig. 7A and

B). Among all the demographics-radiomics models, the one trained on features from Tumor (ROI) + T1-T1C-T2 (MRI sequence) reached the highest mean AUC of 0.88 (CI95%: 0.81–0.96) (Fig. 7C). Detailed ROC, PRC plots, AUC, AUPRC, and ACC results were presented in Supplementary Figs. 4A and B, Supplementary Figs. 5A and B, Supplementary Figs. 6A and B, and Supplementary Table 6.

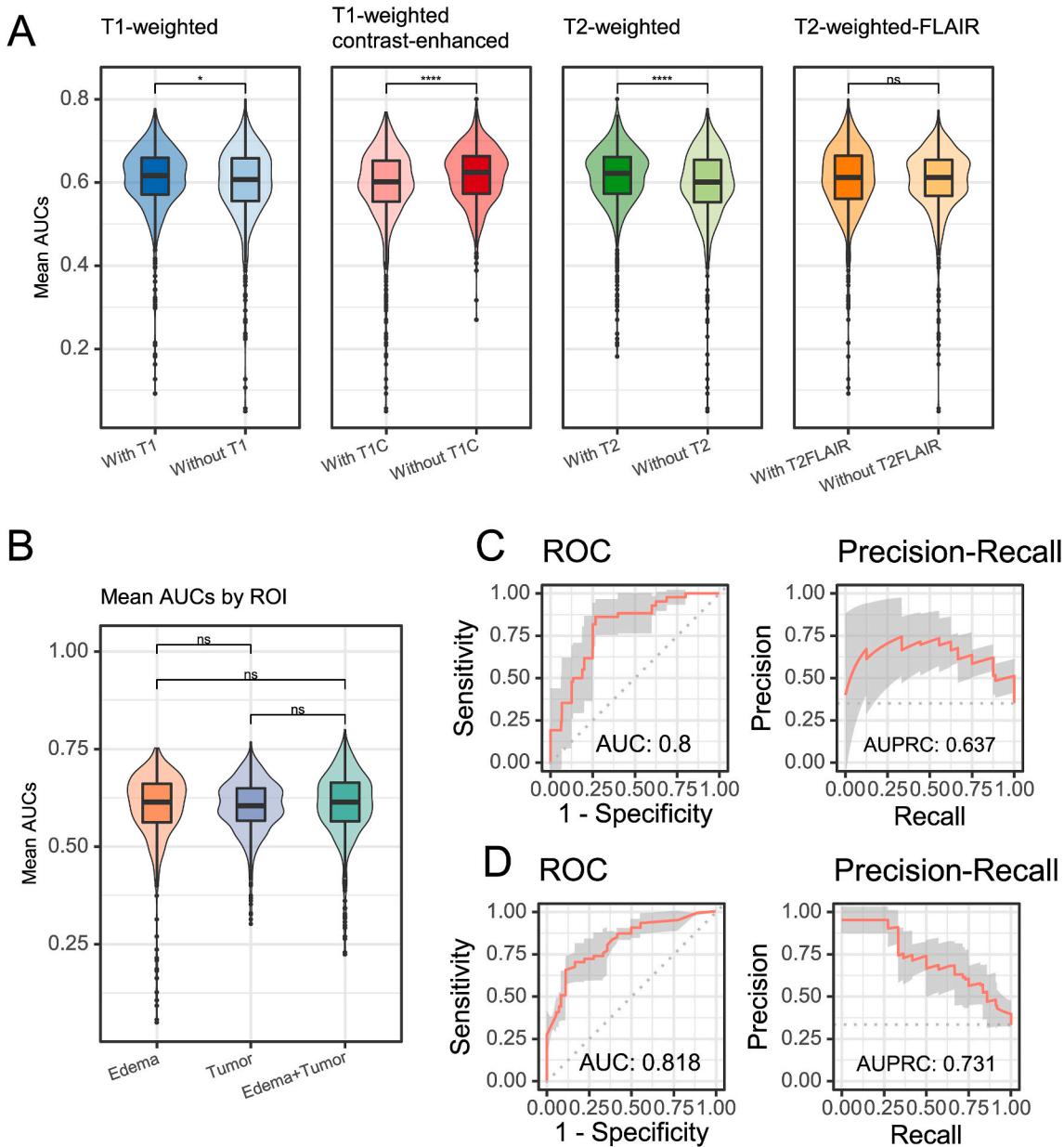


Fig. 6. A, comparison of AUCs with or without a specific MRI sequence. Mean values were compared with the Wilcoxon signed-rank test, **** $p < 0.0001$, *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, ns not significant. B, comparison of AUCs between the three ROI groups: edema, tumor, edema + tumor. C, the ROC, and precision-recall curve of the radiomics-only model with the highest AUC. D, the model's ROC, and precision-recall curve trained only on demographic data.

3.5. Radiomics feature classes preference of models

Radiomics features consist of seven classes: First Order Statistics, Shape-based features, Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), Neighbouring Gray Tone Difference Matrix (NGTDM), Gray Level Dependence Matrix (GLDM). These classes represent different aspects of medical imaging. Features that pass the feature selection algorithms are potentially more critical for this classification task. Therefore, the contribution of each feature class could help us understand how the models work.

For all 45 combinations of ROIs and MRI sequences, we compared the radiomics-only models with the highest mean AUC. First-order statistics were the dominant feature class for almost all models, while Shape-based features were the least used (Fig. 8A, B, C). This indicates that germinoma has a similar shape to other pineal lesions.

4. Discussion

Most previous studies on PRTs are case reports or literature reviews. To our best knowledge, this study is the first systemic benchmark to test all the possible sequence combinations of pre-operative MRI in developing a discriminator for germinoma of the pineal region. In the current study, we reported one of the largest unselected natural cohorts of 122 PRTs, including 42 germinomas, 21 mixed GCTs, 9 NGGCTs, 7 diffused astrocytomas, 6 meningiomas, 6 PPTs, 5 pineocytomas, 4 pilocytic astrocytomas, 4 pineoblastomas and 18 others lesions. Previous quantitative MRI analyses have demonstrated that some image characters are significantly different in germinomas than other PRTs [19], proving that it might be possible to utilize machine learning in automatically performing differential diagnosis on routine MRI images. With radiomics features only, our model reached an AUC of 0.8 AUC. With demographics information, our model reached an AUC of 0.82 AUC, and the combined demographics-radiomics model reached an AUC of 0.88.

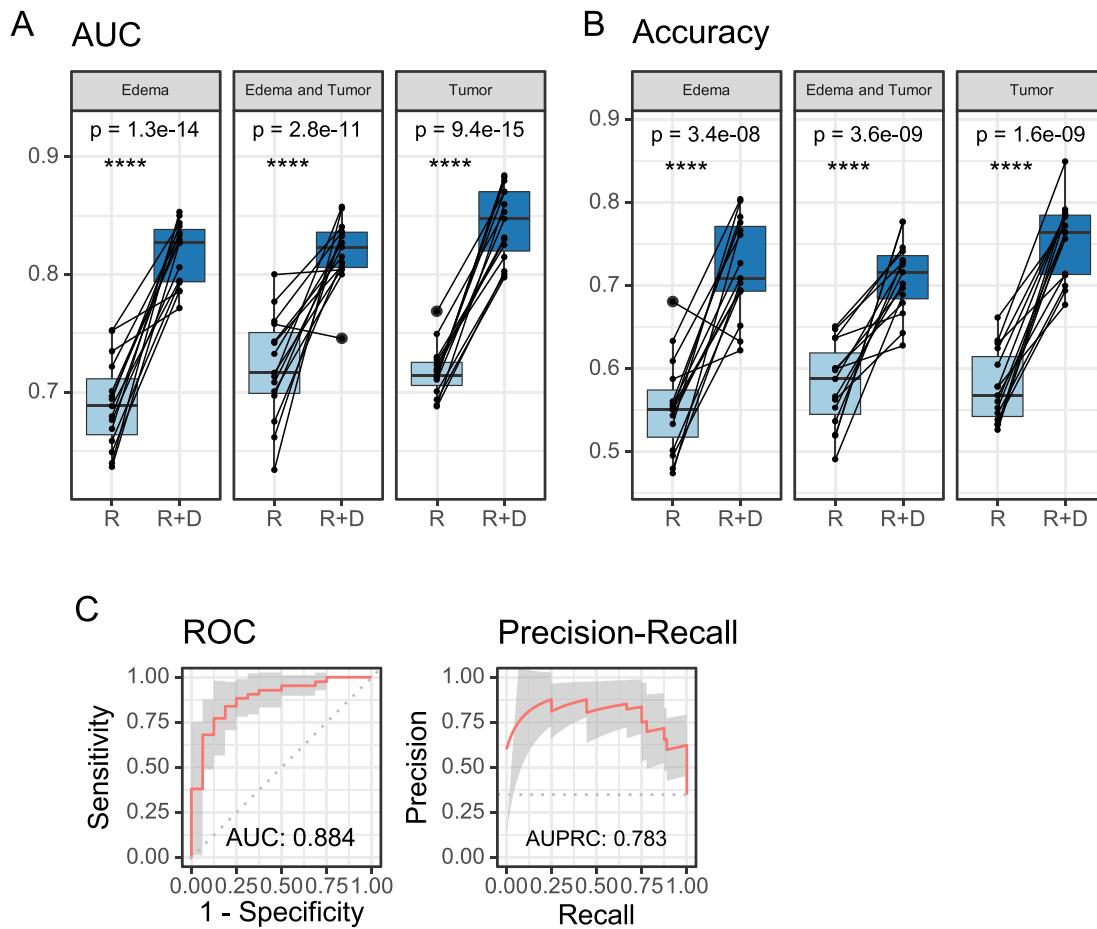


Fig. 7. A, comparison of AUCs between the two groups: the model trained on radiomics features alone (R), the model trained with radiomics features and demographics (R + D). B, comparison of accuracy between the two groups: the model trained with only radiomics features (R), the model trained with radiomics features and demographics (R + D). Mean values were compared with the paired *t*-test. The model with the same input MRI sequences are treated as a pair and connected by a line in the graph, **** $p < 0.0001$. C, the ROC, and precision-recall curve of the demographics-radiomics model with the highest AUC.

Yanghua F. et al. attempted to differentiate germinoma and pinealoblastoma with clinic-radiomics-based machine learning [11]. Their study cohort included 134 patients diagnosed with pineal region germinoma or pinealoblastoma from 2008 to 2020 in Tiantan hospital (germinoma: 69; pinealoblastoma: 65). 1562 radiomics features were extracted from T2 and T1C MRI images. Clinical information such as age, gender, hydrocephalus, AFP, CEA, and β -hCG was collected. The Elastic Net and Support Vector Machine was used as feature selection and classification algorithms. The AUCs of radiomics-only and clinical-radiomics models reached AUCs of 0.88 and 0.94, respectively, which is higher than that of the current study. However, epidemiological reports of PRTs showed that germinoma has a ~ 5-fold incidence rate over pinealoblastoma in the pineal region [20,21]. The performance of models trained with a selected cohort that contains only two diseases might decrease substantially on heterogeneous real-life datasets, especially for PRTs. The demographics-radiomics model in this study does not require any elaborate or time-consuming lab test. At the same time, demographic information like age or gender can be easily acquired from patients before the prescription of a brain MRI. These make it easier for our model to translate into clinical practice. Germinomas of the pineal region have a higher incidence rate in males [2,22]. Clinical information improved the performance of radiomics-only models both in our study and Yanghua F. et al. [11]. Other information, like initial symptoms, might further improve model performance.

Our analyses showed that models with T2 images have higher performance than their counterparts. On the contrary, T1C images hurt the

performance of models, which could provide a valuable reference for future implementation of machine learning based on MRI images. Compared with other PRTs, germinomas are less frequently cystic (germinoma vs others: 27.3% vs 64%) and more frequently surrounded by thick peritumoral edema (germinoma vs others: 40.6% vs 15.8%), as reported by Ryuji Awa et al. [23]. Cystic lesions and edema have distinctly high intensity in the T2 image, which might explain why T2 images improved the performance for radiomics-only models, and a non-cystic lesion with heavy peritumoral edema might favor the diagnoses of germinoma of pineal region. The features from T1C images hurt model performance since enhancement is not a distinctive feature for germinomas, as most PRTs are heterogeneously enhanced [19, 24–27].

Hundreds to thousands of radiomics features can be extracted from single-modal medical images. The number of features is so large that it often outnumbers the patients [28]. Moreover, radiomics features generally have low information entropy, and many might be redundant [29,30]. An excessed number of features challenges computational resources and causes overfitting. In our study, we demonstrated that feature selection and classification algorithms have non-negligible impacts on model performance. Random Forest was the best feature selection method for this classification task, and it is one of the most popular feature selection algorithms based on assembled decision trees. Still, it might not be the same for other similar tasks. As for the selection of a classification algorithm, many previous researchers built two-category classifiers with a single classification algorithm like

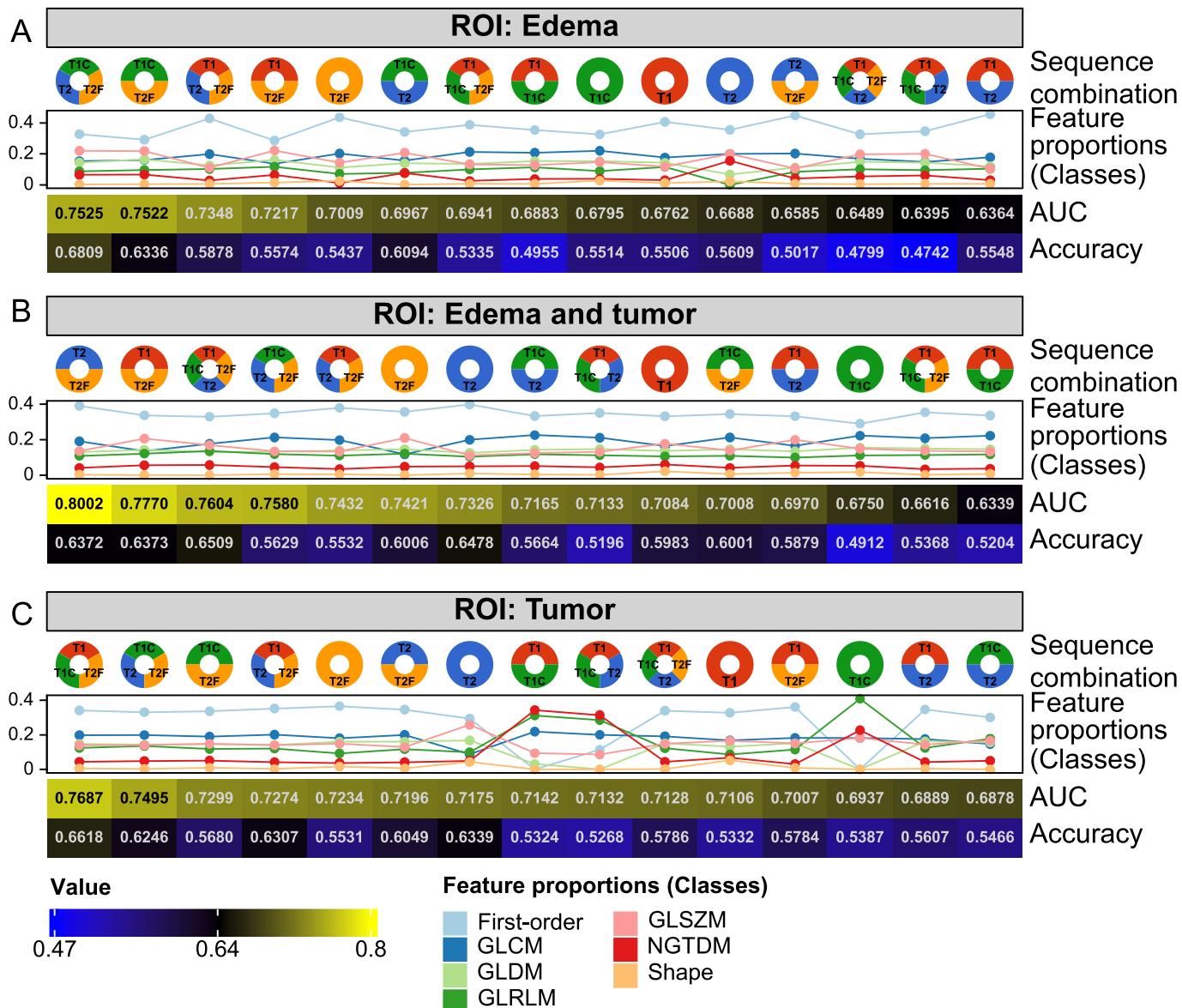


Fig. 8. For models trained on radiomics features alone, with the previously described three groups of ROIs: peritumoral edema alone(A), peritumoral edema combined with tumor bulk (B), and tumor bulk alone (C), the proportion of features from seven feature groups are compared. Sequence combination: T1, T1-weighted scan. T1C, T1-weighted contrast-enhanced scan. T2, T2-weighted scan. T2F, T2 FLAIR. Feature proportions (Classes): First-order, First Order statistics. GLCM, Gray Level Cooccurrence Matrix. GLDM, Gray Level Dependence Matrix. GLRLM, Gray Level Run Length Matrix. GLSVM, Gray Level Size Zone Matrix. NGTDM, Neighbouring Gray Tone Difference Matrix. Shape, Shape-based features.

logistic regression and Random Forest on radiomics features [11,31,32], but there is no pre-existing consensus on how and why the specific algorithm is picked. In this work, we found no particular outstanding classification algorithm. It is worth noting that certain combinations of feature selection and classification algorithms (like Random Forest + Logistic Regression, Random Forest + MLP) showed stably superior performance. In future investigations, these combinations can be prioritized to save computing resources.

Radiomics features consist of seven classes that represent different aspect of the texture information and shape information. An imbalanced proportion of feature classes was observed in this study. For radiomics-only models, first-order features seem to be the most significant category. First-order radiomics features are simple statistics of the input images, including entropy, energy, total energy, minimal value, maximum value, mean value, median value, 10th percentile, 90th percentile, interquartile range, range, mean absolute deviation, robust mean absolute deviation, root mean squared, standard deviation,

skewness, kurtosis, variance, uniformity. These features describe the global distribution of voxel intensity values over the ROI, which means that the difference between germinomas and other lesions in the pineal region lies in global features, while first-order features are mostly unidentifiable by human eyes. Another interesting finding in this study is that shape-based radiomics features are the least important for this classification task. Shape-based features are descriptors for both size and shape of the segmented ROI, including features like the length of the major/minor axis, sphericity and mesh surface area. Though shape-based features can be eyeballed or measured in reading MRI images, they might not help identify germinomas, according to our findings, which is also consistent with the report by Ryuji Awa et al. [23].

There're some limitations of this study. First, the cohort used for training and validation of the model comes from a single medical center. Multi-center data could further validate the robustness and generalization ability of our model. Second, this is a retrospective study, thus requiring a prospective cohort to validate. Finally, with a future

prospective cohort, we could explore whether additional lab tests can further improve model performance.

5. Conclusion

A non-invasive system that helps identify germinoma in the pineal region could reduce lab exams and traumatic therapies. Through the comprehensive evaluation of possible sequence combinations in the differential diagnosis of pineal tumor, T1 and T2 emerged as the most informative sequences for the task. The demographical-radiomics model can accurately and efficiently identify germinomas in the pineal region. In addition, the models trained with unselected real-life datasets might be better for clinical translation in the future and aid decision-making toward a more personalized therapy.

Author contributions

XL, designed and supervised this study. NY, QY, ZC performed the analysis. NY, QY wrote the manuscript. NY, PL finished the segmentation of all MRI images. QY, NY contributed in acquisition of the data used in this study. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China [grant numbers 81770781, 81472594]. Special funds for innovation in Hunan Province [grant number 2020SK2062]. High talent project of Hunan Province [grant number 2022WZ1031].

Declaration of competing interest

None Declared.

Acknowledgements

Fig. 2 was partly generated using Servier Medical Art, provided by Servier, licensed under a Creative Commons Attribution 3.0 unported license.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.combiomed.2022.106366>.

References

- [1] Q.T. Ostrom, et al., CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2014–2018, *Neuro Oncol.* 23 (12 Suppl 2) (2021) iii1–iii105.
- [2] K. Nomura, Epidemiology of germ cell tumors in Asia of pineal region tumor, *J. Neuro Oncol.* 54 (3) (2001) 211–217.
- [3] W. Li, et al., Gamma knife radiosurgery (GKRS) for pineal region tumors: a study of 147 cases, *World J. Surg. Oncol.* 13 (2015) 304.
- [4] A.N. Konovalov, D.I. Pitskhelauri, Principles of treatment of the pineal region tumors, *Surg. Neurol.* 59 (4) (2003) 250–268.
- [5] D. Frappaz, et al., EANO, SNO and Euracan consensus review on the current management and future development of intracranial germ cell tumors in adolescents and young adults, *Neuro Oncol.* 24 (4) (2022) 516–527.
- [6] M.A. Zaazoue, L.C. Goumnerova, Pineal region tumors: a simplified management scheme, *Childs Nerv Syst.* 32 (11) (2016) 2041–2045.
- [7] B.J. Erickson, et al., Machine learning for medical imaging, *Radiographics* 37 (2) (2017) 505–515.
- [8] A. Hosny, et al., Artificial intelligence in radiology, *Nat. Rev. Cancer* 18 (8) (2018) 500–510.
- [9] S.P. Singh, et al., 3D deep Learning on medical images: a review, *Sensors* 20 (18) (2020).
- [10] D. Shen, G. Wu, H.I. Suk, Deep learning in medical image analysis, *Annu. Rev. Biomed. Eng.* 19 (2017) 221–248.
- [11] Y. Fan, et al., Non-invasive preoperative imaging differential diagnosis of pineal region tumor: a novel developed and validated multiparametric MRI-based clinicoradiomic model, *Radiother. Oncol.* 167 (2022) 277–284.
- [12] N. Ye, et al., Classification of gliomas and germinomas of the basal ganglia by transfer learning, *Front. Oncol.* 12 (2022) 844197.
- [13] B. Chen, et al., Differentiation between germinoma and craniopharyngioma using radiomics-based machine learning, *J. Personalized Med.* 12 (1) (2022).
- [14] P.A. Yushkevich, et al., User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability, *Neuroimage* 31 (3) (2006) 1116–1128.
- [15] J.J.M. van Griethuysen, et al., Computational radiomics system to decode the radiographic phenotype, *Cancer Res.* 77 (21) (2017), e104–e107.
- [16] F. Pedregosa, et al., Scikit-learn: machine learning in Python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [17] I. Guyon, L. Sun-Hosoya, Boull, Analysis of the AutoML Challenge series 2015–2018, 2017, pp. 177–219, https://link.springer.com/chapter/10.1007/978-3-030-05318-5_10#chapter-info.
- [18] T. Saito, M. Rehmsmeier, The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets, *PLoS One* 10 (3) (2015), e0118432.
- [19] N. Dumrongpisutikul, J. Intrapromkul, D.M. Yousem, Distinguishing between germinomas and pineal cell tumors on MR imaging, *AJNR Am J Neuroradiol* 33 (3) (2012) 550–555.
- [20] C. Mottolese, A. Szathmary, P.A. Beuriat, Incidence of pineal tumours, A review of the literature, *Neurochirurgie* 61 (2–3) (2015) 65–69.
- [21] J. Hirato, Y. Nakazato, Pathology of pineal region tumors, *J. Neuro Oncol.* 54 (3) (2001) 239–249.
- [22] V. Cuccia, M. Galarza, Pure pineal germinomas: analysis of gender incidence, *Acta Neurochir.* 148 (8) (2006) 865–871, discussion 871.
- [23] R. Awa, et al., Neuroimaging diagnosis of pineal region tumors—quest for pathognomonic finding of germinoma, *Neuroradiology* 56 (7) (2014) 525–534.
- [24] A.G. Solomou, Magnetic resonance imaging of pineal tumors and drop metastases: a review approach, *Rare Tumors* 9 (3) (2017) 6715.
- [25] S. Komakula, et al., Pineal parenchymal tumor of intermediate differentiation: imaging spectrum of an unusual tumor in 11 cases, *Neuroradiology* 53 (8) (2011) 577–584.
- [26] F. Reis, et al., Neuroimaging in pineal tumors, *J. Neuroimaging* 16 (1) (2006) 52–58.
- [27] S. Dahiya, A. Perry, Pineal tumors, *Adv. Anat. Pathol.* 17 (6) (2010) 419–427.
- [28] J.E. Park, et al., Reproducibility and generalizability in radiomics modeling: possible strategies in radiologic and statistical perspectives, *Korean J. Radiol.* 20 (7) (2019) 1124–1137.
- [29] Y. Balagurunathan, et al., Reproducibility and prognosis of quantitative features extracted from CT images, *Transl. Oncol.* 7 (1) (2014) 72–87.
- [30] R. Berenguer, et al., Radiomics of CT features may be nonreproducible and redundant: influence of CT acquisition parameters, *Radiology* 288 (2) (2018) 407–415.
- [31] G. Li, et al., An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas, *Brain* 145 (3) (2022) 1151–1161.
- [32] H.C. Kniep, et al., Radiomics of brain MRI: utility in prediction of metastatic tumor type, *Radiology* 290 (2) (2019) 479–487.