BMC Medical Imaging

RESEARCH

Open Access

Machine learning decision support model construction for craniotomy approach of pineal region tumors based on MRI images

Ziyan Chen[1,2,3], Yinhua Chen[1,2,3], Yandong Su[4], Nian Jiang[1,2,3], Siyi Wanggou[1,2,3] and Xuejun Li[1,2,3,5*]

Abstract

Background Pineal region tumors (PRTs) are rare but deep-seated brain tumors, and complete surgical resection is crucial for effective tumor treatment. The choice of surgical approach is often challenging due to the low incidence and deep location. This study aims to combine machine learning and deep learning algorithms with pre-operative MRI images to build a model for PRTs surgical approaches recommendation, striving to model clinical experience for practical reference and education.

Methods This study was a retrospective study which enrolled a total of 173 patients diagnosed with PRTs radiologically from our hospital. Three traditional surgical approaches of were recorded for prediction label. Clinical and VASARI related radiological information were selected for machine learning prediction model construction. And MRI images from axial, sagittal and coronal views of orientation were also used for deep learning craniotomy approach prediction model establishment and evaluation.

Results 5 machine learning methods were applied to construct the predictive classifiers with the clinical and VASARI features and all methods could achieve area under the ROC (Receiver operating characteristic) curve (AUC) values over than 0.7. And also, 3 deep learning algorithms (ResNet-50, EfficientNetV2-m and ViT) were applied based on MRI images from different orientations. EfficientNetV2-m achieved the highest AUC value of 0.89, demonstrating a significant high performance of prediction. And class activation mapping was used to reveal that the tumor itself and its surrounding relations are crucial areas for model decision-making.

Conclusion In our study, we used machine learning and deep learning to construct surgical approach recommendation models. Deep learning could achieve high performance of prediction and provide efficient and personalized decision support tools for PRTs surgical approach.

Clinical trial number Not applicable.

Keywords Pineal region tumors, Craniotomy approaches, Magnetic resonance imaging, Machine learning, Deep learning

*Correspondence:
Xuejun Li
lxjneuro@csu.edu.cn
[1]Department of Neurosurgery, Xiangya Hospital, Central South University, No.87 Xiangya Road, Changsha, Hunan, P. R. China
[2]National Clinical Research Center for Geriatric Disorders, Xiangya Hospital, Central South University, Changsha, Hunan, P. R. China
[3]Xiangya Hospital, Hunan International Scientific and Technological Cooperation Base of Brain Tumor Research, Central South University, Changsha, Hunan, P. R. China
[4]Department of Neurosurgery, Changhai Hospital, Naval Medical University, Shanghai, P. R. China
[5]Xiangya School of Medicine, Central South University, Changsha, Hunan, P. R. China

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Introduction

Pineal region tumors (PRTs) are a type of tumors rare, and deeply located in the brain including various pathological classification such as germ cell tumors, parenchymal tumors, gliomas, brain metastases, etc. Common clinical manifestations of PRTs have headache, obstructive hydrocephalus, with some endocrinological symptoms [1]. In the treatment of PRTs, surgical procedure is the preferred treatment for most types of tumors in this area except for germinomas and lymphomas which are highly sensitive to radiotherapy [2, 3]. With the improvement of precision preoperative imaging and surgical technique, these deeply-seated tumors could be approached with various surgical routes and strategies to achieve maximal resection while preserving surrounding neural structures and blood vessels, ultimately improving patient prognosis, survival time, and quality of life.

Currently, there are still no standard choice for PRTs surgical approach, but fundamentally, they can be classified as supratentorial or infratentorial [4]. The most used approaches include infratentorial supracerebellar approach (ISA), poppen approach (PA) (also known as occipital transtentorial approach, OTA) and interhemispheric transcallosal approach (ITA) [5–7] (Figure S1). Pre-operative magnetic resonance imaging (MRI) could help to identify the relationships between lesions and surrounding tissues [8]. Selecting the appropriate approach is essential for completely tumor resection in clinical practice, the decision, however, usually differs for different surgeons which is influenced by various factors including tumor volume, location, growth direction and the relationship between the tumor and nearby blood vessels, as well as neural structures like the oculomotor nerve. Additionally, it can also be influenced by the clinical experience, traditions, and education passed down within specific medical institutions. Previous studies, such as those by Mottolese [9, 10] et al., have summarized experience with different approaches for PRTs. However, these studies primarily focused on the procedural aspects, with the rationale for choosing specific approaches often being qualitative and not clearly articulated. To address this gap, we are exploring methods to quantify the "experience of approach choice" for future reference. This has significant implications for surgical decision-making and medical education.

Recent advances in artificial intelligence (AI) applied for MRI have greatly enhanced lesion detection, segmentation, characterization and even prognosis detection [11]. The state-of-art algorithms pretrained on large-scale datasets could be effectively transferred to small datasets, delivering excellent predictive performance. And explainable methods have been developed to explore the "black box" mechanisms of these models, making it easier for clinical interpretation and application [12, 13].

Therefore, in this study, we retrospectively collected PRTs patients in our single center, and plan to establish surgical approach decision machine learning and deep learning models by utilizing pre-operative MRI images. To achieve this goal, we firstly extracted clinical and radiological features and built predictive models using various machine learning algorithms. Subsequently, we employed deep learning models incorporating different mechanisms, including convolutional neural networks (CNNs) and transformer networks, trained directly on MRI images. Although there is no universally accepted "gold standard" for surgical approach selection, our study provides a quantitative representation of clinical decision-making through machine learning techniques.

Methods

Data collection

This retrospective study collected clinical and radiological data of all patients diagnosed in pineal regions from January 2012 to December 2022 in Xiangya hospital. And the project was approved by the Ethics Committee of our hospital (No.202202033) and all procedures followed the Declaration of Helsinki. Finally, a total of 173 patients were selected in this study. All surgical patients signed informed consent forms.

Patients were enrolled followed the inclusion criteria: (1) All patients went through the MRI scan including T1WI (T1-weight imaging), T2WI and T1C(T1-contrast) in our hospital and confirmed the pineal regions tumor by radiologists. (2) Patients who underwent neurosurgical procedures at our hospital were followed up for at least one year. None experienced serious complications such as cerebral edema, stroke, hemorrhage, or recurrence. (3) Tumors were gross total or sub-total resected after post-operative MRI scanning. (4) Availability of MRI sequence on the Picture archiving and communication system (PACS). And the exclusion criteria include: (1) MRI data with deformation, noise, or poor quality. (2) Patients with incomplete clinical and MRI data. (3) Severe complications, that is cerebral edema or cerebral infarction resulting from significant nerve or blood vessel damage following surgery, secondary procedures, or tumor resection was deemed suboptimal after evaluation by two experienced surgeons. (4) The surgical approach was not one of ISA, PA and ITA. The collection workflow was shown in Fig. 1.

Clinical and MRI information preprocessing

Through looking up surgical records of these patients, three approaches were labeled as outcome measures and categorized as variables for subsequent analysis. Gender and age were used as clinical data for machine learning.

For MRI imaging preprocessing, raw data were downloaded and converted into NIFTI format (Supplemental

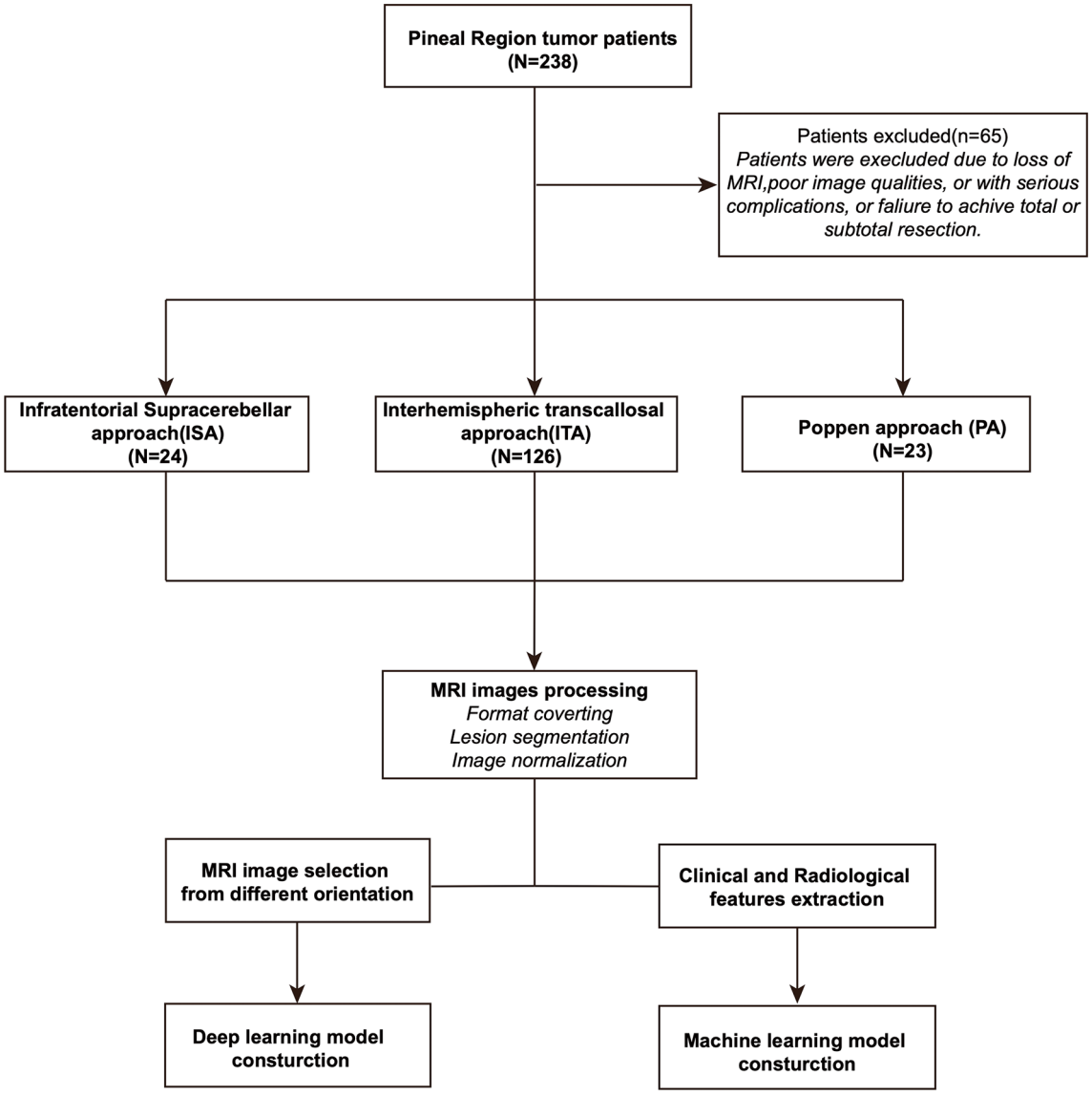

Fig. 1 The flow chart of this project illustrating the data collection and process

information). ITK-SNAP [14] (v4.0.0) was applied for tumor region of interest (ROI) segmentation by 2 neurosurgeons with 10 years of experience in neuro-oncology. The overlapping was used to calculated the volume of tumors. It was supervised and guided by a neurosurgeon with 20 years of experiencs. To extract the radiological features from MRI images, some of VASARI (Visually AcceSAble Rembrandt Images) features were used including enhancement quality (f4), proportion enhancing (f5), cysts (f8), multifocal or multicentric (f9), thickness of enhancing margin (f11), definition of the enhancing margin (f12), definition of the non-enhancing margin (f13), proportion of edema (f14), and hemorrhage (f16). Details of these features could be referred at https://radiopaedia.org/articles/45816. Additionally, the spatial relationship between tumor and tentorium cerebelli, the angles of tentorium cerebelli, as well as hydrocephalus status without preop ventriculostomy were counted as the input features for machine learning. Totally, 15 clinical and radiological features were selected.

To obtain MRI images for deep learning input, we selected interface of the lesions from T1C sequence, including axial, sagittal and coronal views, with no more than 5 images per orientation. No any additional preprocessing steps were applied because clinical observations of MRI images also come from different machines, which would mirror real-world clinical scenarios. Finally, 1134, 229 and 241 images were selected for ITA, ISA and PA, respectively. As for the imbalanced data between ITA and others, random choice strategy was implemented for ITA images to keep the data balance and avoid overfitting for deep learning model.

Machine learning model construction

5 Machine learning algorithms were used for including Logistic Regression (LR), Random Forest (RF), AdaBoost Classifier (AdaBoost), XGBoost Classifier (XGBoost) and Gaussian Processor (GP). Train validation datasets were divided by 7:3 ratio. Grid search for hyper-parameters tunning, and stratified 5-fold training strategy were applied. For model evaluation, ROC (Receiver Operating Characteristic) curve and DET (Detection error tradeoff) curve were used, as well as AUC (Area under the ROC curve) value, accuracy, f1-score, precision, recall and confusion matrix (Fig. 2A).

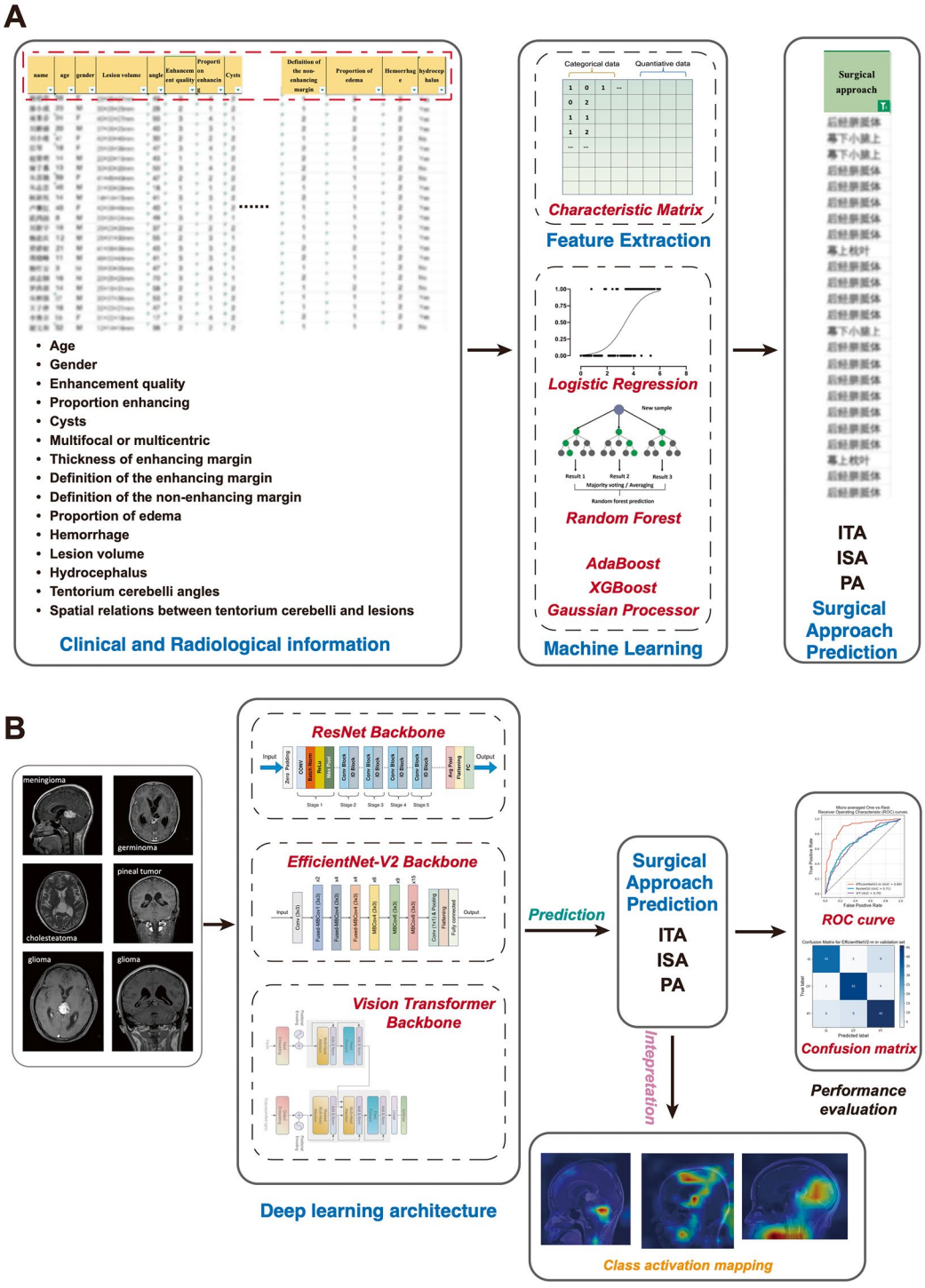

Fig. 2 (**A**) The pipeline explaining the machine learning prediction model construction. (**B**) The pipeline explaining the deep learning prediction model construction

Machine learning was deployed on scikit-learn (v1.3.1), xgboost (v1.6.1) based on Python (v3.8.10).

Deep learning model construction

3 deep learning architectures ResNet-50 [15], EfficientNetV2-m [16], and vision transformer (ViT) [17] were used in our study. Among them, ResNet-50 and EffiicientNetV2-m were based on CNN, while ViT was inspired by attention mechanism which serves as the foundational concept of large-scale models. And we also used transfer learning strategies. To address class imbalance and enhance model robustness, we applied data augmentation techniques during training. These included image flips, random rotations, random affine, Gaussian blurring, and random sharpness adjustments through torchvision (v0.11.0). 80% data were used as training and the rest left for validation. To compare the difference of input images, we also separately used sagittal and coronal images as input evaluate the final prediction performance. Evaluation was the same as machine learning model, and for a better visualization and interpretation, gradient class activation mapping (Grad-CAM) [18] was used for explainable model. *CrossEntropyLoss* was set as loss function and *AdamW* was set as optimization function. Grid search for hyper-parameters (learning rate, weight decay, batch size) tunning was also applied during model training (Fig. 2B). Deep learning was deployed on pytorch (v1.10.1), CUDA (v10.1.24) based on Python with a Nividia Tesla T4 GPU.

Statistical analysis and visualization

For quantitative data, a normality test was firstly conducted. Parametric tests were used if the data followed a normal distribution and if not, non-parametric tests were employed. In parametric tests, ANOVA or Welch's ANOVA was used according to the homogeneity of variances. For non-parametric testing, we use the Kruskal-Wallis test, and post-hoc pairwise comparisons were performed by Bonferroni correction. Chi-squared (χ^2) contingency table analysis was conducted for categorical data. $P<0.05$ was considered as statistical significance. All statistical analysis and visualization were performed on Python and related packages.

Results

Demographic information

Table 1 showed the demographic and radiological information of all 173 PRTs patients. We observed the significant differences in gender ($P=0.003$), age ($P<0.001$), tumor volume (cm^3) ($P=0.008$) among the three groups. As for VASARI and related radiological features, statistical analysis across three different groups showed no significant differences in the angle of the tentorium cerebelli ($P=0.423$), tumor location relative to tentorium cerebelli ($P=0.106$), enhancement quality ($P=0.560$), proportion enhancing ($P=0.121$), cysts ($P=0.343$), multifocal ($P=0.566$), thickness of enhancing margin ($P=0.112$), definition of the enhancing margin ($P=0.815$), definition of the non-enhancing margin ($P=0.247$), proportion of edema ($P=0.693$), and hemorrhage ($P=0.807$). Additionally, the PA group had the lowest incidence of hydrocephalus among PRTs patients ($P<0.001$).

Machine learning

Figure 3A and B showed the ROC and DET curves of the 5 machine learning models, respectively. Among those methods, the GP classifier achieved highest AUC value of 0.869 for prediction, with an accuracy of 0.73, and XGBboost showed a slightly lower AUC of 0.855 and an accuracy of 0.70. Detailed performance metrics of these methods were listed in Table 2. And for tree-based models (RF, AdaBoost, XGBoost), we also calculated the feature importance scores which represents the weights of features of prediction (Figure S2). The results indicated that the most relevant imaging features associated with surgical approach in the RF are the proportion enhancing, the angle of the tentorium cerebelli, and the tumor volume. For XGBoost, the most relevant imaging features associated with surgical approach are the proportion of edema, the relative position of the tumor, and the presence of hydrocephalus. And in the AdaBoost, the presence of hydrocephalus and tumor volume were the most relevant imaging features.

Deep learning

In the machine learning section, although these classifiers could achieve relative high AUC, the presence of class imbalance still remained a challenge. Therefore, we then applied deep learning algorithms by using the MRI images directly as inputs. We randomly selected 300 MRI images from the ITA to match images from the ISA and PA groups. The ROC and DET curves of these three models were shown in Fig. 3C and D, respectively. Off all the deep learning methods, EfficientNetV2-m achieved the highest AUC value of 0.89 with an accuracy of 0.797, whereas ResNet50 and ViT only got AUCs value of 0.709 and 0.614. This demonstrates a significant improvement by EfficientNetV2-m compared to other two methods ($P<0.005$ for both comparison). The training and validation curves for loss and accuracy also revealed a stable but high performed model of EfficientNetV2-m (Fig. S3). And Table 3 showed other evaluation metrics.

In above analysis, we used MRI images from all three orientations (axial, sagittal and coronal) as inputs. In clinical work, coronal and sagittal images often play a more crucial role in surgical approach selection. Therefore, we separately used coronal and sagittal images, matching the number of images across the three groups using the same

Table 1 Demographic and radiological information for the PRTs patients

Variable	All ($n=173$)	ITA($n=126$)	ISA ($n=24$)	PA ($n=23$)	*P*
Gender					0.003
Male	116(67.1)	93(73.8)	14(58.3)	9(39.1)	
Female	57(32.9)	33(26.2)	10(41.7)	14(60.9)	
Age(year)	18 (12,40.5)	16(12,32)	22 (14,47.5)	44 (27.5,58)	<0.001
Tumor Volume (cm^3)	21.6(9.2,40.8)	18.6(8,18.6)	30.6(12.4,58.7)	37.0(23.7,53.5)	0.008
Tentorium Tilt Angle	43 (36,52)	45.5(38,53)	40.5 (36,50)	43 (36.5,47)	0.423
Tumor location to the tentorium cerebellum					0.106
Upper	31 (17.9)	28(22.2)	1(4.2)	2(8.7)	
Lower	42 (24.3)	30(23.8)	8(33.3)	4(17.4)	
Both	100 (57.8)	68(54.0)	15(62.5)	17(73.9)	
Enhancement Quality					0.560
None	41(23.7)	34(27.0)	4(16.7)	3(13.0)	
Mild/minimal	68(39.3)	47(37.3)	11(45.8)	10(43.5)	
Marked/avid	64(37.0)	45(35.7)	9(37.5)	10(43.5)	
Proportion enhancing					0.121
0–5%	58(33.5)	47(37.3)	7(29.2)	4(17.4)	
6–33%	47(27.2)	28(22.2)	8(33.3)	11(47.8)	
34–67%	39 (22.5)	31(24.6)	3(12.5)	5(21.7)	
68–95%	29 (16.8)	20(15.9)	6(25.0)	3(13.0)	
Cysts					0.343
Yes	57(32.9)	42(33.3)	10(41.7)	5(21.7)	
No	116(67.1)	84(66.7)	14(58.3)	18(78.3)	
Multifocal					0.566
Yes	3(1.7)	3(2.4)	0(0.0)	0(0.0)	
No	170(98.3)	123(97.6)	24(100.0)	23(100.0)	
Thickness of enhancing margin					0.112
Thin	91(52.6)	67(53.2)	12(50.0)	12(52.2)	
Thick	15(8.7)	6(4.8)	3(12.5)	6(26.1)	
Solid	67 (38.7)	53(42.1)	9(37.5)	5(21.7)	
Definition of the enhancing margin					0.815
Well-defined	49 (28.3)	36(28.6)	8(33.3)	5(14.8)	
Poorly-defined	124 (71.7)	90(61.4)	16(66.7)	18(78.3)	
Definition of the non-enhancing margin					0.247
Well-defined	97(56.1)	66(52.4)	15(62.5)	16(69.6)	
Poorly-defined	76(43.9)	60(47.6)	9(37.5)	7(30.4)	
Proportion of edema					0.693
0–5%	108(62.4)	81(64.3)	15(62.5)	12(52.2)	
6–33%	56(32.4)	40(31.7)	7(29.2)	9(39.1)	
34–67%	9(5.2)	5(4.0)	2(8.3)	2(8.7)	
Hemorrhage					0.807
Yes	5(2.9)	3(2.4)	1(4.2)	1(4.3)	
No	168(97.1)	123(97.6)	23(95.8)	22(95.7)	
Hydrocephalus					<0.001
Yes	135(78.0)	102(81.0)	22(91.7)	11(47.8)	
No	38(22.0)	24(19.0)	2(8.3)	12(52.2)	

random selection strategy for model training. Figure 4 displayed the ROC curves for different input image sets, as well as the confusion matrix and ROC curves for each class (ITA, ISA and PA) using sagittal and coronal MRI as inputs, separately. The results showed that using images from multiple imaging planes as input led to the highest predictive performance.

For model interpretation, Grad-CAM was employed to visualize the specific areas within MRI images that the model focused on when making predictions related

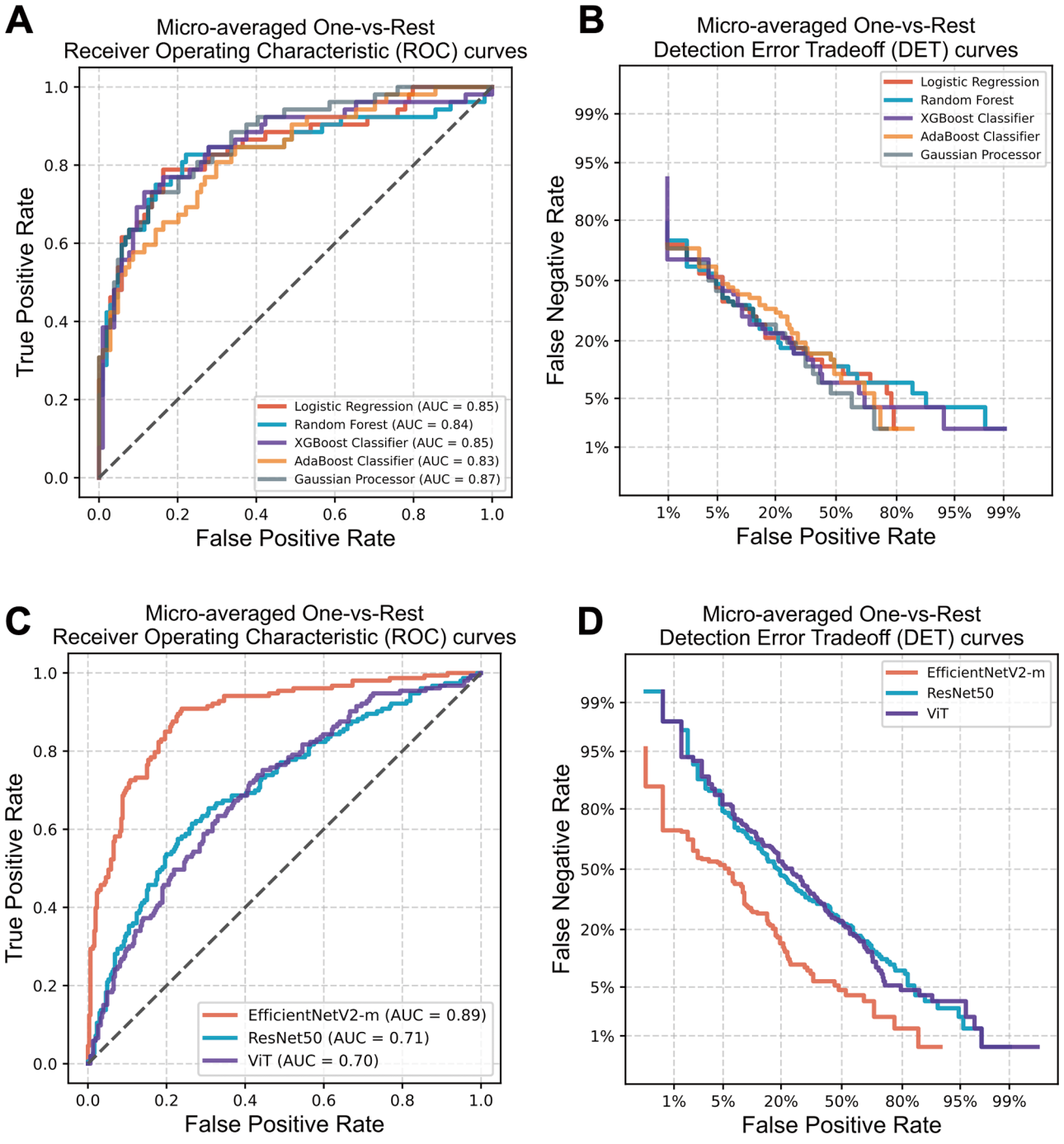

Fig. 3 (**A**) and (**B**) are the ROC and DET curves of machine learning models, respectively. (**C**) and (**D**) are the ROC and DET curves of deep learning models, respectively

Table 2 The performance metrics of machine learning models

	Logistic Regression	Random Forest	XGBoost Classifier	AdaBoost Classifier	Gaussian Processor
Accuracy	0.73	0.75	0.70	0.70	0.73
Weighted Average AUC	0.850	0.836	0.855	0.830	0.869
Weighted Average Precision	0.53	0.61	0.61	0.62	0.53
Weighted Average Recall	0.73	0.73	0.73	0.65	0.73
Weighted F1-score	0.62	0.66	0.66	0.63	0.62

Table 3 The performance metrics of deep learning models

	EfficientNetV2-m	ResNet50	Vision Transformer
Accuracy	0.797	0.601	0.699
Weighted Average AUC	0.891	0.709	0.614
Weighted Average Precision	0.8	0.65	0.65
Weighted Average Recall	0.8	0.6	0.6
Weighted F1-score	0.8	0.59	0.59

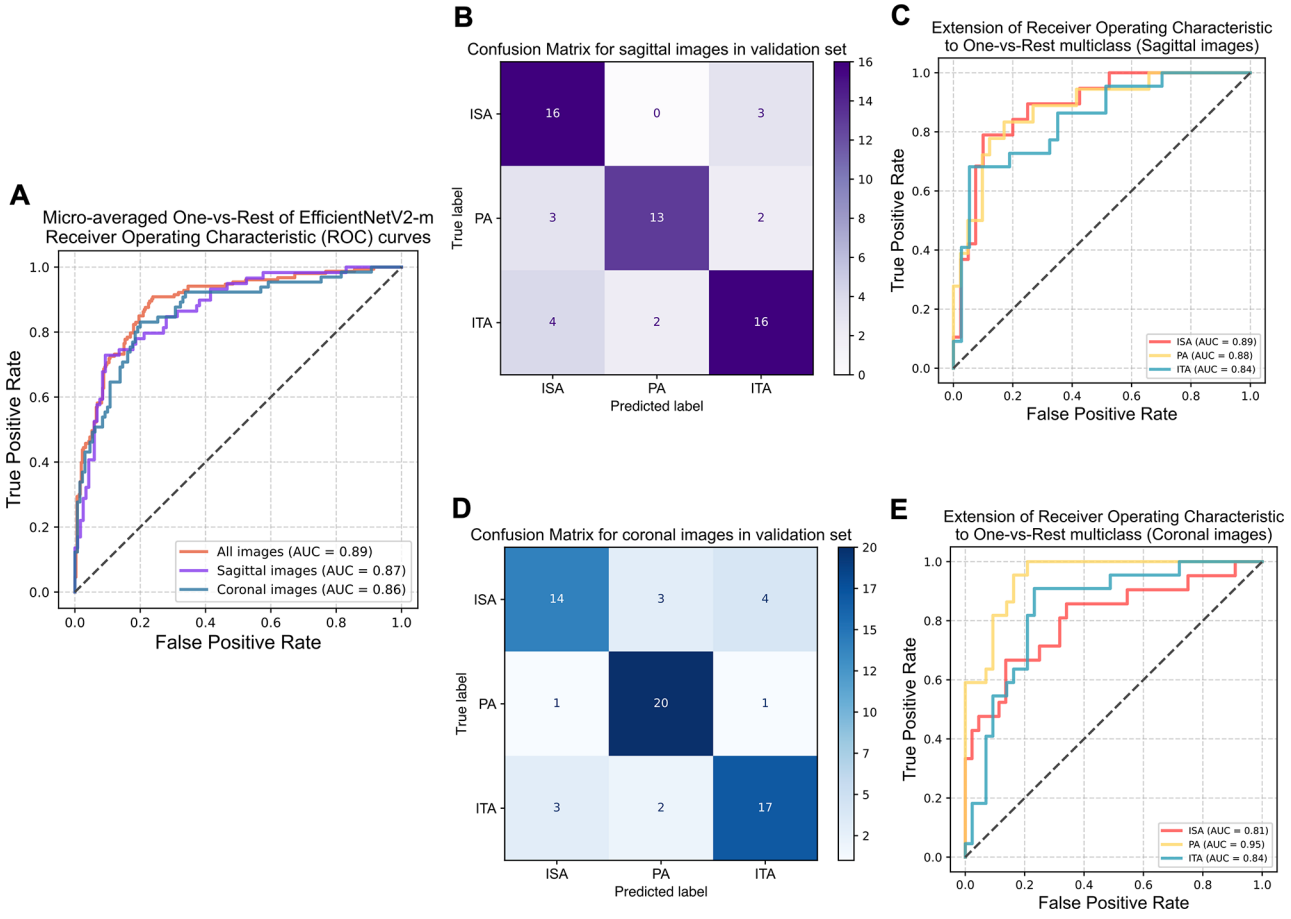

Fig. 4 (**A**) The ROC curves for EfficientNetV2-m using all images, sagittal images only and coronal images only. (**B**) The confusion matrix for EfficientNetV2-m with sagittal images. (**C**) The macro-average ROC curves for each class by taking the average over them for EfficientNetV2-m with sagittal images. (**D**) The confusion matrix for EfficientNetV2-m with coronal images. (**E**) The macro-average ROC curves for each class by taking the average over them for EfficientNetV2-m with coronal images

to surgical approach. Generally, the warmer area in Fig 5 were the contrast enhancing tumor cores and surrounding areas, demonstrating tumors and the surrounding relations are crucial areas for model decision-making, aligning with clinical judgments made by experienced surgeons. However, the visualization effect of CAM needs to be further improved, and its corresponding clinical significance has not been fully explained yet, requiring further research and discussion.

Discussion

In this study, we proposed a workflow with integration of clinical and radiological information, as well as AI methods to investigate the proper surgical approach for PRTs patients. In our deep learning model, we implemented the EfficientNetV2-m architecture using different orientations of MRI images, achieving AUC values consistently over 0.8. These results suggested the potential application value in surgical approach recommendation. Moreover, we facilitate better patient communication and understanding and medical education. Importantly, by integrating conventional radiological features with deep learning-derived representations, this study presents a

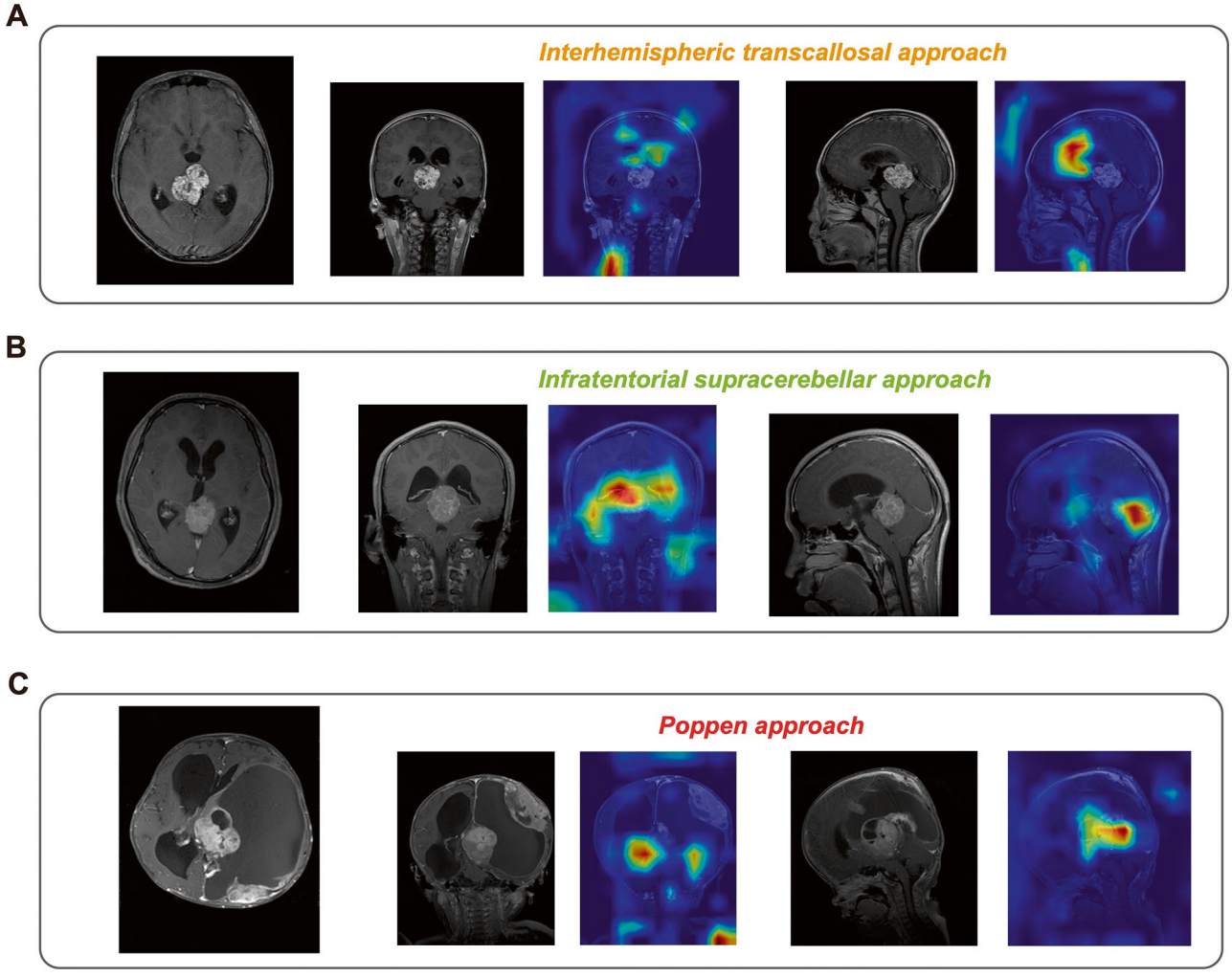

Fig. 5 Grad-CAM visualization of EfficientNetV2-m for PRTs is depicted in (**A**), (**B**), and (**C**), showcasing MRI images from different orientations along with the corresponding class activation maps for patients undergoing ITA, ISA, and PA surgeries, respectively

novel strategy to support surgical planning for pineal region tumors, bridging traditional clinical insight with advanced computational techniques. Combined with the relatively large sample size of 173 cases—the largest in this field to our knowledge—this work provides a novel and practical contribution to neurosurgical decision-making.

Currently, there is no established gold standard for selecting surgical approaches, and the choice of approach largely relies on the surgeon's prior clinical experience. While this experience can be effective, it is subjective and challenging to quantify and standardize. Specifically, the ISA does not fully preserve the bridging veins that drain from the tentorial surface to the tentorium. These veins often need to be partially or completely removed during surgery, and occasionally the anterior central cerebellar vein must also be cut. Consequently, the protection of the bridging veins is relatively poor, increasing the risk of cerebellar edema and infarction [19, 20]. Larger lesions that extensively invade the pineal region and its posterior part may obstruct the Galen vein and its branches when the ITA is used, potentially causing intraoperative injury. In our study, most patients did not undergo Magnetic Resonance Angiography scans due to the retrospective design, which limited our ability to evaluate the tumor's relationship with the vein of Galen. This limitation will be addressed in the following prospective analyses, where vascular imaging has been incorporated to provide more comprehensive anatomical assessment. Such injury can lead to postoperative cerebral blood flow obstruction, severe cerebral edema, and brain herniation [21]. When PA is employed, the internal occipital vein is often compromised due to the surgical field requirements. This can lead to occipital lobe infarction and edema after surgery, which may result in hemianopsia [20, 22]. Moreover, given that tumors in the pineal region are deep-seated, the primary goal of surgery is to decompress and alleviate symptoms of hydrocephalus [23]. For some patients with

extensive tumor growth, achieving complete resection is challenging regardless of the surgical approach used. Therefore, even if only subtotal resection is achieved, the choice of surgical approach can still be deemed reasonable if it fulfills the primary surgical objectives. Therefore, our model demonstrates the potential of machine learning approaches in translating and quantifying clinical experience to facilitate more informed surgical planning. Several VASARI features such as enhancing intensity, hydrocephalus status, edema, tumor volume, et al. were extracted and used for machine learning model construction and these features showed great importance in surgical-decision making in machine learning. However, some VASARI features did not show significant differences in our analysis. This may be because the VASARI system was originally developed for radiological assessment of tumor malignancy in brain tumors. In the case of pineal region tumors, however, the degree of malignancy is not the primary consideration when determining the surgical approach. In addition to VASARI features, some non-VASARI characteristics, such as the presence of hydrocephalus, also influence the choice of surgical approach. For instance, we observed a lower incidence of hydrocephalus in patients who underwent the PA, which may be associated with smaller tumor size and reduced aqueductal compression. In such cases, the PA is often preferred due to its minimally invasive nature and lower risk of damage to surrounding brain structures compared to ISA or ITA. A well-planned surgical approach not only ensures maximal exposure of the tumor for extensive resection but also facilitates protection of surrounding brain tissue [24]. According to previous studies [10, 25, 26], the ISA was the most ideal and common choice for tumors located in the midline, below the plane of the tentorium, growing into the third ventricle. However, there was no superiority of one over the others and surgeons need to determine the better one based on their experience [25].

Another interesting finding was that the angles of tentorium play an important role in approach decision. Hasan [27] et al. found that patients with infratentorial surgeries had a tentorial angle < 55°, whereas patients who underwent transtentorial surgeries had angle > 67°. This indicated that a steeper tentorium cerebelli angle is less favorable for infratentorial operations and provided a novel method for surgical planning evaluation.

However, the assessment of VASARI features often relies on clinical experience and may not fully enable end-to-end application. Therefore, we further advanced by employing deep learning to directly establish surgical approach recommendation models. Among them, EfficientNetV2-m performed better than others and reached highest AUC of 0.89 for all planes of images, the result was similar to other tasks based on MRI [28, 29]. This was due to its efficiency in memory and computational resources. ViT, however, did not converged well in our study because of the small scale of datasets which limited it to extract valuable internal features from the chaotic information [30]. This result highlights the differences in model architecture and training dynamics: transformer-based models like ViT are generally designed to leverage large-scale datasets, whereas CNN-based architectures can perform more robustly in data-limited scenarios, especially when combined with transfer learning techniques [31]. Considering the rare incidences of PTRs, we think our findings remained meaningful and offer practical implications for analysis in low-prevalence conditions. This supported the observation that CNN architectures are generally more data-efficient and better suited for small-scale datasets compared to data-hungry transformer models. Also, Grad-CAM showed that the main focus of EfficientNetV2-m in on the tumor location and its surrounding areas, indicating that the model has established a relatively robust 'attention' range based on MRI images from different orientations and could make decision for surgical approach according to these distinct ranges. The identified attention areas demonstrate that the model is capable of making informed decisions regarding the surgical approach. They can serve as a teaching and demonstration tool, facilitating explanations of surgical procedures to patients and training for junior doctors. Notably, some of the Grad-CAM visualizations revealed that the model did not primarily focus on the tumor itself when classifying the surgical approach based on MRI images. This suggested that the intrinsic characteristics of the tumor may not be the main determinants of surgical decision-making. Instead, the model tended to focus on regions surrounding the tumor, such as the size of the ventricles and the spatial relationship between the tumor and anatomical landmarks like the tentorium. This observation also provides an explanation for the limited utility of using VASARI features alone to assist in surgical route prediction. It highlighted the importance of incorporating contextual anatomical features beyond tumor morphology when developing decision-support models for neurosurgical planning. Additionally, the surgical pathways derived from our past experiences can provide valuable references for designing surgical approaches for new patients, optimizing surgical plans, and improving patient outcomes [32]. However, the interpretability of the models remains challenging specially for clinical implementation [33], and more interpretation methods need to explore in the future work.

Deep learning have been widely integrated into the diagnosis and treatment processes of neuro-oncology diseases, including pre-operative differentiation diagnosis [34, 35], prediction of molecular characteristics [36], surgical assistance [37], and prognosis evaluation [38].

However, to our literature review, the majority of AI in surgery were focused on intraoperative processes. For example, deep learning models could be used for assessing surgical actions in video clips [39], detecting surgical tools [40, 41], or surgical phase recognition [42]. To investigate how to use AI algorithms to assist the surgical planning, Tolga [43] et al. proposed a reinforcement learning method to recommend surgical approach by providing anatomical structures and pathological areas based on MRI. This study was still limited by computational and labelling cost. In our study, we used the simpler and more understandable surgeon 'experience' as labels for surgical approach prediction. This method not only conserves computational resources but also facilitates the translation and application from experience to model.

This study still has several limitations. First, as a retrospective study conducted at a single center, the subject's ratio we enrolled were imbalanced even though we controlled the bias by using different ways such as images resampling or stratified learning strategy. However, our hospital, as one of the tertiary centers for neurosurgery in China, routinely takes care of dozens to hundreds these relatively rare PRTs surgeries each year. Both the expertise of our surgeons and the substantial number of cases contribute significantly to our ability to conduct this research. Looking ahead, future plans will focus on expanding datasets through multicenter collaboration, which will help to improve the generalizability and external validity of our findings. In addition, we aim to incorporate multimodal information—such as preoperative imaging, intraoperative videos, and clinical data—to enrich the model inputs. By leveraging unsupervised or self-supervised learning techniques, we hope to further enhance the robustness and generalizability of our models, especially in scenarios with limited labeled data. Second, the choice of craniotomy approach of PRTs does not have a "***gold standard***", the predictive labels we defined according to achieving gross total or near-total resection of the tumor without significant complications might be biased by our center's experience. However, it is worth noting that our study also provides valuable assistance to surgeons with less experience and holds value for future developments in AI-assisted surgery. Finally, we only used VASARI features and T1C images as input for machine learning. And we will combine multi-sequence MRI analysis to gain a deeper understanding of the relationship between PRTs and adjacent tissues, thereby overcoming limitations of single MRI sequences [44].

Conclusion

In this study, we constructed machine learning and deep learning models based on radiological features and MRI images for PRT craniotomy approach prediction. Specially, deep learning model EfficientNetV2-m could achieve 0.89 AUC of prediction and provide efficient and personalized decision support tools for PRTs surgical approach selection, with promising potential for clinical application. These models assist clinicians in quickly learning and mastering complex surgical skills, shorten the learning time, and contribute to standardizing and improving the quality of medical training and practical skills.

Abbreviations

AI	Artificial intelligence
AUC	Area under the curve
CAM	Class activation mapping
CNN	Conventional neural networks
DET	Detection error tradeoff
MRI	Magnetic resonance imaging
PRT	Pineal region tumor
ROC	Receiver operation curve
ROI	Region of interest
VASARI	Visually AcceSAble rembrandt images
ViT	Vision transformer

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12880-025-01712-2.

Supplementary Material 1

Acknowledgements

We thank all who contributed to this research.

Author contributions

Chen ZY: Formal analysis, Methodology, Visualization, Writing - original draft, Writing - review & editing; Chen YH: Data collection, Writing - original draft, Writing – review & editing; Su YD & Jiang N: Data review, Writing – review; Wanggou SY: Conceptualization, Supervision, Writing - review; Li XJ: Conceptualization, Writing - review, Project administration, Supervision, all authors revised and approved the manuscript.

Funding

The research was funded by the National Natural Science Foundation of China (Grant No. 82270825), the Special funds for innovation in Hunan Province (Grant No. 2020SK2062), and High talent project of Hunan Province (Grant No. 2022WZ1031).

Data availability

The data underlying this article cannot be shared publicly due to the privacy of individuals that participated in the study. The processing code and data that support the findings of this study are available from the corresponding authors upon reasonable request.

Declarations

Ethics approval and consent to participate

The project was approved by the Ethics Committee of Xiangya hospital, Central South University (No.202202033). Given the retrospective nature of the study and the use of fully anonymized patient data, the requirement for informed consent was waived by the ethics committee. All procedures were conducted in accordance with the ethical standards outlined in the Declaration of Helsinki and its later amendments. All patients and their families underwent detailed preoperative discussions and signed surgical informed consent forms prior to treatment.

Consent for publication

Not applicable.

Competing interests
The authors declare no competing interests.

Received: 7 January 2025 / Accepted: 6 May 2025
Published online: 27 May 2025

References

1. Rousselle C, des Portes V, Berlier P, Mottolese C. Pineal region tumors: clinical symptoms and syndromes. Neurochirurgie. 2015;61(2–3):106–12.
2. Azab WA, Nasim K, Salaheddin W. An overview of the current surgical options for pineal region tumors. Surg Neurol Int. 2014;5:39.
3. Qi S, Fan J, Zhang XA, Zhang H, Qiu B, Fang L. Radical resection of nongerminomatous pineal region tumors via the occipital transtentorial approach based on arachnoidal consideration: experience on a series of 143 patients. Acta Neurochir (Wien). 2014;156(12):2253–62.
4. Bruce JN, Ogden AT. Surgical strategies for treating patients with pineal region tumors. J Neurooncol. 2004;69(1–3):221–36.
5. Hu X, Ren YM, Yang X, Liu XD, Huang BW, Chen TY, Jv Y, Lan ZG, Liu WK, Liu XS, et al. Surgical treatment of pineal region tumors: an 18 year-Experience at a single institution. World Neurosurg. 2023;172:E1–11.
6. Pettorini BL, Al-Mahfoud R, Jenkinson MD, Avula S, Pizer B, Mallucci C. Surgical pathway and management of pineal region tumours in children. Childs Nerv Syst. 2013;29(3):433–9.
7. Richards O, Gelder C, Nisar S, Wang K, Goodden J, Chumas P, et al. A comparison of the extent of resection in pineal region tumours via the occipital transtentorial and supracerebellar infratentorial approaches. Br J Neurosurg. 2021;38(3):568–72.
8. Fang AS, Meyers SP. Magnetic resonance imaging of pineal region tumours. Insights Imaging. 2013;4(3):369–82.
9. Mottolese C, Szathmari A, Ricci-Franchi AC, Beuriat PA, Grassiot B. The sub-occipital transtentorial approach revisited base on our own experience. Neurochirurgie. 2015;61(2–3):168–75.
10. Mottolese C, Szathmari A, Ricci-Franchi AC, Gallo P, Beuriat PA, Capone G. Supracerebellar infratentorial approach for pineal region tumors: our surgical and technical considerations. Neurochirurgie. 2015;61(2–3):176–83.
11. Khalighi S, Reddy K, Midya A, Pandav KB, Madabhushi A, Abedalthagafi M. Artificial intelligence in neuro-oncology: advances and challenges in brain tumor diagnosis, prognosis, and precision treatment. NPJ Precis Oncol. 2024;8(1):80.
12. Chen H, Gomez C, Huang CM, Unberath M. Explainable medical imaging AI needs human-centered design: guidelines and evidence from a systematic review. NPJ Digit Med. 2022;5(1):156.
13. Borys K, Schmitt YA, Nauta M, Seifert C, Kramer N, Friedrich CM, et al. Explainable AI in medical imaging: an overview for clinical practitioners-Saliency-based XAI approaches. Eur J Radiol. 2023;162:110786.
14. Yushkevich PA, Yang G, Gerig G. ITK-SNAP: an interactive tool for semi-automatic segmentation of multi-modality biomedical images. Annu Int Conf IEEE Eng Med Biol Soc. 2016;2016:3342–5.
15. He KM, Zhang XY, Ren SQ, Sun J. Deep residual learning for image recognition. Proc Cvpr Ieee. 2016;770–8.
16. Tan M, Le Q. Efficientnetv2: Smaller models and faster training. In: *International conference on machine learning: 2021*: PMLR; 2021: 10096–10106.
17. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. ArXiv Preprint arXiv: 2020;2010:11929.
18. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via Gradient-based localization. Ieee Int Conf Comp Vis. 2017;2017:618–26.
19. Tahta A, Akalan N. Supracerebellar infratentorial approach, indications, and technical pitfalls. Adv Tech Stand Neurosurg. 2023;46:53–64.
20. Richards O, Gelder C, Nisar S, Wang K, Goodden J, Chumas P, Tyagi A. A comparison of the extent of resection in pineal region tumours via the occipital transtentorial and supracerebellar infratentorial approaches. Br J Neurosurg. 2024;38(3):568–72.
21. Sonabend AM, Bowden S, Bruce JN. Microsurgical resection of pineal region tumors. J Neurooncol. 2016;130(2):351–66.
22. Li D, Zhang H, Jia W, Zhang L, Zhang J, Liu W, Ni M, Jia G. Significance of the tentorial alignment in protecting the occipital lobe with the Poppen approach for tentorial or pineal area meningiomas. World Neurosurg. 2017;108:453–9.
23. Shepard MJ, Haider AS, Prabhu SS, Sawaya R, DeMonte F, McCutcheon IE, Weinberg JS, Ferguson SD, Suki D, Fuller GN, et al. Long term outcomes following surgery for pineal region tumors. J Neurooncol. 2022;156(3):491–8.
24. Lombardi G, Poliani PL, Manara R, Berhouma M, Minniti G, Tabouret E, Razis E, Cerretti G, Zagonel V, Weller M et al. Diagnosis and treatment of pineal region tumors in adults: A EURACAN overview. Cancers (Basel) 2022, 14(15).
25. Cavalheiro S, Valsechi LC, Dastoli PA, Nicacio JM, Cappellano AM, Saba da Silva N. Silva Da Costa MD: outcomes and surgical approaches for pineal region tumors in children: 30 years' experience. J Neurosurg Pediatr. 2023;32(2):184–93.
26. Kondo A, Suzuki M, Shimizu Y, Akiyama O. The surgical intervention for pineal region tumors. Childs Nerv Syst. 2023;39(9):2341–8.
27. Syed HR, Jean WC. A novel method to measure the tentorial angle and the implications on surgeries of the pineal region. World Neurosurg. 2018;111:e213–20.
28. Yang Y, Zhang L, Du M, Bo J, Liu H, Ren L, Li X, Deen MJ. A comparative analysis of eleven neural networks architectures for small datasets of lung images of COVID-19 patients toward improved clinical decisions. Comput Biol Med. 2021;139:104887.
29. Zulfiqar F, Bajwa UI, Mehmood Y. Multi-class classification of brain tumor types from MR images using efficientnets. Biomed Signal Proces 2023; 84(9354):10477.
30. Lee SH, Lee S, Song BC. Vision Transformer for Small-Size Datasets. *ArXiv* 2021, abs/2112.13492.
31. Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jegou H. Training data-efficient image transformers & distillation through attention. In: *Proceedings of the 38th International Conference on Machine Learning*. Edited by Marina M, Tong Z, vol. 139. Proceedings of Machine Learning Research: PMLR; 2021: 10347–10357.
32. Huff DT, Weisman AJ, Jeraj R. Interpretation and visualization techniques for deep learning models in medical imaging. Phys Med Biol 2021, 66(4).
33. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). Comput Methods Programs Biomed. 2022;226:107161.
34. Gauriau R, Bizzo BC, Kitamura FC, Landi Junior O, Ferraciolli SF, Macruz FBC, Sanchez TA, Garcia MRT, Vedolin LM, Domingues RC, et al. A deep Learning-based model for detecting abnormalities on brain MR images for triaging: preliminary results from a multisite experience. Radiol Artif Intell. 2021;3(4):e200184.
35. Chakrabarty S, Sotiras A, Milchenko M, LaMontagne P, Hileman M, Marcus D. MRI-based identification and classification of major intracranial tumor types by using a 3D convolutional neural network: A retrospective Multi-institutional analysis. Radiol Artif Intell. 2021;3(5):e200301.
36. Li YM, Wei D, Liu X, Fan X, Wang K, Li SW, Zhang Z, Ma K, Qian TY, Jiang T, et al. Molecular subtyping of diffuse gliomas using magnetic resonance imaging: comparison and correlation between radiomics and deep learning. Eur Radiol. 2022;32(2):747–58.
37. Davids J, Makariou SG, Ashrafian H, Darzi A, Marcus HJ, Giannarou S. Automated Vision-Based microsurgical skill analysis in neurosurgery using deep learning: development and preclinical validation. World Neurosurg. 2021;149:e669–86.
38. Huang J, Shlobin NA, DeCuypere M, Lam SK. Deep learning for outcome prediction in neurosurgery: A systematic review of design, reporting, and reproducibility. Neurosurgery. 2022;90(1):16–38.
39. Khalid S, Goldenberg M, Grantcharov T, Taati B, Rudzicz F. Evaluation of deep learning models for identifying surgical actions and measuring performance. Jama Netw Open 2020, 3(3).
40. Alshirbaji TA, Jalal NA, Docherty PD, Neumuth T, Möller K. A deep learning spatial-temporal framework for detecting surgical tools in laparoscopic videos. Biomed Signal Proces. 2021;68:102801.
41. Namazi B, Sankaranarayanan G, Devarajan V. A contextual detector of surgical tools in laparoscopic videos using deep learning. Surg Endosc. 2022;36(1):679–88.
42. Jumah F, Raju B, Nagaraj A, Shinde R, Lescott C, Sun H, Gupta G, Nanda A. Uncharted waters of machine and deep learning for surgical phase recognition in neurosurgery. World Neurosurg. 2022;160:4–12.
43. Dundar TT, Yurtsever I, Pehlivanoglu MK, Yildiz U, Eker A, Demir MA, Mutluer AS, Tektas R, Kazan MS, Kitis S, et al. Machine Learning-Based surgical planning for neurosurgery: artificial intelligent approaches to the cranium. Front Surg. 2022;9:863633.
44. Yang M, Wang J, Zhang L, Liu J. Update on MRI in pediatric intracranial germ cell tumors-The clinical and radiological features. Front Pediatr. 2023;11:1141397.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.