

RESEARCH

Open Access



A deep learning model for differentiating paediatric intracranial germ cell tumour subtypes and predicting survival with MRI: a multicentre prospective study

Yanong Li^{1,2†}, Zhizheng Zhuo^{2†}, Jinyuan Weng², Sven Haller³, Harrison X. Bai⁴, Bo Li¹, Xing Liu⁵, Mingwang Zhu⁶, Zheng Wang⁷, Jane Li⁸, Xiaoguang Qiu^{1*} and Yaou Liu^{2*}

Abstract

Background The pretherapeutic differentiation of subtypes of primary intracranial germ cell tumours (iGCTs), including germinomas (GEs) and nongerminomatous germ cell tumours (NGGCTs), is essential for clinical practice because of distinct treatment strategies and prognostic profiles of these diseases. This study aimed to develop a deep learning model, iGNet, to assist in the differentiation and prognostication of iGCT subtypes by employing pretherapeutic MR T2-weighted imaging.

Methods The iGNet model, which is based on the nnUNet architecture, was developed using a retrospective dataset of 280 pathologically confirmed iGCT patients. The training dataset included 83 GEs and 117 NGGCTs, while the retrospective internal test dataset included 31 GEs and 49 NGGCTs. The model's diagnostic performance was then assessed with the area under the receiver operating characteristic curve (AUC) in a prospective internal dataset ($n = 22$) and two external datasets ($n = 22$ and 20). Next, we compared the diagnostic performance of six neuroradiologists with or without the assistance of iGNet. Finally, the predictive ability of the output of iGNet for progression-free and overall survival was assessed and compared to that of the pathological diagnosis.

Results iGNet achieved high diagnostic performance, with AUCs between 0.869 and 0.950 across the four test datasets. With the assistance of iGNet, the six neuroradiologists' diagnostic AUCs (averages of the four test datasets) increased by 9.22% to 17.90%. There was no significant difference between the output of iGNet and the results of pathological diagnosis in predicting progression-free and overall survival ($P = .889$).

Conclusions By leveraging pretherapeutic MR imaging data, iGNet accurately differentiates iGCT subtypes, facilitating prognostic evaluation and increasing the potential for tailored treatment.

Keywords Deep learning (DL), Magnetic resonance imaging (MRI), Diagnostic model, Prognostic evaluation, Primary intracranial germ cell tumours (iGCTs)

[†]Yanong Li and Zhizheng Zhuo contributed equally to this work.

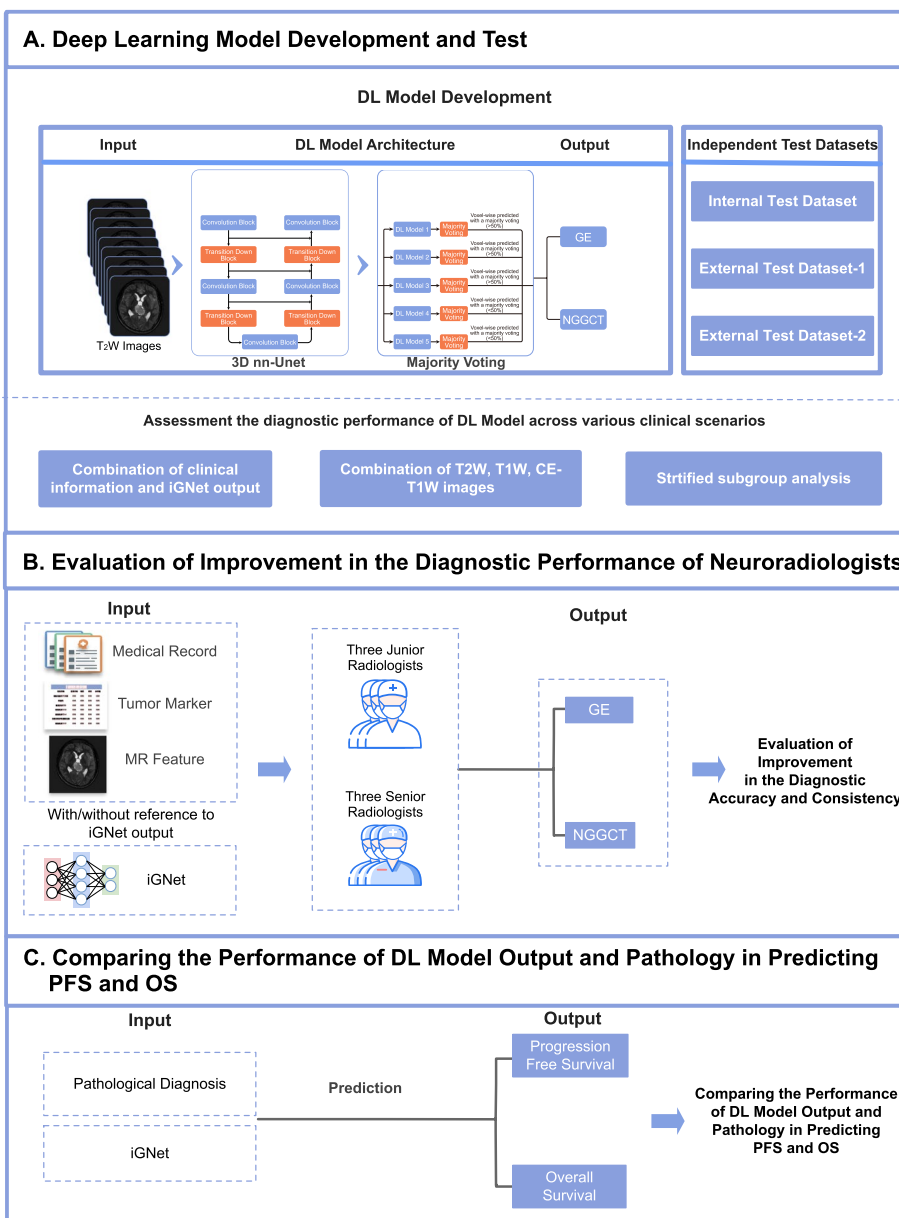
*Correspondence:

Xiaoguang Qiu
qxiaoguang@bjtth.org
Yaou Liu
liuyaou@bjtth.org

Full list of author information is available at the end of the article



Graphical Abstract



Background

Primary intracranial germ cell tumours (iGCTs) are the third most prevalent primary brain tumour in the paediatric and adolescent populations in East Asia, particularly in China, Korea, and Japan [1–3]. The incidence of iGCTs in these regions is approximately 2.7 per million individuals per year, trailing only that of gliomas and medulloblastoma [4]. Notably, iGCTs constitute approximately 8% to 15% of all paediatric brain tumours [5, 6]. According to

the 2021 World Health Organization (WHO) classification, these tumours can be categorized into germinomas (GEs), teratomas, yolk sac tumours, choriocarcinomas, embryonal carcinomas, and mixed germ cell tumours [7]. Clinically, GEs and non-germinomatous germ cell tumours (NGGCTs) demonstrate differences in radiosensitivity and prognosis [8–10]. Among nonmetastatic iGCTs, GEs are highly radiosensitive, and 4 cycles of chemotherapy combined with low-dose radiotherapy

can achieve a 5-year overall survival (OS) rate of 90% [6]. However, radiotherapy can cause side effects such as brain atrophy, white matter degeneration, and impaired or lost fertility (especially in females). Therefore, the primary treatment goal for GEs is to reduce the radiation dose and improve the quality of life of the patients while achieving a high survival rate [4]. NGGCTs are less radiosensitive than GEs; the standard treatment regimen consists of 6 cycles of chemotherapy followed by an assessment of the response. If the residual lesion is greater than 1 cm in diameter, surgical resection is recommended, followed by high-dose radiotherapy. If the residual lesion is less than or equal to 1 cm in diameter, high-dose radiotherapy is administered directly [6]. The 5-year OS rate for NGGCTs ranges from 50 to 90% (excluding mature teratomas, which can be cured by surgery directly). The primary treatment goal for NGGCTs is to prolong survival time [11–13]. Therefore, the differentiation of GEs and NGGCTs is essential for clinically stratified treatment.

Pathology is traditionally regarded as the gold standard for identifying iGCT subtypes. However, approximately 30% of iGCTs are unsuitable for biopsy or surgical resection due to the high risk of significant neurological deficits, especially when located in sensitive regions such as the sellar area and basal ganglia, or for patients harbouring potentially adverse conditions such as intracranial haemorrhage and metastasis [14–16]. Currently, the clinical consensus [6, 17] in the field of paediatric neurooncology suggests that iGCTs can be diagnosed through a combination of characteristic radiological features and elevated levels of tumour markers, including beta-human chorionic gonadotropin (β -HCG) and alpha-fetoprotein (AFP). However, the diagnostic values for these biomarkers are subject to international variations (β -HCG ≥ 50 IU/L and AFP ≥ 25 ng/mL in Europe, β -HCG ≥ 100 IU/L and AFP ≥ 10 ng/mL in the USA), leading to a lack of consistency across different populations [6]. Compared with the use of tumour markers, imaging is a promising approach that is more objective and stable across populations. A study by Wu et al. [18] on the MR features of 85 iGCT patients suggested that GEs have lower apparent diffusion coefficient (ADC) values than NGGCTs. Li et al. [16] combined quantitative and semiquantitative data from MR images (e.g. perfusion and diffusion images) to differentiate GEs and NGGCTs and revealed that conventional MR features combined with ADC and perfusion-weighted imaging (PWI) values achieved good differentiation (area under the curve (AUC) = 0.950). Nevertheless, the generalizability of these findings is constrained by the study's single-centre nature, limited sample size and low dimensionality of the features. Therefore, the determination of specific

imaging features to assist in the diagnosis of GEs and NGGCTs remains challenging. Deep learning (DL), with its ability to harness high-dimensional imaging features related to brain tumour grade, pathology, and molecular markers, has shown promise in enhancing the accuracy of brain tumour diagnosis. Although previous studies have used machine learning algorithms to distinguish iGCTs from other brain tumours [19–21], thus far, they have not been applied in differentiating iGCT subtypes.

This study aims to develop a DL model, iGNet, to accurately and independently distinguish GEs from NGGCTs, providing clinicians with crucial imaging-based differentiation of the two types of tumours. We utilized a comprehensive dataset of widely used MR T2-weighted (T2W) images from multiple centres to ensure the model could be seamlessly integrated into clinical practice. We then evaluated the performance of iGNet in facilitating clinical decision-making and predicting patient survival.

Methods

Study design and participants

This study aimed to develop and test an end-to-end automated pipeline for differentiating GEs and NGGCTs based on the most clinically available preoperative brain T2W images with a DL algorithm to aid in clinical decision-making and prognostic evaluation. First, we trained and tested the DL model (iGNet) with a retrospective dataset from Beijing Tiantan Hospital, Capital Medical University. We subsequently tested iGNet in an independent internal prospective test dataset from Beijing Tiantan Hospital, Capital Medical University, and two independent external test datasets from Beijing Sanbo Hospital, Capital Medical University, and Tianjin Huanhu Hospital, Tianjin Medical University. A series of additional analyses were conducted to interpret model performance. Finally, we evaluated the clinical role of the developed iGNet by assessing whether it could help neuroradiologists improve clinical diagnoses and predicting prognosis.

The Institutional Review Board of Beijing Tiantan Hospital, Capital Medical University, approved this study (KY2021-142-02). The need for informed consent was waived since MR images were retrospectively collected and anonymized. Written informed consent was obtained from patients whose MR images were prospectively collected. The local Institutional Review Board approved the use of data from external centres.

Figure 1 illustrates the flowchart of patient enrolment. The study consisted of multiple datasets: a retrospective dataset, a prospective internal test dataset, and two independent external test datasets. We collected data from 296 patients who underwent MR T2W imaging with pathology-confirmed GEs or NGGCTs from

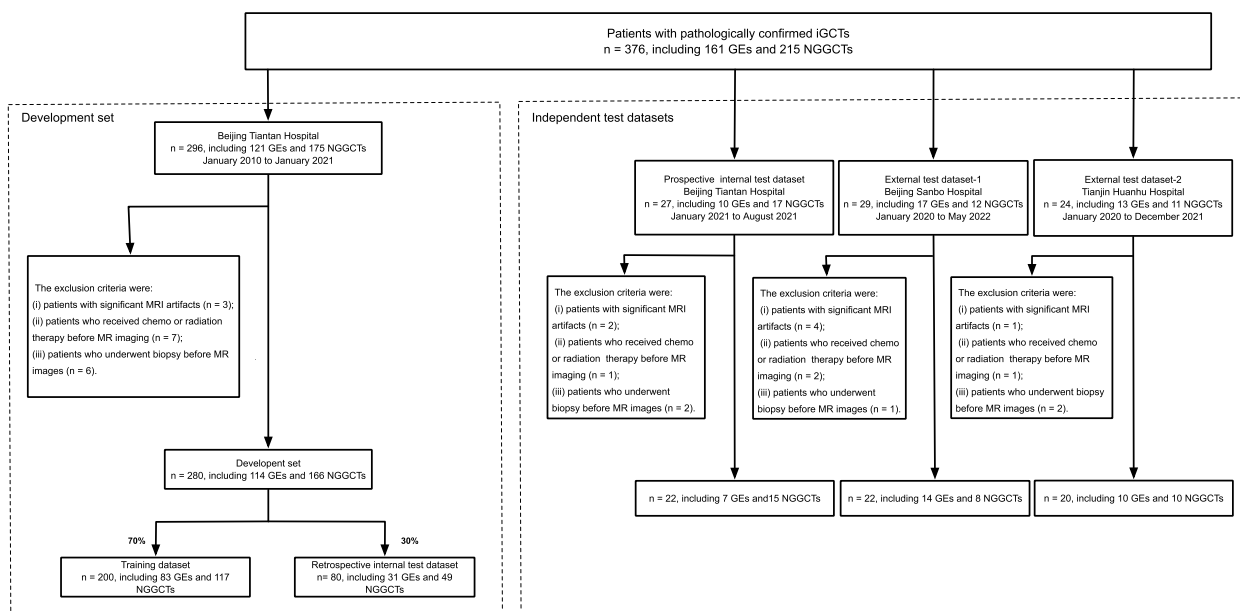


Fig. 1 Patient enrolment process and dataset distribution for the study. The figure shows the total number of patients with pathologically confirmed intracranial germ cell tumours (IGCTs), including in the development cohort ($n=280$), training cohort ($n=200$), and retrospective internal test cohort ($n=80$). The exclusion criteria are as follows: significant MRI artefacts, preimaging chemotherapy or radiation therapy, and preimaging biopsy. The figure also describes the prospective internal test dataset and two external test datasets from Beijing Sanbo Hospital and Tianjin Huanhu Hospital, specifying the number of patients and exclusion criteria for each

January 2010 to January 2021 at Beijing Tiantan Hospital, Capital Medical University, for developing iGNet. In addition, we used three independent datasets for testing the model, including an internal prospective test dataset, consisting of the data of 27 patients from Beijing Tiantan Hospital; external test dataset-1, consisting of the data of 29 patients from Beijing Sanbo Hospital; and external test dataset-2, consisting of the data of 24 patients from Tianjin Huanhu Hospital.

The inclusion criteria were as follows: (i) pathology-confirmed GEs or NGGCTs through biopsy or surgical resection, (ii) pretherapeutic axial T2W images obtained within 2 weeks prior to biopsy and surgical intervention, and (iii) age 0–18 years. We excluded patients with significant MR image artefacts, those who had received chemotherapy or radiation therapy before imaging, and those who underwent brain surgery targeting the tumour before imaging. Ultimately, we included 280 patients (development set) and randomly split them into training ($n=200$, 83 GEs, and 117 NGGCTs) and test datasets ($n=80$, 31 GEs, and 49 NGGCTs). The independent internal prospective test dataset included the data of 22 patients, whereas external test dataset-1 and external test dataset-2 included the data of 22 and 20 patients, respectively, all subjected to the same inclusion criteria as the main study group.

Outcomes

The primary outcome was the diagnostic AUC of iGNet. The secondary outcomes were the AUC improvement rate of the neurologists’ diagnosis with iGNet assistance and the difference in the predictive abilities for progression-free survival (PFS) and OS between DL outputs and pathological findings.

MR image acquisition

T2W images for both the internal and external datasets were acquired with MR scanners with field strengths of 1.5 T and 3.0 T and from different vendors, including Philips, Siemens, GE, and Toshiba. Additionally, conventional 2D or 3D T1-weighted (T1W) and gadolinium contrast-enhanced T1W (CE-T1W) images were collected. For the iGCT MR images, axial turbo-spin-echo 2D-T2W images were acquired with the following protocol parameters: repetition time/echo time = 3030–6711/84–119 ms; flip angle = 90°–160°; slice thickness = 3–6 mm; and matrix size = 256–328 × 512–512. Additional file 1: Fig. S1 and Table S1 provide further details of the protocol parameters.

Histological determination

Two neuropathologists (X.L. and Y.J.H., each with 10 years of experience in neuropathology) independently reviewed the integrated histological diagnosis for

the development and internal test datasets and reached a consensus according to the 2021 WHO classification of tumours of the central nervous system. The specimens that did not receive a diagnosis according to the 2021 WHO classification were re-reviewed by the same neuropathologists. The interrater agreement was high (Kappa = 0.932, $P < 0.0001$). Disagreements were resolved by discussion with a third senior neuropathologist (J.D., with more than 30 years of experience). The immunohistochemical and diagnostic criteria for the external test datasets were consistent with those used for the development set (Additional file 1: Method S1) [22] showed the criteria for the pathological diagnosis of iGCTs).

Conventional MR feature assessment

Two neuroradiologists (X.K. and T.T.H., with 6 years of experience in neuroradiology) independently assessed conventional MR characteristics, including T1W hyperintensity, T2W hypointensity, enhancement, haemorrhage, and cystic/necrosis, in all datasets (Kappa = 0.877–0.932, all $P < 0.0001$). Furthermore, a senior neuroradiologist (Y.Y.D., with 15 years of experience in neuroradiology) resolved any disagreements. For performance comparison with iGNet, we constructed a logistic regression model using conventional clinical information and MR characteristics that were significantly different between GEs and NGGCTs in the training dataset. This approach was employed to evaluate the effectiveness of these variables in distinguishing between GEs and NGGCTs.

Blinded manual tumour labelling for generating the ground truth

Two neuroradiologists (X.K. and T.T.H.), blinded to the clinical data, independently manually delineated the tumour region with open-source software (ITK-SNAP, version 3.8.0; <http://www.itksnap.org>). The average interrater segmentation Dice score was 0.84 ± 0.12 . The manual segmentation results were then reviewed and modified by another experienced neuroradiologist (Y.Y.D.) when necessary. The overlap of the reviewed and confirmed tumour segmentations of the two raters was used as the ground truth for DL model development (Additional file 1: Fig. S2).

Deep learning model development

Image preprocessing

T2W images were first skull-stripped with the 'BET' tool in the FMRIB Software Library (FSL v6.0, <https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/>). The skull-stripped T2W images were subjected to N4 bias correction with the ANTs package (<http://stnava.github.io/ANTs/>). Subsequently, the T2W images were cropped to the size of a nonzero intensity

area. Finally, the intensity of each image was normalized by subtracting the mean and dividing by the standard deviation.

DL model training and testing

We developed iGNet to differentiate GEs from NGGCTs according to the information within the T2W images. In this study, we modified the state-of-the-art 3D nnU-Net [23] framework to simultaneously segment the tumour and predict the tumour subtype. We developed the model in the training dataset and tested in the test dataset. Fivefold cross-validation of the training dataset was adopted to improve the robustness of the model. The DL network segmentation and classification architectures are shown in Additional file 1: Method S2, Fig. S3, and Fig. S4. Method S3 [23–26] shows the details of iGNet development. Furthermore, we independently tested the model in the internal prospective test dataset and two independent external test datasets. Gradient-weighted class activation mapping (Grad-CAM) [27] was used to generate saliency maps to improve model interpretability. These saliency maps highlight important regions in the input image by leveraging the gradient information of the model during classification tasks. This capability aids in elucidating the model's decision-making process, thereby addressing the black box nature of deep learning.

DL model evaluation

We evaluated iGNet in terms of its performance in the test datasets. The evaluation included the following: the diagnostic performance of iGNet across various clinical scenarios (details shown below); the improvement in the diagnostic performance of six neuroradiologists when assisted by iGNet; and the potential role of the output of iGNet in predicting the PFS and OS of GEs and NGGCTs with respect to the pathological findings.

(1) Overview of model performance evaluation methods: We integrated a logistic regression model with the output of iGNet to assess the added value of clinical information and conventional MR characteristics in distinguishing between GEs and NGGCTs; conducted additional analyses using multimodal MR images as inputs; and performed sensitivity analyses on subgroups stratified by eleven clinical and imaging phenotypes, particularly tumour location. Previous studies have demonstrated that various tumour sites exhibit subtype biases; for example, the pineal region tends to have a higher incidence of NGGCTs, whereas the basal ganglia are more commonly associated with a higher incidence of GEs [9]. (2) Evaluation of the clinical role of iGNet: six neuroradiologists conducted diagnostic assessments with and without iGNet assistance to assess its efficacy. (3) Assessment of the output of iGNet as an

independent prognostic factor for PFS and OS: The performance of the model was evaluated with univariable Cox proportional hazards models and visualized with Kaplan–Meier survival analysis. More details of the DL model evaluation are provided in Additional file 1: Method S4 [4, 12, 28–30].

Statistical analysis

We used the Statistical Package for the Social Sciences (SPSS) software (version 22, IBM, USA) and Python (version 3.6, <http://www.python.org>) for the statistical analyses. Categorical variables are displayed as frequencies and percentages and were compared between groups with Pearson's chi-square test or Fisher's exact test. Continuous variables are displayed as the means and standard deviations (SDs) and were compared between groups with the two-sample *t* test or Mann–Whitney *U* test (according to the normality of the distribution of the data). Normality was assessed with the Kolmogorov–Smirnov test for large samples ($n > 50$) and the Shapiro–Wilk test for smaller samples ($n \leq 50$).

To compare the model constructed with T2W images with that constructed with multimodal MRI data, we used the DeLong test to compare their AUCs. Additionally, the DeLong test was employed to compare the AUCs between multivariate logistic regression and iGNet and between neuroradiologists' performance without and with iGNet assistance. A two-sided *P* value < 0.05 was considered to indicate statistical significance. Method S5 provides more details of the statistical analysis.

Results

Patient characteristics

A total of 344 patients were included in this study, including 280 patients (188 males and 92 females, with an age range of 3–19 years, mean age [SD] = 13 [6] years) in the training and test cohort; 22 patients (17 males and 5 females, with an age range of 9–23 years, mean age [SD] = 15 [5] years) in the independent prospective internal test cohort; 22 patients (18 males and 4 females, with an age range of 6–21 years, mean age [SD] = 14 [5] years) in external test cohort-1; and 20 patients (15 males and 5 females, with an age range of 6–16 years, mean age [SD] = 10 [3] years) in external test cohort-2. The demographics, tumour marker levels (β -HCG and AFP), and conventional MR features of these patients are summarized in Table 1.

The multivariate logistic regression model, created with demographic data, tumour marker levels (β -HCG and AFP), and conventional MR characteristics, achieved an accuracy of 76.77% (95% CI: 68.69–84.85%), a specificity of 64.86% (95% CI: 48.72–79.41%), a sensitivity of 83.87% (95% CI: 74.14–92.31%), and an AUC of 0.835 (95% CI:

0.636–0.942) in differentiating GEs from NGGCTs in the retrospective test dataset (Additional file 1: Fig. S5A).

Accurate differentiation of GEs from NGGCTs with the DL model

iGNet automatically segmented the entire tumour in the retrospective internal test dataset, achieving a Dice score of 0.83 ± 0.04 . The model demonstrated an accuracy of 90.91% (95% CI: 84.85–95.96%), a specificity of 78.40% (95% CI: 64.10–90.60%), a sensitivity of 98.39% (95% CI: 94.74–100.00%), and an AUC of 0.950 (95% CI: 0.863–0.994) in differentiating GEs from NGGCTs (Additional file 1: Fig. 2A and Table 2). The results of the fivefold cross-validation are shown in Additional file 1: Table S2.

In the prospective internal test dataset, iGNet achieved an accuracy of 87.10% (95% CI: 68.00–96.00%), specificity of 84.64% (95% CI: 67.42–100.00%), sensitivity of 76.47% (95% CI: 53.85–94.12%), and AUC of 0.921 (95% CI: 0.657–0.974) in differentiating GEs from NGGCTs.

In external test dataset-1, iGNet achieved an accuracy of 80.00% (95% CI: 60.00–95.00%), a specificity of 88.24% (95% CI: 70.59–100.00%), a sensitivity of 78.33% (95% CI: 56.65–95.14%), and an AUC of 0.869 (95% CI: 0.757–0.941) in this differentiation task.

In external test dataset-2, iGNet achieved an accuracy of 89.56% (95% CI: 71.22–100.00%), a specificity of 76.47% (95% CI: 53.85–94.12%), a sensitivity of 89.79% (95% CI: 76.84–98.62%), and an AUC of 0.905 (95% CI: 0.774–0.971) in differentiating GEs from NGGCTs. Additional file 1: Fig. 2B shows the visualization of the outputs of iGNet for the various datasets. The results of the DeLong test comparing the performance of multivariate logistic regression with that of iGNet are presented in Additional file 1: Table S3.

Robust performance of the DL model in various clinical scenarios

Evaluation of the DL model revealed that combining conventional information with the output of iGNet achieved an AUC of 0.963 (95% CI: 0.921–0.990) in the retrospective internal dataset, 0.899 (95% CI: 0.769–0.967) in the prospective internal dataset, 0.891 (95% CI: 0.760–0.943) in external dataset-1, and 0.917 (95% CI: 0.776–0.972) in external dataset-2. The integration of multimodal (T2W, T1W, and CE-T1W) MR images resulted in an AUC of 0.901 (95% CI: 0.800–0.938) in the retrospective internal dataset, 0.896 (95% CI: 0.785–0.980) in the prospective internal dataset, 0.856 (95% CI: 0.753–0.936) in external dataset-1, and 0.903 (95% CI: 0.773–0.969) in external dataset-2 (Fig. 2C and Additional file 1: Table S4). There was no statistically significant difference in diagnostic performance between the use of T2W images alone and the use of multimodal MR images (0.919 vs. 0.901,

Table 1 The main demographic characteristics, tumour marker levels and conventional MR features of the training, retrospective internal test, and prospective internal test cohorts, external test cohort-1, and external test cohort-2

| Variable | Development cohort | | | | | | Independent test cohorts | | | | | | | | | |
|--|----------------------|--------------------|---------|------------------------------------|---------------------|---------|----------------------------------|---------------------|--------|------------------------|---------------------|---------|------------------------|---------------------|---------|--|
| | Training cohort | | | Retrospective internal test cohort | | | Prospective internal test cohort | | | External test cohort-1 | | | External test cohort-2 | | | |
| | GEs (n=83) | NGGCTs (n=117) | P | GEs (n=31) | NGGCTs (n=49) | P | GEs (n=7) | NGGCTs (n=15) | P | GEs (n=14) | NGGCTs (n=8) | P | GEs (n=10) | NGGCTs (n=10) | P | |
| Demographic characteristics | | | | | | | | | | | | | | | | |
| Age (mean ± SD), years | 17 ± 8 | 10 ± 5 | <0.001* | 16 ± 6 | 11 ± 4 | <0.001* | 16 ± 5 | 15 ± 5 | 0.511 | 15 ± 1 | 13 ± 2 | 0.599 | 11 ± 1 | 10 ± 1 | 0.416 | |
| Sex, n | | | | | | | | | | | | | | | | |
| Female:male | 51:32 | 23:94 | <0.001* | 10:21 | 8:41 | 0.031* | 3:4 | 2:13 | 0.136 | 2:12 | 2:6 | 0.584 | 3:7 | 2:8 | 0.987 | |
| Tumour marker level | | | | | | | | | | | | | | | | |
| β-HCG (mIU/L) | 6.63 ± 2.79 | 395.29 ± 237.22 | 0.015* | 12.31 ± 5.59 | 1420.86 ± 721.69 | 0.005* | 25.06 ± 2.42 | 182.49 ± 65.8 | 0.220 | 6.65 ± 1.48 | 10.40 ± 1.04 | 0.021* | 9.13 ± 4.32 | 378.99 ± 21.03 | 0.039* | |
| AFP (ng/mL) | 3.42 ± 1.99 | 1019.22 ± 398.91 | 0.012* | 0.68 ± 0.22 | 1227.39 ± 793.10 | 0.048* | 2.17 ± 0.72 | 190.34 ± 40.31 | 0.049* | 0.599 ± 0.29 | 24.27 ± 8.54 | <0.001* | 5.35 ± 4.52 | 159.59 ± 78.90 | <0.001* | |
| Conventional MRI features | | | | | | | | | | | | | | | | |
| Tumour locations (sellar/suprasellar/pineal region/basal ganglia), n | 49:24:10 | 19:84:14 | 0.213 | 11:15:5 | 9:29:11 | 0.430 | 2:1:4 | 2:11:2 | 0.813 | 4:8:2 | 3:5:0 | 0.441 | 4:5:1 | 2:6:2 | 0.785 | |
| T1W image Foci-hyperintensity (present:absent) | 8:75 | 17:100 | 0.403 | 6:25 | 8:41 | 0.432 | 1:6 | 11:4 | 0.027* | 8:6 | 3:5 | 0.097 | 4:6 | 4:6 | 0.921 | |
| Enhancement degree on CE-T1W images (mild/moderate/obvious) | 19:36:28 | 6:41:70 | 0.210 | 4:8:19 | 5:14:30 | 0.781 | 3:1:3 | 3:5:7 | 0.786 | 6:4:4 | 3:3:2 | 0.811 | 3:5:2 | 4:3:3 | 0.656 | |
| Enhancement appearance (homogeneous/heterogeneous) | 50:33 | 26:91 | 0.078 | 14:17 | 10:39 | 0.049* | 5:2 | 6:9 | 0.442 | 11:3 | 2:6 | 0.036* | 7:3 | 5:5 | 0.492 | |
| T2W image foci-hypointensity (present:absent) | 6:77 | 5:112 | 0.366 | 1:30 | 3:46 | 0.644 | 1:6 | 11:4 | 0.032* | 8:6 | 3:5 | 0.279 | 3:7 | 5:5 | 0.479 | |
| Haemorrhage (present:absent) | 29:54 | 47:70 | 0.688 | 10:21 | 16:33 | 0.576 | 2:5 | 9:6 | 0.281 | 5:9 | 3:5 | 0.329 | 1:9 | 6:4 | 0.077 | |
| Cyst or necrosis (present:absent) | 45:38 | 67:50 | 0.614 | 18:13 | 27:22 | 0.540 | 2:5 | 8:7 | 0.062 | 3:11 | 3:5 | 0.107 | 6:4 | 8:2 | 0.717 | |
| Tumour volume (mm ³) | 19,987.76 ± 14,96.23 | 17,653.36 ± 996.53 | 0.796 | 15,993.66 ± 1543.25 | 19,647.53 ± 1965.33 | 0.675 | 17,540.32 ± 1236.23 | 20,012.21 ± 1500.25 | 0.325 | 18,866.85 ± 1742.32 | 20,365.84 ± 1985.32 | 0.074 | 16,325.23 ± 1302.32 | 18,693.29 ± 1652.44 | 0.056 | |

The P values were determined by the one-way analysis of variance test for age and tumour volume and the χ^2 test for sex, T1WI performance, T2WI performance, enhancement appearance, degree of enhancement, haemorrhage, and necrosis

β-HCG beta-human chorionic gonadotropin, AFP α-fetoprotein

* is used to indicate statistical significance

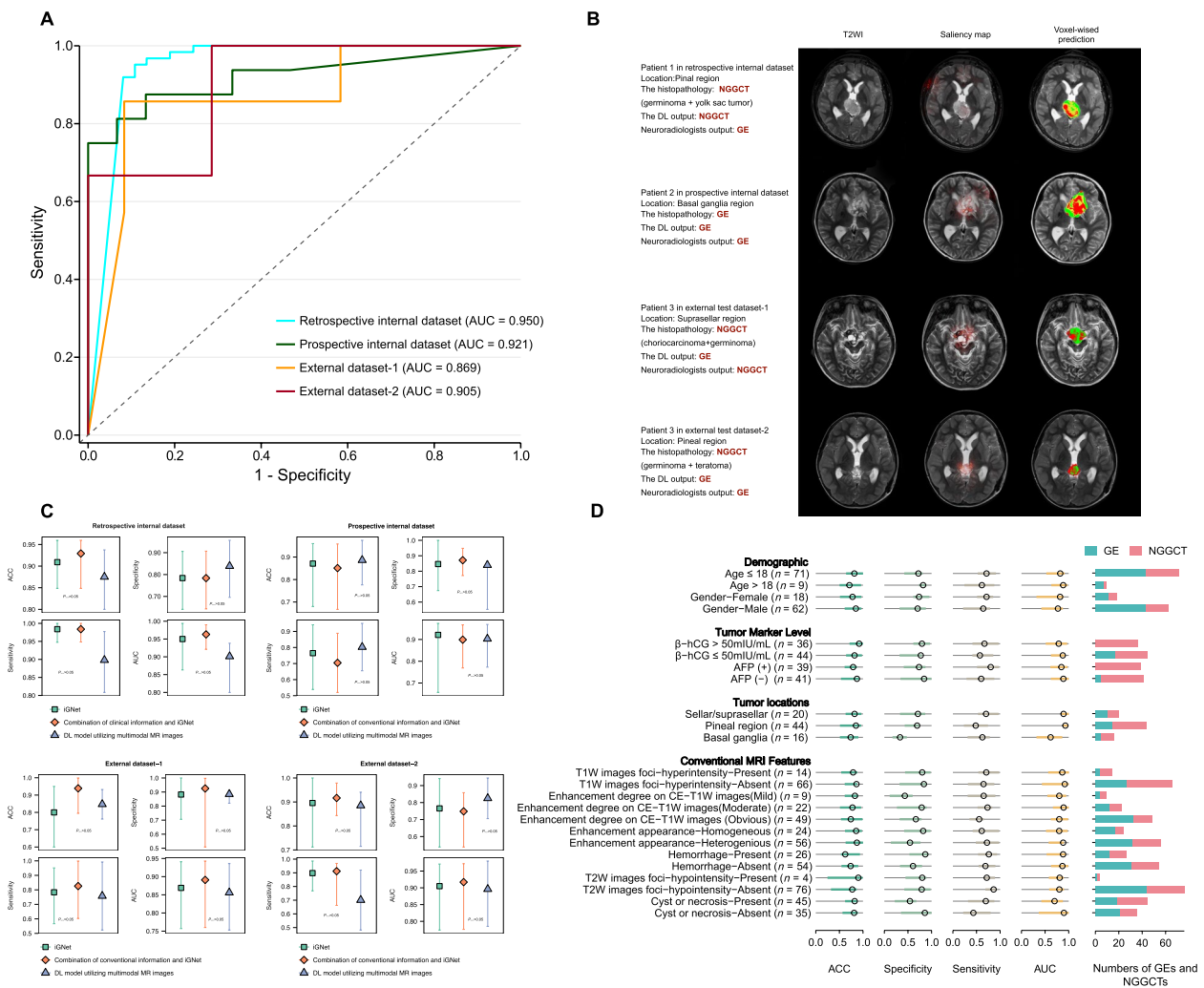


Fig. 2 Performance and application results of iGNet. **A** ROC curves displaying iGNet’s discriminative ability in the retrospective internal test dataset and three independent test datasets. **B** Representative T2-weighted MR image examples and corresponding iGNet predictions across the four independent datasets. Saliency maps highlight regions that influenced the performance of the model, with colour-coded voxel predictions for GEs (red) and NGGCTs (green). **C** Comparison of the performance of iGNet against a model that integrates conventional clinical information with iGNet outputs, as well as a DL model using multimodal MR images. **D** Subgroup performance metrics for iGNet, presented as the accuracy, specificity, sensitivity, and AUC values, alongside their 95% CIs from bootstrap analysis ($N=2000$ replicates). Sensitivity and specificity were calculated at a threshold matched to average reader sensitivity. The frequency of GEs and NGGCTs per subgroup is visualized with bar plots. The full numerical values for each subgroup are available in Table S4

$P=0.676$) in the retrospective internal test dataset (Additional file 1: Fig. S5B). In terms of tumour location, iGNet achieved an AUC of 0.936 (95% CI: 0.843–1.000) for the sellar/suprasellar region, 0.929 (95% CI: 0.875–0.983) for the pineal region, and 0.617 (95% CI: 0.318–0.864) for the basal ganglia region (Table 2, Additional file 1: Fig. S5 C, and Table S5). Moreover, Fig. 2D illustrates the performance of iGNet in subgroup analyses within the retrospective internal test dataset, showing results consistent with the primary, nonstratified findings. The results of the DeLong test comparing the performance

of neuroradiologists without iGNet assistance with those with iGNet assistance are presented in Additional file 1: Table S6.

Improvement in neuroradiologists’ diagnostic performance with the assistance of the DL model

Without iGNet assistance, the average AUC values of the six neuroradiologists (T.T.H., L.N.W., Z.W., Y.Y.D., L.X.H., and M.W.Z.) were 0.717, 0.679, 0.716, 0.778, 0.817, and 0.801, respectively, across the retrospective test dataset and three independent test datasets. When the output

Table 2 Performance of the DL model in the retrospective internal test dataset, prospective internal test dataset, external test dataset-1, and external test dataset-2; evaluation of the DL model after combining iGNet with other information; performance of the DL model developed with multimodal MR images; and performance of iGNet for different tumour locations

| | ACC (% (95% CI) ^a) | Specificity (% (95% CI)) | Sensitivity (% (95% CI)) | AUC (95% CI) |
|---|--------------------------------|--------------------------|--------------------------|---------------------|
| iGNet performance | | | | |
| Retrospective internal dataset | 90.91 (84.85–95.96) | 78.40 (64.10–90.60) | 98.39 (94.74–100.00) | 0.950 (0.863–0.994) |
| Prospective internal dataset | 87.10 (68.00–96.00) | 84.64 (67.42–100.00) | 76.47 (53.85–94.12) | 0.921 (0.657–0.974) |
| External dataset-1 | 80.00 (60.00–95.00) | 88.24 (70.59–100.00) | 78.33 (56.65–95.14) | 0.869 (0.757–0.941) |
| External dataset-2 | 89.56 (71.22–100.00) | 76.47 (53.85–94.12) | 89.79 (76.84–98.62) | 0.905 (0.774–0.971) |
| DL model evaluation | | | | |
| Combination of clinical information and iGNet in the retrospective internal dataset | 92.91 (84.85–95.96) | 78.38 (64.29–90.70) | 98.39 (94.83–100.00) | 0.963 (0.921–0.990) |
| DL model developed using multimodal MR images (T2W-, T1W-, and CE T1W-images) in the retrospective internal dataset | 87.50 (80.00–93.75) | 83.87 (69.57–95.65) | 89.80 (80.85–97.73) | 0.901 (0.800–0.938) |
| DL performance for tumours with different locations in the retrospective internal dataset | | | | |
| Sellar/suprasellar region | 86.84 (76.32–97.37) | 85.19 (70.37–96.55) | 90.91 (70.00–100.00) | 0.936 (0.843–1.000) |
| Pineal region | 87.80 (80.49–93.90) | 78.57 (62.50–92.59) | 92.59 (84.91–98.31) | 0.929 (0.875–0.983) |
| Basal ganglia region | 61.11 (38.89–83.33) | 77.78 (50.00–100.00) | 44.44 (11.11–80.00) | 0.617 (0.318–0.864) |

ACC accuracy, AUC area under the curve

^aWide CIs typically suggest a small sample size or high variability in the estimates

of iGNet was used, the average AUC values increased to 0.841, 0.827, 0.825, 0.900, 0.926, and 0.899 across the four datasets, with improvements of 14.74%, 17.90%, 12.82%, 13.56%, 9.22%, and 10.90%, respectively (Fig. 3A and Additional file 1: Table S6 provide further details).

Regarding diagnostic consistency, the Kappa coefficients ranged from 0.472 to 0.680 across the four datasets for the six neuroradiologists when not referring to the iGNet output and increased to 0.697 to 0.922 when the output of iGNet was consulted. This represents an absolute improvement from 10.90 to 40.49% in diagnostic consistency (Fig. 3B and Additional file 1: Table S7). Furthermore, we assessed the six neuroradiologists' reliance on the iGNet outputs, and the results are presented in Additional file 1: Fig. S6.

Comparable prognostic value of the DL model for pathological diagnosis

For the iGNet output, the hazard ratio (HR) for the 5-year PFS was 1.06 (95% CI: 0.85–1.57, $P=0.061$), and that for the 5-year OS was 1.21 (95% CI: 0.87–1.62, $P=0.035$) in the retrospective internal test dataset (Fig. 3C). For the pathological diagnosis, the HR for 5-year PFS was 1.09 (95% CI: 0.83–1.42, $P=0.057$), and the HR for 5-year OS was 1.33 (95% CI: 0.78–1.58, $P=0.019$) in the same dataset (Fig. 3D). The log-rank test indicated that the prognostic utility of iGNet in the prediction of PFS and OS was comparable to that of the pathological diagnosis, with a P value of 0.889.

Finally, the output results of three real cases generated by iGNet in the retrospective internal test dataset and their pathological verification results are shown in Fig. 4A–C.

Discussion

In this study, we developed and validated an automated DL model to differentiate GEs from NGGCTs. This model demonstrated considerable generalizability across multiple centres and high robustness in various clinical scenarios. Furthermore, it was capable of helping neuro-radiologists significantly improve their clinical diagnostic accuracy and assisting in evaluating patient prognosis, showing comparable performance to pathology.

A significant advantage of iGNet is its ability to simultaneously complete iGCT segmentation and differentiation. The employs a Hough voting-based approach [31, 32], facilitating precise tumour localization and segmentation areas by aggregating feature votes within the neural network architecture, as detailed in the aforementioned references. This strategy allows automatic localization and segmentation of the anatomies of interest, differing from models in previous studies in which image segmentation and differentiation were typically divided into separate steps. Using the outcomes of 3D nnU-Net-based [23] classification, our approach exploits the features from the deepest network layers [33–35], which extract more detailed imaging information than manually selected features such as conventional and quantitative imaging features, as described in previous studies [16, 18]. The performance of iGNet

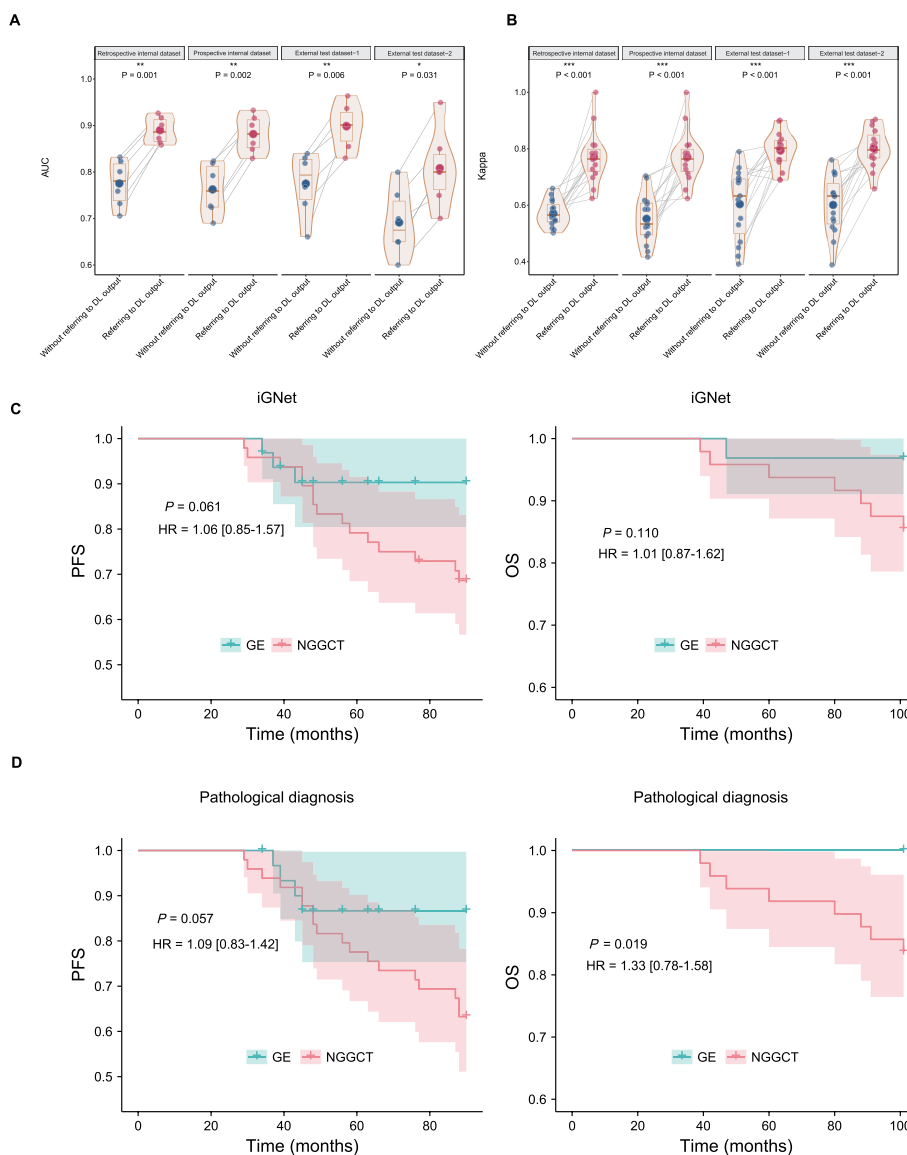


Fig. 3 **A** Enhancement of neuroradiological diagnosis with iGNet assistance. Improvement in the average AUCs for the neuroradiologists’ diagnoses was observed for the retrospective internal test and independent test datasets upon referencing the output of iGNet, with significant percentage increases ($P < .05$). **B** Kappa coefficient enhancements indicating improved diagnostic consistency among neuroradiologists utilizing iGNet, with specific percentage improvements noted ($P < .001$). **C** Kaplan–Meier plots for PFS and OS categorized by iGNet’s MR-based predictions. **D** Kaplan–Meier plots for PFS and OS categorized by pathological diagnosis

(AUC of 0.950 in the test dataset) indicated that features extracted from T2W images alone could effectively differentiate GEs from NGGCTs, which is consistent with previous findings using advanced imaging techniques (including diffusion and perfusion imaging) [16, 18]. The network’s robustness was confirmed in multicentre datasets, representing a crucial step towards understanding the generalizability and clinical value of the neural network, which was lacking in previous single-centre studies [36–39].

Grad-CAM [27] enhances the interpretability of DL models by generating saliency maps that highlight regions of interest, helping neuroradiologists understand the decision-making process of the model. Additionally, on MR images, Grad-CAM precisely localizes tumours by highlighting potential lesion areas, thereby assisting neuroradiologists in making more accurate diagnoses and assessments, ultimately increasing the credibility and acceptability of the diagnosis of the model. We also performed a comprehensive analysis of the discrepancies

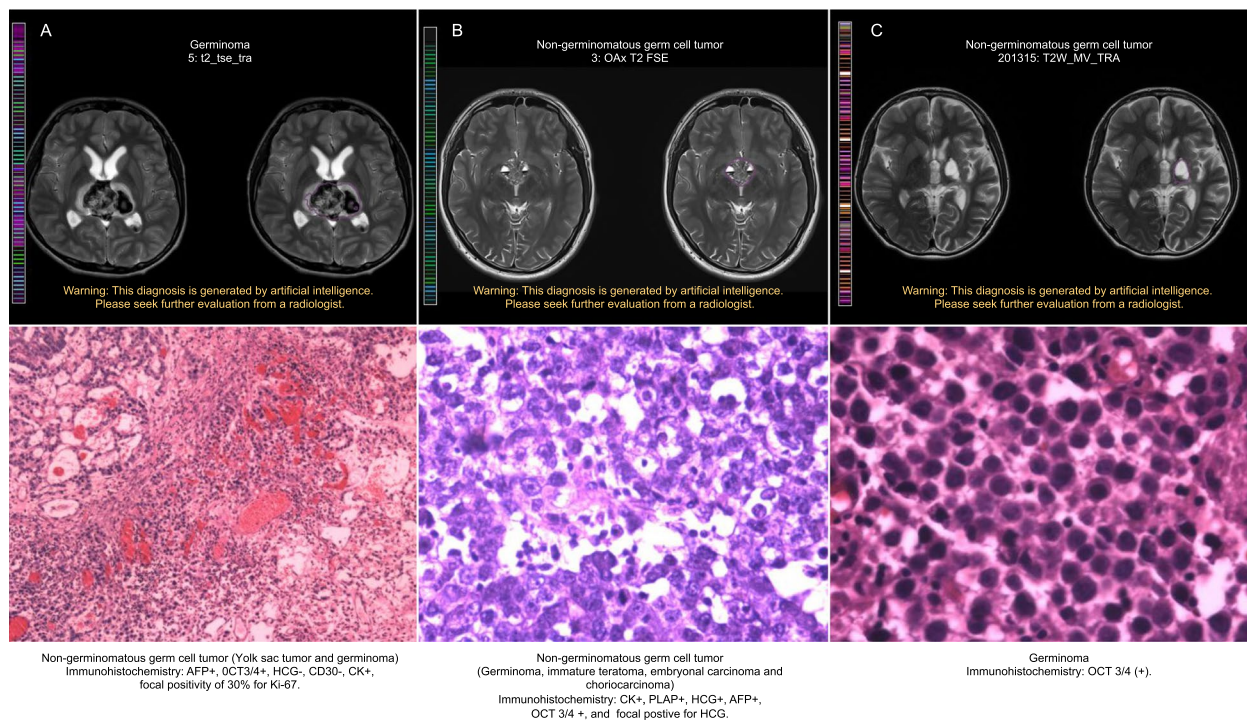


Fig. 4 **A** A 20-year-old male who experienced headaches for 3 months presented with a pineal lesion. In October 2020, his serum AFP level was 8.69 ng/ml (normal level: < 7 ng/ml). By November 2020, his serum AFP had increased to 10.05 ng/ml. Additionally, in November 2020, his CSF AFP level was 7.11 ng/ml, which decreased to < 0.5 ng/ml following one cycle of chemotherapy. **B** An 8-year-old female suffering from headaches and diminished vision for 5 months presented with a suprasellar lesion. In May 2021, her serum AFP was significantly elevated, at 353.1 ng/ml. After two cycles of chemotherapy, her serum AFP had decreased to 266.4 ng/ml, and her CSF AFP level was < 0.5 ng/ml by May 2021. **C** An 11-year-old male who had been experiencing weakness in one limb for 11 months presented with a serum AFP level of 1.53 ng/ml and a CSF AFP level of < 0.5 ng/ml in February 2019. After four cycles of chemotherapy, his serum AFP level was increased to 64.2 ng/ml by June 2019

between iGNet's predictions and the ground truth, as detailed in Additional file 1: Fig. S7A. Approximately 50% of the iGNet diagnostic errors coincided with those of the logistic regression model or the neuroradiologists. A review of some MR images revealed that these discrepancies were due primarily to atypical tumour presentations and inadequate tumour segmentation (Additional file 1: Fig. S7B). The pathological type commonly involved was NGGCT mixed with GE and mature teratoma. By examining these cases, we aimed to identify potential areas for model improvement and gain a deeper understanding of the limitations and strengths of iGNet. This detailed analysis offers valuable insights into specific scenarios where iGNet's performance could be enhanced and underscores the importance of ongoing validation and refinement of the model.

When the DL model was evaluated with conventional information alongside its output, the AUC was 0.963, reflecting iGNet's standalone performance (0.963 vs. 0.950, $P=0.087$) in the test dataset. This finding suggests that the DL model alone can provide reliable and precise clinical diagnostic results. When unimodal and

multimodal MR data inputs were compared, the DL models trained on T2W images and multimodal MR images showed similar performances, indicating that additional MR sequences did not significantly enhance the model's effectiveness over T2W images alone. Considering that a standard MR scan [40] usually includes T1W and CE-T1W sequences, prioritizing T2W imaging is clinically advantageous, especially for patients such as restless children or those receiving platinum-based chemotherapy (platinum-based chemotherapeutic agents induce nephrotoxicity and exhibit increased sensitivity to contrast agents), who have limited tolerance for extended scans [41]. Thus, T2W imaging as the primary MR sequence is clinically beneficial, enabling quicker adoption, eliminating the need for contrast agents, reducing the scan duration to under 3 min, and still providing high-resolution tumour visualization [42].

In the stratified subgroup analysis, we focused on the model's performance across different iGCT locations because of the known bias associated with the tumour site. For example, NGGCT is more common in the pineal region, and GE is more common in the basal ganglia [5].

iGNet yielded AUCs of 0.936, 0.929, and 0.617 in the sellar/suprasellar region, pineal region, and basal ganglia region, respectively, indicating variable performance across tumour locations, with lower sensitivity in the basal ganglia region than in other regions (e.g. the sellar/suprasellar and pineal regions). The heterogeneous imaging features of iGCTs in the basal ganglia present challenges and often require more attention and manual correction [43, 44] (Additional file 1: Fig. S8A and B).

iGNet significantly improved the neuroradiologists' diagnostic performance by over 18% and consistency by up to 40%, confirming its reliability and value for neuroradiologists in clinical settings. It was particularly effective in reducing diagnostic discrepancies in rare paediatric brain tumours such as iGCTs, thereby enhancing clinicians' precision and efficiency. A comparative analysis between iGNet and neuroradiologists' diagnoses revealed that less experienced doctors leaned more on the DL model, whereas experienced doctors relied more on their expertise. However, the results of this study suggest that junior doctors should prioritize gaining clinical experience over solely relying on artificial intelligence for diagnosing relatively rare tumours.

Pathological analyses provide insights into iGCT tumour type, grade, and immunohistochemical characteristics, all of which are correlated with PFS and OS [11]. If a DL- and MR-based diagnostic tool can accurately predict the pathological type, artificial intelligence-enhanced MR images could potentially offer prognostic information similar to traditional pathology. In our study, the prognostic predictions from iGNet demonstrated a strong correlation with those from pathological evaluations, suggesting that iGNet may have potential as a supplementary prognostic tool. The ability of iGNet's output to predict PFS and OS indicates that the model can assist in monitoring treatment response and adjusting therapeutic strategies. For example, a low likelihood of disease progression predicted by iGNet might allow extended intervals between follow-up imaging scans, releasing patient stress and exposure to unnecessary procedures. Conversely, a high risk of progression would prompt more frequent monitoring and timely adjustments to the treatment regimen. By integrating these predictions into routine clinical practice [9, 45], clinicians can potentially tailor treatment plans more effectively, optimize resource allocation, and ultimately improve patient care and outcomes [6].

This study has certain limitations. First, there is potential methodological bias in the comparison of neuroradiologists' diagnostic accuracy. Despite the three-month interval between assessments, the neuroradiologists were exposed to the same images during both diagnoses with and without the assistance from iGNet. Second, iGNet

performed effectively for tumours in the sellar/suprasellar and pineal areas, but its effectiveness was less desirable for tumours in the basal ganglia region, likely due to the limited sample size. Therefore, more tumours in the basal ganglia regions should be included in further studies to improve the model's performance.

Conclusions

In summary, the developed iGNet demonstrated notable performance in the differentiation of iGCTs based on pretherapeutic MR images, as evidenced by its high AUC in identifying GEs and NGGCTs. Given that its predictive capacity closely aligned with that of pathological findings, iGNet promises to be a noninvasive tool that could be employed in parallel with traditional biopsy analyses.

Abbreviations

| | |
|--------|--|
| iGCTs | Intracranial germ cell tumours |
| GEs | Germinomas |
| NGGCTs | Nongerminomatous germ cell tumours |
| AUC | Area under the receiver operating characteristic curve |
| DL | Deep learning |
| PFS | Progression-free survival |
| OS | Overall survival |

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-024-03575-w>.

Additional file 1: Method S1. The pathological diagnosis of GEs and NGGCTs. Method S2. DL model architecture. Method S3. DL model training and testing. Method S4. Additional details of DL model evaluation. Method S5. Additional details of statistical analysis. Table S1. Image acquisition parameters of 2D T2-weighted turbo-spin-echo sequence. Table S2. The outputs of five folds in the retrospective internal test dataset. Table S3. The results of DeLong test comparing multivariate logistic regression vs. iGNet. Table S4. Output of DL model trained on multimodal MR images in different datasets. Table S5. Assessing the performance of the DL model in stratified datasets based on clinical features. Table S6. The AUC value, the improvement rate, and the DeLong test results of six neuroradiologists with and without referring to the iGNet output in different datasets. Table S7. The Kappa coefficients and the improvement rate of six neuroradiologists with and without referring to the DL model output in different datasets. Fig. S1. The scanner information of the development set. Fig. S2. The training phase and testing phase of the DL model. Fig. S3. Details of the 3D nnU-Net for the tumour segmentation model. Fig. S4. Details of the 3D nnU-Net for the tumour differentiation model. Fig. S5. A. The receiver operating characteristic curves using demographic data, tumour marker levels, conventional MR features, a combination of clinical information and iGNet output, and the model performance. B. The ROC curves of the DL model using T2W, T1W, CE-T1W images, and T2W images. C. The iGNet performance on tumours at different locations. Fig. S6. The degree of reliance on the iGNet outputs of six neuroradiologists. Fig. S7. The analysis of discrepant predictions. A. A comprehensive analysis of the discrepancies between iGNet and other models' predictions and the ground truth. B. A representative case of iGNet diagnosis error involves Patient 1, who had NGGCT but was misclassified by iGNet as GE. Upon reviewing the patient's MRI, we found that the error was related to atypical imaging findings; T2-weighted imaging showed heterogeneous intensity with a soap bubble appearance, which is consistent with GE MR characteristics, though pathological results indicated the presence of a small amount of mature teratoma components. In the case of Patient 2, who was diagnosed with

GE but also misclassified by iGNet, we found that the tumour segmentation results from iGNet did not correspond to the actual tumour location, which likely contributed to the incorrect classification. Fig. S8. A. Distinct MR imaging findings of intracranial germ cell tumours in the basal ganglia. A. Patient 1 was diagnosed with germinoma via biopsy, with T2W images revealing a patchy hyperintense lesion with blurred boundaries. B. Patient 2 was diagnosed with a non-germinomatous germ cell tumour via resection, with T1W images showing a heterogeneous-intensity tumour in the left basal ganglia, accompanied by necrosis and fluid-fluid level. Contrast-enhanced T1W images highlighted patchy hyperintensity in the tumour with obvious enhancement

Acknowledgements

We extend our acknowledgment to all colleagues who offered technical guidance and constructive feedback on the study (X.K., T.T.H., and Y.Y.D. assisted with tumor segmentation mask delineation and interpretation of conventional MR images. Y.J.H. and J.D. confirmed the pathological diagnosis. L.X.H., L.N.W. evaluated the conventional MRI features. T.S. and M.G. corrected the grammar.) as well as those who provided invaluable assistance with patient recruitment and MR imaging.

Authors' contributions

Y.N.L.: Data curation, Writing-original draft, and Visualization. Z.Z.Z.: Methodology and Review & editing. J.Y.W.: Software, Validation, Investigation, and Formal analysis. S.H., H. X. B., and B. L.: Methodology and Review & editing. Xing Liu, Z. W., and M. Z.: Reader study participant. J. L.: Writing-review & editing. Y.O.L. and X.G.Q.: Conceptualization, Data curation, Writing-review & editing, and Supervision. All authors reviewed the manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (81870958).

Availability of data and materials

De-identified individual participant data (including data dictionaries) will be made available, along with the study protocols, the statistical analysis plan, and the informed consent form, upon publication to researchers who provide a methodologically sound proposal for use in achieving the objectives of the approved proposal. Proposals should be submitted to liyanong@bjtth.org. Codes are available via <https://github.com/YanongL/DL-model-code-for-differentiation-of-GE-from-NGGCT.git>.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

The Institutional Review Board of Beijing Tiantan Hospital, Capital Medical University, approved this study (KY2021-142-02). Written informed consent for the publication of clinical details and clinical images was obtained from all the patients (or their legal guardians) involved in this study.

Consent for publication

The MR images are completely unidentifiable, and no individual details are reported within the manuscript.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Radiation Oncology, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China. ²Department of Radiology, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China. ³UCL Institutes of Neurology and Healthcare Engineering, London WC1E 6BT, UK. ⁴Department of Radiology and Radiological Science, Johns Hopkins University School of Medicine, Baltimore, MD 21287, USA. ⁵Department of Neuropathology, Beijing Neurosurgery Institute, Beijing 100070, China. ⁶Department of Radiology,

Beijing Sanbo Hospital, Capital Medical University, Beijing 100093, China.

⁷Department of Radiation Oncology, Tianjin Huanhu Hospital, Tianjin Medical University, Tianjin 300350, China. ⁸Department of Radiology, New York Presbyterian, Lower Manhattan Hospital, New York, NY 10038, USA.

Received: 7 May 2024 Accepted: 20 August 2024

Published online: 11 September 2024

References

- Gittleman H, Cioffi G, Vecchione-Koval T, Ostrom QT, Kruchko C, Osorio DS, Finlay JL, Barnholtz-Sloan JS. Descriptive epidemiology of germ cell tumors of the central nervous system diagnosed in the United States from 2006 to 2015. *J Neurooncol*. 2019;143(2):251–60.
- Lee SH, Jung KW, Ha J, Oh CM, Kim H, Park HJ, Yoo H, Won YJ. Nationwide population-based incidence and survival rates of malignant central nervous system germ cell tumors in Korea, 2005–2012. *Cancer Res Treat*. 2017;49(2):494–501.
- McCarthy BJ, Shibui S, Kayama T, Miyaoka E, Narita Y, Murakami M, Matsuda A, Matsuda T, Sobue T, Palis BE, et al. Primary CNS germ cell tumors in Japan and the United States: an analysis of 4 tumor registries. *Neuro Oncol*. 2012;14(9):1194–200.
- Murray MJ, Bartels U, Nishikawa R, Fangusaro J, Matsutani M, Nicholson JC. Consensus on the management of intracranial germ-cell tumours. *Lancet Oncol*. 2015;16(9):e470–7.
- Packer RJ, Cohen BH, Cooney K. Intracranial germ cell tumors. *Oncologist*. 2000;5(4):312–20.
- Frappaz D, Dhall G, Murray MJ, Goldman S, Faure Conter C, Allen J, Kortmann RD, Haas-Kogen D, Morana G, Finlay J, 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*. 2022;24(4):516–27.
- Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, Hawkins C, Ng HK, Pfister SM, Reifenberger G, et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro Oncol*. 2021;23:1231–51.
- Kreutz J, Rausin L, Weerts E, Tebache M, Born J, Hoyoux C. Intracranial germ cell tumor. *JBR-BTR*. 2010;93(4):196–7.
- Lo AC, Hodgson D, Dang J, Tyldesley S, Bouffet E, Bartels U, Cheng S, Hukin J, Bedard PL, Goddard K, et al. Intracranial germ cell tumors in adolescents and young adults: a 40-year multi-institutional review of outcomes. *Int J Radiat Oncol Biol Phys*. 2020;106(2):269–78.
- Hong KT, Han JW, Fuji H, Byun HK, Koh KN, Wong RX, Lee HL, Yoon HI, Lee JH, Phi JH, et al. Outcomes of intracranial non-germinomatous germ cell tumors: a retrospective Asian multinational study on treatment strategies and prognostic factors. *J Neurooncol*. 2022;160(1):41–53.
- Takami H, Fukuoka K, Fukushima S, Nakamura T, Mukasa A, Saito N, Yanagisawa T, Nakamura H, Sugiyama K, Kanamori M, et al. Integrated clinical, histopathological, and molecular data analysis of 190 central nervous system germ cell tumors from the iGCT Consortium. *Neuro Oncol*. 2019;21(12):1565–77.
- Finlay J, da Silva NS, Lavey R, Bouffet E, Kellie SJ, Shaw E, Saran F, Matsutani M. The management of patients with primary central nervous system (CNS) germinoma: current controversies requiring resolution. *Pediatr Blood Cancer*. 2008;51(2):313–6.
- Alapetite C, Brisse H, Patte C, Raquin MA, Gaboriaud G, Carrie C, Habrand JL, Thiesse P, Cuilliere JC, Bernier V, et al. Pattern of relapse and outcome of non-metastatic germinoma patients treated with chemotherapy and limited field radiation: the SFOP experience. *Neuro Oncol*. 2010;12(12):1318–25.
- Sonehara K, Kimura Y, Nakano Y, Ozawa T, Takahashi M, Suzuki K, Fujii T, Matsushita Y, Tomiyama A, Kishikawa T, et al. A common deletion at BAK1 reduces enhancer activity and confers risk of intracranial germ cell tumors. *Nat Commun*. 2022;13(1):4478.
- Huang X, Zhang R, Mao Y, Zhou LF, Zhang C. Recent advances in molecular biology and treatment strategies for intracranial germ cell tumors. *World J Pediatr*. 2016;12(3):275–82.
- Li Y, Wang P, Zhang J, Li J, Chen L, Qiu X. Multiparametric framework magnetic resonance imaging assessment of subtypes of intracranial germ cell tumors using susceptibility weighted imaging, diffusion-weighted

- imaging, and dynamic susceptibility-contrast perfusion-weighted imaging combined with conventional magnetic resonance imaging. *J Magn Reson Imaging*. 2022;56(4):1232–42.
17. Calaminus G, Kortmann R, Worch J, Nicholson JC, Alapetite C, Garre ML, Patte C, Ricardi U, Saran F, Frappaz D. SIOP CNS GCT 96: final report of outcome of a prospective, multinational nonrandomized trial for children and adults with intracranial germinoma, comparing craniospinal irradiation alone with chemotherapy followed by focal primary site irradiation for patients with localized disease. *Neuro Oncol*. 2013;15(6):788–96.
 18. Wu CC, Guo WY, Chang FC, Luo CB, Lee HJ, Chen YW, Lee YY, Wong TT. MRI features of pediatric intracranial germ cell tumor subtypes. *J Neurooncol*. 2017;134(1):221–30.
 19. Ye N, Yang Q, Chen Z, Teng C, Liu P, Liu X, Xiong Y, Lin X, Li S, Li X. Classification of gliomas and germinomas of the basal ganglia by transfer learning. *Front Oncol*. 2022;12:844197.
 20. Ye N, Yang Q, Liu P, Chen Z, Li X. A comprehensive machine-learning model applied to MRI to classify germinomas of the pineal region. *Comput Biol Med*. 2023;152:106366.
 21. Supbumrung S, Kaewborisutsakul A, Tunthanathip T. Machine learning-based classification of pineal germinoma from magnetic resonance imaging. *World Neurosurg* X. 2023;20:100231.
 22. Biermann K, Klingmuller D, Koch A, Pietsch T, Schorle H, Buttner R, Zhou H. Diagnostic value of markers M2A, OCT3/4, AP-2gamma, PLAP and c-KIT in the detection of extragonadal seminomas. *Histopathology*. 2006;49(3):290–7.
 23. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203–11.
 24. Savjani R. nnU-Net: further automating biomedical image autosegmentation. *Radiol Imaging Cancer*. 2021;3(1):e209039.
 25. Randhawa K, Loo CK, Seera M, Lim CP, Nandi AK. Credit card fraud detection using AdaBoost and majority voting. *IEEE Access*. 2018;6:14277–84.
 26. Sunnetci KM, Alkan A. Biphasic majority voting-based comparative COVID-19 diagnosis using chest X-ray images. *Expert Syst Appl*. 2023;216:119430.
 27. Font-Clos F, Zanchi M, Hiemer S, Bonfanti S, Guerra R, Zaiser M, Zapperi S. Predicting the failure of two-dimensional silica glasses. *Nat Commun*. 2022;13(1):2820.
 28. Allen J, Chacko J, Donahue B, Dhall G, Kretschmar C, Jakacki R, Holmes E, Pollack I. Diagnostic sensitivity of serum and lumbar CSF bHCG in newly diagnosed CNS germinoma. *Pediatr Blood Cancer*. 2012;59(7):1180–2.
 29. Matsutani M, Japanese Pediatric Brain Tumor Study G. Combined chemotherapy and radiation therapy for CNS germ cell tumors—the Japanese experience. *J Neurooncol*. 2001;54(3):311–6.
 30. Kortmann RD. Current concepts and future strategies in the management of intracranial germinoma. *Expert Rev Anticancer Ther*. 2014;14(1):105–19.
 31. Samet N, Hicsonmez S, Akbas E. HoughNet: integrating near and long-range evidence for visual detection. *IEEE Trans Pattern Anal Mach Intell*. 2023;45(4):4667–81.
 32. Gall J, Yao A, Razavi N, Van Gool L, Lempitsky V. Hough forests for object detection, tracking, and action recognition. *IEEE Trans Pattern Anal Mach Intell*. 2011;33(11):2188–202.
 33. Havaii M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin PM, Larochelle H. Brain tumor segmentation with deep neural networks. *Med Image Anal*. 2017;35:18–31.
 34. Prastawa M, Bullitt E, Ho S, Gerig G. A brain tumor segmentation framework based on outlier detection. *Med Image Anal*. 2004;8(3):275–83.
 35. Konar D, Bhattacharyya S, Gandhi TK, Panigrahi BK, Jiang R. 3-D quantum-inspired self-supervised tensor network for volumetric segmentation of medical images. *IEEE Trans Neural Netw Learn Syst*. 2023;35(8):10312–25.
 36. Falk T, Mai D, Bensch R, Cicek O, Abdulkadir A, Marrakchi Y, Bohm A, Deubner J, Jackel Z, Seiwald K, et al. U-Net: deep learning for cell counting, detection, and morphometry. *Nat Methods*. 2019;16(1):67–70.
 37. Yogananda CGB, Shah BR, Nalawade SS, Murugesan GK, Yu FF, Pinho MC, Wagner BC, Mickey B, Patel TR, Fei B, et al. MRI-based deep-learning method for determining glioma MGMT promoter methylation status. *AJNR Am J Neuroradiol*. 2021;42(5):845–52.
 38. Zhang G, Yang Z, Huo B, Chai S, Jiang S. Multiorgan segmentation from partially labeled datasets with conditional nnU-Net. *Comput Biol Med*. 2021;136:104658.
 39. van der Velden BHM, Kuijff HJ, Gilhuijs KGA, Viergever MA. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med Image Anal*. 2022;79:102470.
 40. Khajehim M, Christen T, Tam F, Graham SJ. Streamlined magnetic resonance fingerprinting: fast whole-brain coverage with deep-learning based parameter estimation. *Neuroimage*. 2021;238:118237.
 41. Cheng P, Liu H, Li Y, Pi P, Jiang Y, Zang S, Li X, Fu A, Ren X, Xu J, et al. Inhibition of thioredoxin reductase 1 correlates with platinum-based chemotherapeutic induced tissue injury. *Biochem Pharmacol*. 2020;175:113873.
 42. Langen KJ, Galldiks N, Hattingen E, Shah NJ. Advances in neuro-oncology imaging. *Nat Rev Neurol*. 2017;13(5):279–89.
 43. Zhang Y, Wang L, Ma W, Pan H, Wang R, Zhu H, Yao Y. Basal ganglia germ cell tumors with or without sellar involvement: a long-term follow-up in a single medical center and a systematic literature review. *Front Endocrinol (Lausanne)*. 2021;12:763609.
 44. Chiba K, Aihara Y, Kawamata T. Precise detection of the germinomatous component of intracranial germ cell tumors of the basal ganglia and thalamus using placental alkaline phosphatase in cerebrospinal fluid. *J Neurooncol*. 2021;152(2):405–13.
 45. Khatua S, Dhall G, O'Neil S, Jubran R, Villablanca JG, Marachelian A, Nastia A, Lavey R, Olch AJ, Gonzalez I, et al. Treatment of primary CNS germinomatous germ cell tumors with chemotherapy prior to reduced dose whole ventricular and local boost irradiation. *Pediatr Blood Cancer*. 2010;55(1):42–6.

Publisher's Note

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