Child's Nervous System
https://doi.org/10.1007/s00381-024-06400-0

RESEARCH

Diagnostic support in pediatric craniopharyngioma using deep learning

Giovanni Castiglioni[4,5] · Joaquín Vallejos[1] · Jhon Intriago[4,5] · María Isabel Hernández[2,3] · Samuel Valenzuela[2] · José Fernández[2] · Ignacio Castro[6] · Sergio Valenzuela[2] · Pablo A. Estévez[4,5] · Cecilia Okuma[1,3]

Received: 11 March 2024 / Accepted: 8 April 2024

Abstract

Purpose We studied a pediatric group of patients with sellar-suprasellar tumors, aiming to develop a convolutional deep learning algorithm for radiological assistance to classify them into their respective cohort.

Methods T1w and T2w preoperative magnetic resonance images of 226 Chilean patients were collected at the Institute of Neurosurgery Dr. Alfonso Asenjo (INCA), which were divided into three classes: healthy control (68 subjects), craniopharyngioma (58 subjects) and differential sellar/suprasellar tumors (100 subjects).

Results The PPV among classes was 0.828±0.039, and the NPV was 0.919±0.063. Also explainable artificial intelligence (XAI) was used, finding that structures that are relevant during diagnosis and radiological evaluation highly influence the decision-making process of the machine.

Conclusion This is the first experience of this kind of study in our institution, and it led to promising results on the task of radiological diagnostic support based on explainable artificial intelligence (AI) and deep learning models.

Keywords Craniopharyngioma · MRI · Deep learning · Classification

Introduction

Craniopharyngiomas are congenital, low-grade (WHO grade I) epithelial brain tumors that affect the sellar and suprasellar regions. Its annual incidence rate is reported about 0.05–0.2 per 100,000 individuals [1] and they account for 1–4% of brain tumors in children but comprise approximately 54% of tumors in this location in children under 15 years old, without gender or race distinction [2]. Craniopharyngioma occurs in a bimodal age distribution, with peak onset ages ranging from 5 to 14 years and 50 to 74 years [3]. Five-year survival rates are high (91–95%), but due to its locally aggressive growth pattern and their tendency to infiltrate the hypothalamus, optic chiasm, and vascular structures, they are associated with a high percentage of morbidity [4, 5]. The clinical manifestations of craniopharyngiomas are diverse, depending on the tumor location, size, growth pattern, and relation with adjacent brain tissue.

The main histological subtype is adamantinomatous. These types of craniopharyngiomas, more common in children, and typically appear heterogeneous on imaging with cysts, nodules, fibrosis, calcifications, and hemorrhagic changes. The other histological variant is papillary and accounts for 10–15% of all craniopharyngiomas.

Adamantinomatous craniopharyngiomas are mostly caused by mutations in the Wingless pathway (WNT pathway) involving the CTNNB1 gene encoding β-catenin. Activating mutations in this pathway have been described in more than 2/3 of patients with adamantinomatous craniopharyngiomas [6]. The remaining cases are proposed to involve the MAPK/ERK pathway, offering the opportunity for treatment with Mek inhibitors [7]. In papillary craniopharyngiomas, BRAF V600E mutations are detected in 90% of patients [6]. Molecular subtype differentiation based on classic neuroimaging criteria is not possible [8].

✉ Cecilia Okuma
cecilia.okuma@gmail.com

1 Department of Neuroradiology, Institute of Neurosurgery Dr. Alfonso Asenjo, Santiago, Chile

2 Institute of Neurosurgery Dr. Alfonso Asenjo, Santiago, Chile

3 Department of Neurological Sciences, Faculty of Medicine, University of Chile, Santiago, Chile

4 Department of Electrical Engineering, Faculty of Physical and Mathematical Sciences, University of Chile, Santiago, Chile

5 Millennium Institute for Intelligent Healthcare Engineering, Santiago, Chile

6 Faculty of Medicine, Mayor University, Santiago, Chile

Published online: 22 April 2024

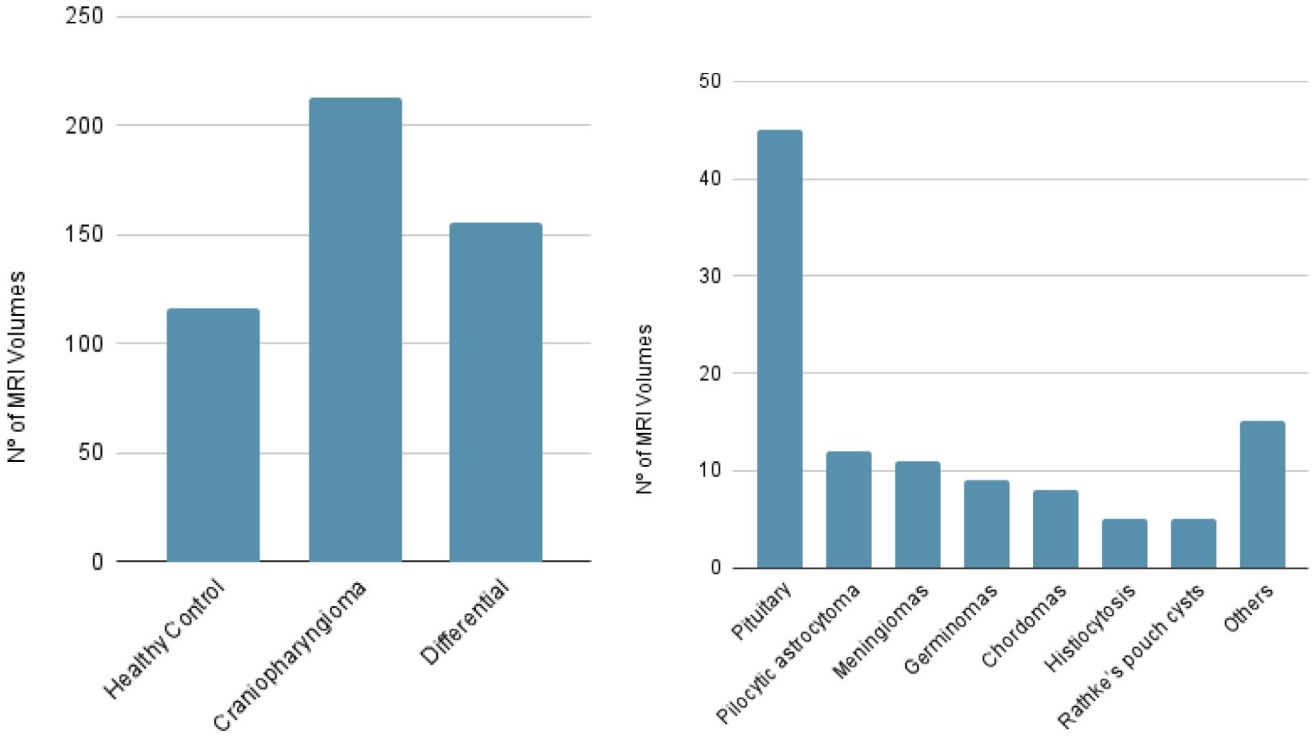
The main diagnostic tool for sellar-suprasellar tumors is magnetic resonance imaging (MRI). Computerized Tomography (CT) is also an important diagnostic tool, being more specific for calcifications. However, MRI is more sensible to determine intrinsic features of the tumors (tumor location, size, shape, composition, tumor cysts signal, enhancement pattern, pituitary stalk morphology, and internal carotid artery encasement) and surrounding regions damage (hypothalamus, pituitary gland, optic chiasm, 3rd ventricle and mammillary bodies). Thus, compared to CT, MRI can clearly display the tumor location and the anatomical relationship of adjacent tissues through multi-directional imaging. Prognosis is linked to the extent and involvement of the floor of the third ventricle, mammillary bodies, and hypothalamus.

AI has been proven useful in the radiogenomic diagnosis and prognosis of this condition [9], and other types of radiological tasks [10, 11], although there are several challenges when working with these tools within medical fields using publicly available datasets made by centers from all around the globe to handle more data, which can lead to neural networks learning non-medically relevant features to achieve the task they are being trained to solve, so for the purposes of this work, only images obtained at the Chilean organization "Institute of Neurosurgery Dr. Alfonso Asenjo" (INCA) were used, avoiding such difficulties by only considering studies from one center. This results in a trade-off between model generalizability for multi-center integration of studies, and control over the images, patient information, modalities, acquisition equipment and characteristics.

Materials and methodology

Patients

226 subjects from the INCA database were included, with studies collected between 2011 and 2023: 58 histologically confirmed craniopharyngioma (CPG) patients with 213 MRI volumes (30 females and 28 males with an average age of 14, with individual ages ranging from 2 to 72; 115 T1-weighted images and 98 T2-weighted images); 100 patients with histologically confirmed differential sellar/suprasellar tumors (45 pituitary macroadenomas, 12 pilocytic astrocytoma, 11 meningiomas, 9 germinomas, 8 chordomas, 5 histiocytosis/Rathke's pouch cysts, among others) with 155 MRI volumes (72 females and 28 males with an average age of 37, with individual ages ranging from 1 to 74; 111 T1-weighted images and 44 T2-weighted images); and 68 healthy control (HC) subjects with 116 MRI volumes (37 females and 31 males with an average age of 21, with individual ages ranging from 0 to 61; 107 T1-weighted images and 9 T2-weighted images), as shown in Fig. 1, where it is possible to see the



(a) MRI volumes per each class

(b) Number of MRI volumes per tumor included in the differential class.

Fig. 1 Histograms of MRI volumes from the dataset used to train and evaluate the classifier algorithm

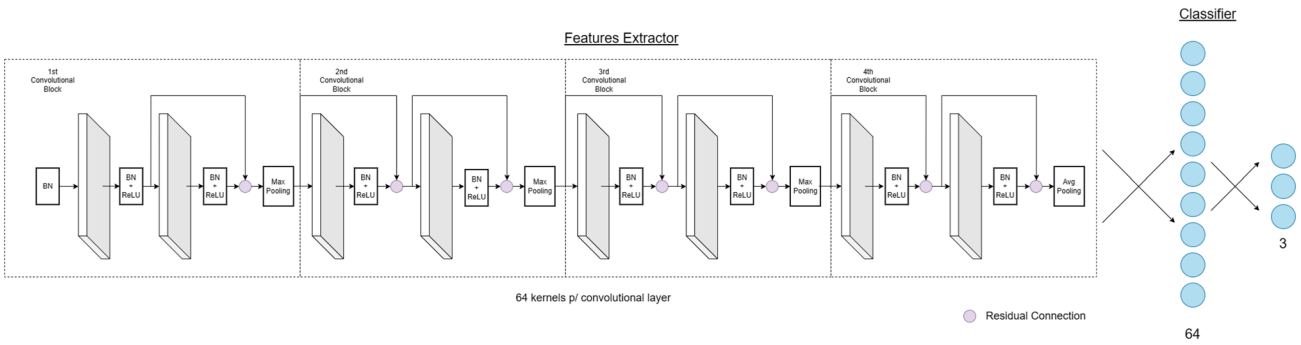


Fig. 2 CNN architecture. Eight convolutional layers were included for features extraction and a MLP was used for classification

imbalance between classes. There was agreement between the radiological evaluation and the histological study in all previous cases, meaning that for all the results that will be presented in the following section be considered that they are calculated with respect to professional appreciation, which was equivalent to ground truth.

MRI preprocessing

Every patient and control subject was studied with MRI of either whole brain or sella turcica standard protocols with contrast. Each volume was independently processed according to a commonly used standardization of volumes with dimension 256x256x256 and 1x1x1mm^3 voxels with RAS orientation. Only the sagittal views were considered (this choice was purely based on image availability due to the protocols used during the imaging acquisition, so with more examples and MRI volumes, axial and coronal views can also be included using the same methodology). Only a 16x105x105 volume was extracted from the center of the field of view (256x256x256 isotropic cube), because the sellar-suprasellar structures are placed within this space after normalization. Given the previous, it is not necessary to perform additional preprocessing methods.

Deep learning network and training

A convolutional neural network (CNN) was developed to extract features from the input images, along with a multilayer perceptron network to classify these representations of the original slices. This network receives slices from the sellar-suprasellar region of the preprocessed volumes, and classifies them into three possible classes: healthy control, craniopharyngioma and differential. The architecture of the network can be seen in Fig. 2.

After training with 80% of the total dataset (the remaining 20% of subjects were reserved as a validation set) we implemented Grad-Cam [12], an explainable AI technique that allows us to better understand the decision-making process of the machine, to see if there is a relation between the known relevant regions and structures for these kinds of tumors, and the focus regions for the neural network.

This network was trained using Focal Loss [13] as cost function, batch size of 32 and AdamW optimizer [14] using a NVIDIA GTX 1080Ti GPU with 11GB of VRAM. During the training process data augmentation was considered, including additive white gaussian noise, coarse dropout and random rotation.

Results

As for the classification task, each volume from each subject (healthy control, craniopharyngioma and differential sellar/suprasellar tumors) contained in the validation set was independently evaluated, considering the final probability distribution to be the average prediction between all the slices

Table 1 Results per subject show high sensitivity and specificity to detect craniopharyngiomas. Differential Sellar/Suprasellar Tumors (DSST), Positive Predictive Value (PPV), Negative Predictive Value (NPV). F1-Score is the harmonic mean between PPV and Recall

Metric	Healthy Control	Craniopharyngioma	DSST
PPV	0.881	0.818	0.786
NPV	0.912	1.000	0.845
Sensitivity	0.820	1.000	0.670
Specificity	0.945	0.890	0.910
F1-Score	0.848	0.900	0.721

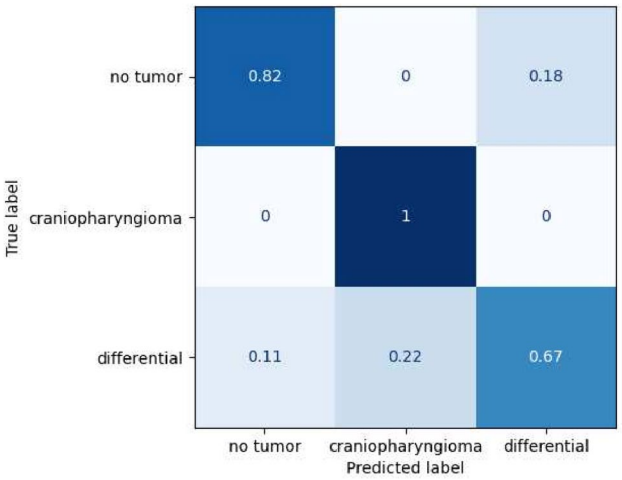


Fig. 3 Confusion matrix. It shows good performance for craniopharyngioma detection, with 100% of them identified

from them. This way it is possible to incorporate the predictions of all the slices of a subject to determine their class. The results are shown in Table 1 and Fig. 3 for Positive Predictive Value (PPV), Negative Predictive Value (NPV) and F1-Score per class. F1-Score is a metric that addresses the imbalanced properties of the classes within the dataset in the performance measurement by including PPV and Recall within its information. Recall measures the percentage of well classified positives (True Positives) among all real positive cases per class (True Positives and False Negatives). Therefore, F1-Score is defined as follows:

$$F1 = \frac{2 \times PPV \times Recall}{PPV + Recall}. \quad (1)$$

To better understand the decision-making process of the machine, we analyzed the results after applying Grad-Cam to the validation set and extracted information about regions of interest for the output layer's neurons of each class. We observed medically known relevant structures such as sella turcica, pituitary gland, optic chiasm and cisterns to be consistently considered by the machine to identify healthy subjects; for craniopharyngiomas, the tumor region was the most commonly highlighted structure; and for DSST patients, surrounding structures, such as corpus callosum, ventricles and pons, stood among the most important ones for classification. A representative example of the observed tendencies can be seen in Fig. 4.

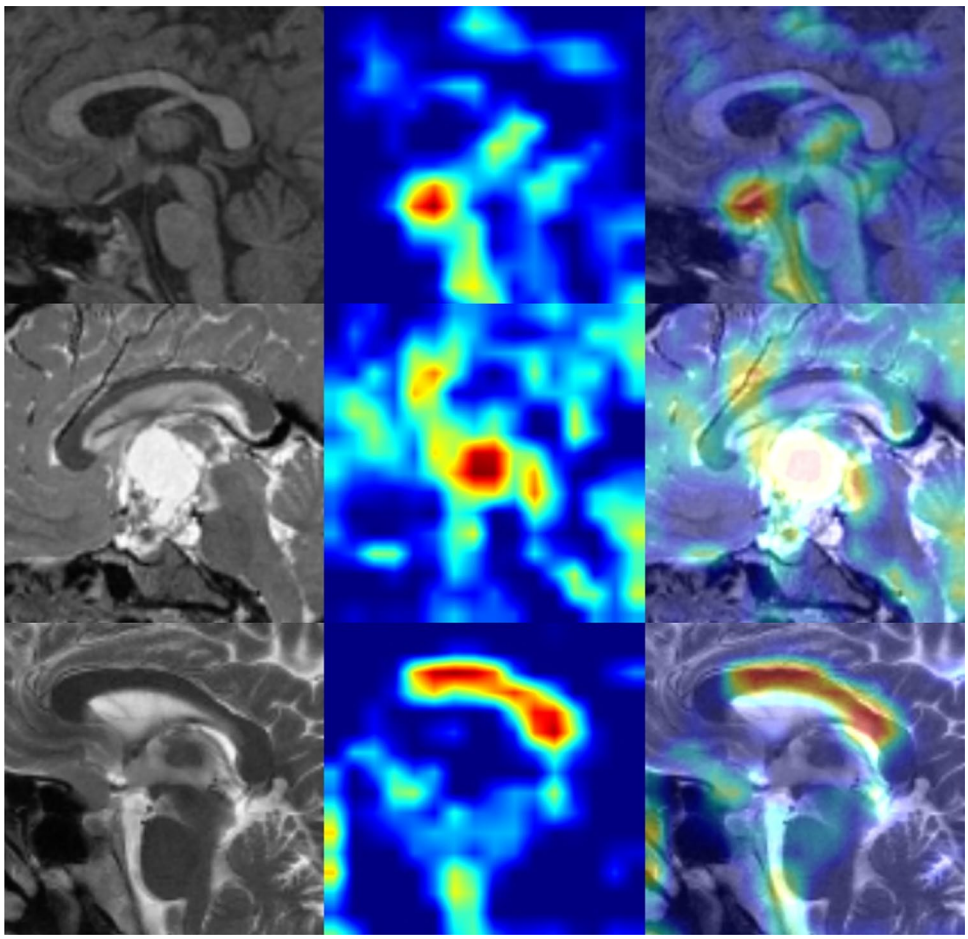


Fig. 4 Examples of regions of interest per class (healthy control, craniopharyngioma and differential respectively in descending order). Columns show the original slice, the generated heatmap, and their overlap, respectively

Discussion

The multi-sequential assessment of brain MRIs using deep learning techniques has successfully discriminated between the presence and absence of tumors in pediatric patients, given sets of slices from various studies. The discrimination between craniopharyngioma and differential tumors is correctly performed in most cases, showing a non-despicable level of confusion between both classes due to the wide variability of the differential subjects.

The use of explainable AI showed a tendency of the machine to focus on known relevant structures for the classification of craniopharyngiomas and other sellar-suprasellar tumors, which is a key and challenging aspect to validate when using black-box behaving algorithms such as neural networks. Particularly, the pituitary gland and sella turcica, along with neighboring cisterns, were important for the classification of healthy controls, whilst the tumor and its mass effect on surrounding areas were relevant to identify craniopharyngiomas. As of differentials, adjacent structures such as lateral ventricles and corpus callosum were the main aspects to focus for the machine, probably due to the wide variability of this class in the used dataset and hydrocephalus being a recurrent symptom for these kinds of tumors.

In considering the implications of our findings, this method offers a promising tool for diagnostic support, especially for medical centers facing challenges such as low patient volumes or constrained access to specialized neurosurgical and neuroradiologic resources. This underscores its potential applicability and value in contexts where such expertise may be limited, especially for their pediatric branch.

The primary limitation of our study arises from the utilization of a small database confined to a single institution, impacting the application of AI tools. Inevitably, overfitting becomes a concern, and the generalization performance is often scrutinized, particularly in high-risk tasks such as tumor diagnosis. On the one hand, an imbalanced dataset introduces variations in the costs of misclassification for both majority and minority categories, with a higher cost associated with misclassifying the minority category. It is crucial to acknowledge that these challenges represent initial explorations in the diagnosis of craniopharyngioma using this kind of technology.

This is the first research within this topic in our institution. In future work, we aim to further develop the AI methods and their applications, extending them to prognosis and morbidity inference. Other deep learning techniques, particularly self-supervised learning, will be explored using multi-center data. Thus, this work stands as a first approach and a robust supervised baseline.

Acknowledgements ANID-Chile Millennium Science Initiative Program ICN2021-004, Fondecyt 1220829, ANID / Scholarship Program / DOCTORADO BECAS CHILE / 2022 - 21221120

Author contributions G.C. wrote the main manuscript and was in charge of designing, developing and implementing the presented algorithm. J.V. studied and selected the patients for this study, contributed with radiological evaluation, and supported the research guidelines. J.I. and P.E. made substantial contributions during the development of the algorithm, and gave several feedback during the whole research and writing processes. M.I.H. studied endocrinological features of the patients and was in charge of the patient follow-up. Sa.V. and Se.V. were in charge of evaluation, surgical treatment and follow-up of the patients. J.F. made general clinical evaluations, patient follow-up and contributed with constant feedback during research. I.C. reviewed the clinical records of the patients and collected all demographic data. C.O. conducted the study with constant supervision, feedback, radiological evaluations, ideas, and made substantial contributions to the writing of this paper.

Funding G.C., J.I., and P.E. acknowledge financial support from ANID-Chile through the Millennium Science Initiative Program ICN2021-004 and Fondecyt 1220829. J.I. was supported by ANID / Scholarship Program / DOCTORADO BECAS CHILE / 2022 - 21221120

Data availability The data used in this study was collected in the "Institute of Neurosurgery Dr. Alfonso Asenjo" located in Chile.

Declarations

Conflict of interest The authors declare no competing interests.

References

1. Erfurth EM, Holmer H, Fjalldal SB (2012) Mortality and morbidity in adult craniopharyngioma. Pituitary. https://doi.org/10.1007/s11102-012-0428-2
2. Miller KD, Ostrom QT, Kruchko C, Patil N, Tihan T, Cioffi G, Fuchs HE, Waite KA, Jemal A, Siegel RL, Barnholtz-Sloan JS (2021) Brain and other central nervous system tumor statistics, 2021. CA A Cancer J Clinicians. https://doi.org/10.3322/caac.21693
3. Hölsken A, Sill M, Merkle J, Schweizer L, Buchfelder M, Flitsch J, Fahlbusch R, Metzler M, Kool M, Pfister SM, von Deimling A, Capper D, Jones DTW, Buslei R (2016) Adamantinomatous and papillary craniopharyngiomas are characterized by distinct epigenomic as well as mutational and transcriptomic profiles. Acta Neuropathol Commun. https://doi.org/10.1186/s40478-016-0287-6
4. Wijnen M, van den Heuvel-Eibrink MM, Janssen JAMJL, Catsman-Berrevoets CE, Michiels EMC, van Veelen-Vincent M-LC, Dallenga AHG, van den Berge JH, van Rij CM, van der Lely A-J, Neggers SJCMM (2017) Very long-term sequelae of craniopharyngioma. Eur J Endocrinol. https://doi.org/10.1530/eje-17-0044
5. Müller HL, Merchant TE, Puget S, Martinez-Barbera J-P (2017) New outlook on the diagnosis, treatment and follow-up of childhood-onset craniopharyngioma. Nat Rev Endocrinol. https://doi.org/10.1038/nrendo.2016.217
6. Prieto R, Pascual JM (2018) Can tissue biomarkers reliably predict the biological behavior of craniopharyngiomas? A comprehensive overview. Pituitary. https://doi.org/10.1007/s11102-018-0890-6
7. Apps JR, Carreno G, Gonzalez-Meljem JM, Haston S, Guiho R, Cooper JE, Manshaei S, Jani N, Hölsken A, Pettorini B, Beynon RJ, Simpson DM, Fraser HC, Hong Y, Hallang S, Stone TJ, Virasami A, Donson AM, Jones D, Aquilina K, Spoudeas H, Joshi AR, Grundy

R, Storer LCD, Korbonits M, Hilton DA, Tossell K, Thavaraj S, Ungless MA, Gil J, Buslei R, Hankinson T, Hargrave D, Goding C, Andoniadou CL, Brogan P, Jacques TS, Williams HJ, Martinez-Barbera JP (2018) Tumour compartment transcriptomics demonstrates the activation of inflammatory and odontogenic programmes in human adamantinomatous craniopharyngioma and identifies the MAPK/ERK pathway as a novel therapeutic target. Acta Neuropathol. https://doi.org/10.1007/s00401-018-1830-2

8. Zoicas F, Schöfl C (2012) Craniopharyngioma in Adults. Front Endocrin. https://doi.org/10.3389/fendo.2012.00046
9. Chen B, Chen C, Zhang Y, Huang Z, Wang H, Li R, Xu J (2022) Differentiation between Germinoma and Craniopharyngioma using radiomics-based machine learning. JPM. https://doi.org/10.3390/jpm12010045
10. Mazurowski MA, Buda M, Saha A, Bashir MR (2018) Deep learning in radiology: an overview of the concepts and a survey of the state of the art with focus on MRI. Magn Reson Imaging. https://doi.org/10.1002/jmri.26534
11. Yao AD, Cheng DL, Pan I, Kitamura F (2020) Deep learning in neuroradiology: a systematic review of current algorithms and approaches for the new wave of imaging technology. Radiol Artif Intell. https://doi.org/10.1148/ryai.2020190026
12. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-CAM: visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/ICCV.2017.74
13. Lin T-Y, Goyal P, Girshick R, He K, Dollar P (2017) Focal loss for dense object detection. 2017 IEEE International Conference on Computer Vision (ICCV). https://doi.org/10.1109/iccv.2017.324
14. Loshchilov I, Hutter F (2019) Decoupled weight decay regularization. 2019 International Conference on Learning Representations (ICLR). https://doi.org/10.48550/ARXIV.1711.05101