

A multi-task two-path deep learning system for predicting the invasiveness of craniopharyngioma



Lin Zhu^{a,b,1}, Lingling Zhang^{c,1}, Wenxing Hu^d, Haixu Chen^e, Han Li^{b,f}, Shoushui Wei^{a,2,*}, Xuzhu Chen^{c,2,*}, Xibo Ma^{b,g,2,*}

^a School of Control Science and Engineering, Shandong University, Jinan, Shandong, 250061, China

^b CBSR&NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China

^c Department of radiology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China

^d University of New South Wales, Sydney, Australia

^e Institute of Geriatrics&National Clinical Research Center of Geriatrics Disease, The Second Medical Center, Chinese PLA General Hospital, Beijing, 100853, China

^f Department of Orthopedics, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China

^g School of Artificial Intelligence, University of the Chinese Academy of Sciences, Beijing, 100049, China

ARTICLE INFO

Article history:

Received 26 August 2021

Revised 19 January 2022

Accepted 19 January 2022

Keywords:

Craniopharyngioma

MRI Imaging

Deep learning

Invasiveness diagnosis

Lesion segmentation

ABSTRACT

Background and Objective: Craniopharyngioma is a kind of benign brain tumor in histography. However, it might be clinically aggressive and have severe manifestations, such as increased intracranial pressure, hypothalamic-pituitary dysfunction, and visual impairment. It is considered challenging for radiologists to predict the invasiveness of craniopharyngioma through MRI images. Therefore, developing a non-invasive method that can predict the invasiveness and boundary of CP as a reference before surgery is of clinical value for making more appropriate and individualized treatment decisions and reducing the occurrence of inappropriate surgical plan choices.

Methods: The MT-Brain system has consisted of two pathways, a sub-path based on 2D CNN for capturing the features from each slice of MRI images, and a 3D sub-network for capturing additional context information between slices. By introducing the two-path architecture, our system can make full use of the fusion of the above 2D and 3D features for classification. Furthermore, position encoding and mask-guided attention also have been introduced to improve the segmentation and diagnosis performance. To verify the performance of the MT-Brain system, we have enrolled 1032 patients with craniopharyngioma (302 invasion and 730 non-invasion patients), segmented the tumors on postcontrast coronal T1WI and randomized them into a training dataset and a testing dataset at a ratio of 8:2.

Results: The MT-Brain system achieved a remarkable performance in diagnosing the invasiveness of craniopharyngioma with the AUC of 83.84%, the accuracy of 77.94%, the sensitivity of 70.97%, and the specificity of 80.99%. In the lesion segmentation task, the predicted boundaries of lesions were similar to those labeled by radiologists with the dice of 66.36%. In addition, some explorations also have been made on the interpretability of deep learning models, illustrating the reliability of the model.

Conclusions: To the best of our knowledge, this study is the first to develop an integrated deep learning model to predict the invasiveness of craniopharyngioma preoperatively and locate the lesion boundary synchronously on MRI. The excellent performances indicate that the MT-Brain system has great potential in real-world clinical applications.

© 2022 Published by Elsevier B.V.

* Corresponding authors.

E-mail addresses: zhulin_sdu@mail.sdu.edu.cn (L. Zhu), zhanglingling_mail@163.com (L. Zhang), sswei@sdu.edu.cn (S. Wei), radiology888@aliyun.com (X. Chen), xibo.ma@ia.ac.cn (X. Ma).

¹ Lin Zhu and Lingling Zhang contributed equally to this work and should be considered co-first authors.

² Shoushui Wei, Xuzhu Chen and Xibo Ma approved the manuscript and should be considered as corresponding authors.

1. Introduction

Craniopharyngioma (CP) is a type of rare neurologically devastating brain tumor, with an incidence rate of 0.5–2 cases per million persons per year [1–3]. In the WHO criterion, CP has two histological subtypes called adamantinomatous CP (ACP) and papillary CP (PCP) [4]. Although as a histologically benign (WHO grade 1) brain tumor, CP often locally invades the perisellar neurovascular structures, such as the hypothalamus, cavernous sinus, and visual system [5,6]. This biological malignant behavior makes the treatment, especially the surgical resection of the tumor, to be more difficult and complicated [6–9]. Once the invaded structures are damaged or resected, there will be severe manifestations and neuroendocrinological deficits after operation. On the other hand, if the tumor is not totally resected, the remained part is a potential site for tumor recurrence as a hidden danger [10,11]. In pathological studies, the histology of the interface between CP and nearby brain tissue is classified into invasion and non-invasion [12,13], which is deemed to be associated with the resection quality and prognosis effect. However, except for the final pathology report after surgery and the intraoperative frozen section diagnosis, there are few methods that can identify the invasiveness of CP and predict the accurate boundary of the lesion before resection. For radiologists, invasion cannot be predicted preoperatively based on radiological examinations, including magnetic resonance imaging (MRI) [12]. At the same time, the edges of some CP lesions are blurred, due to the similar gray level to surrounding brain tissue on MRI images. In addition, CP lesions vary greatly in volume. Therefore, craniopharyngioma invasiveness diagnosis and lesion segmentation are challenging and clinically demanding tasks.³ The data is available at <http://intelligentmed.net:128/data.zip>

MRI, with the advantage of excellent soft-tissue resolution, is a common examination used in the preoperative assessment of the CP. However, the diagnostic accuracy of MRI images depends on the experiences of radiologists. As illustrated in Fig. 1, it is difficult for radiologists to visually differentiate the invasiveness of CP relying on MRI images. Therefore, developing a non-invasive method that can predict the invasiveness and boundary of CP as a reference before surgery is of clinical value for making more appropriate and individualized treatment decisions and reducing the occurrence of inappropriate surgical plan choices.

Radiomics, extracting plenty of image features from high-quality radiative or pathological medical images manually through feature engineering, has been widely used in clinical practice. Zhan et al. combined radiomics features and machine-learning classifiers for diagnosing common lesions located in the anterior skull base [14]. Chen et al. applied radiomics approach to the diagnosis of CP histological subtypes between ACP and PCP [15]. To predict the invasiveness of ACP, Ma et al. designed a radiomics-clinical nomogram based on radiomics features extracted from MRI images [16]. The radiomics approaches show better performance than traditional morphological signs interpreted by radiologists [17]. However, the radiomic methods often take the form of regional summary measures, and thus are tedious and time-consuming [18]. Considering the consistent anatomical location of CP, developing a diagnosis system based on deep learning is a good choice.

Deep learning can learn appropriate feature representations automatically through optimization without complicated feature engineering. Due to these advantages, currently, the deep learning methods have been widely applied in medical imaging analysis, such as lung pneumonia screening [19–22], pulmonary nodule segmentation [23,24], the invasiveness prediction for lung adenocarcinomas [25–27], and brain disease detection [28,29]. As demon-

strated in previous studies, it is feasible to apply deep learning in the diagnosis of ACP and the invasiveness classification of lung adenocarcinoma. Eric W. et al. presented a robust deep learning system to identify ACP from other masses of the same region in the brain [30]. Zhao et al. developed an AI scheme based on 3D densely connected blocks to implement pulmonary adenocarcinomas invasiveness prediction and nodule segmentation [25]. Xia et al. constructed two diagnostic models based on the LeNet and the DenseNet architecture respectively, and compared their performances for the invasiveness prediction within lung adenocarcinoma [27]. Gong et al. built a residual deep learning system for predicting the invasive degree of lung adenocarcinoma [26].

Although substantial success has been achieved in the invasiveness prediction for lung adenocarcinoma, existing studies on predicting the invasiveness of CP based on deep learning are relatively few. Hence, we aim to develop a diagnostic system based on deep learning that can non-invasively predict the invasiveness of CP before surgery. In this work, we present a Multi-task Two-path deep learning system (MT-Brain system) for CP preoperative invasiveness diagnosis and lesion segmentation. In summary, the main contributions of our work are three-fold:

- This paper presents a multi-task network based on deep learning to tackle challenging tasks of preoperative invasiveness diagnosis and locating the accurate lesion boundary synchronously of CP from brain routine MRI images, which fills the research gap of deep learning for invasive diagnosis of CP.
- The proposed two-path network framework, composed of a sub-pathway based on 2D CNN and a sub-pathway based on 3D CNN, combines spatial features obtained from each slice and context features between slices of the MRI for improving patient-level diagnostic accuracy.
- In order to establish a brain MRI dataset for studying the invasiveness of CP, we enrolled 302 patients with invasive CP and 730 patients with non-invasive CP. The experimental results, conducted on the dataset, demonstrates that the MT-Brain system outperforms the existing methods and has the potential to be used in clinical applications.

2. Dataset and annotation

As a retrospective study, it was approved by our institute committee without the informed consent of the patients. CP confirmed by pathology was searched in the neuropathological records of our institute. According to the inclusion criteria (Appendix A), a total of 1032 patients were enrolled in the study, including 302 invasion cases (which were confirmed by postoperative pathology) and 730 non-invasion cases. The clinical characteristics of the 1032 patients are summarized in Table 1.

In order to fairly demonstrate the performance of the algorithm, we only enrolled one 3D volumetric MR exam per patient in our dataset. The ground-truth mask of CP lesion is segmented by one neuroradiologist with 13 years of experience on postcon-

Table 1
Summary of brain MRI dataset.

Category	Num	Age ^a	Sex ^b	Volume(mm^3) ^c
Invasion	302	15[8–40]	169/133	11935[6650–21812]
Non-invasion	730	29[8–48]	426/304	8952[4589–18421]
Total	1032	26[8–45]	595/437	9975[5244–19866]

^aAge(median[25%–75%]).

^bSex(M/F), M:male, F:female.

^cVolume(median[25%–75%]). In invasion cases, the minimum tumor volume was $720mm^3$, the maximum tumor volume was $265785mm^3$. In non-invasion cases, the minimum tumor volume was $702mm^3$, the maximum tumor volume was $234726mm^3$.

³ Code is available at <https://github.com/abc008/MT-Brain-Network>.

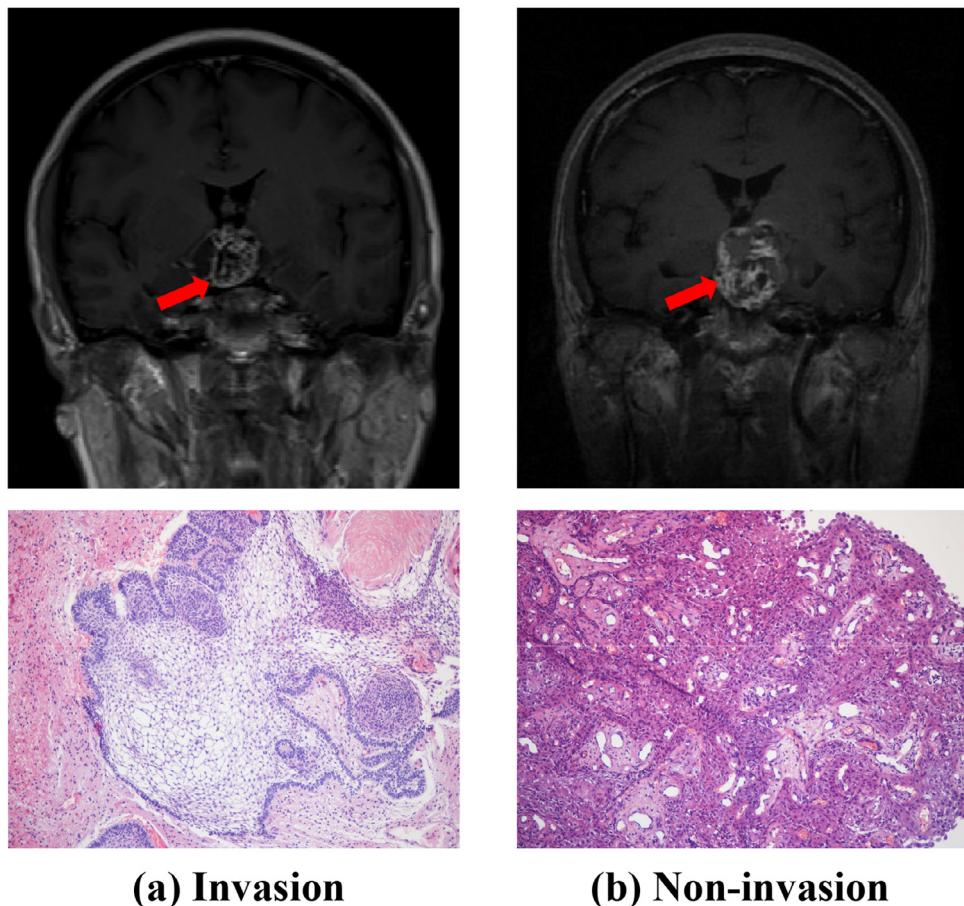


Fig. 1. MRI and corresponding pathological section images of (a) invasion CP and (b) non-invasion CP. Those pathological section images were captured using light microscopy (stain, hematoxylin and eosin; magnification, $\times 100$).

trast coronal T1WI images by using ITK-SNAP (version 3.8.0, www.itk-snap.org). The detailed MRI scanning protocol and parameters are shown in Supplementary S1. As shown in Fig. 2, all eligible patients were randomly divided into a training set and a testing set at a ratio of 8:2, using a computerized random number generator. Specifically, 828 patients were enrolled for the training set (240 invasion and 588 non-invasion patients) and the testing data contained 204 patients (62 invasion and 142 non-invasion patients).

3. Methods

As shown in Fig. 3, the workflow of MT-Brain system consists of two components, including the image preprocessing and deep neural network based on residual blocks. More concretely, the image preprocessing component receives raw MRI images and processes them for model training or inference (in Section 3.1). The deep neural network is designed as a multi-task two-path network, with the purpose of fusing the within-slice spatial features and between-slice context features, and predicting the invasiveness of CP and the boundary of lesion synchronously (in Section 3.2).

3.1. Image preprocessing

MRI images were acquired from 10 different MR scanners. Meanwhile, the age of enrolled patients was spread over a broad range, which led to different proportions and locations of the brain in MRI images. Using raw MRI images for classification directly inevitably introduces irrelevant information or noise, influencing the accuracy of the model. Therefore, according to the priors from radiologists and the characteristics of images, brain cropping, contrast

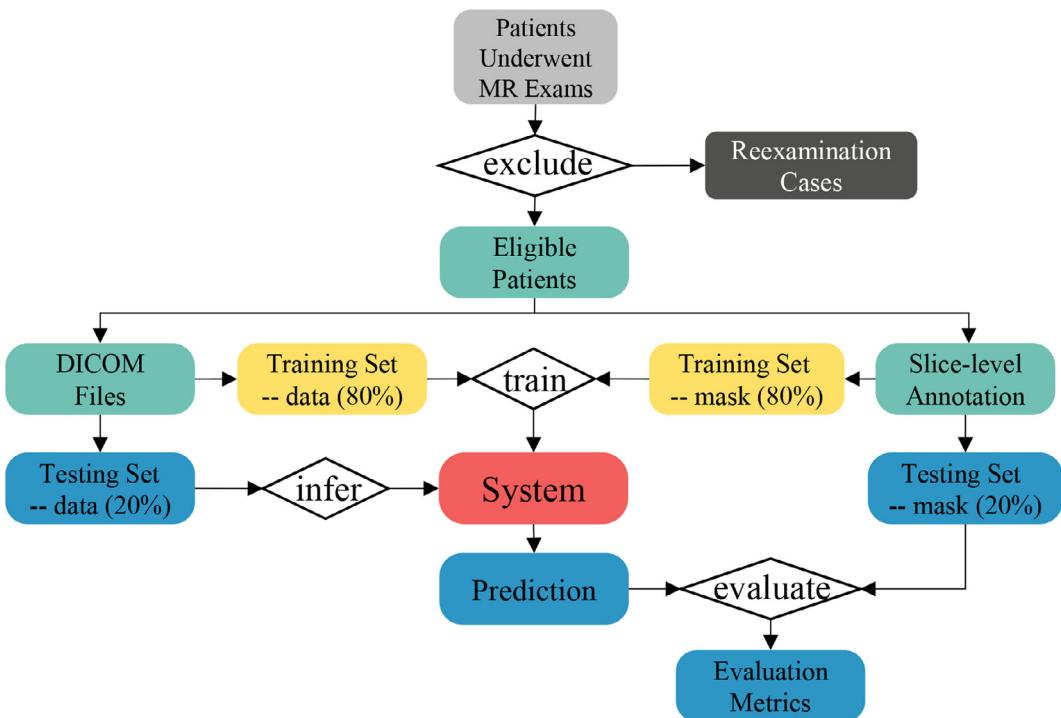
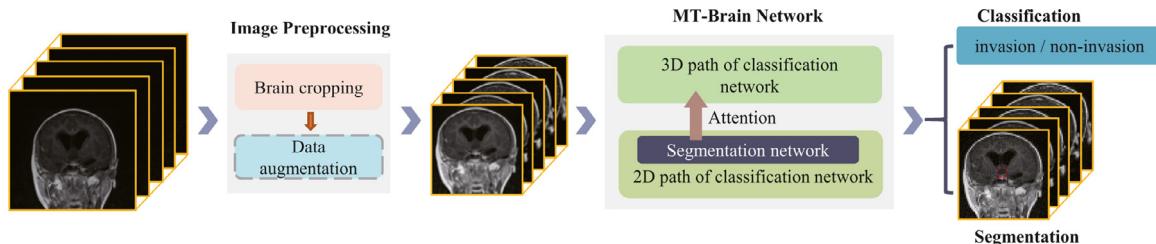
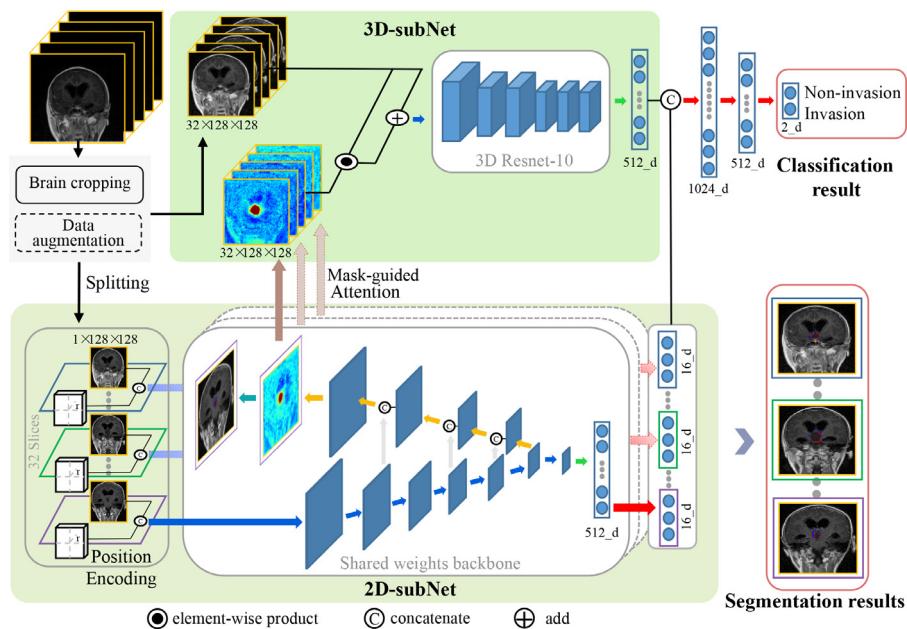
adjustment, and data augmentation are used in this paper. Those three effective and straightforward approaches for image preprocessing are detailed described in Appendix B.

3.2. MT-Brain Network framework

As indicated in Fig. 4, the MT-Brain Network consists of two pathways, a sub-path based on 3D CNNs for capturing additional context information between the slices, and a sub-path based on 2D CNN for capturing the textural and shape-based spatial feature representations from each slice of MRI images. In addation, a multi-task learning strategy is applied in this network to make the system predict the invasiveness of CP and accurate boundary of lesion synchronously. What's more, position encoding is introduced to learn the location information of lesions. Mask-guided attention is used to guide the 3D network to pay more attention to the lesion area and its nearby region.

3.2.1. Two-path network architecture

Nowadays, 2D and 3D CNNs have been widely applied in medical imaging analysis, and they focus on extracting different types of features. To be specific, 2D CNNs seek spatial knowledge among pixels in each slice of MRI images, and 3D CNNs can provide additional context information in the direction orthogonal of 2D patches. However, both 2D and 3D models have their disadvantages when applied to analyzing MRI images. 2D CNNs ignore between-slice context information and cannot understand the spatial dependencies among 2D image slices in a 3D MRI volume [31]. Due to the inconsistency of the pixel spacing and thickness

**Fig. 2.** The flowchart for building the dataset.**Fig. 3.** The schematic of the MT-Brain system. The details of the MT-Brain network are illustrated in Fig. 4.**Fig. 4.** The framework of the proposed Multi-task Two-path (MT-Brain) network. We add different improvements from four aspects: (1) two-path architecture to fuse 2D spatial features and 3D context features; (2) multi-task learning for providing more knowledge to radiologists for diagnosis; (3) position encoding for learning the location information; (4) mask-guided attention for guiding the 3D network to pay more attention to the lesion area and its nearby region.

of MRI images, and the relatively small number of MRI images, it is difficult for 3D CNNs to ensure a sufficient learning of all features, especially the boundary features [32,33]. Those features would be deemed to be important aspects for medical image analysis. As shown in [Appendix D.3](#), the 2D network and the 3D network might provide different and complementary information in classifying between invasion and non-invasion cases. To take full advantage of volumetric data and improve the prediction capability of the model, we introduce a two-path framework, named as Two-pathNet. As shown in [Fig. 4](#), the two-path network consists of two pathways and fuses 2D features and 3D features. Details will be described in [Appendix C](#).

3.2.2. Multi-Task learning

Locating the accurate boundary of the CP lesion before surgery is significant for the clinical diagnosis and treatment planning. However, the borders of CP lesions are often blurred in appearance, making it hard to distinguish them from healthy tissues. Moreover, manual segmentation of the tumor from MRI images is difficult and time-consuming. Therefore, designing a robust and accurate lesion segmentation model is of great value for surgeons.

Considering the low resolution of our data at the third dimension, we conduct the strategy of performing the lesion segmentation slice by slice. Inspired by U-net, which consists of a contracting path to capture context and a symmetric expanding path that enables precise localization [34], we utilize the first three Res-blocks in 2D patient-level model presented in [Appendix C.2](#) as the contracting path and establish the expansive path imitating U-net. The segmentation loss combines the Dice and cross-entropy losses:

$$L_{CE} = \frac{-\sum_{w,h}\{y_{w,h} \log(F(x)_{w,h}) + (1-y_{w,h}) \log(1-F(x)_{w,h})\}}{N} \quad (1)$$

$$L_{Dice} = 1 - \frac{2 \sum_{w,h} F(x)_{w,h} y_{w,h} + 1}{\sum_{w,h} F(x)_{w,h} + \sum_{w,h} y_{w,h} + 1} \quad (2)$$

$$L_{seg} = L_{CE} + L_{Dice} \quad (3)$$

where $F(x)$ is the sigmoid output of the segmentation network, $F(x) \in R^{W \times H}$, N indicates the number of pixels in the image.

To this end, we design a multi-task learning framework to predict the invasiveness of CP and accurate boundary of lesion synchronously. A mutual constraint is set between the classification task and segmentation task to ensure the consistency of the output between those two tasks. The multi-task learning loss combines the classification and segmentation losses:

$$L_{cls} = -y \log \hat{y} - (1-y) \log(1-\hat{y}) \quad (4)$$

$$\text{Loss} = \alpha L_{cls} + \beta L_{seg} \quad (5)$$

where \hat{y} is the classification result after softmax, α and β are the weights of L_{cls} and L_{seg} .

3.2.3. Position encoding

It is known that CP grows in sellar/suprasellar region of the brain, with a consistent anatomical location. As a result, we try to embed this useful prior information into our model. Inspired by [20,35], we give convolution access to its own input coordinates by using the coordinate maps ($H \times W \times 3$), where H and W are the height and the width of the input image, to facilitate locating the lesion. Those coordinate maps consist of three channels respectively. The first channel is instantiated with the coordinates of $x \in [0, W]$, and the second one is filled with the coordinates of $y \in [0, H]$. And they are both linearly normalized to range $[-1, 1]$. The last one represents the Euclidean distance between the point (x, y) and the center $(0, 0)$, that is $r = \sqrt{x^2 + y^2}$. As [Fig. 4](#) shows,

the coordinate maps and MRI slices are fed into the model together, and the in-channels parameter of the first CNN layer has changed to 4 in the 2D ResNet-18 backbone.

3.2.4. Mask-guided attention

Except for the 2D-subNet using coordinate maps mentioned in [Section 3.2.3](#), mask-guided attention mechanism is used in 3D-subNet as well, which can make the network pay attention to a focused location and enhance different representations of the object at that location. Existing soft attention modules used in feed-forward convolutional neural networks [36–38] infer attention maps based on an intermediate feature map, and adaptively refine feature by multiplying the attention maps and the intermediate feature maps. Different from them, we use the segmentation results as attention maps and multiply them by the raw MRI images to get the attention-weighted images. Subsequently, the raw MRI images plus the attention-weighted images are taken as the input of the 3D-subNet. Comparing with the existing attention module, this mask-guided attention method has advantages in controllability and reliability. As can be observed in [Fig. 4](#), the attention maps focus not only on the tumor lesion area, but also on the image information nearby the tumor lesion. In addition, the value of attention maps is very small in the region away from the lesion area. Thus, mask-guided attention achieves the goal of emphasizing meaningful features and suppressing unnecessary ones.

4. Experiments and results

To demonstrate the effectiveness of the proposed two-path architecture and other added strategies, we conduct a series of ablation experiments. Details about ablation studies are shown in [Appendix D](#).

4.1. Evaluation

For the quantitative evaluation of classification performance of the established MT-Brain system on the testing set, we utilize accuracy, sensitivity, specificity, F1-score, and Matthews correlation coefficient (MCC) of the testing dataset. Considering the imbalance between the positive and negative samples, the area under the receiver operating characteristic (ROC) curve (AUC) and the area under the precision-recall (PR) curve (PR-AUC) is also calculated as evaluation indicators. Among them, accuracy, sensitivity, and specificity are widely known. F1-score is used to evaluate the accuracy of the model using precision ($Precision = \frac{TP}{TP+FP}$) and recall ($Recall = \frac{TP}{TP+FN}$). F1-score is derived as follows:

$$F1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (6)$$

The formula for MCC is formulated as follows:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (7)$$

To assess the segmentation performance of the proposed MT-Brain system, we compute the three indicators namely: dice similarity coefficient (DSC), sensitivity (Sen), and Precision (PPV) for segmentation regions. Detailed definitions of those evaluation metrics are as follows:

$$DSC = \frac{2TP}{FP + 2TP + FN} \quad (8)$$

$$Sen = \frac{TP}{TP + FN} \quad (9)$$

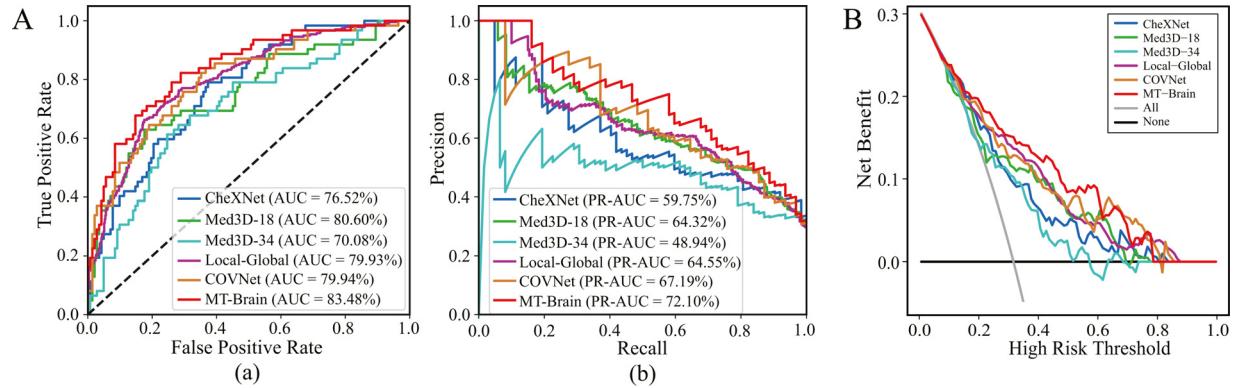
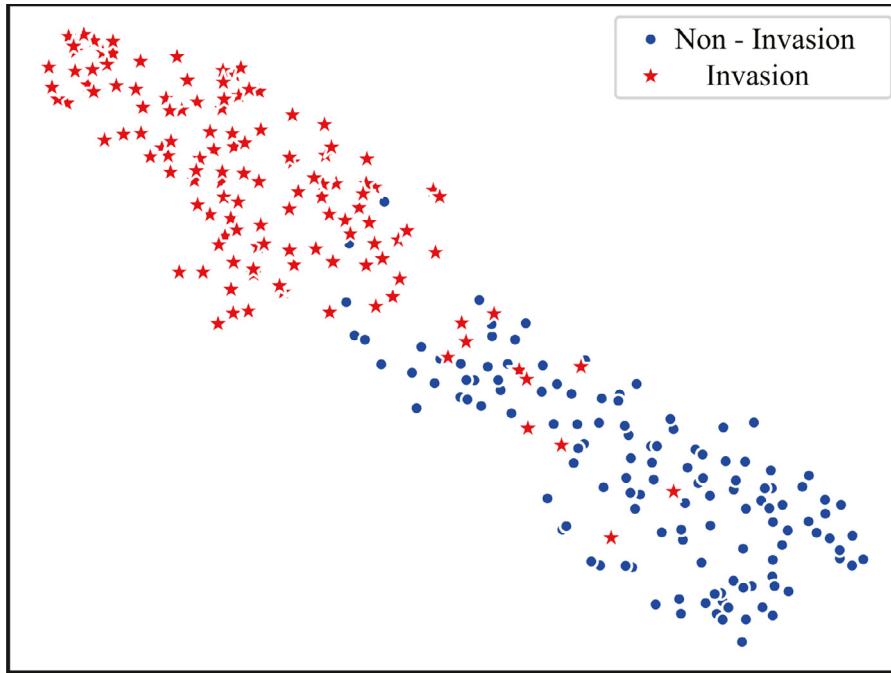
$$PPV = \frac{TP}{TP + FP} \quad (10)$$

Table 2

Comparison between MT-Brain system with several competitors on patient-level diagnosis.

Method	Dataset	Classification						
		AUC	PR-AUC	Accuracy	Sensitivity	Specificity	F1-score	MCC
CheXNet [19] (2018)	Training set: 828 patients	0.7652	0.5975	0.7206	0.5323	0.8028	0.5366	0.3366
Med3D-18 [23] (2019)	Testing set: 204 patients	0.7514	0.5864	0.6912	0.6452	0.7113	0.5594	0.3351
Med3D-34 [23] (2019)	Total: 1032 patients (ACP and PCP)	0.7008	0.4894	0.7109	0.5000	0.8028	0.5124	0.3072
Local-Global [24] (2019)		0.7993	0.6455	0.7500	0.6935	0.7746	0.6277	0.4466
COVNet [21] (2020)		0.7994	0.6719	0.7598	0.6452	0.8099	0.6202	0.4456
MT-Brain (ours)		0.8348	0.7210	0.7794	0.7097	0.8099	0.6617	0.5017

*The boldfaced numbers denote the best results among the methods.

**Fig. 5.** Performances of the proposed MT-Brain system and other competitors. A: The receiver operating characteristic (ROC) curves and the precision-recall (PR) curves of MT-Brain system and other competitors. (a) The ROC curve. (b) the PR curve. B: The decision curves of MT-Brain system and other competitors.**Fig. 6.** The t-SNE visualization results of our proposed MT-Brain system.

4.2. Comparison with existing methods

In this paper, we introduce an MT-Brain system, which aims to preoperatively assist the diagnosis of CP invasiveness at a patient level and predict the boundary of lesion synchronously. Thus, this system is more significant and practical in real-world clinical applications. To illustrate the advantages of our system, we conduct a series of experiments to compare the MT-Brain system with several competitors mentioned in previous studies. The competitors

include two 3D CNN-based models named Med3D-18 [23], Med3D-34 [23], and some models based on 2D CNN, such as CheXNet [19], Local-Global [24], and COVNet [21]. All the results are displayed in Table 2 and Fig. 5.

As Table 2 and Fig. 5 show, the proposed MT-Brain system surpasses the other competitors. Although it is almost impossible for radiologists to predict the invasiveness of CP, the designed system achieves 83.48% and 77.94% on AUC and accuracy respectively, with the PR-AUC, sensitivity, and specificity of 72.10%, 70.97% and

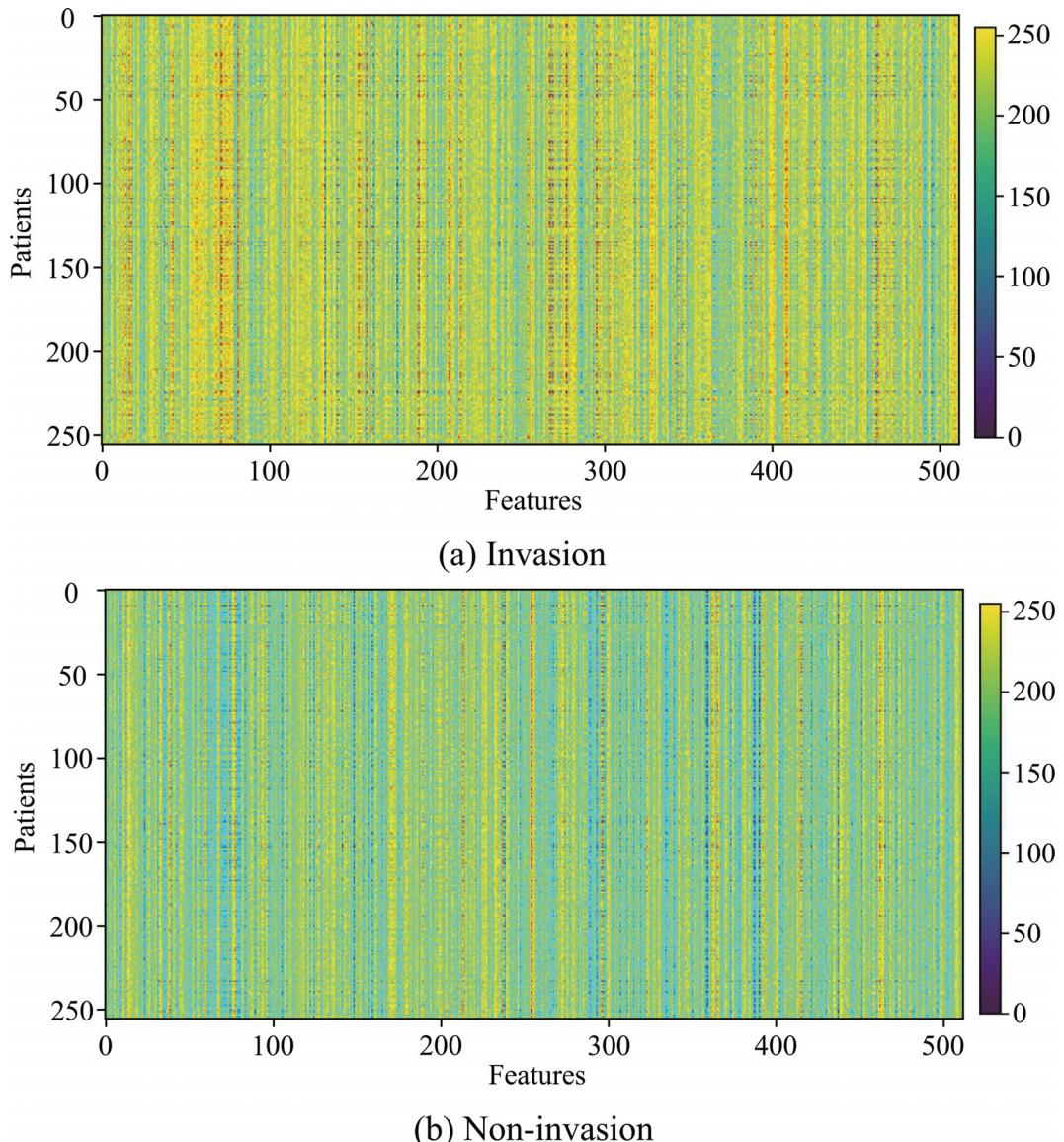


Fig. 7. The heat map of the extracted features of the proposed MT-Brain system. (a) Features extracted from 256 invasion patients. (b) Features extracted from 256 non-invasion patients. The obvious difference between (a) and (b) demonstrates the strong discriminative ability of the features.

80.99%, respectively. However, it can be directly observed that Med3D-34 achieves poor performance at most metrics, even inferior to Med3D-18. Through comparing the difference between Med3D-34 and Med3D-18, we infer it may be that due to the large learning parameters in Med3D-34 and limited dataset used, sufficient knowledge for the training model can not be learned by Med3D-34. This also indicates the validity of our motivation of using lightweight sub-network in the branches of our system.

To evaluate the clinical utility of the MT-Brain system and the other competitors, and prove their superiority when considering different threshold probabilities, we conduct a decision curve analysis [39,40] for testing dataset. As depicted in Fig. 5B, when the threshold probability of the clinical decision is in a large range, patients can benefit. Especially, when the threshold probability of the clinical decision is between 8% to 78%, patients can benefit more from the invasiveness prediction of the MT-Brain system than treating either all or no patients as invasiveness cases. At the same time, when the threshold probability of the clinical decision is between 18% to 60%, patients can benefit more from the invasiveness prediction of the MT-Brain system than the other competitors.

5. Visualizations

To visualize the features extracted from the established MT-Brain system, we randomly select 256 cases (128 invasion patients, 128 non-invasion patients) as the dataset for visualization, and then use the t-distributed stochastic neighbor embedding (t-SNE) algorithm [41]. As illustrated in Fig. 6, the learned features of our proposed system are successfully affined to two clusters intuitively, indicating that the proposed method can learn discriminative deep features and has great potential to predict the invasiveness of CP.

Except for t-SNE, we also compare the learned features from two categories of patients to prove the effectiveness of our proposed system. To show more intuitively, we randomly select 256 non-invasion patients and 256 invasion patients, and then depict the extracted features by painting the feature heat map respectively, as shown in Fig. 7. In the heat map, the grid with bright red color indicates that the exploited feature has a higher value than the dark blue grid. Comparing the Fig. 7(a) and (b), we can intuitively observe that the feature values extracted by the MT-

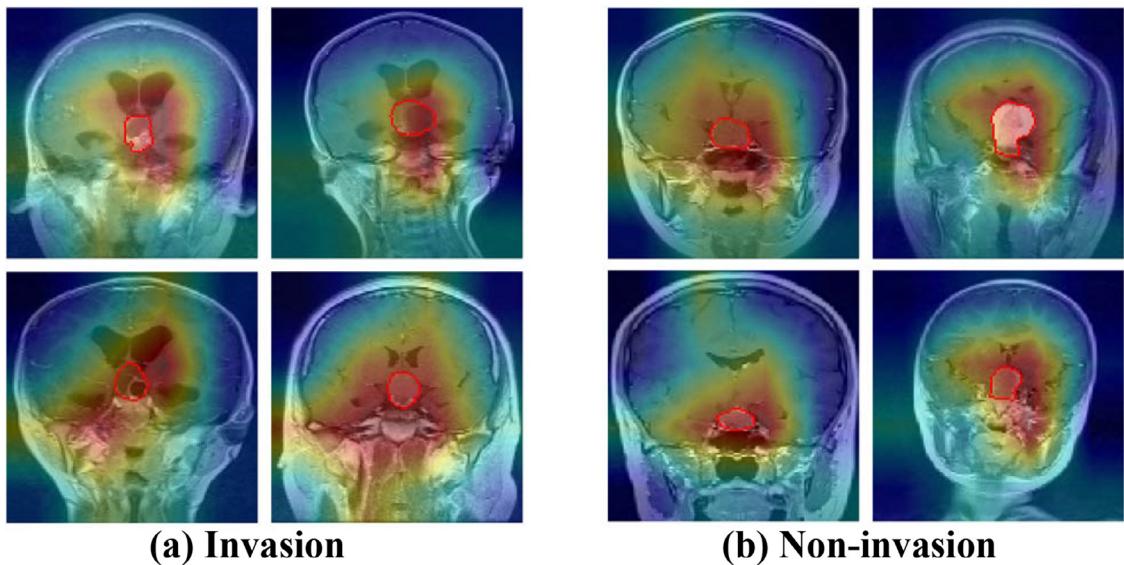


Fig. 8. The activation map of 2D-subNet in the trained MT-Brain system for different patients. (a) Invasion patients. (b) Non-invasion patients. In the activation map, areas with bright red color are more important than dark blue areas. The red line is the lesion boundary annotated by radiologists. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

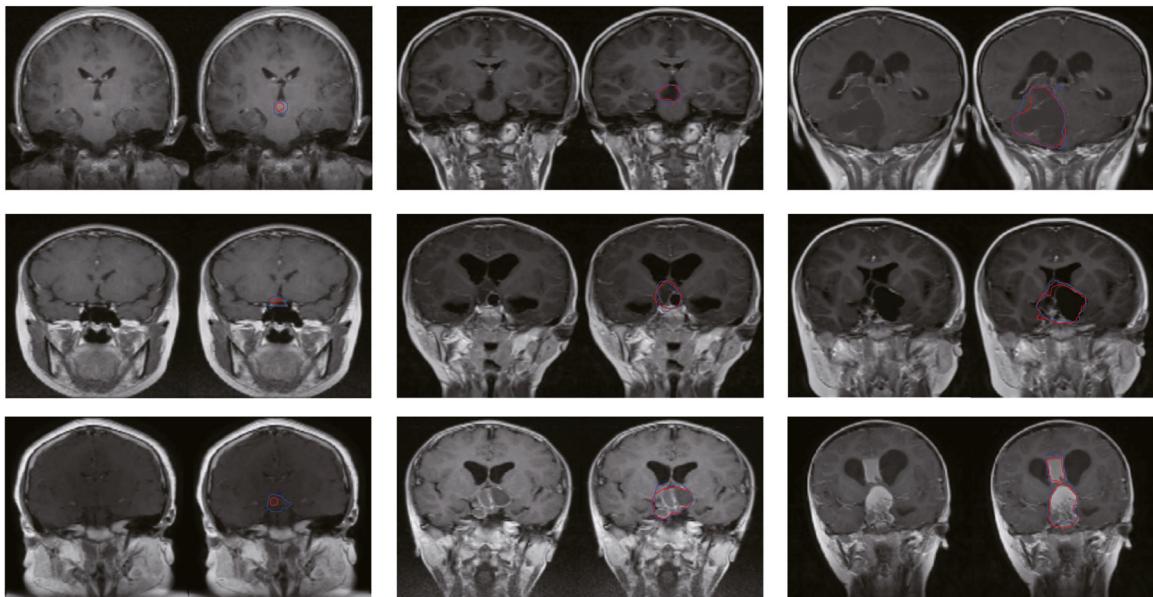


Fig. 9. The lesions segmentation results of different patients. The red line is the lesion boundary annotated by radiologists. The blue line is the boundary of segmentation result from MT-Brain system. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Brain system have obvious differences, demonstrating the learned features have a strong discriminative ability for classification.

In order to demonstrate the MT-Brain system, as a black box, correctly learns information from lesions, we visualize the brain area that draws most attention to the MT-Brain system by the gradient weighted class activation maps (Grad-CAM) [42] algorithm. Those activation maps reflect the regions which play important roles during the inference. Fig. 8 displays slices of the MRI images from different patients in the test cohort, and with the Grad-CAM overlaid on it. As illustrated in Fig. 8, the suspicious brain areas discovered by MT-Brain show high overlap with the actual CP lesions and peritumoral regions, both for invasion cases and non-invasion cases. It reveals that the MT-Brain system has learned the abstract mappings between raw MRI images and clinical diagnosis.

In other words, the MT-Brain system captures features from the lesion areas and peritumoral regions for diagnosing the invasiveness of CP.

As a multi-task learning system, the proposed MT-Brain system not only can diagnose the invasiveness of CP, but also can predict the accurate boundary of lesions synchronously. Fig. 9 illustrates the segmentation results of different patients in the testing dataset. The red line is the lesion boundary annotated by radiologists, and the blue line is the boundary of segmentation results from the MT-Brain system. As Table 1 shows, the upper quartile of the volume is about 4 times larger than the lower quartile, reflecting the large difference in the volume of the CP lesions. Even so, the proposed MT-Brain system performs well in lesion segmentation. As we can see from Fig. 9, for small lesions, our segmentation results are big-

ger than the area labeled by radiologists. But for most lesions, the predictions and annotations are similar correspondingly.

6. Discussion

In this paper, we proposed a novel multi-task two-path deep learning system (MT-Brain system) for predicting the invasiveness of craniopharyngioma and the accurate boundary of lesion synchronously on MRI. The previous study [16] used a radiomics approach and extracted regional features manually, which was tedious and time-consuming. The MT-Brain system eliminated the manually designed feature engineering. It took raw MRI images as input and gained the diagnosis and segmentation results. Although existing 3D networks extracted spatial and context features, they extracted limited spatial information in each slice, especially the lesion boundary features. Hence, they performed poorly in predicting the invasiveness of CP. To make up the defects of existing networks and capture rich information for classification, we designed a two-path framework consisting of a patient-level 2D sub-network (2D-subNet) and a 3D sub-network (3D-subNet). Though utilizing the two-path architecture, the MT-Brain system made full use of both 2D spatial features in each slice and 3D context features between slices simultaneously to enhance the discriminative ability of deep features. The introduced multi-task learning mechanism made the MT-Brain system predict the invasiveness of CP and accurate boundary of lesion synchronously, providing more knowledge to radiologists for diagnosis and of great value in clinical applications.

In addition, considering the consistent anatomical location of CP, the coordinate knowledge was used for position encoding. The position encoding strategy made the MT-Brain system learn the location information of CP lesions. Meanwhile, by using mask-guided attention, the segmentation results from 2D-subNet were introduced into 3D-subNet, which can guide the 3D network to pay more attention to the lesion area and its nearby region. The attention trick used in this paper provided awareness of lesions, which emphasized meaningful features and suppressed unnecessary ones.

What's more, we built a brain MRI dataset by enrolling 1032 patients (302 invasion patients, 730 non-invasion patients) for a series of experiments. With the improvements provided by the above components, the MT-Brain system achieved a remarkable performance of 83.84% AUC, 72.10% PR-AUC, 77.94% accuracy, 70.97% sensitivity, and 80.99% specificity. As illustrated in those experiments, the results of the MT-Brain system were significantly better than those competitors, indicating that the MT-Brain system has great potential in real-world clinical applications.

There are still several limitations in this study. First, the architecture of the network remains to be optimized. For example, the segmentation sub-network based on U-net was trained using imperfect ground-truth masks, without utilizing spatial information. Second, the cases enrolled in this study came from a single hospital, without performing cross-center validations.

7. Conclusion

In this paper, we presented a new multi-task two-path deep learning system (MT-Brain system) to tackle those challenging and clinically-demanding tasks of craniopharyngioma invasiveness diagnosis and lesion segmentation. MT-Brain system was composed of 2D-subNet and 3D-subNet for fusing the 2D spatial features and 3D context features. On this basis, position encoding and mask-guided attention were also introduced. We enrolled 1032 patients to form a craniopharyngioma MRI dataset and evaluated the MT-Brain system on it. The experiment results demonstrated the effectiveness of the proposed MT-Brain system, which is of great potential in clinical applications as a computer-aided diagnostic tool.

Funding

This work was supported by the National Key Research Program of China [grant number 2016YFA0100902]; the National Natural Science Foundation Projects of China [grant numbers 82072014, 82090051, 81871442, 81772005]; the Shandong Province Natural Science Foundation [grant number ZR2020MF028]; the Youth Innovation Promotion Association CAS [grant number Y201930]; and the Collaborative Innovative Major Special Project Supported by Beijing Municipal Science & Technology Commission [grant number Z191100006619088].

Declaration of competing interest

The authors declare that they have no conflict of interest.

Acknowledgment

The authors would like to thank Ziliang Dong for the help of paper revision, and LiPing Han for the assistance of data sorting.

Appendix A. Data inclusion criteria

The inclusion criteria for this study were as follows: (1) histologically confirmed as craniopharyngioma; (2) available preoperative MRI data; and (3) available clinical characteristics. Patients were excluded if (1) recurrent craniopharyngioma; (2) without preoperative MRI in our institute; (3) treated with bypass surgery or drainage before preoperative MRI; (4) the pathological diagnosis was not definite; (5) preoperative MRI without post-contrast images; (6) preoperative MR images with obvious artifact; (7) combined with lesions around the sellar/suprasellar area which may affect the analysis of craniopharyngioma; (8) The period between preoperative MRI and a surgical operation was too long (more than 2 months); and (9) Operation record was not available.

Appendix B. Image preprocessing details

B1. Brain cropping

Due to the wide range of the age of patients, the area of brains takes different proportions in different MRI images. However, training an extra deep learning network for segmenting brain regions needs extreme computational costs, and manual labeling is time-consuming. Therefore, a simple threshold segmentation algorithm is designed to automatically crop the area of brains using the minimum bounding rectangle. We select the threshold as $T = 300$ to separate the brain area from the background. Then, the Minimum bounding rectangle is computed for brain cropping and the cropped images are resized to the same size ($128 \times 128 \times 32$) by using the bicubic interpolation.

B2. Data augmentation

Data augmentation is a key component of any network training, which can increase the limited sample size to avoid the overfitting problem and alleviate the presence of data imbalance. At the same time, data augmentation can increase the artificial diversity of the training dataset, strengthening the generalization capability of the model. Through observing the characteristics of MRI images, we find that some of the images show oblique anatomical structures.

With the purpose of simulating the process of window-leveling when a radiologist is looking at MRI Images, we further adjust the contrast of all MRI images using the Gamma correction method

[43]. The simple formula of the gamma correction (GC) is formulated as:

$$V_{\text{out}} = V_{\text{in}}^{\gamma}. \quad (11)$$

The intensity contrast can be enhanced with $\gamma < 1$. More concretely, we assign the value of the gamma parameter randomly in the range from 0.7 to 1 during the training process. This preprocessing can not only generate more MRI images for training but also make the assessment more accurate and reliable.

In addition to adjusting the contrast of images randomly, we also apply other data augmentation approaches during the training process for 3D images, including random rotation, random alpha affine, and random zoom. Significantly, during testing, those mentioned data augmentation strategies are not applied.

Appendix C. Two-path network architecture

C1. 3D-subNet

The top pathway is 3D-subNet based on the 3D CNN model. 3D convolutional models take 3D images as input and are able to extract more complex features by spanning across three dimensions. Considering the huge amount of 3D CNN model parameters, we utilize the 3D Resnet-10 [44,45] as the backbone network of the 3D-subNet to prevent overfitting. As indicated in the top half of Fig. 4, the 3D MRI image after image preprocessing is fed into this 3D network. Afterward, a global average pooling (GAP) layer is placed to pool the entire feature map of each channel to a single weighting coefficient. Therefore, feature maps are reshaped to a 512-length, 1-dimension vector. Other significant parameters of the network are defined by experiments.

C2. 2D-subNet

The bottom pathway is the patient-level model based on 2D CNN, named as 2D-subNet. As demonstrated by the bottom half of Fig. 4, we introduce a shared weights backbone and a fully connected layer. Each slice of the 3D MRI images is fed into the backbone network in order and the 2D Resnet-18 [44] is selected as the shared weights backbone. Meanwhile, we initialize it with the ImageNet [46] pre-trained weights to facilitate training. After the 2D CNN, a 16-length, 1-dimension vector can be obtained for each slice. Subsequently, those vectors transformed from MRI slices are concatenated into a 512-length, 1-dimension vector.

In the final classification decision, those two 512-length, 1-dimension vectors obtained from the two pathways are concatenated and fed into two fully connected layers with 50% dropout. A binary Softmax classifier is used to identify the invasiveness of CP on patient-level. By using the multilayer perceptron (MLP) as the classification decision, the extracted features are mapped to target categories.

Appendix D. Ablation studies

D1. Implementation

The implementation of the MT-Brain system is based on PyTorch [47]. During the procedure of training, the MT-Brain system is trained with one NVIDIA GeForce RTX 2080Ti. Stochastic gradient descent (SGD) is chosen as the optimization solver to update the weights of the model. The batch size, weight decay, and momentum are set to 16, 0.001, and 0.9, respectively. Since the sub-networks for the classification task and the segmentation task in the MT-Brain system share some parameters, we adopt fine-tuning training tricks in the training process. To speed up training convergence and improve accuracy, we train the segmentation network named as U-net+coord network with a learning rate of 0.01

Table 3
Improvements of position encoding in U-net.

Method	Evaluation Metrics		
	Dice	PPV	Sen
U-net	0.6472	0.6468	0.6659
U-net+Coord	0.6615	0.6581	0.6808

*The boldfaced numbers denote the best results among the methods.

and 3D-subNet network with a learning rate of 0.0005 respectively, and fine-tune them to the MT-Brain model with a learning rate of 0.0001.

D2. Effectiveness of position encoding

It is worth proving the effectiveness of using the coordinate maps as guidance for position encoding. Considering the consistent anatomical location of CP, applying the position encoding can achieve a more effective transformation of the positional information. To demonstrate this, we compare the segmentation performance of U-net [34] and 'U-net+Coord', which adds three extra coordinate channels respectively to represent x, y, and the Euclidean distance to the center of each slice. As can be seen in Table 3, the performances of all segmentation metrics are improved by using coordinate information, which clearly verifies the benefits of the position encoding strategy. In the meanwhile, the segmentation performance satisfies the need to locate the lesion and provides guidance for clinical surgeons.

D3. Effectiveness of two-path architecture

To demonstrate the effectiveness of the proposed two-path architecture and the discrimination of our extracted features, we conduct several ablation experiments as indicated in Table 4. We design and train three networks named as 2D-subNet, 3D-subNet, and Two-pathNet, and evaluate them using evaluation metrics in the testing set. More specifically, '2D-subNet' refers to the patient-level 2D sub-pathway of the two-path framework based on 2D CNNs. '3D-subNet' indicates the 3D sub-pathway of the two-path architecture composed of 3D CNNs. 'Two-pathNet' donates the basic structure of the designed MT-Brain system, consisting of the 2D sub-pathway and the 3D sub-pathway. To acquire the classification results of the 2D-subNet and the 3D-subNet, the classification decision head consisting of two fully connected layers with 50% dropout and a binary Softmax classifier is used during experiments, which is similar to the decision head in the Two-pathNet.

Fig. 10 shows the scatter plot of prediction score distributions of invasion cases and non-invasion cases in the testing set. As can be seen in Fig. 10, a large number of prediction scores generated by 2D-subNet and 3D-subNet are inconsistent and scattered in both invasion and non-invasion cases. It demonstrates that the 2D sub-network and 3D sub-network might obtain different and complementary information in diagnosing the invasiveness of CP. As illustrated in Table 4 and Fig. 11A, ROC curves and precision-recall (PR) curves also show the trend that fusing the obtained 2D and 3D features can improve the classification performance of the model.

Specifically speaking, Fig. 10(a) displays the prediction probabilities of invasion cases in the testing set generated by 2D-subNet and 3D-subNet, and Fig. 10(b) shows the prediction probabilities of non-invasion cases in the testing set. In Fig. 10(a), the prediction score of the 2D-subNet is generally lower than that of the 3D-subNet, which is consistent with the phenomenon in Table 4 that the sensitivity (Recall) of 2D-subNet is lower than that of 3D-subNet. It indicates that the additional context information in the direction orthogonal of 2D slices captured by 3D-subNet might

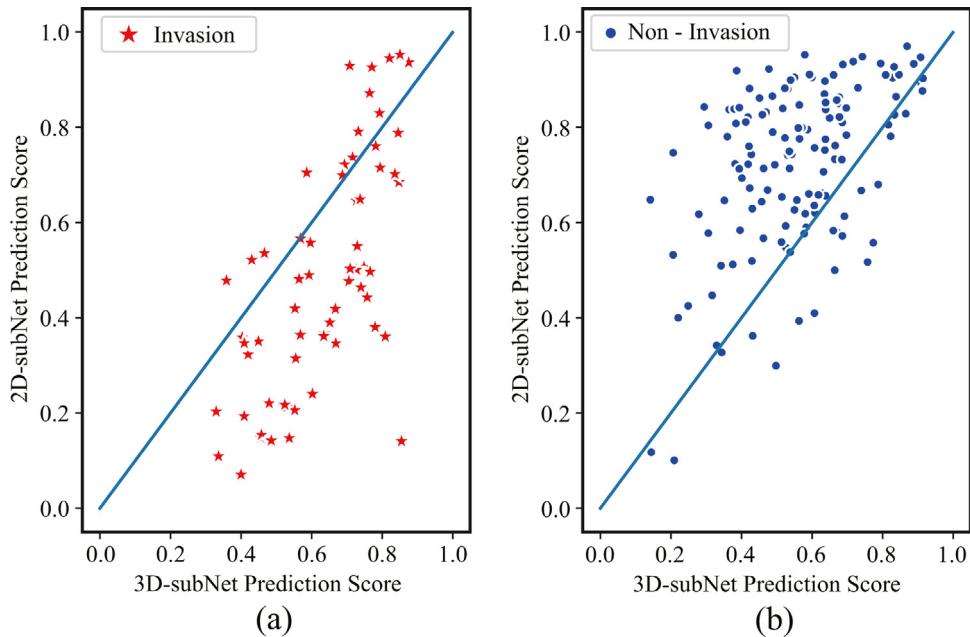
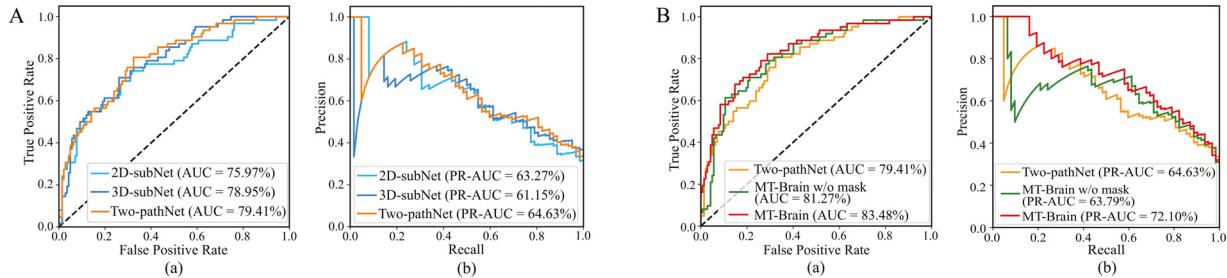
Table 4

Performance improvements of different components.

Method	TP ^a	MT ^b	MG-att ^c	Classification						Segmentation			
				AUC	PR-AUC	Accuracy	Sensitivity	Specificity	F1-score	MCC	Dice	PPV	Sen
2D-subNet				0.7597	0.6327	0.7696	0.4355	0.9155	0.5347	0.4105	—	—	—
3D-subNet				0.7895	0.6115	0.6765	0.7580	0.6408	0.5875	0.3672	—	—	—
Two-pathNet	✓			0.7941	0.6463	0.7157	0.6452	0.7465	0.5797	0.3726	—	—	—
MT-Brain w/o mask	✓	✓		0.8127	0.6379	0.7647	0.6935	0.7958	0.6418	0.4710	0.6623	0.6678	0.6772
MT-Brain	✓	✓	✓	0.8348	0.7210	0.7794	0.7097	0.8099	0.6617	0.5017	0.6636	0.6668	0.6809

^aTwo-path architecture.^bMulti-task learning.^cMask-guided attention.

*The boldfaced numbers denote the best results among the methods.

**Fig. 10.** Scatter plots of prediction score from 2D-subNet and 3D-subNet. (a) Scatter plots of prediction score distribution of invasion CP. (b) Scatter plots of prediction score distribution of non-invasion CP.**Fig. 11.** The receiver operating characteristic (ROC) curve and the precision-recall (PR) curve. (a) The ROC curve. (b) the PR curve. A: The Two-path architecture is proven to be effective. B: The strategy of using multi-task learning and mask-guided attention is proven to be effective.

play an important role in diagnosing invasion cases. However, the prediction score of the 2D-subNet is generally higher than that of the 3D-subNet in Fig. 10(b), and the specificity of the 3D-subNet is lower than that of the 2D-subNet. It demonstrates the 3D-subNet might not sufficiently learn all features, due to the inconsistency of the pixel spacing and thickness of MRI images.

Several observations can also be made from Table 4 and Fig. 11A: (1) Only using the patient-level 2D-subNet or the 3D-subNet, the model distinguishes the invasiveness of CP with AUCs of 75.97% and 78.95% in the test cohorts, respectively. However, the big performance gap between the sensitivity and specificity of those two networks suggests their inferior robustness. (2) Through

comparing the performance gap between sensitivity and specificity, we can find that the performance gap of the Two-pathNet is smaller than 2D-subNet and 3D-subNet. At the same time, the performance on PR-AUC and F1-score improves in Two-pathNet, both reflecting the effectiveness of the strategy of fusing 2D and 3D features. (3) It catches our attention that the performance in accuracy and MCC of Two-pathNet is worse than 2D-subNet. We argue that this does not reflect the inferiority of the Two-pathNet. This is due to the abundance of the features extracted from the Two-pathNet. In other words, the 2D-subNet and the 3D-subNet extract abundant features but they are looking at different features, as indicated in Fig. 10. However, many invasive cases are incorrectly diagnosed by

2D-subNet, which can be inferred from the bad performance of sensitivity.

D4. Effectiveness of multi-task learning and mask-guided attention

In order to explicitly verify the advantages of our improvements including multi-task learning and mask-guided attention, we conduct further experiments by adding different improvements step by step, including multi-task learning and Mask-guided attention. All results are listed in **Table 4** and **Fig. 11B**.

As illustrated in the third and fourth rows of **Table 4**, by comparing the classification evaluation metrics, we can find that the model added multi-task learning achieves significant improvement in all seven metrics and narrows the gap between sensitivity and specificity. It verifies the effectiveness of multi-task learning in making the system less prone to be overfitting, and improving the accuracy and training convergence. Moreover, after establishing the multi-task learning architecture, the model achieves the goal to predict the invasiveness of CP together with the boundary of lesion, which is of great value in clinical applications.

As demonstrated in the last two rows of **Table 4**, by comparing the classification evaluation metrics and segmentation evaluation metrics, we can clearly observe that after adding mask-guided attention, the MT-Brain system can improve all the classification metrics and maintain high segmentation results. This proves that mask-guided attention can guide the 3D sub-pathway network to pay more attention to the lesion and surrounding areas, achieving the goal of emphasizing meaningful features and suppressing unnecessary ones.

Reviewing the last three rows of **Table 4** and **Fig. 11B**, the performance on all metrics improves gradually with adding different improvements, demonstrating the effectiveness of our conducted strategies and reflecting the excellent robustness of the MT-Brain system. Intuitively, with adding multi-task learning and mask-guided attention, the sensitivity and specificity are improved consistently. It means that more invasive samples are correctly diagnosed as positive, and fewer non-invasive samples are misclassified as invasiveness.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.cmpb.2022.106651](https://doi.org/10.1016/j.cmpb.2022.106651).

References

- [1] M. L. Garr, A. Cama, Craniopharyngioma: Modern concepts in pathogenesis and treatment, *Curr Opin Pediatr* 19 (2007) 471–479 <https://pubmed.ncbi.nlm.nih.gov/17630614/>, doi:[10.1097/MOP.0b013e3282495a22](https://doi.org/10.1097/MOP.0b013e3282495a22).
- [2] E.H. Nielsen, U. Feldt-Rasmussen, L. Poulsgaard, L. Kristensen, J. Astrup, J.O. Jrgensen, P. Bjerre, M. Andersen, C. Andersen, J. Jrgensen, J. Lindholm, P. Laurberg, Incidence of craniopharyngioma in Denmark ($n = 189$) and estimated world incidence of craniopharyngioma in children and adults, *J. Neurooncol.* 104 (2011) 755–763, doi:[10.1007/s11060-011-0540-6](https://doi.org/10.1007/s11060-011-0540-6).
- [3] G.R. Bunin, T.S. Surawicz, P.A. Witman, S. Preston-Martin, F. Davis, J.M. Bruner, The descriptive epidemiology of craniopharyngioma, *J. Neurosurg.* 89 (1998) 547–551, doi:[10.3171/jns.1998.89.4.0547](https://doi.org/10.3171/jns.1998.89.4.0547).
- [4] D.N. Louis, A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W.K. Cavenee, H. Ohgaki, O.D. Wiestler, P. Kleihues, D.W. Ellison, The 2016 world health organization classification of tumors of the central nervous system: a summary, *Acta Neuropathol.* 131 (2016) 803–820 <https://pubmed.ncbi.nlm.nih.gov/27157931/>, doi:[10.1007/s00401-016-1545-1](https://doi.org/10.1007/s00401-016-1545-1).
- [5] C. Stache, A. Hlsken, R. Fahibusch, J. Flitsch, S.M. Schlaffer, M. Buchfelder, R. Buslei, Tight junction protein claudin-1 is differentially expressed in craniopharyngioma subtypes and indicates invasive tumor growth, *Neuro-oncology* 16 (2014) 256–264, doi:[10.1093/neuonc/not195](https://doi.org/10.1093/neuonc/not195).
- [6] M.R. Garnett, S. Puget, J. Grill, C. Sainte-Rose, Craniopharyngioma, *Orphanet J Rare Dis.* 2 (2007), <http://ojrd.biomedcentral.com/articles/10.1186/1750-1172-2-18>, doi:[10.1186/1750-1172-2-18](https://doi.org/10.1186/1750-1172-2-18).
- [7] K.C. Wang, S.H. Hong, S.K. Kim, B.K. Cho, Origin of craniopharyngiomas: implication on the growth pattern, *Childs Nerv Syst* 21 (2005) 628–634, doi:[10.1007/s00381-005-1203-8](https://doi.org/10.1007/s00381-005-1203-8).
- [8] S. Takano, H. Akutsu, M. Mizumoto, T. Yamamoto, K. Tsuboi, A. Matsumura, Neuroendoscopy followed by radiotherapy in cystic craniopharyngiomas – a long-term follow-up, *World Neurosurg.* 84 (2015) 1305–1315.e2, doi:[10.1016/j.wneu.2015.06.022](https://doi.org/10.1016/j.wneu.2015.06.022).
- [9] J.R. Apps, J.C. Hutchinson, O.J. Arthurs, A. Virasami, A. Joshi, B. Zeller-Plumhoff, D. Moulding, T.S. Jacques, N.J. Sebire, J.P. Martinez-Barbera, Imaging invasion: Micro-ct imaging of adamantinomatous craniopharyngioma highlights cell type specific spatial relationships of tissue invasion, *Acta Neuropathol Commun* 4 (2016), doi:[10.1186/s40478-016-0321-8](https://doi.org/10.1186/s40478-016-0321-8).
- [10] H.L. Muller, The diagnosis and treatment of craniopharyngioma, *Neuroendocrinology* 110 (2020) 753–766, doi:[10.1159/000504512](https://doi.org/10.1159/000504512).
- [11] M. Kili, S.M. Can, B. zdemir, C. Tanik, Management of craniopharyngioma, *Journal of Craniofacial Surgery* 30 (2019) 178–183, doi:[10.1097/SCS.00000000000005136](https://doi.org/10.1097/SCS.00000000000005136).
- [12] T. Kawamata, O. Kubo, T. Hori, Histological findings at the boundary of craniopharyngiomas, *Brain Tumor Pathol.* 22 (2005) 75–78, doi:[10.1007/s10014-005-0191-4](https://doi.org/10.1007/s10014-005-0191-4).
- [13] Y. Liu, S.-T. Qi, C.-H. Wang, J. Pan, J. Fan, J.-X. Peng, X. Zhang, Y. Bao, Y.-W. Liu, Pathological relationship between adamantinomatous craniopharyngioma and adjacent structures based on qst classification, *J Neuropathol Exp Neurol* 77 (2018) 1017–1023, doi:[10.1093/jnen/nly083](https://doi.org/10.1093/jnen/nly083).
- [14] Y. Zhang, L. Shang, C. Chen, X. Ma, X. Ou, J. Wang, F. Xia, J. Xu, Machine-learning classifiers in discrimination of lesions located in the anterior skull base, *Front Oncol* 10 (2020) 1–9, doi:[10.3389/fonc.2020.00752](https://doi.org/10.3389/fonc.2020.00752).
- [15] X. Chen, Y. Tong, Z. Shi, H. Chen, Z. Yang, Y. Wang, L. Chen, J. Yu, Noninvasive molecular diagnosis of craniopharyngioma with mri-based radiomics approach, *BMC Neurotol* 19 (2019) 1–11, doi:[10.1186/s12883-018-1216-z](https://doi.org/10.1186/s12883-018-1216-z).
- [16] G. Ma, J. Kang, N. Qiao, B. Zhang, X. Chen, G. Li, Z. Gao, S. Gui, Non-invasive radiomics approach predict invasiveness of adamantinomatous craniopharyngioma before surgery, *Front Oncol* 10 (2021) 1–7, doi:[10.3389/fonc.2020.599888](https://doi.org/10.3389/fonc.2020.599888).
- [17] L. Fan, M.J. Fang, Z.B. Li, W.T. Tu, S.P. Wang, W.F. Chen, J. Tian, D. Dong, S.Y. Liu, Radiomics signature: a biomarker for the preoperative discrimination of lung invasive adenocarcinoma manifesting as a ground-glass nodule, *Eur Radiol* 29 (2019) 889–897, doi:[10.1007/s00330-018-5530-z](https://doi.org/10.1007/s00330-018-5530-z).
- [18] S.K. Zhou, H. Greenspan, C. Davatzikos, J.S. Duncan, B. van Ginneken, A. Madabhushi, J.L. Prince, D. Rueckert, R.M. Summers, A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises, *Proceedings of the IEEE* 109 (5) (2021) 820–838, doi:[10.1109/JPROC.2021.3054390](https://doi.org/10.1109/JPROC.2021.3054390).
- [19] P. Rajpurkar, J. Irvin, R.L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C.P. Langlotz, B.N. Patel, K.W. Yeom, K. Shpanskaya, F.G. Blankenberg, J. Seekins, T.J. Amrhein, D.A. Mong, S.S. Halabi, E.J. Zucker, A.Y. Ng, M.P. Lungren, Deep learning for chest radiograph diagnosis: retrospective comparison of the chexnext algorithm to practicing radiologists, *PLoS Med.* 15 (2018), doi:[10.1371/journal.pmed.1002686](https://doi.org/10.1371/journal.pmed.1002686).
- [20] X. Qian, H. Fu, W. Shi, T. Chen, Y. Fu, F. Shan, X. Xue, M3lung-Sys: a deep learning system for multi-class lung pneumonia screening from ct imaging, *IEEE J Biomed Health Inform* 24 (2020) 3539–3550, doi:[10.1109/JBHI.2020.3030853](https://doi.org/10.1109/JBHI.2020.3030853).
- [21] L. Li, L. Qin, Z. Xu, Y. Yin, X. Wang, B. Kong, J. Bai, Y. Lu, Z. Fang, Q. Song, K. Cao, D. Liu, G. Wang, Q. Xu, X. Fang, S. Zhang, J. Xia, J. Xia, Using artificial intelligence to detect covid-19 and community-acquired pneumonia based on pulmonary ct: evaluation of the diagnostic accuracy, *Radiology* 296 (2020) E65–E71, doi:[10.1148/radiol.2020200905](https://doi.org/10.1148/radiol.2020200905).
- [22] R. Zhang, Z. Guo, Y. Sun, Q. Lu, Z. Xu, Z. Yao, M. Duan, S. Liu, Y. Ren, L. Huang, F. Zhou, Covid19xraynet: a two-step transfer learning model for the covid-19 detecting problem based on a limited number of chest x-ray images, *Interdiscip Sci* 12 (2020) 555–565, doi:[10.1007/S12539-020-00393-5](https://doi.org/10.1007/S12539-020-00393-5).
- [23] S. Chen, K. Ma, Y. Zheng, *Med3d: transfer learning for 3d medical image analysis*, arXiv (2019) 1–12.
- [24] M. Al-Shabi, B.L. Lan, W.Y. Chan, K.H. Ng, M. Tan, Lung nodule classification using deep local-global networks, *Int J Comput Assist Radiol Surg* 14 (2019), doi:[10.1007/s11548-019-01981-7](https://doi.org/10.1007/s11548-019-01981-7).
- [25] W. Zhao, J. Yang, Y. Sun, C. Li, W. Wu, L. Jin, Z. Yang, B. Ni, P. Gao, P. Wang, Y. Hua, M. Li, 3D deep learning from ct scans predicts tumor invasiveness of subcentimeter pulmonary adenocarcinomas, *Cancer Res.* 78 (2018) 6881–6889, doi:[10.1158/0008-5472.CAN-18-0696](https://doi.org/10.1158/0008-5472.CAN-18-0696).
- [26] J. Gong, J. Liu, W. Hao, S. Nie, B. Zheng, S. Wang, W. Peng, A deep residual learning network for predicting lung adenocarcinoma manifesting as ground-glass nodule on ct images, *Eur Radiol* 30 (2020) 1847–1855 <https://pubmed.ncbi.nlm.nih.gov/31811427/>, doi:[10.1007/s00330-019-06533-w](https://doi.org/10.1007/s00330-019-06533-w).
- [27] H. Ding, W. Xia, L. Zhang, Q. Mao, B. Cao, Y. Zhao, L. Xu, F. Jiang, G. Dong, Ct-based deep learning model for invasiveness classification and micropapillary pattern prediction within lung adenocarcinoma, *Front Oncol* 10 (2020) 1–9, doi:[10.3389/fonc.2020.01186](https://doi.org/10.3389/fonc.2020.01186).
- [28] M. Nishio, S. Koyasu, S. Noguchi, T. Kiguchi, K. Nakatsu, T. Akasaka, H. Yamada, K. Itoh, Automatic detection of acute ischemic stroke using non-contrast computed tomography and two-stage deep learning model, *Comput Methods Programs Biomed* 196 (2020) 105711, doi:[10.1016/j.cmpb.2020.105711](https://doi.org/10.1016/j.cmpb.2020.105711).
- [29] M.A. Ebrahimighahnavieh, S. Luo, R. Chiong, Deep learning to detect alzheimer's disease from neuroimaging: a systematic literature review, *Comput Methods Programs Biomed* 187 (2020) 105242, doi:[10.1016/j.cmpb.2019.105242](https://doi.org/10.1016/j.cmpb.2019.105242).
- [30] E.W. Prince, R. Whelan, D.M. Mirsky, N. Stence, S. Staulcup, P. Klimo, R.C. Anderson, T.N. Niazi, G. Grant, M. Souweidane, J.M. Johnston, E.M. Jackson, D.D. Limbrick, A. Smith, A. Drapeau, J.J. Chern, L. Kilburn, K. Ginn, R. Naf-

- tel, R. Dudley, E. Tyler-Kabara, G. Jallo, M.H. Handler, K. Jones, A.M. Donson, N.K. Foreman, T.C. Hankinson, Robust deep learning classification of adamantinomatous craniopharyngioma from limited preoperative radiographic images, *Sci Rep* 10 (2020) 1–13, doi:[10.1038/s41598-020-73278-8](https://doi.org/10.1038/s41598-020-73278-8).
- [31] A. Ebrahimi, S. Luo, R. Chiong, Deep sequence modelling for alzheimer's disease detection using mri, *Comput. Biol. Med.* 134 (2021), doi:[10.1016/j.combiomed.2021.104537](https://doi.org/10.1016/j.combiomed.2021.104537).
- [32] J. Wu, Y. Zhang, X. Tang, A joint 3d+2d fully convolutional framework for subcortical segmentation, in: in: Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention(MICCAI 2019), 11766, Springer, Cham, 2019.
- [33] S. Liu, D. Xu, S.K. Zhou, O. Pauly, S. Grbic, T. Mertelmeier, J. Wicklein, A. Jerebko, W. Cai, D. Comaniciu, 3d anisotropic hybrid network: Transferring convolutional features from 2d images to 3d anisotropic volumes, in: in: Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention(MICCAI 2018), 11071, Springer, Cham, 2018.
- [34] O. Ronneberger, P. Fischer, T. Brox, U-Net: convolutional networks for biomedical image segmentation, in: Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention(MICCAI 2015) 9351 (2015) 234–241, doi:[10.1007/978-3-319-24574-4_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- [35] R. Liu, J. Lehman, P. Molino, F.P. Such, E. Frank, A. Sergeev, J. Yosinski, An intriguing failing of convolutional neural networks and the coordconv solution, in: Advances in Neural Information Processing Systems (NIPS 2018), 2018, Curran Associates Inc., 2018, pp. 9628–9639.
- [36] S. Woo, J. Park, J.Y. Lee, I.S. Kweon, Cbam: Convolutional block attention module, in: in: Proc. European Conference on Computer Vision (ECCV 2018), 11211, Springer, Cham, 2018.
- [37] X. Li, W. Wang, X. Hu, J. Yang, Selective kernel networks, in: in: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2019), IEEE, 2019, pp. 510–519. <https://doi.org/10.1109/CVPR.2019.00060>.
- [38] J. Hu, L. Shen, S. Albanie, G. Sun, E. Wu, Squeeze-and-excitation networks, *IEEE Trans Pattern Anal Mach Intell* 42 (2020) 2011–2023, doi:[10.1109/TPAMI.2019.2913372](https://doi.org/10.1109/TPAMI.2019.2913372).
- [39] K.F. Kerr, M.D. Brown, K. Zhu, H. Janes, Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use, *J Clin Oncol* 34 (2016) 2534–2540, doi:[10.1200/JCO.2015.65.5654](https://doi.org/10.1200/JCO.2015.65.5654).
- [40] A.J. Vickers, E.B. Elkin, Decision curve analysis: a novel method for evaluating prediction models, *Med Decis Making* 26 (2006) 565–574, doi:[10.1177/0272989X06295361](https://doi.org/10.1177/0272989X06295361).
- [41] L.V. Maaten, G. Hinton, Visualizing data using t-sne, *Journal of Machine Learning Research* 9 (2008) 2579–2605.
- [42] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-cam: visual explanations from deep networks via gradient-based localization, *Int J Comput Vis* 128 (2020) 336–359, doi:[10.1007/s11263-019-01228-7](https://doi.org/10.1007/s11263-019-01228-7).
- [43] S.C. Huang, F.C. Cheng, Y.S. Chiu, Efficient contrast enhancement using adaptive gamma correction with weighting distribution, *IEEE Trans. Image Process.* 22 (2013) 1032–1041, doi:[10.1109/TIP.2012.2226047](https://doi.org/10.1109/TIP.2012.2226047).
- [44] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: in: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2016), IEEE, 2016, pp. 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- [45] C. Yang, A. Rangarajan, S. Ranka, Visual explanations from deep 3d convolutional neural networks for alzheimer's disease classification, *AMIA Annu Symp Proc.* (2018) 1571–1580.
- [46] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: a large-scale hierarchical image database, in: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009) (2009) 248–255 <https://doi.org/10.1109/CVPR.2009.5206848>, doi:[10.1109/CVPR.2009.5206848](https://doi.org/10.1109/CVPR.2009.5206848).
- [47] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, Pytorch: An imperative style, high-performance deep learning library, *Advances in Neural Information Processing Systems (NIPS 2019)*, 32, Curran Associates Inc., 2019.