Original Article

Deep-learning model for diagnostic clue: detecting the dural tail sign for meningiomas on contrast-enhanced T1 weighted images

Hyunmin Kim, Hyug-Gi Kim, Jang-Hoon Oh[#], Kyung Mi Lee[#]

Department of Radiology, Kyung Hee University Hospital, Kyung Hee University College of Medicine, Seoul, Republic of Korea

Contributions: (I) Conception and design: KM Lee, H Kim; (II) Administrative support: KM Lee, HG Kim; (III) Provision of study materials or patients: HG Kim, KM Lee; (IV) Collection and assembly of data: HG Kim, JH Oh; (V) Data analysis and interpretation: H Kim, JH Oh; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

[#]These authors contributed equally to this work.

Correspondence to: Kyung Mi Lee, MD, PhD; Jang-Hoon Oh, PhD. Department of Radiology, Kyung Hee University Hospital, Kyung Hee University College of Medicine, #23 Kyungheedae-ro, Dongdaemun-gu, Seoul 02447, Republic of Korea. Email: bandilee@khu.ac.kr; roineri5@gmail.com.

Background: Meningiomas are the most common primary central nervous system tumors, and magnetic resonance imaging (MRI), especially contrast-enhanced T1 weighted image (CE T1WI), is used as a fundamental imaging modality for the detection and analysis of the tumors. In this study, we propose an automated deep-learning model for meningioma detection using the dural tail sign.

Methods: The dataset included 123 patients with 3,824 dural tail signs on sagittal CE T1WI. The dataset was divided into training and test datasets based on specific time point, and 78 and 45 patients were comprised for the training and test dataset, respectively. To compensate for the small sample size of the training dataset, 39 additional patients with 69 dural tail signs from the open dataset were appended to the training dataset. A You Only Look Once (YOLO) v4 network was trained with sagittal CE T1WI to detect dural tail signs. The normal group dataset, comprised of 51 patients with no abnormal finding on MRI, was employed to evaluate the specificity of the trained model.

Results: The sensitivity and false positive average were 82.22% and 29.73, respectively, in the test dataset. The specificity and false positive average were 17.65% and 3.16, respectively, in the normal dataset. Most of the false-positive cases in the test dataset were enhancing vessels, misinterpreted as dural thickening.

Conclusions: The proposed model demonstrates an automated detection system for the dural tail sign to identify meningioma in general screening MRI. Our model can facilitate and alleviate radiologists' reading process by notifying the possibility of incidental dural mass based on dural tail sign detection.

Keywords: Meningioma; dural tail sign; magnetic resonance imaging (MRI); convolutional neural networks (CNNs); deep learning

Submitted Jan 30, 2023. Accepted for publication Sep 06, 2023. Published online Nov 03, 2023.
doi: 10.21037/qims-23-114
View this article at: https://dx.doi.org/10.21037/qims-23-114

Introduction

Meningiomas are the most common primary tumors of the central nervous system, accounting for 37% of all intracranial neoplasms (1). Most meningiomas (>80%) are benign, slow-growing lesions (WHO grade I) that can be effectively treated with complete surgical excision (2). Radiotherapy is another viable option when there is a residual tumor after surgery, when the location and size of the tumor are unfavorable for the surgical approach, or when the patient does not want surgery. For active

surveillance with or without radiotherapy, routine imaging follow-up and accompanying growth assessment are crucial. Magnetic resonance imaging (MRI) is the gold standard for diagnosis, surveillance, and treatment planning (3).

The "dural tail sign" is a classic radiologic finding of meningioma. The term "dural tail sign" was first used by Wilms *et al.* (4) in reference to meningioma, referring to the thickening of the dura adjacent to the tumor in contrast-enhanced T1 weighted image (CE T1WI) (5). This thickening has a tapering rim, with a smooth or slightly nodular internal surface and a length ranging from 0.5 to 3 cm (4). When the sign was first described, it was considered pathognomonic of meningioma, but it has since been reported in other diseases, including sarcoidosis, lymphoma, and metastases, as well as infectious, autoimmune, and vascular diseases (6). The prevalence of the dural tail sign in brain tumors ranges from 22% to 32% (7,8). Data regarding the exact prevalence of dural lesions is limited since most examples in the literature exist in case reports and small case series (9). Ghosal *et al.* (10), the largest study consist of 1,000 cases of dural-based lesions initially diagnosed as meningiomas, showed that only 2% of resected dural masses are other pathologies. Rokni-Yazdi *et al.* (8) reported that among 22 cases with intracranial masses exhibiting the dural tail sign, 81.8% (18 out of 22 cases) were meningiomas, and the prevalence of the dural tail sign in biopsy-proven meningioma was about 58.6%. Dural tail sign is not pathognomonic of meningioma, nor all meningiomas have dural tail sign. However, with high incidence of meningioma among brain tumors and high prevalence of the meningioma in the dural tail sign-exhibiting intracranial tumors, the detection of the dural tail sign on screening MRI is highly indicative of meningioma.

Recently, deep-learning technology, which has shown remarkable performance in medical image analyses, has been applied in neuroimaging analyses (11). Previous studies on meningioma have reported pronounced performance in tasks such as classifying different brain tumors (12-14) including glioma, meningioma, and pituitary gland tumor (15). Some recent studies investigated on detection and segmentation of meningiomas using deep learning model and have shown very high precision as high as 98% and strong correlation with manual segmentation (16-18). However, despite their outstanding performances, the deep learning models have a limitation in that humans cannot understand the intermediate processes of the models, which is referred as the black-box problem (19). In medical imaging, this lack of understanding in intermediate process, or lack of interpretability for humans in predictive models, undermine trust in such models and limit the clinical actionability of model predictions, which further undermines their usefulness to clinicians. In response to the significant challenges posed by black-box models, there has been a number of research in recent years in the field of explainable medical machine learning (20). Much of this research (21) is focused on creating intelligible explanations of how a model works and why it makes specific individual predictions by identifying the variables most driving model predictions.

The objective of this study is to develop a deep learning model for the dural tail sign using CE T1WI. To evaluate the performance of the dural tail sign detection model, both meningioma patients and healthy individuals were tested. Moreover, we observed the performance of the model in a different brain tumor group other than meningioma, and conducted external validation using an open dataset. To our knowledge, no studies have developed a fully automated deep-learning model for meningioma detection with primary use of the dural tail sign on CE T1WI. We present this article in accordance with the TRIPOD reporting checklist (available at https://qims.amegroups.com/article/view/10.21037/qims-23-114/rc).

Methods

Patient population

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This retrospective, single-center study was approved by the Institutional Review Board of Kyung Hee University, and the requirement for informed consent was waived. *Figure 1* shows a diagram of the study population. All the data were from an afflicted university hospital's registry (tertiary care) with the approval from the Institutional Review Board. It consists of a total of 183 patients with 248 CE T1WI scans who were incidentally diagnosed with intracranial meningioma between December 2009 and September 2022. The inclusion criteria were preoperative contrast-enhanced brain MRI including sagittal CE T1WI scans, and pathologic confirmation of meningioma after surgical resection. Sixty patients with 93 CE T1WI scans were excluded because their contrast-enhanced sagittal scans did not show the dural tail sign. For more convenient and consistent training purpose for the model, only sagittal CE T1WI images were used in the study. The dural tail sign dataset included 123 patients, with 155 sagittal CE T1WI

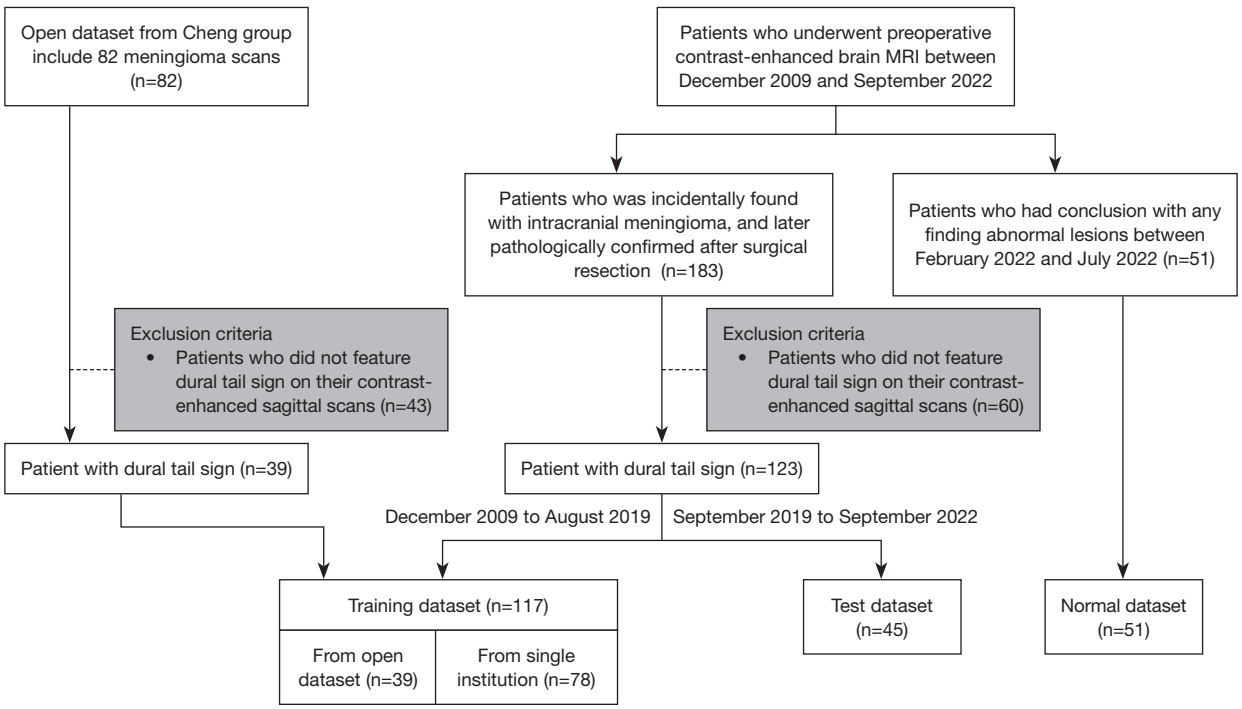

Figure 1 Diagram of the study population. MRI, magnetic resonance imaging.

Table 1 Characteristics of the training and temporal test datasets

Characteristics	Training set	Test dataset	Normal group
No. of individuals	78	45	51
No. of scans	110	45	51
No. of images	26,464	16,220	16,592
No. of labeled images	1,863	897	–
No. of dural tail sign labels	2,623	1,201	–
Age (year), mean ± SD	61.1±12.49	58.69±12.02	55.73±11.69
Age (year), range	27–82	26–76	29–78
Sex (female/male)	71/39	33/12	27/24

SD, standard deviation.

scans comprising a total of 42,684 slices. Among them, 2,760 slices contained 3,824 labeled boxes for dural tail signs. Then, the dural tail sign dataset was divided into the training and test datasets based on specific time points. For the training dataset, 78 patients (110 scans; 26,464 slices; including 1,863 slices with 2,623 labeled boxes for dural tail signs) were selected between December 2009 and August 2019. Only slices with dural tail signs were used for training. For the test dataset, 45 patients (45 scans; 16,220 slices; including 897 slices with 1,201 labeled boxes for dural tail signs) were included between September 2019 and September 2022. The details of these datasets are listed in *Table 1*. To compensate for the small sample size of the training dataset, we performed the labeling for confirmed meningioma cases from the Cheng group dataset (22-24). Additional 39 patients and their selected 69 CE 3DWI slices with the dural tail signs were included in the training dataset. To evaluate the specificity of the trained model, 51 patients (51 scans; 16,592 slices) with no abnormal finding on MRI were randomly selected as the normal group dataset between February 2022 and July 2022.

In this study, we performed two additional experiments to investigate distinguishability for other tumor and the generalizability of the model. To assess the ability to distinguish meningiomas from other types of tumors, a non-meningioma tumor dataset was collected, and details were described in Appendix 1. For generalizability, the meningioma dataset (25) from The Cancer Imaging Archive (TCIA) (26) was used as an external validation dataset and the details were listed in Appendix 2.

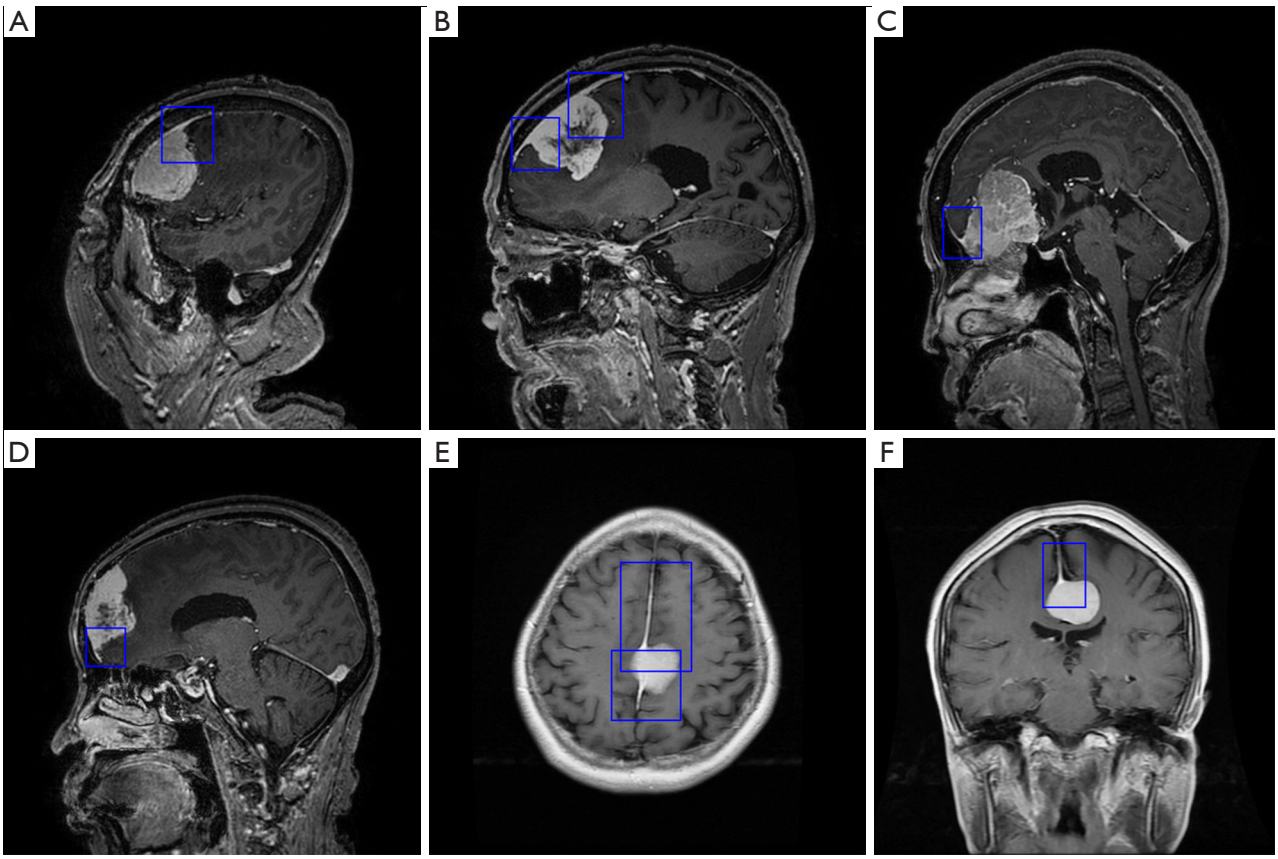

Figure 2 Labeling example of the dural tail sign on contrast-enhanced T1-weighted images. (A-D) Examples of dural tail sign labeling obtained from a medical institution and (E,F) obtained from an open dataset from the Cheng group. Dural tail sign labeling is represented by the blue-boundary boxes.

Image acquisition

For the in-house dataset, most MRI scans were acquired using a 3T MR system (MAGNETOM VIDA, Siemens, Erlangen, Germany). The MRI protocol for brain tumor included CE T1WI sequence, and for gadolinium (Gd)-enhanced imaging, a dose of 0.1 mmol/kg body weight of gadobenate dimeglumine (MultiHance, Bracco Diagnostics, Princeton, NJ, USA) was intravenously administered. The imaging parameters were as follows: repetition time (TR), 2,000 ms; echo time (TE), 3 ms; slice thickness, 0.9 mm; flip angle, 120°; matrix size, 512×512; field of view, 230×230 mm^2; voxel size, 0.8×0.8 mm^2.

Data labeling

For the object detection deep-learning algorithm, two expert neuroradiologists with 10 and 30 years of experience reviewed the images independently, detected extra-axial brain tumors on the CE T1WI in the training and external validation dataset, and labeled dural tail signs in every single slice with a rectangular region of interest boundary box using the Image Labeler application included in the Matlab program (Math Works, R2020b, Natick, MA, USA). One reader, of 10 years of the experience, reviewed the images and labelled the dural tail sign with rectangular region of interest boundary box first. Then, the second reader, of 30 years of the experience, reviewed the images and the pre-drawn region of interest boundary boxes. Any conflicts or discrepancies in the labeling were settled down through discussion between the two readers. *Figure 2* shows an example of dural tail sign labeling. To prevent the misdetection of enhancing vessels as false positives, dural tail sign label boxes were drawn to include a portion of the meningioma lesion.

Training the proposed deep-learning algorithm

A major drawback of utilizing deep-learning algorithms in meningioma detection is its varying anatomic locations, especially for lesions in proximity to important vascular structures such as the venous sinus, prominent cortical veins, or the circle of Willis. To reduce such false-positive cases, skull stripping was performed using SPM 12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12/) for all test dataset cases before evaluation. The skull stripping process comprises segmenting the skull on non-enhanced T1WI, generating brain mask using the gray matter, white matter, and cerebrospinal fluid maps. After that, the T1WI with segmented images and brain mask were co-registered to the segmented image on CE T1WI. As a final step, the trained model predicted the dural tail sign on CE T1WI, and the prediction boxes in locations other than brain tissue were eliminated.

You Only Look Once (YOLO) v4 (27), a state-of-the-art convolutional neural network (CNN) object detection algorithm that can simultaneously detect the location of objects in input images and classify them into different categories, was used for the deep-learning algorithm architecture. We used the "Deep Learning Toolbox" in Matlab to train the YOLOv4 model, and a YOLO v4 network was initialized using the transfer learning method based on a pre-trained DarkNet53 (27) with the following parameters: nine anchor boxes, Adam optimizer, mini-batch size of 4, initial learning rate of 1×10^{-3}, factor for L2 regularization of 5×10^{-4}, and 50 epochs at maximum. All parameters not described in the manuscript were set to default values. The data augmentation was randomly applied including left-to-right flipping (RandXReflection in MATLAB) and image rotation (0, 90, 180, 270 degrees; in-house function). All images with at least one dural tail sign were used as the input data for training the YOLO v4 network.

All preprocessing, labeling, and training processes were performed on a single-server computer running a Windows operating system (Windows Server 2016) with a double NVIDIA V100 GPU with 32 GB of memory (Nvidia Corporation). All image labeling, processing, and training networks were based on Matlab (MathWorks, R2020a, Natick, MA, USA).

Performance evaluation of the deep-learning algorithm

We assumed that radiologists do not require every slice with the dural tail sign to make a diagnosis and that only one slice is sufficient to recognize the dural tail sign. To develop the dural tail sign detection model as an assistant software under this assumption, the algorithm's performance was evaluated at the subject level, not at the slice level. For the meningioma test dataset, any subjects in which the ground truth overlapped with the prediction boxes in one or more than one slices were evaluated as true positive; otherwise, it was evaluated as a false negative. For the normal dataset, a subject that had none of the predicted boxes by the trained model was evaluated as a true negative; otherwise, it was evaluated as a false positive (i.e., if one or more dural tail signs were predicted by the model, then it was considered a false positive). For evaluation of the model performance, the sensitivity and false-positive average were calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN} \quad [1]$$

$$Specificity = \frac{TN}{TN + FP} \quad [2]$$

$$FP_{avg} = \frac{FP_{slice}}{N} \quad [3]$$

(TP = true positive, FP = false positive, FN = false negative, TN = true negative, FP_{avg} = false positive in slice levels averaged by the number of individuals, FP_{slice} = number of false positive in slice level, and N = the number of individuals)

TP was determined when the intersection over union (IoU) between the predicted box and ground truth was >0.1. To investigate the performance related to the confidence score for each boundary box that is predicted by the trained model, the performance was evaluated with a confidence score cutoff between 0.5 and 0.95 with 0.05 increments.

Results

Table 2 presents the results of dural tail sign detection for the test and normal datasets. The sensitivity and false-positive average were 82.22% and 29.73, respectively, for the test dataset, and the specificity and false-positive average were 17.65% and 3.16, respectively, for the normal dataset, with confidence score cutoff of 0.5. The IoU between the predicted DTS boxes and ground truth were measured in the range from 0.59 to 0.86, and the mean IoU was 0.72. *Figures 3*,*4* show examples of true and false positives for the test and normal datasets, respectively, at the slice level. The majority of false-positive cases in the test dataset were enhancing vessels, misinterpreted as dural thickening. For

Table 2 Performance on applying the cutoff predicted confidence scores

Confidence score cutoff	Test dataset		Normal group	
	Sensitivity, %	FP_{avg}	Specificity, %	FP_{avg}
0.5	82.22	29.73	17.65	3.16
0.55	80.00	27.93	25.49	2.63
0.6	80.00	25.89	29.41	2.28
0.65	80.00	23.82	37.25	1.88
0.7	77.78	21.36	45.10	1.63
0.75	75.56	19.00	52.94	1.24
0.8	75.56	16.82	58.82	0.96
0.85	73.33	13.89	62.75	0.69
0.9	68.89	10.58	68.63	0.49
0.95	55.56	5.44	94.12	0.08

FP_{avg}, false positive average.

the normal dataset, most false-positives were detected around the transverse sinus and superior sagittal sinus regions. There was no missing data during the process.

The confidence score of the predicted results, which can affect the performance by applying the cutoff threshold, shows different distributions between true positives in the test dataset and false positives in the normal dataset (*Figure 5*). The average confidence scores of the true positives in the test dataset and false positives in the normal dataset were 0.85 and 0.71, respectively, and the sensitivity was decreased to 26.67% for the test dataset, and the specificity was increased to 76.47% for the normal dataset, while increasing the cutoff threshold from 0.5 to 0.95.

For the external validation, a sensitivity and false-positive average were measured as 36.84% and 9.21, respectively. Table S1 demonstrated the manufacturer and scan parameters for internal and external validation datasets.

For the other tumor dataset, false positives were observed around the brain tumor lesions in 26 scans out of 45 scans and some false positives were due to peritumoral enhancing vessels in clearly parenchymal locations, not near the dura. Figure S1 shows examples of these particular false-positive cases.

Discussion

Artificial intelligence-based image analysis has been extensively applied in neuro-oncology in recent years, primarily in gliomas, through the use of radiomics (16). Such models (28-30) have been developed by extracting features from MR images through a CNN and then applying supervised learning to predict the tumor grade, genotype, and prognosis. Few studies (31-34) have addressed meningiomas, and existing deep-learning models for meningioma have focused on distinguishing malignant meningiomas from nonmalignant meningiomas or differentiating meningiomas from other tumors. More recent studies (16,35) on meningioma have attempted to achieve an automated volumetric measurement of meningioma for accurate tumor volume measurement in serial MR images. Laukamp *et al.* (16) investigated the automated detection and segmentation of meningiomas using a deep-learning model and reported excellent detection accuracy as high as >98%. However, even these studies dealt with only pathologically confirmed meningioma cases, thus the true detection accuracy of any routine brain MR scan remains still unknown.

In this study, we proposed a fully automated detection program that can directly facilitate the radiologic reading process by rendering an easy and fast screening tool. A meningioma is an extra-axial dural-based tumor with broad dural attachments on the edges. Based on this characteristic feature of the tumor along with its high incidence, the presence of a dural tail sign is highly indicative of meningioma, and is suitable as differential diagnostic imaging evidence for the tumor. Our deep-learning model detected the dural tail sign using YOLO v4, and showed reasonable sensitivity (82.22%) with a false-positive average of 29.73 for the test dataset. Our model was not only tested on pathologically confirmed meningioma cases, but also on the normal dataset to better demonstrate its superior initial detection performance.

Our model can be an efficient adjuvant tool for the radiologists in routine screening MRI, since the model can point out the priority patients and lesions before the actual manual reading process. In clinical settings where large amount of screening brain MRI is done, such as in South Korea where screening brain MRI in symptomatic patients is mostly covered with public medical insurance, radiologists are exposed to laborious and tedious process of detecting incidental lesions through many of normal MRI scans. Our model can be used before actual manual reading process, and give an advance warning, if present, of possible incidental dural lesion. It can significantly reduce the searching times and alleviate the fatigue of the radiologists from the large workload.

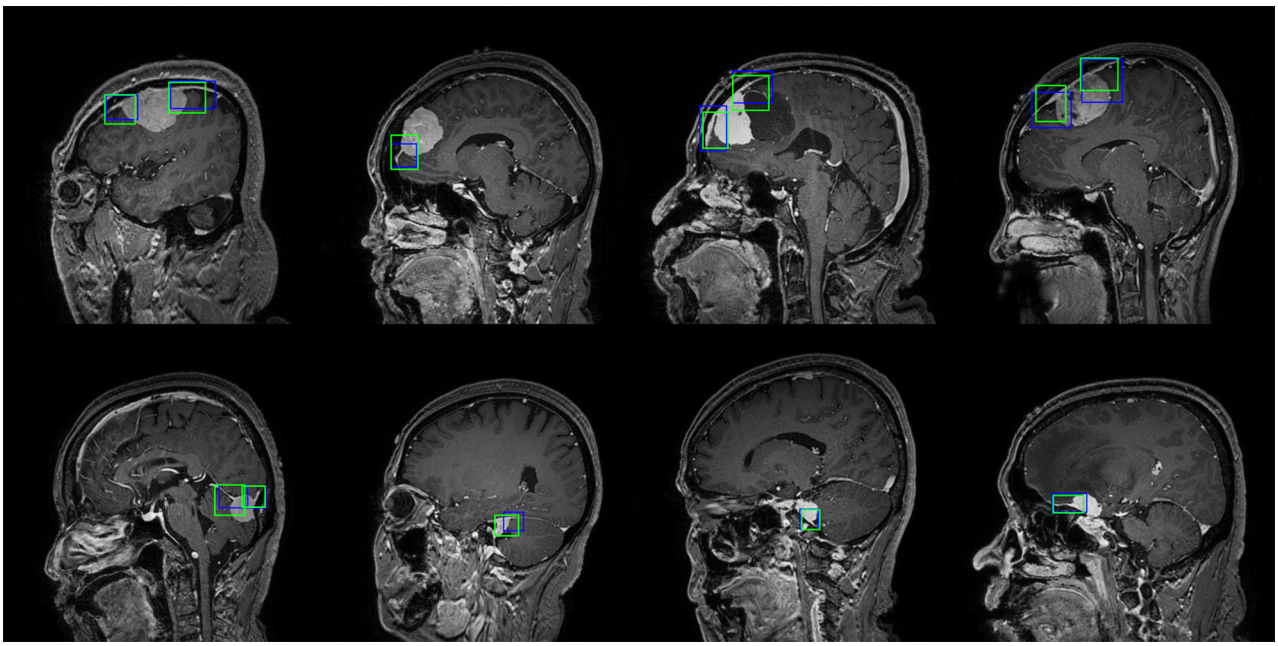

Figure 3 Examples of true positive cases from the test dataset. The ground truth and predicted box are represented by the blue- and green-boundary boxes, respectively.

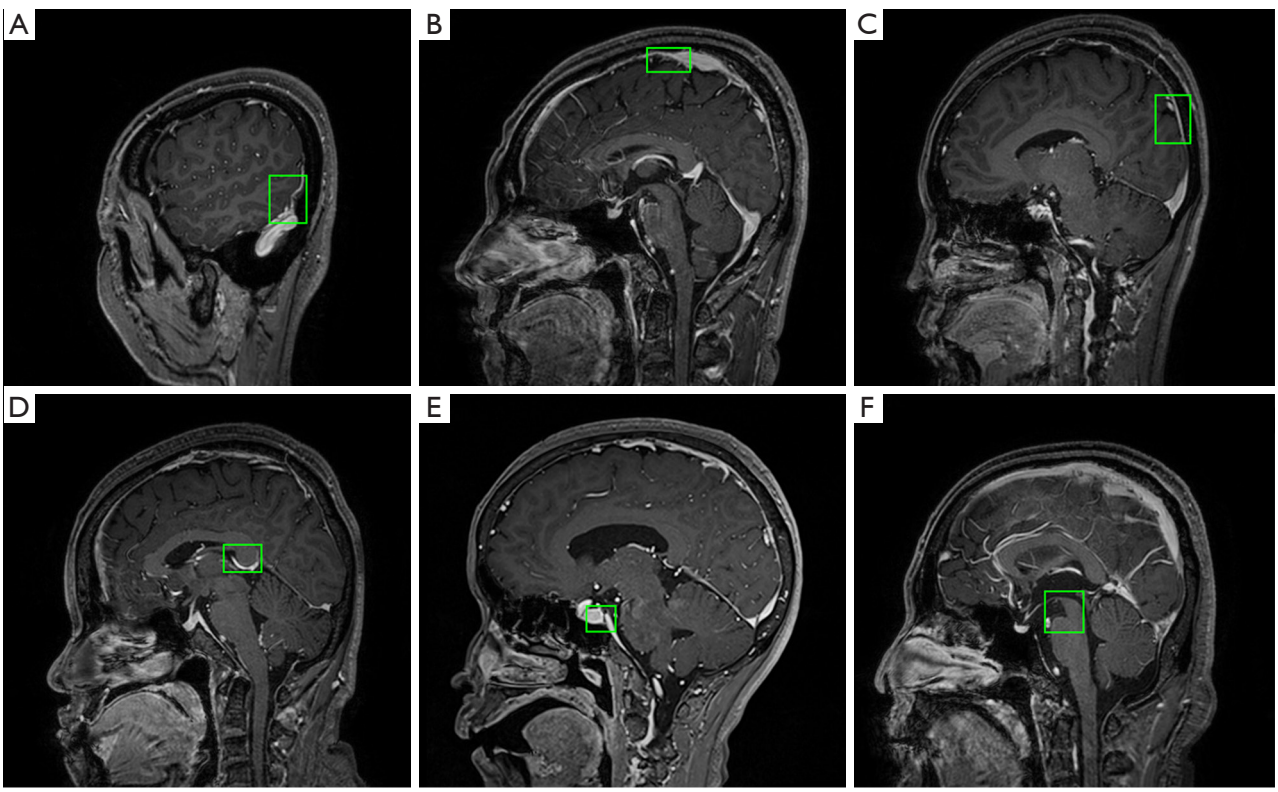

Figure 4 Examples of false-positive cases from the normal dataset. False positive around the (A) transverse sinus, (B) superior sagittal sinus, and (C) enhanced meningeal vessel are represented by the green-boundary boxes; (D-F) other examples observed at low frequencies.

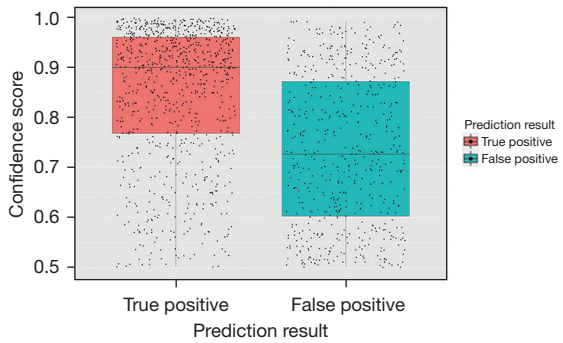

Figure 5 Distribution of predicted confidence scores for the meningioma and normal group datasets. The confidence score of the predicted box evaluated as true positive in the test dataset is recorded as a point (left), and the confidence score of the predicted box evaluated as false positive in the normal dataset is recorded as a point (right).

Accurate detection of a cerebral tumor is clinically important as it permits the preselection of priority lesions and patients, and the model's performance can affect patient's clinical decision. A well-trained deep-learning model can provide reliability with good performance. According to one recent study (21) about deep-learning model utilized in medical imaging, a developed model gives out prediction results very similar to long-experienced radiologic experts, and the model's performance was even more concordant with the expert group than less experienced group. By developing deep learning model focusing on dural tail sign detection, we are trying to lay the foundation stone for the development of a brain tumor detection model based on diagnostic clue. Recently, there has been much effort to develop a medical deep learning model with transparency and apprehensibility, reliable enough to be used in real clinical setting. One of the ways is providing some kind of explainability measure like class activation maps (CAM) (36) to demonstrate how the model is making the classification. Many XAI (Explainable Artificial Intelligence) studies provide explanations by providing the activated regions within the layers of deep learning models, such as CAM. According to the recent research (37), 65% of XAI studies reported using CAM. However, even if we know which regions are activated by using CAM, it can still be difficult to understand why a certain decision was made. For this reason, we plan to develop a framework that makes decisions based on diagnostic clues, and explainable AI that provides evidence-based reasoning for the final decision can be a solution to the black-box problem. Our model itself does not provide a solution to the black-box problem, and it is aiming for further research which focuses on detection and classification meningioma based on the dural tail sign detection.

The main purpose of this research is to present a deep-learning model capable of detecting dural tail sign. Our goal is to develop a framework that makes decisions based on diagnostic clues, and this model is an initiative point for us to develop a meningioma classification model with medical relevance and transparency. In the prospective studies, we are planning to further expand the model's role from simple detection of dural tail sign to a screening tool for meningioma. With improved diagnostic performance of the model, the model's efficacy can be further evaluated through comparison of reading time and diagnostic performance between reader groups with and without deep learning model.

Skull stripping (38) or brain extraction of MR images is often a fundamental step in many neuroimaging processing pipelines. It usually results in a binary brain mask of an MR image after the removal of nonbrain structures, such as eyes, fat, bone, and marrow. In this study, we generated the brain mask by skull strip processing, and applied brain mask to eliminate false positives from nonbrain regions. Combining the object detection algorithm with skull stripping, 121 and 115 false positives were eliminated from the test and normal datasets, respectively, and the specificity was increased by 5.8%.

For the experiment for other tumor dataset, we hypothesized that with relatively low incidence of dural tail sign in non-meningioma tumors, and the model would not detect dural tail signs in the other tumor dataset. However, false positives were detected in more than half of the other tumor dataset. For those cases in which the model mistakenly detects peritumoral vessels as dural tail sign, we can solve this problem by training the model to learn the overall anatomy of brain and recognize the constitutional structure of dura in prospective studies. Also, there are some dural-based brain lesions, other than meningioma, also featuring the dural tail sign, and the differential diagnosis can be difficult. In such confusing cases, other sequences or other imaging modalities can be helpful. For example, most dural metastases exhibit reduced perfusion when compared with meningiomas, with typical rCBV values of less than 2. On MRS, metastases have low NAA:creatine ratios and high lipid:creatine rations without the alanine peak characteristic of meningiomas. Primary dural lymphomas are more commonly associated with larger

extent of vasogenic edema than meningiomas, and tend to show more diffusion restriction due to high cellularity (9). Moreover, meningiomas and surrounding tissues may appear heterogeneous and less distinguishable when peritumoral edema and/or necrosis are present in proximity to the enhancing tumor. Prospective studies with our dural tail sign detection model combined with other analyzing models trained on other MR sequences or imaging modalities would better distinguish meningioma from other brain tumors, and can serve as a clinically relevant assistant tool with even more radiological evidences.

For the external validation, our model shows poor performances, and the performance degradation in external validation can be attributed to inductive bias (39), which refers to the assumptions that a model relies on when inferring on unseen data during training process. Based on the inductive bias of CNN (40), differences in the pattern of the dural tail sign between the validation dataset and the test dataset could result in differences in pixel locality, and the difference of image resolution can be a cause of performance degradation in evaluating deep learning models. Furthermore, differences in the manufacturer and scan parameters (41,42) may also affect the non-biological variance. Prospective studies applying the fine-tuning technique (43), training the model with a larger dataset from multiple institutions, including various manufacturers, and using the model based on vision transformer (44), which minimizes the inductive bias, would improve the model's generalizability.

This study had several limitations. First of all, our model was trained primarily with MRI data acquired from a single institution. Although an open dataset was added to the training dataset, the sample size was still small. Second, the dural tail sign is not always present in the brain MR images of patients with meningiomas. In fact, significant portions of our model's study population (60 out of 183 patients from our hospital database, and 43 out of 82 for the open dataset from Cheng group) were excluded because these patients' images did not show the dural tail sign. If the radiologists are not meticulous, the relatively low sensitivity of the model may entail a considerably high risk of missing the tumor. While enjoying the convenience of the model's preselection of priority patients and lesions on screening MRI, the radiologists should still stay focused and be able to detect any missing lesions through manual reading process. Further, if the radiologists are notified by the model that there is a positive dural tail sign, they may be inclined to describe the lesion as a meningioma even though numerous other intracranial pathologies can exhibit dural tail sign. However, the main purpose of our model was to detect the dural tail sign in screening MRI, if present, highly indicative of meningioma. This model is an assistant tool for radiologists in routine screening, not a definite diagnostic tool. Prospective studies with larger dataset consist of a variety of other intracranial entities featuring the dural tail sign would improve the model's performance. Another limiting factor in this study was that only sagittal scans were used for training and test datasets, except for the open dataset. Finally, hyperparameter optimization was not performed in this study. The hyperparameters used in this deep learning model training were default parameters, and the model performance could be improved by performing hyperparameter optimization.

Conclusions

In this study, we proposed an automated deep-learning model to detect dural tail signs, which is highly indicative of meningiomas. It can be used as a medically evident and apprehensible artificial intelligence model for meningioma detection. Through preselection of priority patients and lesions suspicious for such dural mass, our model can reduce the workload of radiologists and facilitate the manual reading process with convenience and high reproducibility.

Acknowledgments

This research was the result of a study on the "HPC Support" Project, supported by the "Ministry of Science and ICT" and National IT Industry Promotion Agency (NIPA).

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Government of South Korea (Ministry of Science and ICT, MSIT) (NRF-2020R1C1C1006623 to KML, NRF-2021R1F1A1050515 to HGK).

Footnote

Reporting Checklist: The authors have completed the TRIPOD reporting checklist. Available at https://qims.amegroups.com/article/view/10.21037/qims-23-114/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://qims.amegroups.com/article/view/10.21037/qims-23-114/coif).

The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). This retrospective, single-center study was approved by the Institutional Review Board of Kyung Hee University, and the requirement for informed consent was waived.

References

1. Ostrom QT, Cioffi G, Waite K, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014-2018. Neuro Oncol 2021;23:iii1-105.
2. Neromyliotis E, Kalamatianos T, Paschalis A, Komaitis S, Fountas KN, Kapsalaki EZ, Stranjalis G, Tsougos I. Machine Learning in Meningioma MRI: Past to Present. A Narrative Review. J Magn Reson Imaging 2022;55:48-60.
3. Goldbrunner R, Stavrinou P, Jenkinson MD, Sahm F, Mawrin C, Weber DC, Preusser M, Minniti G, Lund-Johansen M, Lefranc F, Houdart E, Sallabanda K, Le Rhun E, Nieuwenhuizen D, Tabatabai G, Soffietti R, Weller M. EANO guideline on the diagnosis and management of meningiomas. Neuro Oncol 2021;23:1821-34.
4. Wilms G, Lammens M, Marchal G, Van Calenbergh F, Plets C, Van Fraeyenhoven L, Baert AL. Thickening of dura surrounding meningiomas: MR features. J Comput Assist Tomogr 1989;13:763-8.
5. Rokni-Yazdi H, Azmoudeh Ardalan F, Asadzandi Z, Sotoudeh H, Shakiba M, Adibi A, Ayatollahi H, Rahmani M. Pathologic significance of the "dural tail sign". Eur J Radiol 2009;70:10-6.
6. Gu H, Zhang X, di Russo P, Zhao X, Xu T. The Current State of Radiomics for Meningiomas: Promises and Challenges. Front Oncol 2020;10:567736.
7. Weber DC, Lovblad KO, Rogers L. New pathology classification, imagery techniques and prospective trials for meningiomas: the future looks bright. Curr Opin Neurol 2010;23:563-70.
8. Rokni-Yazdi H, Sotoudeh H. Prevalence of "dural tail sign" in patients with different intracranial pathologies. Eur J Radiol 2006;60:42-5.
9. Lyndon D, Lansley JA, Evanson J, Krishnan AS. Dural masses: meningiomas and their mimics. Insights Imaging 2019;10:11.
10. Ghosal N, Dadlani R, Gupta K, Furtado SV, Hegde AS. A clinicopathological study of diagnostically challenging meningioma mimics. J Neurooncol 2012;106:339-52.
11. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, Geis JR, Pandharipande PV, Brink JA, Dreyer KJ. Current Applications and Future Impact of Machine Learning in Radiology. Radiology 2018;288:318-28.
12. Chakrabarty S, Sotiras A, Milchenko M, LaMontagne P, Hileman M, Marcus D. MRI-based Identification and Classification of Major Intracranial Tumor Types by Using a 3D Convolutional Neural Network: A Retrospective Multi-institutional Analysis. Radiol Artif Intell 2021;3:e200301.
13. Gao P, Shan W, Guo Y, Wang Y, Sun R, Cai J, Li H, Chan WS, Liu P, Yi L, Zhang S, Li W, Jiang T, He K, Wu Z. Development and Validation of a Deep Learning Model for Brain Tumor Diagnosis and Classification Using Magnetic Resonance Imaging. JAMA Netw Open 2022;5:e2225608.
14. Dixon L, Jandu GK, Sidpra J, Mankad K. Diagnostic accuracy of qualitative MRI in 550 paediatric brain tumours: evaluating current practice in the computational era. Quant Imaging Med Surg 2022;12:131-43.
15. Rao SKV, Lingappa B. Image Analysis for MRI Based Brain Tumour Detection Using Hybrid Segmentation and Deep Learning Classification Technique. Int J Intell Eng Syst 2019;12:53-62.
16. Laukamp KR, Thiele F, Shakirin G, Zopfs D, Faymonville A, Timmer M, Maintz D, Perkuhn M, Borggrefe J. Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI. Eur Radiol 2019;29:124-32.
17. Sukumaran A, Abraham A. Automated Detection and Classification of Meningioma Tumor from MR Images Using Sea Lion Optimization and Deep Learning Models. Axioms 2022;11:15.
18. Liang J, Yang C, Zeng M, Wang X. TransConver:

transformer and convolution parallel network for developing automatic brain tumor segmentation in MRI images. Quant Imaging Med Surg 2022;12:2397-415.
19. von Eschenbach WJ. Transparency and the Black Box Problem: Why We Do Not Trust AI. Philos Technol 2021;34:1607-22.
20. Rudin C, Chen C, Chen Z, Huang H, Semenova L, Zhong C. Interpretable machine learning: Fundamental principles and 10 grand challenges. Stat Surv 2022;16:1-85.
21. Oh JH, Kim HG, Lee KM, Ryu CW. Reliable quality assurance of X-ray mammography scanner by evaluation the standard mammography phantom image using an interpretable deep learning model. Eur J Radiol 2022;154:110369.
22. Cheng J, Huang W, Cao S, Yang R, Yang W, Yun Z, Wang Z, Feng Q. Enhanced Performance of Brain Tumor Classification via Tumor Region Augmentation and Partition. PLoS One 2015;10:e0140381. Erratum in: PLoS One 2015;10:e0144479.
23. Cheng J, Yang W, Huang M, Huang W, Jiang J, Zhou Y, Yang R, Zhao J, Feng Y, Feng Q, Chen W. Retrieval of Brain Tumors by Adaptive Spatial Pooling and Fisher Vector Representation. PLoS One 2016;11:e0157112.
24. Cheng J. Brain tumor dataset. figshare; 2017. DOI: 10.6084/m9.figshare.1512427.v5. [Date of access: Apr 5, 2023].
25. Vassantachart A, Cao Y, Shen Z, Cheng K, Gribble M, Ye JC, Zada G, Hurth K, Mathew A, Guzman S, Yang W. Segmentation and Classification of Grade I and II Meningiomas from Magnetic Resonance Imaging: An Open Annotated Dataset (Meningioma-SEG-CLASS) (Version 1). Cancer Imaging Arch 2023. Doi: 10.1002/mp.16763. [Date of access: Apr 12, 2023].
26. Clark K, Vendt B, Smith K, Freymann J, Kirby J, Koppel P, Moore S, Phillips S, Maffitt D, Pringle M, Tarbox L, Prior F. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. J Digit Imaging 2013;26:1045-57.
27. Bochkovskiy A, Wang CY, Liao HYM. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv preprint 2020 Apr 22; arXiv: 2004.10934. [Date of access: Jan 27, 2022].
28. Choi YS, Bae S, Chang JH, Kang SG, Kim SH, Kim J, Rim TH, Choi SH, Jain R, Lee SK. Fully automated hybrid approach to predict the IDH mutation status of gliomas via deep learning and radiomics. Neuro Oncol 2021;23:304-13.
29. Choi KS, Choi SH, Jeong B. Prediction of IDH genotype in gliomas with dynamic susceptibility contrast perfusion MR imaging using an explainable recurrent neural network. Neuro Oncol 2019;21:1197-209.
30. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, Zhai G. A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. Sci Rep 2017;7:10353.
31. Chen C, Cheng Y, Xu J, Zhang T, Shu X, Huang W, Hua Y, Zhang Y, Teng Y, Zhang L, Xu J. Automatic Meningioma Segmentation and Grading Prediction: A Hybrid Deep-Learning Method. J Pers Med 2021;11:786.
32. Lu Y, Liu L, Luan S, Xiong J, Geng D, Yin B. The diagnostic value of texture analysis in predicting WHO grades of meningiomas based on ADC maps: an attempt using decision tree and decision forest. Eur Radiol 2019;29:1318-28.
33. Bouget D, Pedersen A, Hosainey SAM, Vanel J, Solheim O, Reinertsen I. Fast meningioma segmentation in T1-weighted magnetic resonance imaging volumes using a lightweight 3D deep learning architecture. J Med Imaging (Bellingham) 2021;8:024002.
34. Hale AT, Stonko DP, Wang L, Strother MK, Chambless LB. Machine learning analyses can differentiate meningioma grade by features on magnetic resonance imaging. Neurosurg Focus 2018;45:E4.
35. Kang H, Witanto JN, Pratama K, Lee D, Choi KS, Choi SH, Kim KM, Kim MS, Kim JW, Kim YH, Park SJ, Park CK. Fully Automated MRI Segmentation and Volumetric Measurement of Intracranial Meningioma Using Deep Learning. J Magn Reson Imaging 2023;57:871-81.
36. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning Deep Features for Discriminative Localization. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2016: 2921-9.
37. Groen AM, Kraan R, Amirkhan SF, Daams JG, Maas M. A systematic review on the use of explainability in deep learning systems for computer aided diagnosis in radiology: Limited use of explainable AI? Eur J Radiol 2022;157:110592.
38. Roy S, Butman JA, Pham DL; Alzheimers Disease Neuroimaging Initiative. Robust skull stripping using multiple MR image contrasts insensitive to pathology. Neuroimage 2017;146:132-47.
39. Battaglia PW, Hamrick JB, Bapst V, Sanchez-Gonzalez A, Zambaldi V, Malinowski M, et al. Relational inductive biases, deep learning, and graph networks. arXiv preprint 2018 Jun 4;1-40. arXiv: 1806.01261. [Date of access: Apr 12, 2023].

40. Sabottke CF, Spieler BM. The Effect of Image Resolution on Deep Learning in Radiography. Radiol Artif Intell 2020;2:e190015.
41. Dinsdale NK, Jenkinson M, Namburete AIL. Deep learning-based unlearning of dataset bias for MRI harmonisation and confound removal. Neuroimage 2021;228:117689.
42. Yan W, Huang L, Xia L, Gu S, Yan F, Wang Y, Tao Q. MRI Manufacturer Shift and Adaptation: Increasing the Generalizability of Deep Learning Segmentation for MR Images Acquired with Different Scanners. Radiol Artif Intell 2020;2:e190195.
43. Yin X, Chen W, Wu X, Yue H. Fine-tuning and visualization of convolutional neural networks. In: 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE; 2017:1310-5. DOI: 10.1109/ICIEA.2017.8283041. [Date of access: Apr 20, 2023].
44. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv preprint 2020 Oct 22. arXiv: 2010.11929. [Date of access: Apr 20, 2023].

Cite this article as: Kim H, Kim HG, Oh JH, Lee KM. Deep-learning model for diagnostic clue: detecting the dural tail sign for meningiomas on contrast-enhanced T1 weighted images. Quant Imaging Med Surg 2023;13(12):8132-8143. doi: 10.21037/qims-23-114

Appendix 1 Additional experiment to investigate the performance of dural tail sign detection model for other tumors

Our model was also tested on other tumor groups to verify its ability to distinguish meningiomas from other tumors. The hypothesis of the additional experiment was that most other types of brain tumors do not exhibit the dural tail sign, unlike meningiomas, and therefore the trained model may not be able to predict the dural tail sign in other tumor dataset.

The other tumor group dataset consists of 41 patients (45 scan; 15,260 images) who were pathologically confirmed as having non-meningioma tumors (e.g. metastatic carcinoma, glioblastoma, and adenoma), with brain magnetic resonance imaging (MRI) taken from single institution between March 2016 and June 2022. The MRI acquisition for the other tumor dataset was the same as the dural tail sign dataset and normal dataset in the manuscript, and the preprocessing such as applying the brain mask for false positive reduction were performed the same as the test and normal dataset.

For evaluation, to investigate whether that the trained model detects dural tail signs around other tumor lesions, we manually evaluated the predicted boxes where the dural tail sign was detected around the lesion as false positives, instead of evaluating entire false positives for each scan.

As a result, false positives were observed around the tumor lesion in 26 patients, and some false positives around other tumor lesion were due to peritumoral enhancing vessels in clearly parenchymal locations, not near the dura.

Appendix 2 External validation

The total 96 scans in meningioma dataset from The Cancer Imaging Archive, we excluded 20 scans without the observation of dural tail sign through the labeling process, and selected 76 scans for the external validation set. The external validation dataset consisted mostly of axial contrast-enhanced 3 dimensional T1 weighted image (CE 3D T1WI) with an image matrix size of 256×256. Each image was resliced into the sagittal plane for evaluation using the trained model. We performed the evaluation process for the external validation set using the same procedure as that used for the internal test dataset, excluding the brain masking process because the external validation dataset does not include 3D T1WI. As a results for the external validation, a sensitivity and false positive average were measured as 36.84% and 9.21, respectively.

Table S1 Manufacture and scan parameters for internal and external validation datasets

Manufacturer	Internal validation dataset (testset; 45 scans)	External validation dataset (76 scans)				
	Philips Medical Systems	GE Medical Systems	GE Medical Systems	GE Medical Systems	Siemens	Toshiba
Manufacturer model	Achieva	DISCOVERY MR750	SIGNA EXCITE	Signa HDxt	Symphony	Titan
Study description	T1 3D TFE SAG-Prohence	Ax FSPGR BRAVO post	AX 3D FSPGR, AX 3D stealth	AX 3D stealth	t1_mpr_ns_ax stealth	3D T1 AX STEALTH
Number of scans	45	1	23	50	1	1
Field strength (T)	3.0	3.0	3.0	3.0	1.5	1.5
Repetition time (ms)	9.43	8.13	7.94	8.00	2.14	6.20
Echo time (ms)	4.60	3.16	3.06	3.06	3.93	3.20
Inversion time (ms)	–	450.00	450.00	450.00	1,100.00	600.00
Resolution	512×512	256×256	256×256	256×256	256×256	256×256
Pixel spacing	0.49	1.02	1.02	1.01	1.02	1.00
Intensity range	1,933.04	16,091.00	18,513.52	19,620.82	352.00	4,295.00

https://dx.doi.org/10.21037/qims-23-114

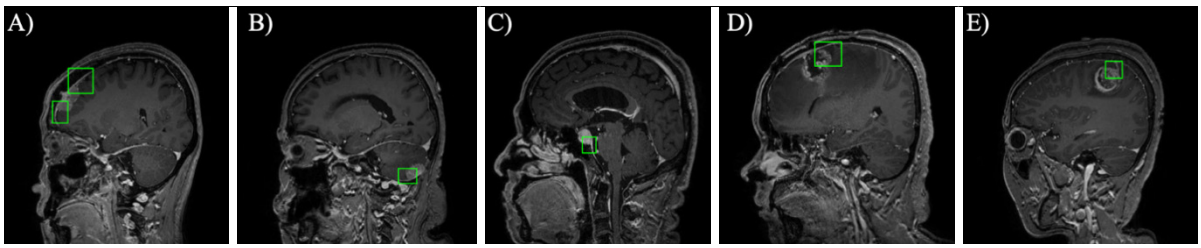

Figure S1 Examples of false-positive cases from the normal dataset. False positive around the (A) transverse sinus, (B) superior sagittal sinus, and (C) enhanced meningeal vessel are represented by the green-boundary boxes; (D-F) other examples observed at low frequencies.

https://dx.doi.org/10.21037/qims-23-114