Deep neural networks allow expert-level brain meningioma detection, segmentation and improvement of current clinical practice

Deep learning for brain meningioma segmentation

Alessandro Boaro[1,3,8], Jakub R. Kaczmarzyk[2,4,8], Vasileios K. Kavouridis[1], Maya Harary[1,5], Marco Mammi[1], Hassan Dawood[1,6], Alice Shea[7], Elise Y. Cho[1], Parikshit Juvekar[6], Thomas Noh[6], Aakanksha Rana[1,2], Satrajit Ghosh[2,9]*, Omar Arnaout[1,6,9]

[1] Computational Neuroscience Outcomes Center, Brigham and Women's Hospital, Harvard Medical School, Boston MA
[2] McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge MA
[3] Institute of Neurosurgery, Department of Neurosciences, Biomedicine and Movement Sciences, University of Verona, Verona, Italy
[4] Medical Scientist Training Program, Stony Brook University School of Medicine, Stony Brook NY
[5] Department of Neurosurgery, University of California Los Angeles, Los Angeles CA
[6] Department of Neurosurgery, Brigham and Women's Hospital, Harvard Medical School, Boston MA
[7] Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston MA
[8] These authors share first authorship
[9] These authors share senior authorship

*** Corresponding author:**
Satrajit Ghosh PhD
McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge MA
Phone:
Email: satra@mit.edu

Abstract

Background. Accurate brain meningioma detection, segmentation and volumetric assessment are critical for serial patient follow-up, surgical planning and monitoring response to treatment. Current gold standard of manual labeling is a time-consuming process, subject to inter-user variability. Fully-automated algorithms for meningioma detection and segmentation have the potential to bring volumetric analysis into clinical and research workflows by increasing accuracy and efficiency, reducing inter-user variability and saving time. Previous research has focused solely on segmentation tasks without assessment of impact and usability of deep learning solutions in clinical practice.

Methods. We developed a three-dimensional convolutional neural network (3D-CNN) to perform expert-level, automated meningioma segmentation and volume estimation on MRI scans. A 3D-CNN was initially trained by segmenting entire brain volumes using a dataset of 10,099 healthy brain MRIs. Using transfer learning, the network was then specifically trained on meningioma segmentation using 806 expert-labeled MRIs. The algorithm tumor-labeling performance was assessed with standard metrics of tumor segmentation performance (i.e., Dice score). To evaluate clinical applicability, we compared volume estimation accuracy and segmentation time based on current practice versus the use of our automated algorithm.

Findings. The final model achieved a median performance of 88.2% reaching the spectrum of current inter-expert variability (82.6% - 91.6%). Compared to current workflows, the use of the algorithm reduced processing time by 99% and produced tumor volume calculations with an almost perfect correlation with the expert manual segmentations ($r=0.98$, $p<0.001$), significantly more accurate compared to volume estimation techniques used in practice.

Conclusions. We demonstrate through a prospective trial conducted in a simulated setting that a deep learning approach to meningioma segmentation is feasible, highly accurate and can substantially improve current clinical practice.

Introduction

Meningiomas are the most common primary intracranial neoplasms accounting for 37% of all primary brain tumors.[1] They arise from the meninges and are benign in the vast majority of cases.[2] Magnetic resonance imaging (MRI), and T1-weighted gadolinium contrast-enhanced sequences in particular, are the cornerstone of the diagnostic evaluation of meningiomas. A large number of scans are produced and examined throughout the life of any given patient for the purposes of initial diagnosis, clinical surveillance, surgical planning and post-operative assessment of residual tumor.[3-5] Currently, assessment of meningioma imaging relies on manual techniques for tumor size and growth estimation, most commonly using unidimensional measurements in two or three orthogonal planes,[6] which are often disregarded in favor of visual estimation of the tumor's dimensions. Such approaches commonly lead to misjudgment of tumor growth and true dimensions and in some cases, due to their small volume, tumors are even completely missed. Manual volumetric segmentation is possible, typically with the help of third-party software, but is subject to considerable inter-rater variability[7] and represents a time-consuming task that is often incompatible with the busy workflows of clinicians.

Advances in computing power and a gradual refinement of algorithm architectures have resulted in increased use of machine learning (ML) and deep learning (DL) techniques in healthcare.[8] A specific class of DL architectures known as deep convolutional neural networks (CNN), in particular, has revolutionized imaging analysis.[9] The impressive success of these networks in disparate tasks, such as diabetic retinopathy or skin cancer classification,[10,11] has sparked an intense interest in employing them for other medical applications.[12] Although there has been a considerable body of research focusing on the implementation of CNN segmentation algorithms for a number of brain pathologies, most notably glioblastoma,[13,14] there is still a dearth of evidence in current literature on the application of end-to-end DL solutions for meningioma segmentation and management. Moreover, most brain tumor segmentation algorithms have focused on the algorithm's labeling performance alone, and have not assessed the applicability and impact of such systems in real clinical scenarios.[15,16] It is important to note that the readiness of a tumor segmentation algorithm for clinical use is not defined by 100% accuracy, but rather by demonstrating performance at the same or higher level as the human experts who currently perform the task, taking into account the naturally occurring inter-expert variability. We therefore designed a study with the goal of developing a 3D-CNN algorithm able to automatically segment meningiomas from MRI scans at clinical expert level, and specifically offer objective measures of impact in a real-world clinical setting, focusing on the accuracy of tumor segmentation, volume estimation and time saving compared to current practice.

Methods

Image acquisition and preprocessing

Healthy brain dataset: A dataset of 10,099 high-resolution, T1-weighted MRI scans of normal human brains was assembled from public and private sources.[17,18] This dataset contained images from a heterogenous human population and of varying acquisition quality. All scans were conformed to have 256 slices in each dimension and $1mm^3$ isotropic voxels using the image processing software FreeSurfer.[19] FreeSurfer's recon-all tool was used to generate parcellations and segmentations for brain structures in each scan. The dataset was separated into a training/validation set.

Meningioma dataset: We screened all adult patients diagnosed with intracranial meningioma in two major US academic hospitals during the time period 2004-2018 for inclusion in our study. Patients with no evidence of meningioma on MRI, unavailable pre-surgical MRI scans, radiation-induced meningioma, unavailable contrast-enhanced images, or no high-resolution sequences were excluded, resulting in a final cohort of 806 exams, containing a total of 936 tumors. The exams were downloaded in DICOM format and stored in an institutionally protected, HIPAA compliant shared drive. Per protocol, high resolution (at least 100 slices on axial plane), T1 contrast-enhanced sequences were identified from our final cohort and converted to NIfTI format in order to ensure patient de-identification.
The meningioma dataset was divided in a training/validation set (746 scans) and a test dataset of 60 exams containing a total of 67 tumors. The test set was constructed in order to ensure adequate representation of all the most relevant tumor locations, sizes and shapes.
All relevant ethical regulations were followed as part of the conduct of this study. We operated under institutional review board approval (protocol number 2015P002352).

Ground truth tumor labeling
We used the open-source image-processing software 3D Slicer to produce expert-level meningioma segmentations. 3D Slicer provides both manual and semi-automated labeling tools that the experts could choose from to perform the task. All tumors were segmented by two experts separately and inter-rater variability was calculated. The labels were reviewed by a third expert in order to produce the final dataset. The segmentations were stored in an institutional shared drive and named using progressive numeration; the link between each exam and the corresponding patient was securely stored in a separate file. Location, size and number of tumors per scan were stored in a dedicated spreadsheet for subsequent analysis.

CNN model design and evaluation
Brain extraction network: A standard three-dimensional U-Net was used as the underlying architecture for our segmentation algorithm. Olaf Ronneberger et al. originally designed the UNet architecture for a bio-medical image segmentation problem. It has been observed to be more successful than other conventional CNN models, in terms of architecture and in terms pixel/voxel-based segmentations from convolutional neural network layers.[20]
The U-Net architecture is primarily composed of an encoder and a decoder. The encoder can be seen as a contraction path - a stack of convolutional and max pooling layers which captures and encodes the contextual information present in the input. The decoder is an expansion path that is used to decode the accurate localization using the learned feature mapping from the input and up-convolutional layers (Fig. 1).
The model was initially trained with random weight initialization for a brain-extraction task on the dataset of 10,000 T1-weighted MRIs of healthy brains. To improve generalizability and robustness of the network, each pair of MRI and labels had a fifty-percent chance of being augmented using random rotation and translation (the transformation for pairs of features and labels was the same). The augmented MRIs were interpolated tri-linearly, and the augmented reference labels were interpolated using nearest neighbor. The voxels of each MRI scan were standard scored. Due to the GPU computational limitations, the MRI scans and labels were separated into eight non-overlapping cubes of equal size per volume. Model training was done on these cubes in batches of six. The model was trained using the Adam optimizer[21] with an initial learning rate of 0.00001 and Jaccard loss for five epochs across three NVIDIA GTX 1080Ti graphical processing units (GPUs). Each batch was distributed evenly to each GPU. The model was implemented using the Nobrainer framework, which wraps the Keras API of TensorFlow.[22] The random augmentation and interpolations were implemented in the Nobrainer framework using TensorFlow. Dice similarity scores were

calculated and used to evaluate the algorithm's performance in brain extraction. The model is publicly available in the Keras format at https://github.com/neuronets/trained-models/tree/master/neuronets/ams/0.1.0.

Meningioma extraction network: The refined brain-extraction network was re-trained for the specific task of meningioma segmentation. The meningioma training/validation dataset underwent the same preprocessing pipeline as the normal brain scans, with the only difference that the tumor-containing scans were not randomly augmented. The MRI scans and labels were separated into eight non-overlapping cubes of equal size per volume. The cubes of the MRI scans that did not contain any meningioma voxels were excluded from the training set. To prevent large deviations from the pre-trained weights, L2 regularization with coefficient 0.001 was applied to the weights of each layer and a low initial learning rate of 0.00001 was used. The model was trained using the Adam optimizer[21] and the Jaccard loss for 194 epochs with a batch size of two cubes using a single NVIDIA GTX 1080Ti GPU. Validation on whole MRIs was performed by standard-scoring the MRI voxel values, separating the MRI into eight non-overlapping cubes of equal size, running a forward pass of the model on each cube, and combining the eight cubes of predictions into a complete volume. The algorithm's segmentation performance was assessed by computing the Dice coefficient between the tumor model prediction and the expert label for each MRI in the test set.
The model was also trained using the meningioma dataset alone from random weight initialization (i.e., without transfer learning). This model was abandoned at its early stage due to its clear inferiority in the segmentation performance compared to the algorithm adopting transfer learning described in this work.

Expert segmentation variability range
Three different experts (one neurosurgeon and two experts in neuroimaging) were given the task to independently segment the tumors of the test set, and the segmentations were compared to the ground truth. Mean and median tumor segmentation results were calculated for each expert and used to create a range of expert segmentation variability compatible with a real-world clinical scenario. The model performance was compared to the obtained range of values.

Clinical simulation
To obtain the most clinically meaningful metrics as to whether our deep learning segmentation algorithm can effectively impact the current clinical practice by improving physicians' accuracy in calculating meningioma volume and reducing the time needed to produce high-quality, readily usable tumor segmentations, we performed a prospective trial in a simulated clinical setting. In the trial we pitted the algorithm against currently accepted methods of tumor volume assessment and measured the difference in time between clinical experts and our algorithm to produce volumetric tumor segmentations (described below).

Tumor volume estimation: We compared our algorithm's volume estimation accuracy to the two most widely employed tumor volume estimation techniques, based on two-dimensional (2D) and three-dimensional (3D) tumor volume calculation. The 2D technique is based on the selection of the MRI axial slice containing the largest tumor cross-section, followed by measurement of the tumor's two largest diameters, called tumor length and tumor width, respectively. The volume is then calculated as follows:

$$V = L * W * W / 2$$

where V is the estimated tumor volume, L is the tumor length and W is the tumor width.
In the 3D-based technique, three slices containing the largest tumor cross-section in the three traditional MRI visualization planes (sagittal, axial, coronal) are chosen and the largest diameter on each plane is measured, called length, width and height of the tumor, respectively. The volume is then calculated as follows:

$$V = pi / 6 * L * W * H$$

Where V is the tumor volume, L is the largest sagittal diameter, W is the largest axial diameter, and H is the largest coronal diameter. These two techniques were applied to each MRI scan of the test set. The tumor volumes from the algorithm's output labels on the same test set were calculated as well. The volumes resulting from these three methods were compared to those obtained from the ground truth labels and the correlation between each pair of volumes was calculated using the r statistic (2D to ground truth, 3D to ground truth, algorithm-volume to ground truth).

Segmentation time analysis: Two certified experts, one neuro-radiologist and one neurosurgeon, who regularly work with brain MRIs containing meningiomas for diagnostic, surgical planning and patient follow-up purposes were given the task of timing themselves in producing an accurate, readily usable meningioma label starting from an unedited MRI scan for each exam included in the test set. They were instructed to open the exam and work on the segmentation of the lesion as they would normally do, with the software tools they were most comfortable with (brush, region-growing tool, pencil and eraser). Once they finished with an exam, they would save the file, close it, and proceed to the next one. Each event was timed from the moment the exam was opened and ready in the viewing software (i.e., 3D Slicer) until the segmentation was considered complete by the expert. Concurrently, our algorithm was used to produce the tumor labels of the same MRI scans and the related run time was stored. The average time difference was calculated between each expert and the algorithm. Given the non-normal distribution of the three sets of measurements, the Wilcoxon signed-rank test was applied to assess statistical significance.

Results

Dataset creation and algorithm design

Our dataset consisted of 10,099 T1-weighted healthy brain MRIs, assembled from public and private sources, and 806 contrast-enhanced T1-weighted meningioma MRIs, representing 936 unique tumors, from the radiological repositories of two major academic hospitals under institutional review board approval. Details on data extraction, preprocessing and tumor labeling are found in the Methods section. In brief, tumor-containing MRIs were screened based on radiological or histological evidence of meningioma; high resolution brain scans were independently segmented by two experts. Sixty MRIs containing a total of 67 tumors were randomly held out as a test set; the rest of the database was used for training. A three-dimensional U-Net was used as the underlying architecture for our segmentation algorithm[23] (Fig. 1). The 10,099 normal brain MRIs served to train and validate the 3D-CNN for the task of brain extraction. The model was then fine-tuned via transfer learning to label brain meningiomas using 746 meningioma scans. The algorithm tumor-labeling performance was assessed with standard metrics of tumor segmentation performance (i.e., tumor Dice score). Three additional experts independently segmented the tumors of the test set in order to provide a measure of inter-expert variability.

Algorithm performance

The tumors in the test set had a mean volume of 13.11 cc (range 0.37 - 85.0 cc) and all intracranial locations commonly harbouring meningiomas were adequately represented: cranial vault (56.7%), skull base (25.3%), falx (13.4%), posterior fossa (4.5%)(Fig. 2). Five MRIs contained two tumors and one MRI contained three. The final model achieved a mean tumor segmentation Dice Score of 84% and a median of 88.2% on the entire test set, comparable to the inter-expert variability in segmenting the same tumors with means ranging from 79.5% to 89.9% and medians between 82.6% to 91.6% (Table 1 and Supplementary Table 1). Overall, the model missed two small tumors (volume < 1cc) whereas two out of three independent experts missed four each.

	> 2cc tumors		**All tumors**	
Pairs for comparison	Mean Dice score (%)	Median Dice Score (%)	Mean Dice score (%)	Median Dice Score (%)
Performance				
Model/Ground	87.7	89.6	84	88.2
Model/Expert 1	86.9	88.7	82.7	86.4
Model/Expert 2	87	90.2	83.6	88
Model/Expert 3	85.2	89.3	80	85.1
Inter-Expert Variability				
Ground/Expert 1	87.6	89.1	85.7	86.8
Ground/Expert 2	91.2	92.5	89.9	91.6
Ground/Expert 3	89.5	90.5	86	89.5
Expert 1/Expert 2	86.9	88.9	84.6	86
Expert 1/Expert 3	84.1	86.9	79.5	82.6
Expert 2/Expert 3	90.1	92.7	87	89.7

Table 1. **Algorithm performance and inter-expert variability.** Comparison of tumor segmentation performance (Dice scores) between algorithm output, ground truth and clinical experts, expressed as mean and median for all tumors and tumors of volume >2cc. *Ground*: Ground truth

We identified the patients who would be most likely to benefit from the use of a segmentation algorithm (i.e. those with the need of close radiological monitoring or higher chance of undergoing surgical resection) as those with tumors larger than 2cc; this cutoff is significantly stricter compared to the size of a clinically relevant meningioma based on the literature (~4 cc).[24-26] In this group of 40 patients, model and experts obtained equivalent performance, with the model demonstrating a mean tumor segmentation accuracy of 87.7% (median 89.6) and an inter-expert variability ranging between 84.1 and 91.2% (median range: 86.9 - 92.5) (Table 1, Fig. 3). There was a significantly positive correlation between tumor Dice scores and tumor volume with a steep performance increase in the 0-2 cc volume range ($r=0.61$, $p<0.001$, Fig. 4a).

Measures of clinical impact

To evaluate the effect of the use of an expert-level meningioma segmentation algorithm on real-world clinical workflow, we designed a simulated clinical scenario to compare volume estimation accuracy and segmentation time based on current practice versus the use of our automated algorithm. In brief, two certified experts, one neuro-radiologist and one neurosurgeon, were asked to manually segment each tumor in the test set while timing

themselves, and tumor volumes were calculated with 2D/3D estimation techniques currently used in clinical practice (setting and details reported in the Methods section). Notably, the incorporation of our algorithm reduced segmentation time by 99% (two seconds per segmentation on average, $p<0.001$, Table 2), and produced tumor volume calculations with an almost perfect correlation with the expert manual segmentations ($r=0.98$, $p<0.001$, Fig. 4b, Supplementary Table 2), significantly more accurate compared to the gold standard 2D and 3D volume estimation techniques ($r=0.88$ $p<0.001$ and $r=0.96$, $p<0.001$, respectively, Fig. 4c, 4d, Supplementary Table 2). Currently used 2D/3D estimation techniques were found to overestimate tumor volume by a minimum of 1.5 to a maximum of 7 times compared to the algorithm-based calculation (Table 3).

	Mean (s)	**SD (s)**	**Time reduction (%)**	**p value**
All tumors				
Manual 1	142.4	93.9	98.7	**< 0.001**
Manual 2	299.8	274.9	99.4	**< 0.001**
Automated	1.88	0.001	-	ref
>2cc				
Manual 1	178.2	96.2	98.9	**< 0.001**
Manual 2	391.5	296.9	99.5	**< 0.001**
Automated	1.88	0.001	-	ref
<2cc				
Manual 1	70.9	17.3	97.3	**< 0.001**
Manual 2	116.4	23.6	98.3	**< 0.001**
Automated	1.88	0.001	-	ref

Table 2. Manual versus automated tumor segmentation times. Comparison between the average time needed by experts **(Manual 1** and **Manual 2)** and our algorithm (**Automated**) to produce high-quality tumor segmentations. The algorithm saved 98% of time on average and the difference between the algorithm and each expert in each tumor group reached significance ($p<0.001$). *s:* seconds, *ref*: reference

Model output	**Avg volume difference (cc)**		
VS	All tumors	>2cc tumors	<2cc tumors
Ground truth	1.56	2.2	0.3
Expert 1	1.65	2.34	0.27
Expert 2	1.81	2.58	0.28
Expert 3	1.63	2.3	0.29
2D estimation	12.2	18	0.6
3D estimation	8.3	12.2	0.5

Table 3. Average tumor volume differences between techniques. The table presents the average tumor volume differences between algorithm output and respectively the ground

truth (**Ground truth**), expert manual segmentations (**Expert 1, Expert 2, Expert 3**) and 2D/3D volume estimation techniques (**2D estimation, 3D estimation**).

Discussion

Here, we demonstrate the development of an end-to-end DL solution for automated brain meningioma segmentation. The system was trained on the most heterogeneous MRI dataset of expert-labeled meningiomas available in the literature and found to have excellent detection and segmentation performance compared to human experts. Efficiency metrics were calculated in a simulated clinical environment compared to a radiologist, a neurosurgeon and two clinical experts in neuroimaging. The 3D-CNN algorithm performed at the same level of human experts while outperforming comparable automated or semiautomated segmentation algorithms.[27-29] Specifically, the model has well surpassed current practices in tumor volume estimation while bringing the time needed for tumor segmentation virtually to zero. The implementation of this algorithm in the clinic, in the operative room and in the radiology visualization suite would allow the automation of the tumor segmentation process while at the same time provide both physicians and patients with higher quality and significantly more accurate tumor information (Fig. 5).

Although others have described segmentation algorithms with good performance, these have been based on smaller datasets, which either only include patients with larger tumors (average volume 30 cc) and/or those undergoing surgical resection,[28,29] thereby excluding almost 50% of patients affected by meningiomas who normally undergo conservative MRI surveillance.[30] Such inclusion criteria, alongside complex pre-processing pipelines, limit the implementation of these algorithms in a real-world clinical population. In order to optimize the applicability and generalizability of our system, we (1) limited the artificial curation of the dataset by keeping image pre-processing to a minimum, simply ensuring standardization between MRI scans produced by different machines (more details in the Methods section), (2) included MRIs with no restrictions on tumor number, size, and location and (3) included patients who underwent both conservative management or surgical treatment. These features make our dataset more representative of the variability inherent in the meningioma patient population.

Despite a growing interest in the application of DL segmentation algorithms to medical imaging, few studies have evaluated the feasibility of applying these tools in daily clinical scenarios.[31,32] Our goal was to bridge the gap between quantitative research and clinical practice. Thus, we chose to evaluate our system not purely based on its performance in the segmentation task but also with real-world metrics. With a tumor detection and segmentation performance comparable to that obtained by clinical experts, a superior volumetric calculation accuracy compared to current 2D/3D estimations and a rapid, completely automated process, our algorithm could represent a drastic improvement in meningioma patient management. It could allow for more accurate tumor information acquisition and growth monitoring which would in turn improve clinical decision-making as well as monitoring of treatment response.

As with all DL algorithms, performance and generalizability improve with enrichment of the dataset. In the context of this work, this will be accomplished through validation on data from other institutions to help control for any institutional biases in terms of MR acquisition and patient population. An additional step towards full implementation of our tool into clinical work-flow would entail the evaluation of the algorithm performance in segmenting residual

or recurrent meningioma in the setting of early and late post-operative changes. Both these steps to optimize the algorithm are currently underway.

In addition to the immediate relevance of a fully automated DL segmentation tool both in everyday practice and clinical trials, such models are important upstream components in more complicated ML pipelines that have the potential to assist in predicting parameters such as tumor consistency, histological grade, growth trajectory and probability of recurrence even before a therapeutic plan is made and discussed with the patient.[33-35] Such systems could enhance clinical decision-making strategies, patient counseling and follow-up in unprecedented ways. To ensure that the development of DL algorithms for medical imaging analysis is more translational in nature, the impact of these systems needs to be validated on simulations of actual clinical tasks in the setting of a well-designed clinical trial. This will allow the selection and eventual implementation of the best performing algorithms which, with the necessary consideration of patient privacy and safety and the clearance provided by the dedicated certification and standardization pathways,[36,37] will be able to effectively advance medical care and improve patient outcomes.[38,39] Most important, the medical community will have to be an active participant in this process, working alongside computer scientists in order to steer the development towards the most relevant clinical questions, guarantee the rigorous validation of these algorithms and their seamless integration in clinical practice.

References:

1. Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, Barnholtz-SloanQ.T. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the united states in 2011-2015. Neuro Oncol. 2018;20: 1-86
2. Backer-Grondahl, T., Moen, B.H., Torp, S.H. The histopathological spectrum of human meningiomas. Int J Clin Exp Pathol. 2012;5: 231-42
3. Wu A, Garcia MA, Magill ST, Chen W, Vasudevan HN, Perry A. et al. Presenting symptoms and prognostic factors for symptomatic outcomes following resection of meningioma. World Neurosurg. 2018;111: e149-e59
4. Oya, S., Kim, S.H., Sade, B., Lee, J.H. The natural history of intracranial meningiomas. J Neurosurg. 2011;114: 1250-6
5. Islim AI, Mohan M, Moon RDC, Srikandarajah N, Mills SJ, Brodbelt AR, et al. Incidental intracranial meningiomas: a systematic review and meta-analysis of prognostic factors and outcomes. J Neuro-Oncol. 2019;142: 211-21
6. Zeidman, L.A., Ankenbrandt, W.J., Du, H., Paleologos, N., Vick, N.A. Growth rate of non-operated meningiomas. J Neurol. 2008;255: 891-5
7. Huang RY, Unadkat P, Bi WL, George E, Preusser M, McCracken JD et al. Response assessment of meningioma: 1D, 2D and volumetric criteria for treatment response and tumor progression. Neuro-Oncology 2018;21: 234-241
8. Yu, K.H., Beam, A.L., Kohane, I.S. Artificial intelligence in healthcare. Nat Biomed Eng. 2018;2: 719-31
9. LeCun, Y., Bengio, Y., Hinton, G. Deep learning. Nature 2015;521: 436-44
10. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A et al. Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs. JAMA 2016;316: 2402-10
11. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017;542: 115-8
12. Senders JT, Zaki MM, Karhade AV, Chang B, Gormley WB, Broekman ML et al. An introduction and overview of machine learning in neurosurgical care. Acta Neurochir. 2018;160: 29-38
13. Akkus, Z., Galimzianova, A., Hoogi, A., Rubin, D.L., Erickson, B.J. Deep learning for brain MRI segmentation: state of the art and future directions. J Digit Imaging. 2017;30: 449-59
14. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med Image Anal. 2017;36: 61-78
15. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y et al. Brain tumor segmentation with Deep Neural Networks. Med Image Anal. 2017;35: 18-31
16. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging 2015;34: 1993-2024
17. Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K The autism brain imaging data exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. Mol Psychiatry 2014;19: 659-67
18. Taylor JR, Williams N, Cusack R, Auer T, Shafto MA, Dixon M et al. The Cambridge centre for ageing and neuroscience (Cam-CAN) data repository: structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. NeuroImage 2017;144: 262-269
19. Fischl, B. Freesurfer. Neuroimage 2012;62: 774-81

20. Ronneberger, O., Fischer, P., Brox, T. U-Net: convolutional networks for biomedical image segmentation. Preprint at: https://arxiv.org/abs/1505.04597v1
21. Kingma, D.P. and Ba, J. Adam: A method for stochastic optimization. Preprint at: https://arxiv.org/abs/1412.6980v9
22. Abadi, M., et al. Tensorflow: A system for large-scale machine learning. Preprint at: https://arxiv.org/abs/1605.08695v2
23. Çiçek, Ö., et al. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. Preprint at: https://arxiv.org/abs/1606.06650v1
24. Romani, R., Ryan, G., Benner, C., Pollock, J. Non-operative meningiomas: long-term follow-up of 136 patients. Acta Neurochir. 2018;160: 1547-1553
25. Romani, R., Laakso, A., Kangasniemi, M., Lehecka, M., Hernesniemi, J. Lateral supraorbital approach applied to anterior clinoidal meningiomas: experience with 73 consecutive patients. Neurosurgery 2011;68: 1632-47
26. Romani, R., Laakso, A., Kangasniemi, M., Niemelä, M., Hernesniemi, J. Lateral supraorbital approach applied to tuberculum sellae meningiomas: experience with 52 consecutive patients. Neurosurgery 2012;70: 1504-18
27. Zhang H, Mo J, Jiang H, Li Z, Hu W, Zhang C et al. Deep learning model for the automated detection and histopathological prediction of meningioma. Neuroinformatics 2020; doi: 10.1007/s12021-020-09492-6
28. Laukamp KR, Thiele F, Shakirin G, Zopfs D, Faymonville A, Timmer M, Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI. Eur Radiol. 2019;29: 124–132
29. Hsieh TM, Liu YM, Liao CC, Xiao F, Chiang IJ, Wong JM. Automatic segmentation of meningioma from non-contrasted brain MRI integrating fuzzy clustering and region growing. BMC Med Inform Decis Mak. 2011; doi: 10.1186/1472-6947-11-54, 11-54
30. Cahill, K.S., Claus, E.B. Treatment and survival of patients with nonmalignant intracranial meningioma: results from the surveillance, epidemiology, and end results program of the national cancer institute. clinical article. J Neurosurg. 2011;115: 259-267
31. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y et al. Brain tumor segmentation with deep neural networks. Med Image Anal. 2017;35: 18-31
32. Yasaka K, Abe O. Deep learning and artificial intelligence in radiology: Current applications and future directions. PLoS Med. 2018;15(11):e1002707
33. Chartrand G, Cheng PM, Vorontsov E, Drozdzal M, Turcotte S, Pal CJ et al. Deep learning: a primer for radiologists. Radiographics 2017;37: 2113-2131
34. Zhu Y, Man C, Gong L, Dong D, Yu X, Wang S et al. A deep learning radiomics model for preoperative grading in meningioma. Eur J Radiol. 2019;116: 128–134
35. Banzato T, Causin F, Della Puppa A, Cester G, Mazzai L, Zotti A. Accuracy of deep learning to differentiate the histopathological grading of meningiomas on MR images: a preliminary study. J Magn Reson Imaging. 2019;50: 1152-1159
36. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med. 2019;25: 30-36
37. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K et al. A guide to deep learning in healthcare. Nat Med. 2019;25: 24-29
38. Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H., Aerts, H.J.W.L. Artificial intelligence in radiology. Nat Rev Cancer. 2018;18: 500-510
39. Wang, S., Summers, R.M. Machine learning and radiology. Med Image Anal. 2012;16: 933-51

Acknowledgements
J.R.K was partially supported by R01EB020740, P41EB019936, and T32GM008444. S.G. was partially supported by R01EB020740, P41EB019936, and RF1MH121885. A.R. was supported by RF1MH121885. Shreya Chawla, Sharmila Devi and Ali Ansaripour from the Computational Neuroscience Outcomes Center provided support for data extraction and pre-processing. We are grateful to Margherita De Marzio and Enrico Maiorino for their independent and precious feedback.

Financial Disclosure statement
The authors received no specific funding for this work.

Competing interests
The authors have no competing interests to disclose related to this work.

Data Availability
The majority of the healthy brain MRI dataset used and analyzed during the current study is publicly available at https://sensein.github.io/open-data-processing/. Restrictions apply to part of the healthy brain MRI dataset and the whole meningioma MRI dataset as derived from private, institutional databases and contain information that could enable patient identification. The authors will consider request to access to training and testing data on individual basis. Any data use will be restricted to non-commercial research purposes, and the data will only be made available upon execution of appropriate data use agreements.

Code availability
As reported in the Methods section of the manuscript, the code for the algorithm development and evaluation is publicly available at: https://github.com/neuronets/ams and the pretrained model is available at https://github.com/neuronets/trained-models/tree/master/neuronets/ams/0.1.0.

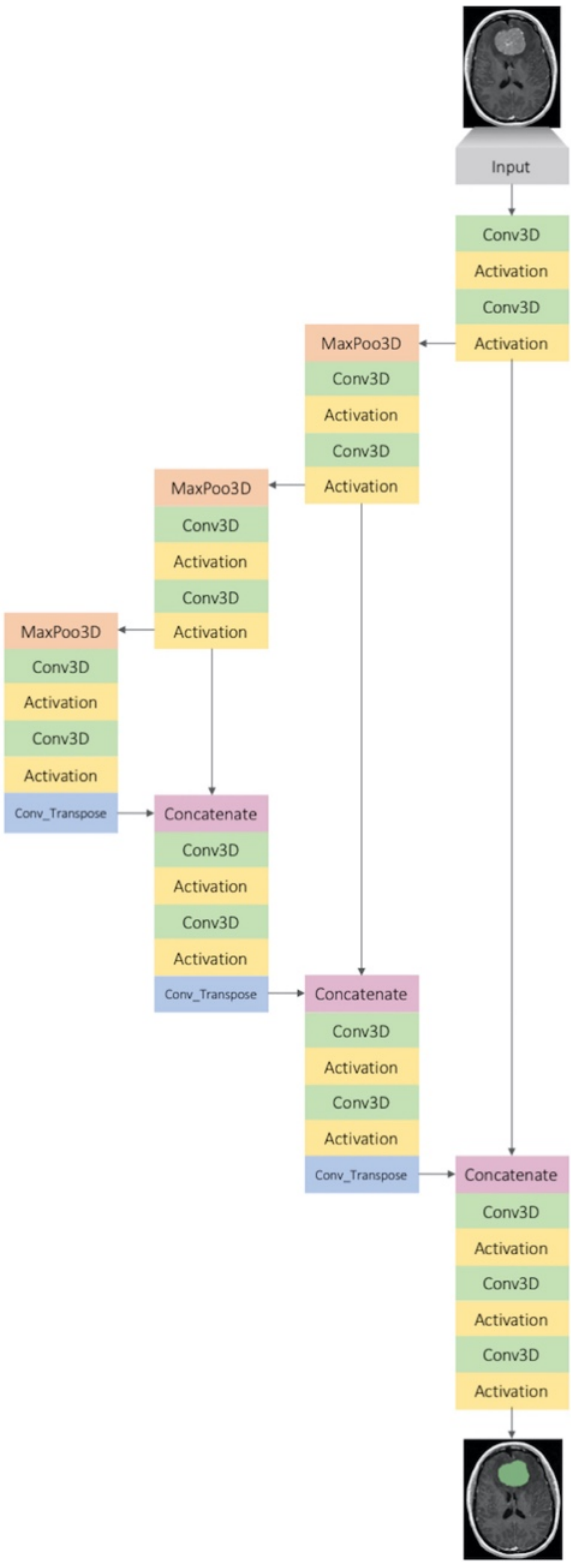

Fig. 1. Deep Neural Network Architecture. The architecture of our deep neural network consisted of a 3D-UNet structured with an encoder and a decoder arm. The network accepts 3D MRI data as input and outputs a 3D segmentation map.

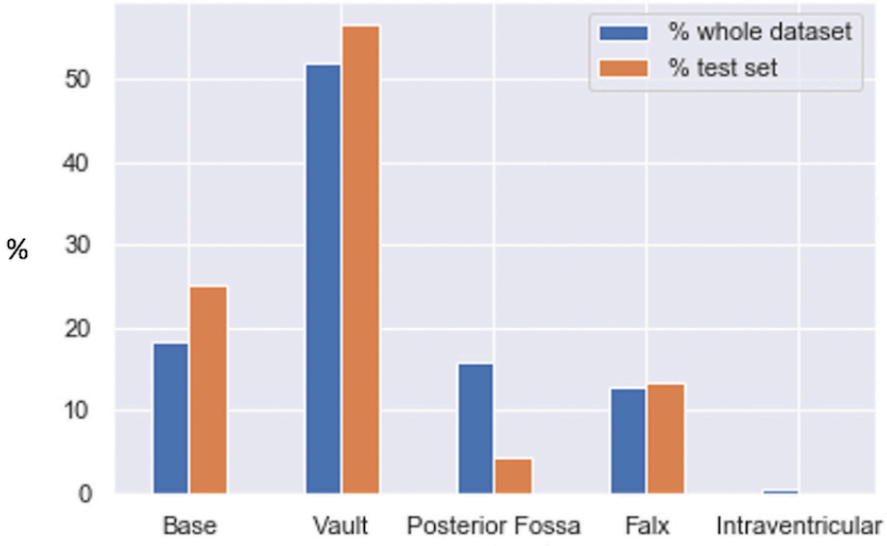

Fig. 2. Tumor location distribution. Bar plot showing tumor location distribution (% of the total) in the whole meningioma dataset and test set. All the main anatomical locations are adequately represented both in the general dataset and in the test set.

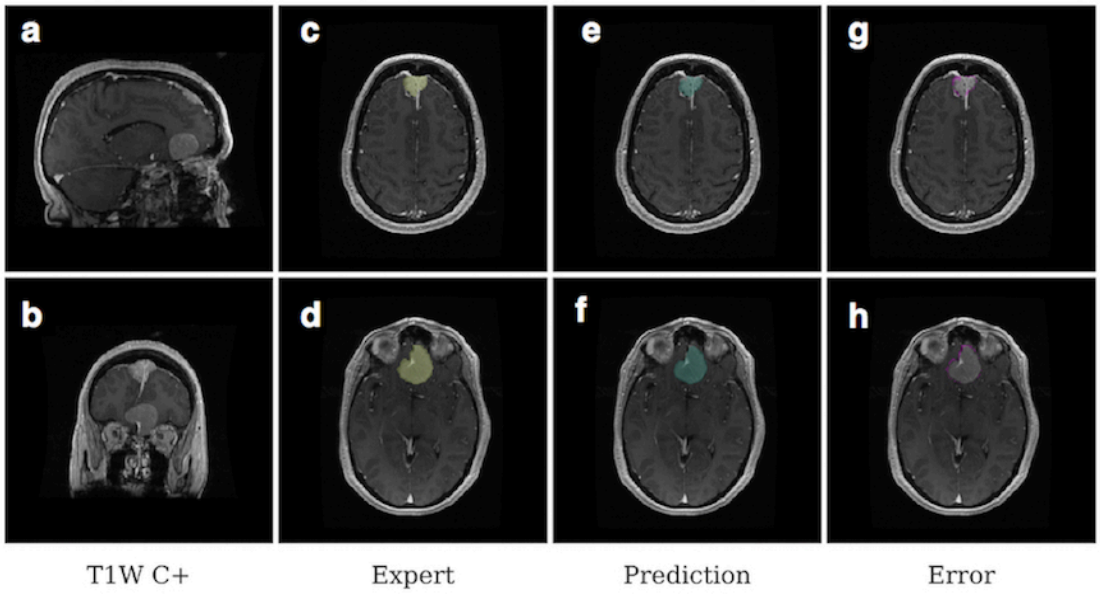

Fig. 3. Example of meningioma segmentation algorithm output. Sagittal and coronal views (**a** and **b**) of a brain MRI scan containing two distinct meningiomas, one located in the convexity at the midline, the other located on the anterior skull base. Display of expert label vs computer-generated segmentation respectively of the meningioma of the convexity (**c** and **e**) and the meningioma of the skull base (**d** and **f**). Display of the mismatch between the expert label and the computer-generated segmentation on the meningioma of the convexity (**g**) and the meningioma of the skull base (**h**).

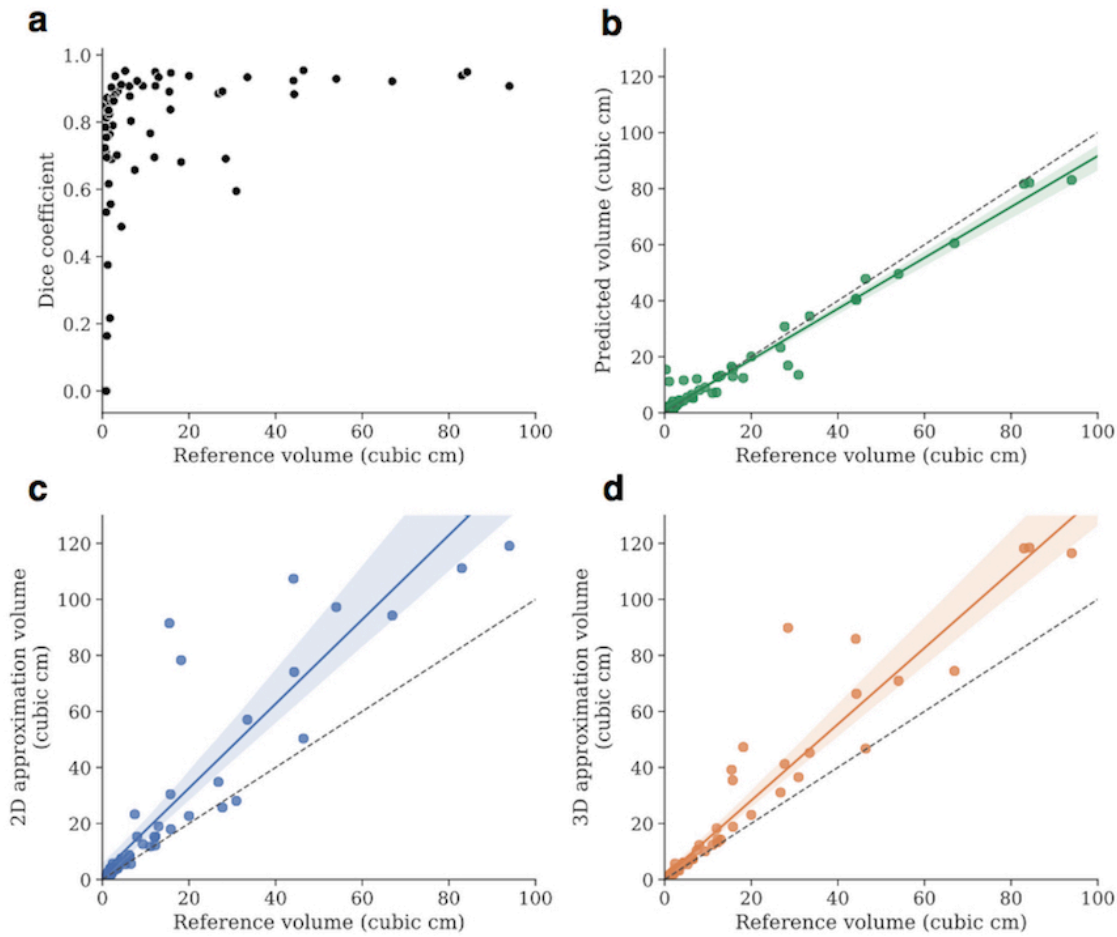

Fig. 4. Algorithm's tumor segmentation and volume estimation accuracy. (**a**) Scatter plot showing the algorithm's segmentation performance expressed as Dice scores on the test set as a function of the tumor volume. The Dice score correlated with the size of the tumor, quickly reaching a mean of 0.87 and median of 0.89 for tumors > 2cc. (**b**, **c**, **d**) Volume estimations by the algorithm, and the 2-D and 3-D traditional estimation techniques, respectively. The algorithm's predicted volumes constitute almost perfect approximations of the real tumor volumes with a correlation of 0.98, $p<0.001$ (**b**), whereas the 2D and 3D techniques present lower values of respectively 0.88 (**c**) and 0.96 (d) $p<0.001$, with evidence of overestimation.

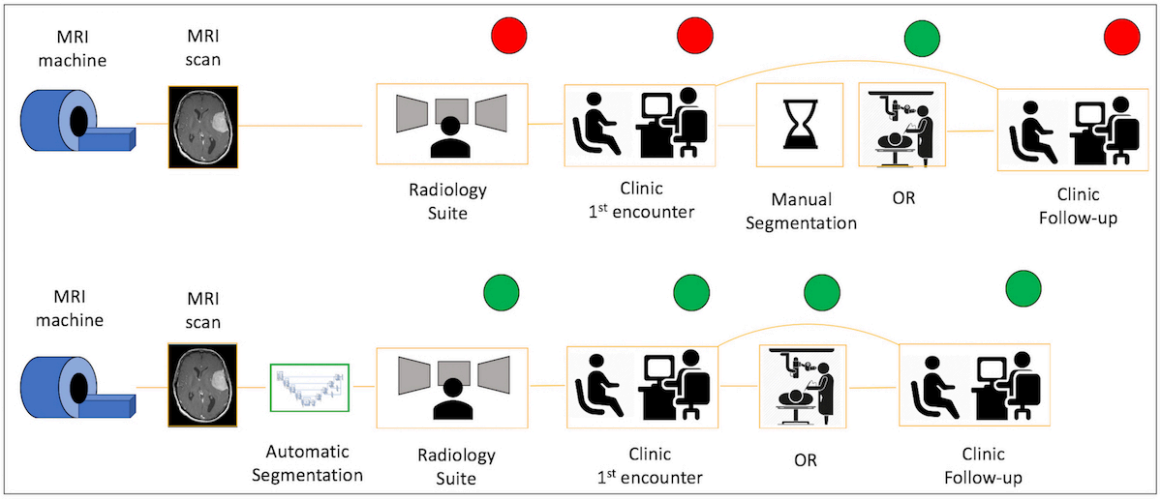

Fig. 5. Current clinical workflow and clinical workflow implementing automatic tumor segmentation. *Red light*: Volumetric tumor mask is not available; *Green light*: Volumetric tumor mask is available. Above: Volumetric tumor segmentation is currently implemented manually and only for surgical planning purpose, while tumor shape and volume information are not routinely available. Below: The implementation of an automated, expert-level tumor segmentation algorithm provides a tumor mask with accurate shape and volume information at each point of patient care, from the day of the first MRI scan in the radiology suite to the operative room, to each radiologic and clinical follow-up patient encounter.