NeuroImage 306 (2025) 121002

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Development and routine implementation of deep learning algorithm for automatic brain metastases segmentation on MRI for RANO-BM criteria follow-up

Loïse Dessoude [a,1], Raphaëlle Lemaire [b,1], Romain Andres [b,1], Thomas Leleu [a], Alexandre G. Leclercq [b], Alexis Desmonts [a], Typhaine Corroller [b], Amirath Fara Orou-Guidou [b], Luca Laduree [b], Loic Le Henaff [c], Joëlle Lacroix [c], Alexis Lechervy [d], Dinu Stefan [a], Aurélien Corroyer-Dulmont [b,e,*]

[a] Radiotherapy Department, Centre François Baclesse, Caen 14000, France
[b] Medical Physics Department, Centre François Baclesse, Caen 14000, France
[c] Radiology Department, Centre François Baclesse, Caen 14000, France
[d] ENSICAEN, CNRS, GREYC UMR6072, Normandie Université, Université Caen Normandie, Caen F-14000, France
[e] CNRS, ISTCT UMR6030, GIP CYCERON, Normandie Université, Université de Caen Normandie, Caen 14000, France

ARTICLE INFO

Keywords:
Deep learning
Radiology
Brain metastases
RANO-BM
Clinical routine

ABSTRACT

Rationale and objectives: The RANO-BM criteria, which employ a one-dimensional measurement of the largest diameter, are imperfect due to the fact that the lesion volume is neither isotropic nor homogeneous. Furthermore, this approach is inherently time-consuming. Consequently, in clinical practice, monitoring patients in clinical trials in compliance with the RANO-BM criteria is rarely achieved. The objective of this study was to develop and validate an AI solution capable of delineating brain metastases (BM) on MRI to easily obtain, using an in-house solution, RANO-BM criteria as well as BM volume in a routine clinical setting.
Materials (patients) and methods: A total of 27,456 post-Gadolinium-T1 MRI from 132 patients with BM were employed in this study. A deep learning (DL) model was constructed using the PyTorch and PyTorch Lightning frameworks, and the UNETR transfer learning method was employed to segment BM from MRI.
Results: A visual analysis of the AI model results demonstrates confident delineation of the BM lesions. The model shows 100 % accuracy in predicting RANO-BM criteria in comparison to that of an expert medical doctor. There was a high degree of overlap between the AI and the doctor's segmentation, with a mean DICE score of 0.77. The diameter and volume of the BM lesions were found to be concordant between the AI and the reference segmentation. The user interface developed in this study can readily provide RANO-BM criteria following AI BM segmentation.
Conclusion: The in-house deep learning solution is accessible to everyone without expertise in AI and offers effective BM segmentation and substantial time savings.

1. Introduction

Brain metastases are a common occurrence in patients with cancer, affecting between 20 and 40 % of individuals. They represent the most prevalent form of brain malignancy (Achrol et al., 2019). In some cases, these metastases demonstrate responsiveness to local treatments, including stereotactic radiotherapy, which has been shown to have an excellent local control rate (exceeding 80 % local control after two years) (Ene et al., 2024).

The advent of novel systemic therapies has led to a notable improvement in the prognosis of patients with brain metastases. Following the administration of an initial localised treatment, patients may be monitored for several years, with the potential for further localised treatment to be beneficial. The preparation of stereotactic

* Corresponding author at: Medical Physics Department, CLCC François Baclesse, Caen 14000, France.
E-mail address: a.corroyer-dulmont@baclesse.unicancer.fr (A. Corroyer-Dulmont).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.neuroimage.2025.121002
Received 5 November 2024; Received in revised form 13 December 2024; Accepted 6 January 2025
Available online 10 January 2025

brain radiotherapy treatments and the subsequent monitoring of patients following treatment represent a significant and growing aspect of the work of radiotherapists. The establishment of consensus criteria for patient follow-up represents a significant challenge, particularly in terms of standardising practice across different centres and facilitating comparisons between clinical trials. The RANO BM (Response Assessment in Neuro-Oncology Brain Metastases) criteria, introduced in 2015 by the international and multidisciplinary RANO BM working group, represent the current gold standard for the assessment of brain metastases post-treatment response (Lin et al., 2015).

In addition to the clinical criteria and a global vision of the disease at the cerebral level, the RANO BM criteria provide the clinician with the capacity to undertake one-dimensional measurement of the largest diameter of so-called target brain metastases. The aforementioned target brain metastases are measured on a post-Gadolinium T1 sequence. The efficacy of treatment is determined based on the observed change in lesion size. This ranges from a 20 % increase, indicative of disease progression, to complete response, characterised by lesion disappearance. The measurement of the largest diameter of all target lesions is a time-consuming process, particularly in the context of multiple MRI follow-ups for different stereotaxic radiotherapy treatments. Consequently, monitoring patients in clinical trials can be a significant burden for radiologists. In clinical practice, compliance with the RANO BM criteria is often challenging to achieve.

Furthermore, brain metastases are a complex entity with significant heterogeneity, including areas of necrosis, areas of progression, pseudoprogression, and other characteristics. The evolving environment of brain metastases is heterogeneous, with various interfaces (meninges, bones, ventricles, etc.) present. Consequently, the growth of a metastasis is not necessarily isotropic (Hessen et al., 2017; Ohtakara et al., 2023).

A recent retrospective study indicates that one-dimensional measurement is imperfect and may not be as effective in detecting progressions as three-dimensional measurement, particularly volumetric measurement (Ocaña-Tienda et al., 2024).

In light of these considerations, the potential benefits of integrating an automatic contouring tool into the clinical workflow, both before and after treatment, are twofold.

Firstly, it could facilitate the preparation of stereotactic radiotherapy treatments by the radiotherapist, assisting in the identification of metastases and reducing delineation time.

Secondly, it could enable the radiologist to monitor treated patients rapidly, accessing numerous metrics, some of which have already been validated by RANO BM, and others which show promise and may offer more efficient solutions.

Artificial intelligence (AI) algorithms for the automatic contouring of brain metastases are currently being developed (Cho et al., 2021a; Xue et al., 2020; Chartrand et al., 2022; Dikici et al., 2020; Li et al., 2023). Notably, UNETR type models have achieved the best results for brain metastases detection and segmentation (Pang et al., 2024; Shaker et al., 2022). However, there are still very few trials evaluating the use of these models for patient monitoring (Cho et al., 2021b; Kickingereder et al., 2019; Hsu et al., 2023) and concrete solutions which can be used for clinical routine are still awaited.

The objective of this study was to develop an algorithm that can accurately detect and segment brain metastases and be readily integrated into the clinical workflows of radiologists and radiotherapists.

2. Materials and methods

2.1. Patients

The present retrospective study has been approved by the local institutional review board. A total of 27,456 2D post-Gadolinium (Gd) T1 MRI scans from 132 patients with a total of 386 brain metastases who were referred to our oncology centre between January 2019 and March 2023 were included in the study. This study was conducted in accordance with the guidelines set forth by MR-004, a national French institution that defines health research conduct and the Declaration of Helsinki. All patients provided informed consent for the use of their data. The characteristics of the study population are outlined in Table 1.

2.2. Magnetic resonance imaging (MRI) acquisition

MRI was performed on an AREA SIEMENS 1.5 Tesla magnet using a brain dedicated 16 channels coil with the patient in a supine position. Prior to the examination, patients were injected with 0.2 mL/kg of DOTAREM (500μmol/ml). After a shimming process and scout imaging scan, tumor gadolinium enhancement was detected with a post-Gd T1 brain sequence with the following parameters: TR/TEeff=2070/3.15 msec; Angle=15°; NEX=1; 208 contiguous slices; 3D resolution=0.5 × 0.5 × 1 mm; acquisition matrix = 512 × 512 pixels and acquisition time=4min48). A total number of 27,456 2D MR images were acquired from the 132 patients.

2.3. Deep learning algorithms

2.3.1. Deep learning (DL)

The deep learning model consisted of a fine-tuned UNETR architecture (Hatamizadeh et al., 2021). This model incorporates the strengths of both the UNet and Vision Transformer models, addressing the challenge of segmenting multiple regions of interest within an image. The UNETR architecture output was modified removing the last 14 output in the last layer by two outputs for the purpose of distinguishing between lesion and healthy tissue.

UNETR was pre-trained in the segmentation task for brain tumor on a set of 484 multi-modal multi-site MRI data with three class (1: tumor, 2: hemorrhagic part of the tumor, 3: eudeme). Architecture was described in "UNETR: Transformers for 3D Medical Image Segmentation by Ali Hatamizadeh et al." Last layer was 1 × 1 × 1 convolutional layer that has been modified for binary segmentation called UnetOutBlock. It is the only layer not freeze, other weights are saves from the pretrain model of structure classification of UNETR. The MLP was the one of MONAI in the Vision Transformer. We applied the following layers: Linear layer 1, Dropout Layer 1, Linear layer 2 and Dropout layer 2. The convolution head was composed of a Convolution 3D, a Prelu, a Dropout and a Layer norm. More information can be found here: https://docs.monai.io/en/1.0.1/_modules/monai/networks/blocks/mlp.html.

UNETR initial model before fine tuning was obtained using MONAI (Cardoso et al., 2022). Fine-tuning technic used the methodology of Yosinski and collaborators (Yosinski et al., 2014). During the training process, only the weights of the final layer were removed and trained, while the weights of the preceding layers were maintained at their original values. A deep learning model was developed from 27,456 unique post-Gd T1 brain images obtained from 132 patient acquisitions with a total of 386 BM. These images were split into three datasets: a training set comprising 19,219 images (70 % of the total), a test set comprising 2746 images (10 % of the total), and a validation set

Table 1
Description of the patient cohort.

Included patients (N)	132	Number
Sex	50 %	Female %
Age (Y)	63.4	Mean
Total number of BM	386	Number
Number of BM *per patient*	2.93 ± 2.32 (min=1; max=13)	
Lesion origin		
- *From Lung cancer*	83 (61 %)	Number (%)
- *From Melanoma cancer*	28 (20 %)	
- *From Breast cancer*	8 (6 %)	
- *From Kidney cancer*	7 (5 %)	
- *From Colorectal cancer*	3 (4 %)	
- *From Head and Neck cancer*	2 (3 %)	
- *From Digestive cancer*	1 (1 %)	

comprising 5491 images (20 % of the total). To avoid bias, we carefully check that each patient cannot be part of the training and validation and test dataset, all slices of each patient was in the same dataset. The input data comprised brain MRI images and the delineation of the tumour lesion, designated as GTV (Gross Tumour Volume). The GTV region of interest was initially transformed into a mask image. Prior to training, specific MRI images underwent normalisation, with bias field correction employed (Masoudi et al., 2021). Data augmentation was conducted through the application of flips along the sagittal axis and 180° rotations. To reduce the computational time required, intensity normalisation was performed between 0 and 255, and the image background was removed. The deep learning model was developed using PyTorch Lightning. The loss function used in this study was the binary cross entropy, using the PyTorch function "binary_cross_entropy_with_logits". The Pydicom and dicompylercore libraries were used to manage the MRI and RTSTRUCT DICOM files (Mason, 2011). Dice index was used to evaluate the performance of the model during the training process. The AI model was trained on two NVIDA A6000 GPU 48Go. All the code used to develop and train the model is available at: https://github.com/AurelienCD/BrainMetaSegmentatorUI-Back (accessed on 01 November 2024).

2.3.2. Image analysis and processing

Quantitative analysis: In accordance with the established workflow within the radiotherapy department, another radiation oncologist delineated a three-dimensional volume of interest (VOI) encompassing 31 lesions utilising the Raystation™ solution (V11.B) for ten patients (not for the purpose of AI model training). Subsequently, an expert radiation oncologist evaluated the RANO-BM criteria on the reference and AI brain metastases segmentation. The concordance of the RANO-BM criteria between AI and reference was then evaluated. To evaluate the ability of the AI model to detect BM, F1 score were evaluated. Several quantitative metrics, which are commonly used in the literature to evaluate the spatial overlap, were employed (Taha and Hanbury, 2015), were used to compare the VOI delimited by the radiation oncologist and the one created by the AI models:

- **Dice Similarity Coefficient (DSC)**: Measures the overlap between two volumes, providing a statistical validation of segmentation precision;
- **Mean Surface Distance (MSD)**: Calculates the average Euclidean distance between the surfaces of two volumes, offering insights into the contour accuracy;
- **Volume Overlap Error (VOE)**: Represents the proportion of the total volume that is over-segmented or under-segmented relative to the reference, complementing the Dice coefficient by providing error rates;
- **Hausdorff Distance**: Evaluates the maximum distance of the dataset boundary points between the predicted and reference segmentations, highlighting the worst-case scenario of boundary prediction;
- **Jaccard Index**: Quantifies the similarity and diversity between sample sets, indicating the proportionate size of the intersection divided by the union of the sample sets;
- **Variation of Information (VI)**: Measures the amount of information lost and gained in the segmentation process, reflecting the complexity and precision of the information captured by the segmentation;
- **Cosine Similarity**: Assesses the cosine angle between the multidimensional representations of the segmented volumes, useful for understanding the orientation and agreement in the segmented shapes.

To compare reference and AI brain metastases ROI, first order intensity evaluation was performed using mean, standard deviation, min, max. Subsequently, the specificity and sensitivity were evaluated in order to ensure the accuracy of the RANO-BM AI prediction in comparison to the radiation oncologist segmentation.

2.4. Statistical analyses

All data are expressed as mean ± SD. The correlation between the first-order intensity values derived from the reference and those obtained from the AI-based brain metastases segmentation was analysed with the concordance correlation coefficient (CCC) (Lin, 1989). A CCC value of 1 indicates a perfect positive or negative correlation, whereas a value of 0 indicates no correlation. Features with a minimum CCC of 0.85 were deemed to be statistically reproducible and concordant, and the values were considered to be stable (Peerlings et al., 2019). All statistical analyses were performed using Python (Anaconda Software Distribution, 2020) and SciPy library. Data visualization used Seaborn library (Waskom, 2021; Hunter, 2007). All Python code used in the analysis is available at https://github.com/AurelienCD/MetIA and "Quantitative analysis.ipynb" (accessed on 01 November 2024).

3. Results

3.1. Deep learning brain metastases segmentation model

The optimisation process resulted in an AI model constructed with binary accuracy validation metrics over 1079 epochs. The training process, which spanned three days, yielded 93 million parameters. As illustrated in Fig. 1, the training and validation loss functions, which represent the model's error rate throughout the training phase, indicate that epoch 1079 was the most optimal.

3.2. Visual analysis

As presented in Fig. 2, the AI model is able to delineate both large BM lesions (black arrows) and small lesions of 4 mm diameter (white arrows). The delineation closely follows the hyperintensity seen on T1 gadolinium enhancement.

3.3. RANO-BM concordance

As evaluated by the expert physician in the reference segmentation, the RANO-BM criteria evaluation were as follow: two complete responses, two partial responses, three stable diseases and two partial diseases. The RANO-BM obtained from AI segmented lesion were 100 % agreement with the above RANO-BM criteria evaluation.

3.4. Quantitative analysis

The spatial overlap of the reference and AI brain metastases segmentation was first analyzed using several metrics. As presented in Table 2, the overlap between the AI and the physician's segmentation volumes with DICE coefficient was 0.77 and the Euclidean distance between the two volumes was 4.13, representing a reliable overlap.

The diameters and volumes obtained from the AI segmentation were then compared with the reference. As shown in Fig. 3 for each brain metastasis, very few differences were observed between the diameters (Fig. 3A) and volumes (Fig. 3B) obtained from the AI segmentation compared to the reference segmentation. Volume and diameter differences from AI and radiotherapist segmentation were 0.15 ± 0.18 mm^3 and 1.38 ± 1.19 mm, respectively. The AI model have shown good BM detectability with F1 score of 95.5 %.

To go deeper, the stability of volume, diameter and first order signal intensity values between AI and reference segmentations were evaluated using the Concordance Correlation Coefficient (CCC). As presented in Fig. 4, the CCC values were: 0.93, 0.97, 0.76, 0.98, 0.99 and 0.93 for diameter, volume, minimal intensity, mean intensity, maximum intensity and standard deviation intensity, respectively. Only, the minimum intensity variable was below the 0.8 threshold, showing a

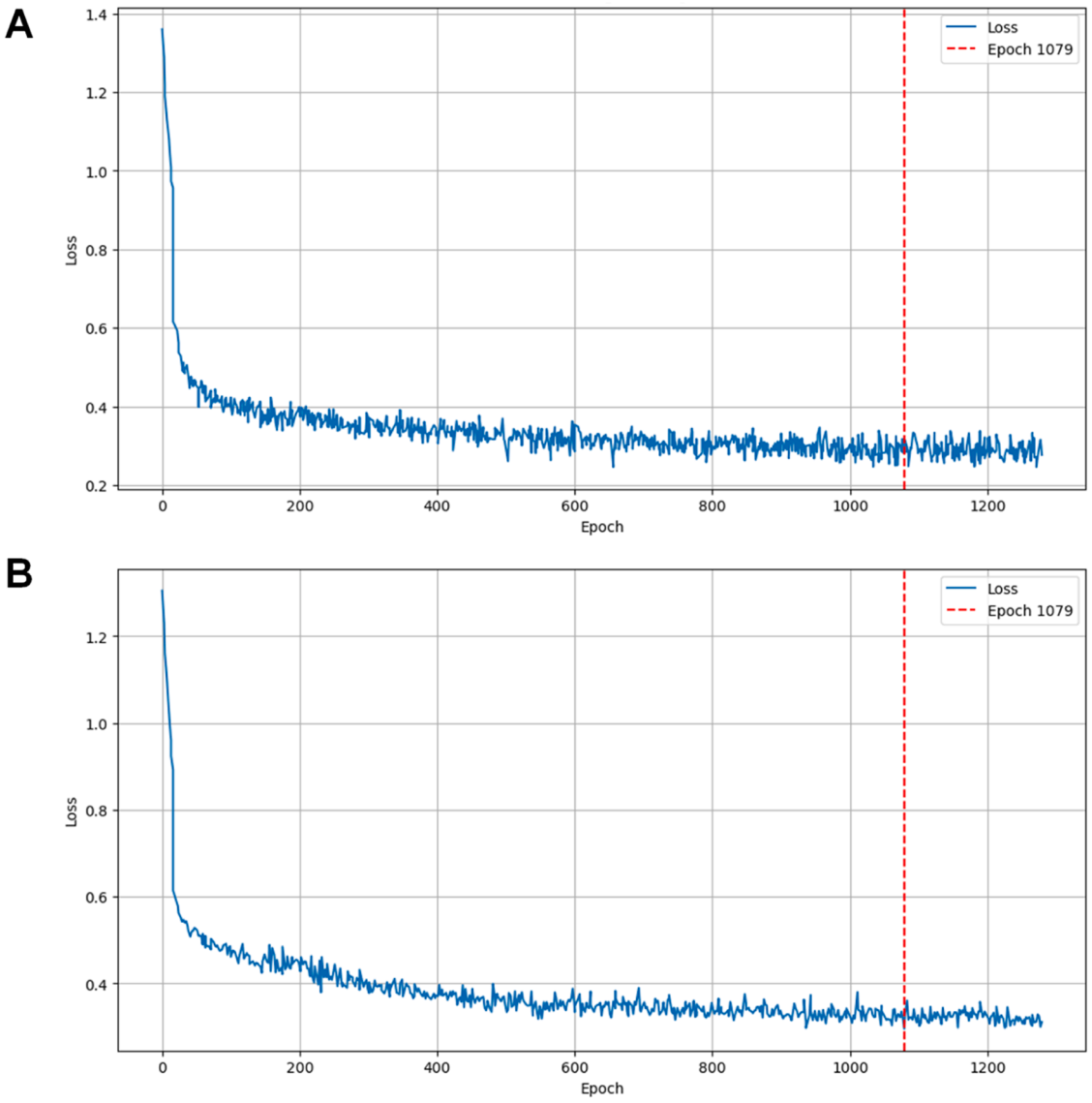


Fig. 1. Deep learning model performance during the training process through the epochs. (A) Training loss and (B) Validation loss functions during the training process.

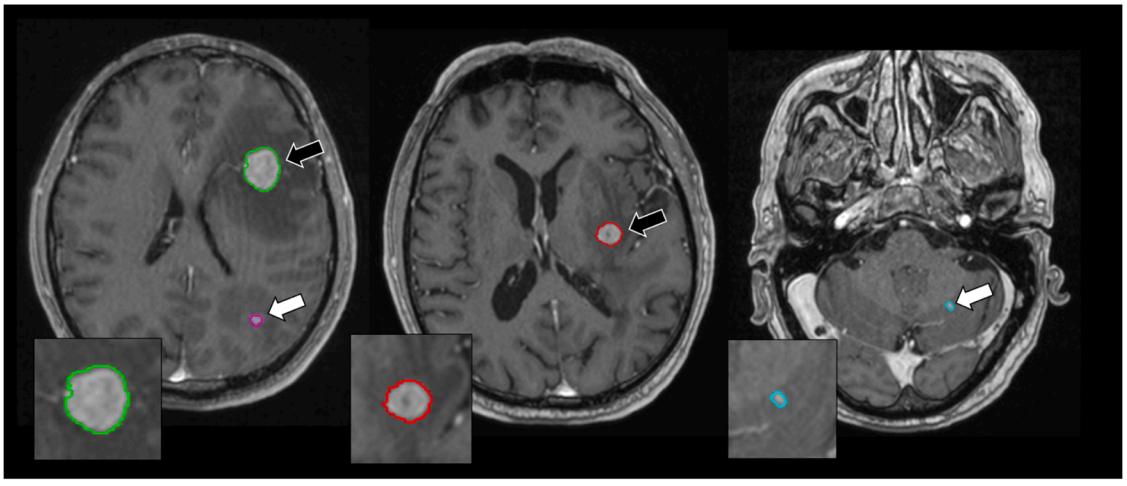


Fig. 2. Three representative MRI with brain metastases segmented by the deep learning model. Large (black arrows) as well as small lesions (white arrows) are detected by the algorithm.

Table 2
Quantitative analysis of reference and AI predicted region of interest similarity.

DICE coefficient (SD)	Mean surface distance (SD)	Volume overlap error (SD)	Haussdorf distance (SD)	Jaccard index (SD)	Variation of information (SD)	Cosine Similarity (SD)
0.77 (0.15)	4.13 (7.32)	0.43 (0.24)	32.67 (60.05)	0.63 (0.19)	0.001 (0.0004)	0.77 (0.13)

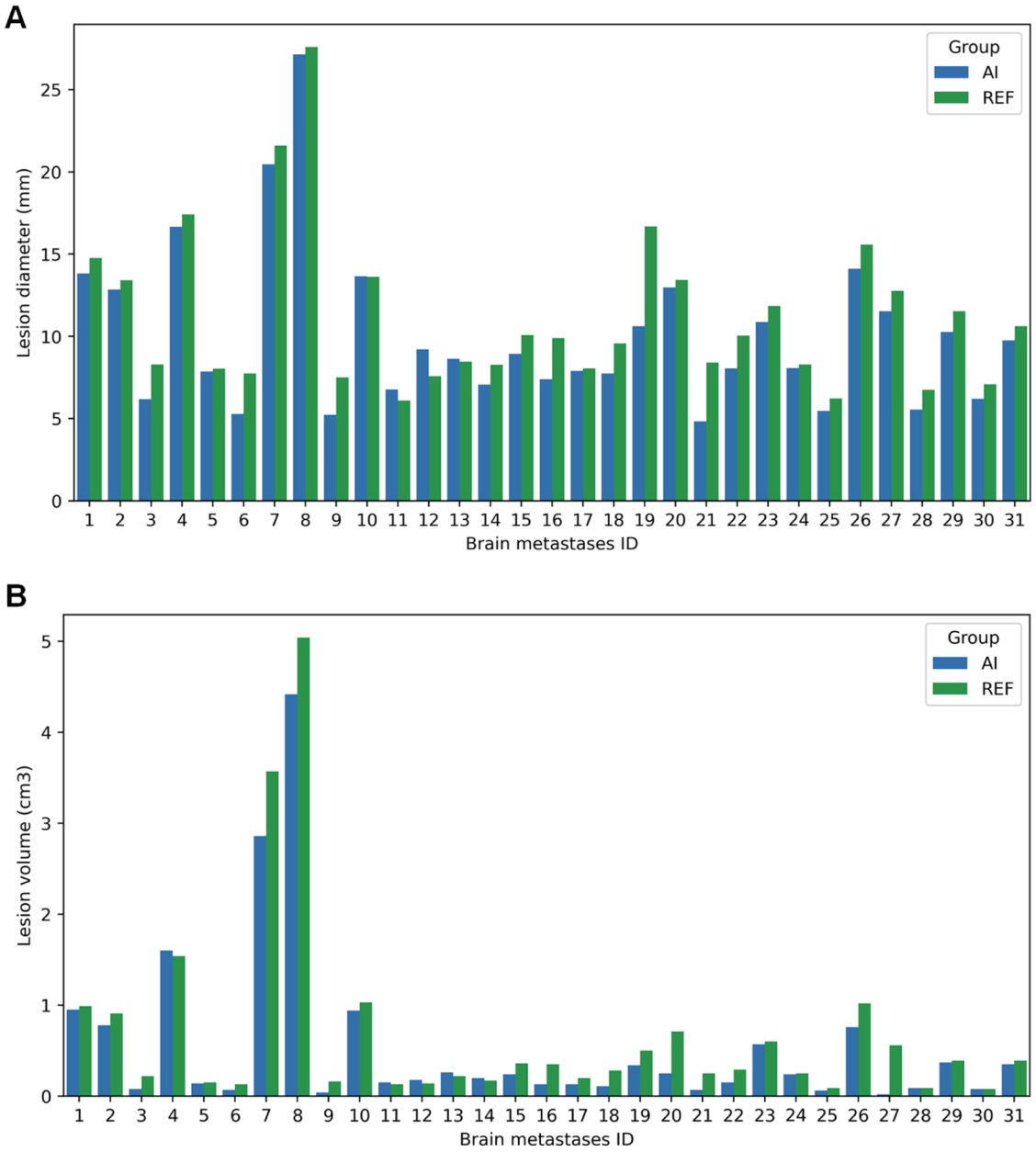


Fig. 3. Impact of deep learning segmentation on brain metastases diameters (A) and volumes (B) evaluation.

discordance between AI and reference segmentation minimum intensity.

3.5. Application of the solution in clinical practice

The aim of this study was to develop an approach that works well but can also be easily implemented in clinical practice. To achieve this, a user interface was developed using ORTHANC and Open Health Imaging Foundation (OHIF) (Ziegler et al., 2020) solutions that can be used in a clinical setting. This user interface interacts with a back-end API to retrieve medical data in DICOM format, start the deep learning model, and visualize the results. Further details on the back-end and front-end parts of the user interface as well as a tutorial can be found at https://github.com/AurelienCD/BrainMetaSegmentatorUI-Back (accessed on 1 November 2024) and https://github.com/AurelienCD/BrainMetaSegmentatorUI-Front (accessed on 1 November 2024). The interface requires as input, a brain MRI and after an average of 30 s in average of processing (with standard RTX 4080 GPU), provide AI brain metastases segmentation in RTSTRUCT format which can be easily uploaded and used in conventional Treatment Planning Software (Raystation™ solution (V11.B) for this study). The solution can also be used to view the AI

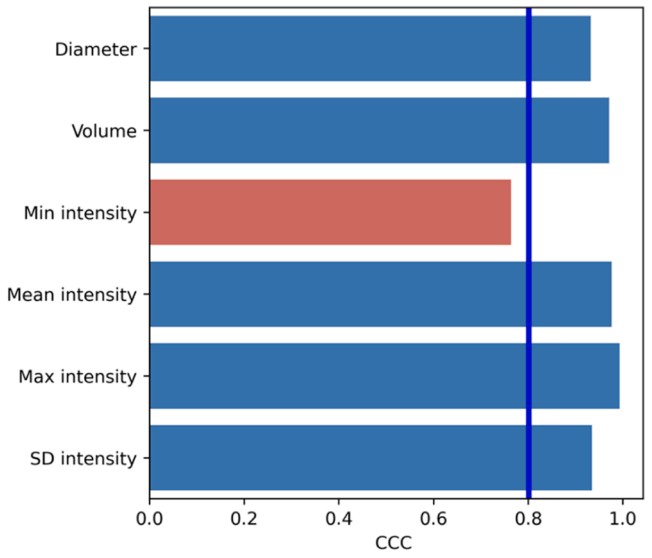


Fig. 4. Stability of the first order statistics between AI and reference segmented ROI.

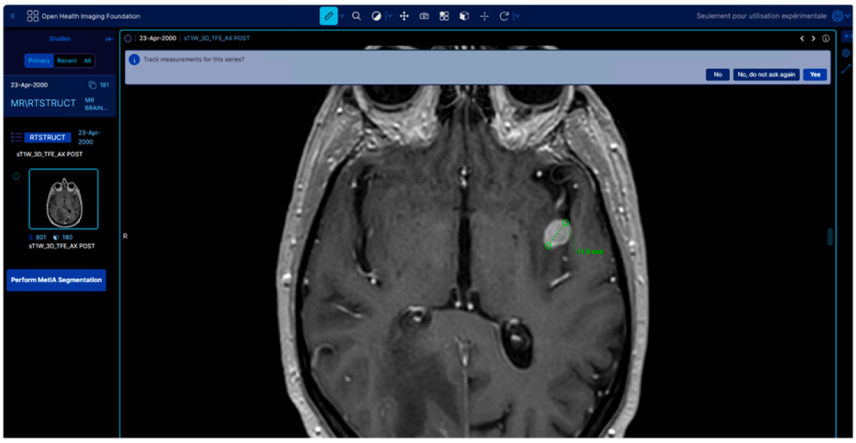
segmentation result and track the diameter and volume of brain lesions for RANO-BM purposes (Fig. 5A and B). The deep learning model and user interface codes are freely available upon reasonable request. However, please note that the performance of the model has only been optimized for our data and needs to be fully validated before external use.

4. Discussion

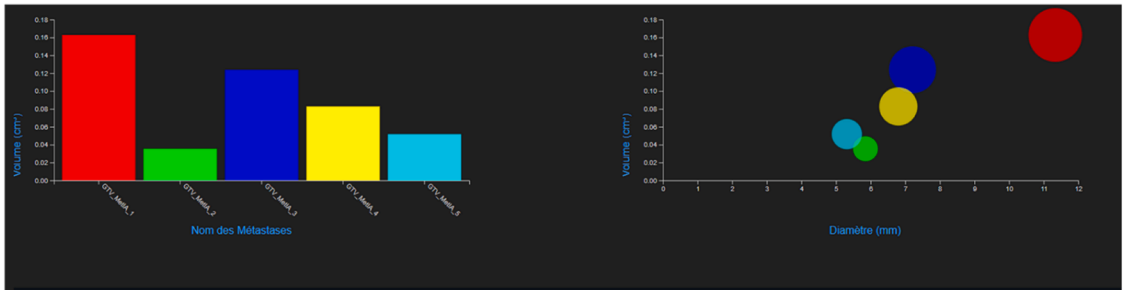
Brain metastases occur in 20–40 % of patients with cancer and represent the most common manifestation of brain malignancies (Achrol et al., 2019). Due to this high number of lesions and to the human resource difficulties in the medical field, patient follow-up during clinical routine or for clinical trials is often difficult to underwent. For these reasons, compliance with the RANO BM criteria is rarely achieved in clinical practice. A highly robust and easy-to-implemented solution that could automatically and quickly extract BM lesion's diameter and volume as well as RANO BM criteria could be an interesting insight for patient therapeutic management.

In this study we have developed and evaluated a deep learning model using the transfer learning method of UNETR to automatically extract BM lesion segmentation. The model was trained using >27,000 unique post-Gd T1 brain images acquired from 132 patient's acquisitions. The number of patients in our study was similar to several previous studies

A



B



Nom	Diamètre (mm)	Volume (cm³)	Slice de début	Slice de fin
GTV_MetIA_1	11.33578410562243	0.1627500016328631	75	85
GTV_MetIA_2	5.852349984754453	0.03537500035519092	86	89
GTV_MetIA_3	7.211102587087033	0.12375000124166917	73	88
GTV_MetIA_4	6.800735288542701	0.0827500008305033	61	70
GTV_MetIA_5	5.315072932995939	0.05175000051901722	140	143

Fig. 5. Example of RANO-BM with diameter and volume follow-up using deep learning segmentation results with the integration of the model within OHIF solution. (A) Example of diameter measurement and (B) patient statistics follow-up.

(Dikici et al., 2020; Luo et al., 2024; Huang et al., 2022), but lower compared to few others (Luo et al., 2024). Furthermore, the quality of the data of our training and validation set were all reviewed by a radiotherapist to deleted incomplete data or complex cases that could lead to confusion due to patient movements, presence of artifacts…

Our model showed close BM segmentation compared to experienced physician segmentation with a mean DICE score of 0.77. In the literature, DICE scores remain below 0.82 (Cho et al., 2021a, 2021b; Xue et al., 2020; Hsu et al., 2023) and our results are consistent with the DeepMedic approaches (Huang et al., 2022). However, in a recent study conducted by Luo and co-workers, the DICE score was of 0.91 possibly due to the size of the cohorts which were 312 and 156 patients for training and validation respectively (Luo et al., 2024).

Lesion diameter and volume are concordant between AI and reference segmentation. More specifically, Supplementary Fig. 1 shows the correlation between AI and reference lesion diameter and volume. The pearson correlation shows a significant correlation with R^2 values of 0.92 ($p < 0.001$) and 0.98 ($p < 0.001$) for lesion diameter and volume respectively.

Here we can confirm that, for lesions larger than 1 cm in diameter, AI and reference values were highly correlated. However, below this threshold of 1 cm, which is exactly threshold imposed by the RECIST criteria, more important heterogeneity was observed. This last point highlights the interest of an AI solution for the assessment of very small lesions below 1 cm, which are currently not assessed by radiologists.

It is interesting to note that the mean volume and diameter as well as the minimum intensity are slightly smaller in the AI segmentation (not significantly discordant for BM diameter and volume but significantly discordant for minimum intensity). This probably highlights the fact that the model is trained to detect the hyperintensity signal revealed by the gadolinium injection in order to delimitate the BM lesion. The model is potentially stricter on the tumor boundary and does not include the area without T1 enhancement in the lesion area which may be done by an experienced physician as they know that tumor cells invade the surrounding healthy tissue close to the area of T1 enhancement. As shown in Supplementary Fig. 2, AI and radiotherapist segmentation can be completely concordant (A), but in some cases the AI segmentation seems to follow the tumor boundary more precisely than the radiotherapist one (B), and in some other cases the AI segmentation was smaller than the reference segmentation. It appeared that applying smoothing could be more realistic if the invasion process of brain metastases is known.

BM are not always well delineated with homogeneous high signal intensity. Supplementary Fig. 3 shows the example of AI segmentation of BM with central necrosis (Supplementary figure 3A), diffuse BM (Supplementary figure 3B) and BM close to an area of high signal intensity without being a tumor (Supplementary figure 3C).

Our patient dataset is representative of the patient population with BM, as lung cancer is the most common primary source of BM in the training dataset. This could introduce a bias and not allow good delineation in other primary histologies (from breast, renal or melanoma primary cancers). In our study, no difference in performance was observed with respect to the different primary histologies, as shown in the Supplementary Fig. 4.

Despite these slight non-significant and significant differences, the RANO-BM criteria obtained from the AI segmentation are 100% concordant with those obtained from the physician segmentation. Patient monitoring with RANO-BM follow-up, which is rarely addressed in the literature, was an important aspect of our study. We found only one study that investigated the concordance of RANO-BM criteria obtained by an AI model with those defined by radiologists (Cho et al., 2021b). The kappa coefficients calculated in this study were equal to 0.52 based on largest diameters and 0.68 based on volumes. Obtaining a radiological response according to the RANO BM criteria is a challenge that resonates with the daily concerns of radiologists and radiotherapists.

In this study, we have developed an easy-to-use interface to exploit AI BM segmentation. To date, no industrial solutions have been validated and proposed for the clinical routine. Raystation and TheraPanacea are examples of two treatment planning software that are developed highly innovative algorithms to optimize therapeutic management in the radiotherapy department. Both are able, in a clinical routine setting, to delineate organs at risk in order to accelerate radiotherapy planning (Bondiau et al., 2022; Stathakis et al., 2022; Mekki et al., 2024). However, to date, there is no fully validated and routinely proposed AI solution for the delineation of tumors as BM.

From the perspective of our study, it would be interesting to fine-tune very recent large models developed for medical purposes such as UNETR++ and nnFormer to improve performance (Shaker et al., 2022; Zhou et al., 2021), which have not yet been used for BM segmentation.

The reproducibility and robustness of the AI models in different clinical settings and at different centers is a key factor for their implementation into clinical practice. The use of a federated learning approach can lead to the development of a global model based on data from different centers (Pati et al., 2022; Ahamed et al., 2023). The next step for this project would be to use federated learning with volunteer centers to improve our model and make it more relevant to other centers. Finally, supporting clinicians in monitoring their patients according to the RANO-BM criteria will facilitate inter-operator reproducibility and the standardisation of practice. This is in line with the objectives proposed by the international RANO-BM group. Indeed, the heterogeneity of follow-up is a major challenge in clinical trials on patients with brain metastases. This provides an opportunity to explore alternative approaches to assessing patient response and, subsequently, differentiating radionecrosis from progression.

5. Conclusion

Together with experienced radiotherapists and radiologists, we have developed and validated a fully automated deep learning solution capable of accurately delineating BM using RANO-BM criteria. Our in-house user interface solution, easily accessible to non-experts in AI, provides sufficient BM segmentation and significant time savings.

Data availability statement

The data presented in this study can be sent upon reasonable request. Python code used for this study are openly available at https://github.com/AurelienCD/MetIA.

CRediT authorship contribution statement

Loïse Dessoude: Validation, Investigation, Formal analysis, Data curation, Methodology, Conceptualization, Writing – review & editing, Writing – original draft. **Raphaëlle Lemaire:** Methodology, Investigation, Conceptualization, Software, Writing – review & editing, Validation. **Romain Andres:** Methodology, Investigation, Data curation, Conceptualization, Software, Writing – review & editing, Validation, Formal analysis. **Thomas Leleu:** Investigation, Methodology, Conceptualization, Validation, Formal analysis, Writing – review & editing, Data curation. **Alexandre G. Leclercq:** Validation, Writing – review & editing, Investigation. **Alexis Desmonts:** Data curation, Formal analysis, Validation, Writing – review & editing, Investigation. **Typhaine Corroller:** Investigation, Software, Writing – review & editing, Validation. **Amirath Fara Orou-Guidou:** Methodology, Software, Validation, Writing – review & editing, Investigation. **Luca Laduree:** Methodology, Software, Validation, Writing – review & editing, Investigation. **Loic Le Henaff:** Writing – review & editing, Visualization, Validation, Investigation. **Joëlle Lacroix:** Investigation, Conceptualization, Writing – review & editing, Validation, Project administration. **Alexis Lechervy:** Investigation, Methodology, Writing – review & editing, Validation. **Dinu Stefan:** Investigation, Methodology, Conceptualization, Writing – review & editing, Validation, Project administration. **Aurélien**

Corroyer-Dulmont: Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization, Software, Visualization.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgments

This study was funded by the Région Normandie thought “Booster IA” grant.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2025.121002.

References

Achrol, A.S., et al., 2019. Brain metastases. Nat. Rev. Dis. Primer 5 (n. 1), 5. https://doi.org/10.1038/s41572-018-0055-y janv.

Ahamed, M.F., et al., 2023. A review on brain tumor segmentation based on deep learning methods with federated learning techniques. Comput. Med. Imaging Graph. Off. J. Comput. Med. Imaging Soc. 110, 102313. https://doi.org/10.1016/j.compmedimag.2023.102313 déc.

Anaconda Software Distribution. Anaconda Inc., 2020. [En ligne]. Disponible sur: https://docs.anaconda.com.

Bondiau, P., et al., 2022. PD-0330 AI-based OAR annotation for pediatric brain radiotherapy planning. Radiother. Oncol. 170, S293-S295. https://doi.org/10.1016/S0167-8140(22)02823-7 mai.

M.J. Cardoso et al., MONAI: an open-source framework for deep learning in healthcare, nov. 2022.

Chartrand, G., et al., 2022. Automated detection of brain metastases on T1-weighted MRI using a convolutional neural network: impact of volume aware loss and sampling strategy. J. Magn. Reson. Imaging JMRI 56 (6), 1885-1898. https://doi.org/10.1002/jmri.28274 ndéc.

Cho, J., et al., 2021b. Deep learning-based computer-aided detection system for automated treatment response assessment of brain metastases on 3D MRI. Front. Oncol. 11. https://doi.org/10.3389/fonc.2021.739639 oct.

Cho, S.J., Sunwoo, L., Baik, S.H., Bae, Y.J., Choi, B.S., Kim, J.H., 2021a. Brain metastasis detection using machine learning: a systematic review and meta-analysis. Neuro-Oncol. 23 (2), 214-225. https://doi.org/10.1093/neuonc/noaa232 nfévr.

Dikici, E., et al., 2020. Automated brain metastases detection framework for T1-weighted contrast-enhanced 3D MRI. IEEE J. Biomed. Health Inform. 24 (n. 10), 2883-2893. https://doi.org/10.1109/JBHI.2020.2982103 oct.

Ene, C.I., et al., 2024. Response of treatment-naive brain metastases to stereotactic radiosurgery. Nat. Commun. 15 (n. 1), 3728. https://doi.org/10.1038/s41467-024-47998-8 mai.

A. Hatamizadeh et al., UNETR: transformers for 3D medical image segmentation, mars 2021, [En ligne]. Disponible sur: http://arxiv.org/abs/2103.10504.

Hessen, E.D., et al., 2017. Significant tumor shift in patients treated with stereotactic radiosurgery for brain metastasis. Clin. Transl. Radiat. Oncol. 2, 23-28. https://doi.org/10.1016/j.ctro.2016.12.007 févr.

Hsu, D.G., et al., 2023. Automatically tracking brain metastases after stereotactic radiosurgery. Phys. Imaging Radiat. Oncol. 27, 100452. https://doi.org/10.1016/j.phro.2023.100452 juill.

Huang, Y., et al., 2022. Deep learning for brain metastasis detection and segmentation in longitudinal MRI data. Med. Phys. 49 (n. 9), 5773-5786. https://doi.org/10.1002/mp.15863 sept.

Hunter, J.D., 2007. Matplotlib: a 2D graphics environment. Comput. Sci. Eng. 9 (n. 3), 90-95. https://doi.org/10.1109/MCSE.2007.55.

Kickingereder, P., et al., 2019. Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study. Lancet Oncol. 20 (5), 728-740. https://doi.org/10.1016/S1470-2045(19)30098-1 nmai.

Li, R., et al., 2023. MRI-based two-stage deep learning model for automatic detection and segmentation of brain metastases. Eur. Radiol. 33 (5), 3521-3531. https://doi.org/10.1007/s00330-023-09420-7 nmai.

Lin, L.I., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45 (1), 255-68 nmars.

Lin, N.U., et al., 2015. Response assessment criteria for brain metastases: proposal from the RANO group. Lancet Oncol. 16 (6). https://doi.org/10.1016/S1470-2045(15)70057-4 np. e270-8juin.

Luo, X., et al., 2024. Automated segmentation of brain metastases with deep learning: a multi-center, randomized crossover, multi-reader evaluation study. Neuro-Oncol. https://doi.org/10.1093/neuonc/noae113 p. noae113, juill.

Mason, D., 2011. SU-E-T-33: pydicom: an open source DICOM library. Med. Phys. 38, 3493-3493. https://doi.org/10.1118/1.3611983 n. 6Part10juin.

Masoudi, S., et al., 2021. Quick guide on radiology image pre-processing for deep learning applications in prostate cancer research. J. Med. Imaging Bellingham Wash 8 (1), 010901. https://doi.org/10.1117/1.JMI.8.1.010901 njanv.

et Mekki, L., Acharya, S., Ladra, M., Lee, J., 2024. Deep learning segmentation of organs-at-risk with integration into clinical workflow for pediatric brain radiotherapy. J. Appl. Clin. Med. Phys. 25 (n. 3). https://doi.org/10.1002/acm2.14310. mars.

Ocaña-Tienda, B., et al., 2024. Volumetric analysis: rethinking brain metastases response assessment. Neuro-Oncol. Adv. 6 (1). https://doi.org/10.1093/noajnl/vdad161 njanv.

et Ohtakara, K., Tanahashi, K., Kamomae, T., Miyata, K., Suzuki, K., 2023. Correlation of brain metastasis shrinking and deviation during 10-fraction stereotactic radiosurgery with late sequela: suggesting dose ramification between tumor eradication and symptomatic radionecrosis. Cureus 15 (n. 1), e33411. https://doi.org/10.7759/cureus.33411. janv.

Pang, Y., et al., 2024. Slim UNETR: scale hybrid transformers to efficient 3D medical image segmentation under limited computational resources. IEEE Trans. Med. Imaging 43 (3), 994-1005. https://doi.org/10.1109/TMI.2023.3326188 nmars.

Pati, S., et al., 2022. Federated learning enables big data for rare cancer boundary detection. Nat. Commun. 13 (n. 1), 7346. https://doi.org/10.1038/s41467-022-33407-5 déc.

Peerlings, J., et al., 2019. Stability of radiomics features in apparent diffusion coefficient maps from a multi-centre test-retest trial. Sci. Rep. 9 (1), 4800. https://doi.org/10.1038/s41598-019-41344-5 nmars.

A. Shaker, M. Maaz, H. Rasheed, S. Khan, M.-H. Yang, et F.S. Khan, UNETR++: delving into efficient and accurate 3D medical image segmentation, 2022, *arXiv*. doi:10.48550/ARXIV.2212.04497.

Stathakis, S., et al., 2022. Evaluation of AI vs. clinical experts SBRT-thorax computed tomography OARs delineation. Int. J. Radiat. Oncol. 114 (n. 3), e102-e103. https://doi.org/10.1016/j.ijrobp.2022.07.897 nov.

et Taha, A.A., Hanbury, A., 2015. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med. Imaging 15, 29. https://doi.org/10.1186/s12880-015-0068-x. août.

Waskom, M., 2021. Seaborn: statistical data visualization. J. Open Source Softw. 6 (60), 3021. https://doi.org/10.21105/joss.03021 navr.

Xue, J., et al., 2020. Deep learning-based detection and segmentation-assisted management of brain metastases. Neuro-Oncol. 22 (n. 4), 505-514. https://doi.org/10.1093/neuonc/noz234 avr.

J. Yosinski, J. Clune, Y. Bengio, et H. Lipson, How transferable are features in deep neural networks?, nov. 2014.

H.-Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, et Y. Yu, nnFormer: interleaved transformer for volumetric segmentation, 2021, *arXiv*. doi:10.48550/ARXIV.2109.03201.

Ziegler, E., et al., 2020. Open health imaging foundation viewer: an extensible open-source framework for building web-based imaging applications to support cancer research. JCO Clin. Cancer Inform. 4, 336-345. https://doi.org/10.1200/CCI.19.00131 avr.