medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fully automated brain metastases segmentation using T1-weighted contrast-enhanced MR images before and after Stereotactic Radiosurgery

Hemalatha Kanakarajan[1], Wouter De Baene[1], Patrick Hanssens[2,3], Margriet Sitskoorn[1]

1 Department of Cognitive Neuropsychology, Tilburg University, Tilburg, The Netherlands
2 Gamma Knife Center, Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands
3 Department of Neurosurgery, Elisabeth-TweeSteden Hospital, Tilburg, The Netherlands

Abstract:

Background: Brain metastases (BM) represent the most common intracranial tumor in adults. An estimated 20% of all patients with cancer will develop BM. Stereotactic Radiosurgery (SRS) is a major treatment option for BM. For SRS treatment planning and outcome evaluation, magnetic resonance imaging (MRI) are acquired before and at multiple stages during the follow-up. Accurate segmentation of brain tumors on MRI is crucial for treatment planning and response evaluation. Detection and segmentation of BM is a tedious and time-consuming task for many radiologists that could be optimized with machine learning methods. Previous studies evaluated the segmentation performance of several deep learning algorithms, but focused mainly on training and testing the models on the planning MR images only. The purpose of this study was to investigate a well-known deep learning approach (nnU-Net) for BM segmentation and to evaluate its performance on both planning MR images and follow-up MR images based on training on planning MR images only and testing with both planning MR and follow-up MR images.

Method: Pre-treatment contrast-enhanced T1-weighted brain MRIs(i.e. the planning MRI) were collected retrospectively for 263 patients with BM. Scans were made as part of clinical care at the Gamma Knife Center of the Elisabeth-TweeSteden Hospital (Tilburg, the Netherlands). This total of 263 patients were split into 203 patients for model training/validation and 60 patients for testing. For these 60 patients used for testing, the post treatment contrast-enhanced follow-up T1-weighted brain MRI scans(i.e. follow-up MRI) were also retrospectively collected. These 60 patients who were part of the testing set are from the set of patients included in the Cognition And Radiation Study A(CAR-Study A) at ETZ. The follow-up (FU) scans were made at 3, 6, 9, 12, 15, and 21 months after treatment. The nnU-Net model was trained with the planning MR images, and then tested separately against the planning and follow-up MR images.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Results: When tested with planning MR images, the model obtained a dice similarity coefficient (DSC) of 0.940, a False Negative Rate (FNR) of 0.065 and a sensitivity of 0.934. When tested with the follow-up MR images 3, 6, 9, 12, 15 and 21 months after treatment , the model obtained, respectively, a DSC of 0.759, 0.667, 0.604, 0.589, 0.666 and 0.574, an FNR of 0.288, 0.379, 0.445, 0.470, 0.409, and 0.487 and a sensitivity of 0.711, 0.620, 0.554, 0.529, 0.590, and 0.512.

Conclusion: The model achieved a good performance score for planning MR images. The nnU-Net model can automatically detect and segment brain metastases with high sensitivity, and low FNR. Though there is a decline in the DSC and an increase in the FNR of the model for the follow-up MR images, the algorithm could be a beneficial tool for clinicians and assist them for diagnosis, treatment planning and treatment response evaluations during follow-ups of BM patients.

Introduction:

Brain metastases represent the most common intracranial tumor in adults [1]. An estimated 20% of all patients with cancer will develop brain metastases [2]. Although some patients who develop brain metastases remain asymptomatic, many patients show neurological symptoms including headaches, nausea, vomiting, dizziness, focal neurological deficits, epileptic seizures, and cognitive impairment [4, 5]. Advances in the treatment of primary tumors have led to prolonged life expectancy and therefore increased the probability of developing brain metastases [1]. The overall prognosis for patients with BM remains poor [6]. Brain metastases account for a disproportionately high percentage of morbidity and mortality among patients with cancer [3], with dismal 2- and 5-year survival rates of 8.1% and 2.4% after diagnosis [4].

Conventional therapies for treating brain metastases include surgical resection, Whole Brain Radiotherapy (WBRT), SRS, and their combination. The combination of these modalities remains the preferred treatment for patients with good systematic performance [7]. Surgery is a treatment option for patients with a single, surgically accessible metastasis, good systemic control, an expected survival of at least three months, and with tumor size larger than 3-4 cm [8, 9]. WBRT has been used as adjuvant therapy after surgical resection or as primary therapy for patients with multiple metastases who are not suitable for surgical resection [7]. With WBRT, the entire brain, including healthy brain tissue, is irradiated multiple times. WBRT has long been the standard of care for multiple BM. However, the

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

toxicity of WBRT remained the concern of clinical treatment. Adverse cognitive effects are common neurotoxic effects in patients who have undergone WBRT [10]. Therefore, SRS is generally performed to avoid the neurocognitive side effects of WBRT [11]. With SRS, the brain metastases are targeted very precisely, whereby the dose of radiation to the healthy brain tissue is limited.

For SRS treatment planning, the physician must manually contour the multitude of presenting lesions on co-registered, three-dimensional MR or CT images. This process is labor-intensive and prone to significant variability among physicians[12]. An automatic and robust system for detecting and contouring brain metastases would facilitate more precise and efficient treatment delivery in the radiotherapy clinic. Automated tools that assist radiologists and radiation oncologists in their respective roles in detection and delineation of multiple metastases can positively impact both the efficiency as well as efficacy of management of patients with multiple BMs.

Deep learning models (DLMs) have shown great potential in detection, segmentation and classification tasks in medical image analysis while having the potential to improve clinical workflow[26]. Several approaches have been introduced for brain metastasis segmentation in MRI using deep learning[16]. The first application which produced state-of-the-art results in automated segmentation of BM in MRI was published in 2015 by Losch et al. [27]. Since then, a large variety of network architectures for deep learning such as convolutional neural networks (CNNs) and DeepMedic have been tested. One limitation of these studies is that they were mainly focused on training and testing the models on the planning MR images only (e.g. [15, 17, 19, 21, 22, 24, 25]). This is a limitation because the performance of the deep learning algorithms on follow-up MR images might not be the same as their performance on the planning MR images. This could be because of the shrinkage of the tumors due to the radiation effect. The evaluation of the performance on the follow-up MRI images is necessary to establish the applicability of these algorithms to assist the clinicians for the response evaluation during follow-ups. Jalalifar et al[28]. evaluated the performance of a deep learning model on the follow-up MR images but presented the performance results for five sample patients only.

One of the popular deep learning network architectures is the so-called nnU-Net[23]. Isensee et al [20] demonstrated how this architecture achieved state of the art performance on different challenges in segmentation of medical images by applying it to 10 international biomedical image segmentation challenges comprising 19 different datasets and 49 segmentation tasks across a variety of

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

organs, organ substructures, tumors, lesions and cellular structures in MRI, computed tomography scans (CT) as well as electron microscopy (EM) images. Ziyaee et al[18] evaluated this algorithm for BM by training and testing it with planning MR images only. The model achieved an overall DSC of 82.2%, which shows good segmentation performance. By comparison to other algorithms, the model achieved the best detection performance. But the performance of this nnU-Net algorithm for the segmentation of the follow-up images is not yet evaluated. In the present work we addressed this gap and assessed the applicability of nnU-Net for automated segmentation of both planning and follow-up images. Hence the objective of this study is to evaluate the effectiveness of the nnU-Net algorithm for the segmentation of planning and follow up images. At ETZ, the segmentations are done only for the planning MRI scans and not for the follow-up scans. This could be the case in other hospitals also. This lack of ground truth segmentations creates a limitation for training the deep learning algorithms with follow-up scans. In this work, we evaluated the performance of the nnU-Net algorithm by training it with planning images only and testing it with both planning and follow-up images. This evaluation will help to understand if this state of the art deep learning algorithm can assist the clinicians in detection and segmentation of BM images for treatment planning and treatment response evaluation during follow-ups.

Method:

For this study, pre-treatment contrast-enhanced (with triple dose gadolinium) T1-weighted brain MRIs of 263 BM patients were used. These planning MRI scans were collected using a 1.5T Philips Ingenia scanner (Philips Healthcare, Best, The Netherlands). The voxel size was 0.82 x 0.82 x 1.5mm. Scans were made as part of clinical care at the Gamma Knife Center of the Elisabeth-TweeSteden Hospital (ETZ) at Tilburg, The Netherlands. The total of 263 patients were split into 203 patients for model training and 60 patients for testing. For the 203 patients in the training data set, the treatment type was decided by assessing the volume of the tumors in the planning MRI. The patients underwent either GKRS at the Gamma Knife Center or were referred to WBRT or surgery at the other departments. The 60 patients who were part of the testing set are from the set of patients included in the Cognition And Radiation Study A (CAR-Study A) at ETZ [29]. Our test set is a random subset of the set of the patients included in this CAR-Study A. All the patients in the test data set underwent GKRS. Patients with other brain tumor types in addition to brain metastases were excluded from the test data set (n=6). For example, some patients also had a meningioma in addition to brain metastases and hence they were excluded from the test set. After this exclusion, there were 54 patients in the test data set. The

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

segmentations of the baseline ground truth were manually delineated by expert oncologists and neuroradiologists at ETZ. The manually delineated ground truth for follow-up scans were only available for the patients who were part of the CAR-Study A.

For the 54 patients used for testing, the post treatment contrast enhanced (with single dose gadolinium) T1-weighted brain follow-up MRI scans were also retrospectively collected. Though the slice thickness of the follow-up scans ranged from 0.21 mm to 1.5 mm, the majority of the scans had a slice thickness of 0.8 mm. The images from 6 follow-up sessions were available. The follow-up (FU) scans were made at 3, 6, 9, 12, 15, and 21 months after treatment. For these follow-ups, scans of 54(FU1), 41(FU2), 32(FU3), 27(FU4), 19(FU5) and 14(FU6) patients were available.

As a preprocessing step, all the MRI scans were registered to standard MNI space using Dartel in SPM12, implemented in Python using the Nipype(Neuroimaging in Python–Pipelines and Interfaces) software package (Gorgolewski et al., 2011). The voxel size of the normalized image was set to 1*1*1. For all other normalization configurations, the default values offered by SPM12 were used. One other preprocessing step was to combine the ground truth labels for patients with more than one BM to create a single ground truth mask with all the BMs. FSL library was used for this integration[30].

The nnU-Net algorithm was used to automatically segment the brain images[23].It is a framework built on top of the U-Net[23]. Based on properties of the dataset, it makes key design choices for pre-and post-processing, data augmentation, network architecture, training scheme, and inference[23]. These automatic design choices allow nnU-Net to perform well on many medical segmentation tasks. The nnU-Net model was trained with the planning MR images in 3d full resolution mode. The trained model was then tested separately against the planning and follow-up images.

To assess the quality of the resulting segmentations, multiple metrics were employed. The dice similarity coefficient (DSC) measures the overlap with the ground truth (ranging from 0 for no overlap to 1 for perfect overlap) per patient. It is calculated by dividing the double of the area of overlap by the sum of the areas of the predicted and the ground truth segmentation. The algorithm's performance in detecting individual metastases was measured by sensitivity (number of pixels in the detected metastases divided by the number of pixels in all metastases contained in ground truth), and by the false negative rate (FNR). The FNR is the probability that a true metastasis will be missed by the model. In the results section, these metrics are presented for the predictions done for baseline and for the follow up test data.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Results:

Table 1 shows the characteristics of patients included in our study.

	Training set	Baseline test set	FU1 test set	FU2 test set	FU3 test set	FU4 test set	FU5 test set	FU6 test set
Number of patients	210	54	54	41	32	27	19	14
Gender Male Female	 100 110	 26 28	 26 28	 19 22	 15 17	 12 15	 7 12	 5 9
Age Average Min Max	 62 34 85	 62 32 81	 62 32 81	 63 44 81	 63 44 81	 63 44 81	 62 49 81	 62 49 75
Total number of BM	1321	203	208	182	134	81	65	39
Highest tumor volume (cm3)	64.26	20.8	16.21	20.85	11.14	14.479	14.29	30.04
Smallest tumor volume (cm3)	1	3	3	1	1	2	1	2

Table 1: Patient characteristics

The tumor segmentation results obtained for the baseline and the FU tests are shown in table 2. The table presents the performance metrics for the baseline and follow up tests. The mean DSC when tested with the planning MR images is 0.940 and it is lower for the tests conducted with follow-up MRI images. The mean FNR for the planning MR images is 0.065 and it is higher for the tests conducted with

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

follow-up MRI images. The mean sensitivity is 0.934 for the test with planning MR images and has lower values for the tests with follow-up MRI images.

	Baseline	FU1	FU2	FU3	FU4	FU5	FU6
DSC	0.940	0.759	0.667	0.604	0.589	0.666	0.574
FNR	0.065	0.288	0.379	0.445	0.470	0.409	0.487
Sensitivity	0.934	0.711	0.620	0.554	0.529	0.590	0.512
Percentage of missed BMs	8.86	22.11	28.57	47.76	28.39	18.46	35.89
Avg. missed BM per patient	0.333	0.851	1.268	2.000	0.851	0.631	1.000

Table 2: Performance of the segmentation model

The tumor segmentation results obtained for a representative patient with good results at baseline and at six follow-up sessions are shown in Figure 1. The tumors in the ground truth and segmented output are marked with green and red respectively and overlaid on each other. The overlapping region is marked in yellow. There are 2 tumors in the ground truth in the baseline and in all follow-up scans. The images show that the tumors have shrunk over time. Both the tumors are predicted correctly in the tests with baseline images and in all follow-ups except for the FU2. In the generated segmentation outcome for FU2, only one of the tumors is visible. The DSC, FNR and sensitivity for this patient in the baseline and in the follow-ups tests are shown in Table 3.

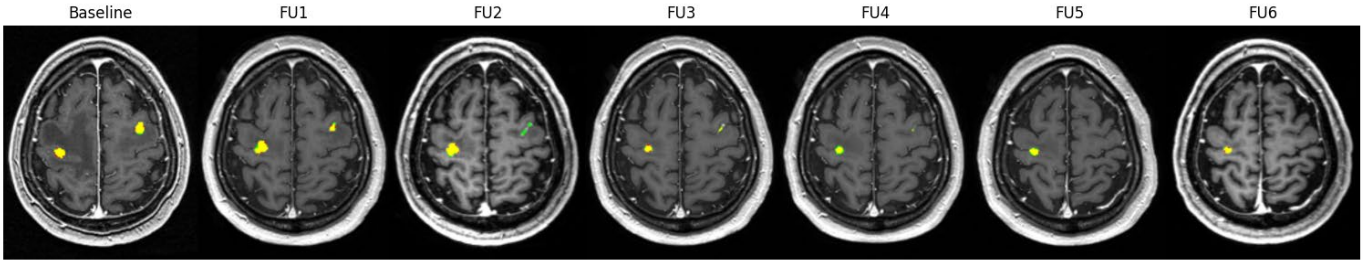

Figure 1: Segmentation result for representative patient 1

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

DSC	0.931	0.929	0.775	0.914	0.858	0.862	0.811
FNR	0.079	0.083	0.336	0.131	0.224	0.206	0.208
Sensitivity	0.920	0.916	0.663	0.868	0.775	0.793	0.791

Table 3: Performance for the sample patient

The tumor segmentation results obtained for another representative patient at the baseline and at six follow-up sessions are shown in Figure 2. The model showed good performance for this patient at baseline, FU1, FU2 but showed poor performance results for the subsequent follow-ups. There is 1 tumor in the ground truth in the baseline and in all follow-up scans. Similar to Figure 1, the tumors in the ground truth and segmented output are marked with green and red respectively and overlaid on each other. The overlapping region is marked in yellow. The images show that the tumor has shrunk over time. This tumor is predicted correctly in the tests with baseline images and in FU1, FU2 and not predicted for the subsequent follow-up tests. The DSC of this patient was 0.949, 0.946 and 0.836 for baseline, FU1 and FU2 respectively. For the subsequent tests, the DSC was 0 because of the missed tumor.

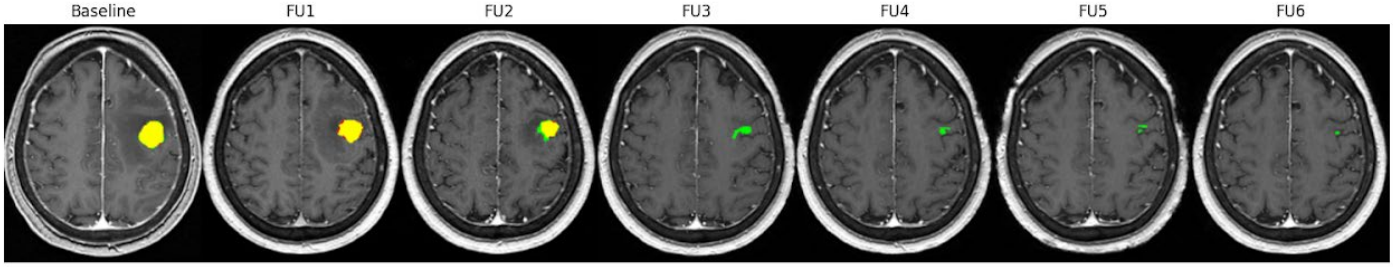

Figure 2: Segmentation result for representative patient 2

Another interesting finding is that the model also detected some tumors that are missing in the ground truth. Some of the extra tumors detected by the model are part of the ground truth of subsequent scans. For example, for some patients the model detected an extra tumor in the baseline test. The ground truth masks of the baseline did not contain this tumor but the ground truth masks of FU1 contained these tumors.

We also compared the performance for the patients who received the GKRS only once at the baseline with the patients who also received GKRS for local recurrence during follow-ups. Table 4 shows the performance for a sample patient (patient 1) who received GKRS only at the baseline and the performance for another sample patient (patient 2) who received the GKRS at the baseline and at two

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

follow-ups. We did not observe a significantly higher drop in performance for the patient who received multiple treatments when compared to the patient who received the treatment only once.

	Baseline	FU1	FU2	FU3	FU4	FU5	FU6
DSC for patient 1	0.949	0.920	0.865	0.812	0.860	0.707	0.581
DSC for patient 2	0.974	0.921	0.913	0.921	0.926	0.929	0.458

Table 4: performance comparison for single treatment vs multiple treatments

Discussion:

In this study, the nnU-Net deep-learning algorithm was evaluated for automatic segmentation of brain tumors on T1- weighted MR images before and after radiation therapy. When tested on the pre-treatment test data set, the model achieved a DSC of 0.940, FNR of 0.065 and a sensitivity of 0.934. The performance of the model for the baseline and the FU tests are shown in Table 2. The performance of the model for the follow up was lower than the performance for the planning MRI scans. The performance of the model when tested for the follow-up images obtained after 3 months (T3) was closer to the performance for the planning MRI images. The performance of the follow-up after FU1 was lower than compared to the baseline and FU1.

The performance of the model for the baseline is higher when compared to other similar studies. For example, Dylan g et al[21], expounded a fully 3D deep learning approach capable of automatically detecting and segmenting brain metastases using T1 contrast and CT images. The DSC of this algorithm was found to be 0.76. Endre et al[25] observed a DSC of 0.79 while evaluating a deep learning algorithm for detection and segmentation of BM on multisequence MRI.

When tested on planning MR images, the model did miss some of the tumors that were present in the ground truth. We observed that the model tends to miss metastases that are either near a blood vessel or located near the tentorium. Additionally, some false positive segmentations turned out to be blood vessels. However the model did detect and segment some tumors that were missed in the ground truth. Some of these extra tumors that were detected by the model were part of the ground truth of the

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

subsequent follow-up scans. This shows that the model can assist the clinicians in the early detection and segmentation of the tumors.

The performance of the model for the follow up was lower than for the planning MR scans. This could be due to the fact that the amount of contrast admistered to the followup images was less than the amount of contrast adminstered to the baseline images. The baseline images were contrast-enhanced with triple-dose gadolinium and the follow-up images were contrast-enhanced with single-dose gadolinium. The decrease in performance for the follow-ups may also be due to the radiation effect which causes the tumors to shrink over time. The detection and segmentation performances of the deep learning algorithms tend to decrease for smaller lesions[22]. Hence, the shrinkage of the tumors over time due to the radiation effect could be a reason for the lower performance for successive follow-up scans. Alternatively, the change in the texture of the tumor at follow-up months after multiple sessions of treatment could make it harder for the algorithm to detect. We compared the performance of those patients with only one treatment with the patients who had multiple treatments over time for local recurrence. The decline in performance for these two categories are similar and we did not observe a higher drop in performance for patients who received multiple treatments when compared to patients who received the treatment only once. This shows that the declining performance is less likely due to the treatments over time for local recurrence. The difference in contrast or slice thickness between the planning and follow-up images might also cause some difference in the performance of the models on the follow-up MR images.

A limitation of this work is that our test samples only included BM patients who were included for GKRS. This means that the performance of the algorithm could be different for the follow-up images after a different treatment approach. Future work could be to evaluate the performance of the algorithm for a larger sample size which also includes patients treated with other treatment types.

Results from this study showed that the algorithm achieved a good performance score for planning MR images. The nnU-Net model can automatically detect and segment brain metastases with high sensitivity, and low FNR for treatment planning. It could therefore be a beneficial tool for clinicians and assist them in diagnosis and treatment planning.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

In the present work we assessed the applicability of nnU-Net for automated segmentation of both planning and follow-up MR images for BM patients. At ETZ, the segmentations are done only for the baseline scans and not for the follow-up scans. This could be the case in other hospitals also. This lack of ground truth segmentations creates a limitation for training the deep learning algorithms with follow-up scans. In this work, we evaluated the performance of the nnU-Net algorithm by training it with planning MR images only and testing it with both planning and follow-up images. To the best of our knowledge, the performance of the algorithm exceeded the performance reported by other similar studies for segmentation of planning MR images. Though there is a decline in the performance of the model for the follow-up images, the algorithm could be a beneficial tool for clinicians and assist them in diagnosis, treatment planning and treatment response evaluations during follow-ups.

Acknowledgment:

This research is supported by KWF Kankerbestrijding and NWO Domain AES, as part of their joint strategic research programme: Technology for Oncology IL. The collaboration project is co-funded by the PPP Allowance made available by Health Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships.

We would also like to acknowledge the support provided by Eline Verhaak for this research and thank her for helping us with the manually delineated ground truth for follow-up scans from the CAR study.

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References

1. Donofrio CA, Cavalli A, Gemma M, et al. Cumulative intracranial tumour volume prognostic assessment: a new predicting score index for patients with brain metastases treated by stereotactic radiosurgery. Clinical & Experimental Metastasis. 2020;37(4):499-508. doi:https://doi.org/10.1007/s10585-020-10037-z
2. Barnholtz-Sloan JS, Sloan AE, Davis FG, Vigneau FD, Lai P, Sawaya RE. Incidence Proportions of Brain Metastases in Patients Diagnosed (1973 to 2001) in the Metropolitan Detroit Cancer Surveillance System. Journal of Clinical Oncology. 2004;22(14):2865-2872. doi:https://doi.org/10.1200/jco.2004.12.149
3. Gavrilovic IT, Posner JB. Brain metastases: epidemiology and pathophysiology. Journal of Neuro-Oncology. 2005;75(1):5-14. doi:https://doi.org/10.1007/s11060-004-8093-6
4. Achrol AS, Rennert RC, Anders C, et al. Brain metastases. Nature Reviews Disease Primers. 2019;5(1). doi:https://doi.org/10.1038/s41572-018-0055-y
5. Kaal EC, Niël CG, Vecht CJ. Therapeutic management of brain metastasis. The Lancet Neurology. 2005;4(5):289-298. doi:https://doi.org/10.1016/s1474-4422(05)70072-7
6. Rastogi K, Bhaskar S, Gupta S, Jain S, Singh D, Kumar P. Palliation of Brain Metastases: Analysis of Prognostic Factors Affecting Overall Survival. Indian Journal of Palliative Care. 2018;24(3):308-312. doi:https://doi.org/10.4103/IJPC.IJPC_1_18
7. Liu Q, Tong X, Wang J. Management of brain metastases: history and the present. Chinese Neurosurgical Journal. 2019;5(1). doi:https://doi.org/10.1186/s41016-018-0149-0
8. Navarro-Olvera JL, Ariñez-Barahona E, Esqueda-Liquidano MA, Muñoz-Cobos A. Brain metastases: Literature review. Revista Médica del Hospital General de México. 2017;80(1):60-66. doi:https://doi.org/10.1016/j.hgmx.2016.04.006
9. Enrique GV, Irving SR, Ricardo BI, et al. Diagnosis and management of brain metastases: an updated review from a radiation oncology perspective. Journal of Cancer Metastasis and Treatment. 2019;2019. doi:https://doi.org/10.20517/2394-4722.2019.20
10. Jani A, Rozenblat T, Yaeh AM, et al. The Energy Index Does Not Affect Local Control of Brain Metastases Treated by Gamma Knife Stereotactic Radiosurgery. Neurosurgery. 2015;77(1):119-125. doi:https://doi.org/10.1227/neu.0000000000000750
11. Abel RJ, Lederman A, Ji L, et al. Stereotactic Radiosurgery to the Resection Cavity for Brain Metastases: Prognostic Factors and Outcomes. International Journal of Radiation Oncology*Biology*Physics. 2014;90(1):S331. doi:https://doi.org/10.1016/j.ijrobp.2014.05.1091
12. Rudie JD, Weiss DJ, Colby JB, et al. Three-dimensional U-Net Convolutional Neural Network for Detection and Segmentation of Intracranial Metastases. Radiology. 2021;3(3):e200204-e200204. doi:https://doi.org/10.1148/ryai.2021200204
13. LeCun Y, Bengio Y, Hinton G. Deep Learning. Nature. 2015;521(7553):436-444. doi:https://doi.org/10.1038/nature14539
14. Shin IG, Kim HC, Ahn SH, et al. Development and Validation of a Deep Learning–Based Model to Distinguish Glioblastoma from Solitary Brain Metastasis Using Conventional MR Images. American Journal of Neuroradiology. 2021;42(5):838-844. doi:https://doi.org/10.3174/ajnr.a7003
15. Lenhard Pennig, Shahzad R, Luciana Albuquerque Caldeira, et al. Automated Detection and Segmentation of Brain Metastases in Malignant Melanoma: Evaluation of a Dedicated Deep Learning Model. American Journal of Neuroradiology. 2021;42(4):655-662. doi:https://doi.org/10.3174/ajnr.a6982

medRxiv preprint doi: https://doi.org/10.1101/2023.07.07.23292387; this version posted July 8, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

16. Nayak L, Lee EQ, Wen PY. Epidemiology of Brain Metastases. Current Oncology Reports. 2011;14(1):48-54. doi:https://doi.org/10.1007/s11912-011-0203-y
17. Liang Y, Lee K, Bovi JA, et al. Deep Learning-Based Automatic Detection of Brain Metastases in Heterogenous Multi-Institutional Magnetic Resonance Imaging Sets: An Exploratory Analysis of NRG-CC001. International Journal of Radiation Oncology*Biology*Physics. 2022;114(3):529-536. doi:https://doi.org/10.1016/j.ijrobp.2022.06.081
18. Hamidreza Ziyaee, Ibbott GS, Debra Nana Yeboa, et al. Automated Brain Metastases Segmentation With a Deep Dive Into False-positive Detection. Advances in radiation oncology. 2022;8(1):101085-101085. doi:https://doi.org/10.1016/j.adro.2022.101085
19. Jünger ST, Hoyer UCI, Schaufler D, et al. Fully Automated MR Detection and Segmentation of Brain Metastases in Non-small Cell Lung Cancer Using Deep Learning. Journal of magnetic resonance imaging: JMRI. 2021;54(5):1608-1622. doi:https://doi.org/10.1002/jmri.27741
20. Isensee F, Jager PF, Simon, Petersen J, Klaus Maier-Hein. Automated Design of Deep Learning Methods for Biomedical Image Segmentation. arXiv: Computer Vision and Pattern Recognition. Published 2019. Accessed April 30, 2023. https://www.semanticscholar.org/paper/Automated-Design-of-Deep-Learning-Methods-for-Image-Isensee Jager/4692b1eefa9bc43c23075412aeb4181a81e0bbe3
21. Hsu DG, Ballangrud Å, Shamseddine A, et al. Automatic segmentation of brain metastases using T1 magnetic resonance and computed tomography images. Physics in Medicine & Biology. 2021;66(17):175014. doi:https://doi.org/10.1088/1361-6560/ac1835
22. Yoo Y, Ceccaldi P, Liu S, et al. Evaluating deep learning methods in detecting and segmenting different sizes of brain metastases on 3D post-contrast T1-weighted images. Journal of Medical Imaging. 2021;8(03). doi:https://doi.org/10.1117/1.jmi.8.3.037001
23. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods. 2020;18(2):203-211. doi:https://doi.org/10.1038/s41592-020-01008-z
24. Bousabarah K, Ruge M, Brand JS, et al. Deep convolutional neural networks for automated segmentation of brain metastases trained on clinical data. Radiation Oncology. 2020;15(1). doi:https://doi.org/10.1186/s13014-020-01514-6
25. Grøvik E, Yi D, Iv M, Tong E, Rubin D, Zaharchuk G. Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI. Journal of Magnetic Resonance Imaging. Published online May 2, 2019. doi:https://doi.org/10.1002/jmri.26766
26. Hwang EJ, Park CM. Clinical Implementation of Deep Learning in Thoracic Radiology: Potential Applications and Challenges. Korean Journal of Radiology. 2020;21(5):511. doi:https://doi.org/10.3348/kjr.2019.0821
27. Charron O, Lallement A, Jarnet D, Noblet V, Clavier JB, Meyer P. Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network. Computers in Biology and Medicine. 2018;95:43-54. doi:https://doi.org/10.1016/j.compbiomed.2018.02.004
28. Jalalifar A, Soliman H, Sahgal A, Sadeghi-Naini A. A Cascaded Deep-Learning Framework for Segmentation of Metastatic Brain Tumors Before and After Stereotactic Radiation Therapy. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). Published online July 2020. doi:https://doi.org/10.1109/embc44109.2020.9175489
29. Verhaak E, Gehring K, Hanssens PEJ, Sitskoorn MM. Health-related quality of life of patients with brain metastases selected for stereotactic radiosurgery. Journal of Neuro-Oncology. 2019;143(3):537-546. doi:https://doi.org/10.1007/s11060-019-03186-z
30. Jenkinson M, Beckmann CF, Behrens TEJ, Woolrich MW, Smith SM. FSL. NeuroImage. 2012;62(2):782-790. doi:https://doi.org/10.1016/j.neuroimage.2011.09.015