Radiotherapy and Oncology 178 (2023) 109425

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Development and external validation of an MRI-based neural network for brain metastasis segmentation in the AURORA multicenter study

Josef A. Buchner [a,1,*], Florian Kofler [b,c,d,e,1], Lucas Etzel [a,f], Michael Mayinger [g], Sebastian M. Christ [g], Thomas B. Brunner [h], Andrea Wittig [i], Björn Menze [b], Claus Zimmer [c], Bernhard Meyer [j], Matthias Guckenberger [g], Nicolaus Andratschke [g], Rami A. El Shafie [k,l,m], Jürgen Debus [k,l], Susanne Rogers [n], Oliver Riesterer [n], Katrin Schulze [o], Horst J. Feldmann [o], Oliver Blanck [p], Constantinos Zamboglou [q,r,s], Konstantinos Ferentinos [s], Robert Wolff [t,u], Kerstin A. Eitz [a,f,v], Stephanie E. Combs [a,f,v], Denise Bernhardt [a,f], Benedikt Wiestler [c,d,1], Jan C. Peeken [a,f,v,1]

[a] *Department of Radiation Oncology, Klinikum rechts der Isar;* [b] *Department of Informatics;* [c] *Department of Diagnostic and Interventional Neuroradiology, Klinikum rechts der Isar;* [d] *TranslaTUM - Central Institute for Translational Cancer Research, Technical University of Munich;* [e] *Helmholtz AI, Helmholtz Zentrum Munich;* [f] *Deutsches Konsortium für Translationale Krebsforschung (DKTK), Partner Site Munich, Munich, Germany;* [g] *Department of Radiation Oncology, University Hospital of Zurich, University of Zurich, Zurich, Switzerland;* [h] *Department of Radiation Oncology, University Hospital Magdeburg, Magdeburg;* [i] *Department of Radiotherapy and Radiation Oncology, University Hospital Jena, Friedrich-Schiller University, Jena;* [j] *Department of Neurosurgery, Klinikum rechts der Isar, Technical University of Munich, Munich;* [k] *Department of Radiation Oncology, Heidelberg University Hospital;* [l] *Heidelberg Institute for Radiation Oncology (HIRO), National Center for Radiation Oncology (NCRO), Heidelberg;* [m] *Department of Radiation Oncology, University Medical Center Göttingen, Göttingen, Germany;* [n] *Radiation Oncology Center KSA-KSB, Kantonsspital Aarau, Aarau, Switzerland;* [o] *Department of Radiation Oncology, General Hospital Fulda, Fulda;* [p] *Department of Radiation Oncology, University Medical Center Schleswig Holstein, Kiel;* [q] *Department of Radiation Oncology, University of Freiburg - Medical Center;* [r] *German Cancer Consortium (DKTK), Partner Site Freiburg, Freiburg, Germany;* [s] *Department of Radiation Oncology, German Oncology Center, European University of Cyprus, Limassol, Cyprus;* [t] *Saphir Radiosurgery Center Frankfurt and Northern Germany, Guestrow;* [u] *Department of Neurosurgery, University Hospital Frankfurt, Frankfurt; and* [v] *Institute of Radiation Medicine (IRM), Department of Radiation Sciences (DRS), Helmholtz Center Munich, Munich, Germany*

ARTICLE INFO

Article history:
Received 29 September 2022
Received in revised form 17 November 2022
Accepted 18 November 2022
Available online 26 November 2022

Keywords:
Brain metastasis
Neural network
Stereotactic radiotherapy
MRI
External testing

ABSTRACT

Background: Stereotactic radiotherapy is a standard treatment option for patients with brain metastases. The planning target volume is based on gross tumor volume (GTV) segmentation. The aim of this work is to develop and validate a neural network for automatic GTV segmentation to accelerate clinical daily routine practice and minimize interobserver variability.

Methods: We analyzed MRIs (T1-weighted sequence ± contrast-enhancement, T2-weighted sequence, and FLAIR sequence) from 348 patients with at least one brain metastasis from different cancer primaries treated in six centers. To generate reference segmentations, all GTVs and the FLAIR hyperintense edematous regions were segmented manually. A 3D-U-Net was trained on a cohort of 260 patients from two centers to segment the GTV and the surrounding FLAIR hyperintense region. During training varying degrees of data augmentation were applied. Model validation was performed using an independent international multicenter test cohort (n = 88) including four centers.

Results: Our proposed U-Net reached a mean overall Dice similarity coefficient (DSC) of 0.92 ± 0.08 and a mean individual metastasis-wise DSC of 0.89 ± 0.11 in the external test cohort for GTV segmentation. Data augmentation improved the segmentation performance significantly. Detection of brain metastases was effective with a mean F1-Score of 0.93 ± 0.16. The model performance was stable independent of the center (p = 0.3). There was no correlation between metastasis volume and DSC (Pearson correlation coefficient 0.07).

Conclusion: Reliable automated segmentation of brain metastases with neural networks is possible and may support radiotherapy planning by providing more objective GTV definitions.

Brain metastases (BMs) are the most frequent type of brain tumor affecting approximately 25% of patients with cancer as shown in autopsy studies [1,2]. An effective treatment for patients with a limited number of BMs is stereotactic radiotherapy (SRT) [3,4]. In preparation and planning of the SRT, precise segmentation of the BMs in magnetic resonance imaging (MRI) is of utmost

* Corresponding author at: Klinik und Poliklinik für RadioOnkologie und Strahlentherapie, Klinikum rechts der Isar, Technische Universität München (TUM), Ismaninger Str. 22, 81675 München, Germany.
E-mail address: j.buchner@tum.de (J.A. Buchner).
[1] Shared Authorships.

https://doi.org/10.1016/j.radonc.2022.11.014
0167-8140/

importance [5]. Manual segmentation of BMs is a time-consuming task in the daily clinical routine and prone to inter-observer variability [6]. Therefore, a tool for fast and reliable segmentation of BMs is needed.

As a possible solution to this problem, the automatic analysis of imaging data offers great potential and has been successfully applied in primary brain tumors [7]. This can be attributed in part to the development of new neural network architectures like the 3D U-Net [8]. Furthermore, the required computing power is widely available nowadays and neural networks for image segmentation can be run on recent consumer-grade hardware.

Recently, multiple authors have demonstrated that the quantitative analyses of MRI studies of BMs as well as primary tumors paired with machine learning prediction models ("radiomics") may provide valuable information regarding prognosis, grade, or histological properties [9–12]. To allow for radiomic analysis, segmentation of the BMs as the volume of interest is necessary. An automized segmentation approach provides the benefits of better reproducibility independent of the operator.

Most published approaches focused on the detection, analysis, and segmentation of the contrast-enhancing metastasis itself. According to Priya et al. however, the most important radiomic features for differentiating glioblastoma from BMs were extracted from both the whole tumor segmentation and the edema segmentation [13]. Another study by Cao et al. concluded that valuable information is held within the peritumoral edema area [14]. Therefore, the segmentation of peritumoral edema complementary to the contrast-enhancing BM may be useful for further analysis of radiomic features.

MRI is an important tool in the diagnosis of BMs. Unlike the intensity values on computed tomography (CT) scans, which are made comparable using the Hounsfield Scale, the intensities on MRIs are non-standardized. They are, among other factors, dependent on the manufacturer and model of the specific MRI scanner [15]. Because of that, multicentric datasets created using different MRI scanners are needed to train reproducible and generalizable models.

The goal of this project is to train a neural network on the exact automatic segmentation of BMs as well as their surrounding edema and subsequently validate the network using a multicentric external test cohort. The automatically created segmentations could then not only be used for radiation treatment planning, but also as a basis for future radiomics models.

Materials and methods

AURORA study

The data was collected within *A Multicenter Analysis of Stereotactic Radiotherapy to the Resection Cavity of Brain Metastases* (AURORA) retrospective trial by the *"Radiosurgery and Stereotactic Radiotherapy Working Group"* of the *German Society for Radiation Oncology* (DEGRO). The inclusion criteria were: known primary tumor, resected BM, and SRT with a radiation dose of > 5 Gy per fraction. Exclusion criteria were: any previous cranial radiation therapy (RT), interval between surgery and RT > 100 days, and premature discontinuation of RT. Synchronous non-resected BMs were allowed but had to be treated simultaneously with SRT. Approval from the ethic committees was received at each institution (main approval at Technical University of Munich: 119/19 S-SR). While the trial is focusing on the postoperative situation, we only analyzed the preoperative imaging.

Data set

In total, we received imaging data from 453 patients from six different centers (FD: General Hospital Fulda, FFM: Saphir Radiochirurgie/University Hospital Frankfurt, HD: Heidelberg University Hospital, KSA: Kantonsspital Aarau, TUM: Klinikum rechts der Isar of the Technical University of Munich, USZ: University Hospital of Zurich) in three different countries. All patients received surgical resection of at least one BM. Pre-operative MRIs were analyzed. For further analysis, four sequences were selected, which are included in the currently recommended brain tumor MRI protocol [16]. These include a T1-weighted sequence without (T1) and with contrast enhancement (T1-CE), as well as a T2-weighted sequence (T2) and a T2 fluid-attenuated inversion recovery (T2-FLAIR) sequence. If the T1-CE sequence was missing, the patient was excluded. This is because the missing information about contrast enhancement cannot be reliably synthesized from the remaining sequences. The patient was also excluded if more than one sequence was missing. Of the 453 patients, 348 (77%) had the required minimum number of 3 sequences in sufficient quality available and were included.

MRI acquisition parameters are listed in Supplementary Table 1. In total, 28 different MRI scanner models were used to acquire the imaging data. The scanners are listed in Supplementary Table 2.

The patients were divided into a training cohort consisting of 260 patients from two centers and a multicentric external test cohort consisting of 88 patients from the remaining four centers.

Data preprocessing

The data was preprocessed using BraTS-Toolkit [17] implementing rigid registrations with niftyreg [18]. First, the MRI sequences were rigidly co-registered and transformed into the T1-CE space. To anonymize the patients, a brain mask was created by skull stripping the T1-CE with HD-BET [19]. Subsequently, the mask as well as all volumes in the T1-CE space were transformed to BraTS space using the SRI-24 atlas [20] and the mask was applied to all available sequences. The SRI-24 atlas was created using 24 adult control subjects and therefore represents normal brain anatomy. This yielded co-registered, skull-stripped sequences in a 1-millimeter isotropic resolution in BraTS space.

In total, 140 missing sequences were synthesized using a generative adversarial network (GAN) [21], originally developed for synthesizing missing sequences from glioma imaging. Therefore, the existing three sequences were fed to the GAN to generate the fourth missing sequence. In the training and test cohorts, one sequence was missing in 127 (49%) and 13 (15%) patients, respectively. The synthetic sequences successfully passed visual inspection.

Annotation

We manually segmented two classes using the open-source software 3D Slicer (Version 4.13.0, stable release, https://www.slicer.org/) [22]. First, the metastasis itself comprising the contrast-enhancing metastasis and necrosis, and second, the surrounding T2-FLAIR hyperintense edematous region were segmented. The images were annotated by a doctoral student (JAB) after undergoing extensive training by an expert radiation oncologist (JCP). To ensure a high segmentation quality, the segmentations of the test cohort were additionally checked and manually adapted by a board-certified trained radiation oncologist (JCP, 6 years of experience).

Intra-observer and inter-observer reliability

For comparison with real-life results, we randomly selected ten cases. For the computation of intra-observer reliability, these cases were segmented a second time by JAB. In addition, to calculate inter-observer reliability, a resident radiation oncologist (LE) seg-

mented the same ten cases. Annotation similarity was assessed using the Dice similarity coefficient (DSC).

Neural network training

A basic *U-Net* architecture inspired by Falk et al [23] implemented via *MONAI* [24] was used for all training runs. This implementation represents a standard U-Net architecture with an encoder and decoder connected by skip connections. However, unlike the original U-Net suggested by Ronneberger [25], our model features 3D convolutions.

The 3D network features four input channels, a dropout of 0.1 and employs *mish* as an activation function [26]. Apart from that, we used the default parameters of the implementation [27]. We selected the last checkpoint after 500 epochs of training with *AdamW optimizer* using an initial learning rate of 10^{-3} [28].

We trained with a batch size of six on three random cuboid-shaped samples with 192 × 192 × 32 dimensions per batch element. The input images were normalized on a channel basis with a percentile-based normalization applying to the 0.5 and the 99.5 percentile. This resulted in voxel intensities between 0.0 and 1.0. Inspired by BraTS segmentation networks [29], the networks were trained on two label channels. The first channel united both annotations, meaning the enhancing metastasis plus necrosis, as well as the edema. The second channel consisted only of the enhancing metastasis plus necrosis label. To obtain binary segmentations, the network outputs were thresholded at 0.5. The edema label was computed by subtracting the second from the first output channel. This way we ensured gapless segmentations.

An equally weighted SOFT DICE + Binary Cross Entropy (BCE) loss also used by Isensee et al. served as loss function for our training runs [30].

We compared three different training conditions: *none, basic, and extensive*, where we varied the amount of training augmentations. The *basic* condition implemented spatial flips, Gaussian noise, and random affine transformations, while the *none* condition omitted augmentation completely. In contrast, the *extensive* condition featured random elastic transformations and employed batch generators [31] for standard imaging augmentations such as gamma, brightness, contrast variations, and blurring. Furthermore, MRI-specific imaging artifacts such as spikes, ghosting, motion, and bias fields were simulated with TorchIO [32].

Additionally, for all three conditions, we explored the added value of introducing test time augmentations (TTA), which improved results in similar tasks, such as glioma segmentation [33]. We integrated TTA in the form of spatial flips and subtle Gaussian noise to create 13 variations of the original input data.

Our experiments were conducted on a workstation equipped with an Intel 9940X CPU in combination with two NVIDIA RTX 8000 GPUs using CUDA version 11.3 in conjunction with Pytorch version 1.11.0 and MONAI version 0.8.1. The final model can be accessed under this link: https://github.com/neuronflow/AURORA.

Metrics

All volumetric segmentation metrics were computed with *pymia* [34]. The connected component analysis for counting metastases was implemented with connected-components-3d [35]. This approach distinguishes instances of metastasis based on spatial proximity. Consequently, neighboring voxels were counted as coherent structures. To compute the individual metastasis detection metrics F1-score (F1), lesion sensitivity (LS), and lesion precision (LP) we use a proven pipeline from Pan et al [36]. A lesion was determined as correctly detected if there was any overlap between the manual segmentation and the output of our neural network. To assess volumetric segmentation performance, the DSC, surface DSC and Hausdorff distance are reported [37]. If not otherwise specified, the segmentation metrics were derived from all segmented BMs. To determine the influence of metastasis size on the volumetric DSC we calculated the Pearson correlation coefficient.

Results

Patient demographics and the number and size of BMs were similar between both cohorts (Table 1). In both cohorts, around one-quarter of patients had multiple metastases: In the training cohort, 78 patients (30%) had up to five BMs. In the test cohort, up to six BMs were found in 21 patients (24%). We segmented 364 and 121 BMs in the training and test cohort, respectively.

Table 1
Cohort demographics:

	Training-Cohort			Test-Cohort				
Characteristic	**Overall, N = 260**[a]	TUM, N = 170[a]	USZ, N = 90[a]	**Overall, N = 88**[a]	FD, N = 6[a]	FFM, N = 13[a]	HD, N = 44[a]	KSA, N = 25[a]
Sex								
F	**127 (49%)**	85 (50%)	42 (47%)	**48 (55%)**	2 (33%)	8 (62%)	25 (57%)	13 (52%)
M	**133 (51%)**	85 (50%)	48 (53%)	**40 (45%)**	4 (67%)	5 (38%)	19 (43%)	12 (48%)
Age at RT start	**62 (53, 71)**	62 (53, 71)	62 (54, 69)	**62 (54, 67)**	62 (57, 64)	59 (53, 66)	61 (54, 65)	63 (54, 70)
Primary Diagnosis								
NSCLC	**94 (36%)**	38 (22%)	56 (62%)	**42 (48%)**	4 (67%)	7 (54%)	19 (43%)	12 (48%)
SCLC	**1 (0.4%)**	0 (0%)	1 (1.1%)	**1 (1.1%)**	0 (0%)	0 (0%)	0 (0%)	1 (4.0%)
Melanoma	**47 (18%)**	24 (14%)	23 (26%)	**9 (10%)**	1 (17%)	1 (7.7%)	2 (4.5%)	5 (20%)
RCC	**11 (4.2%)**	9 (5.3%)	2 (2.2%)	**6 (6.8%)**	0 (0%)	1 (7.7%)	3 (6.8%)	2 (8.0%)
Breast	**35 (13%)**	34 (20%)	1 (1.1%)	**15 (17%)**	0 (0%)	4 (31%)	9 (20%)	2 (8.0%)
GI	**26 (10%)**	26 (15%)	0 (0%)	**7 (8.0%)**	0 (0%)	0 (0%)	5 (11%)	2 (8.0%)
Other	**46 (18%)**	39 (23%)	7 (7.8%)	**8 (9.1%)**	1 (17%)	0 (0%)	6 (14%)	1 (4.0%)
Number of brain metastases	**1.40 ± 0.72**	1.33 ± 0.69	1.53 ± 0.77	**1.38 ± 0.88**	1.00 ± 0.00	2.00 ± 1.63	1.25 ± 0.61	1.36 ± 0.70
Total brain tumor burden (ml)	**11 (5, 21)**	10 (5, 20)	12 (7, 23)	**13 (5, 23)**	32 (14, 47)	17 (8, 23)	9 (4, 15)	14 (6, 33)

The median age in both cohorts was similar (p = 0.3). We differentiated between seven different histologies: non-small cell lung carcinoma (NSCLC), small-cell lung carcinoma (SCLC), melanoma, renal cell carcinoma (RCC), breast cancer, gastrointestinal cancer (GI), and others. All seven histologies were present in the training as well as the test cohort. The number of BMs as well as the total brain tumor burden was similar in both cohorts (p = 0.4 and 0.7, Pearson's Chi-squared test and Wilcoxon rank sum test).
Abbreviations:
FD: General Hospital Fulda, FFM: Saphir Radiochirurgie/ University Hospital Frankfurt, HD: Heidelberg University Hospital, KSA: Kantonsspital Aarau, TUM: Klinikum rechts der Isar Technical University Munich, USZ: University Hospital of Zurich.
[a] n (%); Median (IQR); Mean ± SD.

There was a trend for a significantly different histology distribution ($p = 0.08$, two-sided Fisher's Exact test).

Fig. 1 shows an exemplary segmentation created by our neural network.

Table 2, Figs. 2 and 3 and Supplementary Table 3 summarize the results of our model evaluation. Regardless of the level of augmentation, the use of TTA did not significantly impact the volumetric segmentation or individual metastasis detection results as seen in Supplementary Table 3. For further analysis of the influence of augmentation during training, we only compared the networks tested with TTA.

Using any amount of augmentation during training significantly improved the volumetric results: the mean DSC for the metastasis label increased from 0.83 achieved by the model trained without augmentation to 0.92, regardless of which level of augmentation was used ($p < 0.001$, Wilcoxon rank sum test). There was no significant difference in DSC or Surface-Dice between the models trained with basic and extensive augmentation ($p > 0.9$). The DSC for the edema label increased similarly by using augmentation during training. While the network without augmentation reached a mean DSC of 0.87, the networks with augmentation both reached a mean DSC of 0.91 for the edema label.

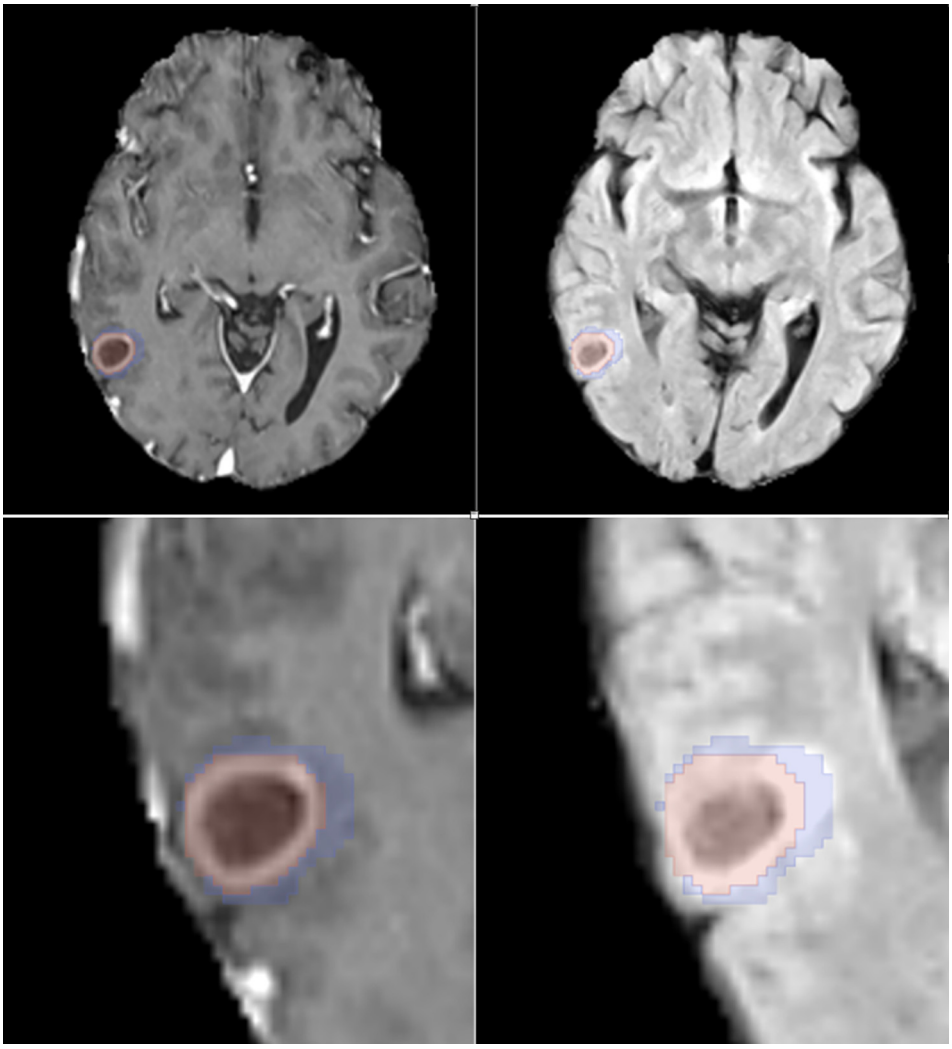

Fig. 1. Example for automatic segmentation by our proposed neural network: Example of a patient with 2 metastases in total (only one is shown here, volume: 1.9 cm^3). A full view as well as a zoomed in view of the T1-CE on the left and the T2-FLAIR on the right is shown. The segmentation of the metastasis in red and the edema in blue was created by our proposed neural network.

Table 2
Volumetric and instance-based segmentation performance:

Augmentation	Label	DSC	Surface-Dice	HD	F1	LS	LP
none	metastasis	0.827 ± 0.231	0.817 ± 0.267	NA	0.878 ± 0.224	0.895 ± 0.232	0.938 ± 0.189
none	edema	0.869 ± 0.176	0.871 ± 0.153	46.0 ± 40.0			
basic	metastasis	0.916 ± 0.079	0.907 ± 0.150	11.8 ± 21.0	0.928 ± 0.161	0.980 ± 0.081	0.919 ± 0.211
basic	edema	0.909 ± 0.094	0.913 ± 0.111	17.7 ± 21.9			
extensive	metastasis	0.922 ± 0.073	0.922 ± 0.140	25.0 ± 33.1	0.863 ± 0.200	0.975 ± 0.097	0.827 ± 0.266
extensive	edema	0.910 ± 0.083	0.913 ± 0.099	25.8 ± 27.7			

We report the mean and standard deviation of Dice similarity coefficient (DSC), Surface-Dice, Hausdorff distance (HD), F1-Score (F1), lesion sensitivity (LS) and lesion precision (LP) for variations of augmentation on each segmented label. The best overall performance was seen in the model trained with basic augmentation.

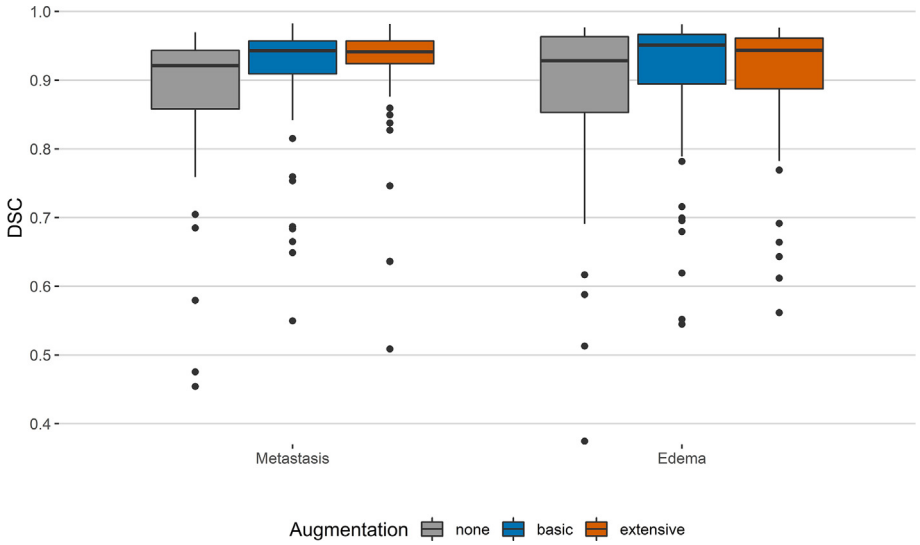

Fig. 2. Performance and augmentation: Using any amount of augmentation during training significantly improved volumetric segmentation performance. There was no significant difference in mean Dice similarity coefficient (DSC) between models trained with basic or extensive augmentation. The model trained with extensive augmentation seemed to achieve a lower interquartile range.

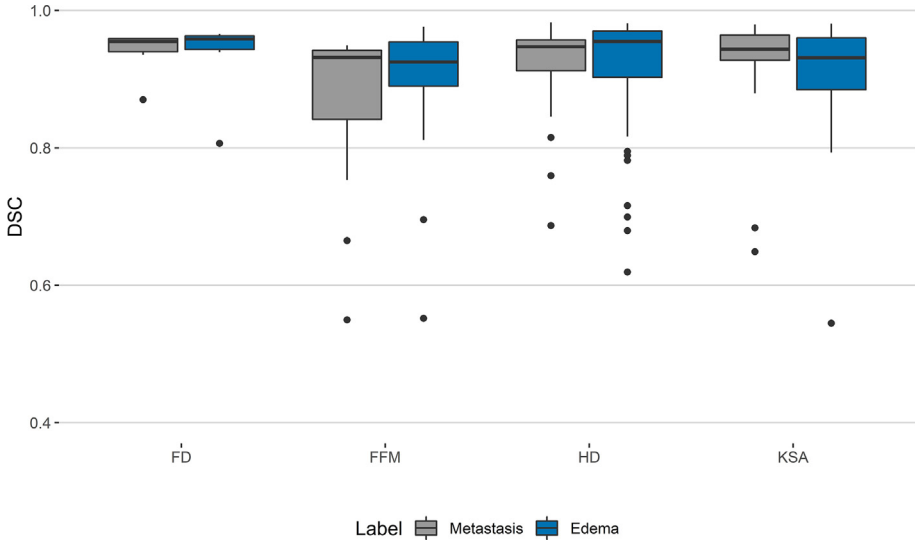

Fig. 3. Performance across the four different centers of the test set: Our proposed model provided a stable performance regardless of the center as shown by the consistently high Dice similarity coefficient (DSC) across all four locations included in our external test cohort. There was no statistically significant relationship between the center and DSC ($p = 0.3$, Kruskal–Wallis test).

Analysis of the individual metastasis detection performance showed that the basic model was able to outperform the extensive model in mean lesion precision and mean F1-Score with significantly better results of 0.92 and 0.93 compared to 0.83 and 0.86 ($p = 0.007$ and 0.013). Both models reached a mean lesion sensitivity of 0.98 ($p > 0.9$). The proposed model was able to generalize, as indicated by consistently high performance across different centers of our test set (see Fig. 3). Based on these metrics, we found that for our test cohort, the model trained with basic augmentation and tested with TTA performed the best.

Comparing the volumetric segmentations of our two raters resulted in a mean intra-observer DSC of 0.95 and a mean inter-observer DSC of 0.94.

In our test set, 13 patients had only three sequences present. In nine cases, the T2 sequence and in four cases, the T2-FLAIR sequence was missing. The mean volumetric DSC in patients with one sequence missing was 0.89 compared to 0.92 in patients with all four sequences present ($p = 0.082$).

To simulate an exemplary use in everyday clinical practice, we manually measured the individual metastasis-wise DSC and size of all segmented BMs (true positives) while leaving out false positive lesions. In total, 115 of 121 BMs with a median size of 7.29 cm^3 were found. The mean metastasis-wise DSC was 0.89. Fig. 4 shows the relationship between metastasis size and DSC. The segmentation performance was independent of metastasis size with a Pearson correlation coefficient of 0.07.

Discussion

We developed a neural network-based segmentation algorithm for BMs and their surrounding edema. The neural networks were evaluated in a multicentric and international external test cohort consisting of four centers. The model trained with basic augmentation and tested with TTA showed the best performance with a mean volumetric DSC of 0.92 for the metastasis label and a mean

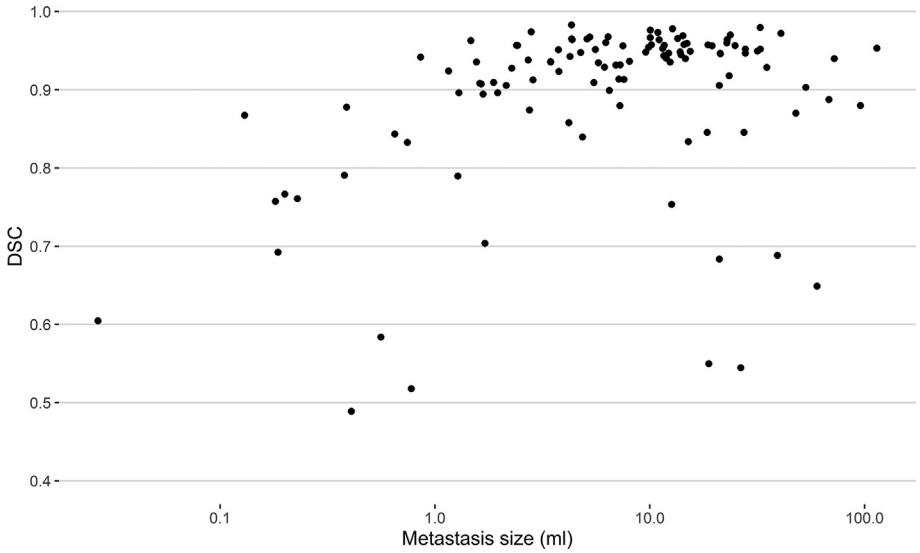

Fig. 4. Segmentation performance depending on the metastasis size: We report the individual metastasis-wise Dice similarity coefficient (DSC) and brain metastasis (BM) size of the metastases segmented by our proposed model. The volumetric performance was independent of BM size with a Pearson correlation coefficient of 0.07. The increased variation in small BMs can be attributed to the high proportion of edge voxels.

F1-Score of 0.93. The model is thus able to generalize well in a multicentric test cohort, an important prerequisite for clinical deployment. Also, the performance was independent of metastasis size. We furthermore showed that synthesizing missing sequences with a GAN has only a small, non-significant impact on segmentation performance, further advocating the broad availability of our model in clinical practice, where missing (or motion-corrupted) sequences are common. The performance of our proposed model was comparable to that achieved within the intra- and inter-observer comparisons. Therefore, the quality of segmentation was equivalent to that achieved in everyday clinical practice.

As the segmentation of a single BM takes less than a minute using recent consumer-grade hardware, automated segmentation of BMs with our proposed network can lead to a considerable acceleration of processes in the clinical routine.

Furthermore, exact segmentation is not only needed for SRT of small BMs, but can also be useful for subsequent automatic analysis of metastases of all sizes using radiomics analysis. Through the standardized segmentation process, the reproducibility of radiomic features may be improved.

Multiple other publications have reported deep learning-based algorithms for BM segmentation [38–44]. Most of them were monocentric studies without external testing. They reported DSC values between 0.55 and 0.85. To assess the true value of detection and segmentation performance and to prove generalizability of the proposed models external validation is necessary. So far, one recent study by Pflüger and colleagues developed a BMs segmentation tool and externally validated it in a monocentric test cohort of 30 patients from a different hospital, which is, however, part of the same university [45]. The test set was limited to patients with lung cancer. Their algorithm achieved an overall (case-specific) median DSC of 0.84, which is slightly inferior to our testing result (median DSC 0.94, mean DSC 0.92). With an individual metastasis-wise DSC of 0.79 however, the segmentation performance was even lower compared with our testing results (median metastasis-wise DSC 0.94, mean metastasis-wise DSC 0.89). The authors demonstrated a strong influence of volume on segmentation performance. Their datasets had significantly smaller mean volumes with a mean volume of 1.24 cm^3 (ours: 7.3 cm^3). The large influence of a few voxels on segmentation performance in small metastases may explain the lower individual metastasis-wise DSC in the study by Pflüger et al. [45] and the large differences between mean and median DSC values observed in our study. In terms of metastasis detection performance, sensitivity (median 0.85) and precision (median 0.76) were also lower compared with our results (median 1 and 1; mean 0.98 and 0.92, respectively). Our results remain in the range of the two largest internally validated studies with 934 and 1652 patients, which achieved mean volumetric DSCs between 0.81 and 0.84 and sensitivity values between 0.88 and 0.95 [46,47].

Whilst we focused on basic MRI sequences, other authors have developed their models based on Black Blood sequences [48,49] and achieved similar performance metrics, with a metastasis detection sensitivity of 0.92 to 0.93 and a DSC of 0.82 in internal validation. Others have focused on single oncological disease entities such as melanoma [50] or non-small-cell lung cancer [51]. They were not able to greatly improve the volumetric segmentation or metastasis detection metrics in internal validation (DSC: 0.72–0.75, metastasis detection sensitivity: 0.85–0.88) compared with similar studies focusing on multiple entities.

There are several limitations to this work: As main inclusion criteria of the multicentric patient cohort at least one metastasis was surgically resected. As a consequence, the GTVs of the metastases that were used for training were actually never used for treatment planning. Therefore, our median metastasis volume of 7.3 cm^3 was relatively large compared to the volume used in other studies of patients that received primary RT. We, thus, show a proof of concept on the segmentation on large BMs. While symptomatic large BMs are often surgically resected, asymptomatic large BMs can also be treated via SRT and are well represented by the BM size range included in the training cohort. In addition to the resected BMs, additional small metastases were present in around 25% of patients. Hence, our neural network showed a consistently high performance with no correlation to metastasis size as seen in Fig. 4. This may improve the transferability to BM that receive primary RT. The reference segmentations were all made by the same person. For this reason, the neural network has adopted the personal segmentation style of this person. To still ensure more objective segmentations, the segmentations of the test cohort were checked by a board-certified trained radiation oncologist. Another issue that can complicate adoption in clinical practice is the need for four MRI sequences in total. Sometimes not all of these sequences were acquired in clinical routine or were corrupted by

(motion) artifacts. While one missing sequence can be synthesized with only a minor non-significant loss in performance (DSC of 0.89 compared to 0.92), patients with more than one missing sequence, however, cannot be segmented with our intended and tested workflow. Comparable studies often used one or two sequences and were therefore less affected by such problems. However, Pflüger et al. showed that a model trained with only T1-CE and T2-FLAIR sequences performed slightly worse compared to the full configuration [45].

Despite these limitations, we created a neural network capable of exact automatic segmentation of brain metastases as well as the surrounding edema, which reached promising results in a multicentric, international external test cohort with a diverse set of MRI scanner types and cancer histologies. We published the network on Github to leverage its broad application. This network could be used as a basis for clinical gross tumor volume segmentations of brain metastases and also supports further computational analysis of brain metastases such as radiomic feature extraction [52].

Conflict of interest statement

The authors declare no potential conflicts of interest.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation - PE 3303/1-1 (JCP), WI 4936/4-1 (BW)). SMC received support from the "Young Talents in Clinical Research" Beginner's Grant from the Swiss Academy of Medical Sciences (SAMW) and the Bangerter-Rhyner Foundation. This funding was not directly related to this project.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radonc.2022.11.014.

References

[1] Mielczarek M, Kubica A, Szylberg M, Zielińska K, Przybył J, Sierzputowska A, et al. An update on the epidemiology, imaging and therapy of brain metastases. Nowotwory 2020;70:111–7. https://doi.org/10.5603/NJO.2020.0024.
[2] Soffietti R, Abacioglu U, Baumert B, Combs SE, Kinhult S, Kros JM, et al. Diagnosis and treatment of brain metastases from solid tumors: Guidelines from the European Association of neuro-oncology (EANO). Neuro Oncol 2017;19:162–74. https://doi.org/10.1093/neuonc/now241.
[3] Kocher M, Wittig A, Piroth MD, Treuer H, Seegenschmiedt H, Ruge M, et al. Stereotactic radiosurgery for treatment of brain metastases: A report of the DEGRO Working Group on Stereotactic Radiotherapy. Strahlenther Onkol 2014;190:521–32. https://doi.org/10.1007/s00066-014-0648-7.
[4] Rogers S, Baumert B, Blanck O, Böhmer D, Boström J, Engenhart-Cabillic R, et al. Stereotactic radiosurgery and radiotherapy for resected brain metastases: current pattern of care in the Radiosurgery and Stereotactic Radiotherapy Working Group of the German Association for Radiation Oncology (DEGRO). Strahlenther Onkol 2022. https://doi.org/10.1007/S00066-022-01991-6.
[5] Putz F, Mengling V, Perrin R, Masitho S, Weissmann T, Rösch J, et al. Magnetic resonance imaging for brain stereotactic radiotherapy : A review of requirements and pitfalls. Strahlenther Onkol 2020;196:444–56. https://doi.org/10.1007/S00066-020-01604-0.
[6] Abdel Razek AAK, Alksas A, Shehata M, AbdelKhalek A, Abdel Baky K, El-Baz A, et al. Clinical applications of artificial intelligence and radiomics in neuro-oncology imaging. Insights. Imaging 2021:12. https://doi.org/10.1186/s13244-021-01102-6.
[7] Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, et al. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge 2018.
[8] Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation 2016.
[9] Mouraviev A, Detsky J, Sahgal A, Ruschin M, Lee YK, Karam I, et al. Use of radiomics for the prediction of local control of brain metastases after stereotactic radiosurgery. Neuro Oncol 2020;22:797–805. https://doi.org/10.1093/neuonc/noaa007.
[10] Kniep HC, Madesta F, Schneider T, Hanning U, Schönfeld MH, Schön G, et al. Radiomics of brain MRI: Utility in prediction of metastatic tumor type. Radiology 2019;290:479–87. https://doi.org/10.1148/radiol.2018180946.
[11] Kocher M, Ruge MI, Galldiks N, Lohmann P. Applications of radiomics and machine learning for radiotherapy of malignant brain tumors. Strahlenther Onkol 2020;196:856–67. https://doi.org/10.1007/S00066-020-01626-8.
[12] Lohmann P, Bousabarah K, Hoevels M, Treuer H. Radiomics in radiation oncology-basics, methods, and limitations. Strahlenther Onkol 2020;196:848–55. https://doi.org/10.1007/S00066-020-01663-3.
[13] Priya S, Liu Y, Ward C, Le NH, Soni N, Pillenahalli Maheshwarappa R, et al. Machine learning based differentiation of glioblastoma from brain metastasis using MRI derived radiomics. Sci Rep 2021:11. https://doi.org/10.1038/s41598-021-90032-w.
[14] Cao R, Pang Z, Wang X, Du Z, Chen H, Liu J, et al. Radiomics evaluates the EGFR mutation status from the brain metastasis: a multi-center study. Phys Med Biol 2022:67. https://doi.org/10.1088/1361-6560/ac7192.
[15] Simmons A, Tofts PS, Barker GJ, Arridge SR. Sources of Intensity Nonuniformity in Spin Echo Images at 1.5 T. 1994.
[16] Villanueva-Meyer JE, Mabray MC, Cha S. Current clinical brain tumor imaging. Clin Neurosurg 2017;81:397–415. https://doi.org/10.1093/neuros/nyx103.
[17] Kofler F, Berger C, Waldmannstetter D, Lipkova J, Ezhov I, Tetteh G, et al. BraTS Toolkit: translating BraTS brain tumor segmentation algorithms into clinical and scientific practice. Front Neurosci 2020:14. https://doi.org/10.3389/fnins.2020.00125.
[18] Ourselin S, Roche A, Subsol G, Pennec X, Ayache N. Reconstructing a 3D structure from serial histological sections. n.d.
[19] Isensee F, Schell M, Pflueger I, Brugnara G, Bonekamp D, Neuberger U, et al. Automated brain extraction of multisequence MRI using artificial neural networks. Hum Brain Mapp 2019;40:4952–64. https://doi.org/10.1002/hbm.24750.
[20] Rohlfing T, Zahr NM, Sullivan EV, Pfefferbaum A. The SRI24 multichannel atlas of normal adult human brain structure. Hum Brain Mapp 2010;31:798–819. https://doi.org/10.1002/HBM.20906.
[21] Thomas MF, Kofler F, Grundl L, Finck T, Li H, Zimmer C, et al. Improving automated glioma segmentation in routine clinical use through artificial intelligence-based replacement of missing sequences with synthetic magnetic resonance imaging scans. Invest Radiol 2022;57:187–93. https://doi.org/10.1097/RLI.0000000000000828.
[22] Kikinis R, Pieper SD, Vosburgh KG. 3D Slicer: A platform for subject-specific image analysis, visualization, and clinical support. Intraoperat Imag Image-Guided Therapy 2014:277–89. https://doi.org/10.1007/978-1-4614-7657-3_19.
[23] Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, et al. U-Net: deep learning for cell counting, detection, and morphometry. Nat Methods 2019;16:67–70. https://doi.org/10.1038/s41592-018-0261-2.
[24] MONAI Consortium: MONAI: Medical open network for AI (3 2020). https://doi.org/10.5281/zenodo.4323058, https://github.com/Project-MONAI/MONAI; n.d.
[25] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation 2015.
[26] Misra D. Mish: A Self Regularized Non-Monotonic Activation Function. n.d.
[27] Implementation: https://docs.monai.io/en/stable/networks.html#basicunet; n.d.
[28] Loshchilov I, Hutter F. Decoupled Weight Decay Regularization; 2017.
[29] Kofler F, Ezhov I, Isensee F, Balsiger F, Berger C, Koerner M, et al. Are we using appropriate segmentation metrics? Identifying correlates of human expert perception for CNN training beyond rolling the DICE coefficient; 2021.
[30] Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 2021;18:203–11. https://doi.org/10.1038/s41592-020-01008-z.
[31] Isensee Fabian, Jäger Paul, Wasserthal Jakob, Zimmerer David, Petersen Jens, Kohl Simon, et al. batchgenerators - a python framework for data augmentation. https://doi.org/10.5281/zenodo.3632567 2020.
[32] Pérez-García F, Sparks R, TorchIO OS. A Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. Comput Methods Programs Biomed 2021;208. https://doi.org/10.1016/j.cmpb.2021.106236.
[33] Wang G, Li W, Ourselin S, Vercauteren T. Automatic Brain Tumor Segmentation using Convolutional Neural Networks with Test-Time Augmentation, 2018. https://doi.org/10.1007/978-3-030-11726-9_6.
[34] Jungo A, Scheidegger O, Reyes M, Balsiger F. pymia: A Python package for data handling and evaluation in deep learning-based medical image analysis. Comput Methods Programs Biomed 2021:198. https://doi.org/10.1016/j.cmpb.2020.105796.
[35] Silversmith W, Kemnitz N. 2020seung-lab/connected-components-3d. seung-lab. See https://github.com/seung-lab/connected-components-3d. n.d.
[36] Pan C, Schoppe O, Parra-Damas A, Cai R, Todorov MI, Gondi G, et al. Deep learning reveals cancer metastasis and therapeutic antibody targeting in the entire body. Cell 2019;179:1661–1676.e19. https://doi.org/10.1016/j.CELL.2019.11.013.
[37] Nikolov S, Blackwell S, Zverovitch A, Mendes R, Livne M, de Fauw J, et al. Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy. 2018.
[38] Yoo SK, Kim TH, Chun J, Choi BS, Kim H, Yang S, et al. Deep-learning-based automatic detection and segmentation of brain metastases with small volume for stereotactic ablative radiotherapy. Cancers (Basel) 2022;14:2555. https://doi.org/10.3390/cancers14102555.

[39] Liu Y, Stojadinovic S, Hrycushko B, Wardak Z, Lau S, Lu W, et al. A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. PLoS One 2017;12. https://doi.org/10.1371/journal.pone.0185844.

[40] Zhou Z, Sanders JW, Johnson JM, Gule-Monroe MK, Chen MM, Briere TM, et al. Computer-aided detection of brain metastases in T1-weighted MRI for stereotactic radiosurgery using deep learning single-shot detectors. Radiology 2020;295:407–15. https://doi.org/10.1148/radiol.2020191479.

[41] Zhang M, Young GS, Chen H, Li J, Qin L, McFaline-Figueroa JR, et al. Deep-learning detection of cancer metastases to the brain on MRI. J Magn Reson Imaging 2020;52:1227–36. https://doi.org/10.1002/jmri.27129.

[42] 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. J Magn Reson Imaging 2020;51:175–82. https://doi.org/10.1002/jmri.26766.

[43] Bousabarah K, Ruge M, Brand JS, Hoevels M, Rueß D, Borggrefe J, et al. Deep convolutional neural networks for automated segmentation of brain metastases trained on clinical data. Radiat Oncol 2020;15. https://doi.org/10.1186/s13014-020-01514-6.

[44] Rudie JD, Weiss DA, Colby JB, Rauschecker AM, Laguna B, Braunstein S, et al. Three-dimensional u-net convolutional neural network for detection and segmentation of intracranial metastases. Radiol Artif Intell 2021;3. https://doi.org/10.1148/ryai.2021200204.

[45] Pflüger I, Wald T, Isensee F, Schell M, Meredig H, Schlamp K, et al. Automated detection and quantification of brain metastases on clinical MRI data using artificial neural networks. n.d. https://doi.org/10.1093/noajnl/vdac138/6674032.

[46] Xue J, Wang B, Ming Y, Liu X, Jiang Z, Wang C, et al. Deep learning-based detection and segmentation-assisted management of brain metastases. Neuro Oncol 2020;22:505–14. https://doi.org/10.1093/neuonc/noz234.

[47] Zhou Z, Sanders JW, Johnson JM, Gule-Monroe M, Chen M, Briere TM, et al. MetNet: Computer-aided segmentation of brain metastases in post-contrast T1-weighted magnetic resonance imaging. Radiother Oncol 2020;153:189–96. https://doi.org/10.1016/j.radonc.2020.09.016.

[48] Kikuchi Y, Togao O, Kikuchi K, Momosaka D, Obara M, van Cauteren M, et al. A deep convolutional neural network-based automatic detection of brain metastases with and without blood vessel suppression. Eur Radiol 2022;32:2998–3005. https://doi.org/10.1007/s00330-021-08427-2.

[49] Won Park Y, Jun Y, Lee Y, Han K, An C, Soo Ahn S, et al. Robust performance of deep learning for automatic detection and segmentation of brain metastases using three-dimensional black-blood and three-dimensional gradient echo imaging, n.d. https://doi.org/10.1007/s00330-021-07783-3/Published.

[50] Pennig L, Shahzad R, Caldeira L, Lennartz S, Thiele F, Goertz L, et al. Automated detection and segmentation of brain metastases in malignant melanoma: Evaluation of a dedicated deep learning model. Am J Neuroradiol 2021;42:655–62. https://doi.org/10.3174/AJNR.A6982.

[51] Jünger ST, Hoyer UCI, Schaufler D, Laukamp KR, Goertz L, Thiele F, et al. Fully Automated MR detection and segmentation of brain metastases in non-small cell lung cancer using deep learning. J Magn Reson Imaging 2021;54:1608–22. https://doi.org/10.1002/jmri.27741.

[52] Peeken JC, Wiestler B, Combs SE. Image-guided radiooncology: the potential of radiomics in clinical application. Recent Results Cancer Res 2020;216:773–94. https://doi.org/10.1007/978-3-030-42618-7_24/COVER.