Automated Detection and Segmentation of Brain Metastases in Malignant Melanoma: Evaluation of a Dedicated Deep Learning Model

L. Pennig, R. Shahzad, L. Caldeira, S. Lennartz, F. Thiele, L. Goertz, D. Zopfs, A.-K. Meißner, G. Fürtjes, M. Perkuhn, C. Kabbasch, S. Grau, J. Borggrefe, and K.R. Laukamp

ABSTRACT

BACKGROUND AND PURPOSE: Malignant melanoma is an aggressive skin cancer in which brain metastases are common. Our aim was to establish and evaluate a deep learning model for fully automated detection and segmentation of brain metastases in patients with malignant melanoma using clinical routine MR imaging.

MATERIALS AND METHODS: Sixty-nine patients with melanoma with a total of 135 brain metastases at initial diagnosis and available multiparametric MR imaging datasets (T1-/T2-weighted, T1-weighted gadolinium contrast-enhanced, FLAIR) were included. A previously established deep learning model architecture (3D convolutional neural network; DeepMedic) simultaneously operating on the aforementioned MR images was trained on a cohort of 55 patients with 103 metastases using 5-fold cross-validation. The efficacy of the deep learning model was evaluated using an independent test set consisting of 14 patients with 32 metastases. Manual segmentations of metastases in a voxelwise manner (T1-weighted gadolinium contrast-enhanced imaging) performed by 2 radiologists in consensus served as the ground truth.

RESULTS: After training, the deep learning model detected 28 of 32 brain metastases (mean volume, 1.0 [SD, 2.4] cm^3) in the test cohort correctly (sensitivity of 88%), while false-positive findings of 0.71 per scan were observed. Compared with the ground truth, automated segmentations achieved a median Dice similarity coefficient of 0.75.

CONCLUSIONS: Deep learning–based automated detection and segmentation of brain metastases in malignant melanoma yields high detection and segmentation accuracy with false-positive findings of <1 per scan.

ABBREVIATIONS: CNN = convolutional neural network; DLM = deep learning model; GT = ground truth

Malignant melanoma is an aggressive skin cancer associated with high mortality and morbidity rates.[1,2] Brain metastases are common in malignant melanoma,[3,4] subsequently causing potential severe neurologic impairment and worsened outcome. Therefore, it is recommended that melanoma patients with an advanced stage undergo MR imaging of the head for screening purposes to detect metastases.[5-8]

Owing to an increased workload of radiologists, repetitive evaluation of MR imaging scans can be tiresome, hence bearing an inherent risk of missed diagnosis for subtle lesions, with satisfaction of search effects leading to decreased sensitivity for additional lesions.[9,10] Automatization of detection could serve as an adjunct tool for lesion preselection that can support image evaluation by radiologists and clinicians.[11,12] Furthermore, automated segmentations may be used as a parameter to evaluate therapy response in oncologic follow-up imaging.[13,14] Additionally, exact lesion determination and delineation of size are required for stereotactic radiosurgery.[15,16] In clinical routine, brain lesions have to be segmented manually by the radiosurgeon. This task proves to be time-consuming, in particular if multiple metastases are present. Furthermore, manual segmentation is potentially hampered by interreader variabilities with reduced reproducibility, hence resulting in inaccuracies of lesion delineation.[17,18] In this

context, accurate objective and automated segmentations of brain metastases would be highly beneficial.[17-19]

Recently, 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.[20-25] The models apply multiple processing layers that result in deep convolutional neural networks (CNNs). Training data are used to create complex feature hierarchies.[26-28] In general, a DLM includes different layers for convolution, pooling, and classification.[28] The required training data are supplied by manual segmentations, which usually serve as the segmentation criterion standard.[18,28,29]

Previous studies on brain metastases from different tumor entities have demonstrated promising results, reporting a sensitivity for automated deep learning–based detection of lesions of around 80% or higher.[17,30-32] However, the often reported relatively high number of false-positive findings questions their applicability in clinical routine.[17,30]

The purpose of this study was to develop and evaluate a DLM for automated detection and segmentation of brain metastases in patients with malignant melanoma using heterogeneous MR imaging data from multiple vendors and study centers.

MATERIALS AND METHODS

The local institutional review board (Ethikkommission, Medizinische Fakultät der Universität zu Köln) approved this retrospective, single-center study (reference No: 19–1208) and waived the requirement for written informed patient consent.

Patient Population

MR imaging of patients treated for malignant melanoma at our tertiary care university hospital between May 2013 and October 2019 was reviewed using our institutional image archiving system. Ninety-two patients could be identified by applying the following inclusion criteria: 1) MR imaging scans at primary diagnosis of brain metastases; 2) distinct therapy following diagnosis of brain metastases, eg, stereotactic radiosurgery, resection, extended biopsy, targeted chemotherapy; and 3) a complete MR image set, being defined as T1-/T2-weighted, T1-weighted gadolinium contrast-enhanced imaging, and T2-weighted FLAIR. Patients with unclear lesions in which follow-up imaging could not confirm metastatic spread to the brain were not included ($n = 11$).

We applied the following exclusion criteria: 1) the presence of a second malignant tumor ($n = 3$); 2) large intracranial extralesional bleeding (the definition of extralesional bleeding was based on reviewing prior/follow-up imaging, $n = 3$); 3) acute ischemic stroke ($n = 1$) impeding delineation of brain metastases; 4) severe MR imaging artifacts impairing image quality ($n = 3$); and 5) insufficient contrast media application ($n = 2$).

The 69 enrolled patients were randomly split into a training cohort consisting of 55 patients and a test cohort with 14 patients, ensuring that there was no overlap of data between the 2 cohorts. The training cohort was used for training and performing 5-fold cross-validation of the DLM. On the contrary, the test cohort was used for independent testing of the DLM. MR images were anonymized and exported to IntelliSpace Discovery (ISD, Version 3.0; Philips Healthcare).

Image Acquisition

MR images were acquired on different scanners from our ($n = 48$) and referring institutions ($n = 21$), ranging between 1T and 3T. Detailed MR imaging parameters are given in the Online Supplemental Data. The imaging protocol of our institution included intravenous administration of gadolinium (gadoterate meglumine, Dotarem; Guerbet; 0.5 mmol/mL, 1 mL = 279.3 mg of gadoteric acid = 78.6 mg of gadolinium) with a concentration of 0.1 mmol/kg of body weight. Contrast medium application at referring institutions was not standardized.

Ground Truth

To establish the reference standard and lesion count, 2 radiologists (each with at least 3 years of experience in neuro-oncologic imaging) confirmed all metastases. A board-certified neuroradiologist with 13 years of experience in neuro-oncologic imaging was consulted when uncertainties occurred. They conducted a review of the original radiology report and double-reviewed the included MR imaging scans as well as prior/follow-up imaging.

By assessing unenhanced T1- and T2-weighted, T1-weighted gadolinium contrast-enhanced imaging, and FLAIR images on ISD, the 2 radiologists performed manual segmentations of lesions on T1-weighted gadolinium contrast-enhanced imaging in a voxelwise manner in consensus, which served as the ground truth (GT). First, initial segmentations of the metastases were performed by 1 radiologist and then presented to/discussed with the second radiologist to define the final segmentations of the lesions in consensus.

Deep Learning Model

Before passing the sequences (T1/T2-weighted, T1-weighted gadolinium contrast-enhanced imaging, and FLAIR) to the DLM, we performed preprocessing of data, which included the following: bias field correction of all 4 sequences, coregistration of T1/T2-weighted and FLAIR to T1-weighted gadolinium contrast-enhanced imaging, skull-stripping, resampling to an isotropic resolution of $1 \times 1 \times 1 \text{ mm}^3$, and z score normalization.[24]

In this study, a 3D CNN based on DeepMedic (Biomedical Image Analysis Group, Department of Computing, Imperial College London) was used. In recent studies, the DeepMedic architecture has demonstrated encouraging results for detection and segmentation of different brain tumors.[24,33]

The network consists of a deep 3D CNN architecture with 2 identical pathways. 3D image patches provide input to the 2 pathways. For the first pathway, original isotropic patches are used. For the second pathway, the patches are down-sampled to a third of their original size. This approach helps to capture higher contextual information. The deep CNN model comprises 11 layers with size 3^3 kernels. The model consists of residual connections for layers 4, 6, and 8. Each layer is followed by batch normalization and a parametric rectified linear unit as the activation function. Layers 9 and 10 are fully connected. The last prediction layer has a kernel size of 1^3 and uses sigmoid as the activation function.[34]

For training of the DLM, multichannel GT 3D image patches with a size of 25^3 were fed to the 3D CNN. These image patches

Detection and segmentation accuracy on the independent test cohort

	Majority Voting Scheme	First Fold	Second Fold	Third Fold	Fourth Fold	Fifth Fold
Detection						
Sensitivity	88%	88%	88%	81%	91%	91%
Precision	74%	49%	35%	40%	41%	51%
F1 score	0.80	0.63	0.50	0.53	0.57	0.65
Missed	4	4	4	6	3	3
FPs/scan	0.71	2.1	3.8	2.9	2.9	2.0
Segmentation						
Dice coefficient	0.75	0.78	0.73	0.73	0.76	0.78

Note:—Missed indicates missed brain metastases in the test cohort; FPs/scan, false-positive lesion findings per patient; Dice coefficient, similarity score reported as median; Majority Voting Scheme, fusion of the 5 deep learning models from the 5-fold cross-validation; First Fold, first deep learning model from the 5-fold cross-validation.

were extracted with a distribution of 50% between background and metastases, ensuring class balance. To increase the number of training samples, image augmentation was used by randomly flipping the image patches along their axes. The Dice similarity coefficient was used as the loss function, and root mean square propagation, as the optimizer. An adaptive learning rate schedule was used, in which the initial learning rate was halved every time the accuracy did not improve for >3 epochs. The training batch size was set to 10, and the number of training epochs was set to 35.

Training was performed on the training set ($n = 55$) using a 5-fold cross-validation approach using an 80%–20% training-validation split without overlapping data, which resulted in 5 trained models.

During inference on the independent test set ($n = 14$), 3D image patches of 45^3 in size are extracted. Larger patch sizes reduced the time spent during inference. The 5 individual models from the 5-fold cross-validation training were applied to the independent test data. The segmentation results from each of the 5 DLMs were fused using a majority voting scheme to reduce false lesion detections.[35] By default, automatically detected lesions of <0.003 cm^3 (2 voxels on average) during inference of both the training and test sets were regarded as image noise and discarded. This threshold was based on the resolution of T1-weighted gadolinium contrast-enhanced sequences (in which a volume of 0.003 cm^3 is approximately 2 voxels) and is determined by referring to the smallest annotated metastases on training (0.0035 cm^3) and test (0.0041 cm^3) sets. Due to limitation of scan resolution, lesions smaller than this volume cannot be accurately detected or segmented by image readers.

Including image preprocessing, the average time needed to run a complete pipeline on a dataset is about 8 minutes: <1 second for bias field correction, 7 minutes for coregistration and skull-stripping, <1 minute for image standardization, and around 10 seconds to run the inference (using a Tesla-P100 GPU card (NVIDIA).

Statistical Analysis
Statistical analysis was performed using JMP Software (Release 12; SAS Institute). Tumor volumes are displayed as mean [SD], and Dice similarity coefficients are reported as median with a 10–90 percentile range. The Wilcoxon rank sum test was applied for determination of a statistical difference with statistical significance being set to $P < .05$. To determine the detection accuracy of the metastases, we computed sensitivity (recall), precision (positive predictive value), and F1 score. Because no scans

without metastases were included, a true specificity could not be determined; hence, precision was calculated.

To evaluate the segmentation accuracy of the DLM on a voxelwise basis, we compared automatically obtained segmentations with the GT annotations with overlap measures between the segmentations being computed using the Dice similarity coefficient.[23,24,35] For quantitative volumetric measurements, the Pearson correlation coefficient (r) was calculated.

RESULTS
Patient Characteristics
The 69 enrolled patients (mean age, 61.5 [SD, 13.4] years; 30 women) had a total of 135 brain metastases on MR imaging, of which 45 patients presented a single brain metastasis. Most ($n = 48$) patients received stereotactic radiosurgery using the CyberKnife System (Accuray). The Online Supplemental Data provide detailed patient information, including distribution of brain metastases and treatment received.

Evaluation of the DLM on the Training Cohort
In the training cohort, 103 metastases with a mean volume of 2.6 [SD, 8.1] cm^3 were identified as the GT.

Using 5-fold cross-validation, the DLM achieved a sensitivity of 87% with a median corresponding Dice similarity coefficient of 0.75 (range, 0.19–0.93). The DLM missed 13 metastases, yielding a mean volume of 0.06 [SD, 0.1] cm^3. On average, the DLM produced 4 false-positive lesions per scan with a mean volume of 0.05 [SD, 0.17] cm^3.

Evaluation of the DLM on the Independent Test Cohort
In the test cohort, 32 metastases with a mean volume of 1.0 [SD, 2.4] cm^3 were identified as the GT, being smaller than in the training cohort, though without a significant difference ($P > .05$). The 5 DLMs from the 5-fold cross-validation as well as their fusion using the majority voting scheme were tested on the independent test cohort. Detailed results of the DLM on the test set are given in the Table.

After we applied the majority voting scheme, the fused DLM detected 28 of 32 brain metastases correctly and missed 4, corresponding to a sensitivity of 88% and an F1 score of 0.80 (Figs 1 and 2 and Online Supplemental Data depict examples of true-positive findings of the DLM). Missed brain metastases were small and yielded a volume between 0.004 and 0.16 cm^3 (Fig 3 provides the metastases, which were missed by the DLM). Compared with manual segmentations, the fused DLM provided a median Dice
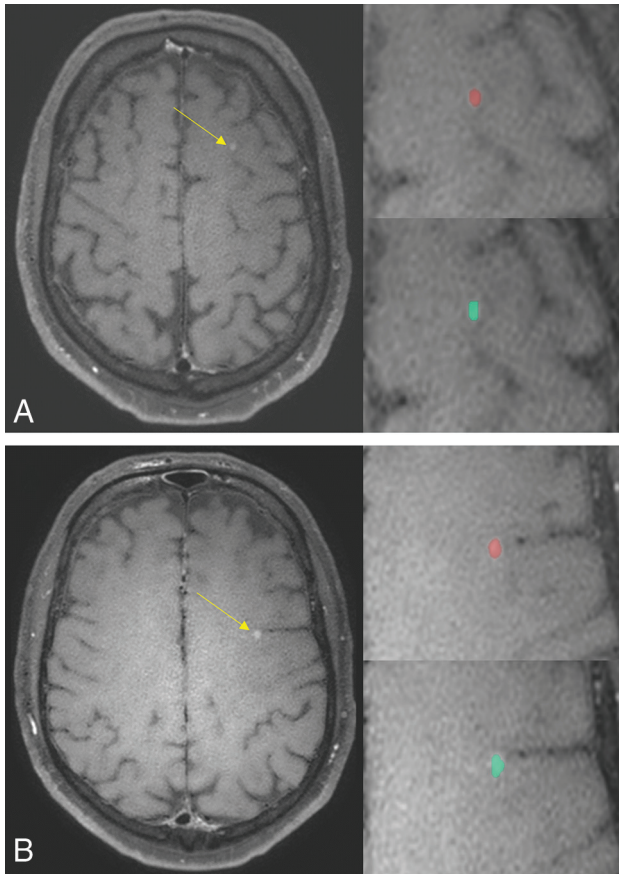
FIG 1. A 55-year-old male patient with malignant melanoma. The DLM (turquoise) detects and segments the metastases of the left frontal lobe (*yellow arrows, A* and *B*) comparable to the manual segmentations (red).

similarity coefficient of 0.75 (range, 0.09–0.93) and a volumetric correlation of $r = 0.97$. The Online Supplemental Data display the relationship between obtained Dice similarity coefficients and the volume of the metastases.

Figure 4A depicts a histogram demonstrating the volume of metastases in the training and test groups as well as the size of missed metastases and false-positive lesions. Figure 4B shows a boxplot comparing Dice similarity coefficients, false-positives, and false-negatives for the 5 different DLMs using the 5-fold cross-validation and the combined DLMs, applying the majority voting scheme. Figure 4C provides the volumetric correlation between automated detection of metastases using the fused DLM and the GT.

In addition, the fusion of all 5 DLMs reduced the number of false-positive lesions to 0.71 per scan (compared with 3.8 of the second fold, as seen in the Table) and increased the precision (74%). Examples of false-positive detections by the DLM are provided in Figs 3 and 4D, which show a free-response receiver operating characteristic curve displaying the relationship between the lesion-detection sensitivity and the average number of false-positive lesions per scan.

DISCUSSION
In this study, we developed and trained a dedicated DLM for automated detection and segmentation of brain metastases in
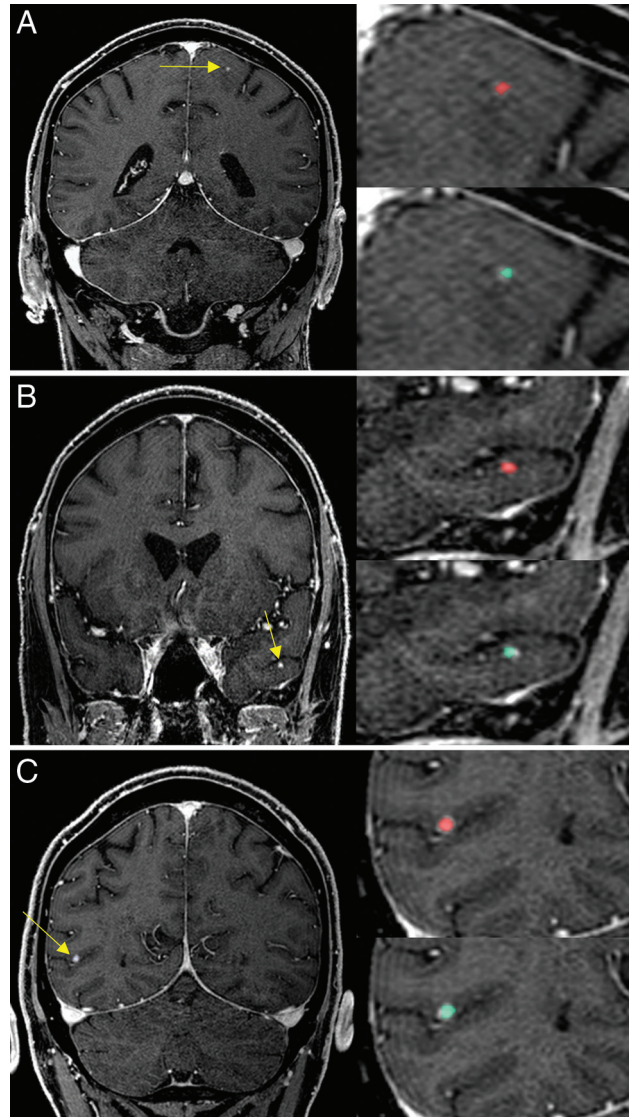
FIG 2. A 67-year-old male patient with malignant melanoma. The DLM (turquoise) detects the metastases (*yellow arrows*) of the left frontal lobe (*A*), the left temporal lobe (*B*), and the right parietal lobe (*C*) accurately and provides manual segmentations (red) comparable to segmentation performance.

malignant melanoma and evaluated its performance on an independent test set. On heterogeneous scanner data, the proposed DLM provided a detection rate of 88%, while producing an error of <1 false-positive lesion per scan. Furthermore, a high overlap between automated and manual segmentations was observed (Dice similarity coefficient = 0.75).

Recent studies investigating automated detection of brain metastases have not focused on a certain underlying pathology and reported lesion sizes between 1.3 and 1.9 cm^3 (Bousabarah et al[32]) and 2.4 cm^3 (Charron et al[17]) for various primary tumors, which are comparable with the average tumor sizes in our training (2.6 [SD, 8.1] cm^3) and test cohorts (1.0 [SD, 2.4] cm^3). Despite the small lesion size, the DLM provided a high detection sensitivity (88%), similar to that in the aforementioned studies.[16,17,30-32] Compared with the GT, the DLM obtained a median Dice similarity coefficient of 0.75, which is in line with recent studies,
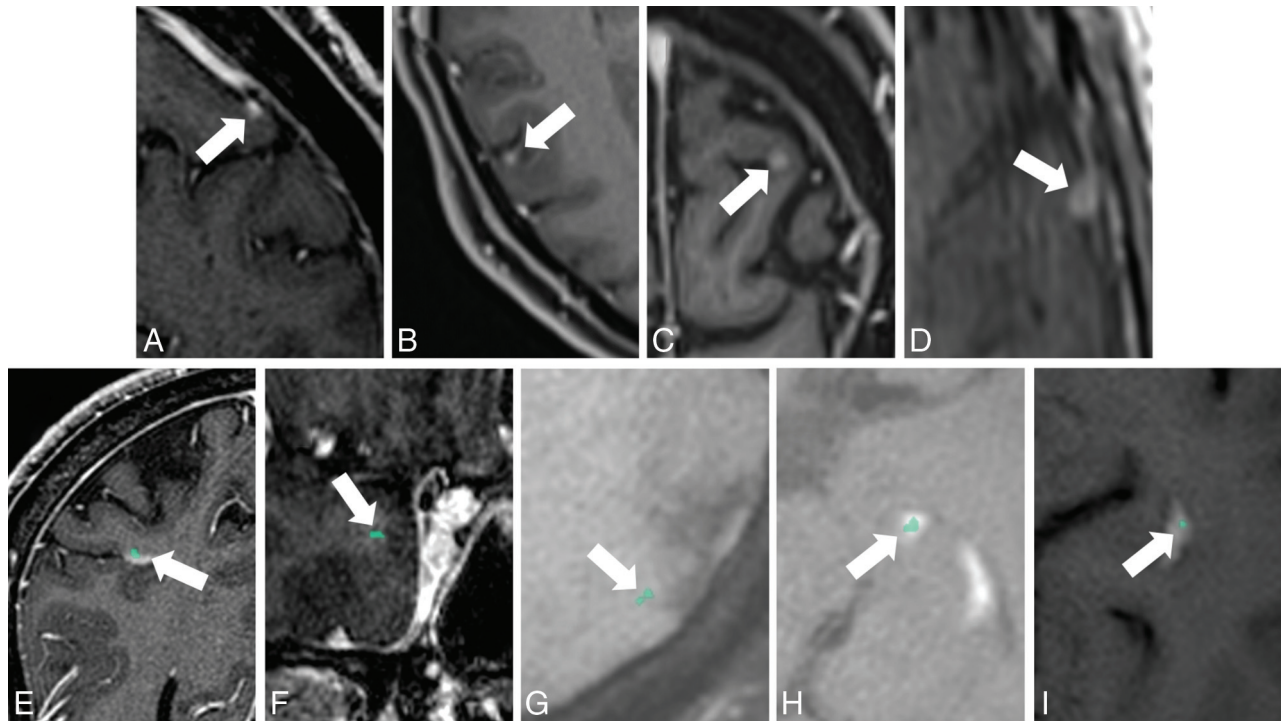
FIG 3. False-negative findings of the DLM (*A–D, white arrows*) as shown in a 67-year-old male patient (*A*, same patient as in Fig 2; metastasis volume: 0.004 cm³), a 56-year-old male patient (*B* and *C*, metastases volume: 0.008 and 0.01 cm³), and a 62-year-old male patient (*E*, metastasis volume: 0.016 cm³) with malignant melanoma. As demonstrated, the DLM missed small metastases. Examples of false-positive findings of the DLM (*E–I, white arrows*) as shown in a 50-year-old female patient (*E*), a 67-year-old male patient (*F*, same patient as in Fig 2), a 55-year-old male patient (*G* and *H*, same patient as in Fig 1), and a 62-year-old female patient (*I*) with malignant melanoma. False-positive findings (turquoise) were related to blood vessels (*E*, developmental venous anomaly), variations in brain tissue contrast (*F* and *G*), and the choroid plexus (*H* and *I*).

which reported Dice similarity coefficients between 0.67 and 0.79.[16,30] The high number of false-positive lesions poses a common drawback in automated detection of brain metastases, which have been reported to be around 7–8 per scan.[17,30] By combining 5 DLMs using a majority voting scheme, false-positive findings of <1 per patient were obtained in the present study, as could also recently be achieved by Bousabarah et al.[32]

Given the high risk of metastatic spread, screening examinations are warranted in patients with malignant melanoma and are suggested according to current guidelines.[5,7,8] For lung cancer, regular screening has also been proposed recently.[36] However, when diagnosed at an early stage in an asymptomatic patient, metastases are often small and more difficult to detect, even by experienced radiologists.[1,2,5,6] Despite the small size of the metastases in the test set, the trained DLM yielded a sensitivity of 88%. Of note, the metastases in the test set were smaller compared with those in the training cohort without reaching a statistical significance. In part, this difference could be explained by the higher number of patients treated by surgery in the training cohort (18.2% versus 14.0%), who usually present with larger metastases.[37]

Brain metastases screening examinations are increasing in number, making evaluation tiresome while bearing an inherent risk of missed diagnoses, in particular for subtle lesions.[9,38] In this context, our DLM can provide assistance for detection of brain metastases in malignant melanoma. Compared with a human reader, the DLM is not impaired by "satisfaction of search," which means that the physician may miss a second

metastasis when a first one has been found.[9,10,38] Additionally, automation of brain metastasis segmentation by a DLM could serve as an accurate mechanism of lesion preselection, in particular when the number of false-positive lesions is <1 per scan, as obtained by the DLM of the present study.[16,17,30,31] Automated segmentation may also provide assistance in evaluating treatment response during oncologic follow-up and may support radiologists in coping with an increased number of image readings, while maintaining high diagnostic accuracy.

Compared with manual segmentations, the proposed DLM achieved a high volumetric correlation despite the small size of the metastases. Automated segmentation of brain tumors such as metastases, being possible with the DLM of the current study, has several applications to potentially improve patient care. For instance, volumetric assessment proves to be a promising tool for quantification of tumor burden.[14,39,40] Furthermore, volumetric assessment has advantages over user-dependent conventional linear measurements because metastatic lesions are not entirely spherical.[18]

Stereotactic radiosurgery requires reliable and objective lesion segmentation.[15,16] Manual segmentation of multiple lesions proves to be time-consuming and is impeded by inter- and intra-reader variabilities. Next to increased efficiency, higher reproducibility of lesion delineation potentially boosts reliability of radiation therapy while improving patient outcome.[17]

Regarding automatic detection and segmentation of brain metastases, one must consider the following challenges: 1) multifocal lesion occurrence; 2) very small and subtle lesions; 3) more complex tumor structures when lesions enlarge
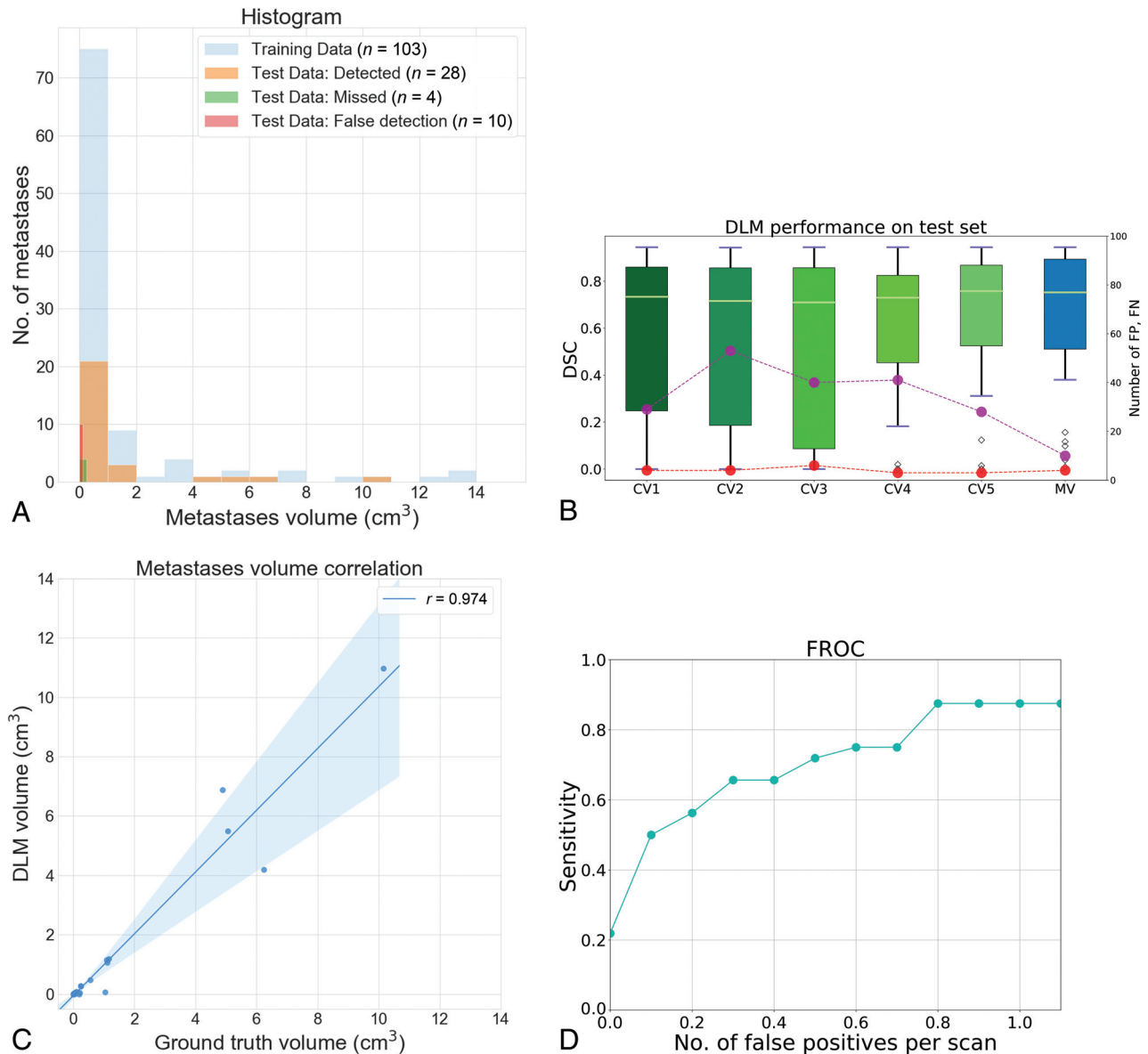
FIG 4. *A*, Histogram depicting the distribution of metastases volumes in the training and test cohorts. Furthermore, the volumes of missed metastases and false-positive findings in the test group are also depicted, all of which were small (mean missed metastases volume of 0.01 [SD, 0.005] cm^3; mean volume of false-positive lesions of 0.02 [SD, 0.02] cm^3). To better visualize the small volumes of false-positive and false-negative findings in the independent test set, we limited the x-axis to 15 cm^2. Hence, 5 metastases of the training data larger than this volume are not shown. *B*, Performance of the 5 different DLMs obtained using the 5-fold cross-validation training and the combined DLM using the majority voting scheme on the independent test cohort. *Magenta circles* represent the number of false-positives (FP) and *red circles* indicate the number of false-negatives (FN). DSC indicates the Dice similarity coefficient; CV1–5, the cross-validation folds; and MV, majority voting. *C*, Volume correlation of the metastases between the automatically segmented lesions and the ground truth on the independent test set on a lesion level. *D*, Free-response receiver operating characteristic (FROC) curve of the DLM on the independent test cohort.

(contrast-enhancing tumor, necrosis, bleeding, and edema); 4) variations in patient anatomy; and 5) heterogeneous imaging data due to varying vendors, MR imaging manufacturers, scanner generations, scan parameters, and unstandardized contrast media application.[16,17,25,28,30,34,41,42] In the present study, our DLM provides high detection accuracy on heterogeneous scanner data as reflected by a large number of scans from referring institutions and examinations performed over a wide range of field strengths.

The results of this study indicate that training of an already established deep-learning architecture initially used for other

tumor entities, ie, glioma and glioblastoma,[24,34] can be successfully applied to other brain tumors[16,43,44] but dedicated retraining is usually warranted.[16,32,33] Still, previous studies have also suggested that dedicated training might be omitted if tumor appearance is similar, although accuracy will/might be negatively impacted by the missing dedicated training.[23,44] Therefore, our DLM, though dedicated to patients with melanoma, might also be applied, for example, to metastases of different origins, which may nurture further investigations.

The following limitations need to be discussed. The study has typical drawbacks of a retrospective setting, not allowing

evaluation if detection and segmentation accuracies are sufficient for clinical needs. This drawback may be addressed in future studies with a focus on specified clinical necessities and tasks. Although almost one-third of included scans were acquired at referring institutions, the application of the DLM should be investigated in a true multicenter setting. Our relatively small number of patients, which resulted from focusing exclusively on malignant melanoma, needs to be considered. This is especially important regarding our test cohort, which consisted of 14 patients only. Future studies, preferably including more cases from differing institutions, are warranted to further validate our DLM. Only patients with melanoma were included, which potentially limits the transferability of our DLM to brain metastases of other primary tumors. In this context, future studies are needed. Because no posttreatment MR images were included, the performance of the DLM in this setting is unknown and requires future research.

The applied DLM operates on 4 MR images, ie, FLAIR, T1-/T2-weighted, and T1-weighted gadolinium contrast-enhanced images. Consequently, this feature limits the application of the DLM if one of these sequences is unavailable. Our study included a relevant amount of imaging data from referring institutions where contrast media application was different and not standardized with our application protocol, potentially reflecting more inhomogeneous imaging data. Because we did not include MR images without any findings, our study did not capture the proper target population of interest. This bias might underestimate the false-positive rate in a true population. For our evaluation, we excluded 10% of initially identified patients due to, for example, a second cerebral tumor, strong artifacts, or insufficient contrast media application. Hence, images of these patients might not be suited to the proposed DLM.

CONCLUSIONS

Despite small lesion size and heterogeneous scanner data, our DLM detects brain metastases in malignant melanoma on multiparametric MR imaging with high detection and segmentation accuracy, while yielding a low false-positive rate.

REFERENCES

1. Davies MA, Liu P, McIntyre S, et al. **Prognostic factors for survival in patients with melanoma with brain metastases.** *Cancer* 2011; 117:1687–96 CrossRef Medline

2. Jakob JA, Bassett RL, Ng CS, et al. **NRAS mutation status is an independent prognostic factor in metastatic melanoma.** *Cancer* 2012; 118:4014–23 CrossRef Medline

3. Goyal S, Silk AW, Tian S, et al. **Clinical management of multiple melanoma brain metastases a systematic review.** *JAMA Oncol* 2015; 1:668–76 CrossRef Medline

4. Sperduto PW, Kased N, Roberge D, et al. **Summary report on the graded prognostic assessment: an accurate and facile diagnosis-specific tool to estimate survival for patients with brain metastases.** *J Clin Oncol* 2012;30:419–25 CrossRef Medline

5. Michielin O, van Akkooi A, Ascierto P, et al; ESMO Guidelines Committee. **Cutaneous melanoma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.** *Ann Oncol* 2019;30:1884–1901 CrossRef Medline

6. Schlamann M, Loquai C, Goericke S, et al. **Cerebral MRI in neurological asymptomatic patients with malignant melanoma** [in German]. *Rofo* 2008;180:143–47 CrossRef Medline

7. Garbe C, Amaral T, Peris K, et al. **European consensus-based interdisciplinary guideline for melanoma. Part 1: Diagnostics - Update 2019.** *Eur J Cancer* 2020;126:141–158 CrossRef Medline

8. Trotter SC, Sroa N, Winkelmann RR, et al. **A global review of melanoma follow-up guidelines.** *J Clin Aesthet Dermatol* 2013;6(9):18–26 Medline

9. Berbaum KS, Franken EA, Dorfman DD, et al. **Satisfaction of search in diagnostic radiology.** *Invest Radiol* 1990;25:133–40 CrossRef Medline

10. Brady AP. **Error and discrepancy in radiology: inevitable or avoidable?** *Insights Imaging* 2017;8:171–82 CrossRef Medline

11. Conson M, Cella L, Pacelli R, et al. **Automated delineation of brain structures in patients undergoing radiotherapy for primary brain tumors: from atlas to dose-volume histograms.** *Radiother Oncol* 2014;112:326–31 CrossRef Medline

12. Xue Y, Chen S, Qin J, et al. **Application of deep learning in automated analysis of molecular images in cancer: a survey.** *Contrast Media Mol Imaging* 2017;2017:9512370 CrossRef Medline

13. Fountain DM, Soon WC, Matys T, et al. **Volumetric growth rates of meningioma and its correlation with histological diagnosis and clinical outcome: a systematic review.** *Acta Neurochir (Wien)* 2017; 159:435–45 CrossRef Medline

14. Chang V, Narang J, Schultz L, et al. **Computer-aided volumetric analysis as a sensitive tool for the management of incidental meningiomas.** *Acta Neurochir* 2012;154:589–97 CrossRef Medline

15. Suh JH. **Stereotactic radiosurgery for the management of brain metastases.** *N Engl J Med* 2010;362:1119–27 CrossRef Medline

16. Liu Y, Stojadinovic S, Hrycushko B, et al. **A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery.** *PLoS One* 2017;12: e0185844 CrossRef Medline

17. Charron O, Lallement A, Jarnet D, et al. **Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network.** *Comput Biol Med* 2018;95:43–54 CrossRef Medline

18. Bauknecht HC, Romano VC, Rogalla P, et al. **Intra-and interobserver variability of linear and volumetric measurements of brain metastases using contrast-enhanced magnetic resonance imaging.** *Invest Radiol* 2010;45:49–56 CrossRef Medline

19. Zhou Z, Sanders JW, Johnson JM, 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 CrossRef Medline

20. Larson DB, Chen MC, Lungren MP, et al. **Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs.** *Radiology* 2018;287:313–22 CrossRef Medline

21. Lakhani P, Sundaram B. **Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks.** *Radiology* 2017;284:574–82 CrossRef Medline

22. Park A, Chute C, Rajpurkar P, et al. **Deep learning-assisted diagnosis of cerebral aneurysms using the HeadXNet model.** *JAMA Netw open* 2019;2:e195600 CrossRef Medline

23. Laukamp KR, Thiele F, Shakirin G, et al. **Fully automated detection and segmentation of meningiomas using deep learning on routine multiparametric MRI.** *Eur Radiol* 2019;29:124–32 CrossRef Medline

24. Perkuhn M, Stavrinou P, Thiele F, et al. **Clinical evaluation of a multiparametric deep learning model for glioblastoma segmentation using heterogeneous magnetic resonance imaging data from clinical routine.** *Invest Radiol* 2018;53:647–54 CrossRef Medline

25. Kickingereder P, Isensee F, Tursunova I, et al. **Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study.** *Lancet Oncol* 2019;20:728–40 CrossRef Medline

26. Kooi T, Litjens G, van Ginneken B, et al. **Large-scale deep learning for computer aided detection of mammographic lesions.** *Med Image Anal* 2017;35:303–12 CrossRef Medline

27. LeCun Y, Bengio Y, Hinton G. **Deep learning.** *Nature* 2015;521:436–44 CrossRef Medline

28. Akkus Z, Galimzianova A, Hoogi A, et al. **Deep learning for brain MRI segmentation: state of the art and future directions.** *J Digit Imaging* 2017;30:449–59 CrossRef Medline

29. Mazzara GP, Velthuizen RP, Pearlman JL, et al. **Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation.** *Int J Radiat Oncol Biol Phys* 2004;59:300–12 CrossRef Medline

30. Grøvik E, Yi D, Iv M, et al. **Deep learning enables automatic detection and segmentation of brain metastases on multisequence MRI.** *J Magn Reson Imaging* 2020;51:175–82 CrossRef Medline

31. Noguchi T, Uchiyama F, Kawata Y, et al. **A fundamental study assessing the diagnostic performance of deep learning for a brain metastasis detection task.** *Magn Reson Med Sci* 2020;19:184–94 CrossRef Medline

32. Bousabarah K, Ruge M, Brand JS, et al. **Deep convolutional neural networks for automated segmentation of brain metastases trained on clinical data.** *Radiat Oncol* 2020;15:87 CrossRef Medline

33. Laukamp KR, Pennig L, Thiele F, et al. **Automated meningioma segmentation in multiparametric MRI: comparable effectiveness of a deep learning model and manual segmentation.** *Clin Neuroradiol* 2020 Feb 14. [Epub ahead of print] CrossRef Medline

34. Kamnitsas K, Ledig C, Newcombe VFJ, et al. **Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation.** *Med Image Anal* 2017;36:61–78 CrossRef Medline

35. Crum WR, Camara O, Hill DL. **Generalized overlap measures for evaluation and validation in medical image analysis.** *IEEE Trans Med Imaging* 2006;25:1451–61 CrossRef Medline

36. Schoenmaekers J, Hofman P, Bootsma G, et al. **Screening for brain metastases in patients with stage III non–small-cell lung cancer, magnetic resonance imaging or computed tomography? A prospective study.** *Eur J Cancer* 2019;115:88–96 CrossRef Medline

37. Narita Y, Shibui S. **Strategy of surgery and radiation therapy for brain metastases.** *Int J Clin Oncol* 2009;14:275–80 CrossRef Medline

38. Bruno MA, Walker EA, Abujudeh HH. **Understanding and confronting our mistakes: the epidemiology of error in radiology and strategies for error reduction.** *Radiographics* 2015;35:1668–76 CrossRef Medline

39. Henson JW, Ulmer S, Harris GJ. **Brain tumor imaging in clinical trials.** *AJNR Am J Neuroradiol* 2008;29:419–24 CrossRef Medline

40. Lin NU, Lee EQ, Aoyama H, et al; Response Assessment in Neuro-Oncology (RANO) group, **Response assessment criteria for brain metastases: proposal from the RANO group.** *Lancet Oncol* 2015;16:e270–78 CrossRef Medline

41. Laukamp KR, Lindemann F, Weckesser M, et al. **Multimodal imaging of patients with gliomas confirms [11] C-MET PET as a complementary marker to MRI for noninvasive tumor grading and intraindividual follow-up after therapy.** *Mol Imaging* 2017;16 CrossRef Medline

42. Menze BH, Jakab A, Bauer S, et al. **The multimodal Brain Tumor Image Segmentation Benchmark (BRATS).** *IEEE Trans Med Imaging* 2015;34:1993–2024 CrossRef Medline

43. Laukamp KR, Pennig L, Thiele F, et al. **Automated meningioma segmentation in multiparametric MRI.** *Clin Neuroradiol* 2020 Feb 4. [Epub ahead of print] CrossRef Medline

44. Pennig L, Hoyer UC, Goertz L, et al. **Primary central nervous system lymphoma: clinical evaluation of automated segmentation on multiparametric MRI using deep learning.** *J Magn Reson Imaging* 2021;53:259–68 CrossRef Medline