Primary Central Nervous System Lymphoma: Clinical Evaluation of Automated Segmentation on Multiparametric MRI Using Deep Learning

Lenhard Pennig, MD,1* Ulrike Cornelia Isabel Hoyer, MD,1 Lukas Goertz, MD,1,2 Rahil Shahzad, PhD,1,3 Thorsten Persikehl, MD,1 Frank Thiele, MSc,1,3 Michael Perkuhn, MD,1,3 Maximilian I. Ruge, MD,4 Christoph Kabbasch, MD,1 Jan Borggrefe, MD,1 Liliana Caldeira, PhD,1† and Kai Roman Laukamp, MD1,5,6†

**Background:** Precise volumetric assessment of brain tumors is relevant for treatment planning and monitoring. However, manual segmentations are time-consuming and impeded by intra- and interrater variabilities.

**Purpose:** To investigate the performance of a deep-learning model (DLM) to automatically detect and segment primary central nervous system lymphoma (PCNSL) on clinical MRI.

**Study Type:** Retrospective.

**Population:** Sixty-nine scans (at initial and/or follow-up imaging) from 43 patients with PCNSL referred for clinical MRI tumor assessment.

**Field Strength/Sequence:** T1−/T2-weighted, T1-weighted contrast-enhanced (T1 CE), and FLAIR at 1.0, 1.5, and 3.0T from different vendors and study centers.

**Assessment:** Fully automated voxelwise segmentation of tumor components was performed using a 3D convolutional neural network (DeepMedic) trained on gliomas (n = 220). DLM segmentations were compared to manual segmentations performed in a 3D voxelwise manner by two readers (radiologist and neurosurgeon; consensus reading) from T1 CE and FLAIR, which served as the reference standard.

**Statistical Tests:** Dice similarity coefficient (DSC) for comparison of spatial overlap with the reference standard, Pearson’s correlation coefficient (r) to assess the relationship between volumetric measurements of segmentations, and Wilcoxon rank-sum test for comparison of DSCs obtained in initial and follow-up imaging.

**Results:** The DLM detected 66 of 69 PCNSL, representing a sensitivity of 95.7%. Compared to the reference standard, DLM achieved good spatial overlap for total tumor volume (TTV, union of tumor volume in T1 CE and FLAIR; average size 77.16 ± 62.4 cm³, median DSC: 0.76) and tumor core (contrast enhancing tumor in T1 CE; average size: 11.67 ± 13.88 cm³, median DSC: 0.73). High volumetric correlation between automated and manual segmentations was observed (TTV: r = 0.88, P < 0.0001; core: r = 0.86, P < 0.0001). Performance of automated segmentations was comparable between pretreatment and follow-up scans without significant differences (TTV: P = 0.242, core: P = 0.177).

**Data Conclusion:** In clinical MRI scans, a DLM initially trained on gliomas provides segmentation of PCNSL comparable to manual segmentation, despite its complex and multifaceted appearance. Segmentation performance was high in both initial and follow-up scans, suggesting its potential for application in longitudinal tumor imaging.

**Level of Evidence:** 3

**Technical Efficacy Stage:** 2

J. MAGN. RESON. IMAGING 2021;53:259–268.
PRIMARY CENTRAL NERVOUS SYSTEM LYMPHOMA (PCNSL) accounts for ~2% of primary intracranial tumors, with an increasing incidence over the past decades.\(^1\) PCNSL represents an aggressive type of extranodal lymphoma without coexisting systemic disease at diagnosis, separating it from CNS involvement occurring in systemic lymphoma (secondary CNS lymphoma).\(^2\)–\(^4\)

Meningiomas proved feasible and provided good segmentation for a different tumor entity, namely, PCNSL, as well.\(^6\)–\(^7\) Usually located in the periventricular white matter and adjacent to subarachnoid and ependymal surfaces, PCNSL lesions are often surrounded by edema.\(^6\)–\(^7\) However, atypical PCNSL, as in immunocompromised patients, may appear differently, with central necrosis, ring-like enhancement, and hemorrhage.\(^8\)–\(^9\)

Precise assessment of size is clinically relevant for therapy planning, prognosis, and monitoring in many brain tumors, such as glioblastoma, meningioma, or PCNSL.\(^5\)–\(^10\)–\(^12\)

Although recent studies indicated that 2D conventional measurements proved to be less reliable than volumetric assessment, manual brain tumor segmentation is time-consuming and suffers from relatively high intra- and interrater variabilities.\(^13\)–\(^16\) On the contrary, automated segmentation of PCNSL might facilitate precise and objective assessment of tumor burden and response in longitudinal imaging, which is often difficult to determine given its multifaceted appearance consisting of multiple scattered lesions.\(^5\)–\(^7\)

Previous studies have already demonstrated the feasibility of deep-learning models (DLMs) for completely automatic brain tumor segmentation, eg, for meningioma and glioblastoma, yielding excellent reproducibility.\(^10\)–\(^15\)–\(^16\) Recent technical innovations have further improved DLM-based tumor detection and segmentation, implying readiness for clinical application.\(^10\)–\(^15\)–\(^17\)

In a previous work, the use of a DLM, which was initially trained on gliomas and validated on glioblastomas,\(^17\) on meningiomas proved feasible and provided good segmentation performance for this tumor.\(^18\) Given these results, we postulated that the same DLM might provide accurate segmentations for a different tumor entity, namely, PCNSL, as well.

The purpose of this study was to evaluate the efficacy of a DLM initially trained on gliomas for fully automated detection and segmentation of PCNSL on multiparametric MRI including heterogeneous imaging data from multiple scanner vendors and several study centers.

**Materials and Methods**

The local Institutional Review Board approved this retrospective, single-center study (reference number: 19–1208) and waived the requirement for written informed consent for the patient cohort.

**Patient Population**

MRI scans of patients treated for PCNSL at our tertiary-care university hospital between November 2010 and July 2019 were reviewed using our institutional image archiving system and clinical information system.

Eighty-one scans of 53 patients, regardless of initial or follow-up imaging, were identified applying the following inclusion criteria: 1) a complete MRI-dataset, consisting of \(T_1-/T_2\)-weighted, \(T_1\)-weighted contrast-enhanced \((T_1, CE)\), and fluid-attenuated inversion recovery (FLAIR) sequences, 2) histologically proven PCNSL, and 3) a clearly identifiable presence of tumor, defined as a minimum volumetric threshold of 0.25 cm\(^3\) for tumor core. Exclusion criteria were: a) severe artifacts (seven scans of five different patients), b) extensive leukoencephalopathy (Fazekas III, three datasets of three different patients), and c) second intracranial tumor (meningioma: two scans of two different patients) as determined by a board certified neuroradiologist (J.B.) with 13 years of experience in neuro-oncologic imaging. This senior neuroradiologist also assessed PCNSLs regarding their imaging characteristics (multifocal tumor spread, supra-/infratentorial location, necrosis, and hemorrhage) and potential tumor recurrence. Applying the aforementioned exclusion criteria, 69 examinations from 43 patients (mean age at initial imaging: 62.6 ± 13.3 years, 23 female) were included, consisting of 36 scans at primary diagnosis and 33 follow-up scans. Twenty-two patients received primary imaging only, 14 patients both primary and follow-up, and seven patients follow-up imaging only.

Included datasets (unenhanced \(T_1\)- and \(T_2\)-weighted, \(T_1\ CE\), and FLAIR images) were anonymized and exported to IntelliSpace Discovery (ISD, v. 3.0, Philips Healthcare, Best, The Netherlands). The following data were collected from the patients’ medical charts: procedure (stereotactic biopsy or surgery) to obtain material for histopathological diagnosis of PCNSL, available additional imaging and bone marrow biopsy to rule out systemic lymphatic disease, infection with human immunodeficiency virus (HIV), and treatment received.

**Imaging**

MRI was performed at 1.0 (\(n = 3\), 1.5 (\(n = 30\), and 3.0T (\(n = 36\) with examinations either performed at our institution (\(n = 54\)) or referring centers (\(n = 15\)). All scans were conducted for clinical purposes and included unenhanced \(T_1\)- and \(T_2\)-weighted, \(T_1\ CE\), and FLAIR images. Table 1 provides detailed data on the scanners used and Table 2 lists the imaging parameters for each field strength. The standardized imaging protocol at our institution included automatic intravenous administration of gadolinium-based contrast agent (Dotarem, 0.5 mmol/mL, 1 mL = 279.3 mg gadoteric acid = 78.6 mg gadolinium; Guerbet, Villepinte, France) with a concentration of 0.1 mmol/kg body weight. In referring institutions, contrast medium application was not standardized.

**Reference Standard**

A radiologist and a neurosurgeon (L.P., L.G.), each with 4 years of experience in neuro-oncologic imaging, together assessed PCNSLs on unenhanced \(T_1\) and \(T_2\)-weighted, \(T_1\ CE\), and FLAIR images on ISD. To establish the reference standard, both readers performed segmentations on ISD simultaneously (using consensus reading).

Segmentation was conducted interactively using a 2-step, semiautomatic approach. First, a rough segmentation of tumor tissue
contrast-enhancing tumor volume in T1 CE. Total tumor volume tumor parts, and surrounding FLAIR-hyperintense tumor as well as FLAIR, including solid contrast-enhancing tumor parts, necrotic classi
des of size 33. The model also consists of residual connections third of its original size. The model comprises 11 layers with ker-
ner, for the second pathway, the image was downsampled to a same image location, provide inputs to the two pathways. How-
tumor appearances. 3D image-segments (patches), centered at the which helps in capturing short- and long-range characteristics of two identical pathways that apply different image resolutions,
the loss function and the number of training epochs was set to 35. Each epoch consisted of an additional 20 subepochs, for which around 1000 3D image-segments of size 25^3 where randomly extracted from the training data. To ensure class-balance, the extracted patches had a distribution of 50% between background and tumor.

During inference on the PCNSL data, the size of the 3D image-segments was set to 45^3 to increase the kernel receptive field and reduce the inference time. Before the multiparametric scans of the PCNSLs (unenhanced T1- and T2-weighted, T1 CE, and FLAIR images) were passed to the CNN, automatic preprocessing was applied. The image preprocessing pipeline included 1) skull stripping employing a brain mask, 2) coregistration of unenhanced T1- and T2-weighted sequences as well as FLAIR to the reference space defined by T1 CE, 3) bias field correction of all four sequences applying a proprietary method (Philips Healthcare) for brain MRI,20 4) resampling to isotropic resolution of 1 × 1 × 1 mm^3, and 5) nor-
mization to zero-mean and a standard deviation of one.17

### Statistical Analysis

Statistical analysis was performed using JMP Software (release 14, SAS Institute, Cary, NC), with P < 0.05 considered statistically signi
cant. Volumes of tumor components (TTV and tumor core) are given as mean ± standard deviation. Successful detection of PCNSL was defined if the DLM obtained a spatial overlap with the manual segmentation of the tumor core (at least one voxel, DSC >0). Detection sensitivity was calculated applying the following formula:

\[
Sensitivity = \frac{TP}{TP + FN} \times 100
\]

With TP being true positive and FN false negative. To evaluate the segmentation (S) performance of the DLM compared to manual segmentations, the DSC, which provides a measure of spatial overlap for each voxel of segmented tissue, was calculated for each tumor component.17,18,21:

\[
DSC(S_{RS}, S_{DLM}) = \frac{2 | S_{RS} \cap S_{DLM} |}{| S_{RS} | + | S_{DLM} |}
\]

where \( S_{RS} \) is segmentation provided by human readers (reference standard) and \( S_{DLM} \) is segmentation provided by the DLM. Resulting DSCs are reported as median with 10/90 percentiles. Comparison of DSCs obtained in initial and follow-up scans was performed with Wilcoxon rank-sum test. Corresponding 3D render-
ing images of automated and manual tumor components were cre-
ted using ISD. To evaluate volumetric agreement between manual and automated segmentations for TTV and tumor core, Pearson’s correlation coefficient (r) was calculated and Bland–Altman analysis

---

**TABLE 1. MRI Scanners Used in This Study, Manufactured by Philips Healthcare, Best, The Netherlands and Siemens Healthineers, Erlangen, Germany**

| Model               | Number of scans |
|---------------------|----------------|
| Philips Achieva 1.5T | 6              |
| Philips Achieva 3.0T | 16             |
| Philips Gyroscan 1.0T | 1             |
| Philips Ingenia 3.0T | 17             |
| Philips Interia 1.5T | 19             |
| Philips Panorama 1.0T | 3             |
| Siemens Avanto 1.5T | 4              |
| Siemens Magneton 1.5T | 1             |
| Siemens Symphony 1.5T | 1             |
| Siemens TrioTrim 3.0T | 1             |

| Location of examination | Number of scans |
|-------------------------|-----------------|
| Local institution       | 54              |
| Referring institutions  | 15              |

was obtained by 3D voxelwise regional thresholding. Then the initial segmentation was further edited manually using a combination of 2D editing tools. Contrast-enhancing tumor and iso- to hyperintense tumor, including surrounding hyperintense edema, were segmented separately in T1 CE and FLAIR. The tumor core was defined as the contrast-enhancing tumor volume in T1 CE. Total tumor volume (TTV) was defined as the union of tumor volumes in T1 CE and FLAIR, including solid contrast-enhancing tumor parts, necrotic tumor parts, and surrounding FLAIR-hyperintense tumor as well as edema.

**Deep-Learning Model**

In this study, a 3D convolutional neural network (CNN) based on DeepMedic (Biomedical Image Analysis Group, Department of Computing, Imperial College London, UK), trained on an independent dataset of 220 glioma cases available through the Brain Tumor Image Segmentation (BRATS) 2015 challenge, was used.19 As defined in the BRATS benchmark, the DLM performs voxelwise classifications of four tumor components (edema, contrast-enhancing tumor, necrosis, and nonenhancing tumor).19

The network consists of a deep 3D CNN architecture with two identical pathways that apply different image resolutions, which helps in capturing short- and long-range characteristics of tumor appearances. 3D image-segments (patches), centered at the same image location, provide inputs to the two pathways. How-
however, for the second pathway, the image was downsampled to a third of its original size. The model comprises 11 layers with ker-
nels of size 3^3. The model also consists of residual connections for layers 4, 6, 8, and 10, whereas layers 9 and 10 are fully connected.

During training of the glioma data,17 on-the-fly image aug-
mentation was employed by flipping the images along their axes. The training batch size was set to 15; batch normalization was applied and a parametric rectified linear unit was used as the activa-
tion function. A Dice similarity coefficient (DSC) was provided as the loss function and the number of training epochs was set to 35. During training on the PCNSL data, the size of the 3D image-segments was set to 45^3 to increase the kernel receptive field and reduce the inference time. Before the multiparametric scans of the PCNSLs (unenhanced T1- and T2-weighted, T1 CE, and FLAIR images) were passed to the CNN, automatic preprocessing was applied. The image preprocessing pipeline included 1) skull stripping employing a brain mask, 2) coregistration of unenhanced T1- and T2-weighted sequences as well as FLAIR to the reference space defined by T1 CE, 3) bias field correction of all four sequences applying a proprietary method (Philips Healthcare) for brain MRI, 4) resampling to isotropic resolution of 1 × 1 × 1 mm^3, and 5) nor-
mization to zero-mean and a standard deviation of one.17

---

January 2021 261

Pennig et al.: Automated Segmentation of Cerebral Lymphoma
| Parameter       | FLAIR          | T₁       | T₂       | T₁ CE      |
|-----------------|----------------|----------|----------|------------|
| Acquisition matrix | 224–320 × 185–320 | 256–404 × 254–333 | 256–512 × 228–450 | 256–412 × 205–414 |
| Average slice thickness (mean, mm ± SD) | 3.65 ± 1.62 | 4.93 ± 0.35 | 4.77 ± 0.49 | 2.81 ± 1.88 |
| Average in-plane resolution (mean, mm² ± SD) | 0.40 ± 0.11 × 0.40 ± 0.11 | 0.47 ± 0.05 × 0.47 ± 0.05 | 0.42 ± 0.11 × 0.42 ± 0.11 | 0.46 ± 0.11 × 0.46 ± 0.11 |
| Repetition time (msec) | 4800–1200 | 295–910 | 2150–4250 | 8.5–910 |
| Echo time (msec) | 132–340 | 4.6–13.4 | 16.1–107 | 2.4–13.4 |
| Inversion time (msec) | 1650–2850 | — | — | — |
| Flip angle (°) | 90–140 | 80–90 | 18–150 | 8–90 |

### 1.5T

| Parameter       | FLAIR          | T₁       | T₂       | T₁ CE      |
|-----------------|----------------|----------|----------|------------|
| Acquisition matrix | 232–512 × 140–203 | 244–388 × 174–308 | 224–556 × 179–308 | 200–388 × 205–308 |
| Average slice thickness (mean, mm ± SD) | 5.50 ± 0.69 | 5.77 ± 0.39 | 5.53 ± 0.73 | 3.94 ± 1.92 |
| Average in-plane resolution (mean, mm² ± SD) | 0.82 ± 0.18 × 0.82 ± 0.18 | 0.76 ± 0.19 × 0.76 ± 0.19 | 0.56 ± 0.22 × 0.56 ± 0.22 | 0.66 ± 0.22 × 0.66 ± 0.22 |
| Repetition time (msec) | 6000–11000 | 174–764 | 635–8086 | 8–764 |
| Echo time (msec) | 89–142 | 2.4–23 | 15–120 | 1.8–17 |
| Inversion time (msec) | 2000–2925 | — | — | — |
| Flip angle (°) | 90–180 | 18–90 | 90–180 | 12–90 |

### 1.0T

| Parameter       | FLAIR          | T₁       | T₂       | T₁ CE      |
|-----------------|----------------|----------|----------|------------|
| Acquisition matrix | 256–280 × 169–172 | 276 × 218–220 | 352–384 × 297–318 | 280–336 × 224–336 |
| Average slice thickness (mean, mm ± SD) | 6 | 6 | 6 | 3.8 ± 2.2 |
| Average in-plane resolution (mean, mm² ± SD) | 0.89 ± 0.01 × 0.89 ± 0.01 | 0.87 ± 0.87 | 0.45 ± 0.01 × 0.45 ± 0.01 | 0.65 ± 0.22 × 0.65 ± 0.22 |
| Repetition time (msec) | 6000 | 625–707 | 4403–4994 | 25–707 |
| Echo time (msec) | 120 | 15 | 100 | 9.8–15 |
| Inversion time (msec) | 2000 | — | — | — |
| Flip angle (°) | 90 | 69 | 90 | 30–69 |

FLAIR: fluid-attenuated inversion recovery; T₁ CE: T₁-weighted contrast-enhanced; SD: standard deviation.
was performed with graphs being created with Python (release 3.7, Python Software Foundation, Wilmington, DE) Matplotlib.

Results

Characteristics of Patients and PCNSL

The diagnosis of PCNSL was confirmed by stereotactic biopsy in 40 cases and by surgical specimen in three. Systemic lymphatic disease was ruled out by additional imaging and bone marrow biopsy in all cases. All patients were HIV-negative, with the majority (95.3%) receiving chemotherapy. For six patients, a recurrence of the tumor was noted during follow-up imaging.

The majority (94.2%) of the PCNSLs showed supratentorial location and multifocal tumor spread (69.6%). Table 3 provides detailed information regarding patients’ treatment. Table 4 lists the localization of PCNSLs and additional imaging characteristics. Mean TTV and tumor core volume, based on manual segmentation, were $77.16 \pm 62.40 \text{ cm}^3$ and $11.67 \pm 13.88 \text{ cm}^3$, respectively.

Evaluation of the DLM

DETECTION. The DLM detected 66 of 69 PCNSL correctly, referring to a sensitivity of 95.7%. The missed PCNSLs were small (average core volume of $1.1 \pm 0.55 \text{ cm}^3$) and were located directly adjacent to the ventricles. Figure 1 shows an example of a missed PCNSL.

SEGMENTATION. The DLM yielded a mean volume of $6.11 \pm 7.12 \text{ cm}^3$ for tumor core and of $67.26 \pm 46.14 \text{ cm}^3$ for TTV. Compared to manual segmentations, the DLM had a DSC of 0.76 (10/90: 0.40/0.90) for TTV and of 0.73 (10/90: 0.37/0.88) for core tumor. Comparing volumetric assessment by automated and manual segmentations, correlation coefficients were $r = 0.88$ ($P < 0.0001$) for TTV and $r = 0.86$ ($P < 0.0001$) for tumor core (Fig. 2). Bland–Altman analysis revealed a good agreement between manual and automated segmentations (Fig. 3).

In pretreatment scans, DSC of 0.80 (10/90: 0.36/0.90) for TTV and of 0.74 (10/90: 0.45/0.88) for core were noted. In follow-up scans, DLM segmentations achieved a DSC of 0.66 (10/90: 0.39/0.90) for TTV ($P = 0.242$) and of 0.68 (10/90: 0.34/0.91) for core ($P = 0.177$) compared to manual segmentations. Figures 4–7 show exemplary comparisons of manual and automated segmentations of tumor components with corresponding 3D rendering.

Discussion

In this study we evaluated the performance of a DLM initially trained on gliomas for fully automated detection and segmentation of PCNSL on clinical routine MRI. Although PCNSL inherently shows a heterogeneous and complex tumor structure on MRI, the DLM we used detected and segmented PCNSL with high segmentation performance independent of scan timing (initial vs. follow-up).

Apart from atypical forms, glioblastoma and PCNSL often present with distinct differences in MRI features enabling potential differentiation.22–24 PCNSL usually consists of several scattered homogenous enhanced tumor parts.

| Initial imaging | FU imaging | Multifocal | Supratentorial | Infratentorial | Necrosis | Hemorrhage |
|-----------------|------------|------------|----------------|----------------|----------|------------|
| 36              | 33         | 48         | 65             | 14             | 22       | 23         |

FU: follow-up.
FIGURE 2: Volumetric correlation between manual and automated segmentations for total tumor volume (TTV, union of tumor volume in T1 CE and FLAIR; left) and for tumor core (contrast-enhancing tumor in T1 CE, right) per scan using Pearson correlation ($r$). The shaded area represents the 95% confidence interval of the fitted line, which is quite narrow, indicating high certainty about the calculated $r$-values. DLM: deep-learning model.

FIGURE 3: Volumetric agreement between manual and automated segmentations for TTV (left) and for tumor core (right) per scan using Bland–Altman analysis. The middle solid line indicates the mean difference between the segmented volumes, the dotted lines represent the 95% limits of agreement. DLM: deep-learning model.

FIGURE 4: Pretreatment MRI (a–c: T1 CE; e–g: FLAIR) of a 72-year-old male patient with partially hemorrhagic PCNSL of the splenium of the corpus callosum. Compared to manual segmentations of T1 CE tumor (b, blue) and FLAIR iso- to hyperintense tumor including edema (f, purple), the DLM provides similar segmentations of tumor core (c, red; DSC: 0.92) and TTV (g, green; DSC: 0.84). Corresponding 3D rendering of the PCNSL indicates a good overlap of tumor components between manual (d; tumor core: blue, TTV: purple) and deep-learning generated (h; tumor core: red, TTV: green) segmentations. L refers to the patient’s left side.
whereas glioblastoma mostly presents as a tumor with a ring-like zone of contrast-enhancement surrounding necrosis and intraleisional hemorrhage.\textsuperscript{22–25} However, PCNSL and glioblastoma show similarities in their features, such as contrast-enhancing tumor, potential central necrosis, FLAIR-hyperintense tumor parts, and surrounding edema.\textsuperscript{22–25} Therefore, we hypothesized that a DLM that had already performed well in detection and segmentation of

![Figure 5](image5.png)

FIGURE 5: Follow-up MRI of a 73-year-old male patient after receiving 2 months of chemotherapy for a PCNSL of the right parietal lobe (a–c: T\textsubscript{1} CE; e–g: FLAIR). Compared to segmentations of tumor core (b, blue) and TTV (f, purple) by human readers, the DLM achieves similar segmentations of tumor core (c, red; DSC 0.91) and TTV (g, green; DSC 0.88). Corresponding 3D rendering of the PCNSL indicates a strong overlap of tumor components between manual (d; tumor core: blue; TTV: purple) and deep-learning generated (h; tumor core: red; TTV: green) segmentations. L refers to the patient’s left side.

![Figure 6](image6.png)

FIGURE 6: Pretreatment MRI (a–c: T\textsubscript{1} CE; e–g: FLAIR) of a 63-year-old female patient diseased with PCNSL of the rostrum of corpus callosum and adjacent frontal lobes. Compared to manual segmentations of contrast-enhancing tumor in T\textsubscript{1} CE (b, blue) and iso- to hyperintense tumor including edema in FLAIR (f, purple), the DLM provides similar segmentations of tumor core (c, red; DSC: 0.77) and TTV (g, green; DSC: 0.89). Corresponding 3D rendering of the PCNSL indicates a similar overlap of tumor components between manual (d; tumor core: blue; TTV: purple) and deep-learning generated (h; tumor core: red; TTV: green) segmentations, despite the scattered contrast-enhancing lesions of the tumor. L refers to the patient’s left side.
glioblastoma might also be applied to PCNSL. The potential to apply the DLM trained for glioblastoma on a different tumor type (meningiomas) has previously been demonstrated. While Laukamp et al showed that dedicated retraining on meningioma cases allowed for a significant improvement of accuracy, this retraining may not be necessary for PCNSL since glioblastoma tends to be more similar in appearance on MRI to PCNSL than meningioma.

In this study the DLM provided high segmentation performance, with the results being comparable to research focusing on automated glioblastoma and glioma segmentation. In this context, Perkuhn et al reported DSCs between 0.62 and 0.86 for the differing tumor components of glioblastoma while using the same DLM as in the present study. Using a different CNN, Menze et al and Hoseini et al calculated DSCs of 0.74–0.85 and 0.84–0.90, for glioma segmentations, respectively. Kickingereder et al reported DSCs between 0.89 and 0.93 for an artificial neural network trained on a larger cohort of patients with glioblastoma.

Given the comparable results of the study by Perkuhn et al and the present study, a DLM trained on gliomas and achieving accurate segmentation of glioblastomas can be applied on PCNSL, despite their often different appearance and overall complex tumor structure without noticeably decreasing segmentation performance. Additionally, the DSCs we observed are comparable to those reported by Laukamp et al, in which the same DLM we used was applied to meningiomas without any training (DSC between 0.78 and 0.81).

Further, segmentation performance of the DLM in the present study is comparable to interrater variabilities of human readers for glioblastoma, which were reported to be as high as 20–30%, which corresponds to a DSC between 0.70 and 0.80.

Automated detection of cerebral masses, being possible with the proposed DLM of the present study, has high clinical relevance, as it may enable preselection of lesions and/or serve as a control mechanism for the radiologist and treating physician. Following stereotactic biopsy, the procedure of choice for histopathological diagnosis, standard treatment of PCNSL consists of different approaches, including chemotherapy based on high-dose methotrexate with or without radiotherapy, high-dose chemotherapy combined with autologous stem-cell therapy, or immunotherapy using rituximab. However, an optimal standard of treatment has not been established for PCNSL. The ability to precisely assess tumor response is necessary to study the efficacy of various PCNSL therapy regimes, and the DLM we studied here may be useful for longitudinal assessment of tumor response in oncological follow-up imaging.

2D diameter measurement methods underestimate slow and multifocal tumor growth or shrinkage, the latter being typical for PCNSL, with possible multiple scattered small enhancing tumor parts in periventricular localization. In contrast, volumetric assessment of tumor size proved to be of higher sensitivity to detect tumor progression or regression. Further, it can be used for planning of

![Figure 7](image-url)
stereotactic biopsy and evaluation of tumor features based on radiomics. Additionally, reliable automated segmentation would lead to increased availability of volumetric data by omitting time-consuming manual volumetric segmentations. Automated tumor segmentation also improves objectivity and offers far higher reproducibility of tumor margins, thereby improving comparability of measurements in oncological follow-up imaging.

Further, we evaluated whether the segmentation performance of the DLM would differ on initial scans compared to follow-up imaging of treated PCNSL, which may include smaller and potentially hemorrhagic tumors, as well as post-operative/postinterventional. Remarkably, the DLM performed independently of scan timing. Additionally, by including datasets from referring institutions acquired on different scanners, we demonstrated the robustness of the DLM on multivendor, heterogeneous data with differing protocols considering contrast media application and field strength, therefore indicating its applicability in clinical routine.

The combined approach of radiomics and machine learning has shown potential to differentiate between PCNSL and glioblastomas, which is clinically especially important for atypical forms of PCNSL due to their similarity to glioblastomas and vice versa. However, manual segmentation for radiomics analysis and/or deep-learning-based characterization of brain tumors suffer from inter- and intrarater variabilities as well as lower reproducibility, both being potentially highly problematic for these kind of analyses. Feature extraction from automated segmentations of brain tumors, eg, PCNSL and glioblastoma, might overcome this challenge by omitting inter- and intrarater variabilities, hence leading to increased robustness of tumor characterization.

Limitations
Given its retrospective study design, whether segmentation performance was sufficient for clinical needs and further investigation focusing on dedicated clinical tasks and parameters is needed. We used a DLM initially trained on gliomas for segmentation of PCNSL. This needs to be considered as a potential drawback of this study design with a newly designed/trained DLM potentially achieving superior results. Given the fact that 22% of included scans were acquired at referring institutions with unknown protocols of imaging acquisition and contrast media application, the DLM should be investigated in a multicenter setting evaluating its true performance on different scanners and protocols.

Conclusion
Deep learning offered comparable detection and segmentation of PCNSL on routine clinical MRI compared to manual segmentation, despite the complex and multifaceted appearance of this tumor type and applied DLM being initially trained on gliomas. Performance of the DLM was good for both scans at initial examination and scans after therapy, implying potential for this method to be used in follow-up assessment of PCNSL.

Disclosures
The authors declare relationships with the following company: Philips Healthcare; Rahil Shahzad, Frank Thiele, and Michael Perkuhn: Employees; Jan Borggreve: Speaker.

References
1. Ostrom QT, Gittleman H, Fulop J, et al. CBTRUS statistical report: Primary brain and central nervous system tumors diagnosed in the United States in 2008-2012. Neuro Oncol 2015;17:v1-v62.
2. Löw S, Han CH, Batchelor TT. Primary central nervous system lymphoma. Ther Adv Neurol Disord 2018;11:1756286418793562.
3. Camilleri-Broët S, Criniè E, Broët P, et al. A uniform activated B-cell-like immunophenotype might explain the poor prognosis of primary central nervous system lymphomas: Analysis of 83 cases. Blood 2006;107:190-196.
4. Swerdlow SH, Campo E, Harris NL, et al. WHO Classification Tumours Haematop Lymphoid Tissue Lyon, France: International Agency for Research on Cancer; 2017. https://doi.org/10.1017/CBO9781107415324.004.
5. Abrey LE, Batchelor TT, Ferreri AJM, et al. Report of an international workshop to standardize baseline evaluation and response criteria for primary CNS lymphoma. J Clin Oncol 2005;23:5034-5043.
6. Küker W, Nägele T, Korfel A, et al. Primary central nervous system lymphomas (PCNSL): MRI features at presentation in 100 patients. J Neuroradiol 2005;22:169-177.
7. Laharie K, Schilling A, Heidenreich J, et al. Radiologic morphology of low-grade primary central nervous system lymphoma in immunocompetent patients. Am J Neuroradiol 2005;26:2446-2454.
8. Louis DN, Ohgaki H, Wiestler OD, et al. The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol 2007;114:97-109.
9. Kieckenereder P, Wiestler B, Sahm F, et al. Primary central nervous system lymphoma and atypical glioblastoma: Multiparametric differentiation by using diffusion-, perfusion-, and susceptibility-weighted MR imaging. Radiology 2014;272:843-850.
10. Xue Y, Chen S, Qin J, Liu Y, Huang B, Chen H. Application of deep learning in automated analysis of molecular images in cancer: A survey. Contrast Media Mol Imaging 2017;2017:9512370.
11. 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;154:589-597.
12. Taboure E, Houiller C, Martin-Duverneuil N, et al. Patterns of response and relapse in primary CNS lymphomas after first-line chemotherapy: Imaging analysis of the ANOCEF-GOEALMS prospective randomized trial. Neuro Oncol 2017;19:422-429.
13. Sorensen AG, Patel S, Harmath C, et al. Comparison of diameter and perimeter methods for tumor volume calculation. J Clin Oncol 2001;19:551-557.
14. Chow DS, Qi J, Guo X, et al. Semiautomated volumetric measurement on postcontrast MR imaging for analysis of recurrent and residual disease in glioblastoma multiforme. Am J Neuroradiol 2014;35:498-503.
15. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: State of the art and future directions. J Digit Imaging 2017;30:449-459.
16. Mazara GP, Velthuizen RP, Pearlman JL, Greenberg HM, Wagner H. Brain tumor target volume determination for radiation treatment
planning through automated MRI segmentation. Int J Radiat Oncol Biol Phys 2004;59:300-312.

17. Perkuhn M, Stavinou P, Thiele F, et al. Clinical evaluation of a multi-parametric deep learning model for glioblastoma segmentation using heterogeneous magnetic resonance imaging data from clinical routine. Invest Radiol 2018;53:647-654.

18. Laukamp KR, Thiele F, Shakirin G, et al. Fully automated detection and segmentation of meningiomas using deep learning on routine multi-parametric MRI. Eur Radiol 2019;29:124-132.

19. Menze BH, Jakab A, Bauer S, et al. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging 2015;34:1993-2024.

20. Zhang B, Peeters H: Efficient MR inhomogeneity correction by regularized entropy minimization and proximal alternations. IEEE 12th International Symposium on Biomedical Imaging (ISBI), New York; 2015, p 1069-1072.

21. Crum WR, Camara O, Hill DLG. Generalized overlap measures for evaluation and validation in medical image analysis. IEEE Trans Med Imaging 2006;25:1451-1461.

22. Al-Okaili RN, Krejza J, Woo JH, et al. Intraaxial brain masses: MR imaging-based diagnostic strategy — Initial experience. Radiology 2007;243:539-550.

23. Koeller KK, Smirniotopoulos JG, Jones RV. Primary central nervous system lymphoma: Radiologic-pathologic correlation. Radiographics 1997;17:1497-1526.

24. Rees JH, Smirniotopoulos JG, Jones RV, Wong K. Glioblastoma multiforme: Radiologic-pathologic correlation. Radiographics 1996;16:1413-1438.

25. 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 intracranial follow-up after therapy. Mol Imaging 2017;16:1536012116687651.

26. Laukamp KR, Pennig L, Thiele F, Reimer R, Görtz L, Shakirin G, Zopfs D, Timmer M, Perkuhn M, Borggrefe J: Automated meningioma segmentation in multiparametric MRI: Comparable effectiveness of a deep learning model and manual segmentation. Clin Neuroradiol 2020. doi: https://doi.org/10.1007/s00062-020-00884-4 [epub ahead of print].

27. Hoseini F, Shahbahrami A, Bayat P. An efficient implementation of deep convolutional neural networks for MRI segmentation. J Digit Imaging 2018;31:738-747.

28. 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;20:728-740.

29. Fyllingen EH, Stensgåen AL, Berntsen EM, Solheim O, Reinertsen I. Glioblastoma segmentation: Comparison of three different software packages. PLoS One 2016;11:e0164891.

30. Diepenbrock S, Prašni J, Lindemann F, Bothe H, Ropinski T. Interactive visualization techniques for neurosurgery planning. Eurographics 2011. https://doi.org/10.2312/EG2011/med/013-016.

31. Schlegel U. Primary CNS lymphoma. Ther Adv Neurol Disord 2009;2:93-104.

32. Rubenstein JL, Fridlyand J, Abrey L, et al. Phase I study of intraventricular administration of rituximab in patients with recurrent CNS and intraocular lymphoma. J Clin Oncol 2007;25:1350-1356.

33. Fountain DM, Soon WC, Matys T, Guilfoyle MR, Kirollos R, Santarius T. Volumetric growth rates of meningioma and its correlation with histological diagnosis and clinical outcome: A systematic review. Acta Neurochir 2017;159:435-445.

34. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology 2016;278:563-577.

35. Laukamp K, Shakirin G, Baeßler B, et al. Accuracy of radiomics-based feature analysis on multiparametric MR images for non-invasive meningioma grading. World Neurosurg 2019;132:e366-e390.

36. Bühring U, Herrlinger U, Krings T, Thiex R, Weller M, Küker W. MRI features of primary central nervous system lymphomas at presentation. Neurology 2001;57:393-396.

37. Suh HB, Choi YS, Bae S, et al. Primary central nervous system lymphoma and atypical glioblastoma: Differentiation using radiomics approach. Eur Radiol 2018;28:3832-3839.

38. Kang D, Park JE, Kim YH, et al. Diffusion radiomics as a diagnostic modal for atypical manifestation of primary central nervous system lymphoma: Development and multicenter external validation. Neuro Oncol 2018;20:1251-1261.