A Deep Learning Approach for Placing Magnetic Resonance Spectroscopy Voxels in Brain Tumors

Sangyoon Lee[1, 2] [0009-0005-8825-5175], Francesca Branzoli[3] [0000-0001-9792-0492], Thanh Nguyen[4], Ovidiu Andronesi[5] [0000-0002-7412-0641], Alexander Lin[6], Roberto Liserre[7] [0000-0003-1536-0183], Gerd Melkus[4] [000-0001-5387-8958], Clark Chen[8, 9] [0000-0001-9544-2570], Małgorzata Marjańska[1] [0000-0002-4727-2447], and Patrick J. Bolan[1] [0000-0002-4194-3975]

[1] Center for Magnetic Resonance Research, Department of Radiology, University of Minnesota, MN, USA
[2] Department of Radiation Oncology, University of Minnesota Medical School, MN, USA
[3] Paris Brain Institute - ICM, Inserm U 1127, CNRS UMR 7225, Sorbonne University, Paris, France
[4] Department of Radiology, Radiation Oncology and Medical Physics, University of Ottawa, Ottawa, ON, Canada
[5] Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA
[6] Center for Clinical Spectroscopy, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[7] ASST Spedali Civili University Hospital, Brescia, Italy
[8] Department of Neurosurgery, University of Minnesota, USA
[9] Department of Neurosurgery, Brown University, USA

bola0035@umn.edu

Abstract. Magnetic resonance spectroscopy (MRS) of brain tumors provides useful metabolic information for diagnosis, treatment response, and prognosis. Single-voxel MRS requires precise planning of the acquisition volume to produce a high-quality signal localized in the pathology of interest. Appropriate placement of the voxel in a brain tumor is determined by the size and morphology of the tumor, and is guided by MR imaging. Consistent placement of a voxel precisely within a tumor requires substantial expertise in neuroimaging interpretation and MRS methodology. The need for such expertise at the time of scan has contributed to low usage of MRS in clinical practice. In this study, we propose a deep learning method to perform voxel placements in brain tumors. The network is trained in a supervised fashion using a database of voxel placements performed by MRS experts. Our proposed method accurately replicates the voxel placements of experts in tumors with comparable tumor coverage, voxel volume, and voxel position to that of experts. This novel deep learning method can be easily applied without an extensive external validation as it only requires a segmented tumor mask as input.

Keywords: Single-voxel spectroscopy, Voxel placement, Brain cancer, Tumor, Deep Learning.

1 Introduction

Magnetic resonance spectroscopy (MRS) provides clinically-relevant metabolic information from brain tumors by assessing over 20 brain metabolites relevant to tumor type and progression [1]. For example, MRS can accurately detect the presence of the compound D-2-hydroxyglutarate, which indicates a mutation in the isocitrate dehydrogenase (*IDH*) gene which is associated with increased patient survival [2-6]. Integrating this non-invasive method for biochemical assessment with existing imaging modalities has shown promise in enhancing diagnosis and treatment response monitoring [7].

However, one challenge hindering widespread adoption of MRS in clinical practice is its acquisition process, which demands expertise in precisely positioning the acquisition volume prospectively at the time of scanning [8]. This volume, commonly called a voxel, is a 3D cuboid region that should be positioned within the tumor while avoiding other tissues. Placement of the voxel is critical for producing high quality MRS data, yet there is no objective way to determine an optimal placement. Placement can vary significantly depending on tumor size, location, and stage, making it subjective and leading to low spatial overlap of voxels even among experts [8, 9].

While several studies have proposed automated voxel placement methods, accurately predicting voxels aligned with expert decisions remains challenging [8, 10-12]. Most prior works focused on voxel placement in normal brain regions, which can be defined *a priori* using atlases or other imaging landmarks. These pre-defined placements are translated to an individual subject using image registration [10-12], but this approach does not work for brain tumors since the location is unknown until scan time [10-12]. In contrast, the method of Bolan et al. identified the tumor extent at scan time using a convolutional neural network (CNN), and then optimized a voxel placement within that tumor (position, size, orientation) by optimizing a hand-crafted objective function [8]. An objective function-based approach such as this may be effective in some cases, but may also fail to match experts' decisions entirely [8]. To address this challenge, we developed a deep learning-based model trained on expert-defined voxel placements to automatically generate voxel placements rather than optimizing objective function or hand-crafted models. We selected a CNN architecture for the model for its ability to capture high-dimensional information distributions [13]. Our proposed method also uses tumor masks as input, generated rapidly and automatically using a high precision tumor segmentation method.

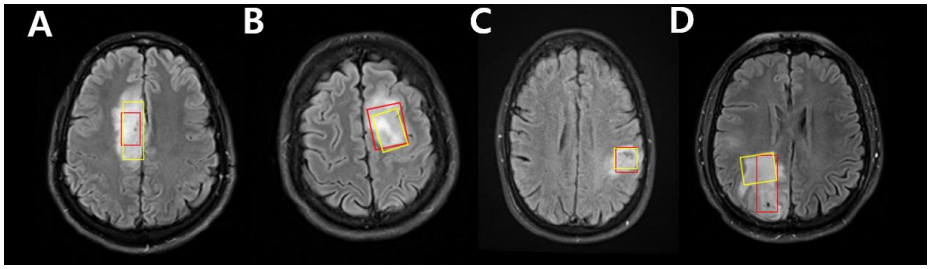

Fig. 1. Examples of experts' manual voxel placements (red and yellow boxes) overlaid on T_2-weighted FLAIR images.

2 Method

2.1 Data Collection and Processing

MR Images and Voxel Placement. MR images from 125 patients with gliomas were collected and anonymized for this study. Each case included standard clinical T_1-weighted and T_2-weighted FLAIR images, acquired at either 1.5 T (n=29) or 3 T (n=96) without an exogenous contrast agent. These cases were randomly distributed into five datasets of 25 cases. We recruited five experts (neuroradiologists or physicists with substantial clinical MRS expertise) to perform voxel placements. Each expert was randomly assigned two datasets of 25 cases, ensuring that each case was evaluated by two experts. Voxel placements were carried out using custom MRS localization software, referring to specific instructions: (i) position the voxel in the solid portion of the tumor; (ii) avoid black holes in the images (cysts, necrosis); (iii) rotate in one or two directions to better fit the voxel in the lesion; (iv) use a minimum volume of 6 mL. Examples of expert voxel placements are shown in **Figure 1**.

Data Preprocessing. Brain images were skull stripped using HD Brain Extraction [13], and whole tumor masks were created using a nnUNet model trained from BRATS 2016 and 2017 datasets [14-15] (train set, n = 484; test set, n = 266; test set dice metric = 0.90). All images were resampled to 128×128×128 to fit into our available computational capacity in model training.

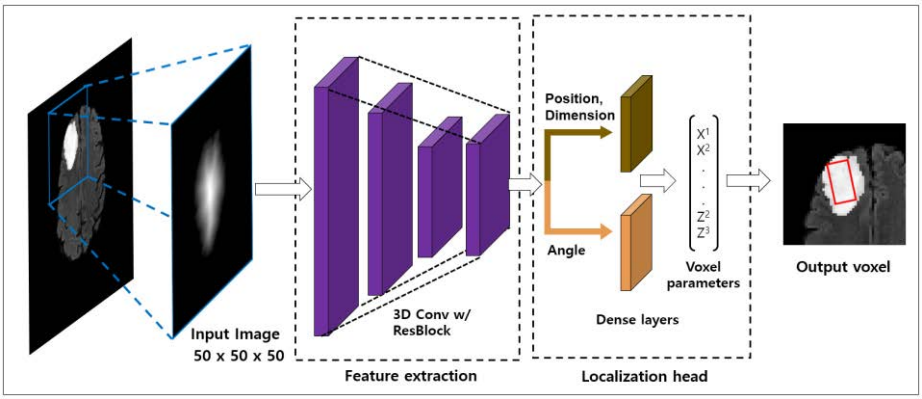

Fig. 2. Overview of deep learning model. 3D conv is an encoder and ResBlock is a residual block. Inputs to the model are tumor mask and Euclidean distance map of tumor mask.

2.2 Model Training and Testing

Model Description and Training. A voxel placement is completely described by the vector parameter $\theta = [x_1, x_2, x_3, \ell_1, \ell_2, \ell_3, \alpha_1, \alpha_2, \alpha_3]$, where x_i are the 3D spatial coordinates of the cuboid center, ℓ_i are the three dimensions of the cuboid, and α_i the

three Euler angles defining its rotation. Given a 3D binary tumor mask **M**, our task is to train a CNN f so that $\theta = f$ **(M),** which minimizes a loss function $\mathcal{L}$ **(**θ, $\boldsymbol{M}$**)** over the training set.

Our model simply consists of four 3D convolutional layers to extract spatial features, a residual block after each convolutional layer to preserve initial spatial features, and linear transformations with activation functions to predict the parameters for voxel placement (see **Figure 2** and **Supplementary Figure 1**). The tumor mask **M** was cropped to a 50 × 50 × 50 pixel volume, centered on the centroid of the original tumor mask, and converted to a scalar distance map **D** using a Euclidean distance transform to provide increased spatial information. Fivefold cross-validation was used to assess the performance of the model to compensate for the small dataset. Each fold had 100 cases for training and 25 for testing; the first fold was used as a model development set for hyperparameter adjustment. The model was trained for 60 epochs in all groups and the model from the last epoch was used for evaluation on each test set.

Training Loss Function. The loss function used to train the CNN used the weighted sum of square losses for each of the voxel parameters relative to the true (expert) values, plus two additional terms to encourage geometric consistency with the tumor mask. The total loss function was

$$\mathcal{L} = \lambda_{pos}\mathcal{L}_{pos} + \lambda_{dim}\mathcal{L}_{dim} + \lambda_{ori}\mathcal{L}_{ori} + (1 - f_{tumor}) + \mathcal{L}_{center}, \quad (1)$$

where $\mathcal{L}_{pos} = \sum_{i=1}^{3}(\hat{x}_i - x_i)^2$ is the sum of square errors between true and estimated position coordinates, and $\mathcal{L}_{dim}$ and $\mathcal{L}_{ori}$ are the sum of square errors for the size and orientation parameters. The function f_{tumor} $(\theta, \boldsymbol{M})$ is the fraction of the voxel contained within the tumor map **M**, and $\mathcal{L}_{center}$ is a binary value indicating the center of the cuboid is inside the tumor map:

$$\mathcal{L}_{center} = \begin{cases} 1, if\ \boldsymbol{M}[x_1, x_2, x_3] \neq 1 \\ 0, if\ \ \boldsymbol{M}[x_1, x_2, x_3] = 1, \end{cases} \quad (2)$$

We selected $\lambda_{pos} = 2, \lambda_{dim} = 3$, and $\lambda_{ori} = 3$ for all subsequent results based on initial trial and error.

Evaluation Metrics. In the test set, the dice similarity coefficient (DSC) was used to assess the spatial overlap of different voxels [16]. Since each image set had two expert voxel placements, the higher of the two DSC values was used for assessment. The f_{tumor} metric and total voxel volume were also used to evaluate performance. Tumor size and morphology were characterized by tumor volume. The tendency to rotate voxels was measured by the mean number of rotations per voxel; i.e., the mean count of $|\alpha_i| > 0$ per voxel. One-way ANOVA was used to compare the DSC, voxel volume and f_{tumor} among the different placements, and results were considered significant if P < 0.05. The intraclass correlation coefficient (ICC) was used to measure the variability of voxel volumes between $\mathcal{V}_M$ and $\mathcal{V}_{DL}$.

2.3 Related Work

A previous study of automated voxel placement in brain tumors using an objective function optimization was implemented for comparison [8]. That method maximized the objective function

$$F_{obj}(\theta, \boldsymbol{M}) = \exp\left(-\frac{1}{2}\left(\frac{V_{tumor}(\theta,\mathbf{M})-\mu_V}{\sigma_V}\right)^2\right)\exp\left(-\frac{1}{2}\left(\frac{f_{tumor}(\theta,\mathbf{M})-\mu_f}{\sigma_f}\right)^2\right), \quad (3)$$

where $V_{tumor}(\theta, \boldsymbol{M})$ is the intersection volume of the voxel and tumor mask $\mathbf{M}$, and $f_{tumor}(\theta, \mathbf{M})$ is as defined above. The tuning parameters (μ_V, σ_V) and (μ_f, σ_f) are the mean and standard deviation of the preferred $V_{tumor}(\theta, \boldsymbol{M})$ and $f_{tumor}(\theta, \mathbf{M})$ distributions. We used the same parameters as the previous work [8], but scaled the preferred voxel size and tumor fraction to match that of the expert placements in the model development set (μ_f = 0.95, σ_f = 0.05, μ_V = 12.9 mL and σ_V = 8.7 mL).

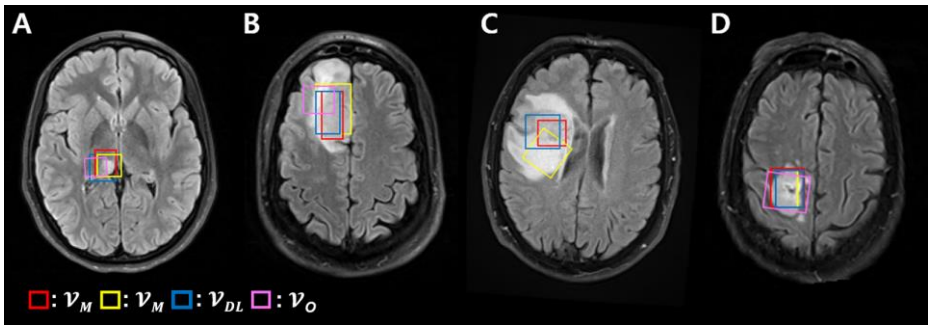

Fig. 3. Comparisons of $\mathcal{V}_M$ (red and yellow), $\mathcal{V}_{DL}$ (blue) and $\mathcal{V}_O$ (pink) overlaid on T_2w-FLAIR images. Note in (C) the objective function-based placement ($\mathcal{V}_O$) is in a different region of the tumor and does not intersect the selected plane.

3 Results

3.1 Qualitative Evaluation of Voxel Placements

Examples of voxel placements produced manually by MRS experts ($\mathcal{V}_M$), the proposed deep learning method ($\mathcal{V}_{DL}$) and the objective function method of Ref [8] ($\mathcal{V}_O$)are shown in **Figure 3**. The two manual voxel placements were generally similar, although in larger tumors the two placements sometimes diverged, sampling different regions of the tumor. **Figure 3C** shows an example of divergent manual placements in a large tumor. In general, it appeared that the deep learning method produced voxel placements comparable to that of at least one of the manual placements, with similar size, f_{tumor}, and alignment with lesion morphology. The prior objective function method also provided reasonable voxel placements, but voxel was more often rotated and had different orientation than manually placed voxels (example in **Figure 3D**).

3.2 Quantitative Evaluation of Voxel Placements

Overall Performance. Prediction of voxel parameters from tumor masks was rapid, requiring an average of 3 ms for $\mathcal{V}_{DL}$, while $\mathcal{V}_O$ averaged > 20 s. Both methods required skull stripping and tumor mask generation (averaging 10 s and 10 s respectively). The summary performance metrics for the three methods are shown in Table 1. Experts generated voxels containing a high portion of tumor (93.2%), but with a low average dice score between experts of 0.46. $\mathcal{V}_{DL}$ produced voxel placements similar to one of the two manual placements in each case, with a dice of 0.47 with the closest placement and a lower 0.39 to the further placement while the objective function method $\mathcal{V}_O$ showed lower dice scores than $\mathcal{V}_{DL}$. Both $\mathcal{V}_{DL}$ and $\mathcal{V}_O$ showed a greater tendency to rotate voxels than the experts. Fold-wise performance metrics for $\mathcal{V}_{DL}$ are provided in **Supplementary Table 1**.

The differences in voxel volumes are shown in the Bland-Altman plots of **Figure 4**. **Figure 4A** plots the differences between the two manual placements, which shows a low bias and a trend of higher agreement for smaller voxels. **Figure 4B** compares the $\mathcal{V}_{DL}$ volumes to the $\mathcal{V}_M$ volumes, showing the same trends (low bias, better agreement with small voxels) as seen in the inter-expert analysis of **Figure 4A**. The interclass correlation coefficients (ICC) for volumes with different placements were ICC $(\mathcal{V}_{M1}, \mathcal{V}_{M2})$ = 0.298, ICC $(\mathcal{V}_{DL}, \mathcal{V}_{M1})$ = 0.275, ICC $(\mathcal{V}_{DL}, \mathcal{V}_{M2})$ = 0.362, ICC $(\mathcal{V}_O, \mathcal{V}_{M1})$ = 0.102 and ICC $(\mathcal{V}_O, \mathcal{V}_{M1})$ = 0.102 suggesting that the similarity between $\mathcal{V}_{DL}$ and $\mathcal{V}_M$ is comparable to inter-expert variation. The f_{tumor} values were also similar between manual and deep learning placements, as shown in **Table 1** and **Figure 5A**.

Table 1. Comparisons between manual, deep learning-based, and objective function-based voxel placements. Values for experts are the mean of two experts per case. Values are mean ± S.D. unless stated otherwise.

Method	f_{tumor} (%)	Voxel Volume (mL)	Dice (w/ closest $\mathcal{V}_M$) / Dice (w/ two $\mathcal{V}_M$)	# of rotated axes
Experts	93.2 ± 11.5	12.38 ± 6.63	N/A / 0.46 ± 0.22	1.22 ± 1.02
Deep learning	91.6 ± 14.1	12.05 ± 4.08	0.47 ± 0.21 / 0.39 ± 0.23	1.44 ± 0.88
Objective function	92.6 ± 10.4	12.34 ± 3.31	0.40 ± 0.21 / 0.34 ± 0.23	1.48 ± 0.99

Dependence on Tumor Size. Voxel placement metrics were seen to depend on tumor size, as shown in **Figure 5.** As the tumor volume increased, f_{tumor} ($\mathcal{V}_M$, r = 0.31; $\mathcal{V}_{DL}$, r = 0.36) and voxel volume ($\mathcal{V}_M$, r = 0.50; $\mathcal{V}_{DL}$, r = 0.49) also increased. In contrast the dice score between placements was lower in larger tumors (**Fig. 5C**). Together these plots show that the $\mathcal{V}_{DL}$ placements have similar characteristics as the experts' placements, and show greater agreement in smaller voxels. The volume of tumor in the voxel (V_{tumor}) also increased as the f_{tumor} increased. **Supplementary Figure 2** shows that the volume of $\mathcal{V}_O$ is less dependent on tumor volume ($\mathcal{V}_M$, r = 0.50; $\mathcal{V}_{DL}$, r = 0.18) and has systematic rule-based trends in its voxel volume changes.

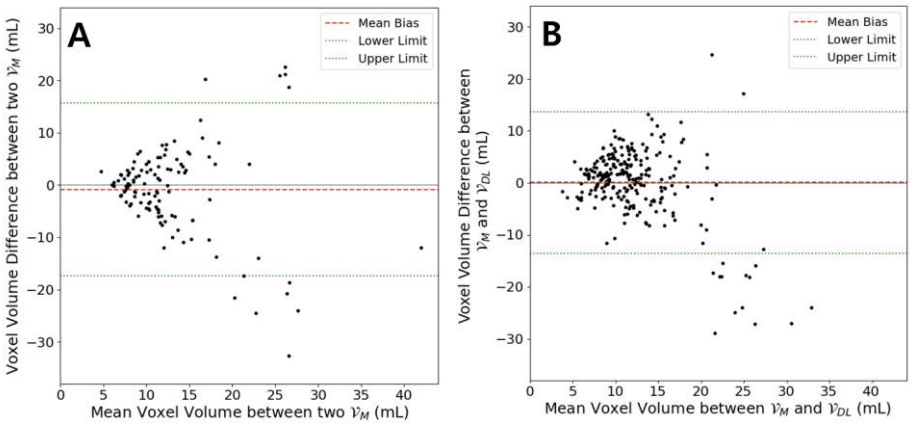

Fig. 4. Bland-Altman plot of voxel volume difference of (A) two $\mathcal{V}_M$ and (B) $\mathcal{V}_M$ and $\mathcal{V}_{DL}$. (A) Mean bias = -0.84 mL, lower limit = -17.34 mL and upper limit = 15.65 mL. (B) Mean bias = 0.06 mL, lower limit = -13.57 mL and upper limit = 13.69 mL.

4 Discussion and Conclusion

In this study, we proposed a deep learning-based method for MRS voxel placement in malignant brain tumors. The need for manual voxel placement in single-voxel MRS has been a major limitation for clinical adoption as it requires expertise to size, position, and orient the voxel while rapidly interpreting the anatomy of normal tissues, cerebrospinal fluid, and tumors. An automated voxel placement method must be able to generate a voxel placement fast enough with characteristics (volume, f_{tumor}, etc.) comparable to that of the experts if it is to be useful in a clinical practice.

However, visual inspection of the previous objective function-based voxel placement method [8] revealed limitations in precisely locating the voxel where the experts would place them, especially for large tumors. In small tumors, the prior objective function method showed good performance in matching the voxel to lesion, but placements did not always reflect experts' tendencies. Compared to that method, experts produced larger voxels and were less likely to rotate the voxel, possibly motivated by concerns about signal-to-noise ratio or potential artifacts due to multiply-rotated voxels.

Quantitative analysis revealed discrepancies between volumes of objective function-based voxels ($\mathcal{V}_O$) and manual voxels ($\mathcal{V}_M$), with $\mathcal{V}_O$ volumes less dependent on tumor size. This disparity could result in important signals being missed in measurements when voxel volumes are insufficient. The deep learning approach addressed these challenges, achieving a high dice and f_{tumor} comparable for $\mathcal{V}_M$. $\mathcal{V}_{DL}$ not only matched $\mathcal{V}_M$ in voxel volume, but also reflected tumor morphology in voxel dimension ratios, as observed during visual inspection. Importantly, $\mathcal{V}_{DL}$ demonstrated high consistency in orientation control, aligning with $\mathcal{V}_M$ orientation trends based on tumor size, crucial for maintaining MRS measurement quality. Moreover, $\mathcal{V}_{DL}$ exhibited similar distributions

of voxel volume and f_{tumor} compared to $\mathcal{V}_M$ from experts with a small mean bias difference and stayed within the variance observed among experts. $\mathcal{V}_{DL}$ was also generated fast enough within a second compared to the previous method of $\mathcal{V}_O$, which required approximately half a minute for optimization.

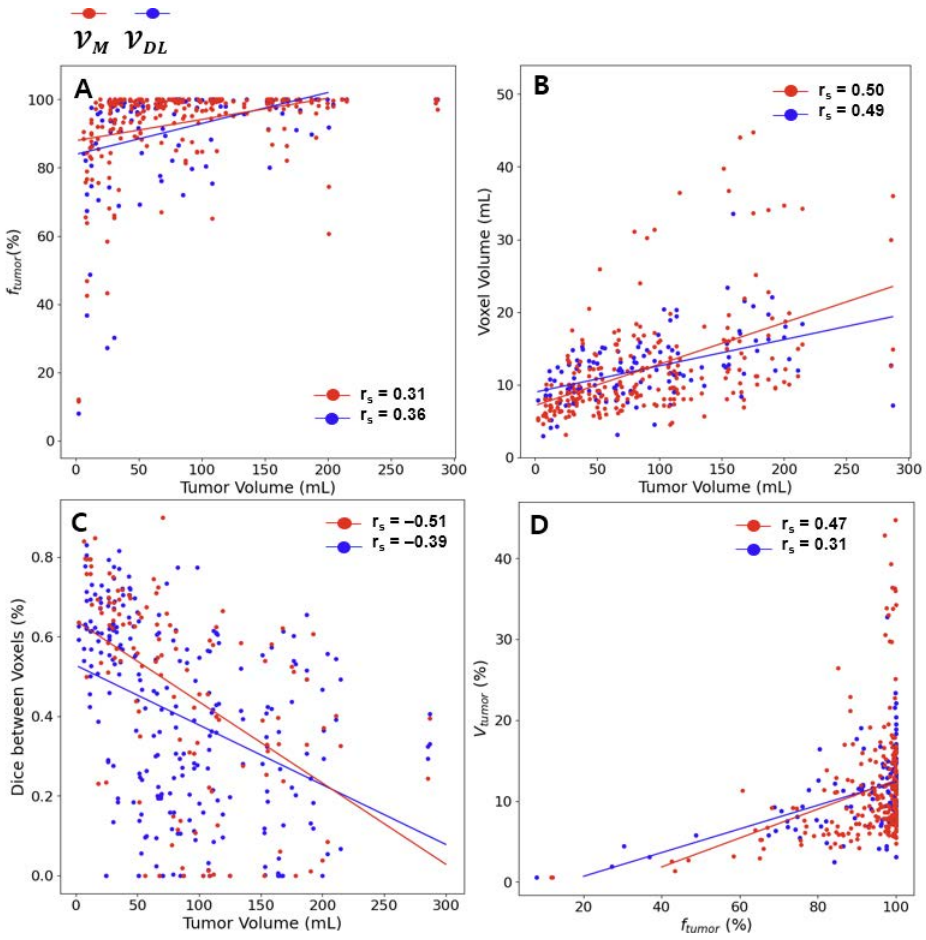

Fig. 5. Correlation plots between tumor volume and voxel placement characteristics, f_{tumor}, and V_{tumor} for $\mathcal{V}_M$ and $\mathcal{V}_{DL}$.

It is also worthwhile to mention that our study utilized a specifically curated dataset tailored for this research, with $\mathcal{V}_{DL}$ trained on meticulously controlled voxel placements. This contrasts with routine voxel placements in clinical practice, which are often performed quickly to maintain workflow efficiency. Thus, our model benefited from the experts' precise control over voxel centers, dimensions, and orientations. Our model can be implemented in the scanner by acquiring T_2w-FLAIR and T_1w scans and automatically sending them to a dedicated DICOM receiver. Image segmentation and voxel

placement will be performed automatically, and coordinates will be exported to a text file on a shared drive, where it will be imported by a custom MRS pulse sequence.

There are several limitations in our study that can be improved. First, the voxel placement dataset was small and increasing the cases may be helpful to improve performance and better evaluate the effect of tumor size and morphology. Second, including additional information as input to the model, such as masks of normal anatomy (cerebrospinal fluid, skull) and regions to avoid (necrosis, cysts), could lead to better matching of experts' performance, as their placements may consider these regions. Third, using MR images acquired after gadolinium injection may provide more vascular information such as tumor core about spatial relationship between the voxels and tumors to the model.

Acknowledgments. Authors would like to thank Sarah Bedell, Noam Harel, Ph.D., and Henry Braun, Ph.D. for help with the de-identification of images. This work was sponsored by the NIH grants: U01CA269110, R01EB034231, and P41EB027061. TBN was sponsored by Bayer HealthCare/RSNA Research Seed Grant and Cancer Research Society. FB acknowledges support from Investissements d'avenir [grant number ANR-10-IAIHU-06 and ANR-11-INBS-0006] and from Agence Nationale de la Recherche [grant number ANR-20-CE17-0002-01].

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Choi C, Raisanen JM, Ganji SK, et al. Prospective longitudinal analysis of 2-hydroxyglutarate magnetic resonance spectroscopy identifies broad clinical utility for the management of patients with *IDH* -mutant glioma. Journal of Clinical Oncology. 2016; 34(33): 4030-4039
2. Tanaka K, Sasayama T, Mizukawa K, et al. Combined IDH1 mutation and MGMT methylation status on long-term survival of patients with cerebral low-grade glioma. Clin Neurol Neurosurg. 2015; 138: 37-44
3. Yah H, Parsons DW, Jin G, et al. IDH1 and IDH2 Mutations in Gliomas. N Engl J Med. 2009; 19;360(8): 765-73
4. Choi C, Ganji SK, DeBerardinis RJ, et al. 2-hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated glioma patients. Nat Med. 2012; 18(4): 624-9
5. Andronesi OC, Kim GS, Gerstner E, et al. Detection of 2-hydroxyglutarate in IDH-mutated glioma patients by in vivo spectral-editing and 2D correlation magnetic resonance spectroscopy. Sci Transl Med. 2012;4(116): 116
6. Pope WB, Prins RM, Albert TM, et al. Noninvasive detection of 2-hydroxyglutarate and other metabolites in IDH1 mutant glioma patients using magnetic resonance spectroscopy. J Neuro Oncol 2012; 107: 197–205.
7. Toh CH, Castillo M, Wei KC, Chen PY. MRS as an Aid to Diagnose Malignant Transformation in Low-Grade Gliomas with Increasing Contrast Enhancement. AJNR Am J Neuroradiol. 2020 Sep;41(9):1592-1598.
8. Bolan PJ, Branzoli F, Di Stefano AL, et al. Automated Acquisition Planning for Magnetic Resonance Spectroscopy in Brain Cancer. Medical Image Computing and Computer-Assisted Intervention. 2020; 12267, 730.

9. Lee S, Branzoli F, Nguyen T, et al. Analysis of MRS voxel placements in brain tumors performed by MRS experts. International Society of Magnetic Resonance in Medicine. 2024 April 19; 1545-4428
10. Bishop JH, Geoly A, Khan N, Tischler C, Krueger R, Keshava P, Amin H, Baltusis L, Wu H, Spiegel D, Williams N, Sacchet MD. Real-Time Semi-Automated and Automated Voxel Placement using fMRI Targets for Repeated Acquisition Magnetic Resonance Spectroscopy. J Neurosci Methods. 2023 May 15;392:109853.
11. Dou W, Speck O, Benner T, Kaufmann J, Li M, Zhong K, Walter M. Automatic voxel positioning for MRS at 7 T. MAGMA. 2015 Jun;28(3):259-70.
12. Lee H, Caparelli E, Li H, Mandal A, Smith SD, Zhang S, Bilfinger TV, Benveniste H. Computerized MRS voxel registration and partial volume effects in single voxel 1H-MRS. Magn Reson Imaging. 2013 Sep;31(7):1197-205.
13. Yamashita, R., Nishio, M., Do, R.K.G. *et al.* Convolutional neural networks: an overview and application in radiology. *Insights Imaging* **9**, 611–629 (2018). Isensee F, Schell M, Tursunova I, et al. Automated brain extraction of multi-sequence MRI using artificial neural networks. Hum Brain Mapp. 2019; 1–13
14. Isensee F, Jaeger, PF, Kohl, SA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods. 2021; 18(2), 203-211
15. Menze BH, Jakab A, Bauer S, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEE Trans Med Imaging. 2015; 34(10): 1993-224
16. Zou KH, Warfield SK, Bharatha A, et al. Statistical Validation of Image Segmentation Quality Based on a Spatial Overlap Index. Academic Radiology. 2004; 11(2): 178-189