Duman A
Cardiff University
School of Engineering

Powell J
Velindre NHS Trust
Department of Oncology, Cardiff

Thomas S
Velindre NHS Trust
Department of Oncology, Cardiff

Sun X
Cardiff University
School of Computer Science and Informatics

Spezi E
Cardiff University
School of Engineering

ENGINEERING FOR HEALTH

Generalizability of Deep Learning Models on Brain Tumour Segmentation

Brain tumour segmentation is a hard and time-consuming task to be conducted in the process of radiotherapy planning. Deep Learning (DL) applications have a significant improvement in image segmentation tasks. In this work, we apply DL models such as 2D and 2.5D U-NET to the segmentation task of a brain tumour on the BraTS 2021 dataset and our local dataset. The 2.5D network is a modified version of 2D U-NET by using three slices as an input for each magnetic resonance imaging (MRI) sequence. We achieve the best segmentation results with 2.5D U-NET on BraTS with Dice scores of 86.97%, 91.27% and 94.42% for enhancing tumour, tumour core and whole tumour respectively. On the other hand, our best segmentation result of the GTV delineation on the local dataset is a Dice score of 78.51% for 2D U-NET. Although the result of GTV contours is not improved by 2.5D for the local dataset due to non-fixed voxel size, the Dice scores of ET, TC and WT are improved by the proposed 2.5D U-NET for the BraTS dataset.

Keywords:
Glioblastoma, autosegmentation, brain tumour segmentation, radiotherapy planning.

Corresponding author:
DumanA@cardiff.ac.uk

A. Duman, J. Powell, S. Thomas, X. Sun, and E. Spezi, 'Generalizability of Deep Learning Models on Brain Tumour Segmentation', *Proceedings of the Cardiff University Engineering Research Conference 2023*, Cardiff, UK, pp. 3-5.

doi.org/10.18573/conf1.b

INTRODUCTION

Brain tumours are one of the most lethal types of cancer [1]. Glioblastoma (GBM) is a aggressive, rapidly developing and fatal type of glioma that originates from glial cells. The median survival time for GBM patients is 15 months after diagnosis [2]. Brain tumour segmentation is the process of separating the tumour from healthy brain tissue. Brain tumour segmentation is a challenging task due to the heterogeneous nature of the tumour tissue [3] and Magnetic Resonance Imaging (MRI) is a commonly used imaging technique in brain because it provides high-quality, high-contrast and detailed images of soft tissues. Accurate tumour segmentation is critical for diagnosis and therapy planning [4]. In radiotherapy, the optimal delivery of high-dose radiation to the tumour while sparing healthy tissues depends on the accuracy of segmentation. Manual segmentation is time-consuming, subjective, and non-reproducible. Automated segmentation is proposed to decrease labour-intensive work, provides objective and reproducible results [5]. Deep Learning (DL) models are one of solutions developed to segment tumours automatically.

In clinical practice the poor generalizability of DL models is a major barrier [6]. Particularly for medical image segmentation tasks, models have demonstrated good performance when using uniform datasets, but their ability to generalise to new and unseen data remains a challenge. In brain tumour segmentation, the generalizability of DL models needs to be evaluated for different datasets consisting of MRI scans with variable settings such as resolution (pixel spacing and slice thickness) and matrix size. Such validation is essential for enhancing the robustness and usefulness of DL models in clinical applications. Ideally, models with high generalizability could be developed when training datasets contain a significant number of high-quality images from different centres using variety of imaging settings.

Despite high number of publications dealing with DL-based tasks including segmentation, due to generalizability as a primary issue, the models transferred to clinic are very few [7]. In this study, we investigated the generalizability of DL models with different configurations on GBM segmentation task across different datasets.

MATERIALS AND METHODS

In this study we utilised a large dataset of publicly accessible MRI scans and reference segmentations (BraTS 2021) and a local dataset. GBMs are segmented according to the BraTS challenge as enhancing tumour (ET), tumour core (TC), and entire tumour (WT) regions [8]. The BraTS 2021 dataset has a fixed voxel size (1x1x1mm) and fixed matrix size (240x240x155). However, clinical datasets including our local glioma scans have variable voxel sizes and matrix sizes for each MRI sequence.

In this research, we used U-NET, a convolutional neural network architecture designed for biomedical image segmentation tasks. The U-NET architecture consists of an encoder network, which down samples the input image extracting salient features, and a decoder network, which up samples the feature maps back to the original image resolution and generates the segmentation mask. This arrangement gives the network its characteristics U shape [9]. Tumour segmentation was carried out using a region-focused selection (RFS) method [10] that combines several single 2D U-NET and 2.5D U-NET architectures trained on BraTS tumour regions and validated with Gross Tumour Volumes (GTV) of the local dataset. Fig. 1 shows 4 image sequences namely: FLAIR, T1, T1ce and T2. The BraTS and the local dataset cover 1251 patients and 53 patients respectively.

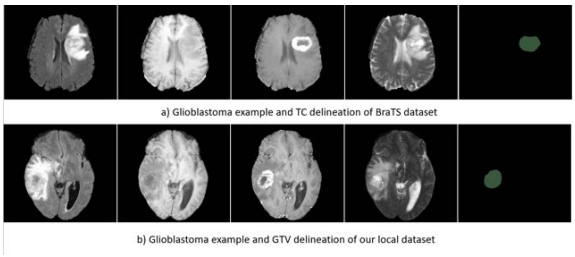

Fig. 1. Examples in BraTS 2021 (Fig. 1a) and our local dataset (Fig. 1b). From left to right; FLAIR, T1, T1ce, T2 sequences and their corresponding segmentation masks.

In the 2.5D model, for each sequence we used 3 channels including: (a) current slice (n), (b) previous slice (n-1) and (c) following slice(n+1). This is shown in Fig. 2. Due to the local dataset having different voxel and matrix sizes for each patient, each scan of the local dataset was resampled to match the BraTS 2021 data format and specifications. Each channel used a 2D image of size (240x240).

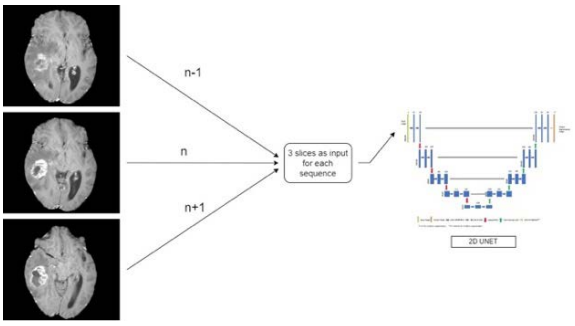

Fig. 2. A 3-slice example of the T1ce sequence for the local dataset.

The RFS method use three types of U-NET models for image segmentation, i.e., Binary class, Multi-label, and Multiclass models. In the case of binary-class models, three separate binary-class models were used for segmenting three regions (ET, TC, and WT), separately. The input and output of each binary-class model corresponded to only one of the regions. In both cases of multilabel and multiclass models, only a single model was used for the segmentation of all the three regions. The difference between the multilabel and multiclass models is that the former uses the overlapping class masks among ET, TC and WT, while the latter takes non-overlapping class masks. Both binary and multilabel models used the sigmoid function, while the multiclass model used the softmax function at the last layer of the U-NET.
The z-score technique was used to normalise images. Multiclass, Multi-label, and Binary class single 2D and 2.5D U-NET architectures were trained on individual tumour regions with the BraTS 2021 dataset. The Dice Similarity Coefficient (DSC) was utilised to assess the similarity between the contours generated by the DL models and the reference contours.

The trained models were then combined with a union RFS (u-RFS) model. The u-RFS model was used to improve the DL-based GTV segmentation for both 2D and 2.5D models.

RESULTS

For our local dataset, the comparison of 2D U-NET and 2.5D U-NET predictions is shown in Fig. 3.

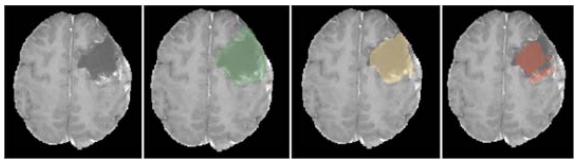

Fig. 3. GTV segmentation of the local dataset. From left to right; T1ce sequence, reference contour, 2D U-NET prediction and 2.5D U-NET prediction for.

In Table 1, the results show that 2.5D U-NET models outperform 2D U-NET models for each DL method and for both ET and WT regions.

DL method	ET(%) (2D)	ET(%) (2.5D)	WT(%) (2D)	WT(%) (2.5D)
Multiclass	84.99	86.76	91.65	93.81
Multilabel	82.29	84.01	92.24	94.42
Binary class	85.19	86.97	92.18	94.35

Table 1. DSC results of ET and WT (BraTS dataset).

In Table 2, the results show that 2.5D U-NET models outperform 2D U-NET models for the TC region. However, the 2D model outperforms the 2.5D model for the GTV region. u-RFS provided the best DSC score with 78.51% when applied to the 2D model.

DL method	TC(%) (2D)	TC(%) (2.5D)	GTV(%) (2D)	GTV(%) (2.5D)
Multiclass	89.71	91.27	78.43	70.35
Multilabel	87.27	88.78	77.91	69.88
Binary class	89.48	91.03	78.22	70.16
u-RFS		78.51	70.42	

Table 2. DSC results for TC (BraTS dataset) and for GTV (local dataset).

DISCUSSION AND CONCLUSIONS

The results presented in this work show that 2.5D models outperform 2D models for the segmentation of GBM on the BraTS dataset. This is because of the additional information provided by the imaging data feeding into the 3 channels used in the model equating to 12 channels in total compared to 4 channels used in the 2D model. However, the 2.5D model did not generalise well when applied to our local dataset that included scans with different voxel and matrix sizes compared to the BraTS 2021 dataset.

If only one model will be developed, the RFS method can be utilised for the selection of the best method when considering only one region. The u-RFS method proved to be useful in improving the performance of both 2D and 2.5D models. The u-RFS can be applied to any DL model to increase segmentation results.

Acknowledgments
This work was sponsored by the Turkish Ministry of National Education.

Conflicts of interest
The authors have no conflict of interest to declare.

REFERENCES

[1] L. M. DeAngelis, 'Brain Tumors', *N Engl J Me*, vol. 344, no. 2, pp. 114–123, 2001. doi.org/10.1056/nejm200101113440207

[2] R. L. Siegel, K. D. Miller, and A. Jemal, 'Cancer statistics, 2019', *CA Cancer J Clin*, vol. 69, no. 1, pp. 7–34, Jan. 2019. doi.org/10.3322/CAAC.21551

[3] S. Bakas *et al*., 'Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features', *Sci Data*, vol. 4, no. 1, p. 170117, 2017. doi.org/10.1038/sdata.2017.117

[4] P. Kickingereder *et al*., 'Automated quantitative tumour response assessment of MRI in neuro-oncology with artificial neural networks: a multicentre, retrospective study', *Lancet Oncol*, vol. 20, no. 5, pp. 728–740, May 2019. doi.org/10.1016/S1470-2045(19)30098-1

[5] G. P. Mazzara, R. P. Velthuizen, J. L. Pearlman, H. M. Greenberg, and H. Wagner, 'Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation', *Int J Radiat Oncol Biol Phys*, vol. 59, no. 1, pp. 300–312, 2004. doi.org/10.1016/j.ijrobp.2004.01.026

[6] K. Yasaka and O. Abe, 'Deep learning and artificial intelligence in radiology: Current applications and future directions', *PLoS Med*, vol. 15, no. 11, p. e1002707, 2018. doi.org/10.1016/j.ijrobp.2004.01.026

[7] C. J. Kelly, A. Karthikesalingam, M. Suleyman, G. Corrado, and D. King, 'Key challenges for delivering clinical impact with artificial intelligence', *BMC Med*, vol. 17, no. 1, p. 195, 2019. doi.org/10.1186/s12916-019-1426-2

[8] B. H. Menze *et al*., 'The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)', *IEEE Trans Med Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015. doi.org/10.1109/TMI.2014.2377694

[9] O. Ronneberger, P. Fischer, and T. Brox, 'U-net: Convolutional networks for biomedical image segmentation', in *International Conference on Medical image computing and computer-assisted intervention, 2015*, pp. 234–241. doi.org/10.1007/978-3-319-24574-4_28

[10] A. Duman, P. Whybra, J. Powell, S. Thomas, X. Sun, and E. Spezi, 'Transferability of deep learning models to the segmentation of gross tumour volume in brain cancer', in *ESTRO 2023 Conference* (Accepted), 2023.

Cynhadledd Ymchwil Peirianneg Caerdydd
Cardiff Engineering Research Conference

E. Spezi and M. Bray (eds.) 2024. *Proceedings of the Cardiff University Engineering Research Conference 2023.* Cardiff: Cardiff University Press.
doi.org/10.18573/conf1

Cardiff University Engineering Research Conference 2023 was organised by the School of Engineering and held from 12 to 14 July 2023 at Cardiff University.

The work presented in these proceedings has been peer reviewed and approved by the conference organisers and associated scientific committee to ensure high academic standards have been met.

Cardiff University Press | Gwasg Prifysgol Caerdydd

First published 2024

Cardiff University Press
Cardiff University, PO Box 430
1st Floor, 30-36 Newport Road
Cardiff CF24 0DE

cardiffuniversitypress.org

Editorial design and layout by
Academic Visual Communication

ISBN: 978-1-9116-5349-3 (PDF)