



A low resource 3D U-Net based deep learning model for medical image analysis

Girija Chetty¹ · Mohammad Yamin² · Matthew White³

Received: 21 September 2021 / Accepted: 11 December 2021 / Published online: 5 January 2022

This is a U.S. government work and not under copyright protection in the U.S.; foreign copyright protection may apply 2022

Abstract The success of deep learning, a subfield of Artificial Intelligence technologies in the field of image analysis and computer can be leveraged for building better decision support systems for clinical radiological settings. Detecting and segmenting tumorous tissues in brain region using deep learning and artificial intelligence is one such scenario, where radiologists can benefit from the computer based second opinion or decision support, for detecting the severity of disease, and survival of the subject with an accurate and timely clinical diagnosis. Gliomas are the aggressive form of brain tumors having irregular shape and ambiguous boundaries, making them one of the hardest tumors to detect, and often require a combined analysis of different types of radiological scans to make an accurate detection. In this paper, we present a fully automatic deep learning method for brain tumor segmentation in multimodal multi-contrast magnetic resonance image scans. The proposed approach is based on light weight UNET architecture, consisting of a multimodal CNN encoder-decoder based computational model. Using the publicly available Brain Tumor Segmentation (BraTS) Challenge 2018 dataset, available from the Medical Image Computing and Computer Assisted Intervention (MICCAI) society, our

novel approach based on proposed light-weight UNet model, with no data augmentation requirements and without use of heavy computational resources, has resulted in an improved performance, as compared to the previous models in the challenge task that used heavy computational architectures and resources and with different data augmentation approaches. This makes the model proposed in this work more suitable for remote, extreme and low resource health care settings.

Keywords AI · Fusion · Deep learning · Multimodal · Medical · Segmentation

1 Introduction

Segmenting brain tumours automatically from 3D magnetic resonance images (MRIs) is necessary for diagnosis, monitoring and treatment planning of the disease. Manual segmentation and delineation methods done in clinical settings require expert anatomical knowledge, and are time consuming, expensive and prone to human errors. Automatic computer based semantic segmentation approached for tumor subregion segmentation from 3D MRIs based on deep learning architectures can lead to availability of decision support tools that can help alleviate the manual and laborious task of traditional segmentation approaches in clinical settings, allowing radiologist to focus on more important tasks of treatment planning and interventions for the patients. Magnetic resonance imaging (MRI) is one of the most efficient radiology scan technique for detecting brain lesions and tumours, as it is a non-invasive detection technique, and when used in conjunction with other sensor modalities, such as computer tomography (CT), and positron emission tomography (PET), can provide better

✉ Mohammad Yamin
myamin@kau.edu.sa

Girija Chetty
girija.chetty@canberra.edu.au

Matthew White
matthewwhite360@gmail.com

¹ Faculty of SciTech, University of Canberra, Canberra, Australia

² King AbdulAziz University, Jeddah, Saudi Arabia

³ IIG R & D Pty. Ltd, Melbourne, Australia

understanding of the lesion or tumour structure in the brain. However, using several of these modalities concurrently could be more expensive, and in some cases invasive (PET for example). Therefore different MRI coils such T1, T1ce (contrast enhanced), T2 and FLAIR (all being variants of magnetic resonance imaging), can serve better for providing concurrent multimodal radiological imaging information for analysis the deep and complex structure of the lesions and tumours in the brain. These modalities (T1, T1ce, T2 and FLAIR) capture brain images with varying intensities, and can show different tissue contrasts through different pulse sequences, and allow better visualization of the regions of interest in the human brain. If some or all of these MRI modalities (T1, T1ce, T2 and FLAIR) are combined to produce multi-modal images, they can provide better information about irregular shaped tumors and lesions, which would have been difficult to localize with a single modality. This multi-modal data, with modalities including T1-weighted MRI (T1), T1-weighted MRI with contrast enhancement (T1ce), T2-weighted MRI (T2) and T2-weighted MRI with fluid attenuated inversion recovery (T2-Flair) contain rich information for segmenting complex irregular shaped structures and detect benign and malignant tumours, and their severity, leading to improved diagnosis in clinical settings. Rest of the paper is organized as follows. Next section presents the background and related work, and details of the proposed multimodal CNN based 3D U-Net deep learning architecture in presented in Section three. Experiment details are provided in section four along with outcomes. The research presented in this paper is concluded in section five. The conclusions include a plan for further research.

2 Background and related work

One of the most promising industry currently for implementing revolutionary data science solutions is medicine and healthcare. Recently data science, machine learning and AI approaches based on deep learning, radiological imaging and natural language processing are growing very fast. The area grew at an astronomical pace during the Covid-19 pandemic, due the complex challenges associated with disease, particularly the fast spread of different variants, low resource settings and availability of trained workforce, and lack of efficient computer based decision support technologies to assist the physicians. Recently, several research works to address this short coming have been addressed by several researchers [1–7]. As an old adage goes, prevention is better than cure, and health care systems based on some of these research works based on AI can help in both prevention and cure, due to the ready availability of these AI based health technologies. Having

this support can enable health care professionals to focus on disease management and leave the job of running the mathematical algorithms to AI. Also, for managing long chronic side effects of Covid-19, and co-morbidities associated with aggressive treatment regimens used during pandemic. All this could be easily possible with AI as an assistive technology, and the doctors could then focus on working with the patients and focus on disease management and control the after effects and side effects of aggressive treatment regimen used to control the deadly virus.

Medical images form a rich source of data for understanding the disease complexity, and when captured with several radiological sensors, can contain a wealth of information that can explain the patient health and disease status. Medical images are often complex data sources and require experts to unlock the mysterious information embedded in the images, and to differentiate the healthy tissue from diseased tissue. The first step is usually to segment, or trace, important structures. Segmentation is the most important step in medical image analysis, and one of the biggest challenges facing researchers in the field of medical imagin—and data science in general—how do we define what is a “true” segmentation? It may appear as a simple problem on the surface, but getting a ground truth, or expert labels/annotation on which part of the tissue is health and which part is diseased is easier said than done. This is because, firstly, doctors are incredibly busy. Getting an expert radiologist to trace out a few hundred (or even a few dozen) scans is a big task. Secondly, if there are multiple radiologists working on the same project, their opinions differ. This is where AI can come to the rescue. If we can build an AI based assistant, which is built using the novel data science and machine learning algorithms and the combined knowledge of several experts as the ground truth, it can be used as a support tool by anyone in the health care settings. The grand challenge tasks along with large data sets along with ground truths currently being facilitated by MICCAI society [12], is one of the ways that Artificial Intelligence (AI) can help provide radiology support to doctors and physicians in enhancing the state of the are in AI enabled health care.

This article proposes a novel method for automatic segmentation of tumorous tissue regions in the brain is proposed, using a multimodal CNN based 3D U-Net deep learning architecture, for segmentation of necrotic tumours (NCR), enhancing and non-enhancing tumor (ET/NET), edematous tissue (ET) and the whole tumor (WT) region. The performance of the proposed approach was evaluated on a publicly available benchmark MICCAI Brain Tumor Segmentation (BraTS) 2018 challenge dataset [12] and has resulted in exciting outcomes relevant to tumorous tissue segmentation performance, when compared to other AI

based approaches using the same dataset and same task, as reported in challenge competition [8–10]. While the method reported in the top performing method, and challenge winner, was based on an encoder-decoder architecture for tumor subregion segmentation from 3D MRIs, and used an approach based on variational auto-encoder strategy to address the limited training data size. The second place winning solution [8], used a generic U-net based architecture with autoencoder regularization, and showed that it is enough to achieve the competitive performance. However, in this method additional training data sourced from their own institution was used by the authors (not publicly available). The third place winning model [9], was based on DenseNet architecture [11] with dilated convolutions embedded in a U-net-like network. Another third place winning method [10], was based on multi-scale context information modelling approach with an ensemble of different networks, and involves cascade segmentation of three tumor subregions, and an attention block with shared backbone weights. While each of these top performing methods are based on computationally heavy architectures requiring data augmentation, additional private data or ensemble of several networks, we propose a lightweight computation model based on minimalist UNET architecture, and rely on leveraging the complementarity and full multimodality of heterogenous data sources, in terms of using all four modalities T1, T1Gd, T2 and FLAIR, and achieved an improved performance without any additional data augmentation, computational heavy architectures or ensemble models, or use of private (not publicly available data sources).

Compared to the related works and previously proposed models, our light-weight UNET model uses all MRI modalities, with flattening of 3D volumes into two dimensional images cropped to 128×128 , and a patch size ranging from 64×64 at the first CNN and batch normalization stage through to 1024×1024 patches at the 5th CNN stage of the encoder leg in the UNET model. This is followed by a stack of 5 deconvolution stages and concatenation with encode part to complete the decoder leg of the UNET model. This light-weight architecture results in the improved model performance. Also, there was no additional training data used for model building, and only the provided training set was used, nor did we use any data augmentation, a traditional practice in most of the previously proposed deep learning models, making it an energy efficient, low footprint deep learning model, suitable for low resource, extreme environment settings, and can be deployed on mobile, edge devices.

3 Proposed multimodal deep learning scheme

Our approach is based on fusion of multiple MRI modalities, including FLAIR, T1, TGd, and T2 MRI modalities and a novel light-weight deep learning architecture, for multi-class segmentation of different tumor tissues including necrotic tumor region (NCR), non-enhancing and enhancing tumor regions (ET/NET), Edematous tissue (ED) and whole tumor tissue (WT), using decomposition of 3D volumes into one dimensional tensors, being propagated between CNN layers, making them computationally light in the feature extraction and for processing of multiple input modalities, and segmenting the regions corresponding to each class. Figure 1 shows the method, and the proposed low-resource and light weight CNN based 3D U-Net deep learning model.

The proposed CNN based 3D U-Net architecture consists of a 5×5 down sampling and an up-sampling processing stack, as in the traditional 2D U-Net architecture [7]. We resized the 3D input image volumes into 128×128 size image vectors from all four modalities. The ReLU activation was used for all layers, except for the final layer, where sigmoid activation was used, and a batch normalization step was used for regularization after each convolution layer [8]. For training the CNN 3D U-Net model, batch size of 64, and a learning rate of 10^{-4} for the adaptive moment estimation (ADAM) optimizer was used, with dice coefficient loss minimization as the optimization objective, and dice-coefficient as the performance metric for assessing the performance of the model. This led to a total of 31,055,873 parameters, with 31,044,097 Trainable params and 11,776 Non-trainable params. Figure 2 show the details of the model layers and organization.

The technology stack used for conducting the experiments included free tier of Google co-laboratory which uses a GPU: 1xTesla K80, compute 3.7, with 2496 CUDA cores, 12 GB GDDR5 VRAM, and a CPU: 1xsingle core hyper threaded Xeon Processors @2.3 GHz i.e. (1 core, 2 threads). Moreover, it must be noted that there was no data augmentation step, and only 60% of the data was used for model building, 20% for validation and hyper-parameter tuning, and 20% of the data as the independent test set, making it a very efficient model architecture suitable for low-resource settings.

The details of the processing steps for multi-class segmentation from different modalities of input image volumes is outlined below:

- **Image pre-processing:**

This step involved 2D patch extraction from 3D volumes, 2D convolution, batch normalization, max pooling during down-sampling/decoding leg and concatenation of residual features steps from down

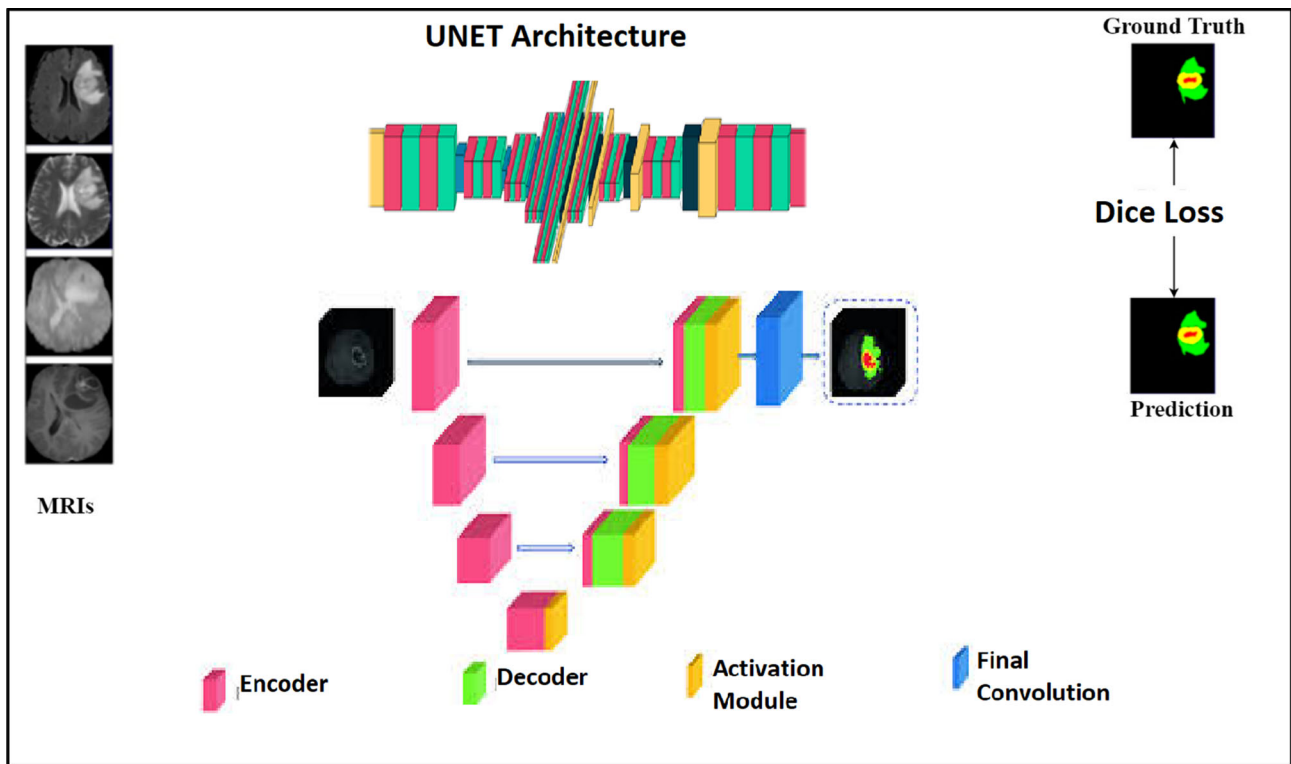


Fig. 1 The proposed light weight CNN 3D U-Net model

sampling leg in the up-sampling/encoding leg. The MRI volumes corresponding to modalities (T1, T1-post, T2, FLAIR) were processing with intensity normalization step to have zero mean and unit variance.

- **Multi-class segmentation:**

The multiclass segmentation task involved segmenting the whole tumor region (WT), necrotic/non-enhancing tumor core (NCR/NET), peri-tumoral edema (ED), and enhancing tumor (ET) by combining all 4 MRI modality images. This step involved downsampling of pre-processed patches from the training set to isotropic voxels of 2 mm size. By using the images from all 4 modalities (T1, Tgd, T2 and FLAIR), a U-Net model was built, and network converges to a WT (Whole Tumor) probability map at 2 mm resolution. This is then thresholded further and up-sampled to 1 mm resolution using naïve nearest-neighbour interpolation, to obtain separate label maps for each class.

- **Post-processing step:**

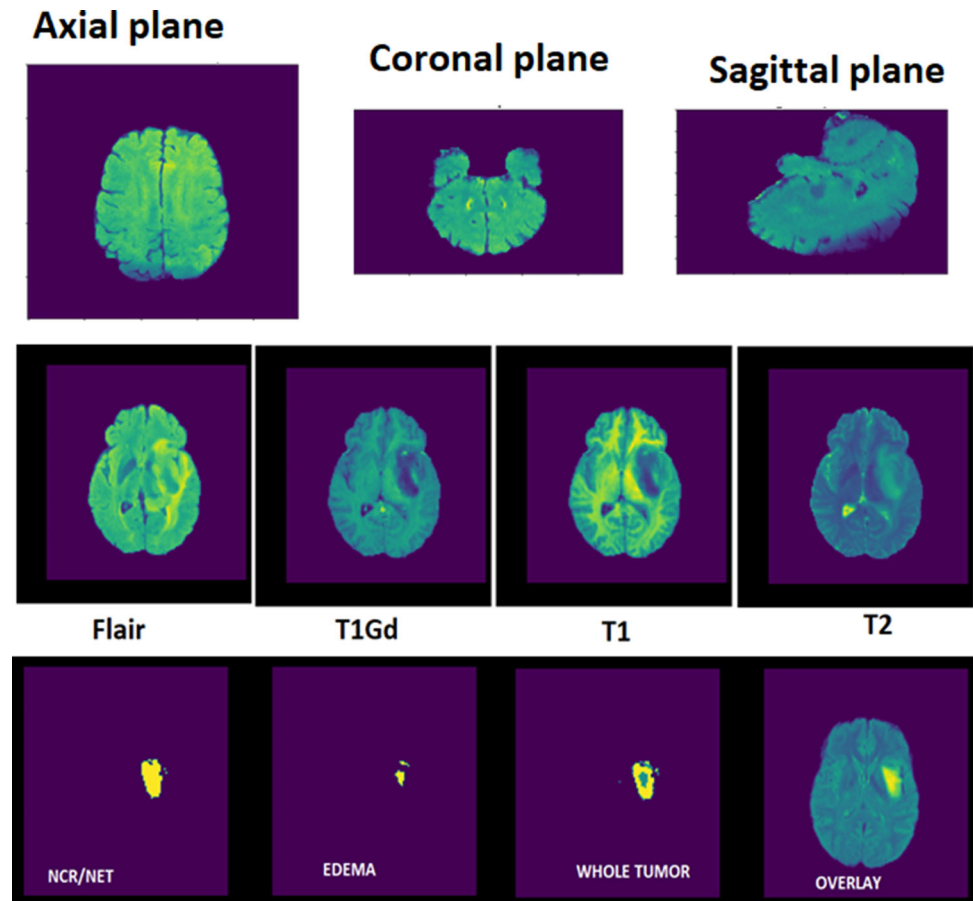
The post-processing requirements with the CNN based U-Net architecture was minimal, as the model can learn the label maps without any need for post-processing, for segmenting different tumor tissues corresponding to each class.

4 Experimental results and discussion

The dataset used for model building and evaluation comprised of LGG (Low Grade Glioma) subset of the BraTS 2018 challenge dataset. This subset contained data from 65 subjects, with each subject data containing 3 dimensional volumes of image data from FLAIR, T1, T1Gd, T2 MRI modalities, as well as the ground truth slices of size $155 \times 240 \times 240$ (155 slices of 240×240 images for each modality). Figure 2 shows the 3D volume for one of the subject in the data set, in terms of axial, coronal and sagittal plane visualization, the multimodal image slices corresponding to FLAIR, T1Gd, T1, and T2 modalities, and segmentation ground truth/label maps for each of the classes and overlay of segmentation map on the training image for a subject.

The model building stage involved resizing of images to 128×128 , with a data split of 60% training set, and 40% validation and test set (20% validation, and 20% test set) from a total of 4550 images, with 3185 images used for training, 1365 images for validation, and 1365 images for test set with each subset comprising data corresponding to different subjects. The hyperparameters used for training involved a batch size of 32, with 30 epochs, and ADAM optimizer with dice loss as optimization function, and dice coefficient as the performance metric (since dice

Fig. 2 3D image volumes, MRI image data slices for each modality, and segmentation label maps/ground truth



coefficient is a common metric used to assess the segmentation performance as compared to the classification accuracy).

It must be noted, that there was no data augmentation done, and to the best of our knowledge, this is the most light-weight deep learning model for this problem, with minimal overhead on computational resources requirements, as compared to other methods proposed in literature, and relies entirely on the power of light weight CNN based 3D U-Net model proposed in this paper, for learning the feature representations for pixel segmentation task. Figure 3 shows the quantitative performance of the model for segmentation of each class [enhancing tumor (ET), whole tumor (WT) and tumor core (TC)], in terms of training and validation dice coefficient and dice loss for 30 epochs. Tables 1 and 2 show the validation and independent test set performances and compares with the previous proposed three top performing challenge participants' methods. The independent test set performance for our method, was superior, with a dice coefficient of 0.9385 ($\sim 94\%$), and dice loss of 0.0614 ($\sim 6\%$). Figurew 4, 5, 6 and 7 show the qualitative assessment in terms of visualization of the ground truth vs. predicted segmentation labels for each class (ET, WT, and TC) for training set

(label 1) in Fig. 4, and validation and independent test set performance for each label in Figs. 5, 6 and 7.

5 Conclusion and further plan

A novel low-resource and light weight CNN based 3D U-Net architecture is proposed in this work for multi-class tumor tissue segmentation. The proposed approach is a light weight deep learning model requiring minimal computational resources as compared to traditional large deep learning architectures proposed for this problem in earlier work. The fully automated pipeline with a stack of 5-CNN down-sampling/decoding stages and 5-CNN up-sampling/encoding stages of a UNet architecture along with multi-modal inputs corresponding to Flair, T1Gd, T1 and T2 modalities, turns out to be a powerful architecture, does not rely on intensive data augmentation as prevalent in other models proposed for this task in the literature, and results in improved performance as compared to previous other participating teams in the challenge, making it suitable for low resource, remote and extreme environment settings. Further work will focus towards developing sparse data models built with small dataset for segmentation of other

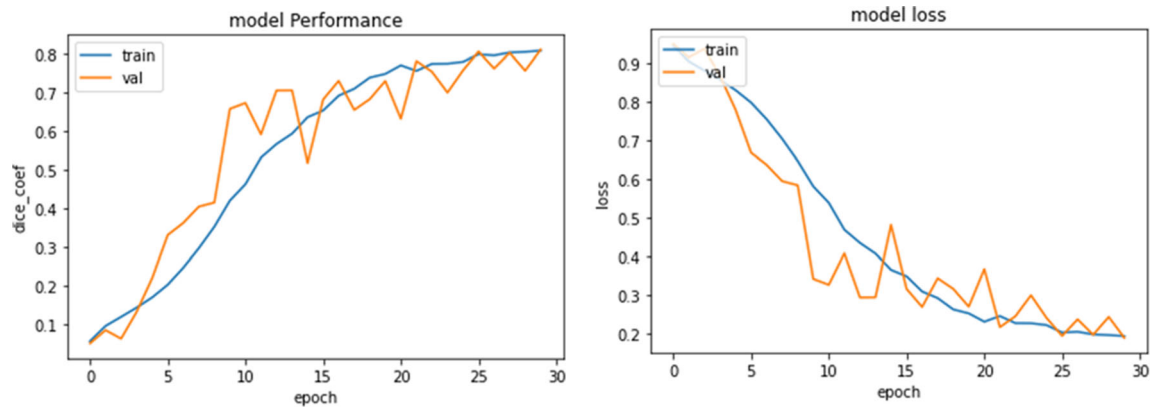


Fig. 3 Proposed segmentation model performance (dice coefficient and dice loss) for label 1 [enhancing tumor (ET)], label 2 (whole tumor), and label 3 (tumor core)

Table 1 Validation set quantitative performance comparison (dice coefficient and dice loss) for multiple classes (ET, WT, TC)

Segmentation method	Performance metric 1 Dice coefficient			Performance metric 2 Dice loss		
	ET	WT	TC	ET	WT	TC
Our method	0.8233	0.9337	0.8695	0.021	0.00914	0.0151
Method 1 (top ranked) [8]	0.7664	0.9100	0.8154	0.029	0.0603	0.0508
Method 2 (2nd ranked) [9]	0.7788	0.8781	0.8062	0.032	0.0612	0.0577
Method 13 (3rd ranked) [10]	0.8173	0.9068	0.8602	0.034	0.0687	0.0586

Fig. 4 Qualitative performance comparison (predicted vs. ground truth segmentation maps) for training set [label 1/ET (enhancing tumor)]



pathologies based on similar medical imaging data, which is a requirement for low resource settings with limited labelled data availability, lack of appropriate deep learning models for imbalanced class distributions, and reliance on

heavy computational resource requirements, which had been the focus of the research community in this area so far.

Fig. 5 Qualitative performance comparison (predicted vs. ground truth segmentation maps) for validation set (top) and test set [bottom (label 1/ET)]

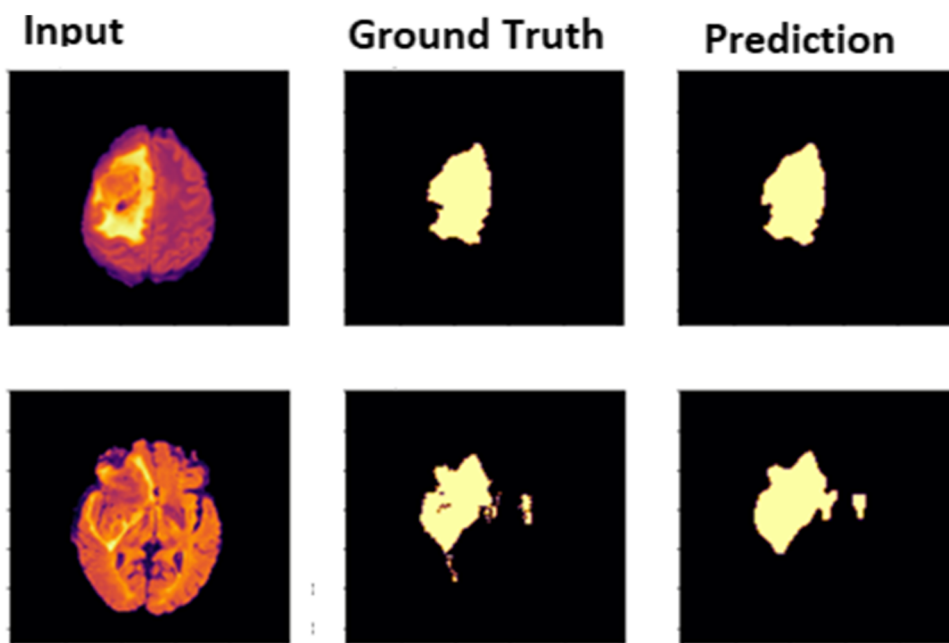


Table 2 Test set quantitative performance comparison (dice coefficient and dice loss) for multiple classes (ET, WT, TC)

Segmentation method	Performance metric 1 Dice coefficient			Performance metric 2 Dice loss		
	ET	WT	TC	ET	WT	TC
Our method	0.8422	0.9412	0.8988	0.022	0.00913	0.0152
Method 1 (top ranked) [8]	0.7464	0.9010	0.8099	0.031	0.0692	0.0523
Method 2 (2nd ranked) [9]	0.7081	0.8892	0.8367	0.034	0.0645	0.0599
Method 13 (3rd ranked) [10]	0.8243	0.9001	0.8403	0.0331	0.0615	0.0525

Fig. 6 Qualitative performance comparison (predicted vs. ground truth segmentation maps) for validation set (top) and test set [bottom (label 2/WT)]

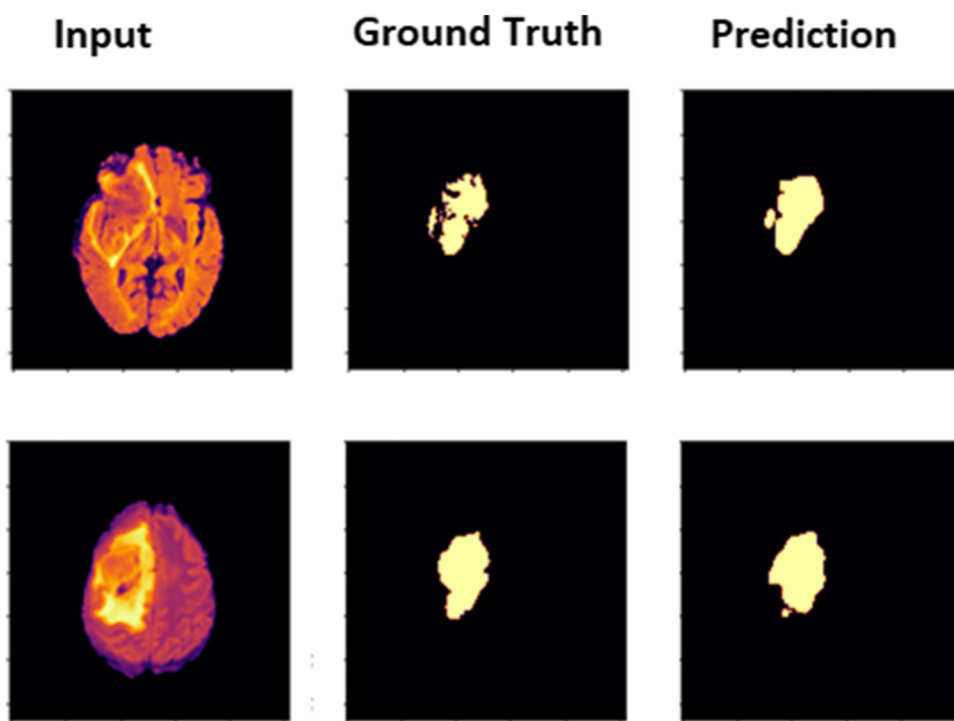
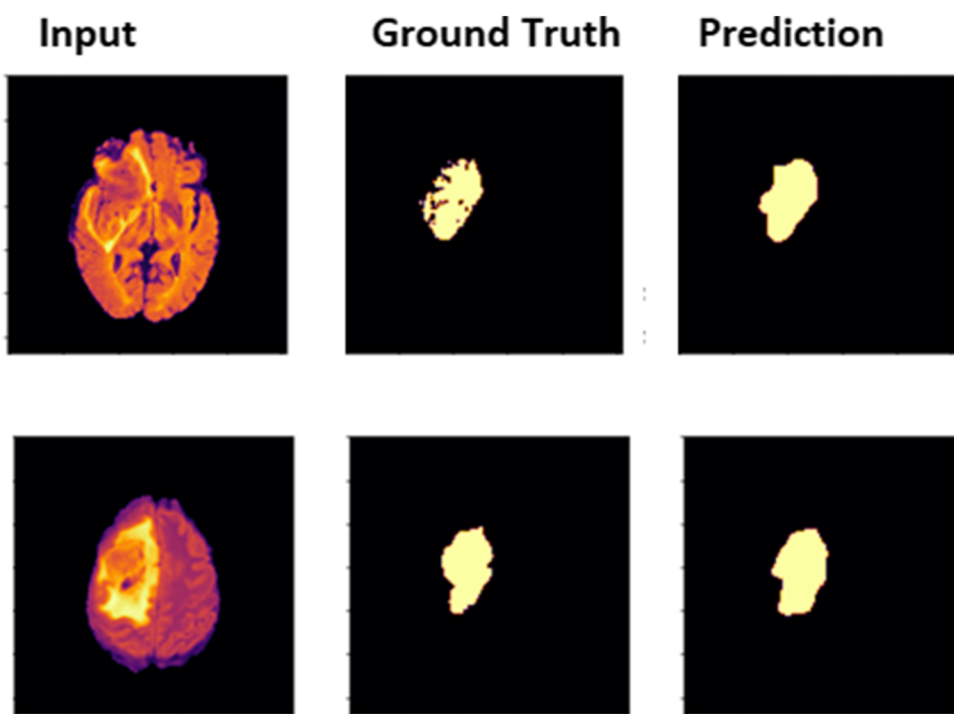


Fig. 7 Qualitative performance comparison (predicted vs. ground truth segmentation maps) for validation set (top) and test set [bottom (label 3/TC)]



Acknowledgements The authors are thankful for the publicly available challenge dataset provided by MICCAI Society <https://www.med.upenn.edu/sbia/brats2018/data.html>, and our preliminary findings in [1] and performance benchmarks from the challenge organizers [12], used a baseline performance reference in this study.

References

- Chetty G, Singh M, White M (2019) Automatic brain image analysis based on multimodal deep learning scheme. In: Rhee PK, Hwa K-Y, Pai T-W, Howard D, Bashar R (eds), Proceedings—2019 international conference on machine learning and data engineering (iCMLDE 2019) [8995755]. IEEE, Institute of Electrical and Electronics Engineers, pp 97–100. <https://doi.org/10.1109/iCMLDE49015.2019.00028>
- Alharbi A, Halikias G, Yamin M et al (2021) Web-based framework for smart parking system. *Int J Inf Technol.* <https://doi.org/10.1007/s41870-021-00725-8>
- Yamin M (2020) Counting the cost of COVID-19. *Int J Inf Technol* 12:311–317. <https://doi.org/10.1007/s41870-020-00466-0>
- Sen AA, Yamin M (2020) Advantages of using fog in IoT applications. *Int J Inf Technol.* <https://doi.org/10.1007/s41870-020-00514-9>
- Yamin M (2018) IT applications in healthcare management: a survey. *Int J Inf Technol* 10(4):503–509. <https://doi.org/10.1007/s41870-018-0203-3>
- Chetty G, Yamin M (2017) A distributed smart fusion framework based on hard and soft sensors. *Int J Inf Technol.* <https://doi.org/10.1007/s41870-017-0008-9>
- Chetty G, Yamin M (2015) Intelligent human activity recognition scheme for e-health applications. *Malays J Comput Sci* 28(1):59–69
- Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein, KH (2018) No newnet. In: International conference on medical image computing and computer assisted intervention (MICCAI 2018). Multimodal brain tumor segmentation challenge (BraTS 2018). BrainLes 2018 workshop. LNCS, Springer (2018)
- McKinley R, Meier R, Wiest R (2018) Ensembles of densely-connected CNNs with label-uncertainty for brain tumor segmentation. In: International conference on medical image computing and computer assisted intervention (MICCAI 2018)
- Zhou C, Chen S, Ding C, Tao D (2018) Learning contextual and attentive information for brain tumor segmentation. In: International conference on medical image computing and computer assisted intervention (MICCAI 2018). multimodal brain tumor segmentation challenge (BraTS 2018). BrainLes 2018 workshop. LNCS, Springer
- Huang G, Liu Z, van der Maaten L, Weinberger KQ (2017) Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2261–2269
- Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R, Lanczi L, Gerstner E, Weber M-A, Arbel T, Avants BB, Ayache N, Buendia P, Collins DL, Cordier N, Corso JJ, Criminisi A, Das T, Delingette H, Demiralp C, Durst CR, Dojat M, Doyle S, Festa J, Forbes F, Geremia E, Glocker B, Golland P, Guo X, Hamamci A, Iftekharuddin KM, Jena R, John NM, Konukoglu E, Lashkari D, Mariz JA, Meier R, Pereira S, Precup D, Price SJ, Raviv TR, Reza SMS, Ryan M, Sarikaya D, Schwartz L, Shin H-C, Shotton J, Silva CA, Sousa N, Subbanna NK, Szekely G, Taylor TJ, Thomas OM, Tustison NJ, Unal G, Vasseur F, Wintermark M, Ye DH, Zhao L, Zhao B, Zikic D, Prastawa M, Reyes M, Van Leemput K (2015) The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans Med Imaging* 34(10):1993–2024. <https://doi.org/10.1109/TMI.2014.2377694>