Brain tumour segmentation using U-Net based fully convolutional networks and extremely randomized trees

Hai Thanh Le[1*], Hien Thi-Thu Pham[2]

[1]Faculty of Mechanical Engineering, Ho Chi Minh city University of Technology, VNU Ho Chi Minh city
[2]Department of Biomedical Engineering, International University, VNU Ho Chi Minh city

Received 12 April 2018; accepted 27 July 2018

Abstract:

In this paper, we present a model-based learning for brain tumour segmentation from multimodal MRI protocols. The model uses U-Net-based fully convolutional networks to extract features from a multimodal MRI training dataset and then applies them to Extremely randomized trees (ExtraTrees) classifier for segmenting the abnormal tissues associated with brain tumour. The morphological filters are then utilized to remove the misclassified labels. Our method was evaluated on the Brain Tumour Segmentation Challenge 2013 (BRATS 2013) dataset, achieving the Dice metric of 0.85, 0.81 and 0.72 for whole tumour, tumour core and enhancing tumour core, respectively. The segmentation results obtained have been compared to the most recent methods, providing a competitive performance.

Keywords: **brain tumour, convolutional neural network, extremely randomized trees, segmentation, U-Net.**

Classification number: **2.3**

Introduction

Accurate brain tumour segmentation plays a key role in cancer diagnosis, treatment planning, and treatment evaluation. Since the manual segmentation of brain tumours is laborious, the development of semi-automatic or automatic brain tumour segmentation methods makes enormous demands on researchers [1]. Ultrasound, computed tomography (CT) and magnetic resonance imaging (MRI) acquisition protocols are standard image modalities that are used clinically. Many previous studies have shown that the multimodal MRI protocols can be used to identify brain tumours for treatment strategy, as the different image contrasts of these MRI protocols can be used to extract important complementary information. The multimodal MRI protocols include T2-weighted fluid-attenuated inversion recovery (FLAIR), T1-weighted (T1), T1-weighted contrast-enhanced (T1c) and T2-weighted (T2).

In recent years, an annual workshop and challenge, called Multimodal Brain Tumour Image Segmentation (BRATS), is held to different benchmark methods that have been developed to segment the brain tumour [2]. The previous studies on brain tumour segmentation can be categorised into unsupervised learning [3] and supervised learning [4, 5] methods. We only reviewed some of the most recent and closely relevant studies to our method.

Unsupervised learning-based clustering has been successfully applied for the brain tumour segmentation.

*Corresponding author: Email: lthai@hcmut.edu.vn

In [3], the Szilagyi group proposed a multi-stage c-means framework for segmenting brain tumours using multimodal MRI scans and received promising results, although limited by the considered scope of the data.

On the other hand, supervised learning-based methods demand a pair of training data and its label to train a classifier that can then be segmented new data without training. Pinto, et al. [4] proposed an algorithm based on a random decision forest (RDF), using a k-fold cross-validation approach. They extracted features for RDF which is intensity complemented and context based features for every voxel represented. Morphological filters were used for post-processing to reduce misclassification errors. Recently, Soltaninejad, et al. [5] applied extremely randomized trees (ExtraTrees) [6] classification with superpixel based segmentation using a single FLAIR scan in four modalities of MRI dataset. Their results achieved an overall 0.88 Dice score of the complete tumor segmentation for both high-grade glioma (HGG) and low-grade glioma (LGG) cases. However, the final segmentation of this method could be influenced by the final delineation caused by the tuning of superpixel size. Additionally, the Soltaninejad group [7] presented a different method by using random forests classifier to segment the brain tumour. This method is based on the features extracted from a fully convolutional neural network (FCN), namely FCN-8s architecture.

Besides, our previous method [8] trained ExtraTrees classifier for brain tumour segmentation based on a region of interest (ROI) of tumour in FLAIR sequence. This method obtained a 0.9 Dice score of the complete tumour but received a low score of enhancing and core tumour with the BRATS 2013 dataset [2].

In the recent years, a lot of researchers have used the convolutional neural networks (CNNs) to classify images, specifically deep CNNs, which makes it possible to train extremely deep neural networks from the random initialised weights with complex and big data. The deep CNNs are constructed by combining many convolutional layers, which convolve an image with kernels to extract features that are more robust and adaptive for discriminative models. Currently, various deep learning methods have achieved the high score in BRATS challenges [9-11]. A detailed review of various medical image classification, segmentation, and registration methods can be found in [12]. Biomedical images have many patterns of the object such as the tumours, and their intensities are usually variable. Ronneberger, et al. [13] developed the U-Net-based fully convolutional networks (FCNs), which consist of a down-sampling (encoding) pathway and an up-sampling (encoding) pathway with residual connections between the two that concatenate feature maps at different spatial scales in order to segment the cell cancer. Based on the original U-Net architecture, some groups [14, 15] proposed a method for brain tumour segmentation and achieved the competitive performance of those built models with BRATS datasets.

However, there are still several challenges: (1) most methods obtain the promising results for HGG cases, but the performance of LGG cases is still poor; (2) especially, the segmentation of enhancing and core tumor always has a low score compared to complete tumor score; (3) finally, the demand for reducing computation time and memory is still unsatisfied.

In this study, we propose a novel segmentation method that uses the U-Net architecture [13] to extract features and then inputs these to train ExtraTrees classifier [8]. Furthermore, we apply a simple filter in a postprocessing step to eliminate misclassified labels.

Methods

Discriminative models create a decision function that describes the input vectors and assigns each vector to a class. The decision function aims to make the needful informational relation based on the training samples. Additionally, the performance of segmentation depends on the quality of the input data and the extraction of effective features. The models for segmentation tasks create the relational space based on the intensity information of input images to ground truth images.

The general structure of our model is shown in Fig. 1. In the following part, we will describe the role of each part of brain tumour segmentation.

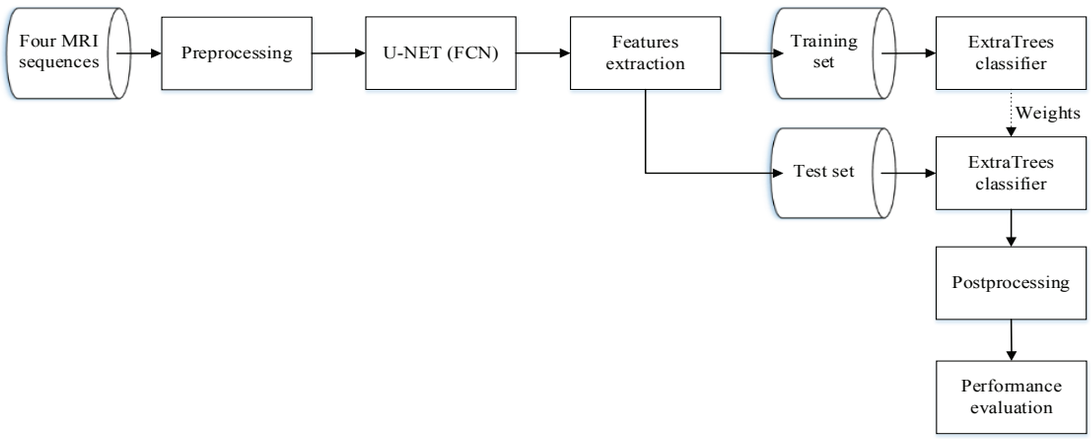

Fig. 1. The proposed discriminative model.

Dataset

The proposed method is trained and validated on the BRATS 2013 dataset [2], which consists of 30 patient MRI scans, of which 20 are HGG and 10 are LGG. Each patient has four MRI sequences including FLAIR, T1c, T2 and T1. This dataset with multimodal MRI data has already been skull-stripped, registered into the T1c scan and interpolated into 1×1×1 mm^3 with a sequence size of 240×240×155. Moreover, the ground truth images of dataset were manually labeled into four types of intra-tumoral classes (labels): 1-necrosis (red), 2-edema (green), 3-non-enhancing (blue) and 4-enhancing tumour (yellow) and the others are 0-normal (healthy) tissue (black) as shown in Fig. 2 (GT). The ground truth data have been used in two steps: model training and performance evaluation for final segmentation.

Pre-processing

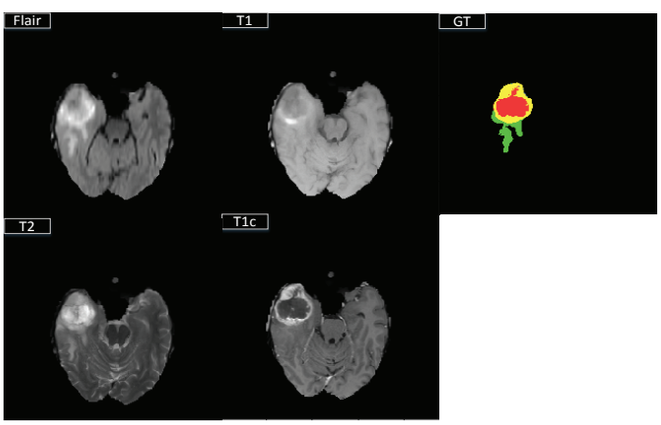

Fig. 2. Four MRI modalities and their ground truth from HGG patient.

In this study, we applied the N4ITK method [16] to reduce inhomogeneity in MR images. A histogram normalisation method [17] was then employed to ensure that addresses data heterogeneity caused by multi-scanners acquisitions of MR images. Finally, the intensities of each MRI sequence were normalised by subtracting the average of intensities of each sequence and then dividing them by its standard deviation. Fig. 2 shows the sample of four MRI modalities and their ground truth from HGG patient 0001 after pre-processing.

U-Net based deep convolutional neural networks

Our network is similar in spirit to the U-Net [14], which is different from the original U-Net [11]. The U-Net [14] described in Fig. 3 uses the deconvolution operator instead of an up-sampling operator in the decoding pathway and applies zero padding to keep the same resolution of output images as the input images. Therefore, the network does not need a cropping operator of the border regions. Every block in the encoding pathway has two convolutional layers with a 3×3 filter, a stride of 1 and rectified linear unit (ReLU) activation, which increases the number of feature maps from 1 to 1024. For the down-sampling, max pooling with stride 2×2 is used to the end of every block except the last block. Therefore, the size of feature maps decrease from 240×240 to 15×15. In the decoding pathway, every block starts with a deconvolutional layer with same size filter in the decoding pathway and a stride of 2×2, which doubles the size of feature maps in both directions but decreases the number of feature maps by two. Thus, the size of feature maps increases from 15×15 to 240×240. In every up-

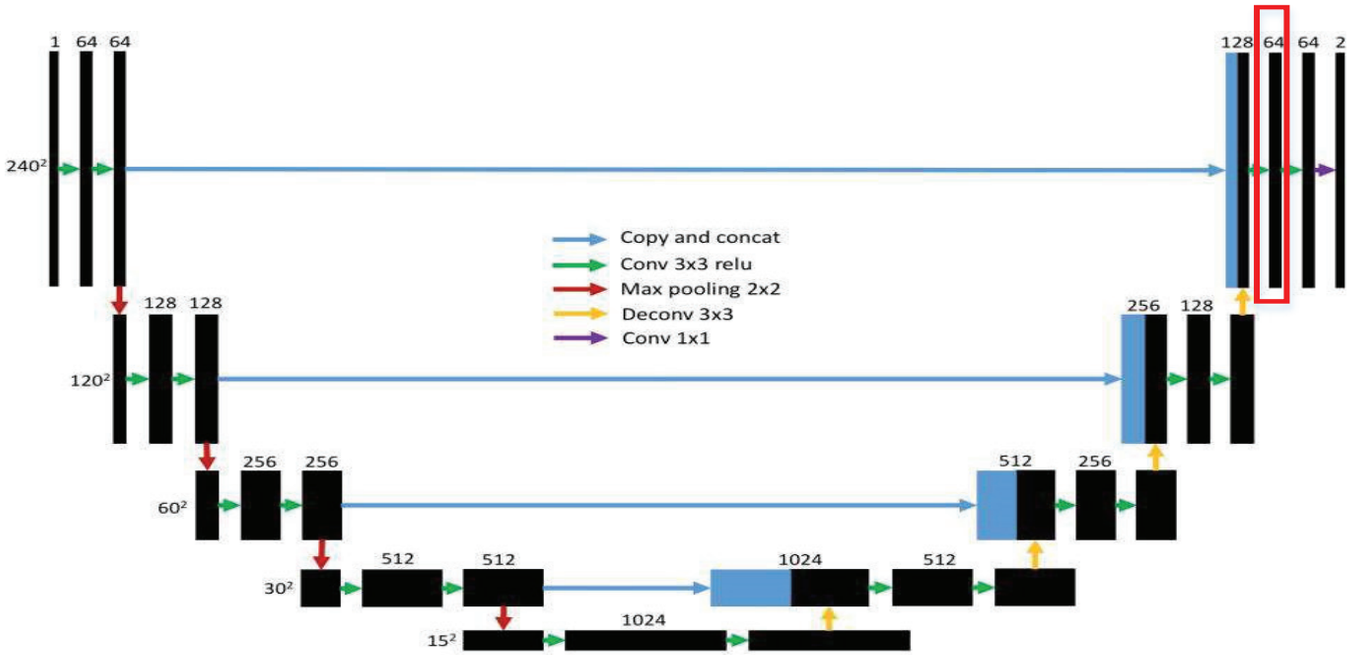

Fig. 3. The U-Net architecture [14].

sampling block, two convolutional layers reduce a half of the feature maps after concatenating the deconvolutional feature maps and the feature maps from the encoding path.

Our proposed network is then added to the batch normalization [18] layer after each convolutional layer for regularization purposes.

Feature extraction

Image processing provides many algorithms for the extraction of characteristics from images. In the field of biomedical image analysis, many studies are trying to find the tumour characteristics with a high correlation to the appearance of the brain images. Nonetheless, no proper feature sets have been extracted yet, which is why various groups need to use a large feature set based on many feature extraction methods such as texture features, spatial context features and higher order operators.

The U-Net model uses the powerful CNN to filter the useful features from input data in encoding pathway and then embeds these features in the output map with the same position in the decoding pathway. It makes the collected features easier to calculate for the next step or compare with the desired output. In this study, we extracted the features in all MRI protocols from the U-Net model, but we did not obtain the output of the model from a top layer, as it was only two values. We collected the features from the convolutional layer next to the concatenated layer in the final block of the decoding pathway as shown a red rectangle in Fig. 3. This

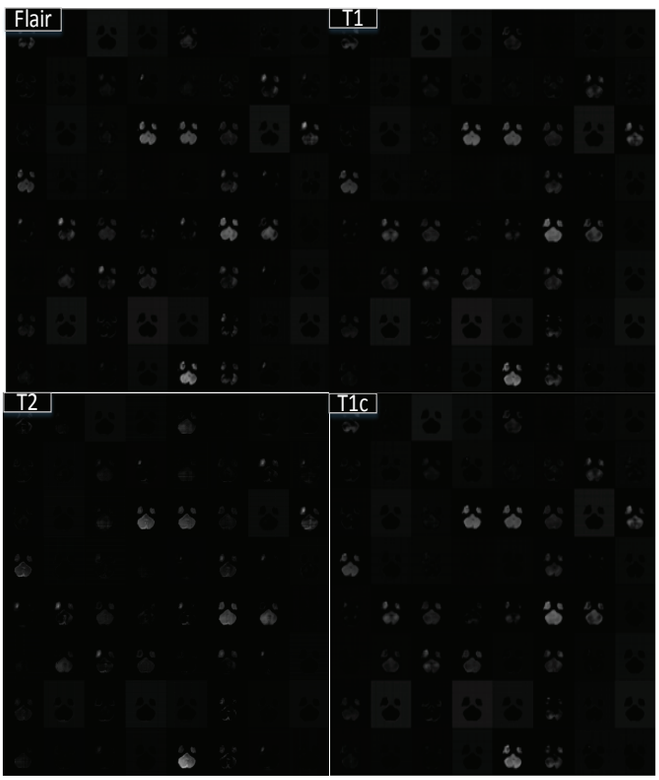

Fig. 4. Feature maps from four MRI multimodalities.

output has 64 feature maps with the size of 240×240 and total parameters of 73792 for each image of MRI scans. Fig. 4 shows the feature maps of each image of FLAIR, T1c, T2 and T1 sequences extracted from the U-Net model.

Training set and test set

From the BRATS 2013 dataset, we used the first half of HGG and LGG cases with all MRI modalities for the training set and the second half of dataset including 10 HGG and 5 LGG cases to evaluate the performance of our method. In this study, the HGG and LGG training sets are combined, trained and cross-validated together.

Classifier

In our method, the Extremely Randomized Trees (ExtraTrees) [6] classifier is the main part of the brain tumour segmentation system. In our previous work [8], we had described the reason for choosing this classifier with the following advantages:

- High accuracy
- Easy handling of large datasets
- Estimating feature importance.

In the ExtraTrees classifier, the splitting rule differs from the Random Decision Forests in how the randomness is applied to choose the cut-points for each candidate feature during the training. It means that a single threshold is chosen at random instead of searching the best threshold for each feature. This classifier usually allows to reduce the variance of the model a bit more. Thus, it can provide slightly better results than the Random Decision Forests.

The main parameters of the ExtraTrees classifier are the number of trees, depth of tree and the set of attributes (K) that performs the random split. For the classification tasks, the optimum value of K is $K=\sqrt{n}$, with n being the total number of features; in our study, K=16. After that calculation, we tuned the other parameters with different number of trees and depths of the tree on the training set and evaluated the accuracy of classification. The highest accuracy was achieved with the number of trees N_{tree}=50 and depth D_{tree}=15 as in [7]. Finally, the ExtraTrees classifier was trained by combining the features extraction described above to a 256-dimensional feature vector.

Postprocessing

Our model is applied without a priori information about the classified objects; hence, the obtained results have to be refined by postprocessing. In this step, we employ simple morphological filters including dilation and erosion with a structuring element of a 3×3 square to remove small false positives (the misclassified labels or 'salt' noises) in the segmented image while keeping the large tumorous regions unaffected.

Performance evaluation

The final step of segmentation is an evaluation of the obtained results. In this study, we evaluate the tumour segmentation on three sub-tumoral regions, following [2], which are the enhancing tumour, the core (necrosis + non-enhancing tumour + enhancing tumour) and the complete tumour (all classes combined), by using the measurements in Dice coefficient and Sensitivity [19]. The Dice score provides the overlap measurement between the ground truth images from the BRATS 2013 dataset and the segmentation results of our proposed method:

$$Dice = \frac{2TP}{FP+2TP+FN} \quad (1)$$

in which, TP, FP and FN denote the true positive, false positive and false negative measurements, respectively.

Additionally, sensitivity is used to determine the number of TP and FN:

$$Sensitivity = \frac{TP}{TP+FN} \quad (2)$$

Results and discussion

In this study, we proposed using the ExtraTrees classifier with features learned from U-Net-based fully convolutional neural networks for solving the brain tumour segmentation challenge. For HGG and LGG training sets, the images were selected from each MRI sequence that depends on their ground truth's energy with a threshold value of HGG greater than LGG. Therefore, this step helped in reducing the number of images that are put into the U-Net model to extract features for training data.

Our U-Net model and ExtraTrees classifier were implemented in Keras [20] with a TensorFlow [21] backend and open source library provided by [22]. The best advantage of our proposed method is that the training

Table 1. Dice and sensitivity scores of our proposed method compared to the results from other groups recently published random forests, ExtraTrees and U-Net based methods for the BRATS 2013 dataset.

Method	Dice score			Sensitivity		
	Complete	*Core*	*Enhancing*	*Complete*	*Core*	*Enhancing*
Proposed	0.85	0.81	0.72	0.87	0.85	0.82
Pinto [4]	0.86	0.71	0.74	0.82	0.66	0.72
Soltaninejad [7]	0.88	0.80	0.73	0.89	0.77	0.70
Our previous [8]	0.90	0.63	0.61	0.87	0.72	0.65
Dong [14]	0.86	0.86	0.65	0.88	0.90	0.78

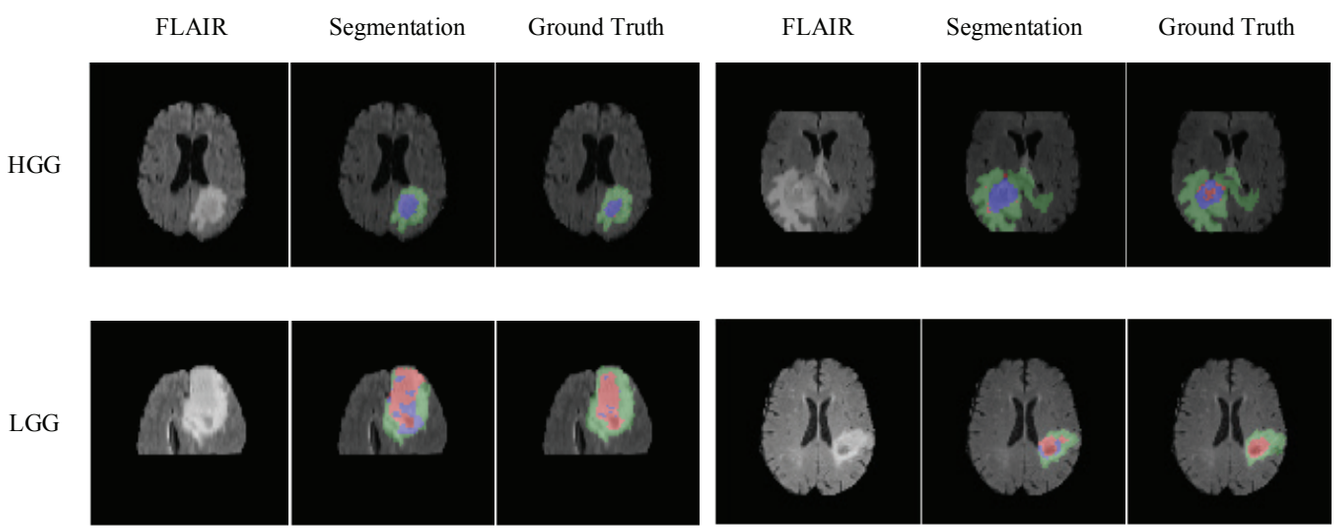

Fig. 5. Segmentation results for the HGG and LGG cases compared to their ground truth.

time is only around one hour, but for the prediction, the computation time is about 3–4 minutes per case. Compared to some studies, our computational time is more efficient than [7-8] and less efficient than [14].

The results of our proposed model and the recent state-of-the-art methods validated on the BRATS 2013 dataset is shown in Table 1. These results are uploaded on the BRATS 2015 server, which evaluates the segmentation and provides measurements in Dice and sensitivity scores of whole tumour, tumour core and enhancing tumour core. Table 1 shows that our method achieves competitive results in the Dice score and performs slightly better in sensitivity measurement for all types of brain tumour with the smaller data for learning.

Figure 5 shows some examples of our qualitative overlaid segmentation results for both HGG and LGG cases on FLAIR MR images compared to the ground truth images. The segmented results are coloured as described in the Dataset section.

Due to the limitation of computational resource, our proposed model is only trained and evaluated on the BRATS 2013 dataset, which contains much less HGG and LGG patient cases than the BRATS 2015 dataset. Furthermore, our model segmenting the enhancing tumour for LGG cases is less successful than for HGG cases because there are fewer LGG cases than HGG cases and because most of the LGG cases rarely have regions of enhancing tumour.

Conclusions

In this paper, we developed a learning-based automatic method for brain tumour segmentation in MR images.

This method used the features extracted from the U-Net-based deep convolutional networks and applied them to the ExtraTrees classifier as the input data. Additionally, we refined the segmentation results by removing the false labels using the simple morphological filters. Based on the BRATS 2013 dataset, in comparing to other state-of-the-art methods, we demonstrated that our approach can achieve comparable results with average Dice scores of 0.85, 0.81 and 0.72 for whole tumour, tumour core and enhancing tumour core, respectively.

ACKNOWLEDGEMENT

This research was carried out in part at the Saijo Laboratory of Professor Yoshifumi Saijo, Department of Biomedical Engineering, Tohoku University. This research is funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 103.03-2016.86.

REFERENCES

[1] S. Bauer, R. Wiest, L.P. Nolte, and M. Reyes (2013), "A survey of MRI-based medical image analysis for brain tumor studies", *Physics in Medicine and Biology*, **58**, pp.97-129.

[2] B.H. Menze, et al. (2015), "The multimodal brain tumor image segmentation benchmark (BRATS)", *IEEE Transitions on Medical Imaging*, **34(10)**, pp.1993-2024.

[3] L. Szilagyi, L. Lefkovits, and B. Benyo (2015), "Automatic Brain Tumor Segmentation in multispectral MRI volumes using a fuzzy c-means cascade algorithm", *The 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pp.285-291.

[4] A. Pinto, S. Pereira, H. Dinis, C.A. Silva, and D.L.M.D Rasteiro (2015), "Random decision forests for automatic brain tumor segmentation on multi-modal MRI images", *Bioengineering (ENBENG) IEEE 4th Portuguese Meeting on IEEE*, pp.1-5.

[5] M. Soltaninejad, G. Yang, T. Lambrou, N. Allinson, T.L. Jones, T.R. Barrick, F.A. Howe, and X. Ye (2017), "Automated brain tumor detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI", *International Journal of Computer Assisted Radiology and Surgery*, **12(2)**, pp.183-203.

[6] P. Geurts, D. Ernst, and L. Wehenkel (2006), "Extremely randomized trees", *Machine Learning*, **63(1)**, pp.3-42.

[7] M. Soltaninejad, L. Zhang, T. Lambrou, N. Allinson, and X. Ye (2017), "Multimodal MRI brain tumor segmentation using random forests with features learned from fully convolutional neural network", *arXiv preprint arXiv:1704.08134v1*.

[8] H.T. Le, H.T.T. Pham and H.H Tran (2018), "Automatic brain tumor segmentation using extremely randomized trees", *Journal of Science and Technology (Technical Universities)* (accepted).

[9] S. Pereira, A. Pinto, V. Alves, and C.A. Silva (2016), "Brain Tumor Segmentation using Convolutional Neural Networks in MRI Images", *IEEE Transactions Medical Imaging*, **35(5)**, pp.1240-1251.

[10] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.M. Jodoin, and H. Larochelle (2017), "Brain tumor segmentation with Deep Neural Networks", *Medical Image Analysis*, **35**, pp.18-31.

[11] K. Kamnitsas, C. Ledig, V.F.J. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, and B. Glocker (2017), "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation", *Medical Image Analysis*, **36**, pp.61-78.

[12] G. Litjens, T. Kooi, B.E. Bejnordi, A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W.M. van der Laak, B.V Ginneken, and C.I. Sanchez (2017), "A survey on deep learning in medical image analysis", *Medical Image Analysis*, **42**, pp.60-88.

[13] O. Ronneberger, P. Fischer, and T. Brox (2015), "U-Net: Convolutional networks for biomedical image segmentation", *Medical Image Computing and Computer-Assisted Intervention*, **9351**, pp.234-241.

[14] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo (2017), "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks", *arXiv preprint arXiv:1705.03820v3*.

[15] A. Beers, K. Chang, J. Brown, E. Sartor, C.P. Mammen, E. Gerstner, B. Rosen, and J.K. Cramer (2017), "Sequential 3D U-Nets for biologically-informed brain tumor segmentation", *arXiv preprint arXiv:1709.02967v1*.

[16] N.J. Tustison, B.B. Avants, P.A. Cook, Y. Zheng, A. Egan, P.A. Yushkevich, and J.C. Gee (2010), "N4ITK: Improved N3 bias correction", *IEEE Transactions Medical Imaging*, **29(6)**, pp.1310-1320.

[17] C.P. Loizou, M. Pantziaris, C.S. Pattichis, and I. Seimenis (2013), "Brain MR image normalization in texture analysis of multiple sclerosis", *Journal of Biomedical Graphics and Computing*, **3(1)**, pp.20-34.

[18] S. Ioffe, and C. Szegedy (2015), "Batch normalization: Accelerating deep network training by reducing internal covariate shift", *The 32nd International Conference on Machine Learning*, **37**, pp.448-456.

[19] D.M. Powers (2011), "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation", *Journal of Machine Learning Technologies*, **2(1)**, pp.37-63.

[20] F. Chollet, and others (2015), "Keras", *GitHub*, https://github.com/keras-team.

[21] https://www.tensorflow.org/

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay (2011), "Scikit-learn: Machine learning in Python", *Journal of Machine Learning Research*, **12**, pp.2825-2830.