

Brain Tumor Segmentation Using Random Forest Trained on Iteratively Selected Patients

Abdelrahman Ellwaa^(✉), Ahmed Hussein, Essam AlNaggar, Mahmoud Zidan, Michael Zaki, Mohamed A. Ismail, and Nagia M. Ghanem

Faculty of Engineering, Alexandria University, Alexandria, Egypt
abdelrahmanellwaa@gmail.com, michaelzaki1212@gmail.com

Abstract. This paper extends a previously published brain tumor segmentation methods based on Random Decision Forest (RDF). An iterative approach is used in training the RDF in each iteration some patients are added to the training data using some heuristics approach instead of randomly selected training dataset. Feature extraction and selection were applied to select the most discriminative features for training our Random Decision forest on. The post-processing phase has a morphological filter to deal with misclassification errors. Our method is capable of detecting the tumor and segmenting the different tumorous tissues of the glioma achieving competitive results.

Keywords: Brain tumor segmentation · Random forests

1 Introduction

Gliomas are the most frequent primary brain tumors in adults. They are originated from glial cells and infiltrate the surrounding tissues. Gliomas can be divided into Low Grade Gliomas (LGG) and High Grade Gliomas (HGG). Although the former are less aggressive, the later can be very deadly [8,9]. Despite considerable advances in glioma research, patient diagnosis remains poor. Segmentation of brain tumors from MR images is important in cancer treatment planning as well as for cancer research. In current clinical practice, the analysis of brain tumor images is mostly done manually. Apart from being time-consuming, this has the additional drawback of significant intra- and inter-rater variability. Accurate brain tumour segmentation is difficult, because in MR images, brain tumors may have the same appearance with gliosis and stroke, have a variety of shapes, appearances and sizes, and may appear in any position in the brain, invade the surrounding tissue rather than displacing it, causing fuzzy boundaries and also there exists intensity inhomogeneity in MR images. The main goal of brain tumor segmentation is to identify areas of the brain whose configuration deviates from normal tissues. Segmentation methods typically look for active tumorous tissues, necrotic tissues, and edema by exploiting several Magnetic resonance imaging (MRI) modalities, such as T1, T2, T1-Contrasted (T1C) and Flair.

In this paper, we introduce a random forest approach which chooses the patients to be used in training according to a cost function instead of randomly selecting them from our dataset (BRATS 2016 dataset), training is iterative at each iteration some patients are added to the training set to be used in the next iteration, this approach tries to prevent over fitted random forest by choosing patients that get the worst results, patients with brain tumors having various shapes, appearances, and sizes, and may appear in any position in the brain, in the previous iteration. Through the paper, we will illustrate in details the approach and the parameters of the approach.

In the past years there are many approaches that used the random forest in brain tumor segmentation those approaches varies in the features selected and the training approach: five class random forest classifier [4] or cascaded random forest that classifies each voxel on two stages the first is two class classifier tumorous or not and the second classifies tumorous voxels to four tumor classes, so this approach tries to balance the training data in each classifier [5].

The paper is organized in the following way, Sect. 2 contains a description of the training pipeline of the random forest. Section 3 refers to different models used in experiments and the obtained results. Finally, Sect. 4 presents the main conclusions.

2 Training Pipeline

The training pipeline consists of four main steps: pre-processing, feature extraction and selection, training random forest, and post-processing step. In the following, we will introduce each step in details respectively. Figure 1 shows the training pipeline of the random forest.

2.1 Preprocessing

- Bias field signal is a low-frequency and very smooth signal that corrupts MRI images, especially those produced by old MRI machines. Image processing algorithms such as segmentation, texture analysis or classification that use the gray level values of image pixels will not produce satisfactory results. A pre-processing step is needed to correct for the bias field signal. Bias field correction on the MR images is applied using open source code N4ITK [1].
- The second step in pre-processing is histogram matching [2,3] which is proposed to correct the variations in scanners sensitivity. This is because quantitative comparisons of abnormalities in MRI scans between patients or within patients serially are affected by variations in MR scanners performance.

2.2 Feature Extraction and Selection

In this phase, we extracted 328 features from the MR images after being pre-processed. Mainly, three categories of features were extracted: gradient features, appearance features, and context aware features. Most of them were from other

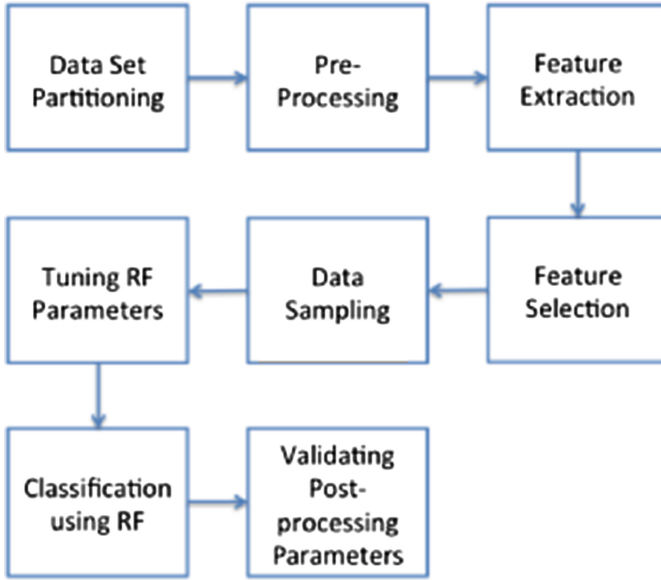


Fig. 1. Random forest training pipeline

published papers from BRATS challenge [4–6]. The first type of features is gradient features which include gradient filter at different Sigma values of 0.5, 1, 2, 3 in each of the three directions x, y, and z and their resultant, difference gradient features, Laplace features, and Recursive Gaussian features. The second type of features is Appearance features which include the voxels intensities, its logarithmic and exponential transformations. The last type of features is the context aware features, which is intensity based, it includes features extracted from the neighbouring voxels, the surrounding cube of the voxel, as most similar, most different, Minimum, Maximum, Range, Kurtosis, Skewness, Standard deviation, and Entropy for all modalities, as well as the local histogram of the cube surrounding the voxel and partitioned into eleven bins, all the previous features were extracted for all modalities Flair, T1, T1c, and T2.

Random forest was used in feature selection using mean decrease impurity, as when training a tree, it can be computed how much each feature decreases the weighted impurity in a tree. For a forest, the impurity decrease from each feature can be averaged and the features are ranked according to this measure.

After feature extraction and selection, each patient will consist of a set of tuples, where each tuple consists of some features which correspond to a voxel in the four modalities of the brain. On average each patient has 1,500,000 tuples. Random sampling is used without replacement, since the health voxels are dominating, to balance healthy and unhealthy tuples. 60,000 health voxels and 15,000 from each other tumor label are randomly sampled from each patient. Each patient finally will contain 100,000 voxels.

2.3 Training Random Decision Forest

Before training the random forest there are some parameters of the random forest that should be determined to be used in training, as the number of trees and number of attributes to split on at each node. So we trained a model by the validation dataset and determined the best values for those parameters based on the k-fold cross validation error, which turns out to be using a random forest with number of trees equals to 45 and number of attributes to split on equals to the square root of the number of features, as we found that the gain in the accuracy is negligible compared to the huge computation it takes when increasing the Random forest parameters values.

Random forest implementation on H2O [7] was used as it is fast, distributed, using the full processing power of the machine, and work on different platforms like python and R.

2.4 Post Processing

The post-processing step is applying binary morphology filter to the output image of the classifier, three binary morphology filters were applied to reduce misclassification errors by connecting large tumorous regions and removing small isolated regions.

The radii used in binary morphology were validated using the validation dataset and they were found to be 8, 8, 0 for complete, core, and enhanced tumors respectively for high-grade gliomas and 1, 8, 2 for complete, core, and enhanced tumors respectively for low-grade gliomas.

3 Experiment and Results

3.1 Experiment

This section explains the models used in classification. BRATS 2016 dataset was used and partitioned into training (70%), testing (20%), and validation (10%) datasets.

Iterative Model. The iterative model mainly addresses the problem of choosing a subset of training patients to train the random forest, so the model is trained through number of iterations, in each iteration the number of patients increases according to a cost function so that the worst N patients are added, there is also a maximum number of patients that is selected according to the hardware resources available. The flowchart in Fig. 2 explains how the iterative model works.

There are many parameters that must be specified first as the number of patients added in each iteration that is set to 5 patients per iteration, initial set of patients that are 30 patients (BRATS 2013 dataset 20 HGG and 10 LGG), maximum number of LGG patients to prevent overfitting LGG patients which

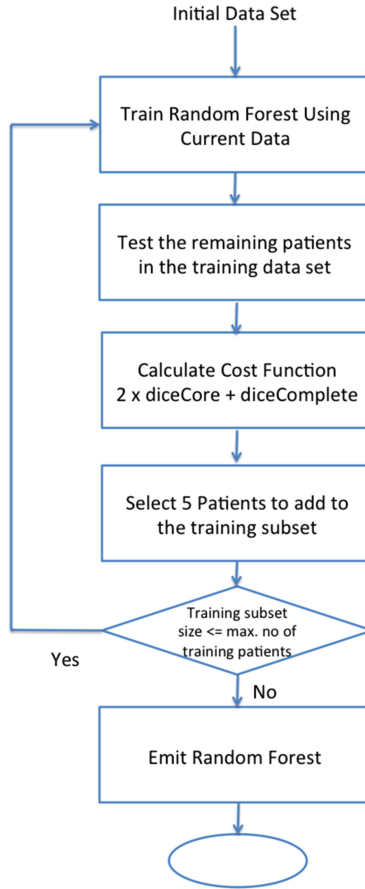


Fig. 2. Flowchart of selecting patients for training the iterative model

is set to 18 patients, the maximum number of patients which is limited to 50 patients, and the cost function which is equal to

$$costFunction = 2 * coreDice + completeDice$$

Cascaded Model. This model consists of two random forests, the first random forest classifies only health voxels (label 0) and non-health voxels (all non-health labels are merged into one representative label for non-health labels), then the second random forest takes the output results from the first random forest and tries to classify the non-health voxels. This approach mainly tries to enhance non-health voxels classification. Firstly, by making a dedicated classifier for health versus non-health voxels which will have many advantages as by merging all non-health labels into one label, this will increase the balancing of the dataset input to the random forest to train on, which is supposed to decrease the number

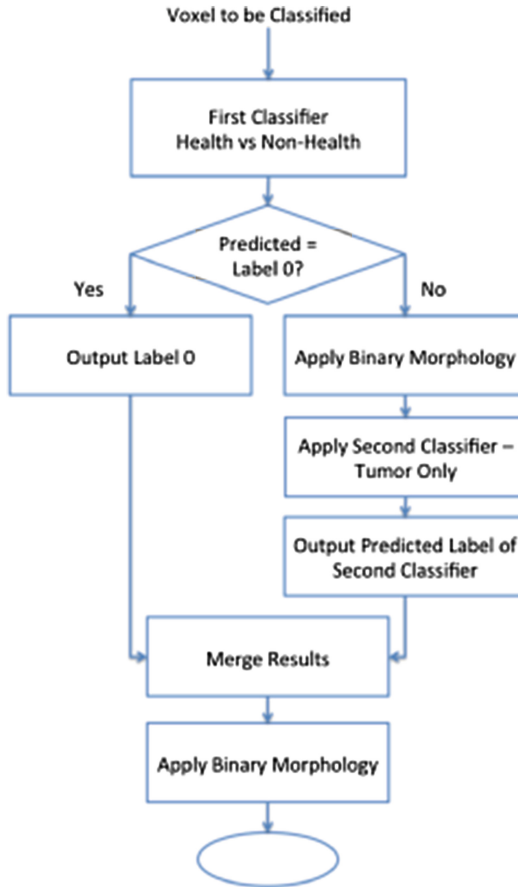


Fig. 3. Flowchart of testing patients on the cascaded model

of non-health voxels classified as health voxels. Secondly, by making a dedicated classifier for non-health labels.

This model was trained on 50 randomly chosen patients from the training data set, the random forest consists of 100 trees each of depth 45. The flowchart in Fig. 3 explains the testing of patients on the cascaded model.

One-Phase Model. This model was trained on 50 randomly chosen patients from the training data set, the random forest consists of 100 trees each of depth 45.

3.2 Results

Those models are tested on 20 (15 HGG and 5 LGG) randomly unseen patients, the results are shown in Table 1 by the dice, specificity and sensitivity scores.

Table 1. The table contains the dice, specificity and sensitivity scores of testing 20 (15 HGG and 5 LGG) randomly unseen patients on our different models.

Model	Complete			Core			Enhancing		
	Dice	Sens	Spec	Dice	Sens	Spec	Dice	Sens	Spec
One-phase HGG	0.81 ± 0.09	0.84 ± 0.10	0.98 ± 0.01	0.62 ± 0.23	0.83 ± 0.16	0.98 ± 0.01	0.75 ± 0.15	0.81 ± 0.12	1.00 ± 0.00
One-phase LGG	0.80 ± 0.06	0.76 ± 0.14	0.99 ± 0.01	0.53 ± 0.17	0.74 ± 0.20	0.99 ± 0.01	0.36 ± 0.32	0.62 ± 0.43	1.00 ± 0.00
Cascaded HGG	0.78 ± 0.10	0.89 ± 0.08	0.97 ± 0.02	0.63 ± 0.23	0.88 ± 0.11	0.98 ± 0.02	0.71 ± 0.18	0.83 ± 0.10	1.00 ± 0.00
Cascaded LGG	0.76 ± 0.11	0.91 ± 0.06	0.98 ± 0.01	0.55 ± 0.20	0.79 ± 0.22	0.99 ± 0.01	0.24 ± 0.31	0.57 ± 0.48	1.00 ± 0.00
Iterative HGG	0.80 ± 0.10	0.84 ± 0.11	0.98 ± 0.01	0.72 ± 0.19	0.79 ± 0.18	0.99 ± 0.01	0.73 ± 0.15	0.75 ± 0.12	1.00 ± 0.00
Iterative LGG	0.84 ± 0.02	0.83 ± 0.09	0.99 ± 0.01	0.72 ± 0.12	0.82 ± 0.14	0.99 ± 0.01	0.39 ± 0.33	0.51 ± 0.42	1.00 ± 0.00

From our results, We found that our one phase model which was trained on 50 random patients including patients with high-grade gliomas and low-grade gliomas, with depth 45 performs well in case of Complete tumor and enhanced

Table 2. The table contains the random forest parameters and training datasets descriptions of our different models used in experiments

Model id	Maximum depth	Mtries	Ntrees	Dataset description
Model 1	45	13	100	50 patients are HGG (20 from 2013 BraTS dataset (subset of 2016 BraTS data set) and 30 randomly sampled from the rest of HGG)
Model 2	45	13	100	50 patient from both HGG and LGG (20 HGG, and 10 LGG from 2013 BraTS dataset (subset of 2016 BraTS data set), and 20 random sampled patients from the rest of patients)
Model 3	30	13	100	50 patient from both HGG and LGG (20 HGG, and 10 LGG from 2013 BraTS dataset (subset of 2016 BraTS data set), and 20 random sampled patients from the rest of patients)
Cascade	45	13	100	50 patient from both HGG and LGG (20 HGG, and 10 LGG from 2013 BraTS dataset (subset of 2016 BraTS data set), and 20 random sampled patients from the rest of patients)
Iterative 1	45	13	100	The initial dataset is 2013 BraTS data set (subset of 2016 BraTS data set), maximum number of patients is limited to 50 patient
Iterative 2	45	13	100	The initial dataset is subset of 2013 BraTS data set (15 HGG and 5 LGG) to give the model more flexibility to select more patients, maximum number of patients is limited to 50 patient

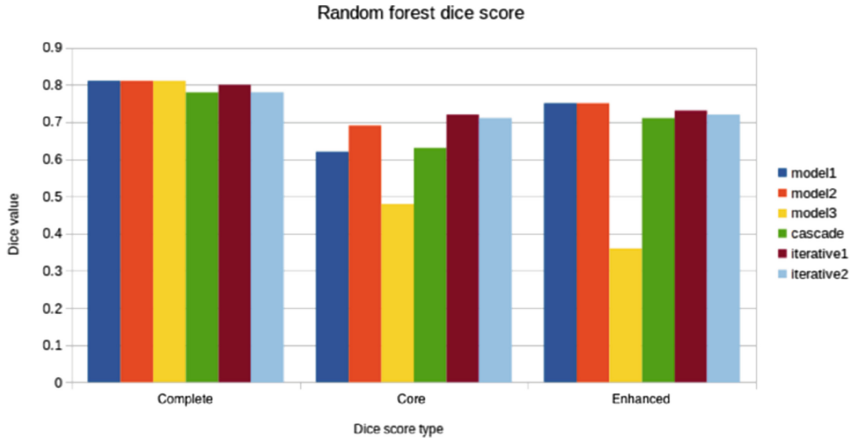


Fig. 4. The figure shows the dice scores of our different models described in Table 2

tumour hitting 81% and 74% respectively for high graded glioma, while our iterative model performs well in case of core tumor by passing 70% which was because the dataset of that model was mainly selected to include patients having different cases for core tumors. Also, we found that training a random forest at depth 30 on the same data on a depth of 45 performed very bad for core and enhanced tumors. Actually, we tried many other approaches like using all our data set as patients with high-grade gliomas and using cascaded approach by classifying first health and non-health voxels, then applying binary morphology and finally classifying the non-health voxels but it turns out that all the previous approaches don't out perform the iterative model.

The Graph in Fig. 4 shows the dice scores of our different models described in Table 2.

4 Conclusion

In this paper, we proposed an approach based on Random Forest that differs from past years' submissions as we mainly tried to extract as much information as we can from our large dataset (BRATS 2016 dataset). We achieved this by applying our iterative selection method to choose the best patients to train our Random Forest with them and then by extracting as much information as we can then applying feature selection, and our proposed method improves the performance over the cascaded method and over training the RF using randomly selected patients.

References

1. Tustison, N., Gee, J.: N4ITK: Nicks N3 ITK implementation for MRI bias field correction. *Insight J.* (2009)
2. Nyu, L.G., Udupa, J.K.: On standardizing the MR image intensity scale image, vol. 1081 (1999)
3. Nyu, L.G., Udupa, J.K.: New variants of a method of MRI scale standardization. *IEEE Trans. Med. Imaging* **19**(2), 143–150 (2000)
4. Dinis, H., Pinto, A., Pereira, S., Silva, C.A.: Random decision forests for automatic brain tumor segmentation in multi-modal MRI images
5. Peyrat, J.-M., Abinahed, J., Malmi, E., Parambath, S., Chawla, S.: CaBS: a cascaded brain tumor segmentation approach
6. Wilms, M., Maier, O., Handels, H.: Highly discriminative features for glioma segmentation in MR volumes with random forests
7. <http://www.h2o.ai/>
8. Bauer, S., Wiest, R., Nolte, L.P., Reyes, M.: A survey of MRI-based medical image analysis for brain tumor studies. *Phys. Med. Biol.* **58**(13), R97 (2013)
9. Menze, B., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* (2014)