

## Article

# A Fully Automatic Procedure for Brain Tumor Segmentation from Multi-Spectral MRI Records Using Ensemble Learning and Atlas-Based Data Enhancement

Ágnes Gyórfi <sup>1,2</sup> , László Szilágyi <sup>1,2,3\*</sup>  and Levente Kovács <sup>1,3</sup> 

<sup>1</sup> Physiological Research Control Center, Óbuda University, Bécsi Street 96/B, H-1034 Budapest, Hungary; gyorfi.agnes@phd.uni-obuda.hu (A.G.); kovacs@uni-obuda.hu (L.K.)

<sup>2</sup> Computational Intelligence Research Group, Sapiientia University, Șos. Sighișoarei 1/C, 540485 Corunca, Romania

<sup>3</sup> Biomatics Institute, John von Neumann Faculty of Informatics, Óbuda University, Bécsi Street 96/B, H-1034 Budapest, Hungary

\* Correspondence: lalo@ms.sapiientia.ro; Tel.: +36-70-242-8204



**Citation:** Gyórfi, Á.; Szilágyi, L.; Kovács, L. A Fully Automatic Procedure for Brain Tumor Segmentation from Multi-Spectral MRI Records Using Ensemble Learning and Atlas-Based Data Enhancement. *Appl. Sci.* **2021**, *11*, 564. <https://doi.org/10.3390/app11020564>

Received: 2 December 2020

Accepted: 1 January 2021

Published: 8 January 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 European Union. Licensee MDPI, Basel, Switzerland. This is an open access article distributed under the terms of the Creative Commons Attribution IGO License (<http://creativecommons.org/licenses/by/3.0/igo/legalcode>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. In any reproduction of this article there should not be any suggestion that European Union or this article endorse any specific organisation or products. The use of the European Union logo is not permitted. This notice should be preserved along with the article's original URL.

**Abstract:** The accurate and reliable segmentation of gliomas from magnetic resonance image (MRI) data has an important role in diagnosis, intervention planning, and monitoring the tumor's evolution during and after therapy. Segmentation has serious anatomical obstacles like the great variety of the tumor's location, size, shape, and appearance and the modified position of normal tissues. Other phenomena like intensity inhomogeneity and the lack of standard intensity scale in MRI data represent further difficulties. This paper proposes a fully automatic brain tumor segmentation procedure that attempts to handle all the above problems. Having its foundations on the MRI data provided by the MICCAI Brain Tumor Segmentation (BraTS) Challenges, the procedure consists of three main phases. The first pre-processing phase prepares the MRI data to be suitable for supervised classification, by attempting to fix missing data, suppressing the intensity inhomogeneity, normalizing the histogram of observed data channels, generating additional morphological, gradient-based, and Gabor-wavelet features, and optionally applying atlas-based data enhancement. The second phase accomplishes the main classification process using ensembles of binary decision trees and provides an initial, intermediary labeling for each pixel of test records. The last phase reevaluates these intermediary labels using a random forest classifier, then deploys a spatial region growing-based structural validation of suspected tumors, thus achieving a high-quality final segmentation result. The accuracy of the procedure is evaluated using the multi-spectral MRI records of the BraTS 2015 and BraTS 2019 training data sets. The procedure achieves high-quality segmentation results, characterized by average Dice similarity scores of up to 86%.

**Keywords:** ensemble learning; binary decision trees; image segmentation; multi-atlas methods; magnetic resonance imaging

## 1. Introduction

Cancers of the brain and the central nervous system cause the death of over two hundred thousand people every year [1]. Life expectancy after the diagnosis depends on several factors like: being a primary tumor or a metastatic one; being an aggressive form of tumor (also called high-grade glioma (HGG)) or a less aggressive one (low-grade glioma (LGG)); and of course, a key factor is how early the tumor is diagnosed [2]. Patients with HGG live fifteen months on average after diagnosis. With an LGG, it is possible to live for several years, as this form of the tumor does not always require aggressive treatment immediately after the diagnosis.

Magnetic resonance imaging (MRI) is the technology that has become the most frequently utilized in the diagnosis of gliomas. MRI is preferred because it is much less

invasive than other imaging modalities, like positron emission tomography (PET) or computed tomography (CT). With its high contrast and good resolution, it can provide accurate data about the structure of the tumor. Multi-modal or multi-spectral MRI, through its T1-weighted (with or without contrast improvement) and T2-weighted (with or without fluid attenuated inversion recovery) data channels, can significantly contribute to the better visibility of intracranial structures [3].

The quick development of computerized medical devices and the economic rise of several underdeveloped countries both contribute to the fast spreading of MRI equipment in hospitals worldwide. These MRI devices produce more and more image data. Training enough human experts to process these records would be very costly, if possible at all. This is why there is a strong need for automatic algorithms that can reliably process the acquired image data and select those records that need to be inspected by the human experts, who have the final word in establishing the diagnosis. Although such algorithms are never perfect, they may contribute to cost reduction and allow for screening large masses of the population, leading to several early detected tumors.

The brain tumor segmentation problem is a difficult task in medical image processing, having major obstacles like: (1) the large variety of the locations, shapes, and appearances of the tumor; (2) displacement and distortion of normal tissues caused by the focal lesion; (3) the variety of imaging modalities and weighting schemes applied in MRI, which provide different types of biological information; (4) multi-channel data are not perfectly registered together; (5) numerical values produced by MRI do not directly reflect the observed tissues; they need to be interpreted in their context; (6) intensity inhomogeneity may be present in MRI measurements due to the turbulence of the magnetic field.

The history of automatic brain tumor segmentation from MRI records can be divided into two eras: the pre-BraTS and the BraTS era, where BraTS refers to the Brain Tumor Segmentation Challenges [3,4] organized every year since 2012 by the MICCAI society, which had an enormous impact with the introduction of a multi-spectral MRI data set that can be used as a standard in the evaluation of segmentation methods. Earlier solutions were usually developed for single data channels and even for 2D data (or reduced number of distant slices) and mostly validated with private collections of MRI records that do not allow for objective comparison. A remarkable review on early segmentation methods, including manual, semi-automatic, and fully automatic solutions, was provided by Gordillo et al. [5]. Methods developed in the BraTS era are usually fully automatic and employ either one or a combination of: (1) advanced general-purpose image segmentation techniques (mostly unsupervised); (2) classical machine learning algorithms (both supervised and unsupervised); or (3) deep learning convolutional neural networks (supervised learning methods).

All methods developed and published in the pre-BraTS era belong to the first two groups or their combination. They dominate the first years of the BraTS era as well, and they are still holding significant research interest. Unsupervised methods have the advantages that they do not require large amounts of training data and provide segmentation results in a relatively short time. They organize the input data into several clusters, each consisting of highly similar data. However, they either have difficulty with correctly labeling the clusters or they require manual interaction. Unsupervised methods involving active contours or region growing strongly depend on initialization as well. Sachdeva et al. [6] proposed a content-based active contour model that utilized both intensity and texture information to evolve the contour towards the tumor boundary. Njeh et al. [7] proposed a quick unsupervised graph-cut algorithm that performed distribution matching and identified tumor boundaries from a single data channel. Li et al. [8] combined the fuzzy *c*-means (FCM) clustering algorithm with spatial region growing to segment the tumors, while Szilágyi et al. [9] proposed a cascade of FCM clustering steps placed into a semi-supervised framework.

Supervised learning methods deployed for tumor segmentation first construct a decision model using image-based handcrafted features and use this for prediction in the

testing phase. These methods mainly differ from each other in the employed classification technique, the features involved, and the extra processing steps that apply constraints to the intermediary segmentation outcome. Islam et al. [10] extracted so-called multi-fractional Brownian motion features to characterize the tumor texture and compared its performance with Gabor wavelet features, using a modified AdaBoost classifier. Tustison et al. [11] built a supervised segmentation procedure by cascading two random forest (RF) classifiers and trained them using first order statistical, geometrical, and asymmetry-based features. They used a Markov random field (MRF)-based segmentation to refine the probability maps provided by the random forests. Pinto et al. [12] deployed extremely randomized trees (ERTs) and trained them with features extracted from local intensities and neighborhood context, to perform a hierarchical segmentation of the brain tumor. Soltaninejad et al. [13] reformulated the simple linear iterative clustering (SLIC) [14] for the extraction of 3D superpixels, extracted statistical and texton feature from these, and fed them to an RF classifier to distinguish superpixels belonging to tumors and normal tissues. Imtiaz et al. [15] classified the pixels of MRI records with ERTs, using superpixel features from three orthogonal planes. Kalaiselvi et al. [16] proposed a brain tumor segmentation procedure that combined clustering techniques, region growing, and support vector machine (SVM)-based classification. Supervised learning methods usually provide better segmentation accuracy than unsupervised ones and require a longer processing time.

In recent years, convolutional neural networks (CNNs) and deep learning have been attempting to conquer a wide range of application fields in pattern recognition [17]. Features are no longer handcrafted, as these architectures automatically build a hierarchy of increasingly complex features based on the learning data [18]. There are several successful applications in medical image processing, e.g., detection of kidney abnormalities [19], prostate cancer [20], lesions caused by diabetic retinopathy [21] and melanoma [22], and the segmentation of liver [23], cardiac structures [24], colon [25], renal artery [26], mandible [27], and bones [28]. The brain tumor segmentation problem is no exception. In this order, Pereira et al. [29] proposed a CNN architecture with small  $3 \times 3$  sized kernels and accomplished a thorough analysis of data augmentation techniques for glioma segmentation attempting to compensate the imbalance of classes. Zhao et al. [30] applied conditional random fields (CRF) to process the segmentation output produced by a fully convolutional neural network (FCNN) and assure the spatial and appearance consistency. Wu et al. [31] proposed a so-called multi-feature refinement and aggregation scheme for convolutional neural networks that allows for a more effective combination of features and leads to more accurate segmentation. Kamnitsas et al. [32] proposed a deep CNN architecture with 3D convolutional kernels and a double pathway learning, which exploits a dense inference technique applied to image segments, thus reducing the computational burden. Ding et al. [33] successfully combined the deep residual networks with the dilated convolution, achieving fine segmentation results. Xue et al. [34] proposed a solution inspired by the concept of generative adversarial networks (GANs). They set up a CNN to produce pixel level labeling, complemented it with an adversarial critic network, and trained them to learn local and global features that were able to capture both short and long distance relationships among pixels. Chen et al. [35] proposed a dual force-based learning strategy, employed the DeepMedic (Version 0.7.0, Biomedical Image Analysis Group, Imperial College London, London, UK, 2018) and U-Net architectures (Version 1.0, Computer Science Department, University of Freiburg, Freiburg, Germany, 2015) for glioma segmentation, and refined the output of deep networks using a multi-layer perceptron (MLP). In general, CNN architectures and deep learning can lead to slightly better accuracy than well designed classical machine learning techniques in the brain tumor detection problem, at the cost of a much higher computational burden in all processing phases, especially at training.

The goal of this paper is to propose a solution to the brain tumor segmentation problem that produces a high-quality result with a reduced amount of computation, without needing special hardware that might not be available in underdeveloped countries, employing classification via ensemble learning assisted by several image processing tasks designed

for MRI records that may contain focal lesions. Further, the proposed procedure only employs machine learning methods that fully comply with the explainable decision making paradigm, which is soon going to be required by law in the European Union [36]. Multiple pre-processing steps are utilized for bias field reduction, histogram normalization, and atlas-based data enhancement. A twofold post-processing is employed to improve the spatial consistency of the initial segmentation produced by the decision ensemble.

Ensemble learning generally combines several classifiers and aggregates their predictions into a final one, thus allowing for more accurate decisions than the individual classifiers are capable of [37,38]. The ensembles deployed in this study consist of binary decision trees (BDTs) [39]. BDTs are preferred because of their ability to learn any complex patterns that contain no contradiction. Further, with our own implementation of the BDT model, we have full control of its functional parameters.

In image segmentation problems, atlases and multiple atlases are generally employed to provide shape or texture priors and to guide the segmentation toward a regularized solution via label fusion [40–42]. Our solution uses multiple atlases in a different manner: before proceeding to ensemble learning and testing, atlases are trained to characterize the local appearance of normal tissues and applied to transform all feature values to emphasize their deviation from normal.

The main contributions consist of: (1) the way the multiple atlases are involved in the pre-processing, to prepare the feature data for segmentation via ensemble learning; (2) the ensemble model built from binary decision trees with unconstrained depth; (3) the two-stage post-processing scheme that discards a large number of false positives.

Some components of the procedure proposed here, together with their preliminary results, were presented in previous works. For example, we can mention some part of the feature generation and the classification with ensembles of binary decision trees [43] or the multi-atlas assisted data enhancement of the MRI data [44]. However, the procedure proposed here puts them all together and proposes additional components like the post-processing steps, and a thorough evaluation process is performed using the BraTS training data sets for the years 2015 and 2019. The evaluation process demonstrates that the proposed brain tumor segmentation procedure is in competition with state-of-the-art methods both in terms of accuracy and efficiency.

The rest of this paper is structured as follows: Section 2 presents the proposed brain tumor segmentation procedure, with all its steps starting from histogram normalization and feature generation, followed by initial decision making via an ensemble learning technique, and a twofold post-processing that produces the final segmentation. Section 3 evaluates the proposed segmentation procedure using recent standard data sets of the BraTS challenge. Section 4 compares the performance of the proposed method with several state-of-the-art algorithms, from the point of view of both accuracy and efficiency. Section 5 concludes this study.

## 2. Materials and Methods

### 2.1. Overview

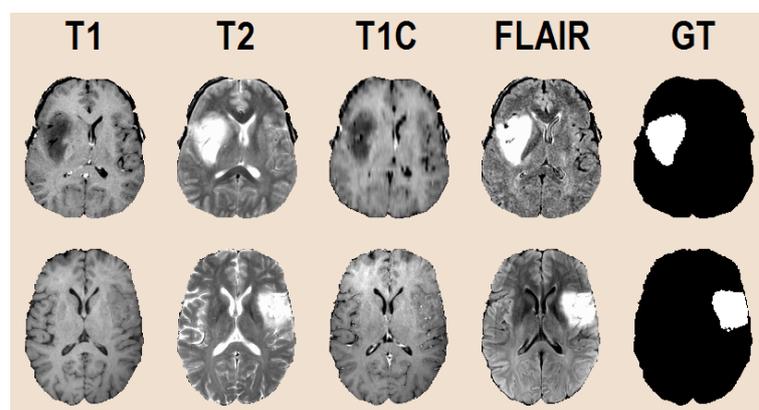
The steps of the proposed method are exhibited in Figure 1. All MRI data volumes involved in this study go through a multi-step pre-processing, whose main goal is to provide uniform histograms to all data channels of the involved MRI records and to generate further features to all pixels. After separating training data records from evaluation (test) data records, the former are fed to the ensemble learning step. Trained ensembles are used to provide an initial prediction for the pixels of the test data records. A twofold post-processing is employed to regularize the shape of the estimated tumor. Finally, statistical markers are used to evaluate the accuracy of the whole proposed procedure.



third of the pixels, in one of the four data channels. An overview of these cases is also reported in Table 1.

**Table 1.** The main properties of the MRI data sets involved in this study.

Property	BraTS 2015 Data		BraTS 2019 Data	
	LGG	HGG	LGG	HGG
Number of records	54	220	76	259
Average size of whole tumor (cm <sup>3</sup> )	101.1	110.6	111.5	95.1
Minimum size of whole tumor (cm <sup>3</sup> )	18.9	8.5	17.1	7.3
Maximum size of whole tumor (cm <sup>3</sup> )	256.3	318.3	361.8	256.9
Total number of negative pixels	71.4 M	308.6 M	101.6 M	351.5 M
Total number of positive pixels	5.46 M	24.3 M	8.48 M	24.6 M
Pixels with missing data	2.375%	9.855%	1.679%	0.426%
Pixels with more than 1 value missing	0.225%	4.838%	0.163%	0.092%
Records with missing data at >1% of pixels	9	112	9	18
Records with missing data at >10% of pixels	3	99	3	1



**Figure 2.** Input data after pre-processing (T1, T1C, T2, FLAIR) and the human expert made ground truth (GT) for the whole tumor segmentation problem.

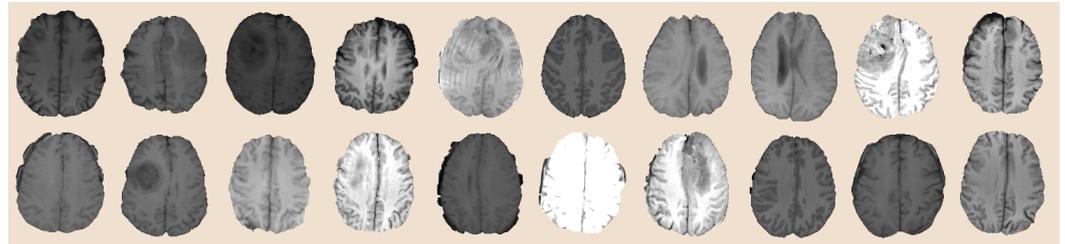
Intensity non-uniformity (INU) is a low-frequency noise with possibly high magnitude [45,46]. In the early years of the BraTS challenges, INU was filtered from the training and test data provided by the organizers. However, in later data sets, this facility is not granted. This is why we employed the method of Tustison et al. [47] to check the presence of INU and to eliminate it when necessary.

There are several non-brain pixels in the volume of any record, a fact indicated by zero intensity on all four data channels. Furthermore, there are some brain pixels in most volumes for which not all four nonzero intensity values exist; some of them are missing. Our first option to fill missing values is to replace them with the average computed from the  $3 \times 3 \times 3$  cubic neighborhood of the pixel, if there are any pixels in the neighborhood with valid nonzero intensity. Otherwise, the missing value becomes  $\gamma_{0.5}$  after pre-processing, which is the middle value of the target intensity interval; see the definition in Section 2.3.

### 2.3. Histogram Normalization

Although very popular among medical imaging devices due to its high contrast and relatively good resolution, MRI has a serious drawback: the numerical values it produces do not directly reflect the observed tissues. The correct interpretation of the MRI data requires adapting the numerical values to the context, which is usually accomplished via a histogram normalization or standardization. Figure 3 gives a convincing example why the normalization of histograms is necessary. This is what we get if we represent on the same

scale the observed intensities from twenty different MRI records, extracted from the same data channel (T1). In some cases, the spectra of intensities from two different records do not even overlap. There is no classification method that could deal with such intensities without transforming the input data before classification.



**Figure 3.** Slice Number 100 from the T1 volume of twenty MRI records from the BraTS 2015 LGG data, without histogram normalization. These pixel intensities are obviously not suitable for any supervised classification.

Several algorithms exist in the literature for this purpose (e.g., [48–50]), more and more sophisticated histogram matching techniques, which unfortunately are not designed for records containing focal lesions. These lesions may grow to 20–30% of the brain volume, which strongly distorts the histogram in MRI channels of any weighting. The most popular histogram normalization method proposed by Nyúl et al. [48] works in batch mode, registering the histogram of each record to the same template. Modifying all histograms to look similar, no matter whether they contain tumor or not, is likely to obstruct the accurate segmentation. The method of Nyúl et al. uses some predefined landmark points (percentiles) to accomplish a two-step transformation of intensities. Many recent works [51–55] reported using this method, omitting any information on the employed landmark points. Pinto et al. [12] and Soltaninejad et al. [13] used 10 and 12 landmark points, respectively. Alternately, Tustison et al. [11] deployed a simple linear transformation of intensities, because it gives slightly better accuracy, but they did not specify any details. This finding was confirmed later in a comparative study by Györfi et al. [56], which also showed that if we stick to the above-mentioned popular method of Nyúl et al., it provides better segmentation of focal lesions when used with a smaller number (no more than five) of landmark points, or in other words, with less restrictions.

Our histogram normalization method relies on a context dependent linear transform. The main goal of this transform is to map the pixel intensities from each data channel of a volume independently of each other onto a target interval  $[\alpha, \beta]$  in such a way that target intensities satisfy some not very restrictive criteria. The inner values of the target interval can be written as  $\gamma_\lambda = \alpha(1 - \lambda) + \beta\lambda$ , with  $0 \leq \lambda \leq 1$ . Obviously,  $\alpha = \gamma_0$  and  $\beta = \gamma_1$ . To compute the two coefficients of the linear transform applied to the intensities from a certain data channel of any record, first we identify the 25th-percentile and 75th-percentile value of the original intensities and denote them by  $p_{25}$  and  $p_{75}$ , respectively. Then, we define the  $I \rightarrow aI + b$  transform of the intensities in such a way that  $p_{25}$  is mapped onto  $\gamma_{0.4}$  and  $p_{75}$  becomes  $\gamma_{0.6}$ . This assumption leads to the coefficients:

$$a = \frac{\gamma_{0.6} - \gamma_{0.4}}{p_{75} - p_{25}} \text{ and } b = \frac{\gamma_{0.4}p_{75} - \gamma_{0.6}p_{25}}{p_{75} - p_{25}} . \quad (1)$$

Finally, all pixel intensities from the given data channel of the current MRI record are transformed according to the formula:

$$\hat{I} = \min\{\max\{\alpha, aI + b\}, \beta\} , \quad (2)$$

using the values  $a$  and  $b$  given by Equation (1), where  $I$  and  $\hat{I}$  stand for the original and transformed intensities, respectively. This formula assures that the transformed intensities are always situated within the target interval  $[\alpha, \beta]$ .

### 2.4. Feature Generation

Feature generation is a key component of the procedure, because the four observed features, namely the T1, T2, T1C, and FLAIR intensities provided by the MRI device, do not contain all the possible discriminative information that can be employed to distinguish tumor pixels from normal ones. A major motivation for feature generation represents the fact that the automated registration algorithm used by the BraTS experts to align the four data channels never performs a perfect job, so for an arbitrary cubic millimeter of brain tissues, represented by the pixel situated at coordinates  $(x, y, z)$  in volume T1, the corresponding information in the other volumes is not in the same place, but somewhere in the close neighborhood of  $(x, y, z)$ . A further motivation consists of the usual property of image data that neighbor pixels correlate with each other.

All four MRI data channels equally contribute to the feature generation process. Observed features are treated independently of each other: 20 computed features are extracted from each of them. The inventory of extracted features is presented in Table 2. Minimum, maximum, and average values are extracted in such a way that only the brain pixels are taken into consideration from the neighborhoods indicated in the table. The computation of the gradient features involves the masks presented in Figure 4. The current pixel is part of all masks to avoid division by zero in Equation (3). The 16 gradient features of an arbitrary pixel  $p$  are computed with the formula:

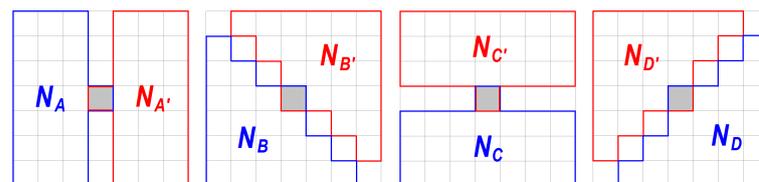
$$g_m^{(c)}(p) = \gamma_{0.5} + k_g \left( \frac{\sum_{q \in N_m(p) \cap \bar{\Omega}} I_q^{(c)}}{|N_m(p) \cap \bar{\Omega}|} - \frac{\sum_{q \in N_{m'}(p) \cap \bar{\Omega}} I_q^{(c)}}{|N_{m'}(p) \cap \bar{\Omega}|} \right), \tag{3}$$

followed by  $g_m^{(c)}(p) \leftarrow \min\{\max\{\alpha, g_m^{(c)}(p)\}, \beta\}$  to stay within the target interval  $[\alpha, \beta]$ , where  $c \in \{T1, T2, T1C, FLAIR\}$  is one of the four data channels,  $m \in \{A, B, C, D\}$  is the index of the current gradient mask,  $\bar{\Omega}$  stands for the set of all brain pixels in the volume,  $\gamma_{0.5}$  is the middle of the target interval of intensities,  $k_g$  is a globally constant scaling factor, and  $|Q|$  represents the cardinality of any set denoted by  $Q$ .

**Table 2.** Inventory of computed features. Observed data channels were equally involved. Together with the four observed features, the total feature count is  $n_\varphi = 84$ .

Neighborhood	AVG	MAX	MIN	GRD	GBW	Total
$3 \times 3 \times 3$	4	4	4			12
$3 \times 3$	4					4
$5 \times 5$	4					4
$7 \times 7$	4			16		20
$9 \times 9$	4					4
$11 \times 11$	4				32	36
Total	24	4	4	16	32	80

AVG—average, MAX—maximum, MIN—minimum. GRD—gradient, GBW—Gabor wavelets.



**Figure 4.** Neighborhoods used for the extraction of gradient features, with respect to the current pixel indicated by grey color.

### 2.5. Multi-Atlas-Based Data Enhancement

In image segmentation, atlases are usually used as approximative maps of the objects that should be present in the image in the normal case. Atlases are usually established based on prior measurements. The image data before the current segmentation is registered to the atlas, and the current segmentation outcome is fused with the atlas to obtain a final, atlas assisted segmentation result.

In this study, we use atlases in a different manner. We build a spatial atlas for each feature using the training records, which contains the local average and standard deviation of the intensity values taken from normal pixels only. Of course, the training records need to be registered to each other so that we can build consistent atlases. These atlases are then used to refine the feature values in both the training and test records, before proceeding to the ensemble learning-based classification.

First of all, it is important to define the spatial resolution of atlases, which depends on a single parameter denoted by  $S$ . Each atlas is a discrete spatial array defined on  $\mathcal{S}^3$ , where  $\mathcal{S} = \{-S, -S + 1, \dots, 0, \dots, S - 1, S\}$ . Within the atlas array, the neighborhood of atlas point  $\hat{\pi}$  having the coordinates  $(\hat{x}, \hat{y}, \hat{z})$  is defined as:

$$C_\delta(\hat{\pi}) = \{(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) \in \mathcal{S}^3, |\hat{\alpha} - \hat{x}| \leq \delta \wedge |\hat{\beta} - \hat{y}| \leq \delta \wedge |\hat{\gamma} - \hat{z}| \leq \delta\}, \quad (4)$$

where  $\delta$  is a small positive integer, typically one or two, which determines the size of the neighborhood.

The mathematical description of the atlas building process requires the introduction of several notations. Let  $\mathcal{M}$  stand for the set of all MRI records of the chosen data set, while  $\mathcal{M}^{(T)}$  and  $\mathcal{M}^{(E)}$  represent the current set of training and evaluation (or test) records, respectively. Obviously,  $\mathcal{M} = \mathcal{M}^{(T)} \cup \mathcal{M}^{(E)}$ , and  $\mathcal{M}^{(T)} \cap \mathcal{M}^{(E)} = \Phi$ . Let  $M_i$  be the record with index  $i$ , belonging to either  $\mathcal{M}^{(T)}$  or  $\mathcal{M}^{(E)}$  in a certain scenario. The set of pixels belonging to record  $M_i$  is denoted by  $\Omega_i$ , for any  $i = 1 \dots n_\rho$ . The set of all pixels from all MRI records is  $\Omega = \bigcup_{i=1}^{n_\rho} \Omega_i$ . Any pixel  $\pi \in \Omega$  has a feature vector with  $n_\phi$  elements. For any pixel  $\pi$ , the value of the feature with index  $\phi \in \{1 \dots n_\phi\}$  is denoted by  $I_\pi^{(\phi)}$ . Further, let  $\Gamma^{(v)}$  and  $\Gamma^{(\tau)}$  be the set of negative and positive pixels, respectively, as indicated by the ground truth. Obviously,  $\Omega = \Gamma^{(v)} \cup \Gamma^{(\tau)}$ , and  $\Gamma^{(v)} \cap \Gamma^{(\tau)} = \Phi$ .

A rigid registration is defined in the following, so that we can map all volumes onto the atlas. For each record  $M_i$  ( $i = 1 \dots n_\rho$ ), a function  $f_i : \Omega_i \rightarrow \mathcal{S}^3$  is needed that maps the pixels onto the atlas, to find the corresponding atlas position for all brain pixels. These functions map the gravity center of each brain to the atlas origin, and the standard deviations of the  $x$ ,  $y$ , and  $z$  coordinates are all transformed to  $S/\zeta$ , where  $\zeta = 2.5$  is a predefined constant. From the pixel coordinates, we compute averages and standard deviations as follows:

$$\begin{cases} \mu_x^{(i)} = \frac{1}{|\Omega_i|} \sum_{\pi \in \Omega_i} x_\pi \\ \mu_y^{(i)} = \frac{1}{|\Omega_i|} \sum_{\pi \in \Omega_i} y_\pi \\ \mu_z^{(i)} = \frac{1}{|\Omega_i|} \sum_{\pi \in \Omega_i} z_\pi \end{cases}, \quad (5)$$

and then:

$$\begin{cases} \sigma_x^{(i)} = \sqrt{\frac{1}{|\Omega_i|-1} \sum_{\pi \in \Omega_i} (x_\pi - \mu_x^{(i)})^2} \\ \sigma_y^{(i)} = \sqrt{\frac{1}{|\Omega_i|-1} \sum_{\pi \in \Omega_i} (y_\pi - \mu_y^{(i)})^2} \\ \sigma_z^{(i)} = \sqrt{\frac{1}{|\Omega_i|-1} \sum_{\pi \in \Omega_i} (z_\pi - \mu_z^{(i)})^2} \end{cases}, \quad (6)$$

where  $x_\pi$ ,  $y_\pi$ , and  $z_\pi$  are the coordinates of pixel  $\pi$  in the brain volume. The formula of the mapping  $f_i$  is:

$$f_i(\pi) = \left( \left\langle \frac{S(x_\pi - \mu_x^{(i)})}{\xi \sigma_x^{(i)}} \right\rangle, \left\langle \frac{S(y_\pi - \mu_y^{(i)})}{\xi \sigma_y^{(i)}} \right\rangle, \left\langle \frac{S(z_\pi - \mu_z^{(i)})}{\xi \sigma_z^{(i)}} \right\rangle \right), \tag{7}$$

where  $\langle \cdot \rangle$  stands for the operation of rounding a floating point variable to the closest integer.

For any feature with index  $\varphi \in \{1 \dots n_\varphi\}$ , the atlas function has the form  $A_\varphi : \mathcal{S}^3 \rightarrow \mathbb{R}^3$ , which for any atlas point  $\hat{\pi} \in \mathcal{S}^3$  is defined as:

$$A_\varphi(\hat{\pi}) = \left( \hat{\mu}_{\hat{\pi}}^{(\varphi)}, \hat{\sigma}_{\hat{\pi}}^{(\varphi)}, \hat{\nu}_{\hat{\pi}}^{(\varphi)} \right), \tag{8}$$

where the components  $\hat{\mu}_{\hat{\pi}}^{(\varphi)}$ ,  $\hat{\sigma}_{\hat{\pi}}^{(\varphi)}$ , and  $\hat{\nu}_{\hat{\pi}}^{(\varphi)}$  are established with the following formulas:

$$\begin{cases} \hat{\nu}_{\hat{\pi}}^{(\varphi)} = \sum_{M_i \in \mathcal{M}^{(T)}} |\Psi(i, \hat{\pi})| \\ \hat{\mu}_{\hat{\pi}}^{(\varphi)} = \left( \hat{\nu}_{\hat{\pi}}^{(\varphi)} \right)^{-1} \sum_{M_i \in \mathcal{M}^{(T)}} \left( \sum_{\pi \in \Psi_i(\hat{\pi})} I_\pi^{(\varphi)} \right) \\ \hat{\sigma}_{\hat{\pi}}^{(\varphi)} = \sqrt{\left( \hat{\nu}_{\hat{\pi}}^{(\varphi)} - 1 \right)^{-1} \sum_{M_i \in \mathcal{M}^{(T)}} \left( \sum_{\pi \in \Psi_i(\hat{\pi})} \left( I_\pi^{(\varphi)} - \hat{\mu}_{\hat{\pi}}^{(\varphi)} \right)^2 \right)} \end{cases} \tag{9}$$

where:

$$\Psi_i(\hat{\pi}) = \left\{ \pi \in \Omega_i \cap \Gamma^{(\nu)}, f_i(\pi) \in \mathcal{C}_\delta(\hat{\pi}) \right\}. \tag{10}$$

The feature values of each pixel  $\pi \in \Omega_i$  ( $i = 1 \dots n_\rho$ ), no matter whether  $\pi$  belongs to a record of the training or test data, are updated with the following formula:

$$\tilde{I}_\pi^{(\varphi)} \leftarrow \min \left\{ \max \left\{ \alpha, \left\langle \bar{\mu} + \bar{\sigma} \frac{I_\pi^{(\varphi)} - \hat{\mu}_{f_i(\pi)}^{(\varphi)}}{\hat{\sigma}_{f_i(\pi)}^{(\varphi)}} \right\rangle \right\}, \beta \right\}, \tag{11}$$

where parameters  $\bar{\mu}$  and  $\bar{\sigma}$  represent the target average and standard deviation, respectively, and their recommended values are:

$$\begin{cases} \bar{\mu} = \gamma_{0.5} = (\alpha + \beta) / 2 \\ \bar{\sigma} = (\beta - \alpha) / 10 \end{cases}. \tag{12}$$

Algorithm 1 summarizes the construction and usage of the atlas. The updated intensity values compose the pre-processed data denoted by  $P_2$  in Figure 1. Both  $P_1$  and  $P_2$  follow the very same classification and post-processing steps.

### 2.6. Ensemble Learning

The total number of  $n_\rho$  records is randomly separated into six equal or almost equal groups. Any of these groups can serve as test data, while the other five groups are used to train the decision making ensemble. The ensembles consist of binary decision trees, whose main properties are listed in the following:

1. The size of the ensemble is defined by the number of trees, set to  $n_T = 125$  based on intermediary tests performed on the out-of-bag (OOB) data, which are the set of feature vectors from the training data that are never selected for training the trees of a given ensemble.
2. The training data size, represented by the number of randomly selected feature vectors used to train each tree, and denoted by  $n_F$ , is chosen to vary between 100 thousand and one million in four steps:  $n_F \in \{100k, 200k, 500k, 1000k\}$ . These four values are

- used during the experimental evaluation of the proposed method to establish its effect on the segmentation accuracy and efficiency.
- The selection of feature vectors for training the decision trees is not totally random. The balance of positive and negative feature vectors in the training set of each BDT is defined by the rate of negatives  $p_-$  and rate of positives  $p_+$ . For the LGG and HGG data sets,  $p_- = 93\%$  and  $p_+ = 91\%$  are used, respectively, values set according to the findings of intermediary tests performed on OOB data.
  - Trees are trained using an entropy-based rule, without limiting their depth, so that they can provide perfect separation of feature vectors representing pixels belonging to tumors and normal tissues. The trees of the ensemble have their maximum depth ranging between 25 and 45, depending on  $n_F$  and the random training data itself.

---

**Algorithm 1:** Build the atlas function  $A_\varphi$  for each feature  $\varphi = 1 \dots n_\varphi$ , and apply it to all feature data.

---

**Data:** Set of MRI records involved in the study  $\mathcal{M}$ , its set of pixels  $\Omega$ , each pixel with 4 observed and  $n_\varphi - 4$  computed features.

**Data:** Parameters  $S$  and  $\delta$

**Result:** Atlas functions  $A_\varphi$ ,  $\varphi = 1 \dots n_\varphi$

**Result:** Updated intensities in all records of  $\mathcal{M}$

Define the set of training and test records,  $\mathcal{M}^{(T)}$  and  $\mathcal{M}^{(E)}$ , respectively.

**for each**  $M_i \in \mathcal{M}$  **do**

    Compute the mapping function  $f_i(\pi)$  for every pixel  $\pi \in \Omega_i$  using Equations (5)–(7).

**end**

**for**  $\varphi = 1 \dots n_\varphi$  **do**

    Compute the atlas function  $A_\varphi(\hat{\pi})$  for every discrete point  $\hat{\pi} \in \mathcal{S}^3$  having  $\hat{v}_{\hat{\pi}}^{(\varphi)} > 1$ , using Equations (8)–(10).

**for each**  $M_i \in \mathcal{M}$  **do**

**for each**  $\pi \in \Omega_i$  **do**

            Update the value of feature  $\varphi$  of pixel  $\pi$  using the formula given in Equation (11).

**end**

**end**

**end**

---

When all trees of the ensemble are trained, they can be used to obtain the prediction for the feature vectors from the test data records. Each pixel from the test records receives  $n_T$  votes, one from each BDT. The decision of the ensemble is established by the majority rule, and that gives a label to each test pixel. However, these are intermediary labels only, as they undergo a twofold post-processing. The intermediary labels received by test pixels compose the  $S_1$  and  $S_2$  segmentation results indicated in Figure 1.

### 2.7. Random Forest-Based Post-Processing

The first post-processing step reevaluates the initial label received by each pixel of the test data records. The decision is made by a random forest classifier that relies on morphological features. The details of this post-processing step are listed in the following:

- The RF is trained to separate positive and negative pixels using six features extracted from the intermediary labels produced by the BDT ensemble. Let us consider  $K = 5$  concentric cubic neighborhoods of the current pixel, denoted by  $\mathcal{N}_k$  ( $k = 1 \dots K$ ), each having the size  $(2k + 1) \times (2k + 1) \times (2k + 1)$ . Inside the neighborhood  $\mathcal{N}_k$  of the current pixel, the count of brain pixels  $n_k$  and the count of positive intermediary labeled brain pixels  $n_k^+$  is extracted. The ratio  $\rho_k = n_k^+ / n_k$  is called the rate of positives within neighborhood  $\mathcal{N}_k$ . The feature vector has the form  $(\rho_1, \rho_2, \dots, \rho_K, \eta)$ , where  $\eta$  is

the normalized value of the number of complete neighborhoods of the current pixel, determined as:

$$\eta = \frac{1}{K} \sum_{k=1}^K \delta(n_k, (2k+1)^3), \quad (13)$$

where:

$$\delta(u, v) = \begin{cases} 1 & \text{if } u = v \\ 0 & \text{otherwise} \end{cases}. \quad (14)$$

2. To assure that testing runs on data never seen by the training process, pixels from HGG tumor volumes are used to train the random forest applied in the segmentation of LGG tumor volumes, and vice versa. Each forest is trained using the feature vectors of  $10^7$  randomly selected voxels, whose feature values fulfil  $\sum_{k=1}^K \rho_k > 0$ .
3. Pixels whose features satisfy  $\sum_{k=1}^K \rho_k = 0$  are declared negatives by default; they are not used for training the RF and are not tested with the RF either.
4. The number of trees in the RF is set to 100, while the maximum allowed depth of the trees is eight.

The result of this post-processing step can be seen in Figure 1, represented by segmentation results  $S'_1$  and  $S'_2$ .

### 2.8. Structural Post-Processing

The structural post-processing handles only pixels that are labeled positive in segmentation results  $S'_1$  and  $S'_2$ ; consequently, it has the option to approve or discard the current positive labels. As a first operation, it searches for contiguous spatial regions formed by positive pixels within the volume using a region growing algorithm. Contiguous regions of positive labeled pixels with a cardinality below 100 are discarded, because such small lesions cannot be reliably declared gliomas. Larger lesions are subject to shape-based validation. For this purpose, the coordinates of all positive pixels belonging to the current contiguous region undergo a principal component analysis (PCA), which establishes the three main axes determined by the three eigenvectors and the corresponding radii represented by the square root of the three eigenvalues provided by PCA. We denote by  $\lambda_1 > \lambda_2 > \lambda_3$  the three eigenvalues in decreasing order. Lesions having the third radius below a predefined threshold ( $\sqrt{\lambda_3} < 2$ ) are discarded, as they are considered unlikely shapes for a glioma. All those detected lesions that are not discarded by the criteria presented above are finally declared gliomas, and all their pixels receive final positive labels. This is the final solution denoted by  $S''_1$  and  $S''_2$  in Figure 1.

### 2.9. Evaluation Criteria

The whole set of pixels of the MRI record with index  $i$  is  $\Omega_i$ , which is separated by the ground truth into two disjoint sets:  $\Gamma_i^{(\pi)}$  and  $\Gamma_i^{(v)}$ , the set of positive and negative pixels, respectively. The final segmentation result provides another separation of  $\Omega_i$  into two disjoint sets denoted by  $\Lambda_i^{(\pi)}$  and  $\Lambda_i^{(v)}$ , which represent the positive and negative pixels, respectively, according to the final decision (labeling) of the proposed procedure. Statistical accuracy indicators reflect in different ways how much the subsets  $\Gamma_i^{(\pi)}$  and  $\Lambda_i^{(\pi)}$  and their complementary subsets  $\Gamma_i^{(v)}$  and  $\Lambda_i^{(v)}$  overlap. The most important accuracy markers obtained for any record with index  $i$  ( $i = 1 \dots n_\rho$ ) are:

1. Dice similarity coefficient (DSC) or Dice score:

$$\text{DSC}_i = \frac{2 \times |\Gamma_i^{(\pi)} \cap \Lambda_i^{(\pi)}|}{|\Gamma_i^{(\pi)}| + |\Lambda_i^{(\pi)}|}, \quad (15)$$

2. Sensitivity or true positive rate (TPR):

$$\text{TPR}_i = \frac{|\Gamma_i^{(\pi)} \cap \Lambda_i^{(\pi)}|}{|\Gamma_i^{(\pi)}|}, \quad (16)$$

3. Specificity or true negative rate (TNR):

$$\text{TNR}_i = \frac{|\Gamma_i^{(v)} \cap \Lambda_i^{(v)}|}{|\Gamma_i^{(v)}|}, \quad (17)$$

4. Positive predictive value:

$$\text{PPV}_i = \frac{|\Gamma_i^{(\pi)} \cap \Lambda_i^{(\pi)}|}{|\Lambda_i^{(\pi)}|}, \quad (18)$$

5. Accuracy or rate of correct decisions (ACC):

$$\text{ACC}_i = \frac{|\Gamma_i^{(\pi)} \cap \Lambda_i^{(\pi)}| + |\Gamma_i^{(v)} \cap \Lambda_i^{(v)}|}{|\Gamma_i^{(\pi)}| + |\Gamma_i^{(v)}|}. \quad (19)$$

To characterize the global accuracy, we may compute the average values of the above-defined indicators over a whole set of MRI records. We denote them by  $\overline{\text{DSC}}$ ,  $\overline{\text{TPR}}$ ,  $\overline{\text{TNR}}$ ,  $\overline{\text{PPV}}$ , and  $\overline{\text{ACC}}$ . The average Dice similarity coefficient is given by the formula:

$$\overline{\text{DSC}} = \frac{1}{n_\rho} \sum_{i=1}^{n_\rho} \text{DSC}_i. \quad (20)$$

The average values of other accuracy indicators are computed with analogous formulas.

The overall Dice similarity coefficient ( $\widetilde{\text{DSC}}$ ) is a single Dice score extracted from the set of all pixels from  $\Omega$ , given by the formula:

$$\widetilde{\text{DSC}} = \frac{2 \times \left| \bigcup_{i=1}^{n_\rho} \Gamma_i^{(\pi)} \cap \bigcup_{i=1}^{n_\rho} \Lambda_i^{(\pi)} \right|}{\left| \bigcup_{i=1}^{n_\rho} \Gamma_i^{(\pi)} \right| + \left| \bigcup_{i=1}^{n_\rho} \Lambda_i^{(\pi)} \right|}. \quad (21)$$

All accuracy indicators are defined in the  $[0, 1]$  interval. Perfect segmentation sets all indicators to the maximum value of 1.

### 3. Results

The proposed brain tumor segmentation procedure was experimentally evaluated in various scenarios. Tests were performed using four different data sets: the LGG and HGG records separately from the BraTS 2015 training data set and, similarly, the LGG and HGG records separately from BraTS 2019 training data set. Detailed information on these data sets is given in Table 1. The whole segmentation process was the same for all four data sets, the one indicated in Figure 1.

All records of these four data sets underwent the mandatory pre-processing steps, namely the histogram normalization and feature generation, resulting in pre-processed data denoted by  $P_1$ . At this point, the records were randomly separated into six equal or almost equal groups, which took turns serving as the test data. An atlas was built using the training data only, namely the records from the other five groups. The atlas was then used to extract enhanced feature values for all records, resulting in pre-processed data denoted by  $P_2$ . During the whole processing, there were six different such sets denoted by  $P_2$ , one for the testing of each group.

Preprocessed data sets  $P_1$  and  $P_2$  underwent the ensemble learning process separately, so that we could evaluate the benefits of the atlas-based pre-processing. All ensembles involved in this study consisted of  $n_T = 125$  binary decision trees, but there were four variants for each data set, in terms of training data size. Each tree of an ensemble was trained to separate  $n_F$  feature vectors,  $n_F$  ranging from 100 thousand to on million, as indicated in Section 2.6. The intermediary segmentation obtained at the ensemble output is denoted by  $S_1$  or  $S_2$ , depending on the use of the atlas during pre-processing. The labels of pixels in  $S_1$  or  $S_2$  were treated with two post-processing steps described in detail in Sections 2.7 and 2.8, reaching the final segmentation denoted by  $S'_1$  or  $S'_2$ . Theoretically, all six segmentation results  $S_1, S_2, S'_1, S'_2, S''_1$ , and  $S''_2$  can be equally evaluated with the statistical measures presented in Section 2.9. The main result is represented by the final outcome of the segmentation:  $S''_1$  for the data that are not pre-processed with the atlas and  $S''_2$  for the data with atlas-based enhancement.

Table 3 presents the average ( $\overline{DSC}$ ) and overall ( $\widetilde{DSC}$ ) values of the Dice similarity coefficients obtained for various data sets and training data sizes, at four different phases of the segmentation process. A larger training data size always led to better accuracy: both the average and overall value rose by 0.5–0.8% if the training data size changed from 100k to 1000k. If we consider segmentation accuracy the only important quality marker, then it is worth using the largest training data size. The use of the atlas was beneficial in all scenarios, but there were large differences in the strength of this effect: at the ensemble output, the difference was between 0.3% and 2%, while at the final segmentation, it was between 0.2% and 1%. The difference between the average and overall values in the same scenario usually ranged between 2% and 3%. This was caused by the distribution of the Dice similarity coefficients obtained for individual MRI records. The obtained Dice scores, having most averages between 0.85 and 0.86 and overall values ranging between 0.86 and 0.88, were quite competitive.

**Table 3.** Average and overall Dice similarity coefficients obtained for various data sets and training data sizes, with and without atlas-based data enhancement at pre-processing.

Segmentation	Train	BraTS 2015 LGG	BraTS 2015 HGG	BraTS 2019 LGG	BraTS 2019 HGG					
Result	Data Size	$\overline{DSC}$	$\widetilde{DSC}$	$\overline{DSC}$	$\widetilde{DSC}$	$\overline{DSC}$	$\widetilde{DSC}$	$\overline{DSC}$	$\widetilde{DSC}$	
ensemble output without atlas	$(S_1)$	100k	0.7904	0.8217	0.7827	0.8248	0.7693	0.8065	0.8232	0.8560
		200k	0.7928	0.8239	0.7858	0.8267	0.7748	0.8113	0.8264	0.8578
		500k	0.7961	0.8266	0.7903	0.8290	0.7795	0.8154	0.8300	0.8602
		1000k	0.7980	0.8284	0.7935	0.8310	0.7828	0.8185	0.8324	0.8619
final result without atlas	$(S'_1)$	100k	0.8470	0.8618	0.8217	0.8524	0.8266	0.8495	0.8421	0.8679
		200k	0.8490	0.8635	0.8241	0.8540	0.8309	0.8531	0.8436	0.8692
		500k	0.8515	0.8657	0.8276	0.8557	0.8347	0.8565	0.8452	0.8707
		1000k	0.8536	0.8674	0.8300	0.8571	0.8378	0.8591	0.8464	0.8717
ensemble output with atlas	$(S_2)$	100k	0.8040	0.8316	0.7943	0.8279	0.8015	0.8307	0.8259	0.8602
		200k	0.8063	0.8336	0.7963	0.8301	0.8047	0.8337	0.8281	0.8626
		500k	0.8089	0.8359	0.7990	0.8334	0.8074	0.8367	0.8307	0.8654
		1000k	0.8107	0.8375	0.8013	0.8357	0.8091	0.8386	0.8326	0.8675
final result with atlas	$(S'_2)$	100k	0.8503	0.8646	0.8299	0.8573	0.8412	0.8624	0.8464	0.8738
		200k	0.8519	0.8662	0.8317	0.8588	0.8442	0.8652	0.8485	0.8751
		500k	0.8547	0.8683	0.8337	0.8610	0.8470	0.8681	0.8504	0.8766
		1000k	0.8564	0.8696	0.8355	0.8625	0.8479	0.8691	0.8516	0.8775

Tables 4–7 exhibit the average sensitivity, positive predictive value, specificity, and correct decision rate (accuracy) values, respectively, obtained for different data sets and various training data sizes. Not all these indicators increased together with the training data size, but the rate of correct decisions did, showing that it is worth using a larger amount of

training data to achieve better segmentation quality. Sensitivity values were in the range 0.8–0.85, while positive predictive values around 0.9, which indicate a fine recognition of true tumor pixels. Specificity rates were well above 0.99, which is highly important because it grants a reduced number of false positives. The average rate of correct decisions, with its values mostly above 0.98, indicates that one out of fifty or sixty pixels was misclassified by the proposed segmentation procedure. All values from Tables 4–7 reflect the evaluation of the final segmentation outcomes denoted by  $S_2''$ , obtained from atlas enhanced pre-processed data.

**Table 4.** Average sensitivity values ( $\overline{\text{TPR}}$ )—final result with atlas ( $S_2''$ ).

Train Data Size	BraTS 2015 Data		BraTS 2019 Data	
	LGG	HGG	LGG	HGG
100k	0.8155	0.7990	0.8131	0.8366
200k	0.8191	0.8016	0.8179	0.8401
500k	0.8248	0.8056	0.8234	0.8444
1000k	0.8299	0.8094	0.8279	0.8473

**Table 5.** Average positive predictive values ( $\overline{\text{PPV}}$ )—final result with atlas ( $S_2''$ ).

Train Data Size	BraTS 2015 Data		BraTS 2019 Data	
	LGG	HGG	LGG	HGG
100k	0.9071	0.8935	0.9055	0.8888
200k	0.9064	0.8936	0.9046	0.8878
500k	0.9043	0.8923	0.9032	0.8857
1000k	0.9019	0.8909	0.8986	0.8842

**Table 6.** Average specificity values ( $\overline{\text{TNR}}$ )—final result with atlas ( $S_2''$ ).

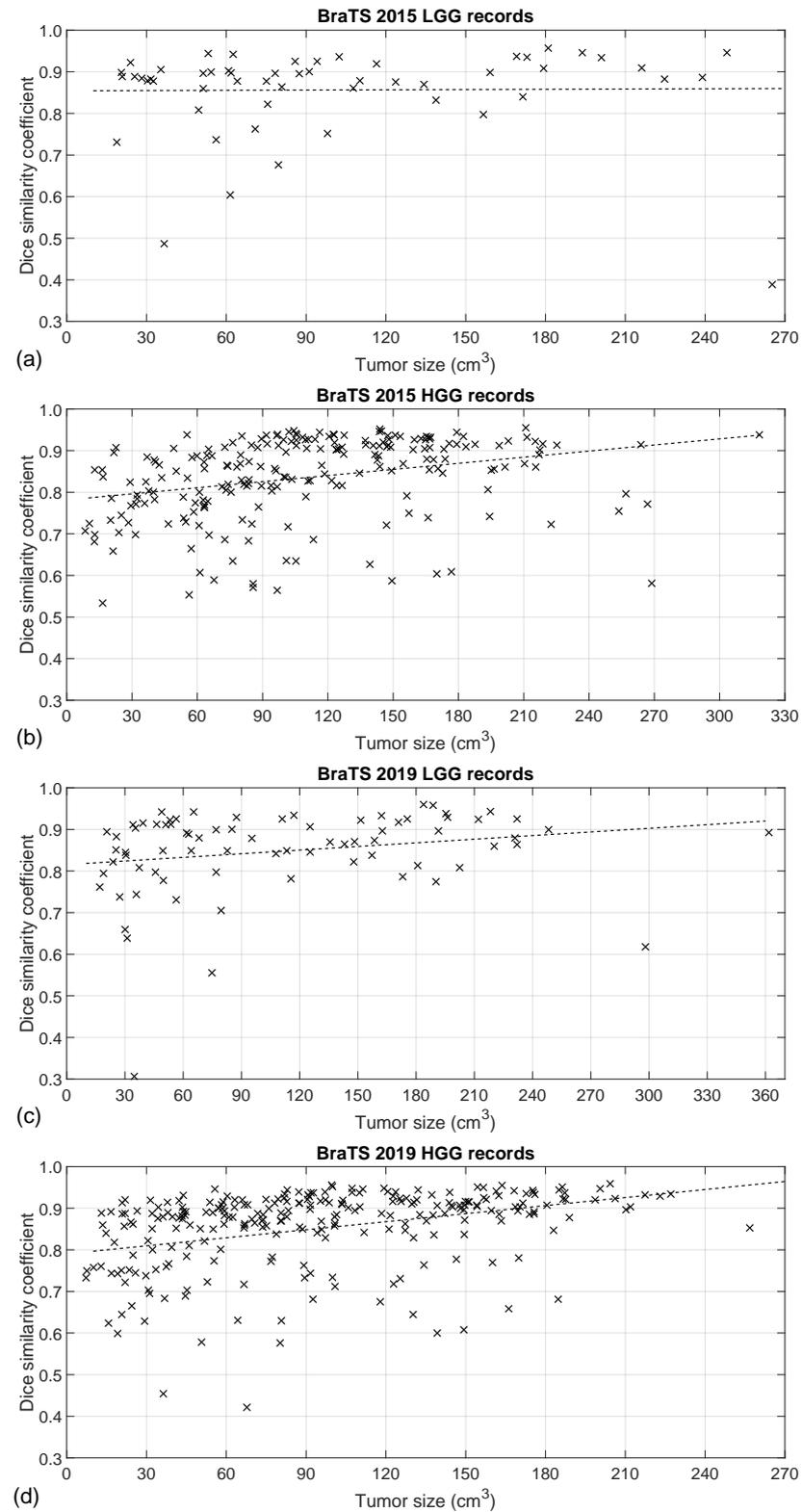
Train Data Size	BraTS 2015 Data		BraTS 2019 Data	
	LGG	HGG	LGG	HGG
100k	0.9936	0.9913	0.9933	0.9923
200k	0.9935	0.9913	0.9932	0.9922
500k	0.9933	0.9912	0.9931	0.9920
1000k	0.9931	0.9910	0.9927	0.9919

**Table 7.** Average accuracy values ( $\overline{\text{ACC}}$ )—final result with atlas ( $S_2''$ ).

Train Data Size	BraTS 2015 Data		BraTS 2019 Data	
	LGG	HGG	LGG	HGG
100k	0.9812	0.9792	0.9796	0.9837
200k	0.9814	0.9794	0.9799	0.9838
500k	0.9816	0.9796	0.9803	0.9840
1000k	0.9817	0.9797	0.9803	0.9840

Figure 5 exhibits the Dice similarity coefficients obtained for individual MRI records ( $\text{DSC}_i$ , for  $i = 1 \dots n_p$ ) plotted against the true size of the tumor according to the ground truth. Each cross represents one of the MRI records, while the dashed lines indicate the linear trend of the  $\text{DSC}_i$  values identified with linear regression. As was expected, the linear trends indicate better segmentation accuracy for larger tumors. It is also visible that for most tumors, we achieved a Dice score above 0.8, and there were some records below that limit where the Dice scores could be very low. This is mostly because not all records in the BraTS data have the same image quality; some of them even contain artificially created obstacles. This distribution of the  $\text{DSC}_i$  values is also the reason why the overall Dice

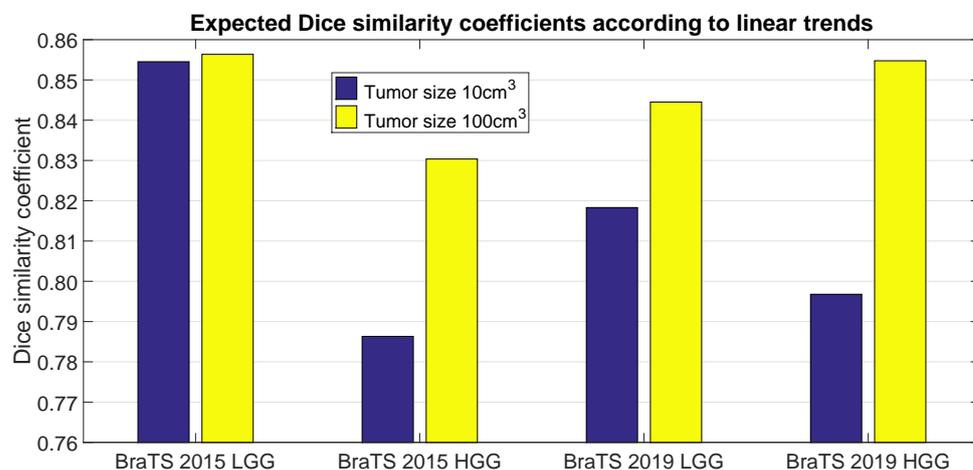
similarity value was higher than the average ( $\widetilde{DSC} > \overline{DSC}$ ) in case of all four data sets. In fact,  $\widetilde{DSC}$  was close to the median value of individual Dice scores  $DSC_i, i = 1 \dots n_p$ .



**Figure 5.** Individual Dice scores plotted against the true size of the tumor: (a) 54 records of BraTS 2015 LGG data; (b) 220 records of BraTS 2015 HGG data; (c) 76 records of BraTS 2019 LGG data; (d) 259 records of BraTS 2019 HGG data.

Figure 6 presents the Dice similarity scores we may expect for a 10 cm³ and a 100 cm³

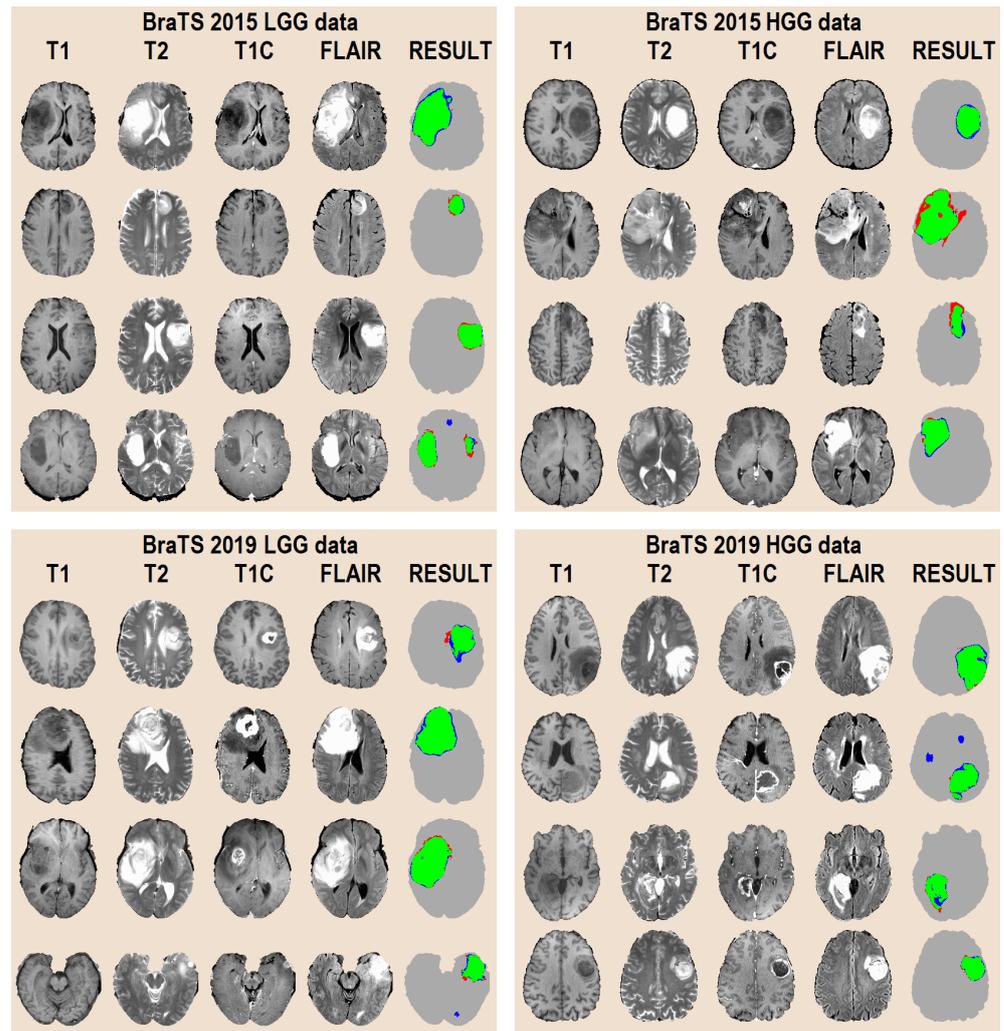
sized tumor according to the linear trends identified from the data presented in Figure 5, in the case of the four data sets separately. Small tumors of  $10\text{ cm}^3$  are not even present in all data sets, but the proposed method apparently learned how to segment them accurately, with an expected Dice similarity value around 0.8, which reportedly represents fine segmentation [57]. The Dice score obtained for an average sized tumor from the data sets was expected in the proximity of 0.85.



**Figure 6.** Expected Dice similarity coefficients for a  $10\text{ cm}^3$  and a  $100\text{ cm}^3$  sized tumor, in the case of the four BraTS data sets, according to the identified linear trends.

Figure 7 exhibits some selected segmentation results. Four MRI volumetric records were selected from each of the four data sets, out of which that slice was chosen that contained the highest number of positive pixels, ground truth positives, and positive labeled pixels combined. Thus, each row in this figure exhibits one slice from a volume in five images, which represent the four observed data channels T1, T2, T1C, and FLAIR and the segmentation outcome, respectively. In the segmentation outcome, true positives are drawn in green, false negatives in red, false positives in blue, and true negatives in grey. These segmentation results suggest that there is still place for improvement in establishing the exact boundary of the tumor, and also in suppressing the patches of false positives.

The proposed procedure does not require special hardware equipment. The whole evaluation process was performed with an Asus Zenbook computer having a quad-core i7 processor with 1.8 GHz nominal frequency, 16 GB RAM, and 1 TB SSD for data storage. The most part of the time consuming processing steps was implemented to run in parallelized tasks. The average processing time of an MRI record (coming in uncompressed .nii files) never seen by the trained procedure was 58 s, out of which: (1) 26 s were needed by the pre-processing including feature generation and atlas based data enhancement; 21 s were required by the trained ensemble to produce the intermediary segmentation; (3) 11 s were the duration of the post-processing steps. No GPU was involved in the computations, so there is still place for improvement in terms of efficiency.



**Figure 7.** One slice from 16 different MRI records, four from each data set, showing the four observed data channels and the segmentation result. The first four columns present the T1, T2, T1C, and FLAIR channel data of the chosen slices. The last column shows the segmented slice, representing true positives ( $|\Gamma_i^{(\pi)} \cap \Lambda_i^{(\pi)}|$ ) in green, false negatives ( $|\Gamma_i^{(\pi)} \cap \Lambda_i^{(v)}|$ ) in red, false positives ( $|\Gamma_i^{(v)} \cap \Lambda_i^{(\pi)}|$ ) in blue, and true negatives ( $|\Gamma_i^{(v)} \cap \Lambda_i^{(v)}|$ ) in grey, where  $i$  is the index of the current MRI record.

#### 4. Discussion

Table 8 presents the accuracy performance of a collection of state-of-the-art methods proposed for the whole brain tumor segmentation problem. Although they are recent solutions, the vast majority still use the BraTS 2013 and BraTS 2015 data sets for evaluation purposes. An important requirement for the methods to be selected is that they apparently report results obtained by processing the whole data set, not only some records that give high accuracy. The Dice similarity coefficients shown in the table are written in the form published by the authors, with the intention to be interpreted correctly. We do not agree with publishing  $\overline{DSC}$  values expressed with two decimals only, because the half percent imprecision can hide an extremely large difference. One important thing reflected by this table is that it is much easier to obtain high accuracy with the BraTS 2013 data, which is probably caused by the reduced number of records in this set that cannot cover enough variation of tumor shapes and appearances. See for example the method of Pereira et al. [29], which obtains average Dice scores of 0.88 and 0.78 on BraTS 2013 and BraTS 2015, respectively. The proposed procedure with its achieved Dice scores belongs to the set of methods with top accuracy.

From the point of view of efficiency, the main question is how much time a method needs to perform a whole segmentation on an MRI record of the same type that is never seen during the training of the method. Table 9 presents the performance against the clock of some selected state-of-the-art methods. Our goal is to report those runtime values, which reflect the total duration needed, starting from uncompressed .nii files and ending when the segmentation result is saved, but we cannot be sure that all methods included in this table report this kind of duration. It is not surprising at all that solutions based on CNN need a longer time even if they run on GPUs, because their architecture is usually more complex. However, other methods can also achieve fine segmentation quality in less time using CPUs only. This is probably because fine tuning complex architectures is much more difficult, and some of them may still run in a suboptimal state. With its reduced time complexity, our proposed method is very competitive and can be deployed without needing any special hardware.

**Table 8.** Accuracy comparison with state-of-the-art methods using BraTS 2013 and BraTS 2015 data.

Method	Year	Classifier	Data	Dice Scores
Tustison et al. [11]	2015	RF, MRF	BraTS 2013	$\overline{DSC} = 0.87$
Pereira et al. [29]	2016	CNN	BraTS 2013	$\overline{DSC} = 0.88$
Lefkovits et al. [58]	2017	RF	BraTS 2013	$\overline{DSC} = 0.868$
Havaei et al. [59]	2017	deep CNN	BraTS 2013	$\overline{DSC} = 0.88$
Pinto et al. [12]	2018	ERT	BraTS 2013	$\overline{DSC} = 0.85$
Pereira et al. [60]	2019	FCNN	BraTS 2013	$\overline{DSC} = 0.86$
Pereira et al. [29]	2016	CNN	BraTS 2015	$\overline{DSC} = 0.78$
Kamnitsas et al. [32]	2017	deep CNN	BraTS 2015	$\overline{DSC} = 0.849$
Xue et al. [34]	2018	CNN, GAN	BraTS 2015	$\overline{DSC} = 0.85$
Zhao et al. [30]	2018	FCNN, CRF	BraTS 2015	$\overline{DSC} = 0.84$
Hussain et al. [61]	2018	deep CNN	BraTS 2015	$\overline{DSC} = 0.86$
Chen et al. [35]	2019	CNN	BraTS 2015	$\overline{DSC} = 0.85$
Ding et al. [33]	2019	deep residual network	BraTS 2015	$\overline{DSC} = 0.86$
Pei et al. [62]	2020	RF, boosting	BraTS 2015	$\overline{DSC} = 0.850$
Wu et al. [31]	2020	CNN	BraTS 2015	$\overline{DSC} = 0.83$
Proposed method	2020	ensembles of binary decision trees	BraTS 2015	$\overline{DSC} = 0.8564, \widehat{DSC} = 0.8696$ (LGG) $\overline{DSC} = 0.8355, \widehat{DSC} = 0.8625$ (HGG)
			BraTS 2019	$\overline{DSC} = 0.8479, \widehat{DSC} = 0.8691$ (LGG) $\overline{DSC} = 0.8516, \widehat{DSC} = 0.8775$ (HGG)

**Table 9.** Efficiency comparison with state-of-the-art methods. Runtime is expressed in seconds, and it represents the total time needed for the whole processing of a multi-spectral BraTS record, averaged over one or more BraTS data sets.

Method	Year	CPU	GPU	Runtime (s)
Pereira et al. [29]	2016	i7 3.5 GHz	GTX 980	480
Havaei et al. [59]	2017	unspecified	unspecified	90–180
Kamnitsas et al. [32]	2017	unspecified	GTX TitanX 12 GB	210
Zhao et al. [30]	2018	unspecified	shared GPU server	120–240
Chen et al. [35]	2019	unspecified	GTX TitanX 12 GB	187
Hu et al. [63]	2019	unspecified	unspecified	90–180
Imtiaz et al. [15]	2019	i7 3.2 GHz	no GPU involved	335
Proposed method	2020	i7 1.8 GHz	no GPU involved	58

The main parameters of the proposed brain tumor segmentation procedure are: (1) the size of training data, or in other words, the number of feature vectors used to train each BDT in the ensemble,  $n_F$ ; (2) the percentage of negatives  $p_-$  and positives  $p_+$  in the set of feature vectors used to train the BDTs; (3) the number of trees in the ensemble  $n_T$ ; (4) the presence or absence of the atlas enhanced pre-processing within the pipeline of the procedure.

The decision on the training data size has to take into account its twofold implications. Larger training data sets enable the procedure to achieve slightly better accuracy in all tested cases. On the other hand, changing  $n_F$  from 100k to 1000k raises the training time of an ensemble from 2 to 40 min, but the total duration of a segmentation after having the ensemble trained only increases by 4–5 s. Once the large ensembles are trained, they can perform the segmentation quickly enough without prohibitive computational load.

The percentage of negatives  $p_-$  in the training data can have a strong implication on the segmentation quality. Experiments show that there are MRI records in the BraTS data sets that give the best segmentation accuracy at  $p_- = 75\%$ , others at  $p_- = 95\%$ , and many of them at various values between these two. It is very likely that the optimal value of  $p_-$  for a given MRI record strongly correlates with the true rate of negative pixels in the volume. However, this true rate is unknown, as we are not allowed to look at the ground truth while testing. The best accuracy at the ensemble output could be achieved if we had a precise estimation for each MRI record, which is the best choice for  $p_-$ , only using the intensity values of the four observed data channels. Such an estimator would allow for 2–3% higher Dice similarity coefficients at the ensemble output. Without such an estimator, a dedicated constant value of  $p_-$  needs to be chosen for all MRI records in each data set. In this study,  $p_-$  was set to 91% for HGG data and 93% for LGG data, based on evaluations performed on OOB data.

The number of trees in the ensemble ( $n_T$ ) theoretically has an impact on the decision error, which should decrease with any additional tree included in the ensemble. Experiments show that the decision accuracy saturates at a certain level due to the noise that is present in the data. Further, the number of trees linearly affects the runtime of both the training and testing process. After several evaluation loops performed on OOB data, with various values of  $n_T$  ranging from five to 255, the size of the ensemble was chosen to be set to  $n_T = 125$  [43,64].

The use of the atlas caused up to a 3% improvement in the average Dice scores obtained at the output of the ensemble (see the difference between solutions  $S_1$  and  $S_2$  in Table 3), but this difference diminished to a maximum of 1% in the accuracy of the final segmentation. Even for this difference, we consider the use of the atlas beneficial, as it does not affect the execution time of the procedure much.

The relatively high accuracy achieved by the proposed procedure is probably caused by the way it handles the imbalanced amounts of positive and negative data. Our BDTs are allowed to grow to unlimited depth during ensemble training, as deep as is found necessary by the entropy-based criterion applied in decision nodes. Unlike the random forest classifier, which limits its depth and sometimes assigns labels to mixtures of positives and negatives when the depth limit is reached, our binary trees has crisp labels assigned to their leaves. This allows for more precise classification unless the trees are over-trained. With the parameter settings we used, the decision trees did not show signs of being over-trained, which is supported by the accuracy indicators that improved as the training data size grew. The number of false positives was not high at the output of the ensemble, and it was still reduced by the twofold post-processing, which discarded those predicted positive structures that were too small or too flat to be reliably called tumors.

The proposed procedure also has some limitations. In its current version, it is trained and tuned to segment the whole tumor only, not the parts of it. Using only a set of handcrafted features, it may not succeed in exploring the input data as thoroughly as a convolutional network can. Further, the approximately three million decision nodes contained by our most complex trained forest of binary decision trees is less than the

number of parameters used in current CNN architectures by up to two orders of magnitude. This might become a visible limitation in accuracy if we increase the number of MRI records used for training and testing.

## 5. Conclusions

This paper proposes a fully automatic procedure for brain tumor segmentation from multi-spectral magnetic resonance image data. The procedure consists of three main phases. The first phase prepares the data to be suitable for supervised classification, via noise suppression, handling missing data, histogram normalization, feature generation, and an optional atlas-based data enhancement. In the second phase, an initial labeling of the pixels is achieved, using supervised classification performed by an ensemble of binary decision trees. In the final phase, the intermediary pixel labels are refined using a random forest-based reclassification and a structural post-processing step that investigates contiguous regions within the detected tumor. The proposed procedure is trained and evaluated using all the MRI records of the BraTS 2015 and BraTS 2019 training data sets.

With average Dice similarity coefficients of up to 85.6% and overall Dice scores approaching 88%, as well as the whole processing time of an MRI record being below one minute without needing a GPU for high-speed computations, the proposed procedure is in competition with state-of-the-art methods. The segmentation quality achieved by the procedure could still improve with the use of more sophisticated handcrafted features in the decision ensemble, while the implementation of a feature selection scheme could be beneficial to the efficiency of the proposed procedure.

**Author Contributions:** Conceptualization, Á.G., L.S., and L.K.; methodology, Á.G. and L.S.; software, Á.G. and L.S.; validation, Á.G. and L.S.; formal analysis, L.K.; investigation, Á.G. and L.S.; resources, L.K.; data curation, Á.G.; writing, original draft preparation, Á.G. and L.S.; writing, review and editing, Á.G. and L.S.; visualization, L.S.; supervision, L.S.; project administration, L.K.; funding acquisition, L.K. All authors read and agreed to the published version of the manuscript.

**Funding:** Research supported in part by Sapientia Institute for Research Programs (KPI). The work of L. Kovács was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (Grant Agreement No 679681). The work of L. Szilágyi was supported by the János Bolyai Fellowship Program of the Hungarian Academy of Sciences and by the ÚNKP-20-5 New National Excellence Program of the Ministry of Innovation and Technology from the Source of the National Research, Development and Innovation Fund.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The brain MRI records processed in this study are available at the BraTS website: <http://www.braintumorsegmentation.org>.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

ACC	accuracy
BDT	binary decision tree
BraTS	Brain Tumor Segmentation (Challenge)
CNN	convolutional neural network
CPU	central processing unit
CT	computed tomography
DICOM	digital imaging and communications in medicine
DSC	Dice similarity score
ERT	extremely randomized trees
FCM	fuzzy <i>c</i> -means

FCNN	fully convolutional neural network
FLAIR	fluid-attenuated inversion recovery
GAN	generative adversarial network
GPU	graphical processing unit
GT	ground truth
HGG	high-grade glioma
INU	intensity non-uniformity
LGG	low-grade glioma
MICCAI	Medical Image Computation and Computer Assisted Intervention
MLP	multi-layer perceptron
MRI	magnetic resonance imaging
MRF	Markov random field
OOB	out-of-bag
PCA	principal component analysis
PET	positron emission tomography
PPV	positive predictive value
RAM	random access memory
RF	random forest
SLIC	simple linear iterative clustering
SSD	solid state drive
SVM	support vector machine
TNR	true negative rate
TPR	true positive rate

## References

- Patel, A.P.; Fisher, J.L.; Nichols, E.; Abd-Allah, F.; Abdella, J.; Abdelalim, A.; Allen, C.A. Global, regional, and national burden of brain and other CNS cancer, 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* **2019**, *18*, 376–393. [\[CrossRef\]](#)
- Mohan, G.; Subashini, M.M. MRI based medical image analysis: Survey on brain tumor grade classification. *Biomed. Sign. Proc. Contr.* **2018**, *39*, 139–161. [\[CrossRef\]](#)
- Menze, B.H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Lanczi, L. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **2015**, *34*, 1993–2024. [\[CrossRef\]](#) [\[PubMed\]](#)
- Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Prastawa, M. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv* **2019**, arXiv: 1181.02629v2.
- Gordillo, N.; Montseny, E.; Sobrevilla, P. State of the art survey on MRI brain tumor segmentation. *Magn. Reson. Imaging* **2013**, *31*, 1426–1438. [\[CrossRef\]](#) [\[PubMed\]](#)
- Sahdeva, J.; Kumar, V.; Gupta, I.; Khandelwal, N.; Ahuja, C.K. A novel content-based active countour model for brain tumor segmentation. *Magn. Reson. Imaging* **2012**, *30*, 694–715. [\[CrossRef\]](#) [\[PubMed\]](#)
- Njeh, I.; Sallemi, L.; Ben Ayed, I.; Chtourou, K.; Lehericy, S.; Galanaud, D.; Ben Hamida, A. 3D multimodal MRI brain glioma tumor and edema segmentation: A graph cut distribution matching approach. *Comput. Med. Imaging Graph.* **2015**, *40*, 108–119. [\[CrossRef\]](#)
- Li, Q.N.; Gao, Z.F.; Wang, Q.Y.; Xia, J.; Zhang, H.Y.; Zhang, H.L.; Liu, H.F.; Li, S. Glioma segmentation with a unified algorithm in multimodal MRI images. *IEEE Access* **2018**, *6*, 9543–9553. [\[CrossRef\]](#)
- Szilágyi, L.; Lefkovits, L.; Benyó, B. Automatic brain tumor segmentation in multispectral MRI volumes using a fuzzy *c*-means cascade algorithm. In Proceedings of the 12th International Conference on Fuzzy Systems and Knowledge Discovery, Zhangjiajie, China, 15–17 August 2015; pp. 285–291. [\[CrossRef\]](#)
- Islam, A.; Reza, S.M.S.; Iftekharuddin, K.M. Multifractal texture estimation for detection and segmentation of brain tumors. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 3204–3215. [\[CrossRef\]](#)
- Tustison, N.J.; Shrinidhi, K.L.; Wintermark, M.; Durst, C.R.; Kandel, B.M.; Gee, J.C.; Grossman, M.C.; Avants, B.B. Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR. *Neuroinformatics* **2015**, *13*, 209–225. [\[CrossRef\]](#)
- Pinto, A.; Pereira, S.; Rasteiro, D.; Silva, C.A. Hierarchical brain tumour segmentation using extremely randomized trees. *Pattern Recogn.* **2018**, *82*, 105–117. [\[CrossRef\]](#)
- Soltaninejad, M.; Yang, G.; Lambrou, T.; Allinson, N.; Jones, T.L.; Barrick, T.R.; Howe, F.A.; Ye, X.J. Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels. *Comput. Method Prog. Biomed.* **2018**, *157*, 69–84. [\[CrossRef\]](#) [\[PubMed\]](#)
- Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Süsstrunk, S. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2274–2282. [\[CrossRef\]](#) [\[PubMed\]](#)

15. Imtiaz, T.; Rifat, S.; Fattah, S.A.; Wahid, K.A. Automated brain tumor segmentation based on multi-planar superpixel level features extracted from 3D MR images. *IEEE Access* **2020**, *8*, 25335–25349. [[CrossRef](#)]
16. Kalaiselvi, T.; Kumarashankar, P.; Sriramakrishnan, P. Three-phase automatic brain tumor diagnosis system using patches based updated run length region growing technique. *J. Digit. Imaging* **2020**, *33*, 465–479. [[CrossRef](#)] [[PubMed](#)]
17. Sengupta, S.; Basak, S.; Saikia, P.; Paul, S.; Tsalavoutis, V.; Atiah, F.; Ravi, V.; Peters, A. A review of deep learning with special emphasis on architectures, applications and recent trends. *Knowl. Based Syst.* **2020**, *194*, 105596. [[CrossRef](#)]
18. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)]
19. Wu, Y.; Yi, Z. Automated detection of kidney abnormalities using multi-feature fusion convolutional neural networks. *Knowl. Based Syst.* **2020**, *200*, 105873. [[CrossRef](#)]
20. Wildeboer, R.R.; van Sloun, R.G.J.; Wijkstra, H.; Mischi, M. Artificial intelligence in multiparametric prostate cancer imaging with focus on deep-learning methods. *Comput. Meth. Prog. Biomed.* **2020**, *189*, 105316. [[CrossRef](#)]
21. Xue, J.; Yan, S.; Qu, J.H.; Qi, F.; Qiu, C.G.; Zhang, H.Y.; Chen, M.R.; Liu, T.T.; Li, D.W.; Liu, X.Y. Deep membrane systems for multitask segmentation in diabetic retinopathy. *Knowl. Based Syst.* **2019**, *183*, 104887. [[CrossRef](#)]
22. Tan, T.Y.; Zhang, L.; Lim, C.P. Adaptive melanoma diagnosis using evolving clustering, ensemble and deep neural networks. *Knowl. Based Syst.* **2020**, *187*, 104807. [[CrossRef](#)]
23. Chung, M.Y.; Li, J.Y.; Lee, M.K.; Lee, J.J.; Shin, Y.G. Deeply self-supervised contour embedded neural network applied to liver segmentation. *Comput. Meth. Prog. Biomed.* **2020**, *192*, 105447. [[CrossRef](#)]
24. Du, X.Q.; Song, Y.H.; Liu, Y.G.; Zhang, Y.P.; Liu, H.; Chen, B.; Li, S. An integrated deep learning framework for joint segmentation of blood pool and myocardium. *Med. Image Anal.* **2020**, *62*, 101685. [[CrossRef](#)]
25. Orellana, B.; Monclús, E.; Brunet, P.; Navazo, I.; Bendejú, A.; Azpiroz, F. A scalable approach to T2-MRI colon segmentation. *Med. Image Anal.* **2020**, *63*, 101697. [[CrossRef](#)] [[PubMed](#)]
26. Wang, C.L.; Oda, M.; Hayashi, Y.; Yoshino, Y.; Yamamoto, T.; Frangi, A.F.; Mori, K. Tensor-cut: A tensor-based graph-cut blood vessel segmentation method and its application to renal artery segmentation. *Med. Image Anal.* **2020**, *60*, 101623. [[CrossRef](#)] [[PubMed](#)]
27. Yan, M.; Guo, J.X.; Tian, W.D.; Yi, Z. Symmetric convolutional neural network for mandible segmentation. *Knowl. Based Syst.* **2018**, *159*, 63–71. [[CrossRef](#)]
28. Noguchi, S.; Nishio, M.; Yakami, M.; Nakagomi, K.; Togashi, K. Bone segmentation on whole-body CT using convolutional neural network with novel data augmentation techniques. *Comput. Biol. Med.* **2020**, *121*, 103767. [[CrossRef](#)] [[PubMed](#)]
29. Pereira, S.; Pinto, A.; Alves, V.; Silva, C.A. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imaging* **2016**, *35*, 1240–1251. [[CrossRef](#)] [[PubMed](#)]
30. Zhao, X. M.; Wu, Y.H.; Song, G.D.; Li, Z.Y.; Zhang, Y.Z.; Fan, Y. A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. *Med. Image Anal.* **2018**, *43*, 98–111. [[CrossRef](#)]
31. Wu, D.Y.; Ding, Y.; Zhang, M.F.; Yang, Q.Q.; Qin, Z.G. Multi-features refinement and aggregation for medical brain segmentation. *IEEE Access* **2020**, *8*, 57483–57496. [[CrossRef](#)]
32. Kamnitsas, K.; Ledig, C.; Newcombe, V.F.J.; Simpson, J.P.; Kane, A.D.; Menon, D.K.; Rueckert, D.; Glocker, B. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **2017**, *36*, 61–78. [[CrossRef](#)] [[PubMed](#)]
33. Ding, Y.; Li, C.; Yang, Q.Q.; Qin, Z.; Qin, Z.G. How to improve the deep residual network to segment multi-modal brain tumor images. *IEEE Access* **2019**, *7*, 152821–152831. [[CrossRef](#)]
34. Xue, Y.; Xu, T.; Zhang, H.; Long, L.; Huang, X.L. SegAN: Adversarial network with multi-scale  $L_1$  loss for medical image segmentation. *Neuroinformatics* **2018**, *16*, 383–392. [[CrossRef](#)] [[PubMed](#)]
35. Chen, S.C.; Ding, C.X.; Liu, M.F. Dual-force convolutional neural networks for accurate brain tumor segmentation. *Pattern Recogn.* **2019**, *88*, 90–100. [[CrossRef](#)]
36. Goodman, B.; Flaxman, S. European Union regulations on algorithmic decision making and “right to explanation”. *AI Mag.* **2017**, *38*, 50–57. [[CrossRef](#)]
37. Waring, J.; Lindvall, C.; Umeton, R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. *Artif. Intell. Med.* **2020**, *104*, 101822. [[CrossRef](#)]
38. Tóth, J.; Tornai, R.; Labancz, I.; Hajdu, A. Efficient visualization for an ensemble-based system. *Acta Polytech. Hung.* **2019**, *16*, 59–75. [[CrossRef](#)]
39. Akers, S. Binary decision diagrams. *IEEE Trans. Comput.* **1978**, *C-27*, 509–516. [[CrossRef](#)]
40. Sun, L.; Shao, W.; Wang, M.; Zhang, D.; Liu, M. High-order feature learning for multi-atlas based label fusion: application to brain segmentation with MRI. *IEEE Trans. Image Proc.* **2019**, *29*, 2702–2713. [[CrossRef](#)]
41. Bao, S.; Chung, A. Feature sensitive label fusion with random walker for atlas-based image segmentation. *IEEE Trans. Image Proc.* **2017**, *26*, 2797–2810. [[CrossRef](#)]
42. Zheng, Q.; Wu, Y.; Fan, Y. Integrating semi-supervised and supervised learning methods for label fusion in multi-atlas based image segmentation. *Front. Neuroinform.* **2018**, *12*, 69. [[CrossRef](#)] [[PubMed](#)]
43. Szilágyi, L.; Iclánzan, D.; Kapás, Z.; Szabó, Z.; Györfi, Á.; Lefkovits, L. Low and high grade glioma segmentation in multispectral brain MRI data. *Acta Univ. Sapientiae Inform.* **2018**, *10*, 110–132. [[CrossRef](#)]

44. Fülöp, T.; Gyórfi, A.; Surányi, B.; Kovács, L.; Szilágyi, L. Brain tumor segmentation from MRI data using ensemble learning and multi-atlas. In Proceedings of the 18th IEEE World Symposium on Applied Machine Intelligence and Informatics, Herlany, Slovakia, 23–25 January 2020; pp. 111–116. [\[CrossRef\]](#)
45. Vovk, U.; Pernuš, F.; Likar, B. A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans. Med. Imaging* **2007**, *26*, 405–421. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Szilágyi, L.; Szilágyi, S.; Benyó, B. Efficient inhomogeneity compensation using fuzzy *c*-means clustering models. *Comput. Method Prog. Biomed.* **2012**, *108*, 80–89. [\[CrossRef\]](#)
47. Tustison, N.J.; Avants, B.B.; Cook, P.A.; Zheng, Y.J.; Egan, A.; Yushkevich, P.A.; Gee, J.C. N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imaging* **2010**, *29*, 1310–1320. [\[CrossRef\]](#)
48. Nyúl, L.G.; Udupa, J.K.; Zhang, X. New variants of a method of MRI scale standardization. *IEEE Trans. Med. Imaging* **2000**, *19*, 143–150. [\[CrossRef\]](#)
49. Leung, K.K.; Clarkson, M.J.; Bartlett, J.W.; Clegg, S.; Jack, C.R., Jr.; Weiner, M.W.; Fox, N.C.; Ourselin, S.; Alzheimer’s Disease Neuroimaging Initiative. Robust atrophy rate measurement in Alzheimer’s disease using multi-site serial MRI: Tissue-specific intensity normalization and parameter selection. *NeuroImage* **2010**, *50*, 516–523. [\[CrossRef\]](#)
50. Shinohara, R.T.; Crainiceanu, C.M.; Caffo, B.S.; Gaitán, M.I.; Reich, D.S. Population-wide principal component-based quantification of blood-brain-barrier dynamics in multiple sclerosis. *NeuroImage* **2011**, *57*, 1430–1446. [\[CrossRef\]](#)
51. Meier, R.; Knecht, U.; Loosli, T.; Bauer, S.; Slotboom, J.; Wiest, R.; Reyes, M. Clinical evaluation of a fully-automatic segmentation method for longitudinal brain tumor volumetry. *Sci. Rep.* **2016**, *6*, 23376. [\[CrossRef\]](#)
52. Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J.S.; Freymann, J.B.; Farahani, K.; Davatzikos, C. Advancing the Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. *Sci. Data* **2017**, *4*, 170117. [\[CrossRef\]](#)
53. Chen, W.; Liu, B.Q.; Peng, S.T.; Sun, J.W.; Qiao, X. Computer-aided grading of gliomas combining automatic segmentation and radiomics. *Int. J. Biomed. Imaging* **2018**, *2018*, 2512037. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Wu, Y.P.; Liu, B.; Lin, Y.S.; Yang, C.; Wang, M.Y. Grading glioma by radiomics with feature selection based on mutual information. *J. Amb. Intell. Human. Comput.* **2018**, *9*, 1671–1682. [\[CrossRef\]](#)
55. Chang, J.; Zhang, L.M.; Gu, N.J.; Zhang, X.C.; Ye, M.Q.; Yin, R.Z.; Meng, Q.Q. A mix-pooling CNN architecture with FCRF for brain tumor segmentation. *J. Vis. Commun. Image R.* **2019**, *58*, 316–322. [\[CrossRef\]](#)
56. Gyórfi, A.; Karetka-Mezsei, Z.; Iclánzan, D.; Kovács, L.; Szilágyi, L. A study on histogram normalization for brain tumour segmentation from multispectral MR image data. In Proceedings of the Ibero-American Congress on Pattern Recognition (LNCS 11896), Havana, Cuba, 28–31 October 2019; pp. 375–384. [\[CrossRef\]](#)
57. Bal, A.; Banerjee, M.; Chakrabarti, A.; Sharma, P. MRI brain tumor segmentation and analysis using rough-fuzzy *c*-means and shape based properties. *J. King Saud Univ. Comput. Inf. Sci.* **2018**. [\[CrossRef\]](#)
58. Lefkovits, L.; Lefkovits, S.; Szilágyi, L. Brain tumor segmentation with optimized random forest. In *2nd International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, LNCS 10154*; Springer: Cham, Switzerland, 2017; pp. 88–99. [\[CrossRef\]](#)
59. Havaei, M.; Davy, A.; Warde-Farley, D.; Biard, A.; Courville, A.; Bengio, Y.; Pal, C.; Jodoin, P.M.; Larochelle, H. Brain tumor segmentation with deep neural networks. *Med. Image Anal.* **2017**, *35*, 18–31. [\[CrossRef\]](#)
60. Pereira, S.; Pinto, A.; Amorim, J.; Ribeiro, A.; Alves, V.; Silva, C.A. Adaptive feature recombination and recalibration for semantic segmentation with fully convolutional networks. *IEEE Trans. Med. Imaging* **2019**, *38*, 2914–2925. [\[CrossRef\]](#)
61. Hussain, S.; Anwar, S.M.; Majid, M. Segmentation of glioma tumors in brain using deep convolutional neural network. *Neurocomputing* **2018**, *282*, 248–261. [\[CrossRef\]](#)
62. Pei, L.M.; Bakas, S.; Vassough, A.; Reza, S.M.S.; Davatzikos, C.; Iftekharuddin, K.M. Longitudinal brain tumor segmentation prediction in MRI using feature and label fusion. *Biomed. Sign. Proc. Contr.* **2020**, *55*, 101648. [\[CrossRef\]](#)
63. Hu, K.; Gan, Q.H.; Zhang, Y.; Deng, S.H.; Xiao, F.; Huang, W.; Cao, C.H.; Gao, X.P. Brain tumor segmentation using multi-cascaded convolutional neural networks and conditional random field. *IEEE Access* **2019**, *7*, 92615–92629. [\[CrossRef\]](#)
64. Gyórfi, A.; Kovács, L.; Szilágyi, L. Brain tumor detection and segmentation from magnetic resonance image data using ensemble learning methods. In Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, Bari, Italy, 6–9 October 2019; pp. 919–924. [\[CrossRef\]](#)