

Comparison of Classifiers for Brain Tumor Segmentation

L. Lefkovits^{1,3}, Sz. Lefkovits², M. F. Vaida³, S. Emerich³ and R. Măluțan³

¹ Department of Electrical Engineering Sapiientia University, Tîrgu-Mureș, Romania

² Department of Computer Science “Petru-Maior” University, Tîrgu-Mureș, Romania

³ Faculty of Electronics, Telecommunication and Information Technology, Technical University of Cluj-Napoca

Abstract— Nowadays numerous efforts and promising results are obtained in medical imaging processing, although reproducible segmentation and classification of tumors is still a challenging task. The difficulty consists of the different shapes, locations and image intensities of these tissues. In this article we present our discriminative segmentation system for brain tumor delimitation from multimodal MR images. The detection of tumor requires a well-defined process-sequence on every analyzed MRI including preprocessing, feature extraction, classification and post-processing. The discriminative models are trained from annotated image databases and build their decision function around a classifier algorithm. In machine learning there are a lot of advanced classifiers that can be used for segmentation task. The choice of the most adequate classifiers is not straightforward. The SVM, the AdaBoost and the Random Forest (RF) are among the most 10 best classifiers and are often used in image segmentation. The goal of this paper is to analyze and compare these three classification techniques and their obtainable performances on the same segmentation framework. First, we present our framework for brain tumor segmentation. In the theoretical part we briefly present each classifier, emphasizing the advantages and disadvantages of them. In practice, we trained all these three classifiers on the same data set and tested them on 10 image sets, which were not used during training phase. The segmentation performance was evaluated with the Dice coefficients, computing them in each test case separately. Finally, we statistically compared the provided results. At the end, we evaluated the dependence of segmentation accuracy on training set size, providing practical information for possible users of the classifiers created.

Keywords— brain tumor segmentation, random forest, AdaBoost, SVM.

1. INTRODUCTION

The spread of MRI equipment creates new possibilities in preventive medicine. The early diagnosis of most diseases can change and improve life quality. In this domain the MRI image processing will play always an important role. The brain tumor segmentation from MRI images is focused only on detection and delimitation of cancerous tissues. The tumor segmentation can be very helpful in clinical diagnosis, drug or radiotherapy treatment planning and in tracking tumor evolution. The automat segmentation system can

enhance the diagnostic capabilities of physicians and reduce the time required for accurate diagnosis. Manual segmentation methods are time consuming, can take around 3 hours to complete. Furthermore, there are large variations between expert segmentation even though they use an accurate annotation protocol. In lots of cases an efficient computer based system is required that accurately examines the boundaries of brain tissues with little any human interaction or even without it. Despite numerous efforts and promising results obtained in medical imaging processing, reproducible segmentation and classification of tumors still remains a challenging task. It is considerably influenced by different shapes, locations and image intensities of the analyzed tissues. In this article we present our discriminative segmentation system for brain tumor delimitation from multimodal MR images. The main part of this system is the classification function built obtained by a learning algorithm. There are lots of algorithms used for the same segmentation task; the choice of the most adequate classifiers is not straightforward. The top 10 classifiers were analyzed in [1]. The ranking based on classifiers performances is very difficult, because the evaluation of them is task dependent. These classifiers were selected by numerous researchers in the domain without any ranking criteria. A. N. Mizil in [2] compared the relationship between the predictions made by different learning algorithms and true posterior probabilities, on 8 binary classification problems. The empirical results show that boosted trees, random forests, and SVMs predict the best probabilities, of course after calibration. Following this work R. Caruana et. al. [3] extended the empirical evaluation of supervised learning on high dimensional data. They concluded that boosted trees do very well in modest dimensions, but lose ground to random forests, neural nets, and SVMs as dimensionality increases.

The AdaBoost [4, 5] with boosting variants, the SVM [6, 7] and the RF [8, 9] are classifiers intensively used in image segmentation and also in brain tumor segmentation.

The paper is organized as follows: after a short introduction of the similar systems in the state of the art we briefly present our approach (section II) and the three algorithms compared in this papers: random forest (section III), AdaBoost (section IV) and Support Vector Machines (section V). The theoretical sections are followed by our compar-

tive experiments and results concerning the MR brain tumor segmentation.

II. THE PROPOSED DISCRIMINATIVE MODEL

The proposed discriminative model is similar to previously used model (fig. 1) in [10].

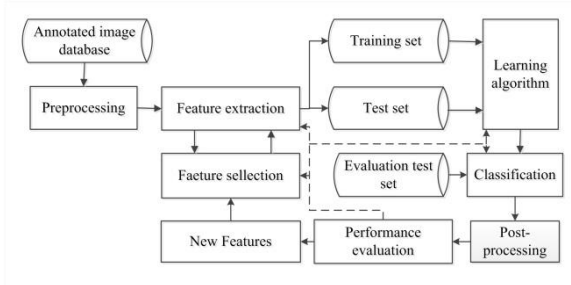


Fig. 1 Components of our system

The performances of our segmentation model, built on a discriminative function, are mainly determined by three important issues: the quality of annotated image-database, the classification algorithm applied and feature set used.

The BRATS 2015 [11] dataset contains 220 high grade (HG) and 54 low grade (LG) brain images with gliomas, assures sufficient diversity, requirement for a performant database. All cases were acquired with similar protocol and contain four types of images: $T1$, $T1c$ (with contrast material Gadolinium), $T2$ and $FLAIR$. A dataset for one patient contains these four types of 3D images and the annotation image. In our approach we consider three classes for performance evaluation used in BRATS Challenges. These classes are: whole tumor - WT (including all four tumor structures), tumor core - TC (including all tumor structures except edema) and active tumor - AT (only the enhancing core).

In our work we have analyzed three important artifacts: inhomogeneity, noise and intensity nonstandardness. For inhomogeneity reduction in MR images, we have applied the $N4$ filter implemented in the ITK package [12]. For noise reduction we used the anisotropic filtering from the same package. Intensity normalization was done by histogram linear transformation in such a way that the first and third quartiles have predefined values.

In our approach we started with defining a large feature set, but this is later reduced. We defined, for each feature, many low-level characteristics that describe the intensities in the neighborhood of the voxels studied. In our application we have used the following features: first order operators (mean, standard deviation, max, min, median, Sobel, gradi-

ent); higher order operator (Laplacian, difference of Gaussian, entropy, curvatures, kurtosis, skewness); texture features (Gabor filter); spatial context features.

By extracting all of these features for every voxel in all modalities, we transform the image segmentation task into a statistical pattern recognition problem. The segmentation process obtained with this statistical model also requires the analysis of variable importance. The appropriate selection of the attributes has to be done according to the target objects. First, we extracted 240 image features of each modality and we obtained a feature vector with 960 elements.

Our algorithm [13] was created to manage this big database and to select a set of adequate features for the given segmentation task. We applied our algorithm several times by evaluating the overall OOB (out-of-bag) error, in order to determine the optimal number of attributes (M) used.

III. RANDOM FOREST ALGORITHM

The random forest is an ensemble classifier built from binary decision trees [14]. The RF classifier is built on two random processes: the random built of bootstrap set and the random feature selection in each node. The creation of RF is based on two sets: the bootstrap set, containing the instances for building a tree and the OOB set (out-of-bag set), containing test instances not included in bootstrap set. The maximization of information gain is the splitting criterion applied in every node. Applying this criterion each node splits the incoming in-instances in two sets. In order to evaluate the information gain, the RF uses in each node only a small number of variables (m_{tries}) out of all existing variables (M). These m_{tries} variables are chosen randomly and the splitting criterion is maximized only with these variables. The RF algorithm has two main parameters: the number of trees K_{trees} used in forest and the number of randomly chosen variables in each node m_{tries} . In the state of the art there are no indications regarding the choice these parameters. They are dataset dependent. Determining the appropriate values of these variables may be the objective of an optimization framework [10]

IV. ADABOOST ALGORITHM

One of the most used algorithms in images processing in the last decade is the AdaBoost algorithm, proposed by Freund and Shapire [15]. It constructs an ensemble of classifiers and uses a voting mechanism for the final classification. The idea of boosting is to use several weak classifiers to form a highly accurate prediction rule by calling the weak classifier repeatedly on different distributions of the training set. AdaBoost was the first adaptive boosting algorithm,

which automatically adjusts its parameters of the algorithm according to the data used in the current iteration. Initially, all the weights are set equally, but each round the weights of incorrectly classified examples are increased so that the images which were incorrectly classified by the previous classifier, will receive greater weight on the next iteration. This forces the learner to concentrate on the hard examples. The weight of the weak learner is computed, in fact, from the classification error of the current weak classifier. The final classifier reduces the statistical dispersion of the decision and at the same time reducing the classification error too.

This algorithm [15] is an iterative method. In each iteration, a weak classifier having the lowest error rate on the training set, is selected. This step is followed by the re-weight of the input instances, putting an accent on hard examples. The restriction of the used weak classifiers is related to their performance which has to be better than the random decision. Thus, weak classifier has to eliminate more than half of the backgrounds. This means that the final classifiers formed of N weak classifiers will have a very small false detection rate.

The most important theoretical propriety of AdaBoost concerns in its ability to reduce the training error. The AdaBoost converts a set of weak classifiers into a strong learning algorithm which can generate an arbitrarily low error rate on the training set. Practically this statement is limited in two ways. First, the number of used classifiers is finite, thus an arbitrary low error rate is not possible to obtain. Second, the computational complexity of training phase increases with each classifier added. By obtaining an equilibrium between the two factors we are able to combine enough weak classifiers in order to create the desired strong classifier [16].

The most important drawback of the algorithm is the overfitting. On the phenomenon of overfitting we understand the fact that the training set is much better approximated. The classification surface lies exactly on the training entities and underperforms for the real data. To avoid overfitting, the task for the algorithm therefore should not be to find the best possible classifier for the underlying training sample, but rather to find the best prediction rule for a set of new observations [17].

V. SUPPORT VECTOR MACHINES

The Support Vector Machines [18] are supervised learning machines for binary classification problems. They are able to separate the inputs linearly, or if they are not linearly separable they can map them in a higher dimensional feature space, where the linear separation is possible. The

learning algorithm finds the best hyperplane, which separates the entities included in the training set. In other words, it finds the hyperplane which maximizes the distance to the nearest entities in each class. The larger the separating margin between the classes the lower is the generalization error of the obtained classifier [19]. The optimization problem consists of maximizing the distance between the closest data points.

Given k inputs in the training data set

$$S = (x_i, y_i) | x_i \in \mathbb{R}^n, y_i \in \{+1, -1\}; i = 1, 2, \dots, k \quad (1)$$

The classification is achieved by a hyperplane of this form:

$$w^T \Phi(x) + b = 0, w \in \mathbb{R}^n, b \in \mathbb{R} \quad (2)$$

where Φ is the transformation of the inputs in a higher-dimensional space, b is the bias, the translation of the hyperplane from origin and the w is the normal vector of the hyperplane [20].

The distance from a point x_i to the hyperplane P is the length of the perpendicular segment from the point x_i . Or it can be computed as the distance of the projection of x_i on the normal vector of the plane and any point on the plane $x \in P$ the distance has the form:

$$dist = \frac{1}{\|w\|} |w^T (\Phi(x_i) - \Phi(x))| = \frac{1}{\|w\|} \quad (3)$$

The optimization problem becomes the maximization of the distance

$$\max \left(\frac{1}{\|w\|} \right) \text{ with the constraint } y_i (w^T \Phi(x_i) + b) \geq 1 \quad (4)$$

where $\|w\|^2 = w^T w$. The solution is obtained from the *Lagrangian* of the problem with respect to k inequality constraints.

$$L(w, b, \lambda) = \frac{1}{2} w^T w - \sum_{i=1}^k \lambda_i [y_i (w^T \Phi(x_i) + b) - 1] \quad (5)$$

The solution will be given by the quadratic programming with Sequential Minimal Optimization SMO [21].

The most important advantages of SVM is its robustness and classification accuracy for the training data set. In case of a sufficiently general training data set it also offers correct classification of the future data instances. It can handle classification in two different classes regardless the dimensionality of the data. If the linear separation is not possible several kernels may be used to transform the original non-

separable data into a higher dimensional space where the linear separation is possible. The classification of new instances is very simple; it only verifies the position of them relative to the hyperplane. SVM has also considerable disadvantages. The optimization of the quadratic programming problem is not obvious. It can be solved only with iterative methods. The analytical solution is not possible. The multiclass classification can be done only with pairwise two class decisions. The final classification is a binary decision which can be transformed in a weighted decision if the distance of the analyzed instance to the hyperplane is taken in consideration.

VI. EXPERIMENTS AND RESULTS

We have measured the obtained performances of the above presented three classifiers. Each algorithm created its classifier by using the same training set. This training dataset is a random subsample 50:1 of the total information from 20 HG images used in BRATS 2013 Challenge [11]. Thus, the training set size is about 2% of the whole dataset. The test set contains all information of 10 HG images. In order to transpose the image segmentation into a statistical pattern recognition task, we applied a voxelwise classification, in each voxel we evaluated 120 low level image features. These features were obtained from a large set by evaluating their importance for brain tumor segmentation [13]. Our optimized RF classifier is composed of $K_{trees}=100$ trees, each having a size of $T_{size}=2048$ nodes [10]. The splitting criterion is evaluated with $m_{tries}=9$ randomly chosen features from the whole $M = 120$ feature set. The classification performance can be easily converted in Dice coefficient, which is used to compare segmentation performances:

$$Dice = 2 \cdot |S_1 \cap S_2| / (|S_1| + |S_2|) \quad (6)$$

where: $||$ is the cardinality, S_1 the true region, S_2 the obtained segmentation. The segmentation results obtained on 10 test images are graphically given in fig. 2.

In order to test the AdaBoost classifier we used decision stamps and 12 iterations, the same depth as used for tree size in RF. The SVM used a polynomial kernel with optimized exponent. For these two classifiers we defined two class classification tasks: WT-vs.-all and TC-vs.-all. The segmentation results obtained on the same test set are shown in fig. 3 and fig. 4. Table 1 presents a statistical evaluation of Dice coefficient obtained with each of the three classifiers. Overall, we can conclude, based on these results, that the RF classifier reaches significantly more accurate performances than the other two classifiers.

Table 1. Statistical parameters of the compared algorithms

	RF		SVM		AdaBoost	
	Mean	Std.	Mean	Std.	Mean	Std.
WT	0.905	0.035	0.736	0.050	0.720	0.094
TC	0.887	0.046	0.817	0.072	0.791	0.094

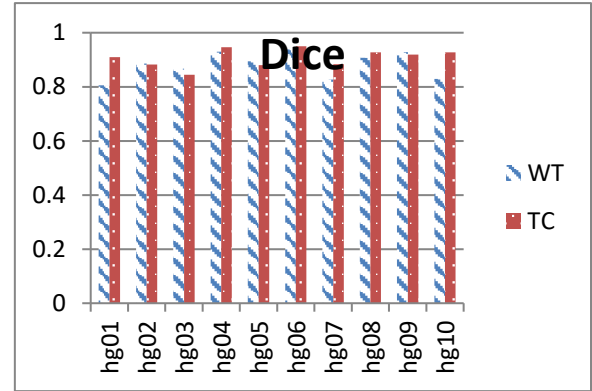


Fig. 2 Dice coefficients of Random Forest

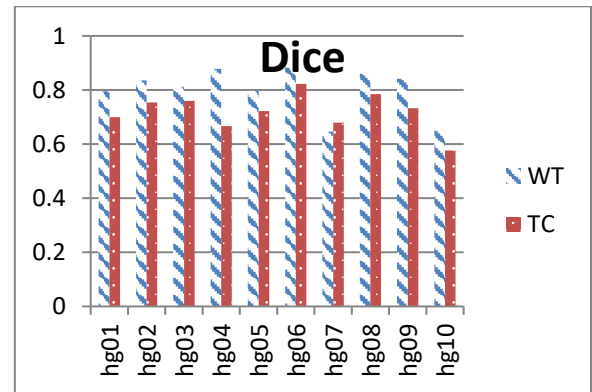


Fig. 3 Dice coefficients of SVM algorithm

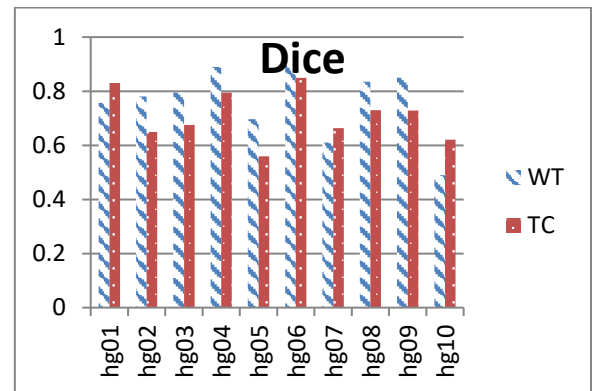


Fig. 4 Dice coefficients of AdaBoost

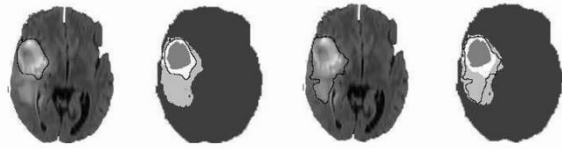


Fig. 5 Segmentation on the test set – usual example

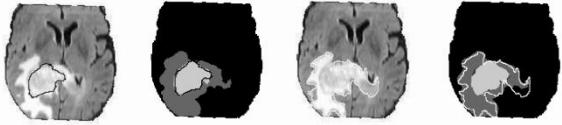


Fig. 6 Segmentation on the test set – hard example

VII. CONCLUSIONS

The segmentation performances obtained by our discriminative model built on RF are explainable by the most important characteristics of this classifier:

- efficient run on big data sets;
- easy handling of multi classification task;
- smooth transition between class borders [22].

Following the results obtained, during our research described in this work and also in the previous articles [10, 13] we have built a segmentation system. In fig.5 and fig.6 we graphically presented our segmentation results on a brain slice of the test set. The black line is the contour of the annotation. The light gray region is the detection of the edema (WT); the white region is the result for the tumor core (TC) and dark grey is the necrotic part of tumor. The segmentation obtained by our approach has performances comparable to state-of-the-art systems (table 2) [11].

In the future we intend to adapt and integrate the proposed system for biometric images, more exactly for dorsal hand vein segmentation.

Table 2. Comparison of segmentation results

	Our clas- sif.	Brats2012 [11]	Brats 2013 [11]
WT HG	75-86%	63-78%	71-87%
CT HG	71-82%	24-37%	66-78%

ACKNOWLEDGMENT

This work was supported by a grant of the Romanian National Authority for Scientific Research and Innovation, CNCS – UEFISCDI, project number PN-II-RU-TE-2014-4-2080.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

REFERENCES

1. Wu X, Kumar V, Quinlan JR, Ghosh J, Yang Q, Motoda H, et al. Top 10 algorithms in data mining. *Knowledge and information systems*. 2008; 14(1): p. 1-37.
2. Niculescu-Mizil A, Caruana R. Predicting good probabilities with supervised learning. In *Proceedings of the 22nd international conference on Machine learning*; 2005. p. 625-632.
3. Caruana R, Karampatziakis N, Yessensalina A. An empirical evaluation of supervised learning in high dimensions. In *Proceedings of the 25th international conference on Machine learning*; 2008. p. 96-103.
4. Nivetha P, Vijayanthi N. Detection and segmentation of brain tumors based on multifractal features. 2015.
5. Rangini M, Jiji GW. Identification of Alzheimer’s Disease Using Adaboost Classifier. In *Proceedings of the International Conference on Applied Mathematics and Theoretical Computer Science*; 2013. p. 229.
6. Nichat AM, Ladhake SA. Brain Tumor Segmentation and Classification Using Modified FCM and SVM Classifier. *Brain*. 2016; 5(4).
7. Shahare PD, Giri RN. Comparative Analysis of Artificial Neural Network and Support Vector Machine Classification for Breast Cancer Detection. 2015.
8. Reza S, Iftekharuddin KM. Improved Brain Tumor Tissue Segmentation Using Texture Features. In *MICCAI-BRATS Challenge on Multimodal Brain Tumor Segmentation*; 2014.
9. Zikic D, Glocker B, Konukoglu E, Shotton J, Criminisi A, Ye D, et al. Context-sensitive classification forests for segmentation of brain tumor tissues. In *MICCAI-BRATS Challenge on Multimodal Brain Tumor Segmentation*; 2012.
10. Lefkovits L, Lefkovits Sz, Szilágyi L. Brain Tumor Segmentation with Optimized Random Forest. In *Medical Image Computing and Computer Assisted Intervention (MICCAI) BrainLes Workshop*; 2016, under review.
11. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). *IEEE Transactions on Medical Imaging*. 2015 Oct; 34(10): p. 1993-2024.
12. The Interactive Learning and Segmentation Toolkit ITK..
13. Lefkovits L, Lefkovits Sz, Emerich S, Vaida MF. Random Forest Feature Selection Approach for Image Segmentation. In *Advanced Concepts for Intelligent Vision Systems ACIVS*; 2016, under review.
14. Breiman L. Random forests. *Machine learning*. 2001; 45(1): p. 5-32.
15. Schapire RE. The boosting approach to machine learning: An overview. In *Nonlinear estimation and classification.*: Springer; 2003. p. 149-171.
16. Lefkovits Sz, Lefkovits L. Distance based k-NN Classification of Gabor Jet Local Descriptors. In *9th International Conference, Interdisciplinarity in Engineering, Tîrgu-Mureş; 2015 8-9 October*. p. 780-785.
17. Mayr A, Binder H, Gefeller O, Schmid M, others. The evolution of boosting algorithms. *Methods of information in medicine*. 2014; 53(6): p. 419-427.
18. Cortes C, Vapnik V. Support-vector networks. *Machine learning*. 1995; 20(3): p. 273-297.

19. Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods: Cambridge university press; 2000.
20. Lefkovits Sz, Lefkovits L. Enhanced Gabor Filter Based Facial Feature Detector. In The 5th International Conference European Integration - Between Tradition and Modernity IETM, Tîrgu-Mureş; 2013 5-7 October. p. 1204-1214.
21. Platt JC. Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. Tech. rep. ADVANCES IN KERNEL METHODS - SUPPORT VECTOR LEARNING; 1998.
22. Criminisi A, Shotton J, Konukoglu E. Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Foundations and Trends® in Computer Graphics and Vision. 2012; 7(2--3): p. 81-227.

Author: László Lefkovits
Institute: Sapientia University, Tîrgu-Mureş
Street: Corunca 1c
City: Tîrgu-Mureş
Country: Romania
Email: lefkolaci@ms.sapientia.ro