The role of atlases and multi-atlases in brain tissue segmentation based on multispectral magnetic resonance image data

David Iclănzan, Rodica Ioana Lung, Zsolt-Levente Kucsván, Béla Surányi,
Levente Kovács, László Szilágyi

2021 IEEE AFRICON | 978-1-6654-1984-0/21/$31.00 ©2021 IEEE | DOI: 10.1109/AFRICON51333.2021.9570952

***Abstract*— Atlas assisted image segmentation has been quite popular in medical imaging during the last two decades. The atlas is able to provide prior information on the imaged organ's shape, appearance, and local texture or intensity distribution. In case of segmenting images via pixelwise classification, the final segmentation result is obtained through a fusion of the classification outcome with the local atlas information. In other words, the atlas guides the classifier towards the shape of local structures normally situated at the given location. This paper proposes to demonstrate the advantages a multi-atlas can bring in a segmentation process of the main tissues in infant brain based on multi-spectral MRI records. Three supervised machine learning methods are deployed to segment brain tissues, with and without the use of the atlas. Differences are evaluated using statistical accuracy indicators. Atlases improved the overall segmentation accuracy by 2.5-3.5%, depending on the deployed classifier method.**

***Index Terms*— magnetic resonance imaging, brain tissue segmentation, atlas-based segmentation, infant brain.**

I. INTRODUCTION

Atlases and multi-atlases are usually involved in medical image segmentation to provide additional information on the investigated object (organ), thus contributing to the quality of segmentation. Without using shape models or atlases, the segmentation process can only rely on local features of pixels, like texture and intensity distributions. Atlases may provide information on what is usually found in the same location in other similar images, or what intensity distributions characterize certain locations in other normal records.

Atlases have many recent applications in medical image segmentation, involving organs like: lung [1], pancreas [2], prostate [3], bones [4], cartilage [5], brain tissues and lesions [6], [7], [8], [9], cardiac structures (e.g. myocardium) [10], [11], and multiple abdominal organs [12]. Atlases are involved in the processing of image data collected through various imaging modalities, including magnetic resonance imaging (MRI) [3], [6], [7], computed tomography (CT) [1], [4], CT angiography [10], positron emission tomography (PET) [11], X-ray [13] and mammography [14]. For further information on atlas-based image segmentation techniques, there are available systematic review papers (e.g. [15], [16]).

This paper proposes to demonstrate the advantages a multi-atlas can bring in a segmentation process of the main tissues in infant brain based on multi-spectral MRI records. The difficulty of the problem consists in the very similar intensity of the white matter and grey matter at the age of 6 months [17]. Under such circumstances, shape priors are likely to provide valuable information into the segmentation process and to assist the improvement of segmentation quality.

The rest of this paper is structured as follows. Section II presents the data, the segmentation framework, and the multi-atlas involved in the comparative study. Section III relates on and discusses the segmentation results achieved with and without the use of the atlas, while section IV concludes the study.

II. MATERIALS AND METHODS

A. Data

This study uses ten infant brain records that served as training data in the iSeg-2017 challenge [17]. Each volume may contain up to 256 slices of 144×192 pixels, but only 100-110 slices contain brain pixels. Pixels are isovolumetric, they represent one cubic millimeter of brain tissues. All records are multispectral, they contain T1 and T2-weighted observed values collected by the MRI equipment. The two observed volumes were registered to each other using an automated algorithm. Each pixel has a label (ground truth) produced by human experts, indicating the tissue type where it belongs. Three tissue types are distinguished in the ground truth: cerebro-spinal fluid (CSF), grey matter (GM), and white matter (WM). The skull and any other non-brain tissues were removed from all volumes. The segmentation

This work was partially supported by a grant of the Romanian Ministry of Education and Research, CNCS - UEFISCDI, project number PN-III-P4-ID-PCE-2020-2360, within PNCDI III.

This work was supported in part by the Sapientia Institute for Research Programs. The work of L. Kovács was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 679681). L. Szilágyi is János Bolyai Fellow of the Hungarian Academy of Sciences.

R. I. Lung is with the Centre for the Study of Complexity, Babeş-Bolyai University, Str. Teodor Mihali 58-60, 400594 Cluj-Napoca, Romania (phone: +40-264-412-570); e-mail: rodica.lung at econ.ubbcluj.ro).

D. Iclănzan, Zs.-L. Kucsván, B. Surányi, and L. Szilágyi are with Computational Intelligence Research Group (CIRG), Sapientia University, Calea Sighişoarei 1/C, 540485 Tîrgu Mureş, Romania (phone: +40-265-206-210; fax: +40-265-206-211; e-mail: {iclanzan, lalo} at ms.sapientia.ro).

Zs.-L. Kucsván is also with Université de Rennes 1, Rennes, France (e-mail: zsolt-levente.kucsvan at etudiant.univ-rennes1.fr).

B. Surányi is also with Doctoral School of Applied Mathematics and Applied Informatics (AIAMDI), Óbuda University, Bécsi út 96/b, H-1034 Budapest, Hungary (phone/fax: +36-1-666-5585; e-mail: bela.suranyi at gmail.com).

L. Kovács and L. Szilágyi are with Biomatics Institute, John von Neumann Faculty of Informatics, Óbuda University, Bécsi út 96/b, H-1034 Budapest, Hungary (phone/fax: +36-1-666-5585; e-mail: {kovacs, szilagyi.laszlo} at uni-obuda.hu).

978-1-6654-1984-0/21/$31.00 ©2021 IEEE

of these volumes has an inherent difficulty consisting in the fact that patients are six-month old, which is the age when the WM and GM tissues have very similar intensity values in the T1 data channel.

B. Preprocessing

The records of the iSeg-2017 challenge consists of relatively good-quality MRI data, which contain no relevant amount of high frequency noise. The presence of intensity inhomogeneity [18], [19] was not investigated. The main preprocessing step in this study consists in the normalization (standardization) of histograms, which is a mandatory step in magnetic resonance imaging, because the absolute numerical intensity values provided by the MRI equipment can only be interpreted together with their context. To accomplish the histogram normalization, the method of Nyúl *et al* [20] was deployed, which registers together the histograms of a batch of MRI records based on some predefined milestones. Our application used five milestones for histogram registration, namely, the 0.5, 25, 50, 75, and 99.5 percentiles. Another preprocessing step consists in handling the very few missing data. Missing intensity values were filled with averaged values extracted from the neighborhood of the pixels in question.

The final preprocessing step consists in feature generation. The two observed features of the pixels, the normalized T1 and T2 intensity values, cannot support a fine segmentation by themselves. Further information of the pixels can be found in their immediate neighborhood, mainly because the automated registration of data channels T1 and T2, performed by the experts of the iSeg-2017 database, is not perfect. Therefore, it is likely to achieve better segmentation if we extract some additional features from the planar and spatial neighborhood of each pixel. Average T1 and T2 values were extracted from planar 3×3, 5×5, $\ldots$, 11×11 neighborhoods (10 features), and minimum, maximum, and average of T1 and T2 from $3 \times 3 \times 3$ spatial neighborhood (6 features). Together with the two observed features, the feature vectors sums up to 18 features. Pixel localization features can be involved in the study by adding normalized x, y and z coordinates to the feature vector.

C. Atlas creation

Involving an atlas in the segmentation process allows us to provide additional information regarding the spatial position of the pixels, and the tissue composition that is likely to be at that position. The initial step of an atlas creation is registering together all MRI volumes from the training data set. Our study relies on rigid registration that uses normalized relative coordinates. First the gravity center $O(x_0, y_0, z_0)$, and the standard deviation of x, y, and z coordinates $(\sigma_x, \sigma_y, \sigma_z)$ are extracted from each volume. The relative coordinates of a pixel situated at (x, y, z) in a given volume are $(\frac{x-x_0}{\sigma_x}, \frac{y-y_0}{\sigma_y}, \frac{z-z_0}{\sigma_z})$. All pixels of the training volumes are mapped onto a discrete mesh of relative coordinates. The distribution of the three labels (CSF, GM, WM) of pixels mapped onto each discrete mesh point of relative coordinates is computed. This way we obtain a spatial map of probabilities for each tissue type. One slice of these three maps are exhibited in Fig. 1. Selecting the maximum probability at each mesh point, it is possible to obtain an average segmentation accuracy of 67-68% for the testing volumes not included in the atlas. This segmentation accuracy is achieved by the use of pixel coordinates, without using the pixel intensities.

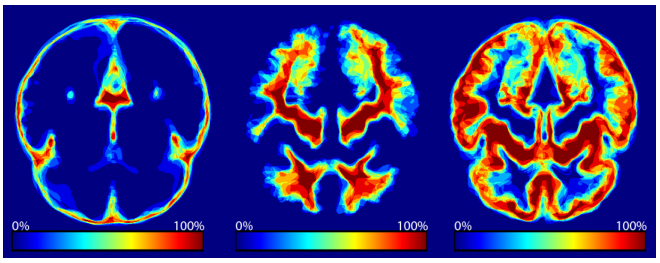

Fig. 1. The atlas obtained for a certain slice: probability maps of CSF (left), white matter (middle) and grey matter (right).

D. Classification

Machine learning techniques were used to perform the classification of pixels into three classes that correspond to the three main tissue types from the brain: CSF, GM, and WM. As the number of multispectral MRI records in the iSeg-2017 dataset is very low, the leave-one-out technique was adopted: one record is selected for testing at a time, and all other nine records were involved in the training of the classifier. Of course, the ten records took turns in serving as testing data, so all ten records were segmented during the evaluation process.

Three machine learning methods were involved in this study: random forest (RF) [21], [22], [23], k-nearest neighbors (kNN), and discrete AdaBoost (ADA) [24].

1) Random Forests: Random forests [21] classify data by via bagging [25], i.e. by building a large collection of trees on samples drawn with replacement from the training data and averaging them by a voting mechanism.

Thus, if the number of trees is B, in order to build each random-forest tree T_b, $b = 1, \ldots, B$ a bootstrap sample Z^* of size N is drawn from the training data and T_b is grown on Z^* in the following recursive manner: for each terminal node in the tree, m out of the p variables are selected, the best split-point among the m is picked and the node is split in two. These steps are repeated until a minimum node size is reached.

The main intuition behind random-forests is to reduce variance by reducing the correlation between trees. This is achieved by randomly selecting the m variables from the bootstrapped dataset for splitting. Typical values for m are $\sqrt{p}$ or even 1 [26].

The output of the algorithm consists of the ensemble of trees $\{T_b\}_{b=\overline{1,B}}$. The random-forests prediction is typically computed using majority voting, but other aggregation methods can be used. The scikit-learn [27] implementation averages the probabilistic predictions of the trees. On many

problems the performance of forests is similar to that of boosting.

2) kNN: k - Nearest Neighbor (kNN) is a classification method that assigns a class to a test instance based on the classes of its k neighbors by using a voting mechanism [28]. If a simple majority is used then:

$$C(x) = \arg\max_{c \in \mathbb{C}} \sum_{(x_i, c_i) \in N_x} \mathbb{I}(c = c_i), \tag{1}$$

where N_x is the neighborhood of x, having k instances defined based on some similarity measure, $\mathbb{C}$ is the set of class labels, c_i is the class label of the i-th neighbor x_i, and $\mathbb{I}$ is the indicator function (returns 1 if condition is true, otherwise 0). For real-valued data, Euclidean distance is mostly used. A weighted voting mechanism can be used instead of eq. (1). The efficiency of kNN is obviously affected by the choice of k, the choice of voting mechanism and that of the similarity measure used to defined the neighbors of an instance. Another drawback consists in the fact that while the model consists in storing the training data for reference, classifying unknown instances can become expensive for large training sets.

3) Discrete AdaBoost: AdaBoost, first proposed as AdaBoost.M1 in [29], is a learning method that efficiently combines the results of many runs of a weak classifier to produce a strong output. AdaBoost repeatedly applies the classifier on weighted training data, adjusting the weights during each iteration in a manner that emphasizes miss-classified observations in each iteration. A weak classifier is a classification method with an error rate slightly better than random guessing. The final output is constructed as a linear combination of the classifiers fitted each iteration.

Based on the original AdaBoost, SAMME - *Stagewise Additive Modeling using a Multi-class Exponential loss function* [30] was adjusted for multi-class problems by tweaking the way weights are updated to take into account the random guessing error of these problems. For a multi-class problem with K classes the error rate for random guessing is $(K-1)/K$.

Given the training data $X \subset \mathbb{R}^{n \times p}$ with response variable $C \subset \mathbb{N}$, with $c_i \in \{1, \ldots, K\}$ and a weak classifier $T(\cdot)$, SAMME will apply T on weighted values of X for M iterations successively by using each iteration different weights w_i. Initial weights w_i are taken equal to $1/n$. Each iteration m, $T^{(m)}(X)$ is fitted to observations $x_i \in X$ weighted by w_i and the corresponding weighted error rate $err^{(m)}$ is computed. Coefficients $\alpha^{(m)}$ that are used to update weights for the next iteration and for the final output of the algorithm are computed as:

$$\alpha^{(m)} = \log \frac{1 - err^{(m)}}{err^{(m)}} + \log (K - 1) \ . \tag{2}$$

Weights w_i are updated as $w_i \leftarrow w_i \exp \alpha^{(m)}$ for all instances i miss-classified by $T^{(m)}$; after that they are re-normalized. The output

$$C(x) = \arg\max_{k} \sum_{m=1}^{M} \alpha^{(m)} \mathbb{I}(T^{(m)}(x) = k) \tag{3}$$

combines the outputs of the M classifiers weighted by $\alpha^{(m)}$. The only difference between SAMME and AdaBoost.M1 is in eq. (2) where the term $\log K - 1$ is added in SAMME in order to ensure that $\alpha^{(m)}$ is always positive. Just like AdaBoost is equivalent to a forward stagewise additive model with an exponential loss function [31], it is shown that adding the term $\log (K-1)$ makes SAMME equivalent to fitting a stagewise additive model with a multi-class exponential loss function [30].

All three classifiers were tested on various sizes of the training data set, ranging from 1k to 200k randomly selected pixels from each train record. More exactly, RF was tested up to 200k, AdaBoost up to 100k, while kNN up to 50k pixels per training record. This difference is due to the computational burden of each algorithm. Nevertheless, the main goal is not to compare the accuracy of the classifiers in the segmentation problem, but to show the beneficial effect of the atlas. Further parameters of the deployed classifier methods are listed below:

- kNN was used with $k = 11$;
- Random forest using maximum 150 trees with maximum depth growing together with the training data size in the range between 12 and 29;
- Discrete AdaBoost used maximum tree depth growing together with the training data size in the range between 12 and 26.

kNN and RF performed directly the classification of pixels into three classes (CSF, GM, WM), while AdaBoost worked in two stages: first it separated CSF from the other two, and then it separated GM from WM.

E. Label fusion

The processing step, called label fusion has the main goal to create a unique decision for each pixels based on two inputs, namely, the estimations given by the classifier and the atlas. Both inputs are probabilistic and non-binary. On the other hand, the output of the label fusion gives each pixel a crisp label with respect to the three classes. Pixels must belong to a single class, either CSF, or GM or WM. Our solution employs the so-called simultaneous truth and performance level estimation (STAPLE) method [32] to perform the label fusion.

F. Evaluation criteria

The segmentation results are evaluated via comparing the decisions made by the classification algorithm with the ground truth. Let us denote the set of labels by Ω: $\Omega = \{\mathrm{CSF}, \mathrm{GM}, \mathrm{WM}\}$. Further on, for any $i \in \Omega$, let Γ_i stand for the set of pixels from a volume that belong to class i by ground truth, and Λ_i the set of pixels assigned to class i by the classifier. Let Ψ be the set of all pixels of a volume. Obviously $\Psi = \Gamma_{\mathrm{CSF}} \cup \Gamma_{\mathrm{GM}} \cup \Gamma_{\mathrm{WM}} = \Lambda_{\mathrm{CSF}} \cup \Lambda_{\mathrm{GM}} \cup \Lambda_{\mathrm{WM}}$. The main accuracy indicators involved in this study are:

- Sensitivity (or recall or true positive rate) with respect to class $i \in \Omega$, defined as $\mathrm{TPR}_i = \frac{|\Gamma_i \cap \Lambda_i|}{|\Gamma_i|}$, which represents the rate of correct decisions regarding pixels that belong to class i by ground truth;

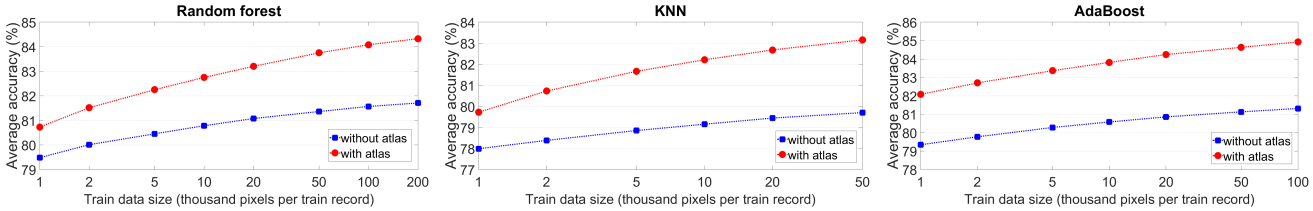

Fig. 2. Average global accuracy achieved by three machine learning methods, without and with the use of the atlas, plotted against the size of the training data set.

TABLE I

AVERAGE VALUES OF THE MAIN ACCURACY INDICATORS (EXPRESSED IN %) OBTAINED BY THE CLASSIFIER ALGORITHMS, USING 50K PIXELS PER TRAINING RECORD

Classifier algorithm	Atlas status	CSF				GM				WM				Overall accuracy
		DSC	TPR	TNR	PPV	DSC	TPR	TNR	PPV	DSC	TPR	TNR	PPV	
Discrete AdaBoost	without	87.02	84.36	97.41	89.95	81.19	84.55	78.88	78.18	76.58	73.82	91.84	79.82	81.13
	with	89.19	87.42	97.66	91.14	84.54	87.57	82.63	81.84	81.31	78.52	93.79	84.67	84.63
Random forest	without	86.99	84.78	97.24	89.42	81.31	85.04	78.52	77.99	76.37	72.82	92.32	80.58	81.14
	with	88.43	86.60	97.48	90.43	83.65	87.74	80.44	80.04	79.44	75.14	94.02	84.62	83.50
K-nearest neighbors	without	85.80	82.45	97.34	89.52	79.62	82.22	78.39	77.30	75.34	74.13	90.23	76.86	79.70
	with	87.54	84.43	97.69	90.99	83.02	85.60	81.72	80.71	80.13	78.87	92.32	81.78	83.16

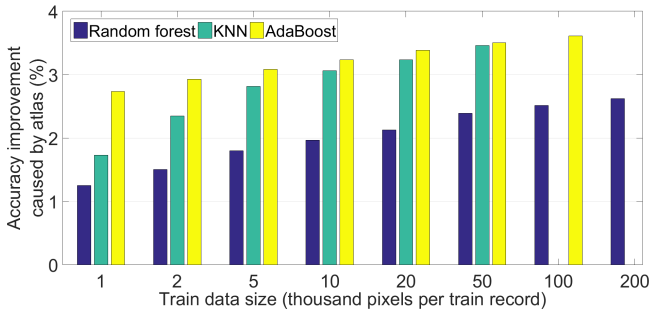

Fig. 3. Improvement of global accuracy caused by the use of atlas, for various machine learning methods and training data sizes.

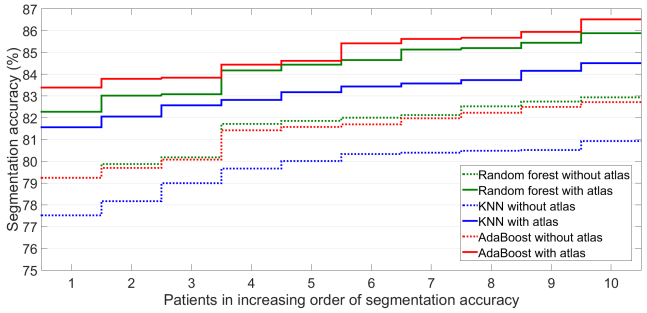

Fig. 4. Accuracy obtained for individual patients by the three machine learning methods, with and without using the atlas.

- Positive predictive value (or precision) with respect to class $i \in \Omega$, defined as $\mathrm{PPV}_i = \frac{|\Gamma_i \cap \Lambda_i|}{|\Lambda_i|}$, which represents the rate of correct decisions regarding pixels that were assigned to class i by the classifier;
- Specificity (or true negative rate) with respect to class $i \in \Omega$, defined as $\mathrm{TNR}_i = \frac{|\Psi \setminus (\Gamma_i \cup \Lambda_i)|}{|\Psi \setminus \Gamma_i|}$;
- Dice similarity coefficient (or Dice score) of class $i \in \Omega$, defined as $\mathrm{DSC}_i = \frac{2 \times |\Gamma_i \cap \Lambda_i|}{|\Gamma_i| + |\Lambda_i|}$;
- and overall accuracy defined as $\mathrm{ACC} = \frac{\sum_{i \in \Omega} |\Gamma_i \cap \Lambda_i|}{|\Psi|}$.

In all above presented formulas, $|X|$ stands for the cardinality of set X. All these accuracy indicators can have positive values up to 1, higher values reflect finer accuracy. In this study we express the accuracy indicators in percentages. The accuracy indicator values are first established for each individual record separately, and then the average value is computed to characterize global accuracy.

III. RESULTS AND DISCUSSION

Three classifier algorithms were deployed to evaluate the segmentation accuracy of brain tumor tissues, with and without using the atlas. For each classifier, a range of training data sizes were tested, keeping the within reasonable bounds. Each test was run multiple times (five to fifteen) using different, randomly sampled training data. Averaged values of accuracy indicators are reported for each test in the following paragraphs.

Figure 2 presents the evolution of the main accuracy indicator, namely the global rate of correct decisions, obtained by the three tested classifier algorithms at various training data sizes, with and without using the atlas. A common property of all three classifiers is the fact that the accuracy improves if larger training data sets are involved. Further on, the use of the atlas is beneficial in all cases, and the difference caused by the use of atlas also grows together with the training data size. The best accuracy was achieved by the discrete AdaBoost classifier, having average accuracy of 85% with the use of atlas, at 100k feature vectors per training record.

Figure 3 shows the improvement of average accuracy caused by the use of atlas, in case of various training data sizes. Also here it is visible that the larger the training data set, the better the improvement brought by the atlas. Accuracy differences in case of the random forest classifier

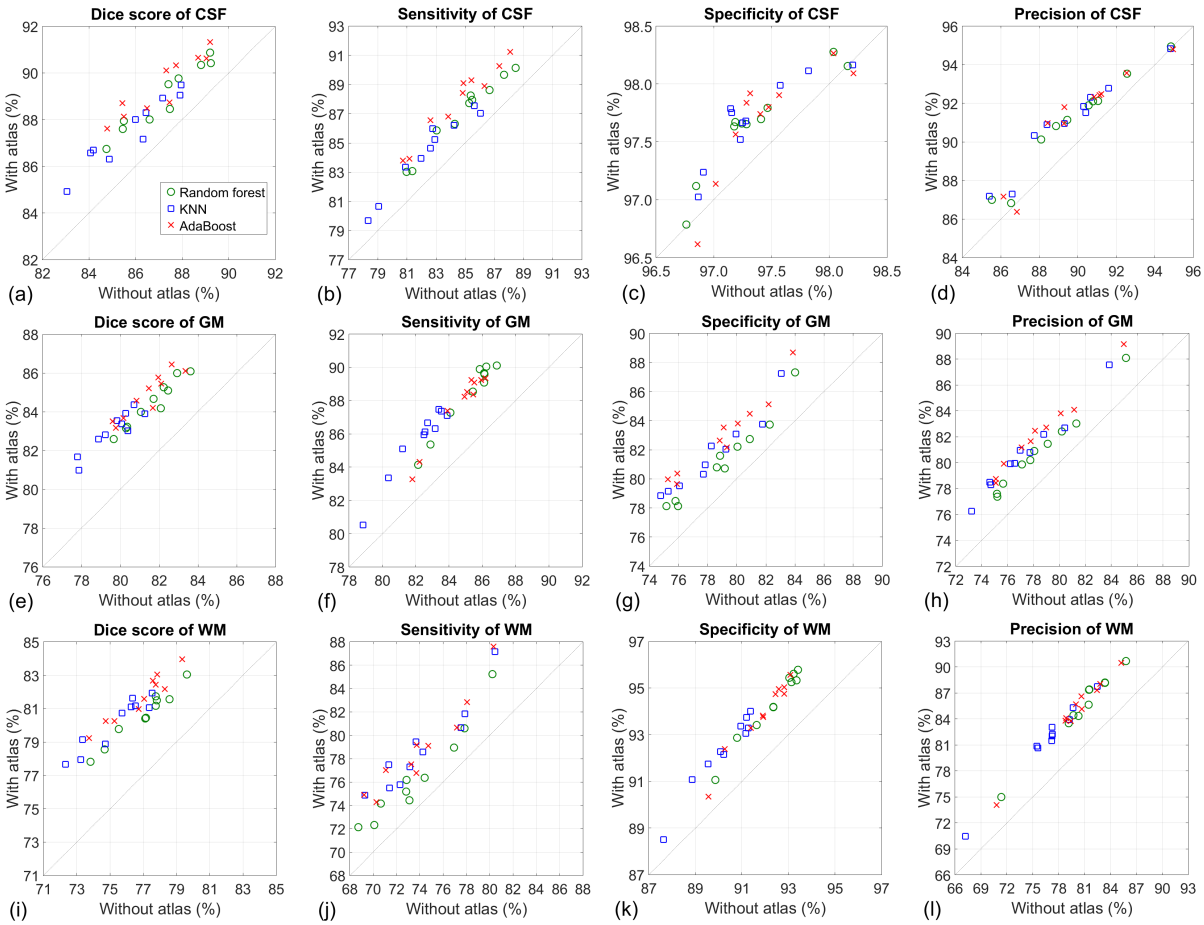

Fig. 5. Scatter plots of various accuracy indicators, with atlas values plotted against without atlas values. Red crosses (×), blue squares (□), and green circles (○) represent segmentation results produced on individual MRI records by the discrete AdaBoost, kNN, and random forest classifiers, respectively.

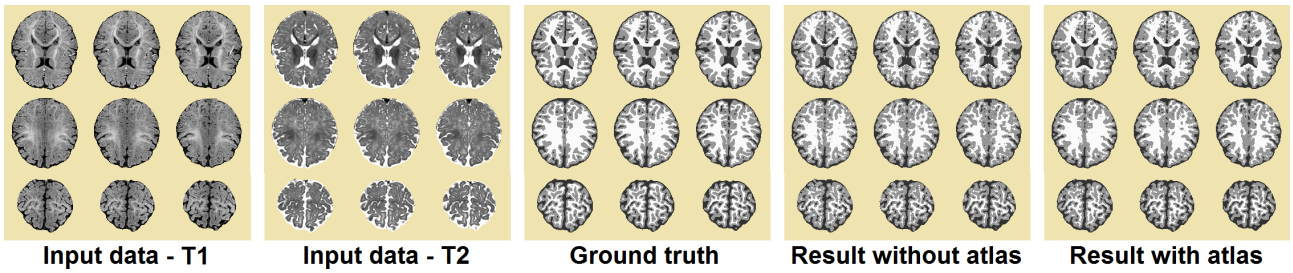

Fig. 6. Nine slices of the brain of Patient no. 8, the normalized input data in channels T1 and T2, the ground truth, and the segmentation result obtained by the AdaBoost classifier at 50k pixels per training volume, without (ACC=82.7%) and with atlas (ACC=86.3%).

exceed 2.5%, while in case of kNN and discrete AdaBoost they are even higher, with values up to 3.5%

Figure 4 shows the rate of correct decision obtained for individual MRI records by the three classifier algorithm, when using their largest tested training data size. For each graph, records were sorted in increasing order of the achieved accuracy. Dotted lines show the results obtained without the use of atlas, while continuous lines the accuracy of atlas based segmentation. These plots clearly show that AdaBoost achieved the best performance, followed by random forest and kNN. Figure 5 presents twelve scatter plots indicating various accuracy indicators obtained for individual MRI records, plotted as values obtained with atlas vs. values obtained without atlas. In each panel, indicators that improved by the presence of the atlas are situated above the diagonal. Dice scores, sensitivity values, specificity values, and precision values are presented in the first, second, third, and fourth column, respectively. Accuracy indicators obtained for CSF, GM, and WM tissues are exhibited in the first, second, and third row, respectively. Only the graphs c) and d) contain a small number of marks below the diagonal, suggesting that the use of atlas improved the very large majority of accuracy indicators. The advantages brought by the use of atlas surely outweigh the disadvantages.

Table I exhibits the average values of the main accuracy indicators, obtained by the three classification algorithms with and without the use of the atlas. These values were obtained by each algorithm using the same number of training data, 50k vectors from each training volume. This table shows that all accuracy indicators values improve with the use of the atlas. It also suggests that out of the three tested classifier methods the discrete AdaBoost provides the best segmentation accuracy.

Figure 6 presents nine selected slices from the brain volume of patient number 8, the normalized input data of channels T1 and T2, the ground truth, and the segmentation result obtained by the AdaBoost classifier without and with using atlas. This result was achieved by the AdaBoost classifier that was using 50k feature vectors per training volume to learn, and the maximum depth of trees was set to 24. The segmentation outcome is better when the atlas is used, the difference is not only measurable by accuracy indicators, but also observable visually.

IV. CONCLUSIONS

This paper proposed to investigate the advantages brought by the inclusion of a multi-atlas into a spatial segmentation problem performed on multi-spectral MRI data. Experiments proved that the accuracy of pixelwise segmentation improves by 2.5-3.5%, depending on the chosen classifier method and parameter adjustment. In future work, we would also wish to study the classification problem within a game theoretic framework [33].

REFERENCES

[1] J. H. Zhou, Z. N. Yan, G. Lasio, J. Z. Huang, B. S. Zhang, N. Sharma, K. Prado and W. D'Souza, "Automated compromised right lung segmentation method using a robust atlas-based active volume model with sparse shape composition prior in CT ," *Comput. Med. Imag. Graph.*, vol. 46, pp. 47–55, 2015.

[2] K. Karasawa, M. Oda, T. Kitasaka, K. Misawa, M. Fujiwara, G. W. Chu, G. Y. Zheng, D. Rueckert and K. Mori, "Multi-atlas pancreas segmentation: Atlas selection based on vessel structure," *Med. Image Anal.*, vol. 39, pp. 18–28, 2017.

[3] H. J. Jia, Y. Xia, Y. Song, W. D. Cai, M. Fulham and D. D. Feng, "Atlas registration and ensemble deep convolutional neural network-based prostate segmentation using magnetic resonance imaging," *Neurocomput.*, vol. 280, pp. 1358–1369, 2018.

[4] H. Arabi and H. Zaidi, "Comparison of atlas-based techniques for whole-body bone segmentation," *Med. Image Anal.*, vol. 36, pp. 98–112, 2017.

[5] L. Shan, C. Zach, C. Charles and M. Niethammer, "Automatic atlas-based three-label cartilage segmentation from MR knee images," *Med. Image Anal.*, vol. 18, pp. 1233–1246, 2014.

[6] Y. Y. Yang, W. J. Jia and Y. N. Yang, "Multi-atlas segmentation and correction model with level set formulation for 3D brain MR images," *Patt. Recogn.*, vol. 90, pp. 450–463, 2019.

[7] S. X. Bao, C. Bermudez, Y. K. Huo, P. Parvathaneni, W. Rodriguez, S. M. Resnick, P. F. D'Haese, M. McHugo, S. Heckers, B. M. Dawant, I. Lyu and B. A. Landman, "Registration-based image enhancement improves multi-atlas segmentation of the thalamic nuclei and hippocampal subfields," *Magn. Res. Imag.*, vol. 59, pp. 143–152, 2019.

[8] J. Huo, J. Wu, J. W. Cao and G. H. Wang, "Supervoxel based method for multi-atlas segmentation of brain MR images," *NeuroImage*, vol. 175, pp. 201–214, 2018.

[9] T. Fülöp, Á. Győrfi, B. Surányi, L. Kovács, and L. Szilágyi, "Brain tumor segmentation from MRI data using ensemble learning and multi-atlas," *18th IEEE World Symposium on Applied Machine Intelligence and Informatics* (SAMI 2020, Herlany, Slovakia), pp. 111–116, 2020.

[10] Y. M. Niu, Q. Lan and X. C. Wang, "Structured graph regularized shape prior and cross-entropy induced active contour model for myocardium segmentation in CTA images," *Neurocomput.*, vol. 357, pp. 215–230, 2019.

[11] S. J. W. Kim, S. Seo, H. S. Kim, D. Y. Kim, K. W. Kang, J. J. Min and J. S. Lee, "Multi-atlas cardiac PET segmentation," *Phys. Med.*, vol. 58, pp. 32–39, 2019.

[12] B. Oliveira, S. Queirós, P. Morais, H. R. Torres, J. Gomes-Fonseca, J. C. Fonseca and J. Vilaça, "A novel multi-atlas strategy with dense deformation field reconstruction for abdominal and thoracic multi-organ segmentation from computed tomography," *Med. Image Anal.*, vol. 45, pp. 108–120, 2018.

[13] D. C. T. Nguyen, S. Benameur, M. Mignotte and F. Lavoie, "Super-pixel and multi-atlas based fusion entropic model for the segmentation of X-ray images," *Med. Image Anal.*, vol. 48, pp. 58–74, 2019.

[14] M. K. Sharma, M. Jas, V. Karale, A. Sadhu and S. Mukhopadhyay, "Mammogram segmentation using multi-atlas deformable registration," *Comput. Biol. Med*, vol. 110, pp. 244–253, 2019.

[15] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet and M. Bach Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images," *Comput. Meth. Prog. Bio.*, vol. 104, pp. e158–e177, 2011.

[16] L. Sun, L. Zhang and D. Q. Zhang, "Multi-atlas based methods in brain MR image segmentation," *Chin. Med. Sci. J.*, vol. 34, no. 2, pp. 110–119, 2019.

[17] L. Wang, D. Nie, G. N. Li, É. Puybareau, J. Dolz, Q. Zhang, *et al* "Benchmark on automatic 6-month-old infant brain segmentation algorithms: the iSeg-2017 challenge", *IEEE Trans. Med. Imag.* vol. 38, no. 9, pp. 2219–2230, 2019.

[18] U. Vovk, F. Pernuš and B. Likar, "A review of methods for correction of intensity inhomogeneity in MRI," *IEEE Trans. Med. Imag.*, vol. 26, pp. 405–421, 2007.

[19] L. Szilágyi, S. M. Szilágyi, and B. Benyó, "Efficient inhomogeneity compensation using fuzzy *c*-means clustering models", *Comput. Meth. Progr. Biomed*, vol. 108, no. 1, pp. 80–89, 2012.

[20] L. G. Nyúl, J. K. Udupa, X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imag.* vol. 19, no. 2, pp. 143–150, 2000.

[21] L. Breiman, "Random forests," *Mach. Learn.* vol. 45, pp. 5–32, 2001.

[22] L. Lefkovits, Sz. Lefkovits, and L. Szilágyi, "Brain tumor segmentation with optimized random forest," *2nd International Workshop on Brain Lesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* (BraTS MICCAI 2016, Athens), *LNCS*, vol. 10154, pp. 88–99, 2017.

[23] J. Lucansky, P. Pistek, and M. Maruniak, "The residual variable in decision diagrams," *Acta Polytech. Hung.* vol. 17, no. 5, pp. 189–205, 2020.

[24] Y. Freund and R.E. Schapire, "Experiments with a new boosting algorithm," *13th International Conference on Machine Learning*, pp. 148–157, 1996.

[25] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.

[26] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, 2016.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[28] E. Fix, J. Hodges, and U. S. of Aviation Medicine, *Discriminatory Analysis: Nonparametric Discrimination, Consistency Properties*, ser. Discriminatory Analysis: Nonparametric Discrimination, Consistency Properties. USAF School of Aviation Medicine, 1985, no. vol. 1-2.

[29] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[30] T. Hastie, S. Rosset, J. Zhu, and H. Zou, "Multi-class adaboost," *Statistics and Its Interface*, vol. 2, pp. 349–360, 2009.

[31] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors)," *The Annals of Statistics*, vol. 28, no. 2, pp. 337 – 407, 2000.

[32] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation," *IEEE Trans. Med. Imag.* vol. 23, no. 7, pp. 903–921, 2004.

[33] M.-A. Suciu and R. I. Lung, "Nash equilibrium as a solution in supervised classification," *Int'l Conf. on Parallel Problem Solving from Nature* (PPSN 2020), *LNCS*, vol. 12269, pp. 539–521, 2020.