



HAL
open science

Supervised machine learning based classification scheme to segment the brainstem on MRI in multicenter brain tumor treatment context

J. Dolz, Anne Laprie, Soléakhéna Ken, Henri-Arthur Leroy, Nicolas Reyns, Laurent Massoptier, Maximilien Vermandel

► To cite this version:

J. Dolz, Anne Laprie, Soléakhéna Ken, Henri-Arthur Leroy, Nicolas Reyns, et al.. Supervised machine learning based classification scheme to segment the brainstem on MRI in multicenter brain tumor treatment context. *International Journal of Computer Assisted Radiology and Surgery*, 2015, pp.1/16. hal-01181348

HAL Id: hal-01181348

<https://hal.science/hal-01181348v1>

Submitted on 30 Jul 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Supervised machine learning based classification scheme to segment the brainstem on MRI in multicenter brain tumor treatment context

Jose Dolz · Anne Laprie, MD, PhD · Soléakhéna Ken, PhD · Henri-Arthur Leroy MD, MSc · Nicolas Reyns MD, PhD · Laurent Massoptier · Maximilien Vermandel, PhD

Received: date / Accepted: date

Abstract *Purpose* To constrain the risk of severe toxicity in radiotherapy and radiosurgery, precise volume delineation of organs at risk (OARs) is required. This task is still manually performed, which is time-consuming and prone to observer variability. To address these issues, and as alternative to atlas-based segmentation methods, machine learning techniques, such as support vector machines (SVM), have been recently presented to segment subcortical structures on magnetic resonance images (MRI).

J. Dolz

AQUILAB. Biocentre A. Fleming, 250 rue Salvador Allende. 59120, Loos les Lille. France

Tel.: +33-369-61-51-51

Fax: +33-369-61-51-50

Univ. Lille, Inserm, CHU Lille, U1189 - ONCO-THAI - Image Assisted Laser Therapy for Oncology, F-59000 Lille, France E-mail: jose.dolz@aquilab.com

A. Laprie

Department of Radiation Oncology, Institut Claudius Regaud, Toulouse, France

S. Ken

Department of Radiation Oncology, Institut Claudius Regaud, Toulouse, France

H.-A. Leroy

Univ. Lille, Inserm, CHU Lille, U1189 - ONCO-THAI - Image Assisted Laser Therapy for Oncology, F-59000 Lille, France

Neurosurgery Department, University Hospital Lille, Lille, France

N. Reyns

Univ. Lille, Inserm, CHU Lille, U1189 - ONCO-THAI - Image Assisted Laser Therapy for Oncology, F-59000 Lille, France

Neurosurgery Department, University Hospital Lille, Lille, France

L. Massoptier

AQUILAB. Biocentre A. Fleming, 250 rue Salvador Allende. 59120, Loos les Lille. France

M. Vermandel

Univ. Lille, Inserm, CHU Lille, U1189 - ONCO-THAI - Image Assisted Laser Therapy for Oncology, F-59000 Lille, France

Methods SVM is proposed to segment the brainstem on MRI in multicenter brain cancer context. A dataset composed by 14 adult brain MRI scans is used to evaluate its performance. Additionally to spatial and probabilistic information, 5 different Image Intensity Values (IIVs) configurations are evaluated as features to train the SVM classifier. Segmentation accuracy is evaluated by computing the Dice Similarity Coefficient (DSC), Absolute Volumes Difference (AVD) and percentage Volume Difference (pVD) between automatic and manual contours.

Results Mean DSC for all proposed IIVs configurations ranged from 0.89 to 0.90. Mean AVD values were below 1.5 cm^3 , where the value for best performing IIVs configuration was 0.85 cm^3 , representing an absolute mean difference of 3.99 % with respect to the manual segmented volumes.

Conclusion Results suggest consistent volume estimation and high spatial similarity with respect to expert delineations. The proposed approach outperformed presented methods to segment the brainstem, not only in volume similarity metrics, but also in segmentation time. Preliminary results showed that the approach might be promising for adoption in clinical use.

Keywords Support Vector Machines · Machine learning · Supervised learning · MRI segmentation · Radiotherapy · Brain cancer

1 Introduction

Brain tumours kill more children and adults under the age of 40 than any other cancer and five-year survival remains still low [1]. Among available techniques to treat brain tumors, radiotherapy and radio surgery have become often the selected treatment, especially when others techniques such as surgery or chemotherapy might not be applicable. During the radiation treatment planning (RTP), the tumor to irradiate, i.e. clinical target volume (CTV), as well as healthy structures to be spared, i.e. the organs at risk (OARs), must be precisely delineated. Because of the high doses used to irradiate the CTV, the risk of severe toxicity of the OARs must be constrained. Therefore, these segmentations are crucial inputs for the RTP, in order to compute the parameters for the accelerators, and to verify the dose constraints. The highly precise irradiation of targets within the brain is planned using information from medical images that are obtained via several image modalities. Among these modalities, and particularly in oncology, magnetic resonance images (MRI) is the preferred modality to contour OARs. Nowadays in clinical practice, OARs delineation on MRI is performed manually by experts, or with very few machine assistance. Consequently, this process is highly time-consuming for radiation oncologists, especially for large datasets, and remains operator dependent [2].

In recent years, numerous methods have been proposed to overcome limitations of manual contouring in RTP [3]. Because of the similar textures and weak boundaries of most of the subcortical brain structures, such as the brainstem (BS), proposed segmentation approaches usually incorporate some prior knowledge. This information is often given as shape or location of the

structure of interest. Among methods using prior information, atlas based have been extensively used [4–6]. In atlas based approaches, anatomical information is exploited by means of atlases to be matched to the patient under examination. To compute such transformation, deformable registration is often used [7]. In [4], an atlas-based approach called Classifier Fusion and Labelling (CFL), was evaluated and compared with other three approaches to segment subcortical brain structures: two statistical based models - Profile Active Appearance Models (PAM) and Bayesian Appearance Models (BAM)- and an Expectation-Maximisation-based approach (EMS). Regarding the BS segmentation, mean Dice Similarity Coefficient (DSC) values for the proposed methods ranged from 0.83 to 0.94, with percentage mean Absolute Volume Difference (AVD) from 21.10% to 3.98%, respectively.

Specifically for the BS segmentation, some atlas-based methods have been evaluated in RTP context [5, 6]. In [5], the BS was automatically segmented on 6 patients by using an atlas base approach. In addition to automatically generated contours, brainstem on these 6 patients was manually segmented by 7 experts, and all segmentations were compared. A mean sensitivity of 0.76 for the automatic proposed method was reported, which lay in the means range for the experts. In the work of [6], an automatic segmentation software (ABAS) was evaluated when segmenting OARs -including the BS - in brain cancer. DSC values for the BS ranged from 0.8 to 0.88 and a mean volume underestimation equal to -14.8% was reported. However, processing time reported in previously presented works to segment the BS was always above several minutes, if not more.

Additionally to atlas-based methods, other approaches such as statistical models [4, 8] have also been employed to segment the BS. In [8], a 2D automatic scheme based on Scale-based Fuzzy Connectedness algorithm and Active Contour models was proposed to segment the BS. In their work, a set of 66 MRT2 images was used, where 33 images were used as training samples and the other 33 formed the testing set. The presented results showed a mean DSC of 0.86 over 5 selected MRI scans.

However, with the presence of tumors, all these methods might fail in the classification, since the anatomical variability introduced by the tumor might not have been captured during the atlas creation or training phase.

In the presented work, an automatic approach based on a supervised machine learning technique, SVM, that outperforms all the presented approaches up to the date to segment the brainstem is proposed. First, it increases the similarity with respect to manual expert segmentations, while also provides a substantial improvement on the processing time. Second, it reduces volume differences between automatic segmentations and expert delineations. Third, no computational expensive registration processes are required. And fourth, a limited training set is sufficient to achieve promising results.

2 Methods

In this section all steps included in the proposed approach are explained. First, the initial alignment performed to the MRIs to be segmented is detailed. In the following subsection, features often used in the literature to train some machine learning classifiers are presented. After this, support vector machine (SVM) technique is formally introduced. Finally, the set-up of the proposed experiment is described.

2.1 Pre-processing

It is a common practice in brain segmentation approaches to apply registration techniques to the MRI images to make them as similar as possible to a common MRI template. Some approaches require a rigid registration step to align the images [9, 10]. However, in the proposed approach, and as in [11] and [12], MRI T1 images were spatially aligned such that the anterior commissure and posterior commissure (AC-PC) line was horizontally oriented in the sagittal plane, and the inter hemispheric fissure was aligned on the two other axes.

2.2 Features extraction

The most influencing factor in realizing a classifier with high generalization ability is the set of features used. A poor selection of features may lead to unsatisfactory results. Unlike in most segmentation methods, image intensity information solely is not good enough to distinguish different subcortical structures since most of these structures share similar intensity patterns on MRI. To address such problem, in learning based segmentation methods, more discriminative features are often extracted from MRI [9-12]. The method presented in [11] used image intensity values (IIVs) of the neighborhood of the voxel to be considered. To provide a computationally efficient survey of the neighboring voxels, signal intensity was sampled as far away as three voxels. A spherical region was additionally used for 3D applications. Only diagonal and orthogonal voxels to the location of the input voxel were used, which provided information from 42 surrounding voxels. Three additional nodes were used to specify the voxel location. The final input feature was the frequency with which the location was found in the search space for the training set. Since image features did not contain a shape representation, however, a large training set was needed.

Recently, Powell et al. [12] have further developed their previous algorithm [11] using 9 IIVs along the largest gradient, a probabilistic map, and the IIVs along each of the three orthogonal axes, leading to an input vector composed by 25 features. In [10], a slightly modified input vector was adapted to increase the performance of an artificial neural network. In their work, a modified spatial location for the spherical coordinates and a neighborhood

connection based on gradient descent were used. While the former aided to reflect symmetry of brain, the latter was used for directional consistency. To improve the differentiation between the brain regions compared to the IIVs, Geometric Moment Invariants (GMIs) were used in [9]. The use of GMIs as feature is interesting because they represent features that are theoretically not sensitive to particular deformation such as rotation, translation or scaling. Additionally to GMI, neighbouring voxel IIVs along the main axis (3 for each direction), as well as the voxel coordinates, completed the input vector.

Several combinations of IIVs were used as subset of the input features vector in the presented work (Table 1). Cubic traditional patches of size 3 and 5 were extracted around the voxel under examination as IIVs for the input features vector in configurations A and B, respectively. Similarly to [9], a cross was used in configuration C, which was composed by the 12 IIVs along each of the orthogonal axis (2 in each direction). Neighborhood information based on gradient descent, as in [10] was used in configuration E, where 9 IIVs along the largest gradient including the voxel to be analyzed composed the IIVs subset. And as in [12], configurations C and E formed the IIVs array in configuration D.

IIVs configuration	3x3x3 Patch	5x5x5 Patch	Cross orthogonal to axes	Conf. C + Conf. E	9 IIVs along max. Gradient
Conf. Name	A	B	C	D	E

Table 1: Names of the different IIVs configurations used in the input features vector used in the SVM classifier.

Additionally to the IIVs, a spatial probabilistic distribution map (SPDM) of the brainstem is used as input feature. SPDM represents the probability of a voxel to be classified as brainstem, according to the training data. The sum of all the brainstem masks contained in the training data set is used to create the SPDM. The resulted map is then smoothed by using a Gaussian filter with a kernel size of 3x3x3 (Fig.1). In addition, to reduce the number of input samples that contains consistent information, the voxel space is pruned. A dilated version of the SPDM is binarized (Fig.1) and its result is used to prune the voxels. To obtain this dilated mask, the SPDM is first binarized by setting its values greater than 0.005 to 1, and the others to 0. Then, a dilation operation with a square kernel type of size 3x3x3 is applied over the binary image, leading to the dilated binary version of the SPDM. Only those voxels that belong to the inner part of the dilated SPDM are kept to feed the prediction algorithm. Last, voxel spherical coordinates completed the input vector.

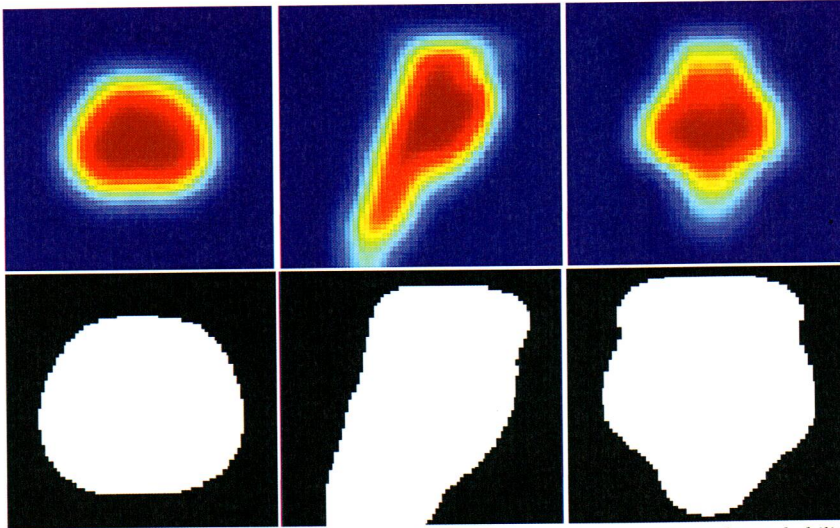


Fig. 1: Brainstem probability map (top) and brainstem mask extracted from the probability map (bottom). From left to right axial, sagittal and coronal views are shown.

2.3 Support vector machines

Support vector machines and their variants and extensions, often called kernel-based methods, have been studied extensively and applied to various pattern classification and function approximation problems. Basically, the main idea behind SVM is to find the largest margin hyperplane that separates two classes. The minimal distance from the separating hyperplane to the closest training example is called margin. Thus, the optimal hyperplane is the one providing the maximal margin, which represents the largest separation between the classes. This will be the line such that the distances from the closest point in each of the two groups will be farthest away. The training samples that lie on the margin are referred as support vectors, and conceptually are the most difficult data points to classify. Therefore, support vectors define the location of the separating hyperplane, being located at the boundary of their respective classes. By employing kernel transformations to map the objects from their original space into a higher dimensional *feature space* [13], SVM can separate objects which are not linearly separable (Figure 2).

Their good generalization ability and their capability to successfully classify non-linearly separable data have led to a growing interest on them for classification problems. By combining both local and global information as input for the classifier, SVM can take decisions at voxel level, even for unseen situations. This makes machine learning techniques in general, and SVM in particular, more suitable segmentation methods for situations where the presence of tumors can modify the anatomy of the patient.

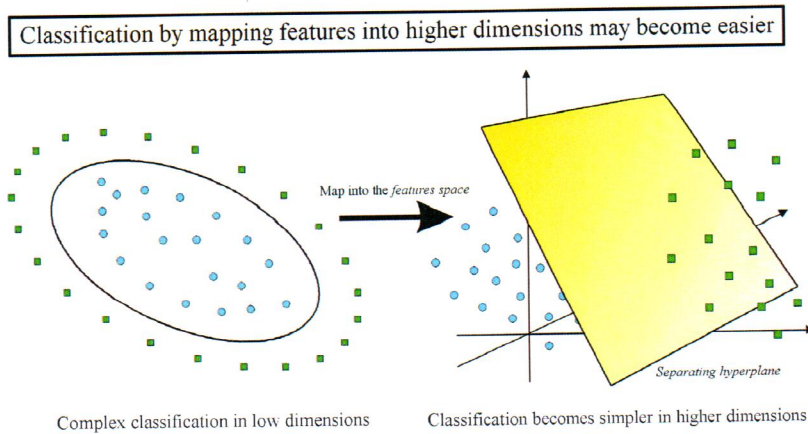


Fig. 2: SVM and hyperplane. Mapping features into a higher dimensionality may make the classification simpler.

2.4 Kernel selection and parameters tuning

As explained in previous section, to map input features into higher dimensionality spaces several kernels can be used, according to the nature of the data. Radial Basis Function (RBF) kernels are one of the most used kernels to separate data in SVM classifiers in complex classification environments. Some previous works have found that RBF kernel generally provides better classification accuracy than many other kernel functions [15]. This kernel nonlinearly maps samples into a higher dimensional space. That means that RBF kernel can handle the cases when the relation between class labels and attributes is nonlinear. Second reason to use this kernel is the number of hyperparameters which influences the complexity of model selection, which is lower than in other non-linear kernels, such as the polynomial kernel.

There are two parameters that can be tuned in the RBF kernel and which depend on the input data: C and γ . While C controls the cost of missclassification on the training data, γ is the parameter of the kernel to handle non-linear classification. Different values for both parameters might lead to soft or hard margins, for the case of C , and the loss of its non-linear power or increase of sensitivity to noise, in the case of γ . Therefore, some kind of model selection, i.e. parameter search, must be done. The goal of this search is to identify good C and γ values so that the classifier can accurately predict unknown data, i.e. testing data.

Best values for parameters C and γ were selected in this work by performing a cross-validation and grid search in the training set of configuration A with $v = 10$ subfolders (Fig. 3). Since performing a fully grid search may become time consuming, a coarse grid search was initially conducted. After identifying the best region on the grid, a finer grid search on that region was

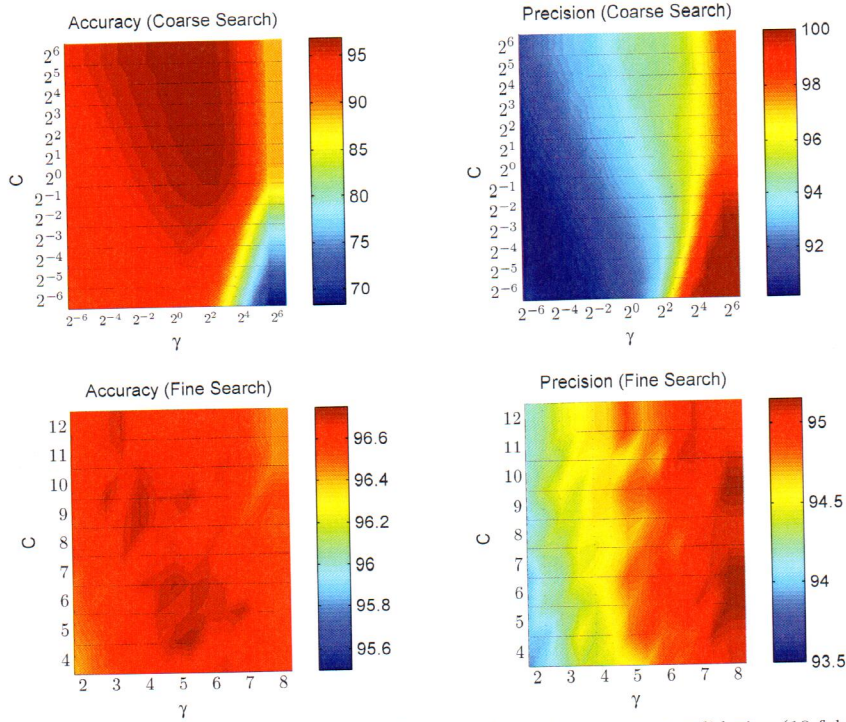


Fig. 3: Optimal parameter selection. The heat map is used for the cross-validation (10-fold) coarse (*top*) and fine (*bottom*) grid search to find the hyper-parameter C and γ in the RBF kernel of the proposed approach. An initial estimation of the search area is found in the hottest spot of the coarse search. This spot is then used to performed a fine search of the best combination of C and γ . A balance between accuracy and precision is employed for the decision of the parameters.

performed. From this search was found that best values for C and γ were approximately 6 and 5.5, respectively, with an accuracy close to 97% and a precision nearly of 95%. These values for C and γ were kept for the training and classification in all the IIVs configurations.

2.5 Materials

The presented approach was evaluated on a dataset containing several data subsets from different institutions and presenting several pathologies. MR 1.5T T1 with and without gadolinium injection brain images were used in this experiment (See Table 2). Pathologies visible in these data sets included trigeminal neuralgia, metastases, brainstem cavernoma and glioblastoma multiforme. According to the brain tumor and radiotherapy center, either radiotherapy or radiosurgery were the techniques used to treat these patients. Additionally,

some of the subjects included in these data sets presented some tumours inside or very close to the brainstem, causing deformations on it.

MRI System	TE(ms)	TR(ms)	Matrix size	Num. Slices	Voxel Size (mm ³)
Philips Achieva 1.5T Scanner	4.602	25	256x256	180	0.9x0.9x1.0
Siemens Avanto MRI Scanner	17	532	256x224	160	1x1x1

Table 2: Acquisition parameters on the 2 MRI devices.

The SVM classifier was trained by using 3 subjects randomly selected from the whole dataset and tested on an independent set of the remaining 14 subjects. From the full set of sampled vectors in the training phase, 50.000 samples were randomly selected to train the classifier. Because of SVM algorithms are not scale invariant, once all samples were added to the input features vectors, and before to start the training, all values were normalized between -1 and 1. Training times for each configuration are shown in Table 3.

Configurations	A	B	C	D	E
Training Time (sec)	68.96	202.62	45.83	59.05	44.03

Table 3: Training times for each of the 5 IIVs configurations used in the proposed approach to train SVM classifier.

MATLAB was the platform used for this experiment. The publicly available library libsvm [16] was used for the classification task. A workstation with 8GB of RAM memory and Intel Xeon processor at 3.06 HGz was employed.

2.6 Data analysis

Performance of the proposed approach was quantitatively assessed by comparing automatic and manual reference contours in terms of shape and volume. DSC [17] was computed to evaluate shapes similarities between automatic and manual BS segmentations. Regarding the volume, absolute volume difference (AVD) between automatic and manual segmented volumes, as well as percentage volume difference (pVD) (Eq. 1) were calculated.

$$pVD(\%) = \frac{V_{auto} - V_{ref}}{V_{ref}} * 100 \quad (1)$$

Both AVD and pVD measure volume similarities. Nevertheless, whilst AVD represents the absolute value of volume differences, in cm³, pVD is the real signed value of such differences in terms of percentage. While the former is important to evaluate the total difference on volume, the later will tell us whether there is underestimation of the volume and how important the

volume difference is. Unlike volume differences, DSC or overlap ratio metrics are sensitive to misplacement of the segmentation label, but they are relatively insensitive to volumetric under- and overestimations.

3 Results

SVM was successfully applied to the 14 independent patients using the 5 different IIVs configurations in the input features vector. A one-way within-subjects analysis of variance (ANOVA) test was performed separately on the DSC and pVD values using the 5 configurations (See Fig. 4). The computed p -values were 0.8312 and 0.583 for DSC and pVD, respectively. This means there were no significant differences between groups in terms of shape and volume similarity. It can be observed that DSC values ranged from 0.82 (conf.A) to 0.93 (conf.D), with means values ranging from 0.89 to 0.91 for all the configurations.

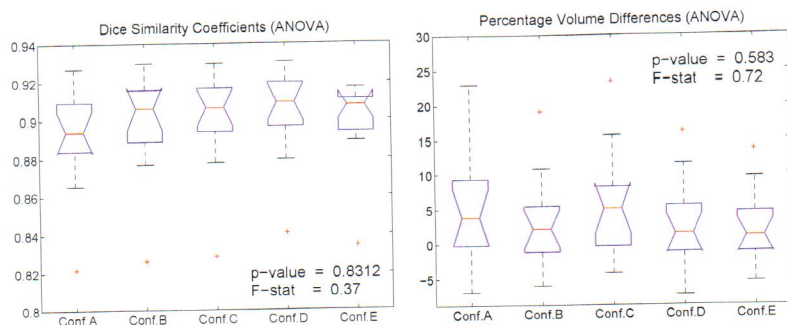


Fig. 4: Statistical analysis of Dice similarity coefficients and percentage volume differences (pVD) for the automatic proposed approach for each of the five IIV configurations.

Concerning volumes differences, the 5 IIVs configurations reported mean AVD values comprised between nearly 1.500 cm^3 and 0.850 cm^3 , being configurations D and E the ones that reported better results, with mean AVD of 0.999 cm^3 ($0.075\text{-}3.194 \text{ cm}^3$) and 0.850 cm^3 ($0.155\text{-}2.679 \text{ cm}^3$), respectively (Figure 5). In terms of percentage of volume differences between automatic and manual segmentations, these volume difference values can be translated to pVD values ranging from -7.43% to 16.19% for conf.D, and from -5.52% to 13.57% for conf.E (Figure 4, right and Table 4).

Table 4 shows the pVD for all the patients when using the 5 IIVs configuration. This makes possible to compare how different IIVs configuration impacted on the pVD for a given patient and evaluate whether the automatic approach underestimated or overestimated the manually segmented volume. Configurations B, D and E showed the highest volume similarities for the majority of patients. However, although mean AVD and pVD values of configurations B and D were closer to those reported by E, standard deviation

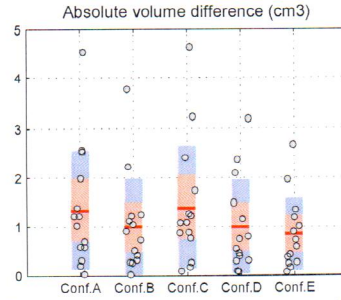


Fig. 5: Absolute volume differences for the automatic proposed approach for each of the five IIV configurations.

values were slightly higher for the former ones (Fig. 4 and 5). Volume difference in *conf.E* represented in most of the cases an absolute pVD ranging from 0 to nearly 5% with respect to the manual volume segmentation. Even in its worst case (Patient002), *conf.E* only over-estimated the brainstem volume 13.57% with respect to the reference volume. These results were very similar to the results observed in Configuration *D*.

IIV Conf.	pVD (%)				
	A	B	C	D	E
Patient001	-6.97	-6.08	-4.28	-7.43	-1.17
Patient002	22.95	19.04	23.45	16.19	13.57
Patient003	12.49	10.82	15.62	11.53	9.50
Patient004	9.37	5.30	8.21	5.49	2.73
Patient005	5.69	1.44	5.05	1.39	0.30
Patient006	5.29	5.44	5.68	2.78	3.75
Patient007	2.28	-1.61	-3.49	-1.14	-3.98
Patient008	2.87	1.08	3.16	0.31	-0.64
Patient009	2.91	4.62	6.33	4.05	4.55
Patient010	-1.22	-2.02	-0.66	-2.99	-5.52
Patient011	-0.12	0.08	1.11	0.37	-1.59
Patient012	12.05	5.96	11.44	9.91	5.65
Patient013	4.86	2.87	5.09	1.55	-1.03
Patient014	-0.86	-1.12	-0.38	-1.97	1.86

Table 4: Percentage volume differences (pVD) between results obtained with the proposed approach for each of the configurations and manual delineations.

In relation with segmentation time, configurations *C* and *E* reported the fastest times, with 40.45 and 36.6 seconds as average, respectively. Processing times for configurations *A* and *D*, although slower, were still under one minute, and configuration *B* was the slowest one, with a processing time close to 5 minutes (See Fig. 6).

Figure 7 shows an example of the automatic segmentation provided by the proposed approach (in red) compared to the manual reference standard (in green).

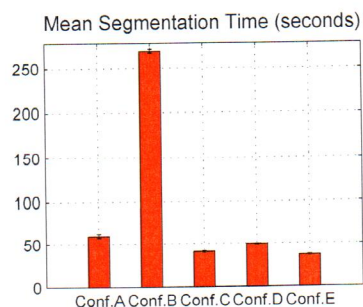


Fig. 6: Mean segmentation times for the automatic proposed approach for each of the five IIV configurations.

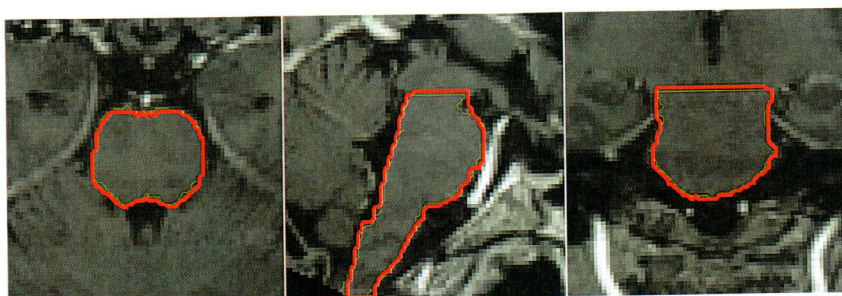


Fig. 7: Brainstem segmentation example. Green represents manual contouring while red bold is the segmentation provided by the proposed approach.

4 Discussion

A supervised machine learning based approach to automatically segment the brainstem on MRI in multicenter brain tumor treatment context has been proposed. The main contribution of this paper is to apply a machine learning method -SVM- to a problem in which it has not been applied before, i.e. brainstem segmentation. To do so, we investigate how different set of features impact on the segmentation performance for this specific problem. Pathologies visible in the analyzed data sets included trigeminal neuralgia, metastases, brainstem cavernoma and glioblastoma multiforme, which were treated either by radiotherapy or radiosurgery. Additionally, some of the subjects included in the study presented some tumours inside or very close to the brainstem, causing deformations on it, which makes the segmentation sometimes more challenging.

Among the features used to train the classifier, 5 different configurations of image intensity values have been evaluated and compared in terms of shape and volume similarity, as well as segmentation processing time. Although the results from the 5 IIVs configurations reported strong agreement compared to the reference segmentations, conf.A provided the worst performance in terms of similarity. It used a 3x3x3 texture patch in the input vector, representing the

weakest neighborhood scope among the five proposed configurations, since only 1 voxel beyond of the voxel under examination was reached. By incorporating a larger patch of size $5 \times 5 \times 5$ to capture images intensity values, conf.*B* improved DSC values, as well as AVD and pVD. However, due to the patch size, 125 elements were added to the input features vector, instead of 27 in conf.*A*, leading to a considerably slower segmentation time.

To reduce the number of elements in the input vector, whilst preserving DSC and AVD values, a cross orthogonal to the axis was presented in conf.*C*, taking ± 2 voxels in each direction (12 IIVs in total). DSC values remained similar to conf.*B* and processing time was substantially reduced. However, AVD and pVD were higher than those provided by conf.*B*. As in [12], 9 IIVs along the maximum Gradient direction were added to the existing IIVs in conf.*C*, which represented the conf.*D*. This resulted in an input vector with 21 IIVs, in addition to other features. Even though processing time augmented around 20% respect to conf.*C*, AVD and pVD values notably decreased, while DSC values did not practically change.

Since the brainstem is a large homogeneous structure, a large scope like in conf.*D* often gives more information than a short neighborhood view, like in conf.*C*. Based on this, a novel texture configuration was proposed in conf.*E*, which only used the 9 IIVs along the maximum Gradient direction, including the voxel under examination, in the input vector. Quantitative results for this IIVs configuration showed that it reported the best trade-off between segmentation similarity and processing time among the five analyzed IIVs configurations. While it reported one of the highest mean DSC values (0.9007), together with configurations *C* (0.9006) and *D* (0.9033), it also represented the fastest segmentation process. Mean segmentation time in conf.*E* was measured to be 36.60 seconds, which represented an improvement of 10% and 25% with respect to configurations *C* and *D*, respectively. From the statistical analysis (Figure 4) it is observed that differences between the results of the 5 configurations were not significantly important in terms of shape and volume similarity. However, segmentation time (Figure 6) was substantially reduced in the last three configurations -i.e., *C*, *D* and *E* -, being the configuration *E* the fastest one. Therefore, it is reasonable to select configuration *E* as the IIVs configuration to train the classifier in the presented approach when segmenting the brainstem on MRI, particularly in relation to improving segmentation times.

Reference	Method	DSC	pVD (%)	Segmentation Time
Babalola et al., (2009)	Atlas-Based	0.94	3.98	120-180 min. (Set of brain structures)
	Statistical-Based (PAM)	0.88	6.80	1 min. + 20 min. ^a
	Statistical-Based (BAM)	0.89	7.80	5 min. + 3min. ^a
	Expectation-Minimization	0.83	21.10	30 min. (Set of brain structures)
Bondiau et al., (2005)	Atlas-Based	-	-13.11	20 min. (7 OARs and 7 normal structures)
Isambert et al., (2008)	Atlas-Based	0.85	-14.8	7-8 min. (6 OARs)
Proposed approach	SVM	0.90	3.99	36.6 seconds

Table 5: Table that summarizes results of previous works which attempted to segment the brainstem on MRI images. DSC and pVD are given as mean values.

Comparing results presented in Table 5 to the ones provided by proposed approach, it can be concluded that SVM was ranked at high to segment the brainstem with relatively fast processing time, reducing substantially the time required in comparison to the other presented methods. Additionally, given the promising results, and following the work of [5] and [6], a larger evaluation involving several clinicians would validate its use in clinical routine to segment the brainstem in brain cancer context.

The main direction for future research is to examine the contribution of other image properties as features during the training and classification. In addition, we aim to extend the presented approach to make it work with a larger set of subcortical structures, which are involved in external radiotherapy and radio-surgery. We are aware that the surrogate ground truth used in the presented work may not be completely satisfactory, since it is based in single manual delineations. Differences in data acquisition often compromise fair comparisons with other works. A solution to perform such comparison would be to have access either to all the code for other methods or to a public database where we may use the same patients than others do and then compare our results. Nevertheless, both options are not straightforward and they are still a topic under discussion. For our current research, which deals with the segmentation of other organs at risk in brain cancer - such as the optic nerves or the optic chiasm -, because of the lack of public datasets with a ground truth, we have no other choice than using manual expert segmentations as gold standards in all our experiments. For this reason, multiple observers delineations will be used during the learning and testing process. Involving several experts in the delineation process will allow to compare automatic segmentations with those that would be clinically used in RTP, helping to validate the presented approach in clinical brain cancer context.

5 Acknowledgment

This project has received funding from the European Unions Seventh Framework Programme for research, technological development and demonstration under grant agreement no PITN-GA-2011-290148.

6 Statements

Jose Dolz, Solakhna Ken, Henri-Arthur Leroy, Nicolas Reyns, Anne Laprie, Laurent Massoptier and Maximilen Vermandel declare that they have no conflict of interest.

^a These two approaches required registration steps which took 20 minutes in the first case, and around 3 minutes for the second method.

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2008 (5).

For this type of study formal consent is not required.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. (2013) GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer; . URL <http://globocan.iarc.fr> (accessed December 16, 2014)
2. Cattaneo GM, Reni M, Rizzo G, Castellone P, Ceresoli GL, Cozzarini C, Ferreri AJ, Passoni P, Calandrino R. (2005) Target delineation in post-operative radiotherapy of brain gliomas: interobserver variability and impact of image registration of MR (pre-operative) images on treatment planning CT scans. *Radiotherapy and oncology*, 75(2): 217–223.
3. Dolz J, Massotier L, Vermandel M. (2015) Segmentation algorithms of subcortical brain structures on MRI for radiotherapy and radiosurgery: a survey. *IRBM*, doi:10.1016/j.irbm.2015.06.001.
4. Babalola KO, and Patenaude B, Aljabar P, Schnabel J, Kennedy D, Crum W, Smith S, Cootes T, Jenkinson M, Rueckert D. (2009) An evaluation of four automatic methods of segmenting the subcortical structures in the brain. *Neuroimage* 47(4):1435–1447.
5. Bondiau PY, Malandain G, Chanalet S, Marcy PY, Habrand JL, Fauchon F, Paquis P, Courdi A, Commowick O, Rutten I and others. (2005) Atlas-based automatic segmentation of MR images: validation study on the brainstem in radiotherapy context. *International Journal of Radiation Oncology Biology Physics* 61: 289–298.
6. Isambert A, Dhermain F, Bidault F, Commowick O, Bondiau PY, Malandain G, Lefkopoulos D. (2008). Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context. *Radiotherapy and oncology* 87: 93–99.
7. Toga AW, Thompson PM. (2001) The role of image registration in brain mapping. *Image and vision computing* 19: 3–24.
8. Lee JD, Tseng YX, Liu LC, Huang CH. (2007) A 2-D Automatic Segmentation Scheme for Brainstem and Cerebellum Regions in Brain MR Imaging. *Conf. Proc. on Fuzzy Systems and Knowledge Discovery* 4: 270–274.
9. Moghaddam MJ, Soltanian-Zadeh H. (2009) Automatic segmentation of brain structures using geometric moment invariants and artificial neural networks. *Information Processing in Medical Imaging*: 326–337.
10. Kim EY, Johnson H. (2010) Multi-structure segmentation of multi-modal brain images using artificial neural networks. *SPIE Medical Imaging*: 76234B–76234B.
11. Magnotta VA, Heckel D, Andreasen NC, Cizadlo T, Corson PW, Ehrhardt JC, Yuh W. (1999) Measurement of Brain Structures with Artificial Neural Networks: Two- and Three-dimensional Applications I. *Radiology* 211(3): 781–790.
12. Powell S, Magnotta VA, Johnson H, Jammalamadaka VK, Pierson R, Andreasen NC. (2008) Registration and machine learning-based automated segmentation of subcortical and cerebellar brain structures. *Neuroimage* 39(1): 238–247.
13. Burges C. (1988) A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery* 2(2): 121–167.
14. Abe S. (2010) Support vector machines for pattern classification. Springer.
15. Abdi MJ, Hosseini SM, Rezghi M. (2012) A novel weighted support vector machine based on particle swarm optimization for gene selection and tumor classification. *Computational and mathematical methods in medicine* 2012.
16. Chang CC, Lin CJ. (2011) LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2: 27:1–27:27.

17. Dice LR. (1945) Measures of the amount of ecologic association between species. *Ecology* 26(3): 297-302.