Radiotherapy and Oncology 87 (2008) 93—99
www.thegreenjournal.com

CNS treatment planning

Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context

Aurélie Isambert[a,*], Frédéric Dhermain[b], François Bidault[c], Olivier Commowick[d,e], Pierre-Yves Bondiau[f], Grégoire Malandain[e], Dimitri Lefkopoulos[a]

[a]*Service de Physique,* [b]*Département de Radiothérapie, and* [c]*Département d'Imagerie Médicale, Institut Gustave Roussy, Villejuif, France,* [d]*DOSIsoft S.A., Cachan, France,* [e]*Institut National de Recherche en Informatique et Automatique, Sophia-Antipolis, France,* [f]*Département de Radiothérapie, Centre Antoine Lacassagne, Nice, France*

Abstract

***Background and purpose*: Conformal radiation therapy techniques require the delineation of volumes of interest, a time-consuming and operator-dependent task. In this work, we aimed to evaluate the potential interest of an atlas-based automatic segmentation software (ABAS) of brain organs at risk (OAR), when used under our clinical conditions.**

***Materials and methods*: Automatic and manual segmentations of the eyes, optic nerves, optic chiasm, pituitary gland, brain stem and cerebellum of 11 patients on T1-weighted magnetic resonance, 3-mm thick slice images were compared using the Dice similarity coefficient (DSC). The sensitivity and specificity of the ABAS were also computed and analysed from a radiotherapy point of view by splitting the ROC (Receiver Operating Characteristic) space into four sub-regions.**

***Results*: Automatic segmentation of OAR was achieved in 7—8 min. Excellent agreement was obtained between automatic and manual delineations for organs exceeding 7 cm^3: the DSC was greater than 0.8. For smaller structures, the DSC was lower than 0.41.**

***Conclusions*: These tests demonstrated that this ABAS is a robust and reliable tool for automatic delineation of large structures under clinical conditions in our daily practice, even though the small structures must continue to be delineated manually by an expert.**

Keywords: Automatic segmentation; Quantitative study; Radiation therapy; Pathological brain; Organs at risk

During the last 10 years, an increasing number of patients have been treated with three-dimensional conformal radiation therapy (3D CRT) with ever more sophisticated techniques such as intensity modulated radiation therapy (IMRT) or proton therapy.

These techniques imply steeper dose gradients. Accurate delineation of several volumes of interest such as the target volume and organs at risk (OAR) is therefore mandatory. This step in radiotherapy treatment planning is highly time-consuming for radiation oncologists and remains operator dependent [5,12,13,20,21]. The use of automatic segmentation could be a solution to such drawbacks.

Several teams have developed tools for automatic segmentation of structures of interest in the brain, such as white matter hyperintensities [1], white and gray matter, cerebrospinal fluid [18], the tumour target volume [9,15], and brain structures including ventricles, thalamus, central nuclei [9,17], and the putamen and cortex for neurosurgical purposes [11]. To our knowledge, only one work clinically evaluated the results obtained with such a tool on human brain images by comparing manual and automatic delineations of the brain stem in 6 patients [4]. Very recently, one team presented the evaluation of a software for automatic delineation of structures of interest of head-and-neck images for 7 patients [23].

The French company, DOSIsoft, commercialises a treatment planning system (TPS) designated ISOgray™. This system includes an atlas-based automatic segmentation (ABAS) software for the delineation of organs at risk (OAR) in the brain. The performance of this software can be evaluated at two levels, either in optimal conditions of use (slice thickness, image quality, etc.) or under the clinical conditions of a given institution that could be different. We aimed to evaluate the possible interest of version 3.1 of this tool when used in our daily practice which differs from the optimal conditions specified by the developers of the ABAS. After reviewing the principle of the ABAS, we will present the pre-clinical retrospective study that was performed in 11 patients by quantitatively

0167-8140/$ - see front matter doi:10.1016/j.radonc.2007.11.030

comparing manual and automatic contours in terms of volume, position and shape.

Materials and methods

Principle of the ABAS

The general principle of the ABAS was described by Bondiau et al. [4]. It is composed of an artificial MRI of the brain [6,7] (http://www.bic.mni.mcgill.ca/brainweb) on which each cerebral structure of interest, from a radiotherapy point of view, was manually delineated by an expert according to the Talairach atlas [22]. Six structures can be automatically segmented: the eyes, optic nerves, optic chiasm, pituitary gland, brain stem, and cerebellum.

The steps of the automatic segmentation procedure are as follows:

Step A: an affine transformation (which includes rotation, translation, shearing and scaling) is used to roughly align the artificial MRI with the patient MRI that is to be segmented into anatomical structures. To that end, the so-called ''Block-Matching'' algorithm, that locally optimises a correlation coefficient, was used [16]. This yields a transformed artificial MRI (t-MRI).

Step B: a second transformation is done to match the t-MRI and the patient MRI.

At this point, the transformation is ''multi-affine'' [8]. At each iteration, the algorithm searches local affine transformations, with the ''Block-Matching'' algorithm, for some structures of interest of the brain, typically the organs at risk that are delineated in the artificial MRI. Then, the fast polyaffine framework [3] allows compounding them to yield a global invertible transformation. The algorithm iterates until stability is reached.

Step C: the transformations applied in step A and step B are applied to the atlas to obtain the segmented cerebral structures on the patient MRI.

The algorithm was developed for T1-weighted MR images without injection of gadolinium contrast agent and the developers recommend using 2-mm thick or smaller slices.

Clinical evaluation

Data acquisition

A retrospective study of a series of 11 adult patients who were treated with conformal radiation therapy in our institution was conducted using the ABAS. These patients had primary brain tumours including high-grade gliomas, unfavourable low-grade gliomas or meningiomas (that were either biopsied, partially or ''completely'' resected). The median age was 47 years old.

Two specialists, a radiation-oncologist (FD), and a neuroradiologist (FB), together manually delineated the six previously described structures of the 11 patients. To help the experts delineate the tumour and OAR, before treatment planning for each patient, a CT/MR image registration (rigid transformation) was performed on previously tested software integrated into the ISOgray™ database [14].

The CT images were acquired on a LightSpeed 16 scanner (General Electric Healthcare), with the patient's head immobilized in a thermoplastic mask. The MR images were acquired on a GE Signa Excite 1.5 T scanner (T1-weighted axial images, sequence AX 3D SPGR with gadolinium contrast injection) with a head coil. The two sets of images (MR and CT) were acquired for each patient on the same day in our department using 3-mm thick slices.

Data analysis

The performance of the automatic segmentation software was assessed by quantitatively comparing manual and automatically generated contours in terms of volume, position and shape. A sensitivity and specificity study was also conducted. Manual segmentation was used as the reference segmentation.

Regarding the volume, the difference between the automatic volume (V_{auto}) and the manual volume, or reference volume (V_{ref}), was calculated for each structure, as follows:

$$\Delta V(\%) = \frac{|V_{auto} - V_{ref}|}{V_{ref}} * 100. \quad (1)$$

Regarding the position, the three coordinates (x, y, z) of the centres of gravity of the automatic and reference structures were compared: $\Delta x = |x_{auto} - x_{ref}|$, $\Delta y = |y_{auto} - y_{ref}|$, $\Delta z = |z_{auto} - z_{ref}|$. To quantitatively compare automatic and reference structures in terms of shape, we computed the Dice similarity coefficient (DSC) [10] defined as:

$$DSC = \frac{2TP}{(2TP + FN + FP)}, \quad (2)$$

where TP (true positive) is the number of voxels common to both automatic and manual segmentations, FN (false negative) is the number of reference volume voxels that are not covered by the automatic segmentation and FP (false positive) is the number of voxels over-segmented by automatic segmentation compared to manual segmentation (Fig. 1). DSC values ranged from 0 to 1, and were identical to 1 if automatic and manual volumes were equal and the intersection was complete.

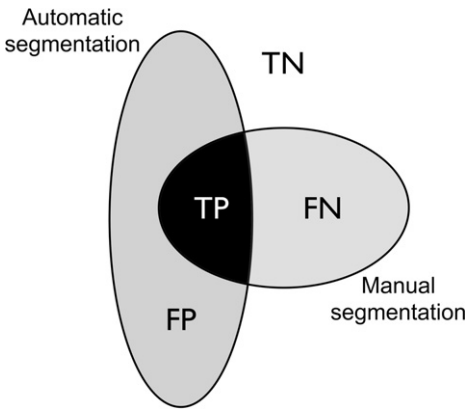

Fig. 1. Definition of true negative (TN), true positive (TP), false negative (FN) and false positive (FP) voxels considering the automatic and manual (reference) segmentations.

The sensitivity (Se) and specificity (Sp) of the ABAS were also computed as follows:

$$\mathrm{Se} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \tag{3}$$

$$\text{and } \mathrm{Sp} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}}, \tag{4}$$

where TN (true negative) is an image voxel not included either in the automatic volume or in the reference volume (Fig. 1). A Receiver Operating Characteristic (ROC) analysis was done by plotting the sensitivity vs. (1 − specificity) for each delineated structure. The best possible result was expected to yield a point in the upper left corner or coordinate (0, 1) of the ROC space, representing 100% sensitivity (all voxels are true positive) and 100% specificity (no false positive voxel is present). This analysis allowed us to visualize the distribution of the points for each structure in the ROC space and to determine the satisfactory automatic delineations and discard the suboptimal ones.

Given the large number of TN voxels in the images relative to FP voxels, when considering Eq. (4) and Fig. 1, the Sp values would probably be close to 1 with weak variations. To obtain consistent specificity results, the TN volume had to be limited to a similar number of voxels to that of a given manually or automatically delineated structure. In our work, we decided to define the TN volume in relation to the reference volume for each structure as follows:

$$\mathrm{TN} + \mathrm{FP} = \alpha \cdot V_{\mathrm{ref}}, \tag{5}$$

where α is a coefficient to be computed. One condition to be fulfilled is $1 - \mathrm{Sp} \leqslant 1$ thus

$$\frac{\mathrm{FP}}{\mathrm{TN} + \mathrm{FP}} \leqslant 1. \tag{6}$$

When Eqs. (5) and (6) were combined, we found that α should not be below $\mathrm{FP}/V_{\mathrm{ref}}$. The ratio $\mathrm{FP}/V_{\mathrm{ref}}$ was computed for all the structures and for the 11 patients. Secondly, the value of α was defined as the integer above the maximum value of $\mathrm{FP}/V_{\mathrm{ref}}$. Finally, the specificity, Sp, was then computed as follows:

$$1 - \mathrm{Sp} = \frac{\mathrm{FP}}{\alpha \cdot V_{\mathrm{ref}}}. \tag{7}$$

Results

The ABAS was used successfully in all the patients: the process was robust, reliable and fast. To automatically delineate the 6 structures, with no user intervention, the average mean time required was 7–8 min using a computer workstation with 4 × 2-GHz CPU and 4-G memory. The time required for manual delineation of the same structures by a radiation oncologist was estimated at about 30 min to one hour.

Fig. 2 shows an example of manual and automatic delineations. Fig. 2a shows a best case congruence for eyes, optic nerves, the brain stem and cerebellum. Fig. 2b shows a worse case congruence for optic nerves and the optic chiasm.

Table 1 shows the comparison between manual and automatic segmentations. The differences between the centre-of-gravity coordinates of the manually and automatically segmented structures were below or equal to 2 mm for the 3 directions (left–right (x), cranio-caudal (y) and antero-posterior (z) directions), except for the optic nerves for which the deviation was greater than 3.8 mm in x- and z-directions. Regarding the volumes, the automatic segmentation underestimated the volume of all the structures from 15% (brian stem) to 50 % (optic chiasm), except for the optic nerves.

DSC values were 0.30 (range, 0–0.72) for the pituitary gland, 0.41 (range, 0–0.58) for the optic chiasm, 0.38 (range, 0.4–0.53) for optic nerves, 0.81 (range, 0.78–0.85) for the eyes, 0.84 (range, 0.79–0.86) for the cerebellum and 0.85 (range, 0.80–0.88) for the brain stem. Fig. 3 shows that DSC values were high (greater than 0.80) for organs larger than 7 cm^3, and tended towards the optimal value of 1 for the eyes, cerebellum and brain stem.

Fig. 4 shows the ROC analysis of the performance of automatic segmentation compared to manual segmentation. The ROC curves for the eyes, brain stem and cerebellum exhibited the same behaviour: all the points were localized in the upper left corner of the ROC space. This means that under our clinical conditions, the ABAS exhibited both high sensitivity and specificity for the larger organs. In the case of the pituitary gland and optic chiasm, the points were mainly localized in the lower left corner of the ROC space: the sensitivity was therefore very low for small structures. In the case of the optic nerves, the points were more dispersed over the ROC space: the results of the ABAS for the optic nerves among the 11 patients were inhomogeneous.

Discussion

In a fundamental study, Rohlfing et al. [19] proposed a complete but complex methodology to assess the performance of atlas-based automatic segmentation software and applied it to images of bee brains. In our work, we tried to estimate in clinical conditions the possible interest of a commercial ABAS (ISOgray™ from DOSIsoft) by using a methodology more adapted for clinical user.

The results obtained with the tested ABAS were satisfactory for all the parameters evaluated (position, volume, DSC) for the larger organs (eyes, brain stem and cerebellum). Contour positioning of the optic chiasm with automatic segmentation was satisfactory but the volume was largely underestimated (−50.6%) and thus the DSC was weak. The results obtained for the pituitary gland were quite similar to those of the optic chiasm but with less underestimation of the volume (−36.8%). All the parameters evaluated for the optic nerves were unsatisfactory: differences in their position exceeded several millimetres and their volumes were overestimated. These results were well correlated with the ROC analysis (Fig. 4).

However, these resulting volume differences must be considered in the light of the following discussion. In the case of the cerebellum, the difference in volume between

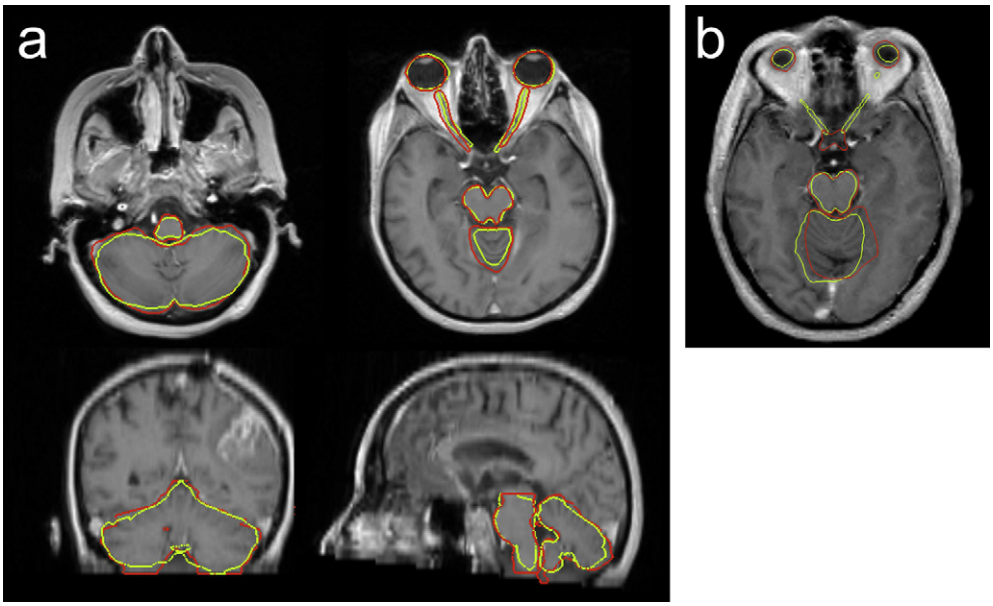

Fig. 2. Manual (red lines) and automatic (green lines) delineations for 2 patients: (a) example of best case congruence for optic nerves, eyes, brain stem and cerebellum (b) example of worse case congruence for optic nerves and optic chiasm.

Table 1
Comparison between automatic and manual segmentations in terms of centre of gravity and volume (mean values for the 11 patients)

Segmented organs	Δx (mm) (SD)	Δy (mm) (SD)	Δz (mm) (SD)	ΔV (%) (SD)
Pituitary gland	1.3 (1.1)	2.0 (1.4)	1.6 (1.3)	−36.8 (35.9)
Optic chiasm	0.8 (0.6)	0.8 (0.9)	1.2 (1.2)	−50.5 (14.2)
Right optic nerve	4.4 (1.8)	1.5 (1.3)	5.3 (3.3)	32.6 (22.2)
Left optic nerve	3.8 (2.1)	1.8 (1.3)	6.5 (2.6)	30.0 (28.1)
Right eye	0.5 (0.3)	0.5 (0.3)	0.4 (0.4)	−24.7 (6.7)
Left eye	0.5 (0.4)	0.5 (0.3)	0.3 (0.2)	−27.8 (5.1)
Brain stem	0.2 (0.2)	0.8 (0.6)	0.6 (0.5)	−14.8 (4.8)
Cerebellum	0.7 (0.8)	0.9 (0.6)	1.0 (0.6)	−16.4 (8.7)

For volumes, a negative value means that the automatic segmentation underestimates the structure volume compared to the manual segmentation.

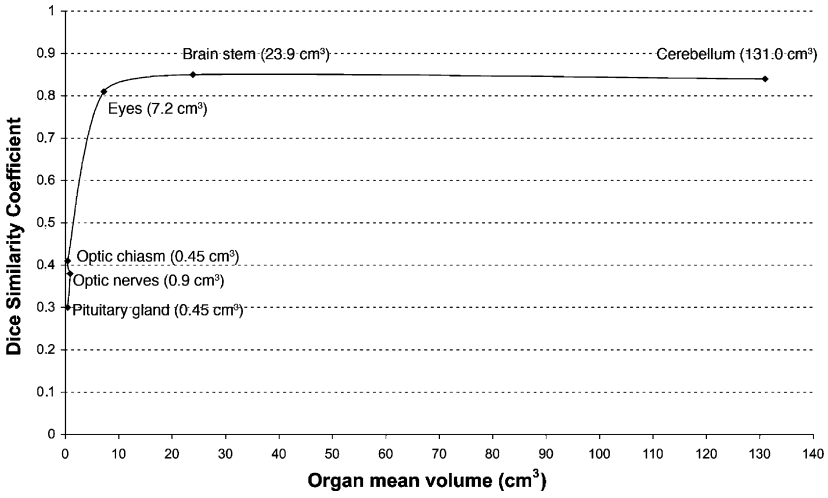

Fig. 3. Dice similarity coefficient values as a function of the manual segmentation mean volume.

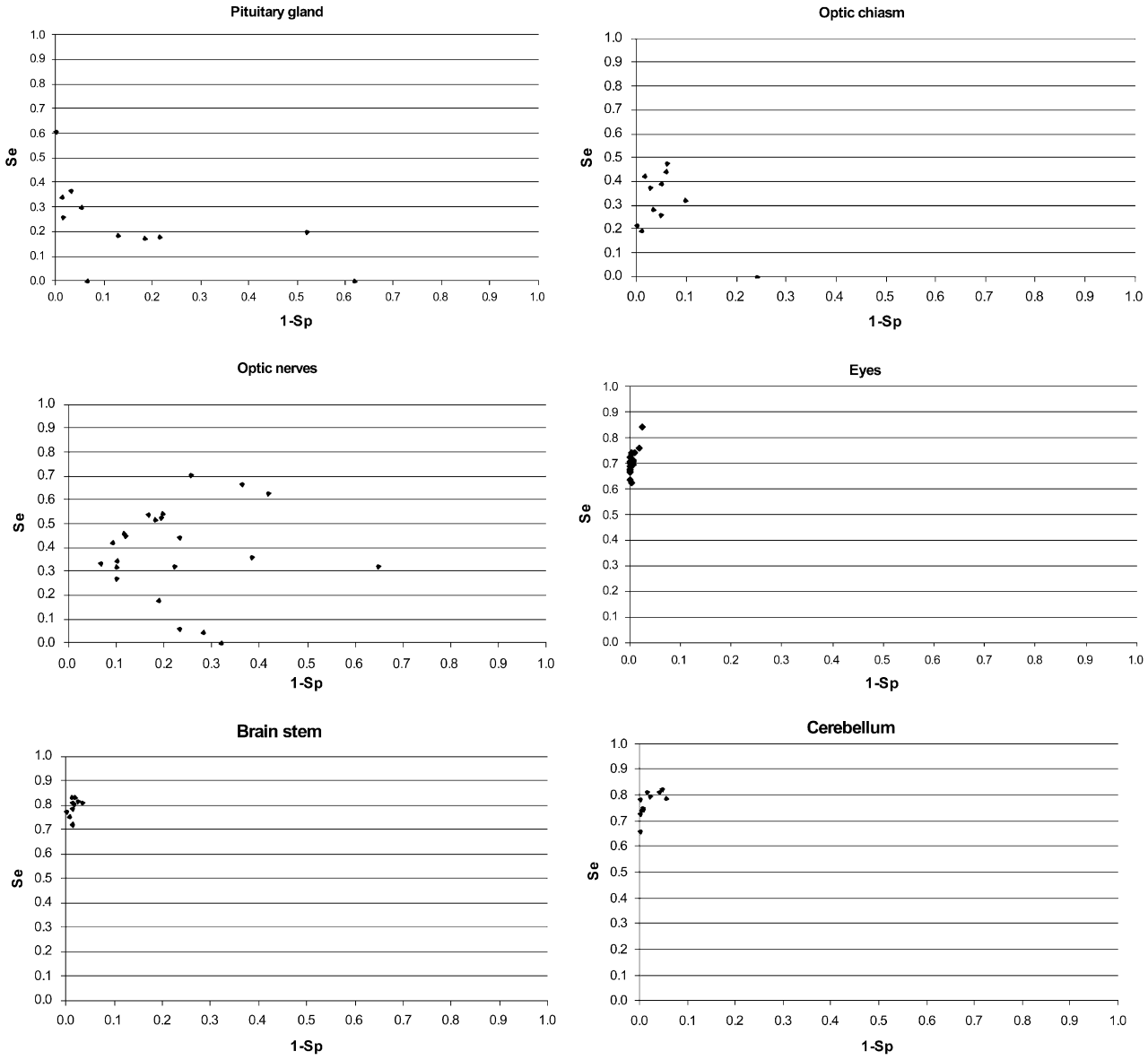

Fig. 4. ROC analysis of the performances of automatic segmentation compared to manual segmentation (=reference segmentation).

automatic and manual segmentations was about −16%. This means that automatic segmentation underestimated the volume compared to manual segmentation. If a 2 mm negative margin is applied to the initial manual contour (V_i = 131.0 cm^3), then the corresponding volume ($V_{i\text{-}2mm}$) would be equal to 111.6 cm^3. The difference between V_i and $V_{i\text{-}2mm}$ in terms of volume would therefore be equal to −15.5%. In other words, the volume errors for the cerebellum correspond to small spatial errors which are equivalent to a contraction of the manual volume by 2 mm. If the same analysis is applied to the eyes, the 26% volume underestimation by the ABAS corresponded to a margin that was slightly larger than 1 mm. In the case of the brain stem, the 15% volume underestimation corresponded to a margin lower than 1 mm.

The same analysis cannot be applied to the smallest structures (optic chiasm, optic nerves, pituitary gland) due to the complexity of the shapes.

Finally, if the ROC analysis is considered from a radiotherapy point of view, the performance of automatic segmentation could result in overirradiation of OAR voxels (FN voxels) or possible underirradiation of target volume voxels (FP voxels). Fig. 4 shows that the higher the sensitivity, the lower the risk of overirradiation of normal tissues (OAR) and the higher the specificity, the lower the risk of underirradiation of tumour tissue: any existing neoplastic tissue would not be falsely considered as normal tissue requiring protection. As suggested by Andrews [2], we subdivided the ROC space into 4 sub-regions, as represented in Fig. 5:

(1) the ''acceptable'' region: both the sensitivity and specificity are high. OAR are adequately segmented, thus well protected, and the number of possibly underirradiated target volume voxels is low,

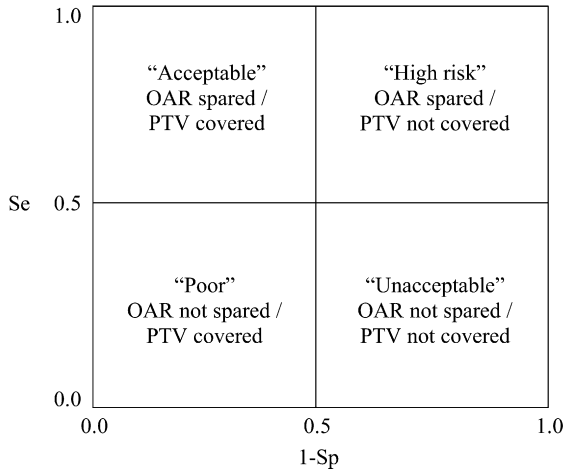

Fig. 5. Subdivision of the ROC space into four sub-regions following the instructions of Andrews [2].

(2) the ''poor'' region: the specificity is high and the sensitivity is low. OAR are not sufficiently protected, but the number of underirradiated target volume voxels is low,
(3) the ''high risk'' region: the sensitivity is high and the specificity is low. The contours of OAR are too large, therefore a neighbouring target volume could be underirradiated,
(4) the ''unacceptable'' region: both the sensitivity and specificity are low. The automatic segmentation quality is completely unsatisfactory. OAR are not adequately protected and a neighbouring target volume could be underirradiated.

If this 4-sub-region scheme is applied to Fig. 4, then the performances of the ABAS would be considered ''acceptable'' for the eyes, brain stem, and cerebellum. The specificity of automatic segmentation is very satisfactory in the case of the pituitary gland and optic chiasm but the sensitivity is low which corresponds to the ''poor'' region. In the case of the optic nerves, most of the points were in the poor region but the specificity was more variable.

The contours delineated with the ABAS were very satisfactory for the larger organs with volumes exceeding 7 cm^3. However, one must bear in mind that the ABAS was developed using non-contrast-enhanced MRI with 1-mm thick slice images and that the developers recommend using 2-mm thick or smaller slice images. In the present retrospective study, we wanted to test the performance of the ABAS under our clinical conditions, i.e., with a slice thickness equal to 3 mm. The pituitary gland and optic chiasm were often only visible on one slice which explains why the results were not as satisfactory as expected. The use of 1-mm thick slices might improve the contour delineation of small structures. However, the complexity of all the surrounding anatomical information could only be analysed by an expert.

Another noteworthy point is that this atlas was developed for adults. Whether the same atlas can be used in a pediatric context (two additional structures, the cochlea and the supra-tentorial parenchyma, would have to be included), or whether a specific atlas should be constructed, warrants investigation.

Conclusion

The aim of this work was to evaluate the possible interest of version 3.1 of the ABAS, included in the TPS designated ISOgray™, from DOSIsoft, for brain OAR when used under clinical conditions of our daily practice, which are different from those specified by the developers.

The present series of tests on the ABAS, used on T1-weighted head MR images, demonstrated that it is robust and reliable in our clinical practice for large structures (brain stem, cerebellum, eyes), even though the small structures (optic chiasm, optic nerves, pituitary gland) still need to be delineated manually by an expert. This ABAS for the brain can thus help radiation oncologists save time during radiation therapy treatment planning and reduce the variability in OAR delineation.

Acknowledgements

The authors thank Dr. J. Coulot, Institut Gustave Roussy, for editorial advice and L. Saint Ange for editing. This work was undertaken in the framework of the MAESTRO Integrated Project (IP CE503564) funded by the European Commission.

* **Corresponding author. Aurélie Isambert**, Service de Physique Institut Gustave Roussy, 39 rue Camille Desmoulins, 94805 Villejuif, France. *E-mail address:* aurelie.isambert@igr.fr

Received 7 August 2007; received in revised form 15 November 2007; accepted 20 November 2007; Available online 26 December 2007

References

[1] Admiraal-Behloul F, van den Heuvel DM, Olofsen H, et al. Fully automatic segmentation of white matter hyperintensities in MR images of the elderly. Neuroimage 2005;28:607—17.
[2] Andrews JR. Benefit, risk, and optimization by ROC analysis in cancer radiotherapy. Int J Radiat Oncol Biol Phys 1985;11:1557—62.
[3] Arsigny V, Commowick C, Pennec X, Ayache N. A log-euclidean polyaffine framework for locally rigid or affine registration. In: Int Workshop Biomed Image Registration. LNCS, vol. 4057, Springer; 2006. p. 120—7.
[4] Bondiau PY, Malandain G, Chanalet S, et al. Atlas-based automatic segmentation of MR images: validation study on the brainstem in radiotherapy context. Int J Radiat Oncol Biol Phys 2005;61:289—98.
[5] Cattaneo GM, Reni M, Rizzo G, et al. 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. Radiother Oncol 2005;75:217—23.
[6] Cocosco CA, Kollokian V, Kwan RKS, Evans AC. BrainWeb: online interface to a 3D MRI simulated brain database (abstract). Neuroimage 1997;5:S425.
[7] Collins DL, Zijdenbos AP, Kollokian V, et al. Design and construction of a realistic digital brain phantom. IEEE Trans Med Imaging 1998;17:463—8.
[8] Commowick O, Arsigny V, Isambert A, et al. An efficient locally affine framework for the smooth registration of anatomical structures. Med Image Anal, in press.
[9] Cuadra MB, Pollo C, Bardera A, Cuisenaire O, Villemure JG, Thiran JP. Atlas-based segmentation of pathological MR brain images using a model of lesion growth. IEEE Trans Med Imaging 2004;23:1301—14.

[10] Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297–302.
[11] Ganser KA, Dickhaus H, Metzner R, Wirtz CR. A deformable digital brain atlas system according to Talairach and Tournoux. Med Image Anal 2004;8:3–22.
[12] Geets X, Daisne JF, Arcangeli S, et al. Inter-observer variability in the delineation of pharyngo-laryngeal tumor, parotid glands and cervical spinal cord: comparison between CT-scan and MRI. Radiother Oncol 2005;77:25–31.
[13] Giraud P, Elles S, Helfre S, et al. Conformal radiotherapy for lung cancer: different delineation of the gross tumor volume (GTV) by radiologists and radiation oncologists. Radiother Oncol 2002;62:27–36.
[14] Isambert A, Dhermain F, Beaudre A, et al. Requirements for the use of an atlas-based automatic segmentation for delineation of Organs at risk (OAR) in conformal radiotherapy (CRT): quality assurance (QA) and preliminary results for 22 adult patients with primary brain tumors. Radiother Oncol 2005;76:S133 (abstract).
[15] Mazzara GP, Velthuizen RP, Pearlman JL, Greenberg HM, Wagner H. Brain tumor target volume determination for radiation treatment planning through automated MRI segmentation. Int J Radiat Oncol Biol Phys 2004;59: 300–12.
[16] Ourselin S, Roche A, Prima S, Ayache N. Block matching: a general framework to improve robustness of rigid registration of medical images. In: Int Conf Med Image Comput Assist Interv. LNCS, vol. 1935, Springer; 2000. p. 557–66.
[17] Popple RA, Griffith HR, Sawrie SM, Fiveash JB, Brezovich IA. Implementation of Talairach atlas based automated brain segmentation for radiation therapy dosimetry. Technol Cancer Res Treat 2006;5:15–21.
[18] Prastawa M, Gilmore JH, Lin W, Gerig G. Automatic segmentation of MR images of the developing newborn brain. Med Image Anal 2005;9:457–66.
[19] Rohlfing T, Russakoff D, Maurer C. Performance-based classifier combination in atlas-based image segmentation using expectation–maximisation parameter estimation. IEEE Trans Med Imaging 2004;23:983–94.
[20] Steenbakkers RJ, Duppen JC, Fitton I, et al. Reduction of observer variation using matched CT-PET for lung cancer delineation: a three-dimensional analysis. Int J Radiat Oncol Biol Phys 2006;64:435–48.
[21] Struikmans H, Warlam-Rodenhuis C, Stam T, et al. Interobserver variability of clinical target volume delineation of glandular breast tissue and of boost volume in tangential breast irradiation. Radiother Oncol 2005;76:293–9.
[22] Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain. Thieme, 1988.
[23] Zhang T, Chi Y, Meldolesi E, Yan D. Automatic delineation of on-line head-and-neck computed tomography images: toward on-line adaptive radiotherapy. Int J Radiat Oncol Biol Phys 2007;68:522–30.