



**HAL**  
open science

## Clinical implementation of deep-learning based auto-contouring tools—Experience of three French radiotherapy centers

C. Robert, A. Munoz, D. Moreau, J. Mazurier, G. Sidorski, A. Gasnier, G. Beldjoudi, V. Grégoire, E. Deutsch, Philippe Meyer, et al.

### ► To cite this version:

C. Robert, A. Munoz, D. Moreau, J. Mazurier, G. Sidorski, et al.. Clinical implementation of deep-learning based auto-contouring tools—Experience of three French radiotherapy centers. *Cancer/Radiothérapie*, 2021, 25 (6-7), pp.607-616. 10.1016/j.canrad.2021.06.023 . hal-03797764

**HAL Id: hal-03797764**

**<https://hal.science/hal-03797764>**

Submitted on 16 Oct 2023

**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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Clinical implementation of deep-learning based auto-contouring tools – Experience of three French radiotherapy centers

## Implémentation clinique d'algorithmes de segmentation automatique basés sur du Deep-Learning – Retour d'expérience de trois départements de radiothérapie français

C. Robert<sup>a</sup>, A. Munoz<sup>b</sup>, D. Moreau<sup>c</sup>, J. Mazurier<sup>d</sup>, G. Sidorski<sup>d</sup>, A. Gasnier<sup>a</sup>, G. Beldjoudi<sup>b</sup>, V. Grégoire<sup>b</sup>, E. Deutsch<sup>a</sup>, P. Meyer<sup>e</sup>, L. Simon<sup>f</sup>

<sup>a</sup>Department of Radiotherapy, Gustave Roussy, Villejuif, France

<sup>b</sup>Department of Radiotherapy, Centre Léon Bérard, Lyon, France

<sup>c</sup>Department of Radiotherapy, Hôpital Européen Georges Pompidou, Paris, France

<sup>d</sup>Department of Radiotherapy, Clinique Pasteur - Oncorad, Toulouse, France

<sup>e</sup>Service d'Oncologie Radiothérapie, Institut de Cancérologie Strasbourg Europe (Icans), Strasbourg, France

<sup>f</sup>Institut Claudius Regaud (ICR), Institut Universitaire du Cancer de Toulouse - Oncopole (IUCT-O), Toulouse, France

Corresponding author:

Charlotte Robert

Département de Radiothérapie,

Gustave Roussy Cancer Campus

114, rue Edouard Vaillant, 94805 Villejuif

Tél. 0142115606

Mail. [ch.robert@gustaveroussy.fr](mailto:ch.robert@gustaveroussy.fr)

Congrès SFRO2021, Session "Session SFRO/SFPM - IA et radiothérapie : vision du physicien médical"

**Running Title:** DL tools for RT auto-contouring

# Abstract

Deep-learning (DL)-based auto-contouring solutions have recently been proposed as a convincing alternative to decrease workload of target volumes and organs-at-risk (OAR) delineation in radiotherapy planning and improve inter-observer consistency. However, there is minimal literature of clinical implementations of such algorithms in a clinical routine. In this paper we first present an update of the state-of-the-art of DL-based solutions. We then summarize recent recommendations proposed by the European society for therapeutic radiology and oncology (ESTRO) to be followed before any clinical implementation of artificial intelligence-based solutions in clinic. The last section describes the methodology carried out by three French radiation oncology departments to deploy CE-marked commercial solutions.

Based on the information collected, a majority of OAR are retained by the centers among those proposed by the manufacturers, validating the usefulness of DL-based models to decrease clinicians' workload. Target volumes, with the exception of lymph node areas in breast, head and neck and pelvic regions, whole breast, breast wall, prostate and seminal vesicles, are not available in the three commercial solutions at this time. No implemented workflows are currently available to continuously improve the models, but these can be adapted/retrained in some solutions during the commissioning phase to best fit local practices. In reported experiences, automatic workflows were implemented to limit human interactions and make the workflow more fluid. Recommendations published by the ESTRO group will be of importance for guiding physicists in the clinical implementation of patient specific and regular quality assurances.

Keywords: Deep-learning, auto-contouring, automatic delineation, clinical implementation, radiotherapy

## Résumé

Les solutions de segmentation automatique basées sur l'apprentissage profond ont été proposées récemment comme une alternative intéressante pour les départements de radiothérapie pour réduire fortement la charge de travail liée à la définition des volumes cibles et des organes à risque dans la planification de traitement et améliorer la cohérence inter-experts. Peu de publications dans la littérature décrivent cependant la mise en œuvre clinique de ces algorithmes. Cet article vise à présenter dans un premier temps une mise à jour de l'état de l'art des solutions de segmentation automatique basées sur l'apprentissage profond. La deuxième partie résume les recommandations récentes publiées par la société européenne de radiologie thérapeutique et d'oncologie (ESTRO) à mettre en œuvre pour une implémentation sécurisée de solutions basées sur l'intelligence artificielle. La dernière partie décrit la méthodologie suivie par trois services de radiothérapie français pour le déploiement de trois solutions commerciales différentes. D'après les informations recueillies, la majorité des OAR proposés par les constructeurs sont retenus par les centres, validant l'apport des modèles basés sur l'apprentissage profond sur la diminution de la charge de travail des cliniciens. Les volumes cibles, à l'exception des aires ganglionnaires dans les régions du sein, de la tête et du cou et pelviennes, de la glande mammaire, de la paroi mammaire, de la prostate et des vésicules séminales, ne sont pas disponibles dans les trois solutions commerciales à l'heure actuelle. Aucune implémentation ne permet aujourd'hui d'améliorer les modèles en continu, mais ceux-ci peuvent être adaptés et réentraînés dans certaines solutions pendant la phase de mise en service pour mieux

correspondre aux pratiques locales. Dans les expériences rapportées, la mise en œuvre est très automatisée et ne requiert quasiment aucune intervention humaine, rendant ainsi très fluide le flux de travail. Les recommandations publiées par l'ESTRO sont importantes pour guider les physiciens dans la mise en œuvre clinique de programmes spécifiques d'assurance qualité.

Mots-clés : Apprentissage profond, segmentation automatique, délinéation automatique, implémentation clinique, radiothérapie.

## Introduction

Workflow in radiation therapy consists of a succession of steps, some of which, such as defining target volumes and organs-at-risk (OAR), are very time-consuming for radiation oncologists. Atlas, multi-atlas and model-based solutions have been proposed first as an alternative to manual contouring to decrease physicians' workload(1). However, clinical adoption has been limited, as the time required to correct automatic contours can take as long as manual contouring from scratch, especially for low-contrast structures or certain lymph node regions(2,3). The recent advent of graphics processing units (GPU) has led to many recent breakthroughs for deep-learning (DL) techniques and their translation to medicine(4). Radiation oncology has rapidly benefited from this trend for several reasons. Firstly, radiotherapy departments have large amounts of annotated retrospective data, which have been of great value in developing models based on artificial intelligence (AI). Delineation of target volumes and OAR, which is a mandatory step for every patient, is in addition highly dependent on the observer. Furthermore, it is an extremely time-consuming and labor-intensive procedure. Therefore, DL solutions provide many strong arguments for the harmonization of practices, which is becoming necessary at a time of millimetric-scale irradiations and image-guided radiotherapy. In this paper we first present an update of the state-of-the-art of DL-based solutions. We then summarize recent recommendations proposed by the European society for therapeutic radiology and oncology (ESTRO) to be followed before any clinical implementation of artificial intelligence-based solutions in clinic. The last section describes the methodology carried out by three French radiation oncology departments to deploy CE-marked commercial solutions.

## Litterature review

A PubMed search based on the following keywords was conducted to summarize state-of-the-art: "radiotherapy" and (("automatic" and "segmentation") or "auto-contouring" or "auto contouring" or "automatic contouring" or "autocontouring" or "auto-segmentation" or "auto segmentation" or "automatic segmentation" or "autosegmentation" or "automatic delineation") and "deep learning". Table 1 summarizes the results of this search. It validates enthusiasm in this new area of research from 2018, however highlights a moderate number of studies detailing methodology or feedback regarding clinical experience. These few papers are briefly described here. In 2021, Cha *et al.* prospectively evaluated performance of an in-house auto-segmentation algorithm for MR-based prostate radiotherapy planning on a cohort of 173 patients, in which 85% of the patients were treated by stereotactic body radiation therapy (SBRT) at Memorial Sloan Kettering Cancer Center (New York, United States)(5). High performance was observed for the clinical target volumes (CTV) including the prostate and the seminal vesicles with a volumetric Dice-Sørensen coefficient (VDSC) equal to 0.89 (interquartile range: 0.83-0.95). For 43 patients, a survey was conducted and showed that 33% of auto-contours required major "clinically significant" adjustments, leading to 12 minutes of time saved (30%) on the whole cohort compared to historic contouring time evaluations. A second evaluation was performed in the Department of Radiation Oncology, Groningen, The Netherlands from April 2019 to April 2020(6). In this experience, a commercial DL-based system (DLC Model, Mirada WorkflowBox 2.0, Mirada Medical Ltd, Oxford, UK) was tested for contouring of 26 OAR on 103 head and neck patients. Authors observed that the software had a tendency to under-estimate the volumes compared to the reference contours of all considered organs. They concluded that the study had allowed them to highlight deviations in manual contouring practices and identify a lack of clarity or uncertainties in guidelines. The latest work detailing a clinical implementation experience

was carried out in the radiotherapy department in Utrecht, the Netherlands and was focused on magnetic resonance imaging-based OAR auto-contouring in prostate cancer(7). In this monocentric study, two 3D convolutional neural networks (CNN) called DeepMedic and dense V-net were trained to segment bladder, rectum and femurs on the first 48 patients. Considering geometric metrics and rating by one expert, the best algorithm, i.e. DeepMedic, was selected, retrained on 97 consecutive patients and implemented in clinical routine in August 2019. On the test set (53 patients), a mean surface Dice similarity coefficient of  $0.98\pm 0.03$ ,  $0.92\pm 0.05$ ,  $0.989\pm 0.008$  and  $0.997\pm 0.003$  was obtained for bladder, rectum, left and right femurs respectively ( $\tau = 2$  mm). As detailed in Table 1, the majority of other papers were dedicated to network development and performance evaluation based either on monocentric (N = 36) or multicentric cohorts (N = 18). Head and neck (N = 21) and prostate (N = 10) were tumor locations encountering the main interest.

## ESTRO recommendations

Following the 3rd **ESTRO** physics workshop on 'Implementation/commissioning/quality assurance (QA) of artificial intelligence techniques' in Budapest (2019), recommendations were proposed by a group of seven medical physicists regarding implementation and quality assurance of artificial intelligence-based applications in radiotherapy(8). Concerning automatic segmentation tools, authors insist on the need to ask for details to the provider of the algorithm about the variability of the clinical data used to train the model, including variability in acquisition parameters/devices, with the objective to evaluate its generalizability. They set the minimal number of patients to be included in the local test set to 10, notifying this number should be increased in case of large variations in



the similarity metrics considered for performance evaluation. Auto-contouring is quite easy to introduce in clinic given that every contour should be reviewed by a radiation oncologist. However, methods have been proposed in the literature to ease quality checks and rate confidence in the proposed contours and were mentioned by the authors as solutions to be implemented for case-specific QA. As an example, statistical models characterizing shape, volume or spatial location of the centroid of generated contours have been proposed(9). Use of a second auto contouring algorithm can be thought about too to identify outliers and generate warnings(10). Last option considers an architecture integrating a network for segmentation and another for QA(11). In this work, the second network predicted either categorial outputs indicating the quality level of each slice in terms of Dice similarity coefficient (DSC) or the DSC score directly. In terms of routine QA, authors recommended setting up a Quality Management Program (QMP) to ensure that the model does not vary with time even in case of software's version update. A reference data set reflecting clinical practices should be selected for this purpose and automatically recontoured on a regular basis to identify and quantify changes. In addition, they suggested the idea of creating a repository of patient cases for which contouring was suboptimal to identify limitations of the proposed model and ease adjustments by the developers of the algorithm. If the acquisition parameters are changed or a new equipment is purchased, a new test set should be implemented and the performance of the model reassessed.

## Experiences of clinical implementation

The present section describes the methodology of clinical implementation of different commercial software for auto-contouring performed in three French radiation oncology departments, i.e. Centre Léon Bérard (Lyon, France), Hôpital Européen Georges Pompidou (Paris, France) and Clinique Pasteur-ONCORAD (Toulouse, France). Table 2 summarizes models deployed in each center, numbers of OAR and target volumes automatically contoured for each model, and percentages of OAR and target volumes retained by the center. The second part of Table 2 describes the number of patient data used to construct the model by the provider (if known). Part 3 details the local commissioning procedures, including the number of patients considered for each tumor location, and the metrics used for performance analysis. The last group of lines is focused on the case-specific QA and regular QA.

## Deployment of ART-Plan Annotate in Centre Léon Bérard (Lyon, France)

ART-Plan Annotate (Therapanacea, Paris, France) is an automatic contouring CE/FDA-cleared solution based on an ensemble of deep learning models trained for contouring of 100+ organs at risk according to ESTRO guidelines on the basis of 25,000 patients after anatomically preserving data augmentation(12).

The solution was deployed one year ago in Centre Léon Bérard, Lyon, France (8 treatment units, more than 3000 patients treated per year) for six tumor locations including male pelvis, abdominal cavity, thorax, breast, head and neck and brain cancers. Since then, approximately 500 prostate patients, 300 digestive patients (abdominal cavity), 386 lung patients, 694 breast patients, 215 head and neck patients, and 260 brain patients have benefitted from this technology. The software

provides lymph node auto-contouring for pelvic, breast and head and neck tumor locations. However, it was not retained by the local medical staff for head and neck tumors. The trained models were directly made available to users, without specific training on local data. No further evolution of the model is currently planned on the basis of the prospective data acquired in clinical routine.

The deployment timeline started with head and neck tumors, for which a first multicentric evaluation was carried out using data from the Centre Léon Bérard and Gustave Roussy (Villejuif, France). In this study, 100 patients were retrospectively selected in both centers (50 from each) and for each patient ART-Plan solution was used to generate full annotations of 14 OAR. After VSDC evaluation, the automatic contours were blended with those corresponding to the clinical experts (manual contours) in a blinded fashion. Each contour was then scored by 5 experts from both centers, as A/ clinically acceptable, B/ clinically acceptable after minor corrections, C/ not acceptable. Results showed that 96% of all manual contours were classified as clinically relevant (A + B). Values were equal to 98% for automatic contouring. The software had difficulty in delineating optical nerves and sub-mandibular glands.

Following these positive results, deployment was pursued for pelvic and thoracic tumor locations, in five patients per location as a local test set. Five OAR were evaluated for the pelvic region and 6 OAR for the thoracic one. The comparison of the contours created by ART-Plan Annotate with the ones drawn by the clinical expert showed a mean Jaccard's index of 86 % for the pelvic region and of 84 % for the thoracic region. Based on the Jaccard's values, the contours that required the strongest corrections were identified to be the rectum in the pelvic region and the trachea in the thoracic area.

Today, batch options, i.e. no human interaction is required, are implemented in order to launch automatically auto-contouring after Computed Tomography (CT) acquisition. As well, contours and associated CT images are sent automatically to the Treatment Planning System (TPS), where they are reviewed and edited when corrections are needed.

Today, the integration of this software into clinical routine has led to significant changes in local practices with more OAR delineated for each patient and the delegation of delineation tasks distributed to dosimetrists. For example, for head and neck tumors, 13 OAR were previously delineated by physicians, including the parotids or the submandibular glands, which are now handled by dosimetrists. The introduction of automatic delineation in a clinical routine has finally allowed the harmonization of the nomenclatures of OAR and target volumes, which constitutes a step forward for the constitution of prospective databases, that is of utmost importance in the growing context of machine learning.

## Deployment of DLCExpert in Hôpital Européen Georges Pompidou (Paris)

DLCExpert (Mirada Medical, Oxford, UK) exploits state-of-the-art DL technology to delineate automatically OAR and lymph node areas for head and neck, thorax, breast and prostate tumor locations. To provide high quality contours, training data integrates hundreds of patient datasets (at least 150 patient data per model), curated by clinical experts according to consensus guidelines. Based on diverse European collaborations, several models have been developed per anatomy, which allows the tool to be configured to fit local practices. Thus, institution can make the choice to

keep model A for organ X and model B for organ Y. A post-processing tool is finally proposed by the provider to interact with the contours after they have been automatically inferred. For example, overlap management can be useful in cases where several models have been used for the contouring of different organs.

The department of radiation oncology of Hôpital Européen Georges Pompidou (4 treatment units, 1600 patients treated per year) is currently deploying DLCExpert for head and neck, male pelvis and breast cancers. However, this feedback focuses on breast cancer only, as the deployment of automatic contouring for this location is the most advanced, even still not used in clinical routine at the time of the writing of this article. For this tumor location, 6 OAR and 12 target volumes, including both breasts and 5 lymph node regions per side, are provided by the software and are all retained by the medical staff. Commissioning phase has evaluated quality of the contours for 10 institutional patients based on the VDSC metric. Subjective rating by experts is in addition in progress using a 4-level scale (A/ no correction, B/ some missing slices at the lower or upper limits of the organ of interest, C/ a part of the organ of interest is missing, D/ a large part of the organ or the whole organ needs correction). In the short term, an evaluation of time saved is also envisaged by the center.

The workflow has been designed so that it does not require human intervention. To this, a DICOM push is used to export the planning CT of the patient from the CT unit or the TPS to the DLCExpert server. The software then automatically identifies the model to be applied thanks to a dedicated DICOM tag. Auto-contours are finally automatically pushed to the TPS where they are retrieved by the user. No specific-case QA is today envisaged. For regular QA, the center will identify test patients for which the automatic contouring will be relaunched at each major change in the workflow (change in the software version, CT device change, etc.).

Questions are now open about the place this type of tool should have in the training of medical residents.

## Deployment of RayStation v9B automatic segmentation in Clinique Pasteur-ONCORAD (Toulouse)

The Oncorad Garonne group is a private center located at the Pasteur Clinic (Toulouse, France) that treats approximately 2500 patients per year with 7 linear accelerators. The group is currently running the commissioning phase of the latest automatic segmentation solution RayStation v9B (RaySearch Laboratories, Stockholm, Sweden), based on fully-convolutional neural networks (FCNN). RayStation offers two alternatives to the users: either the use of its own models (RS models) or the possibility to train a new model (LOCAL model) based on local data. About 100 patients are needed to train RayStation's DL-based models. Today, the Oncorad Group has selected RS models for 7 OAR (bladder, rectum, femoral heads, lungs, heart, esophagus and spinal cord) and the LOCAL models for target volumes (prostate, seminal vesicles and pelvic nodes) to better reproduce local practices. To evaluate the performance of the proposed solution, a first study was carried out, including 250 pelvic patients and 100 patients treated for a thoracic lesion. In this study, DL-based contours were compared to multi-atlas ones using physicians' contours as references. VDSC, sensitivity and specificity metrics were quantified as performance metrics. In addition, a geometric study was conducted with the goal to evaluate the differences between the 3D coordinates of the barycenter of each organ. For pelvic location, shorter times were obtained for DL-based contours, with

about one minute required to generate the OAR, compared to 6 minutes for the multi-atlas solution. DL-based delineations always obtained better scores than the multi-atlas ones with VDSC scores equal to 0.97, 0.87, and 0.88 for the bladder, rectum, and femoral heads respectively, to be compared with 0.71, 0.54 and 0.92 for the multi-atlas algorithm. Using the LOCAL models, VDSC scores of 0.89, 0.81 and 0.71 were obtained for the prostate, seminal vesicles and pelvic nodes respectively. The mean differences between barycenter coordinates were less than 1 cm for all pelvic organs in the head-foot axis. In the right-left and anterior-posterior axes, the differences were all less than 0.4 cm, with a tendency to segment more slices for the LD-based solution than the medical doctors.

The introduction of the use of the software in the clinical routine is underway for pelvic patients, especially those with lymph node involvement.

## Conclusion

This article has summarized the recent literature on DL-based auto-contouring tools and identified the lack of shared experiences regarding clinical implementation. Based on three experiences, we have highlighted the clinical benefit of such tools, at least for the delineation of OAR and lymph node areas. Each center has deployed its own methodology for software implementation and quality assurance, highlighting the usefulness of recently published recommendations. Initial training must evolve rapidly to integrate this major change in practice and make physicians and physicists aware of the basics of AI at least in order to make communication with the industry critical and constructive.

# References

1. Vrtovec T, Močnik D, Strojan P, Pernuš F, Ibragimov B. Auto-segmentation of organs at risk for head and neck radiotherapy planning: From atlas-based to deep learning methods. *Med Phys.* sept 2020;47(9):e929-50.
2. Thomson D, Boylan C, Liptrot T, Aitkenhead A, Lee L, Yap B, et al. Evaluation of an automatic segmentation algorithm for definition of head and neck organs at risk. *Radiat Oncol Lond Engl.* 3 août 2014;9:173.
3. Lim JY, Leech M. Use of auto-segmentation in the delineation of target volumes and organs at risk in head and neck. *Acta Oncol Stockh Swed.* juill 2016;55(7):799-806.
4. Huynh E, Hosny A, Guthier C, Bitterman DS, Petit SF, Haas-Kogan DA, et al. Artificial intelligence in radiation oncology. *Nat Rev Clin Oncol.* déc 2020;17(12):771-81.
5. Cha E, Elguindi S, Onochie I, Gorovets D, Deasy JO, Zelefsky M, et al. Clinical implementation of deep learning contour autosegmentation for prostate radiotherapy. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* 3 mars 2021;159:1-7.
6. Brouwer CL, Boukerroui D, Oliveira J, Looney P, Steenbakkens RJHM, Langendijk JA, et al. Assessment of manual adjustment performed in clinical practice following deep learning contouring for head and neck organs at risk in radiotherapy. *Phys Imaging Radiat Oncol.* oct 2020;16:54-60.
7. Savenije MHF, Maspero M, Sikkes GG, van der Voort van Zyp JRN, T J Kotte AN, Bol GH, et al. Clinical implementation of MRI-based organs-at-risk auto-segmentation with convolutional networks for prostate radiotherapy. *Radiat Oncol Lond Engl.* 11 mai 2020;15(1):104.
8. Vandewinckele L, Claessens M, Dinkla A, Brouwer C, Crijns W, Verellen D, et al. Overview of artificial intelligence-based applications in radiotherapy: Recommendations for implementation and quality assurance. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* déc 2020;153:55-66.
9. Chen H-C, Tan J, Dolly S, Kavanaugh J, Anastasio MA, Low DA, et al. Automated contouring error detection based on supervised geometric attribute distribution models for radiation therapy: a general strategy. *Med Phys.* févr 2015;42(2):1048-59.
10. Court LE, Kisling K, McCarroll R, Zhang L, Yang J, Simonds H, et al. Radiation Planning Assistant - A Streamlined, Fully Automated Radiotherapy Treatment Planning System. *J Vis Exp.* 11 avr 2018;(134):57411.
11. Chen X, Men K, Chen B, Tang Y, Zhang T, Wang S, et al. CNN-Based Quality Assurance for Automatic Segmentation of Breast Cancer in Radiotherapy. *Front Oncol.* 2020;10:524.
12. Ung M, Rouyar-Nicolas A, Limkin E, Petit C, Sarrade T, Carre A, et al. Improving Radiotherapy Workflow Through Implementation of Delineation Guidelines & AI-Based Annotation. *Int J Radiat Oncol.* nov 2020;108(3):e315.
13. Shi J, Ding X, Liu X, Li Y, Liang W, Wu J. Automatic Clinical Target Volume Delineation for Cervical Cancer in CT Images Using Deep Learning. *Med Phys.* 27 avr 2021;



14. Jenkins A, Mullen TS, Johnson-Hart C, Green A, McWilliam A, Aznar M, et al. Novel methodology to assess the effect of contouring variation on treatment outcome. *Med Phys*. 26 mars 2021;
15. Liu C, Zhang X, Si W, Ni X. Multiview Self-Supervised Segmentation for OARs Delineation in Radiotherapy. *Evid-Based Complement Altern Med ECAM*. 2021;2021:8894222.
16. Hague C, McPartlin A, Lee LW, Hughes C, Mullan D, Beasley W, et al. An evaluation of MR based deep learning auto-contouring for planning head and neck radiotherapy. *Radiother Oncol J Eur Soc Ther Radiol Oncol*. mai 2021;158:112-7.
17. Chung SY, Chang JS, Choi MS, Chang Y, Choi BS, Chun J, et al. Clinical feasibility of deep learning-based auto-segmentation of target volumes and organs-at-risk in breast cancer patients after breast-conserving surgery. *Radiat Oncol Lond Engl*. 25 févr 2021;16(1):44.
18. Kim N, Chun J, Chang JS, Lee CG, Keum KC, Kim JS. Feasibility of Continual Deep Learning-Based Segmentation for Personalized Adaptive Radiation Therapy in Head and Neck Area. *Cancers*. 9 févr 2021;13(4).
19. Kieselmann JP, Fuller CD, Gurney-Champion OJ, Oelfke U. Cross-modality deep learning: Contouring of MRI data from annotated CT data only. *Med Phys*. avr 2021;48(4):1673-84.
20. Zhang S, Wang H, Tian S, Zhang X, Li J, Lei R, et al. A slice classification model-facilitated 3D encoder-decoder network for segmenting organs at risk in head and neck cancer. *J Radiat Res (Tokyo)*. 1 janv 2021;62(1):94-103.
21. Brunenberg EJJ, Steinseifer IK, van den Bosch S, Kaanders JHAM, Brouwer CL, Gooding MJ, et al. External validation of deep learning-based contouring of head and neck organs at risk. *Phys Imaging Radiat Oncol*. juill 2020;15:8-15.
22. Oktay O, Nanavati J, Schwaighofer A, Carter D, Bristow M, Tanno R, et al. Evaluation of Deep Learning to Augment Image-Guided Radiotherapy for Head and Neck and Prostate Cancers. *JAMA Netw Open*. 2 nov 2020;3(11):e2027426.
23. Kiljunen T, Akram S, Niemelä J, Löyttyniemi E, Seppälä J, Heikkilä J, et al. A Deep Learning-Based Automated CT Segmentation of Prostate Cancer Anatomy for Radiation Therapy Planning-A Retrospective Multicenter Study. *Diagn Basel Switz*. 17 nov 2020;10(11).
24. Zhu J, Chen X, Yang B, Bi N, Zhang T, Men K, et al. Evaluation of Automatic Segmentation Model With Dosimetric Metrics for Radiotherapy of Esophageal Cancer. *Front Oncol*. 2020;10:564737.
25. Sultana S, Robinson A, Song DY, Lee J. Automatic multi-organ segmentation in computed tomography images using hierarchical convolutional neural network. *J Med Imaging Bellingham Wash*. sept 2020;7(5):055001.
26. Chi W, Ma L, Wu J, Chen M, Lu W, Gu X. Deep learning based medical image segmentation with limited labels. *Phys Med Biol*. 21 oct 2020;
27. Jalalifar A, Soliman H, Sahgal A, Sadeghi-Naini A. A Cascaded Deep-Learning Framework for Segmentation of Metastatic Brain Tumors Before and After Stereotactic Radiation Therapy. *Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Int Conf*. juill 2020;2020:1063-6.

28. Chen Y, Xing L, Yu L, Bagshaw HP, Buyyounouski MK, Han B. Automatic intraprostatic lesion segmentation in multiparametric magnetic resonance images with proposed multiple branch UNet. *Med Phys.* déc 2020;47(12):6421-9.
29. Choi MS, Choi BS, Chung SY, Kim N, Chun J, Kim YB, et al. Clinical evaluation of atlas- and deep learning-based automatic segmentation of multiple organs and clinical target volumes for breast cancer. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* déc 2020;153:139-45.
30. Rhee DJ, Jhingran A, Rigaud B, Netherton T, Cardenas CE, Zhang L, et al. Automatic contouring system for cervical cancer using convolutional neural networks. *Med Phys.* nov 2020;47(11):5648-58.
31. Xue X, Qin N, Hao X, Shi J, Wu A, An H, et al. Sequential and Iterative Auto-Segmentation of High-Risk Clinical Target Volume for Radiotherapy of Nasopharyngeal Carcinoma in Planning CT Images. *Front Oncol.* 2020;10:1134.
32. Ke L, Deng Y, Xia W, Qiang M, Chen X, Liu K, et al. Development of a self-constrained 3D DenseNet model in automatic detection and segmentation of nasopharyngeal carcinoma using magnetic resonance images. *Oral Oncol.* nov 2020;110:104862.
33. Ermiş E, Jungo A, Poel R, Blatti-Moreno M, Meier R, Knecht U, et al. Fully automated brain resection cavity delineation for radiation target volume definition in glioblastoma patients using deep learning. *Radiat Oncol Lond Engl.* 6 mai 2020;15(1):100.
34. Bousabarah K, Ruge M, Brand J-S, Hoevels M, Rueß D, Borggreffe J, et al. Deep convolutional neural networks for automated segmentation of brain metastases trained on clinical data. *Radiat Oncol Lond Engl.* 20 avr 2020;15(1):87.
35. Huang Y-J, Dou Q, Wang Z-X, Liu L-Z, Jin Y, Li C-F, et al. 3-D RoI-Aware U-Net for Accurate and Efficient Colorectal Tumor Segmentation. *IEEE Trans Cybern.* 1 avr 2020;
36. Liang Y, Schott D, Zhang Y, Wang Z, Nasief H, Paulson E, et al. Auto-segmentation of pancreatic tumor in multi-parametric MRI using deep convolutional neural networks. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* avr 2020;145:193-200.
37. Song Y, Hu J, Wu Q, Xu F, Nie S, Zhao Y, et al. Automatic delineation of the clinical target volume and organs at risk by deep learning for rectal cancer postoperative radiotherapy. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* avr 2020;145:186-92.
38. Liu Z, Liu X, Xiao B, Wang S, Miao Z, Sun Y, et al. Segmentation of organs-at-risk in cervical cancer CT images with a convolutional neural network. *Phys Medica PM Int J Devoted Appl Phys Med Biol Off J Ital Assoc Biomed Phys AIFB.* janv 2020;69:184-91.
39. Xue J, Wang B, Ming Y, Liu X, Jiang Z, Wang C, et al. Deep learning-based detection and segmentation-assisted management of brain metastases. *Neuro-Oncol.* 15 avr 2020;22(4):505-14.
40. Wong J, Fong A, McVicar N, Smith S, Giambattista J, Wells D, et al. Comparing deep learning-based auto-segmentation of organs at risk and clinical target volumes to expert inter-observer variability in radiotherapy planning. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* mars 2020;144:152-8.

41. van Dijk LV, Van den Bosch L, Aljabar P, Peressutti D, Both S, J H M Steenbakkers R, et al. Improving automatic delineation for head and neck organs at risk by Deep Learning Contouring. *Radiother Oncol J Eur Soc Ther Radiol Oncol*. janv 2020;142:115-23.
42. Elguindi S, Zelefsky MJ, Jiang J, Veeraraghavan H, Deasy JO, Hunt MA, et al. Deep learning-based auto-segmentation of targets and organs-at-risk for magnetic resonance imaging only planning of prostate radiotherapy. *Phys Imaging Radiat Oncol*. oct 2019;12:80-6.
43. Vaassen F, Hazelaar C, Vaniqui A, Gooding M, van der Heyden B, Canters R, et al. Evaluation of measures for assessing time-saving of automatic organ-at-risk segmentation in radiotherapy. *Phys Imaging Radiat Oncol*. janv 2020;13:1-6.
44. Ahn SH, Yeo AU, Kim KH, Kim C, Goh Y, Cho S, et al. Comparative clinical evaluation of atlas and deep-learning-based auto-segmentation of organ structures in liver cancer. *Radiat Oncol Lond Engl*. 27 nov 2019;14(1):213.
45. Li S, Xiao J, He L, Peng X, Yuan X. The Tumor Target Segmentation of Nasopharyngeal Cancer in CT Images Based on Deep Learning Methods. *Technol Cancer Res Treat*. déc 2019;18:1533033819884561.
46. Dong X, Lei Y, Tian S, Wang T, Patel P, Curran WJ, et al. Synthetic MRI-aided multi-organ segmentation on male pelvic CT using cycle consistent deep attention network. *Radiother Oncol J Eur Soc Ther Radiol Oncol*. déc 2019;141:192-9.
47. Zhong T, Huang X, Tang F, Liang S, Deng X, Zhang Y. Boosting-based Cascaded Convolutional Neural Networks for the Segmentation of CT Organs-at-risk in Nasopharyngeal Carcinoma. *Med Phys*. 16 sept 2019;
48. Rhee DJ, Cardenas CE, Elhalawani H, McCarroll R, Zhang L, Yang J, et al. Automatic detection of contouring errors using convolutional neural networks. *Med Phys*. nov 2019;46(11):5086-97.
49. Schreier J, Attanasi F, Laaksonen H. A Full-Image Deep Segmenter for CT Images in Breast Cancer Radiotherapy Treatment. *Front Oncol*. 2019;9:677.
50. Chan JW, Kearney V, Haaf S, Wu S, Bogdanov M, Reddick M, et al. A convolutional neural network algorithm for automatic segmentation of head and neck organs at risk using deep lifelong learning. *Med Phys*. mai 2019;46(5):2204-13.
51. van der Heyden B, Wohlfahrt P, Eekers DBP, Richter C, Terhaag K, Troost EGC, et al. Dual-energy CT for automatic organs-at-risk segmentation in brain-tumor patients using a multi-atlas and deep-learning approach. *Sci Rep*. 11 mars 2019;9(1):4126.
52. Tappeiner E, Pröll S, Hönig M, Raudaschl PF, Zaffino P, Spadea MF, et al. Multi-organ segmentation of the head and neck area: an efficient hierarchical neural networks approach. *Int J Comput Assist Radiol Surg*. mai 2019;14(5):745-54.
53. Feng X, Qing K, Tustison NJ, Meyer CH, Chen Q. Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images. *Med Phys*. mai 2019;46(5):2169-80.
54. Dong X, Lei Y, Wang T, Thomas M, Tang L, Curran WJ, et al. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. *Med Phys*. mai 2019;46(5):2157-68.

55. Trullo R, Petitjean C, Dubray B, Ruan S. Multiorgan segmentation using distance-aware adversarial networks. *J Med Imaging Bellingham Wash.* janv 2019;6(1):014001.
56. Fu Y, Mazur TR, Wu X, Liu S, Chang X, Lu Y, et al. A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy. *Med Phys.* nov 2018;45(11):5129-37.
57. Yang J, Veeraraghavan H, Armato SG, Farahani K, Kirby JS, Kalpathy-Kramer J, et al. Autosegmentation for thoracic radiation treatment planning: A grand challenge at AAPM 2017. *Med Phys.* oct 2018;45(10):4568-81.
58. Tong N, Gou S, Yang S, Ruan D, Sheng K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. *Med Phys.* oct 2018;45(10):4558-67.
59. Men K, Zhang T, Chen X, Chen B, Tang Y, Wang S, et al. Fully automatic and robust segmentation of the clinical target volume for radiotherapy of breast cancer using big data and deep learning. *Phys Medica PM Int J Devoted Appl Phys Med Biol Off J Ital Assoc Biomed Phys AIFB.* juin 2018;50:13-9.
60. Wang J, Lu J, Qin G, Shen L, Sun Y, Ying H, et al. Technical Note: A deep learning-based autosegmentation of rectal tumors in MR images. *Med Phys.* juin 2018;45(6):2560-4.
61. Charron O, Lallement A, Jarnet D, Noblet V, Clavier J-B, Meyer P. Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network. *Comput Biol Med.* 1 avr 2018;95:43-54.
62. Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiother Oncol J Eur Soc Ther Radiol Oncol.* févr 2018;126(2):312-7.
63. Men K, Chen X, Zhang Y, Zhang T, Dai J, Yi J, et al. Deep Deconvolutional Neural Network for Target Segmentation of Nasopharyngeal Cancer in Planning Computed Tomography Images. *Front Oncol.* 2017;7:315.
64. Men K, Dai J, Li Y. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. *Med Phys.* déc 2017;44(12):6377-89.

		Centre Léon Bérard (Lyon, France)						HEGP (Paris, France)	Clinique Pasteur (Toulouse)
		Male pelvis	Thorax	Breast	Head and neck	Digestive	Brain	Breast	Male pelvis
OAR/Target volumes	Number of OAR automatically contoured by the model/accepted by the doctors (%)	10/100%	12/67%	15/100%	33/85%	8/100%	14/100%	6/100%	4/100%
	Number of target volumes automatically contoured by the model/retained (%)	3/100%	0/NA	12/100%	14/0%	0/NA	0/NA	12/100%	3/100%
Model building and evolutivity	Number of data for model training	1324	information not received	661	1655	1324	1655	> 200	About 100 patients data
	Continuous evolutivity of the model based on local data (yes/no)	No	No	No	No	No	No	No	No
Commissioning	Number of patients	5	5	5	150	Used directly in clinical routine (> 300 patients since implementation)	Used directly in clinical routine (> 250 patients since implementation)	10	250
	Metrics for performances evaluation	Jl (Bladder, Rectum, Femoral heads, whole Patient)	Jl (Lungs, Heart, Trachea, Medullary canal, whole Patient)	Jl (Lungs, Heart, Trachea, Medullary canal, whole Patient)	VDSC, Subjective rating, Time for correction	Jl (Lungs, Spinal Cord, Esophagus)	-	VDSC, Subjective rating, Time for correction	VDSC, Sensitivity, Specificity, Coordinates of the barycenter
Workflow description - Method of interfacing to scanner and TPS		Automatic export and segmentation	Automatic export and segmentation	Automatic export and segmentation	Automatic export and segmentation	Automatic export and segmentation	Automatic export and segmentation	DICOM push from the TPS or CT device, automatic segmentation - application of the	In progress

							right model via a DICOM Tag, application of post-processing, automatic export of the contours to the TPS	
Patient-specific QA	-	-	-	-	-	-	-	In progress
Regular QA	-	-	-	-	-	-	Definition a test set for performance evaluation in case of major changes in the workflow	In progress

Reference	Organs of interest	Imaging modality	Tumor location	Number of patients	Objectives of the study	Metrics used for performance evaluation	Monocentric/Multicentric study
Shi <i>et al.</i> (13)	CTV	CT	Cervix	462	Network refinement	DSC, comparison of performances to those of 3 experts	Multicentric validation
Jenkins <i>et al.</i> (14)	CTV	?	Prostate	232	Analysis of the correlations between contour variations and treatment outcome	Local contour deviation	Monocentric evaluation
Liu <i>et al.</i> (15)	OAR	CT	Head and neck	220	New methodology based on unlabelled data	DSC	Multicentric validation
Cha <i>et al.</i> (5)	OAR + target volumes	MRI	Prostate	173	Clinical implementation	VDSC, SDSC, APL, time for corrections, rates on a 3-point protocol deviation scale	Monocentric evaluation
Hague <i>et al.</i> (16)	OAR	MRI	Head and neck	Training: 100 MR data/ 621 CT data - 38 patients included in the test set in total	Comparison of CT-based and MR-based OAR auto-contouring	DSC, DTA	Multicentric data
Chung <i>et al.</i> (17)	OAR + target volumes	CT	Breast	111	Evaluation of performances of an house network	DSC, HD, qualitative scoring by 2 panels from 10 institutions, interobserver variability, delineation time, dosimetric impact	Monomachine data
Kim <i>et al.</i> (18)	OAR	CT	Head and neck	100	Evaluation of performances of an house network for adaptive RT	DSC, FPD, FND, HD, MSD, Turing test, contouring time	Monocentric evaluation
Kieselmann <i>et al.</i> (19)	Parotid glands	MRI	Head and neck	229	Development of a two stage network for auto-contouring of MRI data from annotated CT data only	DSC, HD, MSD	Monocentric evaluation
Zhang <i>et al.</i> (20)	OAR	CT	Head and neck	170 (20 in the test set)	Evaluation of the performances of a two-step segmentation model including a slice-classification step	DSC and HD	Monocentric study
Brouwer <i>et al.</i> (6)	OAR	CT	Head and neck	103	Evaluation of the manual adjustment of auto-contouring in routine practice	Range adjustments projected on a reference shape for each OAR	Monocentric evaluation
Brunenberg <i>et al.</i> (21)	OAR	CT	Head and neck	58 patients (independent test set)	Evaluation of performances of a commercially available network on an independent test set	DSC and HD, Turing test	Multicentric validation

Oktay <i>et al.</i> (22)	OAR	CT	Prostate, Head and neck cancer	519 (pelvis) + 242 (Head and neck)	Multicentric evaluation of the performances of a modified U-Net	DSC, Surface distance measures, delineation time	Muticentric validation
Kiljunen <i>et al.</i> (23)	OAR + target volumes	CT	Prostate	30	Retrospective evaluation of accuracy and efficiency gain	DSC, SDSC, HD, differences in absolute and relative volumes, delineation time	Muticentric evaluation
Zhu <i>et al.</i> (24)	OAR	CT	esophagus	19	Evaluation of the dosimetric impact of autosegmentation on VMAT treatments	DSC, MDA, Dosimetric evaluation	Monocentric evaluation
Sultana <i>et al.</i> (25)	OAR	CT	Prostate, Head and neck cancer	58 (Head and neck), 15 (pelvis) - Test cohorts	Proposition of a new segmentation strategy: two-step hierarchical CNN	DSC	Muticentric evaluation
Chi <i>et al.</i> (26)	OAR	CT	Head and neck	29 (test cohort)	Proposition of a weakly supervised DL training approach	DSC, HD, ASD	Muticentric evaluation
Jalalifar <i>et al.</i> (27)	Target volume (metastasis)	MRI	Brain	106	Evaluation of a cascaded Deep-Learning Framework	DSC	?
Chen <i>et al.</i> (28)	Subtarget regions	MRI	Prostate	136	Evaluate the performances of DL for subvolume dose escalation	DSC	Monocentric study
Choi <i>et al.</i> (29)	OAR + target volumes	CT	Breast	76 including 14 (+14 other patients) patients in the test set	Comparison of atlas-based and DL-based segmentations	DSC and HD	Monocentric study
Rhee <i>et al.</i> (30)	OAR + target volumes	CT	Cervix	> 2600	Evaluation of an in-house network	DSC, mean surface, HD and expert rating	Multicentric data
Xue <i>et al.</i> (31)	Target volumes	CT	Head and neck	150	Evaluation of the performances of a sequential and iterative U-Net	DSC, JI, ASD, HD	Monocentric study
Ke <i>et al.</i> (32)	Target volumes	MRI	Head and neck	> 4000	To propose a dual task network for diagnosis and segmentation	DSC	Monocentric study
Chen <i>et al.</i> (11)	Target volume	CT	Breast	680	Proposition of a QA tool	Balanced accuracy, F- score, ROC-AUC	Monocentric study
Savenije <i>et al.</i> (7)	OAR	MRI	Prostate	150 (53 included in final evalation)	Investigation of the feasibility of clinical use of DL-based automatic OAR delineation on MR in a clinical workflow	DSC, HD, and mean distances	Monocentric study



Ermis <i>et al.</i> (33)	Resection cavity	MRI	Brain	30	Application of DL on brain post-operative MR images	DSC, volume	Monocentric study
Boursabarah <i>et al.</i> (34)	Target volumes (metastasis)	MRI	Brain	509 (40 patients in the test set)	Evaluation of the performance of DL in automatic segmentation of brain metastases	sensitivity, specificity, average false positive rate, DSC, Bland-Altman analysis and concordance correlation coefficient	Monocentric study
Huang <i>et al.</i> (35)	Target volume	MRI	Colorectal	64	Proposition of a novel encoder-decoder-based framework	DSC	?
Liang <i>et al.</i> (36)	Target volume	MRI	Pancreas	40 (13 in the test set)	Proposition of an automatic segmentation method of the pancreatic GTV based on multi-parametric MRI	DSC, HD, MSD	Monocentric study
Song <i>et al.</i> (37)	OAR + target volumes	CT	Rectal cancer	199	Evaluation of the performance of DL for rectum cancer	Objective grading, time for manual correction	Monocentric study
Liu <i>et al.</i> (38)	OAR	CT	Cervix	105	Evaluation of the performance of DL for cervix cancer	DSC and HD	Monocentric study
Xue <i>et al.</i> (39)	Target volumes (metastasis)	MRI	Brain	1652 (451 in the test set)	Evaluation of the performances of an automatic DL-based detection and segmentation method for brain metastasis	Sensitivity, specificity, and dice ratio	Multicentric study
Wong <i>et al.</i> (40)	OAR + target volumes	CT	Head and neck, prostate, brain	60	Comparison of DL contouring to interobserver variability	DSC and HD	Monocentric study
Van Dijk <i>et al.</i> (41)	OAR	CT	Head and neck	693 (104 in the test set)	Comparison of atlas and DL-based segmentations	DSC, absolute mean and max dose differences, contouring time, Turing test	Monocentric study
Elguindi <i>et al.</i> (42)	OAR + target volumes	MRI	Prostate	100 (50 in the test set)	To evaluate the performances of a MRI-based CNN network for MR-only RT	VDSC and SDSC	Monocentric study
Vaassen <i>et al.</i> (43)	OAR	CT	Lung	20	Analysis of the correlations between quantitative metrics and time saving	SDSC, APL, VDSC, HD, time saving	Monocentric study
Ahn <i>et al.</i> (44)	OAR	CT	Liver	70	Comparison of atlas and DL-based segmentations	HD, DSC, VOE, and relative volume difference	Monocentric study

Li <i>et al.</i> (45)	Target volume	CT	Head and neck	502 (100 in the test set)	Evaluation of performances of an in house model on head and neck data	HD, DSC, VOE, and relative volume difference	Monocentric study
Dong <i>et al.</i> (46)	OAR	CT	Prostate	140	Proposition of a new strategy: generation of sMRI before multi-organ segmentation	DSC and MSD	Monocentric study
Zhong <i>et al.</i> (47)	OAR + target volumes	CT	Head and neck	100 (20 in the test set)	Proposition of a cascaded network structure	DSC, HD, and VOE	Monocentric study
Rhee <i>et al.</i> (48)	OAR	CT	Head and neck	> 3500 (174 + 24 in the test cohort)	Proposition of a QA tool for a multi-atlas auto-contouring strategy used in clinic	DSC, HD, detection of clinically unacceptable contours	Multicentric study
Schreier <i>et al.</i> (49)	3 organs (breasts + heart)	CT	Breast	251 (52 in the test set) + 64 (independent test cohort)	Proposition of a new architecture	Average surface distance, HD, correction time	Multicentric study
Chan <i>et al.</i> (50)	OAR	CT	Head and neck	200 (20 in the test set)	Comparison of network architectures	DSC and RMSE	Monocentric study
Van der Heyden <i>et al.</i> (51)	OAR	DECT	Brain	14	Investigation of the performances of multi-atlas and DL auto contouring methods on DECT images	DSC, HD, center of mass displacement ( $\Delta\text{CoM}$ ), scoring by experts	Monocentric study
Tappeiner <i>et al.</i> (52)	OAR	CT	Head and neck	40	Proposition of a new DL strategy	DSC and HD	Challenge data - multicentric data
Feng <i>et al.</i> (53)	OAR	CT	Lung	90	Development of a novel DCNN method for thoracic OAR segmentation using cropped 3D images	DSC, MSD and HD, correction time	Challenge data - multicentric data + internal cohort
Dong <i>et al.</i> (54)	OAR	CT	Lung	35	Proposition of a new DL strategy	DSC, MSD, dosimetric analysis	Challenge data
Trullo <i>et al.</i> (55)	OAR	CT	Lung	60	Proposition of a new DL strategy including a global localization information	DSC	Challenge data
Fu <i>et al.</i> (56)	OAR	MRI	Abdomen/Pelvis	120 (100 in the test set)	Proposition of a new strategy for MR-based auto-contouring	DSC and HD	Monocentric study
Yang <i>et al.</i> (57)	OAR	CT	Thorax	60 (12 in the test set)	Thoracic Auto-Segmentation Challenge	DSC, HD, and MSD	Challenge data

Tong <i>et al.</i> (58)	OAR	CT	Head and neck	32 (no test set)	Development of a novel method that combines a fully CNN with a shape representation model	DSC, positive predictive value, sensitivity , ASD, and HD	Public data - multicentric
Men <i>et al.</i> (59)	Target volumes	CT	Breast	800	Evaluation of a deep dilated residual network (DD-ResNet)	DSC and HD	Monocentric data
Wang <i>et al.</i> (60)	Target volumes	MRI	Rectal cancer	93	Development of a DL-based autosegmentation algorithm	HD, ASD, DSC, and JI	Monocentric data
Charron <i>et al.</i> (61)	Target volumes (metastasis)	MRI	Brain	182 (18 in the test set)	Adaptation of a CNN for brain lesion segmentation	Sensitivity, false positives, DSC	Monocentric data
Lustberg <i>et al.</i> (62)	OAR	CT	Lung	20 (test set only)	Comparison of atlas and DL-based segmentations	DSC, HD, quality scoring, time saving	Monocentric data
Men <i>et al.</i> (63)	Target volumes	CT	Head and neck	230 (46 in the test set)	Development of a DL-based autosegmentation algorithm	DSC, HD	Monocentric data
Men <i>et al.</i> (64)	OAR + target volumes	CT	Rectal cancer	278 (60 for validation)	Proposition of a deep dilated CNN	DSC	Monocentric data

### Legends:

#### Table 1

Summary table of the properties of the CE-marked automatic contouring software deployed in the 3 radiotherapy departments and description of the methodology carried out for the clinical implementation of these algorithms.

#### Table 2

Studies on Deep-Learning-based auto contouring. The literature search was performed by one reader (see text). Pubmed was the only database queried. The literature search was performed up to April 2021. When available, information were extracted from the abstract of the article.

CTV: Clinical Target Volume, OAR: Organ At Risk, CT : Computed Tomography, MRI: Magnetic Resonance Imaging, DSC: Dice similarity coefficient, VDSC: Volumetric Dice similarity coefficient, SDSC: Surface Dice similarity coefficient, APL: Added Path Length, DTA: Distance To Agreement, FPD: False-Positive DSC, FND: False-Negative DSC, HD: Hausdorff Distance, MSC: Mean Surface Distance, MDA: Mean Distance to Agreement, ASD: Average Surface Distance, JI: Jaccard Index, ROC-AUC: Area Under the Receiving Operator Characteristic Curve, VOE: Volume Overlap Error, RMSE: Root Mean Square Error