



# Key Success Factors for Successful Implementation of AI Based Segmentation Algorithms in Clinical Radiology Practice

**Benoit Charmettant<sup>1</sup> and Nathalie Lassau<sup>1,2\*</sup>**

<sup>1</sup>UMR1281. Université Paris-Saclay, Inserm, CNRS, CEA, Laboratoire d'Imagerie Biomédicale Multimodale Paris-Saclay, France

<sup>2</sup>Département d'imagerie, Institut Gustave Roussy, France

**\*Corresponding author:** Nathalie Lassau, Département d'imagerie, Institut Gustave Roussy, Gustave Roussy Cancer Campus Grand Paris, 39 rue Camille Desmoulins, Villejuif 94800, France.

**To Cite This Article:** Nathalie Lassau, Key Success Factors for Successful Implementation of AI Based Segmentation Algorithms in Clinical Radiology Practice. 2020 - 7(3). AJBSR.MS.ID.001143. DOI: [10.34297/AJBSR.2020.07.001143](https://doi.org/10.34297/AJBSR.2020.07.001143).

**Received:** 📅 January 24, 2020; **Published:** 📅 February 14, 2020

## Introduction

With the rise of large amounts of patient's clinical and imaging data, the development of artificial intelligence tools based on machine learning and deep learning capable of performing several tasks such as image classification or regression, organ segmentation or feature extraction, has soared over the past few years [1]. These developments create many opportunities for radiologists and are likely to impact their routine practice in the long run by providing tools that will improve the accuracy and efficiency of diagnosis and prognosis. It will arguably allow radiologists to spend more time on complex problem solving by rebarbative tasks and help grasping more useful information from medical images [2]. Despite a great research interest, many challenges are getting in the way of an efficient, safe and ethical implementation of those tools in radiologists' daily practice [3]. Nevertheless, not all tasks and modalities of medical AI have reached the same level of maturity nor are they developing at the same pace.

## Overview of Segmentation Tools

Image segmentation is a pixel-wise classification of an image to located specific objects (organs, substructures or lesions) which allows, between other things, quantitative analysis of volume and shape. Segmentation is a very time-consuming procedure, an expert can spend up to four hours for the segmentation of a single case in a complex location such as head and neck [4]. It is also a task that is dependent on the level of expertise of the operator and suffers from a high inter and intra operator variability [5]. In some contexts such as radiotherapy, tumor segmentation can be of critical importance to limit the irradiation of normal tissues. Segmentation can also be

a first step for other image analysis such as local texture feature extraction. For all these reasons, radiologists and patients would benefit a lot from a reliable and automated segmentation tool. This might explain the great interest shown by the research community towards solving this particular problem. Litjens et al. have conducted a review of the major deep learning publication relevant to medical imaging [6]. They have shown that segmentation is the most studied application of deep learning before object detection and image classification. Performances of automated segmentation have been greatly improved by the introduction of the U-Net architecture which won by a large margin the 2015 ISBI cell tracking challenge with a dice score of 0.92.

## Comparison of Segmentation Studies for Head and Neck Structures

We will now go on with a comparison of three studies that have been published concerning the development of automated segmentation tools for head and neck structures. Head and neck organ segmentation is a particularly complex task because of the size and the number of different structures of interest localized in this region.

In a first study, Chan et al. have introduced a neural network based on U-Net and a training methodology capable of reaching dice score up to 0.91 for different organs at risk in the head and neck region [7]. They claimed their methodology surpasses all alternative algorithms. However, the study was conducted on a limited number of patients (around 180) coming from one center and validated on 20 patients from the same database. The ground

truth segmentation used as a reference during the training of the algorithm was performed only by one radiologist and no comparison of the results with experts has been conducted.

A second study illustrated another attempt of automated segmentation of 16 head and neck organs and resulted in a very high dice score of 0.98 for the brain (down to 0.65 for the cochlea) [8]. Compared to the previous study, the advantage of this work is to use a much larger training dataset of around 3495 patients for training and to test the performances on an external multicentric dataset. However the ground truth segmentation was performed by various anonymous operators with no review by an expert and the training exams all came from the same center, which reduces the robustness of the algorithm and increases bias.

The final study considered here brings interesting perspectives on the effort that has been made to reduce the ground truth segmentation's bias [5]. The reference segmentation was decided by consensus between two trained expert radiologists (third expertise was requested in case of disagreement). Another point is made regarding the analysis of the performance of the algorithm. The predicted segmentations were compared with a panel of 8 radiologists who judged that almost 90% of those were satisfying. They also quantitatively analyzed the increase of performance of radiologists assisted by the algorithm and concluded that it reduces intra operator variability by 36%, inter-operator variability by 55% and reduce time dedicated to this task by 40%.

### Key Success Factors for Efficient Transfer from Research to Clinical Practice

A few factors can be highlighted to ensure the reproducibility of results as well as effective and safe transfer to the clinical practice of automated segmentation algorithms. First, it must be noted that current deep learning methods (such as the different versions of the U-Net) have the potential to reach very high dice scores, and will at some point only reproduce the bias of the training database

and ground truth segmentation if those are not built with enough care. Therefore it is essential to promote the development of large qualitative databases with a great variety of cases (multicentric, different operators and imaging machines) to ensure generalization of the results. A special effort should be made for the definition of the ground truth segmentation to avoid any human bias. This can be done by using a consensus from two experts or implementing a systematic and independent review by an expert. Finally, it is important but to also conduct a review of the performances by human experts considering predefined criteria and not only one or two metrics (dice score, Hausdorff distance), which tend to be less meaningful as the algorithm gets better.

### References

1. T Syeda Mahmood (2018) Role of big data and machine learning in diagnostic decision support in radiology. *J Am Coll Radiol* 15(3): 569-576.
2. European Society of Radiology (ESR) (2019) What the radiologist should know about artificial intelligence - an ESR white paper. *Insights into Imaging* 10(1): 44.
3. JH Thrall, Li X, Li Q, Cruz C, Do S, et al. (2018) Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. *J Am Coll Radiol* 15(3): 504-508.
4. PM Harari, S Song, WA Tomé (2010) Emphasizing conformal avoidance versus target definition for imrt planning in head-and-neck cancer. *Int J Radiat Oncol Biol Phys* 77(3): 950-958.
5. L Lin, Dou Q, Jin YM, Zhou GQ, Tang YQ, et al. (2019) Deep learning for automated contouring of primary tumor volumes by mri for nasopharyngeal carcinoma. *Radiology* 291(3): 677-686.
6. G Litjens, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, et al. (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42: 60-88.
7. JW Chan, Kearney V, Haaf S, Wu S, Bogdanov M, et al. (2019) A convolutional neural network algorithm for automatic segmentation of head and neck organs at risk using deep lifelong learning. *Med Phys* (46)5: 2204-2213.
8. DJ Rhee, Cardenas CE, Elhalawani H, McCarroll R, Zhang L, et al. (2019) Automatic detection of contouring errors using convolutional neural networks. *Med Phys* 46(11): 5086-5097.