SASRO 2022

Evaluation of different OARs automatic segmentation techniques for Left Breast Cancer

Joao Rodrigues

3 September 2022

Disclosure

I have no conflicts of interest within the framework of this presentation

About the speaker

- License in radiotherapy, 2001
- Radiation therapist (RTT) and dosimetrist, (2001-2010), Lisbon, Portugal
- Centre Hospitalier Universitaire Vaudois (CHUV), Lausanne, since 2010
- Master in Health management in 2010

Introduction

- Automatic contouring increasing in radiation oncology departments
- Potential to reduce the time required to contouring the organs at risk (OAR)
- Reduce subjectivity linked to the different users ⇨ more homogeneous contours
- An objective evaluation must be made to analyze the results of automatic contouring with respect to their accuracy

Automatic contouring solutions in RayStation

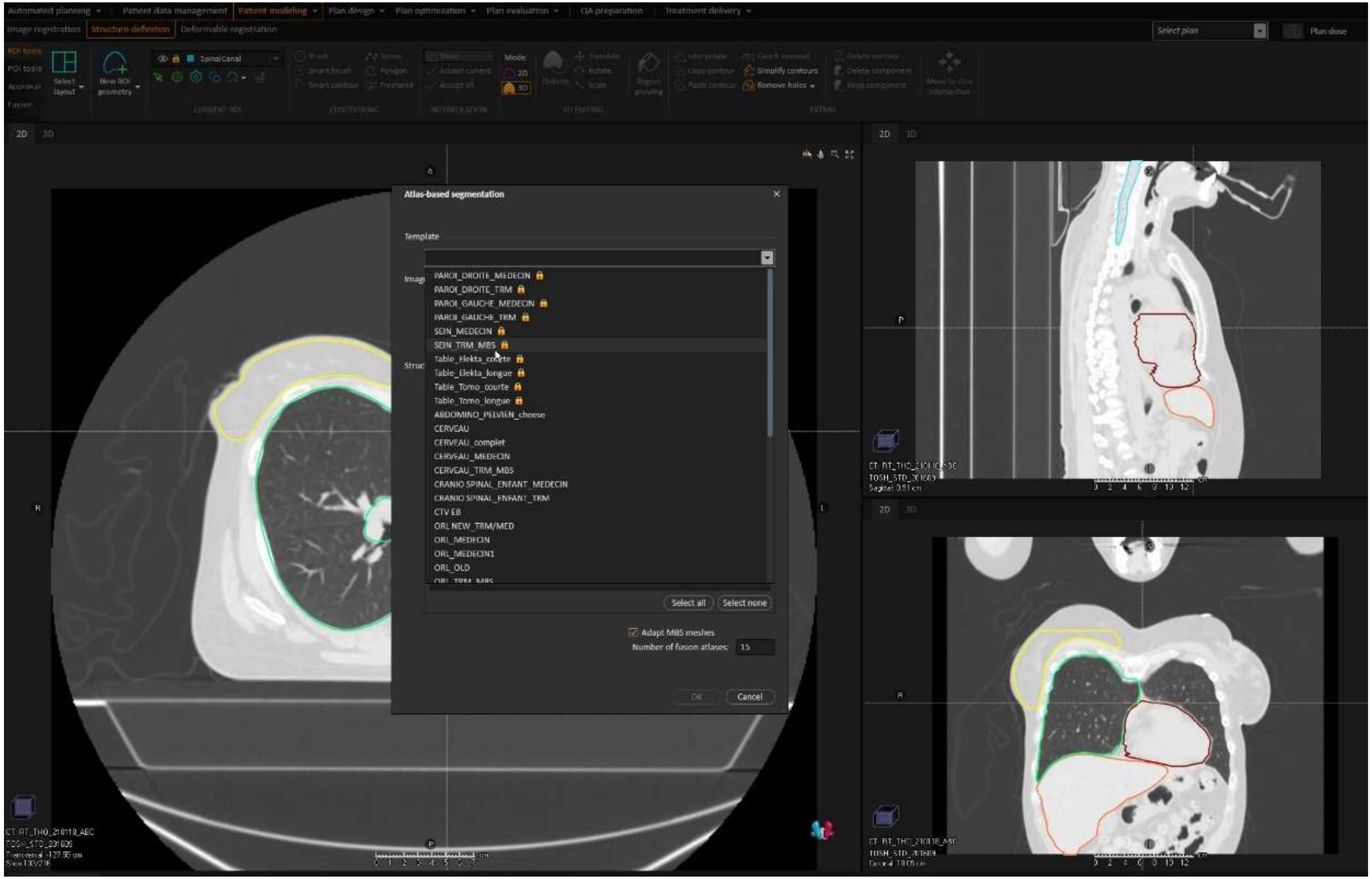
Atlas-Based Segmentation

- Templates with multiple image sets – atlases –
- Best matching atlases through rigid image registration and deformable registration
- The more fusion atlases, the longer the computation time

Deep Learning Segmentation

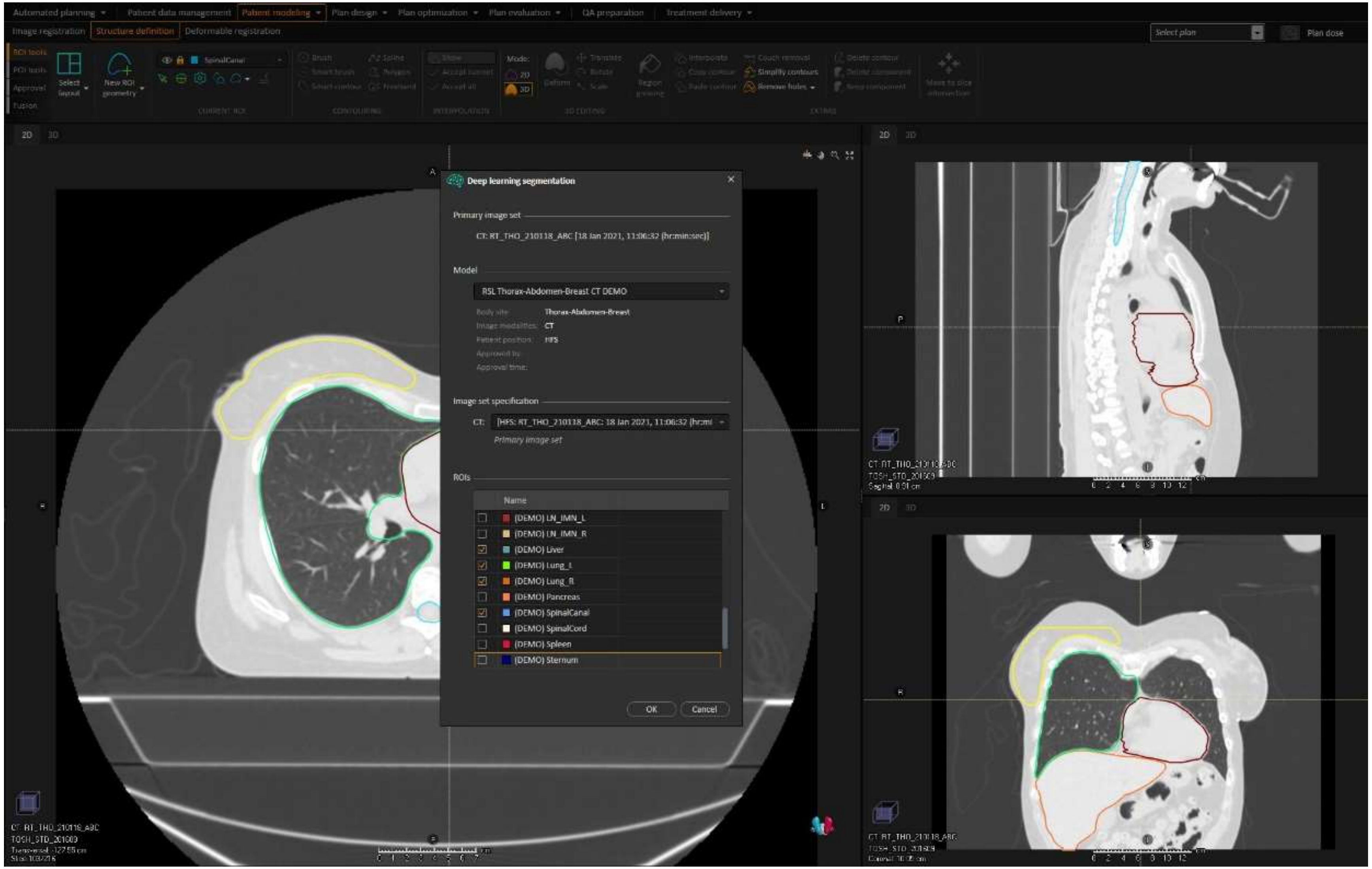
- Neural networks trained on a large number of previously segmented data sets.
- Optimization learning is required
- Garbage In…..Garbage out

Atlas-Based Segmentation



(10 - 15 min)

Deep Learning Segmentation



(1 min)

Materials and Methods

Similarity metrics

- Quantitative evaluation of automatic segmentation vs manual segmentation of OARs by means of the following geometrical metrics: Dice Similarity Coefficient (**DSC**); Overlap index (**OI**); Volume Difference (**Dv**)
- **DSC** is the most used metric in validating medical volume segmentations (2), evaluates the similarity of two delineations by comparing the overlap area (3)
- **OI** normalizes the size of the correctly automatic segmented region over the manual reference segmentation *(4)*
- **DV** measures the absolute size difference of the segmented regions, as a fraction of the size of the manual reference segmentation

Materials and Methods
(cont.)

- Random selection of 20 left breast cancer patients with contours done manually and reviewed by the medical expert. These contours will be the benchmark for comparison (the "ground truth")

  - $DSC = 2(Va \cap VM) / (Va + VM)$
  - $OI = (Va \cap VM) / VM$
  - $Dv = (Va - VM) / VM$

- The closer the **DSC** index and **OI** are to 1, and the closer the **Dv** index is to 0, the better the results of the automatic contouring are

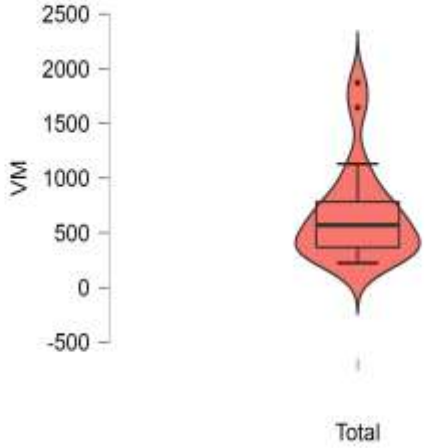
Results

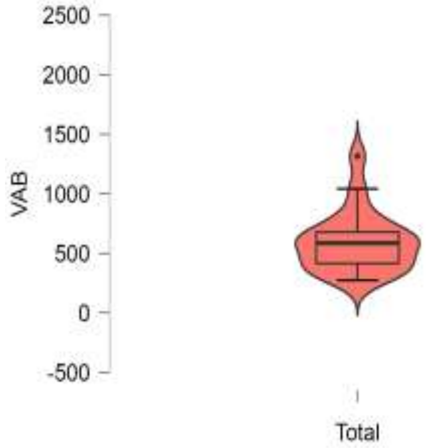
Contra-lateral breast volumes differences

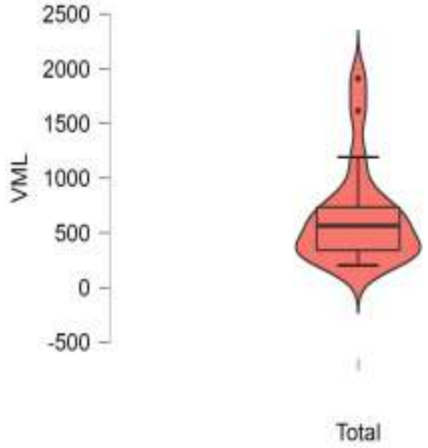
VM (manual reference segmentation)
VAB (Atlas Based segmentation)
VML (Deep Learning segmentation)

Descriptive Statistics

	VM	VAB	VML
Mean	666.4	593.8	648.1
Std. Deviation	447.3	258.9	456.2
Range	1647.8	1044.0	1706.2







Contra-lateral breast similarity

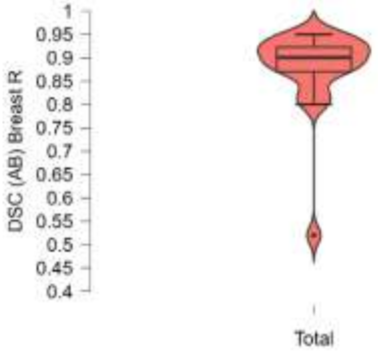
AB (Atlas Based segmentation)

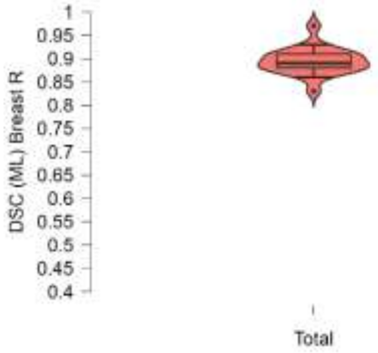
ML (Deep Learning segmentation)

Dice Similarity Coefficient (DSC)

Descriptive Statistics

	DSC (AB) Breast R	DSC (ML) Breast R
Mean	0.88	0.89
Std. Deviation	0.09	0.03
Range	0.43	0.14

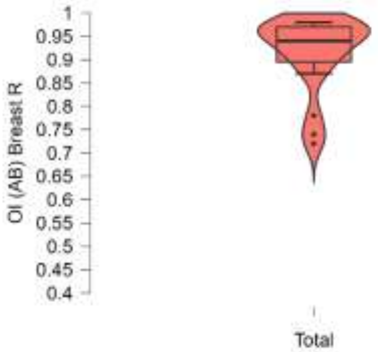


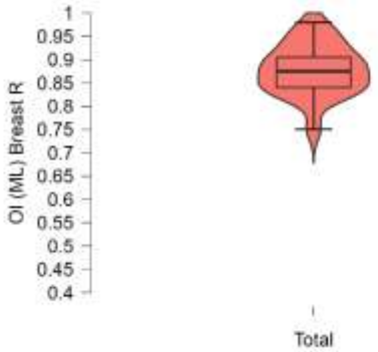


Overlap index (OI)

Descriptive Statistics

	OI (AB) Breast R	OI (ML) Breast R
Mean	0.89	0.88
Std. Deviation	0.15	0.06
Range	0.62	0.23

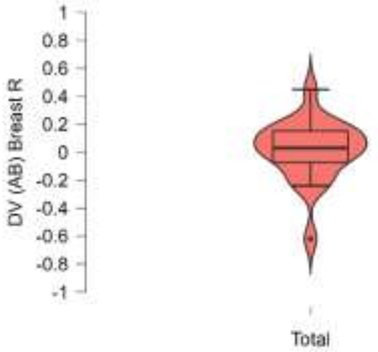


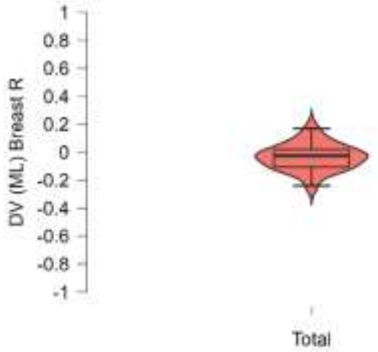


Volume Difference (DV)

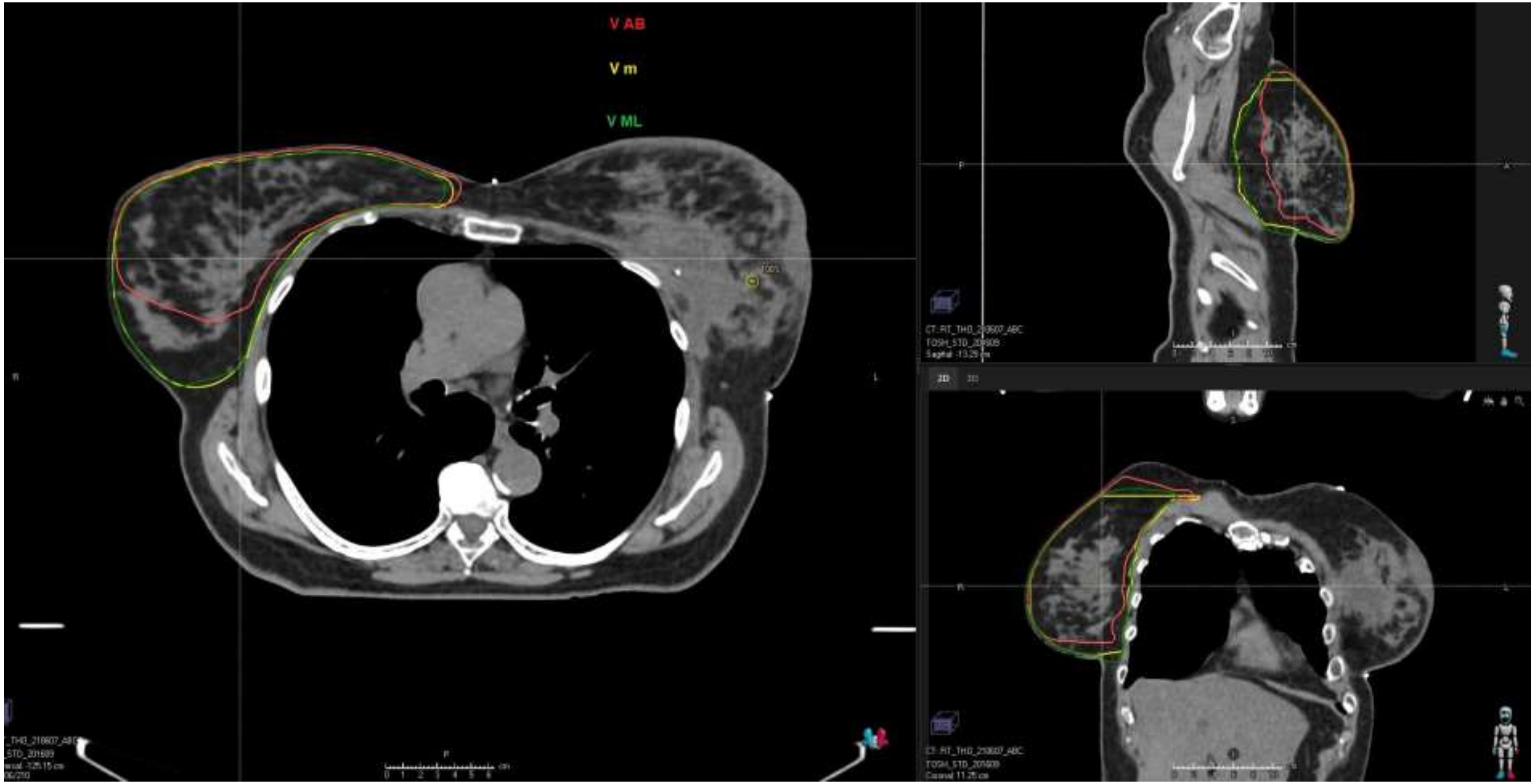
Descriptive Statistics

	DV (AB) Breast R	DV (ML) Breast R
Mean	0.01	−0.03
Std. Deviation	0.22	0.09
Range	1.07	0.41





Contralateral breast - Example

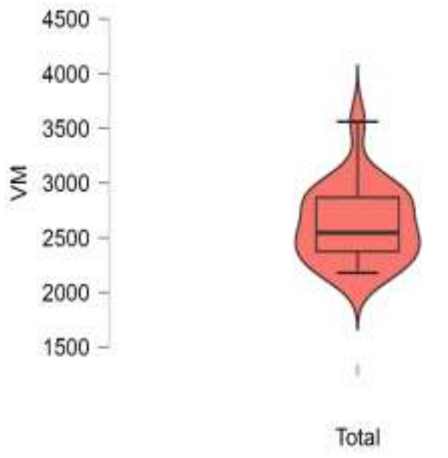


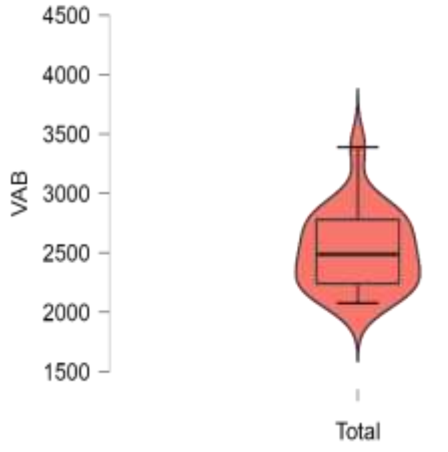
Lung R volumes differences

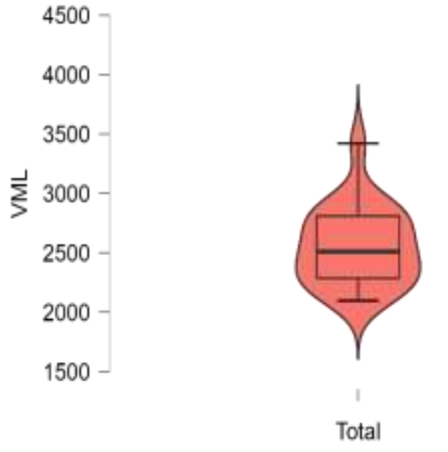
Descriptive Statistics

	VM	VAB	VML
Mean	2630.3	2522.4 P<.001	2552.6 P<.001
Std. Deviation	348.0	329.8	330.8
Range	1380.8	1313.5	1324.5

VM (manual reference segmentation)
VAB (Atlas Based segmentation)
VML (Deep Learning segmentation)







canton de vaud

Lung R similarity

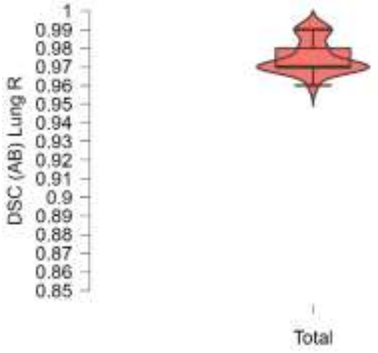
AB (Atlas Based segmentation)

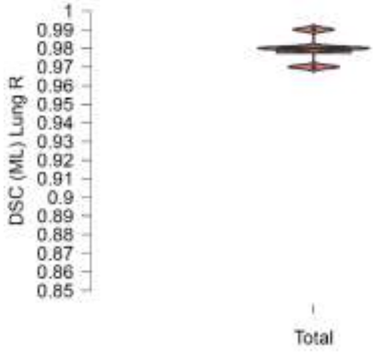
ML (Deep Learning segmentation)

Dice Similarity Coefficient (DSC)

Descriptive Statistics

	DSC (AB) Lung R		DSC (ML) Lung R
Mean	0.97	P<.001	0.98
Std. Deviation	8.89e−3		6.86e−3
Range	0.03		0.02

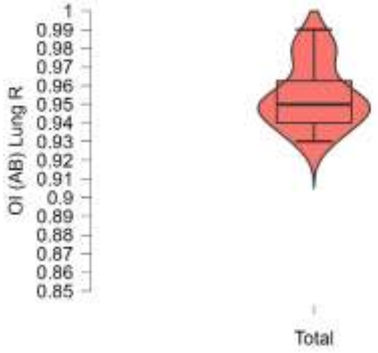


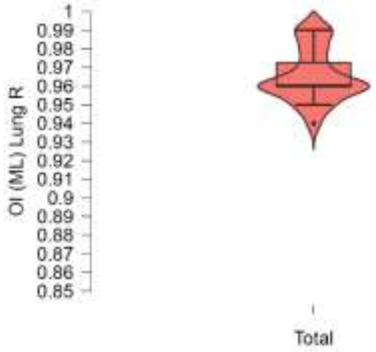


Overlap index (OI)

Descriptive Statistics

	OI (AB) Lung R		OI (ML) Lung R
Mean	0.95	P<.001	0.96
Std. Deviation	0.02		0.01
Range	0.06		0.05

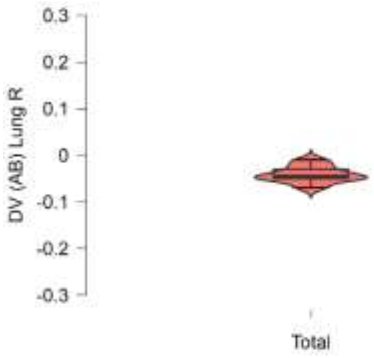


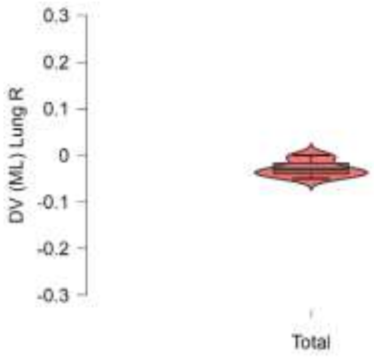


Volume Difference (DV)

Descriptive Statistics

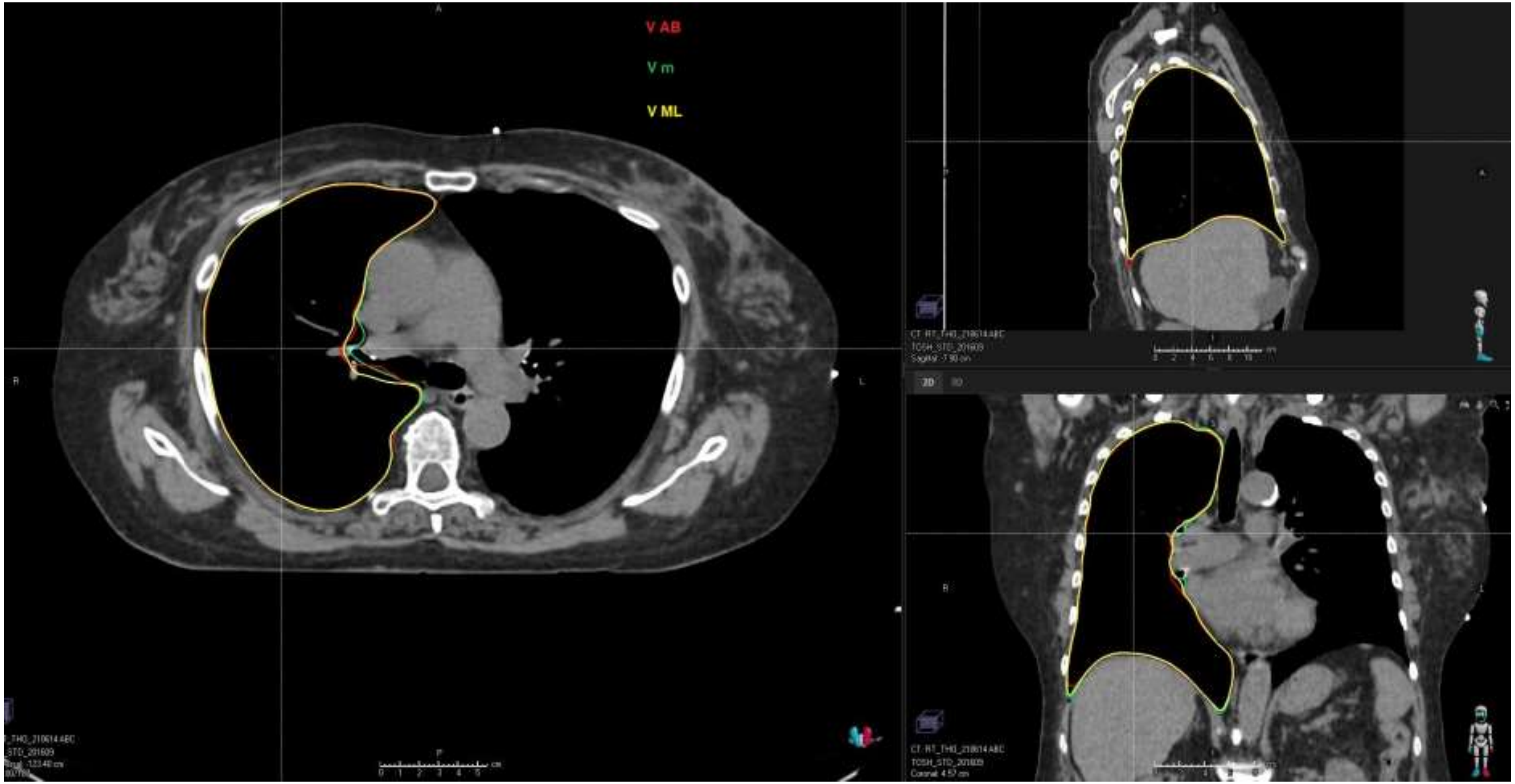
	DV (AB) Lung R		DV (ML) Lung R
Mean	−0.04	P<.001	−0.03
Std. Deviation	0.02		0.02
Range	0.06		0.05





canton de vaud

Lung R - Example

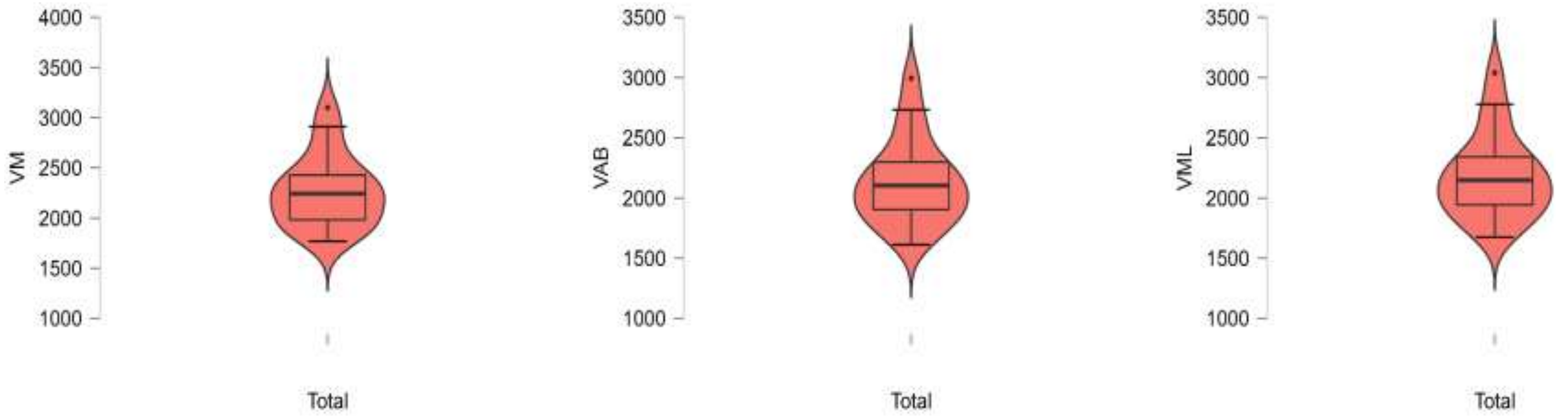


Lung L volumes differences

VM (manual reference segmentation)
VAB (Atlas Based segmentation)
VML (Deep Learning segmentation)

Descriptive Statistics

	VM	VAB	VML
Mean	2262.4	2129.8 (P<.001)	2180.9 (P<.001)
Std. Deviation	359.2	354.3	351.9
Range	1331.4	1382.3	1364.2



Lung L similarity

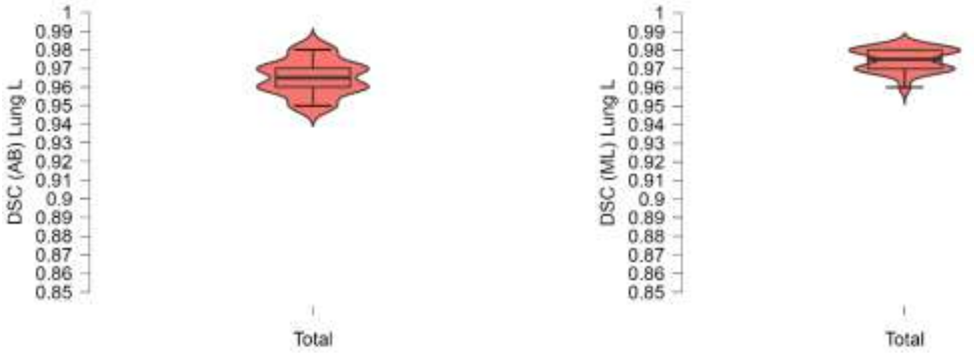
AB (Atlas Based segmentation)

ML (Deep Learning segmentation)

Dice Similarity Coefficient (DSC)

Descriptive Statistics

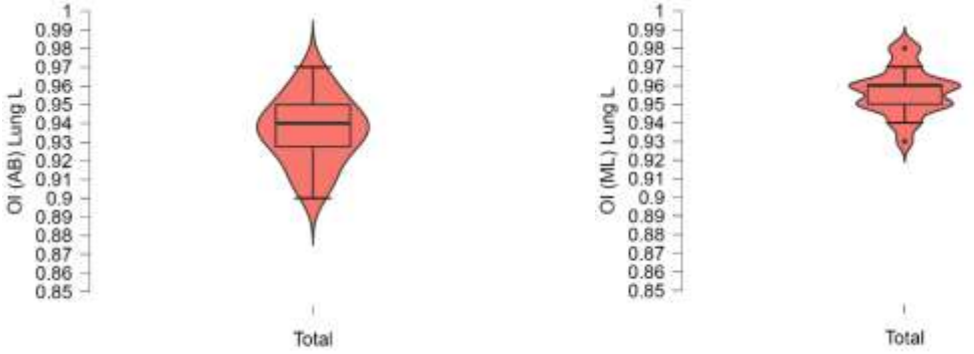
	DSC (AB) Lung L		DSC (ML) Lung L
Mean	0.96	P<.001	0.97
Std. Deviation	9.46×10^{-3}		6.05×10^{-3}
Range	0.03		0.02



Overlap index (OI)

Descriptive Statistics

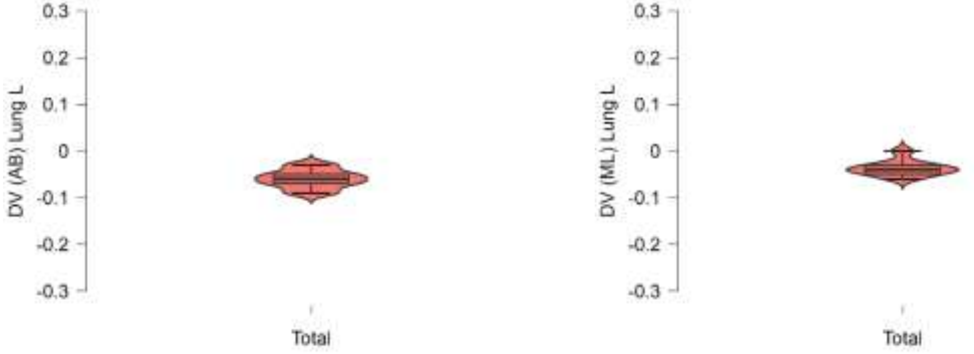
	OI (AB) Lung L		OI (ML) Lung L
Mean	0.94	P<.001	0.96
Std. Deviation	0.02		0.01
Range	0.07		0.05



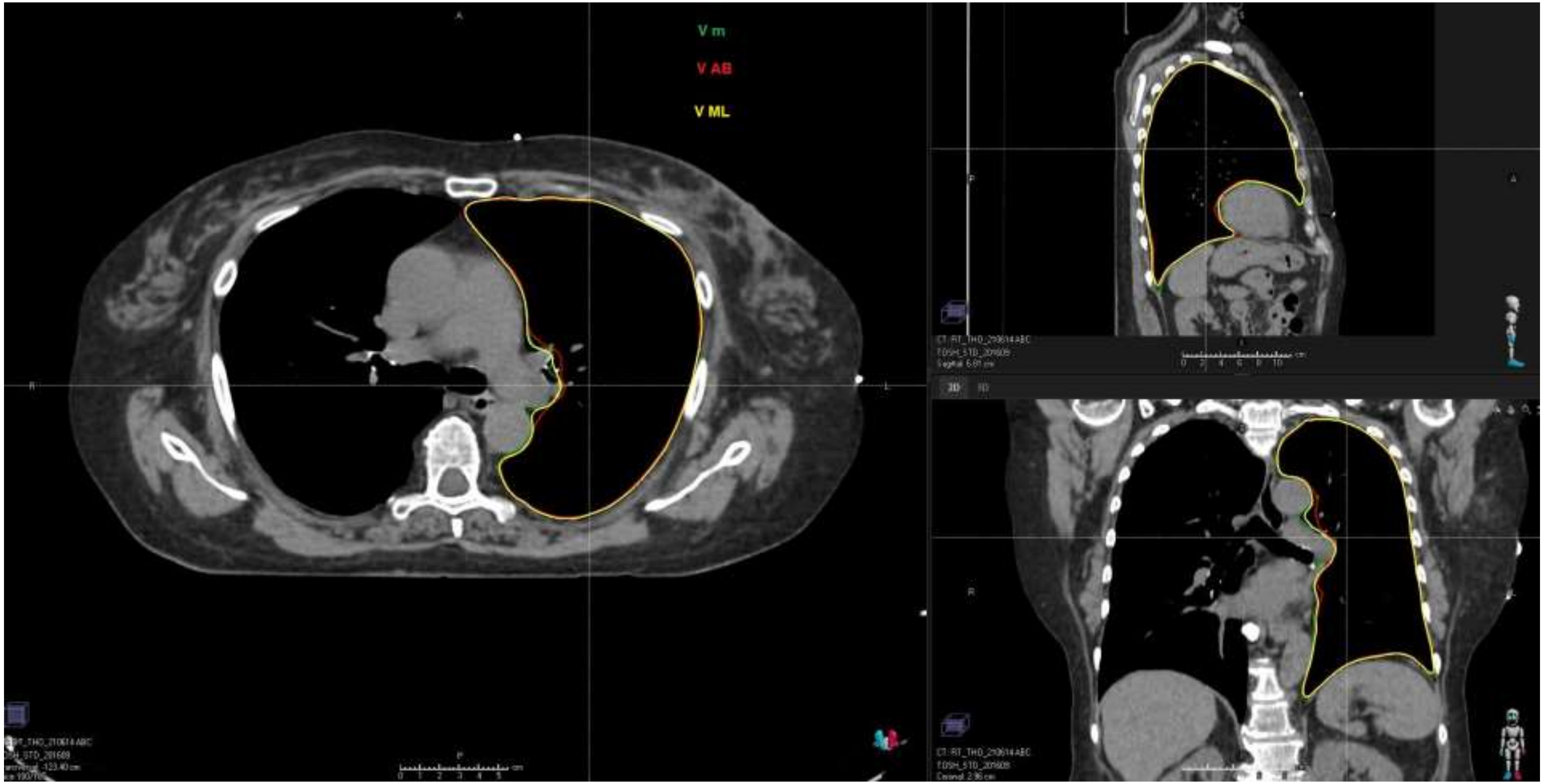
Volume Difference (DV)

Descriptive Statistics

	DV (AB) Lung L		DV (ML) Lung L
Mean	−0.06	P<.001	−0.04
Std. Deviation	0.02		0.02
Range	0.06		0.06



Lung L - Example

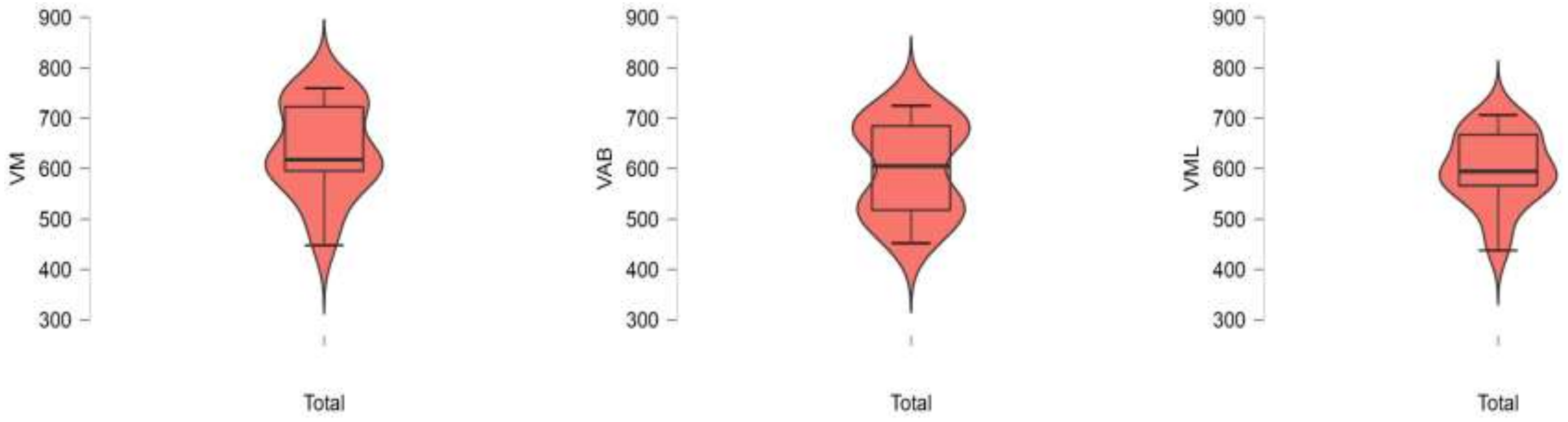


Heart volumes differences

VM (manual reference segmentation)
VAB (Atlas Based segmentation)
VML (Deep Learning segmentation)

Descriptive Statistics

	VM	VAB	VML
Mean	636.0	598.4 (P<.001)	598.2 (P<.001)
Std. Deviation	91.4	92.4	72.4
Range	311.5	272.6	268.4



Heart similarity

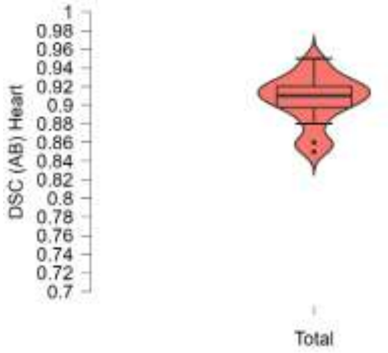
AB (Atlas Based segmentation)

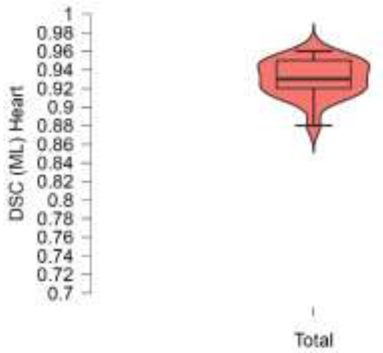
ML (Deep Learning segmentation)

Dice Similarity Coefficient (DSC)

Descriptive Statistics

	DSC (AB) Heart		DSC (ML) Heart
Mean	0.91	P=.002	0.93
Std. Deviation	0.03		0.02
Range	0.10		0.08

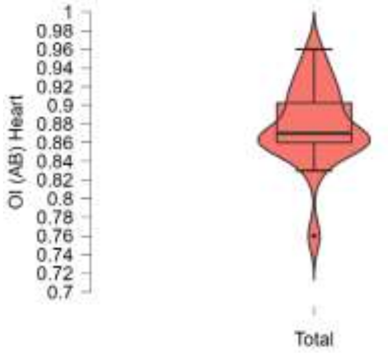


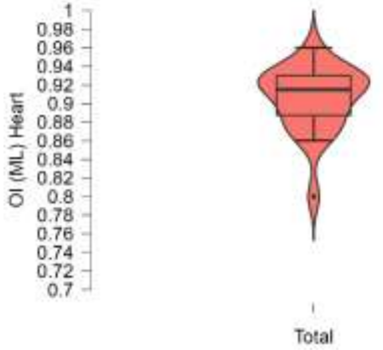


Overlap index (OI)

Descriptive Statistics

	OI (AB) Heart		OI (ML) Heart
Mean	0.88	P=.046	0.90
Std. Deviation	0.04		0.04
Range	0.20		0.16

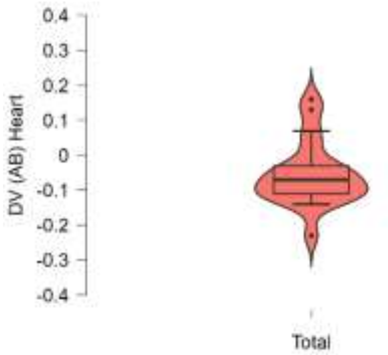


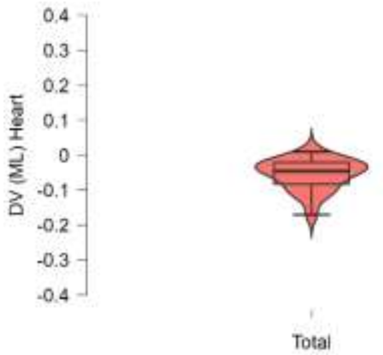


Volume Difference (DV)

Descriptive Statistics

	DV (AB) Heart	DV (ML) Heart
Mean	−0.06	−0.06
Std. Deviation	0.09	0.04
Range	0.39	0.18





canton de vaud

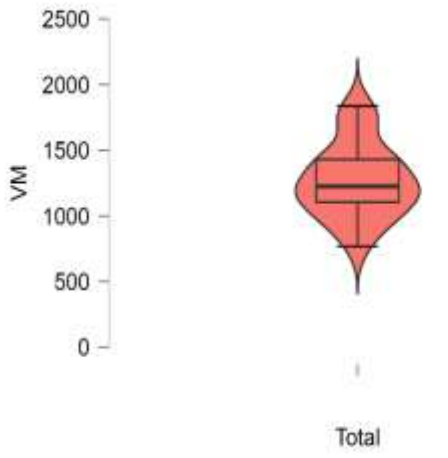
Heart - Example

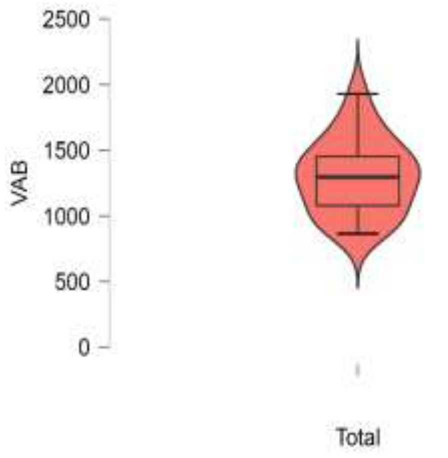
Liver volumes differences

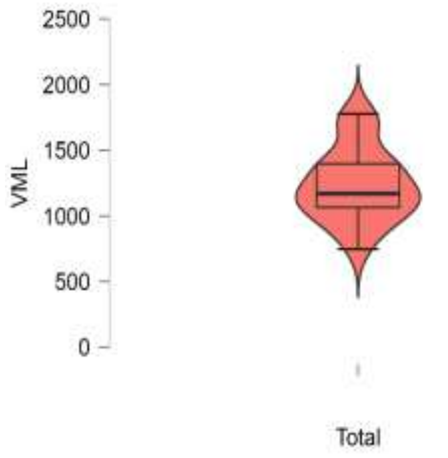
VM (manual reference segmentation)
VAB (Atlas Based segmentation)
VML (Deep Learning segmentation)

Descriptive Statistics

	VM	VAB	VML
Mean	1266.9	1298.5 (P<.040)	1236.5 (P<.001)
Std. Deviation	282.1	287.1	276.7
Range	1069.7	1063.0	1031.5







Liver similarity

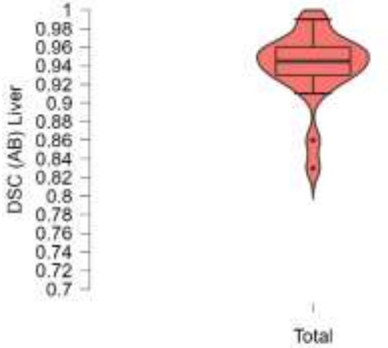
AB (Atlas Based segmentation)

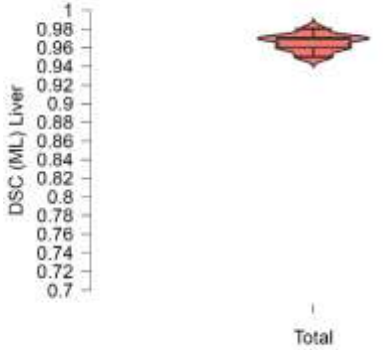
ML (Deep Learning segmentation)

Dice Similarity Coefficient (DSC)

Descriptive Statistics

	DSC (AB) Liver		DSC (ML) Liver
Mean	0.94	P<.004	0.96
Std. Deviation	0.04		8.89×10^{-3}
Range	0.16		0.03

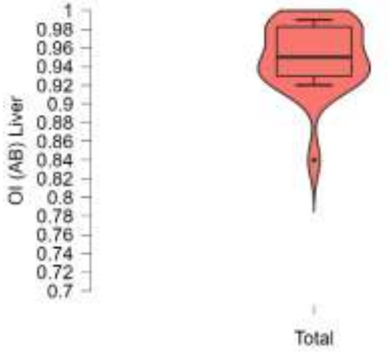


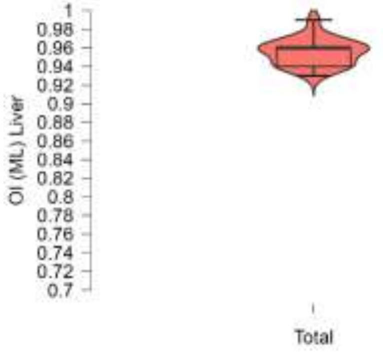


Overlap index (OI)

Descriptive Statistics

	OI (AB) Liver	OI (ML) Liver
Mean	0.95	0.95
Std. Deviation	0.04	0.01
Range	0.15	0.06

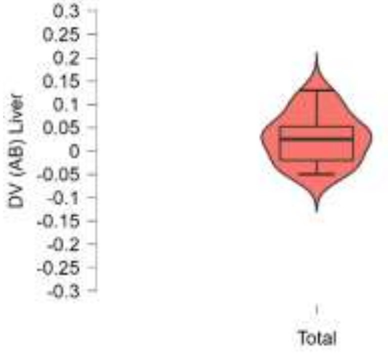


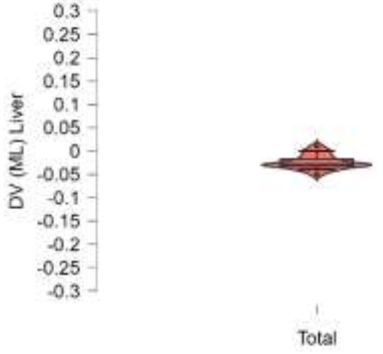


Volume Difference (DV)

Descriptive Statistics

	DV (AB) Liver		DV (ML) Liver
Mean	0.03	P=.001	−0.02
Std. Deviation	0.05		0.02
Range	0.18		0.06





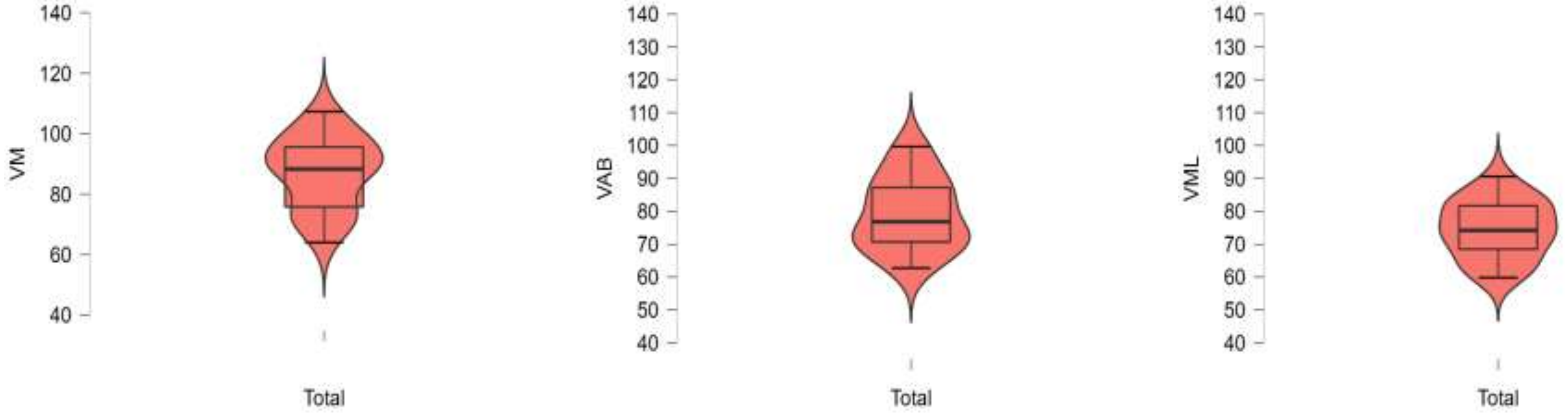
Liver - Example

Spinal Canal volumes differences

Descriptive Statistics

	VM	VAB	VML
Mean	87.3	79.3 P<.001	74.5 P<.001
Std. Deviation	12.1	11.1	8.9
Range	43.3	37.0	30.8

VM (manual reference segmentation)
VAB (Atlas Based segmentation)
VML (Deep Learning segmentation)



Spinal Canal similarity

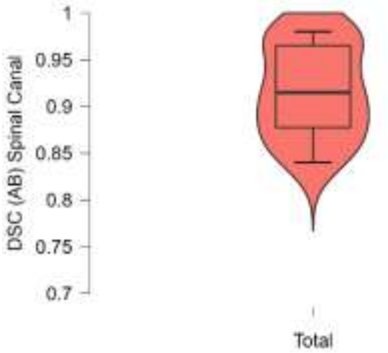
AB (Atlas Based segmentation)

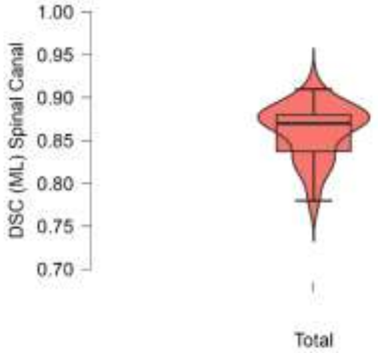
ML (Deep Learning segmentation)

Dice Similarity Coefficient (DSC)

Descriptive Statistics

	DSC (AB) Spinal Canal		DSC (ML) Spinal Canal
Mean	0.92	P=.004	0.86
Std. Deviation	0.05		0.03
Range	0.14		0.13

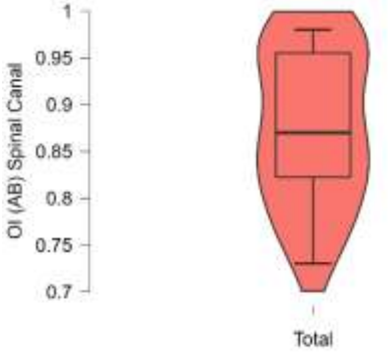


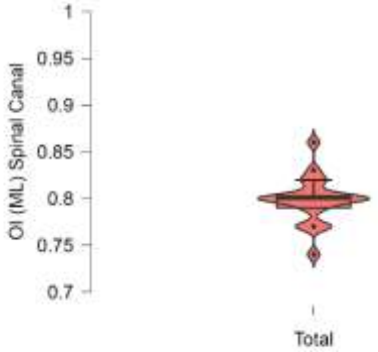


Overlap index (OI)

Descriptive Statistics

	OI (AB) Spinal Canal		OI (ML) Spinal Canal
Mean	0.88	P=.002	0.80
Std. Deviation	0.08		0.02
Range	0.25		0.12

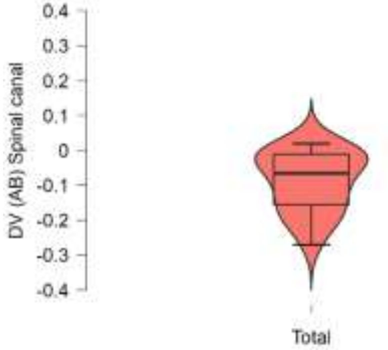


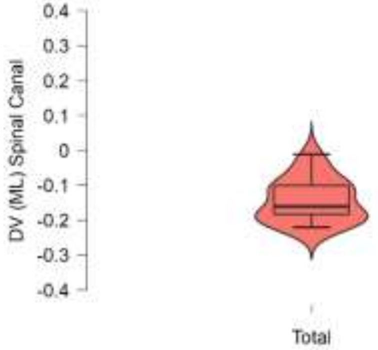


Volume Difference (DV)

Descriptive Statistics

	DV (AB) Spinal canal		DV (ML) Spinal Canal
Mean	−0.09	P=.019	−0.14
Std. Deviation	0.09		0.06
Range	0.29		0.21





canton de vaud

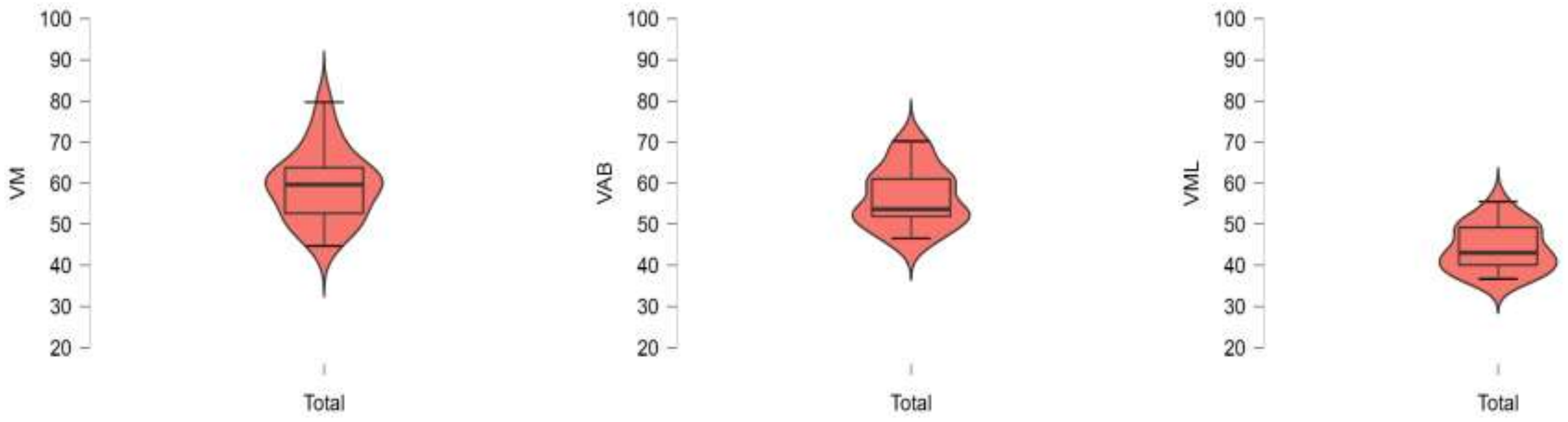
Spinal Canal - Example

Humeral Head L volumes differences

VM (manual reference segmentation)
VAB (Atlas Based segmentation)
VML (Deep Learning segmentation)

Descriptive Statistics

	VM	VAB	VML
Mean	59.4	56.6 (P=.017)	44.1 (P<.001)
Std. Deviation	8.7	7.1	5.6
Range	35.0	23.6	18.8

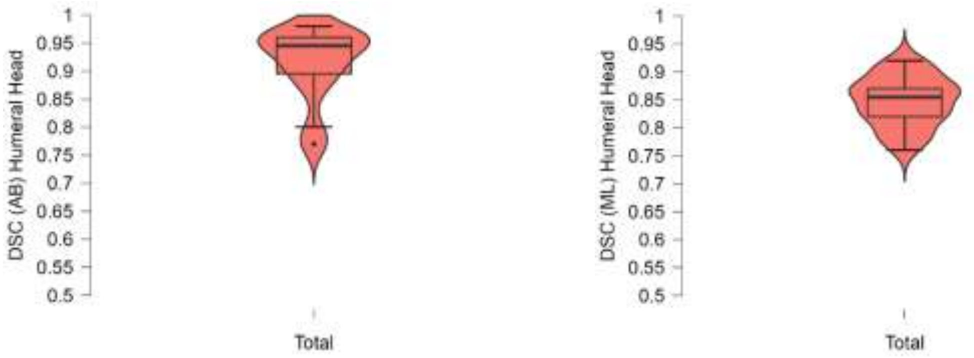


Humeral Head L similarity

Dice Similarity Coefficient (DSC)

Descriptive Statistics

	DSC (AB) Humeral Head		DSC (ML) Humeral Head
Mean	0.91	P<.001	0.85
Std. Deviation	0.07		0.05
Range	0.21		0.16



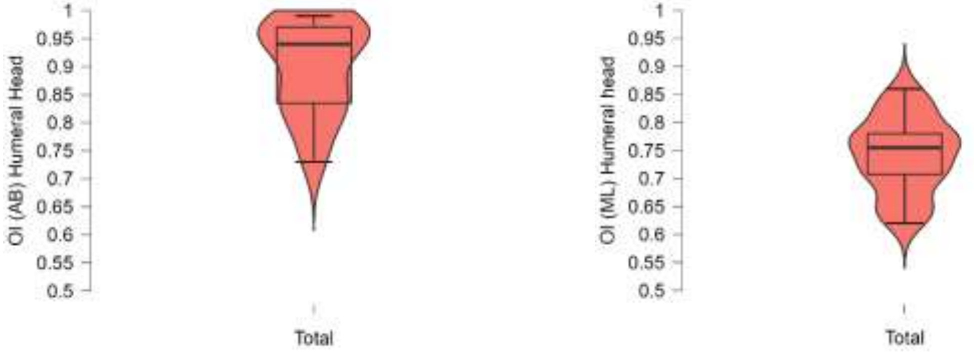
AB (Atlas Based segmentation)

ML (Deep Learning segmentation)

Overlap index (OI)

Descriptive Statistics

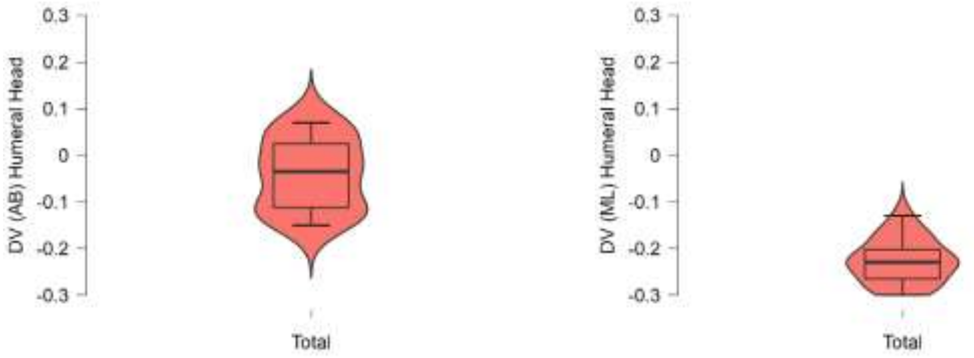
	OI (AB) Humeral Head		OI (ML) Humeral head
Mean	0.90	P<.001	0.74
Std. Deviation	0.08		0.07
Range	0.26		0.24



Volume Difference (DV)

Descriptive Statistics

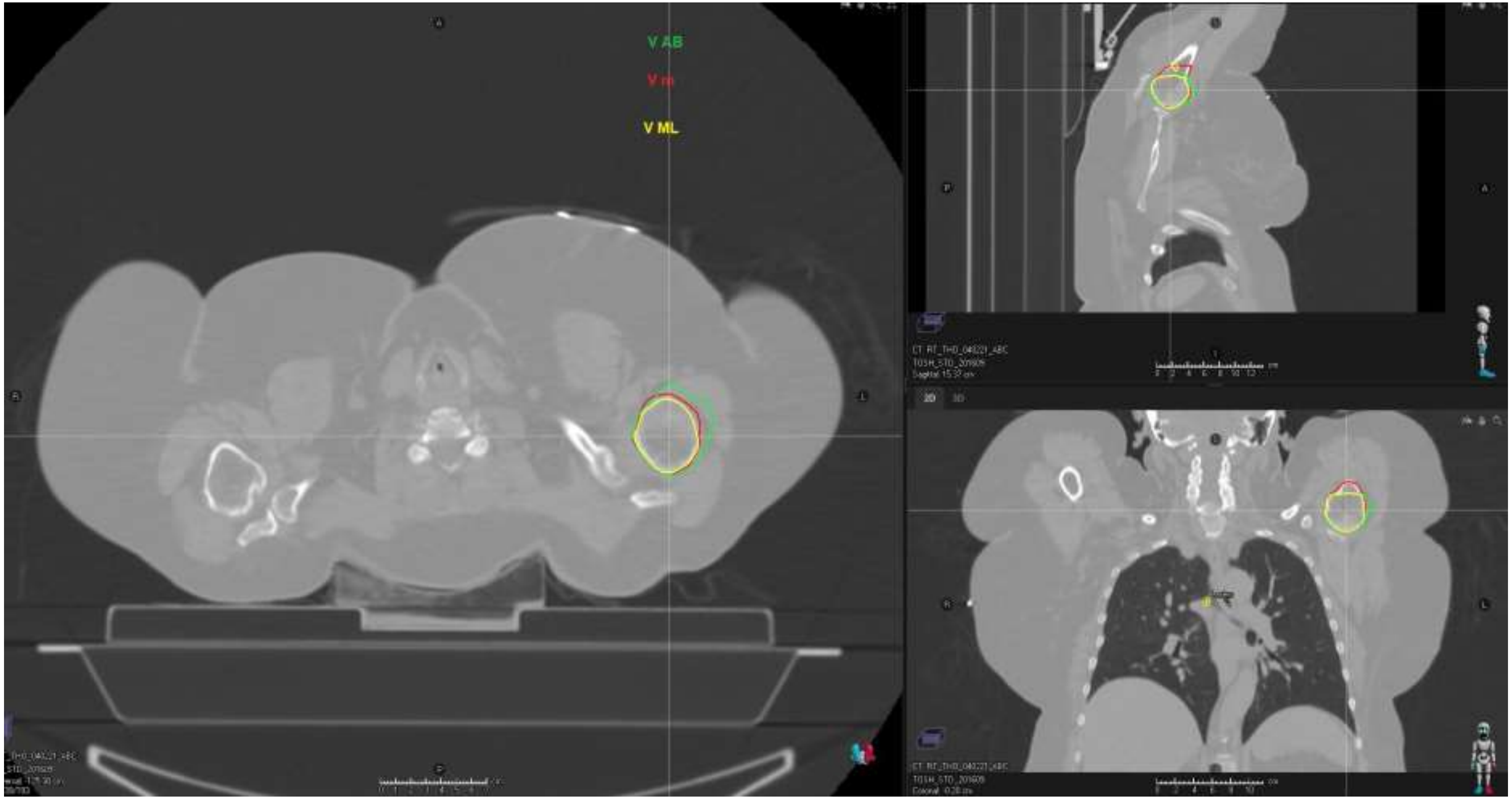
	DV (AB) Humeral Head		DV (ML) Humeral Head
Mean	−0.04	P<.001	−0.25
Std. Deviation	0.08		0.07
Range	0.22		0.24



canton de vaud

CHUV

Humeral Head L - Example



Discussion

- **Breast R**: no significant difference between the volumes. Still, the VAB and the VML tend to be smaller then the VM.

- **Lung R**: significant difference between the volumes. The VAB and the VML tend to be smaller then the VM. For the DSC, OI and DV, there's a significant difference between the VAB and VML. The index are overall better for VML.

- **Lung L**: significant difference between the volumes. The VAB and the VML tend to be smaller then the VM. The difference is higher for Lung L than for the Lung R (Heart anatomy influence?). For the DSC, OI and DV, there's a significant difference between the VAB and VML. The index are overall better for VML. The metrics are less good comparing with the Lung R.

- **Heart:** significant difference between the volumes. The VAB and the VML tend to be smaller then the VM. The mean value for VAB and VML are very similar, but VAB has a higher SD. For the DSC and OI, there’s a significant difference between the VAB and VML, they are better for VML. For the DV, there’s no significant difference between VAB and VML.

- **Liver**: significant difference between the volumes. The VAB it’s larger than VM. For the DSC and DV, there’s a significant difference between the VAB and VML. The DSC and Dv are better for VML. For the OI there's no significant difference.

- **Spinal Canal**: significant difference between the volumes. The VAB and the VML tend to be smaller then the VM. For the DSC, OI and DV, there's a significant difference between the VAB and VML. The index are overall better for AB regarding the DSC and OI. This overall difference can be linked to the fact that the VM as a higher length in the superior-inferior direction.

- **Humeral Head L**: significant difference between the volumes. The VAB and the VML are smaller. The VM it's contoured much higher in the superior direction. The AB and ML always segmented the humeral head without this superior margin. For the DSC, OI and DV, there's a significant difference between the VAB and VML. The index are overall better for AB regarding the DSC and OI.

Conclusion

- As a user, we need to evaluate the quality segmentation output
- Awareness regarding the contouring “deskilling” risk
- Impact of the dose on different segmented OARs needs to be evaluated
- Bad segmentation not always may have an impact on dosimetry (it depends by the treatment site and treatment technique)

- Quality of segmentation was different because of the number of patients used to train each segmentation technique

- It’s important the training datasets for both segmentation techniques, it must include real world patient’s anatomical variability → feedback for vendors

- Importance of collaboration between the radiotherapy multidisciplinary team (RTT’s, RO’s, Phy’s)

References

1. Chen, W., Wang, C., Zhan, W. *et al.* A comparative study of auto-contouring softwares in delineation of organs at risk in lung cancer and rectal cancer. *Sci Rep* **11,** 23002 (2021). https://doi.org/10.1038/s41598-021-02330-y
2. Taha and Hanbury. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool BMC Medical Imaging (2015) 15:29. DOI 10.1186/s12880-015-0068-x
3. Zhu J, Chen X, Yang B, Bi N, Zhang T, Men K and Dai J (2020) Evaluation of Automatic Segmentation Model With Dosimetric Metrics for Radiotherapy of Esophageal Cancer. Front. Oncol. 10:564737. doi: 10.3389/fonc.2020.564737
4. Varduhi Yeghiazaryan, Irina Voiculescu, “Family of boundary overlap metrics for the evaluation of medical image segmentation,” J. Med. Imag. 5(1), 015006 (2018), doi: 10.1117/1.JMI.5.1.015006

Acknowledgments

Special thanks for the medical physicist, Michele Zeverino

Radiation oncologist Wendy Jeanneret Sozzi

RTT Susana Leal

RTT Léonie Heym

Prof. Jean Bourhis

Prof. Raphaël Moeckli

Thank you!

soares-rodrigues.joao-luis@chuv.ch