Breast cancer annotation across genders

Madalina-Liana Costea1, Lorenzo Colombo1, Gorkem Gungor2, Guillaume Klausner3,4, Jean-Baptiste Clavier5, Nicolas Leduc6, Celine Bourgier7, Charlotte Robert8, Chloe Herve9, Enis Ozyar2, Sami Romdhani10, Olivier Teboul10, Gizem Temiz1, Nikos Paragios11
1 TheraPanacea, Clinical Affairs, Paris, France.
2 Acibadem MAA University School of Medicine, Department of Radiation Oncology, Istanbul, Turkey.
3 Centre de Haute Énergie (C.H.E.) Nice - Bastia, Department of Radiation Oncology, Bastia, France.
4 Hôpitaux Universitaires de Genève, Department of Radiation Oncology, Geneva, Swaziland.
5 ICANS - Institut de cancérologie Strasbourg Europe, Department of Radiation Oncology, Strasbourg, France.
6 Centre Catalan d'Oncologie, Department of Radiation Oncology, Perpignan, France.
7 Institut du Cancer de Montpellier, Department of Radiation Oncology, Montpellier, France.
8 Institut Gustave Roussy, Department of Radiation Oncology, Villejuif, France.
9 Centre Hospitalier Départemental Vendée, Department of radiation oncology, La Roche-sur-Yon, France.
10 TheraPanacea, AI engineering, Paris, France.
11 TheraPanacea, CEO, Paris, France

Purpose/Objective:
Male breast cancer is a rare disease that affects 1% of male population with a peak distribution at the age of 71 [1]. AI-based segmentation models can provide precise contours for accurately targeting tumors while sparing healthy tissues and reduce manual contouring time in radiotherapy (RT) treatment planning. However due to high cancer incidence rates, most of the models focus on female patients. The objective of this study was to evaluate the generalizability to male population of a commercial annotation solution trained on female patients data.

Material/Methods:
An AI model was trained and evaluated on CT images from female breast cancer patients who were treated with RT in arms up positioning. The ground truth (GT) data used for evaluation came from experts from different centers with different contouring practices.
To assess the model's performance on male patients, a new set of GT contours were produced by two experts, following ESTRO contouring guidelines [2,3]. All 10 CT images used for this evaluation belonged to patients treated in the arms up position. The time spent on manual contouring was recorded and the inter-expert variation (IEV) was calculated based on Dice Similarity Coefficient (DSC).

Subsequently, the AI model's performance was evaluated by calculating mean DSC between the GT created by the two experts. The results were compared to the IEV results and two physicians qualitatively assessed the AI-based generated contours on A, B, C scoring (A=acceptable without modification, B=acceptable with minor modifications, C=not acceptable, major modifications are needed). When a big discrepancy was observed between the two experts' scores, a third physician was consulted.

Results:

The 16 organs were delineated in an average time of 35 minutes (Figure 1). Per organ, the IEV results ranged from mean DSC of 0.43 for right brachial plexus to 0.79 for the left breast (Table 1). Comparing the AI model predictions to the manual GT contours, the mean DSC results ranged from 0.27 for the right brachial plexus to 0.68 for the right breast. The brachial plexus had low DSC results due to its poor visibility on CT images, reflected in expert-to-expert and expert-to-AI contour comparisons.

Regarding the qualitative evaluation, the raters had close agreement for 11/16 organs, while 5 organs required input from a third rater. For these 5 organs, rater 2 and 3 found the contours acceptable with minor corrections, while rater 1 deemed major corrections necessary for clinical acceptability. Surprisingly the left and the right breast fell into the second category, possibly indicating a gender bias among patients. An interview with rater 1 revealed that breast contours were predominantly accurate, while notable inaccuracies at the upper and lower slices necessitating additional manual correction.

Further perspectives of this work include testing different post processing rules aligned with guidelines and gathering more data from clinics to train a new model with male patients GT contours.

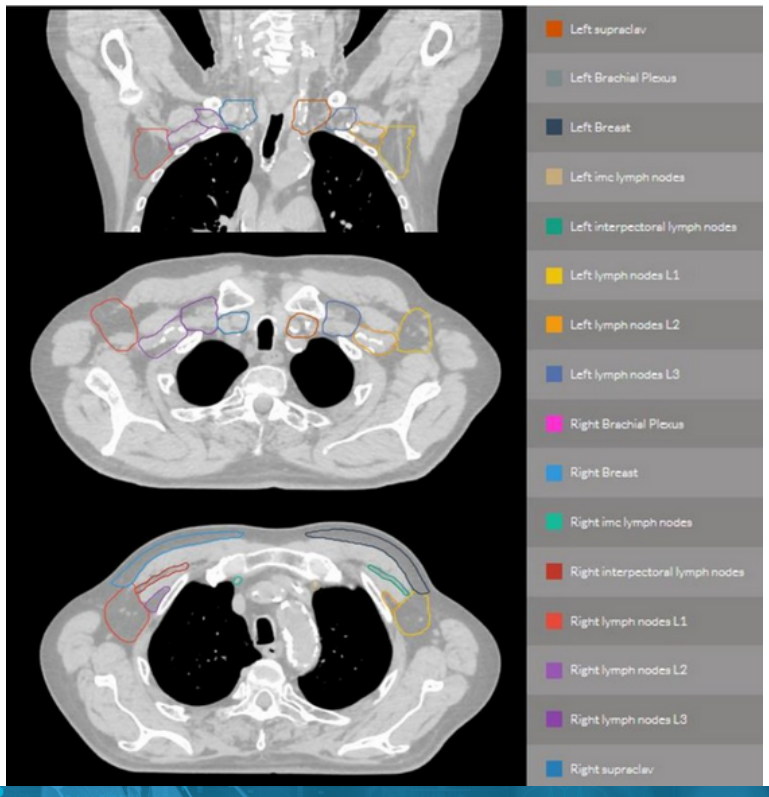

Figure 1. Example of round truth contours on a male breast cancer patient
.

		Quantitative results (mean DSC)		Qualitative results (% of A + B grades)		
	Organ	Expert 1 vs Expert 2	AI vs expert contours	Rater 1	Rater 2	Rater 3
1	Left breast	0.79	0.63	30%	90%	100%
2	Right brachial plexus	0.43	0.27	30%	90%	100%
3	Right breast	0.78	0.68	50%	90%	90%
4	Left lymph node L3	0.62	0.55	80%	80%	100%
5	Left lymph node L1	0.73	0.60	67%	100	100%
6	Left supraclavicular lymph nodes	0.65	0.52	70%	100%	
7	Right supraclavicular lymph nodes	0.68	0.55	70%	100%	
8	Left internal mammary lymph nodes	0.56	0.52	70%	100%	
9	Left interpectoral lymph nodes	0.67	0.54	70%	100%	
10	Right lymph nodes L1	0.73	0.67	78%	100%	
11	Left brachial plexus	0.44	0.33	80%	100%	
12	Right internal mammary lymph nodes	0.54	0.53	90%	90%	
13	Left lymph nodes L2	0.66	0.66	80%	100%	
14	Right interpectoral lymph nodes	0.64	0.54	90%	100%	
15	Right lymph nodes L3	0.59	0.52	90%	100%	
16	Right lymph nodes L2	0.57	0.63	100%	100%	

Table 1. Quantitative and qualitative results of AI model predictions on male population

Conclusion:

In conclusion, the AI model initially trained on female patient cohorts demonstrated a noteworthy level of generalizability to male population. While review and manual refinements remain essential, this underscores the potential for precision in medicine. Therefore, the RT clinics can benefit greatly from fast AI predictions for the treatment of breast cancer regardless of patient gender.

References:

[1] W. F. Anderson et al., "Is Male Breast Cancer Similar or Different than Female Breast Cancer?", Breast Cancer Research and Treatment, vol. 83, no. 1, pp. 77-86, Jan. 2004, doi: 10.1023/b:brea.0000010701.08825.2d.

[2] B. V. Offersen et al., "ESTRO consensus guideline on target volume delineation for elective radiation therapy of early stage breast cancer, version 1.1", Radiotherapy and Oncology, vol. 118, no. 1, pp. 205-208, Jan. 2016, doi: 10.106/j.radonc.2015.12. 027.

[3] C. L. Brouwer et al., "CT-based delineation of organs at risk in the head and neck regio: DAHANCA, GORTEC, HKNPCSG, NCIC CTG, NCRI, NRG Oncology and TROG consensus guidelines", Radiotherapy and Oncology, vol. 117, no. 83-90, Oct. 2015, doi: 10.1016/j.radonc.2015.07.041.