2nd PANHELLENIC CONGRESS OF MEDICAL PHYSICS
4-6 OCTOBER 2024| EUGENIDES FOUNDATION

CLINICAL EVALUATION OF ARTIFICIAL INTELLIGENCE-BASED SOFTWARE FOR AUTO-CONTOURING IN RADIATION THERAPY OF BREAST CANCER

Terzi G. Maria[2,3], Sakellaropoulos George[1,3], Spiridon N. Papatheodorou[3], Despoina A. Spyropoulou[2,4]

[1] Department of Medical Physics, School of Medicine, University of Patras, [2] Department of Radiation Oncology, School of Medicine, University of Patras, [3] Department of Medical Physics, University Hospital of Patras, [4] Department of Radiation Oncology, University Hospital of Patras

Background: The aim of the study was to compare the auto-generated contours by the AI-based software "*Therapanacea*" with *manual - generated contours* by physicians. *Can these software tools eliminate the quality evaluation of the experienced physician?*

Materials and Methods: Four radiation oncologists with *different clinical experience* delineated the Clinical Target Volumes (CTVs) for Breast cancer radiotherapy and Organs at Risk (OARs) on 45 CT cases.

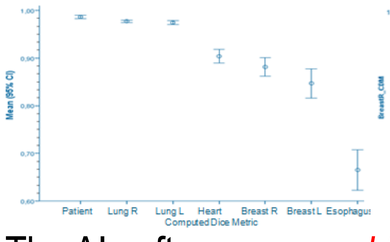
~*Two volumetric overlap metrics*, the Hausdorff distance (*CHD-95%*) and Coefficient Dice Metric (*CDM*), were used. The sample was divided into *14 groups* according to the above metrics and *4 groups*, one for each physician. Statistical analysis was performed.

~ In order to *evaluate the dosimetric impact of AI: Two plans* were created: one with contours by AI and the other using the manual-generated contours.

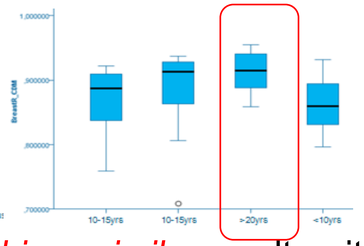
Results: Regarding the CHD-95%, statistically significant differences were found for the *Lung R* ($p<0.05$) and *Breast L* ($p<0.05$). No statistical difference was found for CDM ($p>0.05$).

Conclusion: This software could be used to *reduce* the *workflow time* and the *variability* among physicians. The quality evaluation of the experienced physician is necessary. The physician's experience cannot be replaced by the metrics.

Difference between auto- and manual contours using CDM



Difference AI-Expert with different clinical experience



The Lungs had the highest degreeof geometric overlap

	N	Normal Parameters[a,b] Mean	Std. Deviation	Most Extreme Differences Absolute	Positive	Negative	Test Statistic	Asymp. Sig. (2-tailed)
Patient_CHD	45	11,81	10,80	,170	,170	-,158	,170	,002[c]
Patient_CDM	45	,99	,01	,239	,239	-,149	,239	,000[c]
LungR_CHD	45	4,63	2,70	,206	,206	-,129	,206	,000[c]
LungR_CDM	45	,98	,01	,139	,110	-,139	,139	,028[c]
LungL_CHD	45	6,05	4,13	,220	,220	-,118	,220	,000[c]
LungL_CDM	45	,97	,01	,185	,129	-,185	,185	,001[c]
Heart_CHD	45	9,47	5,57	,188	,188	-,130	,188	,000[c]
Heart_CDM	45	,90	,04	,100	,076	-,100	,100	,200[c,d]
BreastR_CHD	44	12,15	4,97	,121	,121	-,102	,121	,110[c]
BreastR_CDM	44	,88	,05	,169	,099	-,169	,169	,003[c]
BreastL_CHD	45	13,89	5,43	,159	,159	-,085	,159	,006[c]
BreastL_CDM	45	,84	,09	,159	,138	-,159	,159	,006[c]
Esophagus_CHD	36	16,04	12,77	,215	,215	-,161	,215	,000[c]
Esophagus_CDM	36	,66	,12	,135	,082	-,135	,135	,096[c]

The AI software can *achieve similar* results with the *most experienced* physician. The differences between the auto- and manual-generated contours for *patient* and *esophagus* were due to a difference in *length*.

PTV Coverage	
PTV AI	93.83
PTV Expert	93.83

No difference for the PTV (Planning TargetVolume) coverage *despite the low degree* of geometric overlap

	CHD(95%)	CDM
Lung	3.75	0.98
Breast	15.00	0.86
Heart	25.25	0.81

PTV Coverage	
PTV AI	93.00
PTV Expert	95.15

Significant difference for the PTV coverage *and low degree* of geometric overlap

	CHD(95%)	CDM
Lung	9.21	0.98
Breast	10.23	0.94
Heart	3.95	0.93

There is a need to develop a standardized method for the validation of any AI-based software