IL NUOVO CIMENTO **46 C** (2023) 141
DOI 10.1393/ncc/i2023-23141-7

COMMUNICATIONS: SIF Congress 2022

Quantifying multi-Institution KB-plan prediction models transferability in breast cancer radiotherapy: A step forward toward benchmark(*)

A. TUDDA(1)(2), R. CASTRICONI(1), F. DUSI(3), A. SCAGGION(3), L. PLACIDI(4), A. CICCHETTI(5), M. GUERNIERI(6), G. BENECCHI(7), E. CAGNI(8), A. IANIRO(9), V. LANDONI(9), A. MAZZILLI(7), G. MEFFE(4), E. MORETTI(6), C. OLIVIERO(10), G. RAMBALDI GUIDASCI(11), T. RANCATI(5), V. TROJANI(8), A. DEL VECCHIO(1) and C. FIORINO(1)(**)

(1) *Medical Physics Dept, San Raffaele Scientific Institute - Milano, Italy*

(2) *Università Statale di Milano - Milano, Italy*

(3) *Medical Physics Department, Veneto Institute of Oncology IOV-IRCCS - Padua, Italy*

(4) *Fondazione Policlinico Universitario Agostino Gemelli IRCCS - Roma, Italy*

(5) *Fondazione IRCCS Istituto Nazionale dei Tumori - Milano, Italy*

(6) *Department of Medical Physics, University Hospital - Udine, Italy*

(7) *Medical Physics Dept., University Hospital of Parma AOUP - Parma, Italy*

(8) *Medical Physics Unit, Department of Advanced Technology, Azienda USL-IRCCS di Reggio Emilia - Reggio Emilia, Italy*

(9) *IRCCS Istituto Nazionale dei Tumori Regina Elena - Roma, Italy*

(10) *University Hospital, "Federico II" - Naples, Italy*

(11) *UOC di Radioterapia Oncologica, Fatebenefratelli Isola Tiberina, Gemelli Isola Roma, Italy*

received 21 March 2023

(*) This article is a short overview of TUDDA A. *et al.*, *Radiother. Oncol.*, **175** (2022) 10, https://doi.org/10.1016/j.radonc.2022.07.012, @ 2022 Elsevier B. V. All rights reserved.
(**) PI of MIKAPOCo project.

Summary. — Ten Institutes (INST1-INST10) of the MIKAPOCo consortium set their KB plan prediction models of whole breast irradiation (WBI) delivered with tangential fields (TF), by using RapidPlan (Varian Medical System, Inc.) and following the same criteria of contouring and plan model building and validation. DVH prediction bands of organs at risk (heart, ipsilateral lung, contralateral lung and contralateral breast) were exported on 20 new patients from the same Institutes (two patients each). SD of mean predicted DVHs among institutes was assessed as inter-institute variability. The estimated Principal Component (PC1) was considered for transferability models evaluation. Transferability cross-validation was further investigated in detail on a larger population for the model showing the poorest transferability ($\text{Model}_{\text{INST6}}$) against one of the models with high transferability ($\text{Model}_{\text{INST3}}$). Results show a limited inter-institute variability of plan prediction models (1.8% for DVH ipsilateral lung) and a satisfactory inter-institute transferability, excepting one institute, confirmed by the extended analyses on a larger cohort of test patients of INST6 (*vs.* INST3). These achievements pave the way for generating benchmarks for plan prediction in WBI with potentially relevant large-scale applications.

1. – Introduction

Radiation therapy (RT) after surgery represents standard care for breast cancer in the majority of patients. Three-dimensional conformal radiotherapy (3DCRT) planning optimization and its evolution to field-in-field (FiF) delivery dealt with manual optimization, strongly planner-dependent and time consuming [1]. Both inter-planning variability and the time spent for planning could ideally be significantly reduced using AI (Artificial Intelligence) techniques in the planning optimization phase [2-5]. In this scenario, Knowledge-based (KB) planning optimization is a machine learning technique based on modeling of previously optimized clinical plans, aiming to individually predict the expected planning performances in new patients [6]. The MIKAPOCo project (Multi-Institutional Knowledge based Approach for Planning Optimization for the Community - IG23150, approved by San Raffaele institutional ethical committee, no. 248/2021) aims to build consistent KB model's libraries and/or incorporate inter-institute variability into plan predictions with the goal of providing the community with robustly usable KB models. The aim of the current work was to: 1) briefly summarize the main findings, recently published [7], concerning inter-Institute variability and transferability between right breast KB models; 2) more deeply exploring transferability between two institute models (INST3 and INST6) that appeared different in contouring and planning the approach to better quantify a "worst-case scenario".

2. – Materials and methods

Ten Italian institutes set their KB model following commonly previously agreed criteria, using different versions of the Model Configuration tool of RapidPlan (RP, Varian Medical System, Inc). Clinical plans of patients treated with TF with manually optimized wedges or FiF were selected. Contouring was conducted following national guidelines [8],

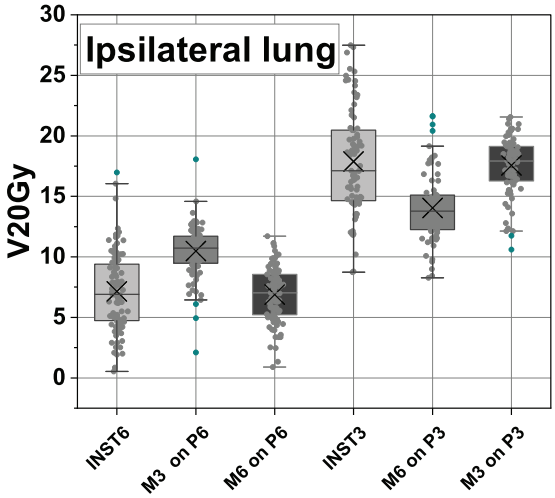

Fig. 1. – Boxplot of V_{20}Gy clinical values for INST3 and INST6 (INST6 and INST3, respectively). The figure also reports estimated V_{20}Gy by $\text{Model}_{\text{INST6}}$ (respectively $\text{Model}_{\text{INST3}}$) on INST6 and INST3 patients.

where ipsilateral lung was always contoured, while heart, left lung and contralateral breast contouring were optional. Using the Principal Component (PC) Analysis, RP generated a regression model correlating geometrical information (patient anatomy) and the Dose-Volume Histogram (DVH), considering the beam arrangement used for the specific plan. Principal Components (PC1) of the training set for DVH were extracted. Then, the system predicted the most probable DVH expressed with a minimum-maximum band, starting from the current anatomic and field geometry situation. The analysis was performed using Model Analytics software (Varian Medical Systems, Inc.) [9], which allows extracting the dosimetric and geometric characteristics of each dataset for every model. Two patients from each center, not included in the models, were randomly selected in order to extract DVH OARs predictions. Considering all test patient (20 patients) and institute models, the mean DVH_{i} was calculated as the average DVH for every patient (i) through every single model, with its corresponding standard deviation. The statistical dialogue box of RP shows how geometric features of the training cohort of a model (PC1 and others) fits the same characteristic of the patient run on the model. The degree of transferability between models was evaluated considering the number of cases in which the ipsilateral lung PC1 of the 18-patient test was out of the 90th percentile of the training set of the other institute models. Transferability cross-validation was further investigated on a larger population considering INST3 and INST6 models ($\text{Model}_{\text{INST3}}$ and $\text{Model}_{\text{INST6}}$) with their dataset patients: $\text{Dataset}_{\text{INST3}} = 79$ patients and $\text{Dataset}_{\text{INST6}} = 100$ patients.

3. – Results

Detailed results have been reported elsewhere [7]. When looking to the predicted ipsilateral lung DVH, mean value of inter-institute SD was 1.8% in the dose range from 20% to 80%, showing relatively similar predictions among Institutes. A clear association between median value PTV D99% and the mean predicted dose on ipsilateral lung by

every institute was found ($R^2 = 0.78$) while excluding INST6, showing the lowest mean dose prediction value and no overlap between PTV and ipsilateral lung and revealing a different contouring attitude compared to the other Institutes. As a consequence, based on the PC1 analysis, 7/18 cases were outside the 90th percentile of the dataset cohort only for INST6, while they were at most 1/18 for all the other Institutes. Transferability cross-validation was further investigated between $\text{Model}_{\text{INST6}}$ against one of the models with high transferability and a different contouring approach ($\text{Model}_{\text{INST3}}$). $\text{Model}_{\text{INST6}}$ versus $\text{Dataset}_{\text{INST3}}$ analysis revealed PC1 out of range values in 41/79 cases, confirming poor transferability of $\text{Model}_{\text{INST6}}$. On the other hand, $\text{Model}_{\text{INST3}}$ versus $\text{Dataset}_{\text{INST6}}$ showed better transferability with PC1 out of range for only 14/100 patients. The median clinical value of V_{20}Gy for INST6 (7.1%) was significantly lower ($p < 0.05$) than the median value of clinical V_{20}Gy INST3 (17.1%), as expected (fig. 1). V_{20}Gy predicted by $\text{Model}_{\text{INST3}}$ on $\text{Dataset}_{\text{INST6}}$ was systematically higher compared to clinical V_{20}Gy and the estimated values. Prediction of V_{20}Gy by $\text{Model}_{\text{INST6}}$ on $\text{Dataset}_{\text{INST3}}$ were lower when compared the $\text{Model}_{\text{INST3}}$ on its dataset patients. Despite better transferability of $\text{Model}_{\text{INST3}}$, a systematic difference of the values predicted by this model on the patients of INST 6 with respect to the original clinical plans was found.

4. – Conclusions

In this study, we compared KB models from ten different Italian institutes. Despite differences between collaborating centers in contouring and sparing OARs approaches, SD% mean calculated as the average ranging prediction of models is 1.8% in dose range from 20% to 80% for ipsilateral lung. A metric of models' transferability between models was chosen as the overlap between model's PC1 and the corresponding test set patients' values. This analysis permitted to find good transferability between models ($< 6\%$ of the unsuccessful), except for INST6 that failed in 39% of the cases. A more detailed analysis was conducted considering INST3 and INST6 with their complete training set data. As expected, a better transferability of $\text{Model}_{\text{INST3}}$ on $\text{Dataset}_{\text{INST6}}$ was found, although to a relatively low degree; instead, $\text{Model}_{\text{INST6}}$ failed in 52% of cases, confirming its very poor transferability. With the exception of 1 in ten institutes, our results show relatively high consistency between KB models expressed in a limited inter-institute variability in models' prediction performances. Results are encouraging showing clear potentials in generating a Benchmark model in whole breast irradiation with tangential fields, incorporating inter-institute variability.

REFERENCES

[1] Penninkhof J. *et al.*, *Int. J. Radiat. Oncol. Biol. Phys.*, **98** (2017) 447.
[2] Fogliata A. *et al.*, *Radiother. Oncol. J. Eur. Soc. Ther. Radiol. Oncol.*, **113** (2014) 385.
[3] Scaggion A. *et al.*, *Phys. Med.*, **53** (2018) 86.
[4] Yuan L. *et al.*, *Phys. Med*, **39** (2012) 6868.
[5] Wu B. *et al.*, *Phys. Med.*, **36** (2012) 5497.
[6] Yan H. *et al.*, *Phys. Med*, **30** (2003) 2675.
[7] Tudda A. *et al.*, *Radiother. Oncol.*, **175** (2022) 10.
[8] Ciabattoni A. *et al.*, *Tumori J.*, **108** (2022) 1.
[9] `https://modelanalytics.varian.com/FAQ`.