Learning to Associate Distances with Historical Patient Data to Enable Fine-grained Studying of Late Adverse Effects of Paediatric Radiotherapy: Data, Methodology, and First Results

Virgolin, Marco; Dijk van, Irma; Wiersma, Jan; Ronckers, Cecile; Witteveen, Cees; Rasch, Coen; Bel, Arjan; Alderliesten, Tanja; Bosman, Peter

Publication date
2016

Document Version
Final published version

Published in
Proceedings of the 18th International Conference on the use of Computers in Radiation Therapy

Learning to Associate Distances with Historical Patient Data to Enable Fine-grained Studying of Late Adverse Effects of Paediatric Radiotherapy: Data, Methodology, and First Results

Marco Virgolin[1†], Irma W.E.M. van Dijk[2], Jan Wiersma[2], Cécile M. Ronckers[3], Cees Witteveen[4], Coen Rasch [2], Arjan Bel[2], Tanja Alderliesten[2] and Peter A. N. Bosman[1]

[1]Centrum Wiskunde & Informatica, Amsterdam, NL; [2]Academic Medical Center (AMC), Amsterdam, NL; [3]Emma Children's Hospital / AMC, Amsterdam, NL; [4]TU Delft, Delft, NL.

Introduction

Children undergoing radiation treatment are among the most sensitive patients prone to develop significant late adverse effects. Because 3D anatomy information could not be collected at the time, current studies on historically-treated childhood cancer survivors typically rely on dichotomous variables, dose categories, or crudely estimated average organ doses to quantify dose-response relationships for late adverse effects. Such relationships provide limited contributions to the improvement of treatment plan design. Several studies deal with this problem through phantom-based dose-reconstructions, e.g., [1]. Phantoms, however, still exhibit limitations in terms of patient individualization.

The work presented here is part of a novel, individualized, data-driven approach that we are developing for 3D dose-distribution reconstruction. This approach is to (i) use relevant features to match a historically-treated patient *p* with a recent patient of whom a 3D computed tomography (CT) scan is available, (ii) adapt this CT to obtain an approximated CT of *p*, and (iii) reconstruct the 3D dose distribution of *p* by applying the treatment plan of *p* on the approximated CT. Here, we present first results on (i): we build a ground-truth notion of distance between patients based on CTs and adopt multivariate linear regression as a first, rough indicator of feature importance and similarity prediction.

Materials & Methods

A cohort of 26 AMC patients was considered (age range at CT acquisition: 2.4y to 5.3y), mainly diagnosed with Wilms' tumor, with features: gender, age (y), diagnosis, weight (kg) and height at intake, anterior-posterior diameter (diamAP) at the center of L2, right-left diameter (diamRL) at the middle of diamAP, distance from top of T12 to bottom of L4 (all lengths in cm), nephrectomy (for each kidney: radical, partial, or none), tumor site, and bending-correction method (see below). CTs were cropped to T10-S1 and pre-processed to limit the influence of anatomically-irrelevant factors on image registration: internal air pockets, foreign objects (like implants, stents), and (spinal) bending due to positioning. We corrected the latter with a separate deformable image registration step or image rotation.

To learn how to match patients based on the aforementioned features, we first need a ground-truth notion of distance. For this, we use deformable image registration software (elastix [2], with guidelines for big structures and visual validation of the results). Based on derivatives of deformation vector fields, we compute a magnitude of deformation that is translation and image-size independent. The final distance is the average of this magnitude for matching patient 1 to 2 and vice versa. To now approximate this ground-truth distance based only on the aforementioned features, as a first rough model, we consider well-known multivariate linear regression. Because distance is defined over a patient couple, differences between features are used. Non-numeric data (i.e., gender, tumor site) is binary encoded (0

same, 1 different). We compute feature relevance with commonly-used metrics LMG and PMVD on a regression trained over all 325 combinations of 2 patients. We finally perform a leave-one-out cross-validation of a model built upon the most relevant features, as follows. For every patient and its two rankings of other patients in predicted and ground-truth distance, we compute 4 quality indicators: *head presence* (*hp*): percentage of patients in top-*k* predicted ranking that are also in top-*k* ground-truth ranking; *tail presence* (*tp*): like *hp*, but in bottom-*k*; *average displacement* (*ad*): average, converted to percentage, over all top-*k* predicted patients, of their position beyond top-*k* ground-truth ranking (0 if in head of ground-truth ranking); *worst displacement* (*wd*): like *ad*, but only largest displacement.

Results

The LMG and PMVD metrics agree on the four most relevant features: diamRL (contributing 0.22 or 0.32 to R^2=0.78 according to LMG or PMVD, respectively), diamAP (0.22 or 0.25), weight (0.21 or 0.18) and height (0.05 or 0.01). Linear regression based on only the first three relevant features (there is a marked drop of relevance to the fourth feature) and trained over all data, gives an adjusted R^2=0.75 ($P < 10^{-6}$). Coefficients are: 0.75 for diamRL (std. err. 0.11), 1.03 for diamAP (std. err. 0.07), 0.2 for weight (std. err. 0.04).

Table on the right shows quality indicators. On average, the model fails 4/5 times in terms of *hp* for *k*=1 (i.e., when considering only the best-matching patient). Performance is not as poor as it seems, however, because the top predicted patient is always within the top 20% (*wd*). Increasing *k* improves the percentages of correct predictions and *ad*, but chances of infiltration of very distant patients (ranked almost half-way among dissimilar ones) increase as well (*wd*).

k	1	2	3	4	5
hp	19.2	30.8	33.3	39.4	44.6
tp	84.6	53.9	73.1	67.3	75.4
ad	18.3	20.9	18.9	16.5	15.4
wd	18.3	32.2	35.0	39.2	42.9

Discussion & Conclusions

Results on multivariate linear regression show that diamRL, diamAP, and weight clearly stand out as explanatory features of patients' dissimilarity. Note, that although diamAP at L2 here is measured from CTs, it is not available for historically treated patients. However, the diameter at isocenter can be adopted as a surrogate. We argue that our proposal can be refined by adopting more complex, non-linear machine learning techniques and by introducing weighted 3D information on organs at risk in ground-truth distances: features like nephrectomy are of great importance in treatment planning and should hence heavily contribute. In case linear regression is to be used, we recommend considering only the top-rank predicted patient (i.e. *k*=1): such choice yields only limited worst-case dissimilarity.

References

[1] Stovall M et al. Dose reconstruction for therapeutic and diagnostic radiation exposures: use in epidemiological studies. In *Radiation Research* **166**(1), pages 141-57, 2006.
[2] Klein S et al. elastix: a toolbox for intensity based medical image registration. In *IEEE Transactions on Medical Imaging* **29**(1), pages 196-205, 2010.

Acknowledgements & Disclosures

Financial support was provided by Stichting Kinderen Kankervrij (KiKa; project # 187).
Dr. Ronckers was supported by the Dutch Cancer Society (KWF; grant # UVA2012-5517).

†Corresponding Author: Marco.Virgolin@cwi.nl