Proceedings Paper:
Bird, D., Teo, M., Casanova, N. et al. (8 more authors) (2020) PH-0410: Multi-centre, deep learning, sCT generation for anorectal cancers with AI robustness assessment. In: Radiotherapy and Oncology. ESTRO 2020, 28 Nov - 01 Dec 2020, Online. Elsevier , s221.

https://doi.org/10.1016/s0167-8140(21)00432-1

Multi-centre, deep learning, synthetic-CT generation for anal & rectal cancers with AI robustness confidence assessment

D Bird, M Teo, N Casanova, R Cooper, A Gilbert, H Mccallum, D Sebag-Montefiore, A Henry, R Speight, B Al-Qaisieh, M Nix

Background/Purpose: Synthetic CT (sCT) generation for radiotherapy dose calculation has been shown to be accurate for prostate cancer treatments, with clinical implementation now achieved in several centres worldwide. However, only limited investigation of s CT generation accuracy for other pelvic cancer sites has occurred.

This study aims to demonstrate the dosimetric accuracy of a deep learning, conditional generative-adversarial-network (cGAN), sCT generation method for anorectal cancers, from a single commercial T2-SPACE MR sequence suitable for delineations, for a large multi-centre patient cohort. This sCT solution also validates sCT generation accuracy through automatic robustness analysis tools following the AutoConfidence approach.

Methods: RT position T2-SPACE MR and planning CT scans were acquired to train and validate the cGAN sCT generation model. All MR and CT sequences were deformably registered prior to use. The cGAN model was developed in-house and trained using 30 rectal cancer patient datasets, 10 female and 20 male, acquired from centre A. The sCT generation model was validated on 10 female and 19 male datasets, 5 rectum from centre A and 24 (7 anus and 17 rectum) from centre B. VMAT treatment plans, following clinical planning protocols, were calculated on the planning CTs and recalculated on sCTs to assess dosimetric differences. HU differences were also assessed. SCT confidence assessments were also undertaken through sCT prediction confidence maps and scores generated by the cGAN.

Results: Mean absolute error (MAE) HU differences were; mean = 65HU, range = 48-93HU. Primary PTV D95 dosimetric differences were found to be clinically insignificant (mean = 0.8%, range = -0.2 to 1.7%) where clinical significance was considered to be a ≥2% (figure 1). A small systematic difference in dosimetric accuracy was noted between centres (mean: centre A = 0.2%, centre B = 0.8%). Similar dosimetric accuracy was found between male (mean =0.8%) and female (mean = 0.7%) and rectal (mean = 0.7%) and anal (mean = 0.9%) cohorts . Local difference hotspots of ≥2% were noted where rectal gas was present on a single dataset.

The AutoConfidence AI robustness feature provided confidence of sufficient sCT generation accuracy in target volumes regions, while discrepancies were highlighted (figure 2). Per-slice confidence scoring identified 11% of test slices as potentially unreliable due to lying beyond the confidence bounds of the training data set.

Conclusion: The deep-learning cGAN model is a feasible and generalisable method for anal and rectal sCT generation using T2-SPACE sequences. SCTs with clinically acceptable dosimetric accuracy were produced from data acquired from multiple centres, for male and female anatomy and for anal and rectal cancer sites. The AutoConfidence AI approach enabled individual patient sCT robustness, through highlighting sCT slices and global sCT datasets with lower HU accuracy.

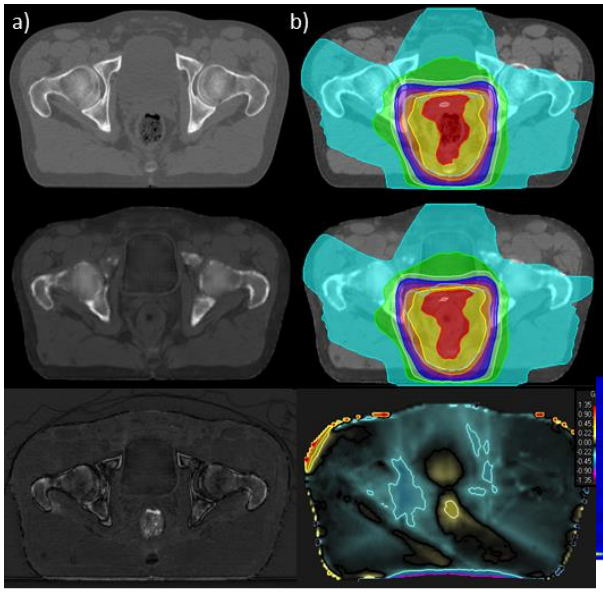

Figure 1: Planning CT (upper), synthetic CT (middle) and their differences (lower) in terms of a) HU and b) dose with a prescription dose of 45Gy. HU and dosimetric differences can be seen to be minimal, where light blue and yellow contours indicate dosimetric differences of ±0.5%.

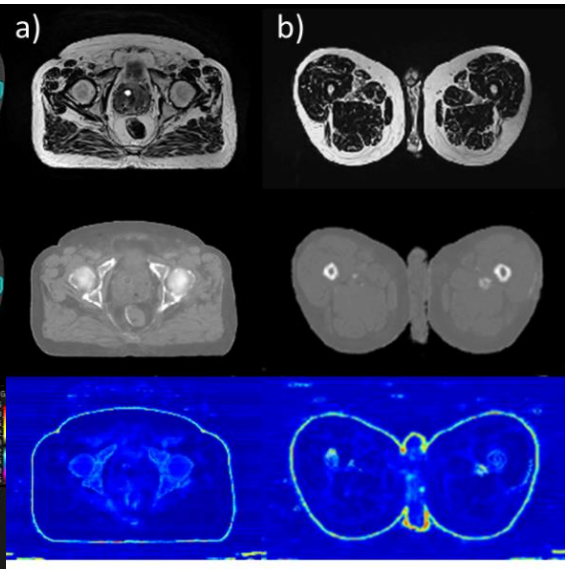

Figure 2: t2w-SPACE MR (upper), synthetic CT (middle) and AutoConfidence discriminator maps (lower) for two representative slices: a) Inlier and b) Outlier classification on training data similarity score. AutoConfidence identifies bone-density is slightly low in sCT and indicates areas of incorrect density assignment which are more prominent in the outlier case.