JULY 12–16 2020 VIRTUAL
JOINT AAPM | COMP MEETING
EASTERN TIME [GMT-4]

A Comprehensive Framework for Radiotherapy Treatment Plan Quality Evaluation in Large Multiple Institution Dataset

L Yuan[a], R Kapoor[a,b] W Sleeman[a.b], M Hagan[a,b], J Palta[a,b], M Rosu-Bubulac[a]
[a]Virginia Commonwealth University Medical Center, Richmond, VA
[b]VHA National Radiation Oncology Program Office, Richmond, VA

INTRODUCTION

With the adaptation of standardized nomenclature and advancement in IT infrastructure in RT, the collection of large amount of annotated data related to RT treatment planning from large number of institutions is becoming available. Effort is underway to perform comprehensive comparative RT plan quality evaluation for large datasets. Mathematical models have been proposed to capture the correlation between plan dosimetric parameters and patient anatomical variations as baseline for plan quality evaluation. However, despite recent development in machine learning methods, there is no systematic study on how to handle the heterogeneity in multi-institution data, e.g. patient anatomy, physician preference, contouring variability, treatment planner experience. This study developed a comprehensive framework to capture plan quality variations in the presence of patient anatomical variation and large data heterogeneity. It will enable evidence based comparative plan quality evaluation among different institutions.

PURPOSE

To develop a framework to evaluate the quality of treatment plans in a multi-institutional setting with the presence of patient anatomical variations and large data heterogeneity.

METHOD

This study analyzed a de-identified dataset consisting of 395 prostate cases from 38 institutions. Five OARs are modeled: bladder, rectum, left/right femoral head and bowel. The models look at anatomical factors such as Distance-to-target Histogram (DTH) and other volumetric factors and utilize a step-wise regression method for feature selection.

In the first step, a single global model was trained for each OAR. However, the single model doesn't work very well for some OAR such as rectum, possibly due to the data heterogeneity among different institutions. In order to improve the modeling, We have developed a workflow to first group the multiple institution data into clusters automatically by hierarchical clustering. Separate models are then trained for each cluster. The similarity between different institutions data was defined as the improvement (or degradation) of the model prediction accuracy trained on the combined data from these institutions compared with the models trained on each individual one. Mathematically, the similarity (or distance) is expressed as the improvement (or degradation) in determination coefficient R2 when we combine the data together:

$$d(\{X_{Inst1}\},(\{X_{Inst2}\}) = R2(\{X_{Inst1},X_{Inst2}\}) - \max(R2(\{X_{Inst1}\}),R2(\{X_{Inst1}\}))$$

The clusters continue growing pair-wise until the model prediction accuracy degrades significantly. The model performances are compared with the global models which were trained by all the institution data together.

RESULTS

Clustered regression for heterogeneous data set: Hierarchical clustering groups the close-by data points pair-wise from bottom up to form a multi-level structure of clusters. The cluster dendrogram for our dataset is shown in figure 1. Note that leaf No. 34 itself consists of 5 institutions. The 5 clusters are determined using an inconsistency coefficient of 0.8. The clustered model divided the 38 institutions into 5 clusters, with 16, 7, 6, 5, 4 institutions in each cluster respectively.

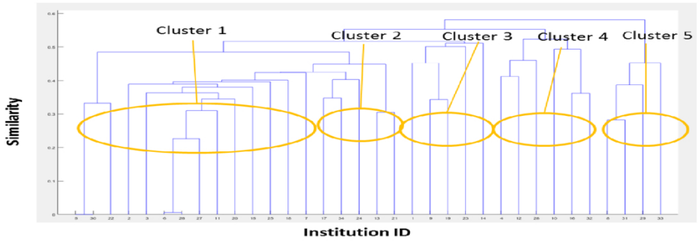

Fig. 1. Cluster dendrogram from the hierarchical clustering of the dataset.

Prediction of dosimetric baseline by the clustered models: Three relevant QUANTEC dosimetric indices D15%, D25%, and mean dose (Dmean) were predicted for bladder, rectum and bowel using the models. The determination coefficients (R2) with the global models for bladder and bowel, and for with both global and clustered models for rectum are listed in table 1. The clustered model significantly improved the prediction accuracy for rectum. This improvement in accuracy likely results from the differences in the contouring, prescription or planning preference for rectum among different institutions.

Table 1. determination coefficients for the OARs with different type of models

	Coefficients of Determination (R2)		
OAR	D15%	D25%	Dmean
Bladder single model	0.5	0.62	0.69
Bowel single model	0.56	0.53	0.51
Rectum single model	0.46	0.46	0.41
Rectum cluster model	0.64	0.58	0.51

The correlation between the predicted bladder and rectum mean dose vs. actual dosimetrics by the global model and the clustered model is shown in figure 2. Each color represents a cluster. The determination coefficients (R2) are shown on top of the figures.

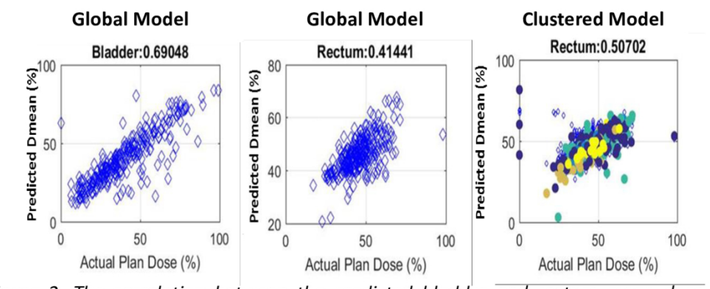

Figure 2. The correlation between the predicted bladder and rectum mean dose vs. actual dosimetrics by the global model and the clustered model

CONCLUSIONS

The proposed plan evaluation framework can capture plan quality variations in the presence of patient anatomical variation and large data heterogeneity.

CONTACT INFORMATION

Lulin.Yuan@vcuhealth.org