AAPM 2022
JULY 10–14 | WASHINGTON, DC
64TH ANNUAL MEETING & EXHIBITION

# The Dataset Heterogeneity Matters: A Machine Learning Study of Dataset Conformation Effects On Model Performance for Dose Deliverability Prediction


**Paulo Quintero[1,2], Yongqiang Cheng[2], David Benoit[2], Craig Moore[1], Andrew W Beavis[1,2,3]**

[1]Medical Physics Department, Queen's Centre for Oncology, Hull University Teaching Hospitals NHS Trust, Cottingham, HU16 5JQ, UK;
[2]Faculty of Sciences and Engineering, University of Hull, Cottingham road, Hull, HU16 7RX, UK;
[3]Faculty of Health and Well Being, Sheffield-Hallam University, Collegiate Crescent, Sheffield, S10 2BP


UNIVERSITY OF HULL
NHS Hull University Teaching Hospitals NHS Trust

## PURPOSE / OBJECTIVES

- Reported machine learning (ML) applications in radiotherapy presented unbalanced datasets for gamma passing rates (GPR) predictions, having plans with different anatomic regions, treatment units, and techniques.
- This work evaluates the influence on ML models prediction performance of various datasets containing plans with different, controlled, treatment factors (number of arcs and treatment unit) and the same anatomical region.

## MATERIAL & METHODS

- 945 prostate plans and 309 predictor features (complexity metrics and radiomics)
- 13 controlled datasets:
  - one randomly assembled dataset as reference
  - 12 datasets controlling:
    - six datasets controlling the number of treatments with one and two arcs
    - six datasets controlling for treatment unit (Halcyon or TrueBeam).
- Models: Random Forest (RF), extreme-gradient boosting (XG-Boost), and neural network (NN)
- Evaluation metrics: The area under the ROC curve (ROC-AUC)

## RESULTS

| Model | Reference Model (Hybrid) | Dataset Condition_1 | Dataset Condition_2 |
|---|---|---|---|
| RF | 0.78 ± 0.15 | 0.84 ± 0.13 | 0.82 ± 0.05 |
| XG-Boost | 0.65 ± 0.13 | 0.85 ± 0.07 | 0.87 ± 0.09 |
| NN | 0.87 ± 0.03 | 0.90 ± 0.05 | 0.96 ± 0.04 |

**ML models** trained with **homogeneous datasets** allow better data generalisation **increasing** the classification accuracy and **prediction reliability**


pquinterome@gmail.com


## RESULTS

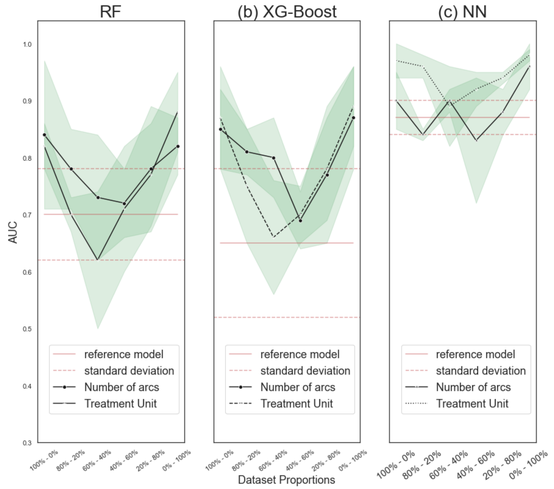


Figure 1. ROC-AUC results and its standard deviations for (a) RF, (b) XG-Boost, and (c) NN models, considering the reference dataset, each heterogeneity source (Number of arcs // Treatment unit)

Figure 2. Features variations for each specific-dataset based models (RF) conformed with plans with one and two arcs