

B-12

Influence of patient cohort size on PFS prediction accuracy using baseline SSTR image models

Victor Santoro-Fernandes¹, Girish Dodda¹, Renuka Iyer², Christos Fountzilas², Robert Jeraj^{1,3}

¹Department of Medical Physics, School of Medicine and Public Health, University of Wisconsin, Madison, WI, USA; ²Division of GI Medicine, Department of Medicine, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA; ³Carbone Cancer Centre, University of Wisconsin, Madison, WI, USA.

BACKGROUND

The prediction of progression-free survival (PFS) after peptide receptor radionuclide therapy (PRRT) in neuroendocrine tumor (NET) patients based on baseline somatostatin receptor (SSTR) imaging holds significant promise for personalized treatment strategies. However, small sample sizes remain common, raising concerns about overfitting and model reliability. This study evaluates the influence of patient cohort size on PFS prediction accuracy using baseline SSTR image models.

METHODS

Eighty-one NET patients underwent baseline PET/CT [⁶⁸Ga]Ga-DOTA-TATE imaging (median: 65 days, IQR:83) prior to initiating PRRT. Patients received between 1 and 4 cycles of [¹⁷⁷Lu]Lu-DOTA-TATE PRRT, and PFS was subsequently monitored. Patients were classified into poor and good responders (PFS=26 months threshold). A quantitative analysis of all lesions was performed on the baseline images, from which patient-level features were extracted. The top four features with the highest concordance index to PFS were selected for training a multivariate linear regression model. This process was repeated with progressively larger training cohorts. The PFS was predicted on a 16-patient hold-out test population and on the training population (using Leave-One-Out cross-validation). Model classification performance was assessed using the root mean squared error (RMSE) and the area under the receiver operating characteristic curve (AUC).

RESULTS

The RMSE on the training cohort decreased from 59(CI=0,120) with 5 patients to 23(CI=18, 28) with 10 patients. From 10 patients onwards, the RMSE decreased monotonically, reaching 16 (CI=15, 17) when 65 patients were used for model training. On the other hand, the AUC increased pronouncedly from 0.73 (CI=0.45,0.95) with 5 patients to 0.89 (CI=0.80, 0.95) with 10 patients. From 10 patients onwards, the AUC also decreased monotonically, reaching 0.70 (CI=0.67, 0.73) when 65 patients were used for model training. The RMSE and AUC confidence interval range decreased substantially as the training cohort increased from 5 to 65 patients: from 120 to 2 for RMSE and from 0.40 to 0.06 for AUC. The predictions on the 16 patients hold-out population were constant with an average RMSE of 20 (CI=21,17.5) and AUC of 0.62 (CI=0.55, 0.70).

CONCLUSIONS

Our findings highlight the overestimation of predictive power in small patient cohorts and emphasize the importance of using larger and hold-out patient populations to improve model generalizability.

Given the relatively rare incidence of NETs, aggregating patient populations from multiple centers is crucial for developing accurate models. Future work will focus on collaboratively expanding patient cohorts to understand the applicability of PFS prediction models for NET patients.

ABSTRACT ID 33461

