Luining W.¹, Cysouw M.C.F.², Oprea-Lager D.E.², Vis A.N.¹, Boellaard R.²

¹Amsterdam University Medical Center, Dept. of Urology, Amsterdam, The Netherlands, ²Amsterdam University Medical Center, Dept. of Radiology and Nuclear Medicine, Amsterdam, The Netherlands

Introduction & Objectives: Prostate-specific membrane antigen (PSMA) PET is a promising imaging modality for radiomics feature extraction and machine learning-based analysis. However, external validation of machine learning models is lacking. The aim of this study was to perform both internal and external validation of machine learning models based on ¹⁸F-DCFPyL PET radiomics for the prediction of lymph node involvement (LNI), extracapsular extension (ECE), and post-operative Gleason Score (GS) in primary prostate cancer (PCa) patients.

Materials & Methods: Patients with intermediate- to high-risk PCa who underwent ¹⁸F-DCFPyL PET/CT imaging followed by radical prostatectomy with extended pelvic lymph node dissection were enrolled. The training dataset included 72 patients, the internal validation dataset 24 patients and the external validation dataset 27 patients. Primary prostate tumors were delineated semi-automatically on PET and radiomic features (n=480) were extracted. Conventional PET metrics were derived for comparative analysis. Segmentation, preprocessing, and machine learning methods were optimized in 10-times repeated 5-fold cross-validation on the training dataset. ComBat harmonization was applied to external radiomics data to mitigate center effects. Optimized models were then tested on the combined internal and external validation dataset. The influence of adding clinical parameters to the radiomics model was investigated. Model performance was assessed using the receiver-operating-characteristics curve integral (AUC).

Results: Despite high cross-validated AUCs in the training dataset for each outcome (0.88 for LNI, 0.79 for ECE, and 0.89 for GS), the radiomics-based machine learning models could only significantly predict GS in the combined validation datasets with an AUC of 0.84 (p<0.05). Validation AUCs for LNI and ECE prediction were not significant (0.57 and 0.63, respectively). Conventional PET metrics-based models had similar AUCs for LNI (0.60, p>0.05) and ECE (0.66, p>0.05), but a lower AUC for GS (0.75, p<0.05). In general, ComBat harmonization improved external validation AUCs (range -0.03 to 0.18). Adding clinical parameters to the radiomics data did not improve model performance.

Conclusions: In multicenter validation, previously proposed machine learning models using ¹⁸F-DCFPyL PET radiomics could predict high post-operative Gleason Score but not LNI or ECE in intermediate- to high-risk prostate cancer patients. These results underline the need for external and/or multicenter validation of PET radiomics-based machine learning models analyses to assess their reproducibility.