SIIM CMIMI24
Oct 21-22 | Boston, MA
Conference On Machine Intelligence In Medical Imaging

Replicating and Validating Radiomics-Based Prediction of PD-L1 Expression Status in NSCLC Patients

Anna Theresa Stüber, MSc, LMU University Hospital, LMU Munich; Maurice Heimer, MD; Clemens C. Cyran, MD; Michael Ingrisch, PhD

Introduction/Background

The purpose of this study is to investigate the predictive value of radiomics in determining PD-L1 expression (positive vs. negative) status among NSCLC patients using an external [18F]FDG PET/CT dataset. Specifically, we aim to replicate and validate the radiomics-based machine learning model proposed by Zhao et al.* to address concerns related to reproducibility and replicability in radiomics research. * Zhao et al. Predicting PD-L1 expression status in patients with non-small cell lung cancer using [18F]FDG PET/CT radiomics. EJNMMI Res. 2023 Jan 22;13(1):4. doi: 10.1186/s13550-023-00956-9. PMID: 36682020; PMCID: PMC9868196.

Methods/Intervention

We analyzed a cohort of 254 NSCLC patients (86 = 33,9 % negative, 168 = 66,1 % positve PD-L1 status) who underwent [18F]FDG PET/CT imaging, utilizing two distinct image segmentation methods: solid component-based segmentation (LUT) with lung tissue window (W1500/L-600) and attenuation-corrected PET volume, and a conservative, smaller segmentation (CON) with soft-tissue window (W400/40) and corresponding PET volume. We replicated two radiomics-based models (“Rad-score” and “complex model”) provided by Zhao et al. for both segmentation sets, along with their clinical stage model. Performance evaluation is based on 10-fold cross-validation and the Area Under the Curve (AUC).

Results/Outcome

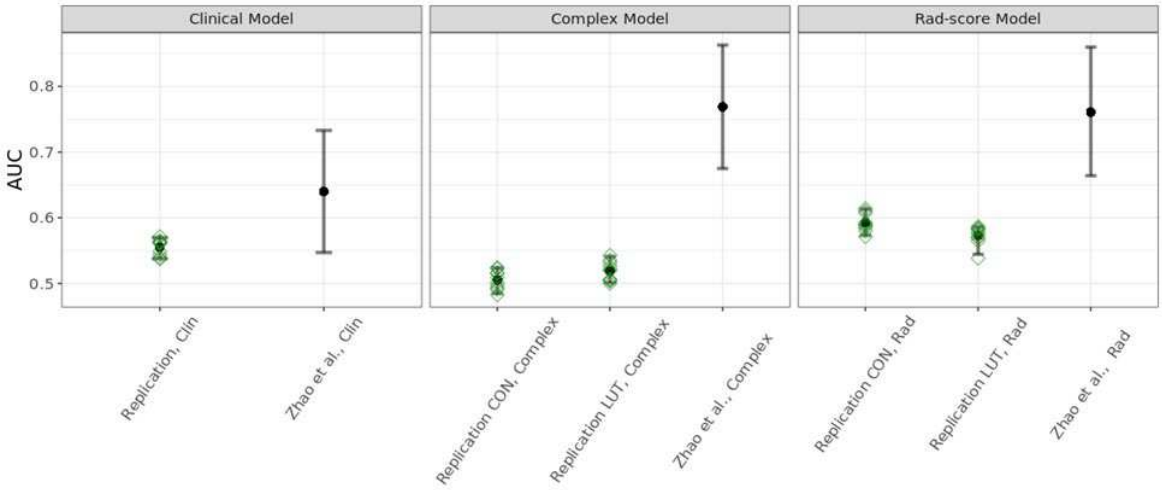
Performance analysis of the Rad-score model revealed a mean AUC of 0.593 (95% CI: 0.573 - 0.613) for CON segmentation and 0.573 (0.544 - 0.586) for LUT segmentation, both falling below the reported mean AUC of 0.761 (0.664 - 0.860) by Zhao et al. Similarly, for the complex model, we achieved mean AUCs of 0.505 (0.485 - 0.524) and 0.519 (0.501 - 0.541), respectively, whereas Zhao et al. reported a mean AUC of 0.769 (0.675 - 0.863).

Conclusion

Our study failed to replicate the findings of the previous study. In particular, the original model achieved very poor prediction performance on our dataset, which is compatible with the original dataset.These findings underscore the challenges in replicating radiomics-based predictive models across different datasets and highlight the importance of rigorous validation in ensuring clinical utility.

Statement of Impact

These findings underscore the challenges in replicating radiomics-based predictive models across different datasets and highlight the importance of rigorous validation in ensuring clinical utility.



Comparison of AUC means and 95% CIs: Replication (STW and LTW) versus reported AUCs by Zhao et al. [1]. Green squares represent AUCs per CV fold, black dots show mean AUC values, and whiskers indicate the 95% CI.

Keywords

replication; radiomics; Machine learning evaluation; PET/CT; lung cancer