TransRP: Transformer-based PET/CT feature extraction incorporating clinical data for recurrence-free survival prediction in oropharyngeal cancer

Baoqiang Ma, Jiapan Guo, Lisanne van Dijk, Peter van Ooijen, Stefan Both, Nanna Maria Sijtsema

Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands

The growing number of subtypes and treatment options for oropharyngeal squamous cell carcinoma (OPSCC), a common type of head and neck cancer (HNC), highlights the need for personalized therapies. Prognostic outcome prediction models can identify different risk groups for investigation of intensified or de-escalated treatment strategies. Convolutional neural networks (CNNs) have been shown to have improved predictive performance compared to traditional clinical and radiomics models by extracting comprehensive and representative features. However, CNNs are limited in their ability to learn global features within an entire volume. In this study, we propose a Transformer-based model for predicting recurrence-free survival (RFS) in OPSCC patients, called TransRP. TransRP consists of a CNN encoder to extract rich PET/CT image features, a Transformer encoder to learn global context features from CNN features, and a fully connected network to incorporate clinical data for outcome prediction. We investigated three different methods for combining clinical features into TransRP. The experiments were conducted using the public HECKTOR 2022 challenge dataset, which includes pretreatment PET/CT scans, Gross Tumor Volume masks, clinical data, and RFS for OPSCC patients. The dataset was split into a test set (n=120) and a training set (n=362) for five-fold cross-validation. The results show that TransRP achieved the highest test concordance index of 0.698 (an improvement > 0.02) in RFS prediction compared to several clinical and CNN-based methods. In addition, we found that concatenating clinical features with those obtained from the Transformer encoder performed better than using the Transformer encoder to extract features from both clinical and image features. This finding contradicts previous studies where Transformer encoders improved HNC survival prediction by better combining multi-modality information from clinical and image data. The code for this study is available at https://github.com/baoqiangmaUMCG/TransRP.