Hubert Szymon Gabryś
Dr. sc. hum.

Machine learning using radiomics and dosiomics for normal tissue complication probability modeling of radiation-induced xerostomia

Fach: Medizinische Physik
Doktorvater: Prof. Dr. Markus Alber

In routine clinical practice, the risk of xerostomia is typically managed by limiting the mean radiation dose to parotid glands. This approach used to give satisfying results. In recent years, however, several studies have reported mean-dose models to fail in the recognition of xerostomia risk. This can be explained by a strong improvement of overall dose conformality in radiotherapy due to recent technological advances, and thereby a substantial reduction of the mean dose to parotid glands. This thesis investigated novel approaches to building reliable normal tissue complication probability (NTCP) models of xerostomia in this context.

For the purpose of the study, a cohort of 153 head-and-neck cancer patients treated with radiotherapy at Heidelberg University Hospital was retrospectively collected. The predictive performance of the mean-dose to parotid glands was evaluated with the Lyman-Kutcher-Burman (LKB) model. In order to examine the individual predictive power of predictors describing parotid shape (radiomics), dose shape (dosiomics), and demographic characteristics, a total of 61 different features was defined and extracted from the DICOM files. These included the patient's age and sex, parotid shape features, features related to the dose-volume histogram, the mean dose to subvolumes of parotid glands, spatial dose gradients, and three-dimensional dose moments. In the multivariate analysis, a variety of machine learning algorithms was evaluated: 1) classification methods, that discriminated patients between a high and a low risk of complication, 2) feature selection techniques, that aimed to select a number of highly informative covariates from a large set of predictors, 3) sampling methods, that reduced the class imbalance, 4) data cleaning methods, that reduced noise in the data set. The predictive performance of the models was validated internally, using nested cross-validation, and externally, using an independent patient cohort from the PARSPORT clinical trial.

The LKB model showed fairly good performance on mild-to-severe (G1+) xerostomia predictions. The corresponding dose-response curve revealed that even small doses to parotid glands increase the risk of xerostomia and should be kept as low as possible. For the patients who did develop moderate-to-severe (G2+) xerostomia, the mean dose was not an informative predictor, even though the efficient sparing of parotid glands allowed to achieve low G2+ xerostomia rates. The features describing the shape of a parotid gland and the shape of a dose proved to be highly predictive of xerostomia. In particular, the parotid volume and the spatial dose gradients in the transverse plane explained xerostomia well. The results of the machine learning algorithms comparison showed that a particular choice of a classifier and a feature

selection method can significantly influence predictive performance of the NTCP model. In general, support vector machines and extra-trees achieved top performance, especially for the endpoints with a large number of observations. For the endpoints with a smaller number of observations, simple logistic regression often performed on a par with the top-ranking machine learning algorithms. The external validation showed that the analyzed multivariate models did not generalize well to the PARSPORT cohort. The only features that were predictive of xerostomia both in the Heidelberg (HD) and the PARSPORT cohort were the spatial dose gradients in the right-left and the anterior-posterior directions. Substantial differences in the distribution of covariates between the two cohorts were observed, which may be one of the reasons for the weak generalizability of the HD models.

The results presented in this thesis undermine the applicability of NTCP models of xerostomia based only on the mean dose to parotid glands in highly conformal radiotherapy treatments. The spatial dose gradients in the left-right and the anterior-posterior directions proved to be predictive of xerostomia both in the HD and the PARSPORT cohort. This finding is especially important as it is not limited to a single cohort but describes a general pattern present in two independent data sets. The performance of the sophisticated machine learning methods may indicate a need for larger patient cohorts in studies on NTCP models in order to fully benefit from their advantages. Last but not least, the observed covariate-shift between the HD and the PARSPORT cohort motivates, in the author's opinion, a need for reporting information about the covariate distribution when publishing novel NTCP models.