SUPPLEMENTARY MATERIAL 2

Nested Cross-Validation for Generalized Linear Model Analysis

Nested cross-validation comprises an inner five-fold cross-validation for feature selection and training of the radiomics-based classifier and an outer five-fold cross-validation to validate the radiomics generalized linear model (GLM) classifier. In this nested cross-validation, the study cohort was semi-randomly divided into five subsets for outer cross-validation, with stratification of the squamous cell carcinoma (SCC) to lymphoma ratio. On the four subsets allocated as the training set, feature selection and training of the radiomics GLM classifier was conducted by inner five-fold cross-validation. Before training, z-score normalization was performed on the radiomic features, based on the means and standard deviations of each feature from the four subsets allocated as the training set. During the GLM training with elastic net regularization, feature selection was conducted by assigning non-zero coefficient values to the relevant features only, while other irrelevant features were nullified by assigning zero coefficient values. Hyperparameters were optimized via grid search during the training. Then, the trained radiomics classifier was validated on the rest of the subsets of the outer cross-validation, which was allocated as the test set, and the predicted probability of SCC or lymphoma diagnosis was acquired for each sample of the test set. This process was repeated so that each study cohort was allocated as a test set once to complete one round of the five-fold cross-validation process, which yielded predicted probabilities based on the radiomics classifier for each cohort sample. These predicted probabilities were used to calculate the area under the receiver operating characteristic curve of the radiomics classifier. Youden’s index was applied to the predicted probabilities to calculate sensitivity, specificity, and accuracy. To assess the overall degree of contribution from each radiomic feature in the GLM classifier, the coefficients of the GLM classifiers and the number of times that each feature was selected during the five-fold outer cross-validation were recorded.