Head & Neck

Privacy-preserving distributed learning of radiomics to predict overall survival and HPV status in head and neck cancer.

Bogowicz M, Jochems A, Deist TM, Tanadini-Lang S, Huang SH, Chan B, Waldron JN, Bratman S, O'Sullivan B, Riesterer O, Studer G, Unkelbach J, Barakat S, Brakenhoff RH, Nauta I, Gazzani SE, Calareso G, Scheckenbach K, Hoebers F, Wesseling FWR, Keek S, Sanduleanu S, Leijenaar RTH, Vergeer MR, Leemans CR, Terhaard CHJ, van den Brekel MWM, Hamming-Vrieze O, van der Heijden MA, Elhalawani HM, Fuller CD, Guckenberger M, Lambin P.

Sci Rep. 2020 Mar 11;10(1):4542. doi: 10.1038/s41598-020-61297-4.

Abstract

A major challenge in radiomics is assembling data from multiple centres. Sharing data between hospitals is restricted by legal and ethical regulations. Distributed learning is a technique, enabling training models on multicentre data without data leaving the hospitals ("privacy-preserving" distributed learning).

This study tested feasibility of distributed learning of radiomics data for prediction of two-year overall survival and HPV status in head and neck cancer (HNC) patients. Pre-treatment CT images were collected from 1174 HNC patients in six different cohorts. A total of 981 radiomic features were extracted using Z-Rad software implementation. Hierarchical clustering was performed to preselect features. Classification was done using logistic regression. In the validation dataset, the receiver operating characteristics (ROC) were compared between the models trained in the centralised and distributed manner. No difference in ROC was observed with respect to feature selection. The logistic regression coefficients were identical between the methods (absolute difference <10-7). In comparison of the full workflow (feature selection and classification), no significant difference in ROC was found between centralised and distributed models for both studied endpoints (DeLong $p > 0.05$).

In conclusion, both feature selection and classification are feasible in a distributed manner using radiomics data, which opens a new possibility for training more reliable radiomics models.