Supplementary

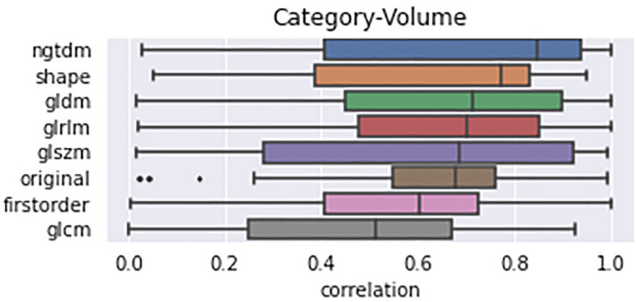

Figure S1 Spearman correlation between tumor volume and different feature categories (aggregated over all sub-features and pre-processing methods).

https://dx.doi.org/10.21037/tlcr-22-248

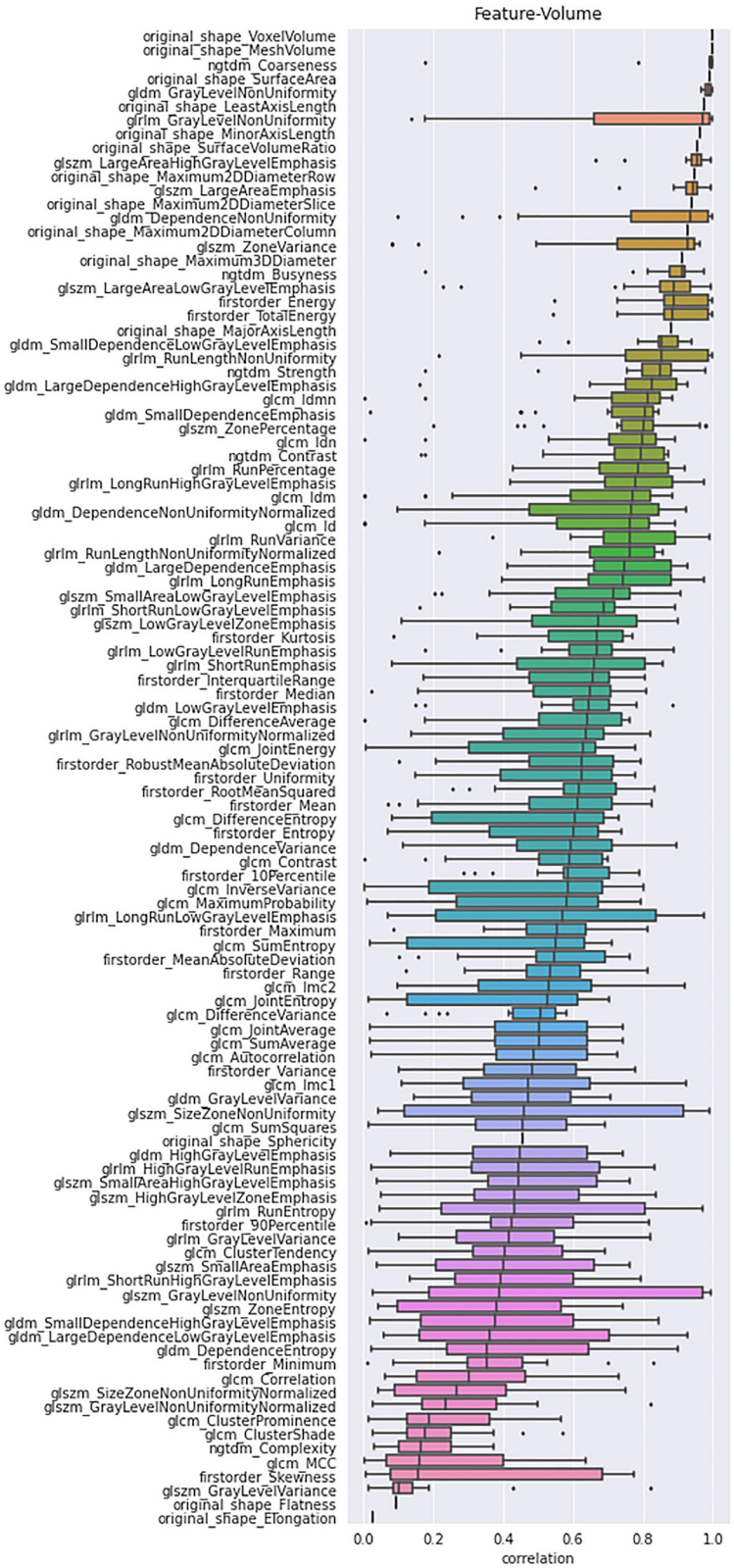

Figure S2 Spearman correlation between tumor volume and different radiomic features (aggregated over the different image preprocessing filters).

https://dx.doi.org/10.21037/tlcr-22-248

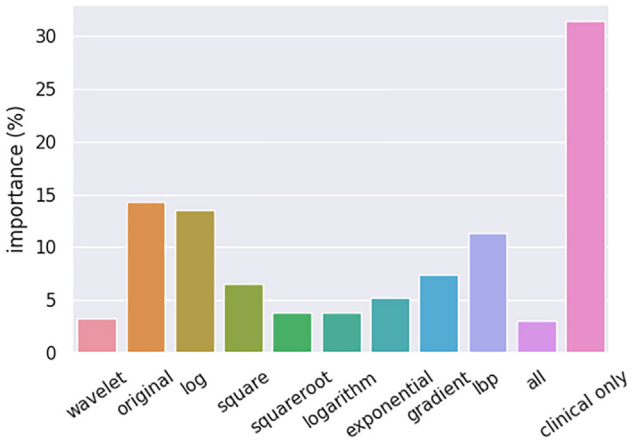

Figure S3 CatBoost volume feature importance from step 4 in the training pipeline.

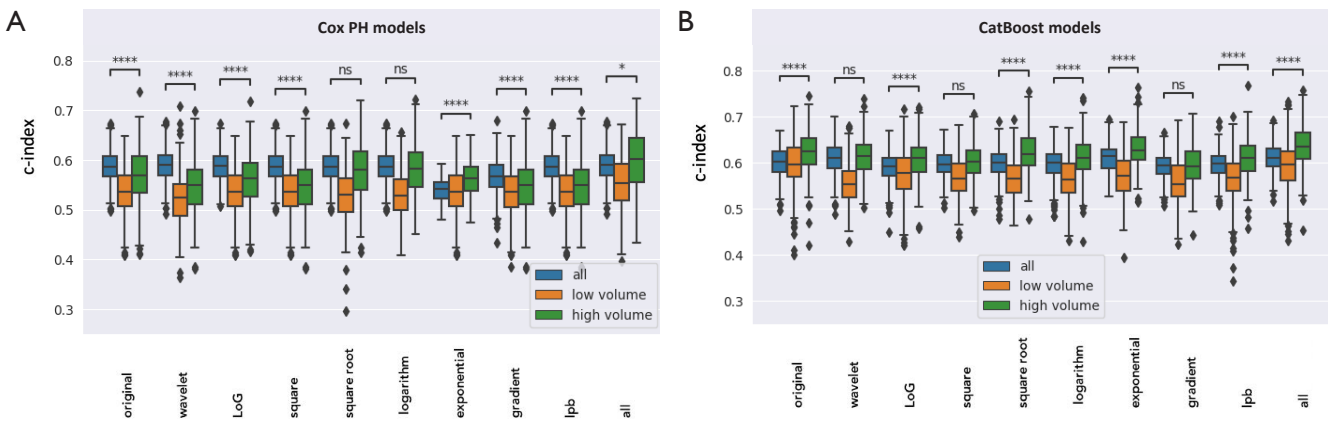

Figure S4 Cox (A) and CatBoost (B) model performance with different sets of patients based on lesion volume. Low-volume and high-volume patients were separated with respect to the median volume (30.3 cm3). All boxes are aggregated from 64 different 5-fold (shuffled) cross validation splits (with constant random seed, so that every different model is trained and validated on the exact same splits). Horizontal bars indicate the significance of the Mann-Whitney U-test ("all" vs "high volume", "ns": not significant). Bonferroni FDR correction has been applied to all P values. All low-volume models have significantly worse performance than the high-volume models with P≤0.0001 (bars not pictured for clarity).

https://dx.doi.org/10.21037/tlcr-22-248