

# Supplement

## **Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma**

Marta Bogowicz<sup>a</sup>, Oliver Riesterer<sup>a</sup>, Luisa Sabrina Stark<sup>a</sup>, Gabriela Studer<sup>a</sup>, Jan Unkelbach<sup>a</sup>, Matthias Guckenberger<sup>a</sup>, Stephanie Tanadini-Lang<sup>a</sup>

*<sup>a</sup>Department of Radiation Oncology, University Hospital Zurich, University of Zurich, Switzerland*

## Radiomic features

The following approaches were used to determine radiomics: histogram of intensities analysis, texture analysis (the Gray Level Co-occurrence Matrix, the Neighborhood Gray Tone Difference Matrix, the Gray Level Size Zone Matrix and the Gray Level Run Length Matrix), shape analysis and wavelet transform analysis.

### Histogram of intensities

The parameters from the histogram of intensities were calculated in the perfusion maps before a discretization. Let  $X$  denotes the intensities of the 3D image with  $N$  voxels.  $\bar{X}$  - mean of  $X$ ,  $N_g$  - number of gray levels in the image,  $p_i$  - the occurrence probability of gray level  $i$ .

- 1)  $mean = \frac{1}{N} \sum_{i=1}^N X_i$
- 2)  $standard\ deviation = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}$
- 3)  $variance = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2$
- 4)  $coefficient\ of\ variation = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}}{\bar{X}}$
- 5)  $skewness = \frac{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^3}{\left(\sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}\right)^3}$
- 6)  $kurtosis = \frac{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^4}{\left(\sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2}\right)^4} - 3$
- 7)  $mean\ absolut\ deviation = \frac{1}{N} \sum_{i=1}^N |X_i - \bar{X}|$
- 8)  $robust\ mean\ absolut\ deviation = \frac{1}{N} \sum_{i=1}^{N_{10-90}} |X_{10-90,i} - \bar{X}_{10-90}|$

where:  $N_{10-90}$  - number of voxel in the range from and 10<sup>th</sup> percentile and 90<sup>th</sup> percentile,  $\bar{X}_{10-90}$  - mean value of voxel in the range from 10<sup>th</sup> percentile and 90<sup>th</sup> percentile

- 9)  $energy = \sum_{i=1}^N X_i^2$
- 10)  $entropy = -\sum_{i=1}^{N_g} p_i \log_2 p_i$
- 11)  $root\ mean\ square = \sqrt{\frac{energy}{N}}$
- 12)  $uniformity = \sum_{i=1}^{N_g} p_i$
- 13)  $minimum\ value = \min(X)$
- 14)  $maximum\ value = \max(X)$
- 15)  $median$  – the median value of  $X$
- 16)  $10^{th}percentile$  – 10<sup>th</sup> percentile of  $X$
- 17)  $90^{th}percentile$  – 90<sup>th</sup> percentile of  $X$
- 18)  $interquartile\ range = 90^{th}percentile - 10^{th}percentile$
- 19)  $range = \max(X) - \min(X)$

### The Gray Level Co-occurrence Matrix (GLCM)

The parameters from the Gray Level Co-occurrence Matrix [1] were calculated in all 26 directions with a distance of one voxel. The final parameters were the average of all directions. If one of the voxels had a 'not a number' value the pair was not taken into account in the calculations. Let  $P_{ij}$  denotes the  $(i, j)$  entry in the Gray Level Co-occurrence Matrix,  $N_g$  - number of gray tones in a studied structure,  $P_{xi} = \sum_{j=1}^{N_g} P_{ij}$ ,  $P_{yj} = \sum_{i=1}^{N_g} P_{ij}$ ,  $P_{x+y}(k) = \sum_{j=1}^{N_g} \sum_{i=1}^{N_g} P_{ij}$ , where  $k = i + j$ ,  $P_{x-y}(k) = \sum_{j=1}^{N_g} \sum_{i=1}^{N_g} P_{ij}$ , where  $k = |i - j|$ .

$$20) \text{ energy} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij}^2$$

$$21) \text{ contrast} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - j)^2 P_{ij}$$

$$22) \text{ correlation} = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ij P_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y}$$

where:  $\mu_x = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} iP_{ij}$ ,  $\mu_y = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} jP_{ij}$ ,  $\sigma_x = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij}(i - \mu_x)^2$ ,

$$\sigma_y = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij}(j - \mu_y)^2$$

$$23) \text{ sum of squares} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 P_{ij}$$

where  $\mu$  – mean of  $P$

$$24) \text{ inverse difference moment (homogeneity)} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P_{ij}}{1 + (i - j)^2}$$

$$25) \text{ inverse difference moment normalized} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P_{ij}}{1 + \frac{(i - j)^2}{N_g^2}}$$

$$26) \text{ inverse difference} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P_{ij}}{1 + |i - j|}$$

$$27) \text{ inverse difference normalized} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P_{ij}}{1 + \frac{|i - j|}{N_g}}$$

$$28) \text{ sum average} = \sum_{i=2}^{2N_g} i \cdot P_{x+y}(i)$$

$$29) \text{ sum entropy} = - \sum_{i=2}^{2N_g} P_{x+y}(i) \log_2 P_{x+y}(i)$$

$$30) \text{ sum variance} = \sum_{i=2}^{2N_g} P_{x+y}(i) \cdot (i - \text{sum average})^2$$

$$31) \text{ entropy} = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} \log_2 P_{ij}$$

$$32) \text{ difference entropy} = - \sum_{i=0}^{N_g-1} P_{x-y}(i) \log_2 P_{x-y}(i)$$

$$33) \text{ information measure of correlation 1 (IMC1)} =$$

$$\frac{- \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} \log_2 P_{ij} - \left( - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} \log_2 P_{xi} P_{yj} \right)}{\max \left\{ \left( - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{xi} \log_2 P_{xi} \right), \left( - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{yj} \log_2 P_{yj} \right) \right\}}$$

$$34) \text{ information measure of correlation 2 (ICM2)} =$$

$$\sqrt{1 - \exp \left[ -2 - \left( - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} \log_2 P_{xi} P_{yj} \right) + \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} \log_2 P_{ij} \right]}$$

$$35) \text{ maximal correlation coefficient (MCC)} = \sqrt{\text{second largest eigenvalue of } \sum_{k=1}^{N_g} \frac{P_{ik} P_{jk}}{P_{xi} P_{ji}}}$$

- 36) *joint maximum* =  $\max(P_{ij})$
- 37) *joint average* =  $\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} i P_{ij}$
- 38) *difference average* =  $\sum_{i=0}^{N_g-1} i P_{x-y}(i)$
- 39) *difference variance* =  $\sum_{i=0}^{N_g-1} P_{x-y}(i) \cdot (i - \text{difference average})^2$
- 40) *dissimilarity* =  $\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i - j| P_{ij}$
- 41) *inverse variance* =  $2 \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P_{ij}}{(i-j)^2}$
- 42) *autocorrelation* =  $\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} ij P_{ij}$
- 43) *cluster tendency* =  $\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^2 P_{ij}$
- 44) *cluster shade* =  $\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^3 P_{ij}$
- 45) *cluster prominence* =  $\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i + j - \mu_x - \mu_y)^4 P_{ij}$

#### *The Neighborhood Gray Tone Difference Matrix (NGTDM)*

The Neighborhood Gray Tone Difference Matrix [2] was calculated based on 26 adjacent voxels. The voxels with 'not a number' value were excluded from the average over the neighborhood region. Let  $s_i$  denotes the  $i^{th}$  entry in the Neighborhood Gray Tone Difference Matrix,  $N_i$  - the number of voxels having gray tone  $i$ ,  $G$  - number of gray tones in a studied structure,  $n$  - number of studied voxels.

$$46) \text{ coarseness} = \left[ \epsilon + \sum_{i=1}^G \frac{N_i}{n} s_i \right]^{-1}$$

where  $\epsilon$  is a small number to prevent coarseness becoming infinite.

$$47) \text{ contrast} = \left[ \frac{1}{N_g(N_g-1)} \sum_{i=1}^G \sum_{j=1}^G \frac{N_i N_j}{n} (i - j)^2 \right] \left[ \frac{1}{n} \sum_{i=1}^G s_i \right]$$

where  $N_g$  is the total number of different gray levels present in the image.

$$48) \text{ busyness} = \frac{\sum_{i=1}^G \frac{N_i}{n} s_i}{\sum_{i=1}^G \sum_{j=1}^G \frac{N_i}{n} \frac{N_j}{n}}$$

for  $\frac{N_i}{n} \neq 0$  and  $\frac{N_j}{n} \neq 0$

$$49) \text{ complexity} = \sum_{i=1}^G \sum_{j=1}^G \frac{|i-j| \left( \frac{N_i}{n} s_i + \frac{N_j}{n} s_j \right)}{N_i + N_j}$$

for  $N_i \neq 0$  and  $N_j \neq 0$

#### *The Gray Level Size Zone Matrix (GLSZM)*

In the Gray Level Size Zone Matrix [3] calculation the voxels with 'not a number' value were excluded from the analysis. Let  $P_{ij}$  denotes the  $(i, j)$  entry in the Gray Level Size Zone Matrix,  $i$  -

gray value,  $j$  - size,  $n_r$  - number of homogeneous areas inside a studied structure and

$$p_{ij} = P_{ij}/n_r, \mu_i = \sum_{j=1}^N \sum_{i=1}^M i P_{ij}, \mu_j = \sum_{i=1}^M \sum_{j=1}^N i P_{ij}.$$

$$50) \text{ gray level non - uniformity (GLNU)} = \frac{1}{n_r} \sum_{i=1}^M (\sum_{j=1}^N P_{ij})^2$$

$$51) \text{ size zone non - uniformity (SZNU)} = \frac{1}{n_r} \sum_{j=1}^N (\sum_{i=1}^M P_{ij})^2$$

$$52) \text{ small zone emphasis (SZE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P_{ij}}{j^2}$$

$$53) \text{ large zone emphasis (LZE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N P_{ij} \cdot j^2$$

$$54) \text{ low gray level zone emphasis (LGZE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P_{ij}}{i^2}$$

$$55) \text{ high gray level zone emphasis (HGZE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N P_{ij} \cdot i^2$$

$$56) \text{ small zone low gray level emphasis (SZLGE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P_{ij}}{i^2 \cdot j^2}$$

$$57) \text{ small zone high gray level emphasis (SZHGE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P_{ij} \cdot i^2}{j^2}$$

$$58) \text{ large zone low gray level emphasis (LZLGE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N \frac{P_{ij} \cdot j^2}{i^2}$$

$$59) \text{ large zone high gray level emphasis (LZHGE)} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N P_{ij} \cdot i^2 \cdot j^2$$

$$60) \text{ size precentage} = \frac{n_r}{\text{number of voxels in the studied structure}}$$

$$61) \text{ gray level variance} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N (i - \mu_i)^2 P_{ij}$$

$$62) \text{ size zone variance} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N (j - \mu_j)^2 P_{ij}$$

$$63) \text{ size zone entropy} = \frac{1}{n_r} \sum_{i=1}^M \sum_{j=1}^N P_{ij} \log(P_{ij})$$

### Shape

In the USZ implementation: to calculate shape features contours were transformed onto 1 mm isotropic grid. The volume and surface estimation was done using marching cubes algorithm implemented in the VTK library [4].

64)  $V$  – volume

65)  $A$  – surface

$$66) \text{ compactness 1} = \frac{V}{(\pi^2 A)^{\frac{2}{3}}}$$

$$67) \text{ compactness 2} = 36\pi \frac{V^2}{A^3}$$

$$68) \text{ spherical disproportion} = \frac{A}{4\pi R^2}$$

where:  $R$  is the radius of a sphere with the same volume as the tumor.

$$69) \text{ sphericity} = \frac{(36\pi V^2)^{\frac{1}{3}}}{A}$$

$$70) \text{ asphericity} = \left( \frac{1}{36\pi} \frac{A^3}{V^2} \right)^{\frac{1}{3}} - 1$$

$$71) \text{ surface to volume ratio} = \frac{A}{V}$$

- 72) *median thickness* - median of distances of each voxel in the region of interest to its surface, calculated using distance transform
- 73) *SD thickness* - standard deviation of distances of each voxel in the region of interest to its surface
- 74) *maximum 3D diameter* - the largest pairwise Euclidian distance between voxels of the region of interest
- 75) *center of mass shift* = geometric center of the mass – gray levels weighted center of the mass
- 76) *fractal dimension* - calculated using box contouring technique and fixed grid scans excluding the voxels with 'not a number' value (Figure 1S) [5].

$$fractal\ dimension = -\frac{\ln(N(r))}{\ln(r)} - I$$

where:  $r$  - size of the contouring box,  $N(r)$  - number of boxes of size  $r$  containing at least one voxel, which belongs the studied structure,  $I$  - intercept.

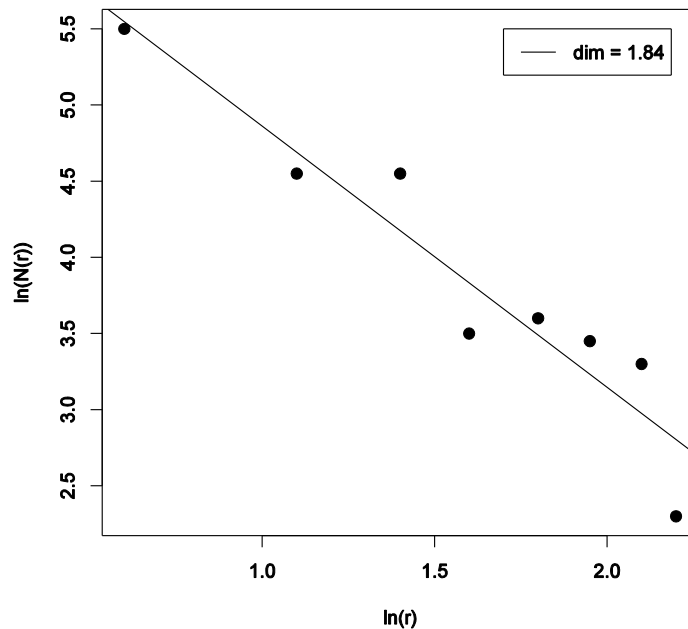


Figure 1S. Example of fractal dimension estimation

- 77) *major axis* – the largest eigenvalue in the principal component analysis
- 78) *minor axis* – the second largest eigenvalue in the principal component analysis
- 79) *least axis* – the smallest eigenvalue in the principal component analysis
- 80) *alongation* =  $\frac{minor\ axis}{major\ axis}$
- 81) *flatness* =  $\frac{least\ axis}{major\ axis}$

### Wavelet

The wavelet analysis was preformed using the Python library PyWavelets version 0.3.0 and the 'Coif1' wavelet as the filter. To account for the boundary effects the entire image set was filtered after the resizing to cubic voxels but before any image segmentation/resegmentation.

## Feature selection

### *a) The principal component analysis combined with univariable Cox regression*

The Horn method [6] was used to define the number of retained principal components. Next each of the radiomic features was assigned to the principal component based on its largest contribution to certain component [7]. Per principal component related group, the features with the highest concordance index (CI) and corresponding false discovery rate  $< 0.2$  in the univariable Cox regression was selected.

### *b) The Pearson correlation between the features and the principal components*

The Horn method [6] was used to define the number of retained principal components. Next, for each principal component one feature was selected to represent it. To that end we determined the feature that correlated the most (the largest Pearson correlation coefficient) with the principal component. In the contrary to the method a) no information about prognostic power of radiomic features was used.

### *c) The average clustering*

The average clustering [8] was performed based on the Pearson correlations between radiomic features. The distance cut-off was set to 0.5. For each of the defined groups, the feature with the highest variance was selected to represent it.

### *d) The mutual information*

The mutual information [9] between the radiomic features and tumor control probability was computed. The feature's selection threshold was defined as 80% of maximum mutual information between tumor control probability and any of the radiomic features.

### *e) The minimum redundancy maximum relevance*

The ensemble maximum relevance minimum redundancy (MRMR) method was used [10]. It was partially combined with the principal component analysis. The feature count was defined as the number of principal components, which explains 95% of data variance. The MRMR was repeated 1000 randomly selecting samples using bootstrap procedure. The redundancy between the features was defined as the Spearman correlation. Features, which achieved at least 80% selection rate were included in the final set.

## Feature classification

### a) *The multivariable Cox regression with backward selection of the variables*

The preselected features were used in the multivariable Cox regression with backward selection of the variables based on the Akaike information criterion [11]. Features, which had at least 80% selection rate in 5-fold cross-validation were chosen to the final model.

### b) *The least absolute shrinkage and selection operator*

The least absolute shrinkage and selection operator [12] (100 times 5-fold cross-validated) was used for variable selection in multivariable Cox regression. Features, which had at least 70% selection rate were chosen to the final model.

### c) *The random forest*

The random forest model was trained using 1000 trees and Breiman-Cutler permutation was calculate variable importance [13]. Features were selected based on the minimization of the out-of-bag error. The additional constrain was set for the maximal number of features  $n=5$ .

## Results

a)

		Cox backward	LASSO	RF
PCA univariable Cox	CI	<b>0.72 (0.64-0.90)</b>	0.69 (0.59-0.85)	0.65 (0.54-0.88)
	n	2	1	3
PCA correlation	CI	0.68 (0.53-0.83)	0.69 (0.56-0.85)	0.55 (0.44-0.77)
	n	5	1	3
average clustering	CI	-	0.67 (0.47-0.90)	0.69 (0.64-0.80)
	n	-	1	5
mutual information	CI	0.71* (0.60-0.90)	0.71* (0.60-0.90)	0.68** (0.63-0.71)
	n	1	1	2
minimum redundancy maximum relevance	CI	0.71* (0.60-0.90)	0.71* (0.60-0.90)	0.68** (0.63-0.71)
	n	1	1	2

b)

		Cox backward	LASSO	RF
PCA univariable Cox	CI	<b>0.74** (0.62-0.80)</b>	0.70 (0.56-0.80)	0.70 (0.57-0.81)
	n	2	2	2
PCA correlation	CI	0.66 (0.54-0.74)	0.62 (0.55-0.71)	0.64 (0.60-0.74)
	n	3	1	3
average clustering	CI	0.67 (0.64-0.76)	0.67 (0.62-0.76)	0.54 (0.39-0.68)
	n	2	2	5
mutual information	CI	0.74** (0.62-0.80)	0.70 (0.66-0.74)	0.68 (0.57-0.80)
	n	2	1	3
minimum redundancy maximum relevance	CI	(0.57-0.81)	0.66 (0.62-0.79)	0.60 (0.50-0.66)
	n	1	1	3

Table 1S. The concordance indexes (CI) for radiomics-based local tumor control models trained with a different combinations of features selection and classifications a) CT radiomics, b) PET radiomics. The results were based on the 5-fold cross-validation in the training cohort. The maximum and minimum CI obtained in the cross-validation are presented in the brackets. Additionally, the number of features (n) included in the multivariable model is shown. The combination of the principal component analysis combined with univariable Cox regression and the multivariable Cox regression with backward selection of the variables resulted in the best models considering both CT and PET features (in bold). The combination of average clustering and Cox multivariable regression failed to optimize a model due to too large number of input variables (n=66). \* the same final model containing 1 feature, \*\* the same final model containing 2 features. PCA univariable Cox - the principal component analysis combined with univariable Cox regression, PCA correlation - the Pearson correlation between the features and the principal components, Cox backward - the multivariable Cox regression with backward selection of the variables, LASSO - the least absolute shrinkage and selection operator, RF - the random forest.

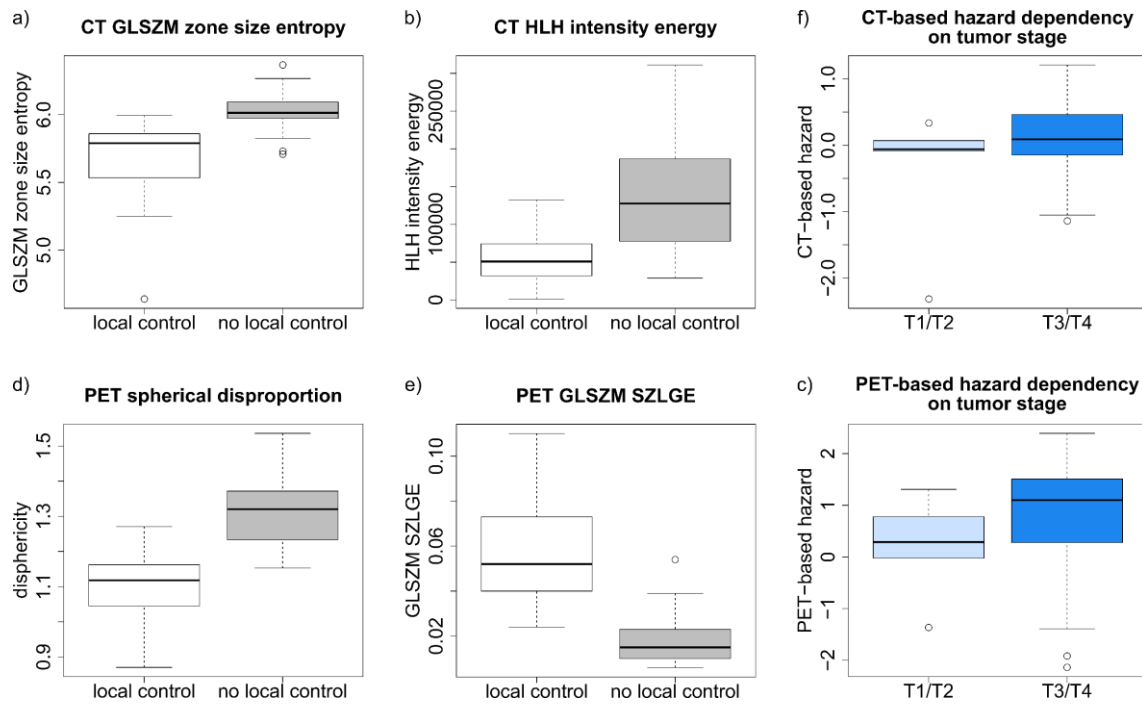


Figure 2S. The CT (a and b) and PET (d and e) radiomics signature prognostic for local tumor recurrence in head and neck cancer. All radiomic features differed significantly (Wilcoxon test  $p$ -value  $< 0.05$ ) between the controlled tumors and recurrences. The predictions based on the radiomics model was independent from the tumor stage (c and f).

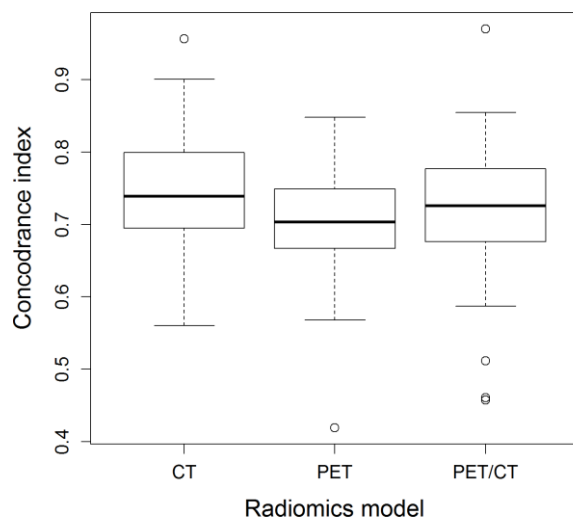


Figure 3. The concordance index distributions in the validation dataset for different local tumor control radiomics models. The distributions were obtained using the bootstrap method. None of the models showed a superior discriminative power.

## References

- [1] Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern.* 1973;3:610-621.
- [2] Amadasun M, King R. Textural features corresponding to textural properties. *IEEE Trans Syst Man Cybern.* 1989;19(5).
- [3] Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, Sequeira J, Mari JL. Texture indexes and Gray Level Size Zone matrix application to cell nuclei classification. *Pattern Recognition Inf Process.* 2009;140-145.
- [4] Schroeder W, Martin K, Lorensen B. The visualization toolkit (4th ed.). Kitware. 2006.
- [5] Napolitano A, Ungania S, Cannata V, Fractal dimension estimation methods for biomedical images in MATLAB - A fundamental tool for scientific computing and engineering applications. *InTech.* 2012 volume 3.
- [6] Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika* 1965;30:179-185
- [7] Bogowicz M, Riesterer O, Ikenberg K, et al. CT radiomics predicts HPV status and local tumor control after definitive radiochemotherapy in head and neck squamous cell carcinoma *Int J Rad Onc Biol Phys.* 2017.
- [8] Jolliffe IT. Discarding variables in a principal component analysis: artificial data. *Journal of the Royal Statistical Society. Series C (Applied Statistics).* 1972;21(2):160-173
- [9] Cover TM, Thomas JA. *Elements of Information Theory.* Wiley. 1991
- [10] De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B. mRMRe: an R package for parallelized mRMR ensemble feature selection. *Bioinformatics.* 2013;29(18):2365-2368
- [11] Sakamoto Y, Ishiguro M, Kitagawa G. Akaike information criterion statistics. KTK Scientific Publishers. 1986
- [12] Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 2011;73:273-282.
- [13] Breiman L. Random forests. *Machine Learning* 2001;45(1):5-32