10.1148/rycan.2020190039

Appendix E1

Development and Validation of Radiogenomics Signatures for Head and Neck Squamous Cell Carcinoma Molecular Subtypes

Radiomic Features Description

We developed an automatic platform extracting a total of 2131 radiomic features to quantify volumetric tumor characteristics from pretreatment CT scans. These radiomic features were primarily categorized into (1) intensity-based features, (2) shape and size features, (3) textural features, and (4) filter-based features.

Group 1. Intensity-based Features

We extracted 23 intensity-based features which are summary statistics of the voxel intensities within the segmented tumors. Let **X** denote the three-dimensional image array with N voxels and **P** the first order histogram with N_l discrete intensity levels.

1. 1 General features (Feature tag: intensity_based_General)

(1) Energy:

$$\text{energy} = \sum_{i}^{N} \mathbf{X}(i)^2$$

(2) Entropy:

$$\text{entropy} = \sum_{i=1}^{N_l} \mathbf{P}(i) \log_2 \mathbf{P}(i)$$

(3) Kurtosis:

$$\text{kurtosis} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{X}(i) - \bar{X}\right)^4}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{X}(i) - \bar{X}\right)^2}\right)^2}$$

(4) Sum:

The sum of intensity values of **X**, ie, sum (**X**).

(5) Maximum:

The maximum intensity value in **X**, ie, max (**X**).

(6) Mean:

$$\text{mean} = \frac{1}{N}\sum_{i}^{N} \mathbf{X}(i)$$

(7) Mean absolute deviation:

The mean of the absolute deviations of all voxel intensities around the mean intensity value.

$$\text{mean absolute deviation} = \frac{1}{N}\sum_{i}^{N}\left|\mathbf{X} - \overline{X}\right|$$

(8) Median:

The median intensity value of **X**, ie, median (**X**).

(9) Minimum:

The minimum intensity value of **X**, ie, min (**X**).

(10) Range:

The range of intensity values of **X**.

(11) Root mean square:

$$\text{root mean square} = \sqrt{\frac{\sum_{i}^{N}\mathbf{X}(i)^{2}}{N}}$$

(12) Skewness:

$$\text{skewness} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{X}(i) - \overline{X}\right)^{3}}{\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{X}(i) - \overline{X}\right)^{2}}\right)^{3}}$$

(13) Standard deviation (std):

$$\text{standard deviation} = \left(\frac{1}{N-1}\sum_{i}^{N}\left(\mathbf{X}(i) - \overline{X}\right)^{2}\right)^{1/2}$$

(14) Uniformity:

$$\text{uniformity} = \sum_{i=1}^{N_l}\mathbf{P}(i)^{2}$$

(15) Variance:

$$\text{variance} = \frac{1}{N-1}\sum_{i=1}^{N}\left(\mathbf{X}(i) - \overline{X}\right)^{2}$$

(16) Median absolute deviation:

The median of the absolute deviations of all voxel intensities around the median intensity value, ie, $\text{median}\left(\left|\mathbf{X} - \text{median}(\mathbf{X})\right|\right)$.

(17) The 10th percentile (p10):

The 10th percentile of the intensity values.

(18) The 90th percentile (p90):

The 90th percentile of the intensity values.

(19) Robust mean absolute deviation:

The mean absolute deviation of intensity values not greater than the 90th percentile and not less than the 10th percentile.

(20) Robust median absolute deviation:

The median absolute deviation of intensity values not greater than the 90th percentile and not less than the 10th percentile.

(21) Interquartile range:

The interquartile range of the intensity values, ie, 75th percentile–25th percentile.

(22) Quartile coefficient of dispersion (coeffDispersion):

A measure of dispersion used to make comparisons within and between data. The quartile coefficient of dispersion is $\frac{\text{75th percentile} - \text{25th percentile}}{\text{75th percentile} + \text{25th percentile}}$.

(23) Coefficient of variation (coeffVariation):

A measure of relative variability, ie, the ratio of the standard deviation to the mean.

1.2 Statistical features based on histogram (Feature tag: intensity_based_Histogram)

We created a histogram of the intensity values with 50 bins, each bin representing the percentage of all the sample points in the bin. Each percentage was considered as a feature resulting in 50 features. In addition, we calculated a set ($n = 21$) of features measuring statistical characteristics of the histogram, these are: variance, skewness, kurtosis, entropy, uniformity, mean, median, range, inter range, median deviation, harmonic mean, geometric mean, central moment, four percentile values (0.025, 0.25, 0.75 and 0.975), and four quantile values (0.25, 0.5, 0.75 and 0.95) of the histogram. In total, 71 features were computed based on histogram.

Group 2. Shape and Size Features (Feature tag: shape_size)

Shape and size features describe the three-dimensional shape and size of the gross tumor. Let V denote the volume and A the surface area of the volume.

(1) Compactness 1:

$$\text{compactness}1 = \frac{V}{\sqrt{\pi}A^{\frac{2}{3}}}$$

(2) Compactness 2:

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

(3) Spherical disproportion:

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

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

(4) Sphericity:

$$\text{sphericity} = \frac{\pi^{\frac{1}{3}}(6V)^{\frac{2}{3}}}{A}$$

(5) Volume in milliliter (volumeML):

The volume (V) of the tumor is calculated by multiplying the number of pixels with the voxel size. The unit is milliliter.

(6) Surface area:

The surface area of a three-dimensional alpha shape of the tumor's voxels.

(7) Surface volume ratio:

Ratio of surface area to volume.

(8) Eccentricity:

$$\text{eccentricity} = \sqrt{1 - \frac{a * b}{c^2}}$$

where c is the longest semiprincipal axes of the ellipsoid, and a and b are the second and third longest semiprincipal axes of the ellipsoid.

(9) Solidity:

The ratio of the number of voxels in the region of interest (ROI) to the number of voxels in the three-dimensional convex hull of the ROI.

Group 3. Textural Features (Feature tag: textural)

We measured texture features from the tumor ROI by incorporating five types of texture features.

3.1 Histogram of Gradient (HOG) features (Feature tag: textural_HOG)

HOG features encode local shape information from the tumor region by counting occurrences of gradient orientation within the ROI. This method is computed on a dense grid of uniformly spaced cells and uses overlapping local contrast normalization for improved accuracy. We choose to define cell size with an 8-by-8 cell and block size with a 4-by-4 square block that allow us to extract totally 576 HOG features (63).

3.2 Local Binary Patterns (LBP) features (Feature tag: textural_LBP)

One subgroup was Local binary patterns (LBP) features, quantifying local intensity texture variations based on local binary patterns encoding intensity neighborhood relationships. We used the default setting on pixel neighborhood structure where 8 samplings lie on a circle of radius 1 (64). In total, we extracted 256 LBP features.

The other textural features were Gray-Level Co-occurrence Matrix (GLCM, n = 23), Gray-Level Run-Length Matrix (GLRLM, n = 11) features, Gray-Level Size Zone Matrix (GLSZM, n = 13) features, and the Neighborhood Gray-Tone Difference Matrix (NGTDM, n = 5) features (17,65).

Group 4. Filter-based Features (Feature tag: filter_based)

4.1 Gabor features (Feature tag: filter_based_Gabor).

We first generated Gabor-filtered images by applying the Gabor-filter with F = 8 different frequencies and D = 8 directions on tumor ROI. We then computed high-order texture features (n = 688) on the generated Gabor filtered images. The two-dimensional Gabor filter was defined as:

$$G(\mathrm{x},\mathrm{y}) = \frac{f^2}{\pi\gamma\eta}\exp\left(-\frac{x'^2+\gamma^2 y'^2}{2\sigma^2}\right)\exp\left(\mathrm{j}2\pi f x' + \phi\right)$$

$$x' = x\cos\theta + y\sin\theta$$

$$y' = -x\sin\theta + y\cos\theta$$

In this study, we used eight frequencies (F) $f = \frac{0.25}{(\sqrt{2})^{i-1}}, \quad i = 1,2,3..,8$, and eight directions (D) $\theta = \frac{j-1}{8}\pi, \quad j = 1,2,....,8$.

4.2 Wavelet Features (Feature tag: filter_based_Wavelet):

Wavelet transform is an important multiresolution analysis tool for texture analysis by decomposing the original image into low-and high-frequencies. A total of 456 wavelet features were computed as Vallières et al (65).

References

63. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, June 20–25 2005. Piscataway, NJ: IEEE, 2005; 886–893.

64. Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans Pattern Anal Mach Intell 2002;24(7):971–987.

65. Vallières M, Kay-Rivest E, Perrin LJ, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep 2017;7(1):10117.