Radiomic Features Description

We developed an automatic platform extracting radiomic features to quantify volumetric tumor characteristics from pre-treatment CT scans. The pipeline is available at https://github.com/gevaertlab/radiomics_pipeline. These radiomic features were primarily categorized into (1) first-order features, (2) shape and size features, (3) Global histogram features, (4) textural features, and (5) filter-based features. Briefly, first-order features are calculated directly from the Hounsfield intensity units, including features such as minimum, maximum, mean or variance. Next, shape and size features quantify the 3D shape and size of the tumors, e.g. volume and surface area. Global histogram features are a group of statistical features based on histograms of the intensity values. Texture features are calculated based on the spatial relationships between voxels and are further subdivided into Gray-Level Co-Occurrence Matrix (GLCM), gray level run length matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM) and Neighboring Gray Tone Difference Matrix (NGTDM) features. Finally, the filter-based features were obtained by applying wavelet transformations to CT images to extract filter-transformed first-order features and shape and size features. The filter-based features are able to capture higher-level abstraction of the imaging objects. All feature extraction was implemented with an in-house developed pipeline using Matlab 2018a (MathWorks, Natick, MA, USA).

Feature description

Group 1. First-order features

We extracted 24 first-order 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) 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:

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

(4) Voxel num:

The total number of voxels.

(5) Sum:

The sum of intensity values of **X**, i.e., sum(**X**).

(6) Maximum:

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

(7) Mean:

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

(8) Mean absolute deviation:

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

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

(9) Median:

The median intensity value of **X**, i.e., median(**X**).

(10) Minimum:

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

(11) Range:

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

(12) Root mean square:

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

(13) Skewness:

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

(14) Standard deviation:

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

(15) Uniformity:

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

(16) Variance:

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

(17) Median absolute deviation:

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

(18) The 10^{th} percentile (p10):

The 10^{th} percentile of the intensity values.

(19) The 90^{th} percentile (p90):

The 90^{th} percentile of the intensity values.

(20) Robust mean absolute deviation:

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

(21) Robust median absolute deviation:

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

(22) Interquartile range:

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

(23) Quartile coefficient of dispersion (coeffDispersion):

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

(24) Coefficient of variation (coeffVariation):

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

Group 2. Shape and size features (Feature tag: shapeSize)

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) Maximum 3D diameter:

The largest pairwise Euclidean distance between voxels of the gross tumour surface.

(2) Compactness 1:

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

(3) Compactness 2:

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

(4) Spherical disproportion:

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

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

(5) Sphericity:

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

(6) 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.

(7) Surface area:

Surface area is approximated with triangulation, that is, dividing the surface into connected triangles and summing up the areas.

$$A = \sum_{i=1}^{N} \frac{1}{2} \left| \overrightarrow{AB_i} \times \overrightarrow{AC_i} \right|$$

A, B, C are vertices of each triangle.

(8) Surface volume ratio:

Ratio of surface area to volume.

(9) Eccentricity:

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

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

(10) Solidity:

The ratio of the number of voxels in the ROI to the number of voxels in the 3D convex hull of the ROI.

Group 3. Global histogram features (Feature tag: Global_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 4. Textural features (Feature tag: textural)

Textural features measured 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.

Group 5. Filter-based features (Feature tag: filter_based)

Wavelet filter-based features were extracted. Wavelet transform is an important multi-resolution analysis tool for texture analysis by decomposing the original image into low- and high-frequencies. A total of 464 wavelet features were computed as Aerts et al.[1]

1. Aerts, H.J.W.L. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* (2014). doi:10.1038/ncomms5006