# Supplementary Methods

# Radiomics Features

We evaluated a total number of 440 CT imaging features, which are divided in four groups as follows:

*Group 1.* First order statistics
*Group 2.* Shape and size based features
*Group 3.* Textural features
*Group 4.* Wavelet features

### *Group 1. First order statistics*

First-order statistics describe the distribution of voxel intensities within the CT image through commonly used and basic metrics. Let $\mathbf{X}$ denote the three dimensional image matrix with $N$ voxels and $\mathbf{P}$ the first order histogram with $N_l$ discrete intensity levels. The following first order statistics were extracted:

1. **Energy:**

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

2. **Entropy:**

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

3. **Kurtosis:**

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

   where $\bar{X}$ is the mean of $\mathbf{X}$.

4. **Maximum:**

   The maximum intensity value of $\mathbf{X}$.

5. **Mean:**

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

6. **Mean absolute deviation:**

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

7. **Median:**

   The median intensity value of $\mathbf{X}$.

8. **Minimum:**

   The minimum intensity value of $\mathbf{X}$.

9. **Range:**

   The range of intensity values of $\mathbf{X}$.

10. **Root mean square (RMS):**

$$RMS = \sqrt{\frac{\sum_{i}^{N}\mathbf{X}(i)^2}{N}}$$

11. **Skewness:**

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

   where $\bar{X}$ is the mean of $\mathbf{X}$.

12. **Standard deviation:**

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

where $\bar{X}$ is the mean of $\mathbf{X}$.

13. **Uniformity:**

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

14. **Variance:**

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

where $\bar{X}$ is the mean of $\mathbf{X}$.

The standard deviation, variance and mean absolute deviation are measures of the histogram dispersion, that is, a measure of how much the gray levels differ from the mean. The variance, skewness and kurtosis are the most frequently used central moments. The skewness measures the degree of histogram asymmetry around the mean, and kurtosis is a measure of the histogram sharpness. As measures of histogram randomness we computed the uniformity and entropy of the image histogram.

***Group 2. Shape and size based features***

In this group of features we included descriptors of the three-dimensional size and shape of the tumor region. Let in the following definitions $V$ denote the volume and $A$ the surface area of the volume of interest. We determined the following shape and size based features:

15. **Compactness 1**:

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

16. **Compactness 2**:

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

17. **Maximum 3D diameter**:

    The maximum three-dimensional tumor diameter is measured as the largest pairwise Euclidean distance, between voxels on the surface of the tumor volume.

18. **Spherical disproportion**:

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

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

19. **Sphericity**:

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

20. **Surface area:**

    The surface area is calculated by triangulation (i.e. dividing the surface into connected triangles) and is defined as:

$$A = \sum_{i=1}^{N} \frac{1}{2} |\mathbf{a}_i\mathbf{b}_i \times \mathbf{a}_i\mathbf{c}_i|$$

    Where $N$ is the total number of triangles covering the surface and $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ are edge vectors of the triangles.

21. **Surface to volume ratio**:

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

22. **Volume:**

    The volume ($V$) of the tumor is determined by counting the number of pixels in the tumor region and multiplying this value by the voxel size.

The maximum 3D diameter, surface area and volume provide information on the size of the lesion. Measures of compactness, spherical disproportion, sphericity and the surface to volume ratio describe how spherical, rounded, or elongated the shape of the tumor is.

### *Group 3. Textural features*

The features shown above that resulted from group 1 (first-order statistics) provide information related to the gray-level distribution of the image; however they do not provide any information regarding the relative position of the various gray levels over the image. In this group we therefore included textural features describing patterns or the spatial distribution of voxel intensities, which were calculated from respectively gray level co-occurrence (GLCM)[1] and gray level run-length (GLRLM)[2] texture matrices. Determining texture matrix representations requires the voxel intensity values within the VOI to be discretized. Voxel intensities were therefore resampled into equally spaced bins using a bin-width of 25 Hounsfield Units. This discretization step not only reduces image noise, but also normalizes intensities across all patients, allowing for a direct comparison of all calculated textural features between patients. Texture matrices were determined considering 26-connected voxels (i.e. voxels were considered to be neighbors in all 13 directions in three dimensions).

*Gray-Level Co-Occurrence Matrix based features*

A GLCM is defined as $\mathbf{P}(i,j;\delta,\alpha)$, a matrix with size $N_g \times N_g$ describing the second-order joint probability function of an image, where the $(i,j)$th element represents the number of times the combination of intensity levels $i$ and $j$ occur in two pixels in the image, that are separated by a distance of $\delta$ pixels in direction $\alpha$, and $N_g$ is the number of discrete gray level intensities. As a two dimensional example, let the following matrix represent a 5x5 image, having 5 discrete gray levels:

$$I = \begin{matrix} 1 & 2 & 5 & 2 & 3 \\ 3 & 2 & 1 & 3 & 1 \\ 1 & 3 & 5 & 5 & 2 \\ 1 & 1 & 1 & 1 & 2 \\ 1 & 2 & 4 & 3 & 5 \end{matrix}$$

For distance $\delta = 1$ (considering pixels with a distance of 1 pixel from each other) in direction $\alpha = 0$, where 0 degrees is the horizontal direction, the following GLCM is obtained:

$$\mathbf{P}(1{,}0) = \begin{matrix} 3 & 3 & 2 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 1 \end{matrix}$$

In this study, distance $\delta$ was set to 1 and direction $\alpha$ to each of the 13 directions in three dimensions, yielding a total of 13 gray level co-occurrence matrices for each 3D image. From these gray-level co-occurrence matrices, several textural features are derived. Each 3D gray level co-occurrence based feature was then calculated as the mean of the feature calculations for each of the 13 directions.

Let:

$\mathbf{P}(i,j)$ be the co-occurrence matrix for an arbitrary $\delta$ and $\alpha$,

$N_g$ be the number of discrete intensity levels in the image,

$\mu$ be the mean of $\mathbf{P}(i,j)$,

$p_x(i) = \sum_{j=1}^{N_g} \mathbf{P}(i,j)$ be the marginal row probabilities,

$p_y(i) = \sum_{i=1}^{N_g} \mathbf{P}(i,j)$ be the marginal column probabilities,

$\mu_x$ be the mean of $p_x$,

$\mu_y$ be the mean of $p_y$,

$\sigma_x$ be the standard deviation of $p_x$,

$\sigma_y$ be the standard deviation of $p_y$,

$p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j),\ i + j = k,\ k = 2{,}3, \dots, 2N_g$,

$p_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j),\ |i - j| = k,\ k = 0{,}1, \dots, N_g - 1$,

$HX = -\sum_{i=1}^{N_g} p_x(i) \log_2[p_x(i)]$ be the entropy of $p_x$,

$HY = -\sum_{i=1}^{N_g} p_y(i) \log_2[p_y(i)]$ be the entropy of $p_y$,

$H = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j) \log_2[P(i,j)]$ be the entropy of $P(i,j)$,

$HXY1 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \mathbf{P}(i,j)\log(p_x(i)p_y(j))$,

$HXY2 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_x(i)p_y(j)\log(p_x(i)p_y(j))$.

23. **Autocorrelation:**

$$autocorrelation = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ij\mathbf{P}(i,j)$$

24. **Cluster Prominence:**

$$cluster\ prominence = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[i+j-\mu_x(i)-\mu_y(j)\right]^4\mathbf{P}(i,j)$$

25. **Cluster Shade:**

$$cluster\ shade = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[i+j-\mu_x(i)-\mu_y(j)\right]^3\mathbf{P}(i,j)$$

26. **Cluster Tendency:**

$$cluster\ tendency = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[i+j-\mu_x(i)-\mu_y(j)\right]^2\mathbf{P}(i,j)$$

27. **Contrast:**

$$contrast = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}|i-j|^2\mathbf{P}(i,j)$$

28. **Correlation:**

$$correlation = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ijP(i,j)-\mu_i(i)\mu_j(j)}{\sigma_x(i)\sigma_y(j)}$$

29. **Difference entropy:**

$$difference\ entropy = \sum_{i=0}^{N_g-1}\boldsymbol{P}_{x-y}(i)\log_2\left[\mathbf{P}_{x-y}(i)\right]$$

30. **Dissimilarity:**

$$dissimilarity = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}|i-j|\mathbf{P}(i,j)$$

**31. Energy:**

$$energy = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}[\mathbf{P}(i,j)]^2$$

**32. Entropy ($H$):**

$$entropy = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\mathbf{P}(i,j)\log_2[\mathbf{P}(i,j)]$$

**33. Homogeneity 1:**

$$homogeneity\ 1 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{\mathbf{P}(i,j)}{1+|i-j|}$$

**34. Homogeneity 2:**

$$homogeneity\ 2 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{\mathbf{P}(i,j)}{1+|i-j|^2}$$

**35. Informational measure of correlation 1 (IMC1):**

$$IMC1 = \frac{HXY - HXY1}{\max\{HX, HY\}}$$

**36. Informational measure of correlation 2 (IMC2):**

$$IMC2 = \sqrt{1 - e^{-2(HXY2 - HXY)}}$$

**37. Inverse Difference Moment Normalized (IDMN):**

$$IDMN = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{\mathbf{P}(i,j)}{1+\left(\frac{|i-j|^2}{N^2}\right)}$$

**38. Inverse Difference Normalized (IDN):**

$$IDN = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{\mathbf{P}(i,j)}{1+\left(\frac{|i-j|}{N}\right)}$$

**39. Inverse variance:**

$$inverse\ variance = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{\mathbf{P}(i,j)}{|i-j|^2} \quad ,i \neq j$$

**40. Maximum Probability:**

$$maximum\ probability = \max\{\mathbf{P}(i,j)\}$$

**41. Sum average:**

$$sum\ average = \sum_{i=2}^{2N_g} [i\mathbf{P}_{x+y}(i)]$$

**42. Sum entropy:**

$$sum\ entropy = -\sum_{i=2}^{2N_g} \mathbf{P}_{x+y}(i) \log_2[\mathbf{P}_{x+y}(i)]$$

**43. Sum variance:**

$$sum\ variance = \sum_{i=2}^{2N_g} (i - SE)^2 \mathbf{P}_{x+y}(i)$$

**44. Variance:**

$$variance = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu)^2 \mathbf{P}(i,j)$$

*Gray-Level Run-Length matrix based features*

Run length metrics quantify gray level runs in an image. A gray level run is defined as the length in number of pixels, of consecutive pixels that have the same gray level value. In a gray level run length matrix $p(i,j|\theta)$, the $(i,j)$th element describes the number of times $j$ a gray level $i$ appears consecutively in the direction specified by $\theta$, and $N_g$ is the number of discrete gray level intensities. As a two dimensional example, consider the following 5x5 image, with 5 discrete gray levels:

$$I = \begin{matrix} 5 & 2 & 5 & 4 & 4 \\ 3 & 3 & 3 & 1 & 3 \\ 2 & 1 & 1 & 1 & 3 \\ 4 & 2 & 2 & 2 & 3 \\ 3 & 5 & 3 & 3 & 2 \end{matrix}$$

The GLRL matrix for $\theta = 0$, where 0 degrees is the horizontal direction, then becomes:

$$p(0) = \begin{array}{ccccc} 1 & 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 \\ 4 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \end{array}$$

In this study, a GLRL matrix was computed for every of the 13 directions in three dimensions, from which the below textural features were derived. Each 3D GLRL feature was then calculated as the mean of the feature values for each of the 13 directions.

Let:

$p(i,j|\theta)$ be the $(i,j)$th entry in the given run-length matrix $p$ for a direction $\theta$,

$N_g$ the number of discrete intensity values in the image,

$N_r$ the number of different run lengths,

$N_p$ the number of voxels in the image.

**45. Short Run Emphasis (SRE)**

$$SRE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\left[\frac{p(i,j|\theta)}{j^2}\right]}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j|\theta)}$$

**46. Long Run Emphasis (LRE)**

$$LRE = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}j^2 p(i,j|\theta)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j|\theta)}$$

**47. Gray Level Non-Uniformity (GLN)**

$$GLN = \frac{\sum_{i=1}^{N_g}\left[\sum_{j=1}^{N_r}p(i,j|\theta)\right]^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j|\theta)}$$

**48. Run Length Non-Uniformity (RLN)**

$$RLN = \frac{\sum_{j=1}^{N_r}\left[\sum_{i=1}^{N_g}p(i,j|\theta)\right]^2}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}p(i,j|\theta)}$$

**49. Run Percentage (RP)**

$$RP = \sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \frac{p(i,j|\theta)}{N_p}$$

### 50. Low Gray Level Run Emphasis (LGLRE)

$$LGLRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[ \frac{p(i,j|\theta)}{i^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta)}$$

### 51. High Gray Level Run Emphasis (HGLRE)

$$HGLRE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} i^2 p(i,j|\theta)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta)}$$

### 52. Short Run Low Gray Level Emphasis (SRLGLE)

$$SRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[ \frac{p(i,j|\theta)}{i^2 j^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta)}$$

### 53. Short Run High Gray Level Emphasis (SRHGLE)

$$SRHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[ \frac{p(i,j|\theta) i^2}{j^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta)}$$

### 54. Long Run Low Gray Level Emphasis (LRLGLE)

$$LRLGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} \left[ \frac{p(i,j|\theta) j^2}{i^2} \right]}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta)}$$

### 55. Long Run High Gray Level Emphasis (LRHGLE)

$$LRHGLE = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta) i^2 j^2}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_r} p(i,j|\theta)}$$

***Group 4. Wavelet features: first order statistics and texture of wavelet decompositions***

Wavelet transform effectively decouples textural information by decomposing the original image, in a similar manner as Fourier analysis, in low –and high-frequencies. In this study a discrete, one-level and undecimated three dimensional wavelet transform was applied to each CT image, which decomposes the original image $X$ into 8 decompositions. Consider $L$ and $H$ to be a low-pass (i.e. a scaling) and, respectively, a high-pass (i.e. a wavelet) function, and the wavelet decompositions of $X$ to be labeled as $X_{LLL}$, $X_{LLH}$, $X_{LHL}$, $X_{LHH}$, $X_{HLL}$, $X_{HLH}$, $X_{HHL}$ and $X_{HHH}$. For example, $X_{LLH}$ is then interpreted as the high-pass sub band, resulting from directional filtering of $X$ with a low-pass filter along x-direction, a low pas filter along y-direction and a high-pass filter along z-direction and is constructed as:

$$X_{LLH}(i,j,k) = \sum_{p=1}^{N_L}\sum_{q=1}^{N_L}\sum_{r=1}^{N_H} L(p)L(q)H(r)X(i+p,j+q,k+r)$$

Where $N_L$ is the length of filter $L$ and $N_H$ is the length of filter $H$. The other decompositions are constructed in a similar manner, applying their respective ordering of low or high-pass filtering in x, y and z-direction. Wavelet decomposition of the image $X$ is schematically depicted in **Supplementary Figure 4**. Since the applied wavelet decomposition is undecimated, the size of each decomposition is equal to the original image and each decomposition is shift invariant. Because of these properties, the original tumor delineation of the gross tumor volume (GTV) can be applied directly to the decompositions after wavelet transform. In this study “Coiflet 1” wavelet was applied on the original CT images. For each decomposition we computed the first order statistics as described in Group 1 and the textural features as described in Group 3.