BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

The details of the selected radiomics features are as follow.

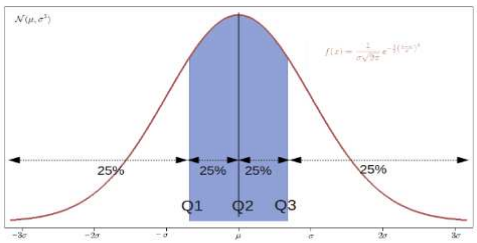
1. Histogram Parameters

Histogram parameters are concerned with properties of individual pixels. They describe the distribution of voxel intensities within the CT image through commonly used and basic metrics. Let X denote the three dimensional image matrix with N voxels and P the first order histogram divided by N_l discrete intensity levels. The following first order statistics were extracted:

1.1 Quantiles

Quantile normalization is a global adjustment method that assumes the statistical distribution of each sample is the same. The normalization is achieved by forcing the observed distributions to be the same and the average distribution, obtained by taking the average of each quantile across samples. They are cut points dividing the range of a probability distribution into contiguous intervals with equal probabilities, or dividing the observations in a sample in the same way.
For a finite population of *N* equally probable values indexed 1, …, *N* from lowest to highest, the *k*-th *q*-quantile of this population can equivalently be computed via the value of:

$$\mathbf{I_p = N\ k/q}$$



The area below the red curve is the same in the intervals (-∞, *Q1*), (*Q1*, *Q2*), (*Q2*, *Q3*), and (*Q3*, +∞).

Selected feature in predicting PD-1 expression model in arterial-phase images: **Quantile0.975.**

1.2 RelativeDeviation

Let $\bar{X}$ denote the mean of a set of quantities X_i , then the relative deviation is defined by:

$$\frac{\Delta X_i}{\bar{X}} = \frac{|X_i - \bar{X}|}{\bar{X}}$$

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

1.3 Cluster Shade

Cluster analysis or **clustering** is the task of grouping a set of objects in such a way that objects in the same group (**cluster**) are more similar (in some sense or another) to each other than to those in other groups (clusters). It is a common technique for statistical data analysis.

Cluster Shade in clustered shading, we group similar view samples according to their position and, optionally, normal into clusters.
Selected feature in predicting PD-1 expression model in arterial-phase images:
ClusterShade_AllDirection_offset4_SD

Formula

$$\sum_{i,j}((i-\mu)+(j-\mu))^3 g(i,j)$$

2. Form Factor Parameters

These group of features includes 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:
Selected feature in predicting PD-1 expression model in arterial-phase images:
Compactness2
Selected feature in predicting PD-L1 expression model in arterial-phase images:
Compactness2

2.1 Compactness 2:

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

3. GLCM Parameters

The Grey level co-occurrence matrix (GLCM) $\mathbf{P}\ (\mathbf{i},\mathbf{j}|\boldsymbol{\theta},\mathbf{d})$ represents the joint probability of certain sets of pixels having certain grey-level values. It calculates how many times a pixel with grey-level **i** occurs jointly with another pixel having a grey value **j**. By varying the displacement vector **d** between each pair of pixels.

The advantage of the co-occurrence matrix calculations is that the co-occurring pairs of pixels can be spatially related in various orientations with reference to distance and angular spatial relationships, as on considering the relationship between two pixels at a time. As a result, the combination of grey levels and their positions are exhibited apparently. Therefore, it is defined as “A two dimensional histogram of gray levels

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

for pair of pixels, which are separated by a fixed spatial relationship". However, the matrix is sensitive to rotation. With the change of different offsets define pixel relationships by varying directions.

The rotation angle of an offset: $0°, 45°, 90°, 135°$ and displacement vectors (distance to the neighbor pixel: 1, 2, 3 ...), different co-occurrence distributions from the same image of reference. GLCM of an image is computed using displacement vector d defined by its radius, (distance or count to the next adjacent neighbor preferably is equal to one) and rotational angles.

Selected feature in predicting PD-1 expression model in arterial-phase images:
RelativeDeviation

3.1 Energy of GLCM

This feature Returns the sum of squared elements in the GLCM. Range = [0 1]
Energy is 1 for a constant image. Is high when image has very good homogeneity or when pixels are very similar The Property Energy is also known as uniformity,

uniformity of energy, and angular second moment.

In AK Software we have 18 parameters related to the GLCM Energy :

(Formula)

$$\sum_{i,j} g(i,j)^2$$

*g is a GLCM
Where i,j are the spatial coordinates of g (i,j).
Selected features in predicting PD-1 expression model in arterial-phase images:
GLCMEnergy_AllDirection_offset1_SD
GLCMEnergy_AllDirection_offset7_SD
Selected features in predicting PD-L1 expression model in arterial-phase images:
GLCMEnergy_AllDirection_offset1_SD
GLCMEnergy_angle90_offset7
Selected features in predicting OS model in arterial-phase images:
GLCMEnergy_AllDirection_offset1_SD

3.2 Entropy of GLCM

Entropy is a measure of randomness of intensity image.
Entropy shows the amount of information of the image that is needed for the image compression. Entropy measures the loss of information or message in a transmitted signal and also measures the image information.
Selected features in predicting PD-1 expression model in arterial-phase images:

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

GLCMEntropy_AllDirection_offset1_SD
GLCMEntropy_AllDirection_offset7_SD
GLCMEntropy_angle45_offset7

Selected features in predicting PD-L1 expression model in arterial-phase images:
GLCMEntropy_AllDirection_offset1_SD
Selected features in predicting OS model in arterial-phase images:
GLCMEntropy_AllDirection_offset1_SD

Formula

$$-\sum_{i,j} g(i,j)\log_2(i,j)$$

3.3 Correlation

Image-based
Correlation measures the similarity of the grey levels in neighboring pixels, tells how correlated a pixel is to its neighbor over the whole image.
Range = [-1 1]. Correlation is 1 or -1 for a perfectly positively or negatively correlated image.
Selected features in predicting PD-L1 expression model in arterial-phase images:
Correlation_AllDirection_offset4_SD
Formula

$$-\sum_{i,j}\frac{(i-\mu)(j-\mu)g(i,j)}{\sigma^2}$$

Selected features in predicting OS model in arterial-phase images:
HaralickCorrelation_angle45_offset7
Formula

$$-\sum_{i,j}\frac{(i,j)g(i,j)-\mu_t^2}{\sigma_t^2}$$

3.4 Inverse Difference Moment

Inverse Difference Moment (IDM) is the local homogeneity. It is high when local gray level is uniform and inverse GLCM is high. IDM weight value is the inverse of the Contrast weight.
Selected features in predicting OS model in arterial-phase images:
InverseDifferenceMoment_AllDirection_offset7_SD

Formula

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

$$\sum\sum \frac{1}{1+(i-j)^2} g(i,j)$$

4. RLM Parameters

The grey level run-length matrix (RLM) $\mathbf{P_r(i,j \mid \theta)}$ is defined as the numbers of runs with pixels of gray level *i* and run length *j* for a given direction θ. RLMs is generated for each sample image segment having directions (0°,45°,90° &135°), then the following ten statistical features were derived: short run emphasis, long run emphasis, grey level non-uniformity, run length non-uniformity, Low Grey Level Run Emphasis, High Grey Level Run Emphasis, Short Run Low Grey Level Emphasis, Short Run High Grey Level Emphasis, Long Run Low Grey Level Emphasis and Long Run High Grey Level Emphasis.

Selected features in predicting PD-1 expression model in arterial-phase images:
HighGreyLevelRunEmphasis_AllDirection_offset1_SD
HighGreyLevelRunEmphasis_AllDirection_offset7_SD
LongRunLowGreyLevelEmphasis_angle0_offset4
RunLengthNonuniformity_AllDirection_offset1_SD
ShortRunEmphasis_AllDirection_offset1_SD
ShortRunHighGreyLevelEmphasis_AllDirection_offset7_SD
ShortRunLowGreyLevelEmphasis_AllDirection_offset1_SD
Selected features in predicting PD-L1 expression model in arterial-phase images:
ShortRunEmphasis_AllDirection_offset1_SD
ShortRunHighGreyLevelEmphasis_AllDirection_offset1_SD
ShortRunHighGreyLevelEmphasis_AllDirection_offset7_SD
ShortRunLowGreyLevelEmphasis_AllDirection_offset1_SD
Selected features in predicting OS model in arterial-phase images:
LongRunEmphasis_AllDirection_offset7_SD
LongRunEmphasis_Angle0_offset1
LongRunHighGreyLevelEmphasis_Angle45_offset7
LongRunLowGreyLevelEmphasis_Angle0_offset4
RunLengthNonuniformity_AllDirection_offset7_SD
ShortRunEmphasis_AllDirection_offset4_SD

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

4.1 High Grey Level Run Emphasis(18Parameters)

$$HGRE(\theta)=\frac{1}{n_r}\sum_{j=i}^{N}\sum_{i=1}^{M}p(i,j,\theta)i^2$$

4.2 Long Run Low Grey Level Emphasis(18Parameters)

$$LRLGE(\theta)=\frac{1}{n_r}\sum_{j=i}^{N}\sum_{i=1}^{M}\frac{p(i,j,\theta)j^2}{i^2}$$

4.3 Run Length Non-uniformity(18Parameters)

$$RLN(\theta)=\frac{1}{n_r}\sum_{j=i}^{N}(\sum_{i=1}^{M}p(i,j,\theta))^2$$

4.4 Short Run Low Grey Level Emphasis(18Parameters)

$$SRLGE(\theta)=\frac{1}{n_r}\sum_{j=i}^{N}\sum_{i=1}^{M}\frac{p(i,j,\theta)}{i^2j^2}$$

4.5 Short Run High Grey Level Emphasis(18Parameters)

$$SRHGE(\theta)=\frac{1}{n_r}\sum_{j=i}^{N}\sum_{i=1}^{M}\frac{p(i,j,\theta)i^2}{j^2}$$

4.6 Short Run Emphasis (18 Parameters)

$$SRE(\theta)=\frac{1}{n_r}\sum_{i=1}^{M}\sum_{j=1}^{N}\frac{p(i,j,\theta)}{j^2}$$

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s)

5. Reference:

[1] CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma Coroller, Thibaud P. et al. Radiotherapy and Oncology, Volume 114, Issue 3, 345 – 350

[2] P. Mohanaiah, P. Sathyanarayana and L. GuruKumar, *Image Texture Feature Extraction Using GLCM Approach*, International Journal of Scientific and Research Publications, Volume 3, Issue 5, May 2013

[3] *NIST/SEMATECH e-Handbook of Statistical Methods*, http://www.itl.nist.gov/div898/handbook/.

[4] Alnihoud J., "Content-Based Image Retrieval System Based on Self Organizing Map, Fuzzy Color Histogram and Subtractive Fuzzy Clustering," International Arab Journal of Information Technology, vol. 9, no. 5, pp. 452- 458, 2012.

[5] Balanda, Kevin P., and H. L. MacGillivray. 1988.
"Kurtosis: A Critical Review". *The American Statistician* 42(2), 111–119.
My thanks to Karl Ove Hufthammer for drawing this article to my attention.

[6] https://en.wikipedia.org/wiki/Kurtosis

[7] Bland, J.M.; Altman, D.G. (1996). "Statistics notes: measurement error" *BMJ*. **312** (7047): 1654.