**S1 File.**

In this study, we explored a feature-based approach to extraction and quantification of meaningful and reliable information from computed tomographic (CT) images. In this section, we describe in detail the different groups of radiomic features assessed in our study. We evaluated a total of 239 CT radiomic features, which were divided into seven groups as follows:

1. Physical features
2. Histogram-based features
3. Shape features
4. Local features
5. Filter-based features (LoG filter)
6. Fractal model-based features
7. Sigmoid features

***Group 1. Physical features***

Let denote the 3D image resolution,denote the tumor volume, and denote the tumor density.

* 1. **Volume:**
  2. **Mass:**
  3. **Density:**

Physical density (g/cm3) was estimated by extrapolating from the mean CT

scan attenuation.

***Group 2. Histogram-based features***

Histogram-based features describe the distribution of voxel intensities within the region of interest (ROI) of the CT image through commonly used basic metrics. Let denote the three-dimensional image matrix with voxels and the first-order histogram with discrete intensity levels.

* 1. **Energy:**
  2. **Entropy:**
  3. **Interquartile range:**

The interquartile range of intensity values of .

* 1. **Kurtosis:**where is the mean of .
  2. **Maximum:**The maximum intensity value of .
  3. **Mean:**
  4. **Median:**The median intensity value of .
  5. **Minimum:**The minimum intensity value of .
  6. **Range:**The range of intensity values of .
  7. **Root mean square (RMS):**
  8. **Skewness:**where is the mean of .
  9. **Standard deviation:**

where is the mean of .

* 1. **Uniformity:**
  2. **Variance:**where is the mean of .

The standard deviation and variance are measures of 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 histogram sharpness. As measures of histogram randomness, we computed the uniformity and entropy of the image histogram. From the histogram, CT numbers or Hounsfield units (HUs) at the 2.5th, 25th, 50th, 75th, and 97.5th percentiles were also computed.

***Group 3. Shape features***

In this group of features, we included descriptors of the three-dimensional size and shape of the tumor region. Let denote the volume and the surface area of the volume of interest. We computed the following shape- and size-based features:

* 1. **Compactness**:
  2. **Convexity:**

Measures ratio of the ROI volume contained within the tumor to the calculated convex hull volume.

where denotes tumor volume and denotes convex hull volume [1].

* 1. **Eccentricity:**

A measure of how close a shape is to an ellipse on the 2D image.where is the distance from the center to a focus and is the distance from that focus to a vertex [2].

* 1. **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.

* 1. **Roundness factor:**

A measure of convexity of a tumor’s profile on the 2D image.

where is the area of the tumor and the length of tumor contour on the 2D image.

* 1. **Solidity:**

A measure of convexity of a tumor’s profile on the 2D image.

where is the area of the tumor and denotes the area of the convex hull at the boundary of the tumor on the 2D image.

* 1. **Spherical disproportion**:

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

* 1. **Sphericity**:
  2. **Surface area:**

The surface area is calculated by triangulation (i.e. dividing the surface into connected triangles) and is defined as   
where is the total number of triangles covering the surface and , and are edge vectors of the triangles.

* 1. **Surface to volume ratio**:

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

***Group 4. Local features***

The features from group 2 (histogram-based features) provide information related to the gray-level distribution of the image. However, they do not provide any information regarding the relative positions of the various gray levels over the image. In this group, we therefore included local features describing patterns or the spatial distribution of voxel intensities, which were calculated from gray-level co-occurrence (GLCM)[3], gray level size zone (GLSZM)[4], and neighborhood gray tone difference (NGTDM)[5] texture matrices. Texture matrices were determined using 26-connected voxels (i.e., voxels considered to be neighbors in all 13 directions in three dimensions).

*Gray-level co-occurrence matrix-based features*

A GLCM is defined as , a matrix with size describing the second-order joint probability function of intensity occurrence pairs of an image, where the th element represents the number of times the combination of intensity levels and occur in two pixels in the image, that are separated by a distance of pixels in direction , and is the number of discrete gray-level intensities. As a two-dimensional example, let the following matrix represent a 5 × 5 image, having five discrete gray levels:

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

In this study, distance was set to 1 and direction to each of the 13 directions in three dimensions, yielding a total of 13 GLCMs for each 3D image. The average of 13 matrices was used as the final GLCM. We then computed various GLCM features from the final GLCM.

Let:

be the co-occurrence matrix for an arbitrary and ,

be the number of discrete intensity levels in the image,

be the mean of ,

be the marginal row probabilities,

be the marginal column probabilities,

be the mean of ,

be the mean of ,

be the standard deviation of ,

be the standard deviation of ,

, , ,

, , ,

be the entropy of ,

be the entropy of ,

be the entropy of ,

,

.

* 1. **Autocorrelation:**
  2. **Cluster Tendency:**
  3. **Contrast:**
  4. **Difference entropy:**
  5. **Dissimilarity:**
  6. **Energy:**
  7. **Entropy:**
  8. **Homogeneity:**
  9. **Informational measure of correlation (IMC):**
  10. **Maximum Probability:**
  11. **Variance:**

*Intensity size zone matrix-based features*

Intensity size zone metrics (ISZM) quantify size zone matrices in an image. Instead of looking in several directions as in a GLCM, these metrics look at a flat zone size for the whole image. A flat zone is a group of connecting pixels with the same gray level. In ISZM , the th element describes the frequency of matrices of size with gray level , and is the number of discrete gray level intensities. As a 2D example, consider the following 5×5 image with five discrete gray levels:

The ISZM matrix becomes

Let

be the th entry in the given size matrix

be the number of discrete intensity values in the image,

be the number of different size matrices, and

be the number of voxels in the image.

* 1. **Intensity variability (IV)**
  2. **Size zone variability (SZV)**

*Neighborhood gray tone difference matrix-based features*

Neighborhood gray tone difference matrix (NGTDM) is a column matrix [5]. It is calculated using neighborhood intensity differences. The entry of the NGTDM can be denoted, defined as



where  is the set of all voxels with gray-level  in the tumor volume (including the peripheral region),  is the number of voxels with gray-level  in the tumor volume, and  is the average gray level of the  connected neighbors around a center voxel  with gray level . In this study, we calculated five texture features from NGTDM. The features were defined previously [6] as follows.

* 1. **Coarseness**

High coarseness means that gray-level differences in the

neighborhood are small.



where  is a small number to prevent coarseness from becoming infinite,and *Ng* is the number of discrete intensity values in theimage.

* 1. **Contrast**

Contrast measures homogeneity in the ROI. A small contrast value means that

the intensity difference between neighborhood pixels is small.



* 1. **Busyness**

Busyness measures how fast the intensity changes in the ROI.



* 1. **Complexity**

Complexity measures the visual information content of a texture in the ROI.



* 1. **Strength**

In general, a high strength value (strong texture) means that texture features are easily defined and clearly visible.



where  is a small number to prevent strength becoming infinite.

***Group 5. Filter-based features (LoG features): Histogram-based features of the Laplacian of Gaussian***

The Laplacian of an image brings out areas of rapid intensity change and is usually used for edge detection. A Gaussian filter is applied prior to the Laplacian to smooth the image and reduce noise.

The equation of LoG with 2D kernels is:

Texture size (fine to coarse) was specified by modifying the Gaussian radius parameter (from 0.5 to 3.5 mm, in 0.5-mm increments). Histogram-based features (mean, maximum, median, minimum, entropy, uniformity, standard deviation, skewness, and kurtosis) described in Group 2 were investigated.

***Group 6. Fractal model-based features***

**6.1. Fractal dimension (Box-counting):**

The fractal dimension describes the relationship between changes in a measuring scale and measurement results at the scale. In this study, a 3D box-counting algorithm was applied to calculate the fractal dimension to quantify the intensity of tumor homogeneity [2, 7-9].

**6.2. Lacunarity:**

The lacunarity is a measure of the texture or distribution of gaps within an image [8]. It is a counterpart to the fractal dimension and has to do with the size distribution of holes. If a fractal has large gaps or holes, it has high lacunarity; in contrast, if a fractal is almost translationally invariant, it has low lacunarity.

**6.3. Fractal signature dissimilarity**

Fractal signature dissimilarity [10] is defined as the slope of the best linear fit of blanket surface area changes. This feature describes tumor heterogeneity. The blanket method [11] was adopted to calculate the fractal signature dissimilarity.

***Group 7. Sigmoid features***

To quantify tumor margins, the sigmoid function is used to fit density change along a sampling line drawn orthogonal to the tumor surface. Each sampling line, going through one voxel on the tumor surface, has a certain length (3, 5, and 7 mm in this work) inside and outside the tumor.

**References**

1. Grove O, Berglund AE, Schabath MB, Aerts HJ, Dekker A, Wang H, et al. Quantitative computed tomographic descriptors associate tumor shape complexity and intratumor heterogeneity with prognosis in lung adenocarcinoma. PloS one. 2015;10(3):e0118261. Epub 2015/03/05. doi: 10.1371/journal.pone.0118261. PubMed PMID: 25739030; PubMed Central PMCID: PMCPMC4349806.

2. Aerts HJ, Grossmann P, Tan Y, Oxnard GR, Rizvi N, Schwartz LH, et al. Defining a radiomic response phenotype: a pilot study using targeted therapy in NSCLC. Scientific reports. 2016;6:33860. Epub 2016/09/21. doi: 10.1038/srep33860. PubMed PMID: 27645803; PubMed Central PMCID: PMCPMC5028716.

3. Haralick R, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;3(6):610-21.

4. Thibault G, Angulo J, Meyer F. Advanced Statistical Matrices for Texture Characterization: Application to Cell Classification. IEEE Trans Biomed Eng 2014;61(3):630–7.

5. Amadasun M, King R. Textural features corresponding to textural properties. IEEE Trans Sys Man Cyb 1989;19:1264-74.

6. Niu L, Qian M, Yang W, Meng L, Xiao Y, Wong KK, et al. Surface roughness detection of arteries via texture analysis of ultrasound images for early diagnosis of atherosclerosis. PloS one. 2013;8(10):e76880. Epub 2013/10/23. doi: 10.1371/journal.pone.0076880. PubMed PMID: 24146940; PubMed Central PMCID: PMCPMC3798305.

7. Mandelbrot B. How long is the coast of britain? Statistical self-similarity and fractional dimension. Science (New York, NY). 1967;156(3775):636-8. Epub 1967/05/05. doi: 10.1126/science.156.3775.636. PubMed PMID: 17837158.

8. Lennon FE, Cianci GC, Cipriani NA, Hensing TA, Zhang HJ, Chen CT, et al. Lung cancer-a fractal viewpoint. Nature reviews Clinical oncology. 2015;12(11):664-75. Epub 2015/07/15. doi: 10.1038/nrclinonc.2015.108. PubMed PMID: 26169924; PubMed Central PMCID: PMCPMC4989864.

9. Kido S, Kuriyama K, Higashiyama M, Kasugai T, Kuroda C. Fractal analysis of internal and peripheral textures of small peripheral bronchogenic carcinomas in thin-section computed tomography: comparison of bronchioloalveolar cell carcinomas with nonbronchioloalveolar cell carcinomas. Journal of computer assisted tomography. 2003;27(1):56-61. Epub 2003/01/25. PubMed PMID: 12544244.

10. Wang C, Subashi E, Yin FF, Chang Z. Dynamic fractal signature dissimilarity analysis for therapeutic response assessment using dynamic contrast-enhanced MRI. Medical physics. 2016;43(3):1335-47. Epub 2016/03/05. doi: 10.1118/1.4941739. PubMed PMID: 26936718; PubMed Central PMCID: PMCPMC4760981.

11. Paskas MP, Reljin IS, Reljin BD. Multifractal framework based on blanket method. TheScientificWorldJournal. 2014;2014:894546. Epub 2014/03/01. doi: 10.1155/2014/894546. PubMed PMID: 24578664; PubMed Central PMCID: PMCPMC3919058.