Supplementary Information

Supplementary Figures

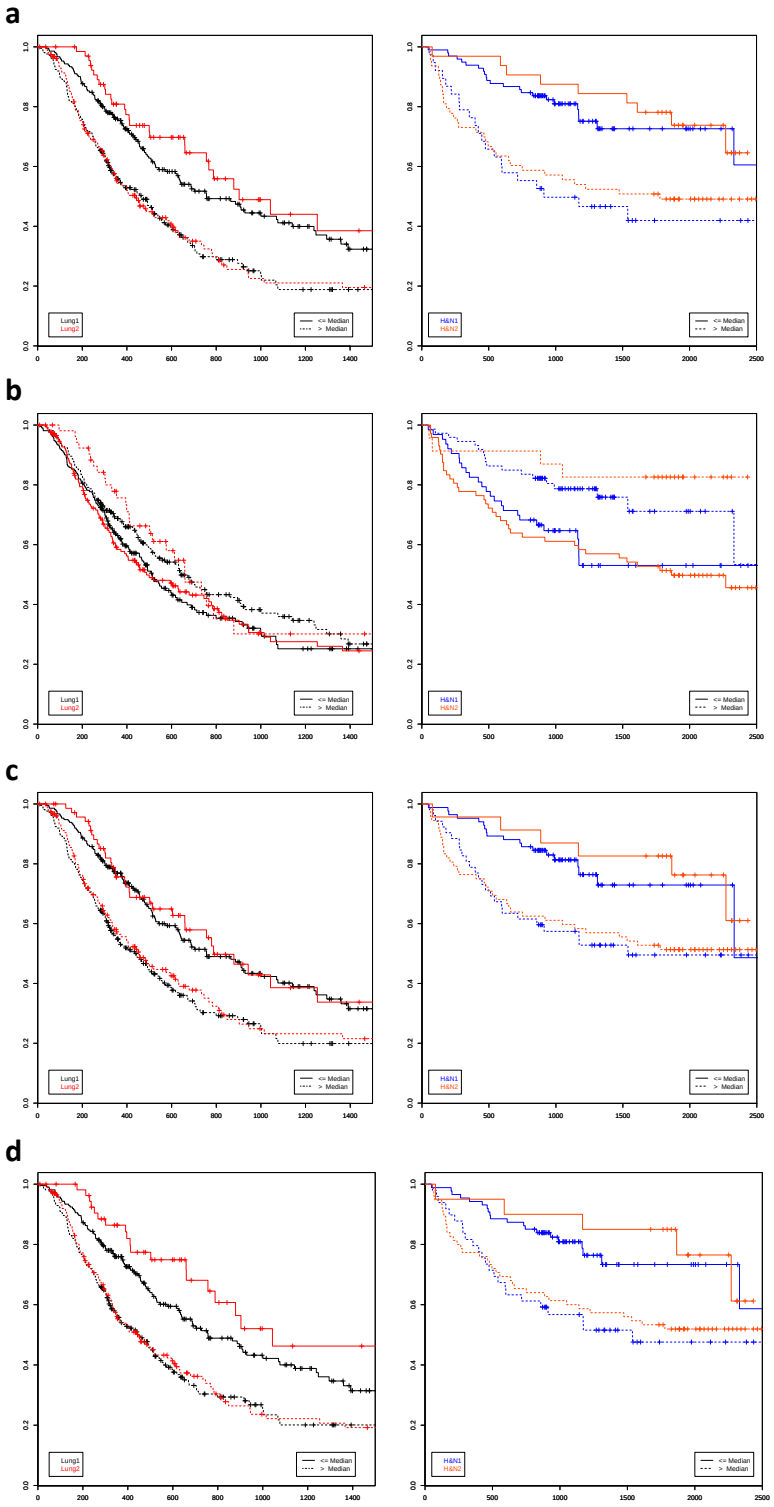

Figure 1 | Kaplan Meier survival curves of single radiomics features. **a** Statistics Energy, **b** Shape Compactness, **c** Gray Level Non-uniformity, and **d** Wavelet HLH Gray Level Non-uniformity. For each feature, the median value was trained on the Lung1 dataset, and selected as the threshold for split in the validation datasets Lung2, H&N1, and H&N2.

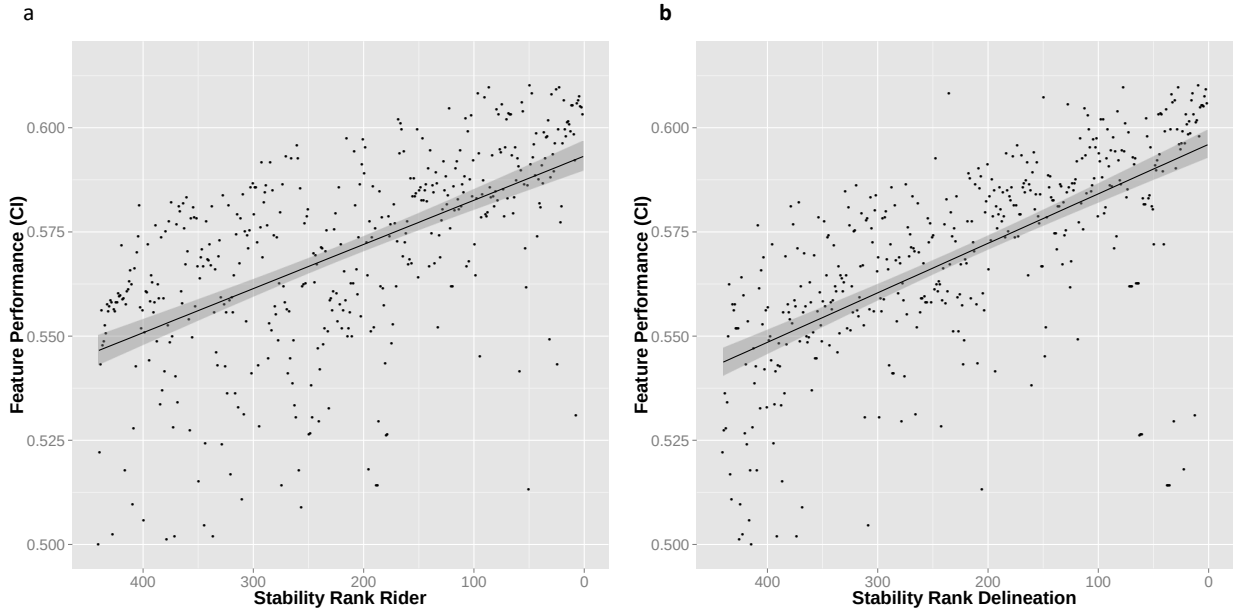

Figure 2 | Stability ranks of the RIDER test retest dataset (**a**), and the Multiple Delineation dataset (**b**) on the x-axis. The prognostic performance of single radiomics features is shown on the y-axis. Every dot indicates one feature. The prognostic performance was assessed by the concordance index (CI) in the Lung1 dataset. A stability rank of 1 indicates the most stable feature and 440 the least stable feature. The stability rank for RIDER test retest (**a**) was calculated using the intra-class correlation coefficient (ICC), giving a measure of how consistent the feature values were between the two scans. The stability rank for delineation inaccuracies was calculated using the Friedman test (non-parametric repeated measurement test) on the Multiple Delineation dataset (**b**). The lines specified a linear model fit, and the grey area the confidence interval. **a**: p-value: $< 2.2 \times 10^{-16}$ **b**: p-value: $< 2.2 \times 10^{-16}$. In general, features with higher stability showed higher prognostic performance, supporting the use of stability ranks for feature selection.

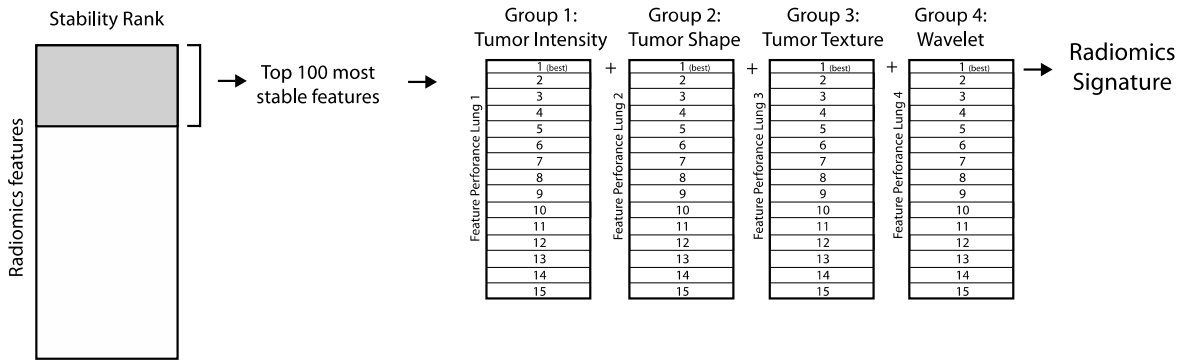

Figure 3 | Schematic representation of building the radiomics signature. First the top 100 most stable radiomic features are selected for further analysis. The stability rank was an average rank of the feature rank based on the RIDER and Multiple Delineation datasets. Next, from each of the four radiomic feature groups the best performing features in the Lung1 dataset were selected. These top four features were combined into a multivariate Cox proportional hazards regression model for prediction survival. The weights of the model were fitted on the Lung1 dataset.

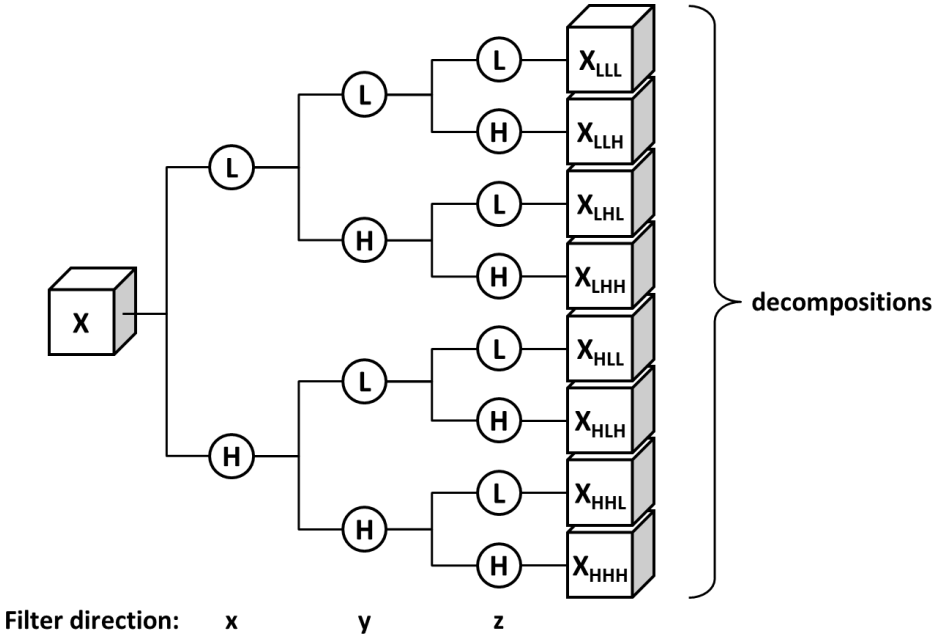

Figure 4 | Schematic of the undecimated three dimensional wavelet transform applied to each CT image. The original image X is decomposed into 8 decompositions, by directional low-pass (i.e. a scaling) and high-pass (i.e. a wavelet) filtering: X_{LLL}, X_{LLH}, X_{LHL}, X_{LHH}, X_{HLL}, X_{HLH}, X_{HHL} and X_{HHH}.

Supplementary Tables

Table 1 | Prognostic performance in validation datasets (Concordance Index CI)

Dataset	TNM	Volume	Radiomics	TNM-Radiomics	Volume-Radiomics	TNM vs. Radiomics	Volume vs. Radiomics	TNM vs. TNM-Radiomics	Volume vs. Volume-Radiomics
Lung2	0.60	0.63	0.65	0.64	0.65	**$1.42x10^{-04}$**	**$6.29x10^{-07}$**	**$1.40x10^{-05}$**	**$7.52x10^{-08}$**
H&N1	0.69	0.68	0.69	0.70	0.69	0.12	**$1.70x10^{-02}$**	**$3.79x10^{-04}$**	**$8.55x10^{-03}$**
H&N2	0.66	0.65	0.69	0.69	0.68	**$6.48x10^{-08}$**	**$3.72x10^{-18}$**	**$3.06x10^{-10}$**	**$2.52x10^{-18}$**

Prognostic performance in validation datasets (Concordance Index CI).

Table 2 | Prognostic performance of Radiomics Signature in treatment groups

	Lung 2		H&N 1		H&N 2	
Treatment	CI	p-value	CI	p-value	CI	p-value
Radiotherapy	0.66	**$1.97x10^{-18}$**	0.72	**$1.98x10^{-18}$**	0.78	**$1.98x10^{-18}$**
Chemo-Radiation	0.64	**$2.17x10^{-18}$**	0.65	**$2.60x10^{-18}$**	0.58	**$9.95x10^{-4}$**

Prognostic performance in validation datasets of Radiomics Signature for patients receiving radiotherapy and chemo-radiation (Concordance Index CI). Statistical prognostic performance was compared to randomness (CI=0.5) using the Wilcoxon test.

Table 3 | Radiomics Signature and HPV in Head and Neck cancer

Dataset	Number of patients HPV-	Number of patients HPV+	Prognostic performance HPV (CI)	Radiomics Signature difference between HPV- and HPV+ (p-value)	Radiomics Signature performance within HPV- group	Radiomics Signature performance within HPV+ group
H&N1	53	22	0.84	0.13	0.61	0.59
H&N2	77	18	0.81	0.29	0.70	0.51
Combined	130	40	0.83	0.17	0.66	0.53

Prognostic performance of Radiomics Signature and Human papillomavirus (HPV) status in the head and neck cancer datasets. The prognostic performance was assessed using the Concordance Index (CI). Statistical difference between the groups was assessed using the Wilcoxon test.

Table 4. Patient, tumour and patient characteristics dataset: Lung 1. MAASTRO NSCLC (n =422)

Variable	Frequency (%)
Gender	
male	68,7
female	31,3
T-stage	
T1	22,0
T2	37,0
T3	12,6
T4	27,7
Tx	0,7
N-stage	
N0	40,3
N1	5,5
N2	33,4
N3	20,1
Nx	0,7
TNM stage grouping	
stage I	22,0
stage II	9,5
stage IIIa	26,5
stage IIIb	41,9
Treatment	
Radiation only	46,5
Chemo-radiation	53,5

Table 5. Patient, tumor and patient characteristics dataset: Head and Neck 1. MAASTRO HNSCC (n =135)

Variable	Frequency (%)
Gender	
male	81,5
female	18,5
Primary Tumor site	
Oropharynx	64,0
Larynx	36,0
T-stage	
T1	25,9
T2	23,0
T3	17,8
T4	33,3
N-stage	
N0	45,2
N1	11,9
N2	40,7
N3	2,2
TNM stage grouping	
stage I	18,5
stage II	8,1
stage III	17,0
stage IVa	54,3
stage IVb	2,2
HPV status (oropharyngeal cancer only)	
Positive	28,4
Negative	71,6
Treatment	
Radiation only	74,1
Chemo-radiation	25,9

Table 6. Patient, tumor and patient characteristics dataset: Head and Neck 2. VUMC OPSCC (n =95)

Variable	Frequency (%)
Gender	
male	65,3
female	34,7
Primary Tumor site	
Oropharynx	100,0
T-stage	
T1	10,5
T2	32,6
T3	35,8
T4	21,1
N-stage	
N0	44,2
N1	11,6
N2	42,1
N3	2,1
TNM stage grouping	
stage I	8,4
stage II	18,9
stage III	18,9
stage IVa	45,3
stage IVb	7,4
stage IVc	1,1
HPV status	
Positive	18,9
Negative	81,1
Treatment	
Radiation only	58,9
Chemo-radiation	41,1

Table 7. Patient, tumor and patient characteristics dataset: Lung 3. MUMC NSCLC dataset (n =89)

Variable	Frequency (%)
Gender	
male	32,6
female	67,4
Primary Tumor site	
Lung	100,0
T-stage	
T1	25,4
T2	47,1
T3	21,1
T4	3,4
Unknown	2,2
N-stage	
N0	67,4
N1	20,2
N2	7,8
Nx	4,4
M-stage	
M0	93,3
M1	5,6
Mx	1,1
Histological sub-type	
Squamous cell carcinoma	37,1
Adenocarcinoma	28,1
Non-small cell	7,9
Adenocarcinoma – papillary	4,5
Adenocarcinoma – bronchiolo-alveolar	4,5
Mixed	17.9
Treatment	
Surgery	100,0

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 δ pixels in direction α, 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 δ was set to 1 and direction α 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 δ and α,

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

μ 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,

μ_x be the mean of p_x,

μ_y be the mean of p_y,

σ_x be the standard deviation of p_x,

σ_y be the standard deviation of p_y,

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

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

$HX = -\sum_{i=1}^{N_g} p_x(i) \log_2[p_x(i)]$ 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:

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

42. Sum entropy:

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

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 θ, 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:

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

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 θ,

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.

Datasets

RIDER NSCLC dataset

Memorial Sloan-Kettering Cancer Center (New York, NY)

This dataset includes 32 consecutive patients (mean age, 62.1 years; range, 29–82 years), with pathologically confirmed non–small cell lung cancer having measurable primary pulmonary tumors of 1 cm or larger, which were recruited between January 2007 through September 2007 at Memorial Sloan-Kettering Cancer Center (New York, NY). Sixteen patients were men (mean age, 61.8 years; range, 29–79 years) and 16 were women (mean age, 62.4 years; range, 45–82 years).

Each patient underwent two thoracic CT scans within 15 minutes of each other, acquired with the same CT scanner and using the same imaging protocol. CT scans were obtained with a 16-detector row LightSpeed 16 (GE Healthcare, Milwaukee, Wis) or a 64-detector row VCT (GE Healthcare) scanner. Thoracic images were acquired without intravenous contrast during a breath hold. Images were reconstructed without overlap by applying the lung convolution kernel. All images had an in-plane resolution of 0.576×0.576 mm/pixel and a slice thickness of 1.25 mm. Further patient and imaging details are described by Zhao et al[3]. The primary tumors were segmented, in both, test and retest scans, using a CT single click ensemble segmentation algorithm, running on Definiens Developer XD with the LuTA extension[4]. Data were downloaded from The Cancer Imaging Archive (TCIA).

Multiple delineation dataset (NSCLC)

Maastro Clinic, (Maastricht, The Netherlands)

Twenty-one consecutive patients with histologically proven non-small cell lung cancer, stages Ib–IIIb, were included. All patients had undergone a diagnostic whole body PET-CT scan (Biograph, SOMATOM Sensation 16 with an ECAT ACCEL PET scanner; Siemens, Erlangen, Germany). Patients were instructed to fast at least 6 h before the intravenous administration of 18F-fluoro-2-deoxy-glucose (FDG) (MDS Nordion, Liège, Belgium), followed by physiologic saline (10 mL). The total injected activity of FDG was

dependent on the patient weight expressed in kg: (weight／4) + 20 Mbq. After a period of 45 min, during which the patient was encouraged to rest, free-breathing PET and CT images were acquired. The CT scan was a spiral CT scan of the whole thorax with intravenous contrast. The PET images were acquired in 5-min bed positions. The CT data set was used for attenuation correction of PET images. The complete data set was then reconstructed iteratively with a reconstruction increment of 5 mm^3.

For all patients the primary tumour was delineated manually on CT/PET scans by five independent radiation oncologists. GTV manual delineations were based on fused PET-CT images using a standard clinical delineation protocol. Briefly, the protocol included fixed window level settings of both CT (lung W 1,700; L –300, mediastinum W 600; L 40) and PET scan (W 30,000; L 15,000) to be used for delineation[5]. All observers were blinded to each other's delineations. The primary gross tumor volume (GTV) was defined for each patient based on combined CT and PET information. Observers were given transversal, coronal, sagittal and 3D views simultaneously. Delineations were performed on a treatment planning system (XiO; Computer Medical System, Inc., St. Louis, MO). Data were downloaded from www.cancerdata.org.

Lung 1. MAASTRO NSCLC dataset

MAASTRO Clinic, (Maastricht, The Netherlands)

Patient population

Four hundred and twenty-two consecutive patients were included (132 women and 290 men), with inoperable, histologic or cytologic confirmed NSCLC, UICC stages I-IIIb, treated with radical radiotherapy alone (n = 196) or with chemo-radiation (n = 226). Mean age was 67,5 years (range: 33–91 years). The study has been approved by the institutional review board. All research was carried out in accordance with Dutch law. The Institutional Review Board of the Maastricht University Medical Center (MUMC+) waved review due to the retrospective nature of this study.

Treatment

During the study period, induction chemotherapy was standard of care for patients with N2/N3 and T4 tumors and consisted of three courses of gemcitabine (1,250 mg/m2 on days 1 and 8) in combination with cisplatin (75 mg/m2) or carboplatin (area under the concentration-time curve [AUC] 5) on day 1. Cycles were repeated every 21 days, and standard dose-reduction rules were applied. An interval between chemotherapy and start of radiotherapy of at minimum 14 days was mandatory.

All patients received an FDG PET-CT scan for radiotherapy treatment planning, in radiotherapy position on a dedicated PET-CT simulator with both arms above the head. For the FDG PET-CT scans a Siemens Biograph (SOMATOM Sensation-16 with an ECAT ACCEL PET scanner) was used. An intravenous injection of (weight * 4 + 20) MBq FDG (Tyco Health Care, Amsterdam, The Netherlands) was followed by 10 ml physiologic saline. After a 45-min uptake period, during which the patient was encouraged to rest, PET and CT images were acquired. A spiral CT (3 mm slice thickness) with or without intravenous contrast was performed covering the complete thoracic region.

Radiotherapy planning was performed on a XiO (Computerized Medical Systems, St Louis, Missouri) treatment planning system, based on a convolution algorithm using inhomogeneity corrections.

Delineation based on fused PET-CT images was performed by the radiation oncologist by using a standard clinical delineation protocol. The protocol included fixed window level settings of both CT (lung W1700; L‑300, mediastinum W600; L40) and PET scan (W30000; L15000) to be used for delineation. For all patients, a gross tumor volume (GTV) was defined based on FDG PET-CT data.

For patients treated with radical radiotherapy, the radiation dose was escalated to an individualized maximal total tumor dose, applying a mean lung dose of 19 Gy while respecting a maximum spinal cord dose of 54 Gy[5]. The maximal total tumor dose allowed was 79.2 Gy. There were no esophageal dose constraints. Radiotherapy was delivered twice a day in fractions of 1.8 Gy, 5 days per week, with a minimum of 8 h

between the two fractions. This protocol was applied as well in patients that received sequential chemo-radiation (n = 104).

Patients that received concurrent chemo-radiation (n = 100), were treated following 2 cycles of carboplatin-gemcitabine, a radiation dose of 45 Gy, in fractions of 1.5 Gy delivered twice a day for the first course, directly followed by an individualized dose ranging from 6 – 24 Gy and delivered in 2.0 Gy fractions once a day. In all patients, individualized patient dosimetry using electronic portal imaging devices was performed[6]. Table 4 provides detailed information about the patients included in this dataset.

Lung 2. Radboud NSCLC Dataset

Radboud University Nijmegen Medical Centre.

Patient population

This dataset included 225 consecutive patients with confirmed NSCLC (mean age, 65.5 years; range, 36–86 years), stages (I-IVa), treated at the Radboud University Nijmegen Medical Centre, The Netherlands, between February 2004 and October 2011.

Treatment

All primary tumors and the mediastinal N2 disease were cytologically or histologically proven. All patients underwent diagnostic work-up, including contrast enhanced CT of the thorax and upper abdomen, whole body 18F-FDG-PET/CT, MRI of the brain, bronchoscopy with transbronchial needle aspiration (TBNA), and/or oesophageal ultrasound fine needle aspiration (EUS-FNA) and/or endobronchial ultrasound with TBNA (EBUS-TBNA) and mediastinoscopy in case of PET-positive, cytologically negative mediastinal lymph nodes. After work up, all patients were discussed in a thoracic oncology multidisciplinary board. Prior to radiotherapy a CT of the thorax was performed in radiotherapy position for radiotherapy planning.

Patients in good general condition were treated with concurrent chemo radiotherapy, those with a contraindication for chemotherapy were treated by radiation alone, and all remaining patients were treated with a sequential chemotherapy and radiotherapy. The planned radiation dose to the primary tumor and metastatic mediastinal lymph nodes

using CRT until March 2008 and IMRT afterwards, was 66Gy in 33 fractions delivered five times per week. Chemotherapeutic agents in the sequential regimen typically consisted of three courses of gemcitabine (1250mg/m2; on day 1 and 8) and cisplatinum (80mg/m2; on day 1). The concurrent schedules varied between referring hospitals; in Radboud University Nijmegen Medical Centre it consisted of two courses of etoposide (100mg/m2; on day 1–3) and cisplatinum (50mg/m2; on day 1 and 8), in Canisius-Wilhelmina Hospital one course of gemcitabine/cisplatinum was administered prior to irradiation and two courses of etoposide/cisplatinum concurrently with radiation therapy. All research was carried out in compliance with the Helsinki Declaration and in accordance with Dutch law. The Institutional Review Board of the Radboud University Medical Center (RUMC) waved review due to the retrospective nature of this study. Follow-up was performed according to national guidelines.

Head and Neck 1. MAASTRO HNSCC

MAASTRO Clinic, (Maastricht, The Netherlands)

Patient population.

Included were 136 consecutive patients with squamous cell carcinoma of the head and neck, stages (I-IVb) treated at MAASTRO Clinic. Excluded from this analysis were patients treated with primary surgery with or without postoperative radiotherapy (n = 11). In addition, patients receiving no or palliative treatment (n = 4). Included patients were treated with curative intent with definitive radiation alone (n = 100) or with chemo-radiation (n = 35). Table 5 provides detailed information about the patients included in this dataset.

Treatment details

The treatment options consisted of either definitive radiotherapy alone or concurrent chemo-radiation. Patients treated with radiotherapy alone received a continuous course of radiotherapy delivered by 4–6 MV linear accelerator. Patients were treated with fractionation schedules in line with the state-of-the-art practices: Early glottic laryngeal cancers (T1N0) were treated locally only with standard fractionated radiation schedules to 60-66 Gy in single daily fractions of 2 - 2.40 Gy, using a technique with standard

parallel opposing wedged beams to the larynx only. Limited T2N0 glottic tumors and T1N0 supraglottic tumors were treated locally only with Accelerated Fractionated RadioTherapy (AFRT) to 68 Gy in 34 fractions over 37-38 days, the first 23 fractions 2 Gy daily, and the last 11 fractions twice daily in fractions of 2 Gy. or the locally more advanced and/or node positive laryngeal carcinomas, the target volume included the primary tumor in the larynx and the elective bilateral neck nodes. The Gross Tumor Volume (GTV) in the larynx and metastatic lymphnodes in the neck were treated up to 68-70 Gy depending on the fractionation scheme. In most cases, patients were treated with AFRT to 68 Gy in 34 fractions over 37-38 days, as described earlier. Patients in moderate general condition, who were deemed unfit for AFRT received standard fractionated radiotherapy to 70 Gy in 35 fractions over 7 weeks.

Patients with early oropharyngeal cancers (stage I-II) were treated with AFRT as described above to the GTV and elective nodal areas. Patients with locoregional advanced disease (stage III-IV) were treated locoregionally with concurrent chemo-radiation to a dose of 70 Gy in 35 fractions of 2 Gy to areas of gross disease over 7 weeks with a concurrent dose of cisplatin of 100 mg/m2 i.v. every 3 weeks[7], if they were in good general condition and below the age of 71 years. Patients in good condition but unfit for chemoradiation were treated with AFRT to 68 Gy in 34 fractions over 37-38 days, as described earlier. Elderly patients in moderate general condition received standard fractionated radiotherapy to 70 Gy in 35 fractions over 7 weeks.

The radiation treatment technique evolved over the years: IMRT was introduced gradually from 2006 with a simultaneous integrated boost technique, whereas the early patients (from 2004 on) were treated with a 3D-conformal technique using compensators.

CT scans

All patients underwent a treatment planning ^{18F}FDG-PET-CT scan (Biograph, SOMATOM Sensation-16 with an ECAT ACCEL PET scanner; Siemens, Erlangen, Germany) made with the patient immobilized using a thermoplastic mask. Patients fasted at least 6 h before the start of the acquisition. A total dose dependent on the weight of the patient (weight x 4 + 20 MBq) of [^{18}F] fluoro-2-deoxy-d-glucose (FDG)

(MDS Nordion, Liège, Belgium), was injected intravenously, followed by physiologic saline (10 mL). Free-breathing PET and CT images were acquired after an uptake period of 45 minutes. A spiral CT (3 mm slice thickness) was performed covering the complete thoracic region. Based on the radiological examinations and clinical findings, the gross tumor volume (GTV) was delineated on the fused PET-CT scan by a radiation oncologist in a radiotherapy treatment planning system (XiO, CMS, St Louis, MO). All research was carried out in accordance with Dutch law. The Institutional Review Board of Maastricht University Medical Center (MUMC+) waved review due to the retrospective nature of this study.

Head and Neck 2. VUMC OPSCC

Free University Medical Center, Amsterdam, The Netherlands.

Patient population

Ninety-five consecutive oropharyngeal squamous cell carcinoma patients, stages (I-IVc), treated with curative intent at the VU Medical Center in the period 2000 – 2006. Treatment options for these patients included definitive radiotherapy alone (n = 56), chemo-radiation with (n = 36) or without surgery (n = 3). Table 6 provides detailed information about the patients included in this dataset.

Treatment details

The definitive radiotherapy regime consisted of standard fractionated radiotherapy to 70 Gy in fractions of 2 Gy over 7 weeks. The concomitant chemo-radiation scheme included daily fractionation of 2Gy up to 70 Gy with a concomitant intra venous administration of cisplatin of with a dose of 100 mg/m, with or without neck dissection.

CT scans

All patients received a treatment planning CT scan of the head and neck (Varian Medical Systems VISION 3253). CT scans were acquired in helical mode with slice thickness of 2.5 mm. The gross tumor volume was delineated by an experienced radiation oncologist on the CT scans. All research was carried out in accordance with Dutch law. The Institutional Review Board of the VU University Medical Center (VUMC) waved review due to the retrospective nature of this study.

Lung 3. MUMC NSCLC dataset

Maastricht University Medical Centre+, (Maastricht, The Netherlands)

Patient population

Eighty-nine histologic confirmed NSCLC patients (29 women and 60 men), stages Ia – IV, treated at the Maastricht University Medical Centre with surgical dissection. Mean age was 65 years (range: 37 – 83). The study has been approved by the institutional review board of Maastricht University Medical Centre+. All research was carried out in accordance with Dutch law. Informed consent was acquired for each patient included in this study. Table 7 provides detailed information about the patients included in this dataset.

All primary tumors were cytologically or histologically proven. All patients underwent diagnostic work-up, including a CT scan of the thorax and upper abdomen and whole body 18F-FDG-PET/CT. After work up, all patients were discussed in a thoracic oncology multidisciplinary board. All patients were referred for surgical dissection.

The CT scans had a slice thickness range between 1.5mm and 5mm. Primary tumors were segmented using the CT-based single-click ensemble segmentation algorithm[4].

Supplementary References

1. Haralick, R.M., Shanmugam, K. & Dinstein, I.H. Textural Features for Image Classification. *Systems, Man and Cybernetics, IEEE Transactions on* **SMC-3**, 610-621 (1973).
2. Galloway, M.M. Texture analysis using gray level run lengths. *Computer Graphics and Image Processing* **4**, 172-179 (1975).
3. Zhao, B.*, et al.* Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. *Radiology* **252**, 263-272 (2009).
4. Rios Velazquez, E.*, et al.* A semiautomatic CT-based ensemble segmentation of lung tumors: comparison with oncologists' delineations and with the surgical specimen. *Radiother Oncol* **105**, 167-173 (2012).
5. van Baardwijk, A.*, et al.* PET-CT-based auto-contouring in non-small-cell lung cancer correlates with pathology and reduces interobserver variability in the delineation of the primary tumor and involved nodal volumes. *Int J Radiat Oncol Biol Phys* **68**, 771-778 (2007).
6. Nijsten, S.M., Mijnheer, B.J., Dekker, A.L., Lambin, P. & Minken, A.W. Routine individualised patient dosimetry using electronic portal imaging devices. *Radiother Oncol* **83**, 65-75 (2007).
7. Forastiere, A.A.*, et al.* Concurrent chemotherapy and radiotherapy for organ preservation in advanced laryngeal cancer. *N Engl J Med* **349**, 2091-2098 (2003).