Electronic Supplementary Material

Robustness of Radiomic Features in $[^{11}C]$Choline and $[^{18}F]$FDG PET/CT Imaging of Nasopharyngeal Carcinoma: Impact of Segmentation and Discretization

Journal: Molecular Imaging and Biology

Lijun Lu[1], Wenbing Lv[1], Jun Jiang[1], Jianhua Ma[1,*], Qianjin Feng[1,*], Arman Rahmim[2,3], Wufan Chen[1]

[1]School of Biomedical Engineering and Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, Guangdong 510515, China

[2]Department of Radiology, Johns Hopkins University, Baltimore, MD 21287, USA

[3]Department of Electrical and Computer Engineering, Johns Hopkins University, Baltimore, MD 21218, USA

[*]Authors to whom any correspondence should be addressed.

Email: ljubme@gmail.com; jhma@smu.edu.cn; qianjinfeng08@gmail.com

Appendix A. Definition of imaging features

● *First order gray level statistics features-22*

Let P define the first-order histogram of tumor volume. $P(i)$ represents the number of voxels with gray level i, and N_g represents the number of gray-level bins set for P. The i^{th} entry of the normalized histogram is then defined as:

$$p(i) = \frac{P(i)}{\sum_{i=1}^{N_g} P(i)}$$

1. SUVmax: the maximum SUV value.
2. SUVmean: the mean SUV value.

3. SUVpeak: defined as the mean SUV within a 26 connected neighborhoods volume centered the maximum SUV voxel.
4. SUVstd: the standard deviation of all SUV values.
5. SUVvar: the variance of all SUV values.
6. SUVenergy: the sum of all voxel SUV values squared.
7. AUC-CSH: Area under the curve of the cumulative SUV-volume histogram describing the percentage of total tumor volume above a percentage threshold of maximum SUV [1]
8. Max_intensity: the maximum intensity value.
9. Mean_intensity: the mean intensity value.
10. Min_intensity: the minimum intensity value.
11. Median_intensity: the median intensity value.
12. Range_intensity: the range of intensity value.
13. MAD_intensity: Mean absolute deviation, the mean of the absolute deviations of all voxel intensities around the mean intensity value.
14. STD_intensity: standard deviation, the standard deviation of all voxel intensities around the mean intensity value.
15. RMS_intensity: root mean square, the quadratic mean, or the square root of the mean of squares of all voxel intensities.

$$RMS = \sqrt{\frac{\sum_{i=1}^{N_g} i^2}{N_g}}$$

16. Mean_hist:

$$\mu = \sum_{i=1}^{Ng} ip(i)$$

17. Variance_hist:

$$\sigma^2 = \sum_{i=1}^{N_g} (i-\mu)^2 p(i)$$

18. Skewness_hist:

$$s = \sigma^{-3} \sum_{i=1}^{N_g} (i-\mu)^3 p(i)$$

19. Kurtosis_hist:

$$k = \sigma^{-4} \sum_{i=1}^{N_g} (i - \mu)^4 p(i) - 3$$

20. Energy_hist:

$$energy\_hist = \sum_{i=1}^{N_g} p(i)^2$$

21. Entropy_hist:

$$entropy\_hist = -\sum_{i=1}^{Ng} p(i) \log_2 [p(i)]$$

22. TLG: total lesion glycolysis, defined as the product of MATV and SUVmean.

● *Shape geometric features-9*

Shape geometric features, describing the shape and size of the volume of interest. Let V be the volume and A the surface area of the volume of interest.

23. MATV: metabolically active tumor volume
24. Eccentricity: find an ellipsoid that best fits the tumor region, and the eccentricity is then given by $(1 - a \times \frac{b}{c^2})^{\frac{1}{2}}$, where c is the longest semi-principal axes of the ellipsoid, a and b are the second and third longest semi-principal axes of the ellipsoid.
25. Solidity: ratio of the number of voxels in the tumor region to the number of voxels in the 3D convex hull of the tumor region (smallest polyhedron containing the tumor region).
26. PI: percent inactive, percentage of the tumor region that is inactive. A threshold of $0.005 \times SUV\max^2$ followed by closing and opening morphological operations were used to differentiate active and inactive regions on PET scans.
27. SurfaceA: the surface area of the volume of interest.
28. SVratio: the surface area divided by the volume.
29. Compactness 1:

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

30. Compactness 2:

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

31. Sphericity：

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

- ***Gray Level Co-occurrence Matrix-based features (GLCM)-26***

Gray level co-occurrence matrix-based features, as described by study [2]. Let: $P(i,j)$ be the co-occurrence matrix, N_g be the number of discrete intensity levels in the image, μ be the mean of $P(i,j)$, $\mu_x(i)$ be the mean of row i, $\mu_x(j)$ be the mean of column j, $\sigma_x(i)$ be the standard deviation of row i, $\sigma_y(j)$ be the standard deviation of column j.

$$p_x\left(i\right) = \sum_{j=1}^{N_g} P\left(i,j\right) \quad p_y\left(j\right) = \sum_{i=1}^{N_g} P\left(i,j\right)$$

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

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

$$HXY1 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P\left(i,j\right)\log_2\left(p_x\left(i\right)p_y\left(j\right)\right) \quad HXY2 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} P_x\left(i\right)P_y\left(j\right)\log_2\left(p_x\left(i\right)p_y\left(j\right)\right)$$

$$HX = -\sum_{i=1}^{N_g} p_x(i)\log_2[p_x(i)] \quad HY = -\sum_{i=1}^{N_g} p_y(i)\log_2[p_y(i)] \quad H = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\log_2[p(i,j)]$$

32. Energy, called Uniformity in [3], also called Angular second moment in [4]:

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

33. Entropy:

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

34. Difference entropy (DifEntropy):

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

35. Sum entropy (SumEntropy):

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

36. Variance1:

$$variance1 = \frac{1}{N_g \times N_g}\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}[(i-\mu_x)^2 p(i,j) + (j-\mu_y)^2 p(i,j)]$$

37. Variance2:

$$variance2 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}(i-\mu)^2 P(i,j)$$

38. Sum variance (SumVariance):

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

Where SE is Sum entropy

39. Maximum probability (MaxPossilility):

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

40. Contrast:

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

41. Dissimilarity:

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

42. Homogeneity 1, also called Inverse difference in [3]:

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

43. Homogeneity 2, also called local homogeneity in [5]:

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

44. Correlation1:

$$correlation1 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{(i-\mu_x)(j-\mu_y)p(i,j)}{\sigma_x\sigma_y}$$

45. Correlation2:

$$correlation2 = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ijP(i,j) - \mu_x(i)\mu_y(j)}{\sigma_x(i)\sigma_y(j)}$$

46. Auto correlation (AutoCorrelation):

$$auto\ correlation = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} ij\,p(i,j)$$

47. Cluster prominence (ClusterPro):

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

48. Cluster shade (ClusterShade):

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

49. Cluster tendency (ClusterTen):

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

50. Informational measure of correlation 1 (IMC1):

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

Where *HX and HY* are the entropies of p_x and p_y.

51. Informational measure of correlation 2 (IMC2):

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

where *H* is the entropy.

52. Inverse difference moment (InvDifMoment) also called inverse variance:

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

53. Inverse Difference Moment Normalized (IDMN):

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

54. Inverse Difference Normalized (IDN):

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

55. Sum average1:

$$sum\ average1 = \frac{1}{N_g \times N_g}\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}[iP(i,j)+jP(i,j)]$$

56. Sum average2:

$$sum\ average2 = \sum_{i=2}^{2N_g}\left[iP_{x+y}(i)\right]$$

57. Agreement:

$$agreement = \frac{P_o - P_e}{1-P_e}$$

$$\text{where}\quad P_o = \sum_{i=1}^{Ng} P(i,i) \qquad P_e = \sum_{i=1}^{Ng} P(i,:)P(:,i)$$

- ***Gray Level Run Length Matrix-based features (GLRLM)-13***

Gray-level run-length matrix-based features, as described by Galloway et al.[6]. Let: $P(i,j)$ be the $(i,j)th$ entry in the given run-length matrix, N_g the number of discrete intensity values in the image, N_r the number of different run lengths, N_p is the number of voxels in the image, and the entry (i,j) of the normalized GLRLM is defined as:

$$p(i,j) = \frac{P(i,j)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}P(i,j)} \qquad \mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{N_r} p(i,j) \qquad \mu_j = \sum_{j=1}^{N_r} j \sum_{i=1}^{N_g} p(i,j)$$

58. Short Run Emphasis (SRE):

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

59. Long Run Emphasis (LRE):

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

60. Gray Leven Non-Uniformity (GLN):

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

61. Run Length Non-Uniformity (RLN):

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

62. Run Percentage (RP):

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

63. Low Gray Level Run Emphasis (LGRE):

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

64. High Gray Level Run Emphasis (HGRE):

$$HGRE = \sum_{i=1}^{N_g}\sum_{j=1}^{N_r} i^2 p(i,j)$$

65. Short Run Low Gray Level Emphasis (SRLGE):

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

66. Short Run High Gray Level Emphasis (SRHGE):

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

67. Long Run Low Gray Level Emphasis (LRLGE):

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

68. Long Run High Gray Level Emphasis (LRHGE):

$$LRHGE = \sum_{i=1}^{N_g}\sum_{j=1}^{N_r} p(i,j)i^2 j^2$$

69. Gray Level Variance (GLV)

$$GLV = \frac{1}{N_g \times N_r}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}(ip(i,j)-\mu_i)^2$$

70. Run length Variance (RLV)

$$RLV = \frac{1}{N_g \times N_r}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}(jp(i,j)-\mu_j)^2$$

- ***Gray Level Size Zone Matrix-based features (GLSZM)-13***

Gray-level size-zone matrix-based features, was described in [2]. Let: $P(i,j)$ be the $(i,j)th$ entry in the given size-zone matrix, N_g the number of discrete intensity values in the image, N_z the size of the largest homogeneous region in the volume of interest, N_α the number homogeneous zones in the image. The entry (i,j) of the GLSZM then normalized as:

$$p(i,j) = \frac{P(i,j)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_z} P(i,j)} \quad \mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{N_z} p(i,j) \quad \mu_j = \sum_{j=1}^{N_z} j \sum_{i=1}^{N_g} p(i,j)$$

71. Small Zone Emphasis (SZE):

$$SZE = \sum_{i=1}^{N_g}\sum_{j=1}^{N_z}\left[\frac{p(i,j)}{j^2}\right]$$

72. Large Zone Emphasis (LZE):

$$LZE = \sum_{i=1}^{N_g}\sum_{j=1}^{N_z} j^2 p(i,j)$$

73. Gray Level Non-uniformity (GLN) also called Intensity Variability (IV) in [7]:

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

74. Zone Size Non-uniformity (ZSN) also called Size Zone Variability (SZV) in [7]:

$$ZSN = \sum_{i=1}^{N_z}\left[\sum_{j=1}^{N_g} p(i,j)\right]^2$$

75. Zone Percentage (ZP):

$$ZP = \sum_{i=1}^{N_z}\sum_{j=1}^{N_g}\frac{p(i,j)}{N_\alpha}$$

76. Low Gray Level Zone Emphasis (LGZE) also called Low Intensity Emphasis (LIE) in [7]:

$$LGZE = \sum_{i=1}^{N_g}\sum_{j=1}^{N_z}\left[\frac{p(i,j)}{i^2}\right]$$

77. High Gray level Zone Emphasis (HGZE) also called High Intensity Emphasis (HIE) in [7]:

$$HGZE = \sum_{i=1}^{N_g}\sum_{j=1}^{N_z} i^2 p(i,j)$$

78. Small Zone Low Gray Level Emphasis (SZLGE) also called Low Intensity Small Area Emphasis (LISAE) in [7]:

$$SZLGE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[\frac{p(i,j)}{i^2 j^2} \right]$$

79. Small Zone High Gray-Level Emphasis (SZHGE) also called High Intensity Small Area Emphasis (HISAE) in [7]:

$$SZHGE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[\frac{p(i,j) i^2}{j^2} \right]$$

80. Large Zone Low Gray-Level Emphasis (LZLGE) also called Low Intensity Large Area Emphasis (LILAE) in [7]:

$$LZLGE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left[\frac{p(i,j) j^2}{i^2} \right]$$

81. Large Zone High Gray-Level Emphasis (LZHGE) also called High Intensity Large Area Emphasis (HILAE) in [7]:

$$LZHGE = \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} p(i,j) i^2 j^2$$

82. Gray Level Variance (GLV)

$$GLV = \frac{1}{N_g \times N_z} \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} (ip(i,j) - \mu_i)^2$$

83. Zone Size Variance (ZSV)

$$ZSV = \frac{1}{N_g \times N_z} \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} (jp(i,j) - \mu_j)^2$$

where zone aforesaid also called area in [7].

● *Neighborhood Gray Tone Difference Matrix–based features (NGTDM)-5*

NGTDM is a column matrix [8], Let i^{th} entry of the NGTDM is $P(i)$, defined as:

$$P(i) = \begin{cases} \sum_{i \in \{N_i\}} \left| i - \overline{A_i} \right| & if \ N_i > 0, \\ 0 & otherwise. \end{cases}$$

where $\{N_i\}$ is the set of all voxels with gray-level i in tumor volume (including the peripheral region), N_i is the number of voxels with gray-level i in tumor volume, and A_i is the average gray level of the 26-connected neighbors around a center voxel $V(i, j, k)$ with gray level i.

$$\overline{A_i} = \overline{A}(j,k,l) = \frac{1}{W}\sum_{m=-d}^{d}\sum_{n=-d}^{d}\sum_{s=-d}^{d} V(j+m,k+n,l+s),(m,n,l) \neq (0,0,0)$$

where $d = 1$, specifies the neighborhood size as $3\times3\times3$, and $W = (2d+1)^3$, The quantity $n_i = \frac{N_i}{N}$ is also defined, where N is the total number of voxels in tumor volume. The NGTDM texture features are then defined as:

84. Coarseness:

$$coarseness = [\varepsilon + \sum_{i=1}^{N_g} n_i P(i)]^{-1}$$

where ε is a small number to prevent coarseness becoming infinite, N_g the number of discrete intensity values in the image.

85. Contrast:

$$contrast = [\frac{1}{N_g \times (N_g - 1)} \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} n_i n_j (i-j)^2] \times [\frac{1}{N}\sum_{i=1}^{N_g} P(i)]$$

86. Busyness:

$$busyness = \frac{\sum_{i=1}^{N_g} n_i P(i)}{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \left| i n_i - j n_j \right|}, n_i \neq 0, n_j \neq 0$$

87. Complexity:

$$complexity = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{\left| i-j \right| \left[n_i P(i) + n_j P(j) \right]}{N(n_i + n_j)}, n_i \neq 0, n_j \neq 0$$

88. Strength:

$$strength = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} (n_i - n_j)(i-j)^2}{\varepsilon + \sum_{i=1}^{N_g} P(i)}, n_i \neq 0, n_j \neq 0$$

where ε is a small number to prevent strength becoming infinite.

Supplementary Appendix B.

Table 1. The Pearson correlation coefficient (r) between the 6 most robust features.

r	Entropy _GLCM	DifEntropy _GLCM	SumEntropy _GLCM	Homogeneity1 _GLCM	Homogeneity2 _GLCM	Coarseness _NGTDM
Entropy _GLCM	1.00	0.94	0.99	-0.99	-0.99	0.19
DifEntropy _GLCM		1.00	0.91	-0.97	-0.97	0.02
SumEntropy _GLCM			1.00	-0.97	-0.97	0.23
Homogeneity1 _GLCM				1.00	0.99	-0.17
Homogeneity2 _GLCM					1.00	-0.16
Coarseness _NGTDM						1.00

Table 2. Clinical data of patients with NPC.

Pt. No.	Scanning Date*		PET scan interval(days)	Sex	Age(y)	Pathology
	FDG	CH				
14641	8.22	8.25	3	M	47	NKUC
14952	10.22	10.24	2	F	42	NKUC
15392	1.5	1.6	1	F	49	NKUC
15677	3.6	3.9	3	M	58	NKUC
15778	3.18	2.4	3	M	46	NKUC
15865	4.2	4.3	1	M	75	NKUC
15866	4.2	4.3	1	M	40	NKUC
15918	4.13	4.14	1	M	41	NKUC
18937	7.20	7.20	0	M	49	NKUC

Abbreviation: Pt. No. = Patient Number, *: Scanning date are formated as month-date
NKUC: nonkeratinizing undifferentiated carcinoma

References:

1. van Velden FH, Cheebsumon P, Yaqub M et al (2011) Evaluation of a cumulative SUV-volume histogram method for parameterizing heterogeneous intratumoural FDG uptake in non-small cell lung cancer PET studies. Eur J Nucl Med Mol Imaging 38:1636-1647
2. Thibault G, Fertil B, Navarro C et al (2009) Texture indexes and gray level size zone matrix application to cell nuclei classification.
3. Gomez W, Pereira WC, Infantosi AF (2012) Analysis of co-occurrence texture statistics as a function of gray-level quantization for classifying breast ultrasound. IEEE Trans Med Imaging 31:1889-1899
4. Lee J, Jain R, Khalil K et al (2015) Texture Feature Ratios from Relative CBV Maps of Perfusion MRI Are Associated with Patient Survival in Glioblastoma. AJNR Am J Neuroradiol
5. El NI, Grigsby P, Apte A et al (2009) Exploring feature-based approaches in PET images for predicting cancer treatment outcomes. Pattern Recognit 42:1162-1171

6. Galloway MM (1974) Texture analysis using grey level run lengths. NASA STI/Recon Technical Report N 75:18555
7. Leijenaar RT, Carvalho S, Velazquez ER et al (2013) Stability of FDG-PET Radiomics features: an integrated analysis of test-retest and inter-observer variability. Acta Oncol 52:1391-1397
8. Amadasun M, King R (1989) Textural features corresponding to textural properties. Systems, Man and Cybernetics, IEEE Transactions on 19:1264-1274