Development and verification of radiomics framework for computed tomography image segmentation

Jiabing Gu[a, b], Baosheng Li[a, b]*, Huazhong Shu [a]*, Jian Zhu[b, c], Qingtao Qiu[b], Tong Bai[b],

* Corresponding author: Baosheng Li: bsli@sdfmu.edu.cn

*Co-corresponding author: Huazhong Shu: shu.list@seu.edu.cn

[a] Southeast University, Laboratory of Image Science and Technology, Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Centre de Recherche en Information Biomédicale Sino-français (CRIBs), Nanjing, 210096, P.R. China

[b] Department of Radiation Oncology Physics and Technology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan250117, China

[c] Shandong Key Laboratory of Digital Medicine and Computer Assisted Surgery, The Affiliated Hospital of Qingdao University, Qingdao 266000, P.R. China

Running Title: CT Image Segmentation Using Radiomics

The mathematical definitions of 53 radiomic features used in this study

Group 1. First-order statistics features (FOS)

Let *P* define the first-order histogram of volume. *P (i)* represents the number of voxels with gray-level *i*, and N_g represents the number of gray-level bins set. 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. Kurtosis (Kur):

$$\text{kurtosis} = \sigma^{-4} \sum_{i=1}^{N_g} [(i-\mu)^4 p(i)] - 3$$

2. Skewness (Ske):

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

3. Variance (Var):

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

4. Entropy[1]:

$$\text{entropy} = -\sum_{i=1}^{N_g} p(i) log_2(p(i))$$

5. MinValue: Minimum CT value of volume.
6. MaxValue: Maximum CT value of volume.
7. Mean Value: Mean CT value of volume.
8. Quantile25: 25%Quantiles of the intensity values among the volume.
9. Quantile75: 75%Quantiles of the intensity values among the volume.
10. InterQuartile Range (IQR): Calculate the quartile distance and equal to quantile_75-quantile_25 in this paper.
11. Standard Deviation (SD): The standard deviation intensity values of volume
12. Range: The intensity range (MaxValue-MinValue) among volume.
13. Mean Absolute Deviation (MAD): The average distance between each data value and the mean.

Group 2. Textural features

The textural feature can provide information regarding the relative position of the various gray levels within the region where features are extracted from. The texture features which describe spatial relationships in Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), Gray Level Size Zone Matrix (GLSZM), and Neighboring Gray Tone Difference Matrix (NGTDM) were calculated in this study. A MATLAB toolbox created by Vallières[1] can be used to facilitate the computation of GLCM, GLRLM, GLSZM, and NGTDM features from the 3D volumes.

2.1 Gray Level Co-occurrence Matrix features

Gray level co-occurrence matrix-based features, as described by Thibault et al. [2] Let: *P(i,j)* represent the co-occurrence matrix of volume with isotropic voxel, N_g represents the predefined number of quantized gray levels in the image.

$$u_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{N_g} P(i,j)$$

$$u_j = \sum_{i=1}^{N_g} j \sum_{j=1}^{N_g} P(i,j)$$

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

$$\sigma_j = \sum_{j=1}^{N_g} \left(j - u_j\right)^2 \sum_{i=1}^{N_g} P(i,j)$$

1. Energy:

$$\text{Energy} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [p(i,j)]^2$$

2. Contrast:

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

3. Entropy:

$$\text{Entropy} = -\sum_{j=1}^{N_g} \sum_{i=1}^{N_g} p(i,j) log_2 [p(i,j)]$$

4. Homogeneity:

$$\text{Homogeneity} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\frac{p(i,j)}{1+|i-j|}$$

5. Correlation:

$$\text{correlation} = \frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)(i-u_i)(j-u_j)}{\sigma_i\sigma_j}$$

6.Variance:

$$\text{variance} = \frac{1}{N_g \times N_g}\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\left[(i-u_i)^2 p(i,j) + (j-u_j)^2 p(i,j)\right]$$

7. Dissimilarity:

$$\text{Dissimilarity} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}|i-j|p(i,j)$$

8. Sum Average:

$$\text{sum average} = \frac{1}{N_g \times N_g}\sum_{i=1}^{N_g}\sum_{j=1}^{N_g}[i\,p(i,j) + j\,p(i,j)]$$

9. Autocorrelation:

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

2.2 Gray Level Run Length Matrix Features

GLRLM quantifies gray level runs, as described by Galloway[3]. Let *P(i,j)* represent the*(i,j)^th^* entry in the given run-length matrix, N_g represents the predefined number of discrete intensity values in the image, N_r represents the number of the different run lengths. The entry (*i*,*j*) of the normalized GLRLM is defined as:

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

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

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

1. GrayLevelNonuniformity (GLN):

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

2. LowGray-LevelRunEmpha (LGRE):

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

3. LongRunLowGray-LevelEmpha (LRLGE):

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

4. LongRunHighGray-LevelEmpha (LRHGE):

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

5. LongRunEmphasis (LRE):

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

6. HighGray-LevelRunEmpha (HGRE):

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

7. RunPercentage (RP):

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

8. RunLengthNonuniformity (RLN):

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

9. ShortRunEmphasis (SRE):

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

10. ShortRunLowGray-LevelEmpha(SRLGE):

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

11. ShortRunHighGray-LevelEmpha (SRHGE):

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

12. Gray Level Variance (GLV):

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

13. Run-Length Variance (RLV):

$$\text{RLV} = \frac{1}{N_g \times N_r}\sum_{i=1}^{N_g}\sum_{j=1}^{N_r}\left(j\,p(i,j) - u_j\right)^2$$

2.3 Gray Level Size Zone Matrix Features

The GLSZM quantifies gray level zones in an image as described by Thibault et al. [2]. Let: *P (i, j)* represent the $(i,j)^{th}$ entry in a given size-zone matrix, N_g represents the predefined number of discrete gray levels set in the images, and N_z represents the size of the largest homogeneous region in the predefined volume, The entry *(i,j)* of normalized GLSZM defined as:

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

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

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

1. Large Zone Emphasis (LZE):

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

2. Small Zone Emphasis (SZE):

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

3. Small Zone High Gray-Level Emphasis (SZHGE):

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

4. Small Zone Low Gray-Level Emphasis (SZLGE):

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

5. Large Zone High Gray-Level Emphasis (LZHGE):

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

6. Large Zone Low Gray-Level Emphasis (LZLGE):

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

7. Zone-Size Variance (ZSV):

$$\text{ZSV} = \frac{1}{N_g \times N_z} \sum_{i=1}^{N_g} \sum_{j=1}^{N_z} \left(j\, p(i,j) - u_j\right)^2$$

8. Zone-Size Non-Uniformity (ZSN):

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

9. Zone Percentage (ZP):

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

10. Gray Level Variance (GLV):

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

11. Low Gray-Level Zone Emphasis (LGZE):

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

12. High Gray-Level Zone Emphasis (HGZE):

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

13. Gray Level Non-Uniformity (GLN):

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

2.4 Neighboring Gray Tone Difference Matrix Features

NGTDM quantifies the difference between a gray value and the average gray value of its neighbors. *P(i)* is the sum of absolute differences for gray level *i*, which is defined as:

$$P(i) = \begin{cases} \sum_{i\epsilon\{N_i\}} |i - \bar{A}_i| & for\ N_i > 0 \\ 0 & for\ N_i = 0 \end{cases}$$

where $\{N_i\}$ is the set of all voxels with gray-level *i* in the predefined volume, N_i is the number of voxels with gray-level *i* in the predefined volume. $\bar{A}_i$ is the average gray level of the 26-connected neighbors around a center voxel (*i, j, k*) with gray-level *i.*

$$\bar{A}_\iota = \bar{A}_\iota(j,k,l) = \frac{\sum_{m=-1}^{1}\sum_{n=-1}^{1}\sum_{s=-1}^{1} w_{m,n,s}\cdot V(j+m,k+n,l+s)}{\sum_{m=-1}^{1}\sum_{n=-1}^{1}\sum_{s=-1}^{1} w_{m,n,s}},\ (m,\ n,\ s)\neq(0,0,0)$$

Where $w_{m,n,s} = \begin{cases} 1 & if\ |j-m| + |k-n| + |l-s| = 1 \\ \frac{1}{\sqrt{2}} & if\ |j-m| + |k-n| + |l-s| = 2 \\ \frac{1}{\sqrt{3}} & if\ |j-m| + |k-n| + |l-s| = 3 \end{cases}$

The gray level probability $n_i = \frac{N_i}{N}$ is also defined, where *N* is the total number of voxels in the volume of interest. N_g represents the predefined number of the number of discreet gray levels in the image.

1. Contrast:

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

2. Coarseness:

$$\text{coarseness} = \left[\epsilon + \sum_{i=1}^{N_g} n_i P(i)\right]^{-1}$$

ϵ is a small number to prevent coarseness from becoming infinite

3. Busyness:

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

4. Complexity:

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

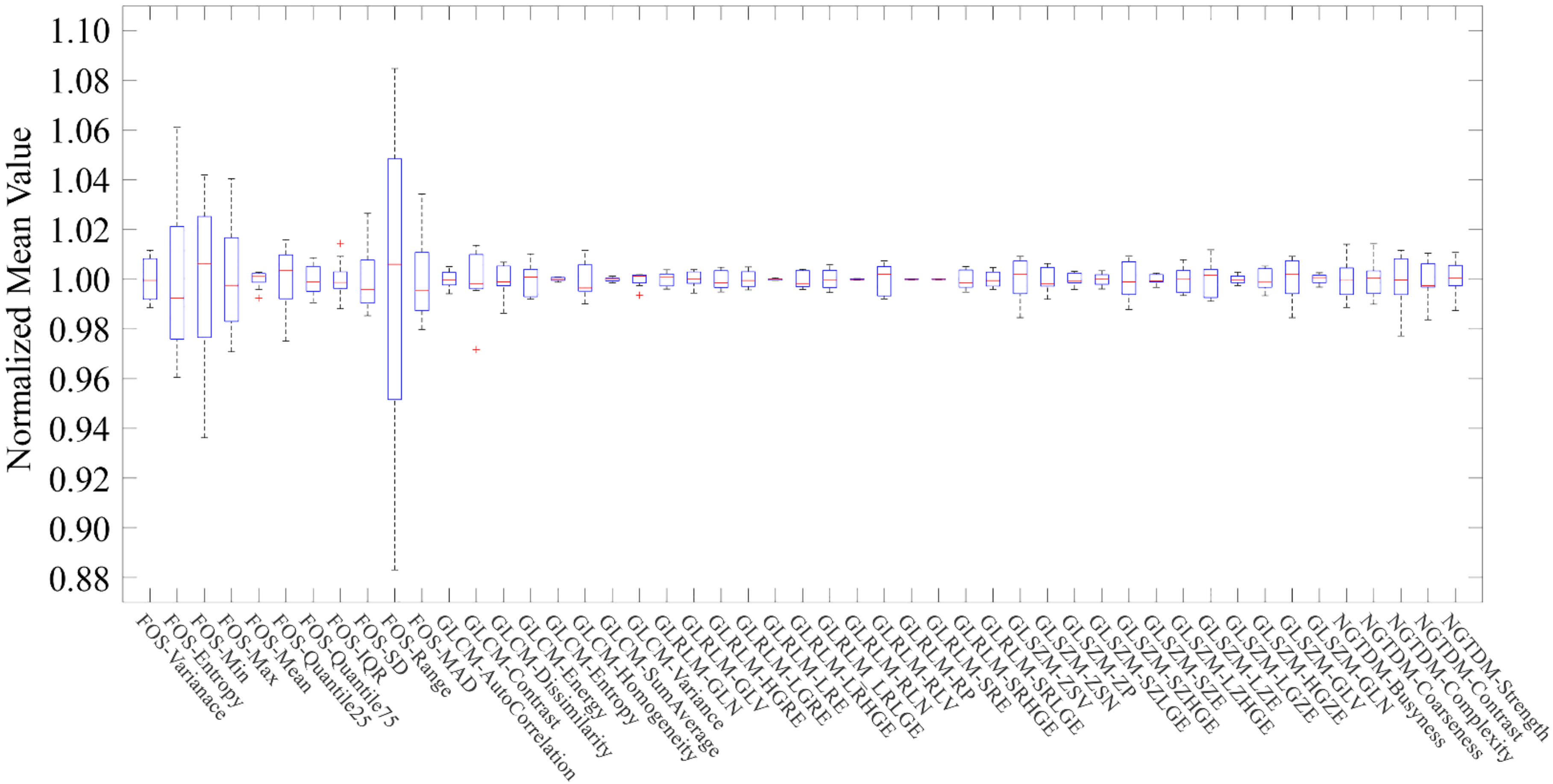
5. Strength:

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

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

Reference:

1. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Physics in Medicine & Biology*. 2015;60(14):5471.
2. Thibault G, Fertil B, Navarro C, et al. Texture indexes and gray level size zone matrix. *Application to Cell Nuclei Classification PRIP*. 2009:140-145.
3. Galloway MM. Texture analysis using grey level run lengths. *NASA STI/Recon Technical Report N*. 1974;75:18555.



FigS1. The normalized mean value distribution of the 49 reproducible svfeatures extracted from 11 CCR phantom image series

Table S1 Image acquisition parameters for the CCR phantom scans

Scanner	Manufacturer	Model	Type	Slice Thickness	Pixel	Spiral Pitch Factor	kVp	Effective mAs	No. Swvolumes
CCR-GE1	GE	Discovery CT750 HD	axial	2.5 mm	0.70 mm	1	120	300	4332
CCR-GE2	GE	Discovery CT750 HD	helical	2.5 mm	0.78 mm	0.98	120	122	3675
CCR-GE3	GE	Discovery ST	helical	2.5 mm	0.98 mm	1.35	120	143	2352
CCR-GE4	GE	LightSpeed RT	helical	2.5 mm	0.98 mm	0.75	120	1102	2352
CCR-GE5	GE	LightSpeed RT16	helical	2.5 mm	0.98 mm	0.94	120	367	2352
CCR-GE6	GE	LightSpeed VCT	helical	2.5 mm	0.74 mm	0.98	120	82	3996
CCR-P1	Philips	Brilliance Big Bore	helical	3 mm	0.98 mm	0.94	120	320	2352
CCR-P2	Philips	Brilliance Big Bore	helical	3 mm	0.98 mm	0.94	120	369	2352
CCR-P3	Philips	Brilliance Big Bore	helical	3 mm	1.04 mm	0.81	120	320	2028
CCR-P4	Philips	Brilliance Big Bore	helical	3 mm	1.04 mm	0.81	120	369	2106
CCR-P5	Philips	Brilliance 64	helical	3 mm	0.98 mm	0.67	120	372	2352

Table S2 Selected Svfeatures for RFIS_{LC} Construction

Categories	Feature Name		
FOS[a]	1. IQR	2. Range	3. Variance
GLCM[b]	4. Entropy	5. SumAverage	
GLRLM[c]	6. Run-Length Variance	7. ShortRunLowGray-LevelEmpha	
	8. Zone-Size Variance	9. Gray Level Variance	
	10.Small Zone High Gray-Level Emphasis	11.Large Zone High Gray-Level Emphasis	
NGTDM[d]	12. Busyness	13. Complexity	

[a] FOS: First-order statistics.

[b] GLCM: Gray-level co-occurrence matrix.

[c] GLRLM: Gray-level run-length matrix.

[d] NGTDM: Neighbourhood gray-tone difference matrix.