

1 **Supplemental Appendix 1: Radiomics System of CAD Scheme**

2 A total of 409 quantitative features were designed and extracted for each MDCT
3 image to quantify and assess tumor characteristics. We classified guideline-based features
4 and radiomics high-throughput features into six specific subcategories: (1) the
5 demographic information, (2) the tumor location, (3) morphology features, (4) intensity
6 features, (5) texture features, and (6) wavelet features. We show the detailed feature list
7 in **Supplemental Table 1**.

- 8 1. The demographic information about the age and sex are obtained from the PACS.
- 9 2. The feature of the tumor location is recorded by radiologists, which reflects the
10 location information of tumors in the pancreas (head, neck, body, or tail).
- 11 3. Morphological features are used to describe the geometric and radiology
12 characteristics of a tumor, including the size, shape, margin, cyst number, cyst
13 size and calcification.
- 14 4. Intensity features reflect the gray-level and histogram distributions of the images.
15 T or H representing this feature is calculated on gray levels or histograms. For
16 example, T-kurtosis is the kurtosis of gray levels while H-kurtosis is the kurtosis
17 of histograms.
- 18 5. Five types of texture features are used to reflect detailed and invisible gray-level

- 1 variations and relationships in an image.
- 2 a. Region of interest-based features depict the distribution of gray levels in the
3 entire region of interest.
- 4 b. Gray-level co-occurrence matrix (GLCM) features show the spatial linear
5 relationship between the frequencies of two gray-level intensities in a certain
6 range.
- 7 c. Gray-level run-length matrix (GLRLM) features are based on the gray-level
8 run lengths. The length of the run is the number of pixels in the run, reflecting the
9 size of the texture elements.
- 10 d. Gray-level size zone matrix (GLSZM) features are based on the co-occurrences
11 of the intensity of each flat zone. They provide a statistical representation by
12 estimating a bivariate conditional probability density function of the image
13 distribution values.
- 14 e. Neighborhood gray-tone difference matrix (NGTDM) features reflect the
15 differences between the pixel and its surrounding neighbors.
- 16 6. Wavelet decomposition features are applied to find more detailed intensity and
17 texture information. The discrete wavelet transform iteratively decomposes a
18 two-dimensional image into four components and all intensity and texture

1 features are calculated in each component.

2

3 **Supplemental Table 1: List of 409 Quantitative Features in the Radiomics System**

Guideline-based features	Demography information	Age
		Sex
	Tumor location	Tumor location
	Morphologic features	Compactness
		Spiculation
		Extreme point number
		Rectangle-fitting factor
		Roundness
		Convexity
		Solidity
		Moment difference
		Diameter of equivalent circle
		SD of normalized radius
		Entropy of normalized radius histogram
		Edge roughness
		Area-circumference ratio
		Number of calcification
		SD of calcification area
		Calcification location
		Central scar density
Boundary area density		
Number of cyst		
Cyst size		
SD of cyst area		
Radiomics high-throughput features	Intensity features	T-energy
		T-entropy
		T-kurtosis
		T-mean
		T-mean absolute deviation
		T-median
		T-range
		T-uniformity
		T-variance
		T-root mean square

		T-skewness
		T-SD
		H-kurtosis
		H-mean
		H-variance
		H-skewness
	ROI-based texture features	Mean of contrast
		SD of contrast
		Mean of covariance
		SD of covariance
		Mean of non-similarity
		SD of non-similarity
		SD of tumor area
	GLCM texture features	Autocorrelation
		Contrast
		Correlation 1
		Correlation 2
		Cluster prominence
		Cluster shade
		Dissimilarity
		Energy
		Entropy
		Homogeneity 1
		Homogeneity 2
		Maximum probability
		Sum of squares: variance
		Sum average
		Sum variance
		Sum entropy
		Difference variance
		Difference entropy
		IMC1
IMC2		
Inverse difference		
Inverse difference normalized		
Inverse difference moment normalized		
GLRLM texture features	Short-run emphasis	
	Long-run emphasis	

		Gray-level nonuniformity
		Run-length nonuniformity
		Run percentage
		Low gray-level run emphasis
		High gray-level run emphasis
		Gray-level variance
		Short-run high gray-level emphasis
		Long-run low gray-level emphasis
		Long-run high gray-level emphasis
		Short-run low gray-level emphasis
		Run-length variance
	GLSZM texture features	Small zone emphasis
		Large zone emphasis
		Gray-level nonuniformity
		Zone-size nonuniformity
		Zone percentage
		Low gray-level zone emphasis
		High gray-level zone emphasis
		Gray-level variance
		Small-zone high gray-level emphasis
		Large-zone low gray-level emphasis
		Large-zone high gray-level emphasis
		Small-zone low gray-level emphasis
		Zone-size variance
	NGTDM texture features	Coarseness
		Contrast
		Busyness
		Complexity
		Strength
	Wavelet decomposition features	All intensity and texture features are calculated for each component.

1 Abbreviations: GLCM = gray-level co-occurrence matrix; GLRLM = gray-level run-
2 length matrix; GLSZM = gray-level size zone matrix; IMC = informational measure of
3 correlation; NGTDM = neighborhood gray-tone difference matrix; ROI = region of
4 interest; SD = standard deviation.

5

6 **Supplemental Appendix 2: Optimal Radiomics Features**

1 The 22 most frequently selected features in all bootstrapping repetitions were used to
 2 construct the final optimal feature subsets (**Supplemental Table 2**).

3
 4 **Supplemental Table 2: Summary of Optimal Feature Subsets**

Category	Feature Name
Guideline-based features	Sex
	Tumor location
	Rectangle-fitting factor
	Moment difference
	Cyst size
Radiomics high-throughput features	Intensity T-range
	Intensity H-kurtosis
	Wavelet intensity H-skewness
	Wavelet mean of ROI contrast
	Wavelet mean of ROI covariance
	Wavelet GLCM sum variance
	NGTDM contrast
	NGTDM busyness
	Wavelet intensity T-median
	Wavelet intensity T-mean
	Wavelet intensity H-mean
	Wavelet GLCM contrast
	Wavelet GLCM sum average
	Wavelet GLCM difference variance
	Wavelet GLCM cluster shade
	Wavelet NGTDM complexity
	Wavelet NGTDM busyness

5 Abbreviations: GLCM = gray-level co-occurrence matrix; NGTDM = neighborhood
 6 gray-tone difference matrix; ROI = region of interest.

7