

Multicollinearity and redundancy of the PET radiomic feature set

ELECTRONIC SUPPLEMENTARY MATERIAL

Supplementary table 1. Cluster analysis of averaged correlation matrix in five independent [¹⁸F]FDG PET cohorts at a very strong Spearman correlation ($r > 0.9$). Table presents representative non-redundant features, correlating redundant features, and an explanation of the cluster. GLCM: grey level cooccurrence matrix, GLRLM: grey level run length matrix, GLSZM: grey level size zone matrix, GLDM: grey level dependence matrix, NGTDM: neighbouring grey tone difference matrix, TLG: total lesion glycolysis.

Non-redundant feature (number non-redundant features/total number of features)	Very strongly correlating redundant features	Cluster explanation
Shape (6/13)		
Elongation		Ratio between longest and second largest axis
Flatness		Ratio between the longest and shortest axis
Least axis length		Shortest axis
Sphericity		Roundness
Surface-to-volume ratio		Ratio between area and volume
Volume	Major axis length (shape) Maximum 2D diameter column (shape) Maximum 2D diameter row (shape) Maximum 2D diameter slice (shape) Maximum 3D diameter (shape) Minor axis length (shape) Surface area (shape) Coarseness (NGTDM)	Size
Intensity (4/18)		
Energy	Total energy (intensity)	Magnitude of voxel values, size-dependent
Kurtosis		'Peakedness' of the voxel value distribution
Maximum	10 th Percentile (intensity) 90 th Percentile (intensity) Entropy (intensity)	Highest voxel intensity

	Interquartile range (intensity) Mean (intensity) Mean absolute deviation (intensity) Median (intensity) Minimum (intensity) Range (intensity) Robust mean absolute deviation (intensity) Root mean squared (intensity) Uniformity (intensity) Variance (intensity) Autocorrelation (GLCM) Cluster prominence (GLCM) Cluster tendency (GLCM) Contrast (GLCM) Difference average (GLCM) Difference entropy (GLCM) Difference variance (GLCM) Inverse difference (GLCM) Inverse difference moment (GLCM) Joint average (GLCM) Joint energy (GLCM) Joint entropy (GLCM) Maximum probability (GLCM) Sum entropy (GLCM) Sum squares (GLCM) Grey level nonuniformity normalized (GLRLM) Grey level variance (GLRLM) High grey level run emphasis (GLRLM) Short run high grey level emphasis (GLRLM) Grey level variance (GLSZM) Grey level variance (GLDM) High grey level emphasis (GLDM)	
--	--	--

	Small dependence high grey level emphasis (GLDM) Complexity (NGTDM)	
Skewness		Asymmetry of the voxel value distribution
GLCM (6/22)		
Cluster shade		Independency of voxel values between neighbouring voxels
Correlation		Dependency of voxel values between neighbouring voxels
Inverse difference moment normalized	Inverse difference normalized (GLCM)	Local homogeneity
Inverse measure of correlation 1		Complexity of the texture
Inverse measure of correlation 2		Complexity of the texture
Inverse variance		Local homogeneity
GLRLM (4/16)		
Grey level nonuniformity	Grey level nonuniformity (GLDM)	Similarity of voxel values
Long run emphasis	Run length nonuniformity normalized (GLRLM) Run percentage (GLRLM) Run variance (GLRLM) Short run emphasis (GLRLM) Large area emphasis (GLSZM) Zone percentage (GLSZM) Zone variance (GLZSM) Dependence nonuniformity normalized (GLDM) Dependence variance (GLDM) Large dependence emphasis (GLDM) Small dependence emphasis (GLDM)	Row of voxels with the same value
Long run high grey level emphasis	Run entropy (GLRLM)	Row of voxels with a high value
Low grey level run emphasis	Long run low Grey level emphasis (GLRLM)	Row of voxels with a low value

	Short run low Grey level emphasis (GLRLM) Low grey level emphasis (GLDM)	
GLSZM (9/16)		
Grey level nonuniformity		Variability of voxel values
High grey level zone emphasis	Grey level nonuniformity normalized (GLSZM) Small area high grey level emphasis (GLZSM)	Clusters of voxels with high value
Large area high grey level emphasis		Large clusters of voxels with high value
Large area low grey level emphasis	Large dependence low grey level emphasis (GLDM)	Clusters of voxels with low value
Low grey level zone emphasis		Clusters of voxels with low value
Size zone nonuniformity	Zone entropy (GLSZM)	Variability of cluster sizes
Size zone nonuniformity normalized		Variability of cluster sizes
Small area emphasis		Fine texture with small clusters
Small area low grey level emphasis		Small clusters with low value
GLDM (2/14)		
Large dependence high grey level emphasis		Large clusters of voxels with high values
Small dependence low grey level emphasis		Small clusters of voxels with low values
NGTDM (3/5)		
Busyness		Rapid changes in intensity between neighbouring voxels
Contrast		Changes in intensity between neighbouring voxels combined with the intensity range
Strength		Slow changes in intensity between neighbouring voxels

Total lesion glycolysis	Run length nonuniformity (GLRLM) Dependence entropy (GLDM) Dependence nonuniformity (GLDM)	Product of intensity and volume
--------------------------------	--	---------------------------------

Supplementary table 2. Cluster analysis of averaged correlation matrix in five independent [¹⁸F]FDG PET cohorts at a strong Spearman correlation ($r > 0.7$). Table presents representative non-redundant features, correlating redundant features, and an explanation of the cluster. GLCM: grey level cooccurrence matrix, GLRLM: grey level run length matrix, GLSZM: grey level size zone matrix, GLDM: grey level dependence matrix, NGTDM: neighbouring grey tone difference matrix, TLG: total lesion glycolysis.

Non-redundant feature (number non-redundant features/total number of features)	Strongly correlating redundant features	Cluster explanation
Shape (4/13)		
Elongation	Flatness (shape)	Ratio between the axes
Sphericity		Roundness
Surface-to-volume ratio		Ratio between area and volume
Volume	Least axis length (shape) Major axis length (shape) Maximum 2D diameter column (shape) Maximum 2D diameter row (shape) Maximum 2D diameter slice (shape) Maximum 3D diameter (shape) Minor axis length (shape) Surface area (shape) Energy (intensity) Total energy (intensity) Total lesion glycolysis Run length nonuniformity (GLRLM) Grey level nonuniformity (GLSZM) Dependence entropy (GLDM) Dependence nonuniformity (GLDM) Coarseness (NGTDM)	Size
Intensity (2/18)		
Kurtosis	Skewness (intensity)	Shape of the voxel value distribution

Maximum	10th Percentile (intensity)	Highest voxel intensity
	90th Percentile (intensity)	
	Entropy (intensity)	
	Interquartile range (intensity)	
	Mean (intensity)	
	Mean absolute deviation (intensity)	
	Median (intensity)	
	Minimum (intensity)	
	Range (intensity)	
	Robust mean absolute deviation (intensity)	
	Root mean squared (intensity)	
	Uniformity (intensity)	
	Variance (intensity)	
	Autocorrelation (GLCM)	
	Cluster prominence (GLCM)	
	Cluster tendency (GLCM)	
	Contrast (GLCM)	
	Difference average (GLCM)	
	Difference entropy (GLCM)	
	Difference variance (GLCM)	
	Inverse difference (GLCM)	
	Inverse difference moment (GLCM)	
	Joint average (GLCM)	
	Joint energy (GLCM)	
	Joint entropy (GLCM)	
	Maximum probability (GLCM)	
	Sum entropy (GLCM)	
	Sum squares (GLCM)	
	Grey level nonuniformity normalized (GLRLM)	
	Grey level variance (GLRLM)	
	High grey level run emphasis (GLRLM)	
	Long run emphasis (GLRLM)	
	Long run high grey level emphasis (GLRLM)	

	Long run low grey level emphasis (GLRLM) Low grey level run emphasis (GLRLM) Run entropy (GLRLM) Run length nonuniformity normalized (GLRLM) Run percentage (GLRLM) Run variance (GLRLM) Short run emphasis (GLRLM) Short run high grey level emphasis (GLRLM) Short run low grey level emphasis (GLRLM) Grey level variance (GLSZM) Grey level nonuniformity normalized (GLSZM) High grey level zone emphasis (GLSZM) Large area emphasis (GLSZM) Size zone nonuniformity (GLSZM) Small area high grey level emphasis (GLSZM) Zone entropy (GLSZM) Zone percentage (GLSZM) Zone variance (GLSZM) Dependence nonuniformity normalized (GLDM) Dependence variance (GLDM) Grey level variance (GLDM) High grey level emphasis (GLDM) Large area low grey level emphasis (GLSZM) Large dependence emphasis (GLDM) Large dependence low grey level emphasis (GLDM)	
--	--	--

	Low grey level emphasis (GLDM) Small dependence emphasis (GLDM) Small dependence high grey level emphasis (GLDM) Busyness (NGTDM) Complexity (NGTDM) Contrast (NGTDM) Strength (NGTDM)	
GLCM (4/22)		
Cluster shade		Independency of voxel values between neighbouring voxels
Correlation		Dependency of voxel values between neighbouring voxels
Inverse measure of correlation 1	Inverse measure of correlation 2 (GLCM)	Complexity of the texture
Inverse variance		Local homogeneity
GLRLM (1/16)		
Grey level nonuniformity (GLRLM)	Inverse difference moment normalized (GLCM) Inverse difference normalized (GLCM) Large area high grey level emphasis (GLSZM) Grey level nonuniformity (GLDM)	Similarity of voxel values
GLSZM (2/16)		
Low grey level zone emphasis	Small area low grey level emphasis (GLSZM)	Clusters of voxels with low value
Size zone nonuniformity normalized	Small area emphasis (GLSZM)	Variability of cluster sizes
GLDM (2/14)		
Large dependence high grey level emphasis		Large clusters of voxels with high values
Small dependence low grey level emphasis		Small clusters of voxels with low values
NGTDM (0/5)		