Appendix 1

SVM and logistic Regression

Support vector machine (SVM) usually proved the best performance when comparing with other popular classification problem in the real-world applications. The outstanding performance of SVM is due to its advantages of regularization and convex optimization (45-48). The kernel function of the support vector machine can find a hyperplane with an N-dimensional space (N is the number of features). In this feature space, the datapoint can be optimally distinct. In SVM, different kernel functions are applied to transform the original data into specific feature space to select support vectors.

The generated hyperplanes provided the decision boundaries which can optimize the classification of the data points which can be distinguished into different classes when they fall on the side of the hyperplane. The data points which closer to the hyperplanes work as the support vectors which can influence the position and orientation of the hyperplane. Based on these support vectors, the margin of the classifier can be maximized to get the best classification performance. Due to the utilization of the hyperplane, the classification performance is relatively better than other methods (48). Also, this strategy can overcome the overfitting issue during training. But due to the complicated settings, the required training dataset needs to be larger compared to using other methods.

Unlike SVM, Linear and logistic regression are popular due to the simple implementation (35). We can estimate a linear model by searching the parameters to fit a model of the straight line in the original data space. Then applying the logistic function to the linear mode, logistic regression model can be used to differentiate binomial distributions. The strategy of logistic function is very simple. The output of the linear model is applied to sigmoid function. All values are nonlinear rescaled to the range between 0 and 1. Logistic regression is one of the simplest methods in ML. With very few inputs, a relatively general model can be established.

Table S1 The model information of MR scanners

Model	Manufacturer	Address	Field strength
Magnetom Sonata	Siemens Healthcare	Erlangen, Germany	1.5T
Optima MR360	GE Medical Systems	Wisconsin, USA	1.5T
Magnetom Trio	Siemens Healthcare	Erlangen, Germany	3.0T
Signa HDx	GE Medical Systems	Wisconsin, USA	3.0T
Discovery MR750	GE Medical Systems	Wisconsin, USA	3.0T
Discovery MR750w	GE Healthcare Japan Corporation	Tokyo, Japan	3.0T
uMR 770	United Imaging Healthcare	Shanghai, China	3.0T

 https://dx.doi.org/10.21037/qims-22-267

Table S2 MRI-Sequence Parameters of the Imaging Protocol

	FOV (cm^2)	Slice thickness (mm)	Slice gap (mm)	TR/TE (ms)
Cervical vertebra				
SAG T2 FRFSE	28×28	3.0	0.5	2700/120
SAG T1 FSE	28×28	3.0	0.5	710/8.0
SAG T2 IDEAL	28×28	3.0	0.5	2500/85
Thoracic vertebra				
SAG T2 FSE	36×36	3.0	0.5	2700/120
SAG T1 FSE	36×36	3.0	0.5	700/9.0
SAG T2 IDEAL	36×36	3.0	0.5	2500/85
Lumbar and sacral vertebra				
SAG T2 FSE	30×30	4.0	0.5	3100/120
SAG T1 FSE	30×30	4.0	0.5	700/10
SAG T2 FS	30×30	4.0	0.5	3300/85

MRI, magnetic resonance imaging; FOV, field of view; TR, rime to repeat; TE, time to echo; SAG, sagittal; FRFSE, fast relaxation fast spin echo; FSE, fast spin echo; IDEAL, iterative decomposition of water and fat with echo asymmetry and least-squares estimation; FS, fat suppression.

Table S3 The model information of CT scanners

Model	Manufacturer	Address
LightSpeed VCT	GE Medical System	Chalfont St Giles, UK
Discovery CT750	GE Medical System	Wisconsin, USA
Sensation	Siemens Healthcare	Erlangen, Germany
SOMATOM Definition Flash	Siemens Healthcare	Erlangen, Germany
uCT790	United Imaging Healthcare	Shanghai, China

https://dx.doi.org/10.21037/qims-22-267

Table S4 List of 107 radiomics features

1st Order	Shape	GLCM	GLSZM	GLRLM	GLDM	NGTDM
(N=18)	(N=14)	(N=24)	(N=16)	(N=16)	(N=14)	(N=5)
10th percentile	Elongation	Autocorrelation	GLN	GLN	DE	busyness
90th percentile	Flatness	Cluster Prominence	GLNN	GLNN	DN	coarseness
Energy	Least Axis Length	Cluster Shade	GLV	GLV	DNN	complexity
Entropy	Major Axis Length	Cluster Tendency	HGLZE	HGLRE	DV	contrast
Interquartile Range	Max 2D diameter (Column)	Contrast	LAE	LGLRE	GLN	strength
Kurtosis	Max 2D diameter (Row)	Correlation	LAHGLE	LRE	GLV	
MAD	Max 2D diameter (Slice)	Difference Average	LALGLE	LRHGLE	HGLE	
Maximum	Max 3D diameter	Difference Entropy	LGLZE	LRLGLE	LDE	
Mean	Mesh Volume	Difference Variance	SAE	RE	LDHGLE	
Median	Minor Axis Length	ID	SAHGLE	RLN	LDLGLE	
Minimum	Sphericity	IDM	SALGLE	RLNN	LGLE	
Range	Surface Area	IDMN	SZN	RP	SDE	
rMAD	Surface Area/Volume ratio	IDN	SZNN	RV	SDHGLE	
RMS	Voxel Volume	IMC1	ZE	SRE	SDLGLE	
Skewness		IMC2	ZP	SRHGLE		
Std		Inverse Variance	ZV	SRLGLE		
Uniformity		Joint Average				
Variance		Joint Energy				
		Joint Entropy				
		Max Probability				
		MCC				
		Sum Average				
		Sum Entropy				
		Sum of Squares				

GLCM, gray-level co-occurrence matrix; GLSZM, gray-level size zone matrix; GLRLM, gray-level run length matrix; GLDM, gray level dependence matrix; NGTDM, neighboring gray tone difference matrix.

References

45. Byvatov E, Fechner U, Sadowski J, Schneider G. Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. J Chem Inf Comput Sci 2003;43:1882-9.
46. Tong S, Chang E. Support vector machine active learning for image retrieval. ACM 2001: Proceedings of the ninth ACM international conference on Multimedia; 2001 Sept 30-Oct 5; Ottawa Canada. New York: Association for Computing Machinery; 2001:107-18.
47. Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273-97.
48. Drucker H, Burges CJC, Kaufman L, Smola A, Vapnik V. Support vector regression machines. Advances in Neural Information Processing Systems 9; 1996 Dec 2-5; Denver, USA: Cambridge: MIT Press; 1996:155-61.

