Appendix 1

Radiomics feature extraction

In this study, 18 first order features, 22 gray-level co-occurrence matrix (GLCM) features, 14 neighboring gray-level dependence matrix (NGLDM), 16 gray-level run length matrix (GLRLM) features, 16 gray-level size zone matrix (GLSZM) features and 14 shape/size features were adopted as the radiomic features (45). A three-dimensional (3D) Coiflet wavelet transform was applied to the DWI images in order to extract the first order features in frequency decomposed images. The frequency components were HHH, HHL, HLH, HLL, LHH, LHL, LLH, and LLL, where "H" and "L" denote high-pass and low-pass filters, respectively (34). To characterize the textural changes on DWI images over different diffusion gradient (different b values), we measured 8 new sequential features from the 21 b values for each texture feature, including mean, max, min, median, variance, kurtosis, skewness, energy. Therefore, a total of 7076 features were extracted on primary dataset for each tumor with 21 b values and 6,100 features for external testing dataset for each tumor with 17 b values. It should be noticed that the number of radiomics features were different between two datasets. This was due to the different number of b values. However, this problem has been solved during the feature selection procedure by choosing the features from the DWI images of which b values were equal between two datasets.

Appendix 2

Feature Selection procedure

All work in this part was accomplished using an open ML library scikit-learn (ver. 0.22), in Python (32). The whole dataset was split into training set (80 percent) and testing set (20 percent). The external test set included 55 cases on five-fold cross-validation sets. Radiomics features with b values that were not included in the external test were excluded. A five-step rigorous selection process has been implemented both on combined DWI-model features and radiomics features:

Step I WMW U-test

All features of the training data were tested by a non-parametric WMW U-test with a significant setting of $P<0.05$.

Step II ML methods

On one way, a learning model-based single feature sequencing approach was involved. The idea of this approach was to use Logistic Regression (LR), Support-vector Machine (SVM), K-nearest neighbors (KNN), Random Forests (RF), Naïve Bayes (NB) and Stacking Methods (Stacks) separately as a learning estimator to build a predictive model for each individual feature filtered by step 1.

On the other way, the top features were selected according to scores derived from Lasso, RF and Recursive feature elimination (RFE) methods. Grid search was used on these estimators to define the hyper-parameters of Lasso and RF. Features ranking in Lasso were determined by the final coefficient, and in RF were sorted by their importance. Recursive feature elimination (RFE) model was also applied for selection. RFE creates a model from all features, and then eliminates the least important features in turn by measuring the contribution of each feature in a given model (26). In this study, RideCV was utilized as an underlying function to stabilize it.

In total, 9 ML based selection methods with 5-fold cross-validation performed were used, each providing the top 20 features of this work.

Step III Voting system

A voting system was proposed to find the common features selected by 9 methods mentioned above. We only reserved features with votes >9/2 and 19 features were left for multiparametric DWI based on ML and 14 features left for DWI radiomics.

 https://dx.doi.org/10.21037/qims-22-145

Step IV Correlation test

The final decision was made by calculating the Pearson correlation coefficient r and eliminate features with r>0.7 according to the rank. After this, 10 and 9 features were selected for multiparametric DWI and radiomics methods, respectively.

Step V Combination and Grouping

Features were combined in accordance with DWI models, where in this study 10 features belong to 5 different DWI models. Based on combination mathematics, there are $2^5-1=31$ types of DWI model combinations. The description of these combinations is simply combining their capitals, such as SC for SEM-model & CTRW-model, SFs for SEM-model & FROC-model & SM model and the like.

In this study, we also aimed to compare with two traditional classification methods. One of traditional method is based on a single DWI model only measuring the average for each parameter shown in Eq. [1] to Eq. [7]. And the other method depends on DWI radiomics features. Therefore, we also trained and tested our estimators on these conventional combinations, which were also grouped following rules stipulated above. The selection procedure for radiomics features followed the same rules except the step 5, we only chose the top 1 to 9 features according to the rank in step 4 on the training set instead. (See *Table S5*, which shows feature combinations of two traditional methods).

References

45. Beylkin G, Coifman R, Rokhlin V. Fast wavelet transforms and numerical algorithms I. Communications on pure and applied mathematics 1991;44:141-83.

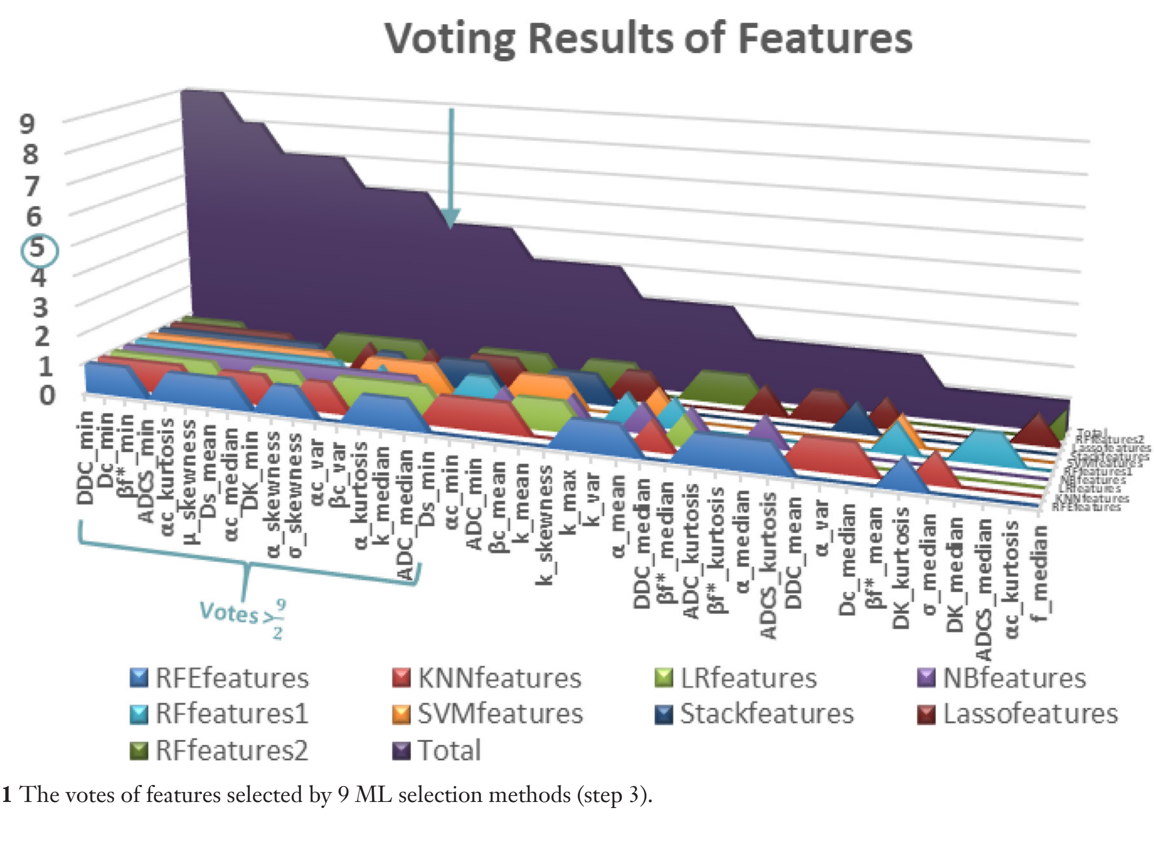

Figure S1 The votes of features selected by 9 ML selection methods (step 3).

https://dx.doi.org/10.21037/qims-22-145

Table S1 Correlation test results

	Pearson correlation matrix																	
	DDC_min	**Dc_min**	**βf*_min**	**αc_kurtosis**	**ADCS_min**	**μ_skewness**	**DK_min**	**α_skewness**	**σ_skewness**	**Ds_mean**	**αc_median**	**αc_var**	**k_median**	**Ds_min**	**αc_mean**	**αc_skewness**	**βc_var**	**α_kurtosis**	**ADC_median**
DDC_min	1	0.8810807	0.5478665	0.3947538	0.8414184	0.253429131	0.8414184	0.321240678	0.490974359	0.104029	0.4855554	0.6610031	0.587563	0.1599478	0.5824501	0.362759941	0.435051	0.182240221	0.536257366
Dc_min	0.8810807	1	0.6570947	0.3494493	0.9132883	0.297697103	0.9132883	0.343171208	0.474311275	0.3551752	0.38070135	0.5948554	0.5674203	0.175077	0.4618118	0.329001535	0.6096523	0.203332977	0.549041176
βf*_min	0.5478665	0.6570947	1	0.2486806	0.6338648	0.129205695	0.6338648	0.175528681	0.488938643	0.2721702	0.265729055	0.3497738	0.4577853	0.0395106	0.2915615	0.274598424	0.5820147	0.032292412	0.46968089
αc_kurtosis	0.3947538	0.3494493	0.2486806	1	0.3678287	0.207394817	0.3678287	0.551987294	0.474235638	0.0065027	0.794399189	0.67698	0.7874129	0.0754365	0.8276838	0.958688564	0.2806428	0.35463391	0.808941161
ADCS_min	0.8414184	0.9132883	0.6338648	0.3678287	1	0.226604814	1	0.292803324	0.455293802	0.3410058	0.375838049	0.6220602	0.5458609	0.1107667	0.463452	0.355409608	0.571527	0.152665357	0.535378527
μ_skewness	0.2534291	0.2976971	0.1292057	0.2073948	0.2266048	1	0.2266048	0.220910547	0.179502319	0.1274533	0.433178985	0.1532537	0.2613249	0.1605273	0.3676551	0.268078053	0.1973063	0.180574349	0.159125801
DK_min	0.8414184	0.9132883	0.6338648	0.3678287	1	0.226604814	1	0.292803324	0.455293802	0.3410058	0.375838049	0.6220602	0.5458609	0.1107667	0.463452	0.355409608	0.571527	0.152665357	0.535378527
α_skewness	0.3212407	0.3431712	0.1755287	0.5519873	0.2928033	0.220910547	0.2928033	1	0.252266711	0.0840269	0.489442806	0.4645366	0.5831461	0.0862979	0.520765	0.559983311	0.101712	0.855274669	0.584011623
σ_skewness	0.4909744	0.4743113	0.4889386	0.4742356	0.4552938	0.179502319	0.4552938	0.252266711	1	0.1555649	0.440414111	0.5304539	0.6002326	0.1253808	0.498797	0.489186612	0.3806164	0.061558816	0.622770263
Ds_mean	0.104029	0.3551752	0.2721702	0.0065027	0.3410058	0.127453316	0.3410058	0.084026859	0.155564925	1	0.126437327	0.2679803	0.1216034	0.0605586	0.0347546	0.00268296	0.7827822	0.061850281	0.101579967
αc_median	0.4855554	0.3807013	0.2657291	0.7943992	0.375838	0.433178985	0.375838	0.489442806	0.440414111	0.1264373	1	0.6523146	0.8254703	0.1773121	0.9639402	0.864013804	0.1804959	0.237974246	0.777779164
αc_var	0.6610031	0.5948554	0.3497738	0.67698	0.6220602	0.153253722	0.6220602	0.464536572	0.530453916	0.2679803	0.652314574	1	0.785901	0.0844488	0.8195331	0.65314769	0.5056715	0.313395889	0.741350071
k_median	0.587563	0.5674203	0.4577853	0.7874129	0.5458609	0.261324905	0.5458609	0.583146072	0.600232625	0.1216034	0.8254703	0.785901	1	0.0079255	0.8647078	0.822287263	0.5385137	0.314511353	0.963861831
Ds_min	0.1599478	0.175077	0.0395106	0.0754365	0.1107667	0.160527294	0.1107667	0.086297872	0.125380784	0.0605586	0.177312116	0.0844488	0.0079255	1	0.1193631	0.100624986	0.0931557	0.021338513	0.016813582
αc_mean	0.5824501	0.4618118	0.2915615	0.8276838	0.463452	0.367655063	0.463452	0.520765006	0.498797003	0.0347546	0.963940177	0.8195331	0.8647078	0.1193631	1	0.860038643	0.2709265	0.289400549	0.815240214
αc_skewness	0.3627599	0.3290015	0.2745984	0.9586886	0.3554096	0.268078053	0.3554096	0.559983311	0.489186612	0.002683	0.864013804	0.6531477	0.8222873	0.100625	0.8600386	1	0.2656724	0.327043339	0.82568054
βc_var	0.435051	0.6096523	0.5820147	0.2806428	0.571527	0.197306291	0.571527	0.101711981	0.380616355	0.6827822	0.180495858	0.5056715	0.5385137	0.0931557	0.2709265	0.265672407	1	0.045724646	0.518963128
α_kurtosis	0.1822402	0.203333	0.0322924	0.3546339	0.1526654	0.180574349	0.1526654	0.855274669	0.061558816	0.0618503	0.237974246	0.3133959	0.3145114	0.0213385	0.2894005	0.327043339	0.0457246	1	0.31758627
ADC_median	0.5362574	0.5490412	0.4696809	0.8089412	0.5353785	0.159125801	0.5353785	0.584011623	0.622770263	0.10158	0.777779164	0.7413501	0.9638618	0.0168136	0.8152402	0.82568054	0.5189631	0.31758627	1

Pink highlight data are highly colinear features (r>0.7) and blue highlight features are finally selected.

Table S2 Combinations of the subgroups of selected 10 features in multiparametric DWI model

Features selected by ML	DDC_min	β_f^*_min	α_c_kurtosis	μ_skewness	Ds_mean	α_skewness	σ_skewness	α_c_var	β_c_var	Ds_min
Subgroups (DWImodels)	**S(SEM)**	**C(CTRW)**	**F(FROC)**	**S(SM)**	**I(IVIM)**					
	DDC_min	α_c_kurtosis	β_f^*_min	σ_skewness	Ds_mean					
	α_skewness	α_c_var	μ_skewness		Ds_min					
		β_c_var								
Combinations of subgroups	s	S	F	I	C					
	Ss	Fs	Is	SF	SI	FI	Cs	SC	CF	CI
	SFs	SsI	FsI	SCs	CFs	CsI	SFI	SCF	SCI	CFI
	SFsI	SCFs	SCsI	CFsI	CFsI	SCFI				
	SCFsI									

Table S3 Integrated training and internal test results for each feature combination and they were sorted according to their AUCs on internal test set

Feature-combination	Feature numbers	Prediction methods	trainCV acc	trainCV auc	test acc	test auc	tpr	tnr	Cut-off
ADC	1	LRfold25	0.7939	0.783	0.6	0.6429	0.4286	1	0.872
DDC_min'	1	'model_KNNs'	0.7	0.7994	0.7333	0.7411	0.5714	0.875	0.5455
Ds min'	1	model KNNs'	0.8333	0.9195	0.6667	0.6875	0.8571	0.5	0.6667
βf* min'	1	model_SVMs'	0.7667	0.7638	0.6	0.5982	0.5714	0.625	—
βc var'	1	'model_Stacks'	0.6	0.727	0.6	0.625	0.5714	0.75	0.5601
α skewness'	1	'model RFs'	0.6333	0.8741	0.6	0.5	0.4286	0.75	0.5847
αc var'	1	'model KNNs'	0.6333	0.823	0.5333	0.5982	0.4286	0.875	0.7143
s	1	'model_RFs'	0.6667	0.9782	0.4667	0.6161	0.7143	0.5	0.4431
μ skewness'	1	'model RFs'	0.7	1	0.4667	0.5357	0.5714	0.75	0.74
αc kurtosis'	1	'model LRCVs'	0.6667	0.7425	0.4667	0.5179	0.8571	0.375	0.4511
Ds_aver'	1	'model_RFs'	0.8	0.9253	0.4667	0.5893	0.5714	0.75	0.721
F	2	LRfold5'	0.8382	0.8184	0.6667	0.7679	1	0.5	0.442
SM	2	'LRfold12'	0.8106	0.7921	0.5333	0.6071	0.4286	1	0.8665
S	2	'LRfold24'	0.807	0.8748	0.6667	0.6786	0.8571	0.625	0.4924
SEM	2	'RFfold20'	0.7909	0.7857	0.7333	0.7143	0.7143	0.75	0.5583
DKI	2	'Stackfold12'	0.8376	0.8341	0.6667	0.6071	0.7143	0.75	0.3458
I	2	'KNNfold8'	0.6845	0.731	0.5333	0.5536	0.2857	1	1
Ss	3	'KNNfold24'	0.8376	0.8743	0.6	0.6696	0.4286	0.875	0.7143
FROC	3	'LRfold2'	0.8112	0.8079	0.6667	0.6429	0.4286	0.875	0.5855
IVIM	3	'Stackfold18'	0.7764	0.735	0.4667	0.5357	0.4286	0.875	0.5655
CTRW	3	'Stackfold15'	0.8042	0.7838	0.5333	0.5357	0.8571	0.375	0.5162
sI	3	'RFfold3'	0.7273	0.8385	0.4	0.5268	0.8571	0.375	0.506
Fs	3	'KNNfold3'	0.787	0.8519	0.6667	0.5179	0.5714	0.75	0.5556
C	3	NIfold5'	0.8182	0.7864	0.5333	0.5179	0.5714	0.625	0.4581
SI	4	'RFfold24'	0.8182	0.8635	0.8667	0.8214	0.7143	0.7143	0.7983
SF	4	'KNNfold24'	0.8312	0.9078	0.7333	0.6518	0.7143	0.7143	0.5294
FI	4	'Stackfold5'	0.7273	0.835	0.5333	0.5714	0.4286	0.4286	0.7126
Cs	4	'KNNfold9'	0.7761	0.8419	0.5333	0.5268	1	1	0.1667
SC	5	'Stackfold1'	0.8921	0.8863	0.7333	0.7857	0.7143	0.875	0.54
CF	5	'KNNfold24'	0.757	0.8509	0.5333	0.6696	1	0.375	0.25
SFs	5	'KNNfold3'	0.8585	0.9007	0.6667	0.625	0.5714	0.75	0.6429
FsI	5	'RFfold24'	0.9091	0.8587	0.5333	0.625	0.8571	0.5	0.4442
CI	5	'RFfold19'	0.6364	0.7798	0.6	0.6161	0.7143	0.625	0.3205
SsI	5	'RFfold12'	0.8182	0.8668	0.6	0.6161	0.7143	0.625	0.4288
SCs	6	'Stackfold4'	0.8955	0.8939	0.6667	0.7679	0.7143	0.875	0.5243
SFI	6	'KNNfold24'	0.8273	0.9013	0.6667	0.6518	0.7143	0.625	0.6
CFs	6	'RFfold3'	0.8182	0.858	0.7333	0.5893	0.5714	0.875	0.5739
CsI	6	'Stackfold4'	0.7273	0.8391	0.5333	0.5714	0.7143	0.625	0.4629
SCF	7	'Stackfold4'	0.9091	0.8898	0.8	0.8393	0.8571	0.75	0.5051
CFI	7	'RFfold13'	0.8182	0.8638	0.6667	0.6964	0.5714	0.875	0.5547
SCI	7	'LRfold24'	0.7976	0.8578	0.6667	0.6786	0.7143	0.75	0.491
SFsI	7	'KNNfold3'	0.8718	0.9037	0.6667	0.5982	0.7143	0.625	0.4667
SCFs	8	'Stackfold7'	0.9091	0.9021	0.7333	0.7679	0.5714	0.875	0.5308
CFsI	8	'RFfold3'	0.9091	0.8731	0.6	0.6429	0.7143	0.625	0.3624
SCsI	8	'Stackfold1'	0.9091	0.8815	0.6	0.6429	0.5714	0.75	0.5474
SCFI	9	'LRfold24'	0.8139	0.8727	0.6667	0.6607	0.8571	0.625	0.1751
All features	10	'RFfold24'	0.8755	0.8873	0.6667	0.7321	0.7143	0.75	0.5698

"tpr" refers to the sensitivity and "tnr" refers to the specificity. The highlighted row represents the best prediction model which achieves the highest AUC in internal test set.

 https://dx.doi.org/10.21037/qims-22-145

Table S4 R-squared of each DWI model to assess the goodness of fitting

DWI model	Mean
ADC	0.9449
IVIM	0.9569
SEM	0.9959
SM	0.9699
DKI	0.9699
FROC	0.9788
CTRW	0.9801

Table S5 Row 'Feature combinations' gives the selected features based on single DWI model and radiomics respectively

DWI models	Traditional methods based on the single DWI model		Traditional DWI radiomics		Feature combinations	Serial numbers of selected features
	Abbreviation	Features	Feature types	Features/serial number		
ADC	Amean	ADC_mean	b =3,500	HHL_Maximum/ ② ; HHL_kurtosis/ ⑧	TOP1	①
IVIM	Imean	f_mean, D_S_mean, D_f_mean	b =4,000	HHH_Kurtosis/ ⑥	TOP2	①②
SEM	Smean	DDC_mean, α_mean	b =0	HHL_Kurtosis/ ④	TOP3	①②③
SM	smean	ADC_S_mean, σ_mean	Sequential	Kurtosis_original_shape_Minor_Axis_Length/ ① ; Skewness_original_shape_Minor_AxisLength/ ③	TOP4	①②③④
DKI	Dmean	D_K_mean, K_mean		Kurtosis_HHH_Inter-quartileRange/ ⑤ ; Skewness_origianl_glszm_LargeAraHigh-GrayLevelEmphasis/ ⑦	TOP5	①②③④⑤
					TOP6	①②③④⑤⑥
FROC	Fmean	D_f_mean, β^*_c_mean, μ_mean		Skewness_original_shape_Maximum2DDiameterColumn/ ⑨	TOP7	①②③④⑤⑥⑦
CTRW	Cmean	D_C_mean, α_c_mean, β_c_aver			TOP8	①②③④⑤⑥⑦⑧
					TOP9	①②③④⑤⑥⑦⑧⑨

The 'serial number ① - ⑨ ' was determined by the ranking after step 4.

https://dx.doi.org/10.21037/qims-22-145

Table S6 Internal test results of two traditional DWI methods

Combination name	Feature-num	Prediction estimator	trainCV_acc	trainCV_auc	test_acc	test_auc	best_tpr	best_tnr	Cut-off
Single DWI model									
ADC	1	LRfold25	0.7939	0.783	0.6	0.6429	0.4286	1	0.872
SM	2	LRfold12	0.8106	0.7921	0.5333	0.6071	0.4286	1	0.8665
SEM	2	RFfold20	0.7909	0.7857	0.7333	**0.7143**	0.7143	0.75	0.5583
DKI	2	Stackfold12	0.8376	0.8341	0.6667	0.6071	0.7143	0.75	0.3458
CTRW	3	Stackfold15	0.8042	0.7838	0.5333	0.5357	0.8571	0.375	0.5162
FROC	3	LRfold2	0.8112	0.8079	0.6667	0.6429	0.4286	0.875	0.5855
IVIM	3	Stackfold18	0.7764	0.735	0.4667	0.5357	0.4286	0.875	0.5655
Radiomics									
TOP6	6	Stackfold3	0.9558	0.9821	0.8	**0.8393**	0.8571	0.875	0.6111
TOP4	4	LRfold3	0.9421	0.944	0.6667	0.75	1	0.625	0.3141
TOP7	7	Stackfold3	0.9624	0.9551	0.7333	0.75	0.8571	0.625	0.429
TOP5	5	LRfold4	0.9321	0.9399	0.7333	0.75	0.8571	0.75	0.3827
TOP8	8	Stackfold2	0.9761	0.9821	0.8	0.75	0.5714	1	0.595
TOP9	9	Stackfold1	0.9794	0.9868	0.7333	0.6964	0.5714	0.875	0.6148
TOP3	3	LRfold1	0.9082	0.9044	0.6667	0.6607	1	0.375	0.3405
TOP2	2	LRfold10	0.8821	0.8801	0.5333	0.5179	0.4286	0.875	0.7174

 https://dx.doi.org/10.21037/qims-22-145