

Table S1 Clinical and pathological characteristics of the patients in dataset A

Characteristics	Training set (n=162)			Validation set (n=45)		
	Benign (n=77)	Malignant (n=85)	P value	Benign (n=15)	Malignant (n=30)	P value
Age (years)	57.4±9.8	60.3±12.6	0.746	58.2±10.6	61.6±14.0	0.633
Gender			0.281			0.831
Male	41 (19.8)	37 (17.87)		9 (4.35)	17 (8.21)	
Female	36 (12.56)	48 (23.18)		6 (2.90)	13 (6.28)	
Location			0.311			0.831
Right kidney	41 (19.80)	53 (25.60)		6 (2.90)	13 (6.28)	
Left kidney	36 (17.39)	32 (15.46)		9 (4.35)	17 (8.21)	
Histological subtype			<0.0001			<0.0001
Simple kidney cyst	77 (37.19)	0 (0.00)		14 (6.76)	0 (0.00)	
Clear cell carcinoma	0 (0.00)	84 (40.5)		0 (0.00)	30 (14.50)	
Other	0 (0.00)	1 (0.4)		1 (0.48)	0 (0.00)	
Bosniak category			<0.0001			<0.0001
IIF	63 (30.43)	16 (7.73)		12 (5.80)	3 (1.45)	
III	14 (6.76)	21 (10.14)		3 (1.45)	10 (4.83)	
IV	0 (0.00)	48 (23.18)		0 (0.00)	17 (8.21)	

Data are expressed as the mean ± SD or frequency (constituent ratio). SD, standard deviation.

Table S2 Clinical and pathological characteristics of the patients in dataset B

Characteristics	Training set (n=65)			Validation set (n=27)		
	CRCC (n=36)	RO (n=29)	P value	CRCC (n=15)	RO (n=12)	P value
Age (years)	56.4±15.7	51.4±13.63	0.10	51.4±13.63	54.5±8.96	0.69
Gender			0.39			0.93
Male	16 (17.58)	11 (12.09)		6 (6.60)	5 (5.50)	
Female	20 (21.98)	18 (19.78)		9 (9.89)	7 (7.69)	
Location			0.62			0.42
Right kidney	17 (18.68)	16 (17.58)		8 (8.79)	5 (5.50)	
Left kidney	19 (20.88)	13 (14.29)		7 (7.69)	7 (7.69)	

Data are expressed as mean ± SD, median (interquartile range), or frequency (constituent ratio). CRCC, chromophobe renal cell carcinoma; RO, renal oncocytoma; SD, standard deviation.

Table S3 Radiomic features used for analysis of datasets A and B

Dataset A radiomic features

Unenhanced of dataset A (4 features)

Original_glcmm_MaximumProbability_Unenhanced
Wavelet-LHH_firstorder_Median_Unenhanced
Wavelet-LLL_firstorder_90Percentile_Unenhanced
Wavelet-LLL_firstorder_Median_Unenhanced

CMP of dataset A (2 features)

Wavelet-LLL_firstorder_Median
Wavelet-LLL_firstorder_Uniformity

NP of dataset A (5 features)

Original_firstorder_Median
Wavelet-LLL_gldm_DependenceEntropy
Original_firstorder_Mean
Original_glcmm_MaximumProbability
Original_gldm_DependenceEntropy

Dataset B radiomic features

CMP of dataset B (9 features)

Log-sigma-1-mm-3D_firstorder_Skewness_CMP
Log-sigma-1-mm-3D_glcmm_Imc2_CMP
Log-sigma-1-mm-3D_gldm_RunLengthNonUniformityNormalized_CMP
Log-sigma-2-mm-3D_glcmm_ClusterShade_CMP
Log-sigma-3-mm-3D_glszm_SmallAreaEmphasis_CMP
Log-sigma-3-mm-3D_glszm_SmallAreaLowGrayLevelEmphasis_CMP
Wavelet-HLL_gldm_DependenceEntropy_CMP
Wavelet-HLL_gldm_LargeDependenceEmphasis_CMP
Wavelet-LLL_firstorder_RobustMeanAbsoluteDeviation_CMP

NP of dataset B (3 features)

Log-sigma-3-mm-3D_glszm_SmallAreaEmphasis_NP
Wavelet-HLL_gldm_DependenceVariance_NP
Wavelet-HLH_glszm_SmallAreaHighGrayLevelEmphasis_NP

Fusion of dataset B (12 features)

12 rad features from the CMP of dataset B (9 features) and NP of dataset B (3 features)

CMP, corticomedullary phase; NP, nephrographic phase.

Table S4 P values and z scores from the DeLong test for pair-wise comparisons of different ML models in the validation set of dataset A

Unenhanced				CMP				NP			
Algorithm A	Algorithm B	z score	P value	Algorithm A	Algorithm B	z score	P value	Algorithm A	Algorithm B	z score	P value
SVM	SGD	1.31	0.19	SVM	SGD	-1.02	0.31	SVM	SGD	3.09*	<0.001*
SVM	KNN	0.71	0.48	SVM	KNN	1.42	0.16	SVM	KNN	0.75	0.45
SVM	RF	-0.27	0.78	SVM	RF	-0.07	0.94	SVM	RF	-0.22	0.83
SVM	XGBoost	1.28	0.20	SVM	XGBoost	0.82	0.41	SVM	XGBoost	1.68	0.09
SVM	LightGBM	0.86	0.39	SVM	LightGBM	0.49	0.62	SVM	LightGBM	-0.13	0.89
SVM	TabPFN	0.33	0.74	SVM	TabPFN	0.84	0.40	SVM	TabPFN	-0.70	0.48
SGD	KNN	-0.85	0.40	SGD	KNN	2.26*	0.02*	SGD	KNN	-2.90*	<0.001*
SGD	RF	-1.82	0.07	SGD	RF	1.69	0.09	SGD	RF	-3.25*	<0.001*
SGD	XGBoost	-0.31	0.76	SGD	XGBoost	2.00	0.05	SGD	XGBoost	-2.26*	0.02*
SGD	LightGBM	-0.89	0.37	SGD	LightGBM	2.13*	0.03*	SGD	LightGBM	-3.00*	<0.001*
SGD	TabPFN	-1.02	0.31	SGD	TabPFN	1.37	0.17	SGD	TabPFN	-3.35*	<0.001*
KNN	RF	-1.11	0.27	KNN	RF	-1.82	0.07	KNN	RF	-0.97	0.33
KNN	XGBoost	0.76	0.44	KNN	XGBoost	-1.43	0.15	KNN	XGBoost	1.46	0.14
KNN	LightGBM	0.25	0.80	KNN	LightGBM	-1.73	0.08	KNN	LightGBM	-0.72	0.47
KNN	TabPFN	-0.03	0.98	KNN	TabPFN	-1.04	0.30	KNN	TabPFN	-1.17	0.24
RF	XGBoost	1.74	0.08	RF	XGBoost	1.27	0.20	RF	XGBoost	1.85	0.06
RF	LightGBM	1.28	0.20	RF	LightGBM	1.25	0.21	RF	LightGBM	0.15	0.88
RF	TabPFN	0.59	0.56	RF	TabPFN	0.78	0.44	RF	TabPFN	-0.45	0.65
XGBoost	LightGBM	-0.71	0.47	XGBoost	LightGBM	-1.28	0.20	XGBoost	LightGBM	-1.55	0.12
XGBoost	TabPFN	-0.66	0.51	XGBoost	TabPFN	-0.21	0.83	XGBoost	TabPFN	-1.94	0.05
LightGBM	TabPFN	-0.25	0.80	LightGBM	TabPFN	0.21	0.83	LightGBM	TabPFN	-0.65	0.51

*, P<0.05. z score presented as positive when algorithm A was better than algorithm B, and z score presented as negative when algorithm A was worse than algorithm B. CMP, corticomedullary phase; KNN, k-nearest neighbor; LightGBM, light gradient boosting machine; ML, machine learning; NP, nephrographic phase; RF, random forest; SGD, stochastic gradient descent; SVM, support vector machine; TabPFN, tabular prior-data fitted network; XGBoost, extreme gradient boosting.

Table S5 P values and z scores from the DeLong test for pair-wise comparisons of different ML models in the validation set of dataset B

CMP				NP				Fusion			
Algorithm A	Algorithm B	z score	P value	Algorithm A	Algorithm B	z score	P value	Algorithm A	Algorithm B	z score	P value
SVM	SGD	0.79	0.43	SVM	SGD	0.17	0.86	SVM	SGD	1.65	0.10
SVM	KNN	-0.39	0.70	SVM	KNN	-0.12	0.90	SVM	KNN	-0.10	0.92
SVM	RF	-1.25	0.21	SVM	RF	1.16	0.24	SVM	RF	-1.03	0.30
SVM	XGBoost	-0.49	0.63	SVM	XGBoost	1.83	0.07	SVM	XGBoost	-0.44	0.66
SVM	LightGBM	-0.50	0.62	SVM	LightGBM	1.60	0.11	SVM	LightGBM	0.14	0.89
SVM	TabPFN	-1.29	0.20	SVM	TabPFN	0.92	0.36	SVM	TabPFN	-0.98	0.33
SGD	KNN	-1.26	0.21	SGD	KNN	-0.30	0.76	SGD	KNN	-1.89	0.06
SGD	RF	-3.38*	<0.001*	SGD	RF	1.18	0.24	SGD	RF	-3.73*	<0.001*
SGD	XGBoost	-1.82	0.07	SGD	XGBoost	1.78	0.07	SGD	XGBoost	-2.50*	0.01*
SGD	LightGBM	-1.57	0.12	SGD	LightGBM	1.50	0.13	SGD	LightGBM	-1.61	0.11
SGD	TabPFN	-3.49*	<0.001*	SGD	TabPFN	1.05	0.29	SGD	TabPFN	-3.06*	<0.001*
KNN	RF	-1.06	0.29	KNN	RF	1.18	0.24	KNN	RF	-1.01	0.31
KNN	XGBoost	-0.33	0.74	KNN	XGBoost	2.67*	0.01*	KNN	XGBoost	-0.37	0.71
KNN	LightGBM	-0.26	0.79	KNN	LightGBM	2.33*	0.02*	KNN	LightGBM	0.22	0.83
KNN	TabPFN	-1.12	0.26	KNN	TabPFN	1.03	0.30	KNN	TabPFN	-0.80	0.42
RF	XGBoost	1.42	0.15	RF	XGBoost	1.11	0.27	RF	XGBoost	1.20	0.23
RF	LightGBM	1.33	0.18	RF	LightGBM	1.16	0.24	RF	LightGBM	2.16*	0.03*
RF	TabPFN	-0.11	0.91	RF	TabPFN	-0.38	0.70	RF	TabPFN	0.52	0.60
XGBoost	LightGBM	0.11	0.91	XGBoost	LightGBM	0.38	0.70	XGBoost	LightGBM	1.12	0.26
XGBoost	TabPFN	-1.40	0.16	XGBoost	TabPFN	-1.17	0.24	XGBoost	TabPFN	-0.64	0.52
LightGBM	TabPFN	-1.46	0.14	LightGBM	TabPFN	-1.21	0.23	LightGBM	TabPFN	-1.54	0.12

*, $P < 0.05$. z score presented as positive when algorithm A was better than algorithm B, and z score presented as negative when algorithm A was worse than algorithm B. CMP, corticomedullary phase; KNN, k-nearest neighbor; LightGBM, light gradient boosting machine; ML, machine learning; NP, nephrographic phase; RF, random forest; SGD, stochastic gradient descent; SVM, support vector machine; TabPFN, tabular prior-data fitted network; XGBoost, extreme gradient boosting.

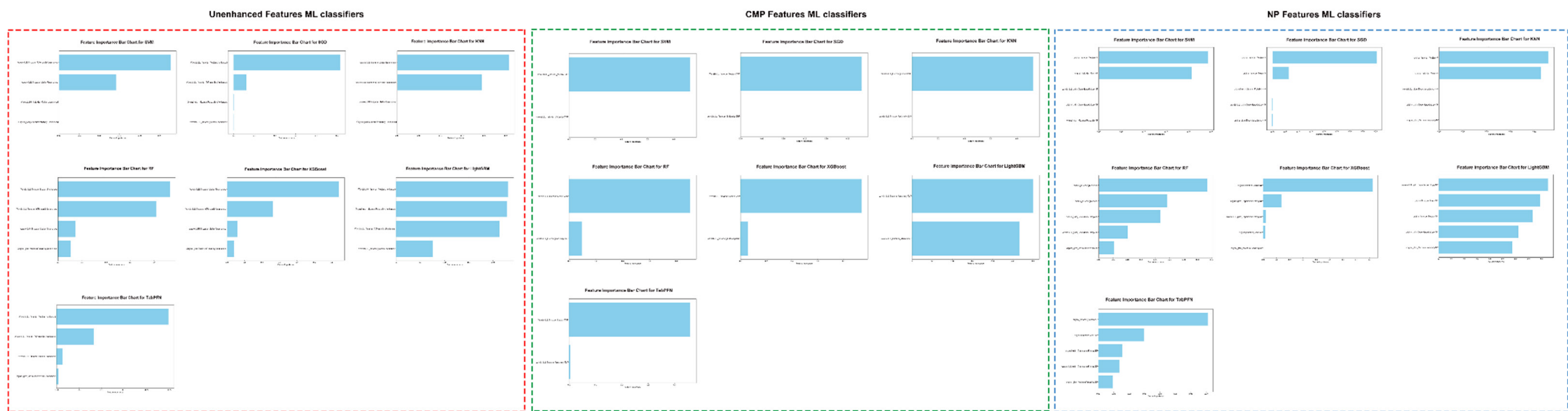


Figure S1 Feature importance bar chart of ML classifiers in unenhanced, CMP, and NP groups of dataset A. CMP, corticomedullary phase; ML, machine learning; NP, nephrographic phase.



Figure S2 Feature importance bar chart of ML classifiers in unenhanced, CMP, and NP groups of dataset B. CMP, corticomedullary phase; ML, machine learning; NP, nephrographic phase.