Supplementary Material

Title: CT-based intratumoral heterogeneity quantification fusing deep learning radiomics for predicting lymph node metastasis in early-stage lung adenocarcinoma: a multicenter study

Author: Min Hu[1,2], Xingyu Hu[2], Wenxiang Chen [2], Chengcheng Xia[2], Weirong Yao[3], Minjing Zuo[4,5], Ze Lin[4,5], Libin Deng [2,*] & Zhaoqiu Wu[1,*]

[1]State Key Laboratory of Natural Medicines, Department of Pharmacology, School of Pharmacy, China Pharmaceutical University, Nanjing, Jiangsu, China
[2]Jiangxi Provincial Key Laboratory of Preventive Medicine, School of Public Health, Jiangxi Medical College, Nanchang University, Nanchang 330006, Jiangxi, China
[3]Department of Oncology, Jiangxi Provincial People's Hospital, The First Affiliated Hospital of Nanchang Medical College, Nanchang 330006, Jiangxi, China
[4]Department of Radiology, The Second Affiliated Hospital, Jiangxi Medical College, Nanchang University, Nanchang 330006, Jiangxi, China
[5]Intelligent Medical Imaging of Jiangxi Key Laboratory, Nanchang 330006, Jiangxi, China
*Correspondence: Libin Deng: lbdeng@ncu.edu.cn; Zhaoqiu Wu: zqwucpu@163.com
Min Hu, Xingyu Hu and Wenxiang Chen contributed equally to this work and share first authorship.

CT Acquisition Parameters

CT scanning equipment used at center 1 included GE MEDICAL SYSTEMS Revolution CT, GE MEDICAL SYSTEMS Optima CT660, Philips iCT 256, Philips Spectral CT, and Siemens SOMATOM Definition. At center 2, the equipment comprised Philips iCT 256, Philips Brilliance 16, Philips Spectral CT, SIEMENS SOMATOM Definition Flash, GE MEDICAL SYSTEMS Revolution CT, and GE MEDICAL SYSTEMS Revolution HD. Patients were positioned supine, and scanning commenced at the end of inhalation, covering the region from the thoracic inlet (apex of the lungs) to the base of the lungs. Image acquisition parameters included a tube voltage of 120 kV, automatic tube current modulation, an acquisition matrix of 512×512, slice thickness ranging from 0.625 to 1.25 mm, and a display field of view of 350 mm.

Intratumoral Subregions Segmentation

To objectively extract subregions within the tumor, simple linear iterative clustering (SLIC) was applied to each layer of the nodule's CT images, with an average subregion size set to 8 and a smoothing factor of 20. Recognizing that radiomics feature extraction may be ineffective in subregions with limited voxel counts, a minimum volume threshold of 3×3 was established to filter out these small subregions.

Table S1 Optimal Hyperparameters of XGBoost Models

model	best_parameters
Radiomics	n_estimators=17, max_depth=3, gamma=2.4, min_child_weight=21, max_delta_step=1, subsample=0.8, colsample_bytree=0.8, learning_rate=0.3, objective='binary:logistic', random_state=2024, scale_pos_weight=4
ITH	n_estimators=13, max_depth=5, gamma=1.8, min_child_weight=3, subsample=1, colsample_bytree=1, learning_rate=0.03, objective='binary:logistic', random_state=2024, scale_pos_weight=4
Deep_R	n_estimators=100, max_depth=4, gamma=2.4, min_child_weight=23, max_delta_step=1, subsample=0.8, colsample_bytree=0.8, learning_rate=0.03, objective='binary:logistic', random_state=2024, scale_pos_weight=4
Radscore-ITHscore	n_estimators=17, max_depth=4, gamma=2.2, min_child_weight=25, max_delta_step=1, subsample=1, colsample_bytree=0.8, learning_rate=0.02, objective='binary:logistic', random_state=2024, scale_pos_weight=4
Deep-Radscore	n_estimators=100, max_depth=3, gamma=0, min_child_weight=10, subsample=0.9, colsample_bytree=0.8, learning_rate=0.01, reg_alpha= 1, objective='binary:logistic', random_state=2024
Deep-ITHscore	n_estimators=100, max_depth=5, gamma=0.2, min_child_weight=5, subsample=0.8, colsample_bytree=0.8, learning_rate=0.01, reg_alpha= 1, reg_lambda=0.1, objective='binary:logistic', random_state=2024, scale_pos_weight=4,
Fusion	n_estimators=125, max_depth=3, gamma=7, min_child_weight=21, max_delta_step=1, subsample=0.8, colsample_bytree=0.9, learning_rate=0.03, objective='binary:logistic', random_state=2024, scale_pos_weight=4

Table S2: The Diagnostic Performance of Each Algorithm

Algorithm	Dataset	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
SVM	Dataset A	0.909 (0.859, 0.955)	0.892	0.855	0.900	0.628	0.969
	Dataset B	0.712 (0.574, 0.844)	0.651	0.833	0.614	0.306	0.947
	Dataset C	0.803 (0.724, 0.863)	0.685	0.882	0.645	0.337	0.964
	Dataset D	0.613 (0.436, 0.796)	0.365	1.000	0.259	0.184	1.000
DT	Dataset A	0.893 (0.863, 0.918)	0.768	0.928	0.736	0.410	0.981
	Dataset B	0.802 (0.684, 0.902)	0.708	0.833	0.682	0.349	0.952
	Dataset C	0.827 (0.753, 0.898)	0.835	0.824	0.837	0.509	0.959
	Dataset D	0.669 (0.477, 0.831)	0.714	0.556	0.741	0.263	0.909
RF	Dataset A	0.920 (0.886, 0.946)	0.873	0.826	0.883	0.582	0.962
	Dataset B	0.764 (0.660, 0.864)	0.679	0.833	0.648	0.326	0.950
	Dataset C	0.819 (0.741, 0.891)	0.745	0.853	0.723	0.387	0.960
	Dataset D	0.682 (0.503, 0.848)	0.635	0.778	0.611	0.250	0.943
KNN	Dataset A	0.901(0.870, 0.929)	0.806	0.841	0.799	0.453	0.962
	Dataset B	0.773 (0.651, 0.874)	0.594	0.889	0.534	0.281	0.959
	Dataset C	0.833 (0.766, 0.899)	0.750	0.882	0.723	0.395	0.968
	Dataset D	0.693 (0.526, 0.858)	0.524	0.889	0.463	0.216	0.962
XGBoost	Dataset A	0.906 (0.869, 0.940)	0.816	0.899	0.799	0.470	0.976
	Dataset B	0.812 (0.714, 0.906)	0.717	0.889	0.682	0.364	0.968
	Dataset C	0.803 (0.702, 0.895)	0.845	0.794	0.855	0.529	0.953
	Dataset D	0.713 (0.534, 0.864)	0.587	0.889	0.537	0.242	0.967

Abbreviations: SVM, Support vector machine; DT, Decision tree; RF, Random forest; KNN, k-Nearest Neighbors; XGBoost, eXtreme Gradient Boosting.

Table S3: The Diagnostic Performance of Each Model

Model	Dataset	AUC (95%CI)	Accuracy	Sensitivity	Specificity	PPV	NPV
Radiomics	Dataset A	0.906 (0.872-0.939)	0.816	0.899	0.799	0.470	0.976
	Dataset B	0.812 (0.720-0.902)	0.717	0.889	0.682	0.364	0.968
	Dataset C	0.803 (0.707-0.888)	0.845	0.794	0.855	0.529	0.953
	Dataset D	0.713 (0.561-0.851)	0.587	0.889	0.537	0.242	0.967
ITH	Dataset A	0.892 (0.854-0.920)	0.806	0.899	0.788	0.456	0.975
	Dataset B	0.813 (0.705-0.917)	0.745	0.889	0.716	0.390	0.969
	Dataset C	0.807 (0.718-0.886)	0.805	0.794	0.807	0.458	0.950
	Dataset D	0.718 (0.509-0.918)	0.857	0.556	0.907	0.500	0.925
Deep_R	Dataset A	0.924 (0.879-0.962)	0.907	0.841	0.920	0.674	0.967
	Dataset B	0.833 (0.723-0.926)	0.877	0.722	0.909	0.619	0.941
	Dataset C	0.834 (0.767-0.894)	0.745	0.853	0.723	0.387	0.960
	Dataset D	0.747 (0.627-0.877)	0.571	1.000	0.500	0.250	1.000
Radscore-ITHscore	Dataset A	0.918 (0.881-0.952)	0.859	0.884	0.854	0.545	0.974
	Dataset B	0.821 (0.719-0.915)	0.726	0.833	0.705	0.366	0.954
	Dataset C	0.836 (0.755-0.910)	0.855	0.735	0.880	0.556	0.942
	Dataset D	0.765 (0.632-0.892)	0.778	0.667	0.796	0.353	0.935
Deep-Radscore	Dataset A	0.924 (0.889-0.950)	0.854	0.870	0.851	0.536	0.971
	Dataset B	0.822 (0.717-0.905)	0.698	0.889	0.659	0.348	0.967
	Dataset C	0.869 (0.813-0.920)	0.845	0.794	0.855	0.529	0.953
	Dataset D	0.767 (0.613-0.889)	0.698	0.889	0.667	0.308	0.973
Deep-ITHscore	Dataset A	0.934 (0.907, 0.960)	0.852	0.928	0.837	0.529	0.983
	Dataset B	0.831 (0.734, 0.917)	0.774	0.778	0.773	0.412	0.944
	Dataset C	0.860 (0.785, 0.919)	0.830	0.765	0.843	0.500	0.946
	Dataset D	0.809 (0.621, 0.946)	0.714	0.889	0.685	0.320	0.974
Fusion	Dataset A	0.934 (0.907-0.957)	0.871	0.913	0.862	0.568	0.980
	Dataset B	0.832 (0.741-0.917)	0.764	0.833	0.750	0.405	0.957
	Dataset C	0.884 (0.836-0.930)	0.800	0.824	0.795	0.452	0.957
	Dataset D	0.815 (0.660-0.935)	0.746	0.778	0.741	0.333	0.952

Table S4: The Importance of Features in the Radiomics Model

Feature	Weight
original_firstorder_Median	0.399
original_glszm_LargeAreaEmphasis	0.354
original_glcm_ClusterProminence	0.088
original_glszm_SizeZoneNonUniformity	0.065
original_firstorder_Range	0.048
original_glszm_SmallAreaHighGrayLevelEmphasis	0.046

Table S5: The Importance of Features in the ITH Model

Feature	Weight
ITH_original_gldm_GrayLevelNonUniformity	0.496
ITH_original_firstorder_90Percentile	0.176
ITH_original_glrlm_RunVariance	0.114
ITH_original_glcm_Idm	0.079
ITH_original_glcm_Imc2	0.043
ITH_original_glcm_ClusterProminence	0.039
ITH_original_glszm_GrayLevelNonUniformity	0.031
ITH_original_glcm_JointEnergy	0.022

Table S6: The Importance of Features in the Deep_R Model

Feature	Weight
f3	0.397
f1	0.139
f10	0.090
f2	0.079
f6	0.078
f4	0.056
f5	0.050
f9	0.041
f7	0.035
f8	0.034

Table S7: The Importance of Features in the Radscore-ITHscore Model

Feature	Weight
Radscore	0.649
ITHscore	0.351

Table S8: The Importance of Features in the Deep-Radscore Model

Feature	Weight
rad_score	0.352
f3	0.329
f1	0.114
f4	0.054
f10	0.050
f2	0.042
f5	0.039
f7	0.012
f6	0.008
f8	0.000
f9	0.000

Table S9: The Importance of Features in the Deep-ITHscore Model

Feature	Weight
ITH_score	0.456
f3	0.140
f1	0.093
f4	0.054
f2	0.051
f5	0.049
f7	0.040
f6	0.032
f9	0.030
f8	0.027
f10	0.026

Table S10: The Importance of Features in the fusion Model

Feature	Weight
ITHscore	0.301
Radscore	0.298
f3	0.073
f7	0.062
f10	0.061
f2	0.059
f4	0.057
f1	0.049
f6	0.041
f5	0.000
f8	0.000
f9	0.000