2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

A Radiomics-based Approach for Predicting Early Recurrence in Intrahepatic Cholangiocarcinoma after Surgical Resection: A Multicenter Study

Xiaohan Hao, Bing Liu, Xiaofei Hu, Jingwei Wei, Yuqi Han, Xianchuang Liu, Zhiyu Chen, Jiaping Li, Jie Bai, Yongliang Chen, Jian Wang, Meng Niu, and Jie Tian, *Fellow, IEEE*

***Abstract*— This work aimed to develop a noninvasive and reliable computed tomography (CT)-based imaging biomarker to predict early recurrence (ER) of intrahepatic cholangiocarcinoma (ICC) via radiomics analysis. In this retrospective study, a total of 177 ICC patients were enrolled from three independent hospitals. Radiomic features were extracted on CT images, then 11 feature selection algorithms and 4 classifiers were to conduct a multi-strategy radiomics modeling. Six established radiomics models were selected as stable ones by robustness-based rule. Among those models, Max-Relevance Min-Redundancy (MRMR) combined with Gradient Boosting Machine (GBM) yielded the highest areas under the receiver operating characteristics curve (AUCs) of 0.802 (95% confidence interval [CI]: 0.727-0.876) and 0.781 (95% CI: 0.655-0.907) in the training and test cohorts, respectively. To evaluate the generalization of the developed radiomics model, stratification analysis was performed regarding different centers. The MRMR-GBM-based model manifested good generalization with comparable AUCs in each hospital ($p > 0.05$ for paired comparison). Thus, the MRMR-GBM-based model could offer a potential imaging biomarker to assist the prediction of ER in ICC in a noninvasive manner.**

***Clinical Relevance*— The proposed radiomics model achieved satisfactory accuracy and good generalization ability in predicting ER in ICC, which might assist personalized surveillance and clinical treatment strategy making.**

I. Introduction

Intrahepatic cholangiocarcinoma (ICC) is the second most common primary hepatic malignancy which accounts for 10-20% of all primary liver tumors [1, 2]. Although ICC is relatively rare, it presents with severely aggressive tumor biology and dismal prognosis [3, 4]. Surgical resection is the mainstay of therapy and the only potentially curative treatment for patients with the resectable disease [5]. However, it is reported that up to 70% of patients developed recurrence after surgery [6], which rendered the median survival is only 12 months [7]. Therefore, preoperative risk assessment is essential that would greatly assist in timely postsurgical adjuvant therapy regime making and personalized surveillance to prolong the survival of ICC patients.

Recent studies reported that clinical factors include tumor size, lymph node metastasis, and vascular invasion were related to postoperative recurrence in ICC [7-9]. The American Joint Committee on the Cancer staging system could also guide postsurgical survival prediction [10]. However, those risk factors and staging schemes are mainly derived by population-based correlation analyses, thus lacking individualized predictive accuracy. Considering that medical imaging including computed tomography (CT) and magnetic resonance imaging (MRI) plays an important role in preoperative assessment of ICC during initial diagnosis, it is well worth exploring an imaging-based biomarker for the prediction of early recurrence (ER) in ICC [5].

A newly emerging technique – radiomics provides a potential approach to solve this targeted clinical issue. Based on machine learning methods, radiomics harnesses mineable quantitative features extracted from encrypted medical images along with clinical or genetic data to produce an evidence-based clinical decision support system [11]. It has been successfully applied in recurrence prediction, lymph node metastasis prediction, and immunotherapy evaluation [12-14]. Regarding ICC, Liang et al. [15] extracted radiomic features from MRI and applied Spearman's rank correlation and logistic regression to predict the ER in ICC. Zhao et al. [16] utilized MRI-based radiomic features with prognostic immunohistochemical markers via univariate analysis and logistic regression to enable the prediction of ER in ICC. The aforementioned studies demonstrated the feasibility and potential benefits of using MRI-based radiomics for ER

This work was supported by the major scientific and technological project of Guangdong Province (No. 2017B030308006); National Natural Science Foundation of China (No. 81601584, 81171441, 82001917, 81930053); the National Key Research and Development Program of China (No. 2017YFC1309100, 2017YFA0205200, 2019YFC0118100); the major program for tackling key problems of Guangzhou City, China (No. 201704020144); Chinese Academy of Sciences (No. GJJSTD20170004, QYZDJ-SSW-JSC005); Project of Innovation Capacity Promotion in Third Military Medical University (No. 2019XLC3007).

[+]X. Hao (hxh045@mail.ustc.edu.cn) is with Centers for Biomedical Engineering, University of Science and Technology of China.

[+]B. Liu and Y. Chen are with Department of Hepatobiliary Surgery, The First Medical Center of Chinese PLA General Hospital.

[+]X. Hu, Z. Chen, J. Bai and J. Wang are with Department of Radiology, Southwest Hospital, Third Military Medical University.

[+]J. Wei is with CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences.

Y. Han is with School of Life Science and Technology, Xidian University.

X. Liu and [*]M. Niu (guanjh@cmu.edu.cn) are with Department of Interventional Radiology, The First Affiliated Hospital of China Medical University.

J. Li is with Department of Interventional Oncology, The First Affiliated Hospital, Sun Yat-sen University.

[*]J. Tian (tian@ieee.org) is with the CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences; University of Chinese Academy of Sciences; Beijing Advanced Innovation Center for Big Data-Based Precision Medicine, School of Medicine, Beihang University; Engineering Research Center of Molecular and Neuro Imaging of Ministry of Education, School of Life Science and Technology, Xidian University.

[*] are corresponding authors [+] are co-first authors.

prediction. However, CT-based and multi-center radiomics studies are still limited for the prediction of ER in ICC.

Thus, we proposed multi-strategy radiomics modeling to predict ER in ICC using CT images in the multi-center cohort. In summary, our contributions are as follows: 1) Complete CT images and follow-up information were collected from 177 patients with ICC in three hospitals, which ensured enough amount to increase the credibility of the developed radiomics model. 2) Eleven feature selection algorithms and 4 classifiers were cross-combined to generate 44 combination modeling strategies, which would optimize the algorithm development for the prediction of ER. 3) Test cohort-based and cross-validation-based compound rules were applied to help identify the most accurate and reliable radiomics model.

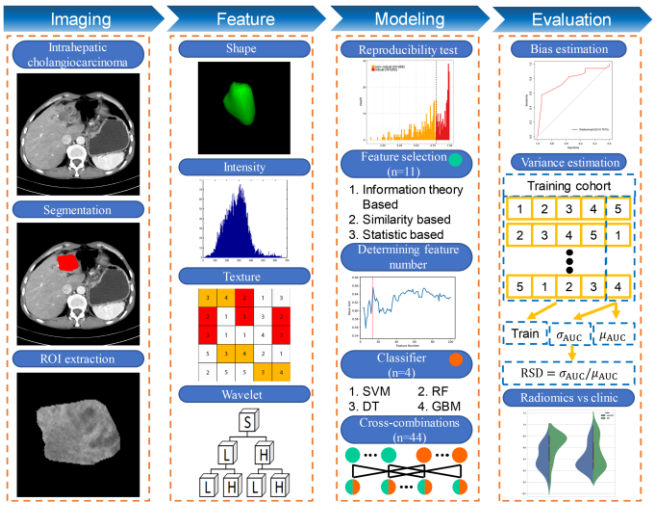

Figure 1. Radiomics model workflow in this study. Firstly, the ROIs were manually segmented by a radiologist from the CT images. Secondly, the radiomic features based on shape, intensity, texture, and wavelet were extracted from ROIs. After the reproducibility test to remove the unstable features, 11 feature selection algorithms and 4 classifiers are combined pair-wisely to construct the radiomics models. Finally, the optimal radiomics model was determined under a comprehensive assessment of accuracy and robustness and compared to the clinical model.

II. METHODS

A. Data

The experimental procedures involving human subjects described in this paper were approved by the Institutional Review Board of the three participating hospitals. A total of 177 patients were collected from August 2010 to August 2017 in three hospitals. The inclusion criteria were as follows: (i) patient who underwent radical resection was pathologically diagnosed as ICC; (ii) receipt of contrast-enhanced CT performed within one month before hepatectomy; and (iii) follow-up with a duration of one month and continued for at least 1 year. The exclusion criteria were as follows: (i) ICC confirmed by biopsy; (ii) disease diagnosed as combined hepatocellular carcinoma plus cholangiocarcinoma; and (iii) patient underwent radiofrequency ablation before surgery.

Patients were divided into a training cohort (from August 13, 2010, to June 30, 2016, n=124) and a time-independent test cohort (from July 1, 2016, to August 27, 2017, n=53). All patients had been followed up at least 1 year unless endpoint events occurred. We defined ER as recurrence within one year after surgical resection.

B. Image acquisition

CT images were acquired via different CT scanners in each hospital. The instruments of the three hospitals were 128-slice CT scanner (SOMATOM Definition AS+, Siemens Medical Solutions USA, Inc), 64-Slice LightSpeed VCT (GE Medical Systems, Milwaukee, WI), and SOMATOM Definition Flash (Siemens Healthcare), respectively. Scanning parameters and contrast media information are also different among those hospitals. After injection of contrast material, plain, arterial, portal venous, and delayed phases were acquired. We utilized portal venous phase CT images to conduct the radiomics analysis.

C. ROI segmentation and features extraction

The workflow of the study is shown in Figure 1. Firstly, the region of interest (ROI) was manually delineated on the largest cross-sectional layer of the tumor by a radiologist with 5-year working experience using ITK-SNAP software. A total of 473 two-dimensional radiomic features were extracted on the ROI areas. The radiomic features could be divided into four categories: (i) 13 shape-based features, (ii) 18 first-order intensity statistics, (iii) 74 texture features derived from gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), neighboring gray-tone difference matrix (NGTDM), and gray level dependence matrix (GLDM), and (iv) 368 wavelet features with the same first-order intensity statistics and textural features extracted from ROIs after different wavelet decomposition in two directions (x, y). A detailed description of the extracted radiomic features is provided at pyradiomics.readthedocs.io/en/latest/features.html [17].

D. Reproducibility test

To achieve reproducibility analysis, we randomly chose 12 patients from the training cohort and re-segmented the CT images by the former radiologist and another radiologist with 7-year working experience in a blinded fashion. The same feature extraction was conducted on newly segmented ROIs. Intra- and inter-class correlation coefficients were calculated to assess the intra- and inter-observer agreement of feature extraction, respectively. The radiomic features of intra-class correlation coefficients larger than 0.8 and inter-class correlation coefficients larger than 0.75 were regarded as reproducible features.

E. Machine learning algorithms

Eleven feature selection algorithms were utilized based on three different principles: (i) information theory-based including Conditional Infomax Feature Extraction (CIFE), Conditional Mutual Info Maximization (CMIM), Mutual Information Maximization (MIM), Joint Mutual Information (JMI), Mutual Information Feature Selection (MIFS) and Max-Relevance Min-Redundancy (MRMR); (ii) similarity-based including Spectral Feature Selection (SPEC) and ReliefF; and (iii) statistics-based including F-score, Gini Index (GI), and t-score.

We investigated two sorts of 4 classifiers: (i) base classifier including Support Vector Machine (SVM) with the radial basis function kernel and Decision Tree (DT) and (ii) ensemble classifier including the Gradient Boosting Machine (GBM) and Random Forest (RF).

F. Radiomics modeling process

All features were normalized by Z-score standardization. The abovementioned feature selection algorithms and classifiers were combined in pairs to acquire 44 modeling strategies. The feature selection algorithm could realize the ranking of features based on corresponding importance to predict the clinical target. The optimal number N of features was determined by the maximal mean area under the receiver operating characteristics curve (AUC) via 5-fold cross-validation based on initial logistics regression in the training cohort. The selected feature set consisted of the top N features was fed into the classifier to predict ER. The parameters of the classifiers were set by the grid search. The cut-off value was derived based on the maximized Youden index to divide all patients into high-risk or low-risk ER groups regarding outcomes of the radiomics model. The performance of each model was assessed by the receiver operator characteristic (ROC) analysis. To estimate the robustness of radiomics models, relative standard deviation (RSD) was calculated in 5-fold cross-validated in the training cohort [18]. RSD is defined as:

$$\mathrm{RSD} = \sigma_{\mathrm{AUC}} / \mu_{\mathrm{AUC}}$$

where σ_{AUC} is the standard deviation of 5-fold cross-validated AUC values, and μ_{AUC} is the mean of 5-fold cross-validated AUC values.

All modeling analyses were performed by Python v3.6.5 (www.python.org). Feature extraction was carried out on the pyradiomics package v2.0.0 (pyradiomics.readthedocs.io). Feature selection algorithms were implemented on the scikit-feature package v1.0.0 (featureselection.asu.edu). Classifier algorithms were implemented on the scikit-learn package v0.19.1 (scikit-learn.org/stable).

G. Stratification analysis

Stratification analysis was performed regarding different centers, the maximum diameter of tumor, and tumor-node-metastasis (TNM) stage. AUC, accuracy (ACC), sensitivity (SEN), specificity (SPE) in each subgroup were calculated to assess the performance of the radiomics model. Delong test was used to compare the AUCs.

H. Comparison to the clinical model

A total of 17 clinical factors were firstly analyzed by univariate analysis with a p-value < 0.1. Then, these factors were analyzed by multivariate analysis with a p-value < 0.05. The final selected factors were fitted by logistic regression to obtain a clinical model. Comparison of clinical and radiomics models was performed by ROC analysis and violin plot.

III. Results

A. Patient characteristics

Baseline characteristics of enrolled patients in the training and test cohorts showed no significant statistical differences in demographic or clinical characteristics between the training and test cohorts (p = 0.269–0.985).

In total, 101 (57.1%) patients developed with ER and 76 (42.9%) patients with non-ER. There was no significant difference for the ER status distribution in the training and test cohorts (p = 0.115).

B. Recurrence predictive performance

The optimal model was defined as a model with an RSD less than 0.05 and the highest AUC. According to this criteria, 6 models were selected and regarded as stable radiomics models. Among the remaining models, MRMR combined with GBM which was defined as the MRMR-GBM-based model, presented the best predictive performance (training AUC: 0.802(95% confidence interval [CI]: 0.727-0.876), training ACC:0.750, training SEN:0.879, training SPE:0.603; test AUC:0.781(95% CI: 0.655-0.907), test ACC:0.698, test SEN:0.714, test SPE:0.667; RSD:0.040). The heatmaps of test AUCs and RSDs of all of the methods are reported in Figure 2.

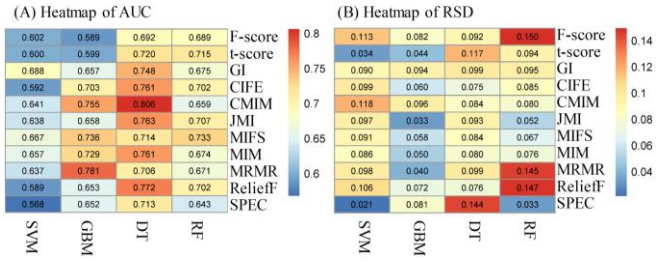

Figure 2. The predictive and robust performance of all radiomics models built by various feature selection algorithms (in rows) and classifiers (in columns). (A) Heatmap of AUC in the test cohort. (B) Heatmap of RSD.

C. Stratification analysis

Stratification analysis showed comparable predictive performance of the MRMR-GBM-based model in subgroups regarding different centers, tumor size, and TNM stage (p > 0.05 for paired comparison). The results in the test cohort are shown in Table I.

D. Comparison to the clinical model

Two clinical factors were selected - the maximum diameter of tumor and tumor number. The performance of the clinical model (test AUC: 0.565 [95% CI: 0.407-0.723]) was statistically lower from the MRMR-GBM-based model (p = 0.029). The distribution of predictive scores of the MRMR-GBM-based model in the test cohort indicated the strong discriminative ability (p = 0.002), but the clinical model was relatively poor (p = 0.281). The ROC curve and violin plot of the two models are shown in Figure 3.

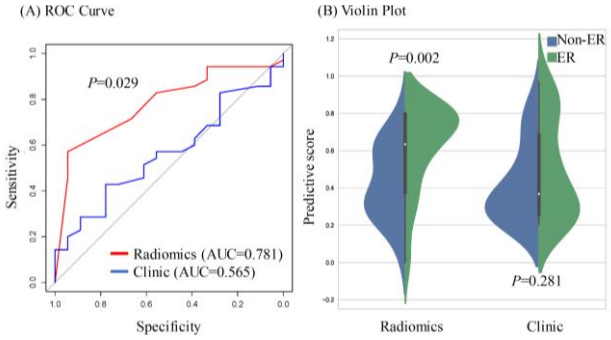

Figure 3. Comparison of the MRMR-GBM-based model and the clinical model. (A) ROC curve of the MRMR-GBM-based model and the clinical model. The p-value of the Delong test is less than 0.05, which indicates a significant statistical difference. (B) The violin plot of the MRMR-GBM-based model and the clinical model, which shows the distribution of predictive scores in the test cohort. The p-value of the Student's t-test less than 0.05 indicates a strong discriminative ability of the model between the ER and non-ER groups, and vice versa.

TABLE I.

STRATIFICATION ANALYSIS OF THE OPTIMAL RADIOMICS MODEL IN THE TEST COHORT

Subgroups		AUC (95% CI)	ACC	SEN	SPE
Center	Southwest Hospital	0.756* (0.544-0.969)	0.682	0.692	0.667
	Chinese PLA General Hospital	0.739* (0.548-0.930)	0.680	0.706	0.625
	The First Affiliated Hospital of China Medical University	1.000* (1.000-1.000)	0.833	0.800	1.000
TNM Stage	I/II	0.731* (0.534-0.927)	0.667	0.600	0.750
	III/IV	0.754* (0.515-0.994)	0.731	0.800	0.500
The maximum diameter of tumor	≤2 cm	0.825* (0.636-1.000)	0.722	0.545	1.000
	2-5 cm	0.769* (0.527-1.000)	0.600	0.462	0.857
	>5 cm	0.830* (0.620-1.000)	0.800	0.727	1.000

* means the AUC of the subgroup no statistically less than the overall AUC ($p > 0.05$)

IV. DISCUSSIONS AND CONCLUSION

In this retrospective study, we conducted a multi-strategy and CT-based radiomics analysis in the multi-center cohort for predicting ER in patients with ICC. Previous studies performed MRI-based methods to predict prognosis in HCC based on single-center cohort, and limited machine learning algorithms were compared in the studies [15-16]. Considering that CT images are more commonly used at the initial diagnosis of ICC in clinical settings, the CT-based radiomics model is worthy to be explored to complement the relevant studies. Besides, identifying the optimal modeling methodology is crucial in radiomics analysis [11]. Machine learning-based methods may behave with significant variation towards different clinical issues and sample sizes. Thus in this study, we fulfilled 44 modeling strategies to figure out the optimal machine learning method for the target of ER prediction in ICC in the multi-center cohort. Our finding manifested that the MRMR-GBM-based strategy was the most suitable modeling strategy for ER prediction in ICC. Meanwhile, the results proved that the noninvasive imaging biomarker derived via radiomics could compete with the clinical model with improved prediction accuracy and a good generalization capability. This study has some limitations. Genetic information is also of great significance for prognosis prediction in ICC, thus, the inclusion of genetic data could be integrated into the prediction system to achieve a more accurate outcome. In addition, although a relatively satisfactory result was achieved, a prospective study should be proposed to validate the effectiveness of the noninvasive imaging biomarker for ER prediction in ICC.

In conclusion, CT-based radiomics could generate a promising imaging biomarker for preoperative prediction of ER in ICC to assist personalized treatment strategy making.

REFERENCES

[1] Aljiffry M, Abdulelah A, Walsh M, et al., "Evidence-Based Approach to Cholangiocarcinoma: A Systematic Review of the Current Literature," *Journal of the American College of Surgeons*, vol. 208, no. 1, pp. 134-147, 2009.

[2] Khan SA, Toledano MB, Taylor-Robinson SD, Epidemiology, "risk factors, and pathogenesis of cholangiocarcinoma," *HPB*, vol. 10, no. 2, pp. 77-82, 2008.

[3] Zhou XD, Tang ZY, Fan J, et al., "Intrahepatic cholangiocarcinoma: report of 272 patients compared with 5,829 patients with hepatocellular carcinoma," *Journal of cancer research and clinical oncology*, vol. 135, no. 8, pp. 1073-1080, 2009.

[4] Nathan H, Pawlik TM, Wolfgang CL, et al., "Trends in survival after surgery for cholangiocarcinoma: A 30-year population-based SEER database analysis," *Journal of gastrointestinal surgery*, vol. 11, no. 11, pp. 1488-1497, 2007.

[5] Weber SM, Ribero D, O'Reilly EM, et al., "Intrahepatic Cholangiocarcinoma: expert consensus statement," *HPB*, vol. 17, no. 8, pp. 669-680, 2015.

[6] Spolverato G, Kim Y, Alexandrescu S, et al., "Management and Outcomes of Patients with Recurrent Intrahepatic Cholangiocarcinoma Following Previous Curative-Intent Surgical Resection," *Annals of surgical oncology*, vol. 23, no. 1, pp. 235-243, 2016.

[7] Hyder O, Hatzaras I, Sotiropoulos GC, et al., "Recurrence after operative management of intrahepatic cholangiocarcinoma," *Surgery*, vol. 153, no. 6, pp. 811-818, 2013.

[8] Fisher SB, Patel SH, Kooby DA, et al., "Lymphovascular and perineural invasion as selection criteria for adjuvant therapy in intrahepatic cholangiocarcinoma: a multi-institution analysis," *HPB*, vol. 14, no. 8, pp. 514-522, 2012.

[9] Ribero D, Rosso S, Pinna AD, et al., "Postoperative nomogram for predicting survival after resection for intrahepatic cholangiocarcinoma," *Journal of Clinical Oncology*, vol. 31, no. 15, pp. 4129-4129, 2013.

[10] Amin MB, Greene FL, Edge SB, et al., "The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population‐based to a more "personalized" approach to cancer staging," *CA: a cancer journal for clinicians*, vol. 67, no. 2, pp. 93-99, 2017.

[11] Lambin P, Leijenaar RTH, Deist TM, et al., "Radiomics: the bridge between medical imaging and personalized medicine," *Nature reviews Clinical oncology*, vol. 14, no. 12, pp. 749-762, 2017.

[12] Li H, Zhu YT, Burnside ES, et al., "MR Imaging Radiomics Signatures for Predicting the Risk of Breast Cancer Recurrence as Given by Research Versions of MammaPrint, Oncotype DX, and PAM50 Gene Assays," *Radiology*, vol. 281, no. 2, pp. 382-391, 2016.

[13] Huang YQ, Liang CH, He L, et al., "Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer," *Journal of Clinical Oncology*, vol.34, no. 18, pp. 2157-2164, 2016.

[14] Sun R, Limkin EJ, Vakalopoulou M, et al., "A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study," *The Lancet Oncology*, vol. 19, no. 9, pp. 1180-1191, 2018.

[15] Liang WJ, Xu L, Yang PF, et al., "Novel Nomogram for Preoperative Prediction of Early Recurrence in Intrahepatic Cholangiocarcinoma," *Frontiers in oncology*, vol. 8, pp. 360, 2018.

[16] Zhao L, Ma X, Liang M, et al., "Prediction for early recurrence of intrahepatic mass-forming cholangiocarcinoma: quantitative magnetic resonance imaging combined with prognostic immunohistochemical markers," *Cancer Imaging*, vol. 19, no. 1, pp. 1-10, 2019.

[17] van Griethuysen JJM, Fedorov A, Parmar C, et al., "Computational Radiomics System to Decode the Radiographic Phenotype," *Cancer Research*, vol. 77, no. 21, pp. 104–107, 2017.

[18] Parmar C, Grossmann P, Bussink J, et al., "Machine Learning methods for Quantitative Radiomic Biomarkers," *Scientific reports*, vol. 5, no. 1, pp. 1-11, 2015.