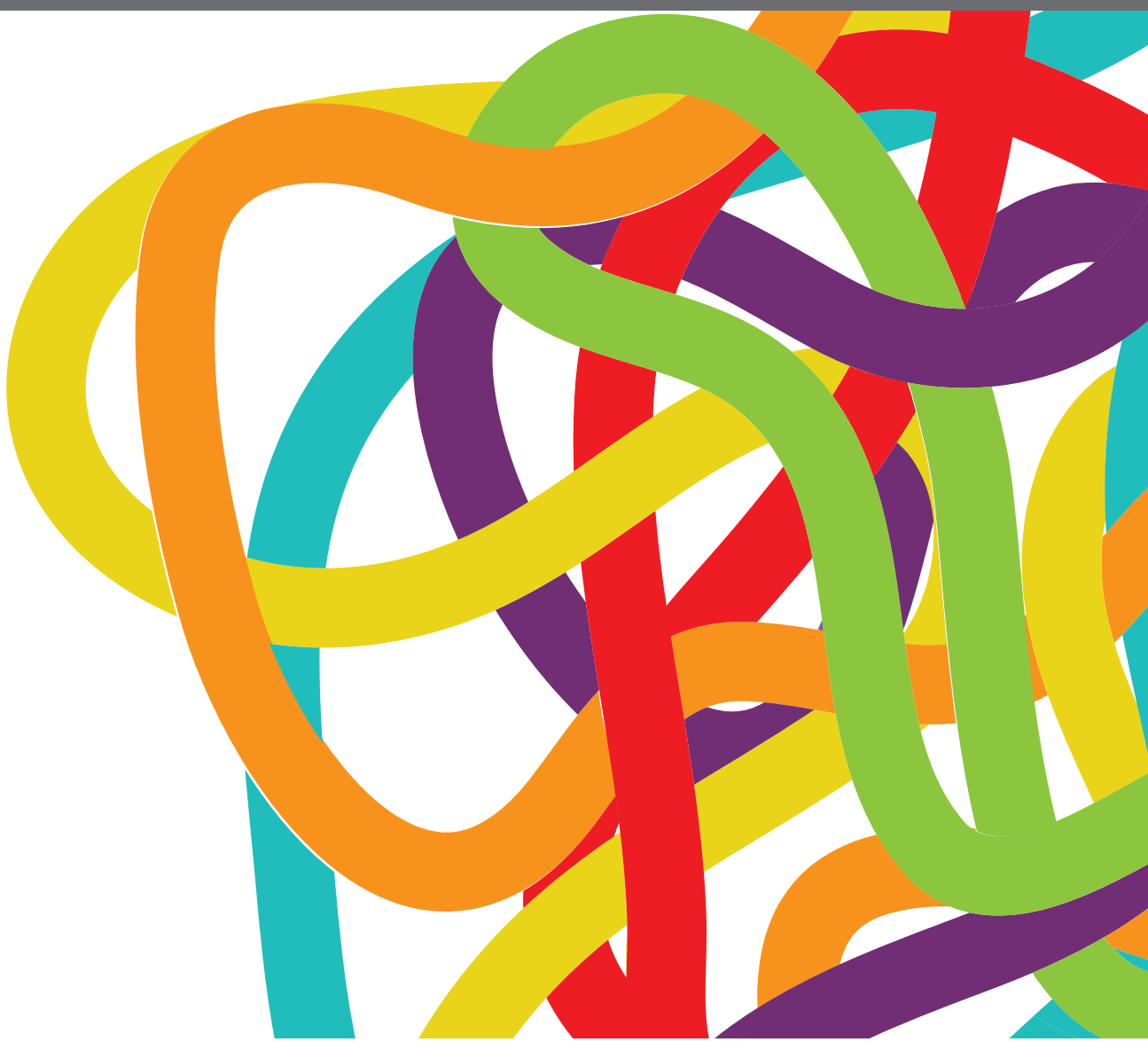
COMPUTATIONAL RADIOMICS FOR CANCER CHARACTERIZATION

EDITED BY: Omar Sultan Al-Kadi, Xujiong Ye, Giorgio Russo and Joseph Ross Mitchell
PUBLISHED IN: Frontiers in Oncology and Frontiers in Bioengineering and Biotechnology



frontiers Research Topics



frontiers

Frontiers eBook Copyright Statement

The copyright in the text of individual articles in this eBook is the property of their respective authors or their respective institutions or funders. The copyright in graphics and images within each article may be subject to copyright of other parties. In both cases this is subject to a license granted to Frontiers.

The compilation of articles constituting this eBook is the property of Frontiers.

Each article within this eBook, and the eBook itself, are published under the most recent version of the Creative Commons CC-BY licence. The version current at the date of publication of this eBook is CC-BY 4.0. If the CC-BY licence is updated, the licence granted by Frontiers is automatically updated to the new version.

When exercising any right under the CC-BY licence, Frontiers must be attributed as the original publisher of the article or eBook, as applicable.

Authors have the responsibility of ensuring that any graphics or other materials which are the property of others may be included in the CC-BY licence, but this should be checked before relying on the CC-BY licence to reproduce those materials. Any copyright notices relating to those materials must be complied with.

Copyright and source acknowledgement notices may not be removed and must be displayed in any copy, derivative work or partial copy which includes the elements in question.

All copyright, and all rights therein, are protected by national and international copyright laws. The above represents a summary only. For further information please read Frontiers' Conditions for Website Use and Copyright Statement, and the applicable CC-BY licence.

ISSN 1664-8714
ISBN 978-2-83250-315-7
DOI 10.3389/978-2-83250-315-7

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: frontiersin.org/about/contact

COMPUTATIONAL RADIOMICS FOR CANCER CHARACTERIZATION

Topic Editors:
Omar Sultan Al-Kadi, The University of Jordan, Jordan
Xujiong Ye, University of Lincoln, United Kingdom
Giorgio Russo, Institute of Bioimaging and Molecular Physiology, Department of Biomedical Sciences, National Research Council (CNR), Italy
Joseph Ross Mitchell, University of Alberta, Canada

Citation: Al-Kadi, O. S., Ye, X., Russo, G., Mitchell, J. R., eds. (2022). Computational Radiomics for Cancer Characterization. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-83250-315-7

Table of Contents

07 ***Prediction of the Growth Rate of Early-Stage Lung Adenocarcinoma by Radiomics***
Mingyu Tan, Weiling Ma, Yingli Sun, Pan Gao, Xuemei Huang, Jinjuan Lu, Wufei Chen, Yue Wu, Liang Jin, Lin Tang, Kaiming Kuang and Ming Li

16 ***Development of CT-Based Imaging Signature for Preoperative Prediction of Invasive Behavior in Pancreatic Solid Pseudopapillary Neoplasm***
Wen-peng Huang, Si-yun Liu, Yi-jing Han, Li-ming Li, Pan Liang and Jian-bo Gao

27 ***CT-Based Radiomics Signature With Machine Learning Predicts MYCN Amplification in Pediatric Abdominal Neuroblastoma***
Xin Chen, Haoru Wang, Kaiping Huang, Huan Liu, Hao Ding, Li Zhang, Ting Zhang, Wenqing Yu and Ling He

38 ***A Deep Multi-Task Learning Framework for Brain Tumor Segmentation***
He Huang, Guang Yang, Wenbo Zhang, Xiaomei Xu, Weiji Yang, Weiwei Jiang and Xiaobo Lai

54 ***Predicting Microsatellite Instability Status in Colorectal Cancer Based on Triphasic Enhanced Computed Tomography Radiomics Signatures: A Multicenter Study***
Yuntai Cao, Guojin Zhang, Jing Zhang, Yingjie Yang, Jialiang Ren, Xiaohong Yan, Zhan Wang, Zhiyong Zhao, Xiaoyu Huang, Haihua Bao and Junlin Zhou

66 ***Radiomics Analysis of Contrast-Enhanced CT Predicts Survival in Clear Cell Renal Cell Carcinoma***
Lei Yan, Guangjie Yang, Jingjing Cui, Wenjie Miao, Yangyang Wang, Yujun Zhao, Ning Wang, Aidi Gong, Na Guo, Pei Nie and Zhenguang Wang

76 ***Prediction of Inflammatory Breast Cancer Survival Outcomes Using Computed Tomography-Based Texture Analysis***
Sung Eun Song, Bo Kyoung Seo, Kyu Ran Cho, Ok Hee Woo, Balaji Ganeshan, Eun Sil Kim and Jaehyung Cha

87 ***Radiomics-Based Texture Analysis of ^{68}Ga-DOTATATE Positron Emission Tomography and Computed Tomography Images as a Prognostic Biomarker in Adults With Neuroendocrine Cancers Treated With ^{177}Lu-DOTATATE***
Charlotte Atkinson, Balaji Ganeshan, Raymond Endozo, Simon Wan, Matthew D. Aldridge, Ashley M. Groves, Jamshed B. Bomanji and Mark N. Gaze

96 ***Performances of Whole Tumor Texture Analysis Based on MRI: Predicting Preoperative T Stage of Rectal Carcinomas***
Jia You and Jiandong Yin

108 ***Overall Survival Prediction for Gliomas Using a Novel Compound Approach***
He Huang, Wenbo Zhang, Ying Fang, Jialing Hong, Shuaixi Su and Xiaobo Lai

128 ***A Multichannel Deep Neural Network for Retina Vessel Segmentation via a Fusion Mechanism***
Jiaqi Ding, Zehua Zhang, Jijun Tang and Fei Guo

142 ***CEUS-Based Radiomics Can Show Changes in Protein Levels in Liver Metastases After Incomplete Thermal Ablation***
Haiwei Bao, Ting Chen, Junyan Zhu, Haiyang Xie and Fen Chen

152 ***Radiomics Models for the Preoperative Prediction of Pelvic and Sacral Tumor Types: A Single-Center Retrospective Study of 795 Cases***
Ping Yin, Xin Zhi, Chao Sun, Sicong Wang, Xia Liu, Lei Chen and Nan Hong

161 ***Radiomic Evaluations of the Diagnostic Performance of DM, DBT, DCE MRI, DWI, and Their Combination for the Diagnosisof Breast Cancer***
Shuxian Niu, Xiaoyu Wang, Nannan Zhao, Guanyu Liu, Yangyang Kan, Yue Dong, E-Nuo Cui, Yahong Luo, Tao Yu and Xiran Jiang

171 ***CT Radiomics for the Preoperative Prediction of Ki67 Index in Gastrointestinal Stromal Tumors: A Multi-Center Study***
Yilei Zhao, Meibao Feng, Minhong Wang, Liang Zhang, Meirong Li and Chencui Huang

177 ***Progress of MRI Radiomics in Hepatocellular Carcinoma***
Xue-Qin Gong, Yun-Yun Tao, Yao–Kun Wu, Ning Liu, Xi Yu, Ran Wang, Jing Zheng, Nian Liu, Xiao-Hua Huang, Jing-Dong Li, Gang Yang, Xiao-Qin Wei, Lin Yang and Xiao-Ming Zhang

190 ***Prediction of Locoregional Recurrence-Free Survival of Oesophageal Squamous Cell Carcinoma After Chemoradiotherapy Based on an Enhanced CT-Based Radiomics Model***
Jie Kong, Shuchai Zhu, Gaofeng Shi, Zhikun Liu, Jun Zhang and Jialiang Ren

202 ***A Novel Radiogenomics Biomarker Based on Hypoxic-Gene Subset: Accurate Survival and Prognostic Prediction of Renal Clear Cell Carcinoma***
Jiahao Gao, Fangdie Ye, Fang Han, Xiaoshuang Wang, Haowen Jiang and Jiawen Zhang

213 ***Exploring CT Texture Parameters as Predictive and Response Imaging Biomarkers of Survival in Patients With Metastatic Melanoma Treated With PD-1 Inhibitor Nivolumab: A Pilot Study Using a Delta-Radiomics Approach***
Antonino Guerrisi, Michelangelo Russillo, Emiliano Loi, Balaji Ganeshan, Sara Ungania, Flora Desiderio, Vicente Bruzzaniti, Italia Falcone, Davide Renna, Virginia Ferraresi, Mauro Caterino, Francesco Maria Solivetti, Francesco Cognetti and Aldo Morrone

225 ***CT Radiomics Model for Predicting the Ki-67 Index of Lung Cancer: An Exploratory Study***
Qing Fu, Shun li Liu, Da peng Hao, Ya bin Hu, Xue jun Liu, Zaixian Zhang, Wen hong Wang, Xiao yan Tang, Chuan yu Zhang and Shi he Liu

235 ***Nomogram Combining Radiomics With the American College of Radiology Thyroid Imaging Reporting and Data System Can Improve Predictive Performance for Malignant Thyroid Nodules***
Xingzhi Huang, Zhenghua Wu, Aiyun Zhou, Xiang Min, Qi Qi, Cheng Zhang, Songli Chen and Pan Xu

246 ***Radiomics Based on Contrast-Enhanced MRI in Differentiation Between Fat-Poor Angiomyolipoma and Hepatocellular Carcinoma in Noncirrhotic Liver: A Multicenter Analysis***
Xiangtian Zhao, Yukun Zhou, Yuan Zhang, Lujun Han, Li Mao, Yizhou Yu, Xiuli Li, Mengsu Zeng, Mingliang Wang and Zaiyi Liu

257 ***Diagnostic Value of CT- and MRI-Based Texture Analysis and Imaging Findings for Grading Cartilaginous Tumors in Long Bones***
Xue-Ying Deng, Hai-Yan Chen, Jie-Ni Yu, Xiu-Liang Zhu, Jie-Yu Chen, Guo-Liang Shao and Ri-Sheng Yu

267 ***Exploring the Value of Radiomics Features Based on B-Mode and Contrast-Enhanced Ultrasound in Discriminating the Nature of Thyroid Nodules***
Shi Yan Guo, Ping Zhou, Yan Zhang, Li Qing Jiang and Yong Feng Zhao

276 ***CT-Based Peritumoral and Intratumoral Radiomics as Pretreatment Predictors of Atypical Responses to Immune Checkpoint Inhibitor Across Tumor Types: A Preliminary Multicenter Study***
Shuai He, Yuqing Feng, Qi Lin, Lihua Wang, Lijun Wei, Jing Tong, Yuwei Zhang, Ying Liu, Zhaoxiang Ye, Yan Guo, Tao Yu and Yahong Luo

288 ***Identification of Stage IIIC/IV EGFR-Mutated Non-Small Cell Lung Cancer Populations Sensitive to Targeted Therapy Based on a PET/CT Radiomics Risk Model***
Dan Shao, Dongyang Du, Haiping Liu, Jieqin Lv, You Cheng, Hao Zhang, Wenbing Lv, Shuxia Wang and Lijun Lu

300 ***Multiple Survival Outcome Prediction of Glioblastoma Patients Based on Multiparametric MRI***
Bin Wang, Shan Zhang, Xubin Wu, Ying Li, Yueming Yan, Lili Liu, Jie Xiang, Dandan Li and Ting Yan

312 ***Multi-Phase CT-Based Radiomics Nomogram for Discrimination Between Pancreatic Serous Cystic Neoplasm From Mucinous Cystic Neoplasm***
Jiahao Gao, Fang Han, Xiaoshuang Wang, Shaofeng Duan and Jiawen Zhang

322 ***Comprehensive Imaging Characterization of Colorectal Liver Metastases***
Drew Maclean, Maria Tsakok, Fergus Gleeson, David J. Breen, Robert Goldin, John Primrose, Adrian Harris and James Franklin

333 ***Maximum Standardized Uptake Value of ^{18}F-deoxyglucose PET Imaging Increases the Effectiveness of CT Radiomics in Differentiating Benign and Malignant Pulmonary Ground-Glass Nodules***
Rong Niu, Jianxiong Gao, Xiaoliang Shao, Jianfeng Wang, Zhenxing Jiang, Yunmei Shi, Feifei Zhang, Yuetao Wang and Xiaonan Shao

342 ***A Comparative Study of Radiomics and Deep-Learning Based Methods for Pulmonary Nodule Malignancy Prediction in Low Dose CT Images***
Mehdi Astaraki, Guang Yang, Yousuf Zakko, Iuliana Toma-Dasu, Örjan Smedby and Chunliang Wang

354 ***Development and Validation of a CT-Based Radiomics Nomogram for Predicting Postoperative Progression-Free Survival in Stage I–III Renal Cell Carcinoma***
Haijie Zhang, Fu Yin, Menglin Chen, Liyang Yang, Anqi Qi, Weiwei Cui, Shanshan Yang and Ge Wen

366 ***CT-Based Radiomics Showing Generalization to Predict Tumor Regression Grade for Advanced Gastric Cancer Treated With Neoadjuvant Chemotherapy***

Yong Chen, Wei Xu, Yan-Ling Li, Wentao Liu, Birendra Kumar Sah, Lan Wang, Zhihan Xu, Michael Wels, Yanan Zheng, Min Yan, Huan Zhang, Qianchen Ma, Zhenggang Zhu and Chen Li

377 ***Response Prediction to Concurrent Chemoradiotherapy in Esophageal Squamous Cell Carcinoma Using Delta-Radiomics Based on Sequential Whole-Tumor ADC Map***

Dianzheng An, Qiang Cao, Na Su, Wanhu Li, Zhe Li, Yanxiao Liu, Yuxing Zhang and Baosheng Li

388 ***Predictive Value of Delta-Radiomics Texture Features in 0.35 Tesla Magnetic Resonance Setup Images Acquired During Stereotactic Ablative Radiotherapy of Pancreatic Cancer***

Garrett Simpson, William Jin, Benjamin Spieler, Lorraine Portelance, Eric Mellon, Deukwoo Kwon, John C. Ford and Nesrin Dogan

397 ***The Value of CT-Based Radiomics for Predicting Spread Through Air Spaces in Stage IA Lung Adenocarcinoma***

Xiaoyu Han, Jun Fan, Yuting Zheng, Chengyu Ding, Xiaohui Zhang, Kailu Zhang, Na Wang, Xi Jia, Yumin Li, Jia Liu, Jinlong Zheng and Heshui Shi

frontiers in Oncology

ORIGINAL RESEARCH
published: 15 April 2021
doi: 10.3389/fonc.2021.658138

Prediction of the Growth Rate of Early-Stage Lung Adenocarcinoma by Radiomics

Mingyu Tan[1†], Weiling Ma[1†], Yingli Sun[1], Pan Gao[1], Xuemei Huang[1], Jinjuan Lu[1], Wufei Chen[1], Yue Wu[2], Liang Jin[1], Lin Tang[3], Kaiming Kuang[4] and Ming Li[1]*

[1] Department of Radiology, Huadong Hospital Affiliated With Fudan University, Shanghai, China, [2] Department of Thoracic Surgery, Huadong Hospital Affiliated With Fudan University, Shanghai, China, [3] Department of Radiology, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China, [4] Dianei Technology, Shanghai, China

Objectives: To investigate the value of imaging in predicting the growth rate of early lung adenocarcinoma.

Methods: From January 2012 to June 2018, 402 patients with pathology-confirmed lung adenocarcinoma who had two or more thin-layer CT follow-up images were retrospectively analyzed, involving 407 nodules. Two complete preoperative CT images and complete clinical data were evaluated. Training and validation sets were randomly assigned according to an 8:2 ratio. All cases were divided into fast-growing and slow-growing groups. Researchers extracted 1218 radiomics features from each volumetric region of interest (VOI). Then, radiomics features were selected by repeatability analysis and Analysis of Variance (ANOVA); Based on the Univariate and multivariate analyses, the significant radiographic features is selected in training set. A decision tree algorithm was conducted to establish the radiographic model, radiomics model and the combined radiographic-radiomics model. Model performance was assessed by the area under the curve (AUC) obtained by receiver operating characteristic (ROC) analysis.

Results: Sixty-two radiomics features and one radiographic features were selected for predicting the growth rate of pulmonary nodules. The combined radiographic-radiomics model (AUC 0.78) performed better than the radiographic model (0.727) and the radiomics model (0.710) in the validation set.

Conclusions: The model has good clinical application value and development prospects to predict the growth rate of early lung adenocarcinoma through the combined radiographic-radiomics model.

Keywords: pulmonary nodules, tomography, X-ray computer, radiomics, volume doubling time, machine learning

OPEN ACCESS

Edited by:
Omar Sultan Al-Kadi,
The University of Jordan, Jordan

Reviewed by:
Pritam Mukherjee,
Stanford University, United States
Shaofeng Duan,
GE Healthcare, China

****Correspondence:***
Ming Li
minli77@163.com

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *25 January 2021*
Accepted: *22 March 2021*
Published: *15 April 2021*

Citation:
Tan M, Ma W, Sun Y, Gao P, Huang X, Lu J, Chen W, Wu Y, Jin L, Tang L, Kuang K and Li M (2021) Prediction of the Growth Rate of Early-Stage Lung Adenocarcinoma by Radiomics. Front. Oncol. 11:658138. doi: 10.3389/fonc.2021.658138

INTRODUCTION

Lung cancer has the highest incidence rate of all cancers in China and worldwide. The incidence rate and mortality rate of lung cancer in 2018 were 11.6% and 18.4%, respectively (1). In recent years, the incidence rate of lung adenocarcinoma was the highest among lung cancers, accounting for 60% of primary lung cancers, and lung adenocarcinoma is the most common histological type of lung cancer

(2, 3). With the widespread use of low-dose CT screening, the increasing early-stage lung cancer are discovered (4). A pulmonary nodule is defined as a rounded opacity that is well or poorly defined measuring up to 3 cm in diameter (5). Although the current guidelines issued by the National Comprehensive Cancer Network (NCCN) (6, 7), Fleischner Society (8), American College of Chest Physicians (ACCP) (9), and the Asian Consensus (10) differ in some respects regarding the diagnosis and treatment of and follow-up strategies for pulmonary nodules, most guidelines are based on radiographic features, such as the type and size of pulmonary nodules. Lung adenocarcinoma is an inert tumor compared with squamous cell carcinoma and small cell carcinoma (11), although it is not uncommon for lung nodules to rapidly grow from the early stage to advanced stage in a short period. However, such nodules are usually followed up with CT examinations, and the best treatment time may be missed. Following the diagnosis and treatment methods recommended by previous guidelines for such fast-growing malignant nodules may lead to untimely diagnosis and treatment of patients, resulting in serious disease. Among all early-stage lung cancers (T1N0M0 stage), the choice of treatment and the prognosis of patients are sometimes quite different in clinical practice. Among similarly sized pulmonary nodules, the prognosis of aggressive lung adenocarcinoma is significantly worse than that of inert nodules.

Radiomics extracts engineering features related to morphology, histogram, intensity, and texture and integrates knowledge from multiple fields of imaging (image interpretation), computers (quantitative feature extraction) and machine learning (model establishment and evaluation) (12, 13). Radiomics can reveal a large amount of invisible, high-dimensional information with potential clinical value hidden behind the image. Currently, radiomics has achieved good results in the diagnosis of benign and malignant pulmonary nodules (14, 15). However, there are few radiomics methods for the prediction of the growth rate of pulmonary nodules. Therefore, the aim of this study was to establish a radiomics-based model for the prediction of the growth rate of early lung adenocarcinoma to assist in clinical decision-making.

MATERIALS AND METHODS

Data Collection

Methods

From January 2012 to June 2018, patients with lung adenocarcinoma confirmed by pathology who had two or more thin-layer CT images were selected. The inclusion criteria were as follows: 1) nodule diameter<3 cm; 2) more than 2 preoperative thin-layer CT scans with an interval of more than 30 days (15); 3) pulmonary nodules confirmed by pathology as lung adenocarcinoma; and 4) all cases with stage T1N0M0. The exclusion criteria were as follows: 1) preoperative surgery or chemoradiotherapy treatment; 2) no complete clinical data; and 3) unclear image due to respiratory motion and other factors and nodules with details that could not be displayed.

In total, 407 pulmonary nodules from 402 patients (mean age 58.45, 22-84 years) were included in this study. The cases were randomly divided into the training set and validation set at a ratio of 8:2. Only no-contrast CT images were included in this study. If the patients had more than two CT images, the first CT scan image and the last preoperative CT scan image were selected. The average interval between the two CT images was 567.56 days (range 30-2813), with a median of 397 days. This study was approved by the institutional review committee of our hospital, and patient informed consent was not required.

Radiographic Features and Postoperative Pathological Evaluation

The first CT images of all patients were independently evaluated by two chest radiologists with 6 and 12 years of chest CT reading experience (evaluation conditions: window width, 1500 Hounsfield units [HU]; window position, -700 HU). Any discrepancies in the interpretation between the observers were resolved by a consensus. The CT findings of each lesion were analyzed, including (1) the lesion location, (2) lesion type (pure ground glass, partial solid nodule, or solid nodule), (3) lesion size, (4) margin (clear or blurred), (5) nodule shape (round, oval, or irregular), (6) pleural attachment, including pleural tag and indentation (absent or present), (7) bubble (absent or present), (8) bronchiole change (absent or present), (9) vascular change (absent or present), (10) and lobulation (absent or present). Solid nodule as a nodule that completely obscures the entire lung parenchyma within it. Part-solid GGN are those having sections that are solid in this sense, and pure GGN are those with no solid parts (16). Vascular changes were defined as the thickening and twisting of blood vessels through the lesion or aggregation of vessels surrounding the lesion (17). Bronchiole changes were defined as enlargement, distortion, or obstruction of the bronchus through the lesion. In the analysis of interobserver reliability, the type (pure GGN, part GGN or solid nodules) were compared between 2 observers, diagnostic concordance was assessed by unweighted kappa values. Two senior pathologists (chest pathologists with more than 5 and 10 years of working experience) evaluated and reviewed the lung tissue according to the classification of lung adenocarcinoma by the International Lung Cancer Research Association, American Thoracic Society and European Respiratory Society (2).

Evaluation of Growth Rate

The tumor volume doubling time (VDT) is a key parameter used to distinguish fast-growing tumors from slow-growing tumors (18). Since few studies investigated the growth rate of pulmonary nodules, we defined the growth rate of pulmonary nodules according to previous studies. Most studies defined pulmonary nodules with a VDT >400 days and VDT ≤400 days as slow-growing and fast-growing pulmonary nodules, respectively (11, 19).

Image Acquisition, Nodule Segmentation and VDT Acquisition

All CT scans were performed with one of the four scanners (GE Discovery CT750HD, 64-slice LightSpeed, VCT, Somatom

Definition Flash; Somatom Sensation 16). The detailed scan and reconstruction parameters are listed in **Table 1**.

An open-source medical image processing and navigation software 3D slicer (version 4.8.0, Brigham and Women's Hospital) was used to manually delineate the volume of interest (VOI) of the 814 nodules by a radiologist with 6 years of experience with chest CT interpretation; then, the VOI was confirmed by another radiologist with 12 years of chest CT interpretation who corrected the boundary of each nodule to avoid the influence of vessels, bronchus, pleura and other structures outside the nodule to the greatest extent possible (20). Finally, all CT data of the VOI of all nodules were exported in NII (desensitization format) for the following analysis.

The volume doubling time (VDT) refers to the time required to calculate the volume doubling based on an exponential growth model (21). Two CT scan images were selected for each patient to analyze and calculate the VDT as follows: the first CT image was selected as the scan image from the first CT scan in our hospital, and the second image was the last preoperative CT scan image. The following formula was used to calculate the VDT:

VDT = (volume doubling time) = $\frac{(T1-T0*log2}{log\ \ (V1/V0)}$,where $V0$ and $V1$ represent the volumes at $T1$ (time 1- the second examination date) and $T0$ (time 0 – the baseline examination date).

Feature Extraction and Repeatability Analysis of Radiomics

Feature Extraction

Pyradiomics Toolkit (Version 2.1.0, https://github.com/Radiomics/pyradiomics) was used to extracted the radiomics features, including the first-order features based on the CT value or the pixel value of the preprocessed image, the shape descriptor features used to describe the shape and size, the gray-level cooccurrence matrix (GLCM), the gray-level run lengths matrix (GLRLM), the gray level size zone matrix (GLSZM), and the gray level dependence matrix (GLDM), to describe the internal and surface texture of the lesion. In total, 1218 radiomics features were extracted from each lesion (22).

Repeatability Analysis

The specific method was as follows: 60 nodules were randomly selected for independent segmentation by a radiologist with 6 years of experience with chest CT interpretation. One month later, the radiologist repeated the same procedure. Then, the interclass correlation coefficient (ICC) was assessed to analyze the correlation of the 1218 features extracted from these 60 nodules. Finally, features with an ICC >0.80 were selected and included in the follow-up study (23) (**Figure 1**).

Feature Selection, Model Establishment, and Verification

Feature Selection

Data collected in practice often have many missing values, duplicate values, and abnormal values, but the final value of the model depends on the amount of useful information. Therefore, the Python 3.7.1 software is used to normalize the feature with Min-Max scaling. Finally, to remove the irrelevant features to prevent over fitting and enhance the robustness of the model, an ANOVA was used to perform a univariate analysis of each radiomics feature, and the radiomics feature that had the most significant impact was selected.

Establishment and Verification of the Model

First, in the training set, the statistically significant differences in the radiographic features between the two groups were analyzed by a univariate analysis, and a multivariate analysis was performed to confirm the independent radiographic predictors that can predict the growth rate of pulmonary nodules. A fast, distributed and high-performance gradient promotion framework (Light Gradient Boosting Machine, LightGBM) based on a decision tree algorithm was used to build a radiomics-based model. Finally, the AUC value was used to evaluate the effectiveness of the model.

Statistical Analysis

All statistical analyses were performed using R statistical software (version 3.4.3; http://www.Rproject.org) and a commercially available software program (SPSS 23.0 for Windows; SPSS, Chicago, IL, USA), and the qualitative data are described as n (%). A chi square test or Fisher's exact test was used to analyze the categorical variables, and an independent-samples t-test or Kruskal Wallis test was used to analyze the continuous variables. The ROC curve, area under curve (AUC) and precision recall (P-R) curve were used to evaluate the predictive effectiveness of the model, and $P<0.05$ indicated statistical significance. The integrated discrimination improvement (IDI) was used to

TABLE 1 | CT scanning parameters.

	GE Discovery CT750 HD	Light Speed VCT	Somatom Definition Flash	Somatom Sensation 16
Tube voltage	120 kVp	120 kVp	120 kVp	120 kVp
Tube current	200 mA	200 mA	200 mA	200 mA
Pitch	0.984:1	0.984:1	1.0	0.8
Collimation	0.625 mm*64	0.625 mm*64	0.6 mm*64	0.75 mm*16
Rotation time	0.5 s/rot	0.5 s/rot	0.33 s/rot	0.35 s/rot
SFOV	50 cm	50 cm	50 cm	50 cm
Slice thickness of reconstruction	1.25 mm	1.25 mm	1 mm	1/1.5 mm
Slice interval of reconstruction	1.25 mm	1.25 mm	1 mm	1/1.5 mm
Reconstruction algorithm	STND	STND	Medium sharp	Medium sharp
Number of nodules	103	138	67	99

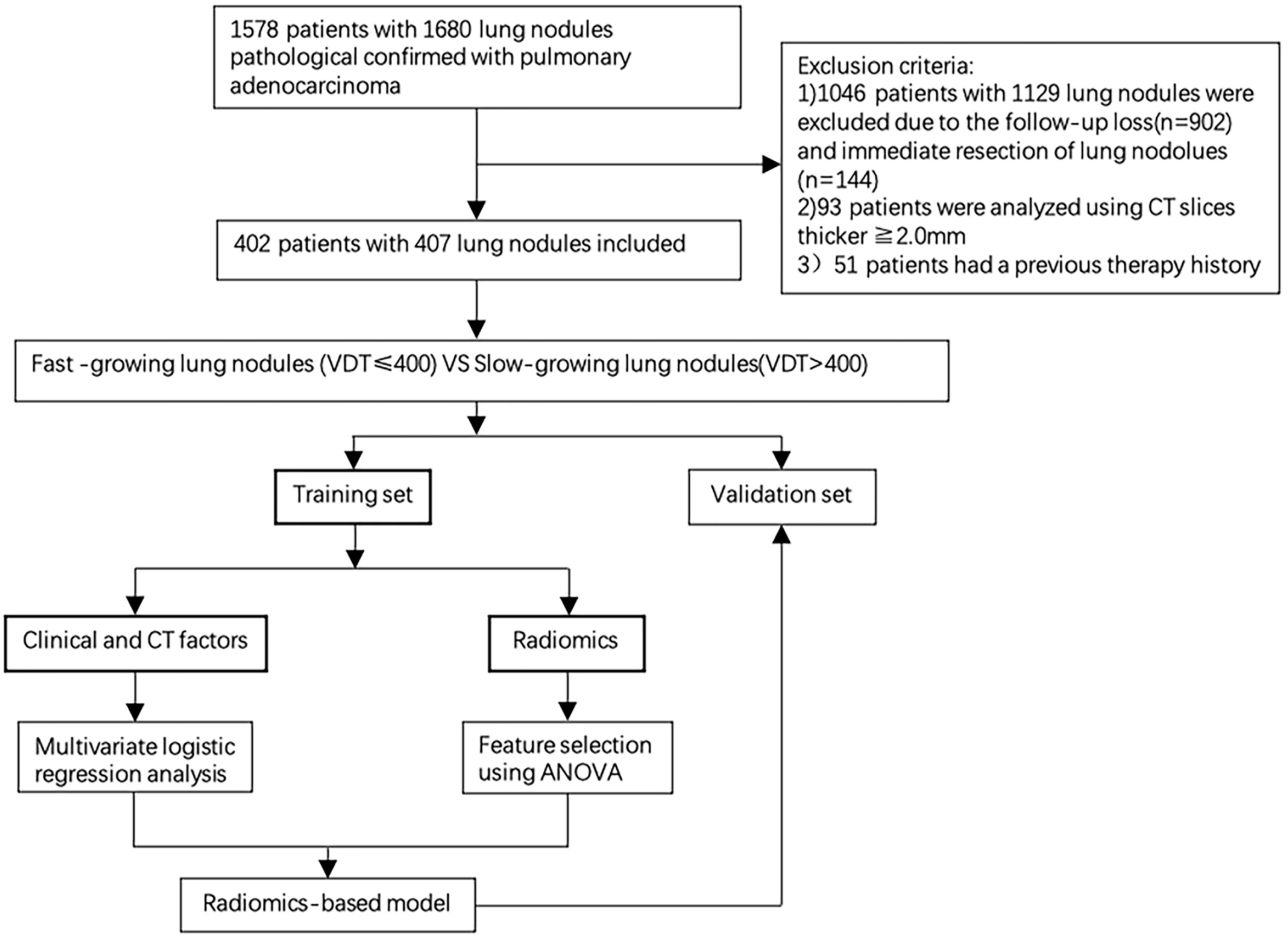


FIGURE 1 | The workflow of the study.

measure the difference in model performance (24), and an IDI>0 indicates that the performance of the model is improved.

RESULTS

Comparison of Radiographic Features Between Slow- and Fast-Growing Nodules

In total, 407 pulmonary nodules from 402 patients with two or more CT images were collected from January 2012 to June 2018, including 41 AIS 42, 158 MIA and 207 IPA. According to the VDTs of the nodules, the 407 nodules were divided into the following two groups: fast-growing nodules (n=77, 18.9%; average VDT=221.78 days) and slow-growing nodules (n=330, 81.1%; average VDT=1722.21 days). In total, 325 cases (fast-growing nodules: n=61; slow-growing nodules: n=264) and 82 cases (fast-growing nodules: n=16; slow-growing nodules: n=66) were randomly divided into the training sets and validation set at a ratio of 8:2. There was no statistically significant difference in the distribution of the clinical and radiographic features between the training sets and validation set, except for the nodule shape and bronchiole change. However, after multivariate analysis in the training set, the two factors were not included in the subsequent analysis. The radiographic features evaluated by the two chest radiologists were almost perfect agreement (unweighted kappa-values, 0.898). The complete patient profiles in the two sets are shown in **Table 2**.

Feature Selection and Establishment of the Radiomics Model

In total, 575 robust radiomics features (ICC: 0.8002-0.9777) were selected for the follow-up analysis. In total, 62 radiomics features (the first 10 radiomics features are shown in **Table 3**) were selected. The details of 62 radiomics features were included in the **Supplement Data**.

In the univariate analysis, four CT factors were statistically significantly associated with the growth rate of pulmonary nodules (**Table 4**).Then, According to the results of the multivariate analysis, the nodule type (pure GGN, part GGN, or solid nodules) was ultimately identified as an independent risk

TABLE 2 | Patient information for the training and validation sets.

Demographic and Clinical Characteristic	Training set (n=325)	Validation set (n=82)	*p*
Sex			0.385
Male	193 (59.4%)	53 (64.6%)	
Female	132 (40.6%)	29 (35.4%)	
Age (years)	57.575±11.652	57.939±11.048	0.532
Size (cm)	0.892±0.498	0.890±0.505	0.928
Location			0.882
Right upper lobe	90 (27.7%)	25 (30.5%)	
Right middle lobe	51 (15.7%)	15 (18.3%)	
Right lower lobe	93 (28.6%)	21 (25.6%)	
Left lower lobe	28 (8.6%)	5 (6.1%)	
Left lower lobe	63 (19.4%)	16 (19.5%)	
Smoking history			0.531
Never smoker	22 (6.8%)	4 (4.9%)	
Current or former smoker	302 (93.2%)	78 (95.1%)	
Family history of cancer			1.000
Present	5 (1.5%)	1 (1.2%)	
Absent	320 (98.5%)	81 (98.8%)	
Margin			0.116
Clear	64 (19.7%)	10 (12.2%)	
Blurred	261 (80.3%)	72 (87.8%)	
Type			0.849
Pure ground glass	130 (40.0%)	30 (36.6%)	
Partial solid nodule	130 (40.0%)	35 (42.7%)	
Solid nodule	65 (20.0%)	17 (20.7%)	
Shape			0.025*
Round	137 (42.2%)	30 (36.6%)	
Oval	105 (32.3%)	39 (47.6%)	
Irregular	83 (25.5%)	13 (15.9%)	
Pleural attachment			0.446
Present	72 (22.2%)	15 (18.3%)	
Absent	253 (77.8%)	67 (81.7%)	
Bubble			0.827
Present	33 (10.2%)	9 (11.0%)	
Absent	292 (89.8%)	73 (89.0%)	
Bronchiole change			0.014*
Present	37 (11.4%)	2 (2.4%)	
Absent	288 (88.6%)	80 (97.6%)	
Vascular change			0.164
Present	37 (11.4%)	14 (17.1%)	
Absent	288 (88.6%)	68 (82.9%)	
Lobulation			0.419
Present	55 (16.9%)	17 (20.7%)	
Absent	270 (83.1%)	65 (79.3%)	
Growth rate			0.878
Fast-growing nodules	61 (18.8%)	16 (19.5%)	
Slow-growing nodules	264 (81.2%)	66 (80.5%)	

Age and size are shown as the mean ± standard deviation; other data are shown as the number of patients, with the percentage in parentheses. The P value is derived from the univariate association analyses between clinical parameters and the growth rate of pulmonary nodules.
**p value < 0.05.*

factor for the prediction of the growth rate of nodules (**Table 5**) and was included in the model establishment.

Evaluation of the Predictive Effectiveness of the Radiomics Models

In the training set, the AUCs of the radiographic model, the radiomics model and the combined radiographic-radiomics model were 0.717 (95%CI: 0.683-0.754), 0.876 (95%CI: 0.855-0.898) and 0.903 (95%CI: 0.884-0.924), respectively (**Figure 2**). And the recall of the radiographic model, the radiomics model and the combined radiographic-radiomics model were 60.7% (precision 71.7%), 85.3% (precision 81.5%) and 80.3% (precision 83.7%), respectively. In the validation set, the AUCs of the radiographic model, the radiomics model and the combined radiographic-radiomics model were 0.727 (95%CI: 0.663-0.792), 0.710 (95%CI: 0.638-0.754) and 0.778 (95%CI: 0.713-0.844), respectively (**Figure 3**). And the recall of the radiographic model, the radiomics model and the combined radiographic-radiomics model were 43.8% (precision 80.5%), 62.5% (precision 74.4%) and 87.5% (precision 65.6%). Compared with the radiographic model, the IDI value of the combined radiographic-radiomics model was 0.086(versus 0), and the difference was statistically significant (z statistic, P<0.05). Compared with the radiomics model, the IDI value of the combined model was 0.094 (versus 0), indicating a statistically significant difference (z statistic, P<0.05). Finally, the results showed that the growth rate of early lung adenocarcinoma could be predicted more effectively by the combined model.

TABLE 3 | Top 10 imaging features after feature screening.

Class	Feature name
First-order features	wavelet-HHH_firstorder_Variance
Gray-level cooccurrence matrix	wavelet-LHH_glcm_ClusterProminence
	wavelet-HHH_glcm_Contrast
	wavelet-HHH_glcm_ClusterTendency'
	wavelet-HHH_glcm_ClusterProminence
	wavelet-HHH_glcm_DifferenceVariance
	wavelet-HHH_glcm_SumSquares
Gray-level run lengths matrix	wavelet-HHH_glrlm_GrayLevelVariance
Gray level dependence matrix	wavelet-HHH_gldm_SmallDependenceHighGrayLevelEmphasis
	wavelet-HHH_gldm_GrayLevelVariance

DISCUSSION

In this study, 407 lung nodules were followed up by chest CT examinations over a long period, and the value of the radiomics features and radiographic features in predicting the growth rate of early-stage lung adenocarcinoma was analyzed. Finally, we established a combined radiographic-radiomics model that can better predict the growth rate of lung adenocarcinoma, with an AUC of 0.778 (95%CI: 0.713-0.844) and a recall of 87.5% (precision 65.9%).

Previously, the prediction of the growth of pulmonary nodules mainly depended on radiographic features (25). However, these CT features are subjective and inaccurate. In this study, we not only analyzed the radiographic features but also combined radiomics features to include more information and establish a more convenient and feasible prediction model, and the model finally achieved good performance.

In this study, the univariate analysis of the clinical data and CT features showed that there were statistically significant

TABLE 4 | Comparison of fast-growing and slow-growing cases in the training set.

Demographic and Clinical Characteristic	Fast-growing nodules (n=61)	Slow-growing nodules(n=264)	P
Sex			0.351
Male	33 (54.1%)	160 (60.6%)	
Female	28 (45.9%)	104 (39.4%)	
Age (years)	56.885±11.986	58.966±11.562	0.238
Size (cm)	0.997±0.529	0.868±0.489	0.030*
Location			0.297
Right upper lobe	14 (23.0%)	76 (28.8%)	
Right middle lobe	7 (11.5%)	44 (16.7%)	
Right lower lobe	18 (29.5%)	75 (28.4%)	
Left lower lobe	9 (14.8%)	19 (7.2%)	
Left lower lobe	13 (21.3%)	50 (18.9%)	
Smoking history			0.180
Never smoker	7 (11.5%)	15 (5.7%)	
Current or former smoker	54 (88.5%)	249 (94.3%)	
Family history of cancer			0.943
Present	1 (1.6%)	4 (1.5%)	
Absent	60 (98.4%)	260 (98.5)	
Margin			0.154
Clear	16 (26.2%)	48 (18.2%)	
Blurred	45 (73.8%)	216 (81.8%)	
Type			0.001*
Pure ground glass	17 (27.9%)	113 (42.8%)	
Partial solid nodule	21 (34.4%)	109 (41.3%)	
Solid nodule	23 (37.7%)	42 (15.9%)	
Shape			0.036*
Round	17 (27.9%)	120 (45.5%)	
Oval	23 (37.7%)	82 (31.1%)	
Irregular	21 (34.4%)	62 (23.5%)	
Pleural retraction			0.605
Present	12 (19.7%)	60 (22.7%)	
Absent	49 (80.3%)	204 (77.3%)	
Bubble lucency			0.396
Present	8 (13.1%)	25 (9.5%)	
Absent	53 (86.9%)	239 (90.5%)	
Bronchiole change			0.384
Present	5 (8.2%)	32 (12.1%)	
Absent	56 (91.8%)	232 (87.9%)	
Vascular change			0.358
Present	5 (8.2%)	32 (12.1%)	
Absent	56 (91.8%)	232 (87.9%)	
Lobulation			0.011*
Present	17 (27.9%)	38 (14.4%)	
Absent	44 (72.1%)	226 (85.6%)	

Age and size are shown as the mean ± standard deviation; other data are shown as the number of patients, with the percentage in parentheses. The P value is derived from the univariate association analyses between clinical parameters and the growth rate of pulmonary nodules.
**p value < 0.05.*

differences in the size, type, shape and lobulation of nodules between the fast-growing and slow-growing nodules ($P<0.05$). However, after the multivariate analysis, only the type (pure GGN, part GGN or solid nodules) significantly differed between the two groups. In total, 77 fast-growing pulmonary nodules were identified, including 19 cases (24.7%) of pure GGN, 28 cases (36.3%) of partial GGN, and 30 cases (39.0%) of solid nodules. From this result, we can get a general conclusion that aggressive nodules may be more commonly observed in patients with solid

TABLE 5 | Multivariate analysis of the radiographic features.

Characteristic	OR (95% CI)	*p*
Size	0.992 (0.932-1.056)	0.806
Type	1.701 (1.112-2.603)	0.014
Shape	1.319 (0.903-1.926)	0.152
Lobulation	1.405 (0.653-3.021)	0.384

OR, odds ratio; CI, confidence interval.

nodules and have a lower VDT; these results are consistent with previous studies (26, 27). Additionally, Oda S et al. found that the growth rate of solid nodules is usually faster than that of part GGNs and pure GGNs (26, 28). We speculate that this finding may be due to most solid components of malignant nodules being mainly proliferative stacks of tumor cells. In this study, although the size, shape and lobulation of the nodules in the univariate analysis were statistically significantly different, these features can predict the growth rate of pulmonary nodules remains to be determined. Kobayashi et al. (29) reported that the nodule size may be robustly associated with the growth of pulmonary nodules; however, this study did find the correlation after multivariate analysis, the reasons may include the following aspects: first, this study only used a univariate analysis and should not be considered authentic. Second, this study used manual measurement of the nodule size to judge the growth of pulmonary nodules; however, the presence of subjective factors, such as inaccurate measurement and error, require a more rigorous analysis.

Radiomics is widely used in clinical research because it can excavate a large amount of invisible information with great clinical value and has achieved good results (20, 30). Due to the instability of the features, a repeatability analysis was carried out. Most studies are based on the assumption that the features are redundant, thus reducing the number of radiomics features used in modeling research. However, each feature may affect the training accuracy to a certain extent. Therefore, to avoid overfitting the model and enhance the robustness, this study adopted data normalization and an ANOVA, and 62 useful features were finally selected for model establishment. According to the theory of radiomics, the features we extracted can reflect the spatial heterogeneity, microenvironment and gene expression of tumors (31). LightGBM is a newly developed gradient lifting algorithm framework based on a decision tree algorithm that can further optimize the tree model and is more conducive to clinical application (32). Therefore, we established a combined radiographic-radiomics model that can better distinguish the growth rate of pulmonary nodules and proved the feasibility of the modeling method. Yooh et al. (33) evaluated 52 lung adenocarcioma patients and demonstrated the potential of margin-related radiomics feature to predict tumor doubling times in lung adenocarcinoma, however, we finally decided that dividing the lung nodules into slow- and fast-growing nodules according to the VDT, because it can be more conducive to the management of nodules in clinical, and has the potential to develop a personalized follow-up strategy for patients (19).

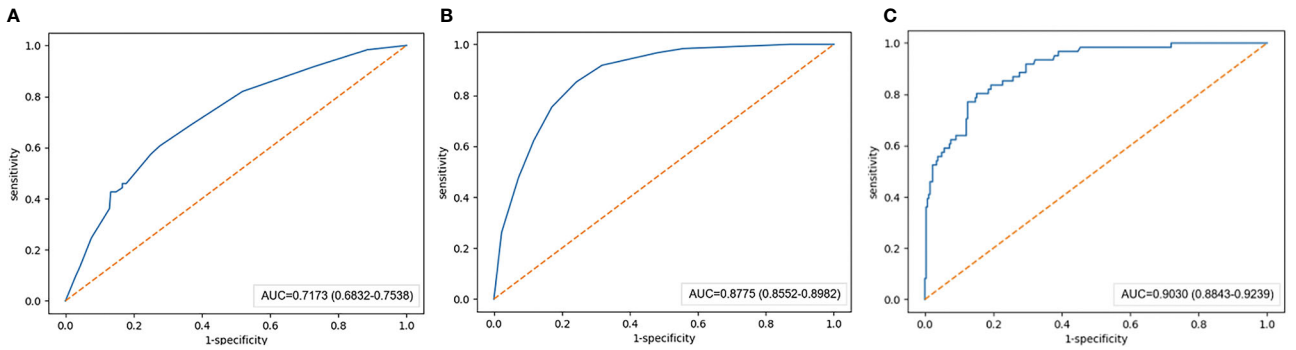


FIGURE 2 | AUC value of the training set. **(A)** The AUC value of the radiographic model with the training set is 0.717. **(B)** The AUC value of the radiomics model with the training set is 0.876. **(C)** The AUC value of the combined radiographic-radiomics model with the training set is 0.903.

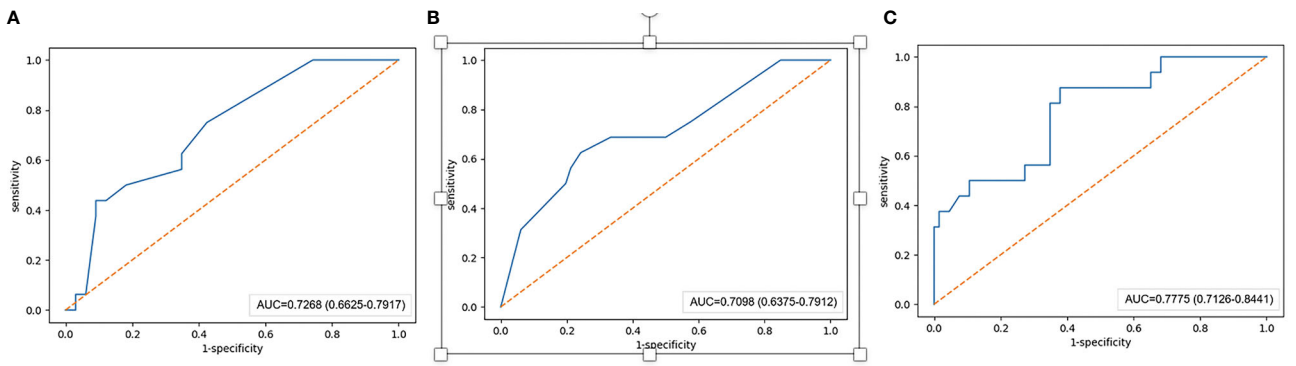


FIGURE 3 | AUC value of the validation set. **(A–C)** The AUC value of the radiographic model with the validation set is 0.727. **(B)** The AUC value of the radiomics model with the validation set is 0.710. **(C)** The AUC value of the combined radiographic-radiomics model with the validation set is 0.778.

There are certainly some underlying limitations to this study. First, this study was a single-center retrospective study that lacked external validation data, and prospective studies are needed to assess the robustness and practical clinical value of the combined model. Second, due to the different reconstruction cores, the CT acquisition protocol is not standardized, which may have had a potential impact on the extraction of the radiomics features. However, all images included in the current study were thin-slice CT image to minimize these variabilities (34). Third, the model achieved good results with the training set but lower results in the validation set, which may be caused by the limited amounts of samples in the study. In the future, we aim to improve the efficacy and feasibility of this model through multicenter data, standardized CT scanning parameters and a larger sample.

In summary, this study analyzed the value of CT features and radiomics features in the diagnosis of lung adenocarcinoma. Considering its practicability and accuracy, we established a combined model that can better predict the growth rate of pulmonary nodules and assist in clinical decision-making.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethic Committee of the Huadong Hospital Affiliated with Fudan University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

ML and YS conceived the presented idea. LJ collected the data. LJ, YW, PG, LT, KK, and XH analyzed the data. MT and WM drafted the manuscript. All authors reviewed the manuscript,

and ML made corrections to the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Nature Science Foundation of China 61976238 (ML), "Future Star" of famous doctors' training plan of Fudan University, Science and Technology Planning Project of Shanghai Science and Technology Commission 20Y11902900 (ML). Medical Imaging Key Program of Wise Information Technology of 120, Health Commission of Shanghai 2018ZHYL0103 (ML).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.658138/full#supplementary-material

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492
2. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/American thoracic society/European respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* (2011) 6:244–85. doi: 10.1097/JTO.0b013e318206a221
3. Yim J, Zhu LC, Chiriboga L, Watson HN, Goldberg JD, Moreira AL. Histologic features are important prognostic indicators in early stages lung adenocarcinomas. *Mod Pathol* (2007) 20:233–41. doi: 10.1038/modpathol.3800734
4. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med* (2011) 365:395–409. doi: 10.1056/NEJMoa1102873
5. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Müller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology* (2008) 246:697–722. doi: 10.1148/radiol.2462070712
6. Wiener RS, Gould MK, Arenberg DA, Au DH, Fennig K, Lamb CR, et al. An official American thoracic society/American college of chest physicians policy statement: implementation of low-dose computed tomography lung cancer screening programs in clinical practice. *Am J Respir Crit Care Med* (2015) 192:881–91. doi: 10.1164/rccm.201508-1671ST
7. Wood DE, Kazerooni EA, Baum SL, Eapen GA, Ettinger DS, Hou L, et al. Lung cancer screening, version 3.2018, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw* (2018) 16:412–41. doi: 10.6004/jnccn.2018.0020
8. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the fleischner society 2017. *Radiology* (2017) 284:228–43. doi: 10.1148/radiol.2017161659
9. Rudin CM, Giaccone G, Ismaila N. Treatment of Small-Cell Lung Cancer: American Society of Clinical Oncology Endorsement of the American College of Chest Physicians Guideline. *J Oncol Pract* (2016) 12:83–6. doi: 10.1200/jop.2015.008201
10. Bai C, Choi CM, Chu CM, Anantham D, Chung-Man Ho J, Khan AZ, et al. Evaluation of pulmonary nodules: clinical practice consensus guidelines for Asia. *Chest* (2016) 150:877–93. doi: 10.1016/j.chest.2016.02.650
11. Henschke CI, Yankelevitz DF, Yip R, Reeves AP, Farooqi A, Xu D, et al. Lung cancers diagnosed at annual CT screening: volume doubling times. *Radiology* (2012) 263:578–83. doi: 10.1148/radiol.12102489
12. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
13. Zhou M, Xiong X, Drummer D, Jiang B. Molecular dynamics simulation and experimental investigation of the geometrical morphology development of injection-molded nanopillars on polymethylmethacrylate surface. *Comput Mater Sci* (2018) 149:208–16. doi: 10.1016/j.commatsci.2018.03.035
14. Zhao W, Xu Y, Yang Z, Sun Y, Li C, Jin L, et al. Development and validation of a radiomics nomogram for identifying invasiveness of pulmonary adenocarcinomas appearing as subcentimeter ground-glass opacity nodules. *Eur J Radiol* (2019) 112:161–8. doi: 10.1016/j.ejrad.2019.01.021
15. Park S, Lee SM, Kim S, Lee JG, Choi S, Do KH, et al. Volume doubling times of lung adenocarcinomas: correlation with predominant histologic subtypes and prognosis. *Radiology* (2020) 295:703–12. doi: 10.1148/radiol.2020191835
16. Henschke CI, Yankelevitz DF, Mirtcheva R, McGuinness G, McCauley D, Miettinen OS. CT screening for lung cancer: frequency and significance of part-solid and nonsolid nodules. *AJR Am J Roentgenol* (2002) 178:1053–7. doi: 10.2214/ajr.178.5.1781053
17. Gao F, Li M, Ge X, Zheng X, Ren Q, Chen Y, et al. Multi-detector spiral CT study of the relationships between pulmonary ground-glass nodules and blood vessels. *Eur Radiol* (2013) 23:3271–7. doi: 10.1007/s00330-013-2954-3
18. Revel MP. Avoiding overdiagnosis in lung cancer screening: the volume doubling time strategy. *Eur Respir J* (2013) 42:1459–63. doi: 10.1183/09031936.00157713
19. Heuvelmans MA, Oudkerk M, de Bock GH, de Koning HJ, Xie X, van Ooijen PM, et al. Optimisation of volume-doubling time cutoff for fast-growing lung nodules in CT lung cancer screening reduces false-positive referrals. *Eur Radiol* (2013) 23:1836–45. doi: 10.1007/s00330-013-2799-9
20. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. *Radiology* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
21. Heuvelmans MA, Vliegenthart R, de Koning HJ, Groen HJM, van Putten M, Yousaf-Khan U, et al. Quantification of growth patterns of screen-detected lung cancers: The NELSON study. *Lung Cancer* (2017) 108:48–54. doi: 10.1016/j.lungcan.2017.02.021
22. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* (1977) 33:159–74. doi: 10.2307/2529310
24. Pencina MJ, D'Agostino RB, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* (2008) 27:157–72. doi: 10.1002/sim.2929
25. Zhang L, Yankelevitz DF, Carter D, Henschke CI, Yip R, Reeves AP. Internal growth of nonsolid lung nodules: radiologic-pathologic correlation. *Radiology* (2012) 263:279–86. doi: 10.1148/radiol.11101372
26. Hasegawa M, Sone S, Takashima S, Li F, Yang ZG, Maruyama Y, et al. Growth rate of small lung cancers detected on mass CT screening. *Br J Radiol* (2000) 73:1252–9. doi: 10.1259/bjr.73.876.11205667
27. Chang B, Hwang JH, Choi YH, Chung MP, Kim H, Kwon OJ, et al. Natural history of pure ground-glass opacity lung nodules detected by low-dose CT scan. *Chest* (2013) 143:172–8. doi: 10.1378/chest.11-2501
28. Oda S, Awai K, Murao K, Ozawa A, Yanaga Y, Kawanaka K, et al. Computer-aided volumetry of pulmonary nodules exhibiting ground-glass opacity at MDCT. *AJR Am J Roentgenol* (2010) 194:398–406. doi: 10.2214/ajr.09.2583
29. Kobayashi Y, Sakao Y, Deshpande GA, Fukui T, Mizuno T, Kuroda H, et al. The association between baseline clinical-radiological characteristics and growth of pulmonary nodules with ground-glass opacity. *Lung Cancer* (2014) 83:61–6. doi: 10.1016/j.lungcan.2013.10.017

30. Lee G, Lee HY, Park H, Schiebler ML, van Beek EJR, Ohno Y, et al. Radiomics and its emerging role in lung cancer research, imaging biomarkers and clinical management: state of the art. *Eur J Radiol* (2017) 86:297–307. doi: 10.1016/j.ejrad.2016.09.005
31. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
32. Gurm HS, Seth M, Kooiman J, Share D. A novel tool for reliable and accurate prediction of renal complications in patients undergoing percutaneous coronary intervention. *J Am Coll Cardiol* (2013) 61:2242–8. doi: 10.1016/j.jacc.2013.03.026
33. Yoon HJ, Park H, Lee HY, Sohn I, Ahn J, Lee SH. Prediction of tumor doubling time of lung adenocarcinoma using radiomic margin characteristics. *Thorac Cancer* (2020) 11:2600–9. doi: 10.1111/1759-7714.13580
34. He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep* (2016) 6:34921. doi: 10.1038/srep34921

Conflict of Interest: KK was employed by Dianei Technology, Shanghai.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Tan, Ma, Sun, Gao, Huang, Lu, Chen, Wu, Jin, Tang, Kuang and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 17 May 2021
doi: 10.3389/fonc.2021.677814

Development of CT-Based Imaging Signature for Preoperative Prediction of Invasive Behavior in Pancreatic Solid Pseudopapillary Neoplasm

*Wen-peng Huang[1], Si-yun Liu[2], Yi-jing Han[1], Li-ming Li[1], Pan Liang[1] and Jian-bo Gao[1]**

[1] Department of Radiology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, [2] Pharmaceutical Diagnostics, General Electric Company (GE) Healthcare, Beijing, China

Purpose: It is challenging for traditional CT signs to predict invasiveness of pancreatic solid pseudopapillary neoplasm (pSPN). We aim to develop and evaluate CT-based radiomics signature to preoperatively predict invasive behavior in pSPN.

Methods: Eighty-five patients who had pathologically confirmed pSPN and preoperative contrasted-enhanced CT imaging in our hospital were retrospectively analyzed (invasive: 24; non-invasive: 61). 1316 radiomics features were separately extracted from delineated 2D or 3D ROIs in arterial and venous phases. 200% (SMOTE) was used to generate balanced dataset (invasive: 72, non-invasive: 96) for each phase, which was for feature selection and modeling. The model was internally validated in the original dataset. Inter-observer consistency analysis, spearman correlation, univariate analysis, LASSO regression and backward stepwise logical regression were mainly applied to screen the features, and 6 logistic regression models were established based on multi-phase features from 2D or 3D segmentations. The ROC analysis and Delong's test were mainly used for model assessment and AUC comparison.

Results: It retained 11, 8, 7 and 7 features to construct 3D-arterial, 3D-venous, 2D-arterial and 2D-venous model. Based on 3D ROIs, the arterial model (AUC: 0.914) performed better than venous (AUC: 0.815) and the arterial-venous combined model was slightly improved (AUC: 0.918). Based on 2D ROIs, the arterial model (AUC: 0.814) performed better than venous (AUC:0.768), while the arterial-venous combined model (AUC:0.893) performed better than any single-phase model. In addition, the 3D arterial model performed better than the best combined 2D model. The Delong's test showed that the significant difference of model AUC existed in arterial models in original dataset ($p = 0.019$) while not in arterial-venous combined model ($p=0.49$) as comparing 2D and 3D ROIs.

Conclusion: The arterial radiomics model constructed by 3D-ROI feature is potential to predict the invasiveness of pSPN preoperatively.

Keywords: pancreatic solid pseudopapillary neoplasm, computed tomography, invasiveness, radiomics, diagnosis

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln, United Kingdom

Reviewed by:
Zhichao Li,
Chongqing West District Hospital, China
Qijun Shen,
Zhejiang University, China

****Correspondence:***
Jian-bo Gao
jianbogaochina@163.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *08 March 2021*
Accepted: *27 April 2021*
Published: *17 May 2021*

Citation:
Huang W-p, Liu S-y, Han Y-j, Li L-m, Liang P and Gao J-b (2021) Development of CT-Based Imaging Signature for Preoperative Prediction of Invasive Behavior in Pancreatic Solid Pseudopapillary Neoplasm. Front. Oncol. 11:677814. doi: 10.3389/fonc.2021.677814

INTRODUCTION

Pancreatic solid pseudopapillary neoplasm (pSPN) is a rare low-grade malignant tumor, accounting for 1%–2% of pancreatic exocrine tumors. It usually occurs in women under the age of 40. Its clinical manifestations are not typical. Most of them come to the hospital with asymptomatic physical examination, abdominal pain or touching abdominal mass as the main complaint. Laboratory examination is of little help in its diagnosis, and the final diagnosis depends on the immunohistochemical results of postoperative pathology (1–3). A 2018 Chinese multicenter retrospective study showed that pSPN accounted for 31.7% of all resected pancreatic cystic tumors (4). According to the classification criteria of digestive system tumors of the World Health Organization, SPN is defined as invasive when the tumor obviously breaks through the capsule or invades the peripancreatic tissue, surrounding organs and blood vessels, vascular invasion, peripheral nerve invasion, lymph node metastasis and distant metastasis (5, 6). Surgery is the only treatment for patients with pSPN, but the traditional radical resection of pancreatic malignant tumor is more traumatic, which is easy to cause postoperative pancreatic secretion insufficiency and high risk. At present, clinicians tend to take smaller surgical methods for patients with pSPN. Non-invasive pSPN is mainly enucleation of the tumor as a whole, and the prognosis is good, but the scope of resection of invasive pSPN is larger, and incomplete resection of the tumor may lead to recurrence and metastasis (7). Gao et al. (1). studies have shown that positive incisal margin increases the risk of postoperative pSPN recurrence, so accurate preoperative judgment of the invasiveness of pSPN is a key factor in making clinical operation plans. However, it is often difficult to obtain pathological results before operation. the pathological diagnosis of puncture biopsy is restricted by the quality and quantity of samples, which can't accurately reflect the heterogeneity of tumor, and the operation of puncture biopsy may cause tumor cells to spread along the needle path. This makes it difficult and controversial for surgeons to choose the mode of operation. Because of its high popularization rate, convenience and few contraindications, CT has become the first choice for pancreatic diseases, the preoperative diagnosis and evaluation of invasiveness of pSPN depend to a large extent on the imaging features of tumors. Some relevant scholars have analyzed the relationship between imaging features and invasiveness of pSPN, but the imaging data of patients in multiple studies are not all analyzed by the same radiologist and concluded that there may be some subjective differences, and the results are not the same. It is controversial to predict the invasiveness of solid pseudopapillary tumor of the pancreas only from CT signs, so it is necessary to explore reliable features to evaluate tumor invasiveness before operation.

Radiomics technique using various automatically extracted data characterization algorithms converts images into a high dimensional mineable feature space (8–10). Numerous studies have applied the emerging radiomics technique to improve diagnostic, identification, prognostic, and predictive accuracy of cancer research (11–14). Some scholars also try to apply radiomics in pancreatic tumor studies, such as malignancy prediction (15), histopathologic characteristics discrimination (16), vascular invasion prediction (17), prognosis prediction (18), and radiogenomics for genetic status prediction (19). However, to the best of our knowledge, there is no literature that has determined whether a radiomics signature derived from CT images would enable superior prediction of invasive behavior in patients with pSPN.

Considering the radiomics feature could be extracted from the single cross section (two dimensional, 2D) or multi-slices (three dimensional, 3D) of the tumor in CT images, the reported radiomics-based pancreatic cancer studies have either applied 2D segmentation (20) or 3D whole-tumor segmentation (14, 21–25). However, whether to select 2D regions of interest (ROIs) or 3D ROIs still remains unclear for invasive behavior prediction in pSPN. In addition, the previous studies also have shown that there is controversy between 2D and 3D radiomics analysis in tumor diagnosis or prognosis (26–30).

In this work, we proposed a CT radiomics-based classification method by considering the performance of 3D or 2D segmentation and multiple CT imaging phases to discriminate invasiveness and non-invasiveness pSPN. The developed CT imaging signature might help treatment decision-making, especially the choice of operation.

MATERIALS AND METHODS

Patient Selection

With institutional review board approval and waiver of the written informed consent, we retrospectively collected 85 patients with pSPN diagnosed by postoperative pathology from January 2012 to April 2020. The patient enrollment criteria included:1) the patient had no history of other malignant tumors before admission; 2) all patients with pSPN underwent surgery and the CT imaging data were complete; 3) abdominal CT plain scan and enhanced examination were performed within 30 days before operation; 4) the lesion covers at least 3 slices on CT cross section, and the maximum plane diameter is not less than 20mm. Exclusion criteria included:1) insufficient data of pathological diagnosis; 2) the patient had been punctured or treated with related tumor before CT examination; 3) poor CT image quality or lack of raw DICOM data; 4) there are a large number of ascites, pancreas or other lesions around the pancreas that cannot be divided. The flow chart of inclusion and exclusion of 85 patients is shown in **Figure 1**.

This cohort contained both screening and symptomatic cases (i.e., 48 and 37 cases respectively). There were 85 lumps on 85 cases, of which 24 tumor masses were invasive and 61 were non-invasive demonstrated by postoperative pathology. Patient characteristics in the invasive and non- invasive cohorts are given in **Table 1**.

CT Image Acquisition

All patients underwent contrast-enhanced CT scan and informed consent forms were signed before inspection. The CT scans were acquired with a 64-row CT scanner (Discovery CT 750 HD, GE Healthcare, Waukesha, WI, United States) or a dual source CT scanner (Somatom Definition Flash, S+iemens Healthineers, Forchheim, Germany). Conventional axial scanning was

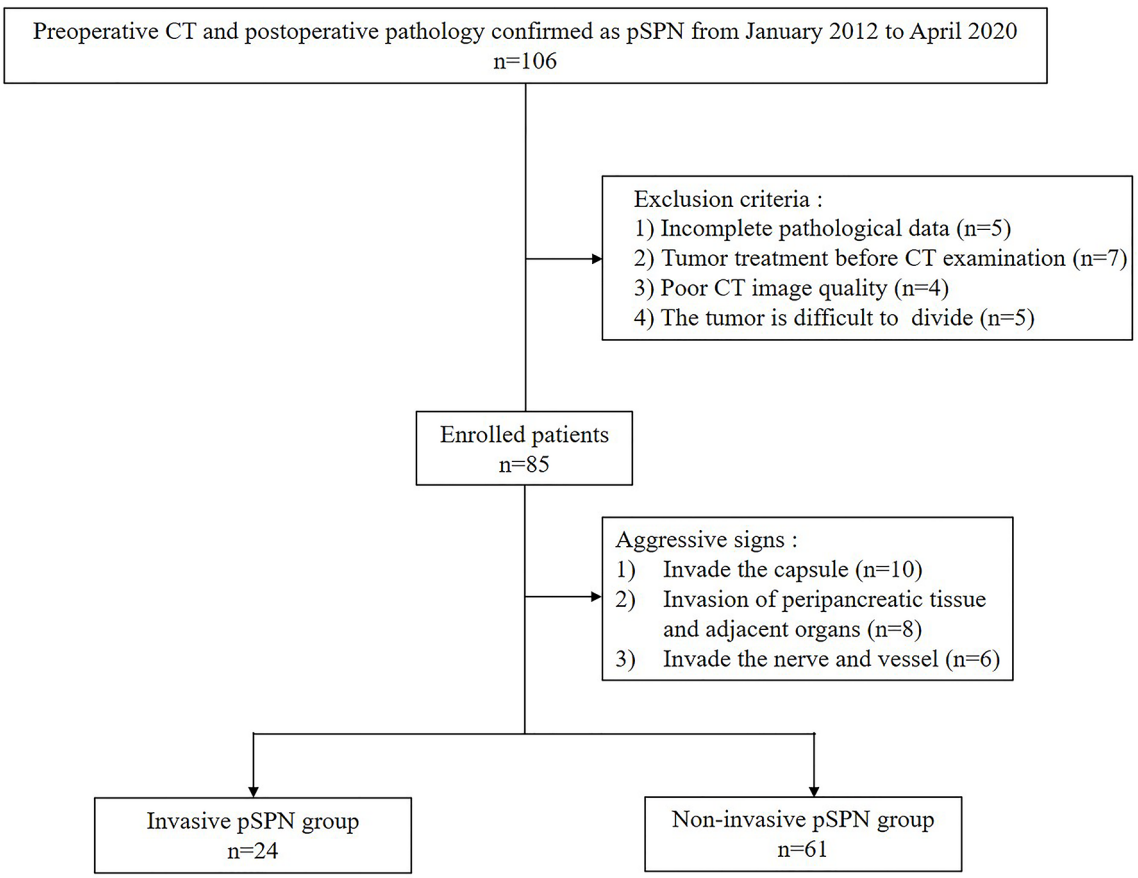


FIGURE 1 | The patient enrollment workflow.

performed before and after an intravenous (i.v.) injection of nonionic iohexol (iopromide, 370 mg/mL, GE Medical Systems, 1.5 mL/kg and 3 mL/s) through a dual-head pump injector (Medrad, Warrendale, PA, United States). The scanning parameters were as follows: tube voltage, 120 kV; automatic mA technology is used for tube current; field of view (FOV), 500 mm; matrix, 512 × 512 mm; slice thickness, 0.625 mm to 5 mm; scan spacing, 0.625 mm to 5 mm. Finally, a 20-mL saline flush was performed at a rate of 3 mL/s. Using low-dose trigger technique, when the descending aorta reached 100 HU after injection of contrast medium, arterial phase images were collected 10 seconds later, and venous phase images were collected at intervals of 30 seconds.

TABLE 1 | The clinical features of pancreatic solid pseudopapillary neoplasm.

Characteristics		Number of cases	Percentage
Age, mean ± SD (years)	32.12 ± 13.66		
Gender	Male	18	21.95%
	Female	64	78.05%
Clinical symptoms	Abdominal pain	43	52.43%
	Abdominal mass	4	4.88%
	Physical examination	35	42.68%

Image Process and Lesion ROI Segmentation

The CT images in arterial and venous phases were firstly resampled isotropically into 1 mm ×1 mm × 1 mm voxel size by using trilinear interpolation, to reduce the heterogeneity resulted from different scanner (24, 25). Then the CT images in respective phases were sequentially imported into A.K. software (Artificial Intelligence Kit, GE Healthcare, version 3.3.0) and the lesions were separately delineated in each imaging phase. 2D segmentation was realized by delineating around the tumor outline for the largest cross-sectional area in the CT axial plane. By conducting slice-by-slice delineation along with the tumor outer contour in the CT axial plane, 3D ROI was automatically merged. Each ROI was outlined by a radiologist (H.WP, 5 years of experience in abdominal imaging diagnosis) and supervised by a radiologist (L.P, 10 years of experience in abdominal imaging diagnosis). All the segmentations were finally accomplished with the consensus of these two radiologists. At the same time, thirty CT image sets were randomly and separately chosen from arterial and venous phase for assessing inter-observer repeatability of radiomics features. The ROIs were outlined by another radiologist (H.YJ, 5 years of experience in abdominal imaging diagnosis) and supervised by another

radiologist (L.LM, 8 years of experience in abdominal imaging diagnosis). The features were then extracted and the features with intra-class correlation coefficient (ICC) greater than 0.75 were retained for further analysis, which meant a good feature reliability (31).

Radiomics Feature Extraction

The radiomics features were automatically extracted by using Python package Pyradiomics (32). And before feature extraction, the CT values included in the ROI were discretized with binWidth = 25 HU (33). 1316 radiomics features were separately extracted from the delineated 2D or 3D ROIs in arterial and venous phases. There were 107 features extracted from the original images including: 32 first-order features (18 intensity statistical and 14 shape features). Among 75 textural features, there were 24 Gray Level Co-occurrence Matrix (GLCM), 16 Gray Level Run Length Matrix (GLRLM), 16 Gray Level Size Zone Matrix (GLSZM), 14 Gray Level Dependence Matrix (GLDM) and 5 Neighboring Gray Tone Difference Matrix (NGTDM) features. By using transformed images, 1209 first-order and textural features were calculated, including 744 wavelet features based on level-1 wavelet decomposition images in three directions and 8 channels of LLL, HHH, LHL, LLH, HLL, HLH, HHL and LHH; 186 Laplacian of Gaussian (LoG) filtered features with sigma 2.0 mm and 3.0 mm and 279 features based on local binary pattern (LBP) filtered images including 2 sets of images based on 2-level spherical harmonics and 1 set of kurtosis image. For each transformed image, the same 6 kinds of features (93 features in total per image) were extracted including: 1) first-order features (18 intensity statistical features); 2) 24 Gray Level Co-occurrence Matrix (GLCM) features; 3) 16 Gray Level Run Length Matrix (GLRLM) features; 4) 16 Gray Level Size Zone Matrix (GLSZM) features; 5) 14 Gray Level Dependence Matrix (GLDM) features; 6) 5 Neighboring Gray Tone Difference Matrix (NGTDM) features. Thirty CT image sets were randomly and separately chosen from arterial and venous phase for assessing inter-observer repeatability of radiomics features.

Feature Selection and Model Construction

The radiomics features extracted from 2D and 3D ROIs in arterial and venous phases were firstly processed. The missing values were firstly replaced with median values. Then z-score normalization was used for standardization. As the class distribution of the original dataset was moderately imbalanced (invasive:24; non-invasive:61; ratio of 2.5), 200% data oversampling based on Synthetic Minority Oversampling Technique (SMOTE)was conducted to obtain equilibrium for class distribution (34, 35). The generated dataset (SMOTE dataset) involve 72 invasive and 96 noninvasive samples respectively (ratio of 1.3), which was used as training set for modeling.

The radiomics models were constructed separately based on 2D and 3D ROIs. While, the same feature selection and modeling methods were applied. The feature selection and final modeling procedure was performed in the SMOTE dataset as follows.

1. The features with agreement coefficient larger than 0.75 during the inter-observer consistency analysis were retained.
2. The features with relatively low variance less than 1.0 were excluded.
3. The features with less collinearity were retained by using correlation analysis at cut-value 0.7.
4. The features with significant difference ($P<0.05$) between invasive and noninvasive groups were selected by using univariate analysis (Mann-Whitney U test or t-test).
5. The least absolute shrinkage and selection operator (LASSO) logistic regression involving 10-fold cross validation was conducted to avoid overfitting (36). The maximum area under the curve (AUC) for model fitting among the 10-folds cross validation was applied to determine the lambda values, at which the remaining features with non-zero coefficients were retained.
6. The retained features after LASSO regression were finally involved into backward stepwise logistic regression with minimum AIC (Akaike Information Criterion) criteria to develop the regression radiomics model and radiomics-derived signature "Radscore" was derived by using the regression coefficients, which could be further transferred into probability by using sigmoid function P (Radscore) = 1/(1+exp (-Radscore)).

Hence, there were four basic radiomics models and corresponding Radscore derived, including the arterial phase model based on 2D ROIs ($Radscore^{AP_2D}$), venous phase model based on 2D ROIs ($Radscore^{VP_2D}$), arterial phase model based on 3D ROIs ($Radscore^{AP_3D}$) and venous phase model based on 3D ROIs ($Radscore^{VP_3D}$). Besides these four basic radiomics models, additional two combined models were also constructed, including the arterial-venous combined model based on 2D ROIs ($Radscore^{AP_VP_2D}$) and arterial-venous combined model based on 3D ROIs ($Radscore^{AP_VP_3D}$).

Evaluation of Model Predictive Performance

The discrimination of the radiomics models were assessed by the receiver operating characteristic (ROC) curve analysis, and the area under the ROC curve (AUC), sensitivity, specificity and accuracy could be derived. In order to validate the constructed model based on the SMOTE dataset, the constructed single or combined model were applied in the original dataset. The regression coefficients and the model cut-off value (when Youden index reached the maximum) derived in the SMOTE dataset were applied in the original dataset. And the AUC, sensitivity, specificity and accuracy in the original dataset could be obtained to validate the model performance. Furthermore, the 1000-times bootstrap was used to assess the optimism and overall performance of radiomics models (37). To investigate the consistency of radiomics model for predicting invasiveness of pSPN in both SMOTE and original datasets, the calibration curves were plotted. Meanwhile, the decision curve analysis (DCA) was also used for assessment of the model clinical usefulness. In order to compare the ROC performance of each same kind of model between 2D and 3D ROIs, or to perform comparison between paired model from different phases, the Delong's test was applied.

Statistical Analysis

Statistical analysis was conducted by R software (version 3.5.3; http://www.r-project.org). Student's t test or Mann-Whitney U

test was used for continuous variables with normal or non-normal distribution (the Shapiro–Wilk test for assessing the normality of distribution) and the categorical variables were tested by Chi-square (or Fisher's exact test). The Delong's test was used for comparison of AUC between each paired model. The statistical significance levels were two-sided with P< 0.05. The following R packages were applied: "DMwR" for SMOTE oversampling; "findCorrelation" in "caret" package for correlation analysis; "glmnet" for logistic regression including LASSO regression algorithm; "pROC" for ROC analysis, and "rmda" for DCA analysis.

RESULTS

Feature Selection and Radiomics Model Development

There were totally 6 radiomics models constructed based on 2D and 3D ROIs in arterial and venous imaging phase.

Arterial Phase Model Based on 3D ROI

By using inter-observer consistency analysis, 825 radiomics features with ICCs>0.75 were retained among 1316 features. After removing features with variance less than 1.0, 409 features were kept. Then, 25 features were retained after correlation analysis by using cut-value 0.7. Among 19 features selected by Mann-Whitney U test, 5 features were further removed by LASSO regression (**Figure S1**). Finally, 11 radiomics features were kept by backward stepwise logistic regression analysis (minimum AIC criteria), and the regression function deriving the $Radscore^{AP_3D}$ was summarized in **Supplementary Equation (S1)**.

Venous Phase Model Based on 3D ROI

By using inter-observer consistency analysis, 810 radiomics features with ICCs>0.75 were retained among 1316 features. After removing features with variance less than 1.0, 385 features were kept. Then, 20 features were retained after correlation analysis by using cut-value 0.7. Among 17 features selected by Mann-Whitney U test, 13 features were retained after LASSO regression (**Figure S2**). Finally, 8 radiomics features were kept by backward stepwise logistic regression analysis (minimum AIC criteria), and the regression function deriving the $Radscore^{VP_3D}$ was summarized in **Supplementary Equation (S2)**.

Arterial Phase Model Based on 2D ROI

By using inter-observer consistency analysis, 1059 radiomics features with ICCs>0.75 were retained among 1316 features. After removing features with variance less than 1.0, 475 features were kept. Then, 23 features were retained after correlation analysis by using cut-value 0.7. Among 12 features selected by Mann-Whitney U test, 8 features were retained after LASSO regression (**Figure S3**). Finally, 7 radiomics features were kept by minimum AIC criteria, and the regression function deriving the $Radscore^{AP_2D}$ was summarized in **Supplementary Equation (S3)**.

Venous Phase Model Based on 2D ROI

By using inter-observer consistency analysis, 1126 radiomics features with ICCs>0.75 were retained among 1316 features. After removing features with variance less than 1.0, 517 features were kept. Then, 32 features were retained after Spearman correlation analysis by using cut-value 0.7. Among 12 features selected by Mann-Whitney U test, no feature was removed after LASSO regression (**Figure S4**). Finally, 7 radiomics features were kept by minimum AIC criteria, and the regression function deriving the $Radscore^{VP_2D}$ was summarized in **Supplementary Equation (S4)**.

Arterial-Venous Combined Model Based on 3D ROI

The derived $Radscore^{AP_3D}$ and $Radscore^{VP_3D}$ were involved directly into multivariate logistic regression to construct the Arterial-Venous combined model based on 3D ROIs and the regression function deriving the $Radscor^{AP_VP_3D}$ was summarized in **Supplementary Equation (S5)**.

Arterial-Venous Combined Model Based on 2D ROI

The derived $Radscore^{AP_2D}$ and $Radscore^{VP_2D}$ were involved directly into multivariate logistic regression to construct the Arterial-Venous combined model based on 2D ROIs and the regression function deriving the $Radscor^{AP_VP_2D}$ was summarized in **Supplementary Equation (S6)**.

A statistically significant difference existed in $Radscore^{AP_3D}$, $Radscore^{VP_3D}$ were $Radscor^{AP_VP_3D}$ between noninvasive and invasive in the original datasets with (-2.85(-4.66, -0.60) vs. 1.90 (1.02, 7.25), $p < 0.001$), (-1.34(-2.65, -0.41) vs. 1.11(-0.44, 2.10), $p < 0.001$), (-2.51(-4.71, -1.02) vs. 2.63(0.92, 8.58), $p < 0.001$). Meanwhile, such significant difference also existed in $Radscore^{AP_3D}$, $Radscore^{VP_3D}$ were $Radscor^{AP_VP_3D}$ between noninvasive and invasive in the original datasets with (-1.24 ± 2.12 vs. 1.01 ± 1.81, $p < 0.001$), (-0.79(-2.15, 0.09) vs. 0.38(-0.34, 1.25), $p < 0.001$), (-2.17(-4.51, -0.27) vs. 2.01(0.53, 4.24), $p < 0.001$). The distribution of each model's Radscore in the invasive and noninvasive and the P values for the statistical difference analysis were also shown in **Figure 2** and their inset.

Radiomics Model Performance

The ROC analysis was used to evaluate the predictive performance of the constructed six models and the ROC curves for each model in SMOTE and original dataset were illustrated in **Figure 3**. The AUC, specificity, sensitivity, accuracy and the determined cut-value for each model performance in SMOTE dataset and the original dataset were summarized in **Tables 2** and **3**. Based on 3D ROIs, the arterial phase model has better performance than venous phase model and the arterial-venous combined model performed slightly better than the others. Based on 2D ROIs, the arterial phase model performed better than venous phase model, while the arterial-venous combined model performed better than any model constructed by independent imaging phase. In addition, the 3D arterial model performed better than the best arterial-venous combined 2D model. The Delong's test result showed that the significant difference of model AUC existed in arterial models in original dataset ($p = 0.019$) while not in arterial-venous combined model ($p=0.49$) as comparing 2D and 3D segmentations (**Table S1**). In addition, the AUC of each selected 2D-based or 3D-based radiomics feature and their statistical differences between noninvasive and invasive groups were also summarized in the **Tables S2–S5**.

The arterial-venous combined models in both of 2D and 3D conditions showed a relatively good agreement between predicted and actual probability as shown by calibration curves in **Figure 4**. The overall performance of radiomics logistic regression model trained in the SMOTE dataset among 1000-times bootstrap were summarized in **Tables S6** and **S7**. The appearing frequency of each radiomics feature in the 3D and 2D logistic regression models during 1000-times bootstrap was respectively illustrated in **Figures S1** and **S2**. The respective optimism-corrected model's AUC (3D arterial model: 0.928; 3D venous model: 0.832; 2D arterial model: 0.815; 2D venous model: 0.78) and their average optimism (3D arterial model: 0.045; 3D venous model: 0.059; 3D arterial model: 0.048; 3D venous model: 0.055) represents a relatively good reliability of the model established from the selected feature. In **Figure S2**, all of the features selected in the final 2D and 3D models appeared over 500 times during 1000-times bootstrap, which also reflected the reliability of the features. As shown by the DCA curves in **Figure 4**, in 3D models, the arterial model and arterial-venous combined model have wider range for risk threshold (0-0.9) than venous model to make model net benefit larger than 50%. And in 2D models, the arterial-venous combined model has wider range for risk threshold (0-0.6) than arterial or venous model alone to reach model net benefit exceed 50%.

DISCUSSION

In this retrospective study, we applied radiomics techniques to predict the invasiveness of pancreatic solid pseudopapillary neoplasm (pSPN), and established radiomics models to evaluate tumor invasiveness before operation based on CT images. The results showed that the 3D-domain radiomics features and 2D-domain features are potential predictors. Therefore, the classifier based on radiomics features could potentially provide a noninvasive and personalized management method for pSPN patients.

Previous studies mainly focused on the characteristic manifestations of CT and/or magnetic resonance imaging (38). However, it is difficult to identify whether pSPN is invasive before operation (39). Radiomic analysis has been proposed as a step towards realization of precision medicine by providing means to interrogate the spatial complexity of tumors *in vivo* (40). Therefore, we try to identify the invasiveness of pSPN by radiomics features and comprehensively consider the effectiveness of different contrast-enhanced phases and 2D or 3D segmentation.

The results showed that when using the single-phase-based radiomics features to predict the invasiveness of pSPN, the arterial phase features were more effective than the venous phase features and there was no statistically significant model's AUC improvement in the original dataset when combining arterial and venous phase compared with sole arterial-phase model. The arterial phase enhancement characteristics of the tumor reflected the characteristics of tumor blood supply and functional capillaries, and the invasive tumor had more obvious blood supply. On the one hand, angiogenesis is closely related to the occurrence, development and prognosis of the tumor, but there are complex substances such as collagen and hyaluronic acid in the cell stroma. During the venous phase, the interstitial components of the tumor will also be enhanced because of the inflow of the contrast medium. As a result, the vascular enhancement of the tumor in the venous phase is not obvious. The arterial phase images can only enhance the blood vessels of the tumor, because the interstitial components will not be enhanced. Thus the arterial phase images can better reflect the characteristics of blood supply in the tumor and reflect the characteristics of the tumor more accurately. In some related studies of pancreatic tumors, Kwon et al. (41) especially emphasized the role of enhanced MRI in arterial phase in the study of the differentiating focal autoimmune pancreatitis and pancreatic ductal adenocarcinoma. Corwin et al. (42). reported that the greater attenuation differential between lesions and normal pancreas during the arterial phase compared to venous. Bian et al. (23). also reported a significant positive association between the arterial-phase-based radiomics Radscore and the risk of LN metastasis in pancreatic ductal adenocarcinoma. In addition, it was also the arterial-phase model could have higher multivariable AUC in predicting malignancy and invasive pathological status of pancreatic intraductal papillary mucinous neoplasms compared with that of venous-phase model (14). On the other hand, the invasive biological behavior of invasive pSPN is more prone to vascular invasion, and peripheral blood vessels are easily involved, especially the venous wall is thin and the pressure is low, and the tumor is easy to cause stenosis of venous branches. If the formation of microtumor thrombus can lead to obstruction of venous branches in microcirculation and obstruction of blood flow, resulting in an increase of compensatory blood supply of arteries. These results might reflect the predictive capability of the arterial-phase-based radiomics features in predicting tumor invasiveness.

Most prior studies have employed either a single slice or whole tumor to extract features for radiomics analysis. However, the actual effect of using features extracted from 2D slices or 3D volumes varies. In a colorectal cancer prognosis study, it has shown that the CT feature extracted from 3D segmentation was superior to the largest cross-sectional segmentation to predict the survival rate (16). Some scholar reports also showed that the 2D and 3D segmentation possessed similar capability in clinical outcome prediction, such as the determination of pathological feature or prognosis in hepatic metastatic colorectal cancer (CRC) (28) and prediction of axillary lymph node metastasis in breast cancer (29). However, it has also been reported that 2D CT radiomics features performed better in prognosis prediction in lung cancer (30). These studies show that the results of 2D or 3D radiological analysis need to be further studied by different patient cohorts and tasks. Our study indicated that the AUC value of 3D-based model is greater than that of 2D-based model in both SMOTE dataset and original dataset. Except for that the AUC value of 2D venous model is 0.768, the AUC values of all models are all greater than 0.80. In the original dataset, the ROC comparison of 2D-based and 3D-based arterial-phase model was statistically significant. Multi-slices (3D) analysis covers the entire tumor volume and can better depict spatial heterogeneity than using a single tumor slice (2D) (27).

In terms of feature contribution to each 2D or 3D model in the current study, most of the selected features were filtered or

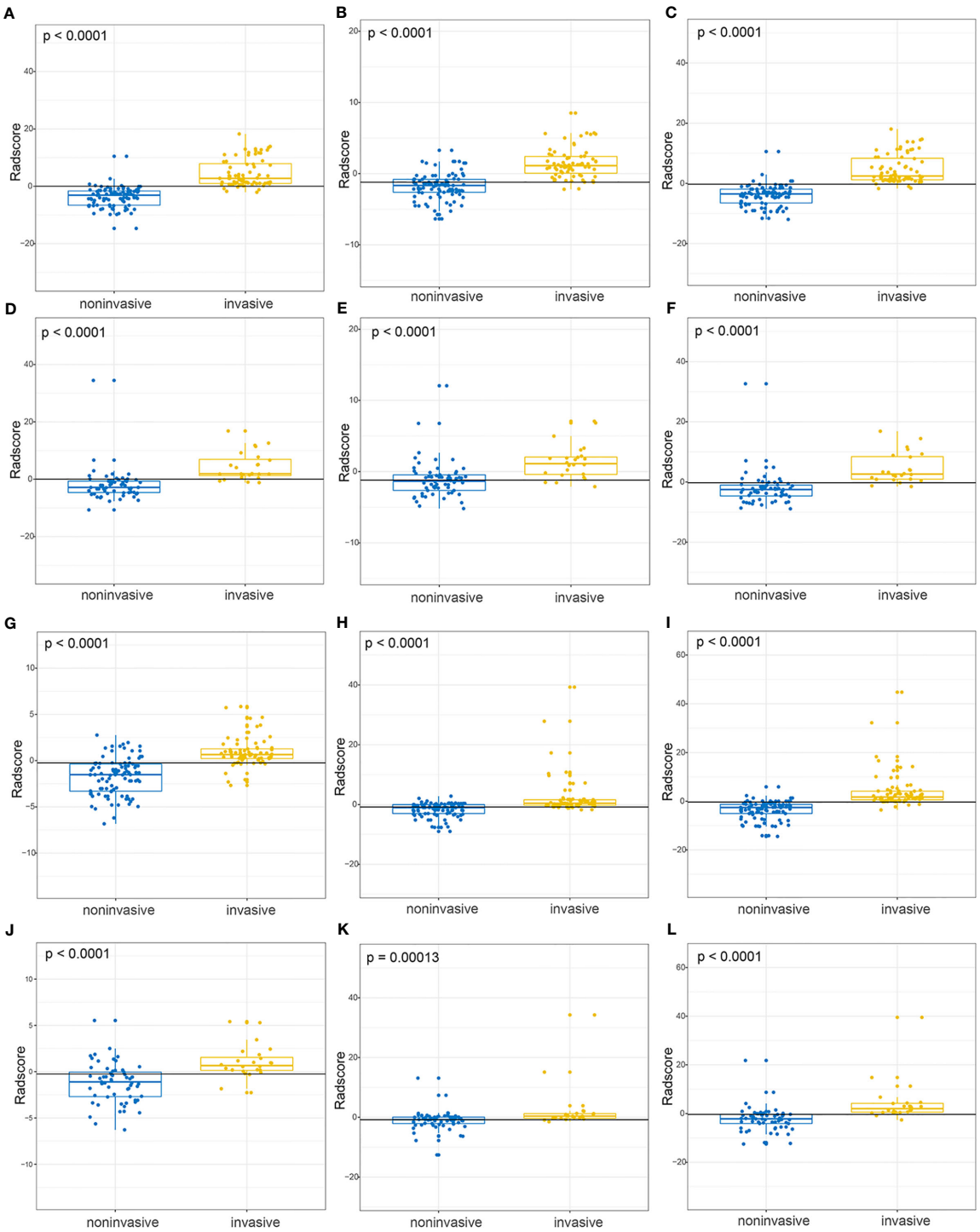


FIGURE 2 | The boxplot for Radscores from 2D-domain and 3D-domain radiomics models and their statistical differences between invasive and noninvasive groups. **(A–C)** The respective distribution of $Radscore^{AP_3D}$, $Radscore^{VP_3D}$, $Radscore^{AP_VP_3D}$ in the invasive (yellow) and noninvasive (blue) groups in the SMOTE dataset. **(D–F)** The respective distribution of $Radscore^{AP_3D}$, $Radscore^{VP_3D}$, $Radscore^{AP_VP_3D}$ in the invasive (yellow) and noninvasive (blue) groups in the original dataset. **(G–I)** The respective distribution of $Radscore^{AP_2D}$, $Radscore^{VP_2D}$, $Radscore^{AP_VP_2D}$ in the invasive (yellow) and noninvasive (blue) groups in the SMOTE dataset. **(J–L)** The respective distribution of $Radscore^{AP_2D}$, $Radscore^{VP_2D}$, $Radscore^{AP_VP_2D}$ in the invasive (yellow) and noninvasive (blue) groups in the original dataset.

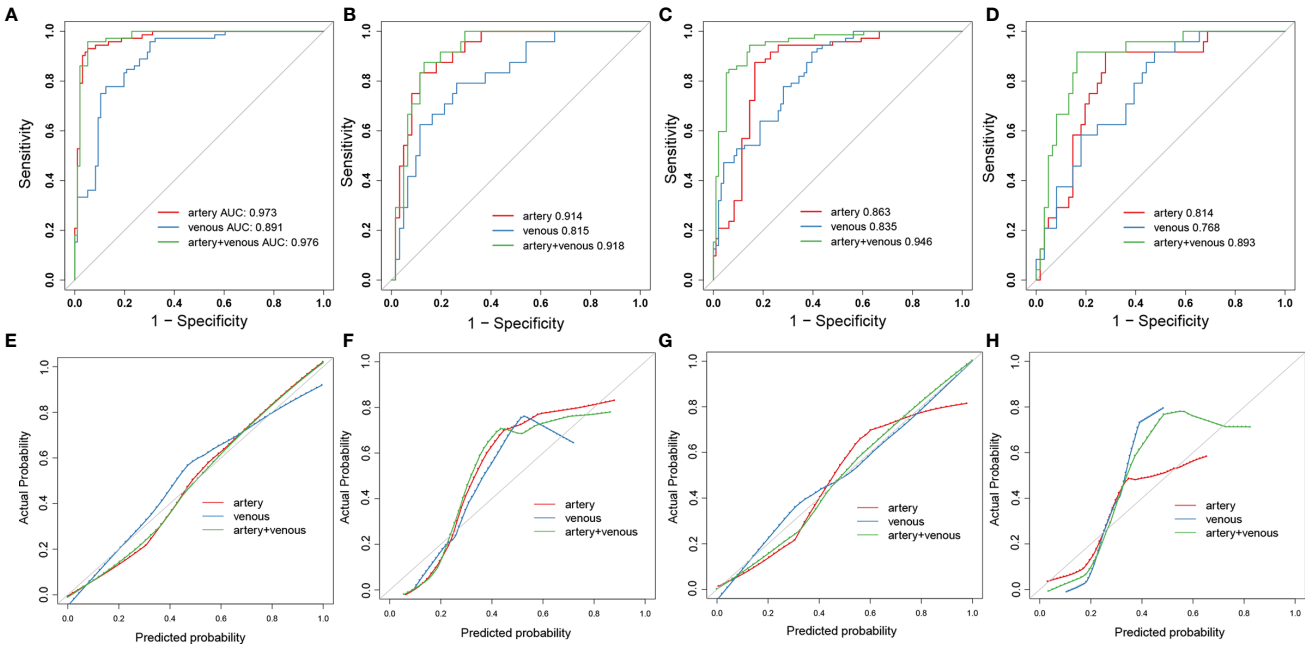


FIGURE 3 | The ROC and calibration curves for 3D- and 2D- radiomics models. The ROC curves of artery-(red), venous- (blue) and combined artery-venous (green) radiomics models based on 3D ROIs in the SMOTE dataset **(A)** and original dataset **(B)**. The ROC curves of artery-(red), venous-(blue) and combined artery-venous (green) radiomics models based on 2D ROIs in the SMOTE dataset **(C)** and original dataset **(D)**. The calibration curves of artery-(red), venous-(blue) and combined artery-venous radiomics (green) models based on 3D ROIs in the SMOTE dataset **(E)** and original dataset **(F)**. The calibration curves of artery-(red), venous-(blue) and combined artery-venous (green) radiomics models based on 2D ROIs in the SMOTE dataset **(G)** and original dataset **(H)**.

TABLE 2 | Performance of 2D-domain radiomics model in predicting invasive behavior of pancreatic solid pseudopapillary neoplasm.

	SMOTE dataset			Original dataset		
	Artery	**Venous**	**Artery +Venous**	**Artery**	**Venous**	**Artery +Venous**
Threshold	-0.248	-0.819	-0.337	-0.248	-0.819	-0.337
AUC (95%CI)	0.863 (0.806-0.921)	0.835 (0.777-0.894)	0.946 (0.913-0.980)	0.814 (0.717-0.910)	0.768 (0.665-0.871)	0.893 (0.821-0.966)
Specificity	0.833	0.604	0.854	0.721	0.492	0.738
Sensitivity	0.875	0.917	0.944	0.875	0.917	0.917
Accuracy	0.851	0.738	0.893	0.765	0.612	0.788
NPV	0.899	0.906	0.953	0.936	0.938	0.957
PPV	0.797	0.635	0.829	0.553	0.415	0.579

TABLE 3 | Performance of 3D-domain radiomics model in predicting invasive behavior of pancreatic solid pseudopapillary neoplasm.

	SMOTE dataset			Original dataset		
	Artery	**Venous**	**Artery +Venous**	**Artery**	**Venous**	**Artery +Venous**
Threshold	0.024	-1.195	-0.262	0.024	-1.195	-0.262
AUC (95%CI)	0.973 (0.950-0.996)	0.891 (0.843-0.939)	0.976 (0.953-0.998)	0.914 (0.854-0.974)	0.815 (0.718-0.912)	0.918 (0.860-0.976)
Specificity	0.948	0.698	0.948	0.820	0.525	0.803
Sensitivity	0.931	0.958	0.958	0.833	0.875	0.875
Accuracy	0.940	0.810	0.952	0.824	0.624	0.824
NPV	0.948	0.957	0.968	0.926	0.914	0.942
PPV	0.931	0.704	0.932	0.645	0.420	0.636

transformed first-order or texture features. It might indicate that the distinguishment between the noninvasive and invasive pSPN might need the emphasized features in the spatial or frequency domains. It could be found that there existed some overlaps for selected 2D and 3D feature types in the artery-phase models, including the RunVariance (GLRLM), Median (First order), Kurtosis (First order) and DependenceVariance (GLDM), which represent similar tendency between invasive and noninvasive

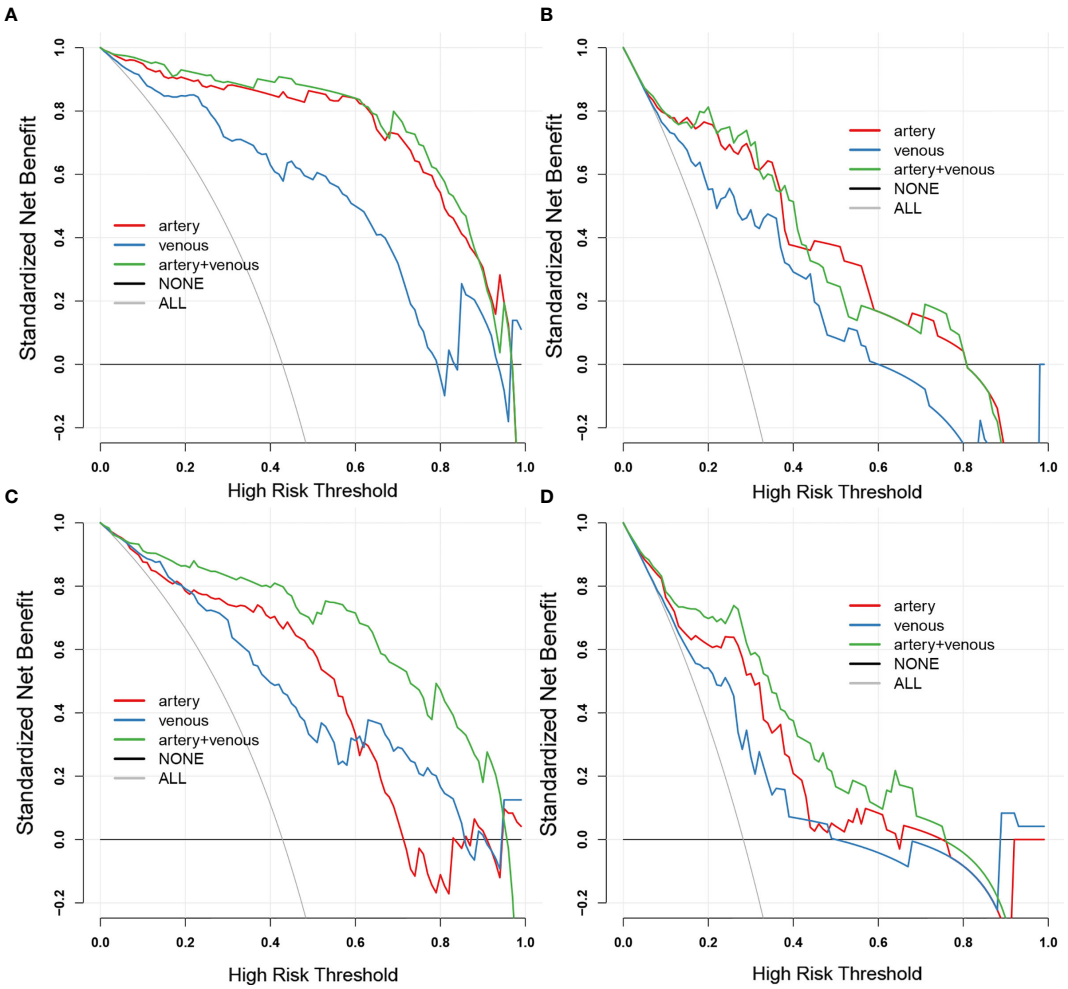


FIGURE 4 | The decision curves for 2D-domain and 3D-domain radiomics models. **(A, B)** The decision curves for 3D-domain radiomics models in SMOTE dataset **(A)** and original dataset **(B)**. **(C, D)** The decision curves for 2D-domain radiomics models in SMOTE dataset **(C)** and original dataset **(D)**. The black horizontal line manifests no patients is invasive type (NONE) and the grey line manifests all patients are invasive type (ALL). The colored lines of each model respectively illustrate the net benefit brought to each patient based on artery-(red), venous-(blue) and combined artery-venous (green) radiomics models. The closer the decision curves to the black and gray curves, the lower the clinical decision net benefit of the model.

groups. However, the 3D features' AUC located in the range of 0.58-0.72, which were slightly higher than that of 2D features (AUC from 0.58 to 0.66). And the features that not involved in the 2D model, such as wavelet.LLL_firstorder_InterquartileRange, original_firstorder_Skewness, and wavelet.HHL_glrlm_Long RunLowGrayLevelEmphasis could even have acceptable AUC values larger than 0.65. First-order statistics features describe the distribution of voxel intensity in the ROI region. GLDM texture features mainly describe the gray level dependence between voxels. GLRLM quantifies the length of consecutive pixels that have the same grey level. Three-dimensional ROI includes the whole lesion and does not avoid cystic necrosis and calcification in the tumor. It could have more chances compared with 2D cross-section to extract intensity or texture details distributed in 3D space or multiple 3D directions and select the representative features describing the internal structure or pathological heterogeneity of tumor which are closely related to invasiveness. This might indicate that intensity or texture details distributed in 3D space or multiple 3D directions could be extracted from the entire volume rather than those from the single slice. In the current study, 3D features can better predict invasiveness from a global point of view.

Thus, the noninvasive radiomics signature could serve as a more convenient biomarker for the prediction of invasive behavior in pSPN. To justify the clinical usefulness, decision curve analysis was applied in this study to confirm the predictive value of the imaging group model. This novel method offers insight into clinical consequences on the basis of threshold probability, from which the net benefit could be derived.

There are still some limitations in this study. Firstly, the disease is rare, all available data have been collected, but the sample size in this study was small and had some class imbalance

with noninvasive-to-invasive sample ratio of 2.5. The data were augmented and balanced by SMOTE method into noninvasive-to-invasive sample ratio of 1.3 and used as the training set of the model, which was further internally validated in the original dataset to test the model performance as much as possible. Secondly, it was not combined with clinical data and pathological immunohistochemical results. A study of radiomics and clinical data is needed in the future to further reveal the biological or clinical meanings or association for the radiomics features, which still could not be well explained in the current study. In addition, this is a single-center study, and we are working to further evaluate our model in a bigger dataset that may come from multiple centers and multiple imaging schemes.

CONCLUSION

In conclusion, a radiomics method based on CT imaging data was developed and validated as a potential method for predicting invasiveness of pSPN before the operation in our study. Radiomics model showed encouraging performance and is expected to provide an intelligent, non-invasive diagnostic tool for predicting the invasiveness of pSPN. Further research is needed to explore the relationship between radiomics features and clinicopathological index and establish more generalized prediction models.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

W-pH and PL: designed the research. L-mL and W-pH: performed the research. Y-jH: collected the data. S-yL and W-pH: analyzed the data and wrote the paper. J-bG: reviewed the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural and Science Fund of China (No. 81971615).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.677814/full#supplementary-material

REFERENCES

1. Gao H, Gao Y, Yin L, Wang G, Wei J, Jiang K, et al. Risk Factors of the Recurrences of Pancreatic Solid Pseudopapillary Tumors: A Systematic Review and Meta-Analysis. *J Cancer* (2018) 9:1905–14. doi: 10.7150/jca.24491
2. Antoniou EA, Damaskos C, Garmpis N, Salakos C, Margonis GA, Kontzoglou K, et al. Solid Pseudopapillary Tumor of the Pancreas: A Single-Center Experience and Review of the Literature. *In Vivo* (2017) 31:501–10. doi: 10.21873/invivo.11089
3. Lubezky N, Papoulas M, Lessing Y, Gitstein G, Brazowski E, Nachmany I, et al. Solid Pseudopapillary Neoplasm of the Pancreas: Management and Long-Term Outcome. *Eur J Surg Oncol* (2017) 43:1056–60. doi: 10.1016/j.ejso.2017.02.001
4. Pancreatic Surgery of Chinese Academic Society of Young Surgeons. [the Current Status of Diagnosis and Treatment of Pancreatic Cystic Neoplasm in China: A Report of 2 251 Cases]. *Zhonghua Wai Ke Za Zhi* (2018) 56:24–9. doi: 10.3760/cma.j.issn.0529-5815.2018.01.007
5. Xu Y, Zhao G, Pu N, Nuerxiati A, Ji Y, Zhang L, et al. One Hundred Twenty-One Resected Solid Pseudopapillary Tumors of the Pancreas: An 8-Year Single-Institution Experience At Zhongshan Hospital, Shanghai, China. *Pancreas* (2017) 46:1023–8. doi: 10.1097/MPA.0000000000000885
6. Wang P, Wei J, Wu J, Xu W, Chen Q, Gao W, et al. Diagnosis and Treatment of Solid-Pseudopapillary Tumors of the Pancreas: A Single Institution Experience With 97 Cases. *Pancreatology* (2018) 18:415–9. doi: 10.1016/j.pan.2017.12.012
7. Ercelep O, Ozdemir N, Turan N, Topcu TO, Uysal M, Tanriverdi O, et al. Retrospective Evaluation of Patients Diagnosed Solid Pseudopapillary Neoplasms of the Pancreas. *Curr Probl Cancer* (2019) 43:27–32. doi: 10.1016/j.currproblcancer
8. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: The Bridge Between Medical Imaging and Personalized Medicine. *Nat Rev Clin Oncol* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
9. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are More Than Pictures, They are Data. *Radiology* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
10. Dong J, Qi X. Liver Imaging in Precision Medicine. *EBioMedicine* (2018) 32:1–2. doi: 10.1016/j.ebiom.2018.05.017
11. Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clin Cancer Res* (2017) 23:7253–62. doi: 10.1158/1078-0432.CCR-17-1038
12. Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma. *Clin Cancer Res* (2017) 23:4259–69. doi: 10.1158/1078-0432.CCR-16-2910
13. Li J, Liu S, Qin Y, Zhang Y, Wang N, Liu H. High-Order Radiomics Features Based on T2 Flair MRI Predict Multiple Glioma Immunohistochemical Features: A More Precise and Personalized Gliomas Management. *PLoS One* (2020) 15:e0227703. doi: 10.1371/journal.pone.0227703
14. Polk SL, Choi JW, McGettigan MJ, Rose T, Ahmed A, Kim J, et al. Multiphase Computed Tomography Radiomics of Pancreatic Intraductal Papillary Mucinous Neoplasms to Predict Malignancy. *World J Gastroenterol* (2020) 26:3458–71. doi: 10.3748/wjg.v26.i24.3458
15. Tobaly D, Santinha J, Sartoris R, Dioguardi Burgio M, Matos C, Cros J, et al. Ct-Based Radiomics Analysis to Predict Malignancy in Patients With Intraductal Papillary Mucinous Neoplasm (IPMN) of the Pancreas. *Cancers (Basel)* (2020) 12:3089. doi: 10.3390/cancers12113089
16. Benedetti G, Mori M, Panzeri MM, Barbera M, Palumbo D, Sini C, et al. Ct-Derived Radiomic Features to Discriminate Histologic Characteristics of Pancreatic Neuroendocrine Tumors. *Radiol Med* (2021). doi: 10.1007/s11547-021-01333-z
17. Chen F, Zhou Y, Qi X, Zhang R, Gao X, Xia W, et al. Radiomics-Assisted Presurgical Prediction for Surgical Portal Vein-Superior Mesenteric Vein Invasion in Pancreatic Ductal Adenocarcinoma. *Front Oncol* (2020) 10:523543. doi: 10.3389/fonc.2020.523543
18. Zhang Y, Lobo-Mueller EM, Karanicolas P, Gallinger S, Haider MA, Khalvati F. Improving Prognostic Performance in Resectable Pancreatic Ductal Adenocarcinoma Using Radiomics and Deep Learning Features Fusion in CT Images. *Sci Rep* (2021) 11:1378. doi: 10.1038/s41598-021-80998-y

19. Iwatate Y, Hoshino I, Yokota H, Ishige F, Itami M, Mori Y, et al. Radiogenomics for Predicting P53 Status, Pd-L1 Expression, and Prognosis With Machine Learning in Pancreatic Cancer. *Br J Cancer* (2020) 123:1253–61. doi: 10.1038/s41416-020-0997-1
20. Chakraborty J, Midya A, Gazit L, Attiyeh M, Langdon-Embry L, Allen PJ, et al. Ct Radiomics to Predict High-Risk Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Med Phys* (2018) 45:5019–29. doi: 10.1002/mp.13159
21. Tobaly D, Santinha J, Sartoris R, Dioguardi Burgio M, Matos C, Cros J, et al. Ct-Based Radiomics Analysis to Predict Malignancy in Patients With Intraductal Papillary Mucinous Neoplasm (IPMN) of the Pancreas. *Cancers (Basel)* (2020) 12:3089. doi: 10.3390/cancers12113089
22. Park S, Chu LC, Hruban RH, Vogelstein B, Kinzler KW, Yuille AL, et al. Differentiating Autoimmune Pancreatitis From Pancreatic Ductal Adenocarcinoma With Ct Radiomics Features. *Diagn Interv Imaging* (2020) 101:555–64. doi: 10.1016/j.diii.2020.03.002
23. Bian Y, Guo S, Jiang H, Gao S, Shao C, Cao K, et al. Relationship Between Radiomics and Risk of Lymph Node Metastasis in Pancreatic Ductal Adenocarcinoma. *Pancreas* (2019) 48:1195–203. doi: 10.1097/MPA.0000000000001404
24. Mackin D, Fave X, Zhang L, Yang J, Jones AK, Ng CS, et al. Harmonizing the Pixel Size in Retrospective Computed Tomography Radiomics Studies. *PLoS One* (2017) 12:e0178524. doi: 10.1371/journal.pone.0178524
25. Ligero M, Jordi-Ollero O, Bernatowicz K, Garcia-Ruiz A, Delgado-Muñoz E, Leiva D, et al. Minimizing Acquisition-Related Radiomics Variability by Image Resampling and Batch Effect Correction to Allow for Large-Scale Data Analysis. *Eur Radiol* (2021) 31:1460–70. doi: 10.1007/s00330-020-07174-0
26. Ahn SJ, Kim JH, Park SJ, Han JK. Prediction of the Therapeutic Response After FOLFOX and FOLFIRI Treatment for Patients With Liver Metastasis From Colorectal Cancer Using Computerized Ct Texture Analysis. *Eur J Radiol* (2016) 85:1867–74. doi: 10.1016/j.ejrad.2016.08.014
27. Ng F, Kozarski R, Ganeshan B, Goh V. Assessment of Tumor Heterogeneity by CT Texture Analysis: can the Largest Cross-Sectional Area be Used as an Alternative to Whole Tumor Analysis? *Eur J Radiol* (2013) 82:342–8. doi: 10.1016/j.ejrad.2012.10.023
28. Lubner MG, Stabo N, Lubner SJ, del Rio AM, Song C, Halberg RB, et al. Ct Textural Analysis of Hepatic Metastatic Colorectal Cancer: Pre-Treatment Tumor Heterogeneity Correlates With Pathology and Clinical Outcomes. *Abdom Imaging* (2015) 40:2331–7. doi: 10.1007/s00261-015-0438-4
29. Arefan D, Chai R, Sun M, Zuley ML, Wu S. Machine Learning Prediction of Axillary Lymph Node Metastasis in Breast Cancer: 2d Versus 3d Radiomic Features. *Med Phys* (2020) 47:6334–42. doi: 10.1002/mp.14538
30. Shen C, Liu Z, Guan M, Song J, Lian Y, Wang S, et al. 2D and 3D CT Radiomics Features Prognostic Performance Comparison in Non-Small Cell Lung Cancer. *Transl Oncol* (2017) 10:886–94. doi: 10.1016/j.tranon.2017.08.007
31. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* (2016) 15:155–63. doi: 10.1016/j.jcm.2016.02.012
32. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472
33. Larue RTHM, van Timmeren JE, de Jong EEC, Feliciani G, Leijenaar RTH, Schreurs WMJ, et al. Influence of Gray Level Discretization on Radiomic Feature Stability for Different Ct Scanners, Tube Currents and Slice Thicknesses: A Comprehensive Phantom Study. *Acta Oncol* (2017) 56:1544–53. doi: 10.1080/0284186X.2017.1351624
34. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. Smote: Synthetic Minority Over-Sampling Technique. *J Artif Intell Res* (2002) 16(1):321–57. doi: 10.1613/jair.953
35. Ren S, Zhao R, Zhang J, Guo K, Gu X, Duan S, et al. Diagnostic Accuracy of Unenhanced Ct Texture Analysis to Differentiate Mass-Forming Pancreatitis From Pancreatic Ductal Adenocarcinoma. *Abdom Radiol (NY)* (2020) 45:1524–33. doi: 10.1007/s00261-020-02506-6
36. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol* (2016) 34:2157–64. doi: 10.1200/JCO.2015.65.9128
37. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (Tripod): Explanation and Elaboration. *Ann Intern Med* (2015) 162:W1–73. doi: 10.7326/M14-0698
38. Yin Q, Wang M, Wang C, Wu Z, Yuan F, Chen K, et al. Differentiation Between Benign and Malignant Solid Pseudopapillary Tumor of the Pancreas by MDCT. *Eur J Radiol* (2012) 81:3010–8. doi: 10.1016/j.ejrad.2012.03.013
39. Kang CM, Choi SH, Kim SC, Lee WJ, Choi DW, Kim SW. Korean Pancreatic Surgery Club. Predicting Recurrence of Pancreatic Solid Pseudopapillary Tumors After Surgical Resection: A Multicenter Analysis in Korea. *Ann Surg* (2014) 260:348–55. doi: 10.1097/SLA.0000000000000583
40. Roy S, Whitehead TD, Quirk JD, Salter A, Ademuyiwa FO, Li S, et al. Optimal Co-Clinical Radiomics: Sensitivity of Radiomic Features to Tumour Volume, Image Noise and Resolution in Co-Clinical T1-Weighted and T2-Weighted Magnetic Resonance Imaging. *EBioMedicine* (2020) 59:102963. doi: 10.1016/j.ebiom.2020.102963
41. Kwon JH, Kim JH, Kim SY, Byun JH, Kim HJ, Lee MG, et al. Differentiating Focal Autoimmune Pancreatitis and Pancreatic Ductal Adenocarcinoma: Contrast-Enhanced MRI With Special Emphasis on the Arterial Phase. *Eur Radiol* (2019) 29:5763–71. doi: 10.1007/s00330-019-06200-0
42. Corwin MT, Lamba R, Wilson M, McGahan JP. Renal Cell Carcinoma Metastases to the Pancreas: Value of Arterial Phase Imaging At Mdct. *Acta Radiol* (2013) 54:349–54. doi: 10.1258/ar.2012.120693

Conflict of Interest: Author S-yL was employed by General Electric Company (GE) Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Huang, Liu, Han, Li, Liang and Gao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 24 May 2021
doi: 10.3389/fonc.2021.687884

CT-Based Radiomics Signature With Machine Learning Predicts MYCN Amplification in Pediatric Abdominal Neuroblastoma

Xin Chen[1†], Haoru Wang[1†], Kaiping Huang[1], Huan Liu[2], Hao Ding[1], Li Zhang[1], Ting Zhang[1], Wenqing Yu[1] and Ling He[1]*

[1] Department of Radiology, Children's Hospital of Chongqing Medical University, National Clinical Research Center for Child Health and Disorders, Ministry of Education Key Laboratory of Child Development and Disorders, Chongqing Key Laboratory of Pediatrics, Chongqing, China, [2] GE Healthcare, Shanghai, China

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln, United Kingdom

Reviewed by:
Yunlang She,
Tongji University, China
Kun Zheng,
Peking Union Medical College Hospital (CAMS), China

****Correspondence:***
Ling He
doctorheling@yeah.net

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *30 March 2021*
Accepted: *03 May 2021*
Published: *24 May 2021*

Citation:
Chen X, Wang H, Huang K, Liu H, Ding H, Zhang L, Zhang T, Yu W and He L (2021) CT-Based Radiomics Signature With Machine Learning Predicts MYCN Amplification in Pediatric Abdominal Neuroblastoma. Front. Oncol. 11:687884. doi: 10.3389/fonc.2021.687884

Purpose: MYCN amplification plays a critical role in defining high-risk subgroup of patients with neuroblastoma. We aimed to develop and validate the CT-based machine learning models for predicting MYCN amplification in pediatric abdominal neuroblastoma.

Methods: A total of 172 patients with MYCN amplified (n = 47) and non-amplified (n = 125) were enrolled. The cohort was randomly stratified sampling into training and testing groups. Clinicopathological parameters and radiographic features were selected to construct the clinical predictive model. The regions of interest (ROIs) were segmented on three-phrase CT images to extract first-, second- and higher-order radiomics features. The ICCs, mRMR and LASSO methods were used for dimensionality reduction. The selected features from the training group were used to establish radiomics models using Logistic regression, Support Vector Machine (SVM), Bayes and Random Forest methods. The performance of four different radiomics models was evaluated according to the area under the receiver operator characteristic (ROC) curve (AUC), and then compared by Delong test. The nomogram incorporated of clinicopathological parameters, radiographic features and radiomics signature was developed through multivariate logistic regression. Finally, the predictive performance of the clinical model, radiomics models, and nomogram was evaluated in both training and testing groups.

Results: In total, 1,218 radiomics features were extracted from the ROIs on three-phrase CT images, and then 14 optimal features, including one original first-order feature and eight wavelet-transformed features and five LoG-transformed features, were identified and selected to construct the radiomics models. In the training group, the AUC of the Logistic, SVM, Bayes and Random Forest model was 0.940, 0.940, 0.780 and 0.927, respectively, and the corresponding AUC in the testing group was 0.909, 0.909, 0.729, 0.851, respectively. There was no significant difference among the Logistic, SVM and Random Forest model, but all better than the Bayes model ($p < 0.005$). The predictive performance of the Logistic radiomics model based on three-phrase is similar to

nomogram, but both better than the clinical model and radiomics model based on single venous phase.

Conclusion: The CT-based radiomics signature is able to predict MYCN amplification of pediatric abdominal NB with high accuracy based on SVM, Logistic and Random Forest classifiers, while Bayes classifier yields lower predictive performance. When combined with clinical and radiographic qualitative features, the clinics-radiomics nomogram can improve the performance of predicting MYCN amplification.

Keywords: children, abdomen, neuroblastoma, MYCN, radiomics, prediction

INTRODUCTION

Neuroblastoma (NB) is one of the most common solid malignancy in children originating from neural crest tissues along the sympathetic chains (1). NB can arise from various anatomical compartments (i.e., neck, chest, abdomen or pelvis), but most frequently arise from the abdomen (adrenal gland or extra-adrenal retroperitoneum), accounting for 73% of all systems (2). As a kind of heterogeneous tumor, the clinical outcome of abdominal NB varies from spontaneous regression to extensive systemic metastasis (3). For the pediatric patients with abdominal NB in advanced stage, the long-term survival rate is less than 50% regardless of the intensive treatment (4). Therefore, risk stratification is vital enough to choose the optimal therapy for individuals in the era of precision medicine (5). Among the various attempts from different international groups aimed to identify factors that can be used to risk stratification and to define an sub-population with poor clinical outcome (6), all groups highlight the significance of MYCN amplification status for defining high-risk group and consider that all patients with MYCN amplified are prone to relapse (2). Clinically, the amplification of MYCN oncogene is significantly correlated to an aggressive phenotype (7). Therefore, the detection of MYCN amplification status is critical to risk-stratify patients. However, as an invasive method, traditional biopsy may cause various complications (8). Meanwhile, the availability of detection of MYCN has been hindered by the limited access to genetic testing methods in many institutions (9), therefore, an alternative non-invasive method is needed to characterize the MYCN amplification status availably.

In recent years, the increasing application of radiomics in solid tumors has resulted in the emergence of radiogenomics. The heart of radiogenomics is to identify and predict the expression of clinically significant molecular biomarkers of tumors by analyzing high-dimensional quantitative signatures extracted from tumor regions of interest (ROIs) in radiographic images (9, 10). Compared with histopathology and genetic testing methods, radiogenomics not only can overcome sampling bias and the possible complications caused by biopsy, but also is expected to provide more comprehensive and accurate information in predicting the biomarkers (9). To date, the application of radiogenomics in pediatric tumors is mainly focused on MRI-based signatures of medulloblastoma, and the CT-based radiogenomics is rarely used (11, 12). Although a recent study has shown the potential of CT-based signature in the prediction of MYCN amplification of NB and ganglioneuroblastoma (GNB) (8), there were some problems with the patients' selection, in which nonabdominal NB and GNB were also enrolled, because previous literatures have demonstrated that MYCN amplification rarely occurs in nonabdominal NB and GNB (13, 14). Meanwhile, due to the heterogeneity of NB, the ROIs selectively delineated on several largest levels of the tumor cannot reflect the biological characteristics of the tumor comprehensively (15). Instead, the whole-tumor ROIs delineated on all slices in other radiomics studies have contributed to reduce sampling bias and improve intra- and inter-observer consistency (16, 17).

In the present study, we developed and validated the CT-based radiomics features combined with various machine learning methods for predicting MYCN amplification of abdominal NB in the cohort of pediatric patients. Besides, we constructed a clinical model based on clinicopathological parameters and radiographic features, and then added the radiomics signature to develop radiomics-clinics model. The predictive performance of clinical model, radiomics model and radiomics-clinics model was finally evaluated and compared according to the AUC and Delong test.

MATERIALS AND METHODS

Patients and MYCN Amplification Characterization

The Ethics Committee of our hospital approved this single-center retrospective study and waived the requirement for informed patient consent. We identified 172 abdominal NB patients with MYCN amplified (n = 47) and non-amplified (n = 125) by searching the medical record management system and radiology picture archiving and communication system (PACS) of our department from May 2012 to August 2020 consecutively according to the inclusion and exclusion criteria. Inclusion criteria were: (1) availability of abdominal contrast-enhanced CT with sufficient image quality, including non-enhanced, arterial and venous phase; (2) patients without any radiotherapy, chemotherapy or surgical treatment before the first CT examination; (3) pathologically confirmed abdominal NB; (4) with the detection of MYCN status. Exclusion criteria were: (1) patients with ganglioneuroma or GNB; (2) patients with

nonabdominal NB (e.g., neck, chest or pelvis); (3) abdominal NB patients absent of three-phrase CT scans; (4) insufficient image quality; (5) without the detection of MYCN status; (6) abdominal NB patients with prior treatments. The detailed workflow of patients' selection is shown in **Figure 1**.

The study cohort was randomly stratified sampling into training group and testing group in a proportion of 7:3. Clinicopathological parameters, including gender, age (month), histopathology, INSS stage, Shimada classification and urinary vanillylmandelic acid (VMA) were collected from medical records. According to the differentiation degree, the histopathological results were categorized into two groups: undifferentiated or poorly differentiated, and differentiated NB (8). The prognostic Shimada classification of patients was defined as favorable histology (FH) and unfavorable histology (UFH) on the basis of age, degree of differentiation and mitotic karyorrhectic index (MKI) of NB (18). The MYCN gene copy number was detected by fluorescence *in situ* hybridization (FISH) method in all specimens using a MYC-N/LAF double probe, and cases with the number of signals exceeding 10-fold MYCN copies were considered to be MYCN amplified (18). The intervals between the MYCN status detection and contrast-enhanced CT scans of the same patient in the present study were less than a month.

CT Scanning

All CT scans were acquired during a single breath-hold in cooperative children or during quiet respiration in children unable to suspend respiration, and those who could not cooperate were sedate by oral administration of 10% chloral hydrate (0.5 ml/kg, body weight) before examination. All abdominal three-phase CT scans, including non-enhanced phase (NP), arterial phase (AP), and venous phase (VP), were performed on Lightspeed VCT 64-slice spiral CT (GE Healthcare, USA) scanner or Brilliance ICT 256-slice spiral CT (Philips, Netherlands) scanner. The CT scanning parameters were (1) tube voltage: 120 kV; (2) tube current: 200 mAs; (3) pitch: 0.984:1; (4) slice thickness: 5.0 mm; (5) slice interval: 5.0 mm; (6) reconstructed slice thickness: 1.25 mm. Nonionic iodinated contrast material (Omnipaque 300 mg I/mL or Visipaque 320 mg I/ml, GE Healthcare) was used. Contrast material (2 ml/kg, body weight) was injected into peripheral vein of the forearm with a power injector at a rate of 1–3 ml/s. AP and VP of post-contrast scanning were performed at 20–35 and 60–70 s respectively after contrast material administration.

Imaging Analysis

All CT examinations were transmitted to the workstation for review and analysis. All images were initially analyzed independently by two experienced pediatric radiologists without knowledge of the MYCN status. The tumor features, including calcification (present or not), infiltrating across midline (exceeding the contralateral edge of the spine, present or not) and necrosis (present or not), were recorded[8]. Disagreements were resolved by negotiation.

Clinical Model Building

Clinicopathological parameters of MYCN-amplified and non-amplified groups included gender, age, histopathology, INSS stage, Shimada classification and urinary vanillylmandelic acid (VMA) and radiographic features. Influence characteristics that were statistically significant with $p<0.05$ in the univariate logistic analyses were included in the multivariate analysis following the stepwise selection method. The Akaike information criterion (AIC) and Log-Likelihood were used as the stopping rules to select the most predictive clinical features.

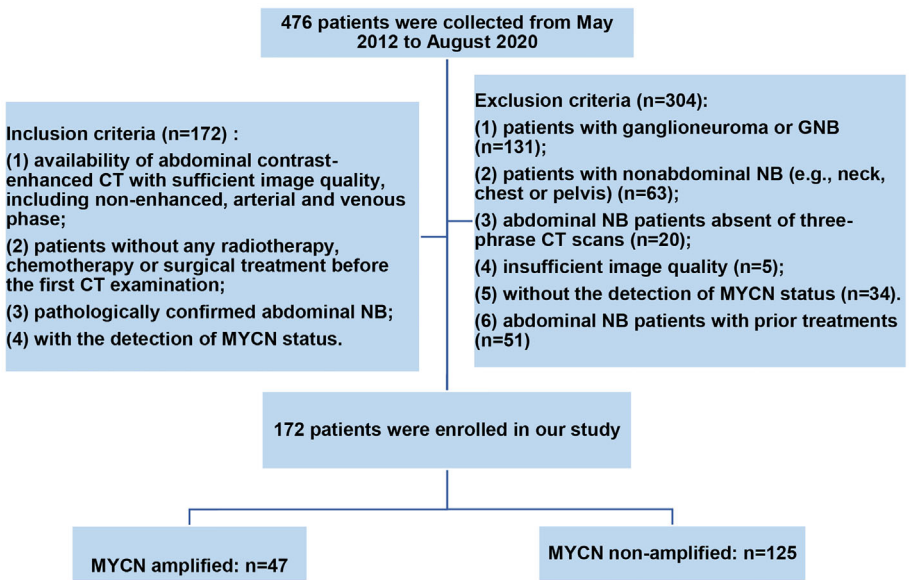


FIGURE 1 | The workflow of patients' selection in our study.

Image Preprocessing and Tumor Segmentation

Before the tumor segmentation, isotropic voxel resampling into 1 mm × 1 mm × 1 mm with linear interpolation was used to image preprocessing for purpose of normalizing the geometry of CT images. The ROIs of whole-tumor were manually 3D-delineated on three phrases respectively using a free open-source software package (ITK-SNAP, ver.3.4.0) by a pediatric radiologist with 2 years of experience, and the ROIs were then reviewed and confirmed by the other pediatric radiologist with 10 years of experience (**Figure 2**). The ROIs included the calcification and necrosis area of the lesion. The ROIs segmentation of each tumor was performed twice by reader 1 (time-interval of 2 weeks) and once by reader 2. The intra-observer class correlation coefficients (ICCs) were calculated based on the features extracted from the ROIs delineated by reader 1 at different time points.

Radiomics Features Extraction and Selection

The images and corresponding ROIs were imported into the in-house software (Artificial Intelligence Kit, AK, Version V3.2.2.R, GE Healthcare) together, and then features extraction was performed with AK software. The radiomics features were classified into seven groups including: first order, shape, gray-level co-occurrence matrix (GLCM), gray-level size-zone matrix (GLSZM), gray-level run-length matrix (GLRLM), neighborhood gray-tone difference matrix (NGTDM) and neighboring gray-level dependence matrix (GLDM). To enhance intricate patterns in the data invisible to the human eye, advanced filters, including Laplacian of Gaussian (LoG; sigma, 2.0 and 3.0 mm), and wavelet decompositions with all possible combinations of high (H) or low (L) pass filter in each of the three dimensions (HHH, HHL, HLH, LHH, LLL, LLH, LHL, HLL), were applied. A total of 3654 radiomic features of each patient (1218 features in each phase) were finally extracted from the ROIs based on NP, AP and VP.

Because many of the extracted high-dimensional features are often redundant and meaningless, a variety of methods were used for dimensionality reduction. To begin with, intra-observer analysis was used to assess the reliability and reproducibility of the features in order to find out the robust features. Features with ICCs higher than 0.80 were considered reliable and selected. Then, two feature selection methods, the maximum relevance minimum redundancy (mRMR) and the least absolute shrinkage and selection operator (LASSO) regression, were applied to eliminate the redundant and irrelevant features and choose the optimized subset of features to construct the radiomics models. Due to the CT scans in our study were performed on two scanners from different manufacturers (Lightspeed VCT and Brilliance ICT), the performance of radiomics features derived from two scanners was evaluated by ROC analysis and Delong

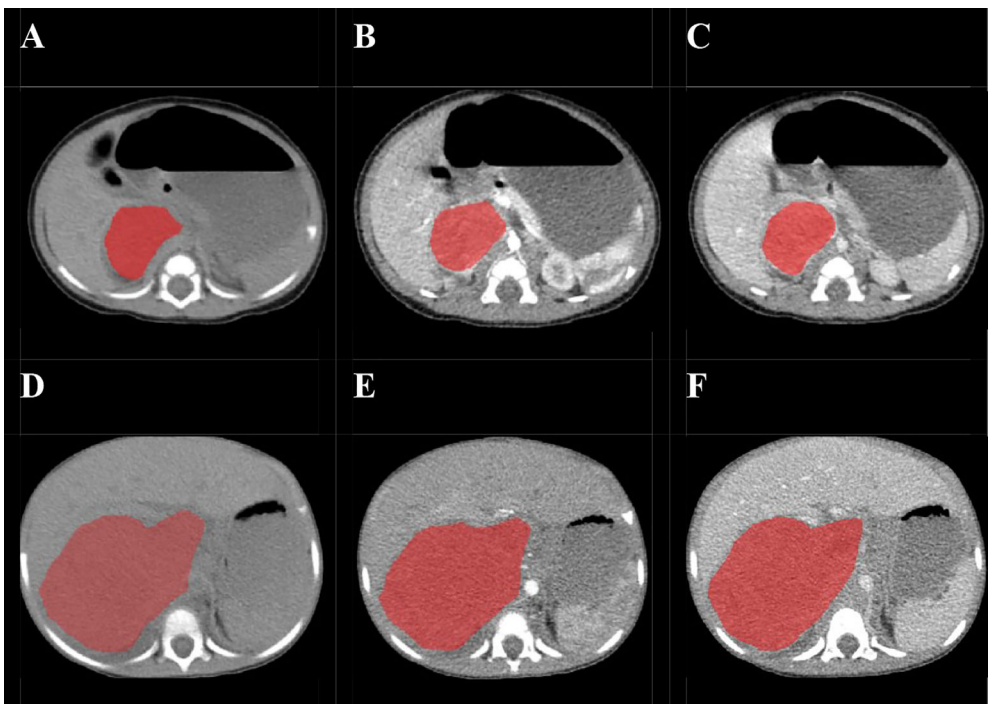


FIGURE 2 | Examples of manual delineated regions of interests (ROIs) of amplified and non-amplified NB. Delineation of the ROI on one slice of a non-amplified NB (Female, 2 months, asymptomatic) on non-enhanced phase (NP) **(A)**, arterial phase (AP) **(B)**, and venous phase (VP) **(C)**; Delineation of the ROI on one slice of an amplified NB (female, 2 years, presented with abdominal palpable mass) on NP **(D)**, AP **(E)**, and VP **(F)**.

test. Rad-score was calculated by summing the selected features weighted by their coefficients.

Machine Learning

The selected features from the training group were used to establish radiomics models based on three-phrase using Logistic regression, Support Vector Machine (SVM), Bayes and Random Forest. The performance of the developed radiomics models were then validated in both training and testing groups according to the area under the receiver operator characteristic (ROC) curve (AUC). The Delong test was used to compare the performance of four different machine learning models.

Nomogram Building and Evaluating

Finally, the radiomics signature was added to build the radiomics-clinics nomogram incorporated of statistically significant clinicopathological parameters and radiographic features on the basis of the results of multivariate logistic regression analysis in the training group. The predictive performance of the clinical model, radiomics models, and nomogram was evaluated according to the area under the receiver operator characteristic (ROC) curve (AUC) in both training and testing groups, and the Delong test was applied to compare the performance of different models.

Statistical Analysis

IPM statistics (IPMs, version 2.4.0, GE healthcare) and R programming language (ver. 3.4.2, http://www.r-project.org) were used to carry out statistical analysis. A chi square test or Fisher's exact test was used for the nominal variables, and a Mann–Whitney test was used for the continuous variables with abnormal distribution between the two cohorts. A two-tailed $p < 0.05$ indicated statistical significance. "mRMRe" and "glmnet" packages were used to carry out the mRMR and LASSO respectively. The "pROC" package was used to perform Delong test and plot the ROC curves of each model. The "rms" package was used to carry out machine learning and build clinical-radiomics nomogram.

RESULTS

Patient Characteristics and Clinical Model Building

According to the inclusion and exclusion criteria, 172 patients were identified in the present study (47 patients with MYCN amplified and 125 patients with MYCN non-amplified). The patients were divided into training group (n = 121) and testing group (n = 51) randomly in a proportion of 7:3, and the characteristics of patients are detailed in **Table 1**. The meaningful

TABLE 1 | Clinicopathologic and radiographic features in training and testing groups.

Features	Training group (n = 121)			Testing group (n = 51)		
	Non-amplified	Amplified	p-value	Non-amplified	Amplified	p-value
Gender (%)			0.736			0.543
Male	51 (57.95)	18 (54.55)		22 (59.46)	7 (50.00)	
Female	37 (42.05)	15 (45.45)		15 (40.54)	7 (50.00)	
Histopathological types (%)			0.629			0.772
Differentiated	28 (31.82)	9 (27.27)		14 (37.84)	4 (28.57)	
Undifferentiated or poorly differentiated	60 (68.18)	24 (72.73)		23 (62.16)	10 (71.43)	
INSS [1] (%)			0.023*			0.43
1	12 (13.64)	0 (0.00)		6 (16.22)	0 (0.00)	
2	7 (7.95)	1 (3.03)		1 (2.70)	0 (0.00)	
3	18 (20.45)	4 (12.12)		8 (21.62)	3 (21.43)	
4	48 (54.55)	28 (84.85)		20 (54.05)	11 (78.57)	
4S	3 (3.41)	0 (0.00)		2 (5.41)	0 (0.00)	
Shimada classification (%)			0.006*			0.071
UFH	58 (65.91)	30 (90.91)		23 (62.16)	13 (92.86)	
FH	30 (34.09)	3 (9.09)		14 (37.84)	1 (7.14)	
Infiltrating across midline (%)			0.17			0.201
Yes	41 (46.59)	20 (60.61)		23 (62.16)	12 (85.71)	
No	47 (53.41)	13 (39.39)		14 (37.84)	2 (14.29)	
Calcification (%)			0.016*			0.454
Yes	61 (69.32)	15 (45.45)		27 (72.97)	8 (57.14)	
No	27 (30.68)	18 (54.55)		10 (27.03)	6 (42.86)	
Necrosis (%)			0.034*			0.099
Yes	51 (57.95)	26 (78.79)		24 (64.86)	13 (92.86)	
No	37 (42.05)	7 (21.21)		13 (35.14)	1 (7.14%)	
Age (P_{25}, P_{75})	24.0 (8.45, 48.0)	25.0 (12.0, 40.3)	0.518	15.00 (7.70, 48.00)	22.0 (13.9, 30.3)	0.598
VMA (P_{25}, P_{75})	21.7 (5.97, 43.36)	5.14 (2.96, 21.76)	0.005*	21.19 (6.39, 41.27)	4.67 (2.36, 11.6)	0.002*

**p <0.05. A chi square test or Fisher's exact test was used for the nominal variable. A Mann–Whitney test was used for the continuous variable with abnormal distribution. [1] INSS, International Neuroblastoma Stage System. Since the morphological characteristics of lesions were firstly interpreted by two radiologists independently and then the difference was resolved by negotiation, the reader agreements were evaluated [Infiltrating across midline: 0.909 (0.878–0.933), Calcification: 0.871 (0.809–0.909), Necrosis: 0.920 (0.890–0.942)].*

characteristics, including the INSS stage, Shimada classification, infiltrating across midline, calcification, necrosis and VMA, were identified as significant with p <0.05 by univariate analyzing. Among them, four characteristics, including Shimada classification (odds ratio (OR) = −2.22, p <0.001), infiltrating across midline (OR = 0.89, p = 0.352), calcification (OR = −1.363, p = 0.0017) and VMA (OR = −0.019, p = 0.0127) were selected using the stepwise selection by multivariate logistic regression analysis (**Table 2**). The AIC criterion was in the multivariate analysis following the stepwise selection method, and the model with smallest AIC value would be chosen. The AIC value of the selected model was 171.4549.

Feature Selection and Machine Learning

A total of 1,218 radiomics features were automatically extracted for each segmented ROI (NP, AP and VP). 734 features were firstly selected with ICCs higher than 0.80 by intra-observer analysis. Before selection of the 734 features, the abnormal or missing values were replaced by the median, and features standardization was applied. And then, mRMR and LASSO were used to select the most optimal features. After the redundant and irrelevant features were removed by mRMR, 30 features from AP, NP and VP were retained. Then LASSO was conducted to identify the final 14 optimal features, including one first-order feature and eight wavelet-transformed features and five LoG-transformed features, to construct the radiomics models. The LASSO includes choosing the regular parameter λ and determining the number of the feature (**Figure 3**). After the number of features determined, the most predictive subset of features was chosen and the corresponding coefficients were calculated (**Figure 4**). The comparison of radiomics signatures derived from two scanners is shown in **Supplementary Figure 1** and **Table 1**, and the performance of the signatures from two scanners was different. Rad-score was calculated by summing the selected features weighted by their coefficients. The final formula of rad-score is showed in **Supplementary Figure 2**.

The ROC curves of the four machine learning models in the training and testing groups are shown in **Figure 5**. In the training group, the AUC among the Logistic, SVM, Bayes and Random Forest was 0.940, 0.940, 0.780 and 0.927, respectively, and the corresponding AUC in the testing group was 0.909, 0.909, 0.729, 0.851, respectively. The Delong test was applied to compare the performance of the four models. There was no significant difference among the Logistic, SVM and Random Forest model (Logistics vs SVM: p = 0.99, Logistic vs Random Forest: p = 0.33, SVM vs Random Forest: p = 0.33), but all better than the Bayes model (p <0.005) (**Table 3**).

TABLE 2 | Univariate and Multivariate logistic analysis in the cohort.

Variable	Univariate analysis		Multivariate analysis	
	OR (95%CI)	P-value	OR (95%CI)	P-value
Gender	0.809 (0.412, 1.589)	0.539		
Histopathological type	1.323 (0.632, 2.771)	0.457		
INSS stage	2.040 (1.255-3.316)	0.004*		
Shimada classification	0.171 (0.058, 0.508)	0.001*	−2.22 (−3.396 to −1.045)	<0.001*
Infiltrating across midline	2.033 (1.003, 4.121)	0.049*	0.890 (0.062–1.718)	0.0352*
Calcification	0.403 (0.202, 0.802)	0.010*	−1.363 (−2.212 to −0.513)	0.0017*
Necrosis	3.250 (1.402, 7.533)	0.006*		
Age	0.996 (0.984, 1.008)	0.467		
VMA	0.981 (0.967, 0.996)	0.013*	−0.019 (−0.035 to −0.004)	0.0127*

*Intercept = 0.24026. *reflected the significant difference with the P value <0.05.*

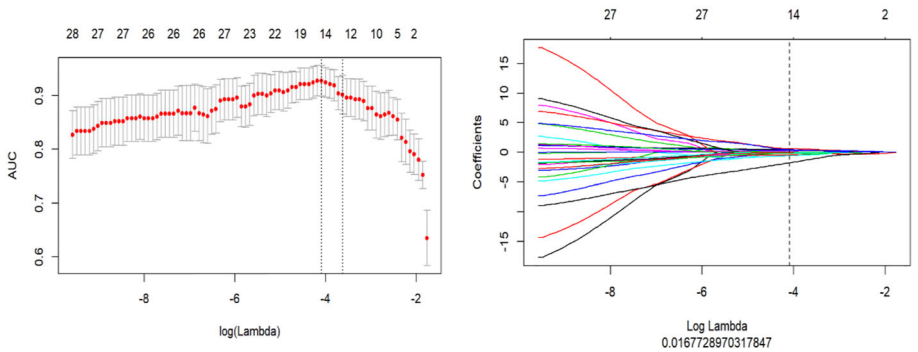


FIGURE 3 | The LASSO includes choosing the regular parameter λ, determining the number of the feature.

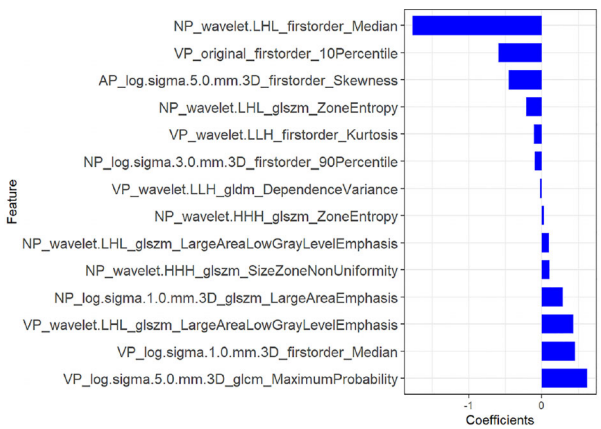


FIGURE 4 | The most predictive subset of feature was chosen and the corresponding coefficients were evaluated in the training group.

Nomogram Building and Evaluating

After performing multivariate logistic regression analysis, the radiomics-clinics nomogram was built by incorporating of clinical-radiological predictors (Shimada classification (odds ratio (OR) = −2.22, p <0.001), infiltrating across midline (OR = 0.89, p = 0.352), calcification (odds ratio (OR) = −1.363, p = 0.0017), VMA (odds ratio (OR) = −0.019, p = 0.0127)) (detailed in **Figure 6**) and the calculated radscore. The ROC analysis of the clinical model, radiomics model, and nomogram is illustrated in **Figure 7** and the comparison of different models is shown in **Table 4**. The nomogram had a superior predictive performance than using the clinical model alone, accompanied with an improved AUC value from 0.770 to 0.946 in the training group and 0.917 to 0.977 in the testing group. The performance of the Logistic radiomics model based on three-phrase is similar to nomogram, but both better than clinical model and radiomics model based on single venous phase (**Table 4**).

DISCUSSION

MYCN amplification status plays a significant role in risk classification of NBs, and NBs with MYCN amplified are usually classified into the high-risk group, where the patients need intensive treatment of operation, radiotherapy and chemotherapy (19). In addition to genetic testing method, radiogenomics, which focusing on establishing the correlation between imaging features and molecular biomarkers, is expected to provide an alternative method to characterize and predict the MYCN amplification status of neuroblastoma noninvasively and inexpensively (8, 20). Previous studies on radiogenomics have demonstrated its potential to predict mutated genes in the solid

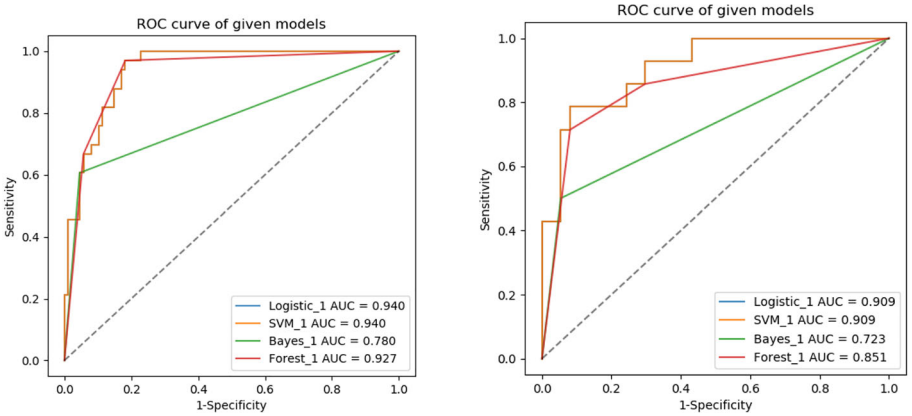


FIGURE 5 | ROC analysis used to evaluate the predictive performance of different radiomics models in the training and testing groups.

TABLE 3 | Comparison of four different machine learning model based on three-phrase in the training and testing groups.

Models	Training group (n = 121)		Validation group (n = 51)	
	AUC (95%CI)	Delong test	AUC (95%CI)	Delong test
Logistic	0.940 (0.901–0.978)	0.99[#]	0.909 (0.824–0.994)	0.99[#]
SVM	0.940 (0.901–0.978)	<0.005[##]	0.909 (0.824–0.994)	<0.005[##]
Bayes	0.780 (0.692–0.867)	<0.005[###]	0.729 (0.581–0.863)	0.051[###]
Forest	0.927 (0.879–0.974)	0.33[####]	0.851 (0.725–0.976)	0.055[####]

[#]indicated the Delong test between the logistic and SVM model; [##]indicated the DeLong test between the SVM and Bayes model; [###]indicated the Delong test between the Bayes and Forest model; [####]indicated the DeLong test between the Logistic and Forest model.

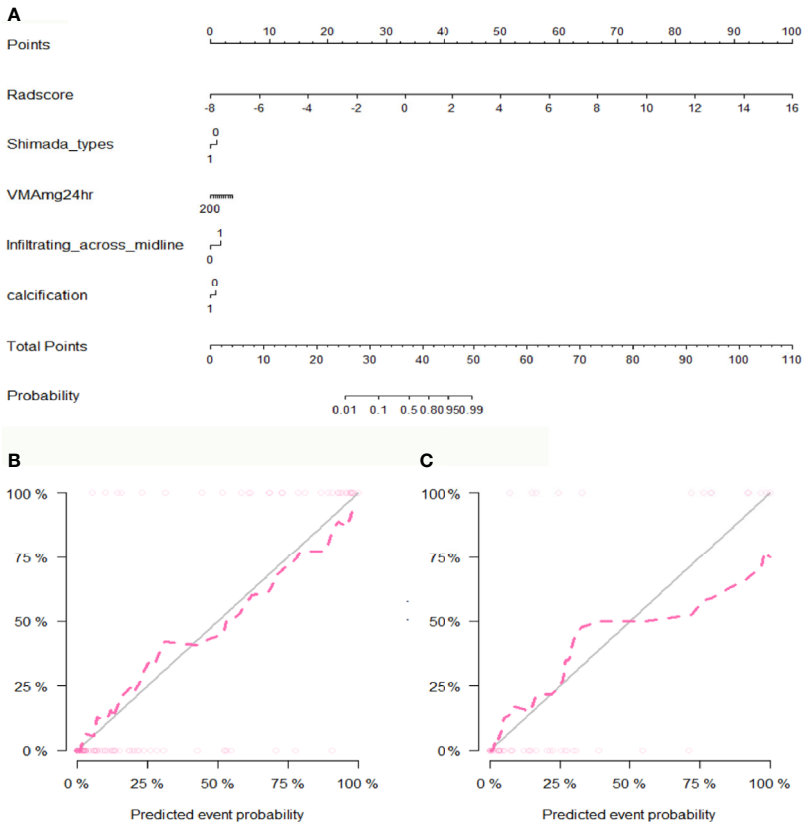


FIGURE 6 | The developed nomogram and calibration curve. **(A)** The nomogram was developed in the training group and incorporated the radiomics signature, Shimada types, VMA, Infiltrating across midline and calcification. **(B)** Calibration plots of the nomogram in the training group. **(C)** Calibration plots of the nomogram in the testing group.

tumors (11, 12, 20, 21). Among adult tumors, CT-based radiogenomics has been widely studied in lung, kidney and liver neoplasms (22–24). However, there have been a few reports on CT radiogenomics in pediatric tumors (8). In this study, data from clinicopathologic parameters (Shimada classification, VMA) and radiographic features (infiltrating across midline, calcification and radiomics features) were selected to develop predictive models for the MYCN amplification of pediatric abdominal NB. Compared to the other study of CT radiogenomics in pediatric NB and GNB (8), we only enrolled the pediatric patients with abdominal NB, because MYCN amplification mostly occurs in abdominal NB (13, 14). Meanwhile, we delineated the whole-tumor ROIs on all slices for the purpose of improving intra- and inter-observer consistency. In addition to first-order and textural features, higher-order features transformed by wavelet and LoG were also extracted to further evaluate the optimal radiomics features correlating with MYCN amplification. Moreover, we also compared the performance of radiomics models developed by four common machine learning methods.

In the present study, quantitative radiomics features, derived from CT images of the whole-tumor ROIs on three-phrase, were extracted and selected by using ICCs, mRMR and LASSO methods. mRMR refers to Maximum Relevance and Minimum Redundancy, which is used to select the optimal features that are most relevant to the classification task but least redundant to each other. mRMR is an algorithm based on mutual information, similar to the Maximum Dependency algorithm. However, unlike Maximum Dependency algorithm, which is not applicable in the case of large number of features, mRMR is especially suitable for high-dimensional data space (25). After the redundant and irrelevant features were removed by mRMR, LASSO regression model was used to prevent overfitting of the selected radiomics features. The main advantage of LASSO method is that it does not compress the variable with larger parameter estimation, while the variable with smaller parameter estimation is compressed exactly to zero. The complexity of the model is controlled through a series of parameters, so as to avoid overfitting. Moreover, the parameter estimation of LASSO analysis has continuity, which is suitable for the model selection of high-dimensional data (26).

Finally, 14 features from three phases were identified as the most predictive subset of feature to construct the radiomics model, including one original first-order feature and eight wavelet-transformed features and five LoG-transformed features. Among the final selected features, the higher-order

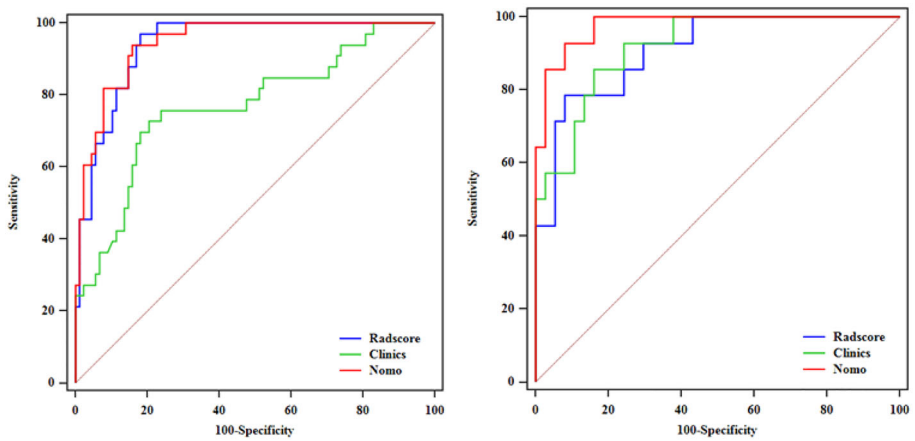


FIGURE 7 | ROC analysis among the clinical model, radiomics model and nomogram in the training and testing groups.

TABLE 4 | Comparison of different models in the training and testing groups.

Variable	Training group				Testing group			
	AUC (95%CI)	Sensitivity	Specificity	Delong	AUC (95%CI)	Sensitivity	Specificity	Delong
VP	0.89 (0.83–0.95)	0.569	0.942		0.88 (0.79–0.97)	0.416	0.852	
Three-phase	0.940 (0.901–0.978)	0.970	0.818	0.0008[#]	0.909 (0.824–0.994)	0.786	0.919	0.884[#]
Clinics	0.770 (0.685–0.841)	0.727	0.795	0.0003[##]	0.917 (0.805–0.976)	0.857	0.834	0.0784[##]
Nomo	0.946 (0.889–0.979)	0.939	0.841	0.343[###]	0.977 (0.89–0.999)	0.928	0.918	0.05[###]

[#]indicated the Delong test between the venous and three-phrase radiomics model; [##]indicated the Delong test between the three-phrase radiomics and Clinics model; [###]indicated the Delong test between the three-phrase radiomics and Nomo combined model.

features filtered by wavelet and LoG filters were obviously superior to the original first-order and textural features. Chen et al. (16) investigated the role of CT-based radiomics to differentiate pelvic rhabdomyosarcoma from yolk sac tumors in children. Among the 10 features selected in their radiomics model based on each phrase, most of the selected features were wavelet-transformed features. Wavelet and LoG are both higher-order statistical methods imposing filter grids on the images, and could possibly reflect more information about vascularity and spiculation of the lesion (27). The principle of wavelet is to put a matrix of linear or radial "waves" on images, while LoG is mostly used to extract features from areas with coarse textural pattern (27). Besides, we evaluated the performance of radiomics features from two scanners, and the results showed that the performance of two signatures is different. One reason for this difference may be that radiomics features are correlated with different scanners from different manufactures, and the other reason may be that the sample size of patients scanned on Lightspeed VCT (GE Healthcare) was relatively too small.

The results of our study showed that radiomics models based on NP, AP and VP images can predict MYCN amplification in pediatric abdominal NB, while the performance of different machine learning radiomics models varies. The AUC in the training group among the Logistic, SVM, Bayes and Random Forest was 0.940, 0.940, 0.780 and 0.927, respectively, and the corresponding AUC in the testing group was 0.909, 0.909, 0.729, 0.851, respectively. The Logistic and SVM models have the best predictive performance with the same value of AUC. According to Delong test, there was no significant difference among the Logistic, SVM and Random Forest model, but all better than the Bayes model. In previous studies, researchers mostly chose one classifier to build radiomics model, and there is no consensus on the best-performing classifier method. Deist et al. (28) compared the performance of different classifiers (decision tree, random forest, neural network, support vector machine, elastic net logistic regression, LogitBoost) in predicting radiotherapy outcomes. In their study, Random forest and elastic net logistic regression performed better than other classifiers. Machine learning classifiers can be used to identify the best combination of radiomics features, while different algorithms have different advantages and disadvantages (29). Therefore, we should choose the optimal machine learning method with overall maximal predictive performance according to the specific clinical application.

In addition, we constructed a nomogram combining clinical parameters, imaging features and radscore. The variables including Shimada classification, VMA, infiltrating across midline and calcification were selected to build the nomogram with the most predictive signatures of radiomics. In this study, UFH was found to show significant correlation with MYCN

amplification. This finding supported the previous study which also found that MYCN-amplified NBs were mostly categorized as UFH group (18). Besides, we found that the majority of NBs infiltrating across midline were MYCN-amplified, which is consistent with the finding of Wu et al. (8), but calcification in NB was found to be related to MYCN amplification. The nomogram had a superior predictive performance than using the clinical model alone, accompanied with an improved AUC value from 0.770 to 0.946 in the training group and 0.917 to 0.977 in the testing group. Besides, we found that the radiomics features used to construct the radiomics models were mostly derived from the NP and VP, so we developed Logistic radiomics model based on single VP. Then, we further evaluated and compared the predictive performance of the nomogram, Logistic radiomics VP model and Logistic radiomics three-phrase model. Although compared with radiomics model, nomogram did not significantly improve the prediction of MYCN amplification, they were both better than clinical model and radiomics model based on single venous phase, which demonstrated that the radiomics features are useful for predicting MYCN amplification and radiogenomics is expected to be involved in risk stratification in NB patients.

Despite our study showed that CT-based radiomics has the potential to predict MYCN amplification in pediatric abdominal NB, there were some limitations. First, this was a retrospective study, which may cause inherent selection bias, especially for those valuable absent clinical indicators that could potentially improve the performance of clinics-radiomics nomogram. Second, we only enrolled 172 patients in the present study because MYCN status has begun to be detected in recent years in our hospital. Previous literatures have shown that MYCN amplification usually occurs in about 20% of neuroblastoma. As a tertiary referral medical center, we have accumulated a certain number of MYCN-amplified cases over the past several years, and the inclusion of more cases will take some time in the future. Third, the CT scans of enrolled patients were performed on two scanners from different manufacturers in the study, from which the derived features have a certain influence on the predictive performance of radiomics models. Fourth, we only choose four common machine learning methods to build radiomics models, and the performance of other classifiers still needs to be evaluated.

CONCLUSIONS

In conclusion, the CT-based radiomics signature is able to predict MYCN amplification of pediatric abdominal NB with high accuracy based on SVM, Logistic and Random Forest classifiers, while Bayes classifier yields lower predictive performance. Thus, one of these three machine learning methods should be the first consideration for researchers to construct predictive models for MYCN amplification of abdominal NB. When combined with clinical and radiographic qualitative features, the clinics-radiomics nomogram can improve the performance of predicting MYCN amplification. With the development of tumor molecular stratification, radiogenomics is expected to provide a promising method to characterize and predict molecular biomarkers noninvasively.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Ethics Committee of the Children's Hospital Affiliated with Chongqing Medical University. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

XC, HW, and LH contributed to conception and design of the study. KH, HD, LZ, TZ, and WY organized the database. HL performed the statistical analysis. HW wrote the first draft of the manuscript. XC and LH reviewed the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was funded by Project Supported by Scientific and Technological Research Program of Chongqing Municipal Education Commission (Grant No.KJQN202000440); Basic Research and Frontier Exploration Project (Yuzhong District, Chongqing, China) (Grant No.20200155); and Science and Health Joint Medical Research Project (Science and Technology Commission and Health Bureau, Chongqing, China) (Grant No.2020FYYX128).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.687884/full#supplementary-material

REFERENCES

1. Van Arendonk KJ, Chung DH. Neuroblastoma: Tumor Biology and Its Implications for Staging and Treatment. *Children (Basel)* (2019) 6(1):12. doi: 10.3390/children6010012
2. Cohn SL, Pearson AD, London WB, Monclair T, Ambros PF, Brodeur GM, et al. The International Neuroblastoma Risk Group (INRG) Classification System: An INRG Task Force Report. *J Clin Oncol* (2009) 27(2):289–97. doi: 10.1200/JCO.2008.16.6785
3. Orr KE, McHugh K. The New International Neuroblastoma Response Criteria. *Pediatr Radiol* (2019) 49(11):1433–40. doi: 10.1007/s00247-019-04397-2
4. Maris JM, Hogarty MD, Bagatell R, Cohn SL. Neuroblastoma. *Lancet* (2007) 369(9579):2106–20. doi: 10.1016/S0140-6736(07)60983-0
5. Sokol E, Desai AV. The Evolution of Risk Classification for Neuroblastoma. *Children (Basel)* (2019) 6(2):27. doi: 10.3390/children6020027
6. Morgenstern DA, Bagatell R, Cohn SL, Hogarty MD, Maris JM, Moreno L, et al. The Challenge of Defining "Ultra-High-Risk" Neuroblastoma. *Pediatr Blood Cancer* (2019) 66(4):e27556. doi: 10.1002/pbc.27556
7. Yanishevski D, McCarville MB, Doubrovin M, Spiegl HR, Zhao X, Lu Z, et al. Impact of MYCN Status on Response of High-Risk Neuroblastoma to Neoadjuvant Chemotherapy. *J Pediatr Surg* (2020) 55(1):130–4. doi: 10.1016/j.jpedsurg.2019.09.067
8. Wu H, Wu C, Zheng H, Wang L, Guan W, Duan S, et al. Radiogenomics of Neuroblastoma in Pediatric Patients: CT-based Radiomics Signature in Predicting MYCN Amplification. *Eur Radiol* (2020) 31(5):3080–9. doi: 10.1007/s00330-020-07246-1
9. Lasocki A, Rosenthal MA, Roberts-Thomson SJ, Neal A, Drummond KJ. Neuro-Oncology and Radiogenomics: Time to Integrate? *AJNR Am J Neuroradiol* (2020) 41(11):1982–8. doi: 10.3174/ajnr.A6769
10. Kuo MD, Yamamoto S. Next Generation Radiologic-Pathologic Correlation in Oncology: Rad-Path 2.0. *AJR Am J Roentgenol* (2011) 197(4):990–7. doi: 10.2214/AJR.11.7163
11. Yan J, Liu L, Wang W, Zhao Y, Li KK, Li K, et al. Radiomic Features From Multi-Parameter MRI Combined With Clinical Parameters Predict Molecular Subgroups in Patients With Medulloblastoma. *Front Oncol* (2020) 10:558162. doi: 10.3389/fonc.2020.558162
12. Iv M, Zhou M, Shpanskaya K, Perreault S, Wang Z, Tranvinh E, et al. MR Imaging-Based Radiomic Signatures of Distinct Molecular Subgroups of Medulloblastoma. *AJNR Am J Neuroradiol* (2019) 40(1):154–61. doi: 10.3174/ajnr.A5899
13. Brisse HJ, Blanc T, Schleiermacher G, Mosseri V, Philippe-Chomette P, Janoueix-Lerosey I, et al. Radiogenomics of Neuroblastomas: Relationships Between Imaging Phenotypes, Tumor Genomic Profile and Survival. *PloS One* (2017) 12(9):e0185190. doi: 10.1371/journal.pone.0185190
14. Vo KT, Matthay KK, Neuhaus J, London WB, Hero B, Ambros PF, et al. Clinical, Biologic, and Prognostic Differences on the Basis of Primary Tumor Site in Neuroblastoma: A Report From the International Neuroblastoma Risk Group Project. *J Clin Oncol* (2014) 32(28):3169–76. doi: 10.1200/JCO.2014.56.1621
15. Meeus E, Zarinabad N, Manias K, Novak J, Rose H, Dehghani H, et al. Diffusion-Weighted MRI and Intravoxel Incoherent Motion Model for Diagnosis of Pediatric Solid Abdominal Tumors. *J Magn Reson Imaging* (2018) 47(6):1475–86. doi: 10.1002/jmri.25901
16. Chen X, Huang Y, He L, Zhang T, Zhang L, Ding H. Ct-Based Radiomics to Differentiate Pelvic Rhabdomyosarcoma From Yolk Sac Tumors in Children. *Front Oncol* (2020) 10:584272. doi: 10.3389/fonc.2020.584272
17. Liu Y, Zhang Y, Cheng R, Liu S, Qu F, Yin X, et al. Radiomics Analysis of Apparent Diffusion Coefficient in Cervical Cancer: A Preliminary Study on Histological Grade Evaluation. *J Magn Reson Imaging* (2019) 49(1):280–90. doi: 10.1002/jmri.26192
18. Altungoz O, Aygun N, Tumer S, Ozer E, Olgun N, Sakizli M. Correlation of Modified Shimada Classification With MYCN and 1p36 Status Detected by Fluorescence in Situ Hybridization in Neuroblastoma. *Cancer Genet Cytogenet* (2007) 172(2):113–9. doi: 10.1016/j.cancergencyto.2006.10.005
19. Pinto NR, Applebaum MA, Volchenboum SL, Matthay KK, London WB, Ambros PF, et al. Advances in Risk Classification and Treatment Strategies for Neuroblastoma. *J Clin Oncol* (2015) 33(27):3008–17. doi: 10.1200/JCO.2014.59.4648
20. Iwatate Y, Hoshino I, Yokota H, Ishige F, Itami M, Mori Y, et al. Radiogenomics for Predicting p53 Status, PD-L1 Expression, and Prognosis With Machine Learning in Pancreatic Cancer. *Br J Cancer* (2020) 123 (8):1253–61. doi: 10.1038/s41416-020-0997-1
21. Khan S, Naim S, Bilwani R, Salem A, Gorlin D, Muhammad A, et al. Radiogenomics and Its Role in Lymphoma. *Curr Hematol Malig Rep* (2020) 15(3):211–24. doi: 10.1007/s11899-020-00577-2
22. Zhou M, Leung A, Echegaray S, Gentles A, Shrager JB, Jensen KC, et al. Non-Small Cell Lung Cancer Radiogenomics Map Identifies Relationships Between Molecular and Imaging Phenotypes With Prognostic Implications. *Radiology* (2018) 286(1):307–15. doi: 10.1148/radiol.2017161845
23. Alessandrino F, Shinagare AB, Bosse D, Choueiri TK, Krajewski KM. Radiogenomics in Renal Cell Carcinoma. *Abdom Radiol (NY)* (2019) 44 (6):1990–8. doi: 10.1007/s00261-018-1624-
24. Banerjee S, Wang D, Kim H, Sirlin C, Chan M, Korn R, et al. A Computed Tomography Radiogenomic Biomarker Predicts Microvascular Invasion and Clinical Outcomes in Hepatocellular Carcinoma. *Hepatology* (2015) 62 (3):792–800. doi: 10.1002/hep.27877
25. Peng H, Long F, Ding C. Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Trans Pattern Anal Mach Intell* (2005) 27(8):1226–38. doi: 10.1109/TPAMI.2005.159
26. Tibshirani RJ. Regression Shrinkage and Selection Via the Lasso. *J ROY STA B* (1996) 58(1):267–88. doi: 10.1111/j.2517-6161.1996.tb02080.x
27. Gillies R, Kinahan P, Hricak H. Radiomics: Images are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
28. Deist TM, Dankers F, Valdes G, Wijsman R, Hsu IC, Oberije C, et al. Machine Learning Algorithms for Outcome Prediction in (Chemo)Radiotherapy: An Empirical Comparison of Classifiers. *Med Phys* (2018) 45(7):3449–59. doi: 10.1002/mp.12967
29. Erickson B, Korfiatis P, Akkus Z, Kline T. Machine Learning for Medical Imaging. *Radiographics* (2017) 37(2):505–15. doi: 10.1148/rg.2017160130

Conflict of Interest: HL is an employee of GE Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Chen, Wang, Huang, Liu, Ding, Zhang, Zhang, Yu and He. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

METHODS
published: 04 June 2021
doi: 10.3389/fonc.2021.690244

A Deep Multi-Task Learning Framework for Brain Tumor Segmentation

He Huang[1†], Guang Yang[2,3†], Wenbo Zhang[1], Xiaomei Xu[1], Weiji Yang[4], Weiwei Jiang[1] and Xiaobo Lai[1*]*

[1] College of Medical Technology, Zhejiang Chinese Medical University, Hangzhou, China, [2] Cardiovascular Research Centre, Royal Brompton Hospital, London, United Kingdom, [3] National Heart and Lung Institute, Imperial College London, London, United Kingdom, [4] College of Life Science, Zhejiang Chinese Medical University, Hangzhou, China

Glioma is the most common primary central nervous system tumor, accounting for about half of all intracranial primary tumors. As a non-invasive examination method, MRI has an extremely important guiding role in the clinical intervention of tumors. However, manually segmenting brain tumors from MRI requires a lot of time and energy for doctors, which affects the implementation of follow-up diagnosis and treatment plans. With the development of deep learning, medical image segmentation is gradually automated. However, brain tumors are easily confused with strokes and serious imbalances between classes make brain tumor segmentation one of the most difficult tasks in MRI segmentation. In order to solve these problems, we propose a deep multi-task learning framework and integrate a multi-depth fusion module in the framework to accurately segment brain tumors. In this framework, we have added a distance transform decoder based on the V-Net, which can make the segmentation contour generated by the mask decoder more accurate and reduce the generation of rough boundaries. In order to combine the different tasks of the two decoders, we weighted and added their corresponding loss functions, where the distance map prediction regularized the mask prediction. At the same time, the multi-depth fusion module in the encoder can enhance the ability of the network to extract features. The accuracy of the model will be evaluated online using the multispectral MRI records of the BraTS 2018, BraTS 2019, and BraTS 2020 datasets. This method obtains high-quality segmentation results, and the average Dice is as high as 78%. The experimental results show that this model has great potential in segmenting brain tumors automatically and accurately.

Keywords: automatic segmentation, brain tumor, deep multi-task learning framework, multi-depth fusion module, magnetic resonance imaging

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln,
United Kingdom

Reviewed by:
Weiping Ding,
Nantong University, China
Chengjin Yu,
Zhejiang University, China

***Correspondence:**
Guang Yang
g.yang@imperial.ac.uk
Xiaobo Lai
dmia_lab@zcmu.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *02 April 2021*
Accepted: *17 May 2021*
Published: *04 June 2021*

Citation:
Huang H, Yang G, Zhang W, Xu X, Yang W, Jiang W and Lai X (2021) A Deep Multi-Task Learning Framework for Brain Tumor Segmentation. Front. Oncol. 11:690244. doi: 10.3389/fonc.2021.690244

INTRODUCTION

Glioma is the most common and aggressive brain tumor. It is diffuse and can spread to any part of the brain, which makes it difficult to detect (1). According to the statistics, the annual incidence of glioma is about 5 cases per 100,000 people (2, 3). Although gliomas are less common than other fatal diseases, they are poorly treated and have a higher mortality rate (4). The World Health

Organization (WTO) divides gliomas into 4 grades, grades 1 and 2 are low-grade gliomas (LGG), and grades 3 and 4 are high-grade gliomas (HGG), which grade 4 glioma patient survival time of less than one year (1). Therefore, early diagnosis plays a vital role.

Magnetic resonance imaging (MRI) is an important method in medical imaging diagnosis (5). It can visualize interesting parts of the brain, generate multi-modal images, and provide more information of the tumor, so it is widely used in clinical diagnosis, treatment, and surgical planning (6). Generally, clinicians manually segment brain tumor regions on two-dimensional MRI slices or planar projections to facilitate the implementation of the next treatment plan (7). However, this method of segmentation is laborious and subjective (6). A patient usually has hundreds of slices, manual segmentation will consume a lot of time and energy for the doctor. Moreover, the boundaries of brain tumors are blurred and it is difficult to distinguish from healthy tissues, which also makes the segmentation results of experts differ (8, 9). Therefore, it is urgent to use computer technology to realize automatic segmentation of medical images. In recent years, with the development of deep learning, convolution neural networks (CNN) has been widely used in the field of medical image segmentation, and CNN has become the main method for its high efficiency and time-saving. In this paper, brain tumor segmentation is a multi-category segmentation task, which often requires a more complex network structure to obtain the ideal segmentation result. Based on this, we propose a new network structure to segment brain tumors, called the deep multi-task learning framework. The contributions of this article are mainly in the following aspects:

- In this paper, we propose a deep multi-task learning framework to improve the discontinuity problem of mask boundary prediction. It adds a distance transform decoder based on V-Net, which regularizes the mask decoder to ensure the smoothness of segmentation prediction.
- We also propose the use of linear weighting to combine the loss function of mask prediction and distance estimation tasks, and discuss the impact of weight changes on segmentation accuracy. Besides, we also set weights in the loss function of mask prediction to reduce the model's attention to the background.
- We integrate the multi-depth fusion module into the down-sampling stage, which can effectively fuse global information and local information, improving the ability of the network to extract features. In the control group experiment, we remove the multi-depth fusion module and find that the accuracy was reduced, which can verify the effectiveness of the module in improving accuracy.

The rest of this article is as follows. Section 2 describes the development of image segmentation, related research on MRI segmentation of brain tumors, and a summary of the work of the article. In Section 3, we elaborated on the proposed method, including data preprocessing, the principle of the model, and the loss function. In the fourth section, we give the details of the experiment. Then we presented and discussed the experimental results in Section 5, and finally summarized the conclusions in Section 6.

RELATED WORKS

Traditional Image Segmentation Method

The earliest and most traditional image segmentation method is based on threshold segmentation. The basic principle of the threshold segmentation method is to divide the image pixels into the target area and the background area by setting the characteristic threshold. Taheri S. et al. used a threshold-based method to segment three-dimensional brain tumors, and adopted two threshold update schemes for searching and self-adaptation, achieving automatic or semi-automatic segmentation according to the complexity of the tumor shape (10). A more widely used algorithm is a segmentation algorithm based on edge detection, which is one of the most studied methods. Max W.K. et al. proposed an edge detection method based on weighted local variance for blood vessel boundary segmentation, which is robust to changes in edge intensity contrast (11). But the edge segmented by this method is not continuous.

Machine Learning

The segmentation method of machine learning is mainly divided into two categories: supervised learning and unsupervised learning. Supervised learning includes KNN, Bayes, and ANN algorithms. Unsupervised learning mainly includes some clustering methods, such as K-means, FCM, etc. Anbeek P. et al. applied the KNN algorithm to skull MRI to segment multiple sclerosis lesions. This method uses voxel position and signal strength to determine the probability of each voxel lesion to generate probabilistic segmentation images (12). In order to improve the segmentation of brain PET images, Xia Y. et al. proposed to incorporate the *a priori* anatomical knowledge represented by the probabilistic brain atlas into the variational Bayes to segment gray matter, white matter, and cerebrospinal fluid in brain PET-CT images (13). Franklin S.W. et al. used ANN technology based on Gabor and moment-invariant features to segment the retinal blood vessels in the fundus to accurately obtain the width of the blood vessels (14). In medical image segmentation, the most commonly used clustering method is k-means clustering. Moftah H.M. et al. proposed an adaptive k-means clustering method, which maintains the best results in the iterative process and can effectively segment MR breast images (15). The fuzzy C-means algorithm based on objective function is also commonly used. Chen W. et al. applied the fuzzy c-means (FCM) clustering method to the wind field to segment breast lesions from MRI-enhanced images (16).

Deep Learning

With the development of deep learning, the CNN is gradually applied to image segmentation, which greatly improves the

accuracy of image segmentation. Fully Convolutional Networks (FCN) is the first network structure that successfully uses deep learning for image semantic segmentation. FCN converts the fully connected layers in classification networks such as AlexNet, VGG Net, and GoogLeNet into convolutional layers and applies them to segmentation tasks to achieve pixel-level semantic segmentation (17). Badrinarayanan V. et al. proposed a new semantic pixel segmentation network structure SegNet, which is based on the DeconvNet. Their innovation is to input feature maps during the up-sampling process to better restore the information lost during the down-sampling process (18). U-Net is a segmentation network with a simple structure proposed by Ronneberger et al., which is widely used in medical image segmentation because it can adapt to a small training set. It is similar to FCN, but the difference is that U-Net uses features to be stitched together in the channel dimension to form a thicker feature, while FCN adds features point by point (19). Although CNN is very popular, most methods can only process two-dimensional images, and most medical data used in clinical practice is composed of three-dimensional volumes. Milletari et al. proposed a three-dimensional image segmentation network based on U-Net. Their network is trained end-to-end on the 3D MRI of the prostate and can predict the segmentation of the entire 3D image at once (20). Liu et al. proposed a new convolutional neural network, which consists of three independent sub-networks, including an improved ResNet50, a feature pyramid attention network and a naive decoder network. The three networks are connected to form an end-to-end prostate segmentation model (21). Ding et al. proposed a fuzzy information deep learning segmentation (Fl-DL-Seq) network to segment infant brain tissue. They use the volumetric fuzzy pooling (VFP) layer to model the local fuzziness of the volume convolution map by fuzzing, accumulating and deblurring the neighborhood of the adjacent feature map (22).

Brain Tumor Segmentation

Although a large number of neural network structures have achieved high segmentation performance in the segmentation field, they are not necessarily adaptable to the field of brain tumors. Due to the complexity of multi-modal brain tumors, in order to obtain clinical segmentation effects, the network structure must be designed according to the characteristics of brain tumor MRI images, for this reason, many experts have done a lot of research. Lachinov D. et al. proposed a deep cascade method for automatic brain tumor segmentation, which modified the three-dimensional U-Net architecture to effectively process multi-modal MRI input. They used multiple encoders to make each individual mode independently generating a corresponding feature map (7). Feng X. et al. proposed to use a set of three-dimensional U-Net with different hyperparameters to segment brain tumors. They trained the six networks with different encoder/decoder block numbers, different input patch sizes, and different loss weights, and finally performed integrated modeling (23). Lele C. et al. proposed a model based on 3D convolutional neural networks. Their innovation lay in extracting features from two different scales to obtain multi-scale context information. They also proposed a new structure, that was, according to the characteristics of the brain tumor lesion area, the lesion sub-regions were stratified (24). Zhou C.H. et al. proposed a lightweight deep model based on the model cascade (MC) strategy, a one-time multi-task network (OM-Net), which could better solve the problem of class imbalance. Besides, they also designed a cross-task guided attention (CGA) module that could adaptively recalibrate the channel characteristics (25). Myronenko A. et al. designed a variational autoencoder (VAE) branch to reconstruct the input image in a network based on the encoder-decoder architecture. The function of the VAE branch was to jointly reconstruct the input image and the segmented image to standardize the shared encoder (5). Zhang et al. proposed a brain tumor segmentation model based on multi-encoders. Each modal image corresponds to a down-sampling path. This one-to-one feature extraction method reduces the complexity of the feature extraction (26).

Our Work

Segmentation of brain tumor MRI data based on multimodality is challenging for the following reasons. First of all, brain tumors may show similar characteristics with glial hyperplasia and stroke, which is easy to cause confusion (1). Second, brain tumors can appear in any part of the brain. Besides, the size, shape, and appearance of brain tumors in different patients are also different, which increases the difficulty of segmentation (27). Finally, the main difficulty of the brain tumor segmentation task is the imbalance between the classes. Since the lesion area is very small in most cases, the background area dominates, resulting in low segmentation accuracy (24). All these make the task of brain tumor segmentation more difficult.

In this article, in order to deal with the above challenges, we propose a deep multi-task learning framework combined with a multi-depth fusion module. It is a derivative of the V-Net network structure. The traditional V-Net has only one decoder, which will produce discontinuous boundaries in the segmentation results. Therefore, we propose a parallel decoder architecture to perform distance estimation while predicting the mask to ensure the smoothness of the prediction result. As shown in **Figure 1A**, the first decoder is used to predict the mask, and the second decoder is used to estimate the distance map. The main function of the distance decoder is to regularize the mask prediction path to make the boundary of the mask smooth and continuous. In order to combine the two tasks, we propose a new loss function, which consists of two parts: the categorical focal loss of the mask decoder block and the mean square error of the distance transform decoder block. The final loss function is the weight of the above two sum up. We also set some weights for Categorical Focal loss to reduce the attention of the model to the background area and alleviate class imbalance. In order to improve the ability of the model to extract features, we integrate the multi-depth fusion module into the encoder (28). This module averages and fuses multi-level feature signals, which can effectively capture global features and local features. Based on the brain tumor data provided by the BraTS 2018, BraTS 2019, and BraTS 2020, we evaluate the model and compare it with the methods proposed by other researchers

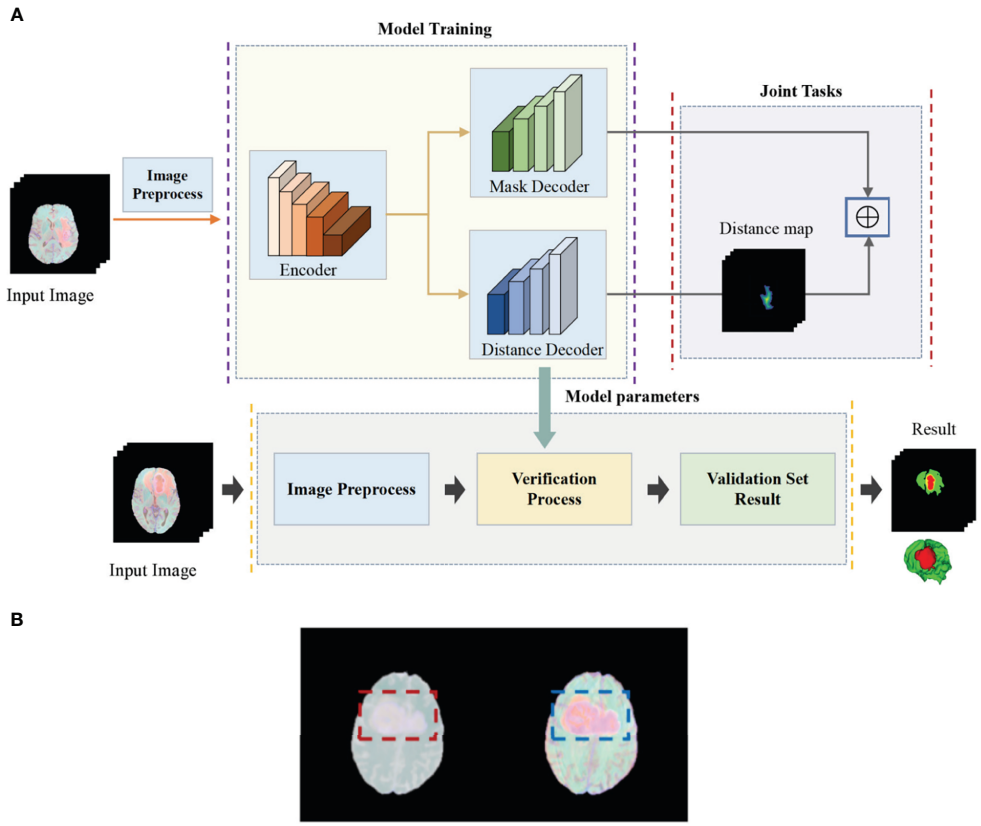


FIGURE 1 | **(A)** Flow chart of our deep multi-task learning framework for brain tumor segmentation. **(B)** Brain tumor images before and after the standardization.

and participating teams. Experimental results show that this method has a good segmentation effect.

METHODS

The task of this paper is to segment brain tumors from three-dimensional MRI images. In order to obtain higher segmentation accuracy, we propose a deep multi-task learning framework based on V-Net and integrate the attention module in the encoder. **Figure 1A** shows the flow chart of all work. In this section, we will introduce in detail the preprocessing process and the structure of the deep multi-task learning framework.

Pre-Processing Steps

The BraTS data all have four MRI sequences with different contrasts: fluid attenuation inversion recovery (FLAIR), T1-weighted (T1), T1-weighted contrast enhancement (T1-CE), and T2-weighted (T2), different contrast may lead to slow convergence or disappearance of gradients during training, so images must be standardized. We standardize the four modes separately and then merge the modes. Here we choose Z-score for standardization, that is, the image minus the mean divided by the standard deviation. In **Figure 1B**, we show the comparison images before and after standardization. The left side is before standardization and the right side is after standardization (**Figure 1B**). It can be seen that the characteristics of the tumor are more obvious after standardization. The standardized formula is as follows:

$$\hat{X} = \frac{X - \bar{X}}{\sigma}, \tag{1}$$

among them, $\hat{X}$ represents the normalized image, X represents the original image, $\bar{X}$ represents the mean value of the image, and σ represents the standard deviation of the image.

After standardization, the MRI images of four modal sequences with a size of 240×240×155×1 are combined to generate a three-dimensional image of four channels, and the combined image size is 240×240×155×4. Then split the Mask image, that is, each type of label image is used as a separate channel image. The original image size is 240×240×155×1, the generated size is 240×240×155×4, and then the one-hot operation is performed on each channel. The non-zero value in channel 0 is the background area, and the non-zero value in channel 1 is the gangrene area, the non-zero value in channel 2 is the edema area, and the non-zero value in channel 3 is the enhanced tumor area. Finally, a patch operation is performed on the image and the mask, and several images and masks with a size of 128×128×64×4 are generated.

The Deep Multi-Task Learning Framework

The basic idea of a deep multi-task learning framework is to implement feature information extraction at different resolution levels through alternately stacked convolutional layers and down-sampling layers. Then, the features extracted by the deconvolution joint encoder are used to realize the step-by-step resolution restoration and feature information restoration. Distance transform encoder helps smooth segmentation prediction and attention module helps feature extraction. Its network structure is shown in **Figure 2A**. We will elaborate on the innovation of the model from three aspects: the multi-depth fusion module, the distance transform decoder, and the loss function.

Multi-Depth Fusion

The multi-depth fusion module was originally used to perform whole-heart segmentation of CT images. Ye C. et al. applied the module to 3D U-Net to obtain the most advanced results (28). It is incorporated in the down-sampling part of our model. The structure of the multi-depth fusion block is shown in **Figure 2A**, and all convolution operations in the structure use a convolution kernel with a size of 3×3×3 and a stride of 1.

The operation of superimposing and averaging the signal that has undergone two convolution operations on the feature map and the signal that has undergone one convolution operation is called the fusion of feature maps of different depths. The input feature map is subjected to two different depth feature map fusion operations, and the input feature map is subjected to a convolution operation, and then the two signals are superimposed and averaged. Finally, the output signal is combined with the input feature map again as the final output of the module, its specific process is shown in **Figure 2B**.

This module averagely merges the characteristic signals of different depths, which can continuously merge local and global information. Compared with simply merging feature maps of different depths, the multi-depth fusion module adopts a better iterative layered fusion method. This architecture ensures that the deep feature map can effectively receive feature information from the shallow feature map. The final signal input and output resolution remain the same. The original encoder block consists of two to three convolutional layers and a downsampling layer. Now we place the multi-depth fusion module in front of each downsampling layer. The convolution still uses batch normalization and ReLU activation functions, just change the location of the ResNet connection. The specific structure is shown in **Figure 2C**.

Distance Transform Decoder

Our proposed network has two decoder modules with similar structures, and each decoder module is assigned a different task. The mask decoder module performs training mask segmentation according to pixel classification tasks, and the distance transform decoder module performs regression tasks to realize distance map estimation.

The distance transform decoder module is similar in structure to the mask decoder module. The image size is restored stage by stage by alternately stacking the deconvolution layer and the convolution layer, and the feature information extracted by the block-level connection joint encoder module is used to improve the predictive performance of brain tumor area contour. The difference with the mask decoder is the output channel in the distance transform decoder block is 3, which is equal to the number of input categories of the mask decoder block. We show in **Figure 3** an example of the mask decoder output and the distance decoder output comparison. In all the figures illustrate the segmentation results, red color represents the tumor core (necrosis), yellow color represents the active tumor and green regions are the edema.

Loss Function

The loss function of the model consists of two parts, the categorical focal loss of the mask decoder block and the mean square error of the distance transform decoder block. The final loss function learned by the optimizer is the weighted sum of the coefficients of the above two losses, where the distance graph prediction regularizes template prediction. The overall loss function formula is as follows:

$$L_{\text{total}} = \lambda_1 L_{\text{mask}} + \lambda_2 L_{\text{dist}}, \tag{2}$$

where λ_1 and λ_2 are the scaling factors, L_{mask} is the loss function of the mask decoder, and L_{dist} is the loss function of the distance transform decoder. The loss function of the mask decoder is shown in (3) and (4),

$$L_{\text{FL}} = -y\log y' \times (1 - y')^{\gamma} \tag{3}$$

$$L_{\text{mask}} = L_{\text{FL}} \times W, \tag{4}$$

where L_{FL} is the pixel-by-pixel classification loss, y represents the true label, y' represents the predicted value after activation by Softmax, and W represents the category weight.

We choose categorical focal loss to solve the problem of imbalance between the foreground and background categories of brain tumor images. It adds a gamma factor to the two-class cross-entropy loss. Here we set the gamma factor to 2, so that the model reduces the loss of background voxels, and makes the model pay more attention to the target voxels that are difficult to segment and easy to error. Finally, in order to further adjust the category imbalance, we assign a specific weight to each type of label, the background label weight is assigned a value of 0.1, and the ET, WT, TC area label weight is assigned a value of 1.0. For the loss function of the distance transform encoder, we refer to the practice of Balamurali M. et al. (29), using the mean square error loss. The loss function of the distance transform decoder is shown in (5),

$$L_{\text{dist}} = \sum_{x \in \Omega} (\hat{D}(x) - D(x))^2, \tag{5}$$

where x represents the pixel, Ω is the number of voxels in the whole brain. $\hat{D}(x)$ is the distance estimation map after activation of the Sigmoid function and $\hat{D}(x)$ is the ground truth distance map.

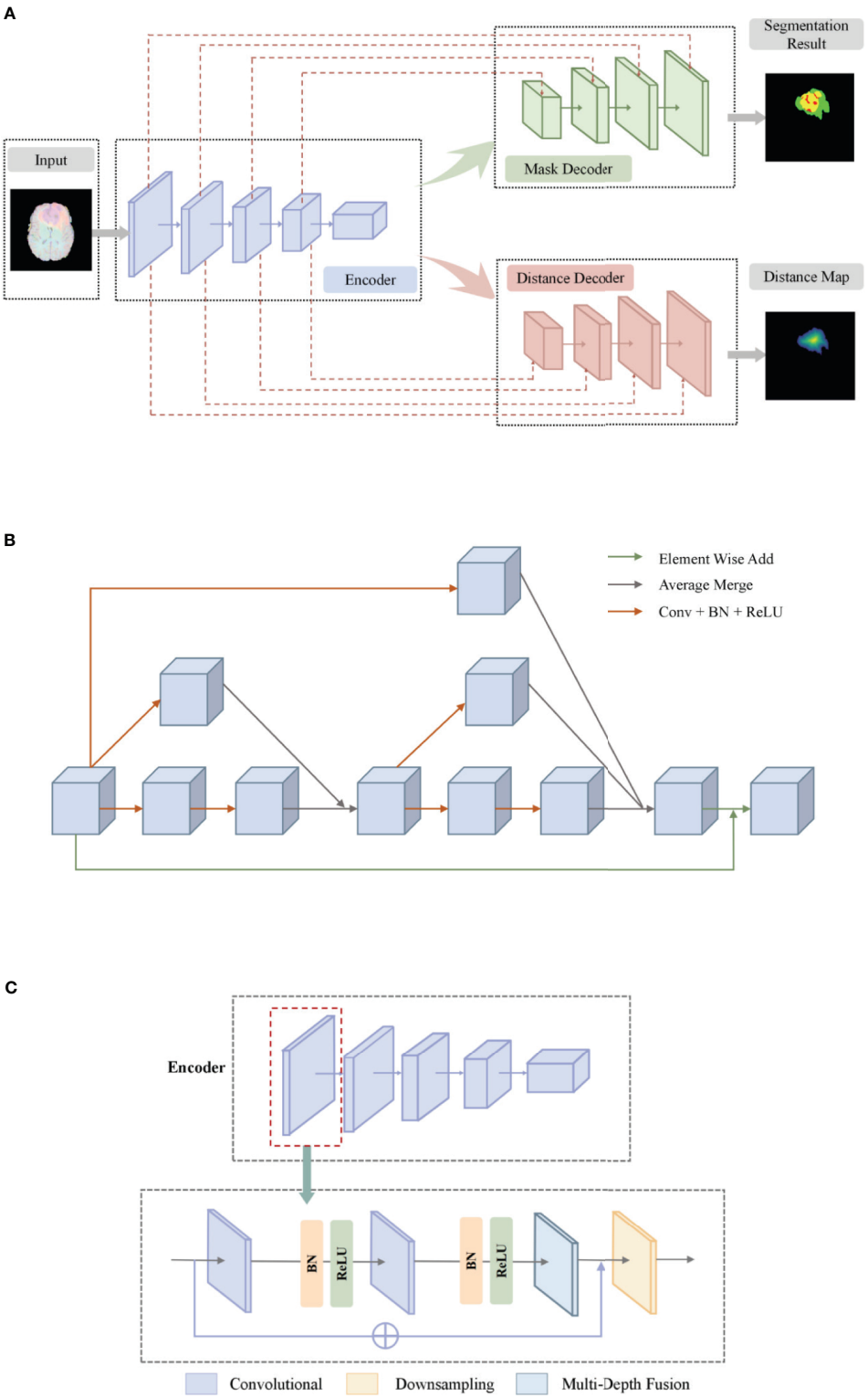


FIGURE 2 | **(A)** The network structure of our deep multi-task learning framework. **(B)** Structure diagram of the multi-depth fusion module. **(C)** A detailed illustration of the encoder.

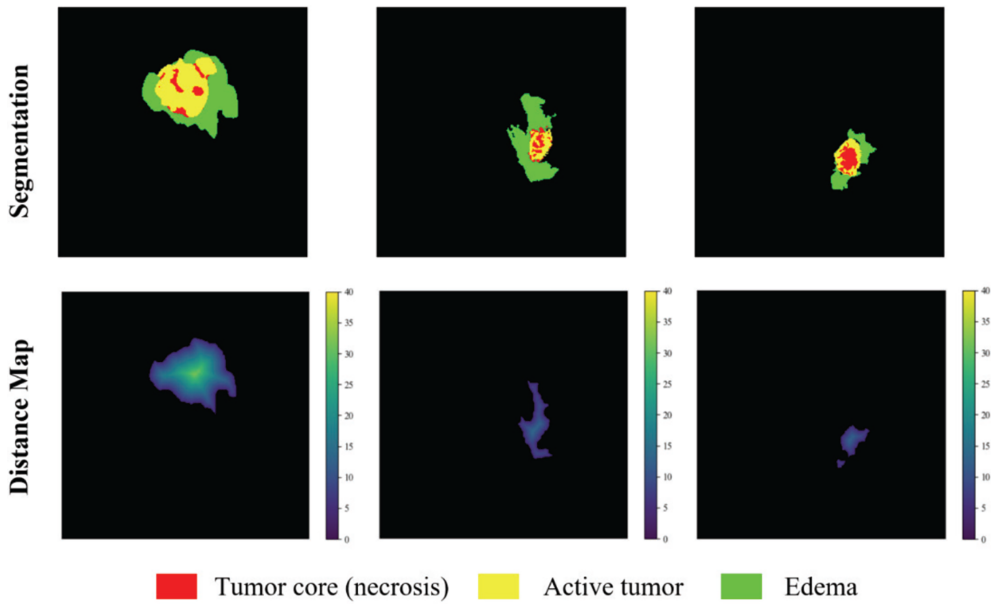


FIGURE 3 | The example output results of the two decoders. Red color represents the tumor core (necrosis), yellow color represents the active tumor and green regions are the edema.

DATASET AND EXPERIMENTS

Datasets

We evaluated our model on three different datasets, BraTS 2018, BraTS 2019, and BraTS 2020. These datasets all have two types of brain tumor data, namely high-grade glioblastoma (HGG) and low-grade glioma (LGG). The MRI of each sample contains four modes: fluid attenuation inversion recovery (FLAIR), T1 weighting (T1), T1-weighted contrast-enhanced (T1-CE), and T2 weighting (T2). The ground truth mask of the data has a necrotic area, edema area, and enhancement area. Our task is to segment the sub-regions formed by the nesting of three targets, the enhancing tumor (ET), the whole tumor (WT), and the tumor core (TC). They are all divided into the training set and unlabeled validation set by the organizer. Among them, BraTS2019 and BraTS2020 use the same test data, while the test data used by BraTS2018 is part of the 125 test data used by the former. The division of each dataset is shown in **Figure 4**.

Evaluation Metrics

We follow the evaluation indicators of the brain tumor segmentation challenge in 2020, using Dice coefficient, sensitivity, specificity, and Hausdorff95 distance to measure the performance of the model. The Dice coefficient is an indicator of the overall evaluation, and its formula is defined as:

$$\text{Dice} = \frac{2\text{TP}}{\text{FN} + \text{FP} + 2\text{TP}}, \tag{6}$$

where TP, FN, and FP represent the number of voxels of true positive, false negative, and false positive respectively. Sensitivity is used to measure the proportion of voxels in the tumor area that are correctly labeled, which can indicate the accuracy of the model segmentation of the target area, which is defined as:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{7}$$

Specificity represents the accuracy of the background voxel being correctly predicted, it can measure the ability of the model to predict the background area, defined as:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \tag{8}$$

where TN is the number of voxels with true negatives. Hausdorff95 distance measures the similarity between actual voxels and predicted voxels. The smaller the value, the closer the prediction is to the reality, it is defined as:

$$\text{Haus95}(X, Y) = \max\left\{ \max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X} d(x, y) \right\} \tag{9}$$

where X is the volume of ground truth, Y is the predicted volume, and d(.,.) is the distance from point x to point y.

Post-Processing Steps

The most difficult part of BraTS dataset segmentation is to distinguish between enhanced tumors and tumor cores, especially when some tumor patients do not have enhanced tumors. If there is no enhanced tumor label in the ground truth and prediction, the BraTS Challenge will set the Dice score to 1. But if the ground truth does not enhance the tumor, even if there is only one false positive voxel in the prediction, the Dice score

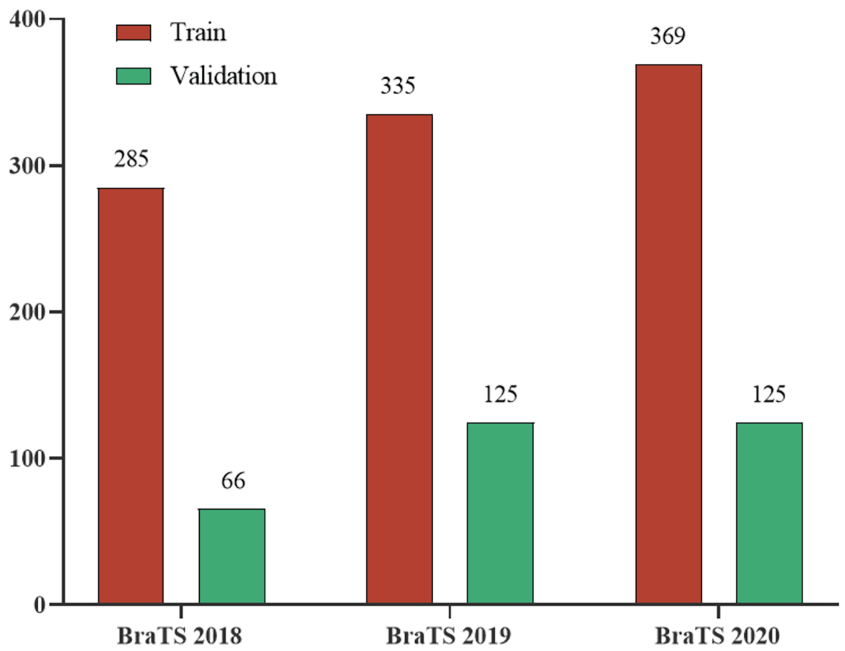


FIGURE 4 | The division of the three datasets in the BraTS Challenge.

will be 0 (30). This greatly affects the Dice value in the ET area. Therefore, we post-process the segmentation results. If the total number of predicted enhanced tumors is less than 500, we replace all enhanced tumor voxels with tumor cores.

Experimental Details

All our implementations are based on Tensorflow 1.13.1, which is currently one of the most mainstream deep learning frameworks. Besides, we use the Adam optimizer to train the model. The specific details are that the entire network is trained for a total of 500,000 steps, and each training set is traversed about 10 times. After each traversal of the training set, the order of the data is randomly shuffled to enhance the robustness of training. The learning rate is initially set to 0.0001, and the training set is traversed twice, reducing to half of the original. Finally, we use Mean Dice as the evaluation index for our training and output the loss value and accuracy index every 10 steps to achieve effective supervision of model training. At the same time, the model outputs the segmentation effect map and the corresponding real label map every 1000 steps and saves the parameter model. In this way, the segmentation effect of the model is further monitored through the visualization method. The segmentation effect map and the real label map during the training process are shown in **Figure 5**.

TensorFlow platform is used for algorithm development using PyCharm with Python 3.6. The runtime platform processor is Intel (R) Xeon (R) Silver 4210 CPU @2.20GHz with 128GB RAM and Nvidia Titan RTX GPU on a 64-bit Windows 10 workstation. All algorithms are trained and tested using the same GPU and environment.

RESULTS AND DISCUSSION

Results

The brain tumor segmentation method proposed in this paper is experimentally evaluated on three different datasets, namely BraTS 2018, BraTS 2019, and BraTS 2020. The preprocessing and segmentation process of these three datasets are the same. There are multi-modal imaging protocols in these datasets: Flair, T1, T1-CE, T2. The data comes from different centers, and the magnetic field strength is also different. The ground-truth was manually created by experts (31), including three nested sub-regions: the enhancing tumor (ET), the whole tumor (WT), and the tumor core (TC). In order to obtain better results, we train on 3D volumes. We use the Dice coefficient, sensitivity, specificity, and Hausdorff95 distance as the evaluation indicators of the model. **Tables 1** and **2** show the average results of the model on the validation sets, among them, "Post" represents the post-processing process added. The results show that the model achieves good segmentation results and has good robustness. Among them the segmentation accuracy of the WT and TC regions is high, but the segmentation accuracy of the ET region is slightly lower. The reason may be that the boundary between the ET region and the WT region is not obvious (32). We can see that post-processing has greatly improved the Dice accuracy in the ET area, with an average increase of 4%. The Dice value of all regions in the verification set of BraTS 2020 is greater than 0.75, especially the WT region is 0.86, which has exceeded the average level of existing methods. In particular, the specific value is stable at a high level, indicating that the model has stable performance in predicting the background area. At the same time, the

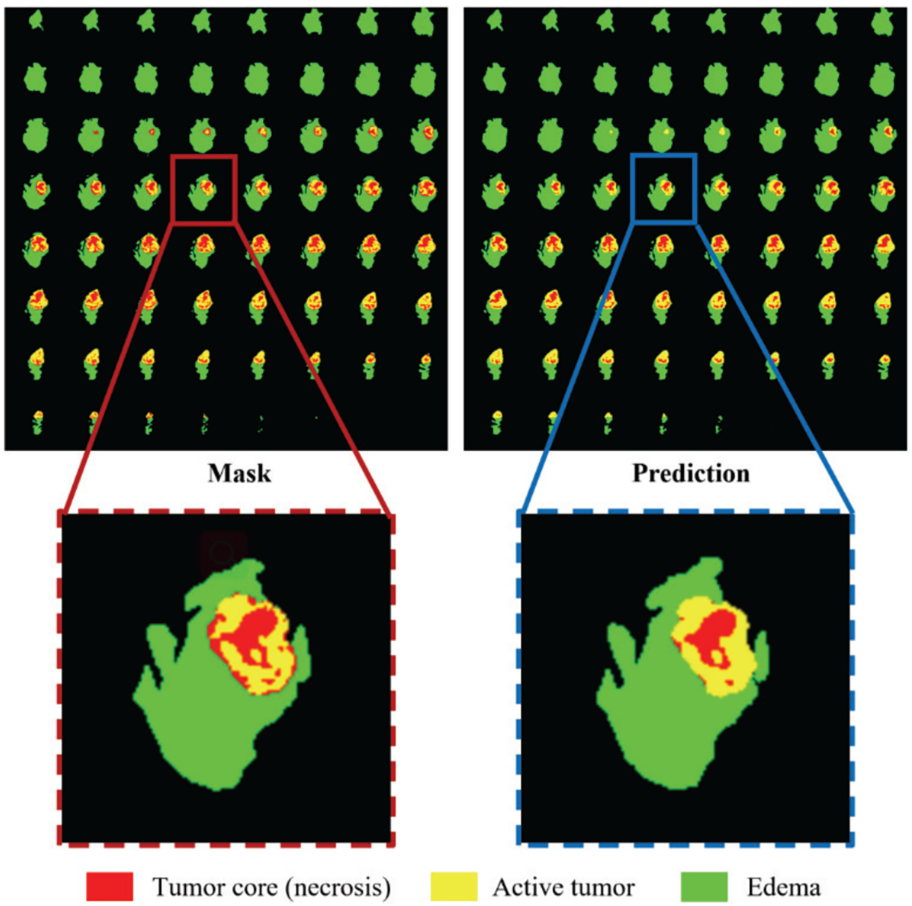


FIGURE 5 | Comparison of the ground truth mask and the prediction during training. Red color represents the tumor core (necrosis), yellow color represents the active tumor and green regions are the edema.

TABLE 1 | Comparison of Dice and Hausdorff95 post-processing of three validation sets.

	Dice			Hausdorff95		
	ET	**WT**	**TC**	**ET**	**WT**	**TC**
BraTS 2018	0.687	0.801	0.759	10.4	13.2	15.2
BraTS 2018+Post	0.717	0.801	0.759	9.9	13.2	15.2
BraTS 2019	0.700	0.827	0.788	6.6	8.5	9.2
BraTS 2019+Post	0.730	0.827	0.788	6.1	8.5	9.2
BraTS 2020	0.700	0.860	0.772	39.1	6.7	15.1
BraTS 2020+Post	0.750	0.860	0.772	34.6	6.7	15.1

TABLE 2 | Comparison of Sensitivity and Specificity post-processing of three validation sets.

	Sensitivity			Specificity		
	ET	**WT**	**TC**	**ET**	**WT**	**TC**
BraTS 2018	0.789	0.962	0.800	0.996	0.977	0.995
BraTS 2018+Post	0.829	0.962	0.800	0.996	0.977	0.995
BraTS 2019	0.758	0.967	0.801	0.997	0.981	0.996
BraTS 2019+Post	0.798	0.967	0.801	0.997	0.981	0.996
BraTS 2020	0.749	0.958	0.791	0.999	0.998	0.999
BraTS 2020+Post	0.800	0.958	0.791	0.999	0.998	0.999

sensitivity has also reached a very high level, and the accuracy of specificity is not much different, indicating that the model has a small difference in predicting target area and predicting background, and can effectively alleviate the problem of class imbalance in brain tumor segmentation. In MRI segmentation, multi-modal brain tumors are one of the most challenging tasks. Although our experimental results have some gaps compared with the top methods, we still achieve high accuracy.

Figure 6 shows a combination of scatter plots and violin plots of the Dice and Hausdorff95 evaluation indicators in the three validation sets of BraTS. From the violin chart, it can be seen that the results are mainly concentrated in one area, and

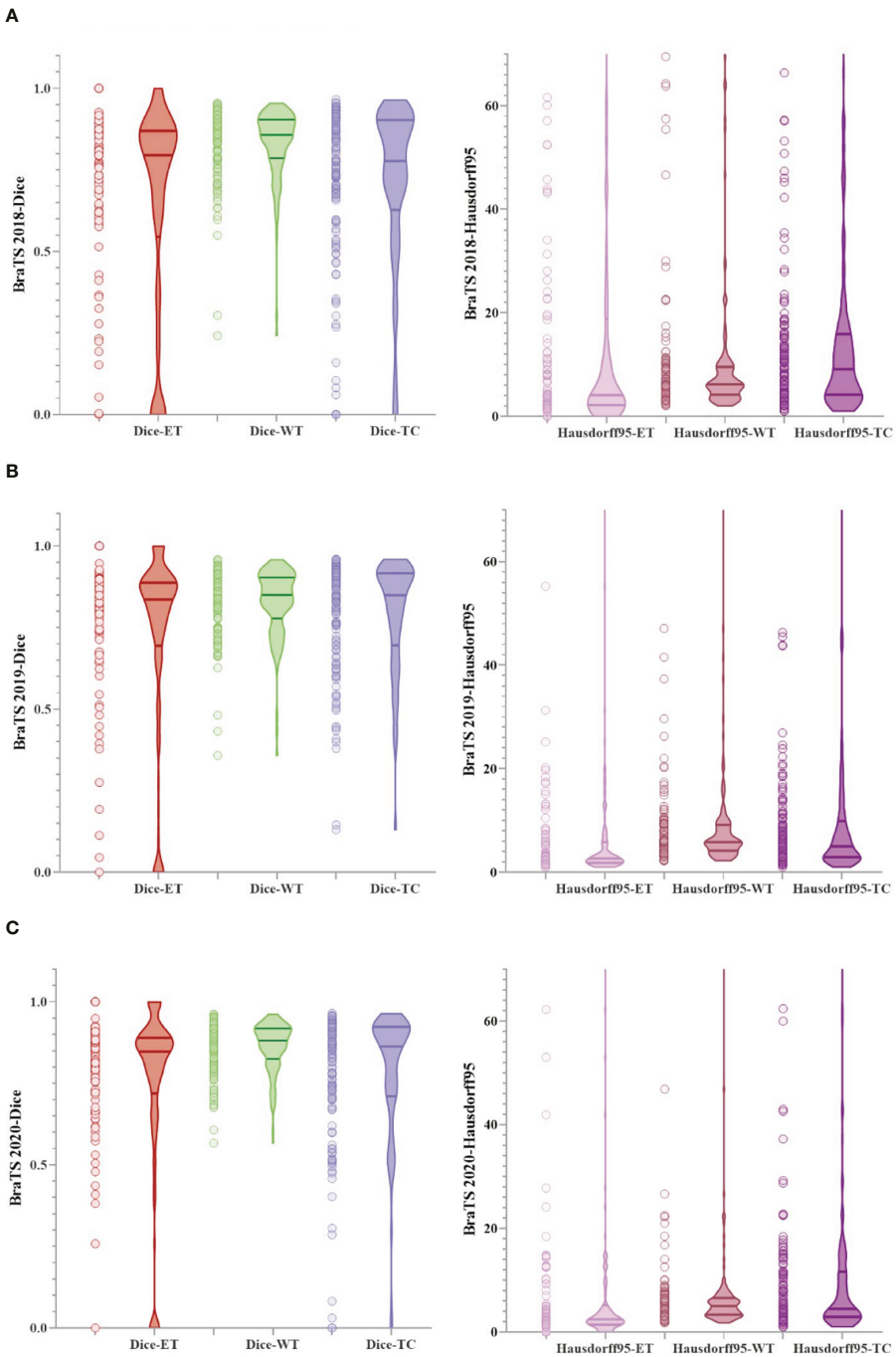


FIGURE 6 | The result of the evaluation metrics of the validation set. **(A)** Dice and Hausdorff95 at BraTS 2018, **(B)** Dice and Hausdorff95 at BraTS 2019 and **(C)** Dice and Hausdorff95 at BraTS 2020.

the median is obviously greater than the average, so the results are leftward and there are outliers. In the scatter plot, the data points are concentrated in areas with higher accuracy, which indicates that the model has a strong ability to predict individual situations.

We selected some representative segmentation results in the validation set for display, as shown in **Figure 7**. Among them (a) is the validation set of BraTS 2018, (b) is the validation set of BraTS 2019, and (c) is the validation set of BraTS 2020. For a more comprehensive display, we marked the Dice value of the ET area. **Figure 7** shows the segmentation results of the model in the validation set samples. From these examples, we can see that the model has good segmentation results for brain tumors of different sizes, positions, and shapes, and the predictive ability of the model is not affected by the intensity of MRI slice scans. Overall, the model has high performance.

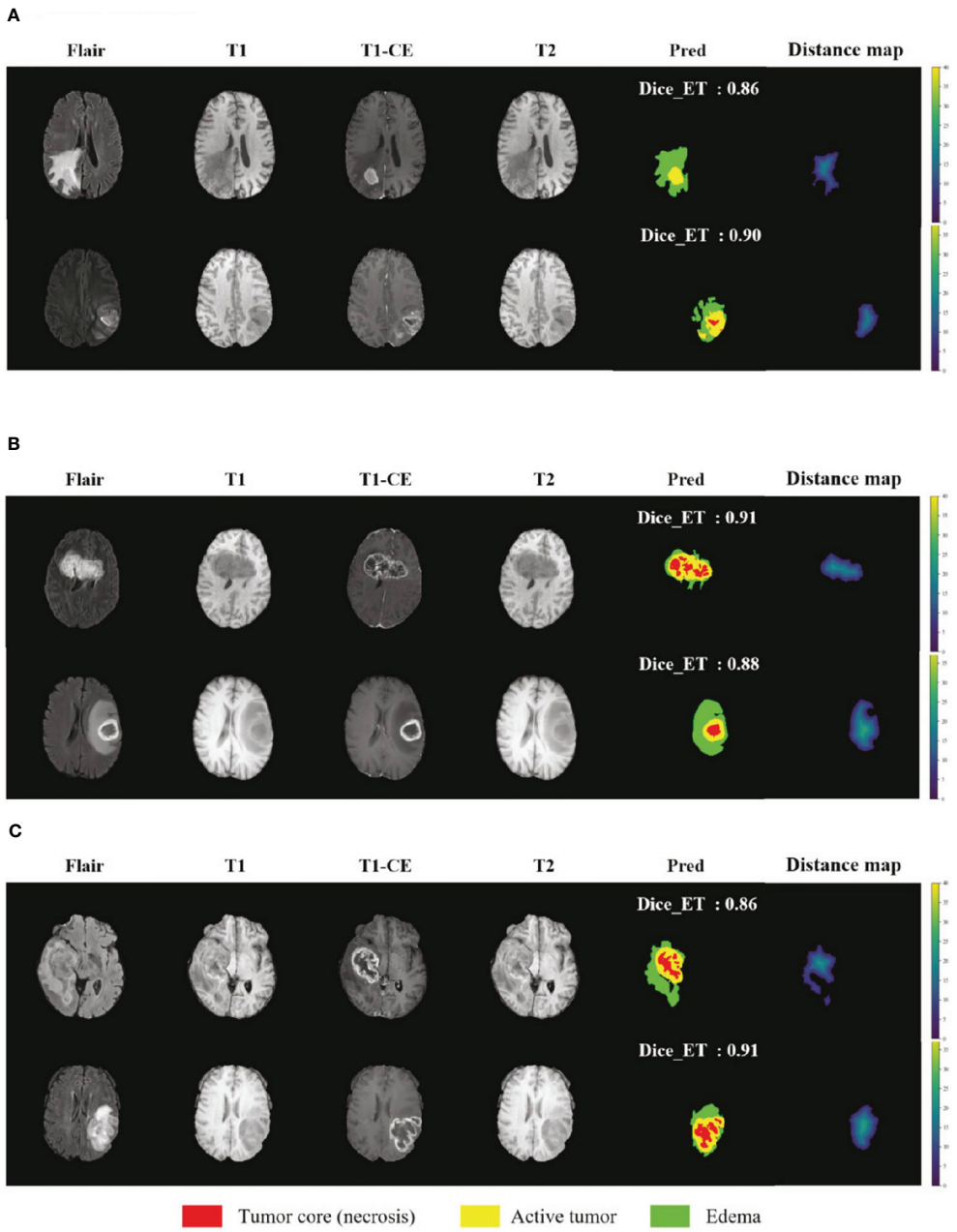


FIGURE 7 | Display of the segmentation results in validation set samples. **(A)** BraTS 2018 Data, **(B)** BraTS 2019 Data and **(C)** BraTS 2020 Data. Red color represents the tumor core (necrosis), yellow color represents the active tumor and green regions are the edema.

TABLE 3 | Comparison with the BraTS 2018 validation set of other methods.

BraTS 2018	Dice			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Proposed	0.72	0.80	0.76	9.9	13.2	15.2
Evan G. et al	0.68	0.80	0.67	14.5	14.4	20.0
Hu Y. et al	0.66	0.87	0.72	7.56	6.73	15.74
Tuan T. et al	0.68	0.82	0.70	7.0	9.4	12.5
Weninger L. et al	0.71	0.86	0.82	5.6	7.0	7.9
Hu X. et al	0.72	0.86	0.77	5.5	10.8	10.0
Serrano-Rubio J. P. et al	0.60	0.84	0.73	11.7	9.0	14.7

Comparison Study Results

In order to further verify the accuracy of the model, we compare the results of BraTS 2018 with other studies, as shown in **Table 3**, here only the values of Dice and Hausdorff95 are shown. Hu Y. et al. proposed a multi-level up-sampling network (MU-Net) to automatically segment multi-modal brain tumors. This model used the global attention GA module to combine the low-level feature maps obtained by the encoder with the high-level feature maps obtained by the decoder stand-up (33). Evan G. et al. used a 3-dimensional CNN constructed by the DeepMedic architecture created by Kamnitsas et al. (34), which contained a low-resolution and normal-resolution path, each with 11 layers (8). Tuan T. et al. proposed to segment all glioma regions using a U-Net model with multiple kernels (35). Weninger L. et al. first determined the location of the tumor and then used 3D U-Net to segment it (36). Hu X et al. proposed a 3D-Usid-Unet architecture, which included a context aggregation path and a localization path (37). Serrano-Rubio J.P. et al. trained an extreme random tree (ERT) algorithm to classify abnormal tissues with multiple labels (38).

We also compare the results of the BraTS 2019 dataset as shown in **Table 4**. Kim S. et al. used a two-step convolutional neural network (CNN) to segment brain tumors in brain MR images. First used three 2D U-Net to obtain the global information on the axial, coronal, and sagittal axes, and then used a 3D U-Net to obtain the local information in the 3D patch (39). The structure proposed by Amian M. et al. contained two parallel streamlines with two different resolutions. One was that the deep convolutional neural network learned the local features of the input data, and the other was to set the entire image for global observation. Then the output of each stream was combined to provide integrated learning of the input image (40). Shi W. et al. proposed a dense channel two-dimensional U-Net segmentation model based on residual units and feature pyramid units (41). Agravat R.R. et al. separately trained three tumor subcomponents and finally combined the three segmentation results to obtain a complete segmentation (42). Hamghalam M. et al. designed a novel pixel-by-pixel segmentation framework through a convolutional 3D to 2D MR patch conversion model (43). Wang F. et al. trained a deep learning model based on 3D U-net in the BraTS 2019 dataset with the help of brain intelligence and patching strategies (44).

We compare the results with the teams participating in the 2020 BraTS Challenge to further demonstrate the effectiveness and generalization capability of our method. These data are available on the official website of the challenge as shown in **Table 5**. Based on the results of these comparisons, we analyze that most of the methods can only achieve a relatively high degree of accuracy in a certain sub-region. At present, there is no method to achieve the highest accuracy in all sub-regions. How to balance the accuracy of all segmented regions is also one of the directions of our future work.

Discussions

We have shown that the proposed multi-task deep framework can be effectively used in multi-modal brain tumor segmentation tasks. At the same time, our integrated multi-depth fusion module can strengthen the feature extraction ability. The results on the 2020 BraTS dataset have shown the excellent performance of our model compared to other network structures. The Dice values of our ET, WT, and TC regions on the validation set are 0.75, 0.86, and 0.77, respectively, which are all above 0.7. The performance of our model is balanced among ET, WT, and TC and ranked #1, #2 and #3 compared to other state-of-the-art methods.

TABLE 4 | Comparison with the BraTS 2019 validation set of other methods.

BraTS 2019	Dice			Hausdorff95		
	ET	WT	TC	ET	WT	TC
Proposed	0.73	0.83	0.79	6.1	8.5	9.2
Kim S. et al	0.67	0.87	0.76	8.8	14.2	11.7
Amian M. et al	071	0.86	0.77	6.9	8.5	11.6
Shi W. et al	0.69	0.87	0.77	5.9	21.2	12.2
Agravat R.R. et al	0.60	0.70	0.63	11.69	14.33	17.10
Hamghalam M. et al	0.72	0.90	0.80	5.4	7.8	8.7
Wang F. et al	0.74	0.90	0.80	6.0	7.4	5.7

TABLE 5 | Comparison of validation set of teams participating in the BraTS 2020 challenge.

BraTS 2020	Dice			Sensitivity			Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Proposed	0.75	0.86	0.77	0.80	0.96	0.79	0.99	0.99	0.99	34.6	6.7	15.1
DLU	0.70	0.87	0.78	0.72	0.89	0.78	0.99	0.99	0.99	39.8	8.9	11.2
MQUNSW	0.70	0.86	0.79	0.70	0.93	0.86	0.99	0.99	0.99	40.2	11.6	13.6
Nico@	0.67	0.87	0.71	0.62	0.84	0.65	0.99	0.99	0.99	41.7	10.1	33.5
agussa	0.59	0.83	0.69	0.60	0.87	0.71	0.99	0.99	0.99	56.6	23.2	30.0
FutureHealth	0.69	0.87	0.79	0.69	0.87	0.78	0.99	0.99	0.99	44.0	10.5	11.7
unet3d	0.70	0.84	0.72	0.71	0.87	0.79	0.99	0.99	0.99	37.4	12.3	13.1
Persistent	0.69	0.82	0.72	0.69	0.85	0.70	0.99	0.99	0.99	36.9	41.5	26.3
Iris	0.68	0.86	0.73	0.67	0.90	0.70	0.99	0.99	0.99	44.1	23.9	20.0
Uncertainty	0.68	0.87	0.78	0.66	0.90	0.77	0.99	0.99	0.99	47.6	12.1	15.7
ovgu_seg	0.60	0.79	0.68	0.66	0.78	0.67	0.99	0.99	0.99	54.1	12.1	19.1

TABLE 6 | Accuracy comparison with other comparative experiments.

BraTS 2020		Dice			Sensitivity			Specificity			Hausdorff95		
		ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Train	Proposed	**0.79**	**0.88**	**0.88**	**0.85**	0.97	**0.92**	**0.99**	**0.99**	**0.99**	28.4	**5.3**	**4.8**
	Model I	0.76	0.87	0.87	0.80	0.97	0.91	0.99	0.99	0.99	**23.1**	5.8	5.9
	Model II	0.73	0.85	0.87	0.82	**0.98**	0.91	0.99	0.99	0.99	28.8	7.0	6.6
Validation	Proposed	**0.75**	**0.86**	**0.77**	**0.80**	**0.96**	**0.79**	**0.99**	**0.99**	**0.99**	**34.6**	**6.7**	**15.1**
	Model I	0.67	0.85	0.75	0.72	0.94	0.76	0.99	0.99	0.99	42.1	7.4	18.8
	Model II	0.67	0.82	0.77	0.73	0.96	0.79	0.99	0.99	0.99	51.5	9.5	11.5

Bold values indicate the best performed method.

In order to show the effectiveness of the multi-depth fusion module in improving the segmentation accuracy, we train the multi-task model that does not include this module (Model I). **Table 6** shows its segmentation accuracy. Among them, the Dice scores in the ET, WT, and TC regions are 0.674, 0.848, and 0.747 respectively. Compared with the method proposed in this paper, it can be seen that the multi-depth fusion module can effectively improve the segmentation effect. Besides, when setting the weights of the loss function of the joint two tasks, we consider that the weight of the auxiliary task of distance estimation should not be greater than the weight of the segmentation task, so we set both λ_1 and λ_2 to 1. In our study, we have also tried different weight coefficients and set them to 1 and 0.1 for training (i.e., in Model II), and we have found that the accuracy of the segmentation has decreased. Therefore, we believe that the weights used in the article are locally optimal. The above comparison experiments are all completed on the BraTS 2020 dataset.

In order to show the discrimination of the three methods more intuitively, we randomly select 4 examples in the training set for display as shown in **Figure 8**. Comparing the prediction results of the three models with the manually segmented labels, we can find that the model proposed in this paper has more advantages in detail prediction and has a more refined contour, which also proves the superiority of the model.

By comparing with some of the leading methods in this field, combined with some analysis of the results in this article, we have found that our method still has some limitations. First, in order to save training costs, we only set the weight coefficients of the loss function based on experience, and only did a set of experiments to adjust the weights instead of a large number of experiments to verify whether the weights are optimal to find the global optimal weights. In future research, we should find the optimal weight for many experiments to improve the accuracy of segmentation. Second, our segmentation results are not the best, and there are still some gaps compared with the top methods, so we propose the following improvement methods that can be developed in future work. Since the most clinically concerning information only occupies a small part of the image, inspired by Chen et al., we can design a method of region clipping, that is, to locate the part of interest (ROI) first and then segment it. This method can make the segmentation more precise. But this will also increase the amount of calculation. How to find a balance between accuracy and amount of calculation is also one of the focuses of our future work. Besides, the module we transplanted in the encoder proved to be suitable for whole-heart CT segmentation, but this is the first time that it is used for brain tumor segmentation, so we are not sure whether it is the most suitable attention module for brain tumor segmentation. We can try to use some other attention modules, such as the SE module, Non-local module, etc. After many comparison experiments, we can verify whether the effect of a multi-depth fusion module is the best.

CONCLUSION

In this article, we propose a deep multi-task learning framework that integrates multi-depth fusion modules, and perform a performance test on multiple BraTS datasets, and obtain satisfactory results. We improved the traditional V-Net framework and proposed a structure of two parallel decoder branches. The original decoder can only perform segmentation,

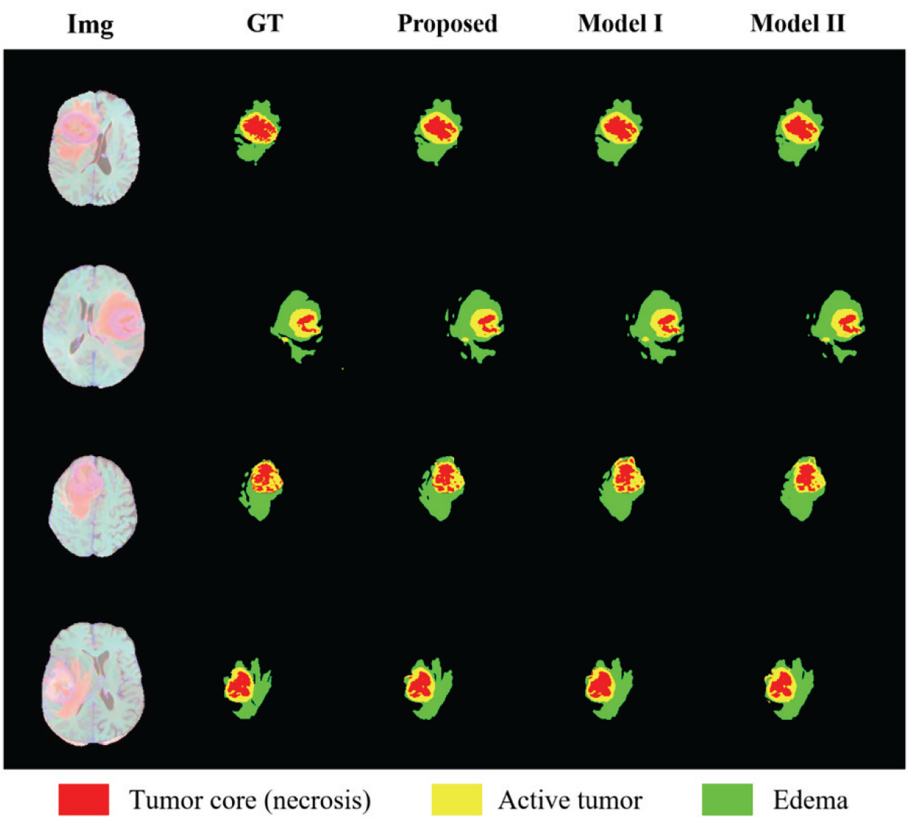


FIGURE 8 | Results of randomly selected examples compared with other model variations. Red color represents the tumor core (necrosis), yellow color represents the active tumor and green regions are the edema.

and the newly added decoder performs the auxiliary task of distance estimation, which can make the segmentation boundary more accurate. A total loss function is introduced to combine the two tasks. At the same time, we added a multi-depth fusion module after each encoder block to enhance the extraction of image features. We added a gamma factor to the loss function of the mask decoder to reduce the focus on the background area and set different weights for each type of label to alleviate the problem of category imbalance. We evaluated the accuracy of the model online for the BraTS 2018, BraTS 2019, and BraTS 2020 datasets. As a result, we obtained high-quality segmentation results, with an average Dice of 78%.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.med.upenn.edu/sbia/brats2018/data.html.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Zhejiang Chinese Medical University. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

HH performed the experiments, analyzed the data, and wrote the original draft of the manuscript. HH, GY, and XL performed the experiments, analyzed the data, and revised the manuscript. WZ and XX contributed to the experiments. WY and WJ contributed to the data analysis. GY and XL designed the project, supervised the experiments, drafted and revised the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work is funded in part by the National Natural Science Foundation of China (Grant No. 62072413), and also supported in part by the AI for Health Imaging Award 'CHAIMELEON: Accelerating the Lab to Market Transition of AI Tools for Cancer Management' [H2020-SC1-FA-DTS-2019-1 952172].

REFERENCES

1. Hussain S, Anwar SM, Majid M. Segmentation of Glioma Tumors in Brain Using Deep Convolutional Neural Network. *Neurocomputing* (2017) 282:248–61. doi: 10.1016/j.neucom.2017.12.032
2. Reardon DA, Wen PY. Therapeutic Advances in the Treatment of Glioblastoma: Rationale and Potential Role of Targeted Agents. *Oncologist* (2006) 11(2):152–64. doi: 10.1634/theoncologist.11-2-152
3. Brandes AA. State-of-the-Art Treatment of High-Grade Brain Tumors. *Semin Oncol* (2003) 30(6):pp. doi: 10.1053/j.seminoncol.2003.11.028
4. Nie J, Xue Z, Liu T, Young GS, Setayesh K, Guo L, et al. Automated Brain Tumor Segmentation Using Spatial Accuracy-Weighted Hidden Markov Random Field. *Computerized Med Imaging Graphics* (2009) 33(6):431–41. doi: 10.1016/j.compmedimag.2009.04.006
5. Myronenko A. 3d MRI Brain Tumor Segmentation Using Autoencoder Regularization. *Lect Notes Comput Sci* (2019) 11384:311–210. doi: 10.1007/978-3-030-11726-9_28
6. Soltaninejad M, Yang G, Lambrou T, Allinson N, Ye X. Supervised Learning Based Multimodal MRI Brain Tumour Segmentation Using Texture Features From Supervoxels. *Comput Methods Programs Biomed* (2018) 157:69–84. doi: 10.1016/j.cmpb.2018.01.003
7. Lachinov D, Vasiliev E, Turlapov V. Glioma Segmentation With Cascaded Unet. *Lect Notes Comput Sci* (2018) 11384:189–98. doi: 10.1007/978-3-030-11726-9_17
8. Evan G J, Gregory P, Dawid S, David F. Glioma Segmentation and a Simple Accurate Model for Overall Survival Prediction. *Int Workshop Brainlesion; International Conf Med Imaging Comput Comput Assisted Intervention* (2018) 476–84. doi: 10.1007/978-3-030-11726-9_42
9. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The Multimodal Brain Tumor Image Segmentation Benchmark (Brats). *IEEE Trans Med Imaging* (2015) 34(10):1993–2024. doi: 10.1109/TMI.2014.2377694
10. Taheri S, Ong SH, Chong VFH. Level-Set Segmentation of Brain Tumors Using a Threshold-Based Speed Function. *Image Vision Comput* (2010) 28 (1):26–37. doi: 10.1016/j.imavis.2009.04.005
11. Law MWK, Chung ACS. Weighted Local Variance-Based Edge Detection and Its Application to Vascular Segmentation in Magnetic Resonance Angiography. *IEEE Trans Med Imaging* (2007) 26(9):1224–41. doi: 10.1109/TMI.2007.903231
12. Anbeek P, Vincken KL, Viergever MA. Automated MS-Lesion Segmentation by K-Nearest Neighbor Classification. *Grand Challenge Work: Mult Scler Lesion Segm Challenge* (2008), 1–12.
13. Xia Y, Wang J, Eberl S, Fulham M, Feng DD. Brain Tissue Segmentation in PET-CT Images Using Probabilistic Atlas and Variational Bayes Inference. In: *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE (2011). p. 7969–72.
14. Franklin SW, Rajan SE. Retinal Vessel Segmentation Employing ANN Technique by Gabor and Moment Invariants-Based Features. *Appl Soft Comput* (2014) 22:94–100. doi: 10.1016/j.asoc.2014.04.024
15. Moftah HM, Azar AT, Al-Shammari ET, Ghali NI, Hassanien AE, Shoman M. Adaptive K-Means Clustering Algorithm for MR Breast Image Segmentation. *Neural Comput Appl* (2014) 24(7). doi: 10.1007/s00521-013-1437-4
16. Chen W, Giger ML, Bick U. A Fuzzy C-Means (FCM) Based Approach for Computerized Segmentation of Breast Lesions in Dynamic Contrast-Enhanced MR Images. *Acad Radiol* (2008) 13(1):63–72. doi: 10.1016/j.acra.2005.08.035
17. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Comput Soc* (2017) 39(4):640–51. doi: 10.1109/TPAMI.2016.2572683
18. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans Pattern Anal Mach Intell* (2017) 39:2481–95. doi: 10.1109/TPAMI.2016.2644615
19. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Lect Notes Comput Sci* (2015) 9351:234–41. doi: 10.1007/978-3-319-24574-4_28
20. Milletari F, Navab N, Ahmadi S. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. *016 Fourth Int Conf 3D Vision (3DV)* (2016) 1:565–71. doi: 10.1109/3DV.2016.79
21. Liu Y, Yang G, Mirak SA, Hosseiny M, Azadikhah A, Zhong X, et al. Automatic Prostate Zonal Segmentation Using Fully Convolutional Network With Feature Pyramid Attention. *IEEE Access* (2019) 7:163626–32. doi: 10.1109/ACCESS.2019.2952534
22. Ding W, Abdel-Basset M, Hawash H, Pedrycz W. Multimodal Infant Brain Segmentation by Fuzzy-informed Deep Learning. *IEEE Trans Fuzzy Syst* (2021). doi: 10.1109/TFUZZ.2021.3052461
23. Feng X, Tustison N, Meyer C. Brain Tumor Segmentation Using an Ensemble of 3D U-Nets and Overall Survival Prediction Using Radiomic Features. *Lect Notes Comput Sci* (2019) 11384:279–88. doi: 10.1007/978-3-030-11726-9_25
24. Lele C, Yue W, Adora MD, Anas ZA, Axel W, Chenliang X. Mri Tumor Segmentation With Densely Connected 3d CNN. In: *SPIE Medical Imaging Conference*. SPIE (2018) p. 105741F.1–105741F.8.
25. Zhou CH, Ding CX, Wang XC, Lu ZT, Tao DC. One-Pass Multi-task Networks With Cross-task Guided Attention for Brain Tumor Segmentation. *IEEE Trans Imaging Process* (2019) 29:4516–29. doi: 10.1109/TIP.2020.2973510
26. Zhang W, Yang G, Huang H, Yang W, Xu X, Liu Y, et al. Me-Net: Multi-encoder Net Framework for Brain Tumor Segmentation. *Int J Imaging Syst Technol* (2021), 1–15. doi: 10.1002/ima.22571
27. Zhao X, Wu Y, Song G, Li Z, Zhang Y, Fan Y. A Deep Learning Model Integrating FCNNs and CRFs for Brain Tumor Segmentation. *Med Image Anal* (2017) 43:98–111. doi: 10.1016/j.media.2017.10.002
28. Ye C, Wang W, Zhang S, Wang K. Multi-Depth Fusion Network for Whole-Heart CT Image Segmentation. *IEEE Access* (2019) 29:4516–29. doi: 10.1109/ACCESS.2019.2899635
29. Balamurali M, Kaushik S, Sharath MS, Keerthi R, Jayaraj J, Mohanasankar S. Psi-Net: Shape and Boundary Aware Joint Multi-Task Deep Network for Medical Image Segmentation. *Annu Int Conf IEEE Eng Med Biol Soc* (2019), 7223–6. doi: 10.1109/EMBC.2019.8857339
30. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein K. No New-Net. *Lect Notes Comput Sci* (2018) 11384:234–44. doi: 10.1007/978-3-030-11726-9_21
31. Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, et al. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. (2018) *arXiv Preprint* arXiv:1811.02629.
32. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, et al. Advancing The Cancer Genome Atlas Glioma MRI Collections With Expert Segmentation Labels and Radiomic Features. *Nat Sci Data* (2017) 4:1–13. doi: 10.1038/sdata.2017.117
33. Hu Y, Liu X, Wen X, Niu C, Xia Y. Brain Tumor Segmentation on Multimodal Mr Imaging Using Multi-level Upsampling in Decoder. *Lect Notes Comput Sci* (2019) 11384:168–77. doi: 10.1007/978-3-030-11726-9_15
34. Konstantinos K, Christian L, Virginia FJN, Joanna PS, Andrew DK, David KM, et al. Efficient Multi-Scale 3D CNN With Fully Connected CRF for Accurate Brain Lesion Segmentation. *Med Image Anal* (2017) 36:61–78. doi: 10.1016/j.media.2016.10.004
35. Tuan T, Bao P. Brain Tumor Segmentation Using Bit-Plane and UNET. *Lect Notes Comput Sci* (2019) 11384:466–75. doi: 10.1007/978-3-030-11726-9_41
36. Weninger L, Rippel O, Koppers S, Merhof D. Segmentation of Brain Tumors and Patient Survival Prediction: Methods for the BraTS 2018 Challenge. *Lect Notes Comput Sci* (2019) 11384:3–12. doi: 10.1007/978-3-030-11726-9_1
37. Hu X, Li H, Zhao Y, Dong C, Menze BH, Piraud M. Hierarchical Multi-Class Segmentation of Glioma Images Using Networks With Multi-Level Activation Function. *Lect Notes Comput Sci* (2019) 11384:116–27. doi: 10.1007/978-3-030-11726-9_11
38. Serrano-Rubio JP, Everson R. Brain Tumour Segmentation Method Based on Supervoxels and Sparse Dictionaries. *Lect Notes Comput Sci* (2019) 11384:210–21. doi: 10.1007/978-3-030-11726-9_19
39. Kim S, Luna M, Chikontwe P, Park S. Two-Step U-Nets for Brain Tumor Segmentation and Random Forest With Radiomics for Survival Time Prediction. In: *International MICCAI Brainlesion Workshop*. Springer, (2020) p. 200–9. doi: 10.1007/978-3-030-46640-4_19
40. Amian M, Soltaninejad M. Multi-Resolution 3d CNN for MRI Brain Tumor Segmentation and Survival Prediction. In: *International MICCAI Brainlesion Workshop*. Springer (2019) (p. 221–30). doi: 10.1007/978-3-030-46640-4_21

41. Shi W, Pang E, Wu Q, Lin F. Brain Tumor Segmentation Using Dense Channels 2d U-Net and Multiple Feature Extraction Network. *Lect Notes Comput Sci* (2020) 11992:273–83. doi: 10.1007/978-3-030-46640-4_26
42. Agravat RR, Raval MS. Brain Tumor Segmentation and Survival Prediction. *Lect Notes Comput Sci* (2020) 11992:338–48. doi: 10.1007/978-3-030-46640-4_32
43. Hamghalam M, Lei B, Wang T. Convolutional 3D to 2D Patch Conversion for Pixel-Wise Glioma Segmentation in MRI Scans. *Lect Notes Comput Sci* (2020) 11992:3–12. doi: 10.1007/978-3-030-46640-4_1
44. Wang F, Jiang R, Zheng L, Meng C, Biswal B. 3d U-Net Based Brain Tumor Segmentation and Survival Days Prediction. *Lect Notes Comput Sci* (2020) 11992:131–41. doi: 10.1007/978-3-030-46640-4_13

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Huang, Yang, Zhang, Xu, Yang, Jiang and Lai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 10 June 2021
doi: 10.3389/fonc.2021.687771

Predicting Microsatellite Instability Status in Colorectal Cancer Based on Triphasic Enhanced Computed Tomography Radiomics Signatures: A Multicenter Study

Yuntai Cao[1,2,3,4†], Guojin Zhang[3,5†], Jing Zhang[3,6†], Yingjie Yang[7†], Jialiang Ren[8], Xiaohong Yan[9], Zhan Wang[10], Zhiyong Zhao[2,3,4], Xiaoyu Huang[2,3,4], Haihua Bao[1] and Junlin Zhou[3,4*]*

[1] Department of Radiology, Affiliated Hospital of Qinghai University, Xining, China, [2] Second Clinical School, Lanzhou University, Lanzhou, China, [3] Department of Radiology, Lanzhou University Second Hospital, Lanzhou, China, [4] Key Laboratory of Medical Imaging, Lanzhou, China, [5] Sichuan Academy of Medical Sciences Sichuan Provincial People's Hospital, Chengdu, China, [6] The Fifth Affiliated Hospital of Zunyi Medical University, Zhuhai, China, [7] Department of Radiology, Second People's Hospital of Lanzhou City, Lanzhou, China, [8] Department of Pharmaceuticals Diagnosis, GE Healthcare, Beijing, China, [9] Department of Critical Medicine, Affiliated Hospital of Qinghai University, Xining, China, [10] Department of Biomedical Engineering, Tsinghua University, Beijing, China

Background: This study aimed to develop and validate a computed tomography (CT)-based radiomics model to predict microsatellite instability (MSI) status in colorectal cancer patients and to identify the radiomics signature with the most robust and high performance from one of the three phases of triphasic enhanced CT.

Methods: In total, 502 colorectal cancer patients with preoperative contrast-enhanced CT images and available MSI status (441 in the training cohort and 61 in the external validation cohort) were enrolled from two centers in our retrospective study. Radiomics features of the entire primary tumor were extracted from arterial-, delayed-, and venous-phase CT images. The least absolute shrinkage and selection operator method was used to retain the features closely associated with MSI status. Radiomics, clinical, and combined Clinical Radiomics models were built to predict MSI status. Model performance was evaluated by receiver operating characteristic curve analysis.

Results: Thirty-two radiomics features showed significant correlation with MSI status. Delayed-phase models showed superior predictive performance compared to arterial- or venous-phase models. Additionally, age, location, and carcinoembryonic antigen were considered useful predictors of MSI status. The Clinical Radiomics nomogram that incorporated both clinical risk factors and radiomics parameters showed excellent performance, with an AUC, accuracy, and sensitivity of 0.898, 0.837, and 0.821 in the training cohort and 0.964, 0.918, and 1.000 in the validation cohort, respectively.

OPEN ACCESS

Edited by:
Giorgio Russo, National Research Council, Italy

Reviewed by:
Xiao-Jian Wu, The Sixth Affiliated Hospital of Sun Yat-sen University, China
Swapnil Ulhas Rane, Advanced Centre for Treatment, Research and Education in Cancer, India

***Correspondence:**
Haihua Bao
baohelen2@sina.com
Junlin Zhou
ery_zhoujl@lzu.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *30 March 2021*
Accepted: *17 May 2021*
Published: *10 June 2021*

Citation:
Cao Y, Zhang G, Zhang J, Yang Y, Ren J, Yan X, Wang Z, Zhao Z, Huang X, Bao H and Zhou J (2021) Predicting Microsatellite Instability Status in Colorectal Cancer Based on Triphasic Enhanced Computed Tomography Radiomics Signatures: A Multicenter Study. Front. Oncol. 11:687771. doi: 10.3389/fonc.2021.687771

Conclusions: The proposed CT-based radiomics signature has excellent performance in predicting MSI status and could potentially guide individualized therapy.

Keywords: microsatellite instability, colorectal cancer, radiomics, CT, triphasic enhanced phase

INTRODUCTION

Colorectal cancer (CRC) is characterized by complex biological features and shows distinct heterogeneity. Even though the clinicopathological characteristics of CRC are similar, there is still significant variability in treatment response and prognosis (1). Two major molecular events are involved in the occurrence and development of CRC (2, 3). The vast majority of CRCs are caused by chromosomal instability events (approximately 85%), including mutations in *APC*, *KRAS*, and *TP53* genes, etc. However, a small percentage of CRCs are caused by microsatellite instability (MSI) (approximately 15%).

Mismatch repair (MMR) genes are highly conserved and are involved in repairing DNA base mismatches. They are beneficial in maintaining genome stability and reducing spontaneous mutations (4). MMR proteins include MLH1, MSH2, MSH6, and PMS2. During DNA replication, minor DNA mismatches occasionally occur, which are recognized by these proteins and then cut and synthesized into new strands for repair (5, 6). When any one of these four proteins are non-functional, they cause accumulation of DNA base mismatches in proliferating cells, a phenomenon known as MSI (6).

MSI status is currently a key predictor for evaluating the treatment strategies and prognosis of CRC patients (7, 8). Compared with microsatellite-stable (MSS) CRC patients, CRC patients with MSI status are more likely to benefit from immunotherapy, but they do not benefit from pyrimidine analogs or fluorouracil-based adjuvant chemotherapy (9–11). In addition, CRC patients with MSI status may have a favorable prognosis (12–14). The National Comprehensive Cancer Network (NCCN), European Society for Medical Oncology (ESMO), and Japanese Society for Cancer of the Colon and Rectum (JSCCR) guidelines recommend testing the MSI status of CRC patients (4, 15, 16).

At present, MSI status detection is mainly done through immunohistochemistry (IHC) and polymerase chain reaction (PCR) methods on biopsy or surgical tissue, both of which are invasive and costly (8, 17). Furthermore, the small part of the tissue captured by biopsy may not be sufficient to accurately reflect the MSI status of tumors (18, 19). In addition, these advanced biological tests can only be performed in qualified tertiary medical centers, as local medical institutions have not widely adopted these methods because of the lack of suitable medical equipment (20). Therefore, developing a non-invasive, cost-effective, and easily repeatable method to identify MSI status could help clinicians to develop more accurate treatment strategies for CRC patients.

Radiomics is a burgeoning field in the era of precision medicine, involving screening, diagnosis, treatment, and prognostic assessment of multiple systemic diseases (21–24). By extracting high-dimensional, mineable data from medical imaging and evaluating its association with clinicopathologic factors or gene expression, radiomics facilitates the formulation of individualized treatment strategies. Radiomics has been widely used in CRC stage assessment (21), tumor differentiation identification (25), post-neoadjuvant chemotherapy efficacy evaluation (26), and *KRAS* mutation status identification (27). A previous study demonstrated a significant correlation between a CT-based radiomics signature and MSI status in CRC patients (28, 29). These results indicate that pretreatment CT may be associated with MSI status and that radiomics analysis may greatly contribute to MSI status identification. However, previous studies have only included a single group and lack external validation. Moreover, the superiority of the venous phase (VP) compared to arterial and delayed phases (AP and DP, respectively) in the prediction of MSI status in CRC patients remains to be confirmed. Therefore, the aim of this study was to investigate whether a CT-based radiomics signature could identify MSI status in CRC patients and to identify the phase with the most robust and high-performing radiomics signature from triphasic enhanced CT.

Abbreviations: CT, computed tomography; CRC, colorectal cancer; MSI/MSS, microsatellite instability/stability; AP, arterial phase; APR, radiomics model of arterial phase; VP, venous phase; VPR, radiomics model of venous phase; DP, delayed phase; DPR, radiomics model of delayed phase; FR, radiomics model of fusion of arterial phase, delayed phase and venous phase features; ROI, region of interest; LoG, Laplacian of Gaussian; CEA, carcinoembryonic antigen; ROC, receiver operating curve; AUC, area under the curve; DCA, decision curve analysis; PR, precision-recall; CI, confidence interval; PPV/NPV, positive/negative predictive value; HL, Hosmer-Lemeshow.

MATERIALS AND METHODS

Patients

Ethical approval was obtained by the medical ethics committee in both participating centers (center I: Lanzhou University Second Hospital; center II: The Second People's Hospital of Lanzhou city), and the requirement for informed consent was waived due to the retrospective nature of the study. Patient inclusion and exclusion details and the patient recruitment pathway are shown in **Figure 1**. The institutional database in center I was searched for eligible patients who underwent curative resection between March 2014 and August 2020, and a total of 441 consecutive patients were enrolled. This study included 255 males (42.2%) and 186 females (57.8%), with an average age of 58.64 ± 12.92 years (range, 20–89 years). Furthermore, 61 patients from center II were also enrolled between July 2018 and August 2020, including 38 males (62.3%) and 23 females (37.7%), with an average age of 56.93 ± 11.94 years (range, 27–84 years). The model for MSI

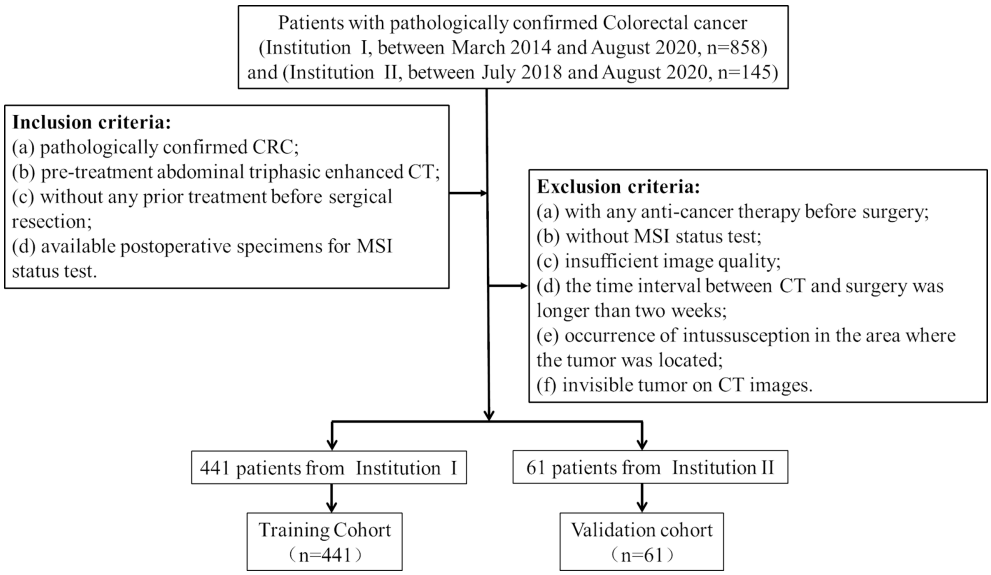


FIGURE 1 | Patient inclusion and exclusion details and the patient recruitment pathway.

prediction was established in the training cohort and evaluated in the external validation cohort. The baseline clinical data of all CRC patients, including age, sex, tumor location, carcinoembryonic antigen (CEA) level, CA125 level, and CA199 level, were collected. Two radiologists (radiologist A, Y.T.C.; radiologist B, J.Z.) with more than 10 years of experience in abdominal imaging collected radiological features on preoperative CT images and recorded the results, including clinical tumor/lymph node (cT/N) stage and tumor maximum diameter (maximum diameter perpendicular to the long axis of the cross-sectional image). In order to minimize the deviation of the measurement results, the quantitative data was taken as the final result by the average of the measurement values of the two radiologists, while the qualitative data is diagnosed by the two radiologists independently, and the disagreement is resolved through negotiation.

Identification of MSI Status

MSI status was evaluated by immunohistochemical staining of MMR proteins (MLH1, MSH2, MSH6, PMS2). The standard streptavidin biotin-peroxidase procedure was performed on postoperative tissues to identify the MSI status. Patients were classified into the MSI or MSS group according to the staining results of MMR proteins. Among the four MMR proteins, negative staining for one or more proteins was defined as MSI. MSS was defined as positive staining for all four MMR proteins (6).

CT Image Acquisition and Segmentation

All patients underwent abdominal and/or pelvic enhanced CT scans in two institutions, which covered the whole tumor. Triphasic enhanced CT images were retrieved from the picture archiving and communication system (PACS, Carestream; Rochester, NY) and stored in corresponding folders in DICOM format for further analysis. The CT scanner and acquisition parameters of the three institutions are listed in **Supplementary Tables S1, S2**.

Two gastrointestinal radiologists (radiologist A and radiologist B) performed three-dimensional (3D) radiomics segmentation on AP, VP, and DP using ITK-SNAP software (version 3.6.0; www.itksnap.org). Radiologist A segmented 300 cases and radiologist B segmented the other 202 cases.

For radiomics segmentation, an ROI was manually delineated on each slice of the tumor. Air and feces in the intestinal tract and pericolonic fat were carefully excluded from the contours. Finally, three ROIs (AP, DP, and VP) were generated for each patient. To evaluate inter-observer reproducibility and robustness of feature extraction, radiologist A and radiologist B randomly selected 30 patients and performed manual segmentation again. We estimated the reproducibility of feature extraction using inter-class correlation coefficients (ICCs), where ICCs greater than 0.80 indicated good reproducibility (30). Additionally, 30 patients were randomly selected from each CT scanner to build the CT scanner set for calculating intra- and interclass correlation coefficients (ICCs).

Feature Extraction

Before feature extraction, we adopted three steps to preprocess the CT images. First, we resampled images to 1 mm × 1 mm × 1 mm using linear interpolation to try to reduce the influence of different layer thicknesses. Second, we transformed the continuous images into discrete integer values using gray-level discretization processing (bin width = 25). Finally, Laplacian of Gaussian (LoG) and wavelet image filters were used to eliminate mixed noise in the processing of image digitization and to obtain low- or high-frequency features.

Radiomics features were extracted using the PyRadiomics package (31). Seven classes of radiomics features were extracted from the original and filtered images (wavelet and LoG). Finally, 1037 3D radiomics features were extracted from each patient. The feature types and their numbers are as follows: (1) first-order (histogram) features (n = 198); (2) shape features (n = 14); (3) gray-level co-occurrence matrix (GLCM) features (n = 264); (4) gray-level run-length matrix (GLRLM) features (n = 176); (5) gray-level size zone matrix (GLSZM) features (n = 176); (6) neighboring gray-tone difference matrix (NGTDM) features (n = 55); (7) gray-level dependence matrix (GLDM) features (n = 154). The specific definitions and descriptions of the features are demonstrated in the **Supplementary Materials**.

Features Selection and Prediction Model Building

After radiomics feature extraction, all missing data in the training cohort were replaced by median value, z-score normalization was performed on each feature, and the same preprocessing procedure was applied to the validation cohort. We performed a binary classification task for MSI status prediction: MSS vs. MSI expression. It is worth noting that the sample numbers of the two groups were unbalanced between the training and validation cohorts. The initial bias adjustment method was used to correct the influence of unbalanced sample size. The adjustment bias b_0 was determined using the following equation:

$$p_0 = \frac{pos}{(pos + neg)} = \frac{1}{1 + e^{-b_0}}$$

$$b_0 = -\log_e \frac{1}{p_0 - 1}$$

The process of radiomics feature selection that is most related to MSI status consists of three steps. First, univariate analysis with the Mann-Whitney U test was performed for feature selection to retain features with $p < 0.05$ for the subsequent process. Second, the least absolute shrinkage and selection operator (LASSO) method was used to retain features closely associated with MSI status. Finally, multivariable stepwise logistic regression further eliminated irrelevant features and retained the most informative features. A ten times five-fold cross-validation method was applied to avoid overfitting and to identify the model with the best performance.

Three radiomics models were established based on the above radiomics signatures in triphasic phase-enhanced CT images (APR, VPR, and DPR models). In order to verify whether the model combining the triphasic enhanced phases can improve the prediction performance of MSI status, the FR model was built based on AP, VP, and DP fusion features from 3D segmentation patterns. The maximum area under the curve (AUC) in the training cohort determined the final regularization parameter. Furthermore, the Radiomics models predicted a radiomics signature demonstrating the likelihood of MSI status for each patient.

Clinical, Combined Model, and Nomogram Construction

For clinical and radiological features, the chi-squared test or Fisher's exact test was used to compare differences in sex, CEA, CA125, CA199, cT stage and cN stage, while the Student's t-test or Mann-Whitney U test was used to compare differences in age, and maximum diameter between the MSS and MSI groups in the training and external validation cohorts. Generally, P-values < 0.05 (two-sided) were considered statistically significant. We performed multivariable analyses to identify the most important features. A clinical model was established based on the inclusion of selected features.

A combined model (clinical Radiomics) was developed based on correlated clinicalradiological features and radiomics features to verify whether the combination of radiomics signatures and clinicalradiological features could improve the prediction of MSI status, and it was presented as an individualized nomogram.

Using multivariate logistic regression coefficients, a nomogram incorporating clinicalradiological characteristics and radiomics signatures was created for the training and external validation cohorts following the selection of clinical characteristics and model comparison. This nomogram provides a more convenient and reliable tool for patients and clinicians. A flowchart of the study is shown in **Figure 2**.

Statistical Analyses

All statistical analyses were conducted using the R statistical software package (version 3.6.3; http://www.Rproject.org). Student's t-test, the Mann-Whitney U test, and the chi-squared test or Fisher's exact test were used to compare continuous and categorical variables, as appropriate. A two-sided P-value < 0.05 was considered statistically significant. ICCs were used to calculate the consistency of measurements between the two radiologists and different CT scanners. Receiver operating characteristic (ROC) analysis was used to evaluate the predictive accuracy of the different models. The AUC, 95% confidence interval (CI), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for each model. Precision-recall (PR) curves and the DeLong test were used to compare the AUC estimates of the discrimination performance between different predictive models. A calibration curve was constructed to assess the goodness-of-fit of the models. The Hosmer-Lemeshow (HL) test was performed to assess the agreement between the predicted MSI status and the observed outcomes. To verify the clinical usefulness of the models, we quantified the net benefit at different threshold probabilities in the dataset using decision curve analysis (DCA).

RESULTS

Clinical Characteristics

Patient and tumor characteristics in the training cohort are listed in **Table 1**. This study included 502 CRC patients (441 patients in center I, 61 patients in center II) in the final analysis. The prevalence of MSI was 15.19% (67/441) in center I and 14.75% (9/61) in center II.

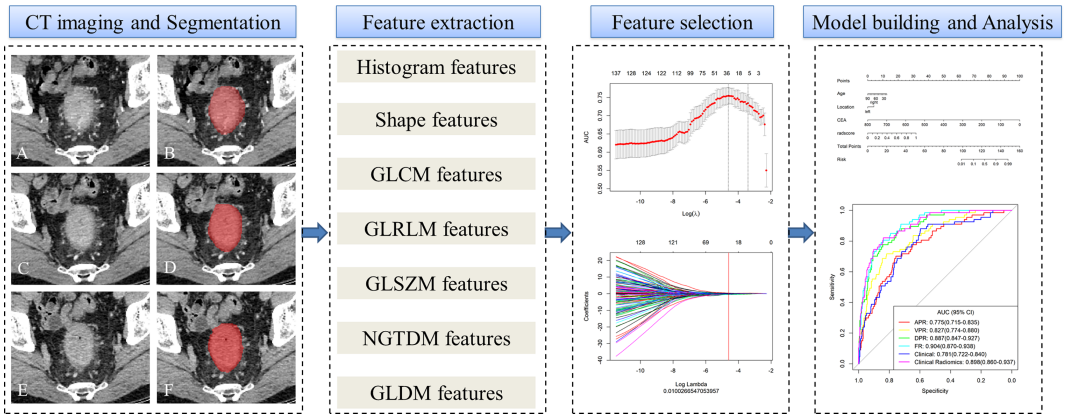


FIGURE 2 | Workflow of microsatellite instability (MSI) prediction building and analysis. The tumors were segmented on arterial phase **(A, B)**, delayed phase **(C, D)** and venous phase **(E, F)** CT images to form volumes of interest (VOIs). One thousand and thirty-seven quantitative radiomics features were extracted from each patient. The least absolute shrinkage and selection operator (LASSO) was used to select the features. Multivariate logistic regression was used to build radiomics, clinical, and clinicoradiomics combined models for MSI prediction. Finally, the radiomics signature and clinical factors were incorporated into a nomogram for individual evaluation. Receiver operating characteristic curves were used to evaluate the clinical usefulness of the nomogram.

TABLE 1 | Characteristics of patients in the training cohort [median (Q1, Q3) or no. (%)].

Characteristics		Training cohort (n=441)		
		MSS	MSI	P value
Age (years)		61.00 (51.00, 68.00)	51.00 (42.50, 63.00)	<0.001
Gender	Female	153 (40.9%)	33 (49.3%)	0.203
	Male	221 (59.1%)	34 (50.7%)	
Tumor Location	Left	267 (71.4%)	22 (32.8%)	<0.001
	Right	107 (28.6%)	45 (67.2%)	
CEA level		4.03 (2.18, 12.82)	2.81 (1.60, 6.37)	0.009
CA125 level		12.02 (8.73, 17.30)	16.71 (9.59, 24.64)	0.004
CA199 level		13.45 (7.74, 26.59)	9.99 (5.94, 25.36)	0.067
cT stage	T1	12 (3.2%)	0 (0.0%)	0.671
	T2	58 (15.5%)	10 (14.9%)	
	T3	236 (63.1%)	47 (70.1%)	
	T4	68 (18.2%)	10 (14.9%)	
cN stage	N0	210 (56.1%)	44 (65.7%)	0.201
	N1	81 (21.7%)	11 (16.4%)	
	N2	83 (22.2%)	12 (17.9%)	
Maximum diameter (mm)		19.80 (15.71, 25.62)	24.70 (18.31, 30.80)	0.001

Predictive Performance of the Clinical Model

Age, tumor location, CEA level, CA125 level, and maximum diameter were found to be significantly different ($P < 0.05$) between the MSI and MSS groups in the training cohort, but other characteristics were not significantly different ($P > 0.05$). Finally, after multivariate analyses, age, tumor location, and CEA were selected as independent predictors of MSI and were enrolled into the clinical model (**Supplementary Table S3**). The clinical model showed moderate performance in predicting MSI both in the training cohort and the validation cohort, with an AUC of 0.781 (95%CI, 0.722-0.840) in the training cohort and 0.919 (95%CI, 0.833-1.000) in the validation cohort (**Table 3**). The accuracy, sensitivity, and specificity were 0.721, 0.716, and 0.722 in the training cohort and 0.869, 0.889, and 0.865 in the validation cohort, respectively.

Radiomics Signature Building and Discrimination Performance Assessment

ICCs were calculated to evaluate the agreement of features extracted by the two radiologists and different CT scanners, and ICC values > 0.80 indicated good agreement. A total of 1037 3D radiomics features from AP, VP, and DP images were extracted for each patient, and irrelevant features were removed as described earlier. Finally, 6 AP, 10 VP, and 16 DP 3D radiomics features were retained as the final signatures. The feature names and distributions are listed in **Table 2**. The values of these features were significantly different between the MSI and MSS groups. Following stepwise regression analysis, 16 features were selected after fusion of the radiomics features from AP, VP, and DP (FR model). Significant differences were found in these features between the MSI and MSS groups (**Supplementary**

TABLE 2 | The final signatures selected from 3D radiomics features.

Arterial phase (n=10)	Venous phase (n=10)	Delayed phase (n=16)	Radiomics (n=16)
A_original_glszm_GrayLevelVariance	V_original_glszm_GrayLevelVariance	D_original_shape_Elongation	V_original_glszm_ZoneEntropy
A_log.sigma.5.0.mm.3D_glszm_LargeAreaHighGrayLevelEmphasis	V_original_glszm_ZoneEntropy	D_original_firstorder_Range	V_wavelet.LHL_glszm_LargeAreaHighGrayLevelEmphasis
A_wavelet.LHL_firstorder_90Percentile	V_log.sigma.5.0.mm.3D_gldm_DependenceNonUniformityNormalized	D_original_ngtdm_Contrast	V_wavelet.HLH_firstorder_Mean
A_wavelet.LHL_firstorder_Skewness	V_wavelet.LHL_glcm_MCC	D_log.sigma.3.0.mm.3D_glszm_LargeAreaLowGrayLevelEmphasis	V_wavelet.HHH_glszm_LargeAreaLowGrayLevelEmphasis
A_wavelet.LHL_gldm_SmallDependenceLowGrayLevelEmphasis	V_wavelet.LHL_glszm_LargeAreaHighGrayLevelEmphasis	D_log.sigma.3.0.mm.3D_gldm_DependenceNonUniformityNormalized	V_wavelet.HHH_gldm_SmallDependenceLowGrayLevelEmphasis
A_wavelet.LHH_glszm_SmallAreaEmphasis	V_wavelet.LHL_gldm_DependenceVariance	D_log.sigma.3.0.mm.3D_ngtdm_Contrast	D_original_firstorder_Range
	V_wavelet.HLL_glszm_SizeZoneNonUniformity	D_log.sigma.5.0.mm.3D_glszm_LargeAreaHighGrayLevelEmphasis	D_original_ngtdm_Contrast
	V_wavelet.HLH_firstorder_Mean	D_log.sigma.5.0.mm.3D_ngtdm_Busyness	D_log.sigma.3.0.mm.3D_glszm_LargeAreaLowGrayLevelEmphasis
	V_wavelet.HHH_glszm_LargeAreaLowGrayLevelEmphasis	D_wavelet.LLH_glcm_InverseVariance	D_log.sigma.3.0.mm.3D_gldm_DependenceNonUniformityNormalized
	V_wavelet.HHH_gldm_SmallDependenceLowGrayLevelEmphasis	D_wavelet.LHL_glszm_GrayLevelNonUniformityNormalized	D_log.sigma.5.0.mm.3D_ngtdm_Busyness
		D_wavelet.LHL_glszm_LargeAreaHighGrayLevelEmphasis	D_wavelet.LLH_glcm_InverseVariance
		D_wavelet.LHH_glcm_InverseVariance	D_wavelet.LHL_glszm_GrayLevelNonUniformityNormalized
		D_wavelet.LHH_glszm_LargeAreaEmphasis	D_wavelet.LHH_glcm_InverseVariance
		D_wavelet.LHH_gldm_DependenceNonUniformityNormalized	D_wavelet.LHH_gldm_DependenceNonUniformityNormalized
		D_wavelet.HLH_glcm_Imc1	D_wavelet.HLH_glcm_Imc1
		D_wavelet.LLL_firstorder_Skewness	D_wavelet.LLL_firstorder_Skewness

Figure 1). As shown in **Supplementary Figure 2**, the feature heatmaps show that the correlation between most of the features is below than 0.9, indicating that the final features are less collinear with each other. Four models were built based on the above radiomics signatures for preoperatively predicting MSI (APR, DPR, VPR, and FR models). The AUC, accuracy, sensitivity, specificity, PPV, and NPV for each model are listed in **Table 3** and **Figure 3**. The DPR model had optimal predictive performance compared to APR or VPR in the training and validation cohorts (**Figures 3A, B**). In addition, the FR model had a higher predictive AUC than APR, DPR, or VPR in the training cohort. In the validation cohort, the FR model had a higher predictive AUC than APR or VPR in the training cohort but slightly lower than the AUC of the DPR model.

Predictive Performance of the Combined Model

As shown in **Figure 4A**, a Clinical Radiomics combined model was developed that incorporates clinical risk factors and DP radiomics signatures, which was presented as a quantitative nomogram. The Clinical Radiomics model showed excellent predictive ability for MSI status, with an AUC, accuracy, and sensitivity of 0.898, 0.837, and 0.821 in the training cohort and 0.964, 0.918, and 1.000 in the validation cohort, respectively. As shown in **Table 3** and **Figure 3**, the Clinical Radiomics model had a better predictive AUC value than either the clinical model or the radiomics models in the training cohort and validation cohort.

The PR curves show that the Clinical Radiomics model had better MSI prediction performance than the clinical or radiomics models (**Figure 5**). DeLong test results showed a significant difference between the AUC of the Clinical Radiomics model and of the APR, VPR, and clinical models in the training cohort (**Figure 6**). The calibration curve of the nomogram showed favorable agreement between prediction and observation in predicting the risk of MSI (**Figures 4B, C**). The HL test yielded non-significant statistics in the training and validation cohorts, indicating goodness-of-fit in the models.

The DCA results for the clinical model, radiomics models, and combined nomogram are presented in **Figures 4D, E**. The nomogram achieved more clinical utility in predicting MSI than the clinical model or radiomics model alone. The DCA curve of the nomogram demonstrated that when the threshold probability of a patient or doctor ranged between 5% and 80%, the use of the nomogram added greater benefit for MSI prediction than the treat-all-patients scheme or the treat-none scheme in the training cohort.

DISCUSSION

In this study, we investigated the association between triphasic enhanced CT radiomics features and MSI status. Six, ten, and sixteen radiomics features showed significant correlation with MSI status in AP, DP, and VP, respectively. Four radiomics models (APR, DPR, VPR, and FR) were proposed using the above radiomics features in the training cohort to predict MSI status for patients with colorectal cancer, and we validated its

TABLE 3 | Predictive performance of different models in training and validation cohorts.

Feature_num	Methods	Training cohort						Validation cohort					
		AUC	Accuracy	Sensitivity	Specificity	PPV	NPV	AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
6	APR	0.775 (0.715-0.835)	0.698 (0.653-0.741)	0.716 (0.612-0.806)	0.695 (0.583-0.810)	0.296 (0.265-0.322)	0.932 (0.920-0.941)	0.786 (0.644-0.929)	0.689 (0.557-0.801)	0.667 (0.333-1.000)	0.692 (0.519-0.962)	0.273 (0.158-0.360)	0.923 (0.900-0.943)
10	VPR	0.827 (0.774-0.880)	0.744 (0.700-0.784)	0.731 (0.612-0.836)	0.746 (0.631-0.869)	0.340 (0.301-0.371)	0.939 (0.929-0.948)	0.810 (0.674-0.946)	0.754 (0.627-0.855)	0.556 (0.222-0.889)	0.788 (0.500-1.000)	0.312 (0.154-0.421)	0.911 (0.867-0.929)
16	DPR	0.887 (0.847-0.927)	0.787 (0.746-0.824)	0.791 (0.701-0.896)	0.786 (0.722-0.909)	0.398 (0.370-0.429)	0.955 (0.951-0.960)	0.953 (0.903-1.000)	0.852 (0.738-0.930)	1.000 (0.778-1.000)	0.827 (0.808-0.981)	0.500 (0.437-0.500)	1.000 (1.000-1.000)
16	FR	0.904 (0.870-0.938)	0.803 (0.762-0.839)	0.836 (0.716-0.925)	0.797 (0.684-0.885)	0.424 (0.387-0.449)	0.964 (0.959-0.968)	0.893 (0.804-0.982)	0.787 (0.663-0.881)	0.778 (0.444-1.000)	0.788 (0.635-0.962)	0.389 (0.267-0.450)	0.953 (0.943-0.962)
3	Clinical	0.781 (0.722-0.840)	0.721 (0.677-0.762)	0.716 (0.567-0.836)	0.722 (0.618-0.799)	0.316 (0.268-0.350)	0.934 (0.924-0.940)	0.919 (0.833-1.000)	0.869 (0.758-0.942)	0.889 (0.442-1.000)	0.865 (0.596-1.000)	0.533 (0.362-0.563)	0.978 (0.969-0.981)
4	Clinical Radiomics	0.898 (0.860-0.937)	0.837 (0.799-0.870)	0.821 (0.672-0.896)	0.840 (0.663-0.912)	0.478 (0.429-0.500)	0.963 (0.954-0.966)	0.964 (0.919-1.000)	0.918 (0.819-0.973)	1.000 (0.667-1.000)	0.904 (0.846-1.000)	0.643 (0.545-0.643)	1.000 (1.000-1.000)

FR, fusion of radiomics features of arterial phase, venous phase, and delayed phase; Clinical, fusion of clinical, and radiological characteristics; Clinical Radiomics, fusion of clinicalradiological features and radiomics features. APR, radiomics model of arterial phase; AUC, area under the curve; D, DPR, radiomics model of delayed phase; NPV, negative predictive value; PPV, positive predictive value; VPR, radiomics model of venous phase.

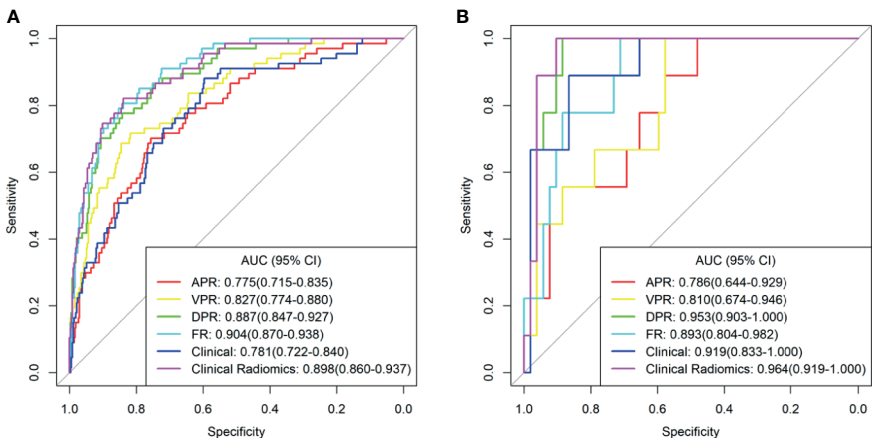


FIGURE 3 | The receiver operating characteristic (ROC) curves of the different models in training cohort **(A)** and validation cohort **(B)**. AUC, area under the curve; APR, radiomics model of arterial phase; DPR, radiomics model of delayed phase; VPR, radiomics model of venous phase; FR, radiomics model of fusion of arterial phase, delayed phase and venous phase features; Clinical Radiomics, fusion of clinical risk factors and radiomics features of delayed phase.

performance in an external validation cohort from another center. Our study showed that the DPR model had a higher outstanding performance than the APR or VPR models in both the training and external validation cohorts. Meanwhile, the nomogram, based on DP radiomics features and clinical risk factors, showed excellent identification ability for MSI status in both training (AUC: 0.898, 95% CI 0.860-0.937) and external validation (AUC: 0.964, 95% CI 0.919–1.000) cohorts. Our nomogram may be useful for predicting the MSI status of CRC patients and, thus, has the potential to aid in the determination of therapeutic strategies. In common studies, the results of external validation cohort are lower than the training cohort due to overfitting. Our results show that the results of external validation cohort are slightly higher than the training cohort, and lack of overlap between the 95% CI of the accuracies between the training and the validation cohorts. Since our validation cohort is external data set, there is often some deviation in distribution between the two data sets due to geographical location and other factors, which may cause the model performance of the validation cohort higher than the training cohort. Validation on additional cohort is required to ensure the model's reproducible and generalizable.

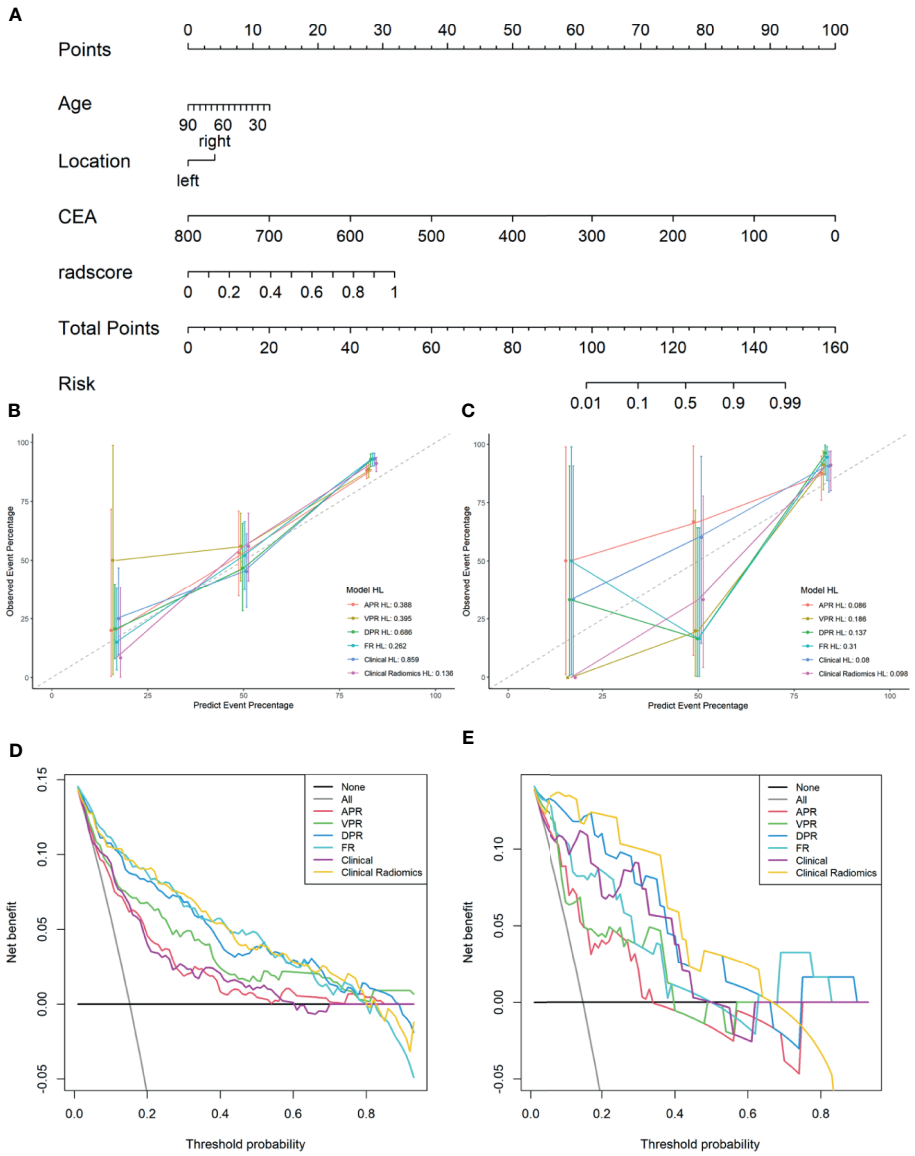


FIGURE 4 | A Clinical Radiomics nomogram for preoperative identification of microsatellite instability status in colorectal cancer patients **(A)**. The nomogram was constructed based on multivariate logistic regression and consisted of three clinical factors and 16 radiomics signatures. Calibration curves of the different models in training cohort **(B)** and validation cohort **(C)**; the y-axis represents the actual microsatellite instability rate and the x-axis represents the predicted microsatellite instability risk. The diagonal dotted line indicates that the predicted outcome perfectly corresponds with the actual outcome. The solid line indicates the bias-corrected accuracy of the different models, with a closer fit to the diagonal dotted line representing a better prediction. Decision curve analysis of the different models in training cohort **(D)** and validation cohort **(E)**; the y-axis represents the net benefit, which is calculated by subtracting the expected harm (false positives) from the expected benefit (gaining true positives) and subtracting expected harm (deleting false positives). The higher curve at any given threshold probability is the optimal prediction to maximize net benefit. The solid colored lines represent the different models. The solid gray line represents the assumption that all patients had microsatellite instability. The solid black line represents the assumption that no patients had microsatellite instability. APR, radiomics model of arterial phase; DPR, radiomics model of delayed phase; VPR, radiomics model of venous phase; FR, radiomics model of fusion of arterial phase, delayed phase and venous phase features; Clinical Radiomics, fusion of clinical risk factors and radiomics features of delayed phase; HL, Hosmer-Lemeshow test.

In the present study, the incidence of MSI was 15.19% (67/441) in the training cohort and 14.75% (9/61) in the external validation cohort, which is consistent with previous literature (32, 33). CRC patients with MSI have distinct prognoses and treatment strategies compared to patients with MSS tumors, including better prognosis and benefits from fluorouracil chemotherapy; moreover, MSI may be a negative marker for immunotherapy. Previous studies (28, 29) have investigated the association between MSI and radiomics features. Fan et al. (28) used CT-based radiomics to predict the MSI status in 119 stage II CRC patients. The predictive AUC of the radiomics model (combination of clinical factors and radiomics features) was

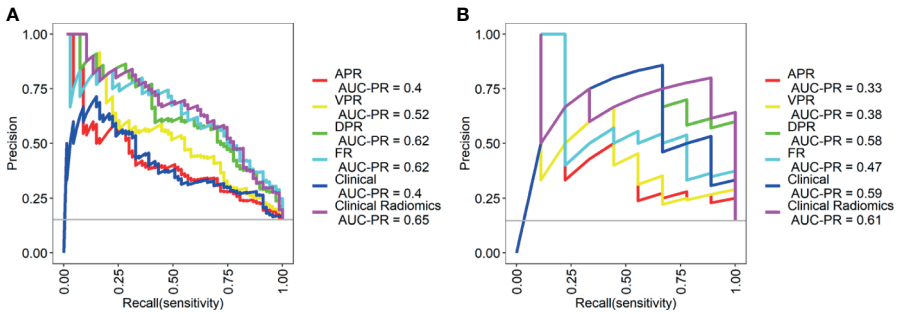


FIGURE 5 | Precision-recall (PR) curves of the different models in the training cohort **(A)** and validation cohort **(B)**. PR represents the relationship between precision and recall. The larger the area under the curve value of the PR curve, the better the model performance. Precision = true positive/(true positive + false positive); recall = true positive/(true positive + false negative). APR, radiomics model of arterial phase; DPR, radiomics model of delayed phase; VPR, radiomics model of venous phase; FR, radiomics model of fusion of arterial phase, delayed phase and venous phase features; Clinical Radiomics, fusion of clinical risk factors and radiomics features of delayed phase.

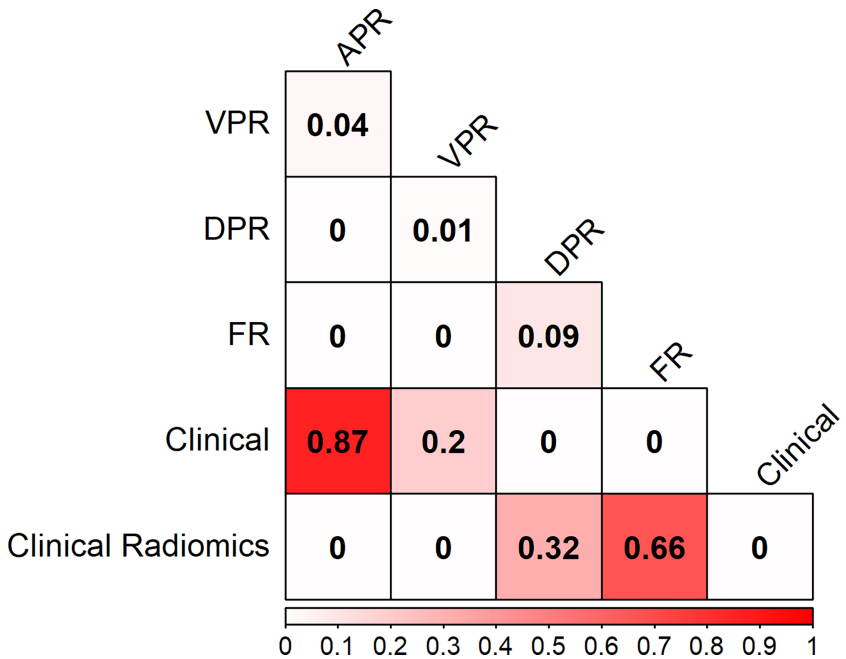


FIGURE 6 | Heat map comparison of the different models in the training cohort. The values in the matrix represent the results of Delong test between two models. APR, radiomics model of arterial phase; DPR, radiomics model of delayed phase; VPR, radiomics model of venous phase; FR, radiomics model of fusion of arterial phase, delayed phase and venous phase features; Clinical Radiomics, fusion of clinical risk factors and radiomics features of delayed phase.

0.752. Pernicka et al. (29) proposed a CT-based radiomics model for the prediction of MSI in stage II–III colon cancer. The combined model (combination of clinical factors and radiomics features) had moderate diagnostic efficacy, with AUC values of 0.80 and 0.79 in the training and validation sets, respectively. Both studies contained small samples and lacked effective validation of external data. Our proposed clinicoradiomics combined model performed better than previous models in both training and external validation cohorts. Therefore, it may be a potential quantitative tool for individualized MSI prediction.

Due to the low incidence of MSI, the data distribution in this study was significantly unbalanced. The unbalanced distribution of data is a common problem in classification. Therefore, the bias adjustment method was used to overcome the training fit error in our study. A previous study used synthetic minority over-sampling technique (SMOTE) methods (28) to resolve data imbalance. The SMOTE method is based on increasing the "artificial" sample to resolve the imbalance of the data set. However, this strategy is prone to model overfitting and is difficult to demonstrate validity.

In our study, 1037 quantitative features were extracted from CT images to build radiomics signatures. During the image preprocessing stage, LoG and wavelet filters (27) were applied to process the original image. Of the 1037 radiomics features, 6, 10, and 16 features were retained in AP, DP, and VP images, respectively, all of which demonstrated high correlations with MSI and were stable across multiple centers. To our surprise, the majority of radiomics features were LoG and wavelet filter features (26/32 in radiomics features) in the present study, which means that LoG and wavelet filters can improve the efficiency of capturing more phenotypic features related to MSI of CRC.

In the present study, the texture feature was the most frequent radiomics feature in triphasic enhanced CT signatures (4/6 in AP, 9/10 in VP, 13/16 in DP). Texture features are microscopic features in an image that have been shown to be highly correlated with tumor heterogeneity (34, 35). However, these features are not easily identified by the human eye and cannot be interpreted as having a clear meaning (36). Our results showed that most texture features were associated with MSI status. Compared with the MSS group, the values of these features were significantly higher in the MSI group, which indicated more homogeneity in the ROI. Our finding is in line with those of previous reports (28, 29) that texture features were also the most frequent radiomics features for MSI prediction. We observed that first-order statistic features including A_wavelet.LHL_firstorder_90Percentile, A_wavelet.LHL_firstorder_Skewness, V_wavelet.HLH_firstorder_Mean, D_original_firstorder_Range, and D_wavelet.LLL_firstorder_Skewness were significantly associated with MSI status, which was consistent with the results of the studies by Fan et al. and Pernicka et al. (28, 29) The results of their studies show that the MSI status is associated with kurtosis and intensity histograms.

Among the triphasic enhanced CT models for the prediction of the MSI status in the training cohort, the DPR model showed the highest performance, with an AUC value of 0.887, followed by 0.827 in the VPR model and 0.775 in the APR model. A similar trend was found in the validation cohort; the predictive AUCs of the DPR, VPR, and APR models were 0.953, 0.810, and 0.876, respectively. Although the VP is the most commonly used phase in gastrointestinal radiomics research, and previous radiomics features for MSI prediction were extracted from portal VP CT images. However, to date, this is the first study to develop a radiomics based model to predict the risk of MSI status in CRC patients based on triphasic enhanced CT with big data. To our surprise, the DPR model showed the best predictive performance in the training and validation cohorts. The triphasic enhanced phase images reflect the uptake and clearance of iodine over time in AP, VP, and DP (37). In AP, the contrast agent is mainly in the intervascular space, which results in focal mucosa enhancement. During VP and DP, the contrast agent is evenly distributed between the intervascular space and the extravascular space, leading to a well-proportioned enhancement (38). The degree of tumor enhancement in AP is positively correlated with the density of microvessels in the tumor, while in VP and DP, the degree of tumor enhancement is related to the content of contrast agent in the tumor interstitial space and vascular space. In addition, CRCs lack normal lymphatic drainage, and the contrast agent tends to remain in the tumor interstitial space for a longer time (38, 39). Therefore, CRCs are significantly enhanced in AP, while VP and DP show continuous enhancement. Previous literature shows that the increase in structure in the enhanced image is proportional to iodine concentration (38). The high content and uniform distribution of contrast agents in tumors may be one of the reasons for the high diagnostic efficiency of the DPR model. This is exactly the same as the number of key features in our study. The numbers of radiomics signatures in DP, VP, and AP were 16, 10, and 6, respectively.

The dynamic changes of CRC from AP, VP to DP showed obvious transmural enhancement from inside the tumor to the outside. For triphasic enhanced CT, AP is mainly used for tumor detection and assessment of the tumor extent along the colorectal wall, VP is used for differentiating CRC from adjacent organs and evaluating lymph nodes, and DP is used to determine the depth of tumor invasion (40). Therefore, the range of tumors detected in DP is larger than that in VP or AP. This means that the ROI delineation range of the DP is the largest during the delineation of the tumor in triphasic enhanced CT images, which is consistent with our observations in the process of delineating tumor ROIs. A positive correlation between increased tumor range and increased diagnostic efficiency has been confirmed by previous studies (41). From the above description, another reason for the high predictive performance of the DPR model could be the large ROI range of tumors in DP images.

Age, location, and CEA were independent predictors of MSI status in the multivariate analysis. CRC patients with an MSI status have distinct clinical characteristics compared to those with MSS tumors, such as a predominance of right-sided colonic tumors, and early age. Our finding is consistent with the results of a previous study (28, 29, 42). CEA levels were significantly lower in the MSI group than in the MSS group, while CA125 was significantly higher in the MSI group than in the MSS group in the present study. A significant correlation between MSI status and the above clinical predictors suggests that genetic alterations may have independent influences on CRC development, thus resulting in distinct tumor biological behavior compared with that of MSS tumors. These parameters could be easily obtained and thus considered as novel approaches for predicting MSI status. Further studies are essential to validate our findings.

As for radiation dose, the average dose length product of triphasic enhanced scans was 1934.76 ± 147.18 mGy*cm, which is slightly higher than the diagnostic reference level for adults (1490 mGy*cm) published by China's National Health Industry standard (WS/T 637-2018) (43). Application of new techniques such as multi-model iterative reconstruction technology could effectively reduce the radiation dose in clinical practice (44).

Several limitations of our study should be noted. First, 501 patients were excluded because they did not meet the inclusion or exclusion criteria, which inevitably produced selection bias. Second, due to the irregular shape of some tumors, manual segmentation is time-consuming and may have observer variability. In future studies, automated segmentation may be a potential tool to resolve this problem. Third, in this study, we

used different imaging instruments and acquisition parameters to complete CT scanning. The influence of different instruments and different parameters on radiomics features is obvious. Therefore, it is important to standardize scanning protocols in different instruments and different institutions.

CONCLUSION

In conclusion, we proposed and validated a CT-based radiomics model, incorporating clinical risk factors and radiomics parameters, which showed a relatively high diagnostic performance for the risk prediction of MSI in patients with CRC. This model may be a potential tool for preoperatively identifying the MSI status and can be used in individualized therapeutic strategy planning and prognostic prediction.

DATA AVAILABILITY STATEMENT

Data are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Lanzhou University Second Hospital medical ethics committee. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

Conception and design: JZho, HB, and YC. Collection and assembly of the data: YC and YY. Development of the methodology: JR. Data analysis and interpretation: All authors. Manuscript writing: All authors. All authors contributed to the article and approved the submitted version.

FUNDING

This study received funding from the National Natural Science Foundation of China (82071872), Open Fun project of Key Laboratory of Medical Imaging of Gansu Province (GSYX202009), Science and Technology Project of Qinghai Province (No. 2017-SF-158) and Qinghai Provincial Key Clinical Specialty Construction Project.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.687771/full#supplementary-material.

REFERENCES

1. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N Engl J Med* (2012) 366(10):883–92. doi: 10.1056/NEJMoa1113205
2. Smith G, Carey FA, Beattie J, Wilkie MJV, Lightfoot TJ, Coxhead J, et al. Mutations in APC, Kirsten-Ras, and P53–Alternative Genetic Pathways to Colorectal Cancer. *Proc Natl Acad Sci U S A* (2002) 99(14):9433–8. doi: 10.1073/pnas.122612899
3. Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, et al. Genetic Alterations During Colorectal-Tumor Development. *N Engl J Med* (1988) 319(9):525–32. doi: 10.1056/NEJM198809013190901
4. Battaglin F, Naseem M, Lenz HJ, Salem ME. Microsatellite Instability in Colorectal Cancer: Overview of its Clinical Significance and Novel Perspectives. *Clin Adv Hematol Oncol* (2018) 16(11):735–45.
5. Mcdermott U, Longley DB, Johnston PG. Molecular and Biochemical Markers in Colorectal Cancer. *Ann Oncol* (2002) 13(Suppl 4):235–45. doi: 10.1093/annonc/mdf665
6. Gelsomino F, Barbolini M, Spallanzani A, Pugliese G, Cascinu S. The Evolving Role of Microsatellite Instability in Colorectal Cancer: A Review. *Cancer Treat Rev* (2016) 51:19–26. doi: 10.1016/j.ctrv.2016.10.005
7. Yang G, Zheng RY, Jin ZS. Correlations Between Microsatellite Instability and the Biological Behaviour of Tumours. *J Cancer Res Clin Oncol* (2019) 145 (12):2891–9. doi: 10.1007/s00432-019-03053-4
8. Sepulveda AR, Hamilton SR, Allegra CJ, Grody W, Cushman-Vokoun AM, Funkhouser WK, et al. Molecular Biomarkers for the Evaluation of Colorectal Cancer Guideline From the American Society for Clinical Pathology, College of American Pathologists, Association for Molecular Pathology, and American Society of Clinical Oncology. *J Clin Oncol* (2017) 35(13):1453–86. doi: 10.1200/JCO.2016.71.9807
9. Ribic CM, Sargent DJ, Moore MJ, Thibodeau SN, French AJ, Goldberg RM, et al. Tumor Microsatellite-Instability Status as a Predictor of Benefit From Fluorouracil-Based Adjuvant Chemotherapy for Colon Cancer. *N Engl J Med* (2003) 349(3):247–57. doi: 10.1056/NEJMoa022289
10. Sargent DJ, Marsoni S, Monges G, Thibodeau SN, Labianca R, Hamilton SR, et al. Defective Mismatch Repair as a Predictive Marker for Lack of Efficacy of Fluorouracil-Based Adjuvant Therapy in Colon Cancer. *J Clin Oncol* (2010) 28 (20):3219–26. doi: 10.1200/JCO.2009.27.1825
11. Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep Learning can Predict Microsatellite Instability Directly From Histology in Gastrointestinal Cancer. *Nat Med* (2019) 25(7):1054–6. doi: 10.1038/s41591-019-0462-y
12. Malesci A, Laghi L, Bianchi P, Delconte G, Randolph A, Torri V, et al. Reduced Likelihood of Metastases in Patients With Microsatellite-Unstable Colorectal Cancer. *Clin Cancer Res* (2007) 13(13):3831–9. doi: 10.1158/1078-0432.CCR-07-0366
13. Sargent DJ, Shi Q, Yothers G, Tejpar S, Bertagnolli MM, Thibodeau SN, et al. Prognostic Impact of Deficient Mismatch Repair (Dmmr) in 7,803 Stage II/III Colon Cancer (CC) Patients (Pts): A Pooled Individual Pt Data Analysis of 17 Adjuvant Trials in the ACCENT Database. *Asco Meeting Abstracts* (2014) 32. doi: 10.1200/jco.2014.32.15_suppl.3507
14. Dienstmann R, Mason MJ, Sinicrope FA, Phipps AI, Tejpar S, Nesbakken A, et al. Prediction of Overall Survival in Stage II and III Colon Cancer Beyond TNM System: A Retrospective, Pooled Biomarker Study. *Ann Oncol* (2017) 28 (5):1023–31. doi: 10.1093/annonc/mdx052
15. Cutsem V E, Cervantes A, Adam R. ESMO Consensus Guidelines for the Management of Patients With Metastatic Colorectal Cancer. *Ann Oncol* (2016) 27(8):1386–422. doi: 10.1093/annonc/mdw235
16. Watanabe T, Muro K, Ajioka Y, Hashiguchi Y, Ito Y, Saito Y, et al. Japanese Society for Cancer of the Colon and Rectum (JSCCR) Guidelines 2016 for the

Treatment of Colorectal Cancer. *Int J Clin Oncol* (2018) 23(1):1–34. doi: 10.1007/s10147-017-1101-6
17. Umar A, Boland CR, Terdiman JP, Syngal S, Srivastava S. Revised Bethesda Guidelines for Hereditary Nonpolyposis Colorectal Cancer (Lynch Syndrome) and Microsatellite Instability. *J Natl Cancer Inst* (2004) 96(4):261–8. doi: 10.1093/jnci/djh034
18. Itakura H, Achrol AS, Mitchell LA, Loya JJ, Liu T, Westbroek EM, et al. Magnetic Resonance Image Features Identify Glioblastoma Phenotypic Subtypes With Distinct Molecular Pathway Activities. *Sci Transl Med* (2015) 7(303):303ra138–303ra138. doi: 10.1126/scitranslmed.aaa7582
19. Sacher AG, Dahlberg SE, Heng J, Mach S, Jn PA, Oxnard GR. Association Between Younger Age and Targetable Genomic Alterations and Prognosis in non–Small-Cell Lung Cancer. *JAMA Oncol* (2016) 2(3):313–20. doi: 10.1001/jamaoncol.2015.4482
20. Wen-Yue Y, Jing H, Li X, Lei C, Mi Y, Li L, et al. Prediction of Biological Behavior and Prognosis of Colorectal Cancer Patients by Tumor MSI/MMR in the Chinese Population. *Onco Targets Ther* (2016) 9:7415–24. doi: 10.2147/OTT.S117089
21. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol* (2016) 34(18):2157–64. doi: 10.1200/JCO.2015.65.9128
22. Nie K, Shi L, Chen Q, Hu X, Jabbour SK, Yue N, et al. Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome Based on Radiomics of Multiparametric MRI. *Clin Cancer Res* (2016) 22(21):5256–64. doi: 10.1158/1078-0432.CCR-15-2997
23. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are More Than Pictures, They are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
24. Zhang S, Zhang B, Tian J, Dong D, Gu DS, Dong YH, et al. Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma. *Clin Cancer Res* (2017) 23(15):4259–69. doi: 10.1158/1078-0432.CCR-16-2910
25. Huang X, Cheng Z, Huang Y, Liang C, He L, Ma Z, et al. CT-Based Radiomics Signature to Discriminate High-Grade From Low-Grade Colorectal Adenocarcinoma. *Acad Radiol* (2018) 25(10):1285–97. doi: 10.1016/j.acra.2018.01.020
26. Horvat N, Veeraraghavan H, Khan M, Blazic I, Zheng J, Capanu M, et al. MR Imaging of Rectal Cancer: Radiomics Analysis to Assess Treatment Response After Neoadjuvant Therapy. *Radiology* (2018) 287(3):833–43. doi: 10.1148/radiol.2018172300
27. Cui Y, Liu H, Ren J, Du X, Wang D. Development and Validation of a MRI-Based Radiomics Signature for Prediction of KRAS Mutation in Rectal Cancer. *Eur Radiol* (2020) 30(4):1948–58. doi: 10.1007/s00330-019-06572-3
28. Fan S, Li X, Cui X, Zheng L, Ren X, Ma W, et al. Computed Tomography-Based Radiomic Features Could Potentially Predict Microsatellite Instability Status in Stage II Colorectal Cancer: A Preliminary Study. *Acad Radiol* (2019) 26(12):1633–40. doi: 10.1016/j.acra.2019.02.009
29. Golia Pernicka JS, Gagniere J, Chakraborty J, Yamashita R, Nardo L, Creasy JM, et al. Radiomics-Based Prediction of Microsatellite Instability in Colorectal Cancer at Initial Computed Tomography Evaluation. *Abdom Radiol (NY)* (2019) 44(11):3755–63. doi: 10.1007/s00261-019-02117-w
30. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5644
31. Griethuysen JJMV, Fedorov A, Parmar C, Hosny A, Aerts HJWL. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339
32. Kawakami H, Zaanan A, Sinicrope FA. Microsatellite Instability Testing and its Role in the Management of Colorectal Cancer. *Curr Treat Options Oncol* (2015) 16(7):30. doi: 10.1007/s11864-015-0348-2
33. Pino MS, Chung DC. The Chromosomal Instability Pathway in Colon Cancer. *Gastroenterology* (2010) 138(6):2059–72. doi: 10.1053/j.gastro.2009.12.065
34. Liu H, Zhang C, Wang L, Luo R, Li J, Zheng H, et al. MRI Radiomics Analysis for Predicting Preoperative Synchronous Distant Metastasis in Patients With Rectal Cancer. *Eur Radiol* (2019) 29(8):4418–26. doi: 10.1007/s00330-018-5802-7
35. Meng X, Xia W, Xie P, Zhang R, Li W, Wang M, et al. Preoperative Radiomic Signature Based on Multiparametric Magnetic Resonance Imaging for Noninvasive Evaluation of Biological Characteristics in Rectal Cancer. *Eur Radiol* (2019) 29(6):3200–9. doi: 10.1007/s00330-018-5763-x
36. Zhang J, Yao K, Liu P, Liu Z, Zhou J. A Radiomics Model for Preoperative Prediction of Brain Invasion in Meningioma non-Invasively Based on MRI: A Multicentre Study. *EBioMedicine* (2020) 58:102933. doi: 10.1016/j.ebiom.2020.102933
37. Wu J, Lv Y, Wang N, Zhao Y, Zhang P, Liu Y, et al. The Value of Single-Source Dual-Energy CT Imaging for Discriminating Microsatellite Instability From Microsatellite Stability Human Colorectal Cancer. *Eur Radiol* (2019) 29(7):3782–90. doi: 10.1007/s00330-019-06144-5
38. Miles KA. Tumour Angiogenesis and its Relation to Contrast Enhancement on Computed Tomography: A Review. *Eur J Radiol* (1999) 30(3):198–205. doi: 10.1016/S0720-048X(99)00012-1
39. Zhang M, Kono M. Solitary Pulmonary Nodules: Evaluation of Blood Flow Patterns With Dynamic CT. *Radiology* (1997) 205(2):471–8. doi: 10.1148/radiology.205.2.9356631
40. Lee JH, Jeong YK, Kim DH, Go BK, Woo YJ, Ham SY, et al. Two-Phase Helical CT for Detection of Early Gastric Carcinoma: Importance of the Mucosal Phase for Analysis of the Abnormal Mucosal Layer. *J Comput Assist Tomogr* (2000) 24(5):777–82. doi: 10.1097/00004728-200009000-00020
41. Beig N, Khorrami M, Alilou M, Prasanna P, Madabhushi A. Perinodular and Intranodular Radiomic Features on Lung CT Images Distinguish Adenocarcinomas From Granulomas. *Radiology* (2018) 290(3):180910. doi: 10.1148/radiol.2018180910
42. Wu J, Zhang Q, Zhao Y, Liu Y, Chen A, Li X, et al. Radiomics Analysis of Iodine-Based Material Decomposition Images With Dual-Energy Computed Tomography Imaging for Preoperatively Predicting Microsatellite Instability Status in Colorectal Cancer. *Front Oncol* (2019) 9:1250. doi: 10.3389/fonc.2019.01250
43. Cao Y, Zhang G, Bao H, Zhang S, Zhang J, Zhao Z, et al. Development of a Dual-Energy Spectral CT Based Nomogram for the Preoperative Discrimination of Mutated and Wild-Type KRAS in Patients With Colorectal Cancer. *Clin Imaging* (2020) 69:205–12. doi: 10.1016/j.clinimag.2020.08.023
44. Jia Y, Zhai B, He T, Yu Y, Yu N, Duan H, et al. The Application of a New Model-Based Iterative Reconstruction in Low-Dose Upper Abdominal CT. *Acad Radiol* (2019) 26(10):e275–83. doi: 10.1016/j.acra.2018.11.020

Conflict of Interest: Author JR was employed by company GE Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Cao, Zhang, Zhang, Yang, Ren, Yan, Wang, Zhao, Huang, Bao and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 25 June 2021
doi: 10.3389/fonc.2021.671420

Radiomics Analysis of Contrast-Enhanced CT Predicts Survival in Clear Cell Renal Cell Carcinoma

Lei Yan[1†], Guangjie Yang[1†], Jingjing Cui[2], Wenjie Miao[1], Yangyang Wang[1], Yujun Zhao[1], Ning Wang[3], Aidi Gong[1], Na Guo[2], Pei Nie[4] and Zhenguang Wang[1*]*

[1] Department of Positron Emission Tomography-Computed Tomography (PET-CT) Center, The Affiliated Hospital of Qingdao University, Qingdao, China, [2] Scientific Research Department, Huiying Medical Technology Co., Ltd., Beijing, China, [3] Department of Radiology, Shandong Provincial Hospital, Jinan, China, [4] Department of Radiology, The Affiliated Hospital of Qingdao University, Qingdao, China

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln, United Kingdom

Reviewed by:
Qingtao Qiu,
Shandong Cancer Hospital, China
Shaofeng Duan,
GE Healthcare, China

***Correspondence:**
Pei Nie
niepei1984@126.com
Zhenguang Wang
doctorwzg2002@hotmail.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *23 February 2021*
Accepted: *07 June 2021*
Published: *25 June 2021*

Citation:
Yan L, Yang G, Cui J, Miao W, Wang Y, Zhao Y, Wang N, Gong A, Guo N, Nie P and Wang Z (2021) Radiomics Analysis of Contrast-Enhanced CT Predicts Survival in Clear Cell Renal Cell Carcinoma. Front. Oncol. 11:671420. doi: 10.3389/fonc.2021.671420

Purpose: To develop and validate the radiomics nomogram that combines clinical factors and radiomics features to estimate overall survival (OS) in patients with clear cell renal cell carcinoma (ccRCC), and assess the incremental value of radiomics for OS estimation.

Materials and Methods: One hundred ninety-four ccRCC cases were included in the training cohort and 188 ccRCC patients from another hospital as the test cohort. Three-dimensional region-of-interest segmentation was manually segmented on multiphasic contrast-enhanced abdominal CT images. Radiomics score (Rad-score) was calculated from a formula generated *via* least absolute shrinkage and selection operator (LASSO) Cox regression, after which the association between the Rad-score and OS was explored. The radiomics nomogram (clinical factors + Rad-score) was developed to demonstrate the incremental value of the Rad-score to the clinical nomogram for individualized OS estimation, which was then evaluated in relation to calibration and discrimination.

Results: Rad-score, calculated using a linear combination of the 11 screened features multiplied by their respective LASSO Cox coefficients, was significantly associated with OS. Calibration curves showed good agreement between the OS predicted by the nomograms and observed outcomes. The radiomics nomogram presented higher discrimination capability compared to clinical nomogram in the training (C-index: 0.884; 95% CI: 0.808–0.940 *vs.* 0.803; 95% CI: 0.705–0.899, $P < 0.05$) and test cohorts (C-index: 0.859; 95% CI: 0.800–0.921 *vs.* 0.846; 95% CI: 0.777–0.915, $P < 0.05$).

Conclusions: The radiomics nomogram may be used for predicting OS in patients with ccRCC, and radiomics is useful to assist quantitative and personalized treatment.

Keywords: radiomics, Rad-score, nomogram, clear cell renal cell carcinoma, overall survival

INTRODUCTION

Renal cell carcinoma (RCC) is one of the most common urological malignancies that accounts for approximately 3.8% of all human cancers (1). Approximately 300,000 patients are diagnosed with RCC every year, resulting in 140,000 RCC-related deaths (2). The overall 5-year survival rate for the majority of patients (65%) diagnosed with localized RCC is 93%, whereas for those with lymph node metastasis or distant metastasis is 66 and 12%, respectively (3).

Clear cell renal cell carcinoma (ccRCC) constitutes the highest proportion (90%) of all diagnosed RCC and is the subtype with the poorest prognosis. In addition, ccRCC is associated with high metastatic potential (4). It is essential to find accurate predictive information for valid prognosis assessment, treatment planning, and implementation of surveillance strategies. Due to the lack of reliable biomarkers, prognostic prediction is primarily based on the combination of stage, grade, and histology (5). The conventional prognostic evaluation relies on the American Joint Committee on Cancers (AJCC) tumor-node-metastasis (TNM) staging system (6). However, the accuracy of the prognostic system for individualized prediction is limited (7). The prognostic models —Memorial Sloan Kettering Cancer Center [MSKCC]; Stage, Size, Grade, and Necrosis [SSIGN]; and the University of California at Los Angeles Integrated Staging System [UISS]— which include more clinical and pathological factors, have been proposed to guide the follow-up in RCC patients, including ccRCC (8–10). However, a prospective study that evaluated 1,647 patients with RCC reported that the above models have a slightly better prediction efficiency compared to the TNM staging system (5). Furthermore, the predictive ability of these models was instable after 2 years after diagnosis. Consequently, more accurate prediction models are needed to achieve precise and individualized prognostic evaluation.

Radiomics is a relatively new approach that extracts features from multimodality medical images using data-characterization algorithms (11). Over the last decade, radiomics features have been applied as imaging biomarkers for prognosis, staging, and prediction of cancer (12). The radiomics approach has been successfully applied to predict metastasis, recurrence, and other clinical outcomes of lung cancer, breast cancer, and colorectal cancer (13–15). To the best of our knowledge, the majority of previous radiomics studies related to RCC focused on the differentiation between malignant and benign renal lesions and the prediction of nuclear grading (16, 17), while only a few studies reported on radiomics-based research for prediction of overall survival (OS) in ccRCC. In this study, we investigated whether the radiomics features extracted from the enhanced CT images could be used to quantitatively assess the OS in patients with ccRCC.

MATERIALS AND METHODS

Patients

Ethical approval for this retrospective study was obtained by the Ethics Committee of the Affiliated Hospital of Qingdao University. A total of 382 patients with ccRCC were enrolled in this two-center study. One hundred ninety-four patients from the Affiliated Hospital of Qingdao University (from May 2011 to December 2016) were collected as the training cohort, and 188 patients from Shandong Provincial Hospital (from June 2012 to May 2017) were collected as the independent test cohort. The detailed recruitment pathway for patients in this study is presented in **Figure 1**.

Three hundred fifty-one patients with TNM group I, II, and III were treated with radical or partial nephrectomy. Among 31 patients with TNM group IV, 29 patients underwent cytoreductive nephrectomy or tumors resection, and two patients with surgically unresectable tumors were treated with tissue sampling. After diagnosis, 20 out of the 31 patients received one or fewer systemic treatments (targeted therapy, immunotherapy, or chemotherapy).

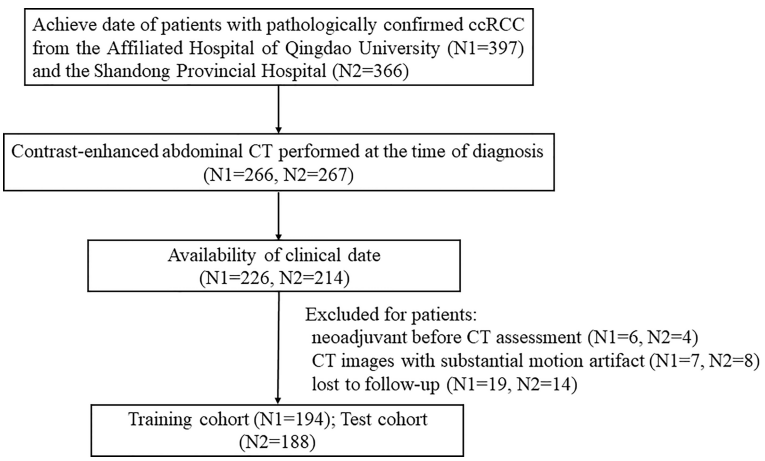


FIGURE 1 | The patient recruitment pathway with inclusion and exclusion details.

The beginning of the follow-up was the date of diagnosis by pathological examination. Patients were reviewed every 6 to 12 months for the first 2 years, then every year after that. The review included clinical physical examination, chest X-ray, abdominal ultrasound, abdominal CT, or MRI. All data were collected in July 2019, and a follow-up phone call was conducted for patients who were not able to visit the hospital. The endpoint of the study was the time of death or the date of the last follow-up.

Baseline data consisted of age, sex, TNM group (I, II, III, or IV), Fuhrman grade, presence of histologic necrosis, hemoglobin (HB), neutrophil count (NE), lymphocyte count (LY), platelet count (PLT), blood urea nitrogen (BUN), creatinine (CREA), Eastern Cooperative Oncology Group Performance Status (ECOG-PS), and calculated neutrophil-lymphocyte ratio (NLR).

CT Image Acquisition, Region-of-Interest Segmentation, and Radiomics Feature Extraction

CT scan protocols are explained in **Supplementary Methods**. The unenhanced abdominal scan was performed first. The enhanced scan was performed after injecting a 90–100 ml iodinated contrast agent (Ultravist 370, Bayer, Germany) into the antecubital vein at a flow rate of 2.5–3.0 ml/s. Images of corticomedullary phase (CMP), nephrographic phase (NP), and excretory phase (EP) were obtained at 30–35, 80–90, and 300–350 s after contrast injection, respectively.

Pre-treatment contrast-enhanced abdominal CT was exported in DICOM form from the picture archiving and communication system (PACS) workstation. Three-dimensional (3-D) ROI segmentation was manually performed using the ITK-SNAP software (Version 3.8.0, www.itksnap.org). Before feature extraction, image resampling and gray-level discretization were applied for the standardization of three-phase CT images. A total of 1,409 quantitative imaging features were extracted from each phase of CT images with the Radcloud platform (Huiying Medical Technology Co., Ltd). Radcloud platform was utilized to process the imaging and clinical data, as well as the spectra of radiomics analysis. The platform feature extraction is based on the "pyradiomics" package in Python (version 2.2.0, https://pyradiomics.readthedocs.io/). The features were grouped into three groups: (1) first-order statistics features describe the distribution of voxel intensities; (2) size- and shape-based features that reflect the size and shape of the region; (3) texture features that can quantify region heterogeneity differences. In addition, several filters were used to calculate the intensity and texture features on original images and derived images. The details of the radiomics features are shown in **Supplementary Methods**.

To obtain the inter-class correlation coefficient, ROIs of the 30 patients were segmented by two radiologists with 5 and 10 years of abdominal imaging experience, respectively. Then, the first radiologist completed the segmentation of the 30 ROIs after 2 weeks to obtain an intra-class correlation coefficient. To enhance the stability and reproducibility, radiomics features derived from the ROIs with both inter- and intra-class >0.75 were retained in the analysis and used in the following study. The first radiologist delineated the remaining ROIs.

Feature Selection and Radiomics Score Calculation

The least absolute shrinkage and selection operator (LASSO) penalized Cox proportional hazards regression, which is appropriate for reducing high-dimensional data, was applied to select the optimal prognostic features in the training cohort. A formula was generated *via* a linear combination of the screened features multiplied by their respective LASSO Cox coefficients. Then, the formula was used to calculate the Rad-score of each individual. The median Rad-score was applied as a cutoff that stratified patients into the high-risk group with short survival time and the low-risk group with long survival time. The association of Rad-score with OS was estimated using the Kaplan-Meier survival analysis, and the difference in survival between the stratified subgroups was determined using the log-rank test.

Rad-Score Assessment

Both 3-year and 5-year OS were described and analyzed in the training and test cohorts. The distribution of Rad-score in the 3-year survival group and the dead group was illustrated in a violin plot (box plot in the middle and a density plot on the side); Wilcoxon rank-sum test was used to analyze the significant difference. In addition, the prognostic accuracy of the Rad-score for the 3-year survival group and the dead group was assessed through the time-dependent receiver operating characteristic (ROC) analysis and the correlated area under the ROC curve (AUC). Rad-score for the 5-year survival group and the dead group was analyzed using the same processes.

Development of the Clinical and Radiomics Nomograms

Clinical factors were assessed for their impact on OS by the univariable and multivariable Cox regression analysis in the training cohort. The clinical nomogram for probability prediction of 3- and 5-year OS was developed based on the multivariable Cox regression analysis. Independent prognostic clinical factors were selected through the last step, and Rad-score was used to develop the radiomics nomogram by the multivariable Cox regression analysis. The variables were no longer excluded in this section. The relative hazard ratio (HR) of each factor from the two nomograms was obtained simultaneously. The severity of multicollinearity among variables was detected using the Variance inflation factor (VIF) before the nomograms development regarding respective multivariable Cox regression. If VIF was <10, the multicollinearity was low.

Performance of the Clinical and Radiomics Nomograms

To assess the performance of the clinical and radiomics nomograms, calibration and discrimination were performed in the training cohort and then validated in the test cohort. The calibration curve was used to indicate the agreement between the OS predicted by the nomograms and the observed outcomes after bias correction in 3- and 5-year OS. The Harrell concordance index (C-index) was measured to quantify the discrimination performance of the nomograms. To evaluate

the goodness-of-fit of the nomograms, the Akaike information criterion (AIC) was generated. The discrimination capability of the two nomograms was compared to assess the incremental value of the Rad-score to the general clinical risk factors for an individualized assessment of OS in patients with ccRCC.

Additionally, the risk group, which was predicted by the nomogram with better discrimination capability, was used as a prediction factor to generate the Kaplan-Meier survival curves. If survival probability at 3 or 5 years predicted by the nomogram was <0.5, the patient was stratified into the nomogram-predicted (Nomo-predicted) high-risk group; if it was >0.5, the patient was stratified into the Nomo-predicted low-risk group. The difference between survival curves was assessed by using the log-rank test. The Kaplan-Meier survival curves were depicted by using actual survival status as prediction factor as well. The workflow of the study is shown in **Supplementary Figure 1**.

Statistical Analysis

The statistical analysis was performed with SPSS version 24.0 and R software (version 3.3.3 https://www.r-project.org). The Student's t-test, Chi-square test, or Non-parametric Mann-Whitney U test were applied to compare the differences in clinical factors between the training and test cohorts using SPSS software. Analysis and figure plots for the remaining data were performed using R software. A two-sided $P < 0.05$ was considered to be statistically significant.

RESULTS

Clinical Factors and OS

By the time of the last follow-up, 60 patients (15.7%) died. The mean OS was 56 months, and the median OS was 59.0 months (interquartile range: 49.8–68.0 months) in the training cohort; the mean and median OS for the test cohort was 46 months and 43.5 months (interquartile range: 32.3–60.0 months), respectively. A significant difference in OS was observed between the two cohorts ($P < 0.05$), which was calculated based on the differences at follow-up time. No difference was found in sex, age, TNM group, presence of histologic necrosis, ECOG-PS, HB, LY, and BUN between the two cohorts (all $P > 0.05$), While differences in Fuhrman grade, NE, NLR, PLT, and CREA distribution were statistically significant ($P < 0.05$) (**Table 1**).

TABLE 1 | Baseline clinical date of the training and test cohorts.

Characteristic	Training cohort (n = 194)	Test cohort (n = 188)	P value
Sex			0.901[a]
Male	133	130	
Female	61	58	
Age (years)	55.94 ± 11.59	56.08 ± 10.96	0.906[b]
TNM group			0.597[a]
I	138	129	
II	29	29	
III	10	16	
IV	17	14	
Fuhrman grade			0.004[a]
I	11	16	
II	124	143	
III	50	27	
IV	9	2	
Necrosis			0.721[a]
Present	84	78	
Absent	110	110	
ECOG-PS			0.282[a]
0	105	112	
≥1	89	76	
Hemoglobin (g/L)	136.57 ± 19.797	138.70 ± 20.989	0.309[b]
Neutrophil count (10^9/L)	3.746 ± 1.702	4.125 ± 1.751	0.032[b]
Lymphocyte count (10^9/L)	1.972 ± 0.658	1.880 ± 0.660	0.171[b]
Neutrophil-lymphocyte ratio	2.200 ± 1.861	2.783 ± 3.183	0.029[b]
Platelet count (10^9/L)	232.78 ± 65.225	248.09 ± 81.951	0.044[b]
Creatinine (umol/L)	88.626 ± 20.399	69.449 ± 18.189	<0.001[b]
Blood urea nitrogen (mmol/L)	6.268 ± 7.200	5.496 ± 1.829	0.154[b]
Overall Survival (month)			<0.001[c]
Median	59.0 (49.8–68.0)	43.5 (32.3–60.0)	
Mean	56.4 ± 17.3	46.0 ± 18.4	

ECOG-PS, Eastern Cooperative Oncology Group Performance Status.
[a]*Chi-square test.*
[b]*Student's t-test.*
[c]*Mann-Whitney U test.*

Feature Extraction, Selection, and Rad-score Calculation

After excluding the subjective difference in ROI segmentation by observers, we only retained the repeatable and stable radiomics features with the inter- and intra-class correlation coefficients >0.75 and obtained 3,485 three-phase CT imaging features. After that, 11 optimal prognostic features were screened out through the LASSO Cox regression algorithm (**Figure 2A**), and the radiomics signature was constructed. The signature calculation equation is depicted in **Supplementary Methods**. The bar chart below showed the contribution of selected features with their LASSO Cox regression coefficients for the signature construction (**Figure 2B**).

The distribution of the Rad-score calculated by the equation for each patient is shown in **Supplementary Methods**. A distribution difference was observed in Rad-score between the training and test cohorts ($P > 0.05$). The cut-off value of the Rad-score was −0.04481. Consequently, patients were stratified into high-risk group (Rad-score ≥ −0.04481) and low-risk group (Rad-score < −0.04481). The Kaplan-Meier survival analysis showed a correlation between Rad-score and OS in the training cohort (**Figure 3A**). The low-risk group has a longer OS compared to a high-risk group ($P < 0.001$, log-rank test). The same finding was demonstrated in the test cohort (**Figure 3B**; $P < 0.001$, log-rank test).

Rad-Score Assessment

Violin plots showed that the median Rad-score in 3- or 5-year dead groups were higher than that in 3- or 5-year survival groups in the training and test cohorts (**Figures 4A, B, D, E**), and the distribution of Rad-score was significantly different (all $P < 0.001$). In addition, the AUC of the time-dependent ROC curves (**Figure 4C**) for 3-year OS in the training and test cohorts was 0.902 (95% CI: 0.851–0.940) and 0.857 (95% CI: 0.798–0.903), respectively. Satisfactory prognostic accuracy was achieved for the 5-year OS. AUC of the time-dependent ROC curves (**Figure 4F**) for 5-year OS in the training and test cohorts was 0.904 (95% CI: 0.854–0.942) and 0.850 (95% CI: 0.791–

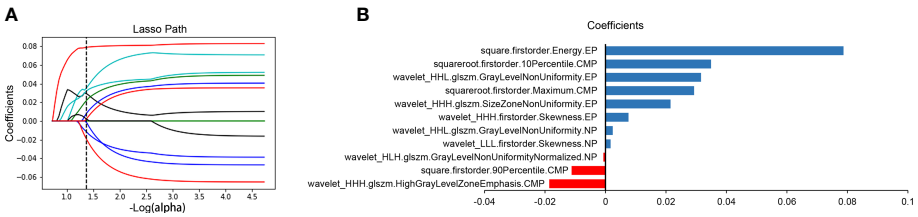


FIGURE 2 | Radiomics feature selection and presentation. Radiomics feature selection using the LASSO Cox regression algorithm. LASSO regression coefficient profiles of survival-associated radiomics features **(A)**. A bar chart shows the contribution of selected features for radiomics signature construction **(B)**. The y-axis represents features that contribute to the signature. The x-axis represents corresponding coefficients in the LASSO Cox analysis. LASSO, least absolute shrinkage and selection operator; CMP, corticomedullary phase; NP, nephrographic phase; EP, excretory phase.

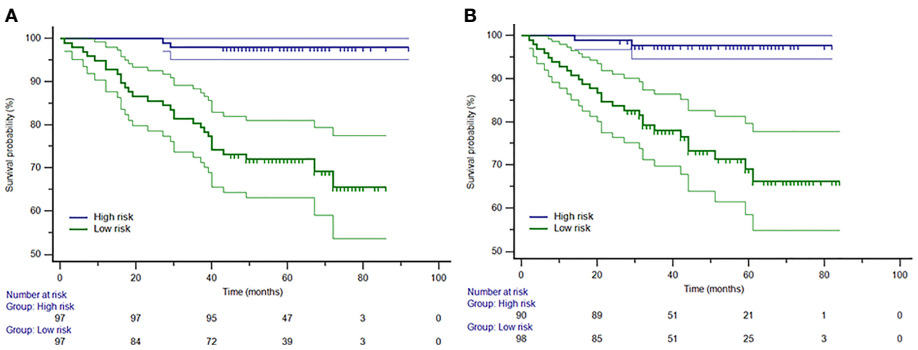


FIGURE 3 | The Kaplan-Meier curves for patients in low-risk and high-risk groups in the training **(A)** and test **(B)** cohorts. The median Rad-score was –0.04481, which was applied as a cutoff that stratified patients into the high-risk group (Rad-score ≥ –0.04481) and the low-risk group (Rad-score < –0.04481).

0.898), respectively. This data proved the discrimination accuracy of OS was reliable and robust when using Rad-score.

Development of the Clinical and Radiomics Nomograms

Clinical factors were selected by the univariable and multivariable Cox regression analysis (**Table 2**). The multivariable Cox regression showed that the TNM group and CREA were independent factors closely correlated with OS. So the clinical nomogram incorporating these two factors was developed (**Figure 5A**). Both of these factors were negatively related to OS for patients (HR for TNM group: 2.912, 95% CI, 2.188–3.875; HR for CREA: 1.020, 95% CI, 1.004–1.036). The radiomics nomogram was developed with Rad-score, TNM group, and CREA by the multivariate Cox regression analysis (**Figure 5B**). All three predictors were negatively correlated to OS (HR for Rad-score: 277.920, 95% CI, 26.039–2,966.297; HR for TNM group: 1.713, 95% CI, 1.149–2.552; HR for CREA: 1.014, 95% CI, 0.998–1.031).

Performance of the Clinical and Radiomics Nomograms

The calibration curves depicted survival probability at 3 and 5 years after diagnosis, showing good agreement between survival probability predicted by the nomograms and observed outcomes in the training and test cohorts (**Figures 6A–D**). Integrating clinical factors and Rad-score, the radiomics nomogram obtained good discrimination performance with a C-index of 0.884 (95% CI: 0.808, 0.940), as well as higher discrimination capability compared with the clinical nomogram ($P < 0.05$) (**Supplementary Methods**). The favorable result was confirmed in the test cohort, which implied the incremental value of the Rad-score for individual OS estimation. AIC and C-index estimates for the two nomograms are listed in **Table 3**.

The Kaplan-Meier survival curves were generated using a risk group based on survival probability at 3 or 5 years predicted by the radiomics nomogram as a prediction factor. A significant difference was confirmed between the stratified Nomo-predicted high-risk and low-risk groups in both the training and test cohorts (**Figures 7B, D, F, H**; $P < 0.001$, log-rank test). These results were consistent with the Kaplan-Meier survival analysis in the actual survival and dead groups (**Figures 7A, C, E, G**; $P < 0.001$, log-rank test).

DISCUSSION

ccRCC is the predominant pathological subtype of renal malignancy associated with aggressive behavior (high invasion and metastasis) and chemoresistance. Patients with ccRCC have the worse OS compared to those with other subtypes (18, 19).

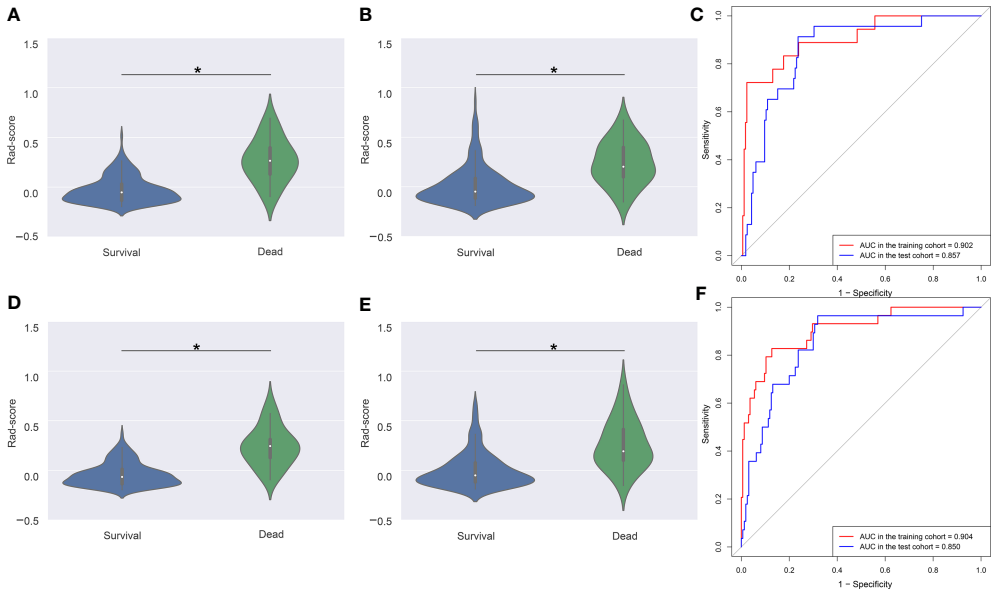


FIGURE 4 | Violin plots and ROC curves of Rad-score in 3- and 5-year survival and dead groups. Violin plots show the distribution of Rad-score in 3-year survival and dead groups in the training **(A)** and test **(B)** cohorts. The median (central white dot), interquartile range (black box), and 95% confidence interval (vertical line) are shown in the middle. The color area represents a density plot on the side. ROC curves **(C)** for 3-year survival were determined to assess prognostic accuracy in the two cohorts. Violin plots **(D, E)** and ROC curves **(F)** in 5-year survival and dead groups. ROC, receiver operator characteristic. AUC, area under the curve. *P < 0.001.

TABLE 2 | Uni- and multivariable COX regression analysis of predictors of OS.

Variable	Univariable Cox regression		Multivariable Cox regression	
	HR (95% CI)	P value	HR (95% CI)	P value
Sex	2.448 (0.940–6.377)	0.066	NA	
Age	1.044 (1.009–1.079)	0.012	1.026 (0.992–1.061)	0.141
TNM stage	2.908 (2.195–3.851)	<0.001	2.431 (1.709–3.459)	<0.001
Fuhrman grade	3.283 (2.037–5.291)	<0.001	1.188 (0.577–2.446)	0.640
Necrosis	2.513 (1.204–5.247)	0.014	1.595 (0.722–3.523)	0.248
ECOG-PS	2.371 (1.191–4.721)	0.014	1.235 (0.610–2.500)	0.557
Hemoglobin	0.964 (0.950–0.987)	<0.001	0.995 (0.974–1.016)	0.633
Neutrophil count	1.190 (1.040–1.361)	0.011	1.311 (0.84–2.044)	0.233
Lymphocyte count	0.358 (0.199–0.644)	0.001	0.395 (0.145–1.072)	0.068
Neutrophil-lymphocyte ratio	1.145 (1.055–1.243)	0.001	0.818 (0.55–1.217)	0.322
Platelet count	1.007 (1.002–1.012)	0.006	1.003 (0.998–1.007)	0.253
Creatinine	1.019 (1.006–1.033)	0.004	1.019 (1.003–1.036)	0.024
Blood urea nitrogen	1.001 (0.956–1.048)	0.954	NA	

NA, not available.

Hence it is essential to find an accurate prediction approach to improve prognosis and treatment in patient with ccRCC. In this study, we developed a radiomics nomogram by combining clinical factors and Rad-score for the prediction of OS in patients with ccRCC.

A number of previous studies have focused on identifying renal tumors and predicting nuclear grade by radiomics analysis, obtaining desirable results (20–22). Yet, so far, only a few studies have reported on prognostic prediction of RCC. Goh et al. assessed the texture parameters of 87 metastases in 39 RCC patients who received tyrosine kinase inhibitors (TKI) treatment and found that uniformity of texture was an independent predictive factor of time to progression (23). Another study showed that texture features consisting of the SD of pixel distribution histogram, entropy, and mean of positive pixels may be used to the prediction of OS for the patients with large RCCs (>7 cm; mean size, 9.9 cm) (24). These two studies indicated that texture analysis has the potential to predict the prognosis of RCC. More recently, Zeng et al. used integrative radiogenomics analysis (by analyzing contrast-enhanced CT images) for predicting molecular features and survival in ccRCC and found that these features could predict molecular subtypes, mutations, and prognosis of ccRCC patients (25). Moreover, Huang et al. suggested a radiomics model for predicting OS (5-year AUC = 0.775) in patients with the ccRCC model (26).

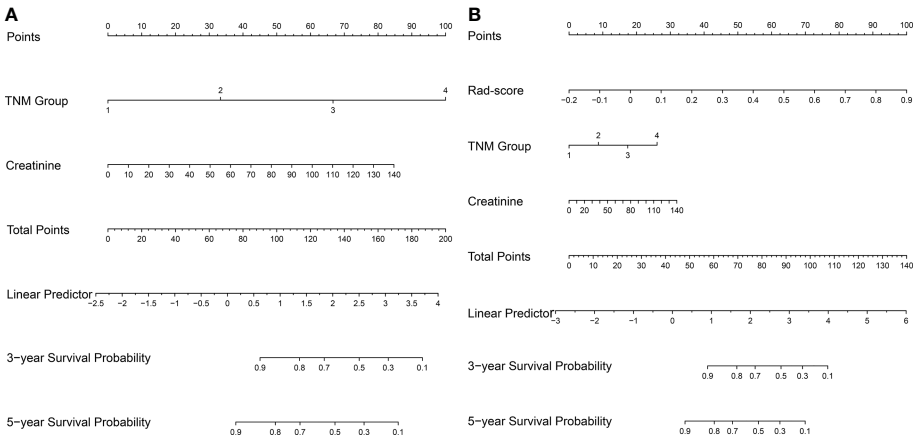


FIGURE 5 | The nomogram for survival estimation. The clinical nomogram **(A)**, combing TNM group and creatinine. The radiomics nomogram **(B)**, combing Rad-score, TNM group, and creatinine.

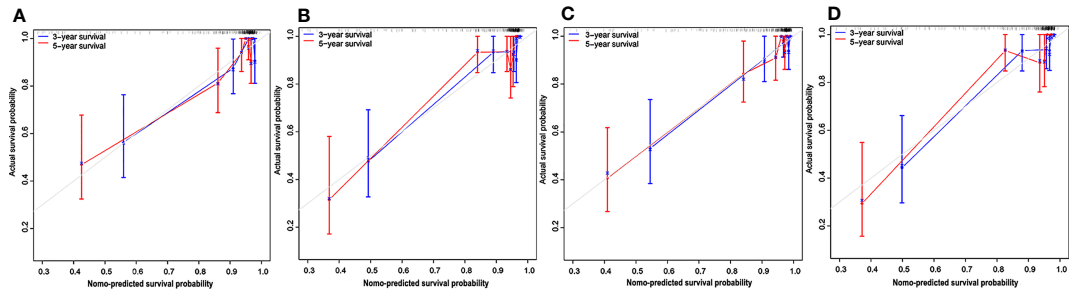


FIGURE 6 | Calibration curves for the nomograms. Calibration curves for the clinical nomogram in the training **(A)** and test **(B)** cohorts. The y-axis indicates the actual probability of survival; the x-axis indicates the predicted probability of survival. The 45-degree gray line represents the ideal prediction; blue and red lines represent the performance of the clinical nomogram to predict 3- and 5-year survival, respectively. Calibration curves for the radiomics nomogram in the training **(C)** and test **(D)** cohorts. Nomo-predicted, nomogram-predicted.

TABLE 3 | Performance of the nomograms.

	Training cohort		Test cohort	
	C-index (95% CI)	**AIC**	**C-index (95% CI)**	**AIC**
Clinical nomogram	0.803 (0.705–0.899)	263.26	0.846 (0.777–0.915)	233.65
Radiomics nomogram	0.884 (0.808–0.940)	243.35	0.859 (0.800–0.921)	234.25

In this study, we discovered that the combination of radiomics and clinical data showed a higher predictive value than the clinical nomogram alone, thus suggesting it might be used to predict OS in patients with ccRCC. This study took valuable clinical factors into account. The TNM staging system is an internationally accepted system used to determine the disease stage, including RCC (27). Tumor grade is one of the strongest elements incorporated into prognostic models for patients with ccRCC (4). Histologic necrosis and ECOG-PS have shown to be independent risk factors for ccRCC patients in the SSIGN and UISS scoring systems, respectively (28). The value of laboratory examination and NLR was evaluated in the prognosis of malignant tumors such as kidney cancer (29, 30). In this study, TNM group and CREA were selected as the independent prognostic factors for survival. As for prognostic scoring systems, the MSKCC system was developed to define risk groups of patients by coalescing independent factors for survival prediction of metastatic RCC, while the UISS is an evidence-based system for predicting recurrence or metastases after surgical treatment in patients with localized or locally advanced RCC (31). The radiomics nomogram developed in this study was applied to both metastatic and locally ccRCC.

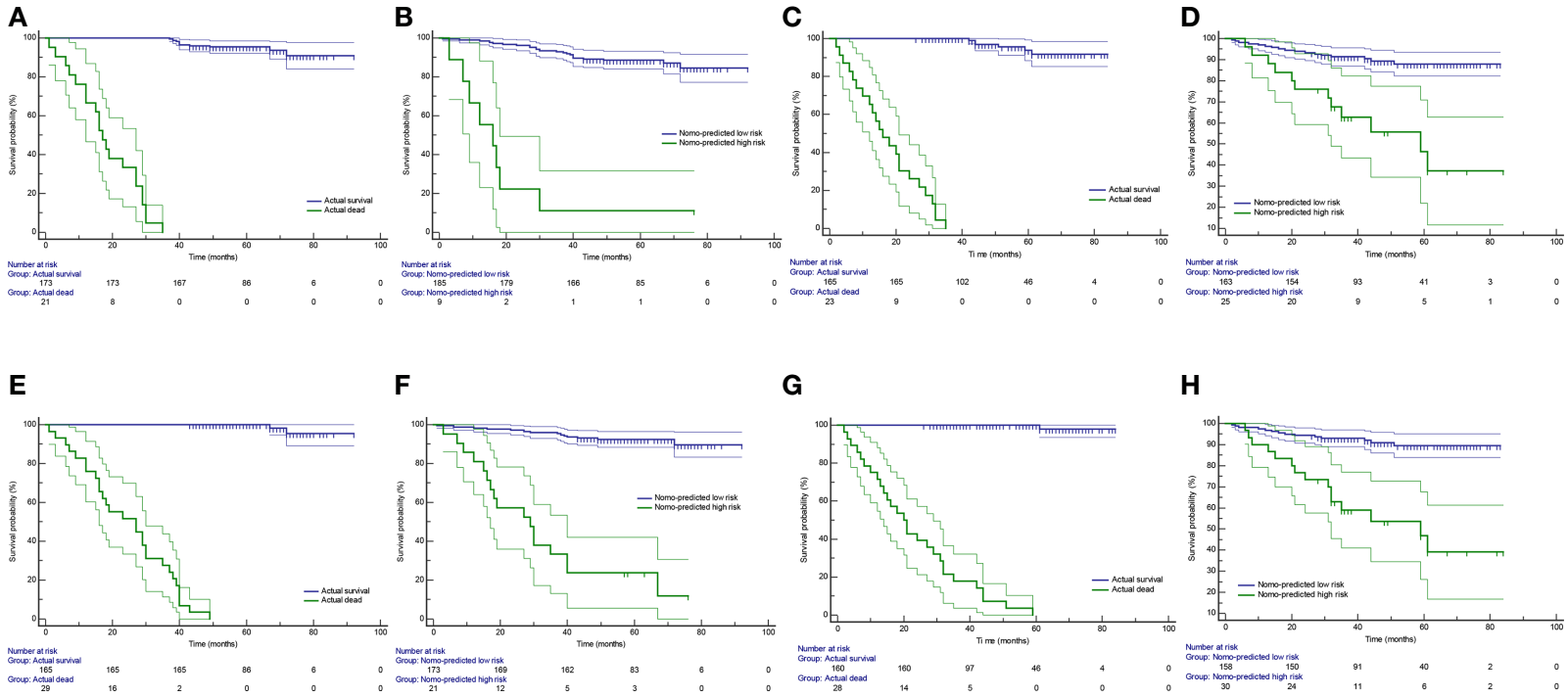


FIGURE 7 | Kaplan-Meier survival curves in group stratification. Based on survival probability predicted by the radiomics nomogram, patients were stratified into the Nomo-predicted high-risk group (survival probability < 0.5) and the Nomo-predicted low-risk group (survival probability > 0.5). The Kaplan-Meier survival curves were depicted by using the actual survival status **(A)** and the Nomo-predicted risk group **(B)** at 3 years as prediction factor in the training cohort. **(C, D)** are corresponding figures in the test cohort. The Kaplan-Meier survival curves were depicted by using the actual survival status **(E)** and the Nomo-predicted risk group **(F)** at 5 years as prediction factor in the training cohort. **(G, H)** are corresponding figures in the test cohort. Areas between thin lines of the same color represent 95% confidence intervals.

The evaluation of markers should depend on their ability to improve an already superior prediction model instead of on their P-value in multivariable analysis (32). Our results showed the radiomics nomogram performed better than the clinical nomogram, thus suggesting the incremental value of radiomics to OS prediction, and demonstrating that this new model is a useful method for outcome prognostication and treatment planning. In 2018, Meng et al. suggested that radiomics model combined with a clinicopathologic Cox model has a prognostic value for locally advanced rectal cancer (33). Another study extracted maximum and mean standard uptake values (SUVmax and SUVmean), total lesion glycolysis (TLG), metabolic tumoral volume (MTV), and texture features into Cox regression analysis in order to obtain prognostic model for identifying patients with more aggressive treatment (34). The importance of establishing comprehensive models was also reflected in the prognostic analysis of brain tumors, head-and-neck cancer, lung cancer, breast cancer, prostate cancer, liver cancer, and gastric cancer (12).

Ideally, an independent external validation dataset should be collected to test the results as with any biomarker analysis (35). Furthermore, developed models subsequently validated on an entirely new validation dataset from different centers can bolster its generalizability. Nonetheless, up to now, most of the external test dataset and the training dataset enrolled in studies were derived from the same center. Fortunately, an independent external test cohort from another hospital was assessed in this study to interpret the generalizability of the reported findings and correctly estimate the empirical error. As for the disparity of CT scanners arising from two institutes, image resampling, and gray-level discretization were implemented to standardize three-phase CT images, minimizing the impact of different scanning machines.

This study has a few limitations. First, this was a retrospective study, and the sample size was relatively small. Second, the clinical efficacy of our nomograms needs to be validated with the multicenter data. Last, patients with ccRCC were not classified into localized ccRCC, locally advanced ccRCC, and metastatic ccRCC groups according to the guidelines for stratified analysis. In this study, we did not compare the predictive ability between the radiomics nomogram and prognostic scoring system such as SSIGN, MSKCC. Thus, a large-scale prospective multicenter investigation is needed to further verify reported findings.

CONCLUSIONS

In conclusion, we developed and validated a non-invasive predictive method for predicting the survival of ccRCC and identified radiomics as a useful biomarker for prognostic prediction. The radiomics analysis may facilitate quantitative and personalized treatment for ccRCC patients, although it still needs to be further validated before being widely applied in clinical practice.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

Conception and design: LY and GY. Collection and assembly of data: LY, WM, YW, YZ, AG and NW. Development of methodology: LY, GY, JC, PN, and ZW. Data analysis and interpretation: GY, JC, WM, LY, and NG. Manuscript writing: LY, GY, and PN. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the National Natural Science Foundation of China (81701688 and 81601527), the Natural Science Foundation of Shandong Province (ZR2017BH096 and ZR2017MH036), the Key Research and Development Project of Shandong Province (2018GSF118078), and the Postdoctoral Science Foundation of China (2018M642617).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.671420/full#supplementary-material

REFERENCES

1. Motzer RJ, Haas NB, Donskov F, Gross-Goupil M, Varlamov S, Kopyltsov E, et al. Randomized Phase Iii Trial of Adjuvant Pazopanib Versus Placebo After Nephrectomy in Patients With Localized or Locally Advanced Renal Cell Carcinoma. *J Clin Oncol* (2017) 35(35):3916–23. doi: 10.1200/jco.2017.73.5324
2. Fitzmaurice C, Dicker D, Pain A, Hamavid H, Moradi-Lakeh M, MacIntyre MF, et al. The Global Burden of Cancer 2013. *JAMA Oncol* (2015) 1(4):505–27. doi: 10.1001/jamaoncol.2015.0735
3. Coy H, Young JR, Pantuck AJ, Douek ML, Sisk A, Magyar C, et al. Association of Tumor Grade, Enhancement on Multiphasic CT and Microvessel Density in Patients With Clear Cell Renal Cell Carcinoma. *Abdom Radiol (NY)* (2019). doi: 10.1007/s00261-019-02271-1
4. Sun M, Shariat SF, Cheng C, Ficarra V, Murai M, Oudard S, et al. Prognostic Factors and Predictive Models in Renal Cell Carcinoma: A Contemporary Review. *Eur Urol* (2011) 60(4):644–61. doi: 10.1016/j.eururo.2011.06.041
5. Correa AF, Jegede O, Haas NB, Flaherty KT, Pins MR, Messing EM, et al. Predicting Renal Cancer Recurrence: Defining Limitations of Existing Prognostic Models With Prospective Trial-Based Validation. *J Clin Oncol* (2019) 37(23):2062–71. doi: 10.1200/jco.19.00107
6. Guo C, Zhao H, Wang Y, Bai S, Yang Z, Wei F, et al. Prognostic Value of the Neo-Immunoscore in Renal Cell Carcinoma. *Front Oncol* (2019) 9:439. doi: 10.3389/fonc.2019.00439

7. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, 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 Cancer J Clin* (2017) 67(2):93–9. doi: 10.3322/caac.21388
8. Frank I, Blute ML, Cheville JC, Lohse CM, Weaver AL, Zincke H. An Outcome Prediction Model for Patients With Clear Cell Renal Cell Carcinoma Treated With Radical Nephrectomy Based on Tumor Stage, Size, Grade and Necrosis: The SSIGN Score. *J Urol* (2002) 168(6):2395–400. doi: 10.1097/01.ju.0000035885.91935.d5
9. Lam JS, Shvarts O, Leppert JT, Pantuck AJ, Figlin RA, Belldegrun AS. Postoperative Surveillance Protocol for Patients With Localized and Locally Advanced Renal Cell Carcinoma Based on a Validated Prognostic Nomogram and Risk Group Stratification System. *J Urol* (2005) 174(2):466–472; discussion 472; quiz 801. doi: 10.1097/01.ju.0000165572.38887.da
10. Motzer RJ, Bacik J, Murphy BA, Russo P, Mazumdar M. Interferon-Alfa as a Comparative Treatment for Clinical Trials of New Therapies Against Advanced Renal Cell Carcinoma. *J Clin Oncol* (2002) 20(1):289–96. doi: 10.1200/jco.2002.20.1.289
11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
12. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. *Theranostics* (2019) 9(5):1303–22. doi: 10.7150/thno.30309
13. Liang C, Huang Y, He L, Chen X, Ma Z, Dong D, et al. The Development and Validation of a CT-based Radiomics Signature for the Preoperative Discrimination of Stage I-II and Stage III-IV Colorectal Cancer. *Oncotarget* (2016) 7(21):31401–12. doi: 10.18632/oncotarget.8919
14. Li H, Zhu Y, Burnside ES, Huang E, Drukker K, Hoadley KA, et al. Quantitative MRI Radiomics in the Prediction of Molecular Classifications of Breast Cancer Subtypes in the TCGA/TCIA Data Set. *NPJ Breast Cancer* (2016) 2. doi: 10.1038/npjbcancer.2016.12
15. Vallieres M, Freeman CR, Skamene SR, El Naqa I. A Radiomics Model From Joint FDG-PET and MRI Texture Features for the Prediction of Lung Metastases in Soft-Tissue Sarcomas of the Extremities. *Phys Med Biol* (2015) 60(14):5471–96. doi: 10.1088/0031-9155/60/14/5471
16. Moldovanu CG, Boca B, Lebovici A, Tamas-Szora A, Feier DS, Crisan N, et al. Preoperative Predicting the WHO/ISUP Nuclear Grade of Clear Cell Renal Cell Carcinoma by Computed Tomography-Based Radiomics Features. *J Pers Med* (2020) 11(1). doi: 10.3390/jpm11010008
17. Uhlig J, Leha A, Delonge LM, Haack AM, Shuch B, Kim HS, et al. Radiomic Features and Machine Learning for the Discrimination of Renal Tumor Histological Subtypes: A Pragmatic Study Using Clinical-Routine Computed Tomography. *Cancers (Basel)* (2020) 12(10). doi: 10.3390/cancers12103010
18. Shuch B, Amin A, Armstrong AJ, Eble JN, Ficarra V, Lopez-Beltran A, et al. Understanding Pathologic Variants of Renal Cell Carcinoma: Distilling Therapeutic Opportunities From Biologic Complexity. *Eur Urol* (2015) 67(1):85–97. doi: 10.1016/j.eururo.2014.04.029
19. Rini BI, Campbell SC, Escudier B. Renal Cell Carcinoma. *Lancet* (2009) 373(9669):1119–32. doi: 10.1016/s0140-6736(09)60229-4
20. Nie P, Yang G, Wang Z, Yan L, Miao W, Hao D, et al. A CT-based Radiomics Nomogram for Differentiation of Renal Angiomyolipoma Without Visible Fat From Homogeneous Clear Cell Renal Cell Carcinoma. *Eur Radiol* (2019). doi: 10.1007/s00330-019-06427-x
21. Varghese BA, Chen F, Hwang DH, Cen SY, Desai B, Gill IS, et al. Differentiation of Predominantly Solid Enhancing Lipid-Poor Renal Cell Masses by Use of Contrast-Enhanced Ct: Evaluating the Role of Texture in Tumor Subtyping. *AJR Am J Roentgenol* (2018) 211(6):W288–w296. doi: 10.2214/ajr.18.19551
22. Feng Z, Shen Q, Li Y, Hu Z. CT Texture Analysis: A Potential Tool for Predicting the Fuhrman Grade of Clear-Cell Renal Carcinoma. *Cancer Imaging* (2019) 19(1):6. doi: 10.1186/s40644-019-0195-7
23. Goh V, Ganeshan B, Nathan P, Juttla JK, Vinayan A, Miles KA. Assessment of Response to Tyrosine Kinase Inhibitors in Metastatic Renal Cell Cancer: CT Texture as a Predictive Biomarker. *Radiology* (2011) 261(1):165–71. doi: 10.1148/radiol.11110264
24. Lubner MG, Stabo N, Abel EJ, Del Rio AM, Pickhardt PJ. Ct Textural Analysis of Large Primary Renal Cell Carcinomas: Pretreatment Tumor Heterogeneity Correlates With Histologic Findings and Clinical Outcomes. *AJR Am J Roentgenol* (2016) 207(1):96–105. doi: 10.2214/ajr.15.15451
25. Zeng H, Chen L, Wang M, Luo Y, Huang Y, Ma X. Integrative Radiogenomics Analysis for Predicting Molecular Features and Survival in Clear Cell Renal Cell Carcinoma. *Aging (Albany NY)* (2021) 13(7):9960–75. doi: 10.18632/aging.202752
26. Huang Y, Zeng H, Chen L, Luo Y, Ma X, Zhao Y. Exploration of an Integrative Prognostic Model of Radiogenomics Features With Underlying Gene Expression Patterns in Clear Cell Renal Cell Carcinoma. *Front Oncol* (2021) 11:640881. doi: 10.3389/fonc.2021.640881
27. Kattan MW, Hess KR, Amin MB, Lu Y, Moons KG, Gershenwald JE, et al. American Joint Committee on Cancer Acceptance Criteria for Inclusion of Risk Models for Individualized Prognosis in the Practice of Precision Medicine. *CA Cancer J Clin* (2016) 66(5):370–4. doi: 10.3322/caac.21339
28. Speed JM, Trinh QD, Choueiri TK, Sun M. Recurrence in Localized Renal Cell Carcinoma: A Systematic Review of Contemporary Data. *Curr Urol Rep* (2017) 18(2):15. doi: 10.1007/s11934-017-0661-3
29. Motzer RJ, Jonasch E, Agarwal N, Bhayani S, Bro WP, Chang SS, et al. Kidney Cancer, Version 2.2017, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* (2017) 15(6):804–34. doi: 10.6004/jnccn.2017.0100
30. Viers BR, Thompson RH, Lohse CM, Cheville JC, Leibovich BC, Boorjian SA, et al. Pre-Treatment Neutrophil-to-Lymphocyte Ratio Predicts Tumor Pathology in Newly Diagnosed Renal Tumors. *World J Urol* (2016) 34(12):1693–9. doi: 10.1007/s00345-016-1821-7
31. Senbabaoglu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, et al. Tumor Immune Microenvironment Characterization in Clear Cell Renal Cell Carcinoma Identifies Prognostic and Immunotherapeutically Relevant Messenger RNA Signatures. *Genome Biol* (2016) 17(1):231. doi: 10.1186/s13059-016-1092-z
32. Kattan MW. Judging New Markers by Their Ability to Improve Predictive Accuracy. *J Natl Cancer Inst* (2003) 95(9):634–5. doi: 10.1093/jnci/95.9.634
33. Meng Y, Zhang Y, Dong D, Li C, Liang X, Zhang C, et al. Novel Radiomic Signature as a Prognostic Biomarker for Locally Advanced Rectal Cancer. *J Magn Reson Imaging* (2018). doi: 10.1002/jmri.25968
34. Lovinfosse P, Polus M, Van Daele D, Martinive P, Daenen F, Hatt M, et al. Fdg PET/CT Radiomics for Predicting the Outcome of Locally Advanced Rectal Cancer. *Eur J Nucl Med Mol Imaging* (2018) 45(3):365–75. doi: 10.1007/s00259-017-3855-5
35. Dalal V, Carmicheal J, Dhaliwal A, Jain M, Kaur S, Batra SK. Radiomics in Stratification of Pancreatic Cystic Lesions: Machine Learning in Action. *Cancer Lett* (2019). doi: 10.1016/j.canlet.2019.10.023

Conflict of Interest: Authors JC and NG were employed by company Huiying Medical Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer QQ declared a shared affiliation with one of the authors, NW, to the handling editor at time of review.

Copyright © 2021 Yan, Yang, Cui, Miao, Wang, Zhao, Wang, Gong, Guo, Nie and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Bioengineering and Biotechnology

ORIGINAL RESEARCH
published: 20 July 2021
doi: 10.3389/fbioe.2021.695305

Prediction of Inflammatory Breast Cancer Survival Outcomes Using Computed Tomography-Based Texture Analysis

Sung Eun Song[1], Bo Kyoung Seo[2], Kyu Ran Cho[1], Ok Hee Woo[3], Balaji Ganeshan[4], Eun Sil Kim[2] and Jaehyung Cha[5]*

[1] Department of Radiology, Korea University Anam Hospital, Korea University College of Medicine, Seoul, South Korea, [2] Department of Radiology, Korea University Ansan Hospital, Korea University College of Medicine, Ansan, South Korea, [3] Department of Radiology, Korea University Guro Hospital, Korea University College of Medicine, Seoul, South Korea, [4] Institute of Nuclear Medicine, University College London Hospitals NHS Trust, London, United Kingdom, [5] Medical Science Research Center, Korea University Ansan Hospital, Korea University College of Medicine, Ansan, South Korea

Background: Although inflammatory breast cancer (IBC) has poor overall survival (OS), there is little information about using imaging features for predicting the prognosis. Computed tomography (CT)-based texture analysis, a non-invasive technique to quantify tumor heterogeneity, could be a potentially useful imaging biomarker. The aim of the article was to investigate the usefulness of chest CT-based texture analysis to predict OS in IBC patients.

Methods: Of the 3,130 patients with primary breast cancers between 2006 and 2016, 104 patients (3.3%) with IBC were identified. Among them, 98 patients who underwent pre-treatment contrast-enhanced chest CT scans, got treatment in our institution, and had a follow-up period of more than 2 years were finally included for CT-based texture analysis. Texture analysis was performed on CT images of 98 patients, using commercially available software by two breast radiologists. Histogram-based textural features, such as quantification of variation in CT attenuation (mean, standard deviation, mean of positive pixels [MPP], entropy, skewness, and kurtosis), were recorded. To dichotomize textural features for survival analysis, receiver operating characteristic curve analysis was used to determine cutoff points. Clinicopathologic variables, such as age, node stage, metastasis stage at the time of diagnosis, hormonal receptor positivity, human epidermal growth factor receptor 2 positivity, and molecular subtype, were assessed. A Cox proportional hazards model was used to determine the association of textural features and clinicopathologic variables with OS.

Results: During a mean follow-up period of 47.9 months, 41 of 98 patients (41.8%) died, with a median OS of 20.0 months. The textural features of lower mean attenuation, standard deviation, MPP, and entropy on CT images were significantly associated with worse OS, as was the M1 stage among clinicopathologic variables (all P-values < 0.05). In multivariate analysis, lower mean attenuation (hazard ratio [HR], 3.26; $P = 0.003$), lower MPP (HR, 3.03; $P = 0.002$), and lower entropy (HR, 2.70; $P = 0.009$) on chest CT images were significant factors independent from the M1 stage for predicting worse OS.

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln, United Kingdom

Reviewed by:
Liu Yiqiang,
Beijing Cancer Hospital, China
Chaoyue Chen,
Sichuan University, China

***Correspondence:**
Bo Kyoung Seo
seoboky@korea.ac.kr;
seoboky@gmail.com

Specialty section:
This article was submitted to Biomaterials, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *14 April 2021*
Accepted: *18 June 2021*
Published: *20 July 2021*

Citation:
Song SE, Seo BK, Cho KR, Woo OH, Ganeshan B, Kim ES and Cha J (2021) Prediction of Inflammatory Breast Cancer Survival Outcomes Using Computed Tomography-Based Texture Analysis. Front. Bioeng. Biotechnol. 9:695305. doi: 10.3389/fbioe.2021.695305

Conclusions: Lower mean attenuation, MPP, and entropy on chest CT images predicted worse OS in patients with IBC, suggesting that CT-based texture analysis provides additional predictors for OS.

Keywords: overall survival, breast neopalsms, computed tomgraphy, texture, histogram

INTRODUCTION

Inflammatory breast cancer (IBC) is an aggressive malignancy and accounts for 1–6% of all breast cancers (Anderson et al., 2005). Despite the lack of molecular markers that distinguish inflammatory and non-inflammatory breast cancer at the molecular level, the two clinical entities are clearly distinct in terms of symptoms, natural history, and survival (Fouad et al., 2017). It is characterized by diffuse edema and erythema, involving one-third or more of the skin of the breast that results from tumor infiltration of the dermal lymphatics. IBC is classified as "cT4d" by the American Joint Committee on Cancer (AJCC) (Giuliano et al., 2018). Thus, patients with non-metastatic IBC are considered stage IIIB, regardless of tumor size or lymph node spread, and the patients with metastatic IBC are considered stage IV. Fouad et al. (2015) reported that the patients with IBC who present with distant metastasis at diagnosis (stage IV) have significantly worse overall survival compared with those who have stage IV non-inflammatory breast cancer. With the introduction of trimodal treatment, including neoadjuvant chemotherapy, mastectomy, and radiation therapy, the prognosis of IBC has improved with a 2-year survival rate of 62% between 1990 and 1995 to 76% between 2006 and 2010 (Ueno et al., 1997; Low et al., 2004; Dawood et al., 2014). However, even after trimodal treatment, IBC still has a very poor outcome with a 5-year overall survival (OS) rate of 25–50% (Hance et al., 2005; Cristofanilli et al., 2007). Therefore, it is important to identify early prognostic factors that predict poor survival outcomes and to determine aggressive treatment in a carefully selected group of IBC patients.

Several studies demonstrated that diffuse erythema, lymph node involvement, chest-wall adherence, hormone receptor-negative status, age >50 years, and incomplete tumor regression after chemotherapy are associated with worse survival outcomes in IBC (Chevallier et al., 1987; Fields et al., 1989; Palangie et al., 1994; Pierga et al., 2017). Therefore, breast imaging is crucial in determining the extent of tissue involvement and predicting the prognosis of IBC. However, few studies have used imaging features to predict the prognosis in IBC patients. To assess AJCC cancer staging and treatment effectiveness according to the response evaluation criteria in solid tumors (RECIST) (Eisenhauer et al., 2009), morphological methods, such as tumor size measurements on imaging, are essential. However, IBC regards as a non-measurable tumor at the RECIST guideline, and tumor size measurement is inappropriate for evaluation of the treatment response. Therefore, there is an unmet need for treatment monitoring and risk stratification of patients with IBC.

Texture analysis, which is a non-invasive technique that quantifies the spatial pattern of pixel intensities on imaging, is an objective measure of tumor heterogeneity (Ganeshan et al., 2013). Tumor heterogeneity has been recognized as a key feature of therapeutic resistance that reflects intra-tumor variation in cell density, necrosis, or angiogenesis (Nelson et al., 2004). Computed tomography (CT)-based texture analysis has been shown to provide predictive information about survival or recurrence or response to chemotherapy in patients with lung, liver, head and neck, and colorectal cancer (Ganeshan et al., 2012; Ng et al., 2013; Zhang et al., 2013; Chee et al., 2017; Mule et al., 2018). However, there is very little published literature on CT-based texture analysis in breast cancer despite several studies have demonstrated that textural features from magnetic resonance imaging (MRI) are associated with survival outcomes (Pickles et al., 2016; Kim et al., 2017) and response to neoadjuvant chemotherapy (Chamming's et al., 2018). Because IBC is considered at least stage IIIB at diagnosis, systemic staging studies, such as chest CT or positron emission tomography, combined with CT scan, are done to assess whether the cancer has spread to lymph nodes or lungs. It was reported that 95% of patients with stages III and IV breast cancer underwent chest CT scan for staging (James et al., 2019). That study shows that chest CT is one of the investigations of choice for staging and metastatic work-up of new breast cancers, and suggests routine performance of chest CT in advanced breast cancers (stages III and IV) because metastases are higher compared with early breast cancers (stages I and II). If simpler measures (e.g., histogram analysis) of CT-derived prognostic biomarkers could be readily incorporated in routine clinical practice, they could be potentially useful non-invasive imaging biomarkers, which could help in making optimal treatment decisions and improve overall clinical outcomes.

The purpose of this study was to investigate the usefulness of CT-based texture analysis and clinicopathologic features to predict OS in patients with IBC.

METHODS

Patients

This retrospective study was approved by our institutional review board (IRB Number: 2020AS0113), and the need for written informed patient consent was waived. After a review of medical records between January 2006 and December 2016, we identified 3,130 consecutive patients with breast cancer who were first diagnosed with tissue biopsies in our institution. Among those, 104 patients (3.3%) had IBC in AJCC breast cancer staging (Giuliano et al., 2018). The inclusion criteria were as follows: (i) pretreatment contrast-enhanced chest CT scan in our institution; (ii) combined modality therapy in our institution, consisting of neoadjuvant chemotherapy, radiotherapy with or without

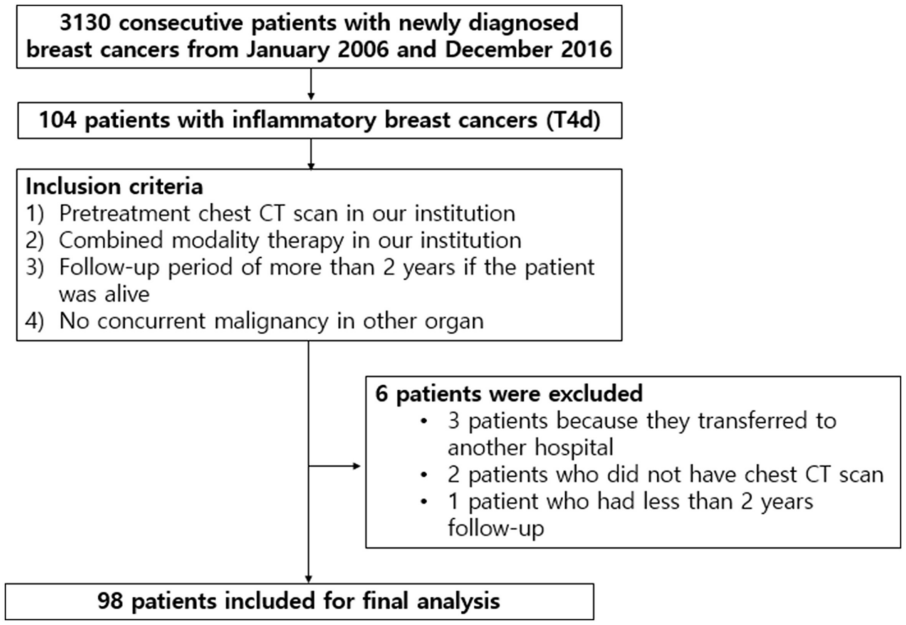


FIGURE 1 | A flow diagram of study population.

surgery, and/or adjuvant chemotherapy; (iii) a follow-up period of more than 2 years if the patient was alive; and (iv) no concurrent malignancy in other organs. Of the 104 patients, six were excluded: three patients because they were transferred to another hospital, two patients who did not have chest CT images, and one patient had <2 years of follow-up. Final study population, consisting of 98 patients, was presented in **Figure 1**. The interval between pathological diagnosis of breast cancer and chest CT scans ranged from 3 to 30 days (median: 7 days).

CT Acquisition

Various CT scanners were used, including Brilliance 64 or Ingenuity 128 or Icon Spectral CT (Philips Medical Systems, Amsterdam, Netherlands), or Somatom Volume Zoom or Somatom Definition Flash (Siemens, Erlangen, Germany). Our chest CT was performed, using tube voltage of 80–120 kVp and tube current ranging from 30 to 300 mAs. The acquired CT images were reconstructed into a lung and a mediastinal window. The reconstruction parameters comprised of a 3-mm-slice thickness and no reconstruction interval. For the post-contrast scan, intravenous contrast medium (120–150 ml of 300–370 mg/ml non-ionic contrast) was injected at a rate of 2–3 ml/s, followed by a 40-m saline bolus. Scanning was commenced 90 s after contrast medium injection.

CT Texture Analysis

Histogram-based texture analysis was done on the axial post-contrast DICOM images with a mediastinal window setting, using commercially available research software (TexRAD software—Feedback Medical Ltd., Cambridge, UK). Two radiologists (S.E.S. and B.K.S., 10 and 22 years of experience in breast imaging), who were blinded to the clinicopathologic findings and survival outcomes, reviewed the CT images and drew a region of interest (ROI) two times on the CT image surrounding the entire enhancing tumor margin at the largest cross-sectional area of the tumor (**Figures 2**, **3**). For diffuse infiltrative tumors, the largest enhancing lesion was selected and measured. The initial step of texture analysis constitutes an image filtration-histogram step where the Laplacian of Gaussian band-pass filter was used to selectively extract textural features of different sizes and intensity variation. For tuning the filter parameter, spatial-scale image filtration (SSF), ranging from 2.0 to 6.0, was used. Each SSF corresponded to the same number of millimeters of pixel scales in radius; fine (SSF 2.0: 2 mm), medium (SSF 3.0–5.0: 3–5 mm), and coarse (SSF 6.0: 6 mm). After an image filtration-histogram step, a texture quantification step, which quantified heterogeneity within ROI with and without image filtration, using SSF, was performed for calculating six textural features (mean attenuation, standard deviation [SD], mean of positive pixels [MPP], entropy, kurtosis, and skewness) of the pixel distribution histogram. Mean attenuation is the average attenuation value (Hounsfield units, HU). SD is a measure of the dispersion from the average value of the histogram. MPP represents the average attenuation value of pixels >0, reflecting the average brightness of only positive pixel values. Entropy is a measure of irregularity or complexity. Kurtosis indicates the pointiness/sharpness of the histogram pixel distribution. Skewness reflects the asymmetry of the histogram pixel distribution. The values initially measured by one reader were used for statistical analysis. CT-based texture

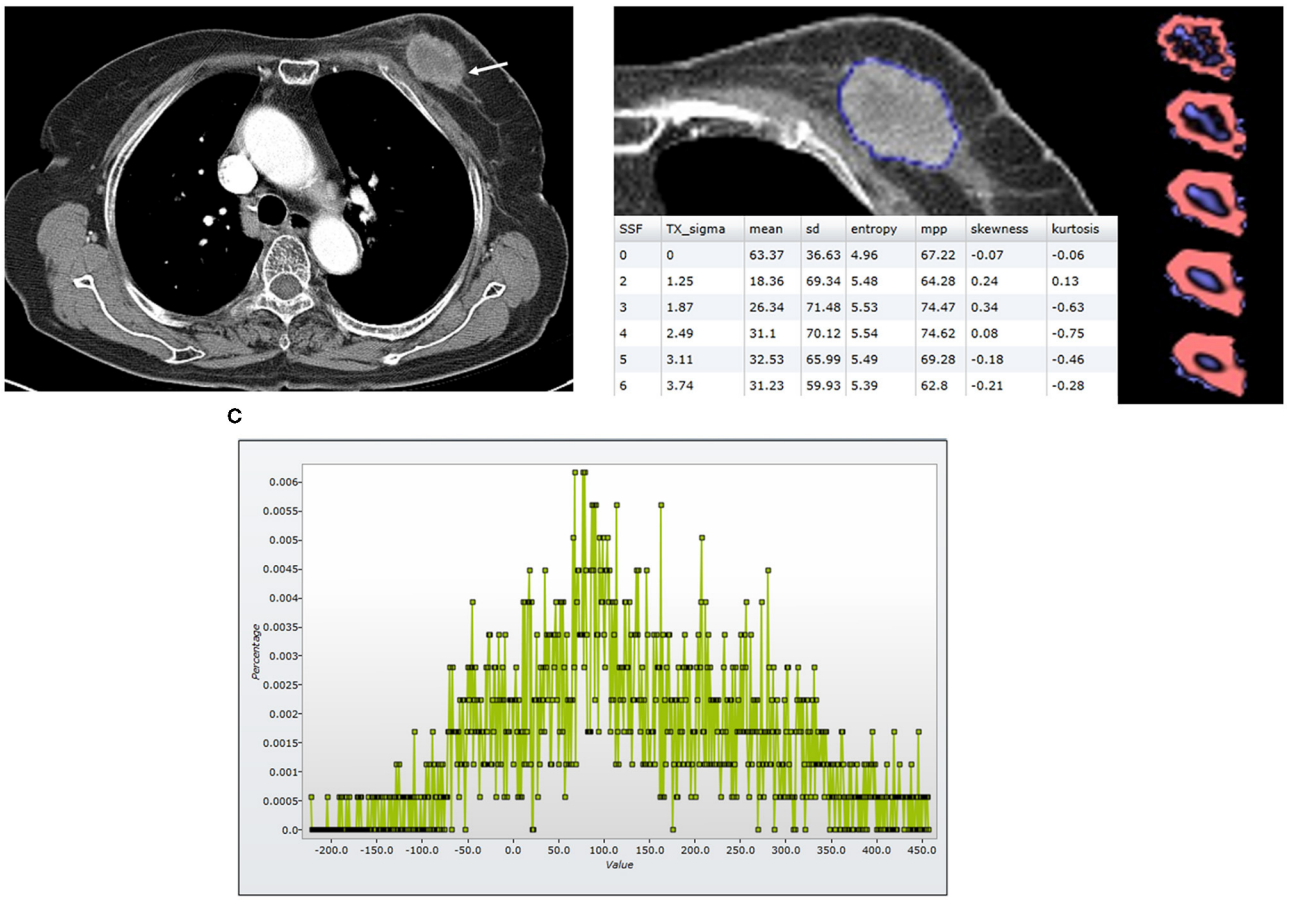


FIGURE 2 | Findings in a 74-year-old woman who died 23 months after the diagnosis of luminal-like breast cancer in the left breast. At the time of diagnosis, she did not have a distant metastasis. **(A)** An axial post-contrast CT image shows an irregular enhancing mass (arrow) with a central area of low attenuation, suggesting necrosis. **(B)** A texture analysis image shows a region of interest (blue line), texture maps, and texture parameters on the CT image. **(C)** A CT texture histogram is obtained; from which, the different statistical-based metrics are extracted. Mean, MPP, entropy, and SD were lower for the patient with poor a prognosis. MPP, mean of positive pixels; SD, standard deviation.

analysis has undergone the qualification and validation processes as evidenced from a number of publications in literature in an oncological setting (Van der Auwera et al., 2004; Ganeshan et al., 2012; Ng et al., 2013).

Clinicopathologic Data

We evaluated AJCC breast cancer staging and pathological biomarkers, using the medical records database. Cancer staging was assessed as tumor, lymph node (N), and the metastasis (M) stage from physical examination and imaging studies, such as breast MRI, chest CT scan, bone scan, and positron emission tomography combined with CT scan. Estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2) status were evaluated from tissue biopsies or surgical specimens. ER- and PR-positivities were determined, using a cutoff value of >1% positively stained nuclei. Silver-enhanced *in situ* hybridization (SISH) was performed to detect the presence of HER2 amplification (gene copy number >6 or HER2/chromosome 17 ratio >2.2) if tumors had HER2 scores of 2+. On the basis of the results of ER, PR, and HER2 analysis, the tumors were categorized into three molecular subtypes: luminal-like (ER/PR-positive and HER2-negative), HER2-like (ER/PR-negative and HER2-positive), and basal-like (ER/PR/HER2-negative) (Giuliano et al., 2018). Treatment information, such as neoadjuvant chemotherapy, adjuvant chemotherapy, and adjuvant radiation therapy, was also recorded.

Statistical Analysis

The primary end point was OS. Patients were followed up until death from disease progression or until March 2020 if they were alive. OS was defined as the time span from the date of diagnosis of IBC to the date of death or the last clinical follow-up.

All data were checked, using Levene's test for equality of variance. The textural features were compared between survivors and non-survivors, using non-parametric Mann–Whitney *U*-test. For survival analysis, significant textural features were

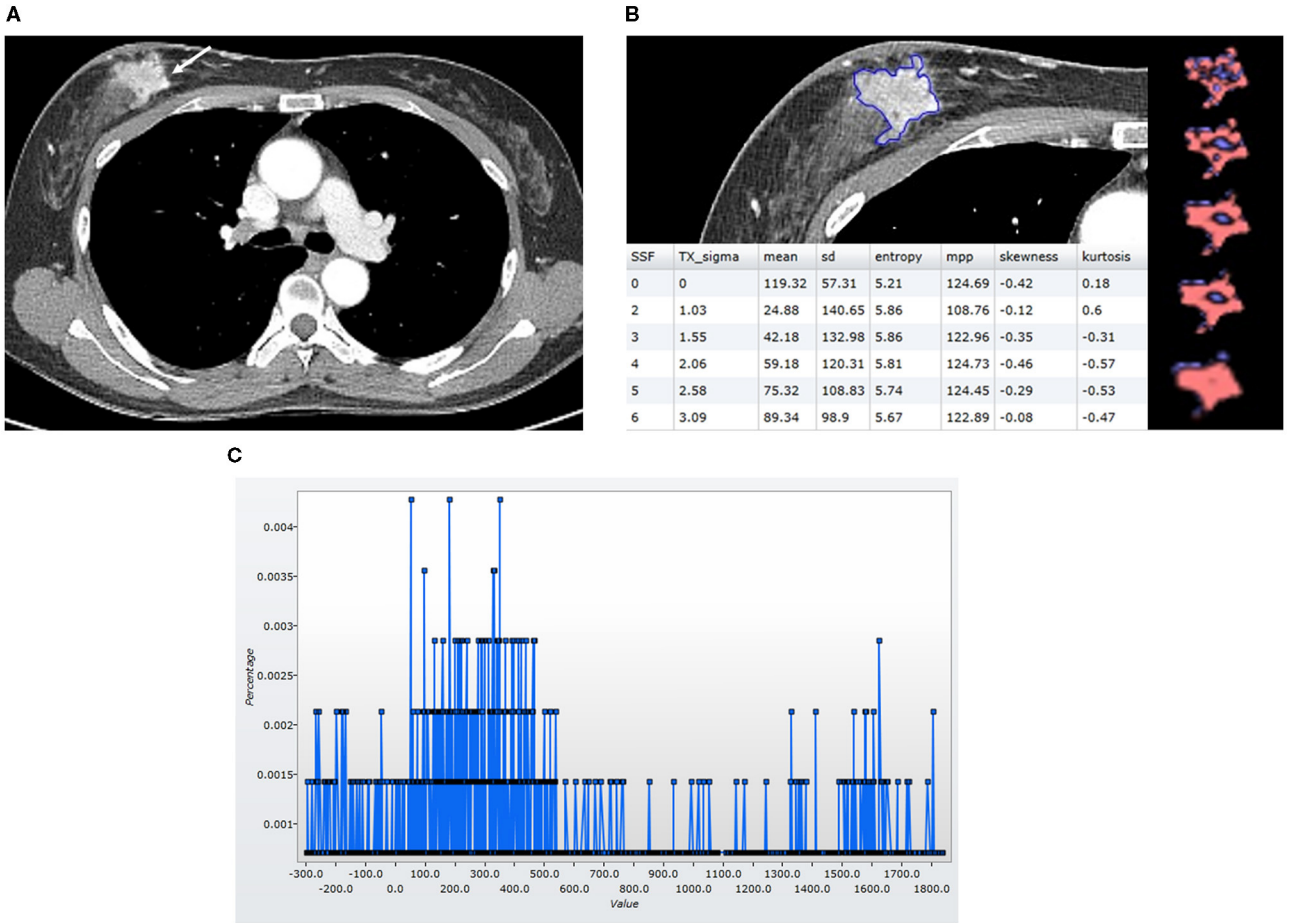


FIGURE 3 | Findings in a 49-year-old woman who was alive 65 months after the diagnosis of luminal-like breast cancer in the right breast. At the time of the diagnosis, she did not have a distant metastasis. **(A)** An axial post-contrast CT image shows an irregular mass (arrow) with strong enhancement. **(B)** A texture analysis image shows a region of interest (blue line), texture maps, and texture parameters on the CT image. **(C)** A CT texture histogram is obtained; from which, the different statistical-based metrics are extracted. Mean, MPP, entropy, and SD were higher for the patient with a good prognosis. MPP, mean of positive pixels; SD, standard deviation.

dichotomized based on the optimal cutoff values identified by receiver operating characteristic curve analysis. The patients with values for textural features below the cutoff values were categorized into one group, while the remaining patients were categorized into another group. Kaplan–Meier curves were generated to compare survival between the two groups. Differences between the curves were evaluated by a log-rank test.

A univariate Cox proportional hazards model was used to analyze the effects on OS of clinicopathologic variables (age, *N* stage, *M* stage, ER, PR, HER2, molecular subtype, neoadjuvant chemotherapy, adjuvant chemotherapy, and adjuvant radiation therapy) and textural parameters. Collinearity between textural features was analyzed, using the Spearman rank correlation test, and a multivariate Cox proportional hazards model assessed the independence of each significant univariate texture predictor from the most significant clinical predictor of OS. Interobserver reliability between two readers and intraobserver reliability were assessed, using the interclass correlation coefficient (ICC). An *r* value of 1. was considered a perfect agreement; 0.81–0.99, almost perfect agreement; 0.61–0.80, substantial agreement; 0.41–0.60, moderate agreement; 0.21–0.40, fair agreement; and ≤0.20, slight agreement. Statistical analyses were performed, using SPSS version 20.0 (IBM Corp., Armonk, NY, USA), with a P-value < 0.05 considered to be significant.

RESULTS

Patient Characteristics

Ninety-eight patients (mean age, 54 ± 12 years; range, 25–87 years) were evaluated. The most common clinical N stage was N3 (35.7%, 35 of 98), followed by N1 (34.7%, 34 of 98), N2 (25.5%, 25 of 98), and N0 (4.1%, 4 of 98). At the time of diagnosis, 38 patients (38.7%, 38 of 98) had M1 stage. Molecular subtypes were as follows: luminal-like (32.6%, 32 of 98), HER2-like (39.8%, 39 of 98), and basal-like (27.6%, 27 of 98). Of the 98 patients, 88 (89.7%) underwent mastectomy, but the remaining 10 (10.3%)

TABLE 1 | Univariate cox proportional hazard analysis of clinicopathologic variables associated with worse overall survival.

Variable	No. of patients	Overall survival	
		Hazard ratio	*P*-value
Age			
≥45 years	81	Reference	
<45 years	17	1.40 (0.64, 3.12)	0.392
***N* stage**			
0–1	38	Reference	
2–3	60	1.67 (0.86, 3.24)	0.125
***M* stage**			
0	60	Reference	
1	38	4.05 (2.14, 7.65)	<0.001
ER			
Positive	44	Reference	
Negative	54	1.12 (0.56, 1.96)	0.891
PR			
Positive	37	Reference	
Negative	61	1.06 (0.56, 2.00)	0.867
HER2			
Positive	39	Reference	
Negative	59	1.54 (0.80, 3.03)	0.196
Molecular subtype			
Luminal-like		Reference	
HER2-like		0.63 (0.30, 1.30)	0.215
Basal-like		0.94 (0.44, 1.99)	0.876
Neoadjuvant chemotherapy			
Not done	24	Reference	
Done	74	1.68 (0.74, 3.80)	0.211
Adjuvant chemotherapy			
Not done	5	Reference	
Done	93	0.91 (0.22, 3.78)	0.899
Adjuvant radiation therapy			
Not done	5	Reference	
Done	93	1.02 (0.51, 2.04)	0.273

Data in parentheses are 95% confidence intervals. ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2.

did not undergo surgery because of their poor general condition. During a mean follow-up period of 47.9 months (range, 3–157 months), 41 of 98 patients (41.8%) died with a median OS of 20.0 months (range, 3–62 months). Among the clinicopathologic variables, M1 was the only significant predictor of OS ($P < 0.001$). These data are summarized in **Table 1**.

CT Texture Analysis and Survival Outcomes

The CT textural features were compared between survivors ($n = 57$) and non-survivors ($n = 41$). When no image filtration (SSF0) was used, non-survivors showed lower mean attenuation, SD, MPP, and entropy on CT images (all $P < 0.05$) (**Table 2**). However, textural features obtained using an SSF of 2–6 showed no differences between survivors and non-survivors (**Supplementary Table 1**). Receiver operating characteristic curve analyses determined that the optimal cutoff values were 66.79 HU (area under the curve [AUC], 0.67; $P = 0.004$) for mean attenuation, 38.84 (AUC, 0.63; $P = 0.024$) for SD, 67.54 (AUC, 0.69; $P = 0.001$) for MPP, and 5.01 (AUC, 0.62; $P = 0.042$) for entropy on CT images (**Supplementary Figure 1**). Using these cutoffs, textural features were categorized into two groups. Kaplan–Meier curves were generated to compare OS between the two groups, and the log-rank test revealed that the textural features were significantly associated with OS outcomes (**Figure 4**). The Kaplan–Meier curve of *M* stage was shown in **Supplementary Figure 2**.

Survival Analysis Using Cox Proportional Hazard Models

Among the textural features, lower mean attenuation (hazard ratio [HR], 3.87; $P = 0.001$), lower SD (HR, 2.18; $P = 0.016$), lower MPP (HR, 3.31; $P = 0.001$), and lower entropy (HR, 2.78; $P = 0.005$) on CT images were associated with worse OS in univariate analysis (**Table 3**). Collinearity was observed between all the textural features (all, $P < 0.05$) (**Supplementary Table 2**). Because of this collinearity, each dichotomized CT textural feature was analyzed with M1 stage, using a multivariate Cox proportional hazards model. In multivariate analysis, lower mean attenuation (HR, 3.26; $P = 0.003$), lower MPP (HR, 3.03; $P = 0.002$), and lower entropy (HR, 2.70; $P = 0.009$) on post-contrast CT images were significant factors in predicting worse OS, independent of M1 stage (**Table 4**).

Interobserver and Intraobserver Agreement

For interobserver agreement, the ICC between two readers was calculated. The ICC was 0.77 for mean, 0.81 for SD, 0.82 for entropy, 0.80 for MPP, 0.28 for skewness, and 0.16 for kurtosis. For intraobserver agreement, the ICC for reader 1 was 0.98 for mean, 0.93 for SD, 0.98 for entropy, 0.99 for MPP, 0.32 for skewness, and 0.18 for kurtosis, and that for reader 2 was 0.95 for mean, 0.90 for SD, 0.91 for entropy, 0.95 for MPP, 0.30 for skewness, and 0.19 for kurtosis.

DISCUSSION

Our results suggest that CT-based texture analysis can assess tumor heterogeneity and provide prognostic information in patients with IBC. Histogram-based statistical parameters, such as lower mean attenuation, lower MPP, and lower entropy on post-contrast CT images, were significant predictors of worse OS in the patients with IBC, independent of M1 stage at the time of diagnosis.

Our study had strengths. First, our exploratory study is the first to report the value of simple CT-based histogram statistics to act as a prognostic indicator for IBC, which can potentially be readily incorporated in routine practice. Since we used commercial software for image analysis, the results can be applied to clinical practice and multicenter studies. Our results suggest that CT-based texture analysis can be used as an adjunct in risk stratification for patients with IBC,

TABLE 2 | Comparison of CT texture features between survivors and non-survivors.

Texture feature	All patients ($n = 98$)	Survivors ($n = 57$)	Non-survivors ($n = 41$)	*P*-value
Mean	60.92 ± 22.24	66.44 ± 23.14	53.25 ± 18.60	0.003
SD	40.91 ± 11.97	43.55 ± 13.48	37.24 ± 8.35	0.005
MPP	68.61 ± 19.65	74.33 ± 20.77	60.66 ± 14.87	<0.001
Entropy	4.96 ± 0.29	5.01 ± 0.32	4.89 ± 0.23	0.031
Skewness	−0.51 ± 0.81	−0.41 ± 0.70	−0.65 ± 0.93	0.157
Kurtosis	2.63 ± 6.95	1.88 ± 6.14	3.66 ± 7.99	0.232

Data are mean ± standard deviation. SSF, spatial scale filters; SD, standard deviation; MPP, mean of positive pixels.

which could help in making optimal treatment decisions and improve overall clinical outcomes. An interesting finding of our study is that when no image filtration was used, the texture features were different between survivors and non-survivors, while, when image filtration was used, there was no significant difference. Our results implicate that simple histogram statistics that do not require a dedicated post-processing platform can be used as prognostic markers in patients with breast cancer. This will increase the utility and applicability of CT-based texture analysis for patients with breast cancer. Second, our study provided further insight into the interpretation of texture results for the design of future CT studies in breast malignancies and non-measurable tumors according to the RECIST guideline (Eisenhauer et al., 2009). At RECIST, tumor lesions are categorized as measurable or non-measurable. Non-measurable lesions include IBC, leptomeningeal disease, ascites, pleural or pericardial effusion, lymphangitic involvement of the lung, and abdominal masses identified by a physical examination that cannot be measured by imaging techniques. Future studies could look at more intrinsic and complex texture analysis and machine-learning algorithms.

IBC is highly angiogenic and lymphangiogenic by gene expression quantification and immunohistochemistry, compared with other breast cancer types (McCarthy et al., 2002; Shan et al., 2013). Therefore, it has the potential to metastasize through both the hematogenous and lymphatic routes, and shows extensive spread to the breast subcutaneous lymphatics. Our results may be related to these pathologic properties of IBC. Lower mean CT attenuation in IBC is indicative of extensive necrosis, resulting from rapid tumor growth and edematous swelling from lymphatic emboli. Shan et al. (2013) evaluated the correlation between the degree of necrosis on CT images and the expression of hypoxic and angiogenesis biomarkers in breast cancers, and demonstrated that lower CT values were significantly positively correlated with microvessel density and hypoxic biomarkers. Rapid expansion of a tumor mass requires neovascularization to provide adequate oxygen and metabolites to the proliferating tumor cells; the absence of which results in hypoxia. Perrone et al. (2008) reported that the mean CT attenuation of 27 non-inflammatory breast cancers was 125 HU on post-contrast images. In our study, the mean attenuation of IBC was 60.92 HU on post-contrast images, lower than the values for non-inflammatory breast cancers (Perrone et al., 2008), which might have resulted from tumor necrosis and hypoxia. A CT texture study of metastatic renal cell carcinomas (Matoori et al., 2017) reported that lower mean attenuation predicted poor OS, and that pretreatment mean attenuation was a predictive marker of treatment response. Similarly, a lower mean CT attenuation of the tumor might be useful as an imaging biomarker for predicting worse OS in patients with IBC.

In our study, lower MPP predicted worse OS in patients with IBC. This finding is consistent with previous reports, demonstrating that lower MPP was associated with poor survival outcomes in patients with non-small cell lung cancer and hepatic metastatic colorectal cancer (Lubner et al., 2015; Hayano et al., 2016). Several studies have reported that MPP is associated with tumor angiogenesis, grade, and necrosis (Ganeshan et al., 2013; Lubner et al., 2015; Tian et al., 2015). In non-small cell lung cancer, MPP showed a significant inverse association with CD34 expression, which is an immunohistochemial marker of angiogenesis (Ganeshan et al., 2013). In hepatic metastatic colorectal cancers, lower MPP, SD, entropy on post-constrast images were associated with higher tumor grade (Lubner et al., 2015). In addition, because MPP was strongly associated with tumor necrosis, it was suggested to be the best predictor for chemotherapy response in soft-tissue sarcomas (Tian et al., 2015). In our study, lower MPP may be related to more extensive lymphovascular proliferation and necrosis in patients with IBC with poor OS. Further studies will be needed to confirm the association between MPP and pathological characteristics of IBC.

Entropy indicates irregularity and a marker of tumor heterogeneity. In our study, lower entropy on post-contrast CT images was associated with poor OS in patients with IBC. Previous CT texture studies of colon cancer and hepatic metastatic colorectal cancers also showed a relationship between lower entropy and poorer survival (Ng et al., 2013; Lubner et al., 2015). Kim et al. (2017) investigated MRI textural features in breast cancer and demonstrated that lower entropy on post-contrast T1-weighted images and higher entropy on T2-weighted images were associated with worse recurrence-free survival. They inferred that tumors with greater heterogeneity (higher entropy on T2-weighted images) could show more homogeneous enhancement on post-contrast T1-weighted images because of increased permeability; greater permeability that leads to lower contrast resolution between parenchyma and adjacent blood vessels may represent as lower heterogeneity on post-contrast images. Therefore, based on the results of previous MRI texture

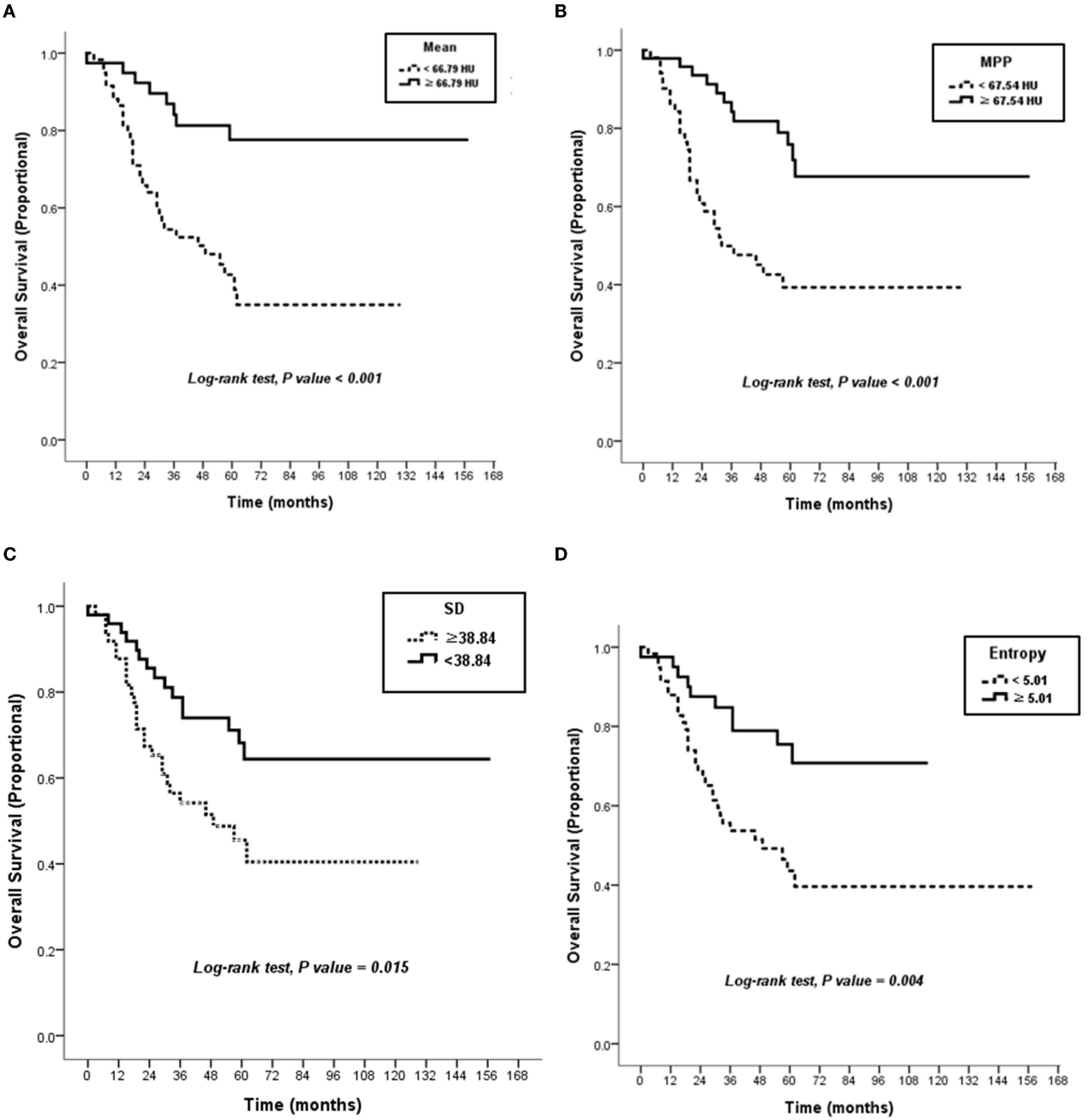


FIGURE 4 | Kaplan–Meier curves. There were significant differences in overall survival according to attenuation of 66.79 HU in mean on post-contrast the CT image ($P < 0.001$) **(A)**, attenuation of 67.54 HU in MPP on the CT image ($P < 0.001$) **(B)**, SD of 38.84 on the CT image ($P = 0.015$) **(C)**, and 5.01 in entropy on the CT image ($P = 0.004$) **(D)**. HU, Hounsfield units; MPP, mean of positive pixels; SD, standard deviation.

studies and our CT study, we suppose that lower entropy on post-contrast CT or MRI may be associated with worse OS in breast cancer patients.

Metastasis is common at the time of diagnosis of IBC because the increased lymphovascular proliferation and vascular permeability of IBC promote a distant metastasis. A large-cohort study of IBC (Wang et al., 2020) revealed that 224 of 635 patients (35.3%) had a distant metastasis, which was very similar to the 38.7% seen in our study. In our study, a metastasis had a significant impact on OS; however, known prognotic factors, such

TABLE 3 | Univariate Cox proportional hazard analysis of CT texture features associated with worse overall survival.

		Overall survival	
Texture feature	**No. of patients**	**Hazard ratio**	***P*-value**
Mean			
≥66.79 HU	39	Reference	
<66.79 HU	59	3.87 (1.78, 8.43)	0.001
SD			
≥38.84	49	Reference	
<38.84	49	2.18 (1.15, 4.13)	0.016
MPP			
≥67.54	47	Reference	
<67.54	51	3.31 (1.67, 6.52)	0.001
Entropy			
≥5.01	40	Reference	
<5.01	58	2.78 (1.36, 5.69)	0.005

Data in parentheses are 95% confidence intervals. SSF, spatial scale filters; HU, Hounsfield unit; SD, standard deviation; MPP, mean of positive pixels.

TABLE 4 | Multivariate Cox proportional hazard analysis of clinicopathologic variables and CT texture features without image filtration associated with worse overall survival.

	Overall survival	
Variable	**Hazard ratio**	***P*-value**
***M* stage**		
0	Reference	
1	3.53 (1.85, 6.74)	<0.001
Mean		
≥66.79 HU	Reference	
<66.79 HU	3.26 (1.48, 7.16)	0.003
***M* stage**		
0	Reference	
1	3.68 (1.93, 7.01)	<0.001
SD		
≥38.84	Reference	
<38.84	1.77 (0.92, 3.38)	0.083
***M* stage**		
0	Reference	
1	3.81 (1.99, 7.28)	<0.001
MPP		
≥67.54	Reference	
<67.54	3.03 (1.52, 6.05)	0.002
***M* stage**		
0	Reference	
1	3.79 (2.00, 7.18)	<0.001
Entropy		
≥5.01	Reference	
<5.01	2.70 (1.28, 5.69)	0.009

Data in parentheses are 95% confidence intervals. HU, Hounsfield unit; SD, standard deviation; MPP, mean of positive pixels.

as lymph node involvement, hormone receptor negativity or age (Chevallier et al., 1987; Fields et al., 1989; Palangie et al., 1994), were not associated with worse OS.

Our study had several limitations. Firstly, the mean follow-up period of the study population was short at 47.9 months; that of non-survivors was 25.8 months, and that of survivors was 63.0 months. A longer follow-up period will be required to acquire a robust estimate of the impact of texture features on the survival outcome. Secondly, we obtained texture features on a single section of the largest cross-sectional diameter of the tumor, rather than on the multi-slice or volume analysis. However, according to previous studies, the texture results from two-dimensional and three-dimensional analysis are similar, and the intraobserver reproducibility of single-section measurements is quite high (Lubner et al., 2015; Smith et al., 2015). Based on our results, ICCs for mean, SD, entropy, and MPP, which were significant predictors for OS, showed almost perfect to perfect agreement. On the other hand, ICCs for skewness and kurtosis showed slight to a fair agreement. Thirdly, our study was retrospective in nature and results were acquired from a single institution, using a single software program on a small sample size. Finally, we included the patients with IBC with the M1 stage in our study population despite the M1 stage is a well-known factor predicting a worse prognosis. We collected the patients with IBC during 10 years, but the number of study population was small, and 38.7% of the patients with IBC (38 of 98) had M1 stage at the time of diagnosis. Nevertheless, our study proved that texture features were significant factors independent from the M1 stage for predicting worse OS in patients with IBC.

In conclusion, our results from an exploratory study highlight the potential of CT-based texture analysis to act as a non-invasive imaging biomarker, predicting the prognosis of patients with IBC. Identifying new risk factors from chest CT scans may allow clinicians to optimize surveillance and treatment strategies based on individual risk of the patients with IBC.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Korea University Ansan Hospital IRB committee. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

BKS was responsible for research design, research process management, and financial support. SES performed image analysis. JC and BG performed the statistical analysis. OHW

and KRC interpreted the results. SES and BKS drafted the manuscript. OHW and ESK made substantial contributions to the experimental design, data interpretation, and manuscript. All authors read and approved the final draft of the manuscript and are accountable for all aspects of the work.

FUNDING

This research was supported by the National Research Foundation of Korea (NRF) grant, funded by the Korean government (MSIT: No. NRF-2021R1A2C1010565), the Korean University grant (K2011241), and the UK National Institute for Health Research University College London Hospitals Biomedical Research Centre.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2021.695305/full#supplementary-material

Supplementary Figure 1 | Receiver operating characteristic (ROC) curve analyses for the optimal cutoff values. An ROC curve (an area under the curve [AUC], 0.67; $P = 0.004$) for mean attenuation **(A)**, an ROC curve (AUC, 0.63; $P = 0.024$) for SD **(B)**, an ROC curve (AUC, 0.69; $P = 0.001$) for MPP **(C)**, and an ROC curve (AUC, 0.62; $P = 0.042$) for entropy **(D)**.

Supplementary Figure 2 | Kaplan–Meier curves of the *M* stage. There were significant differences in overall survival according to the metastasis (M1) stage at the time of the diagnosis.

Supplementary Table 1 | CT texture features at spatial scale filters of 2–6.

Supplementary Table 2 | Spearman rank correlation coefficients for CT texture features.

REFERENCES

Anderson, W. F., Schairer, C., Chen, B. E., Hance, K. W., and Levine, P. H. (2005). Epidemiology of inflammatory breast cancer (IBC). *Breast Dis.* 22, 9–23. doi: 10.3233/BD-2006-22103

Chamming's, F., Ueno, Y., Ferre, R., Kao, E., Jannot, A. S., Chong, J., et al. (2018). Features from computerized texture analysis of breast cancers at pretreatment MR imaging are associated with response to neoadjuvant chemotherapy. *Radiology* 286, 412–420. doi: 10.1148/radiol.2017170143

Chee, C. G., Kim, Y. H., Lee, K. H., Lee, Y. J., Park, J. H., Lee, H. S., et al. (2017). CT texture analysis in patients with locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy: a potential imaging biomarker for treatment response and prognosis. *PLoS ONE* 12:e0182883. doi: 10.1371/journal.pone.0182883

Chevallier, B., Asselain, B., Kunlin, A., Veyret, C., Bastit, P., and Graic, Y. (1987). Inflammatory breast cancer. Determination of prognostic factors by univariate and multivariate analysis. *Cancer* 60, 897–902. doi: 10.1002/1097-0142(19870815)60:4<897::AID-CNCR2820600430>3.0.CO;2-S

Cristofanilli, M., Valero, V., Buzdar, A. U., Kau, S. W., Broglio, K. R., Gonzalez-Angulo, A. M., et al. (2007). Inflammatory breast cancer (IBC) and patterns of recurrence: understanding the biology of a unique disease. *Cancer* 110, 1436–1444. doi: 10.1002/cncr.22927

Dawood, S., Lei, X., Dent, R., Gupta, S., Sirohi, B., Cortes, J., et al. (2014). Survival of women with inflammatory breast cancer: a large population-based study. *Ann. Oncol.* 25, 1143–1151. doi: 10.1093/annonc/mdu121

Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., et al. (2009). New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer* 45, 228–247. doi: 10.1016/j.ejca.2008.10.026

Fields, J. N., Kuske, R. R., Perez, C. A., Fineberg, B. B., and Bartlett, N. (1989). Prognostic factors in inflammatory breast cancer. Univariate and multivariate analysis. *Cancer* 63, 1225–1232. doi: 10.1002/1097-0142(19890315)63:6<1225::AID-CNCR2820630632>3.0.CO;2-5

Fouad, T. M., Barrera, A. M. G., Reuben, J. M., Lucci, A., Woodward, W. A., Stauder, M. C., et al. (2017). Inflammatory breast cancer: a proposed conceptual shift in the UICC-AJCC TNM staging system. *Lancet Oncol.* 18, e228–e232. doi: 10.1016/S1470-2045(17)30192-4

Fouad, T. M., Kogawa, T., Liu, D. D., Shen, Y., Masuda, H., El-Zein, R., et al. (2015). Overall survival differences between patients with inflammatory and noninflammatory breast cancer presenting with distant metastasis at diagnosis. *Breast Cancer Res. Treat.* 152, 407–416. doi: 10.1007/s10549-015-3436-x

Ganeshan, B., Goh, V., Mandeville, H. C., Ng, Q. S., Hoskin, P. J., and Miles, K. A. (2013). Non-small cell lung cancer: histopathologic correlates for texture parameters at CT. *Radiology* 266, 326–336. doi: 10.1148/radiol.12112428

Ganeshan, B., Panayiotou, E., Burnand, K., Dizdarevic, S., and Miles, K. (2012). Tumour heterogeneity in non-small cell lung carcinoma assessed by CT texture analysis: a potential marker of survival. *Eur. Radiol.* 22, 796–802. doi: 10.1007/s00330-011-2319-8

Giuliano, A. E., Edge, S. B., and Hortobagyi, G. N. (2018). Eighth edition of the AJCC cancer staging manual: breast cancer. *Ann. Surg. Oncol.* 25, 1783–1785. doi: 10.1245/s10434-018-6486-6

Hance, K. W., Anderson, W. F., Devesa, S. S., Young, H. A., and Levine, P. H. (2005). Trends in inflammatory breast carcinoma incidence and survival: the surveillance, epidemiology, and end results program at the National Cancer Institute. *J. Natl. Cancer Inst.* 97, 966–975. doi: 10.1093/jnci/dji172

Hayano, K., Kulkarni, N. M., Duda, D. G., Heist, R. S., and Sahani, D. V. (2016). Exploration of imaging biomarkers for predicting survival of patients with advanced non-small cell lung cancer treated with antiangiogenic chemotherapy. *Am. J. Roentgenol.* 206, 987–993. doi: 10.2214/AJR.15.15528

James, J., Teo, M., Ramachandran, V., Law, M., Stoney, D., and Cheng, M. (2019). A critical review of the chest CT scans performed to detect asymptomatic synchronous metastasis in new and recurrent breast cancers. *World J. Surg. Oncol.* 17:40. doi: 10.1186/s12957-019-1584-x

Kim, J. H., Ko, E. S., Lim, Y., Lee, K. S., Han, B. K., Ko, E. Y., et al. (2017). Breast cancer heterogeneity: MR imaging texture analysis and survival outcomes. *Radiology* 282, 665–675. doi: 10.1148/radiol.2016160261

Low, J. A., Berman, A. W., Steinberg, S. M., Danforth, D. N., Lippman, M. E., and Swain, S. M. (2004). Long-term follow-up for locally advanced and inflammatory breast cancer patients treated with multimodality therapy. *J. Clin. Oncol.* 22, 4067–4074. doi: 10.1200/JCO.2004.04.068

Lubner, M. G., Stabo, N., Lubner, S. J., del Rio, A. M., Song, C., Halberg, R. B., et al. (2015). CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. *Abdom. Imaging* 40, 2331–2337. doi: 10.1007/s00261-015-0438-4

Matoori, S., Thian, Y., Koh, D. M., Sohaib, A., Larkin, J., Pickering, L., et al. (2017). Contrast-Enhanced CT density predicts response to sunitinib therapy in metastatic renal cell carcinoma patients. *Transl. Oncol.* 10, 679–685. doi: 10.1016/j.tranon.2017.06.001

McCarthy, N. J., Yang, X., Linnoila, I. R., Merino, M. J., Hewitt, S. M., Parr, A. L., et al. (2002). Microvessel density, expression of estrogen receptor alpha, MIB-1, p53, and c-erbB-2 in inflammatory breast cancer. *Clin. Cancer Res.* 8, 3857–3862.

Mule, S., Thiefin, G., Costentin, C., Durot, C., Rahmouni, A., Luciani, A., et al. (2018). Advanced hepatocellular carcinoma: pretreatment contrast-enhanced ct texture parameters as predictive biomarkers of survival in patients treated with Sorafenib. *Radiology* 288, 445–455. doi: 10.1148/radiol.2018171320

Nelson, D. A., Tan, T. T., Rabson, A. B., Anderson, D., Degenhardt, K., and White, E. (2004). Hypoxia and defective apoptosis drive genomic instability and tumorigenesis. *Genes Dev.* 18, 2095–2107. doi: 10.1101/gad.1204904

Ng, F., Ganeshan, B., Kozarski, R., Miles, K. A., and Goh, V. (2013). Assessment of primary colorectal cancer heterogeneity by using whole-tumor texture analysis: contrast-enhanced CT texture as a biomarker of 5-year survival. *Radiology* 266, 177–184. doi: 10.1148/radiol.12120254

Palangie, T., Mosseri, V., Mihura, J., Campana, F., Beuzeboc, P., Dorval, T., et al. (1994). Prognostic factors in inflammatory breast cancer and therapeutic implications. *Eur. J. Cancer* 30A, 921–927. doi: 10.1016/0959-8049(94)90115-5

Perrone, A., Lo Mele, L., Sassi, S., Marini, M., Testaverde, L., Izzo, L., et al. (2008). MDCT of the breast. *Am. J. Roentgenol.* 190, 1644–1651. doi: 10.2214/AJR.07.3145

Pickles, M. D., Lowry, M., and Gibbs, P. (2016). Pretreatment prognostic value of dynamic contrast-enhanced magnetic resonance imaging vascular, texture, shape, and size parameters compared with traditional survival indicators obtained from locally advanced breast cancer patients. *Invest. Radiol.* 51, 177–185. doi: 10.1097/RLI.0000000000000222

Pierga, J. Y., Bidard, F. C., Autret, A., Petit, T., Andre, F., Dalenc, F., et al. (2017). Circulating tumour cells and pathological complete response: independent prognostic factors in inflammatory breast cancer in a pooled analysis of two multicentre phase II trials (BEVERLY-1 and−2) of neoadjuvant chemotherapy combined with bevacizumab. *Ann. Oncol.* 28, 103–109. doi: 10.1093/annonc/mdw535

Shan, X., Wang, D., Chen, J., Xiao, X., Jiang, Y., Wang, Y., et al. (2013). Necrosis degree displayed in computed tomography images correlated with hypoxia and angiogenesis in breast cancer. *J. Comput. Assist. Tomogr.* 37, 22–28. doi: 10.1097/RCT.0b013e318279abd1

Smith, A. D., Gray, M. R., del Campo, S. M., Shlapak, D., Ganeshan, B., Zhang, X., et al. (2015). Predicting overall survival in patients with metastatic melanoma on antiangiogenic therapy and RECIST stable disease on initial posttherapy images using CT texture analysis. *Am. J. Roentgenol.* 205, W283–W293. doi: 10.2214/AJR.15.14315

Tian, F., Hayano, K., Kambadakone, A. R., and Sahani, D. V. (2015). Response assessment to neoadjuvant therapy in soft tissue sarcomas: using CT texture analysis in comparison to tumor size, density, and perfusion. *Abdom. Imaging* 40, 1705–1712. doi: 10.1007/s00261-014-0318-3

Ueno, N. T., Buzdar, A. U., Singletary, S. E., Ames, F. C., McNeese, M. D., Holmes, F. A., et al. (1997). Combined-modality treatment of inflammatory breast carcinoma: twenty years of experience at M. D. Anderson Cancer Center. *Cancer Chemother. Pharmacol.* 40, 321–329. doi: 10.1007/s002800050664

Van der Auwera, I., Van Laere, S. J., Van den Eynden, G. G., Benoy, I., van Dam, P., Colpaert, C. G., et al. (2004). Increased angiogenesis and lymphangiogenesis in inflammatory versus noninflammatory breast cancer by real-time reverse transcriptase-PCR gene expression quantification. *Clin. Cancer Res.* 10, 7965–7971. doi: 10.1158/1078-0432.CCR-04-0063

Wang, Z., Chen, M., Pan, J., Wang, X., Chen, X. S., and Shen, K. W. (2020). Pattern of distant metastases in inflammatory breast cancer - a large-cohort retrospective study. *J. Cancer* 11, 292–300. doi: 10.7150/jca.34572

Zhang, H., Graham, C. M., Elci, O., Griswold, M. E., Zhang, X., Khan, M. A., et al. (2013). Locally advanced squamous cell carcinoma of the head and neck: CT texture and histogram analysis allow independent prediction of overall survival in patients treated with induction chemotherapy. *Radiology* 269, 801–809. doi: 10.1148/radiol.13130110

Conflict of Interest: BG is the Co-Founder/Co-Inventor of TexRAD texture analysis software used in this study and a shareholder (not an employee) of Feedback Plc., a UK based company which owns, develops and markets the TexRAD texture analysis software.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Song, Seo, Cho, Woo, Ganeshan, Kim and Cha. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 02 August 2021
doi: 10.3389/fonc.2021.686235

Radiomics-Based Texture Analysis of ^{68}Ga-DOTATATE Positron Emission Tomography and Computed Tomography Images as a Prognostic Biomarker in Adults With Neuroendocrine Cancers Treated With ^{177}Lu-DOTATATE

Charlotte Atkinson[1†], Balaji Ganeshan[2†], Raymond Endozo[2], Simon Wan[1], Matthew D. Aldridge[1], Ashley M. Groves[2], Jamshed B. Bomanji[1] and Mark N. Gaze[1]*

[1] Departments of Oncology and Nuclear Medicine, University College London Hospitals NHS Foundation Trust, London, United Kingdom, [2] Institute of Nuclear Medicine, University College London, London, United Kingdom

OPEN ACCESS

Edited by:
Omar Sultan Al-Kadi,
The University of Jordan, Jordan

Reviewed by:
Francesco Fiz,
Université de Genève, Switzerland
Nicolò Gennaro,
Ente Ospedaliero Cantonale (EOC), Switzerland
Bruno De Santi,
Politecnico di Torino, Italy

***Correspondence:**
Balaji Ganeshan
b.ganeshan@ucl.ac.uk

[†]These authors share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 26 March 2021
Accepted: 12 July 2021
Published: 02 August 2021

Citation:
Atkinson C, Ganeshan B, Endozo R, Wan S, Aldridge MD, Groves AM, Bomanji JB and Gaze MN (2021) Radiomics-Based Texture Analysis of ^{68}Ga-DOTATATE Positron Emission Tomography and Computed Tomography Images as a Prognostic Biomarker in Adults With Neuroendocrine Cancers Treated With ^{177}Lu-DOTATATE. Front. Oncol. 11:686235. doi: 10.3389/fonc.2021.686235

Purpose: Neuroendocrine tumors (NET) are rare cancers with variable behavior. A better understanding of prognosis would aid individualized management. The aim of this hypothesis-generating pilot study was to investigate the prognostic potential of tumor heterogeneity and tracer avidity in NET using texture analysis (TA) of ^{68}Ga-DOTATATE positron emission tomography (PET) and non-enhanced computed tomography (CT) performed at baseline in patients treated with ^{177}Lu-DOTATATE. It aims to justify a larger-scale study to evaluate its clinical value.

Methods: The pretherapy ^{68}Ga-DOTATATE PET-CT scans of 44 patients with metastatic NET (carcinoid, pancreatic, thyroid, head and neck, catecholamine-secreting, and unknown primary NET) treated with ^{177}Lu-DOTATATE were analyzed retrospectively using commercially available texture analysis research software. Image filtration extracted and enhanced objects of different sizes (fine, medium, coarse), then quantified heterogeneity by statistical and histogram-based parameters (mean intensity, standard deviation, entropy, mean of positive pixels, skewness, and kurtosis). Regions of interest were manually drawn around up to five of the most ^{68}Ga-DOTATATE avid lesions for each patient. ^{68}Gallium uptake on PET was quantified as SUV_{max} and SUV_{mean}. Associations between imaging and clinical markers with progression-free (PFS) and overall survival (OS) were assessed using univariate Kaplan-Meier analysis. Independence of the significant univariate markers of survival was tested using multivariate Cox regression analysis.

Results: Measures of heterogeneity (higher kurtosis, higher entropy, and lower skewness) on coarse-texture scale CT and unfiltered PET images predicted shorter

PFS (CT coarse kurtosis: p=0.05, PET entropy: p=0.01, PET skewness: p=0.03) and shorter OS (CT coarse kurtosis: p=0.05, PET entropy: p=0.01, PET skewness p=0.02). Conventional PET parameters such as SUV_{max} and SUV_{mean} showed trends towards predicting outcome but were not statistically significant. Multivariate analysis identified that CT-TA (coarse kurtosis: HR=2.57, 95% CI=1.22–5.38, p=0.013) independently predicted PFS, and PET-TA (unfiltered skewness: HR=9.05, 95% CI=1.19–68.91, p=0.033) independently predicted OS.

Conclusion: These preliminary data generate a hypothesis that radiomic analysis of neuroendocrine cancer on ^{68}Ga-DOTATATE PET-CT may be of prognostic value and a valuable addition to the assessment of patients.

Keywords: ^{68}Ga-DOTATATE PET/CT, ^{177}Lu-DOTATATE molecular radiotherapy, neuroendocrine tumor (NET), prognostic biomarker, texture analysis, radiomics

INTRODUCTION

Neuroendocrine tumors (NET) are rare and heterogeneous. This variety presents challenges when making decisions about sequencing of therapies in metastatic disease. Clinical features such as primary tumor, metastatic characteristics, and prior systemic treatments have prognostic potential (1). However, they are inconsistent, and more reliable markers are needed.

Somatostatin receptor (STTR) expression on the tumor cell surface can be measured by ^{68}Ga-DOTATATE positron emission tomography (PET) computed tomography (CT) hybrid-imaging scans. There is evidence that tumor avidity on ^{68}Ga-DOTATATE scans may be linked to outcome in NET patients (2, 3).

De-differentiation of NET is often associated with a more aggressive phenotype. Histologic intratumor heterogeneity is a measure of this (4) but is difficult to evaluate in metastases. Functional imaging is simpler for assessment of metastatic heterogeneity and enables quantitative assessment. Texture analysis (TA) of CT and PET images, where the tumor coarseness and distribution of pixel gray-level intensity can reflect tumor heterogeneity, is one such tool (5). It uses standard non-invasive diagnostic imaging, avoiding the need for extra radiation or additional complex, invasive procedures or expense. TA has been used in various tumors including lung and esophageal cancers: higher indices of heterogeneity are associated with worse outcomes (6–8).

Could PET and CT-TA be useful, in addition to clinical factors, for prognostication in NET? With better prognostic information, management of metastatic NET could be individualized, leading to personalized selection of treatments in those most likely to benefit.

This pilot study aimed to demonstrate the prognostic potential of tumor avidity and heterogeneity in patients with NET from a single institution treated with ^{177}Lutetium DOTATATE peptide receptor radionuclide therapy (PRRT). We hypothesized that tumors with higher heterogeneity scores and lower standardized uptake values (SUV) would be associated with a worse prognosis.

Abbreviations: NET, neuroendocrine tumor; TA, texture analysis; PET, positron emission tomography; CT, computed tomography; PFS, progression-free survival; OS, overall survival; ^{68}Ga, ^{68}Gallium; ^{177}Lu, ^{177}Lutetium; SUV, standardized uptake value; STTR, somatostatin receptor; PRRT, peptide receptor radionuclide therapy; KM, Kaplan Meier; IQR, interquartile range; PR, partial response; SD, stable disease; PD, progressive disease; SSF, spatial scale filter; MTV, metabolic tumor volume; NAT, nor-adrenaline transporter.

MATERIALS AND METHODS

Patients

This retrospective pilot study included all NET patients aged over 18 years treated at our hospital with ^{177}Lu-DOTATATE molecular radiotherapy between 2007 and 2016. It included carcinoid, pancreatic, thyroid, head and neck, catecholamine-secreting, and unknown primary NET. Neuroblastoma and central nervous system NET were excluded. Ethical approval for data collection and analysis was granted by HRA and Health and Care Research Wales (IRAS project number 204124).

For treatment, patients had to have a baseline ^{68}Ga-DOTATATE PET-CT with tumor avidity greater than liver. Other criteria were creatinine <150 μmol L^{-1} or creatinine clearance >50 ml min^{-1} 1.72 m^{-2}, and Eastern Cooperative Oncology Group (ECOG) performance status of 3 or less.

Treatment

^{177}Lu-DOTATATE was obtained commercially from IDB Holland (http://www.idb-holland.com/our-products/lutetium-177-lumark/-Baarle-Nassau, Netherlands). Patients were planned to receive four administrations of ^{177}Lu-DOTATATE at a dose of 7.4 GBq. Those with glomerular filtration rate of <60 ml min^{-1} 1.72 m^{-2} were treated with 50% dose reduction. Premedication was with oral ondansetron and dexamethasone over 5 days. To reduce renal radiation, an intravenous infusion of amino acids (2.5% L-lysine and 2.5% L-arginine) was commenced (1 L/4 h) 30 min before administration of intravenous ^{177}Lu-DOTATATE, given over 30 min. Subsequently, patients were monitored weekly for toxicities. Typically, four courses were administered at 8- to 12-week intervals; however, treatment was discontinued for disease progression or toxicity.

Follow-Up

Patients were followed and imaged 6 monthly with ^{68}Ga-DOTATATE PET-CT. The primary outcome measure was

progression-free survival (PFS) as further cancer therapy might affect overall survival. PFS was calculated from the date of first ^{177}Lu-DOTATATE course to disease progression of any type; classified on ^{68}Ga-DOTATATE PET-CT as development of a new metastasis or a 20% increase in target lesion size (RECIST v1.1) and not increase in SUV of a target lesion. Overall survival (OS) was recorded.

PET-CT Image Acquisition

^{68}Ga-DOTATATE PET-CT scans were carried out on a Discovery STE dual PET-CT machine (GE Healthcare). The activity of ^{68}Ga-DOTATATE injected intravenously was dependent on the yield from the ^{68}Ga/Ge generator, but it was intended that 250 MBq be used. PET-CT scans were acquired after 45 min. No intravenous or oral contrast was used. A low-dose scout projection was used to localize the region of head to thigh for transmission and emission imaging. The low-dose CT component of the PET-CT scan was obtained at 120 kVp with modulated tube current (30–300 mA) and 1.75 pitch. CT slice thickness was 5 mm. In plane pixel resolution was 0.98 mm × 0.98 mm. A three-dimensional mode was used for PET acquisition, spending 4 min in each bed position. Images were iteratively reconstructed with 21 subsets with attenuation correction. PET-CT images were fused but were also available separately for analysis.

PET-CT Image Analysis

Conventional PET-CT Analysis

Regions of interest (ROI) were drawn around the most avid lesions/metastases (up to five tumor foci) for each patient as seen on the ^{68}Ga-DOTATATE PET-CT scan. For each patient within each lesion, standardized Gallium uptake value (SUV) on PET was quantified as the maximum (SUV_{max}) and average (SUV_{mean}) uptake values. Additionally, for each patient, the ^{68}Ga-DOTATATE uptake tumor area on PET for each lesion was quantified as total number of pixels. This measurement is similar to the metabolic tumor area (which has increasingly been shown to be a useful prognostic biomarker ^{18}F-FDG PET imaging), but the term "metabolic" in the setting of ^{68}Ga-DOTATATE PET is a misnomer, hence using the descriptive term ^{68}Ga-DOTATATE uptake tumor area.

CT and PET Texture Analysis (CT-TA and PET-TA)

Tumor heterogeneity was further assessed using texture analysis performed on the non-contrast, low-dose CT, as well as the attenuation-corrected PET components of the pretreatment ^{68}Ga-DOTATATE PET scans, using a commercially available TexRAD research software (TexRAD, part of Feedback Medical Ltd, www.fbkmed.com, Cambridge, UK), which has been developed following a stringent regulatory and quality process based on ISO 13485 and 9001.

For each patient, two-dimensional ROIs were manually drawn around the individual lesions (identified on PET as described above) on the CT and PET slices corresponding to the most avid slice on PET. For CT-TA, each ROI underwent a filtration-histogram technique where the image filtration step (using a Laplacian of Gaussian band-pass filter, which is similar to a non-orthogonal Wavelet) extracted and enhanced objects/features of different sizes corresponding to different spatial scale filter (SSF), varying from 2 mm (fine), 3–5 mm (medium), and 6 mm (coarse) texture scale. This was followed by quantification using statistical and histogram-based analysis comprising mean gray-level intensity, standard deviation, entropy, mean of positive pixels, skewness, and kurtosis. Pixels within the ROI less than −50 HU (e.g., air, gases) were excluded from CT-TA. In addition, CT-TA also comprised of quantification without the use of image filtration as a control. In total, 36 texture features were quantified for CT-TA. Owing to the inherently low resolution associated with PET SUV images, PET-TA did not use image filtration which comprised six texture features. This filtration-histogram-based texture analysis methodology has undergone a qualification process demonstrating biological correlate, technical validation, clinical applications (prognosis, disease severity, treatment response/prediction), and cost-effective analysis evidenced through a number of publications (8–13). **Figure 1** provides an illustration of the filtration-histogram-based TA as applied on an individual lesion for a couple of patients with NET on ^{68}Ga-DOTATATE PET/CT.

The texture analysis was conducted by a single radiographer (RE) with 20 years of experience in radiology and nuclear medicine under the supervision of a consultant radionuclide radiologist/nuclear medicine physician (SW) with 6 years of experience and senior imaging scientist (BG) with 12 years of experience in texture analysis. All of them were blinded to the clinical outcome.

Clinical Parameters

Information about site of primary, location of metastases, and previous treatments was collected to assess the effect of clinical parameters on outcome measures.

Statistical Analysis

For each patient the average value of the different image quantification from all available lesions was derived to provide patient level metrics. The relationships of tumor heterogeneity (quantified by CT-TA and PET-TA), Gallium DOTATATE uptake (SUV_{max}, SUV_{mean}, and ^{68}Ga-DOTATATE uptake tumor area), and clinical markers with patient survival (PFS and OS) were assessed using univariate Kaplan-Meier (KM) survival analysis. Median values for these parameters were used as thresholds to separate the survival plots (poor and good prognostic groups). The significant differences in the survival plots were further evaluated using non-parametric log rank test. KM curves for patients above and below the median cut-off for the significant parameter were constructed.

A Benjamini-Hochberg correction using a false discovery rate of 30% was applied to the analysis in order to reduce the false discovery rate that can occur when making multiple comparisons (14).

Multivariate step-wise forward-Wald Cox regression analysis was used to determine which significant univariate markers (after Benjamini-Hochberg correction) were independent of each other in predicting patient survival along with the hazard ratio, 95% confidence interval. In all analyses, p values <0.05 were considered significant.

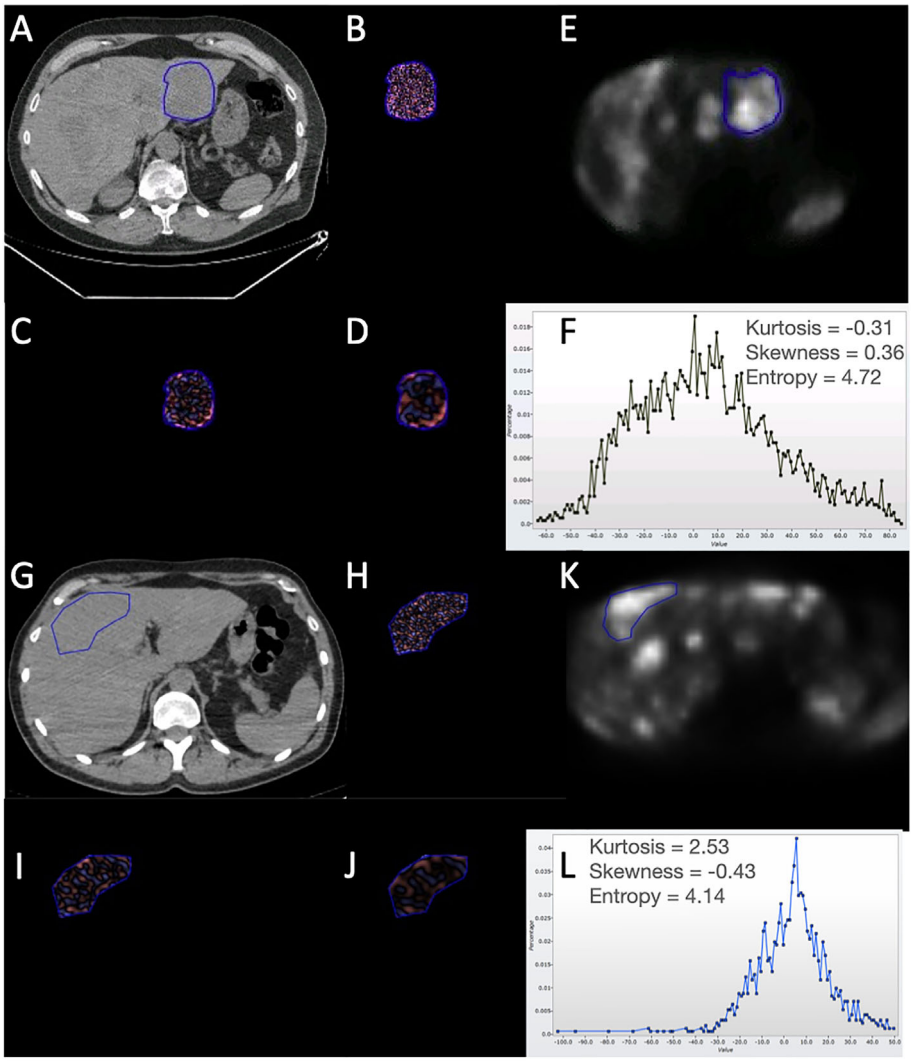


FIGURE 1 | Illustration of the filtration-histogram-based texture analysis as applied on an individual liver lesion for two patients with NET on 68Ga-DOTATATE PET/CT. Top patient had a better response/outcome (PFS=30 months and OS=50 months) compared to the bottom patient (PFS=1 month and OS=6 months) to ^{177}Lu-DOTATATE. **(A, G)** highlight the conventional CT image with the tumor encircled for the two patients, followed by corresponding **(B, H)**, which highlight the fine-texture map; **(C, I)** highlight the medium-texture map; **(D, J)** highlight the coarse-texture map; **(E, K)** highlight the PET image with the tumor encircled; and **(F, L)** highlight the histogram distribution along with the texture features (kurtosis, skewness, and entropy) for the coarse-texture map on CT (as an example). For the texture maps: Pink reflects positive filtered pixel intensity values, i.e., bright areas (the more intense the color, the brighter the area); blue reflects negative filtered pixel intensity values, i.e., dark areas (the more intense the color, the darker the area). For the PET-image: Brighter areas reflect increased avidity.

RESULTS

Forty-four patients were treated with ^{177}Lu-DOTATATE between June 2007 and February 2016. Patient characteristics are listed in **Table 1**. Seventy-five percent of patients (33/44) received at least four treatments, and 18% (8/44) only had one dose of radionuclide.

The median follow-up time was 37 months (range 5–82 months), during which, 39 (80%) patients progressed and 28 (57%) patients died. The median PFS was 22 months (95% confidence interval: 10–34, IQR 8–40). The median OS was 37 months (95% confidence interval: 19–55, IQR 9–74).

Thirty-two patients had a ^{68}Ga-DOTATATE PET-CT at 6 months after treatment to assess disease response. No patients had a complete response to ^{177}Lu-DOTATATE; however, 10 patients (23%) had a partial response (PR), 20 had stable disease (SD) (63%), and 2 had disease progression (PD) (6%). Twelve (27%) patients had PD before 6 months and so did not have a response assessment scan.

Univariate Kaplan-Meier Analysis

The PET-CT, texture, and clinical parameters that influenced PFS and OS at the median threshold values after Benjamini-Hochberg correction are detailed in **Tables 2** and **3**, respectively.

TABLE 1 | Patient characteristics.

	Number of patients (44)	%
Gender		
Female	15	34
Male	29	66
Median Age	56	
Age Range	28–87	
Familial Syndrome	**4**	**9**
SDHB mutation	3	7
MEN1	1	2
Site of Primary Carcinoid	**17**	**39**
- Lung	6	
- Small Bowel	6	
- Appendix	1	
- Large Bowel	3	
- Unknown Primary	1	
Pancreatic NET	**9**	**20**
- Gastrinoma	1	
- Glucagonoma	3	
- Insulinoma	1	
- Non-functioning	4	9
Catecholamine secreting	**7**	**16**
- Phaeochromocytoma	1	
- Paraganglioma	6	
Thyroid	**4**	**9**
- Hürthle Cell	1	
- Medullary Carcinoma	3	
Head and Neck Adenocarcinoma	**3**	**7**
NET of Unknown Primary	**4**	**9**
Metastatic disease	**44**	**100**
- Bone	23	52
- Brain	3	7
- Liver	29	65
- Lung	14	32
- Lymph Node	32	72
- Peritoneal	5	11
Previous Treatments	**44**	**100**
- Surgery	26	59
- PRRT	5	11
- RT	10	22
- Somatostatin Analogue	18	40
- Chemo	13	29
- Nil	4	9
Treatments after Lu-DO	**18**	**40**
- Surgery	2	5
- PRRT	3	7
- Somatostatin Analogue	2	5
- Chemo	7	16

The numbers in bold represent the number of patients for each main category.

Conventional PET-CT Analysis

Larger ^{68}Ga-DOTATATE tumor uptake area was a significant marker of shorter PFS (p=0.016, **Table 2** and **Figure 2**) and OS (p=0.012, **Table 3** and **Figure 3**). Lower PET SUV_{max} and SUV_{mean} values demonstrated a trend towards shorter PFS and OS but did not reach statistical significance (PFS p=0.15 and p=0.17, respectively; and OS p=0.42 and p=0.33, respectively).

CT and PET Texture Analysis (CT-TA and PET-TA)

Among the different CT-TA parameters, higher kurtosis at coarse-texture scale (SSF=6 mm) was a significant marker of shorter PFS and OS [PFS p=0.049 (**Table 2**) and OS p=0.022 (**Table 3**)]. From among the different PET-TA parameters, higher entropy without filtration and lower skewness without filtration were significant markers of shorter PFS [entropy p=0.0088, skewness p=0.026 (**Table 2** and **Figure 2**)] and OS [entropy p=0.0058, skewness p=0.0073 (**Table 3** and **Figure 3**)].

Clinical Parameters

Of the different NET subtypes, patients with thyroid primaries had the worst survival with a mean OS of 6 months [p=0.013 (**Table 3**)]. Presence of lung metastases also negatively affected PFS [p=0.026 (**Table 2**)]. Otherwise, distribution of metastatic disease and previous treatments did not appear to have a significant effect on outcome in our cohort.

Multivariate Cox Regression Analysis

When all of the significant predictors of PFS from the above univariate analysis (with Benjamini-Hochberg correction) were included in a multivariate Cox regression model, CT-TA kurtosis on coarse-texture scale (SSF 6) (HR=2.57, 95% CI=1.22–5.38, p=0.013) and presence of lung metastases (HR=2.97, 95% CI=1.37–6.46, p=0.006) independently predicted PFS. For OS, skewness on unfiltered PET-TA (HR=9.05, 95% CI=1.19–68.91, p=0.033) and having a thyroid primary (HR=6.21, 95% CI=1.92–20.12, p=0.002) were independent predictors on multivariate analysis.

DISCUSSION

The behavior of metastatic NET is varied as shown here where OS ranged between 0 and 82 months. TNM staging doesn't distinguish between patients with one or many metastases, requiring other risk stratification techniques. Our study did not demonstrate that patients with bone metastases, functioning pancreatic NETs, or chemotherapy pretreatment fare worse, as have other series (15, 16). In this study, only a thyroid primary or lung metastases predicted poor survival. These inconsistencies highlight the need for other prognostic biomarkers.

Histological evaluation of the mitotic count and the Ki67 proliferation index are used to estimate aggressiveness, and these factors are crucial to WHO NET grading (17). However, histologically and phenotypically, there can be considerable heterogeneity between different metastatic deposits in one patient and within each focus of tumor (18, 19). Tumor de-differentiation may occur between presentation and recurrence, resulting in a discrepancy between disease behavior and the original histological appearances. Imaging assessments are time-specific and can assess all visible metastases.

We demonstrate how radiomics-based CT texture analysis could be used to assess the heterogeneity of metastatic NET to assist prognostication. Kurtosis on coarse filtered images and skewness and entropy on PET-TA were significantly related to survival outcomes. There is increasing interest in imaging of heterogeneity and radiomics-based texture analysis (20–25). Most of these studies focus predominantly on the PET-TA in NET demonstrating prognostic potential or treatment-related changes, where in one study (20) entropy significantly predicted outcome. However, our analysis, including both CT-TA and

TABLE 2 | Summary of univariate Kaplan-Meier analysis for PFS for the most significant CT and PET imaging parameters using the median value as the cut-off.

Imaging	Median cut-off (direction indicates poor prognosis)	Median survival in months (number of patients)		Kaplan-Meier p-value	Benjamini-Hochberg corrected p-value
		Below threshold	Above threshold		
CT-TA					
Kurtosis SSF 6	>0.29	35 (29)	14 (15)	0.049	0.05
PET-TA (no filter)					
Entropy	>3.79	39 (22)	12 (22)	0.0088	0.01
Skewness	<0.95	16 (37)	58 (7)	0.008	0.03
PET conventional analysis					
PET ^{68}Ga-DOTATATE-uptake tumor area	>52.5 pixels	39 (22)	14 (22)	0.016	0.02
Clinical parameters					
Lung metastases	>1	35 (29)	14 (15)	0.026	0.04

PET-TA, could be considered more comprehensive. Among all texture features analyzed in this research, kurtosis and entropy best distinguished between good and poor prognostic groups. These appear to consistently predict outcome in other TA studies and have been described as the most global measures of heterogeneity (13).

^{68}Ga-DOTATATE PET has superseded octreoscan in being able to predict prognosis in NET (26). In several studies, SUV_{max} and SUV_{mean} have shown prognostic potential where higher values are associated with longer survival (2, 3). In this study, SUV_{max} and SUV_{mean} showed only trends towards predicting survival outcomes. The reliability of these parameters is questionable. For example, marked intra and interpatient variabilities in SUV values have been seen in studies reviewing quantitative SUV analysis in similar patient cohorts to ours (27). In the studies where higher ^{68}Ga-DOTATATE avidity has predicted improved survival, there are large differences in the threshold values for SUV_{max} used in the literature, highlighting the potential for variability of this biomarker. SUV_{max} is a poor representation of the uptake in the tumor as a three-dimensional structure. We have described a new term, ^{68}Ga-DOTATATE uptake tumor area (measured in pixels), because "metabolic" in this setting is not accurate. Larger ^{68}Ga-DOTATATE uptake tumor area was significantly associated with both poor PFS and OS in the univariate analysis where SUV measures were not. MTV (metabolic tumor volume) has been explored in a number of cancers, with promising prognostic ability (28). Perhaps, with more validation, ^{68}Ga-DOTATATE uptake tumor area could be used in a similar way.

CT and PET-TA is a developing technology. The filtration-histogram-based CT and PET-TA employed in this study has undergone rigorous qualification processes, essential for validating any biomarker. Miles et al. have highlighted how texture features extracted from the filtration-histogram technique reflect different components of heterogeneity (object/feature size, number, density in relation to the background tissue) (29). Furthermore, the texture features extracted from this filtration-histogram technique have been investigated in several tumor types where increased heterogeneity has been associated with inferior prognosis (8, 10–13, 30). To better understand NET, we need to correlate the imaging findings with histological parameters. In lung cancer, a number of texture parameters have been found to be associated with histological markers of angiogenesis and hypoxia (31). In a tumor where high Ki67 indicates de-differentiation and worse outcomes, it is plausible that the proliferation index would correlate well with heterogeneity on imaging. Conversely, SSTR2 expression, which is generally increased in well-differentiated tumors (32), would be expected to be inversely associated with tumor heterogeneity. We will investigate these associations in a future analysis, as well as the relationship between the noradrenaline transporter (NAT), which is overexpressed in catecholamine-secreting NETs, and TA.

Limitations

It is important to acknowledge the significant drawbacks to retrospective research. There are several potential biases, including referral bias, selection bias, and information bias. We

TABLE 3 | Summary of univariate Kaplan-Meier analysis for OS for the most significant CT and PET imaging parameters using the median value as the cut-off.

Imaging	Median cut-off (direction indicates poor prognosis)	Median survival in months (number of patients)		Kaplan-Meier p-value	Benjamini-Hochberg corrected p-value
		Below threshold	Above threshold		
CT-TA					
Kurtosis SSF 6	>0.29	50 (29)	16 (15)	0.022	0.05
PET-TA (no filter)					
Entropy	>3.79	53 (22)	22 (22)	0.0058	0.01
Skewness	<0.95	28 (37)	Not reached (7)	0.034	0.02
PET conventional analysis					
PET ^{68}Ga-DOTATATE-uptake tumor area	>52.5 pixels	Not reached (22)	22 (22)	0.011	0.03
Clinical parameters					
Thyroid primary	>1	50 (44)	2(5)	0.013	0.04

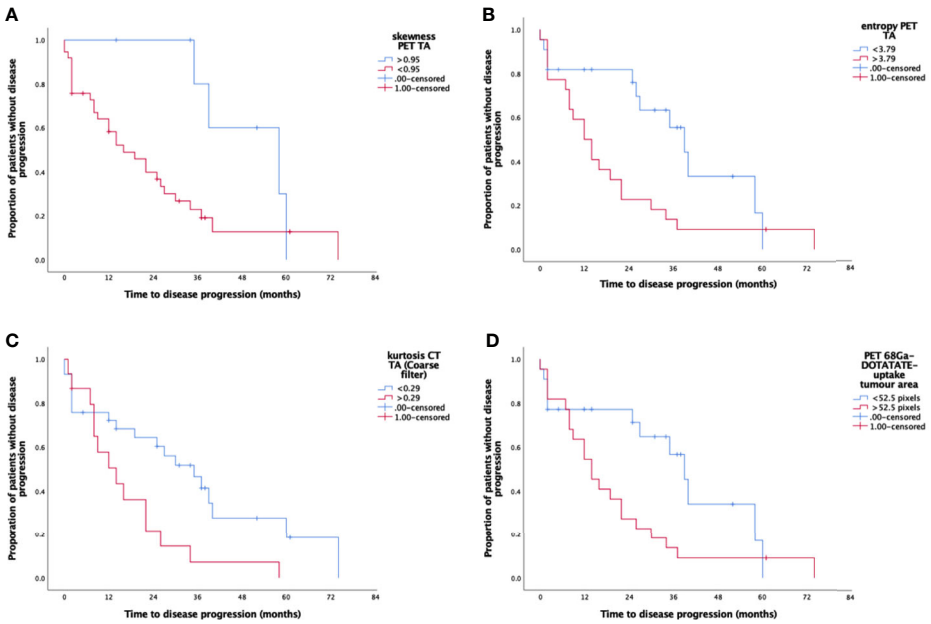


FIGURE 2 | Kaplan-Meier survival curves for significant CT texture analysis (CT-TA), PET texture analysis (PET-TA), and PET uptake parameters against progression-free survival. Panel **(A)** PET-TA Skewness (without filtration), corrected p= 0.03, Panel **(B)** PET-TA Entropy (without filtration), corrected p=0.01, Panel **(C)** CT-TA Kurtosis (coarse), corrected p= 0.05, Panel **(D)** ^{68}Ga-DOTATATE tumor uptake area, corrected p= 0.02. Median threshold values for each parameter are shown on the individual graphs.

also acknowledge that the sample of patients is small, from a single center, and heterogenous including NETs with different biology and included patients who had been heavily pretreated. While this could be perceived to limit the interpretation of this study, in this preliminary, hypothesis-generating study, we chose to include all of these patients to reflect the variable NET population, allowing a practical review of TA. Moreover, due to the referral pathway of the majority of patients, we were unable to centrally review the histopathology and comment on Ki67 or grade of tumor in this cohort. We recognize that these parameters are prognostically important and plan to collect and analyze these data against TA in future prospective studies. Also, as the focus of this "theranostic" research study was to employ the radiomics analysis on the ^{68}Ga-DOTATATE PET-CT scans as a prognostic marker in NET patients treated with ^{177}Lu-DOTATATE, therefore only lesions that were avid on ^{68}Ga-DOTATATE PET-CT were analyzed. In future studies we need to look at a range of lesions. Another limitation of the exploratory pilot study was to analyze only a single slice corresponding to the most avid part of the lesion. Single-slice analysis will potentially be time-efficient (within routine clinical practice) and potentially reduce operator variability (unless the segmentation is semi/fully automated) but may not entirely capture tumor heterogeneity information across the whole tumor volume, which will provide additional biological information potentially important in the clinical decision-making. Although previous studies, e.g., in PET/CT in lymphoma (33), have demonstrated the ability of single-slice texture analysis to provide prognostic information, we nevertheless believe that future studies should employ single-slice and volumetric analysis and compare both the approaches. In addition to the filtration-histogram-based texture analysis employed in this study, future work should evaluate a comprehensive radiomics analysis comprising of filtration-histogram, high-order statistics, and shape, along with their biological intuitiveness, inter- and intra-operator variability, and reproducibility assessment. Also, the use of a more sophisticated approach (e.g., cumulative assessment) for aggregating the individual lesion quantification rather than a simplistic approach (e.g., averaging) to get a patient-specific score needs to be evaluated in a future study. These are all essential for it to be used in clinical practice, so further prospective work is necessary in texture analysis on ^{68}Ga-DOTATATE PET-CT in an independent, larger cohort including more clearly defined subclassifications of NETs and homogeneous treatment regime across several centers to validate our findings. If validated, radiomics-based TA in NET may have clinical value.

CONCLUSION

We have demonstrated in this preliminary study that PET-CT-TA shows promise in the stratification of patients, and important prognostic information can potentially be obtained from one baseline scan. As patient outcomes are very variable, identification of those with the worst prognosis is important;

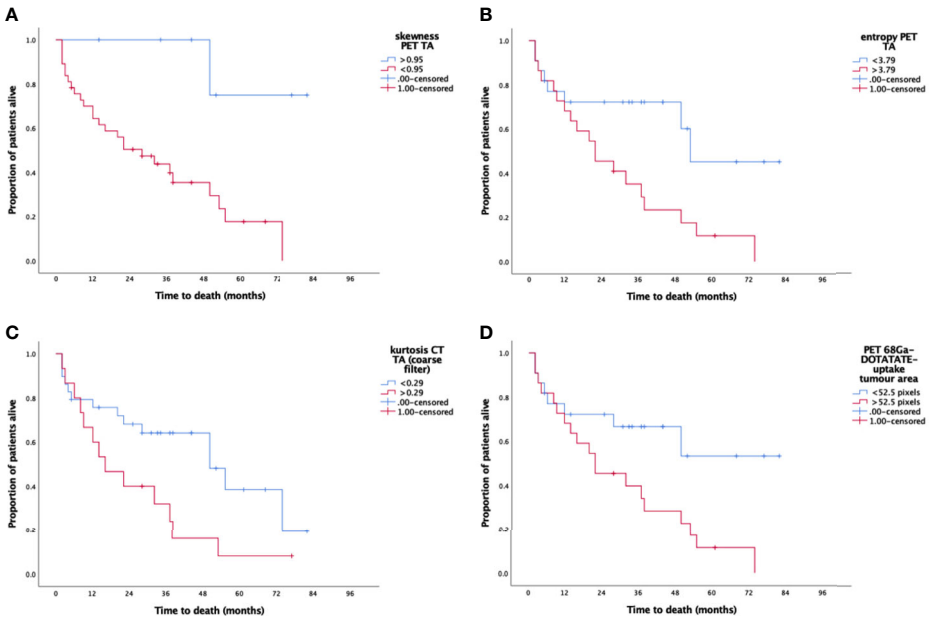


FIGURE 3 | Kaplan-Meier survival curves for significant CT and PET texture parameters (CT-TA and PET-TA) against overall survival. Panel **(A)** PET-TA Skewness (without filtration), corrected p=0.02, Panel **(B)** PET-TA Entropy (without filtration), corrected p=0.01, Panel **(C)** CT-TA Kurtosis (coarse), corrected p=0.05, Panel **(D)** ^{68}Ga-DOTATATE tumor uptake area, corrected p= 0.03. Median threshold values for each parameter are shown on the individual graphs.

they could be spared futile treatment and resources saved. Patients with more heterogeneous tumors, suggesting a more aggressive phenotype, might be offered chemotherapy earlier or could be considered for dual modality treatment with PRRT and chemotherapy as in the Australian Gastrointestinal Trials Group study, CONTROL NETS.

DATA AVAILABILITY STATEMENT

Data not readily available due to institutional permissions, request to access dataset should be directed to the corresponding author, BG at b.ganeshan@ucl.ac.uk.

ETHICS STATEMENT

Ethical approval for data collection and analysis was granted by HRA and Health and Care Research Wales (IRAS project number 204124). The patients/participants provided their written informed consent for treatment.

AUTHOR CONTRIBUTIONS

Guarantor of integrity of the entire study: MG and JB. Study concepts and design: CA, BG, MG, and JB. Literature research: CA and BG. Clinical Studies: CA, BG, RE, MA, and SW. Experimental Studies/Data analysis: CA, BG, RE, MA, SW, AG, and MG. Statistical Analysis: BG. Manuscript Preparation: CA, BG, and MG. Manuscript Editing: CA, BG, MG, and JB. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the University College London Hospitals Charity, supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. Dr M. N Gaze is supported by the Radiation Research Unit at the Cancer Research UK City of London Centre Award [C7893/A28990].

REFERENCES

1. Yao JC, Hassan M, Phan A, Dagohoy C, Leary C, Mares JE, et al. One Hundred Years After "Carcinoid": Epidemiology of and Prognostic Factors for Neuroendocrine Tumors in 35,825 Cases in the United States. *J Clin Onc* (2008) 26:3063–72. doi: 10.1200/JCO.2007.15.4377
2. Ambrosini V, Campana D, Polverari G, Peterle C, Diodato S, Ricci C, et al. Patients With Neuroendocrine Tumors of the Pancreas. *J Nucl Med* (2010) 56:1843–9. doi: 10.2967/jnumed.115.162719
3. Campana D, Ambrosini V, Pezzilli R, Fanti S, Maria A, Labate M, et al. Standardised Uptake Values of ^{68}Ga-DOTANOC PET: A Promising Prognostic Tool in Neuroendocrine Tumors. *J Nucl Med* (2010) 51:353–60. doi: 10.2967/jnumed.109.066662
4. Jögi A, Vaapil M, Johansson M, Pahlman S. Cancer Cell Differentiation Heterogeneity and Aggressive Behavior in Solid Tumors. *Ups J Med Sci* (2012) 117:217–24. doi: 10.3109/03009734.2012.659294
5. Davnall F, Yip CSP, Ljungqvist G, Selmi M, Ng F, Sanghera B, et al. Assessment of Tumor Heterogeneity: An Emerging Imaging Tool for

Clinical Practice? *Insights Imaging* (2012) 3:573–89. doi: 10.1007/s13244-012-0196-6
6. Ganeshan B, Skogen K, Pressney I, Coutroubis D, Miles K. Tumour Heterogeneity in Oesophageal Cancer Assessed by CT Texture Analysis: Preliminary Evidence of an Association With Tumour Metabolism, Stage, and Survival. *Clin Radiol . R Coll Radiologists* (2012) 67:157–64. doi: 10.1016/j.crad.2011.08.012
7. Ganeshan B, Abaleke S, Young RCD, Chatwin CR, Miles KA. Texture Analysis of Non-Small Cell Lung Cancer on Unenhanced Computed Tomography: Initial Evidence for a Relationship With Tumour Glucose Metabolism and Stage. *Cancer Imaging* (2010) 10:137–43. doi: 10.1102/1470-7330.2010.0021
8. Win T, Miles KA, Janes SM, Ganeshan B, Shastry M, Endozo R, et al. Tumor Heterogeneity and Permeability as Measured on the CT Component of PET/CT Predict Survival in Patients With Non-Small Cell Lung Cancer. *Clin Cancer Res* (2013) 19:3591–9. doi: 10.1158/1078-0432.CCR-12-1307
9. Rinke A, Müller H-H, Schade-Brittinger C, Klose K-J, Barth P, Wied M, et al. Placebo-Controlled, Double Blind, Prospective, Randomized Study on the Effect of Octreotide LAR in the Control of Tumor Growth in Patients With Metastatic Neuroendocrine Midgut Tumors : A Report From the PROMID Study Group. *J Clin Onc* (2009) 27:4656–63. doi: 10.1200/JCO.2009.22.8510
10. Goh V, Ganeshan B, Nathan P, Juttla JK, Vinayan A, Miles KA. Assessment of Response to Tyrosine Kinase Inhibitors in Metastatic Renal Cell Cancer: CT Texture as a Predictive Biomarker. *Radiology* (2011) 261:165–71. doi: 10.1148/radiol.11110264
11. Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour Heterogeneity in Non-Small Cell Lung Carcinoma Assessed by CT Texture Analysis: A Potential Marker of Survival. *Eur Radiol* (2012) 22:796–802. doi: 10.1007/s00330-011-2319-8
12. Ng F, Ganeshan B, Kozarski R, Miles KA, Goh V. Assessment of Primary Colorectal Cancer Heterogeneity by Using Whole-Tumor Texture Analysis: Contrast-Enhanced. *Radiology* (2013) 266:177–84. doi: 10.1148/radiol.12120254
13. Ganeshan B, Miles KA, Babikir S, Shortman R, Afaq A, Ardeshna KM. CT-Based Texture Analysis Potentially Provides Prognostic Information Complementary to Interim FDG-PET for Patients With Hodgkin's and Aggressive Non-Hodgkin's Lymphomas. *Eur Radiol* (2017) 27:1012–20. doi: 10.1007/s00330-016-4470-8
14. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B* (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
15. Kwekkeboom DJ, de Herder WW, Kam BL, van Eijck CH, van Essen M, Kooij PP, et al. Treatment With the Radiolabeled Somatostatin Analog [177lu-DOTA0,Tyr3]Octreotate: Toxicity, Efficacy, and Survival. *J Clin Oncol* (2008) 26:2124–30. doi: 10.1200/JCO.2007.15.2553
16. Bodei L, Kidd M, Paganelli G, Grana CM, Drozdov I, Cremonesi M, et al. Long-Term Tolerability of PRRT in 807 Patients With Neuroendocrine Tumours: The Value and Limitations of Clinical Factors. *Eur J Nucl Med Mol Imaging* (2015) 42:5–19. doi: 10.1007/s00259-014-2893-5
17. Bosman F, Carneiro F, Hruban R, Theise N. *WHO Classification of Tumors of Digestive System. 4th Edition* Vol. 3. Geneva, Switzerland: WHO Press (2010).
18. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N Engl J Med* (2012) 366:883–92. doi: 10.1056/NEJMoa1113205
19. Mcgranahan N, Swanton C. Review Clonal Heterogeneity and Tumor Evolution: Past , Present, and the Future. *Cell* (2017) 168:613–28. doi: 10.1016/j.cell.2017.01.018
20. Werner RA, Lapa C, Ilhan H, Higuchi T, Buck AK, Lehner S, et al. Survival Prediction in Patients Undergoing Radionuclide Therapy Based on Intratumoral Somatostatin-Receptor Heterogeneity. *Oncotarget* (2017) 8:7039–49. doi: 10.18632/oncotarget.12402
21. Bezzi C, Mapelli P, Presotto L, Neri I, Scifo P, Savi A, et al. Radiomics in Pancreatic Neuroendocrine Tumors: Methodological Issues and Clinical Significance. *Eur J Nucl Med Mol Imaging* (2021). doi: 10.1007/s00259-021-05338-8
22. Liberini V, De Santi B, Rampado O, Gallio E, Dionisi B, Ceci F, et al. Impact of Segmentation and Discretization on Radiomic Features in ^{68}Ga-DOTA-TOC PET/CT Images of Neuroendocrine Tumor. *EJNMMI Phys* (2021) 8:21. doi: 10.1186/s40658-021-00367-6
23. Liberini V, Rampado O, Gallio E, De Santi B, Ceci F, Dionisi B, et al. ^{68}Ga-DOTATOC PET/CT-Based Radiomic Analysis and PRRT Outcome: A Preliminary Evaluation Based on an Exploratory Radiomic Analysis on Two Patients. *Front Med (Lausanne)* (2021) 7:601853. doi: 10.3389/fmed.2020.601853
24. Weber M, Kessler L, Schaarschmidt B, Fendler WP, Lahner H, Antoch G, et al. Treatment-Related Changes in Neuroendocrine Tumors as Assessed by Textural Features Derived From ^{68}Ga-DOTATOC PET/MRI With Simultaneous Acquisition of Apparent Diffusion Coefficient. *BMC Cancer* (2020) 20:326. doi: 10.1186/s12885-020-06836-y
25. Ibrahim A, Vallières M, Woodruff H, Primakov S, Beheshti M, Keek S, et al. Radiomics Analysis for Clinical Decision Support in Nuclear Medicine. *Semin Nucl Med* (2019) 49(5):438–49. doi: 10.1053/j.semnuclmed.2019.06.005
26. van Adrichem RCS, Kamp K, van Deurzen CHM, Biermann K, Feelders RA, Franssen GJH, et al. Is There an Additional Value of Using Somatostatin Receptor Subtype 2a Immunohistochemistry Compared to Somatostatin Receptor Scintigraphy Uptake in Predicting Gastroenteropancreatic Neuroendocrine Tumor Response? *Neuroendocrinology* (2016) 103:560–6. doi: 10.1159/000441604
27. Mojtahedi A, Thamake S, Tworowska I, Ranganathan D, Delpassand ES. The Value of ^{68}Ga-DOTATATE PET / CT in Diagnosis and Management of Neuroendocrine Tumors Compared to Current FDA Approved Imaging Modalities : A Review of Literature. *Am J Nucl Med Mol Imaging* (2014) 4:426–34.
28. Van De Wiele C, Kruse V, Smeets P, Sathekge M, Maes A. Predictive and Prognostic Value of Metabolic Tumour Volume and Total Lesion Glycolysis in Solid Tumours. *Eur J Nucl Med Mol Imaging* (2013) 40:290–301. doi: 10.1007/s00259-012-2280-z
29. Miles K, Ganeshan B, Hayball MB. CT Texture Analysis Using the Filtration-Histogram Method: What do the Measurements Mean? *Cancer Imaging* (2013) 13:400–6. doi: 10.1102/1470-7330.2013.9045
30. Zhang H, Graham C, Elci O, Griswold M, Zhang X, Khan M, et al. Locally Advanced Squamous Cell Carcinoma of the Head and Neck: CT Texture and Histogram Analysis Allow Independent Prediction of Overall Survival in Patients Treated. *Radiology* (2013) 269:801–9. doi: 10.1148/radiol.13130110
31. Ganeshan B, Goh V, Mandeville HC, Hoskin PJ, Miles KA. Non – Small Cell Lung Cancer : Histopathologic Correlates for Texture. *Radiology* (2013) 266:326–36. doi: 10.1148/radiol.12112428
32. Mehta S, De Reuver PR, Gill P, Andrici J, Urso LD, Mittal A, et al. Somatostatin Receptor SSTR-2a Expression is a Stronger Predictor for Survival Than Ki-67 in Pancreatic Neuroendocrine Tumors. *Medicine* (2015) 94:1–6. doi: 10.1097/MD.0000000000001281
33. Ganeshan B, Miles KA, Babikir S, Shortman R, Afaq A, Ardeshna KM, et al. CT-Based Texture Analysis Potentially Provides Prognostic Information Complementary to Interim Fdg-Pet for Patients With Hodgkin's and Aggressive non-Hodgkin's Lymphomas. *Eur Radiol* (2017) 27(3):1012–20. doi: 10.1007/s00330-016-4470-8

Conflict of Interest: One author, BG, is a Co-Founder/Co-Inventor of TexRAD texture analysis software used in this study and a shareholder (not an employee) of Feedback Plc., a UK based company which owns, develops and markets the TexRAD texture analysis software.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Atkinson, Ganeshan, Endozo, Wan, Aldridge, Groves, Bomanji and Gaze. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 03 August 2021
doi: 10.3389/fonc.2021.678441

Performances of Whole Tumor Texture Analysis Based on MRI: Predicting Preoperative T Stage of Rectal Carcinomas

Jia You and Jiandong Yin *

Department of Radiology, Shengjing Hospital of China Medical University, Shenyang, China

Objective: To determine whether there is a correlation between texture features extracted from high-resolution T2-weighted imaging (HR-T2WI) or apparent diffusion coefficient (ADC) maps and the preoperative T stage (stages T1–2 *versus* T3–4) in rectal carcinomas.

Materials and Methods: One hundred and fifty four patients with rectal carcinomas who underwent preoperative HR-T2WI and diffusion-weighted imaging were enrolled. Patients were divided into training (n = 89) and validation (n = 65) cohorts. 3D Slicer was used to segment the entire volume of interest for whole tumors based on HR-T2WI and ADC maps. The least absolute shrinkage and selection operator (LASSO) was performed to select feature. The significantly difference was tested by the independent sample t-test and Mann-Whitney U test. The support vector machine (SVM) model was used to develop classification models. The correlation between features and T stage was assessed by Spearman's correlation analysis. Multivariate logistic regression analysis was performed to identify independent predictors of tumor invasion. The performance of classifiers was evaluated by the receiver operating characteristic (ROC) curves.

Results: The wavelet HHH NGTDM strength (R_S = -0.364, $P < 0.001$) from HR-T2WI was an independent predictor of stage T3–4 tumors. The shape maximum 2D diameter column (R_S = 0.431, $P < 0.001$), log σ = 5.0 mm 3D first-order maximum (R_S = 0.276, P = 0.009), and log σ = 5.0 mm 3D first-order interquartile range (R_S = -0.229, P = 0.032) from ADC maps were independent predictors. In training cohorts, the classification models from HR-T2WI, ADC maps and the combination of two achieved the area under the ROC curves (AUCs) of 0.877, 0.902 and 0.941, with the accuracy of 79.78%, 89.86% and 89.89%, respectively. In validation cohorts, the three models achieved AUCs of 0.845, 0.881 and 0.910, with the accuracy of 78.46%, 83.08% and 87.69%, respectively.

Conclusions: Texture analysis based on ADC maps shows more potential than HR-T2WI in identifying preoperative T stage in rectal carcinomas. The combined application of HR-T2WI and ADC maps may help to improve the accuracy of preoperative diagnosis of rectal cancer invasion.

Keywords: magnetic resonance imaging, rectal cancer, apparent diffusion coefficient, texture analysis, T stage

OPEN ACCESS

Edited by:
Omar Sultan Al-Kadi,
The University of Jordan, Jordan

Reviewed by:
Hasan Aydin,
Dr. Abdurrahman Yurtaslan Ankara Oncology Training and Research Hospital, Turkey
Jacob Antunes,
MIM Software Inc., United States

***Correspondence:**
Jiandong Yin
jiandongyin@sina.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *26 March 2021*
Accepted: *19 July 2021*
Published: *03 August 2021*

Citation:
You J and Yin J (2021) Performances of Whole Tumor Texture Analysis Based on MRI: Predicting Preoperative T Stage of Rectal Carcinomas. Front. Oncol. 11:678441. doi: 10.3389/fonc.2021.678441

INTRODUCTION

Colorectal cancer is the second most common cancer in females and the third most common cancer in males in the United States, and 30–35% of colorectal cancers occur in the rectum (1, 2). In recent years, the incidence and mortality of rectal cancer have gradually increased, showing the imbalance both in age and region, and because of the occult onset of early rectal cancer, most patients are already in the locally advanced stage at the first diagnosis (3, 4).

Surgical resection has been considered the standard treatment for patients with early rectal cancer (stages T1–2), while locally advanced stage (stages T3–4) requires total mesorectal excision (TME) after neoadjuvant chemoradiotherapy (NCRT) (5). Thus, the precise preoperative stage of rectal cancer is vitally important. However, there are challenges in diagnosing whether the stage is T2 or T3, because the perirectal desmoplastic fibrotic response is similar with tumor penetration through the muscular rectal wall, which blurs the tumor borders (6). As the gold standard for disease diagnosis, the pathological result is usually obtained after surgery and cannot be used as a routine method to guide clinical diagnosis and treatment (7). Therefore, a comprehensive noninvasive method is desired to make the preoperative risk stratification available and dependable.

As a routine examination of rectal cancer, magnetic resonance imaging (MRI) has been widely accepted as the main examination method for preoperative diagnosis of rectal cancer, selection of treatment methods, and postoperative efficacy evaluation (8–10). High-resolution T2-weighted imaging (HR-T2WI) is currently a routine sequence for MRIs to check the degree of rectal tumor invasion (11, 12). Neoplasms show a slightly higher signal than normal rectal tissue on T2WI images. Previous studies have suggested that HR-T2WI, which allows higher spatial or temporal resolution acquisitions and consequently has better signal-to-noise ratio, has better accuracy in distinguishing tumors from normal rectal tissues and diagnosing preoperative T stages of rectal cancers (13, 14).

Apparent diffusion coefficient (ADC) maps were calculated from diffusion-weighted imaging (DWI), which obtained image contrast based on differences in the mobility of water protons between tissues with two different *b* values. ADC values have been approved to be accurate in the discrimination of benign and malignant lesions and tumor assessment (15), and a lower value shows the denser cell structure of the corresponding area. A recent study showed that this quantitative index may reflect the invasion of tumor tissue into normal tissue in patients with rectal cancers (16).

Texture analysis, one of the "radiomics" aspects, is a tool for high-throughput extraction and analysis of quantitative features obtained from medical images, including computed tomography, MRI, or positron emission tomography (17, 18). This technique facilitates the prediction of tumor stage and aggressiveness, which provides a more objective method to support individual treatment options (19). As a radiology signature, the potential benefit of texture analysis has been highlighted in many studies involving clinical diagnosis, therapy selection, treatment response assessment, and so on (20–23).

However, to the best of our knowledge, the number of studies on the texture analysis of rectal ADC maps is small. Meanwhile, in most of the past studies, ROI delineations were mostly carried out on the slice images that showed the largest tumor dimension (6, 24). This study was to extract texture features from the whole tumor volume based on HR-T2WI images and ADC maps, and evaluate the performance of classification models established by three-dimensional (3D) features in predicting the preoperative T stage (stages T1-2 *versus* T3-4) in rectal carcinomas.

MATERIALS AND METHODS

This study was approved by the Ethics Review Board of Shengjing Hospital of China Medical University (2020PS011K). Requirements for written informed consent were waived because of the retrospective nature of the study. The flow chart of this research is shown in **Figure 1**.

Patient Cohort

Between April 2014 and December 2019, the records of 538 patients who underwent rectal MRI were reviewed using the picture archiving and communication system. A total of 283 patients were initially selected using the following inclusion criteria: 1) patients who underwent preoperative rectal HR-T2WI and DWI examinations, and 2) who underwent radical rectal cancer surgery within 1 month after MRI, 3) patients with postoperative pathology involving confirmed rectal carcinomas. One hundred and twenty-nine patients were excluded for the following reasons: 1) incomplete medical records (such as a lack of T staging information) (n = 41), 2) obvious artifacts in the images (n = 9), 3) patients with mucinous carcinoma (such patients having low cell density and high ADC values that may cause larger deviations in the results) (n = 30), and 4) patients who received NCRT before MRI examinations (n = 49). Finally, 154 patients were enrolled in this study. These patients were divided in a training cohort (n = 89; 60 males, 29 females; mean age, 61.5 ± 11.2 years; range, 26–85 years) and a validation cohort (n = 65; 44 males, 21 females; mean age 61.2 ± 11.3; range, 27–82) by different MRI scanners. In each cohort, the patients were divided into a low T stage (stages T1–2) group and a high T stage (stages T3–4) group according to pathological results.

MRI Image Acquisition

The patients lay on the scanning bed in a supine position without bowel preparation or intravenous injection of antispasmodics. An axial HR-T2WI sequence and axial DWI sequence were conducted during the MRI examination, using an eight-channel phased array surface coil. Eighty-nine patients were examined with a 3.0 Tesla (T) MRI scanner (Ingenia 3.0; Philips Medical System, Best, The Netherlands). The acquisition parameters were as follows: repetition time/echo time, 6,000/76 ms; flip angle, 90°; matrix size, 576 × 576; field of view, 450 mm; slices, 48; slice thickness, 5 mm; spacing between slices, 1 mm; *b* values, 0 and 600 s/mm^2. Sixty-five patients were imaged with another 3.0T MRI scanner (Signa HDxt, GE Healthcare). The acquisition parameters were as follows: repetition time/echo time, 6540/130 ms; flip angle, 90°; matrix size, 320 × 320; field of view,

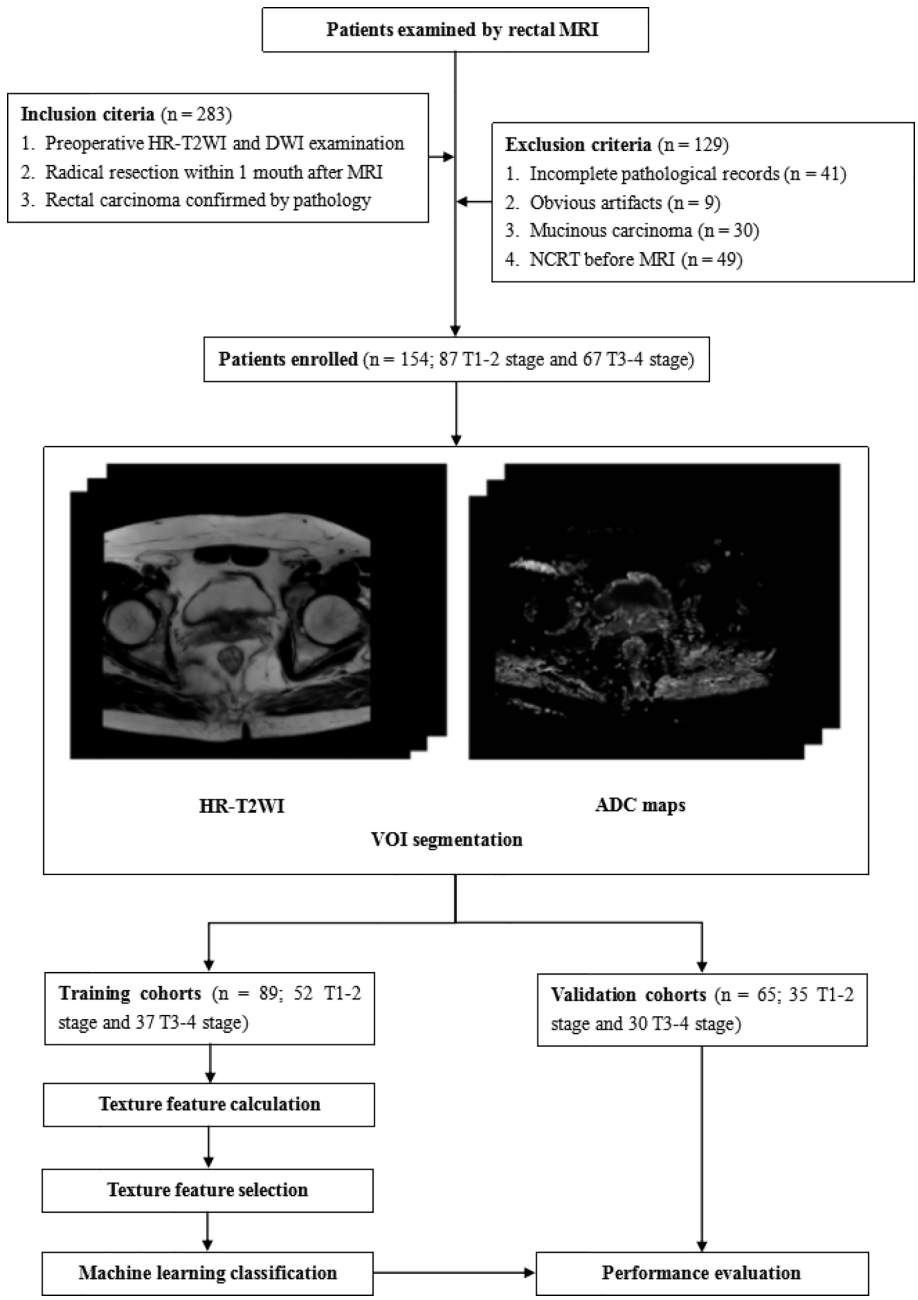


FIGURE 1 | The flow chart of this research. VOI, volume of interest.

360 mm; slices, 45; slice thickness, 5 mm; spacing between slices, 1 mm; b values, 0 and 600 s/mm^2.

ADC Map Acquisition

ADC maps were obtained with MATLAB R2018b (Mathworks, Natick, MA, USA) based on DWI images and calculated using the formula:

$$ADC = (\ln SI_0 - \ln SI_1)/(b - b_0),$$

where SI_0 and SI_1 represent the signal intensity of the pixel when the b value is 0 and 600 s/mm^2, respectively.

Lesion Segmentation

For each patient, two radiologists with more than 10 years of experience in rectal MRI took part in determining the tumor area using HR-T2WI with the DWI images as references. They individually and manually segmented the entire volume of interest (VOI) on HR-T2WI images, slice by slice, using 3D

Slicer (version 4.10.2, www.slicer.org). Both of the two radiologists were blinded to the pathological results. Obvious lumen content areas, necrosis, and gas were excluded from subsequent analysis. The delineated VOIs based on the HR-T2WI images were exactly copied to the same location of ADC maps.

Texture Extraction and Selection

The VOIs were first reconstructed into 3D labels and the reconstructed labels were placed on HR-T2WI and ADC maps to calculate features, and both of the two images were smoothed by the 8 mm Laplacian of Gaussian filter (25). A plug-in from 3D Slicer, Pyradiomics, was used for feature extraction. The extracted features were classified into the following categories: 1) shape-based features, 2) first-order statistics, 3) gray level dependence matrix (GLDM) features, 4) gray level co-occurrence matrix (GLCM) features, 5) gray level run length matrix features (GLRLM), 6) gray level size zone matrix (GLSZM) features, and 7) neighborhood gray-tone difference matrix (NGTDM) features. Features other than shape-based features were extracted from wavelet, Laplacian of Gaussian (Log), and original images. For wavelet transforms, each image was transformed in the x, y, and z directions using a low band pass filter or a high band pass filter. For Log transforms, the sigma (σ) values were 2.0, 3.0, 4.0, and 5.0 mm, respectively. As a result, for each patient, 1226 features were derived from HR-T2WI and ADC maps.

The above features were screened using MATLAB R2018b. First, highly correlated features were removed with coefficients greater than 0.95 using Pearson's correlation analysis. Then, the least absolute shrinkage and selection operator (LASSO) was used to reduce the dimensionality of the remaining features with 10-fold cross-validation to avoid the overfitting. Finally, the optimal subsets of features selected by LASSO with 10 fold cross validation were statistically tested to select significant features.

Statistical Analysis

All statistical analyses were performed using SPSS statistical software for Windows (IBM, Armonk, NY, USA), and $P < 0.05$ was considered statistically significant. The chi-square test was performed on categorical variables between different T stage groups. The Kolmogorov-Smirnov test was first used to check whether quantitative variables satisfied a normal distribution. If it satisfied a normal distribution, the independent sample *t*-test was performed between low and high T stage groups. Otherwise, the Mann-Whitney U test was performed. A support vector machine (SVM) method with 10 fold cross validation was used to establish a classification model based on the statistically significant texture features. Spearman's correlation analysis was conducted to assess the correlation between features and tumor T stage. Multivariate logistic regression analysis was also used to evaluate whether the statistically significant features were independent predictors of T3–4 rectal tumors. The performance of classification models in predicting the different T stages of rectal tumors was evaluated using receiver operating characteristic curve analyses by Medcalc (version 14.10.20, www.medcalc.org) by measuring the area under the ROC curve (AUC). In addition, the corresponding sensitivity and specificity were also determined.

The intraclass correlation coefficients (ICCs) were calculated to evaluate the interobserver variability between two radiologists delineating the VOIs and extracting features (0–0.4, poor agreement; 0.41–0.6, moderate agreement; 0.61–0.8, good agreement; and 0.81–0.9, excellent agreement).

RESULTS

Clinical Characteristics Analysis

A case was randomly selected to display the results of the VOI segmentation, as shown in **Figure 2**. The clinical and pathological characteristics of the cases selected in this study are shown in **Table 1**. There was no significant difference between low and high T stage groups in the terms of age ($P = 0.589$), sex ($P = 0.980$), and location of the lesion ($P = 0.083$). Lymph node invasion showed a significant difference between the two groups ($P = 0.008$).

Interobserver Agreement Evaluation

Features derived from the VOIs segmented separately by two radiologists showed excellent agreement, and the ICCs ranged from 0.837 to 0.945.

Features Extracted From HR-T2WI

Twenty features were selected from all the texture features extracted from HR-T2WI images by LASSO method. Five significant features were obtained by statistical analysis, as shown in **Table 2**. Shape flatness ($P = 0.020$), wavelet HHH NGTDM strength ($P < 0.001$), Log σ = 4.0mm 3D firstorder minimum ($P = 0.041$), Log σ = 3.0mm 3D GLSZM large area high gray level emphasis ($P = 0.018$) and Log σ = 5.0mm 3D firstorder interquartile range ($P = 0.019$) showed the significant difference in the discrimination of preoperative T stage, and achieved AUCs of 0.659, 0.713, 0.640, 0.643 and 0.646, respectively, whereas the other features were not significantly different. The distribution of those features is shown in **Figure 3**. The results of multivariate logistic regression analysis showed that the wavelet HHH NGTDM strength was an independent predictor of stage T3-4 rectal tumors, and it showed the highest correlation ($R_S = -0.364$, $P < 0.001$) with the preoperative T stage of rectal cancer among the significant features.

Features Extracted From ADC Maps

Twenty-three features were selected from all the texture features extracted from ADC maps by LASSO method. Ten significant features were obtained by statistical analysis, as shown in **Table 3**. Shape sphericity ($P = 0.027$), shape maximum 2DDiameter Column ($P < 0.001$), NGTDM Strength ($P = 0.018$), wavelet HLH first order energy ($P = 0.049$), wavelet LLL first order range ($P = 0.024$), Log σ = 2.0 mm 3D NGTDM contrast ($P = 0.003$), Log σ = 4.0 mm 3D firstorder skewness ($P = 0.024$), Log σ = 5.0 mm 3D first order maximum ($P = 0.009$),

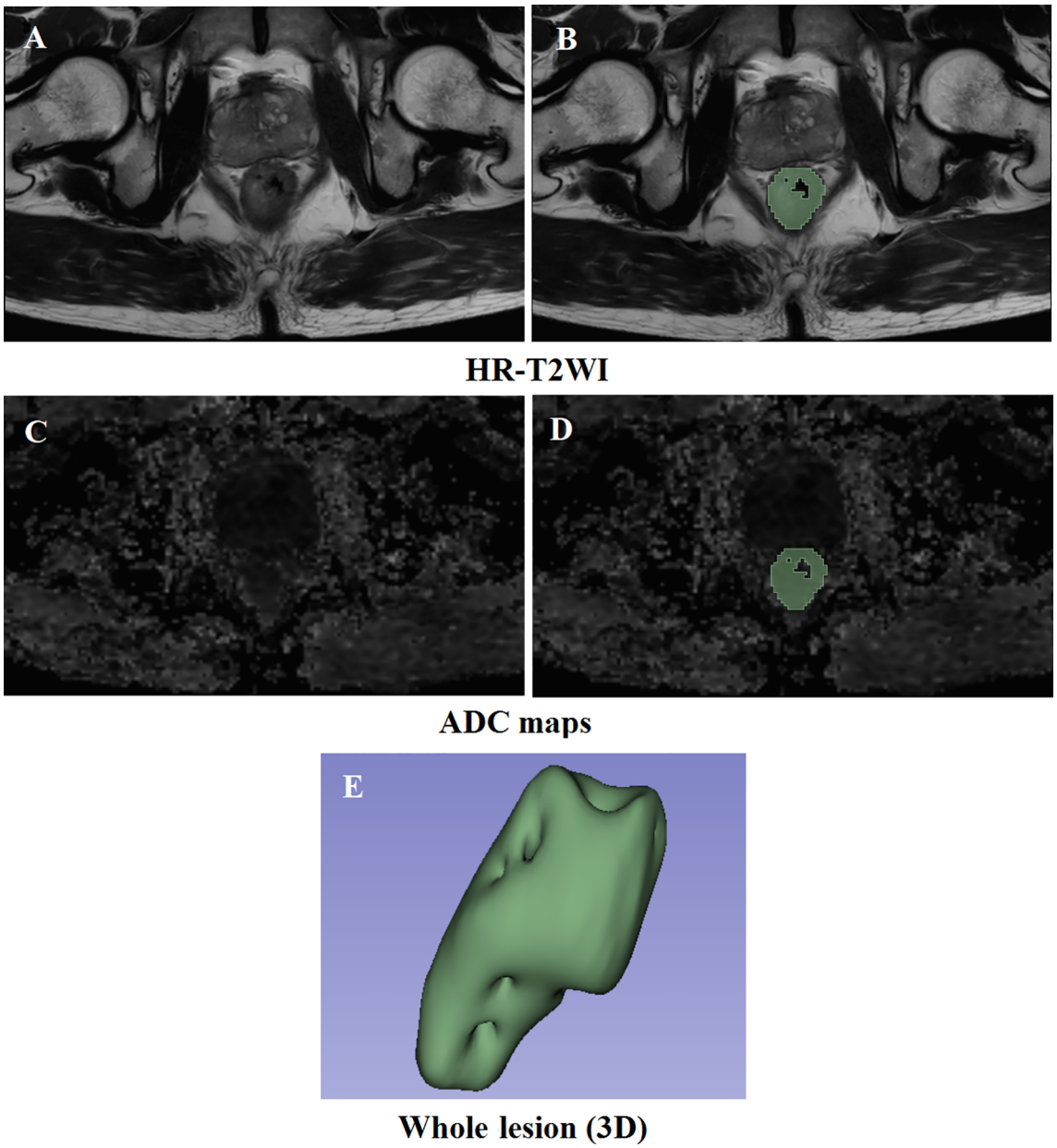


FIGURE 2 | Results of the delineation of the VOI. **(A, C)** High-resolution T2-weighted imaging and apparent diffusion coefficient maps at the same slice level, respectively. **(B, D)** The result of lesion segmentation. The green marked part is the delineated lesion. **(E)** The three-dimensional label of the whole tumor.

Log σ = 5.0 mm 3D first order kurtosis ($P = 0.011$) and Log σ = 5.0 mm 3D first order interquartile range ($P = 0.032$) showed the significant difference in the discrimination of preoperative T stage, and achieved AUCs of 0.641, 0.752, 0.647, 0.623, 0.614, 0.683, 0.598, 0.659, 0.659 and 0.634, respectively, whereas other features did not present significant difference. The results of multivariate logistic regression analysis showed that the shape maximum two-dimensional (2D) diameter column, Log σ = 5.0 mm 3D first-order maximum and Log σ = 5.0 mm 3D first-order interquartile range were independent predictors of stage T3–4 tumors. The shape maximum 2D diameter column showed the highest correlation with the preoperative T stage ($R_S = 0.431$, $P < 0.001$). The distribution of significant features from ADC maps is shown in **Figure 4**.

Performance of Classification Models

The performance of classification models for identifying preoperative rectal cancer T stage was shown in **Table 4**. In training cohorts, the model from HR-T2WI achieved an AUC of 0.877 [95% confidence interval (CI), 0.791-0.937] with a sensitivity of 86.54%, a specificity of 78.38%, and an accuracy of 79.78%. The model from ADC maps achieved an AUC of 0.902 (95% CI, 0.820-

TABLE 1 | Clinical and pathological characteristics of the cases selected for analysis.

Characteristics	Training cohort		*Validation cohort*		*P* value
	T1-2	T3-4	T1-2	T3-4	
Total patients	52 (58.43%)	37 (41.57%)	35 (53.85%)	30 (46.15%)	
Age (years)	61.29 ± 9.645	59.97 ± 13.268	58.86 ± 13.425	61.91 ± 10.262	0.589[a]
Sex					0.980[b]
Male	35 (67.31%)	25 (67.57%)	24 (68.57%)	20 (66.67%)	
Female	17 (32.69%)	12 (32.43%)	11 (31.43%)	10 (33.33%)	
Tumor location					0.083[b]
Proximal rectum	16 (30.77%)	12 (32.43%)	10 (28.57%)	10 (33.33%)	
Middle rectum	19 (36.54%)	15 (40.54%)	13 (37.14%)	12 (40.00%)	
Distal rectum	17 (32.69%)	10 (27.03%)	12 (34.19%)	8 (26.67%)	
Lymph node invasion					0.008[b]
Negative	44 (84.62%)	22 (59.46%)	26 (74.29%)	16 (53.33%)	
Positive	8 (15.38%)	15 (40.54%)	9 (25.71%)	14 (46.67%)	

[a]*Independent sample t-test was used;*
[b]*Chi-square test was used.*
Date are number (%) or mean ± Standard deviation.

TABLE 2 | Significant features between stage T1–2 and T3–4 tumors derived from HR-T2WI.

Feature	T stage		AUC	R_S	*P* value
	T1-2 (n = 52)	T3-4 (n = 37)			
Shape Flatness	0.444 ± 0.139	0.513 ± 0.131	0.659	0.247	0.020[a]
Wavelet HHH NGTDM Strength	0.788 ± 0.582	0.512 ± 0.309	0.713	-0.364	<0.001[b]
Log σ = 4.0mm 3D Firstorder Minimum	-76.35 ± 33.73	-90.52 ± 28.83	0.640	0.249	0.041[a]
Log σ = 3.0mm 3D GLSZM Large Area High Gray Level Emphasis	23925.256 ± 26104.701	47535.971 ± 65621.849	0.643	-0.249	0.018[b]
Log σ = 5.0mm 3D Firstorder Interquartile Range	54.943 ± 21.370	50.147 ± 12.872	0.646	-0.217	0.019[b]

[a]*Independent sample t-test was used, and data are the mean ± SD;*
[b]*Mann–Whitney U test was used, data are the medians ± interquartile range.*
AUC, area under the receiver operating characteristic curve; NGTDM, neighborhood gray-tone difference matrix; Log, Laplacian of Gaussian; GLSZM, gray level size zone matrix.

0.955) with a sensitivity of 81.08%, a specificity of 92.31%, and an accuracy of 89.86%. The combination model from HR-T2WI and ADC maps achieved an AUC of 0.941 (95% CI, 0.870-0.980) with a sensitivity of 89.19%, a specificity of 94.23%, and an accuracy of 89.89%. In validation cohorts, the model from HR-T2WI achieved an AUC of 0.845 (95% CI, 0.734-0.923) with a sensitivity of 76.67%, a specificity of 85.71%, and an accuracy of 78.46%. The model from ADC maps achieved an AUC of 0.881 (95% CI, 0.777-0.948) with a sensitivity of 83.33%, a specificity of 88.57%, and an accuracy of 83.08%. The combination model from HR-T2WI and ADC maps achieved an AUC of 0.910 (95% CI, 0.812-0.966) with a sensitivity of 90.00%, a specificity of 88.57%, and an accuracy of 87.69%. The ROC curves of LASSO_SVM models established by features extracted from HR-T2WI and ADC maps are shown in **Figure 5**.

DISCUSSION

In the present study, 3D VOI texture analysis was applied to rectal HR-T2WI and ADC maps, and the correlations between the whole tumor volume features and T stage were investigated. The results showed that texture features played a potential role in predicting the preoperative T stage of rectal carcinomas.

Preoperative staging are essential for treatment choices and prognosis evaluations of patients with rectal cancers. Especially for locally advanced patients, NCRT is required before accepting TME to ensure smooth progress of the operation and a better treatment response (5). In fact, the differential diagnosis of stages T2 and T3 is still difficult due to the perirectal desmoplastic fibrotic response (6). Texture analysis can describe the relationship between the gray level intensity of pixels and quantify the heterogeneity in images (18, 26, 27). And MRI-derived texture parameters have been proposed as tools for accurate diagnosis, preoperative risk stratification, or assessment of treatment response in several cancer types, including tumors in the brain, breast, prostate, and uterus (28–31). As reported in previous studies, high-resolution MRI has higher soft tissue resolution and can help in the diagnosis and stage-oriented treatment decisions (32–36). He et al. (25) suggested that MRI-based radiomic signatures showed acceptable performance for tumor grading of rectal carcinomas. Sun et al. (34) verified the feasibility of using

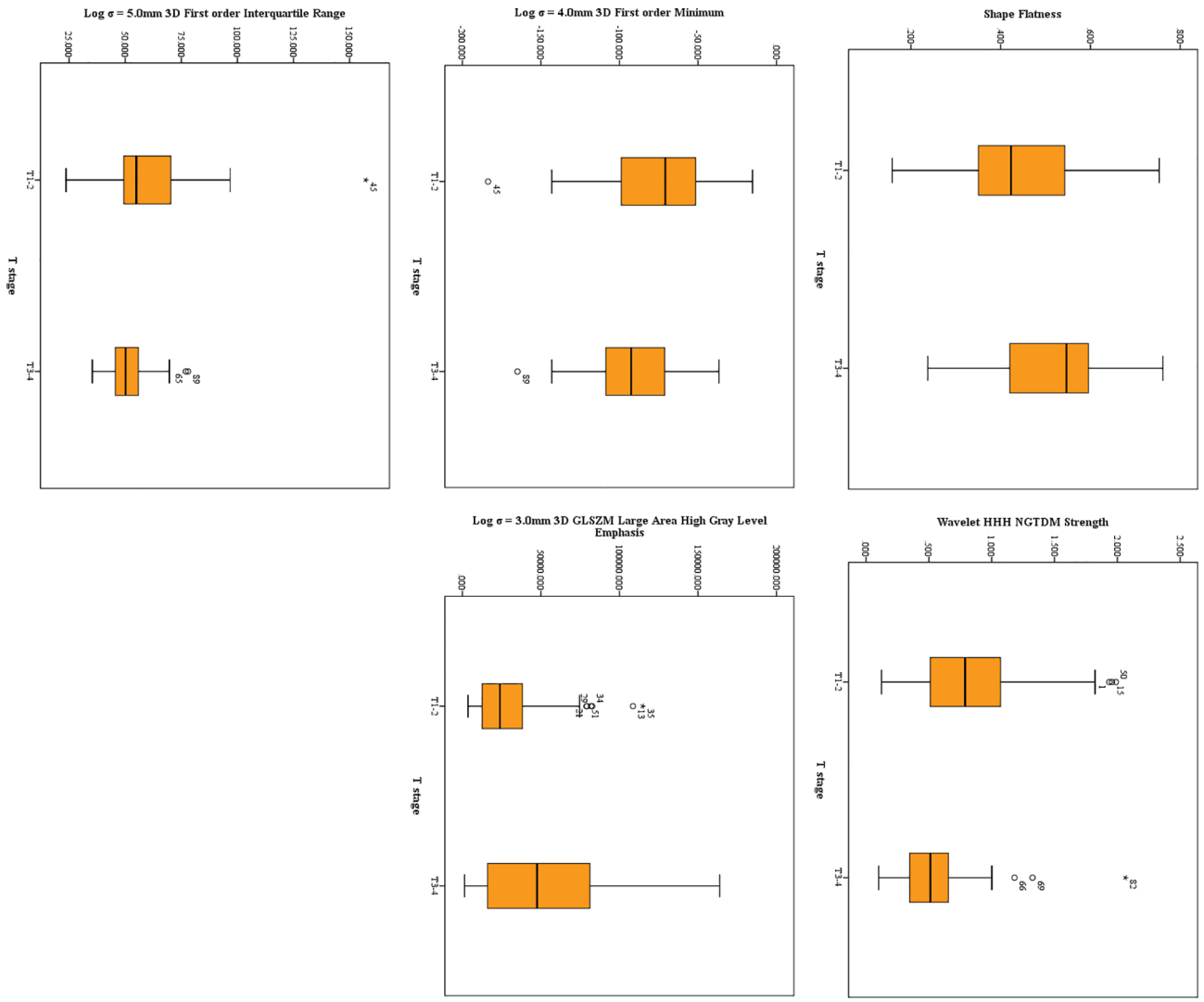


FIGURE 3 | The distribution of significant features derived from HR-T2WI. The symbol ("*") represents the extreme outlier.

radiomic features from T2WI images to identify the T staging of rectal cancer. Meanwhile, texture analysis of ADC maps was proven to have potential in cancer diagnosis, such as breast carcinoma (37), Cervical Carcinoma (38), Renal Carcinoma (39), gliomas (40). A study from Liu et al. (6) indicated that texture analysis on the features from the single slice of ADC map that showed the largest tumor dimension could provide valuable information in identifying locally advanced rectal cancer. However, few studies have focused on the 3D features extracted from whole volume lesion from HR-T2WI and ADC maps to distinguish the preoperative T stage of rectal cancers. Our study conducted the volume texture analysis based on HR-T2WI or ADC maps.

In this study, a total of 1226 texture features were calculated from preoperative HR-T2WI and ADC maps. These features were selected by LASSO to obtain optimum feature subsets, and statistical analysis was carried out to further reduce feature dimensionality. The classification models were established using the SVM method, which has been proven to have better performance and widely used in previous texture analysis reports (41). In training cohorts, the accuracy of the LASSO_SVM models based on HR-T2WI and ADC maps was 79.78% and 89.86% in predicting T stage of rectal cancers. The corresponding AUC values were 0.877 and 0.902, respectively. In validation cohorts, the accuracy of the LASSO_SVM models based on HR-T2WI and ADC maps was 78.46% and 83.08% in predicting T stage of rectal cancers. The corresponding AUC values were 0.845 and 0.881, respectively. The results demonstrated that pretreatment HR-T2WI and ADC maps had high potential application in identifying the preoperative T stage of rectal cancers, which is consistent with the results of Sun et al. (6, 34, 42). Meanwhile, the performance of the 3D features from ADC maps showed better than that of features

TABLE 3 | Significant features between stage T1–2 and T3–4 tumors derived from ADC maps.

Feature	T stage		AUC	R_S	*P* value
	T1-2 (n = 52)	T3-4 (n = 37)			
Shape Sphericity	0.600 ± 0.112	0.547 ± 0.106	0.641	-0.235	0.027[a]
Shape Maximum 2D Diameter Column	36.864 ± 7.143	42.638 ± 11.033	0.752	0.431	<0.001[b]
NGTDM Strength	1768.469 ± 631.326	1546.459 ± 600.620	0.647	-0.251	0.018[b]
Wavelet HLH First order Energy	6534962116 ± 2544354482	7850582575 ± 6464968925	0.623	0.210	0.049[b]
Wavelet LLL First order Range	86766.186 ± 14721.532	78306.717 ± 20030.935	0.614	-0.239	0.024[a]
Log σ = 2.0 mm 3D NGTDM Contrast	148.315 ± 272.076	81.591 ± 127.352	0.683	-0.313	0.003[b]
Log σ = 4.0 mm 3D Firstorder Skewness	-0.048 ± 0.318	-0.225 ± 0.409	0.598	-0.239	0.024[a]
Log σ = 5.0 mm 3D First order Maximum	3466.430 ± 1879.743	4534.088 ± 1810.633	0.659	0.276	0.009[a]
Log σ = 5.0 mm 3D First order Kurtosis	2.637 ± 0.579	2.811 ± 0.617	0.659	0.272	0.011[b]
Log σ = 5.0 mm 3D First order Interquartile Range	4557.574 ± 1560.537	4270.367 ± 1436.946	0.634	-0.229	0.032[b]

[a]*Independent sample t-test was used, and data are the mean ± SD;*
[b]*Mann–Whitney U test was used, data are the medians ± interquartile range.*
AUC, area under the receiver operating characteristic curve; NGTDM, neighborhood gray-tone difference matrix; Log, Laplacian of Gaussian.

from HR-T2WI. In the training cohort, the accuracy of the joint model of HR-T2WI and ADC maps was 89.89% with AUC of 0.941. In the validation cohort, the accuracy of the joint model was 87.69% with AUC of 0.910. It could be concluded that the combination model of HR-T2WI and ADC maps presented better performance than that using each of them alone. Cui et al. (43) established a radiomics predictive model based on pre-treatment multiparameter features from pre-CRT T2-weighted (T2-w), contrast-enhanced T1-weighted (cT1-w) and ADC maps and clinical features to predict a pathological complete response (pCR) in patients with locally advanced rectal cancer. They pointed that 3D features from ADC maps showed better performance than T2-w and cT1-w, and the radiomics signatures from joint T2-w, ADC and cT1-w images achieved better predictive efficacy than those from any of them alone. Their opinion was similar to ours.

There are controversies regarding the use of ADC values in predicting preoperative rectal T stages (6, 16, 44). In the present study, ADC quantification calculated from original images was not statistically significant between different T stages, which is consistent with the results of Liu et al. (6). However, the maximum of ADC value was statistically significant when a Log transform of σ = 5.0 mm was used. It may be because the conventional measurement of the ADC is only a mean value of signal intensity, and the heterogeneous intensity may be weakened due to the mutual influence of signals between various tissues. Compared with routine ADC measurements, texture analysis on ADC maps may therefore better predict local invasion of rectal cancers.

He et al. (25) indicated that not all categories of radiomics features contributed equally in the grading of rectal cancer. The transformed first-order features showed more significance than higher order features, and could diagnose and predict the grading of rectal carcinoma relatively, steadily and accurately (25). This point was similar to that of our study. After the original image was transformed by wavelet and Log, the parameters related to the lesion in the original image may be emphasized. Therefore, the transformed features may have better abilities to identify lesions than the original features, and were retained after screening. Cui et al. (45) develop a T2-weighted image-based radiomics signature for the individual prediction of KRAS mutation status in patients with rectal cancer. Most of the obtained features were from the images filtered by wavelet or log, and many of them appeared to be shape and first order with fewer higher order features. That point was similar to ours.

In addition, interobserver variability in feature calculations between two radiologists was evaluated. The values of ICCs ranging from 0.837 to 0.945, showed excellent agreement. The interobserver variability mainly occurred in delineating the lesion region slice by slice. Thus, it is important to reach a consensus between two radiologists in determining the tumor areas on HR-T2WI and ADC maps.

This study had some limitations. First, texture analysis was usually based on a large number of samples. Therefore, there may be potential biases because of the small sample in this study. Second, stage T2 and T3 accounted for a larger proportion among the enrolled cases, while stage T1 and T4 accounted for a smaller proportion. This study only focused on the differential diagnosis between early stage and local advanced stage, and divided patients into low (stages T1–2) and high (stages T3–4) T stage groups. More cases need to be collected, especially stage T1 and T4 tumors. Finally, whole tumor VOI delineation is a time-consuming operation. Two radiologists who were familiar with the operation of 3D Slicer spent an average of 15 minutes to outline and identify the whole lesion, which played a negative role in the progress and application of texture analysis.

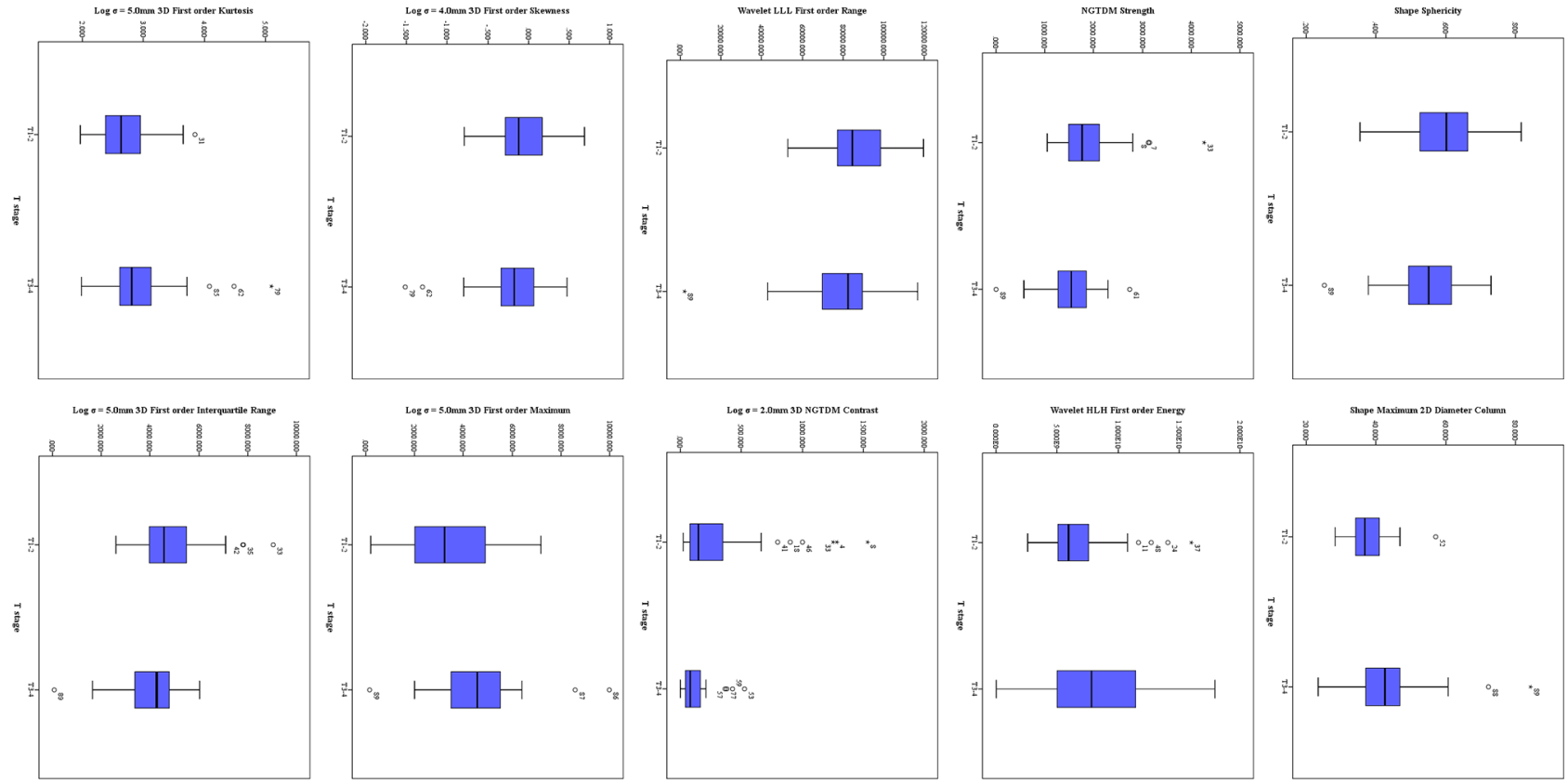


FIGURE 4 | The distribution of significant features derived from ADC maps. The symbol ("*") represents the extreme outlier.

TABLE 4 | Performance of classification models for identifying preoperative T stage of rectal cancer.

Methods		Cohorts	AUC	95% CI	Sensitivity (%)	Specificity (%)	Accuracy (%)
T stage	**HR-T2WI**	Training	0.877	0.791-0.937	86.54	78.38	79.78
		Validation	0.845	0.734-0.923	76.67	85.71	78.46
	ADC maps	Training	0.902	0.820-0.955	81.08	92.31	89.86
		Validation	0.881	0.777-0.948	83.33	88.57	83.08
	Combination	Training	0.941	0.870-0.980	89.19	94.23	89.89
		Validation	0.910	0.812-0.966	90.00	88.57	87.69

Combination means the joint application of classification models based on HR-T2WI and ADC maps.

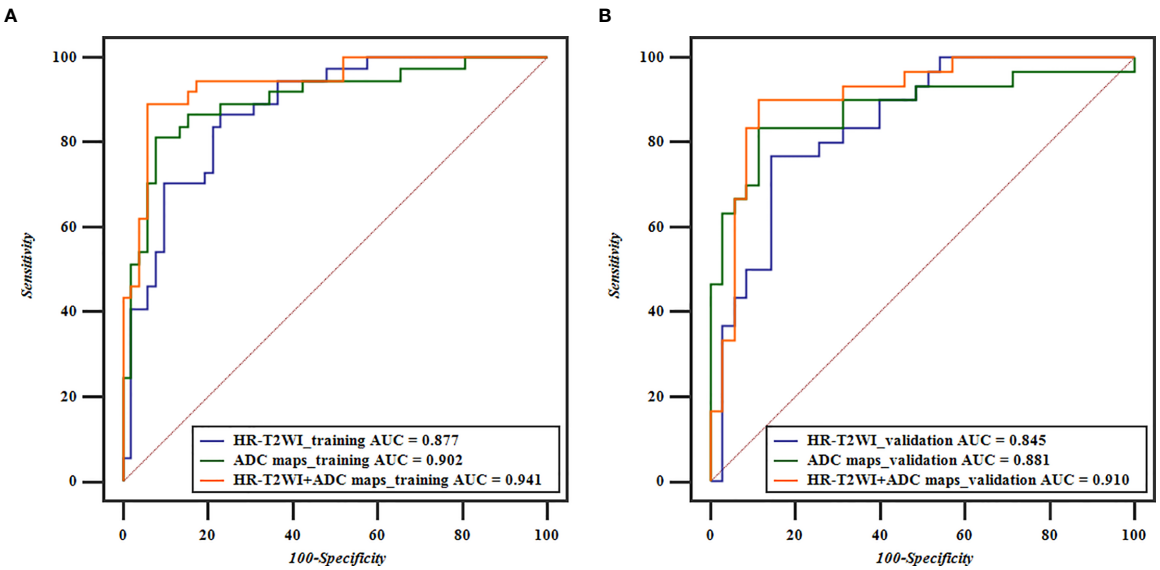


FIGURE 5 | ROC curves based on HR-T2WI and ADC maps for predicting preoperative T1-2 and T3-4 stage of rectal cancer. **(A)** The ROC curves of the training cohorts of T stage, and **(B)** the ROC curves of the validation cohorts of T stage.

In conclusion, texture features with LASSO_SVM models had good performance in predicting local invasion of rectal cancer. Texture analysis based on ADC maps was more potential than that based on HR-T2WI in identifying preoperative T stage in rectal carcinomas. The combined application of HR-T2WI and ADC maps could be used as an auxiliary diagnostic option for preoperative diagnosis of rectal cancer invasion.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

This study was approved by the Ethics Review Board of Shengjing Hospital of China Medical University (2020PS011K). Requirements for written informed consent were waived because of the retrospective nature of the study.

AUTHOR CONTRIBUTIONS

JY analyzed the patient data, and wrote the paper. JDY was a major contributor in designing the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by grants from the Research and Development (R&D) Foundation for Major Science and Technology from Shenyang (No. 19-112-4-105), the Big Data Foundation for Health Care from China Medical University (No. HMB201902105), the Natural Fund Guidance Plan from Liaoning (No. 2019-ZD-0743), and 345 talent project from Shengjing Hospital.

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2020. *CA Cancer J Clin* (2020) 70(1):7–30. doi: 10.3322/caac.21590
2. Glynne-Jones R, Wyrwicz L, Tiret E, Brown G, Rödel C, Cervantes A, et al. Rectal Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann Oncol* (2018) 29(Suppl 4):iv263. doi: 10.1093/annonc/mdy161
3. Saule B, Sajad AB, Arip K, Zhanna G, Olzhas U, Anar T, et al. Trends in Colorectal Cancer Incidence in Western Kazakhstan Through the First Decade of the Screening Implementation, 2009–2018. *J Coloproctol* (2020) 40(1):43–9. doi: 10.1016/j.jcol.2019.10.004
4. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A, et al. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68 (6):394–424. doi: 10.3322/caac.21492
5. Goldenberg BA, Holliday EB, Helewa RM, Singh H. Rectal Cancer in 2018: A Primer for the Gastroenterologist. *Am J Gastroenterol* (2018) 113(12):1763–71. doi: 10.1038/s41395-018-0180-y
6. Liu L, Liu Y, Xu L, Li Z, Lv H, Dong N, et al. Application of Texture Analysis Based on Apparent Diffusion Coefficient Maps in Discriminating Different Stages of Rectal Cancer. *J Magn Reson Imaging* (2017) 45(6):1798–808. doi: 10.1002/jmri.25460
7. Rania AES, Mohammad A, Mohammad EK. Preoperative MRI Evaluation of Mesorectum in Cases of Rectal Carcinoma. *J Menoufia Med J* (2017) 30 (1):122–7. doi: 10.4103/1110-2098.211484
8. Benson AB, Venook AP, Al-Hawary MM, Cederquist L, Chen YJ, Ciombor KK, et al. Nccn Guidelines Insights: Colon Cancer, Version 2.2018. *J Natl Compr Canc Netw* (2018) 16(4):359–69. doi: 10.6004/jnccn.2018.0021
9. Zhang XM, Zhang HL, Yu D, Dai Y, Bi D, Prince MR, et al. 3-T MRI of Rectal Carcinoma: Preoperative Diagnosis, Staging, and Planning of Sphincter-Sparing Surgery. *AJR Am J Roentgenol* (2008) 190(5):1271–8. doi: 10.2214/AJR.07.2505
10. Shaish H, Aukerman A, Vanguri R, Spinelli A, Armenta P, Jambawalikar S, et al. Radiomics of MRI for Pretreatment Prediction of Pathologic Complete Response, Tumor Regression Grade, and Neoadjuvant Rectal Score in Patients With Locally Advanced Rectal Cancer Undergoing Neoadjuvant Chemoradiation: An International Multicenter Study. *Eur Radiol* (2020) 30 (11):6263–73. doi: 10.1007/s00330-020-06968-6
11. Feng Q, Yan YQ, Zhu J, Xu JR. T Staging of Rectal Cancer: Accuracy of Diffusion-Weighted Imaging Compared With T2-Weighted Imaging on 3. *0 tesla MRI. J Dig Dis* (2014) 15(4):188–94. doi: 10.1111/1751-2980.12124
12. Chen XL, Chen GW, Pu H, Yin LL, Li ZL, Song B, et al. DWI and T2-Weighted MRI Volumetry in Resectable Rectal Cancer: Correlation With Lymphovascular Invasion and Lymph Node Metastases. *AJR Am J Roentgenol* (2019) 212(6):1271–8. doi: 10.2214/AJR.18.20564
13. Hoogendam JP, Kalleveen IM, de Castro CS, Raaijmakers AJ, Verheijen RH, van den Bosch MA, et al. High-Resolution T2-Weighted Cervical Cancer Imaging: A Feasibility Study on Ultra-High-Field 7.0-T MRI With an Endorectal Monopole Antenna. *Eur Radiol* (2017) 27(3):938–45. doi: 10.1007/s00330-016-4419-y
14. Kuhl CK, Träber F, Schild HH. Whole-Body High-Field-Strength (3.0-T) MR Imaging in Clinical Practice. Part I. Technical Considerations and Clinical Applications. *Radiology* (2008) 246(3):675–96. doi: 10.1148/radiol.2463060881
15. Iannicelli E, Di Pietropaolo M, Pilozzi E, Osti MF, Valentino M, Masoni L, et al. Value of Diffusion-Weighted MRI and Apparent Diffusion Coefficient Measurements for Predicting the Response of Locally Advanced Rectal Cancer to Neoadjuvant Chemoradiotherapy. *Abdom Radiol (NY)* (2016) 41 (10):1906–17. doi: 10.1007/s00261-016-0805-9
16. Curvo-Semedo L, Lambregts DM, Maas M, Beets GL, Caseiro-Alves F, Beets-Tan RG, et al. Diffusion-Weighted MRI in Rectal Cancer: Apparent Diffusion Coefficient as a Potential Noninvasive Marker of Tumor Aggressiveness. *J Magn Reson Imaging* (2012) 35(6):1365–71. doi: 10.1002/jmri.23589
17. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
18. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are More Than Pictures, They are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
19. Liu H, Zhang C, Wang L, Luo R, Li J, Zheng H, et al. MRI Radiomics Analysis for Predicting Preoperative Synchronous Distant Metastasis in Patients With Rectal Cancer. *Eur Radiol* (2019) 29(8):4418–26. doi: 10.1007/s00330-018-5802-7
20. Fu J, Fang MJ, Dong D, Li J, Sun YS, Tian J, et al. Heterogeneity of Metastatic Gastrointestinal Stromal Tumor on Texture Analysis: DWI Texture as Potential Biomarker of Overall Survival. *Eur J Radiol* (2020) 125:108825. doi: 10.1016/j.ejrad.2020.108825
21. Corino VDA, Montin E, Messina A, Casali PG, Gronchi A, Marchianò A, et al. Radiomic Analysis of Soft Tissues Sarcomas can Distinguish Intermediate From High-Grade Lesions. *J Magn Reson Imaging* (2018) 47(3):829–40. doi: 10.1002/jmri.25791
22. Meng Y, Zhang Y, Dong D, Li C, Liang X, Zhang C, et al. Novel Radiomic Signature as a Prognostic Biomarker for Locally Advanced Rectal Cancer. *J Magn Reson Imaging* (2018) 48(3):605–14. doi: 10.1002/jmri.25968
23. Zhang Y, Oikonomou A, Wong A, Haider MA, Khalvati F. Radiomics-Based Prognosis Analysis for non-Small Cell Lung Cancer. *Sci Rep* (2017) 7:46349. doi: 10.1038/srep46349
24. Yin JD, Song LR, Lu HC, Zheng X. Prediction of Different Stages of Rectal Cancer: Texture Analysis Based on Diffusion-Weighted Images and Apparent Diffusion Coefficient Maps. *World J Gastroenterol* (2020) 26(17):2082–96. doi: 10.3748/wjg.v26.i17.2082
25. He B, Ji T, Zhang H, Zhu Y, Shu R, Zhao W, et al. MRI-Based Radiomics Signature for Tumor Grading of Rectal Carcinoma Using Random Forest Model. *J Cell Physiol* (2019) 234(11):20501–9. doi: 10.1002/jcp.28650
26. Kudou M, Nakanishi M, Kuriu Y, Murayama Y, Arita T, Kishimoto M, et al. Value of Intra-Tumor Heterogeneity Evaluated by Diffusion-Weighted MRI for Predicting Pathological Stages and Therapeutic Responses to Chemoradiotherapy in Lower Rectal Cancer. *J Cancer* (2020) 11(1):168–76. doi: 10.7150/jca.38354
27. Li X, Zhu H, Qian X, Chen N, Lin X. MRI Texture Analysis for Differentiating Nonfunctional Pancreatic Neuroendocrine Neoplasms From Solid Pseudopapillary Neoplasms of the Pancreas. *Acad Radiol* (2020) 27(6):815–23. doi: 10.1016/j.acra.2019.07.012
28. Han Y, Wang T, Wu P, Zhang H, Chen H, Yang C. Meningiomas: Preoperative Predictive Histopathological Grading Based on Radiomics of MRI. *Magn Reson Imaging* (2021) 77:36–43. doi: 10.1016/j.mri.2020.11.009
29. Ahmed A, Gibbs P, Pickles M, Turnbull L. Texture Analysis in Assessment and Prediction of Chemotherapy Response in Breast Cancer. *J Magn Reson Imaging* (2013) 38(1):89–101. doi: 10.1002/jmri.23971
30. Wibmer A, Hricak H, Gondo T, Matsumoto K, Veeraraghavan H, Fehr D, et al. Haralick Texture Analysis of Prostate MRI: Utility for Differentiating non-Cancerous Prostate From Prostate Cancer and Differentiating Prostate Cancers With Different Gleason Scores. *Eur Radiol* (2015) 25(10):2840–50. doi: 10.1007/s00330-015-3701-8
31. Ytre-Hauge S, Dybvik JA, Lundervold A, Salvesen ØO, Krakstad C, Fasmer KE, et al. Preoperative Tumor Texture Analysis on MRI Predicts High-Risk Disease and Reduced Survival in Endometrial Cancer. *J Magn Reson Imaging* (2018) 48 (6):1637–47. doi: 10.1002/jmri.26184
32. Tudyka V, Blomqvist L, Beets-Tan RG, Boelens PG, Valentini V, van de Velde CJ, et al. EURECCA Consensus Conference Highlights About Colon & Rectal Cancer Multidisciplinary Management: The Radiology Experts Review. *Eur J Surg Oncol* (2014) 40(4):469–75. doi: 10.1016/j.ejso.2013.10.029
33. KSAR Study Group for Rectal Cancer. Essential Items for Structured Reporting of Rectal Cancer MRI: 2016 Consensus Recommendation From the Korean Society of Abdominal Radiology. *Korean J Radiol* (2017) 18 (1):132–51. doi: 10.3348/kjr.2017.18.1.132
34. Sun Y, Hu P, Wang J, Shen L, Xia F, Qing G, et al. Radiomic Features of Pretreatment MRI Could Identify T Stage in Patients With Rectal Cancer: Preliminary Findings. *J Magn Reson Imaging* (2018) 48(3):615–21. doi: 10.1002/jmri.25969
35. Just N. Improving Tumour Heterogeneity MRI Assessment With Histograms. *Br J Cancer* (2014) 111(12):2205–13. doi: 10.1038/bjc.2014.512
36. Choudhery S, Gomez-Cardona D, Favazza CP, Hoskin TL, Haddad TC, Goetz MP, et al. MRI Radiomics for Assessment of Molecular Subtype, Pathological Complete Response, and Residual Cancer Burden in Breast Cancer Patients Treated With Neoadjuvant Chemotherapy. *Acad Radiol* (2020) S1076-6332 (20):30607–3. doi: 10.1016/j.acra.2020.10.020

37. Zhao Q, Xie T, Fu C, Chen L, Bai Q, Grimm R, et al. Differentiation Between Idiopathic Granulomatous Mastitis and Invasive Breast Carcinoma, Both Presenting With non-Mass Enhancement Without Rim-Enhanced Masses: The Value of Whole-Lesion Histogram and Texture Analysis Using Apparent Diffusion Coefficient. *Eur J Radiol* (2020) 123:108782. doi: 10.1016/j.ejrad.2019.108782
38. Yamada I, Oshima N, Miyasaka N, Wakana K, Wakabayashi A, Sakamoto J, et al. Texture Analysis of Apparent Diffusion Coefficient Maps in Cervical Carcinoma: Correlation With Histopathologic Findings and Prognosis. *Radiol Imaging Cancer* (2020) 2(3):e190085. doi: 10.1148/rycan.2020190085
39. Li H, Li A, Zhu H, Hu Y, Li J, Xia L, et al. Whole-Tumor Quantitative Apparent Diffusion Coefficient Histogram and Texture Analysis to Differentiation of Minimal Fat Angiomyolipoma From Clear Cell Renal Cell Carcinoma. *Acad Radiol* (2019) 26(5):632–9. doi: 10.1016/j.acra.2018.06.015
40. Kanazawa T, Minami Y, Takahashi H, Fujiwara H, Toda M, Jinzaki M, et al. Magnetic Resonance Imaging Texture Analyses in Lower-Grade Gliomas With a Commercially Available Software: Correlation of Apparent Diffusion Coefficient and T2 Skewness With 1p/19q Codeletion. *Neurosurg Rev* (2020) 43(4):1211–9. doi: 10.1007/s10143-019-01157-6
41. Qian Z, Li Y, Wang Y, Li L, Li R, Wang K, et al. Differentiation of Glioblastoma From Solitary Brain Metastases Using Radiomic Machine-Learning Classifiers. *Cancer Lett* (2019) 451:128–35. doi: 10.1016/j.canlet.2019.02.054
42. Ma X, Shen F, Jia Y, Xia Y, Li Q, Lu J. MRI-Based Radiomics of Rectal Cancer: Preoperative Assessment of the Pathological Features. *BMC Med Imaging* (2019) 19(1):86. doi: 10.1186/s12880-019-0392-7
43. Cui Y, Yang X, Shi Z, Yang Z, Du X, Zhao Z, et al. Radiomics Analysis of Multiparametric MRI for Prediction of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Eur Radiol* (2019) 29(3):1211–20. doi: 10.1007/s00330-018-5683-9
44. Sun Y, Tong T, Cai S, Bi R, Xin C, Gu Y. Apparent Diffusion Coefficient (ADC) Value: A Potential Imaging Biomarker That Reflects the Biological Features of Rectal Cancer. *PLoS One* (2014) 9(10):e109371. doi: 10.1371/journal.pone.0109371
45. Cui Y, Liu H, Ren J, Du X, Xin L, Li D, et al. Development and Validation of a MRI-Based Radiomics Signature for Prediction of KRAS Mutation in Rectal Cancer. *Eur Radiol* (2020) 30(4):1948–58. doi: 10.1007/s00330-019-06572-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 You and Yin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 18 August 2021
doi: 10.3389/fonc.2021.724191

Overall Survival Prediction for Gliomas Using a Novel Compound Approach

He Huang†, *Wenbo Zhang*†, *Ying Fang*, *Jialing Hong*, *Shuaixi Su* and *Xiaobo Lai**

School of Medical Technology and Information Engineering, Zhejiang Chinese Medical University, Hangzhou, China

As a highly malignant tumor, the incidence and mortality of glioma are not optimistic. Predicting the survival time of patients with glioma by extracting the feature information from gliomas is beneficial for doctors to develop more targeted treatments. Magnetic resonance imaging (MRI) is a way to quickly and clearly capture the details of brain tissue. However, manually segmenting brain tumors from MRI will cost doctors a lot of energy, and doctors can only vaguely estimate the survival time of glioma patients, which are not conducive to the formulation of treatment plans. Therefore, automatically segmenting brain tumors and accurately predicting survival time has important significance. In this article, we first propose the NLSE-VNet model, which integrates the Non-Local module and the Squeeze-and-Excitation module into V-Net to segment three brain tumor sub-regions in multimodal MRI. Then extract the intensity, texture, wavelet, shape and other radiological features from the tumor area, and use the CNN network to extract the deep features. The factor analysis method is used to reduce the dimensionality of features, and finally the dimensionality-reduced features and clinical features such as age and tumor grade are combined into the random forest regression model to predict survival. We evaluate the effect on the BraTS 2019 and BraTS 2020 datasets. The average Dice of brain tumor segmentation tasks up to 79% and the average RMSE of the survival predictive task is as low as 311.5. The results indicate that the method in this paper has great advantages in segmentation and survival prediction of gliomas.

Keywords: automatic segmentation, deep learning, gliomas, magnetic resonance imaging, overall survival prediction

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln, United Kingdom

Reviewed by:
Sikandar Shaikh,
Shadan Hospital and Institute of Medical Sciences, India
Vinay Ayyappan,
University of Pennsylvania, United States

****Correspondence:***
Xiaobo Lai
dmia_lab@zcmu.edu.cn

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *12 June 2021*
Accepted: *26 July 2021*
Published: *18 August 2021*

Citation:
Huang H, Zhang W, Fang Y, Hong J, Su S and Lai X (2021) Overall Survival Prediction for Gliomas Using a Novel Compound Approach. Front. Oncol. 11:724191. doi: 10.3389/fonc.2021.724191

INTRODUCTION

Gliomas are the most common primary malignant brain tumors (1). The incidence of primary intracranial tumors is 23 per 10 million, and gliomas account for about 60% (2). According to the degree of malignancy, the World Health Organization divides gliomas into low-grade gliomas (LGG) and high-grade gliomas (HGG). Different grades of gliomas have different levels of invasion and variable prognosis, which seriously threatens human health. The best treatment is complete

surgical resection, however, due to the unresectable nature of normal brain tissue and the widespread infiltration of malignant tumors into the brain, complete resection surgery is extremely difficult (3). Therefore, early detection of tumors and targeted treatment play a vital role in prolonging the survival time of patients.

Magnetic resonance imaging (MRI) is one of the most common methods to obtain brain tumor images, it usually has multiple modes, namely fluid attenuation inversion recovery (Flair), T1 weighted (T1), T1 weighted contrast enhancement (T1-CE) and T2 weighted (T2) (4). Medical image analysis usually needs to segment the tumor first, but the manual segmentation method often depends on the doctor's experience, knowledge and emotions, and the efficiency is low. Moreover, doctors can only observe image slices in a fixed manner, and cannot directly extract feature information in the segmentation results to quantify the diagnosis results, which makes the diagnosis have subjective experience. The development of machine learning and deep learning can change this situation. Through image processing technology, the computer can accurately and quantitatively analyze the tumor area, and make efficient and accurate survival prediction for patients with brain tumors, so as to formulate personalized diagnosis and treatment plans for patients.

Up to now, the overall survival prediction of glioma patients with multimodal MRI has received widespread attention (5). It is observed that most survival prediction models proposed in the literature are based on radiomics (6). Although radiological features extracted from images can be used to predict tumor grade or molecular biomarkers (7, 8), it is difficult to accurately predict survival time without considering other factors such as age. In this article, we propose an automatic prediction method by combining radioactive features, deep features and clinical features to provide accurate survival predictions for patients with glioma. The major contributions of our work are four folds that can be summarized as follows:

- We improved the V-Net and proposed NLSE-VNet to segment brain tumors. The innovation of the NLSE-VNet model lies in porting the Non-Local (NL) module and Squeeze-and-Excitation (SE) module to the V-Net network structure. These attention models can enhance feature extraction capabilities. We have designed multiple ablation experiments to prove that NLSE-VNet can greatly improve the accuracy of brain tumor segmentation.
- We extracted various types of radiological features such as intensity, texture, wavelet, etc., and designed a CNN network to extract deep features. After cross-validation and a large number of comparative experiments, it is proved that when radiological features, deep features and clinical features are combined, the effect of predicting survival is the best.
- We perform three-dimensional reconstruction of the segmentation results to provide clinicians with a visual display. At the same time, in order to improve the interpretability of feature dimensionality reduction, we draw feature heat maps to show how the model finds meaningful features, which provides a reliable basis for the clinical application of survival prediction.

The remainder of this paper is organized as follows. The second section describes the previous research on MRI segmentation and survival prediction of glioma. The third part introduces the method proposed in this paper in detail, including data preprocessing, model description and parameter setting. *Dataset and Experiments* introduces the data set, evaluation indicators and experimental configuration. Then, in *Experimental Results*, the experimental results are presented and discussed and analyzed. Finally, conclusions are drawn in *Conclusion*.

RELATED WORKS

Brain Tumor Segmentation

The segmentation of the tumor area is a prerequisite for prediction. At present, brain tumor segmentation has received extensive attention and in-depth research, and various excellent models have been proposed for segmenting brain tumors. Zhou et al. improved the model cascade strategy and proposed a single multi-task network (OM-Net) (9), which can solve the problem of category imbalance. OM-Net integrates the separated segmentation tasks into a deep model, which is shared by Parameters to learn joint features and task-specific parameters to learn discriminative features. At the same time, they designed a CGA attention module that can adaptively recalibrate the characteristics of the channel direction. S. Pereira et al. designed a deep-level architecture based on a convolutional neural network using a small 3×3 kernel (1). The model has a positive effect on overfitting when the network weight is less. In the training process, the number of LGG classes is increased by rotating training patches and using HGG samples, and the number of training patches is artificially increased. Chen et al. established a model based on 3D convolutional neural network to segment brain tumors (10). The model obtains multi-scale context information by extracting the features of two scales of the receptive domain, and they use hierarchical segmentation to segment different lesion areas such as necrotic and non-enhanced tumors, peritumoral edema, and enhanced tumors, using densely connected The convolution block further improves performance. Sun et al. proposed an anatomical attention-guided deep learning framework for segmenting brain tumors (11). It contains two sub-networks, one is the segmentation sub-network, and the other is the anatomical attention sub-network, so as to combine the anatomical structure information of the brain with The feature information in the segmentation process is combined to improve performance. Lachinov et al. proposed an automatic segmentation algorithm for brain tumors based on deep cascades (12). Their team modified the 3D U-Net architecture and designed 4 down-sampling paths to extract the features of the four modalities of brain tumors. The model can effectively process input images on multiple scales at the same time and extract features of specific scales.

Overall Survival Prediction

Overall survival prediction of cancer patients has also been a hot research topic in recent years. Sun et al. extracted 4,524 radiomic features from the segmented area of the tumor, then used decision trees and cross-validation to select effective features, and finally trained a random forest model to predict the survival of patients (13). Shboul et al. extracted about 31,000 features from the tumor area, representing texture, volume, area, and Euler features. Then they performed recursive feature selection on Euler features separately, and finally used XGBoost to predict survival (14). Baid et al. calculated the first-order statistics, shape features, gray-level co-occurrence matrix and gray-level run length matrix for a total of 679 features (15). The radiation group variables with Spearman correlation coefficients of 0.95 and above that are nearly completely correlated were excluded, and the features were reduced to 56 dimensions, using multilayer perceptrons to train the neural network to predict the number of survival days. Kim et al. extracted a total of 6472 radiological features from multi-mode MR images, applied the Least Absolute Shrinkage and Selection Operator (LASSO) to the training data set to select significant non-zero coefficient radiation features, and constructed a radiation group using a generalized linear model to predict the survival period with a scientific model (16). Weninger et al. used the volume characteristics of all regions of interest, the distance from the brain to the center of mass of the tumor, and input the age into the linear regression model when predicting survival (17). They found that using only the "age" feature to train the regression model achieved higher accuracy on the test set. Banerjee et al. extracted two types of radiological features, namely "semantic" and "agnostic" (18). The former includes attributes such as size, shape, and location, while the latter uses histograms, textures, and other quantitative descriptions to capture the heterogeneity of lesions. A total of 83 features were extracted as the input of the multi-layer perceptron to predict the number of survival days. Wang et al. used internal radiomics analysis software to extract 43 unique quantitative features in 4 categories, selected features with high r values in related tests, used support vector regression SVR to predict OS, and used leave-one-out cross-validation (LOOCV) (19).

Our Work

Although computer technology has made significant progress in the field of brain tumor segmentation and survival prediction, there are still some challenges that prevent this fully automated technology from being well applied in clinical practice. First, the size, location, and shape of brain tumors vary from patient to patient (20). Secondly, the lesion area in MRI is very small in most cases, which leads to the voxel imbalance between the lesion area and the background area (10). The above challenges all increase the difficulty of segmentation. Finally, the information obtained from other data sources, such as genes and age, is usually not used when extracting radiomic features, which further limits the ability to distinguish predictions (21). Our goal is to innovate existing methods to improve the accuracy of segmentation and prediction on the basis of predecessors.

In this article, we first propose a brain tumor segmentation model NLSE-VNet, which is an improvement on the V-Net network structure. We transplant the Squeeze-and-Excitation (SE) module to the front of each down/up sampling layer of V-Net (22). It can clarify the interdependence between channels, and at the same time, the Non-Local (NL) modules are transplanted after the last encoder block in the network to capture long-term dependencies (23). The comparative experiment proves the validity of the model. Secondly, we use the Pyradiomics toolkit to extract radiological features such as intensity, texture, and filtering from the original and derived images, and designed a CNN network to extract deep features. Since most of the extracted features are redundant, we consider reducing the dimensionality of the features. After several experiments, the factor analysis method (FA) was used to reduce the feature dimension. Finally, a random forest model was constructed to predict the survival time by using the features, age, and tumor grade after dimensionality reduction. We have implemented a fully automatic method from brain tumor segmentation to survival prediction. The experimental results show that our proposed method has great potential in clinical application.

METHODS

The task of this article is to automatically segment brain tumors and predict the survival time of glioma patients. All the workflow of this project is shown in **Figure 1**, which is divided into five parts: brain tumor segmentation, feature extraction, feature dimensionality reduction, radiology model selection and model evaluation.

Glioma Segmentation Model

The basic idea of the NLSE-VNet model is to achieve feature information extraction of images of different resolutions through alternating convolutional layers and down-sampling layers, and then use the features extracted by the up-sampling layer joint encoder to gradually achieve resolution restoration. The SE module is placed in front of the down-sampling layer and the up-sampling layer, and the NL module is placed behind the last down-sampling layer. These attention modules bring a significant improvement in segmentation accuracy while slightly increasing the computational cost. Its network structure is shown as in **Figure 2**. We will introduce from four aspects: preprocessing process, network structure, loss function and model training.

Data Preprocessing and Data Augmentation

In addition, we use the Z-score method to standardize the images of the four modalities, and then merge the standardized images as input to the model. Z-score is the image minus the mean divided by the standard deviation; its mathematical formula is as follows:

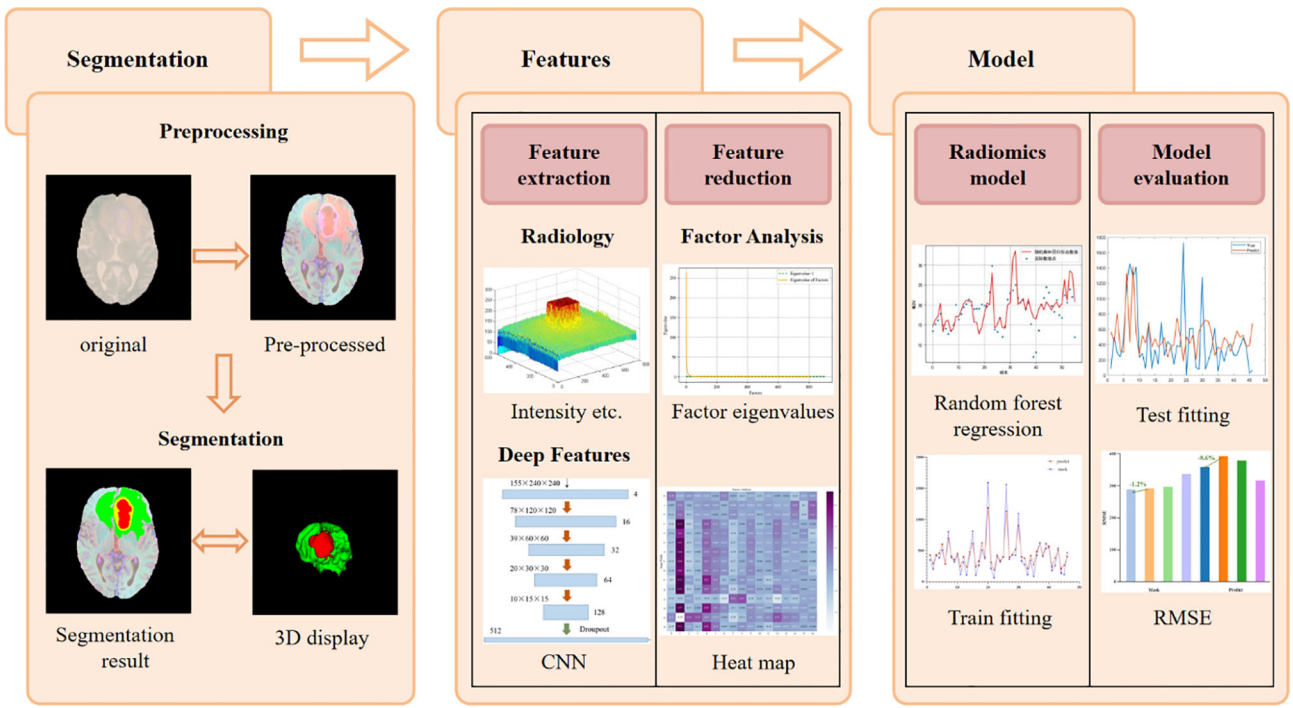


FIGURE 1 | Flow chart of brain tumor segmentation and overall survival prediction.

$$Z = \frac{X - \overline{X}}{s} \tag{1}$$

where Z represents the normalized image, X represents the original image, $\overline{X}$ represents the pixel average, and s represents the pixel standard deviation. **Figure 3** shows the comparison images before and after preprocessing. The first four columns are the comparison before and after the pre-processing of the four modes, Flair, T1, T1-CE, and T2, respectively, and the last column is the comparison after combining the four modes. It can be seen from **Figure 3** that after preprocessing, the contrast of the tumor part is significantly enhanced compared to normal tissue, which facilitates image segmentation.

The data enhancement method we use is patch. Combine the preprocessed MRI of the four modalities to generate a three-dimensional image with four channels, and then divide the original image and the mask into multiple blocks. One case will generate 175 pictures with a size of 128×128×64×4. This makes it possible to process only one patch instead of the entire image, thereby better detecting edge features.

Network Architecture

On the basis of the VNet network, we introduced the Squeeze-and-Excitation (SE) module and the Non-Local (NL) module. The 3D SE module uses feature recalibration to explicitly model the interdependence between feature channels, that is, the importance of each feature channel is automatically obtained through learning. The specific process of the SE block module is as follows, the input is $X \in \mathbb{R}^{Z \times H \times W \times C}$, where Z is the depth, H is the height, W is the width, and C is the number of channels. Then the input image will undergo a global average pooling, called squeeze operation, the formula is as (2):

$$z_c = F_{sq}(x_c) = \frac{1}{Z \times H \times W} \sum_{i=1}^{Z} \sum_{j=1}^{H} \sum_{k=1}^{W} x_c(i,j,k) \tag{2}$$

The excitation operation is used to utilize the information aggregated in the squeeze operation, and the excitation part is composed of two fully connected layers. The first full connection compresses the C channel into a C/r channel to reduce the amount of calculation, and the second full connection returns to the C channel. r is the compression ratio. The calculation of the excitation part is shown in (3):

$$s = F_{ex}(z,\ W) = \sigma\left(g\left(z,\ W\right)\right) = \sigma\left(W_2\ \delta\left(W_1\ z\right)\right) \tag{3}$$

where δ represents the ReLU activation function, $W_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $W_2 \in \mathbb{R}^{C \times \frac{C}{r}}$. Finally, by re-calibrating X by activatings, the final output of the block can be obtained as (4):

$$\hat{x}_c = F_{scale}(x_c,\ s_c) = s_c x_c \tag{4}$$

where $\hat{X} = [\hat{x}_1,\ \hat{x}_2, \ldots, \hat{x}_C]$ and $F_{scale}(x_c,\ s_c)$ refers to channel-wise multiplication. The structure of SE module is shown as in **Figure 4**.

At the same time, in order to quickly capture remote dependencies and improve computational efficiency, we have also introduced a 3D Non-Local module, which is integrated after the model downsampling stage. Its structure is shown in **Figure 5**.

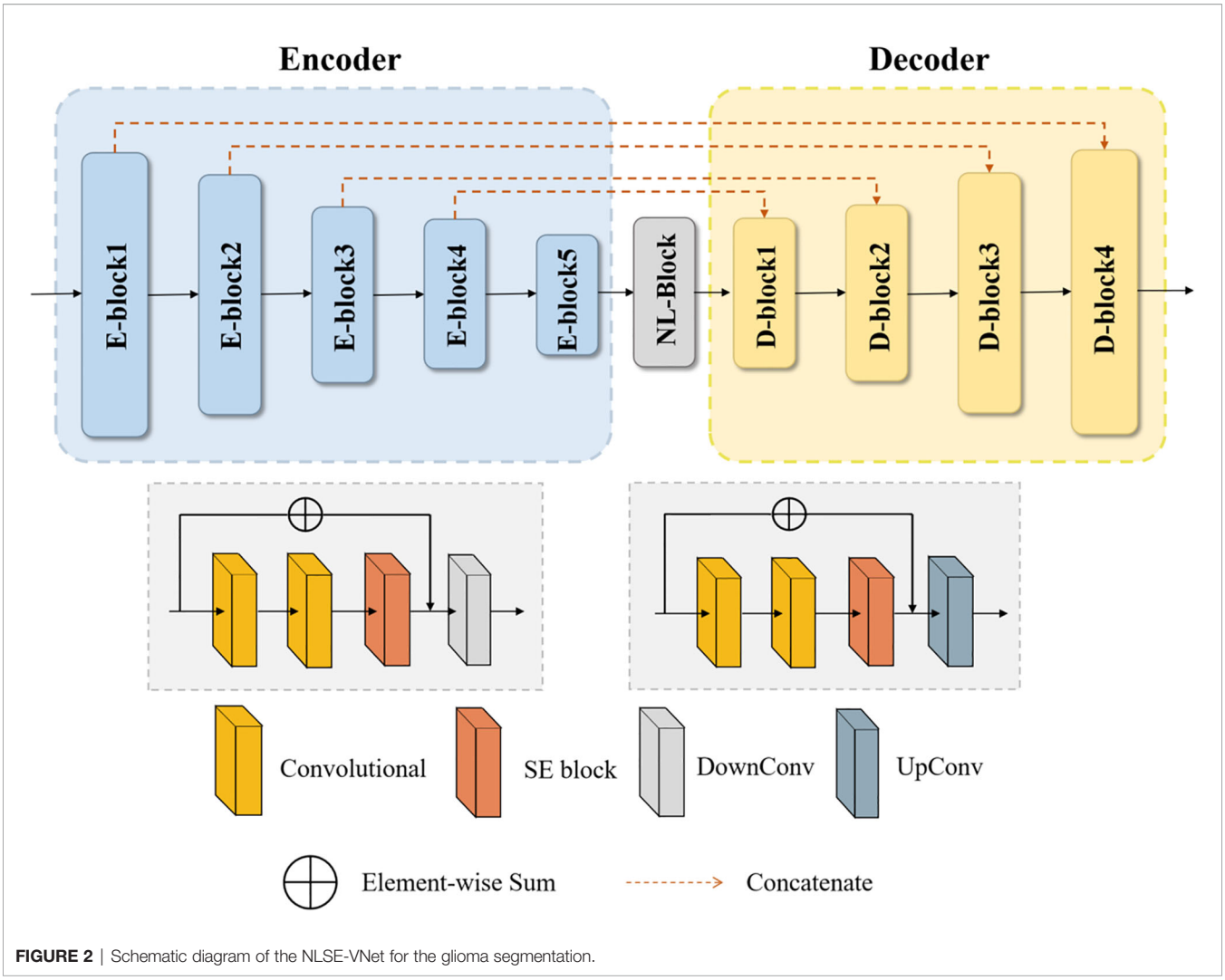


FIGURE 2 | Schematic diagram of the NLSE-VNet for the glioma segmentation.

First of all, the network input $X = (Z, H, W, 320)$, after the weight matrix W_θ, W_ϕ transform respectively, use the 1×1×1 size convolution kernel to perform convolution operation to reduce the number of channels. Then reshape the two outputs to ZHW×160, then perform matrix multiplication and perform softmax processing. At the same time, the weight matrix W_g is performed on X again, and the convolution operation of 1×1×1 is used. Perform a matrix multiplication of this result with the output result of softmax in the previous step. Finally, the output channel is restored through a 1×1×1 convolution operation to ensure that the input and output sizes are exactly the same.

Loss Function

Considering that the tumor area we want to segment only occupies a small part of the entire scanning area, and the proportion of foreground and background area is extremely unbalanced, we choose Categorical Dice as the loss function of the model to prevent the prediction from being strongly biased towards the background area that we are not interested in. The Categorical Dice is an improvement based on the Generalized Dice loss function (GDL) proposed by Sudre C.H. et al. (24). The generalized dice loss function has been proved to be able to effectively solve the problem of brain tumor imbalance. Its calculation formula is:

$$\mathrm{GDL} = 1 - 2\frac{\sum_{l=1}^{2}\omega_l\sum_{i=1}^{N}p_{\mathrm{ln}}g_{\mathrm{ln}}}{\sum_{l=1}^{2}\omega_l\sum_{i=1}^{N}p_{\mathrm{ln}} + g_{\mathrm{ln}}} \tag{5}$$

The weight is defined as $\omega_l = 1/(\sum_{n=1}^{N}g_{\mathrm{ln}})^2$. Where N represents all voxels, l represents the number of categories, p represents the predicted voxel, g represents the real voxel. And we assign the weights of different categories. We set the weight of the background area to 0.1, and the weight of the gangrene, edema and enhanced tumor area to 1.0, and the value of ω is [0.1, 1.0, 1.0, 1.0]. Through this weight distribution, the problem that the weight assigned to the background area tends to 0 when there are too many voxels is avoided. The calculation formula is shown in (6):

$$\mathrm{Loss} = -2\frac{\sum_{l=1}^{4}\omega_l\sum_{i=1}^{N}p_{\mathrm{ln}}g_{\mathrm{ln}}}{\sum_{l=1}^{4}\omega_l\sum_{i=1}^{N}p_{\mathrm{ln}} + g_{\mathrm{ln}}} \tag{6}$$

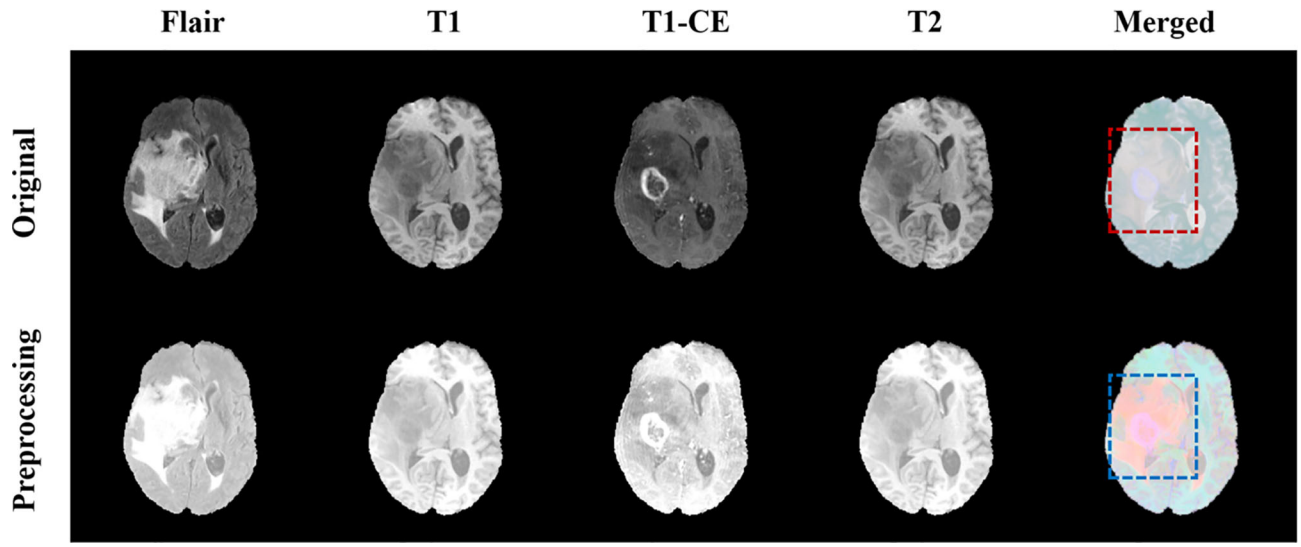


FIGURE 3 | Comparison of MRI before and after pretreatment. The first line is the original image, the second line is the preprocessed image. The first four columns are the images of the four modalities Flair, T1, T1-CE, and T2, and the last column is the merged image.

Model Training

The realization of the segmentation task is based on tensorflow 1.13.1. In addition, we used the Adam optimizer to train the model. The entire network was trained with a total of 500,000 steps and the training set was traversed 10 times. After each traversal of the training set, the order of the data will be randomly shuffled to enhance the robustness of training. The initial learning rate is set to 0.0001, which is reduced to half of the original learning rate after traversing the training set twice. Finally, Mean Dice is used as the evaluation index of training, and the loss value and accuracy index are output every ten steps to realize effective supervision of model training. At the same time, the parameter model is saved every 1000 steps.

The experimental environment is run on TensorFlow. The runtime platform processor is Intel (R) Xeon (R) Silver 4210 CPU @2.20GHz 2.20GHz 128GB RAM, Nvidia Titan RTX, 64-bit Windows10.The development software platform is PyCharm with Python 3.6.

Overall Survival Prediction

In the survival prediction task, we extract radiological features such as intensity, texture, and filtering, and then build a CNN network to extract deep features. This network can also be used to predict survival. Subsequently, factor analysis is used to reduce the dimensions of the above two types of features to remove redundant features. Finally, the dimensionality reduction features combined with clinical factors such as age and tumor

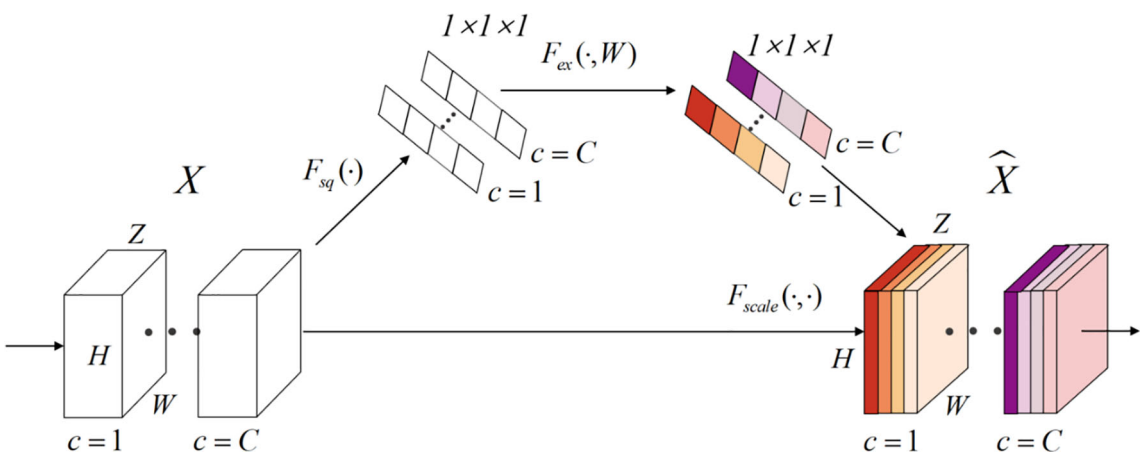


FIGURE 4 | Structure of Squeeze-and-Excitation module (SE).

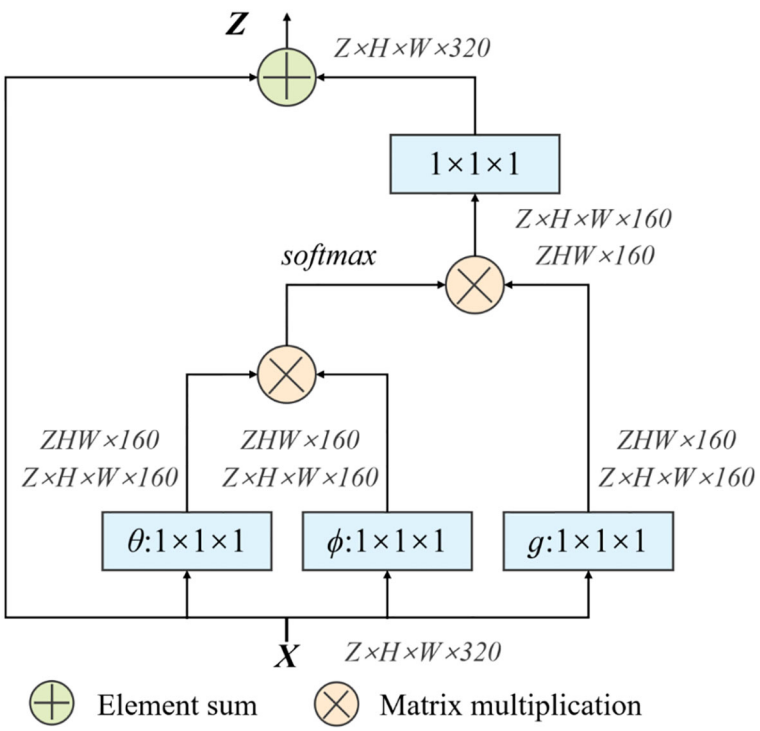


FIGURE 5 | Structure of Non-Local module (NL).

grade are input into the random forest regression model for survival prediction. The entire flow chart is shown in **Figure 6**.

Radiological Feature Extraction

Based on the segmentation results, we use the Pyradiomics toolbox to extract the radiomic features of edema, non-enhanced verification, and necrosis/cystic nucleus areas. We mainly extract three types of features: intensity, texture, and wavelet, as shown in **Figure 7**. Pyradiomics is an open source Python software package that can extract the features of Radiomics from medical images (25).

Then, we further subdivide the extracted radiological features into 7 categories. The first-order statistical features reflect the voxel intensity distribution in the image area defined by the mask. The shape is based on 3D shape features, including a series of tumor shape features, such as sphericity, circumference ratio, spindle length, and elongation. The Gray Level Co-occurrence Matrix (GLCM) defines information about correlation, energy, contrast, deficiency, variance, probability, entropy, sum of squares, etc. In addition, we extracted 16 features of the gray run length matrix (GLRLM), 16 features of the gray size band matrix (GLSZM), 14 features of the gray dependency matrix (GLDM) and neighbor gray tone difference Five features of the matrix (NGTDM). **Table 1** shows the detailed information of feature categories.

We not only extract features from the original image, but also extract the same features from the image after wavelet decomposition. Wavelet decomposition can segment the image into multiple levels of detail components. In the end, we extracted a total of 2500 radiological features.

Deep learning has been used to predict the survival of patients with brain tumors. We tried to build a CNN network for the survival regression task, as shown in **Figure 6**. With the help of the segmentation result, we set the MR sequence to retain only part of the tumor information, and set other pixels to 0 as input. The CNN network consists of four convolutions with a step size of 2 and three fully connected layers. The last fully connected layers are directly used to predict survival time. The model can either extract deep features or directly use neural networks to predict survival days. In the end, we extracted 512 deep features. Since the CNN network can also learn the shape, texture of the brain tumor, these features are the same as part of the radiology features, we confirmed this in *Overall Survival Prediction Results*. So we believe that the deep features and radiological features should perform feature reduced together, which helps filter out repeat features.

Feature Dimensionality Reduction

Some of the features we extract are redundant or have nothing to do with survival prediction, which will increase the degree of model overfitting, here we use factor analysis to filter features. The core of factor analysis is to analyze a series of features and extract common factors to achieve the purpose of reducing feature dimensions. First, the feature is used as a factor, and the feature value of the factor is calculated. The factor with the eigenvalues greater than 1 can be used as the subsequent feature

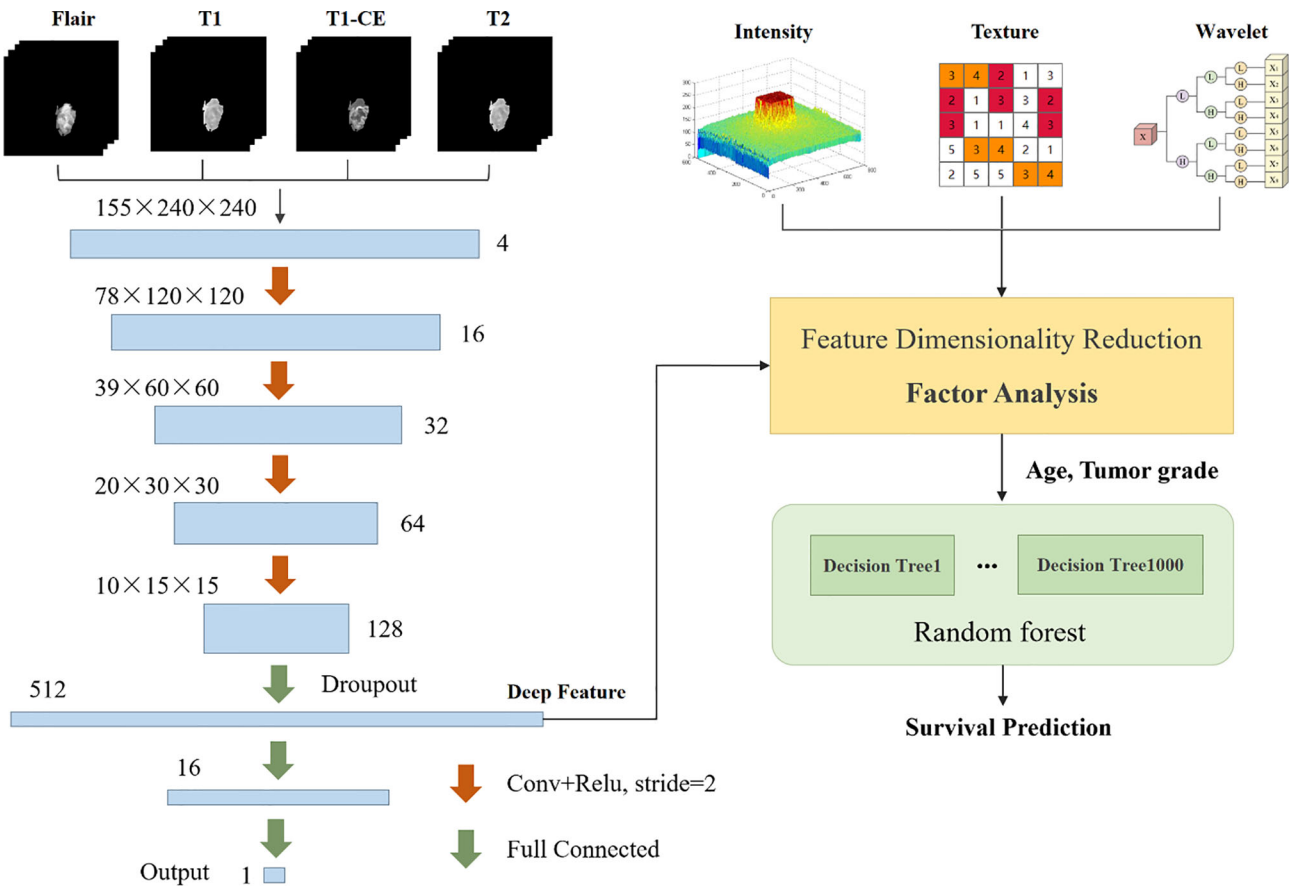


FIGURE 6 | Summary of prediction methods for the survival period of patients with brain tumors.

dimensionality reduction. **Figure 8** shows the feature values of all the factors after sorting.

Then we need to determine the dimensionality of the feature after dimensionality reduction, that is, recursively select within the range of factors whose feature value is greater than 1. Traverse the number of factors, reduce the feature to this dimension every two steps, enter the random forest model for training, and save the feature dimension with the smallest root mean square error. We draw a heat map of the process of feature dimensionality reduction, as shown in **Figure 9**. The abscissa represents the features before dimensionality reduction, and the ordinate represents the features after dimensionality reduction

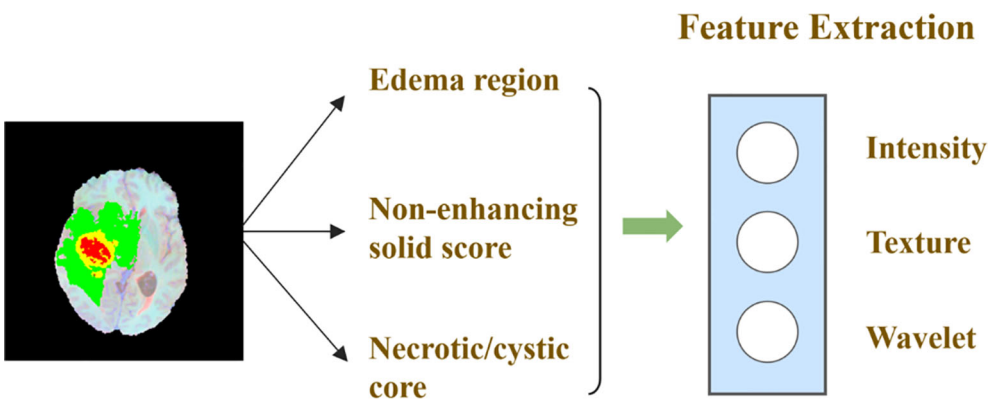


FIGURE 7 | Schematic diagram of image feature extraction.

TABLE 1 | Types and contents of extracted features.

Feature Category	Name of the features
First Order Statistic	Energy, Total Energy, Entropy, Minimum, 10th percentile, 90th percentile, Maximum, Mean, Median, Interquartile Range, Range, Mean Absolute Deviation, Robust Mean Absolute Deviation, Root Mean Squared, Standard Deviation, Skewness, Kurtosis, Variance, Uniformity.
Shape	Mesh Volume, Voxel Volume, Surface Area, Surface Area to Volume ratio, Sphericity, Compactness 1, Compactness 2, Spherical Disproportion, Maximum 3D diameter, Maximum 2D diameter (Slice), Maximum 2D diameter (Column), Maximum 2D diameter (Row), Major Axis Length, Minor Axis Length, Least Axis Length, Elongation, Flatness.
GLCM	Autocorrelation, Joint Average, Cluster Prominence, Cluster Shade, Cluster Tendency, Contrast, Correlation, Difference Average, Difference Entropy, Difference Variance, Joint Energy, Joint Entropy, Informational Measure of Correlation, Informational Measure of Correlation, Inverse Difference Moment, Maximal Correlation Coefficient, Inverse Difference Moment Normalized, Inverse Difference, Inverse Difference Normalized, Inverse Variance, Maximum Probability, Sum Average, Sum Entropy, Sum of Squares.
GLRLM	Short Run Emphasis, Long Run Emphasis, Gray Level Non Uniformity, Gray Level Non-Uniformity Normalized, Run Length Non-Uniformity, Run Length Non-Uniformity Normalized, Run Percentage, Gray Level Variance, Run Variance, Run Entropy, Low Gray Level Run Emphasis, High Gray Level Run Emphasis, Short Run Low Gray Level Emphasis, Short Run High Gray Level Emphasis, Long Run Low Gray Level Emphasis, Long Run High Gray Level Emphasis.
GLSZM	Small Area Emphasis, Large Area Emphasis, Gray Level Non Uniformity, Gray Level Non-Uniformity Normalized, Size Zone Non-Uniformity, Size-Zone Non-Uniformity Normalized, Zone Percentage, Gray Level Variance, Zone Variance, Zone Entropy, Low Gray Level Zone Emphasis, High Gray Level Zone Emphasis, Small Area Low Gray Level Emphasis, Small Area High Gray Level Emphasis, Large Area Low Gray Level Emphasis, Large Area High Gray Level Emphasis.
GLDM	Small Dependence Emphasis, Large Dependence Emphasis, Gray Level Non-Uniformity, Gray Level Non-Uniformity Normalized, Dependence Non-Uniformity, Dependence Non-Uniformity Normalized, Gray Level Variance, Dependence Variance, Dependence Entropy, Dependence Percentage, Low Gray Level Emphasis, High Gray Level Emphasis, Small Dependence Low Gray Level Emphasis, Small Dependence High Gray Level Emphasis, Large Dependence Low Gray Level Emphasis, Large Dependence High Gray Level Emphasis.
NGTDM	Coarseness, Contrast, Busyness, Complexity, Strength.

(Use numbers to represent feature names). It can be seen the degree of correlation between the original feature and the feature after dimensionality reduction.

Random Forest Model

After obtaining the effective feature set, we choose the random forest regression model to predict the survival period. Random forest algorithm is an ensemble technology that combines multiple decision trees, it usually has better generalization capabilities and not sensitive to multiple collinearity. Random Forest randomly selects k new self-service sample sets from the original training data set each time by applying the bootstrap method with replacement to construct k decision trees, and the unselected sample sets are used to estimate the generalization error of the model. Secondly, randomly extract m features at the node of each tree (m is less than the total number of features), and select the best split point to split by calculating the amount of information entropy in each feature. The calculation formula of information entropy is:

$$S = -\sum_{i=1}^{n} P_i \log_2 P_i \tag{7}$$

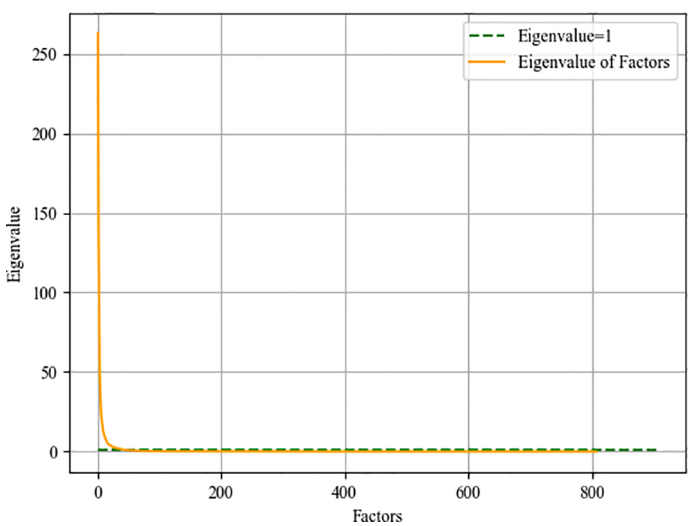


FIGURE 8 | Eigenvalue curve of factor.

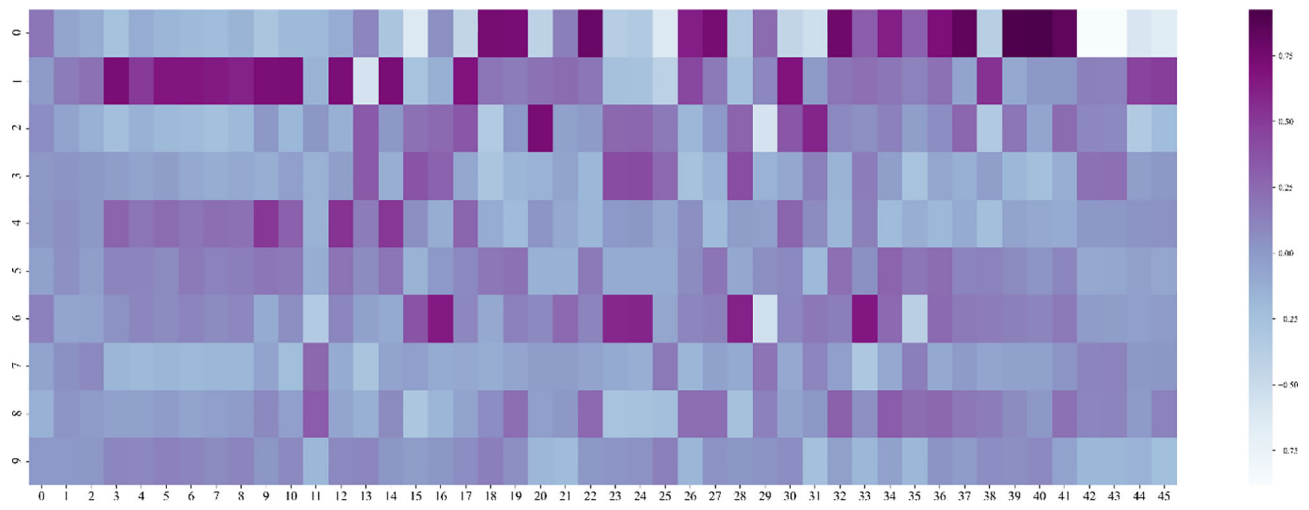


FIGURE 9 | Heat map of feature dimensionality reduction.

where P_i represents the probability of occurrence of the i-th situation. Finally, the generated trees are formed into a random forest, and the random forest is used to classify the new data. The classification result is determined by the number of votes of the tree classifier.

We use 1000 basic decision tree regressors. In order to ensure the robustness of model training, we will cross-validate the model 100 times and randomly divide the data into a training set and a test set each time. The training set data accounted for 0.9 of the total, and the test set data accounted for 0.1 of the total. We make sure to use different data combinations for training and testing every time. Finally, the average of 100 cross-validations is used as the final error loss.

DATASET AND EXPERIMENTS

Datasets

The dataset we use comes from the 2019 and 2020 Brain Tumor Segmentation Challenge. The data set contains two types of tumors, namely high-grade glioblastoma (HGG) and low-grade glioblastoma (LGG). The MRI of each sample in the tumor segmentation task data contains four modalities: fluid attenuation inversion recovery (FLAIR), T1 weighting (T1), T1-weighted contrast-enhanced (T1-CE), and T2 weighting (T2). The tumor sub-regions are edema, non-enhanced verification, necrosis/cystic nucleus and the entire tumor. We need to segment the enhancing tumor (ET), whole tumor (WT), and tumor core (TC) formed by nesting these sub-regions. The organizer provided 335 training samples and 125 validation samples without masks in BraTS 2019. In the survival prediction task, 209 samples containing age, tumor grade (HGG and LGG) and survival time defined in days are provided. Also in BraTS 2020, the training set and validation set sizes are 369 and 125, and 236 survival samples are provided. We divide the survival data into training set and test set by 9:1 respectively for cross-validation.

Evaluation Metrics

For the evaluation indicators of the segmentation model, we follow the 4 indicators used in the Brain Tumor Segmentation Challenge, namely Dice, Sensitivity, Specificity and Hausdorff distance. Dice is the overall evaluation standard, and its formula is defined as:

$$\text{Dice} = \frac{2\text{TP}}{\text{FP} + 2\text{TP} + \text{FN}} \tag{8}$$

where FN, TP and FP represent the number of false negative, true positive and false positive voxels respectively. Sensitivity represents the sensitivity of the model to voxels of the segmented region, and is used to measure the accuracy of segmenting the target region and is defined as:

$$\textit{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

Specificity represents the ability of the model to correctly predict the background and is defined as:

$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} \tag{10}$$

where TN represents the number of true negative. In addition, Hausdorff95 is a measure of Hausdorff distance. It is more sensitive to the segmentation boundary. The smaller the value, the closer the prediction is to the true value.

$$\begin{aligned}&\text{Hausdorff}_95(X, Y)\\ &= \max\left\{\max_{x\in X}\min_{y\in Y} d(x, y), \max_{y\in Y}\min_{x\in X} d(x, y)\right\}\end{aligned} \tag{11}$$

where X is the volume of the mask, Y is the volume predicted by the model, and d(), represents the distance from X to Y.

In addition, for the survival prediction task, we choose MSE, MAE and RMSE as the evaluation indicators of the model. They are all used to measure the deviation between the predicted value and the true value. The calculation formula is as follows:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(X_{obs,i} - X_{pre,i})^2 \tag{12}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|X_{obs,i} - X_{pre,i}\right| \tag{13}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{pre,i})^2}{n}} \tag{14}$$

where i represents the patient, n represents the total number of patients, X_{obs} represents the true survival period of the patient, and X_{pre} represents the survival period predicted by the model.

EXPERIMENTAL RESULTS

Segmentation Result

The brain tumor segmentation method proposed in the article is evaluated experimentally on the BraTS 2019, 2020 dataset. The data has a multi-modal imaging protocol: Flair, T1, T1-CE, T2. The mask is manually segmented by experts, including three nested sub-regions: enhanced tumor (ET), whole tumor (WT) and tumor core (TC) (5, 26). We train on the 3D volume of brain tumors, using dice coefficients, sensitivity, specificity and Hausdorff95 distance to evaluate the performance of the model. **Table 2** shows the average performance of the model on the validation set, and it can be seen that the model has achieved good segmentation results.

At the same time, we conducted ablation experiments on BraTS 2020 dataset, removing the SE module, NL module and the V-Net model without any attention module. Since CNN is the basis for the automatic segmentation of brain tumors, we also designed a CNN network to segment brain tumors as a control group. The CNN network consists of 8 layers of convolution and a sigmoid classification layer. **Table 3** shows the results of these comparative experiments. On the whole, the performance of NLSE-VNet is the best. However, the CNN network only uses a small number of convolutions to reach a medium segmentation level, therefore CNN has an important role for automatic segmentation.

Figures 10 and **11** respectively show the combination of violin chart and scatter plot for each evaluation index in the validation set of the NLSE-VNET model on BraTS 2019 and BraTS 2020. It can be seen from the figure that in all samples, the results are relatively concentrated in the higher area, and there are only a few abnormalities, indicating that the model has strong individual case prediction ability. Since sensitivity measures the model's ability to predict the tumor area, and specificity measures the model's ability to predict the background, it can be seen that the distribution range of sensitivity is close to the specificity, indicating that the model's ability to predict the tumor area is similar to the background, effectively alleviating the problem of category imbalance.

We randomly selected four slices in the training set, comparing the experts to the model prediction, while three-dimensional reconstruction of the division results, as shown in **Figure 12**. Among them, red represents the core of the tumor, the combined area of yellow and red represents the enhanced tumor, and the entire segmented area represents the entire tumor. It can be seen that the results of the model are very similar to the standards in overall and detail, providing a more intuitive diagnostic basis for the doctor.

We randomly selected some examples in the validation set to visually present the segmentation results and marked the Dice of the ET area on the graph, as shown in **Figure 13**. We also perform three-dimensional reconstruction of the segmentation results. From these examples, it can be seen that the model is very effective in segmenting tumors of different sizes, shapes, and positions, which also provides an important guarantee for the accuracy of subsequent survival prediction.

Overall Survival Prediction Results

There are 335 cases in the training set of the BraTS 2019 segmentation task, but the organizer only provided survival labels for 209 cases. BraTS 2020 has 236 cases of lifetime labeling, and due to the error of the segmentation model, the data used in the experiment is 232 cases. The features we use are divided into three categories, clinical factors such as age and tumor grade provided by the organizer, radiological features extracted through the Pyradiomics toolkit, and deep features extracted through CNN. We use these three types of features to obtain the results on the two testsets as shown in **Table 4**.

At the same time, we also conducted a comparative experiment on the BraTS 2020 dataset. In order to ensure effective comparison, we conducted the same experiment on the mask provided by the organizer and the prediction result of our segmentation model. **Table 5** lists the results obtained by combining different types of features. Among them, "CNN (RF)" represents the use of deep features and random forest models to predict the survival period, and "CNN (DL)" represents the use

TABLE 2 | Segmentation results of different models for BraTS 2019 and BraTS 2020.

	Dice			Sensitivity			Specificity			Hausdorff95		
	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
BraTS2019	0.70	0.87	0.74	0.73	0.87	0.70	0.99	0.99	0.99	5.7	7.4	9.8
BraTS2020	0.73	0.88	0.79	0.77	0.89	0.82	0.99	0.99	0.99	36.9	6.5	10.0

TABLE 3 | Segmentation results of ablation experiments on the BraTS 2020 dataset.

Segmentation	Dice			Sensitivity			Specificity			Hausdorff95		
Model	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
NL-VNet	0.68	0.83	0.74	0.78	0.90	0.78	0.99	0.99	0.99	44.66	11.16	15.38
SE-VNet	0.72	0.88	0.78	0.75	0.88	0.79	0.99	0.99	0.99	40.76	6.94	12.46
VNet	0.65	0.86	0.76	0.75	0.89	0.84	0.99	0.99	0.99	55.04	9.21	13.28
CNN	0.69	0.86	0.74	0.70	0.87	0.75	0.99	0.99	0.99	52.88	12.07	18.02

of deep features and neural networks to predict the survival period.

CNN extracting features is usually difficult to explain the biological principles behind it (27). In order to have a deeper understanding of the process of model learning features, we generate activation maps for each activation layer (ReLU) in the network (28), as shown in **Figure 14**. We can observe that after the first activation layer, some features such as texture and intensity are learned, and the second activation layer learns spatial features such as shape and size. It can be seen that the third activation layer focuses attention on In the TC region, we speculate that the features of the TC region are more important for survival prediction than other regions. We can observe that as the number of layers increases, the features extracted by CNN will become more and more abstract.

In order to verify that the radiological features extracted from the original image and the radiological features extracted from the wavelet image can improve the prediction performance, and at the same time prove the importance of the clinical features, we have done comparative experiments, as shown in **Table 6**. "Original" represents the radiological features extracted from the tumor region of the original MR sequence, and "Wavelet"

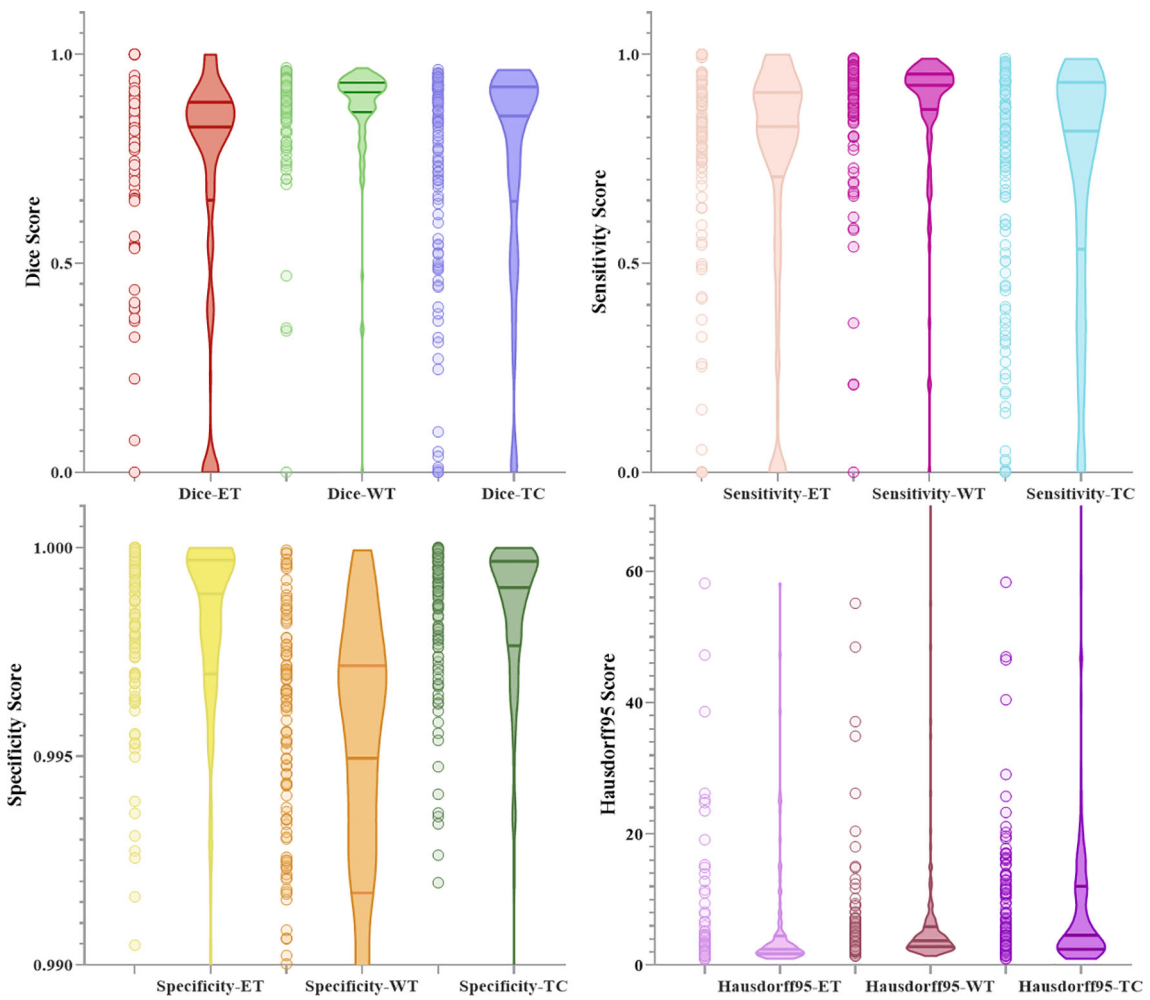


FIGURE 10 | Combination of scatters plot and box plot of the indicators for BraTS 2019.

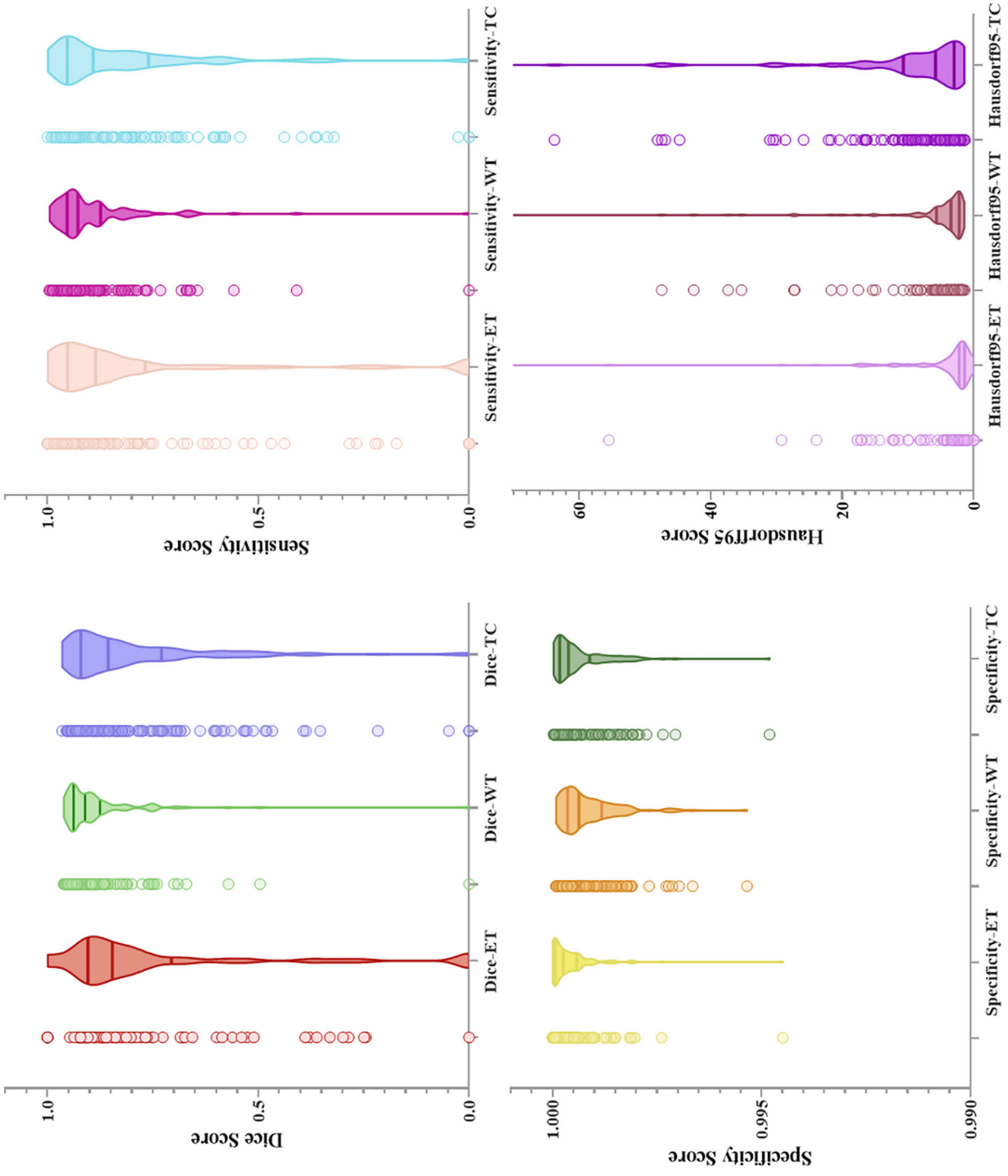


FIGURE 11 | Combination of scatters plot and box plot of the indicators for BraTS 2020.

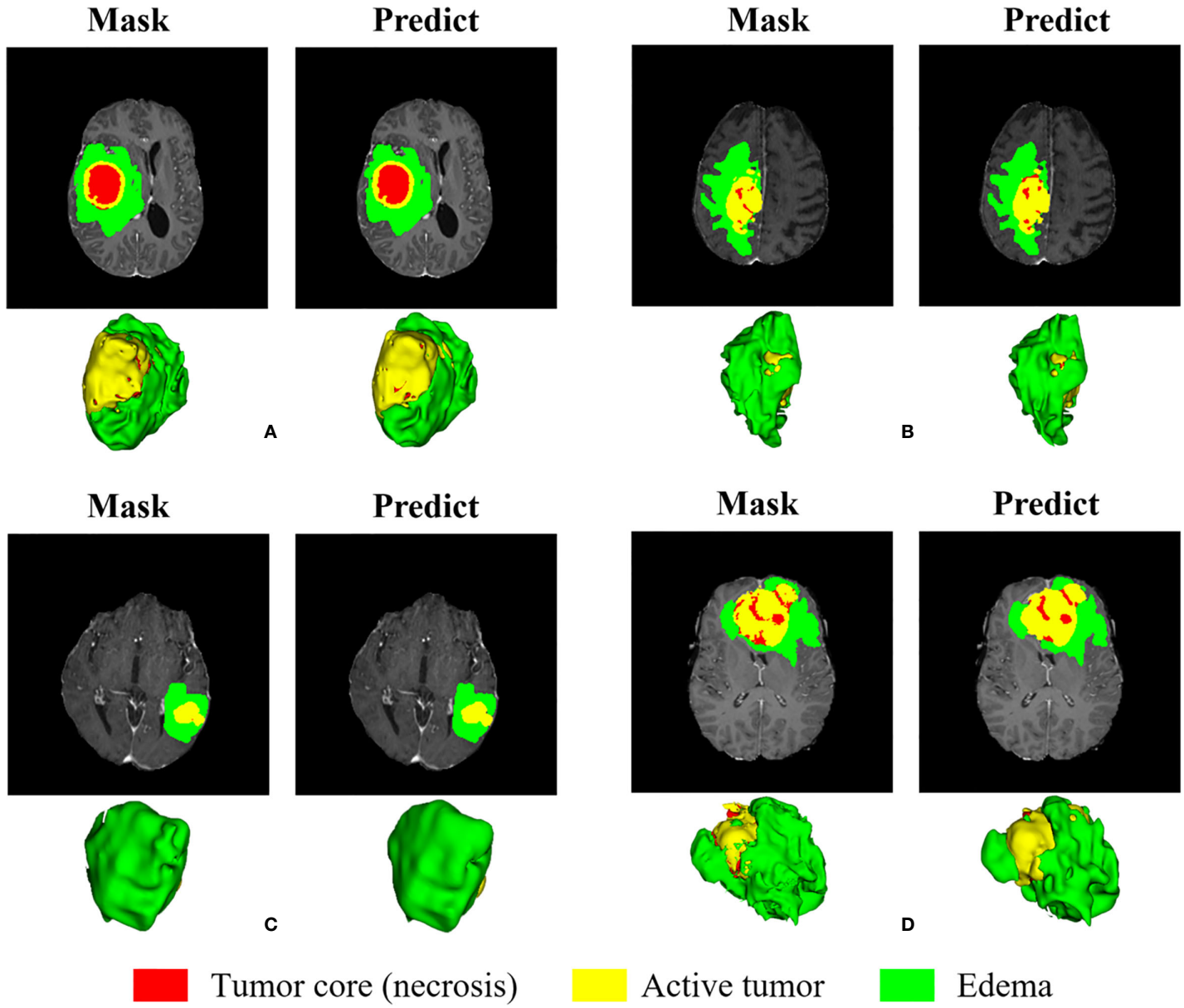


FIGURE 12 | Visual display of brain tumor segmentation results in training set. Red color represents the tumor core (necrosis), yellow color represents the active tumor and green regions are the edema.

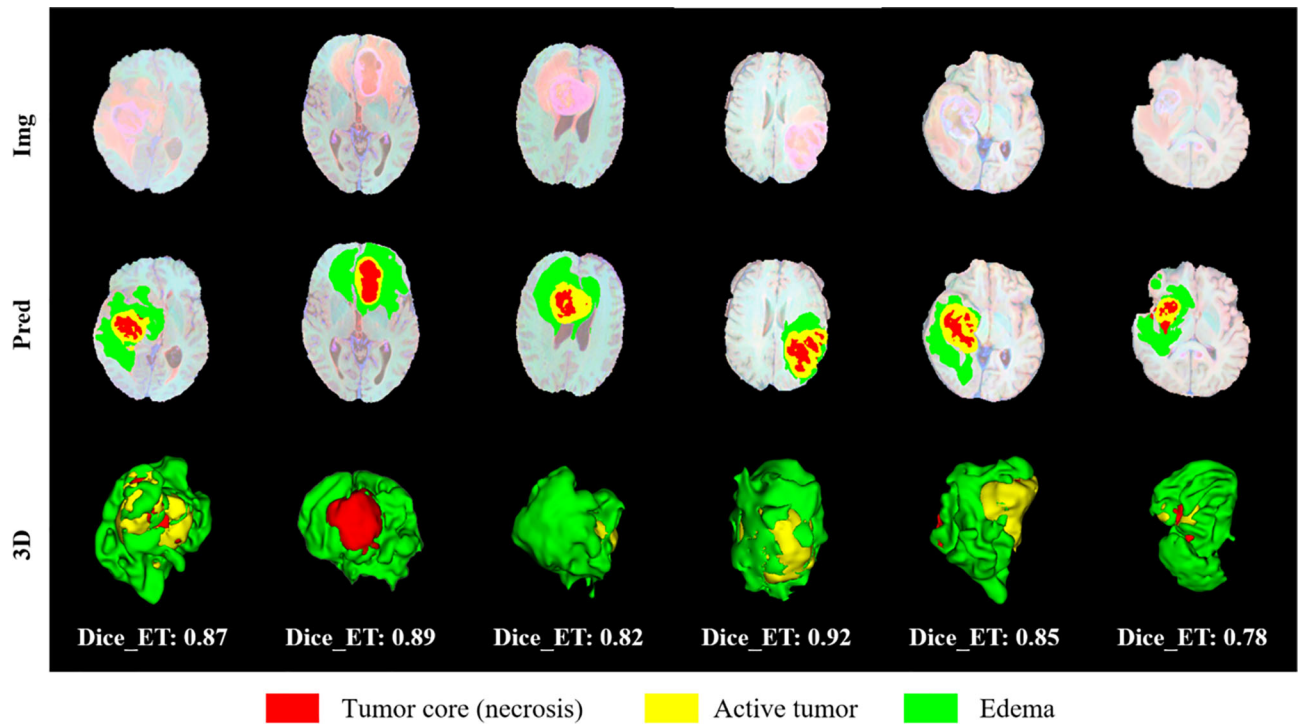


FIGURE 13 | Visual display of brain tumor segmentation results in validation set. Red color represents the tumor core (necrosis), yellow color represents the active tumor and green regions are the edema.

TABLE 4 | The survival prediction result of BraTS 2019 and BraTS 2020.

Dataset	MSE	MAE	RMSE
BraTS 2019	96470.98	229.07	306.78
BraTS 2020	100053.08	240.05	316.31

represents the radiological features extracted from the tumor region of the image after wavelet processing. Through the comprehensive comparison in **Table 6**, it is found that the radiological features extracted from the original image are better than those after wavelet in predicting survival. The effect of clinical factors is not as good as we expected. The reason may be that the amount of data is small and the features of clinical factors cannot be reflected.

For intuitive comparison, we show the histogram of the RMSE of each method in **Figure 15**. In **Figure 15** (a), after adding the deep features extracted by CNN, the RMSE has been greatly reduced. Among them, the RMSE of the automatic segmentation results is reduced by 19.4 percentage points, while the survival prediction results using only the deep features are not stable, due to the limitation of data volume, we cannot get more robust results.

Figure 16 shows the correlation between age and survival time, and the correlation coefficient is -0.35, that is, there is a weak negative correlation between them. The study of Weninger L. et al. also confirmed our conclusions (29). Since the tumor type of the living time data given by the organizer is HGG, we no longer discuss the correlation between tumor type and survival

TABLE 5 | Survival prediction results of different types of feature on BraTS 2020.

	Method	MSE	MAE	RMSE
Mask	CNN+ Radiology+ Clinical	**87679.85**	**225.59**	**288.84**
	Radiology+ Clinical	88899.18	228.04	292.25
	CNN (RF)	91838.87	235.86	297.16
	CNN(DL)	113472.38	257.71	336.86
Predict	CNN+ Radiology+ Clinical	**100053.08**	**240.05**	**316.31**
	Radiology+ Clinical	162904.10	284.54	392.51
	CNN(RF)	153998.46	284.13	378.85
	CNN(DL)	136770.59	269.37	358.92

The bold values provided mean the best performed method.

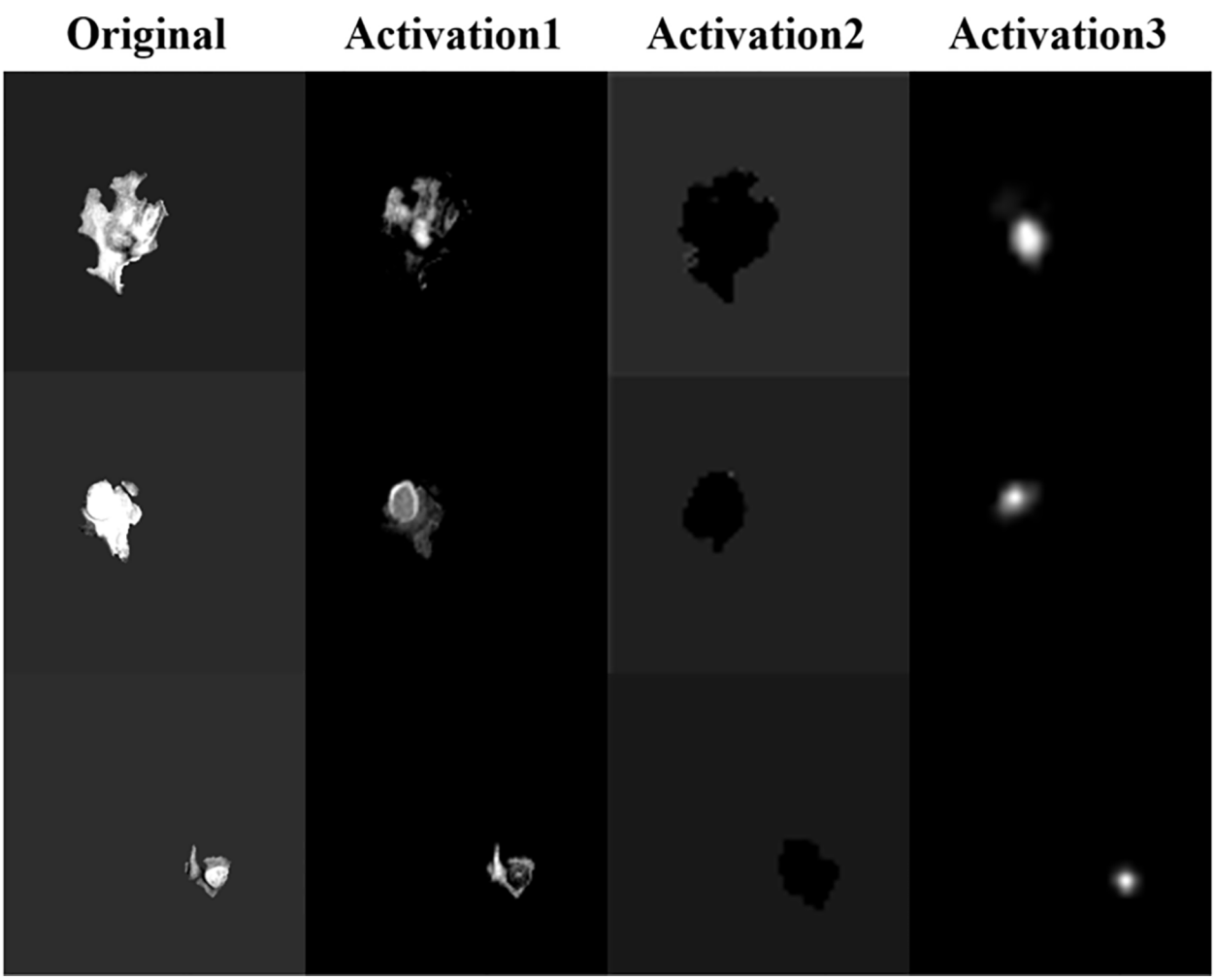


FIGURE 14 | Activation map, the feature map after the ReLU activation layer in the network.

TABLE 6 | Survival prediction results of different radiological features on BraTS 2020.

	Method	MSE	MAE	RMSE
Mask	Original+ Clinical	92542.55	233.08	297.83
	Wavelet+ Clinical	89551.18	236.87	296.02
	Original	95207.94	235.46	301.94
	Wavelet	88840.27	236.17	294.93
Predict	Original+ Clinical	158230.24	280.82	389.52
	Wavelet+ Clinical	167600.90	292.26	400.93
	Original	157355.79	279.59	388.28
	Wavelet	167466.49	292.26	401.19

time. However, it has been proved pathologically that HGG patients with glioma have a poor prognosis, and their survival period is often shorter than that of LGG.

Comparison

We compare the segmentation results with other methods, as shown in **Table 7**. Kim S. et al. proposes to obtain the initial segmentation probability map with 2D U-Net and then input the MR image and the initial segmentation into the 3D U-Net for segmentation (30). Amian M. et al. proposed a 3D deep segmentation method to divide glioma, including two parallel streamlines having two different resolutions, one convolutional neural network for learning local features, another for the entire image global observation (31). Shi W.

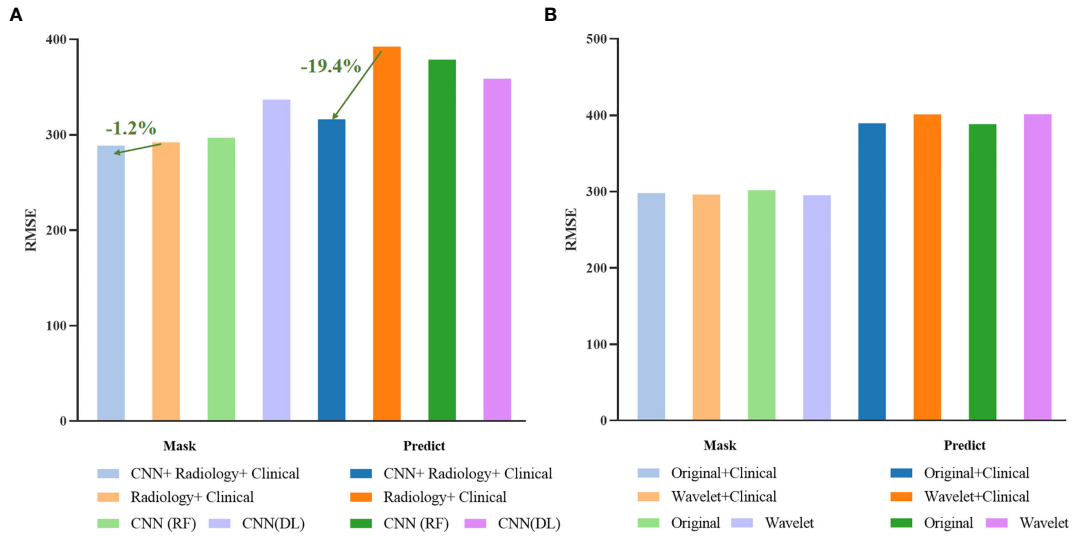


FIGURE 15 | Histogram of survival prediction results.

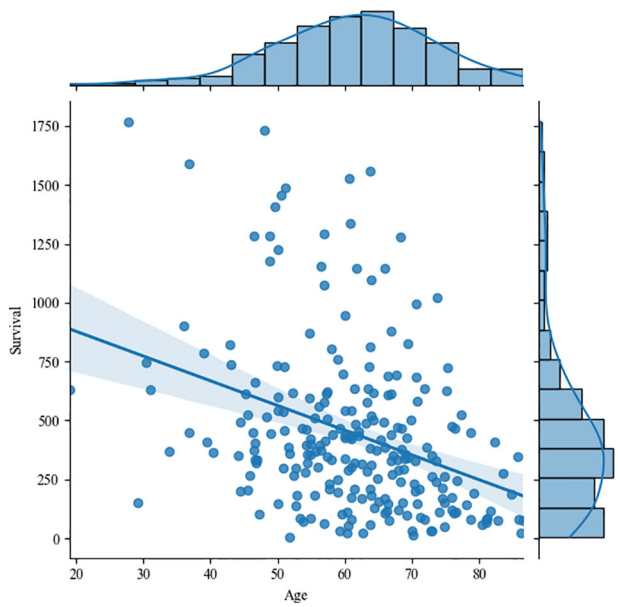


FIGURE 16 | The correlation between age and survival time.

et al. adopted a 2D U-Net network segmentation model based on dense cell and feature pyramid unit (32). Agravat R.R. et al. used the three sub-regions of the tumor in the whole convolutional neural network to fuse the segmentation results (33).

At the same time, we listed the results of using the BraTS 2020 dataset. Tarasiewicz T. et al. used Skinny, a lightweight U-Net-based architecture to segment brain tumors, which was originally used to detect skin from color images (34). Mchugh H. et al. believe that two-dimensional segmentation is more advantageous than three-dimensional segmentation. They use 2D density-UNet to segment brain tumors on two-dimensional slices (35). Zhao C. et al. replaced the convolution in the three-dimensional U-Net with two-dimensional multi-view convolution, and learned features in the axial, sagittal, and coronal respectively (36). Savadikar C. et al. used probabilistic U-Net to explore the effect of sampling different segmentation maps, and at the same time explored the effect of changes in the number of attention modules on segmentation quality (37).

We also compare the results of survival prediction in **Table 8**. Kim S. et al. extracted radiomics features, select a small amount of features from random forest retrogenizer to avoid overfitting, and finally predicted the survival time using a random forest regression model (30). Amian M. et al. extracted the spatial features of the entire tumor and sub-organization, using a random forest model to predict the survival time (31). Kofler F. et al. Only uses age this clinical features to predict the survival period, and three orthogonal polynomials and posetric regression models are used (38). Islam M. et al. according to the geometric shape of the tumor, and the position of the new radiology features is combined with clinical features, using XGBooST predictive survival (39).

Here are some methods for survival prediction using the BraTS 2020 dataset. Soltaninejad M. et al. used the ratio of tumor volume to brain tissue and the average tumor intensity as features and applied a random forest model to predict survival time (40). Agravat R.R. et al. used a random forest regressor to train the three types of features extracted from shape, volume, and age to predict survival (41). Patel J. extracted 2048 deep features from the segmentation network, used principal component analysis to reduce dimensionality, and trained the Cox risk proportional model for survival prediction (42). Ali M.J. et al. extracted multiple radiological features and image features

TABLE 7 | Comparison of the results of the segmentation task.

		Dice			Hausdorff95		
		ET	**WT**	**TC**	**ET**	**WT**	**TC**
BraTS 2019	**Proposed**	0.70	**0.87**	0.74	**5.7**	**7.4**	**9.8**
	Kim S. et al. (30)	0.67	0.87	0.76	8.8	14.2	11.7
	Amian M. et al. (31)	**071**	0.86	**0.77**	6.9	8.5	11.6
	Shi W. et al. (32)	0.69	0.87	0.77	5.9	21.2	12.2
	Agravat R.R. et al. (33)	0.60	0.70	0.63	11.7	14.3	17.1
BraTS 2020	**Proposed**	**0.73**	0.88	**0.79**	**36.9**	6.5	**10.0**
	Tarasiewicz T. et al. (34)	0.70	**0.89**	0.75	40.1	**4.6**	10.7
	Mchugh H. et al. (35)	0.71	0.88	0.79	40.6	6.7	10.2
	Zhao C. et al. (36)	0.67	0.86	0.62	47.3	12.6	50.1
	Savadikar C. et al. (37)	0.69	0.82	0.72	36.9	41.5	26.3

The bold values provided mean the best performed method.

TABLE 8 | Comparison of the results of survival prediction task.

	Team	MSE
BraTS 2019	**Proposed**	**96470.98**
	Kim S. et al. (30)	121778.60
	Amian M. et al. (31)	104253.00
	Kofler F. et al. (38)	101877.80
	Islam M. et al. (39)	127478.65
BraTS 2020	**Proposed**	**100053.08**
	Soltaninejad M. et al. (40)	109564.00
	Agravat R.R. et al. (41)	116083.48
	Patel J. et al. (42)	152467.00
	Ali M.J. et al. (43)	105079.40

The bold values provided mean the best performed method.

from the MRI volume, used random forest recursive features to eliminate, and then used random forest regression factors combined with grid search to predict survival (43).

Discussion

In this section we will discuss the results presented and highlight the shortcomings and solutions in current research.

In the segmentation task, through ablation experiments, we found that the NL module has limited ability to improve segmentation accuracy. It may be that the attention module is not the most suitable model for brain tumor segmentation. We can try to replace the NL module with some other attention modules, such as the SCSE module, which is a variant of the SE module. By focusing on important feature maps or feature channels, it reduces the impact of unimportant features, thereby improving image segmentation results. This module has been applied to brain MR segmentation and achieved excellent results (44). Or Edge Guidance module, which combines edge detection and semantic segmentation, and can use edge information to better supervise and learn semantic segmentation (45). While ignoring a very important concept in medical images, that is, the structure of medical images, we will consider adding structure such as edges and textures when designing segmentation algorithms in the future.

In this article, we design a CNN network to extract the deep features of the image. It contains 4 layers of convolution with a step size of 2 and 3 layers of fully connected neural networks (the last 2 layers of fully connected layers are used to directly predict the survival time). Other network structures were also tried to extract the deep features, and two layers of convolution with a step size of 1 and four layers of convolution with a step size of 1 were added respectively before the fully connected layer. In order to compare which network extracts the best performance of the deep feature, we directly use the fully connected layer to predict the survival time of the deep feature. The results are shown in **Table 9**. Here we only extract deep features from the results of model segmentation. From the results in **Table 9**, it can be seen that the 4-layer convolutional network has the best performance in extracting deep features.

In the survival prediction task, through the comparative experiments in **Table 3**, we found that only using deep features will make the survival prediction results unstable. The experiment on the mask shows that the neural network results are better than the random forest, while the experiment on the automatic segmentation results has the opposite result, because CNN will show a high degree of variability in different periods. We need more data to fully obtain the robustness of the CNN results. Radiological features have better interpretable advantages and generally more robust results can be obtained. Banerjee et al. designed two new radiological features, extracted from the brain segmentation atlas and spatial habitats and proved their effectiveness (46). We consider introducing these two features in future work to further improve survival prediction performance. In addition, we can use clinical knowledge to classify radiological features more finely, in order to find more suitable feature dimensionality reduction methods, so that feature dimensionality reduction can also have strong interpretability and clinical applicability (47).

TABLE 9 | Survival prediction results of different CNN network structures.

Method	MSE	MAE	RMSE
CNN with 4 layers of convolution	**136770.59**	**269.37**	**358.92**
CNN with 6 layers of convolution	138065.90	269.83	363.64
CNN with 8 layers of convolution	144921.16	301.79	380.69

The bold values provided mean the best performed method.

CONCLUSION

This paper proposes a method for automatically segmenting brain tumors and predicting the survival time of tumor patients, and the performance is verified on the BraTS 2019 and BraTS 2020. First of all, the traditional VNet is improved, and the NLSE-VNet model is proposed. It adds an attention module on the basis of the original V-Net, which can increase a small amount of calculation and greatly improve the segmentation accuracy. Secondly, we used the Pyradiomics toolkit to extract the radiological features of the segmented tumor regions, and the CNN network extracted the deep features (48). Then the factor analysis method is used to reduce the feature dimension, and the clinical features such as age are input into the random forest model to predict the survival period. This research combines promising radiology, machine learning, and deep learning methods, and achieves an average segmentation accuracy of 0.79, and a average RMSE of 311.5 for survival prediction. Experimental results show that this method has reached a relatively prominent level and has good advantages in clinical applications.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: https://www.med.upenn.edu/cbica/brats2020/data.html.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board (IRB), Zhejiang Chinese Medical University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HH, WZ, and XL conceived and designed the study. HH, WZ, YF, JH, SS, and XL contributed to the literature search. HH, WZ, JH, and XL contributed to data analysis and data curation. HH, YF, JH, SS, and XL contributed to data visualization. HH and WZ contributed for software implementation. HH, WZ, YF, JH, and XL contributed to the tables and figures. HH, WZ, YF, SS, and XL contributed to writing of the report. HH, WZ, and XL contribute to review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This work is funded in part by the National Natural Science Foundation of China (Grants No. 62072413, 61602419), in part by the Natural Science Foundation of Zhejiang Province of China (Grant No. LY16F010008), and also supported in part by Medical and Health Science and Technology Plan of Zhejiang Province of China (Grant No. 2019RC224).

REFERENCES

1. Pereira S, Pinto A, Alves V, Silva CA. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. *IEEE Trans Med Imaging* (2016) 35(5):1240–51. doi: 10.1109/TMI.2016.2538465
2. Sun L, Zhang S, Chen H, Luo L. Brain Tumor Segmentation and Survival Prediction Using Multimodal MRI Scans With Deep Learning. *Front Neurosci* (2019) 13:810. doi: 10.3389/fnins.2019.00810
3. da Silva B, Mathew RK, Polson ES, Williams J, Wurdak H. Spontaneous Glioblastoma Spheroid Infiltration of Early-Stage Cerebral Organoids Models Brain Tumor Invasion. *Slas Discov Adv Life ences R D* (2018) 23(8):862–8. doi: 10.1177/2472555218764623
4. Myronenko A. 3d MRI Brain Tumor Segmentation Using Autoencoder Regularization. *Lecture Notes Comput Sci* (2019) 11384:311–20. doi: 10.1007/978-3-030-11726-9_28
5. Bakas S, Reyes M, Jakab A, Bauer S, Rempfler M, Crimi A, et al. Identifying the Best Machine Learning Algorithms for Brain Tumor Segmentation, Progression Assessment, and Overall Survival Prediction in the BRATS Challenge. *arXiv* (2018) 1811.02629.
6. Sun L, Zhang ST, Luo L. Tumor Segmentation and Survival Prediction in Glioma With Deep Learning. *Int MICCAI Brainlesion Workshop* (2018) 11384:83–93. doi: 10.1007/978-3-030-11726-9_8
7. Banerjee S, Mitra S, Shankar BU. Synergetic Neuro-Fuzzy Feature Selection and Classification of Brain Tumors. *Proceedings of IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (2017). pp. 1–6. doi: 10.1109/FUZZ-IEEE.2017.8015514
8. Lu CF, Hsu FT, Hsieh KLC, Kao YCJ, Cheng SJ, Hsu JBK, et al. Machine Learning–Based Radiomics for Molecular Subtyping of Gliomas. *Clin Cancer Res* (2018) 24(18):4429–36. doi: 10.1158/1078-0432.CCR-17-3445
9. Zhou C, Ding C, Wang X, Lu ZT, Tao DC. One-Pass Multi-Task Networks With Cross-Task Guided Attention for Brain Tumor Segmentation. *IEEE Trans Image Process* (2019) 29:4516–29. doi: 10.1109/TIP.2020.2973510
10. Chen L, Wu Y, Dsouza AM, Abidin AZ, Wismuller A, Xu C. MRI Tumor Segmentation With Densely Connected 3D CNN. *SPIE Med Imaging Conference* (2018). doi: 10.1117/12.2293394
11. Sun L, Shao W, Zhang D, Liu M. Anatomical Attention Guided Deep Networks for ROI Segmentation of Brain MR Images. *IEEE Trans Med Imaging* (2019) 39(6):2000–12. doi: 10.1109/TMI.2019.2962792
12. Lachinov D, Vasiliev E, Turlapov V. Glioma Segmentation With Cascaded UNet. *Lecture Notes Comput Sci* (2019) 11384:189–98. doi: 10.1007/978-3-030-11726-9_17
13. Sun L, Zhang S, Luo L. Tumor Segmentation and Survival Prediction in Glioma With Deep Learning. *Lecture Notes Comput Sci* (2019) 11384:83–93. doi: 10.1007/978-3-030-11726-9_8
14. Shboul ZA, Alam M, Vidyaratne L, Pei L, Iftekharuddin KM. Glioblastoma Survival Prediction. *Lecture Notes Comput Sci* (2019) 11384:508–15. doi: 10.1007/978-3-030-11726-9_45
15. Baid U, Talbar S, Rane S, Gupta S, Thakur MH, Moiyadi A, et al. Deep Learning Radiomics Algorithm for Gliomas (DRAG) Model: A Novel Approach Using 3d UNET Based Deep Convolutional Neural Network for Predicting Survival in Gliomas. *Lecture Notes Comput Sci* (2019) 11384:369–79. doi: 10.1007/978-3-030-11726-9_33
16. Kim JY, Yoon MJ, Park JE, Choi EJ, Lee J, Kim HS. Radiomics in Peritumoral non-Enhancing Regions: Fractional Anisotropy and Cerebral Blood Volume Improve Prediction of Local Progression and Overall Survival in Patients With Glioblastoma. *Neuroradiology* (2019) 61(11):1261–72. doi: 10.1007/s00234-019-02255-4

17. Weninger L, Rippel O, Koppers S, Merhof D. Segmentation of Brain Tumors and Patient Survival Prediction: Methods for the BraTS 2018 Challenge. *Lecture Notes Comput Sci* (2019) 11384:3–12. doi: 10.1007/978-3-030-11726-9_1
18. Banerjee S, Mitra S, Shankar BU. Multi-Planar Spatial-ConvNet for Segmentation and Survival Prediction in Brain Cancer. *Lecture Notes Comput Sci* (2019) 11384:94–104. doi: 10.1007/978-3-030-11726-9_9
19. Wang C, Sun W, Kirkpatrick J, Chang Z, Yin FF. Assessment of Concurrent Stereotactic Radiosurgery and Bevacizumab Treatment of Recurrent Malignant Gliomas Using Multi-Modality MRI Imaging and Radiomics Analysis. *J Radiosurg SBRT* (2018) 5(3):171–81. doi: 10.1016/j.ijrobp.2017.06.2367
20. Zhao X, Wu Y, Song G, Li Z, Zhang Y, Fan Y. A Deep Learning Model Integrating FCNNs and CRFs for Brain Tumor Segmentation. *Med Image Anal* (2017) 43:98–111. doi: 10.1016/j.media.2017.10.002
21. Afshar P, Mohammadi A, Plataniotis KN, Oikonomou A, Benali H. From Handcrafted to Deep-Learning-Based Cancer Radiomics: Challenges and Opportunities. *IEEE Signal Process Mag* (2019) 36(4):132–60. doi: 10.1109/MSP.2019.2900993
22. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-And-Excitation Networks. *IEEE Trans Pattern Anal Mach Intell* (2019) 42(8):2011–23. doi: 10.1109/TPAMI.2019.2913372
23. Wang X, Girshick R, Gupta A, He K. Non-Local Neural Networks. *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2017). p. 7794–803.
24. Sudre CH, Li W, Vercauteren T, Ourselin S, Jorge Cardoso M. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. *Lecture Notes Comput Sci* (2017) 10553:240–8. doi: 10.1007/978-3-319-67558-9_28
25. Griethuysen J, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339
26. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, et al. Advancing The Cancer Genome Atlas Glioma MRI Collections With Expert Segmentation Labels and Radiomic Features. *Nat Sci Data* (2017) 4:170117. doi: 10.1038/sdata.2017.117
27. Ayyappan V, Chang A, Zhang C, Paidi SK, Bordett R, Liang T, et al. Identification and Staging of B-Cell Acute Lymphoblastic Leukemia Using Quantitative Phase Imaging and Machine Learning. *ACS Sensors* (2020) 5(10):3281–9. doi: 10.1021/acssensors.0c01811
28. Raman R, Hosoya H. Convolutional Neural Networks Explain Tuning Properties of Anterior, But Not Middle, Face-Processing Areas in Macaque Inferotemporal Cortex. *Commun Biol* (2020) 3:221. doi: 10.1038/s42003-020-0945-x
29. Weninger L, Liu Q, Merhof D. Multi-Task Learning for Brain Tumor Segmentation. *LNCS* (2020) 11992:327–37. doi: 10.1007/978-3-030-46640-4_31
30. Kim S, Luna M, Chikontwe P, Park SH. Two-Step U-Nets for Brain Tumor Segmentation and Random Forest With Radiomics for Survival Time Prediction. *LNCS* (2020) 11992:200–9. doi: 10.1007/978-3-030-46640-4_19
31. Amian M, Soltaninejad M. Multi-Resolution 3d CNN for MRI Brain Tumor Segmentation and Survival Prediction. *LNCS* (2020) 11992:221–30. doi: 10.1007/978-3-030-46640-4_21
32. Shi W, Pang E, Wu Q, Lin F. Brain Tumor Segmentation Using Dense Channels 2d U-Net and Multiple Feature Extraction Network. *Lecture Notes Comput Sci* (2020) 11992:273–83. doi: 10.1007/978-3-030-46640-4_26
33. Agravat RR, Raval MS. Brain Tumor Segmentation and Survival Prediction. *Lecture Notes Comput Sci* (2020) 11992:338–48. doi: 10.1007/978-3-030-46640-4_32
34. Tarasiewicz T, Kawulok M, Nalepa J. Lightweight U-Nets for Brain Tumor Segmentation. *LNCS* (2021) 12659:3–14. doi: 10.1007/978-3-030-72087-2_1
35. Mchugh H, Maso TG, Wang A. 2d Dense-UNet: A Clinically Valid Approach to Automated Glioma Segmentation. *LNCS* (2021) 12659:69–80. doi: 10.1007/978-3-030-72087-2_7
36. Zhao C, Zhao Z, Zeng Q, Feng YJ. MVP U-Net: Multi-View Pointwise U-Net for Brain Tumor Segmentation. *LNCS* (2021) 12659:93–103. doi: 10.1007/978-3-030-72087-2_9
37. Savadikar C, Kulhalli R, Garware B. Brain Tumour Segmentation Using Probabilistic U-Net. *LNCS* (2021) 12659:255–64. doi: 10.1007/978-3-030-72087-2_22
38. Kofler F, Paetzold JC, Ezhov I, Shit S, Krahulec D, Kirschke KS, et al. A Baseline for Predicting Glioblastoma Patient Survival Time With Classical Statistical Models and Primitive Features Ignoring Image Information. *LNCS* (2020) 11992:254–61. doi: 10.1007/978-3-030-46640-4_24
39. Islam M, Vibashan VS, Jose VJM, Wijethilake N, Utkarsh U, Ren HL. Brain Tumor Segmentation and Survival Prediction Using 3D Attention UNet. *LNCS* (2021) 11992:262–72. doi: 10.1007/978-3-030-46640-4_25
40. Soltaninejad M, Pridmore T, Pound M. Efficient MRI Brain Tumor Segmentation Using Multi-Resolution Encoder-Decoder Networks. *LNCS* (2021) 12659:30–9. doi: 10.1007/978-3-030-72087-2_3
41. Agravat RR, Raval MS. 3d Semantic Segmentation of Brain Tumor for Overall Survival Prediction. *LNCS* (2021) 12659:215–27. doi: 10.1007/978-3-030-72087-2_19
42. Patel J, Chang K, Hoebel K, Gidwani M, Arun N, Gupta S, et al. Segmentation, Survival Prediction, and Uncertainty Estimation of Gliomas From Multimodal 3d MRI Using Selective Kernel Networks. *LNCS* (2021) 12659:228–40. doi: 10.1007/978-3-030-72087-2_20
43. Ali MJ, Akram MT, Saleem H, Raza B, Shahid AR. Glioma Segmentation Using Ensemble of 2D/3D U-Nets and Survival Prediction Using Multiple Features Fusion. *LNCS* (2021) 12659:189–99. doi: 10.1007/978-3-030-72087-2_17
44. Roy A G N, Nav AB, Wachinger C. Concurrent Spatial and Channel Squeeze & Excitation in Fully Convolutional Networks. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention. LNCS* (2018) 11070:421–9. doi: 10.1007/978-3-030-00928-1_48
45. Zhao J, Liu JJ, Fan DP, Cao Y, Yang JF, Cheng MM. EGNet: Edge Guidance Network for Salient Object Detection. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. (2020). doi: 10.1109/ICCV.2019.00887
46. Wijethilake N, Meedeniya D, Chitraranjan C, Perera I, Islam M, Ren HL. Glioma Survival Analysis Empowered With Data Engineering—A Survey. *IEEE Access* (2021) 9:43168–91. doi: 10.1109/ACCESS.2021.3065965
47. Suter Y, Jungo A, Rebsamen M, Knecht U, Herrmann E, Wiest R, et al. Deep Learning Versus Classical Regression for Brain Tumor Patient Survival Prediction. *LNCS* (2019) 11384:429–40. doi: 10.1007/978-3-030-11726-9_38
48. Pospisil DA, Pasupathy A, Bair W. 'Artiphysiology' Reveals V4-Like Shape Tuning in a Deep Network Trained for Image Classification. *eLife Sci* (2018) 7:e38242. doi: 10.7554/eLife.38242

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Huang, Zhang, Fang, Hong, Su and Lai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Bioengineering and Biotechnology

METHODS
published: 19 August 2021
doi: 10.3389/fbioe.2021.697915

A Multichannel Deep Neural Network for Retina Vessel Segmentation *via* a Fusion Mechanism

Jiaqi Ding[1], *Zehua Zhang*[1], *Jijun Tang*[1] *and Fei Guo*[2]*

[1]School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin, China, [2]School of Computer Science and Engineering, Central South University, Changsha, China

Changes in fundus blood vessels reflect the occurrence of eye diseases, and from this, we can explore other physical diseases that cause fundus lesions, such as diabetes and hypertension complication. However, the existing computational methods lack high efficiency and precision segmentation for the vascular ends and thin retina vessels. It is important to construct a reliable and quantitative automatic diagnostic method for improving the diagnosis efficiency. In this study, we propose a multichannel deep neural network for retina vessel segmentation. First, we apply U-net on original and thin (or thick) vessels for multi-objective optimization for purposively training thick and thin vessels. Then, we design a specific fusion mechanism for combining three kinds of prediction probability maps into a final binary segmentation map. Experiments show that our method can effectively improve the segmentation performances of thin blood vessels and vascular ends. It outperforms many current excellent vessel segmentation methods on three public datasets. In particular, it is pretty impressive that we achieve the best F1-score of 0.8247 on the DRIVE dataset and 0.8239 on the STARE dataset. The findings of this study have the potential for the application in an automated retinal image analysis, and it may provide a new, general, and high-performance computing framework for image segmentation.

Keywords: retina vessel segmentation, multi-objective optimization, multiple probability map fusion mechanism, skeleton extraction, multi-channel DCNN

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln, United Kingdom

Reviewed by:
Yalin Zheng,
University of Liverpool,
United Kingdom
Jemima Jebaseeli,
Karunya Institute of Technology and Sciences, India

***Correspondence:**
Fei Guo
guofeieileen@163.com

Specialty section:
This article was submitted to Bionics and Biomimetics, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *20 April 2021*
Accepted: *06 July 2021*
Published: *19 August 2021*

Citation:
Ding J, Zhang Z, Tang J and Guo F (2021) A Multichannel Deep Neural Network for Retina Vessel Segmentation via a Fusion Mechanism. Front. Bioeng. Biotechnol. 9:697915. doi: 10.3389/fbioe.2021.697915

1 INTRODUCTION

The fundus photography can quickly and noninvasively obtain retinal images, which is usually used as an effective way for diagnosing fundus diseases. Furthermore, by observing retina blood vessels, medical scientists can assess symptoms of diseases, such as hypertension, diabetes, and neurodegenerative diseases. However, many studies based on retinal vascular changes still rely on a manual qualitative assessment, which prevents experts from grasping retinal diseases more accurately and efficiently. For example, narrowed retinal blood vessels is a typical early symptom of hypertension, but disease symptoms can only be assessed subjectively by ophthalmologists through fundus photography or angiography. These early symptoms are not only time-consuming but also hard to be spotted. Therefore, a reliable and quantitative automatic diagnostic method is urgently required to improve diagnosis efficiency, and some related research works have gradually risen in the recent years.

Retina vessel segmentation methods are generally divided into filter-based methods, machine learning algorithms, and deep learning methods. The filter-based technology (Annunziata et al., 2016) is almost consistent with image processing methods, using the filter window to process fundus images. Peter et al. (2012) used a wavelet transform to quickly detect blood vessels and calculated vascular profiles to determine blood vessel boundaries. Fraz et al. (2012) employed the Gabor filter and top-hat transformations of morphological operations for feature extraction and vessel segmentation. Nguyen et al. (2013) performed vessel segmentations by linear operators of different scales. Salazar-Gonzalez et al. (2014) used graph cut technology for vessel segmentation. In addition, machine learning (Roychowdhury et al., 2014) models usually extract feature vectors and then construct a classifier to label pixels. Orlando and Blaschko (2014) used a conditional random field (CRF) with a fully connected model to segment the fundus retina vessels. Gu and Cheng (2014) proposed an iterative two-step learning-based method to boost the segmentation performance by existing basic segmenters. Lupascu et al. (2010) constructed a 41-D vector for each pixel in the image to encode the alignment information, and then classified pixels using the AdaBoost classifier.

With the rapid development of deep learning in recent years, convolutional neural network (CNN) performs well on classification and regression tasks because it can hierarchically abstract representations using local operations. It is very suitable for computer vision–related applications. Especially, since the advent of U-net (Ronneberger et al., 2015) in 2015, it brought great progress to medical image segmentation tasks. It is an encoder–decoder structure, and skip connections inspired many subsequent studies. For example, M2UNet by Laibacher et al. (2018) and LadderNet by Zhuang (2018) obtained excellent results in the fundus retina vessel segmentation. They both are inspired by U-net. In addition, Melinscak et al. (2015) developed a 10-layer CNN for a binary classification based on the patch-wise method. Fu et al. (2016) constructed a deeply integrated network consisting of a convolutional neural network (CNN) and a conditional random field (CRF). In detail, multi-scale and multilevel CNNs were used to extract features, and a CRF was used to model the pixel interaction. In the recent years, many researchers made great progress. CS-Net (Mou et al., 2019) adds two attention mechanisms: spatial attention and channel attention, to the encoder and decoder to better capture the local and global features of images, thereby improving the segmentation results. DUnet (Jin et al., 2019) integrated the deformable convolution into U-net so that it can adaptively adjust the receptive field of the filter during the feature extraction process to extract features of different scales. Vessel-Net (Wu et al., 2019) embedded the inception-residual convolution block into U-net to improve the feature extraction ability of the encoder, and then used multiple supervision paths to train the network to obtain more refined segmentation. Wang et al. (2020) separately trained the "easy" and "hard" parts in the encoder stage to perform targeted vascular segmentation, and added an attention mechanism to the "hard" part for more effective segmentation. NFN+ (Wu et al., 2020) used two networks to achieve more refined segmentation. It exploited the front network to obtain a basic prediction probability map, and then used the followed network for post-processing. In addition, the author applied inter-network skip connections to unite the two networks to make better use of multi-scale features. SCS-Net (Wu et al., 2021) first used a scale-aware feature aggregation (SFA) module to extract multi-scale features, then employed the adaptive feature fusion (AFF) module to fuse different levels of features to obtain richer semantic information, and finally used the multilevel semantic supervision (MSS) module to obtain more refined segmentation results. RV-GAN (Kamran et al., 2021) used a generative network to perform blood vessel segmentation. It employed two generators and two multi-scale discriminators for microvessel segmentation. In addition, it replaced the original adversarial loss with a new weighted loss.

However, the abovementioned methods are more focused on obtaining accurate prediction probability maps rather than binary segmentation features. But only increasing the accuracy of probability maps is very limited for the ability to improve the accuracy of segmentation. In addition, existing methods do not predict thick and thin vessels separately although they have different characteristics, which also leads to the relative neglect of improving accuracy on thin blood vessel segmentation. Therefore, we propose a specific method to skillfully fuse prediction results from original, thick, and thin vessels.

In the task of retina vessel segmentation, there are many difficulties such as a low contrast between blood vessels and background, and central bright band in vessels and the lesion area around blood vessels, as shown in **Figure 1**. But segmenting vascular ends and thin blood vessels is the most difficult part. As we all know, the proportion of thin blood vessels in a retina image is small. So in the deep learning method, the misclassification or omission of some thin vessel pixels does not greatly affect the segmentation accuracy, but it leads to unsatisfactory segmentation maps, which causes the network to pay more attention to segment thick vessels than thin vessels. Inspired by this limitation, according to the original label and the other two extra training objectives made by ourselves, we train original, thick, and thin vessels separately. Thus, we can obtain three different prediction probability maps. Then we use a special fusion method instead of directly choosing a fixed threshold to get the final binary segmentation map. Three kinds of prediction probability maps can exert their own strength so that they can complement each other. Experimental results show that our training strategy and fusion mechanism can get excellent performance. Furthermore, our method can transfer our novel training strategy and fusion mechanism to other deep learning models, and they can perform better than before. Therefore, our proposed method can be commonly applied on any kind of deep learning models for retina image segmentation.

2 RELATED WORKS

With the development of computational technologies, various deep learning models have emerged for solving image calculation

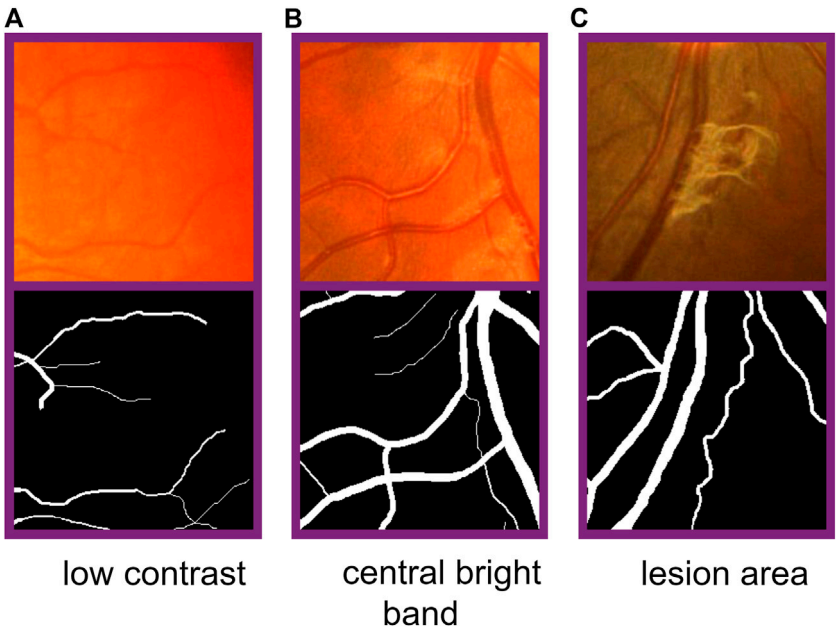


FIGURE 1 | Difficulties in fundus image segmentation task. From the left to right: low contrast between blood vessels and background, central bright band in blood vessels, and lesion area around vessels. From the top to bottom: part of original images and corresponding labels.

problems. AlexNet (Krizhevsky et al., 2012), which won the 2012 ImageNet competition, should be regarded as the first deep learning convolution neural network. It can extract higher dimensional features of images than LeNet (Lecun et al., 1998). The network structure was composed of eight layers, including five convolution layers and three fully connected layers. It also introduced the popular activation function ReLU (Rectified Linear Unit) and the Dropout layer used to prevent over-fitting. Also, VGGNet (Simonyan and Zisserman, 2014) is a very famous deep convolution neural network, which won the runner-up of the 2014 ILSVRC competition. It explored the relationship between the network depth and performance of the convolution neural network. What is more, ResNet (He et al., 2016) that won several championships in ILSVRC 2015 and COCO 2015 was dedicated to solve the model degradation problem caused by the deepening of the network during the layer stacking process. It used a shortcut connection to add the output of previous layers to the current output of this layer, and then the sum can be put into the activation function as the final output of this layer. It was proven that ResNet can effectively alleviate the problem of vanishing gradients.

In the field of image segmentation, there are still some methods that perform very well. The first to mention is fully convolutional networks (FCNs) (Shelhamer et al., 2017), which is a landmark invention in the field of image segmentation. It creatively replaced fully connected layers of the CNN with convolution layers. The FCN classified images at the pixel level, accepted images of any size, and obtained the output with the same size; thereby, it can solve the problem of image segmentation at the semantic level. There was also another characteristic, the skip-level structure, which can take into account local and global information simultaneously. Next one is SegNet (Badrinarayanan et al., 2017), a deep network to solve the problem of image semantic segmentation for autonomous driving or intelligent robots. Based on the semantic segmentation model of the FCN, the framework of VGG16 was used, and it removed the fully connected layer to build an encoder–decoder symmetrical structure to achieve end-to-end pixel-level image segmentation. One of its highlights was the use of max-pooling indexes, which can reduce the amount of parameters for end-to-end training and can be incorporated into any encoding–decoding architecture with only a few modifications. Finally, there is Mask R-CNN (He et al., 2017) that won the championship of COCO 2016 competition; it can perform instance segmentation while performing target detection. Mask R-CNN was based on Faster R-CNN (Ren et al., 2017). In Mask R-CNN, the FCN was used in semantic segmentation for each proposal box of Faster R-CNN. In addition, another important change was the replacement of the ROI pooling module of Faster R-CNN with a more accurate ROI align module.

3 METHODOLOGY

In this article, we propose a novel deep learning framework for fundus diseases diagnosis. First, the fundus images were preprocessed and divided into patches. Second, we perform multi-objective optimization on the network. Explaining in detail, given an image, there is the original label, and then the extra thick and thin vascular training objectives can be obtained by our own algorithms. Based on three different annotation sets, we can obtain the prediction probability maps of original, thick,

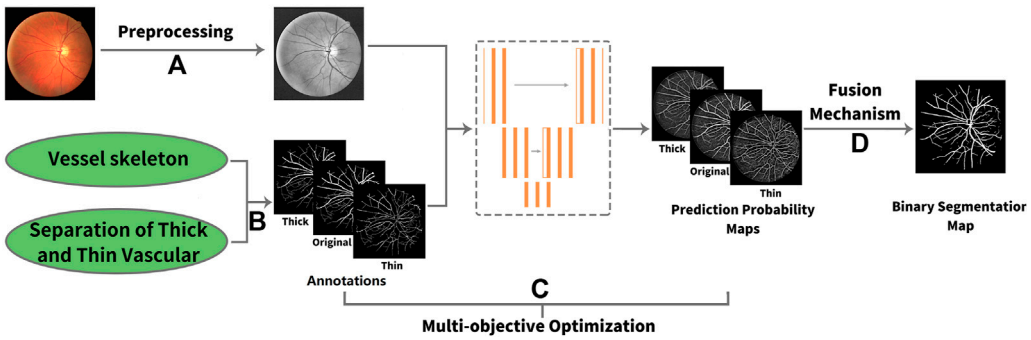


FIGURE 2 | Framework of our method. **(A)** Process of preprocessing. **(B)** Original label, thick, and thin training objectives. **(C)** Multi-objective optimization. Three types of annotations lead to three different prediction probability maps. **(D)** Fusion mechanism.

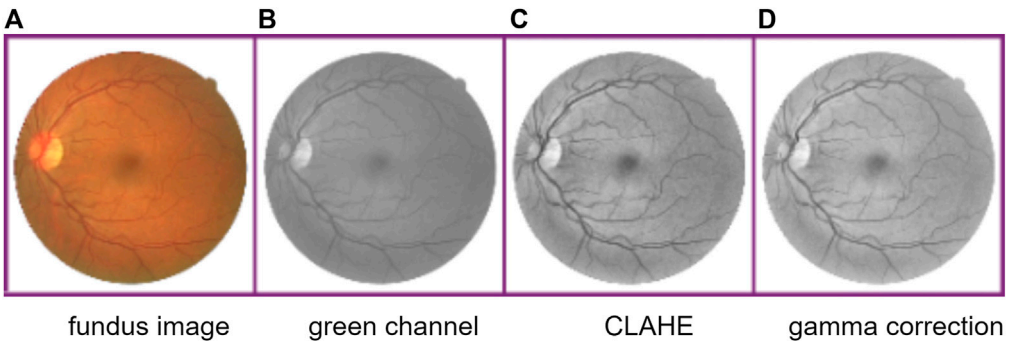


FIGURE 3 | Process of image preprocessing. From the left to right: fundus retina image, image of green channel, CLAHE-processed image, and gamma correction processed image.

and thin vessels. Finally, our new fusion mechanism fuses three different prediction probability maps to obtain the final binary segmentation map. The framework of our method is shown in **Figure 2**.

3.1 Preprocessing

The fundus retina blood vessel images have uneven brightness, image noises, and low contrast between vessels and background. Thus, similar to other methods, we first extract the green channel of original RGB images because the green channel has the highest image contrast. Second, different images are normalized. Third, we apply contrast limited adaptive histogram equalization (CLAHE) (Zuiderveld, 1994) on these images so that the background brightness of images can be equalized without magnifying the image noise. Finally, the gamma correction is used to compress highlight portion and expand dark portion of images. The process of preprocessing is shown in **Figure 3**.

3.2 Network Architecture

U-net Ronneberger et al. (2015) has outstanding performances in medical image segmentation tasks. It was proposed at a medical image conference (MICCAI) in 2015. As shown in **Figure 4**, U-net had a symmetrical encoder–decoder structure. It performed feature extraction in the encoding stage. It uses convolution to gradually extract features of different depths. The convolution operation can be formulated as follows:

$$ConV_k = \mathrm{ReLU}\left\{\sum_{n=0}^{N}\sum_{n=1}^{N} W(k)_{n,n} * X\right\}. \quad (1)$$

Here, N is the size of weight matrix W, X is the tensor from last layer, and ReLU is the rectified linear function; its expression is as follows:

$$ReLU = \begin{cases} x & \mathrm{x} \geq 0 \\ 0 & \mathrm{x} < 0 \end{cases}, \quad (2)$$

Then the pixel-level classification can be obtained in the decoding stage *via* up-sampling operation. U-net also used skip-connection to concatenate features of the corresponding layers of the encoder and decoder on the channel dimension (see the gray arrows in **Figure 4**) so that deep semantic information and shallow representation information can be combined to make the segmentation results more refined. U-net performs very well in segmentation tasks of various organs, especially for fundus blood vessels; it can segment almost all thick blood vessels and most of thin blood

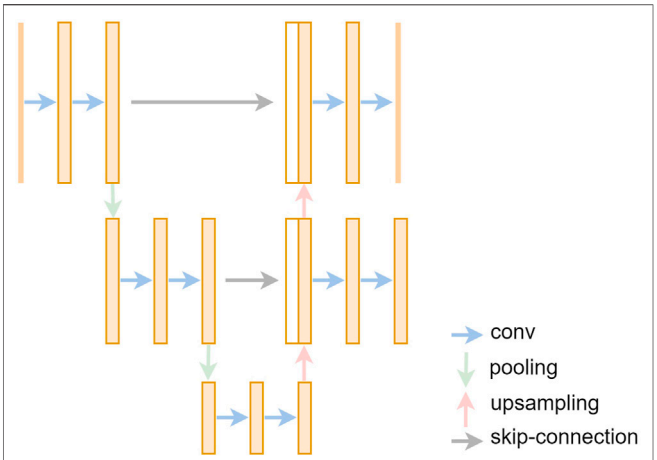


FIGURE 4 | The structure of U-net.

vessels, achieving high accuracy. So we choose a U-net–based network as our segmentation model.

3.3 Multi-Objective Optimization

We construct a train model for original vessel images using preprocessed raw images, original vessel labels, and masks. Then, prediction probability maps of original vessels can be identified. Similarly, in order to achieve the specialized training on thin vessels and thick vessels separately, we use thin (or thick) vessel training objectives made by ourselves, which can emphasize thin (or thick) vessels relatively during the training process. Thus, thin (or thick) vessel training objectives, preprocessed images, and masks can be used as an input, and we can obtain prediction probability maps of thin (or thick) vessels.

Here, the thin (or thick) vessel training objectives are composed of two parts: thin (or thick) blood vessels and vessel skeleton, as shown in **Figure 5**. For the thick vessel training objective, we retain the thick vessels of the original label image, remove the thin blood vessels, and replace it with the vessel skeleton. In the same way, the thin vessel training objectives are composed of the thin vessels and vessel skeleton. The reason why we use skeleton to replace the removed vessels is to preserve the complete vessel topology and keep vessels consistent during the training process.

3.3.1 Vessel Skeleton

There are many existing methods to obtain the skeleton of an object in an image (Zhang and Suen, 1984; Saeed et al., 2010). Due to the uneven thickness and tortuosity of retina blood vessels, some previous methods are not very suitable for obtaining the skeleton of retina vessels. However, the vessel skeleton is needed as a vascular topology consisting of the centerline of blood vessels.

Here, we get the retina vessel skeleton as follows. First, finding the outline of blood vessels. Through searching points with a pixel value of zero around each pixel of blood vessels, the boundary pixels of vessels can be identified. From this, we can define the outline of vascular. In other words, the outline is made up of all the boundary pixels, as shown in **Figure 6**. Next, except for the end pixels of blood vessels, rest of outline pixels are removed, having a new outline according to the above method. Our skeleton method keeps iterating this process until the vessel skeleton is obtained, that is, the horizontal width of the remaining blood vessels is less than two pixels. The vessel skeleton is shown in **Figure 7**.

3.3.2 Separation of Thick and Thin Vessels

Because blood vessels gradually taper from the root to ends, the width of the vessel is gradually reduced. For each pixel in blood vessels, we match a vascular width to it, and then thick and thin vessels can be divided on the pixel level by the vascular width of

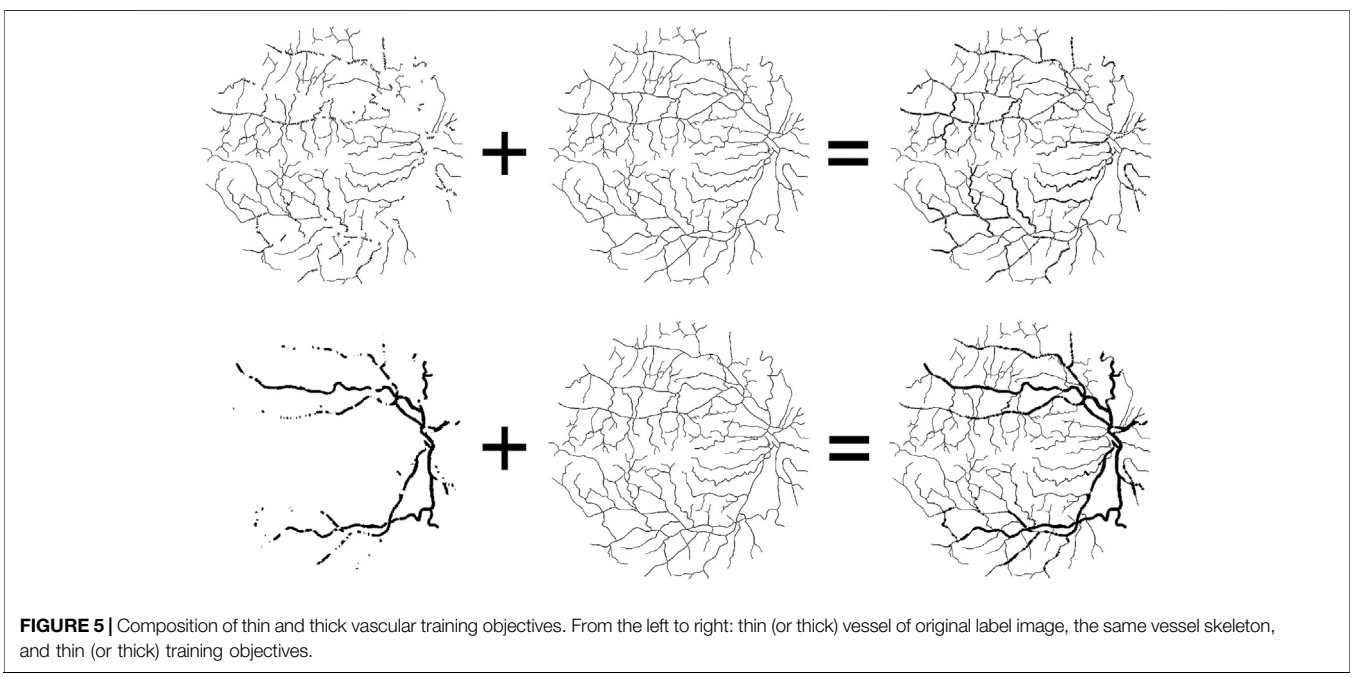


FIGURE 5 | Composition of thin and thick vascular training objectives. From the left to right: thin (or thick) vessel of original label image, the same vessel skeleton, and thin (or thick) training objectives.

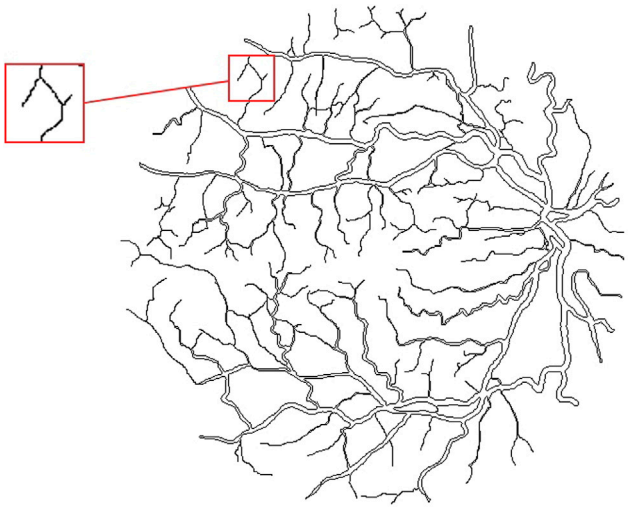


FIGURE 6 | Outline of vessel. An exemplar of non-deletable points is shown in the red box: the vessel in this segment is only about two pixels wide. If removing the outline of this segment, the corresponding vessel in this segment will disappear.

every pixel. If the vascular width of any pixel is greater than or equal to the threshold, we define it as a thick vessel pixel, and naturally the opposite as thin vessel pixels. In detail, the vascular width at a certain point is defined as follows. First, we identify the vascular width of skeleton pixels. We define the twice distance from each skeleton pixel to the nearest outline pixel as the vascular width of this skeleton pixel. Second, we define the vascular width for other vascular pixels. The vascular width of the nearest skeleton pixel from them is used to replace their width. In this way, we can match the vascular width for each pixel. Thereby, thick and thin blood vessels are separated based on the vascular width of each pixel. In our experiments, the separation threshold is 2.2 pixels. This means that those with a width greater than or equal to 2.2 pixels are thick vessel pixels, and those with width less than 2.2 pixels are thin vessel pixels. Due to the slight uneven thickness of the blood vessels, this may lead to the intersection of separated thick and thin vascular pixels and unsmooth appearance of separation profile.

3.4 Fusion Mechanism

We design three fusion methods for prediction probability maps of original, thick, and thin vessels, and then adopt one of them that performs the best. For this method, we apply the pixel-wise classification on prediction probability maps. At a certain pixel point, if the pixel probability of one of three probability maps is greater than or equal to the threshold 0.5, it is defined as a vascular pixel. Furthermore, since the original vascular label is more complete, it has more complete segmentation map, accordingly. So during the fusion process, there is a greater weight on prediction probability maps of original vessels, and the best weight we get on the data set is 1.25; that is, k is 1.25. Therefore, we are relatively more strict with the application of prediction probability maps of thick and thin vessels, as shown in **Algorithm 1**.

Algorithm 1 Fusion Mechanism: The fusion of original, thick and thin prediction probability maps.

Input: prediction probability maps of original vessels P_{org}, thick vessels P_{thick} and thin vessels P_{thin}; threshold th; size of maps $M \times N$.
Output: final binary segmentation map A_{out}.

function FUNCTION FUSION($P_{org}, P_{thick}, P_{thin}$)
  for $i = 1 : M$ **do**
    for $j = 1 : N$ **do**
      Calculate binary segmentation maps:
        $A_{org}[i][j] = (k * P_{org}[i][j] > 0.5)? 1 : 0;$
        $A_{thick}[i][j] = (P_{thick}[i][j] > 0.5)? 1 : 0;$
        $A_{thin}[i][j] = (P_{thin}[i][j] > 0.5)? 1 : 0;$
      Calculate fusion segmentation map:
        $A_{out}[i][j] = \max(A_{org}[i][j], A_{thick}[i][j], A_{thin}[i][j]);$
    end for
  end for
end function

4 EXPERIMENTS

In order to verify the validity of our method, we perform experiments on three datasets: DRIVE, STARE, and IOSTAR. Our experiments are implemented on Keras based on TensorFlow with GeForce RTX2080 Ti GPU. The network models for three tasks use the same parameter settings.

4.1 DRIVE Dataset

The DRIVE dataset contains 40 fundus images, corresponding labels, and binary field of view (FOV) masks. The images have the resolution of 584 × 565 pixels and 45° FOV. We use the standard split: first 20 images belong to the test set and rest 20 images belong to the training set. Each image of the training set has one manual annotation, and each image of the test set has two manual annotations. In our experiments, the first manual annotation is used to be the gold standard.



FIGURE 7 | Vessel skeleton. From the left to right: fundus retina image, ground truth, and vessel skeleton obtained by our method.

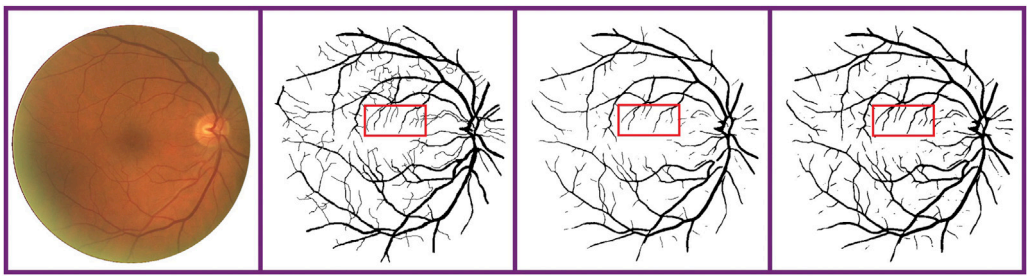


FIGURE 8 | Comparison of a normal FCN-based model and our proposed method on the FCN-based model. From the left to right: fundus retina image, ground truth, binary segmentation map of the normal FCN-based model and our proposed method on the FCN-based model. Obviously, our method can segment more thin blood vessels.

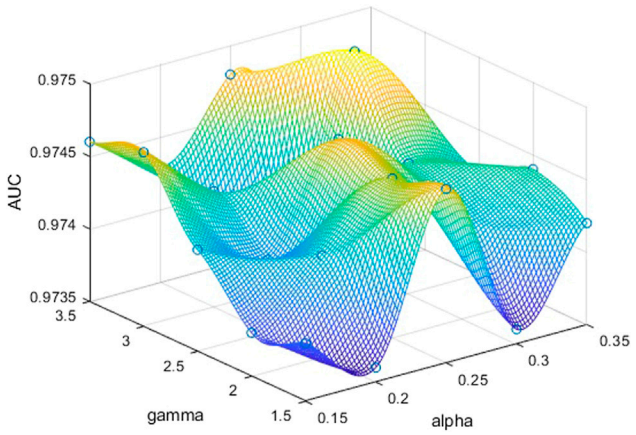


FIGURE 9 | AUCs obtained by binary focal loss on DRIVE according to different α and γ. We can obtain the best performance when α is 0.3 and γ is 3.0.

4.2 STARE Dataset

The STARE dataset contains 20 fundus images and corresponding labels (Hoover et al., 2000). The images have the resolution of 605 × 700 pixels, 8 bits per color channel. This dataset does not contain FOV masks, so we use the masks provided by Marin et al. (2011) for comparison. Due to the small size of the dataset, we use leave-one-out method: select 19 images for training at a time, leave one image for testing, and then calculate the average of various metrics on 20 images as the final results. Similarly to other methods, we use the manual annotation by the first observer as our ground truth.

4.3 IOSTAR Dataset

The IOSTAR dataset contains 30 fundus images, corresponding labels, binary field of view (FOV) masks, and the optic disc (OD) masks. The images have the resolution of 1024 × 1024 pixels and 45° FOV. Since the annotations of vessels within the OD are not available as stated on the dataset website, we take the official recommendation and use the OD mask for the evaluation of the retinal vessel segmentation. To make it easy for comparison, we also use five-fold cross-validation to train and test our model; that is, we select 24 images as the training set each time, then use the remaining six images to test, and finally use the average of the five tests as the final result.

5 RESULTS AND DISCUSSION

In order to evaluate the segmentation method proposed in this article, we test on three datasets: DRIVE, STARE, and IOSTAR. First, so as to prove the model independence of our method, U-net is replaced with the FCN and the relevant experiments performed. Second, we compare the performances of four loss functions and list the relevant comparison results. Third, the other two fusion methods are introduced, and we compare them to our adopted fusion method. Then to verify the robustness of our method, we perform cross-training on two datasets: DRIVE and STARE. Finally, we compare our segmentation results with some existing methods.

5.1 Evaluation Metrics

Through the comparison between segmentation map and ground truth, the pixels which are in the segmentation map can be divided into the following four categories: correctly classified as positive (TP), correctly classified as negative (TN), incorrectly classified as positive (FP), and incorrectly classified as negative (FN). We use some general evaluation metrics such as Acc (accuracy), Se (sensitivity), Sp (specificity), and F1 score, as follows:

$$Acc = \frac{TP + TN}{TN + TP + FN + FP}, \tag{3}$$

$$Se = \frac{TP}{TP + FN}, \tag{4}$$

$$Sp = \frac{TN}{TN + FP}, \tag{5}$$

$$F1 = 2 \times \frac{TP}{TP + FP + FN}. \tag{6}$$

To further evaluate the effectiveness of our method, we also calculate the area under the receiver operating characteristics curve (AUC).

5.2 Performance on Fully Convolutional Networks

To verify the model independence of our method, we replace U-net with the FCN model; that is to say, we transfer our training strategy and fusion mechanism to the FCN: first, training the

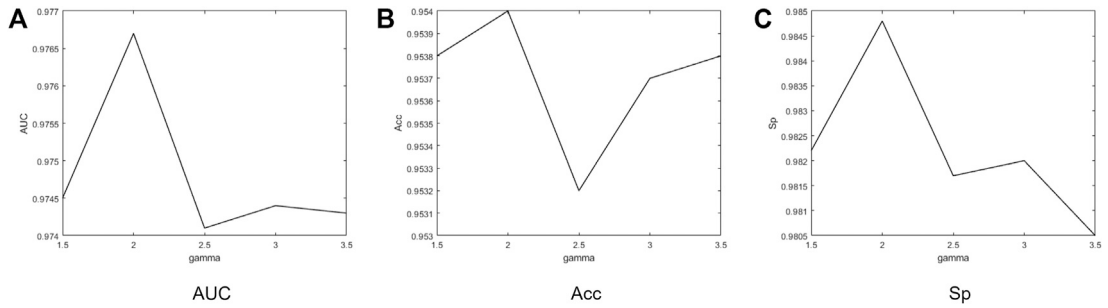


FIGURE 10 | Performance obtained by categorical focal loss on DRIVE according to different γ. When γ is 2.0, the best values of AUC, Acc, and Sp can be obtained.

TABLE 1 | Performance on four loss functions.

Vessel Type	Loss	Se	Sp	F1	Acc	AUC
Original	Binary cross entropy	0.7586	0.9820	0.8061	0.9535	0.9750
	Categorical cross entropy	0.7074	**0.9867**	0.7866	0.9511	0.9735
	Binary focal	0.7637	0.9815	0.8079	0.9538	0.9750
	Categorical focal	**0.7677**	0.9816	**0.8107**	**0.9544**	**0.9782**
Thick	Binary cross entropy	**0.6917**	0.9791	0.7387	0.9493	0.9665
	Categorical cross entropy	0.6317	0.9825	0.7087	0.9462	0.9603
	Binary focal	0.6721	0.9814	0.7333	0.9494	0.9653
	Categorical focal	0.6846	**0.9877**	**0.7404**	**0.9559**	**0.9684**
Thin	Binary cross entropy	**0.6324**	0.9793	**0.6801**	0.9504	**0.9606**
	Categorical cross entropy	0.6091	0.9768	0.6536	0.9462	0.9506
	Binary focal	0.6233	0.9805	0.6785	0.9508	0.9589
	Categorical focal	0.6120	**0.9817**	0.6751	**0.9509**	0.9582

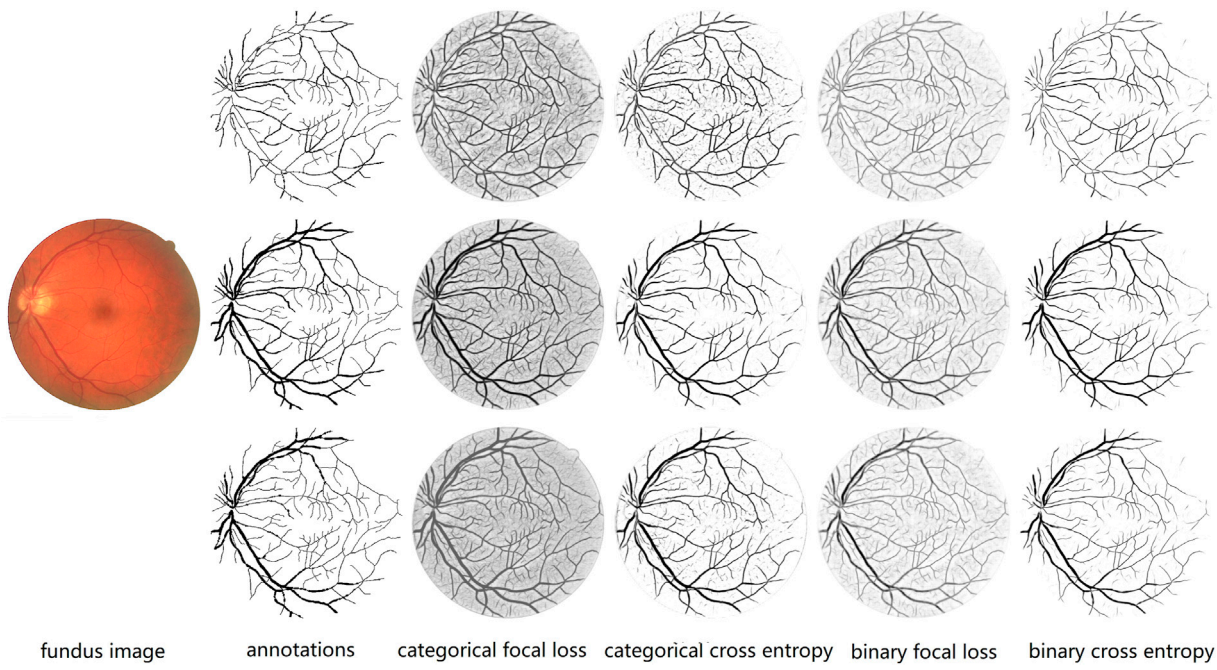


FIGURE 11 | Comparison of different loss functions. From the left to right: fundus retina image, annotations of thin, original and thick vessels, prediction probability maps of thin, original, and thick vessels through categorical focal loss, categorical cross entropy, binary focal loss, and binary cross entropy, respectively. We can notice that categorical cross entropy is the best one for our proposed method because it can help us segment blood vessels more clearly, especially thin blood vessels.

TABLE 2 | Performance on the three fusion mechanisms.

Fusion	Se	Sp	F1	Acc
1st	0.8191	**0.9831**	**0.8201**	**0.9685**
2nd	0.8158	0.9820	0.8138	0.9673
3rd	**0.8201**	0.9821	0.8158	0.9677

TABLE 3 | Cross-training results on DRIVE and STARE datasets.

Dataset	Method	Se	Sp	Acc	AUC
DRIVE (trained on STARE)	Marin et al.	—	—	0.9448	—
	Fraz et al.	0.7242	0.9792	0.9456	**0.9697**
	Li et al.	0.7273	0.9810	0.9486	0.9677
	Yan et al.	0.7292	0.9815	0.9494	0.9599
	Ours	**0.7644**	**0.9843**	**0.9651**	0.9689
STARE (trained on DRIVE)	Marin et al.	—	—	0.9528	—
	Fraz et al.	0.7010	0.9770	0.9495	0.9660
	Li et al.	0.7027	0.9828	0.9545	0.9671
	Yan et al.	0.7211	0.9840	0.9569	**0.9708**
	Ours	**0.7713**	**0.9850**	**0.9684**	0.9694

TABLE 4 | Comparison of thin vessels segmentation.

Method	Se	Sp	Acc
Yan et al. (2019)	**0.8170**	0.8115	0.8127
Our proposed	0.7677	**0.9779**	**0.9602**

original vessel images on the FCN-based model, then just as common methods, taking 0.5 as the threshold of the prediction probability map to turn it into a binary segmentation map. For comparison, we train original, thick, and thin vessels separately on the FCN-based model, and then apply our proposed fusion method to get a final segmentation map. We perform this comparison experiment on the DRIVE dataset, and related experimental results are shown in **Figure 8**. It can be found that our novel method can segment more vessel ends than the original normal FCN model. We achieve 0.8082, 0.9830, 0.8141, and 0.9677 on Se, Sp, F1-score, and Acc, respectively, while the normal FCN achieves 0.7718, 0.9795, 0.8072, and 0.96535 on Se, Sp, F1-score, and Acc, respectively.

5.3 Comparison of Loss Functions

We compare four loss functions on the basis of our framework, namely, binary cross entropy, categorical cross entropy, binary focal loss, and categorical focal loss. In deep learning methods for retinal vessel segmentation, the cross-entropy loss function is generally used. For the binary classification problem such as blood vessel segmentation, the first choice we think of is binary cross entropy. This loss can be defined as follows:

$$L_{bce} = \begin{cases} -\log y', & \mathrm{y} = 1 \\ -\log\left(1 - y'\right), & \mathrm{y} = 0 \end{cases}. \tag{7}$$

Here, y is the ground truth and y' is the prediction.

At the same time, for such problems, we can also use categorical cross entropy loss; the expression is shown as follows:

$$L_{ce} = -y*\log y', \tag{8}$$

where y is the ground truth and y' is the prediction.

The performances of above two cross-entropy loss functions do not have absolute strengths or weaknesses. According to the expression, it can be found that for all pixels in an image, whether they belong to foreground or background, the cross-entropy loss treats them all the same. Therefore, in the retina vessel segmentation tasks, even if thin vascular pixels with the small portion are not well segmented, cross-entropy loss will not be very high. So, it would place the emphasis on thick vessels and ignore thin vessels relatively.

In order to solve above problems, the focal loss (Lin et al., 2017) is a more appropriate choice. It can improve the accuracy of difficult segmented pixels. Focal loss can adjust a loss through two parameters, α and γ. α is the weighting factor; it can control the contribution of positive and negative samples to the total loss. And γ is the focusing parameter; its purpose is to reduce the weight of samples that are easy to classify so that the model can focus more on samples that are difficult to be classified during the training process. Inspired by the cross-entropy loss, we also try two different focal losses: binary focal loss and categorical focal loss. The expressions are as follows:

$$L_{binary_focal} = \begin{cases} -\alpha*\left(1 - y'\right)^{\gamma}*\log y', & \mathrm{y} = 1 \\ -(1-\alpha)*y'^{\gamma}*\log\left(1 - \mathrm{y}'\right), & \mathrm{y} = 0 \end{cases}, \tag{9}$$

$$L_{categorical_focal} = -\alpha*\log y'*\left(1 - \mathrm{y}'\right)^{\gamma}. \tag{10}$$

Here, y is the ground truth and y' is the prediction.

Here, α can play the role of balancing class. In categorical focal loss, α is a 2-dimensional vector because our task is a binary classification. In the binary focal loss, α is a fixed value.

In order to further explore the role of α and γ in two focal loss functions, we take different values of α and γ to determine the best set of parameters. The experiments are performed on the DRIVE dataset. For binary focal loss, we compare α and γ, respectively. In the original article, the best α and γ are 0.25 and 2.5, respectively. Based on this, the range of α is 0.15 – 0.35, and the range of γ is 1.5 – 3.5, as shown in **Figure 9**. Since other metrics are related to the threshold of binary images, we chose to compare AUC that is just related to the predicted probability value. When α is 0.3 and γ is 3.0, the maximum AUC is 0.9750. Otherwise, there is no big difference between AUCs obtained by different parameters. For the categorical focal loss, we set α to 0.25 for positive samples directly and compare the performances of different γ. As shown in **Figure 10**, we can see that when γ is 2.0, the best values of AUC, Acc, and Sp can be obtained. So we can basically conclude that in our experiments, categorical cross entropy performs best when γ is 2.0.

We also compare the performances of four types of loss functions on original, thick, and thin blood vessels. For controlling variables, 0.5 is simply used as the threshold for segmentation instead of our proposed fusion method. For two types of focal loss, we use best parameters discussed above; that is,

TABLE 5 | Comparison of different retina vessel segmentation methods on DRIVE, STARE and IOSTAR datasets.

Dataset	Method	Se	Sp	F1	Acc	AUC
DRIVE	2nd observer	0.7760	0.9724	—	0.9472	—
	Zhang et al. (2010)	0.7120	0.9724	—	0.9382	—
	Marin et al. (2011)	0.7067	0.9801	0.7690	0.9452	0.9588
	Fraz et al. (2012)	0.7406	0.9807	—	0.9480	0.9747
	Roychowdhury et al. (2015)	0.7395	0.9782	—	0.9494	0.9672
	Li et al. (2016)	0.7569	0.9816	—	0.9527	0.9738
	Yan et al. (2018)	0.7653	0.9818	—	0.9542	0.9752
	Yan et al. (2019)	0.7631	0.9820	—	0.9538	0.9750
	Liu et al. (2019)	0.8072	0.9780	0.8225	0.9559	0.9779
	Wang et al. (2019)	0.7940	0.9816	**0.8270**	0.9567	0.9772
	Li et al. (2020)	0.7791	0.9831	0.8218	0.9574	0.9813
	Wu et al. (2020)	0.7996	0.9813	—	0.9582	**0.9830**
	Ours	**0.8140**	**0.9847**	0.8247	**0.9697**	0.9782
STARE	2nd observer	0.8952	0.9384	—	0.9349	—
	Zhang et al. (2010)	0.7177	0.9753	—	0.9484	—
	Marin et al. (2011)	0.6944	0.9819	0.7531	0.9526	0.9769
	Fraz et al. (2012)	0.7548	0.9763	—	0.9534	0.9768
	Roychowdhury et al. (2015)	0.7317	0.9842	—	0.9560	0.9673
	Yin et al. (2015)	**0.8541**	0.9419	—	0.9325	-
	Li et al. (2016)	0.7726	0.9844	—	0.9628	**0.9879**
	Yan et al. (2018)	0.7581	0.9846	—	0.9612	0.9801
	Yan et al. (2019)	0.7735	0.9857	—	0.9638	0.9833
	Liu et al. (2019)	0.7771	0.9843	0.8036	0.9623	0.9793
	Li et al. (2020)	0.7715	**0.9886**	0.8146	0.9701	0.9881
	Wu et al. (2020)	0.7963	0.9863	—	0.9672	0.9875
	Ours	0.8251	0.9859	**0.8239**	**0.9737**	0.9821
IOSTAR	Ronneberger et al. (2015)	0.8044	0.9793	—	0.9675	0.9464
	Azzopardi et al. (2015)	0.7610	0.9670	—	0.9410	0.9550
	Zhao et al. (2018)	0.7720	0.9670	—	0.9480	0.9600
	Alom et al. (2018)	0.8042	0.9779	—	0.9652	0.9530
	Gu et al. (2019)	0.8110	0.9749	—	0.9572	0.9658
	Fu et al. (2019)	0.8298	0.9832	—	0.9720	0.9504
	Mou et al. (2021)	**0.8341**	0.9831	—	**0.9722**	0.9758
	Ours	0.7998	**0.9847**	0.8059	0.9702	**0.9788**

γ is 2.0 in categorical focal loss, and γ is 3.0 and α is 0.3 in the binary focal loss. As described in **Table 1**, for original blood vessels, except for categorical cross entropy with the best performance on Sp, the categorical focal loss realizes the best performances in the remaining metrics. For thick blood vessels, similarly, except for binary cross entropy with the best performance on Se, categorical focal loss achieves the best performances on Sp, F1-score, Acc, and AUC. However, for thin blood vessels, binary cross entropy performs the best on Se, F1-score, and AUC, while categorical focal loss achieves good results on Sp and Acc. In general, categorical cross entropy is the best one for our proposed method. Therefore, we choose categorical focal loss as our loss function. The performances of four loss functions are shown in **Figure 11**. At the same time, it can be noticed that when we train thick and thin vascular images, all results we get are lower than those of original vascular images. This is mainly because training objectives of thick and thin vessels do not exactly coincide with their original retina images.

5.4 Comparison of Fusion Methods

We design two other fusion methods to compare with the first method mentioned earlier. In the other two methods, for a certain pixel in the image, its prediction values are first checked in the original, thick, and thin vascular prediction probability maps. Then for the second fusion method, if two of three prediction values are greater than or equal to the threshold 0.5, the pixel is determined as positive, that is, a blood vessel pixel, and its pixel value is set to 255 to obtain a binary segmentation map. And for the third method, we calculate the average of three prediction pixel values; if the average value is greater than or equal to the threshold 0.5, the pixel is defined as a vascular pixel. Comparison of three fusion methods is shown in **Table 2**. We can find that although the first fusion method is slightly lower than the third fusion method on Se, it performs best on Sp, F1-score, Acc, and AUC. So we finally chose the first fusion method as our fusion mechanism.

5.5 Cross-Training on Different Datasets

To test the robustness of our method, we perform cross-training on two datasets: DRIVE and STARE. Similar to other methods (Yan et al., 2018), we first use the STARE dataset for training, and then test on the DRIVE dataset. In the same way, in turn, we use the DRIVE dataset for training and the STARE dataset for testing. The performances of cross-training are shown in **Table 3**. It can be found that for DRIVE (trained on STARE), our method obtained best results on Se, Sp, and

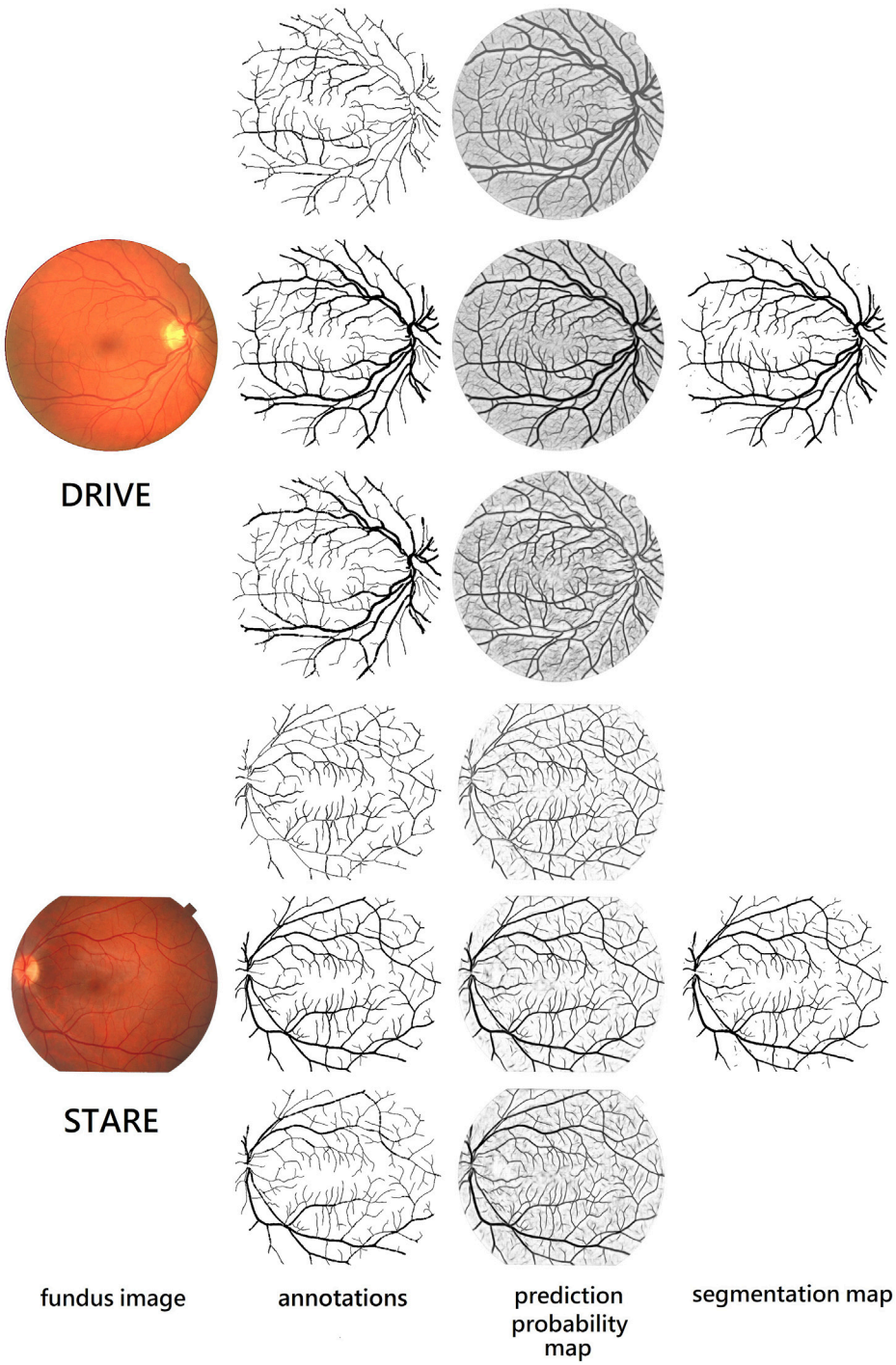


FIGURE 12 | Performance on retina vessel segmentation. From the top to bottom: images from DRIVE and STARE. From the left to right: fundus retina images, ground truths of thin, original, and thick vessels, and prediction probability maps of thin, original, and thick vessels, and binary segmentation map.

Acc. And for STARE (trained on DRIVE), our method still obtained best results on Se, Sp, and Acc. However, we can also notice that results obtained by DRIVE (trained on STARE) are generally lower than those obtained by STARE (trained on DRIVE). Due to differences in manual annotations of two datasets, the model trained on STARE is relatively weak in segmenting images of DRIVE datasets because it is not able to detect thin vessels well.

TABLE 6 | Comparison of standard deviations on the F1-score.

Method	F1-mean	F1-SD
Normal U-net	**0.8157**	**0.02345**
Ours based on U-net	0.8247	0.01744

5.6 Segmentation Results

We analyze the segmentation results of our method on the thin vessels and compare the results with those obtained by Yan et al. (2019). In order to make it easy for comparison, as with their method, we separate the thick vessels and thin vessels with a boundary of three pixels, and then we calculate Se, Sp, and Acc of the thin vessel segmentation results. From **Table 4**, we can see that our Se is lower than the result of the study by Yan et al. (2019), but our Sp and Acc are significantly higher than them, which shows that we have improved the segmentation of thin vessels.

We also compare different retina vessel segmentation methods on three datasets, as shown in **Table 5**. The performances of our proposed method for segmentation retina images are shown in **Figure 12**. For DRIVE, we achieve 0.9697, 0.8140, 0.9847, 0.8247, and 0.9782 on Acc, Se, Sp, F1-score, and AUC, respectively. Compared to the current excellent experimental results, we surpass them on Acc, Se, and Sp. But for AUC, we differ from the best result, 0.9830 of Wu et al. (2020), by 0.0048. And for the F1-score, our result is lower than that obtained using DEU-net (Wang et al., 2019) by 0.0023, which achieves 0.8270. For STARE, as mentioned above, we use FOV masks generated by Marin et al. (2011). Our framework obtains 0.9737, 0.8251, 0.9859, 0.8239, and 0.9821 on Acc, Se, Sp, F1-score, and AUC, respectively. Compared with many existing methods, our method surpasses current state-of-the-art methods on F1-score and Acc. But for Se, Sp and AUC, we are slightly lower. Among them, AUC is 0.0058 lower than the current best result (Li et al., 2016). For the IOSTAR dataset, we get 0.7998, 0.9847, 0.8059, 0.9702, and 0.9788 on Se, Sp, F1-score, Acc, and AUC, respectively. Compared to the current outstanding experimental results, we surpass them on Sp and AUC. Especially for AUC, we are higher than CS^2-net (Mou et al., 2021) by 0.003, which achieves 0.9758.

In addition, we also calculate the standard deviation of the F1-score obtained by using our method based on U-net and using only U-net on the DRIVE test set. According to **Table 6**, our standard deviation is 0.01744, while U-net's standard deviation is 0.02345. This indicates that the F1-score predicted by our method fluctuates less than using U-net alone. For different fundus images, our model is less affected by differences in image quality; thus, we can obtain more stable segmentation maps, which also shows that our method has high reliability.

6 CONCLUSION

This study proposes a novel deep learning method to train original, thick, and thin vessels. At the same time, we design algorithms for extracting vessel skeleton and separating thick or thin blood vessels. Importantly, we make a novel fusion mechanism that can fuse prediction probability maps from three different types of vessels to obtain final binary segmentation map. Experimental results indicate that our proposed method has outperformed many current outstanding retina vessel segmentation methods on DRIVE, STARE, and IOSTAR datasets. The effectiveness and robustness with different image conditions can make this blood vessel segmentation proposal suitable for retinal image computer analyses such as automated screening for early diabetic retinopathy detection.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

JD conceived and designed the experiments; JD and ZZ performed the experiments and analyzed the data; JD and FG wrote the article. FG and JT supervised the experiments and reviewed the manuscript. All authors have participated in study discussion and manuscript preparation.

FUNDING

This work was supported by a grant from the National Natural Science Foundation of China (NSFC 61772362, 61972280) and National Key R&D Program of China (No. 2020YFA0908400).

REFERENCES

Alom, M. Z., Hasan, M., Yakopcic, C., Taha, T., and Asari, V. (2018). Recurrent residual U-Net for medical image segmentation. *J. Med. Inform.* 1, 014006. doi:10.1117/1.JMI.6.1.014006

Annunziata, R., Garzelli, A., Ballerini, L., Mecocci, A., and Trucco, E. (2016). Leveraging Multiscale Hessian-Based Enhancement with a Novel Exudate Inpainting Technique for Retinal Vessel Segmentation. *IEEE J. Biomed. Health Inform.* 20, 1129–1138. doi:10.1109/jbhi.2015.2440091

Azzopardi, G., Strisciuglio, N., Vento, M., and Petkov, N. (2015). Trainable Cosfire Filters for Vessel Delineation with Application to Retinal Images. *Med. Image Anal.* 19, 46–57. doi:10.1016/j.media.2014.08.002

Badrinarayanan, V., Kendall, A., and Cipolla, R. (2017). Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 2481–2495. doi:10.1109/tpami.2016.2644615

Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., et al. (2012). An Ensemble Classification-Based Approach Applied to Retinal Blood Vessel Segmentation. *IEEE Trans. Biomed. Eng.* 59, 2538–2548. doi:10.1109/tbme.2012.2205687

Fu, H., Xu, Y., Lin, S., Kee Wong, D. W., and Liu, J. (2016). "Deepvessel: Retinal Vessel Segmentation via Deep Learning and Conditional Random Field," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Editors S. Ourselin, L. Joskowicz, M. R. Sabuncu, G. Unal, and W. Wells (Cham: Springer International Publishing), 132–139. doi:10.1007/978-3-319-46723-8_16

Fu, J., Liu, J., Tian, H., Li, Y., Bao, Y., Fang, Z., and Lu, H. (2019). "Dual Attention Network for Scene Segmentation," in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, 15-20 June 2019, 3141–3149. doi:10.1109/CVPR.2019.00326

Gu, L., and Cheng, L. (2014). "Learning to Boost Filamentary Structure Segmentation," in International Conference on Computer Vision (ICCV) 2015, Santiago, Chile, December 7-13, 2015.

Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., et al. (2019). Ce-net: Context Encoder Network for 2d Medical Image Segmentation. *IEEE Trans. Med. Imaging* 38, 2281–2292. doi:10.1109/TMI.2019.2903562

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask R-Cnn," in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October 22-29, 2017, 2980–2988. doi:10.1109/iccv.2017.322

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep Residual Learning for Image Recognition," in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, June 27-30, 2016, 770–778. doi:10.1109/cvpr.2016.90

Hoover, A. D., Kouznetsova, V., and Goldbaum, M. (2000). Locating Blood Vessels in Retinal Images by Piecewise Threshold Probing of a Matched Filter Response. *IEEE Trans. Med. Imaging* 19, 203–210. doi:10.1109/42.845178

Jin, Q., Meng, Z., Pham, T. D., Chen, Q., Wei, L., and Su, R. (2019). Dunet: A Deformable Network for Retinal Vessel Segmentation. *Knowledge-Based Syst.* 178, 149–162. doi:10.1016/j.knosys.2019.04.025

Kamran, S. A., Hossain, K. F., Tavakkoli, A., Zuckerbrod, S. L., Sanders, K. M., and Baker, S. A. (2021). *Rv-gan: Segmenting Retinal Vascular Structure in Fundus Photographs Using a Novel Multi-Scale Generative Adversarial Network*.

Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). Imagenet Classification with Deep Convolutional Neural Networks. Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, December 2012, 1097–1105.

Laibacher, T., Weyde, T., and Jalali, S. (2018). M2u-net: Effective and efficient retinal vessel segmentation for real-world applications. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 115–124.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based Learning Applied to Document Recognition. *Proc. IEEE* 86, 2278–2324. doi:10.1109/5.726791

Li, L., Verma, M., Nakashima, Y., Nagahara, H., and Kawasaki, R. (2020). "Iternet: Retinal Image Segmentation Utilizing Structural Redundancy in Vessel Networks," in Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Snowmass, CO, March 1-5, 2020. doi:10.1109/wacv45572.2020.9093621

Li, Q., Feng, B., Xie, L., Liang, P., Zhang, H., and Wang, T. (2016). A Cross-Modality Learning Approach for Vessel Segmentation in Retinal Images. *IEEE Trans. Med. Imaging* 35, 109–118. doi:10.1109/tmi.2015.2457891

Lin, T.-Y., Goyal, P., Girshick, R. B., He, K., and Dollár, P. (2017). "Focal Loss for Dense Object Detection," in 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, October 22-29 2017, 2999–3007. doi:10.1109/iccv.2017.324

Liu, B., Gu, L., and Lu, F. (2019). "Unsupervised Ensemble Strategy for Retinal Vessel Segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention (Cham: Springer), 111–119. doi:10.1007/978-3-030-32239-7_13

Lupascu, C. A., Tegolo, D., and Trucco, E. (2010). Fabc: Retinal Vessel Segmentation Using Adaboost. *IEEE Trans. Inform. Technol. Biomed.* 14, 1267–1274. doi:10.1109/titb.2010.2052282

Marin, D., Aquino, A., Gegundez-Arias, M. E., and Bravo, J. M. (2011). A New Supervised Method for Blood Vessel Segmentation in Retinal Images by Using gray-level and Moment Invariants-Based Features. *IEEE Trans. Med. Imaging* 30 (1), 146–158. doi:10.1109/tmi.2010.2064333

Melinscak, M., Prentasic, P., and Loncaric, S. (2015). Retinal Vessel Segmentation Using Deep Neural Networks. *VISIGRAPP, Proc.* 1, 577–582. doi:10.5220/0005513005770582

Mou, L., Zhao, Y., Chen, L., Cheng, J., Gu, Z., Hao, H., et al. (2019). "Cs-net: Channel and Spatial Attention Network for Curvilinear Structure Segmentation," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Editors D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, et al. (Cham: Springer International Publishing), 721–730. doi:10.1007/978-3-030-32239-7_80

Mou, L., Zhao, Y., Fu, H., Liu, Y., Cheng, J., Zheng, Y., et al. (2021). Cs2-net: Deep Learning Segmentation of Curvilinear Structures in Medical Imaging. *Med. Image Anal.* 67, 101874. doi:10.1016/j.media.2020.101874

Nguyen, U. T. V., Bhuiyan, A., Park, L. A. F., and Ramamohanarao, K. (2013). An Effective Retinal Blood Vessel Segmentation Method Using Multi-Scale Line Detection. *Pattern Recognition* 46, 703–715. doi:10.1016/j.patcog.2012.08.009

Orlando, J. I., and Blaschko, M. (2014). "Learning Fully-Connected Crfs for Blood Vessel Segmentation in Retinal Images," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*. Editors P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe (Cham: Springer International Publishing), 634–641. doi:10.1007/978-3-319-10404-1_79

Peter, B., Norman, S. C., Graham, M. J., M., C. T., and Teresa, S.-G. (2012). Fast Retinal Vessel Detection and Measurement Using Wavelets and Edge Location Refinement. *Plos One* 7, e32435. doi:10.1371/journal.pone.0049632

Ren, S., He, K., Girshick, R., and Sun, J. (2017). Faster R-Cnn: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 1137–1149. doi:10.1109/tpami.2016.2577031

Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Editors N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi (Cham: Springer International Publishing), 234–241. doi:10.1007/978-3-319-24574-4_28

Roychowdhury, S., Koozekanani, D. D., and Parhi, K. K. (2014). Dream: Diabetic Retinopathy Analysis Using Machine Learning. *IEEE J. Biomed. Health Inform.* 18, 1717–1728. doi:10.1109/jbhi.2013.2294635

Roychowdhury, S., Koozekanani, D. D., and Parhi, K. K. (2015). Iterative Vessel Segmentation of Fundus Images. *IEEE Trans. Biomed. Eng.* 62, 1738–1749. doi:10.1109/tbme.2015.2403295

Saeed, K., Tabędzki, M., Rybnik, M., and Adamski, M. (2010). K3m: A Universal Algorithm for Image Skeletonization and a Review of Thinning Techniques. *Appl. Maths. Comput. Sci.* 20, 317–335. doi:10.2478/v10006-010-0024-4

Salazar-Gonzalez, A., Kaba, D., Yongmin Li, Y., and Xiaohui Liu, X. (2014). Segmentation of the Blood Vessels and Optic Disk in Retinal Images. *IEEE J. Biomed. Health Inform.* 18, 1874–1886. doi:10.1109/jbhi.2014.2302749

Shelhamer, E., Long, J., and Darrell, T. (2017). Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 39, 640–651. doi:10.1109/tpami.2016.2572683

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. 3rd International Conference on Learning Representations, San Diego, CA, May 7-9, 2014.

Wang, B., Qiu, S., and He, H. (2019). "Dual Encoding U-Net for Retinal Vessel Segmentation," in International Conference on Medical Image Computing and Computer-Assisted Intervention, Shenzhen, China, October 13-17, 2019 (Cham: Springer), 84–92. doi:10.1007/978-3-030-32239-7_10

Wang, D., Haytham, A., Pottenburgh, J., Saeedi, O., and Tao, Y. (2020). Hard Attention Net for Automatic Retinal Vessel Segmentation. *IEEE J. Biomed. Health Inform.* 24, 3384–3396. doi:10.1109/JBHI.2020.3002985

Wu, H., Wang, W., Zhong, J., Lei, B., Wen, Z., and Qin, J. (2021). Scs-net: A Scale and Context Sensitive Network for Retinal Vessel Segmentation. *Med. Image Anal.* 70, 102025. doi:10.1016/j.media.2021.102025

Wu, Y., Xia, Y., Song, Y., Zhang, D., Liu, D., Zhang, C., et al. (2019). "Vessel-net: Retinal Vessel Segmentation under Multi-Path Supervision," in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*. Editors D. Shen, T. Liu, T. M. Peters, L. H. Staib, C. Essert, S. Zhou, et al. (Cham: Springer International Publishing), 264–272. doi:10.1007/978-3-030-32239-7_30

Wu, Y., Xia, Y., Song, Y., Zhang, Y., and Cai, W. (2020). NFN+: A Novel Network Followed Network for Retinal Vessel Segmentation. *Neural Networks* 126, 153–162. doi:10.1016/j.neunet.2020.02.018

Yan, Z., Yang, X., and Cheng, K.-T. (2019). A Three-Stage Deep Learning Model for Accurate Retinal Vessel Segmentation. *IEEE J. Biomed. Health Inform.* 23. doi:10.1109/jbhi.2018.2872813

Yan, Z., Yang, X., and Cheng, K.-T. (2018). Joint Segment-Level and Pixel-wise Losses for Deep Learning Based Retinal Vessel Segmentation. *IEEE Trans. Biomed. Eng.* 65, 1912–1923. doi:10.1109/tbme.2018.2828137

Yin, B., Li, H., Sheng, B., Hou, X., Chen, Y., Wu, W., et al. (2015). Vessel Extraction from Non-fluorescein Fundus Images Using Orientation-Aware Detector. *Med. Image Anal.* 26, 232–242. doi:10.1016/j.media.2015.09.002

Zhang, B., Zhang, L., Zhang, L., and Karray, F. (2010). Retinal Vessel Extraction by Matched Filter with First-Order Derivative of Gaussian. *Comput. Biol. Med.* 40, 438–445. doi:10.1016/j.compbiomed.2010.02.008

Zhang, T. Y., and Suen, C. Y. (1984). A Fast Parallel Algorithm for Thinning Digital Patterns. *Commun. ACM* 27, 236–239. doi:10.1145/357994.358023

Zhao, Y., Zheng, Y., Liu, Y., Zhao, Y., Luo, L., Yang, S., et al. (2018). Automatic 2-D/3-D Vessel Enhancement in Multiple Modality Images Using a Weighted Symmetry Filter. *IEEE Trans. Med. Imaging* 37, 438–450. doi:10.1109/TMI.2017.2756073

Zhuang, J. (2018). *Laddernet: Multi-Path Networks Based on U-Net for Medical Image Segmentation*. arXiv:1810.07810.

Zuiderveld, K. (1994). *Contrast Limited Adaptive Histogram Equalization*. USA: Academic Press Professional, Inc., 474–485. doi:10.1016/b978-0-12-336156-1.50061-6

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Ding, Zhang, Tang and Guo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 26 August 2021
doi: 10.3389/fonc.2021.694102

CEUS-Based Radiomics Can Show Changes in Protein Levels in Liver Metastases After Incomplete Thermal Ablation

Haiwei Bao[1], *Ting Chen*[1], *Junyan Zhu*[1], *Haiyang Xie*[2] *and Fen Chen*[1]*

[1] Department of Ultrasound, The First Affiliated Hospital, Zhejiang Chinese Medical University, Hangzhou, China, [2] Key Laboratory of Combined Multi-organ Transplantation, The First Affiliated Hospital, Zhejiang University, Hangzhou, China

OPEN ACCESS

Edited by:
Giorgio Russo, Institute of Bioimaging and Molecular Physiology (CNR), Italy

Reviewed by:
Sikandar Shaikh, Shadan Hospital and Institute of Medical Sciences, India
Xin-Wu Cui, Huazhong University of Science and Technology, China
Jianqiao Zhou, Shanghai Jiao Tong University, China

***Correspondence:**
Fen Chen
chenfen0571@qq.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *12 April 2021*
Accepted: *03 August 2021*
Published: *26 August 2021*

Citation:
Bao H, Chen T, Zhu J, Xie H and Chen F (2021) CEUS-Based Radiomics Can Show Changes in Protein Levels in Liver Metastases After Incomplete Thermal Ablation. Front. Oncol. 11:694102. doi: 10.3389/fonc.2021.694102

Objective: To investigate the ability of contrast-enhanced ultrasound (CEUS)-based radiomics combined with machine learning to detect early protein changes after incomplete thermal ablation.

Methods: HCT-26 colorectal adenoma cells were engrafted into the livers of 80 mice, which were randomly divided into 4 groups for palliative laser ablation. Changes in heat shock protein (HSP) and apoptosis-related protein expression in the tumors were assessed. SCID mice subjected to CEUS and ultrasonography were divided into training (n=56) and test (n=24) datasets. Then, 102 features from seven feature groups were extracted. We use the least absolute shrinkage and selection operator (LASSO) feature selection method to fit the machine learning classifiers. The feature selection methods and four classifiers were combined to determine the best prediction model.

Results: The areas under the receiver-operating characteristic curves (AUCs) of the classifiers in the test dataset ranged from 0.450 to 0.932 (median: 0.721). The best score was obtained from the model in which the omics data of CEUS was analyzed in the arterial phase by random forest (RF) classification.

Conclusions: A machine learning model, in which radiomics characteristics are extracted by multimodal ultrasonography, can accurately, rapidly and noninvasively identify protein changes after ablation.

Keywords: CEUS (contrast-enhanced ultrasound), radiomics, machine learning, liver metastases, thermal ablation

INTRODUCTION

Whether liver metastases can be inactivated in patients with liver metastases from colorectal cancer (CRLM) is a key issue influencing the survival and long-term tumor-free survival of patients (1–3). Thermal ablation has been considered an effective method in the treatment of such patients, but sometimes the residual tumor cells cannot be completely inactivated by ablation (4, 5). Various imaging

studies have a good judgment on the necrotic changes at the histological level after ablation and have formed certain clinical guidelines, but there are few reports on the changes at the molecular level, especially at the protein level. The molecular changes before and after ablation and the combination of molecular targeted therapy on this basis are the current research focus (6–8). At the same time, there is an urgent need for noninvasive imaging methods to reveal the molecular changes in the tumor before and after ablation (9).

The development of imaging has enabled the successful transformation of high-dimensional medical images into massive amounts of multilevel quantitative data (10, 11). In theory, imaging has the potential to reveal the tumor phenotype and molecular changes from the level of macroscopic characteristics of organ tissues to the level of local cell and molecular characteristics (12). The accurate, timely and sensitive display of tumor molecular characteristics by imaging is of great significance for treatment (13–16). CT, MRI and PET have been successfully explored in these fields (17–20). There are relatively few reports on multimodal ultrasound, including contrast-enhanced ultrasound (CEUS), and there is no standard for omics exploration methods (21–23). The main difficulty is the lack of universal ultrasound omics analysis methods, especially for CEUS data (24, 25).

Multimodal ultrasound in the imaging-based diagnosis of liver metastases has the advantages of being easy, repeatable, nonradioactive, and highly sensitive and is the main imaging means for guiding thermal ablation. In most interventional treatment centers in China, the rate at which ultrasound can detect liver metastases is an important factor for decision-making regarding ablation. We explored whether CEUS-based multimode ultrasound imaging can detect early molecular changes in incomplete tumor ablation in animal models.

MATERIALS AND METHODS

Animal Model

All experimental procedures followed Zhejiang University Laboratory Animal Operating Regulations. (http://www.lac.zju.edu.cn/cms/12997). This research protocol was approved by the Research Ethics Committee of the First Affiliated Hospital of Zhejiang University. The severe combined immunodeficiency (SCID) mice were housed in a specific pathogen-free (SPF) animal room and were underwent a 12-h light/dark cycle to obtain free food and tap water. The temperature of the room is 20-25°C and the humidity is 50-60%. The 10mm long HCT-26 colorectal adenoma tumor tissue was cut into 2-mm3 pieces and implanted into the left lobe liver of SCID mice by surgical incision. According to the different ablation conditions, a total of 80 mice were divided into 4 groups with 20 mice in each group. The grouping was performed with a random number method. The first group is a blank control group, the second group is a sham puncture group, the third group and the fourth group are incomplete ablation groups. The incomplete ablation experiments were performed after 2 weeks when tumors had grown to an average diameter of 0.6-0.7 cm (**Figure 1A**). The SCID mice were killed 18 h after incomplete ablation, and the largest tumor diameter was 0.8-1.0 cm.

Palliative Ablation Method

Intraperitoneal anesthesia was administered to the SCID mice with approximately 200 μl 4% chloral hydrate at a dose of 400 mg/kg mouse weight. We applied the conditions established in the previous study to construct an incomplete ablation model (26, 27). Mice were fixed on the operating table with local disinfection. An ultrasound-guided 21-G percutaneous transhepatic cholangiography (PTC) needle was used to puncture the tumor. After confirming the position of the needle by high-frequency ultrasound, the operator removed the inner core and inserted a laser fiber into the needle (Mylab^{TM} Twice, Esaote SpA) (**Figure 1B**). After confirming that the fiber was positioned at 1/3 of the length of the tumor, the power was set at 1 W, and the foot switch was activated for continuous laser ablation; the treatment time was 0 seconds, 30 seconds and 60 seconds (total dose of 0 J, 30 J, 60 J). Local hemostasis was conducted by applying gentle pressure following treatment.

Multimodal Ultrasound Examination and Image Evaluation

All ultrasound examinations were performed by two experienced radiologists who had more than 10 years of experience in ultrasound-guided interventional procedures. Conventional sonography was conducted with MyLab Twice (Esaote SpA, Genoa, Italy) with a linear array transducer with a frequency range of 18-22 MHz. Scans were performed under the abdominal preset, and the imaging focal zone was positioned posterior to the level of the lesion. For the target lesion, the largest section in terms of both longitudinal and transverse views was stored. Color Doppler of each lesion was performed by using the same transducer, and the picture was recorded.

The linear array transducer with a frequency range of 1.0 to 4.0 MHz is equipped with a contrast-specific contrast pulse sequencing imaging mode for CEUS inspection. Contrast-enhanced sonography was performed using SonoVue (Bracco SpA, Milan, Italy), a second-generation ultrasound contrast agent consisting of microbubbles of sulfur hexafluoride gas. A 0.1-mL bolus of SonoVue was hand injected through a 25-G intravenous catheter in the caudal vein. The target lesion in the largest plane was continuously observed and documented with a 60-second long clip (**Figures 1C, D**). All the above mentioned ultrasound examinations were completed on the experimental day 8 h after incomplete ablation or sham puncture.

CEUS features (including the enhancement level, enhancement homogeneity, enhancement boundary, and feeding artery) were evaluated and recorded. All digital cine clips of the study population were retrospectively reviewed by two investigators (C.F. and B.H.W.), each of whom has more than 10 years of experience in evaluating liver CEUS scans. They were asked to evaluate and record the imaging features of all the mice using a standardized approach. In cases of discordance, a consensus reading was performed, and the classification judgment was made.

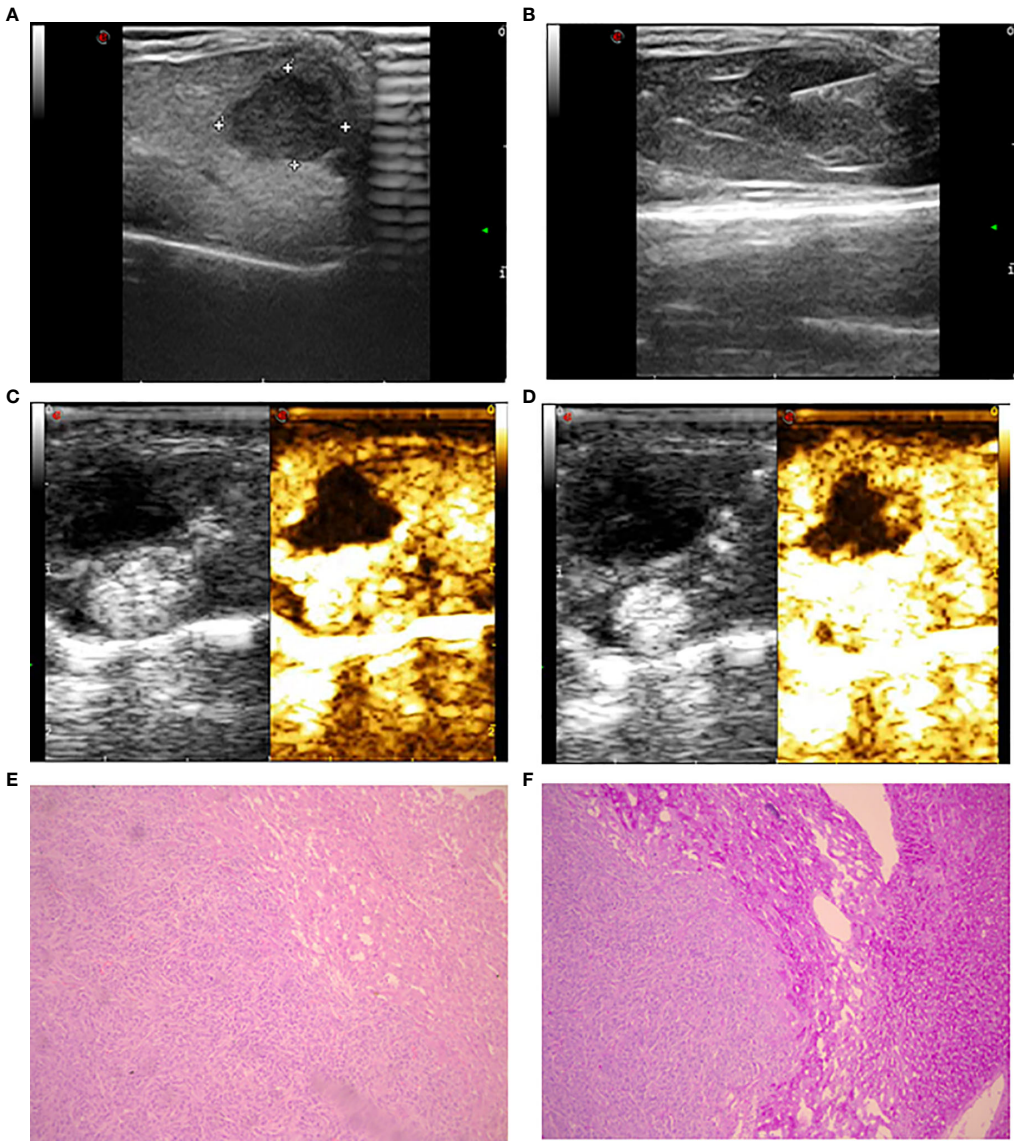


FIGURE 1 | Ultrasonic image and histopathology of the implanted tumor. **(A)** Gray ultrasound showed hypoechoic metastases in the liver parenchyma with clear boundaries. **(B)** Ultrasound showed laser fiber into the center of the tumor. **(C)** The 5-second image of contrast-enhanced ultrasound in liver metastases. **(D)** The 25-second image of contrast-enhanced ultrasound in liver metastases. **(E)** The morphology of intestinal adenocarcinoma cells was uniform with large and deep stained nuclei (HE staining). **(F)** Compared to intratumoral tissue, there was increased staining in liver tissue with clear edges (periodic acid-Schiff staining).

Sample Collection and Molecular Biological Examination

The SCID mice were euthanized *via* intraperitoneal injection of 2% sodium pentobarbital at a dose of 150 mg/kg body weight. Cervical dislocation was used to confirm the death of 80 mice 18 h post-operation. The 80 tumors were cut along their diameter, and one-quarter was fixed in 4% paraformaldehyde for 48 h at room temperature, whereas the remaining three-quarters of each tumor was kept in a liquid nitrogen jar. Histopathology was performed with HE and periodic acid-Schiff (PAS) staining (**Figures 1E, F**).

Terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) was used to determine the apoptosis of the tumor cells in the liver that underwent incomplete ablation. For each section, five or more high-magnification fields (200x) with at least 500 cells were selected to count the number of cells emitting green fluorescence and to subsequently calculate the apoptosis index (Ai). The equation was as follows: Ai = (number of positive cells/number of total cells) 100% ± SD.

Small tumor samples were lysed on ice using RIPA lysis buffer (P0013B, Beyotime, China) containing a protease inhibitor cocktail; cellular debris was pelleted by centrifugation at 12,000 rpm for 5 min

at 4°C, and the supernatants were harvested. Protein concentrations were measured by a bicinchoninic acid (BCA) protein assay kit (P0010 Beyotime, China). The protein lysates (30 μg per lane) were separated *via* SDS-PAGE. Following separation on an 8%~15% acrylamide gel, the proteins were transferred onto PVDF membranes. The membrane containing the proteins was successively incubated with blocking buffer (overnight at 4°C), with a primary antibody (37°C for 1 h) and with a secondary antibody (37°C for 1 h). The following primary antibodies were used: anti-Bax (1:3000; ab32503), anti-caspase-3 (1:1000; ab184787), anti-Hsp70 (1:1000; Ab2787), anti-Hsp90 (1:1000; Ab13492), and anti-GAPDH (1:1000; ab181602), all from Abcam Biotechnology. HRP-conjugated goat anti-mouse IgG H&L (1:3000; SE131, Solarbio, USA) was used as the secondary antibody. Chemiluminescence detection was achieved by exposure to film in a darkroom. Following development and fixation with washing buffer at 20–25°C for 10 min (P0019, Beyotime Institute of Biotechnology), the film was visualized with an enhanced chemiluminescence system (ECL, Beyotime, China). The densities of the protein bands were determined using ImageJ software (v1.46; National Institutes of Health), normalized to actin expression and quantified using Microsoft Excel software (version 2016, Microsoft Corporation).

Feature Extraction and Selection

Dynamic CEUS was used to obtain a series of static images at a frequency of one per second. All stored images were transformed into an 8-bit bitmap. For each lesion, a region of interest (ROI) around the tumor border was delineated on the largest cross section with the ABsnake plugin using ImageJ software semiautomatically. Two doctors performed three times combined with their own judgment. Then, a total of 102 features were extracted from the ROI using PyRadiomics (version 1.3.0; Computational Imaging and Bioinformatics Lab, Harvard Medical School). The CEUS images of each mouse were extracted at 5, 25, and 45 seconds (**Figure 2A**). To reduce the differences in semi-automatic manual segmentation between sonographers, we calculated the intraclass correlation coefficient (ICC) of each feature between two doctors. We extracted the following three types of data for machine learning analysis: the grayscale data, contrast-enhanced arterial phase data (15 seconds), and the whole course of CEUS data. The absolute values of the differences in the 5- and 25- second and the 25-second and 45-second CEUS radiomics data per mouse were added as the data representing the whole course of CEUS for subsequent analysis. To construct nonredundant and robust combined radiomic signatures, the least absolute shrinkage and selection operator (LASSO) regression method was used (**Figure 2B**). The complexity of the LASSO regression is controlled by a tuning parameter lambda (λ) with the rule that as the value of λ increases, the penalty for each variable coefficient also increases. Only nonzero coefficient variables were selected in this method.

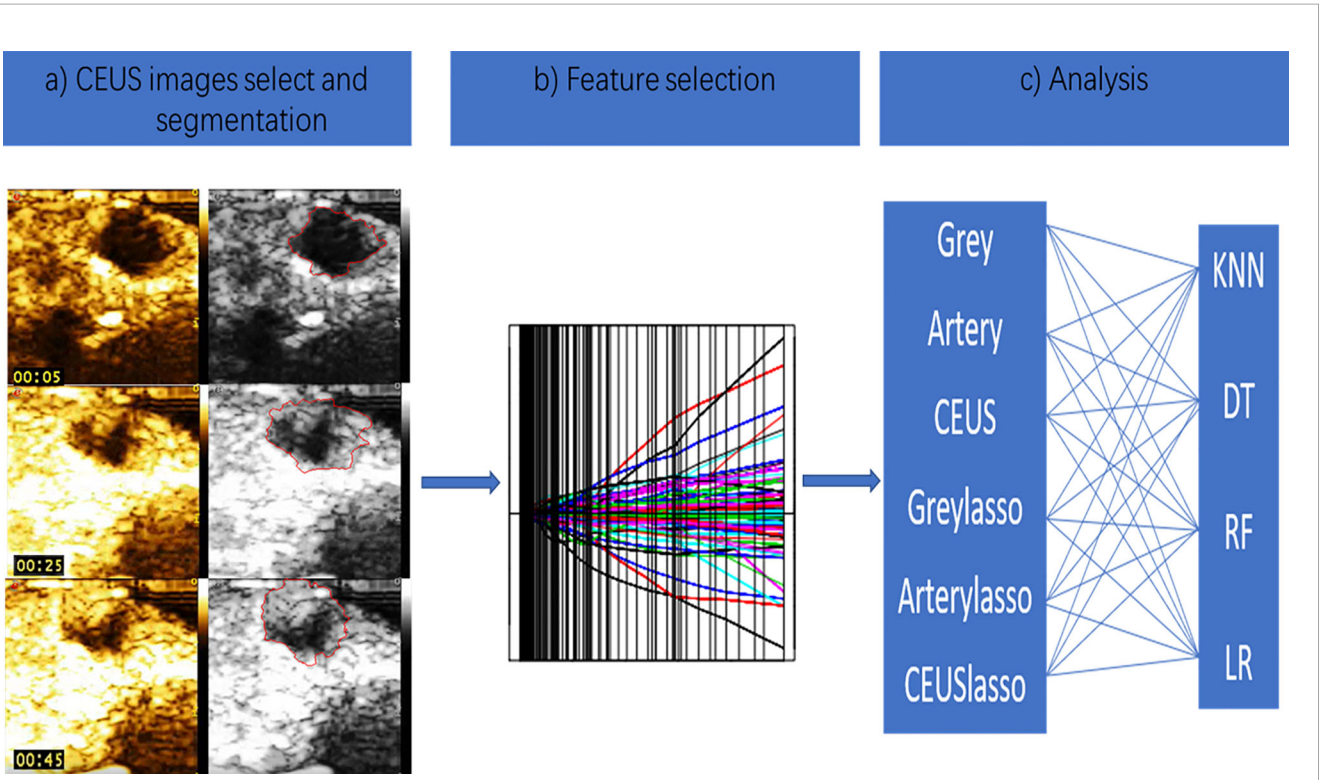


FIGURE 2 | The CEUS-based radiomics analysis schematic. **(A)** The CEUS images of mouse was extracted at 5,25 and 45 seconds. **(B)** Lasso regression was used as feature selection method. **(C)** Four supervised machine learning algorithms were applied.

Machine Learning and Model Performance Evaluation

Data analysis and training of binary classifiers was performed using the Python programming language (Python Software Foundation, version 3.7.4, 2019, available at http://www.python.org), including the packages "numpy," "pandas," "sklearn," and the other package and the Anaconda integrated development environment (version 1.9.12, Anaconda Inc., USA).

We applied 4 supervised machine learning algorithms; these classifiers were k nearest neighbors (KNN),decision tree (DT), random forest(RF) and logistic regression (LR) (**Figure 2C**).

There were six kinds of radiomics data as the inputs of each machine learning model(1. Grayscale ultrasound 2. Contrast-enhanced arterial phase data 3. The whole course of CEUS 4. Grayscale data with LASSO regression 5. Contrast-enhanced arterial phase data with LASSO regression 6. The whole course of CEUS with LASSO regression).These 4 classification methods combined with six kinds radiomics data to establish 24 (6× 4 = 24) models.

Each of the 24 models was trained and 10-fold cross validated in the training set with scikit-learn. The receiver operating characteristic (ROC) curve and area under the ROC curve (AUC) were employed to evaluate the predictive accuracy of the radiomics signatures developed. The model that had the highest AUC value in the test dataset was selected as the final model.

A two-sided p value of < 0.05 was used as the criterion to indicate a statistically significant difference. Statistical analysis was conducted with SPSS 22.0 for Windows (Chicago, IL).

The above is the main process of this study (**Figure 3**).

RESULTS

Molecular Changes After Tumor Ablation

The Expression of HSPs Is Increased After Ablation

Compared with those in groups 1 and 2, the expression levels of two HSPs (HSP70 and HSP90a) were increased in groups 3 and 4. The western blotting results confirmed that the HSP70 and HSP90a expression levels were significantly higher in experimental groups 3 and 4 than in control groups 1 and 2; the relative HSP70 and HSP90a expression levels in group 4 were higher than those in group 3 ($p < 0.01$) (**Figures 4A, B**).

Apoptosis in the Incomplete Ablation Group and the Control Group

The apoptosis rate was higher in groups 3 and 4 than in groups 1 and 2. The TUNEL assay showed more green fluorescence accumulation in groups 3 and 4 than in groups 1 and 2. Ai was higher for the tumors in groups 3 and 4 than for the tumors in groups 1 and 2, and Ai was higher in group 4 than in group 3 (**Figures 4C, D**).

The expression of the apoptosis protein Bcl-2 was lower in groups 3 and 4 than in groups 1 and 2. The expression of the apoptosis protein BAX was higher in groups 3 and 4 than in groups 1 and 2. The relative expression levels of cleaved caspase-3 protein detected by western blot analysis showed were significantly different between groups 1 and 2 *vs*. groups 3 and 4 (**Figures 4E, F**).

Interobserver Agreement

The interobserver reproducibility of feature extraction by the two ultrasonographers was good, with ICCs ranging from 0.789 to 0.932. Therefore, all the outcomes were based on the measurements made by the first sonographer.

Feature Selection

The training dataset includes 102 radiomics features. The features were grouped into first-order statistics (18 features) and shape-based (9 features), gray level dependence matrix (GLDM, 14 features), gray level cooccurrence matrix (GLCM, 24 features), gray level run length matrix (GLRLM, 16 features), gray level size zone matrix (GLSZM, 16 features), and neighboring gray tone difference matrix (NGTDM, 5 features)

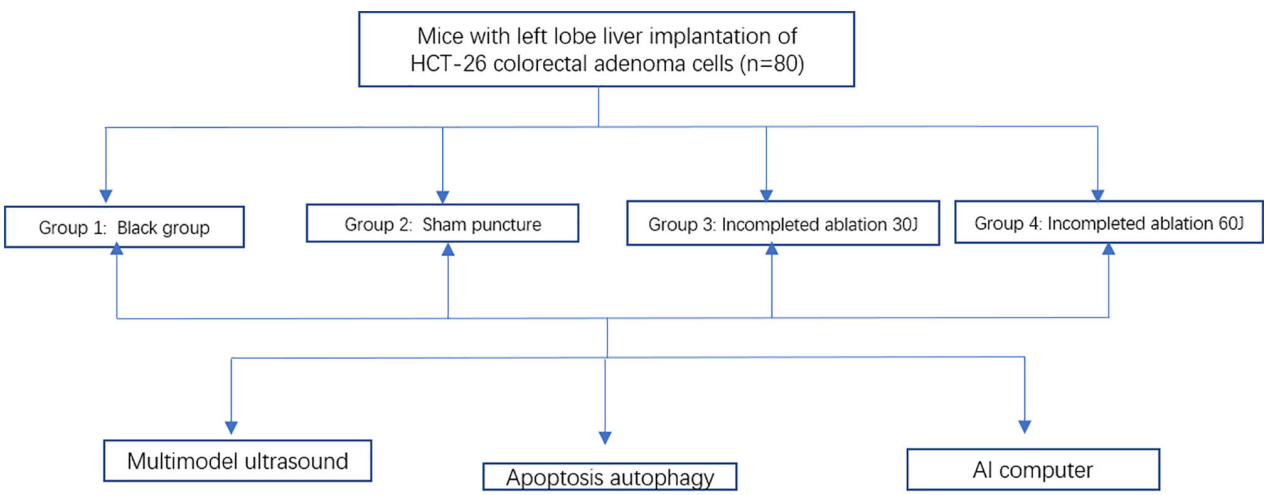


FIGURE 3 | The flow chart shows the three main steps process in our study.

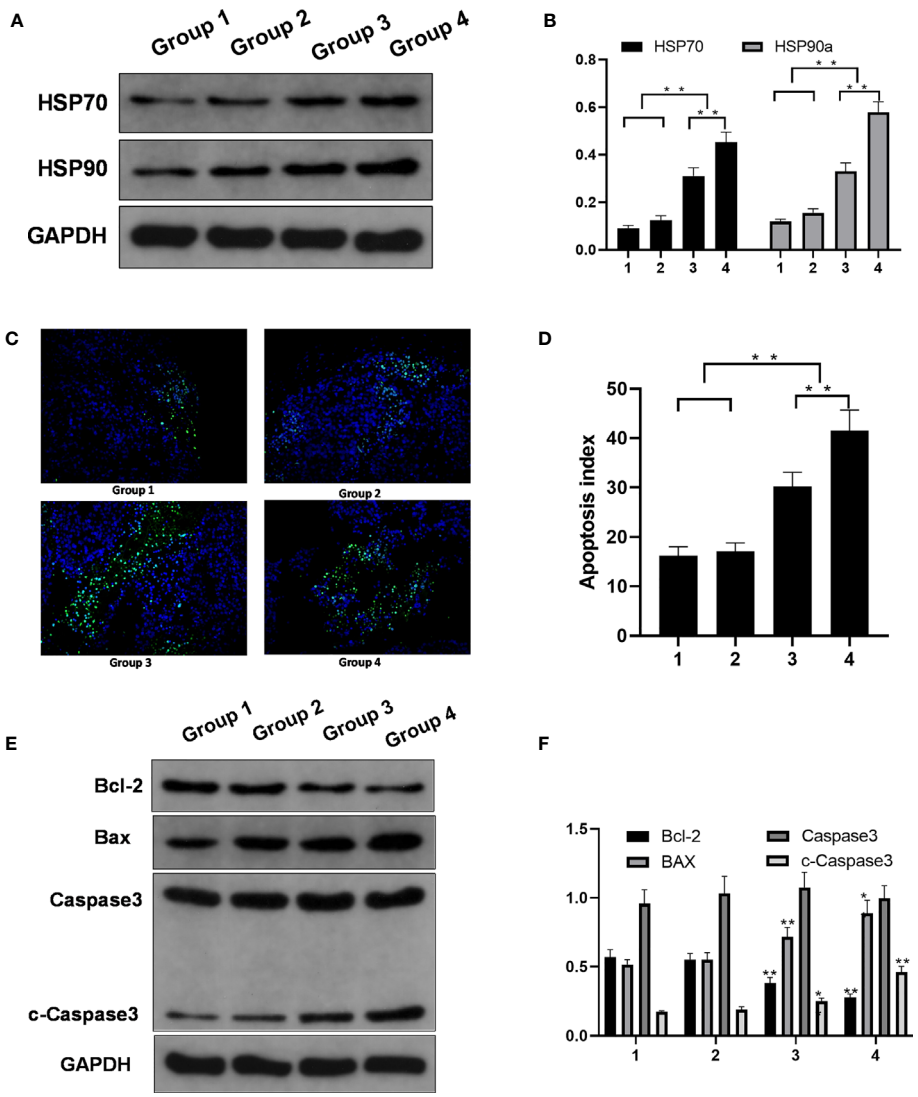


FIGURE 4 | Molecular changes in the control groups and the experimental groups. **(A)** HSP70 and HSP90a expression. **(B)** HSP70 and HSP90a expression levels in four groups. Significant differences were calculated in groups 1 and 2 *vs*. groups 3 and 4, groups 3 *vs*. 4. **(C)** TUNEL expression (magnification). **(D)** Apoptosis index of the four groups. Significant differences were calculated in groups 1 and 2 *vs*. groups 3 and 4, groups 3 *vs*. 4. **(E)** Protein expression. **(F)** Bcl-2, BAX, and c-capase3 protein expression level in four groups. Significant differences were calculated between groups 1 and 2 *vs*. groups 3 and 4. *p < 0.05, **p < 0.01.

features. A radiomics signature was further constructed based on the 102 features with respective nonzero coefficients selected through the LASSO regression method.

Model Performance Evaluation

LR is one of the most commonly used generalized linear mixed models (GLMMs) for two classifications of data. The number of key features selected for building a radiomics signature was determined by the concordance index (C-index) value using 10-fold cross-validation. LR with L1 regularization had an AUC of 0.450-0.829 in the test dataset; LR with LASSO feature selection had an AUC of 0.493-0.850 in the test dataset (**Figure 5A**).

In the tune grid, we varied the max_depth between 1 and 10; min_impurity_decrease between 0-1; the best score was 0.792. The final model had a max_depth of 4 per decision tree. The final AUC of the DT classifier in the test dataset was 0.542-0.875. The final AUC of the DT classifier with LASSO feature selection in the test dataset was 0.546-0.622 (**Figure 5A**).

The selected RF classifiers consisted of 28 decision trees. In the tune grid, we varied the number of features randomly selected for each tree between 1 and 10. At a max-depth of 4, a plateau was observed at an AUC of 0.79, and the accuracy cannot be improved by increasing the tree depth. Consequently, the depth of each decision tree of the final model is 4 randomly selected predictors. The final AUC of the RF classifier was 0.613-0.932.

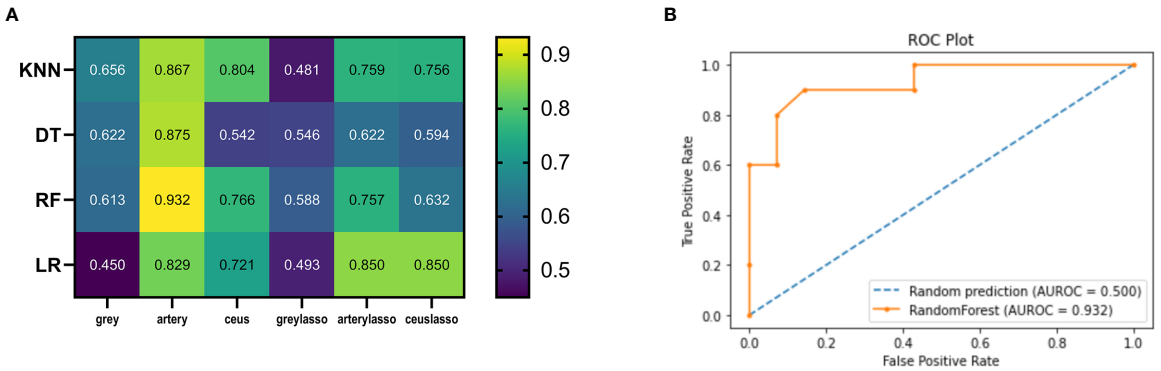


FIGURE 5 | Heatmaps illustrating the predictive performance (AUC) of different combinations of feature selection methods (rows) and classification algorithms (columns). **(A)** Cross-validated AUC values of 24 models on the training and validation datasets. **(B)** ROC curve of RF_artery.

The final AUC of the RF classifier with LASSO was 0.588-0.757 (**Figure 5A**).

We kept a KNN classifier as a reference because it is frequently applied, easy to implement, and robust, although its performance is worse than other classifiers. The number of neighbors used for classification varied from 1 to 11, and gradually increased by 1 during training. The best performance was achieved by the classifier using k = 1 neighbors. The final AUC of the KNN classifier was 0.656-0.867. The final AUC of the KNN classifier with LASSO was 0.481-0.759 (**Figure 5A**).

In **Figure 2A**, the mean AUCs of the 8 models are presented in the heatmap. AUCs ranged from 0.450 to 0.932; the median value was 0.721, and the best score was 0.932 (the RF model in the artery dataset; RF_artery) (**Figure 5B**). The AUCs of grayscale data combined with all the four classifiers ranged from 0.450-0.656 and the AUCs of graylasso data combined with all four classifiers ranged from 0.481-0.588. This shows that the classification value of grayscale ultrasonic image data is not high.

The AUC for the image identification ability of sonographers using conventional analytical methods is 0.705. The best machine learning score was significantly better than the ultrasonographer score (P<0.01).

DISCUSSION

In this preclinical radioproteomics study, we hypothesized that there was a causal relationship between image features and protein expression. To prove our hypothesis, we first employed a model of metastatic cancer implantation in SCID mice and performed incomplete ablation intervention. Our previous study showed that 30J and 60J laser ablation caused incomplete ablation, which induced increased apoptosis of the tumor tissue without causing significant necrosis (26, 27). Heat shock protein(HSP) family is sensitive to heat injury. Increased reactivity of HSP90 and HSP70 after incomplete thermal ablation has been reported in the literature (28, 29). The increase in HSP70 and HSP 90 indicates that the tumor tissue has suffered the corresponding thermal damage. Apoptosis related examinations including the increased expression of TUNEL, BAX protein, cleaved caspase-3 protein and the decrease of Bcl-2 protein, all proved the changes in apoptosis related proteins in tumor tissue (30).The related indicators—HSPs and apoptosis -related proteins—were assessed to confirm the ability of the model to detect protein level changes in a short time after incomplete ablation. The results show that the binary classification machine learning model trained by quantitative radiomics ultrasound data can distinguish changes in protein levels after incomplete ablation.

Radiomics has facilitated some meaningful advances in the field of liver tumor ablation (31). A total of 647 radiologic features of three-phase contrast-enhanced CT within 2 weeks before ablation were extracted from 184 HCC patients. LASSO Cox regression model was used to select valuable indexes. A recurrence prediction model was established based on clinicopathological factors and radiological features The results indicated that among the four radiomic models, the portal venous-phase model performed best in the validation subgroup (C-index = 0.736 (95% confidence interval [CI]: 0.726-0.856)). The predictive ability of the combined clinicopathological and radiological features model was significantly better than that of the simple clinical model (ANOVA, P < 0.05). 0001).

At present, there are several kinds of ultrasound contrast agents for monitoring microcirculation perfusion. Only SonoVue(Bracco Italy) has obtained the approval and license for abdominal imaging procedures in China. SonoVue is a stable sulfur hexafluoride microbubble surrounded by a phospholipid shell with a mean diameter of 2.5–6 μm (32). Guidelines and clinical practice recommendations for CEUS have been developed. CEUS is helpful to improve the detection of liver lesions. It has been suggested in planning and monitoring of liver metastatic tumor ablation.

CEUS with SonoVue is close to the gold standard of contrast-enhanced imaging in evaluating the short-term ablation rate (1-3 months) after ablation. It has the advantage of real-time and quick evaluation after ablation (immediate and 30 min), which can immediately guide re-ablation and improve the complete

ablation rate of single ablation. In China and other countries, more than 50% of patients choose multimodal ultrasound guidance (33, 34). Early multimodality ultrasound has the advantage that incomplete ablation is detected sooner but brings about the challenge of interpreting dynamic processes such as transient edema. Therefore, it is valuable to explore the imaging omics evaluation of ablation based on ultrasound and CEUS immediately after surgery. Based on our limited knowledge, we have not yet seen a study assessing the protein level before and after tumor ablation by ultrasound imaging omics based on CEUS data. Only one study applied a radiomics technique with CEUS to assess tumor heterogeneity (35). In three tumor models of xenografted mice, the morphological and functional characteristics of tumor vessels were extracted by CEUS, and the tumor phenotype was classified by a trained linear support vector machine (SVM) incorporating features including the image intensity median, gray level cooccurrence matrix energy, vascular network length, and run length nonuniformity of the gray run matrix. The fourfold cross validation scheme was used to train the model, and a correct classification rate of 82.1% (95% CI: 0.640-0.92) was obtained. However, the author used an uncommon and hard-to-obtain vessel segmentation algorithm to extract CEUS radiomics data.

In this study, we successfully extracted the static radiomics data of grayscale ultrasound and CEUS at a certain time point by using the general open-source radiomics software PyRadiomics. The numerical differences between three static time CEUS scans were calculated to represent the whole process of CEUS. The data in this study are two-category small sample data, which are not suitable for deep learning methods that require large amounts of data, such as neural networks. We have selected the following four as exploration methods in machine learning classification algorithms. First, LR is one of the most commonly used methods for the two types of data classification, and traditional statistical methods can also be implemented. The KNN classifier is selected as a reference because it is a frequently used and easy to implement. The DT classifier and RF classifier are the most effective classification methods in machine learning. The results also prove that in this study, the effect of RF classification is the best, which has exceeded that of experienced doctors. The four most commonly used machine learning methods were applied to analyze six kinds of radiomics data. The RF_artery model proved to have the highest AUC value of 0.932. The value of CEUS in the arterial phase at 5 seconds was the highest, which is consistent with our clinical experience. The arterial phase of CEUS is often the most representative feature of microcirculation perfusion in tumors. In the LI-RADS classification standard of the liver contrast-enhanced ultrasound guide, the performance of the arterial phase of contrast-enhanced ultrasound occupies a high weight (36). The combination of anatomical and functional information might enable the development of better models for radiomics analysis.

In this study, relatively few animals were assessed, and a standardized acquisition protocol was used to image all mice. Thus, there is a risk that the data were overfitted and that the best image for analysis was not randomly selected for the test dataset. In addition, the small sample size makes the results more susceptible to data variability. Radiomics analysis can be influenced and challenged by the different types of scanners from different manufacturers in the clinical setting. This study provides good confidence that CEUS data can be used to perform radiomic analysis. Future studies should address the above mentioned issues.

CONCLUSIONS

We have shown in *in vivo* preclinical models that radiomics is able to quantify early protein changes in tumors after incomplete ablation and identify differences that are not visible to the human eye. After incomplete ablation, the radiomics profile of CRLM appears to be different from that before surgery. These differences may be identified by binary classification algorithms, especially by RF trained with radiomics features extracted from the arterial-phase CEUS immediately after ablation, which showed the best AUC of 0.932, allowing for an early assessment of protein level changes after ablation. Further studies are needed to further validate its classification ability.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The animal study was reviewed and approved by Research Ethics Committee of the First Affiliated Hospital of Zhejiang University.

AUTHOR CONTRIBUTIONS

Conception and design: FC. Data acquisition: HWB, TC, JYZ, and FC. Data analysis and interpretation: FC and HWB. Drafting the article: FC and HWB. Revising the article: FC, HWB, TC, JYZ, and HYX. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by Zhejiang Province Basic Public Welfare Research Program (grant number LGD19C04007).

ACKNOWLEDGMENTS

Thanks to all participants in this study.

REFERENCES

1. Hasselgren K, Røsok BI, Larsen PN, Sparrelid E, Lindell G, Schultz NA, et al. ALPPS Improves Survival Compared With TSH in Patients Affected of CRLM: Survival Analysis From the Randomized Controlled Trial LIGRO. *Ann Surg* (2021) 273(3):442–8. doi: 10.1097/SLA.0000000000003701
2. Siegel R, Ma J, Zou Z, Jemal A. Cancer Statistics, 2014. *CA Cancer J Clin* (2014) 64(1):9–29. doi: 10.3322/caac.21208
3. Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Patterns and Trends in Colorectal Cancer Incidence and Mortality. *Gut* (2017) 66(4):683–91. doi: 10.1136/gutjnl-2015-310912
4. Evans J. Ablative and Catheter-Delivered Therapies for Colorectal Liver Metastases (CRLM). *Eur J Surg Oncol* (2007) 33(Suppl 2):S64–75. doi: 10.1016/j.ejso.2007.09.027
5. Pawlik TM, Izzo F, Cohen DS, Morris JS, Curley SA. Combined Resection and Radiofrequency Ablation for Advanced Hepatic Malignancies: Results in 172 Patients. *Ann Surg Oncol* (2003) 10(9):1059–69. doi: 10.1245/ASO.2003.03.026
6. Siriwardana PN, Luong TV, Watkins J, Turley H, Ghazaley M, Gatter K, et al. Biological and Prognostic Significance of the Morphological Types and Vascular Patterns in Colorectal Liver Metastases (CRLM): Looking Beyond the Tumor Margin. *Med (Baltimore)* (2016) 95(8):e2924. doi: 10.1097/MD.0000000000002924
7. Bai H, Huangz X, Jing L, Zeng Q, Han L. The Effect of Radiofrequency Ablation vs. Liver Resection on Survival Outcome of Colorectal Liver Metastases (CRLM): A Meta-Analysis. *Hepatogastroenterology* (2015) 62(138):373–7.
8. Reghupaty SC, Sarkar D. Current Status of Gene Therapy in Hepatocellular Carcinoma. *Cancers (Basel)* (2019) 11(9):1265–85. doi: 10.3390/cancers11091265
9. Song J, Zhang F, Ji J, Chen M, Li Q, Weng Q, et al. Orthotopic Hepatocellular Carcinoma: Molecular Imaging-Monitored Intratumoral Hyperthermia-Enhanced Direct Oncolytic Virotherapy. *Int J Hyperthermia* (2019) 36(1):344–50. doi: 10.1080/02656736.2019.1569731
10. Liu F, Ning Z, Liu Y, Liu D, Tian J, Luo H, et al. Development and Validation of a Radiomics Signature for Clinically Significant Portal Hypertension in Cirrhosis (CHESS1701): A Prospective Multicenter Study. *EBioMedicine* (2018) 36:151–8. doi: 10.1016/j.ebiom.2018.09.023
11. Mokrane FZ, Lu L, Vavasseur A, Philippe O, Peron JM, Luk L, et al. Radiomics Machine-Learning Signature for Diagnosis of Hepatocellular Carcinoma in Cirrhotic Patients With Indeterminate Liver Nodules. *Eur Radiol* (2020) 30(1):558–70. doi: 10.1007/s00330-019-06347-w
12. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, 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. *Lancet Oncol* (2018) 19(9):1180–91. doi: 10.1016/S1470-2045(18)30413-3
13. Feng ST, Jia Y, Liao B, Huang B, Zhou Q, Li X, et al. Preoperative Prediction of Microvascular Invasion in Hepatocellular Cancer: A Radiomics Model Using Gd-EOB-DTPA-Enhanced MRI. *Eur Radiol* (2019) 29(9):4648–59. doi: 10.1007/s00330-018-5935-8
14. Gillies RJ, Schabath MB. Radiomics Improves Cancer Screening and Early Detection. *Cancer Epidemiol Biomarkers Prev* (2020) 29(12):2556–67. doi: 10.1158/1055-9965.EPI-20-0075
15. Li Y, Eresen A, Shangguan J, Shangguan J, Lu Y, Chen D, et al. Establishment of a New non-Invasive Imaging Prediction Model for Liver Metastasis in Colon Cancer. *Am J Cancer Res* (2019) 9(11):2482–92.
16. Wang Z, Lin F, Ma H, Shi Y, Dong J, Yang P, et al. Contrast-Enhanced Spectral Mammography-Based Radiomics Nomogram for the Prediction of Neoadjuvant Chemotherapy-Insensitive Breast Cancers. *Front Oncol* (2021) 11:605230. doi: 10.3389/fonc.2021.605230
17. Wu W, Ye J, Wang Q, Luo J, Xu S. CT-Based Radiomics Signature for the Preoperative Discrimination Between Head and Neck Squamous Cell Carcinoma Grades. *Front Oncol* (2019) 9:821. doi: 10.3389/fonc.2019.00821
18. Witt JS, Rosenberg SA, Bassetti MF. MRI-Guided Adaptive Radiotherapy for Liver Tumours: Visualising the Future. *Lancet Oncol* (2020) 21(2):e74–82. doi: 10.1016/S1470-2045(20)30034-6
19. Bressem KK, Adams LC, Vahldiek JL, Erxleben C, Poch F, Lehmann KS, et al. Subregion Radiomics Analysis to Display Necrosis After Hepatic Microwave Ablation-A Proof of Concept Study. *Invest Radiol* (2020) 55(7):422–9. doi: 10.1097/RLI.0000000000000653
20. Elhalawani H, Lin TA, Volpe S, Mohamed ASR, White AL, Zafereo J, et al. Machine Learning Applications in Head and Neck Radiation Oncology: Lessons From Open-Source Radiomics Challenges. *Front Oncol* (2018) 8:294. doi: 10.3389/fonc.2018.00294
21. Sun Q, Lin X, Zhao Y, Li L, Yan K, Liang D, et al. Deep Learning vs. Radiomics for Predicting Axillary Lymph Node Metastasis of Breast Cancer Using Ultrasound Images: Don't Forget the Peritumoral Region. *Front Oncol* (2020) 10:53. doi: 10.3389/fonc.2020.00053
22. Xu EJ, Lv SM, Li K, Long YL, Zeng Q, Su ZZ, et al. Immediate Evaluation and Guidance of Liver Cancer Thermal Ablation by Three-Dimensional Ultrasound/Contrast-Enhanced Ultrasound Fusion Imaging. *Int J Hyperthermia* (2018) 34(6):870–6. doi: 10.1080/02656736.2017.1373306
23. Vogl TJ, Farshid P, Naguib NN, Darvishi A, Bazrafshan B, Mbalisike E, et al. Thermal Ablation of Liver Metastases From Colorectal Cancer: Radiofrequency, Microwave and Laser Ablation Therapies. *Radiol Med* (2014) 119(7):451–61. doi: 10.1007/s11547-014-0415-y
24. Huang Q, Zeng Q, Long Y, Tian L, Zheng R, Xu E, et al. Fusion Imaging Techniques and Contrast-Enhanced Ultrasound for Thermal Ablation of Hepatocellular Carcinoma - A Prospective Randomized Controlled Trial. *Int J Hyperthermia* (2019) 36(1):1207–15. doi: 10.1080/02656736.2019.1687945
25. Kaye EA, Cornelis FH, Petre EN, Tyagi N, Shady W, Shi W, et al. Volumetric 3D Assessment of Ablation Zones After Thermal Ablation of Colorectal Liver Metastases to Improve Prediction of Local Tumor Progression. *Eur Radiol* (2019) 29(5):2698–705. doi: 10.1007/s00330-018-5809-0
26. Chen F, Bao H, Xie H, Tian G, Jiang T. Heat Shock Protein Expression and Autophagy After Incomplete Thermal Ablation and Their Correlation. *Int J Hyperthermia* (2019) 36(1):95–103. doi: 10.1080/02656736.2018.1536285
27. Chen F, Xie H, Bao H, Violetta L, Zheng S. Combination of HSP90 and Autophagy Inhibitors Promotes Hepatocellular Carcinoma Apoptosis Following Incomplete Thermal Ablation. *Mol Med Rep* (2020) 22(1):337–43. doi: 10.3892/mmr.2020.11080
28. Schueller G, Kettenbach J, Sedivy R, Stift A, Friedl J, Gnant M, et al. Heat Shock Protein Expression Induced by Percutaneous Radiofrequency Ablation of Hepatocellular Carcinoma In Vivo. *Int J Oncol* (2004) 24(3):609–13. doi: 10.3892/ijo.24.3.609
29. Nikfarjam M, Muralidharan V, Su K, Malcontenti-Wilson C, Christophi C. Patterns of Heat Shock Protein (HSP70) Expression and Kupffer Cell Activity Following Thermal Ablation of Liver and Colorectal Liver Metastases. *Int J Hyperthermia* (2005) 21(4):319–32. doi: 10.1080/02656730500133736
30. Rai R, Richardson C, Flecknell P, Robertson H, Burt A, Manas DM. Study of Apoptosis and Heat Shock Protein (HSP) Expression in Hepatocytes Following Radiofrequency Ablation (RFA). *J Surg Res* (2005) 129(1):147–51. doi: 10.1016/j.jss.2005.03.020
31. Yuan C, Wang Z, Gu D, Tian J, Zhao P, Wei J, et al. Prediction Early Recurrence of Hepatocellular Carcinoma Eligible for Curative Ablation Using a Radiomics Nomogram. *Cancer Imaging* (2019) 19(1):21. doi: 10.1186/s40644-019-0207-7
32. Correas JM, Bridal L, Lesavre A, Méjean A, Claudon M, Hélénon O. Ultrasound Contrast Agents: Properties, Principles of Action, Tolerance, and Artifacts. *Eur Radiol* (2001) 11(8):1316–28. doi: 10.1007/s003300100940
33. Dill-Macky MJ, Asch M, Burns P, Wilson S. Radiofrequency Ablation of Hepatocellular Carcinoma: Predicting Success Using Contrast-Enhanced Sonography. *AJR Am J Roentgenol* (2006) 186(5 Suppl):S287–95. doi: 10.2214/AJR.04.1916
34. Choi D, Lim HK, Kim SH, Lee WJ, Jang HJ, Lee YJ, et al. Hepatocellular Carcinoma Treated With Percutaneous Radio-Frequency Ablation: Usefulness of Power Doppler US With a Microbubble Contrast Agent in Evaluating Therapeutic Response-Preliminary Results. *Radiology* (2000) 217(2):558–63. doi: 10.1148/radiology.217.2.r00oc07558

35. Theek B, Opacic T, Magnuska Z, Lammers T, Kiessling F. Radiomic Analysis of Contrast-Enhanced Ultrasound Data. *Sci Rep* (2018) 8(1):11359. doi: 10.1038/s41598-018-29653-7
36. Zheng W, Li Q, Zou XB, Wang JW, Han F, Li F, et al. Evaluation of Contrast-Enhanced US LI-RADS Version 2017: Application on 2020 Liver Nodules in Patients With Hepatitis B Infection. *Radiology* (2020) 294(2):299–307. doi: 10.1148/radiol.2019190878

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Bao, Chen, Zhu, Xie and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 09 September 2021
doi: 10.3389/fonc.2021.709659

Radiomics Models for the Preoperative Prediction of Pelvic and Sacral Tumor Types: A Single-Center Retrospective Study of 795 Cases

Ping Yin[1†], Xin Zhi[1†], Chao Sun[1], Sicong Wang[2], Xia Liu[1], Lei Chen[1] and Nan Hong[1]*

[1] Department of Radiology, Peking University People's Hospital, Beijing, China, [2] Department of Pharmaceuticals Diagnosis, GE Healthcare, Shanghai, China

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln, United Kingdom

Reviewed by:
Swapnil Ulhas Rane,
Advanced Centre for Treatment, Research and Education in Cancer, India
Xiaohu Li,
First Affiliated Hospital of Anhui Medical University, China

***Correspondence:**
Nan Hong
hongnan1968@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 14 May 2021
Accepted: 02 August 2021
Published: 09 September 2021

Citation:
Yin P, Zhi X, Sun C, Wang S, Liu X, Chen L and Hong N (2021) Radiomics Models for the Preoperative Prediction of Pelvic and Sacral Tumor Types: A Single-Center Retrospective Study of 795 Cases. Front. Oncol. 11:709659. doi: 10.3389/fonc.2021.709659

Purpose: To assess the performance of random forest (RF)-based radiomics approaches based on 3D computed tomography (CT) and clinical features to predict the types of pelvic and sacral tumors.

Materials and Methods: A total of 795 patients with pathologically confirmed pelvic and sacral tumors were analyzed, including metastatic tumors (n = 181), chordomas (n = 85), giant cell tumors (n =120), chondrosarcoma (n = 127), osteosarcoma (n = 106), neurogenic tumors (n = 95), and Ewing's sarcoma (n = 81). After semi-automatic segmentation, 1316 hand-crafted radiomics features of each patient were extracted. Four radiomics models (RMs) and four clinical-RMs were built to identify these seven types of tumors. The area under the receiver operating characteristic curve (AUC) and accuracy (ACC) were used to evaluate different models.

Results: In total, 795 patients (432 males, 363 females; mean age of 42.1 ± 17.8 years) were consisted of 215 benign tumors and 580 malignant tumors. The sex, age, history of malignancy and tumor location had significant differences between benign and malignant tumors ($P < 0.05$). For the two-class models, clinical-RM2 (AUC = 0.928, ACC = 0.877) performed better than clinical-RM1 (AUC = 0.899, ACC = 0.854). For the three-class models, the proposed clinical-RM3 achieved AUCs between 0.923 (for chordoma) and 0.964 (for sarcoma), while the AUCs of the clinical-RM4 ranged from 0.799 (for osteosarcoma) to 0.869 (for chondrosarcoma) in the validation set.

Conclusions: The RF-based clinical-radiomics models provided high discriminatory performance in predicting pelvic and sacral tumor types, which could be used for clinical decision-making.

Keywords: radiomics, machine learning, neoplasms, classification, diagnostic imaging

INTRODUCTION

Pelvic and sacral tumors have various types, with metastatic tumors being the most common because of their prominent hematopoietic function until late in life (1, 2). Primary sacral tumors are rare and mainly include chordoma, giant cell tumor (GCT), neurogenic tumor, etc. (3, 4). In addition, a small number of GCTs and neurogenic tumors can also occur in the pelvis. Sarcomas, such as chondrosarcoma, osteosarcoma, and Ewing's sarcoma, are also the common primary malignant bone tumors of the pelvis and sacrum (5, 6). Because pelvic and sacral tumors are usually large and surrounded by complex structures, treatment for these tumors is often a challenging procedure that can be accompanied by serious complications, such as massive bleeding (7–10). In clinical practice, these tumors are treated differently. For example, preoperative chemotherapy is important for osteosarcoma, but it is not effective for chondrosarcoma. Therefore, accurate preoperative identification of these tumors is essential for the development of individualized treatment (11, 12).

Pelvic and sacral tumors share many similar clinical and imaging features, they often present heterogeneous masses with different components. This makes it difficult to identify these tumors in clinical practice, especially when they occur in unusual or multiple locations. In addition, the classic semantic assessments of traditional computed tomography (CT) and magnetic resonance (MR) images suffer from strong inter- and intra-reader variations, they do not provide sufficient diagnostic power to identify these tumors (13). Clinically, a simple and accurate method is needed to identify pelvic and sacral tumors.

Sophisticated machine learning methods show promise in complementing human diagnostics (14, 15). In recent years, radiomics has been successfully applied in the classification of tumor types. A few previous studies have identified sacral tumors using radiomics methods, but most of these studies identified only two types of tumors, and their sample sizes were relatively small (3, 11, 12). Yin et al. (3) built a random forest (RF)-based triple-classification radiomics model for the differentiation of primary chordomas, GCTs, and metastatic tumors of sacrum based on MR features. They concluded that their model is feasible to differentiate these tumors and can improve the precision of preoperative diagnosis in clinical practice. RF algorithms have a comparably low tendency to overfit and are well suited for multi-classification discrimination (13). Although machine learning has been applied to the segmentation, lesion detection, evaluation of chemotherapy response and prediction of local recurrence of bone tumors in recent years (16–20), the ability to identify multiple types of pelvic and sacral tumors remains unknown.

The aim of our study was to determine the performance of RF-based radiomics approaches based on CT features and clinical characteristics to predict the multiple types of pelvic and sacral tumors.

Abbreviations: GCT, giant cell tumor; CT, computed tomography; RF, random forest; ICC, intra- and interclass correlation coefficients; RM, radiomics model; AUC, area under the receiver operating characteristic curve; ACC, accuracy.

MATERIALS AND METHODS

Patients and Data Acquisition

After the approval of our local ethics committee, we conducted this single-center retrospective study and waived written informed consent. A total of 1000 patients with pathologically confirmed pelvic and sacral tumors in our institution from April 2006 to December 2019 were retrospectively analyzed. All patients had a single pelvic and sacral tumor that was detected on CT within 1 month before the initial surgery. Patients that had pelvic and sacral tumor types with a sample size less than 30 (n = 111), without preoperative CT images (n = 84), or with obvious artifacts (n = 10) were excluded. Finally, a total of 795 patients with pelvic and sacral tumor types of metastatic tumors (n = 181), chordomas (n = 85), GCTs (n =120), chondrosarcomas (n = 127), osteosarcomas (n = 106), neurogenic tumors (n = 95), or Ewing's sarcomas (n = 81) were included in the study. Sex, age, maximal tumor size, history of malignancy and tumor location (Zone I–IV) (21) of patients were also analyzed. Zone I includes the iliac crest, Zone II includes the acetabulum and its surroundings, Zone III includes the pubis and ischium regions, and Zone IV refers to the sacrum region. A "multi-zone" is a tumor that involves more than one area simultaneously.

All CT images were acquired on each patient using multi-detector row CT systems (Philips iCT 256, Philips Medical System; GE Lightspeed VCT 64, GE Medical System). The acquisition parameters were as follows: 120 kV, 100-370 mAs, slice thickness = 5 mm, matrix = 512 × 512 mm, field of view = 350 × 350 mm. The CT images were reconstructed with soft-tissue and bone kernel.

Tumor Segmentation

MITK software version 2018.04.2 (www.mitk.org) was used for the semi-automatic segmentation of all tumors (22). First, the two authors (PY and XL) manually delineated the edge of the lesion at the axial, sagittal, and coronal sites, respectively. Then, a three-dimensional lesion was automatically formed and manually corrected by a musculoskeletal radiologist with 5 years of experience and a senior musculoskeletal radiologist with 20 years of experience who were blinded to the assessment.

Feature Extraction and Selection

A total of 1316 radiomics features of each patient were extracted from the CT images using the Artificial Intelligence Kit software version 3.3.0 (GE Healthcare, China), including 18 first-order histogram features, 14 shape features, 24 gray-level co-occurrence matrix features, 14 gray-level dependence matrix features, 16 gray-level run-length matrix features, 16 gray-level size-zone matrix features, 5 neighboring gray-tone difference matrix features, 744 wavelet features, 186 Laplacian of Gaussian ($LoG_{sigma=2.0/3.0}$) features, and 279 local binary pattern features.

We preprocessed the data and normalized the extracted features. When the data value exceeded the range of mean value and standard deviation, the median of specific variance vector was used to replace the outliers. In addition, we standardized the data in a specific interval. The standardized formula is as follows: (fi−u)/std, where fi represents a single

characteristic data, u is the average value of the data column, and std pertains to the standard deviation of the data column.

To reduce overfitting or selection bias in our radiomics model, we used two methods to select the features: Spearman correlation as representative of filter models and gradient boosting decision tree (GBDT) as representative of embedded models (23, 24). The features with Spearman correlation > 0.7 were excluded to avoid overfitting. After the number of features was determined, the most predictive radiomics features were chosen to construct the final model.

Model Building and Validation

First, we divided the patients into the benign tumors group (n=215) and malignant tumors group (n=580), and built the first radiomics model (RM1) by using RF. After differentiating benign and malignant tumors, the specific types of benign and malignant tumors were then divided respectively. RM2, a two-class RF-based radiomics model, was built to identify GCTs and neurogenic tumors. After that, we constructed two triple-classification models, namely, RM3 and RM4. RM3 was used to identify metastatic tumors, chordomas, and sarcomas. RM4 was used to identify osteosarcomas, chondrosarcomas, and Ewing's sarcomas.

Clinical features were also compared and variables with P value < 0.05 were included in the clinical model. When combined RM with clinical data, we also constructed four clinical-RMs. In addition, we also constructed a seven-classification model, namely, clinica-RM5, to identify these seven types of tumors. In each model, all patients were randomly divided the patients into the training and validation sets by a ratio of 7:3. Models were trained with the training set by using the repeated 5-fold cross-validation method, and estimation performance was evaluated with the validation set.

The performance of different models was assessed using the area under the receiver operating characteristic curve (AUC) and accuracy (ACC). **Figure 1** shows the workflow of this study.

Statistical Analysis

All statistical analyses were performed with R (version 3.5.1) and Python (version 3.5.6). Mann-Whitney U test or Kruskal-Wallis H test was performed to compare continuous variables, while chi-squared test or Fisher's exact was used for categorical variables between groups. All statistical tests were two-sided, and Bonferroni-corrected P value were used to identify the feature significance of multiple comparisons.

RESULTS

Patient Characteristics

A total of 795 patients (432 males, 363 females; mean age of 42.1 ± 17.8 years, range 4–82 years; mean tumor size of 9.6 ± 4.0 cm) were included in this study (**Table 1**). In the clinical-RM1, sex, age, history of malignancy and tumor location had significant differences between groups (χ^2 = 9.111, Z = -3.962, χ^2 = 62.277, χ^2 = 149.379, P < 0.01). No significant difference was found in terms of tumor size between groups (Z = 0.534, P > 0.05). For the clinical-RM3, age, tumor size, history of malignancy and tumor location had significant differences between groups (Z = 248.6, Z = 55.167, χ^2 = 272.494, χ^2 = 181.17, P < 0.001). However, for the clinical-RM2 and clinical-RM4, significant differences were found in terms of age, tumor size and tumor location (P < 0.01), no significant difference was found in terms of sex and history of malignancy between groups (P > 0.05). There was no significant statistical difference between the training and validation sets in terms of age, sex, tumor location, tumor size, history of malignancy and histology (P > 0.05). The 95 neurogenic tumors were composed of 45 neurofibromas and 50 schwannomas. In the case of metastatic tumor, 43 metastases from lung, 27 from breast, 26 from liver, 25 from kidney, 15 from gastrointestinal, 13 from thyroid, 12 from prostate, 9 from uterus, 4 from bladder, 2 from melanoma, 2 from osteosarcoma, 2 from brain, and 1 from nasopharynx, respectively.

Performance of Different Models

For the two-class models, RM1 achieved an AUC of 0.949 and an ACC of 0.894, while RM2 reached an AUC of 0.974 and an ACC of 0.913 in the training set (**Figure 2** and **Table 2**). In the validating set, we found: RM2 (AUC = 0.863, ACC = 0.800) had a slightly higher performance than RM1 (AUC = 0.834, ACC = 0.782).

For the three-class models, the proposed radiomics classifiers achieved AUCs between 0.742 (for osteosarcoma, RM4) and 0.849 (for chordoma and sarcoma, RM3) in the validation set (**Figure 3** and **Table 3**). The ACCs of RM3 and RM4 were 0.665 and 0.667 in the validation set, respectively.

The AUC of the clinical models ranged from 0.645 to 0.951, and the ACC ranged from 0.583 to 0.833 in the validation set. When combined with clinical features, clinical-RMs performed better than individual RMs and clinical models. The clinical-RM1 exhibited an AUC of 0.899 and an ACC of 0.854 in the validation set. Similarly, the clinical-RM2 (AUC = 0.928, ACC = 0.877) performed better than individual RM2 in the validation set. The clinical-RM3 achieved AUCs between 0.923 (for chordoma) and 0.964 (for sarcoma), and ACC of 0.841 in the validation set. The AUCs of the clinical-RM4 ranged from 0.799 (for osteosarcoma) to 0.869 (for chondrosarcoma) in the validation set.

In addition, the AUC and ACC of clinical-RM5 in the training set were 0.771 and 0.580, and those in the validation set were 0.722 and 0.533, respectively. The Bar chart reflecting the predictive values of RM5 was shown in **Supplemental Figure 1**.

DISCUSSION

In this study, we found significant differences in terms of age, sex, history of malignancy and tumor location for differentiating benign and malignant pelvic and sacral tumors. Both two-class and three-class RMs had good performance in predicting pelvic and sacral tumor types. When combined with clinical data, the clinical-RMs performed better than individual RMs.

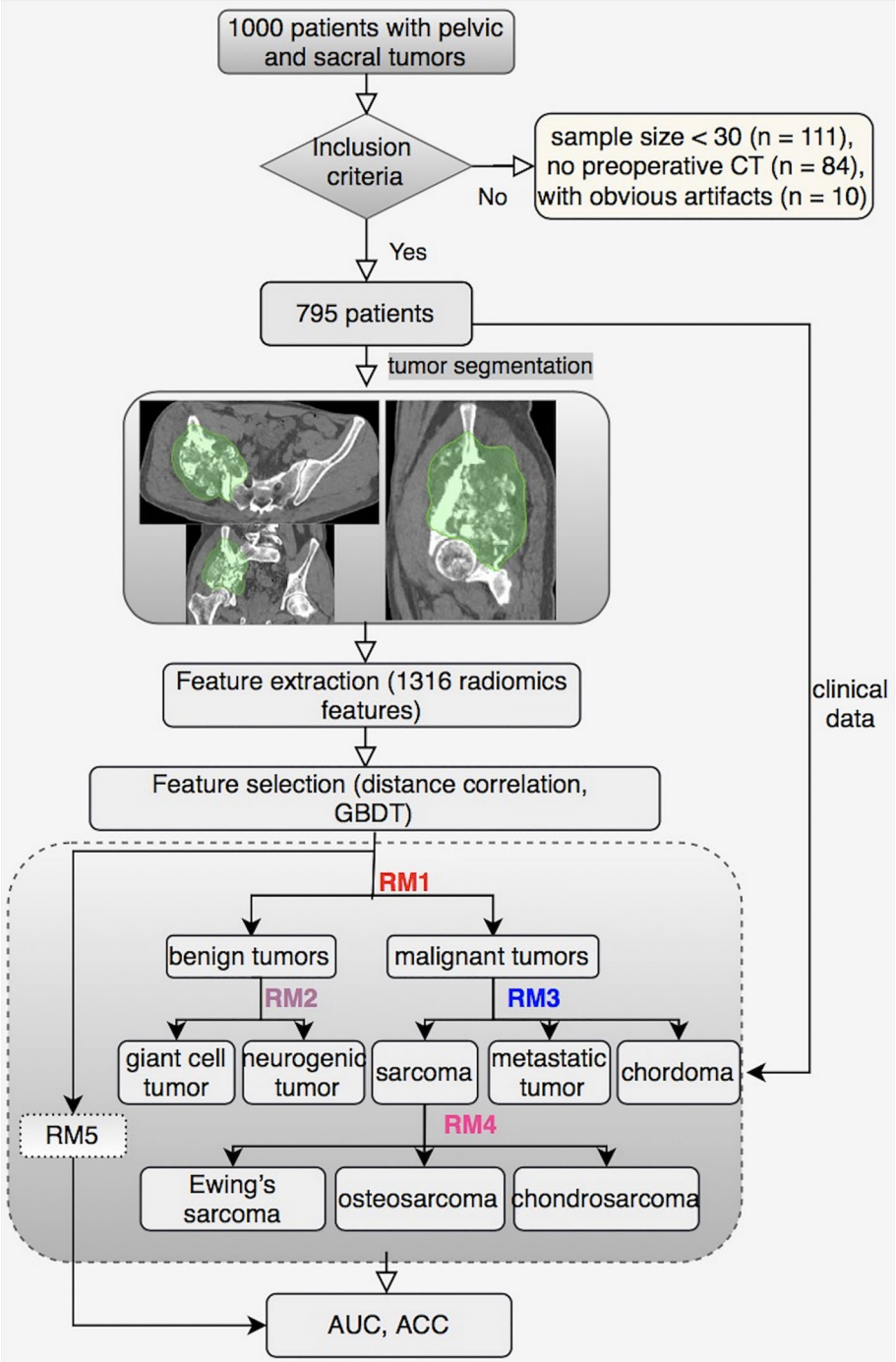


FIGURE 1 | The workflow of this study.

Pelvic and sacral tumors have a tendency to be silent till reaching extreme volumes and involving the adjacent nerve roots, blood vessels and organs (7). They often present as large heterogeneous masses, which are often difficult to identify on conventional imaging. Although metastatic tumors occur frequently in the pelvis and sacrum, a diagnostic dilemma can present when there is a single tumor with no history or evidence of malignancy elsewhere in the body (1). In our study, age, sex, history of malignancy and tumor location were considered to be important clinical features to differentiate benign and malignant pelvic and sacral tumors. The proportion of males in patients with malignant tumors was higher than that in patients with benign

TABLE 1 | Clinical characteristic of patients.

Variable	Class 1	Class 2	Class 3	Statistics	*P* value
Clinical-RM1					
Age (years)	36.00 (27.20, 48.00)	47.00 (27.00, 59.00)	–	-3.962	<0.001
Size (cm)	8.90 (7.02, 11.80)	9.00 (6.50, 12.10)	–	0.534	0.593
Female	117 (54.42%)	246 (42.41%)	–	9.111	0.003
Male	98 (45.58%)	334 (57.59%)	–		
Location I	10 (4.65%)	123 (21.21%)	–	149.379	<0.001
Location II	3 (1.40%)	49 (8.45%)	–		
Location III	3 (1.40%)	47 (8.10%)	–		
Location IV	186 (86.51%)	219 (37.76%)	–		
Multi-location	13 (6.05%)	142 (24.48%)	–		
No history of malignancy	211 (98.14%)	422 (72.76%)		62.277	<0.001
A history of malignancy	4 (1.86%)	158 (27.24%)			
Clinical-RM2					
Age (years)	44.00 (32.00, 53.80)	33.00 (25.00, 43.00)	–	4.726	<0.001
Size (cm)	8.50 (6.40, 11.66)	9.20 (7.80, 12.10)	–	-2.129	0.033
Female	51 (53.68%)	66 (55.00%)	–	0.037	0.847
Male	44 (46.32%)	54 (45.00%)	–		
Location I	1 (1.05%)	9 (7.50%)	–	–	0.005
Location II	0 (0.00%)	3 (2.50%)	–		
Location III	0 (0.00%)	3 (2.50%)	–		
Location IV	91 (95.79%)	95 (79.17%)	–		
Multi-location	3 (3.16%)	10 (8.33%)	–		
No history of malignancy	93 (97.89%)	118 (98.33%)		0.074	0.786
A history of malignancy	2 (2.11%)	2 (1.67%)			
Clinical-RM3					
Age (years)	29.00 (19.00, 45.05)	58.00 (50.00, 65.00)	59.00 (49.70, 68.30)	248.6	<0.001
Size (cm)	10.20 (7.39, 13.11)	7.60 (5.50, 9.80)	8.10 (6.20, 11.06)	55.167	<0.001
Female	133 (42.36%)	85 (46.96%)	28 (32.94%)	4.656	0.098
Male	181 (57.64%)	96 (53.04%)	57 (67.06%)		
Location I	81 (25.80%)	42 (23.20%)	0 (0.00%)	181.17	<0.001
Location II	29 (9.24%)	20 (11.05%)	0 (0.00%)		
Location III	35 (11.15%)	12 (6.63%)	0 (0.00%)		
Location IV	64 (20.38%)	71 (39.23%)	84 (98.82%)		
Multi-location	105 (33.44%)	36 (19.89%)	1 (1.18%)		
No history of malignancy	298 (94.90%)	50 (27.62%)	74 (87.06%)	272.494	<0.001
A history of malignancy	16 (5.10%)	131 (72.38%)	11 (12.94%)		
Clinical-RM4					
Age (years)	17.00 (13.00, 25.30)	26.00 (19.00, 34.05)	44.00 (33.00, 52.00)	127.47	<0.001
Size (cm)	8.00 (5.41, 10.26)	11.75 (8.00, 14.41)	11.00 (8.62, 13.58)	31.792	<0.001
Female	26 (32.10%)	47 (44.34%)	60 (47.24%)	4.904	0.086
Male	55 (67.90%)	59 (55.66%)	67 (52.76%)		
Zone I	23 (28.40%)	28 (26.42%)	30 (23.62%)	24.453	0.002
Zone II	5 (6.17%)	8 (7.55%)	16 (12.60%)		
Zone III	10 (12.35%)	8 (7.55%)	17 (13.39%)		
Zone IV	28 (34.57%)	16 (15.09%)	20 (15.75%)		
Multi-zone	15 (18.52%)	46 (43.40%)	44 (34.65%)		
No history of malignancy	76 (93.83%)	99 (93.40%)	123 (96.85%)	1.687	0.43
A history of malignancy	5 (6.17%)	7 (6.60%)	4 (3.15%)		

Clinical-RM1: class 1 = benign tumor, class 2 = malignant tumor; Clinical-RM2: class 1 = neurogenic tumor, class 2 = giant cell tumor; Clinical-RM3: class 1 = sarcoma, class 2 = metastatic tumor, class 3 = chordoma; Clinical-RM4: class 1 = Ewing's sarcoma, class 2 = osteosarcoma, class 3 = chondrosarcoma.

tumors, which is consistent with previous study (12). For the clinical-RM3, age, tumor size, history of malignancy and tumor location had significant differences between groups. For the clinical-RM2 and clinical-RM4, however, significant differences were found in terms of age, tumor size and tumor location. Chordomas of the pelvis and sacrum occur almost exclusively in the sacrum, but can also involve the pelvis when the tumor is large, increasing the difficulty of identification. Neurogenic tumors and GCTs of bone also occur more frequently in the sacrum than in the pelvis. In this study, 4 neurogenic tumors and 25 GCTs occurred in the pelvis. Nevertheless, more sarcomas occur in the pelvis than in the sacrum. Chondrosarcoma is the most common primary malignant bone tumor in the pelvis, followed by osteosarcoma and Ewing's sarcoma (5, 6). Osteosarcomas and Ewing's sarcomas tend to occur between the ages of 10 and 30, while chondrosarcoma mostly affects people in their 40s to 70s (6). In our study, we found that the mean age of Ewing's sarcoma was the lowest among these tumors, which is consistent with previous study (25). Park et al. (2) reported that osteosarcoma of the pelvic bones was more frequent in older patients, which may be due to their limited sample size. GCTs of bone typically affect younger patients between 20 and 30 years, and neurogenic tumors tend to

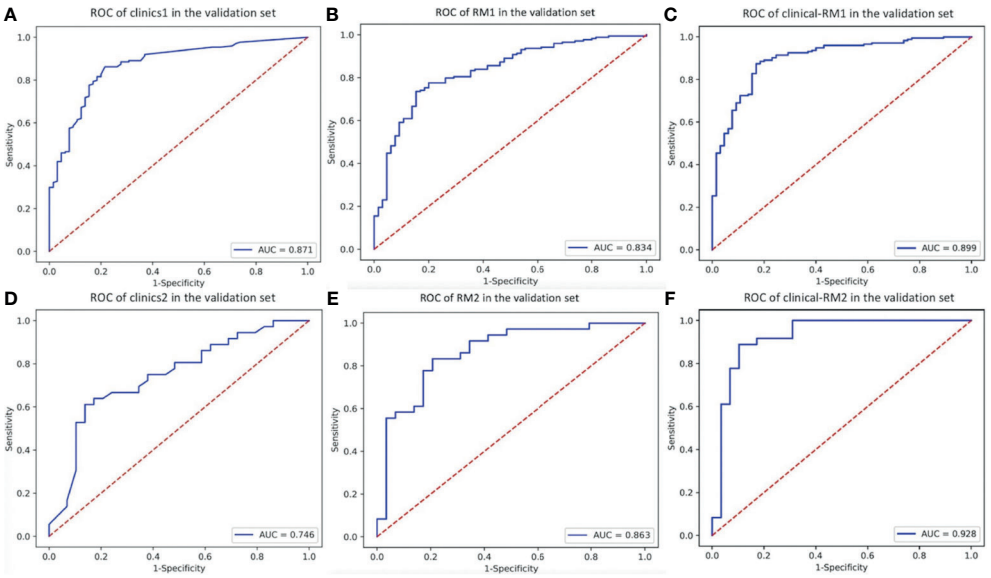


FIGURE 2 | The ROC curve of two-class models in the validation set. **(A)**, clinics1. **(B)**, RM1. **(C)**, clinical-RM1. **(D)**, clinics2. **(E)**, RM2. **(F)**, clinical-RM2.

TABLE 2 | Performance of two-class models in the validation set.

	AUC	ACC	Sensitivity	Specificity	PPV	NPV
RM1	0.834	0.782	0.891	0.492	0.824	0.627
RM2	0.863	0.800	0.917	0.655	0.767	0.864
Clinics1	0.871	0.833	0.874	0.723	0.894	0.681
Clinics2	0.746	0.646	0.833	0.414	0.638	0.667
Clinical-RM1	0.899	0.854	0.948	0.600	0.864	0.812
Clinical-RM2	0.928	0.877	0.889	0.862	0.889	0.862

AUC, area under curve; ACC, accuracy; PPV, positive predictive value; NPV, negative predictive value.

occur between 20 and 50 years old without significant age predominance (4). In line with previous studies (1, 26, 27), we found that metastatic tumors and chordomas were more common in older populations. Furthermore, the size of the GCTs was significantly larger than that of the neurogenic tumor. The size of the sarcoma was significantly larger than that of the metastatic tumor and chordoma. For sarcomas, the size of the Ewing's sarcoma was significantly smaller than that of the osteosarcoma and chondrosarcoma. As expected, the proportion of patients with metastatic tumors who had a history of malignancy was significantly higher than that of other tumor types.

Until now, there are few studies on the differentiation of pelvic and sacral tumors using machine learning methods (3, 11, 12). Yin et al. (12) compared the performance of radiomics model based on CT and MR features to identify sacral tumors. They found that clinical-radiomics nomogram performed better than radiomics nomogram. Most of previous studies investigated two-class problems, which intrinsically achieve higher AUCs compared with multiclass problems (13). In this study, we built four radiomics classifiers to identify seven types of pelvic and sacral tumors using RF. RF was widely used in multi-class machine learning due to its high accuracy and low overfitting (28, 29). It is a relatively efficient model-free method both in variable selection and classification (3). Kniep et al. (13) built a RF-based five-class radiomics model to predict the metastatic tumor type of brain and found it is superior to the radiologist's readings. In our study, we first proposed a four-step models framework to improve the identification efficiency of models. Using this multi-model framework, we identified seven types of pelvic and sacral tumors, beginning with benign and malignant tumors and their subgroups. We found that both two-class and three-class RF-based RMs had good performance in predicting pelvic and sacral tumor types. Although the RMs performed worse than the clinical models alone, except for the RM2. This may be related to the fact that the clinical features included in this study are important features for differentiating these tumors and that they differ significantly. When combined with clinical data, the clinical-RMs performed better than individual RMs, which is consistent with previous studies (6, 12). Furthermore, the AUC of the seven-classification model was lower than that of other models, which may be related to the performance of RF decreases as the number of categories increases.

Our study has certain limitations. First, all images were collected from one center over the past decade or so, and from two different machines. Although we included large sample data for the study, a multicenter prospective study is beneficial to future research. Second, only plain CT image data were available in this study. Multimodal data may be required, such as enhanced CT and MR, which may provide more useful information for the differentiation of lesions. Third, we did not compare the performance of our model with that of the radiologists, and we will conduct further research in the following study. Last, we cannot include all types of pelvic and sacral tumors because some types are extremely rare. Our models

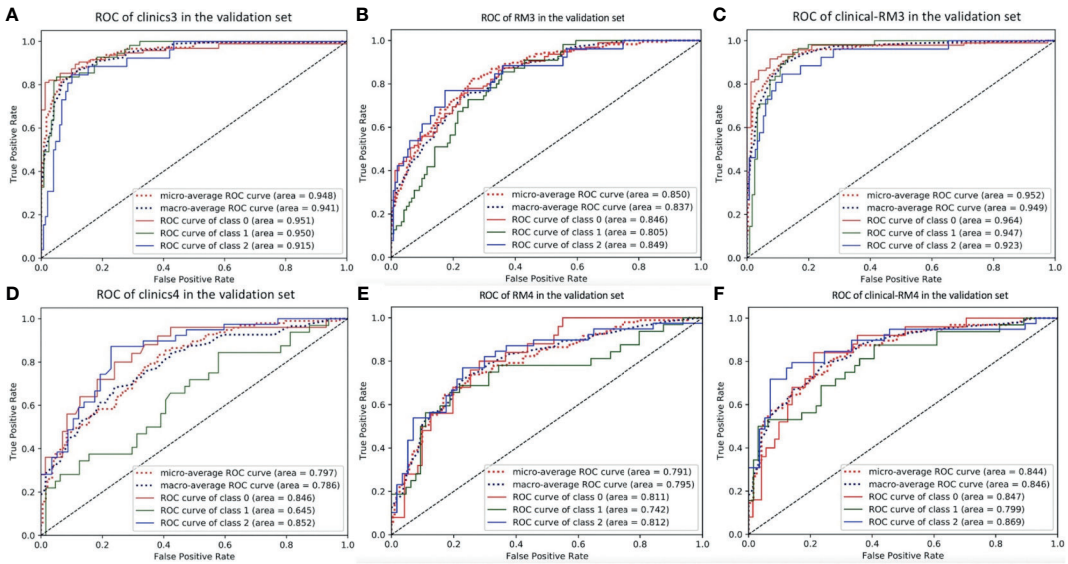


FIGURE 3 | The ROC curve of three-class models in the validation set. **(A)**, clinics3. **(B)**, RM3. **(C)**, clinical-RM3. **(D)**, clinics4. **(E)**, RM4. **(F)**, clinical-RM4. For clinics3, RM3 and clinical-RM3: class 0 = sarcoma, class 1 = metastatic tumor, class 2 = chordoma. For clinics4, RM4 and clinical-RM4: class 0 = Ewing's sarcoma, class 1 = osteosarcoma, class 2 = chondrosarcoma.

TABLE 3 | Performance of three-class models in the validation set.

	AUC	ACC	Precision	Recall	F1-score
RM3					
metastatic tumor	0.805	0.665	0.645	0.364	0.465
chordoma	0.849	0.665	0.818	0.346	0.486
sarcoma	0.846	0.665	0.657	0.926	0.769
RM4					
chondrosarcoma	0.812	0.667	0.682	0.769	0.723
osteosarcoma	0.742	0.667	0.636	0.656	0.646
Ewing's sarcoma	0.811	0.667	0.684	0.520	0.591
Clinics3					
metastatic tumor	0.950	0.824	0.836	0.836	0.836
chordoma	0.915	0.824	0.609	0.538	0.571
sarcoma	0.951	0.824	0.867	0.895	0.881
Clinics4					
chondrosarcoma	0.852	0.583	0.743	0.667	0.703
osteosarcoma	0.645	0.583	0.448	0.406	0.426
Ewing's sarcoma	0.846	0.583	0.531	0.680	0.596
Clinical-RM3					
metastatic tumor	0.947	0.841	0.780	0.836	0.807
chordoma	0.923	0.841	0.737	0.538	0.622
sarcoma	0.964	0.841	0.898	0.926	0.912
Clinical-RM4					
chondrosarcoma	0.869	0.667	0.756	0.795	0.775
osteosarcoma	0.799	0.667	0.625	0.625	0.625
Ewing's sarcoma	0.847	0.667	0.565	0.520	0.542

AUC, area under curve; ACC, accuracy.

include seven types of the most common pelvic and sacral tumors, and we continue to believe that our models can be of great clinical benefit.

In conclusion, the RF-based clinical-RMs provided high discriminatory performance in predicting pelvic and sacral tumor types. Our models can provide a simple, non-invasive and accurate auxiliary diagnostic tool for the differentiation of pelvic and sacral tumors, improving the diagnostic efficiency of clinicians.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Peking University People's Hospital Ethics Committee. Written informed consent was waived.

AUTHOR CONTRIBUTIONS

Guarantor of integrity of the entire study: NH. Study concepts: PY and NH. Study design: PY and NH. Definition of intellectual content: PY and XZ. Literature research: PY and XZ. Clinical studies: PY and XZ. Experimental studies: PY and XL. Data acquisition: PY and CS. Data analysis: PY and SW. Statistical analysis: SW and PY. Manuscript preparation: PY and LC. Manuscript editing: PY and XZ. Manuscript review: NH.

All authors contributed to the article and approved the submitted version.

FUNDING

Roles of funding sources: Project (RDY2020-08) supported by Peking University People's Hospital Scientific Research Development Funds.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.709659/full#supplementary-material

Supplementary Figure 1 | Bar chart shows the predictive values of RM5. Label 1 = metastatic tumor, Label 2 = chordoma, Label 3 = Ewing's sarcoma, Label 4 = osteosarcoma, Label 5 = chondrosarcoma, Label 6 = neurogenic tumors, Label 7 = GCTs.

REFERENCES

1. Thornton E, Krajewski KM, O'Regan KN, Giardino AA, Jagannathan JP, Ramaiya N. Imaging Features of Primary and Secondary Malignant Tumours of the Sacrum. *Br J Radiol* (2012) 85(1011):279–86. doi: 10.1259/bjr/25247602
2. Park SK, Lee IS, Cho KH, Lee YH, Yi JH, Choi KU. Osteosarcoma of Pelvic Bones: Imaging Features. *Clin Imaging* (2017) 41:59–64. doi: 10.1016/j.clinimag.2016.10.013
3. Yin P, Mao N, Zhao C, Wu J, Chen L, Hong N. A Triple-Classification Radiomics Model for the Differentiation of Primary Chordoma, Giant Cell Tumor, and Metastatic Tumor of Sacrum Based on T2-Weighted and Contrast-Enhanced T1-Weighted MRI. *J Magn Reson Imaging* (2019) 49 (3):752–9. doi: 10.1002/jmri.26238
4. Si MJ, Wang CS, Ding XY, Yuan F, Du LJ, Lu Y, et al. Differentiation of Primary Chordoma, Giant Cell Tumor and Schwannoma of the Sacrum by CT and MRI. *Eur J Radiol* (2013) 82(12):2309–15. doi: 10.1016/j.ejrad.2013.08.034
5. Tang X, Guo W, Yang R, Tang S, Ji T. Evaluation of Blood Loss During Limb Salvage Surgery for Pelvic Tumours. *Int Orthop* (2009) 33(3):751–6. doi: 10.1007/s00264-008-0695-8
6. Yin P, Mao N, Liu X, Sun C, Wang S, Chen L, et al. Can Clinical Radiomics Nomogram Based on 3D Multiparametric MRI Features and Clinical Characteristics Estimate Early Recurrence of Pelvic Chondrosarcoma? *J Magn Reson Imaging* (2020) 51(2):435–45. doi: 10.1002/jmri.26834
7. Luo Y, Duan H, Liu W, Min L, Shi R, Zhang W, et al. Clinical Evaluation for Lower Abdominal Aorta Balloon Occluding in the Pelvic and Sacral Tumor Resection. *J Surg Oncol* (2013) 108(3):148–51. doi: 10.1002/jso.23376
8. Freeman AK, Thorne CJ, Gaston CL, Shellard R, Neal T, Parry MC, et al. Hypotensive Epidural Anesthesia Reduces Blood Loss in Pelvic and Sacral Bone Tumor Resections. *Clin Orthop Relat Res* (2017) 475(3):634–40. doi: 10.1007/s11999-016-4858-4
9. Tang X, Guo W, Yang R, Tang S, Ji T. Risk Factors for Blood Loss During Sacral Tumor Resection. *Clin Orthop Relat Res* (2009) 467(6):1599–604. doi: 10.1007/s11999-008-0483-1
10. Ji T, Yang Y, Wang Y, Sun K, Guo W. Combining of Serial Embolization and Denosumab for Large Sacropelvic Giant Cell Tumor: Case Report of 3 Cases. *Med (Baltimore)* (2017) 96(33):e7799. doi: 10.1097/MD.0000000000007799
11. Yin P, Mao N, Zhao C, Wu J, Sun C, Chen L, et al. Comparison of Radiomics Machine-Learning Classifiers and Feature Selection for Differentiation of Sacral Chordoma and Sacral Giant Cell Tumour Based on 3D Computed Tomography Features. *Eur Radiol* (2019) 29(4):1841–7. doi: 10.1007/s00330-018-5730-6
12. Yin P, Mao N, Wang S, Sun C, Hong N. Clinical-Radiomics Nomograms for Pre-Operative Differentiation of Sacral Chordoma and Sacral Giant Cell Tumor Based on 3D Computed Tomography and Multiparametric Magnetic Resonance Imaging. *Br J Radiol* (2019) 92(1101):20190155. doi: 10.1259/bjr.20190155
13. Kniep HC, Madesta F, Schneider T, Hanning U, Schönfeld MH, Schön G, et al. Radiomics of Brain MRI: Utility in Prediction of Metastatic Tumor Type. *Radiology* (2019) 290(2):479–87. doi: 10.1148/radiol.2018180946
14. Truhn D, Schrading S, Haarburger C, Schneider H, Merhof D, Kuhl C. Radiomic Versus Convolutional Neural Networks Analysis for Classification of Contrast-Enhancing Lesions at Multiparametric Breast MRI. *Radiology* (2019) 290(2):290–7. doi: 10.1148/radiol.2018181352
15. Cuocolo R, Caruso M, Perillo T, Ugga L, Petretta M. Machine Learning in Oncology: A Clinical Appraisal. *Cancer Lett* (2020) 481:55–62. doi: 10.1016/j.canlet.2020.03.032
16. Shimizu H, Nakayama KI. Artificial Intelligence in Oncology. *Cancer Sci* (2020) 111(5):1452–60. doi: 10.1111/cas.14377
17. Bini SA. Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? *J Arthroplasty* (2018) 33(8):2358–61. doi: 10.1016/j.arth.2018.02.067
18. Dufau J, Bouhamama A, Leporq B, Malaureille L, Beuf O, Gouin F, et al. Prediction of Chemotherapy Response in Primary Osteosarcoma Using the Machine Learning Technique on Radiomic Data. *Bull Cancer* (2019) 106 (11):983–99. doi: 10.1016/j.bulcan.2019.07.005
19. Lindgren Belal S, Sadik M, Kaboteh R, Enqvist O, Ulén J, Poulsen MH, et al. Deep Learning for Segmentation of 49 Selected Bones in CT Scans: First Step in Automated PET/CT-Based 3D Quantification of Skeletal Metastases. *Eur J Radiol* (2019) 113:89–95. doi: 10.1016/j.ejrad.2019.01.028
20. He Y, Guo J, Ding X, van Ooijen PMA, Zhang Y, Chen A, et al. Convolutional Neural Network to Predict the Local Recurrence of Giant Cell Tumor of Bone After Curettage Based on Pre-Surgery Magnetic Resonance Images. *Eur Radiol* (2019) 29(10):5441–51. doi: 10.1007/s00330-019-06082-2
21. Enneking WF, Dunham WK. Resection and Reconstruction for Primary Neoplasms Involving the Innominate Bone. *J Bone Joint Surg Am* (1978) 60 (6):731–46. doi: 10.2106/00004623-197860060-00002
22. Wolf I, Vetter M, Wegner I, Böttger T, Nolden M, Schöbinger M, et al. The Medical Imaging Interaction Toolkit. *Med Image Anal* (2005) 9(6):594–604. doi: 10.1016/j.media.2005.04.005
23. Zhang Y, Chen C, Cheng Y, Teng Y, Guo W, Xu H, et al. Ability of Radiomics in Differentiation of Anaplastic Oligodendroglioma From Atypical Low-Grade Oligodendroglioma Using Machine-Learning Approach. *Front Oncol* (2019) 9:1371. doi: 10.3389/fonc.2019.01371
24. Fan Y, Chen C, Zhao F, Tian Z, Wang J, Ma X, et al. Radiomics-Based Machine Learning Technology Enables Better Differentiation Between Glioblastoma and Anaplastic Oligodendroglioma. *Front Oncol* (2019) 9:1164. doi: 10.3389/fonc.2019.01164
25. Zhu C, Olson KA, Roth M, Geller DS, Gorlick RG, Gill J, et al. Provider Views on the Management of Ewing Sarcoma of the Spine and Pelvis. *J Surg Oncol* (2018) 117(3):417–24. doi: 10.1002/jso.24860
26. Gerber S, Ollivier L, Leclère J, Vanel D, Missenard G, Brisse H, et al. Imaging of Sacral Tumours. *Skeletal Radiol* (2008) 37:277–89. doi: 10.1007/s00256-007-0413-4
27. Diel J, Ortiz O, Losada RA, Price DB, Hayt MW, Katz DS. The Sacrum: Pathologic Spectrum, Multimodality Imaging, and Subspecialty Approach. *Radiographics* (2001) 21:83–104. doi: 10.1148/radiographics.21.1.g01ja0883
28. Ho TK. Random Decision Forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, Qc. (1995), 278–82.
29. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJ. Machine-Learning Methods for Quantitative Radiomic Biomarkers. *Sci Rep* (2015) 5:13087. doi: 10.1038/srep13087

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in

this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Yin, Zhi, Sun, Wang, Liu, Chen and Hong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 10 September 2021
doi: 10.3389/fonc.2021.725922

Radiomic Evaluations of the Diagnostic Performance of DM, DBT, DCE MRI, DWI, and Their Combination for the Diagnosis of Breast Cancer

Shuxian Niu[1†], Xiaoyu Wang[2†], Nannan Zhao[2†], Guanyu Liu[2], Yangyang Kan[2], Yue Dong[2], E-Nuo Cui[3], Yahong Luo[2], Tao Yu[2] and Xiran Jiang[1*]*

[1] Department of Biomedical Engineering, School of Fundamental Sciences, China Medical University, Shenyang, China, [2] Liaoning Cancer Hospital and Institute, Cancer Hospital of China Medical University, Shenyang, China, [3] School of Computer Science and Engineering, Shenyang University, Shenyang, China

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln, United Kingdom

Reviewed by:
S Senthil Kumaran,
All India Institute of Medical Sciences, India
Sudarshan Ragunathan,
Barrow Neurological Institute (BNI), United States

***Correspondence:**
Tao Yu
dryutao@hotmail.edu.cn
Xiran Jiang
xrjiang@cmu.edu.cn

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 16 June 2021
Accepted: 23 August 2021
Published: 10 September 2021

Citation:
Niu S, Wang X, Zhao N, Liu G, Kan Y, Dong Y, Cui E-N, Luo Y, Yu T and Jiang X (2021) Radiomic Evaluations of the Diagnostic Performance of DM, DBT, DCE MRI, DWI, and Their Combination for the Diagnosis of Breast Cancer. Front. Oncol. 11:725922. doi: 10.3389/fonc.2021.725922

Objectives: This study aims to evaluate digital mammography (DM), digital breast tomosynthesis (DBT), dynamic contrast-enhanced (DCE), and diffusion-weighted (DW) MRI, individually and combined, for the values in the diagnosis of breast cancer, and propose a visualized clinical-radiomics nomogram for potential clinical uses.

Methods: A total of 120 patients were enrolled between September 2017 and July 2018, all underwent preoperative DM, DBT, DCE, and DWI scans. Radiomics features were extracted and selected using the least absolute shrinkage and selection operator (LASSO) regression. A radiomics nomogram was constructed integrating the radiomics signature and important clinical predictors, and assessed with the receiver operating characteristic (ROC) curve, calibration curve, and decision curve analysis (DCA).

Results: The radiomics signature derived from DBT plus DM generated a lower area under the ROC curve (AUC) and sensitivity, but a higher specificity compared with that from DCE plus DWI. The nomogram integrating the combined radiomics signature, age, and menstruation status achieved the best diagnostic performance in the training (AUCs, nomogram *vs.* combined radiomics signature *vs.* clinical model, 0.975 *vs.* 0.964 *vs.* 0.782) and validation (AUCs, nomogram *vs.* combined radiomics signature *vs.* clinical model, 0.983 *vs.* 0.978 *vs.* 0.680) cohorts. DCA confirmed the potential clinical usefulness of the nomogram.

Conclusions: The DBT plus DM provided a lower AUC and sensitivity, but a higher specificity than DCE plus DWI for detecting breast cancer. The proposed clinical-radiomics nomogram has diagnostic advantages over each modality, and can be considered as an efficient tool for breast cancer screening.

Keywords: breast, mammography, MRI, radiomics, nomogram

INTRODUCTION

Breast cancer has been a major concern and the second leading cause of cancer death among women (1). The prevalence of breast cancer has increased in the recent years, mainly due to the implementation of an early screening mammography (2). Although there is still no effective way to prevent breast cancer, studies have shown that early detection and treatment can increase the chance of full recovery for the patients (3).

Digital mammography (DM) using 2D technique, as a widely used tool for detecting breast cancer, has a serious limitation that the visibility of lesions may be decreased since they are frequently obscured by dense fibroglandular and other normal tissues within the breast (4), which often leads to a missed diagnosis or misdiagnosis (5). To address this issue, digital breast tomosynthesis (DBT) rotates the X-ray tubes in a limited angle, thus allowing an improved identification of anomalies obscured by normal tissues (6, 7). Therefore, the DBT is commonly considered to be capable of decreasing the recall rates and increasing the detection rates for breast cancer compared with DM (8). Magnetic resonance imaging (MRI), as another popular tool for breast screening, has been demonstrated to be very sensitive in detecting breast cancer (9). While, the relative low specificity of MRI screening may lead to a high rate of overtreatment (10). Besides, the high examination fees of MRI also hinder the clinical application in early breast screening.

In the clinical practice, the diagnosis of breast cancer based on DM, DBT, or MRI mainly relies on visual inspections of the morphological changes of breast lesions, including size, shape, and gray level changes, and, thus, require experienced clinicians to make decisions. Previous reports have compared the diagnostic capabilities of DM with DBT (11, 12) and mammography with MRI (13, 14), all based on subjective visual examinations and the lack of quantified assessments. Recently, the radiomics-based computer aided diagnosis (CAD) has received increasing attention due to its quantitative advantages (15, 16). By using automated data characterization algorithms, the radiomics can extract and select discriminative and quantified features from a region of interest, which were shown to reflect biological information regarding the tumor and were highly correlated with disease status (17). Subsequent analysis, including statistics, machine learning classifiers, and nomogram can give associations between imaging features and the underlying pathophysiology (18). Radiomics-based studies on breast cancer have been proposed for predicting the axillary lymph node metastasis (19–23), molecular subtypes (24–28), tumor grades (29–31), and treatment responses (32–37). Some recent studies also conducted a radiomics-based quantified analysis for the diagnosis of breast cancer based on DM (38, 39), DBT (40, 41), and MRI (42, 43) separately, and demonstrated improvements of the diagnostic performance using radiomics compared with visual examinations by radiologists. A recent effort evaluated T2W, DCE, and DWI separately and in combination, but ignored the clinical values of mammography screening, and lack of correlating their findings with clinical evaluation, which may limit the clinical applicability (44).

To our knowledge, direct and quantified comparisons among MD, DBT, and MRI have not been reported. Therefore, the present study aims to widen the understanding of mammography and MRI in breast cancer screening by directly and quantitatively comparing the diagnostic efficiency of each modality individually and in combination. Besides, this study aims to propose a visualized clinical-radiomics nomogram based on the optimal imaging combination and important clinical factors for early assessment of suspected breast lesions.

MATERIAL AND METHODS

Patients

This retrospective analysis of breast DM, DBT, and MRI data was approved by the Institutional Research Ethics Board of our institute (Approval No. 2013010). The informed consent requirement was waived. A total of 120 patients [mean age ± standard deviation (SD), 48.81 ± 10.83] were enrolled between September 2017 and July 2018 in our hospital. The number of the patients harboring pathologically confirmed benign or malignant lesions were 50 and 70, respectively. Inclusion criteria were as follows: (i) older than 18 years; (ii) underwent DM, DBT, and MRI screening before surgery; and (iii) underwent surgical resection with pathological confirmation. Exclusion criteria were: (i) combined with other tumor diseases; (ii) during menstruation, pregnancy, or lactation periods; (iii) history of breast surgery, radiotherapy, or chemotherapy, as well as breast implants; and (iv) having artifacts in the images. All patients were randomly divided into training and validation cohorts at a 2:1 ratio using stratified sampling. Clinical factors including age, family history of breast cancer, history of biopsy, and menstruation status were obtained from the electronic medical record system of our hospital.

Digital Mammography, Digital Breast Tomosynthesis, and Magnetic Resonance Imaging Acquisitions

Preoperative DM and DBT examinations were performed by a radiographer with 10 years of work experience using a DBT scanner (Hologic Selenia Dimensions, Hologic, USA). The obtained images of the compressed breast were reconstructed with a 1-mm intersection spacing to give a three-dimensional view of the tissue, slice by slice, and suitably spaced. The number of the slices depends on the compressed breast thickness. The following parameters were used to perform the DBT scanning: The voltage range of the X-ray tubes: 20.0–49.0 kV (step: 1.0 kV), nominal power: 3.0 kW, current time range: 300–400 mAs, scanning time < 4.0 s, reconstruction time: 2.0–5.0 s, and pixel size: 70 μm. The obtained DBT images were interpreted on a Hologic breast computer-aided diagnosis (CAD) workstation (SecureViewDx; Hologic) equipped with two 5-megapixel monitors.

Preoperative MRI scans were performed using a 1.5-T MRI scanner (HDx, GE Healthcare). The axial diffusion-weighted imaging was used with the following parameters: the b-value: 800 s/mm^2, repetition time (TR)/echo time (TE)/inversion time (TI): 5,000 ms/64 ms/0 ms, flip angle: 90°, slice thickness: 6 mm, slice gap: 7.5 mm, field of view: 240 mm, matrix size: 128 × 128. The axially vibrant sequence (a 3D T1-weighted imaging

technique covering bilateral breasts conventional scans or dynamic enhanced scans to obtain axial or sagittal images with high signal-to-noise ratio and high resolution) with the following parameters: TR/TE/TI: 6.2 ms/3.0 ms/13 ms; flip angle: 10°; slice thickness: 3.2 mm; slice gap: 3.2 mm, 48 slices per volume; field of view: 360 mm; matrix size: 350 × 350. The contrast agent was injected intravenously (0.1 mmol/kg of Gd-DTPA-MBA, Omniscan, GE Healthcare), followed by a 20-mL saline flush, both at the rate of 3 ml/s. After the intravenous injection, continuous non-interval scans were performed in eight phases, with a scan time for each phase of 43 seconds. All scanned images were stored in the Picture Archiving and Communication System (PACS) in our hospital in a Digital Imaging and Communications in Medicine (DICOM) format. The details about their scan parameters are shown in **Supplementary Tables S1, S2**.

Breast Lesion Segmentation

Regions of interest (ROIs) were manually segmented slice by slice for each patient using the ITK-SNAP software (version 3.6.0) by a radiologist with 12 years of working experience according to the breast imaging reporting and data system (BI-RADS). The radiologist was blinded to the pathological results for the patients. The ROIs included the breast lesions and edges, exporting as a compressed package in an NII format for further analysis.

Radiomics Feature Extraction

Radiomics features including 18 first-order statistical, 13 shape-based, and 74 textual features were extracted based on the segmented ROIs using the Pyradiomics package in Python 3.6 (https://pyradiomics.readthedocs.io/en/). The texture feature category consists of the gray level cooccurence 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) features. The first-order and texture features were also calculated from the original images that were filtered with eight types of filters: logarithm, square, gradient, exponential, laplacian of Gaussian, wavelet, and localbinarypattern2D (45). Detailed descriptions of the features and calculation protocols can be found in a previous report (46).

Feature Selection

To obtain reliable and discriminative features, 30 patients were randomly selected to perform the intraclass correlation coefficient (ICC) analysis (47), 15 from the training group and 15 from the validation group. The ROIs were double-blind segmented by another radiologist with 8 years of working experience. Features with ICC > 0.75 were retained, then further selected by the Mann-Whitney *U* test. Features with $P < 0.05$ were considered significant variables between the benign and malignant groups. Finally, the least absolute shrinkage and selection operator (LASSO) logistic regression was used to identify the most discriminative features with a 10-fold cross-validation for selecting the parameter lambda using the "glmnet" package in R language v3.6 (available from URL: https://www.r-project.org) (48).

Development of the Radiomics Signature, Clinical Model, and Nomogram

The radiomics signature formula was calculated for each patient by a linear combination of the selected features weighted by the respective LASSO coefficients. The logistic regression was used to identify the discriminative clinical predictors. A clinical model was established using the multivariate logistic regression with the Akaike's Information Criterion (AIC) as the stopping rule (49). A radiomics nomogram for differentiating benign and malignant lesions was constructed incorporating the radiomics signature and the most important clinical factors using the "rms" package in R v.3.6.

Statistical Analysis

The Mann-Whitney *U*-test, t-test, Chi-Square test, and Shapiro-Wilk test were performed on continuous and discrete variables, respectively. All hypothesis tests were two-sided. The ROC curve analysis was performed to evaluate the diagnostic performance of each model, with the area under the ROC curve (AUC), accuracy, sensitivity, and specificity calculated as comparison metrics. The optimal cutoff value was obtained on the ROC curve with the maximum Youden index (50). ROC curves were evaluated with the DeLong test using the "pROC" package in R. Calibration curves were plotted to assess the calibration of the model-predicted results with truth values. The decision curve analysis (DCA) (51) was performed using the "rmda" package to assess the potential clinical usefulness of the models.

RESULTS

Patient Characteristics

The clinical characteristics of the patients were statistically analyzed and shown in **Table 1**. The age and menstruation status were significantly different between the benign and malignant groups ($P < 0.05$). No statistical difference was observed in the types of family history and history of biopsy. A clinical model was built integrating the age and menstruation status for detecting malignant lesions.

Evaluation of Diagnostic Performance of Digital Mammography, Digital Breast Tomosynthesis, and Magnetic Resonance Imaging

Diagnostic performance of the radiomics signature derived from the DM, DBT, DCE, and DWI individually and in combination were assessed (**Table 2**). **Figure 1** shows the ROC curves of each radiomics signature. The results indicated that the DCE generated

TABLE 1 | >Statistical analysis results of clinical characteristics.

Characteristic	Training cohort		*P*	Validation cohort		*P*
	Benign (n = 33)	Malignant (n = 46)		Benign (n = 17)	Malignant (n = 24)	
Age (years)			0.008			0.009
<40	10 (30.3)	4 (8.7)		6 (35.2)	2 (8.3)	
40–49	14 (42.4)	19 (41.3)		8 (47.1)	9 (37.5)	
50–59	8 (24.2)	11 (23.9)		3 (17.6)	4 (16.7)	
>=60	1 (3.0)	12 (26.1)		0 (0.0)	9 (37.5)	
Family history of breast cancer, n (%)			0.693			1.000
+	2 (6.1)	5 (10.9)		2 (11.8)	2 (8.3)	
–	31 (93.9)	41 (89.1)		15 (88.2)	22 (91.7)	
History of biopsy, n (%)			0.171			1.000
+	2 (6.1)	0 (0.0)		1 (89.7)	1 (4.2)	
–	31 (93.9)	46 (1.0)		16 (10.3)	23 (95.8)	
Menstruation status, n (%)			0.001			0.002
+	6 (76.8)	20 (49.3)		1 (89.7)	13 (30.3)	
–	27 (23.2)	26 (50.7)		16 (10.3)	11 (69.7)	
BI-RADS (DM plus DBT), n (%)			<0.001			<0.001
0, 1, 2, 3	8 (24.2)	0 (0.0)		6 (35.3)	1 (4.2)	
4A, 4B, 4C	24 (72.7)	32 (69.6)		11 (64.7)	15 (62.5)	
5, 6	1 (3.0)	14 (30.4)		0 (0.0)	8 (33.3)	
BI-RADS (MRI), n (%)						<0.001
1, 2, 3	18 (54.5)	0 (0.0)		8 (47.1)	0 (0.0)	
4, 5	15 (45.5)	46 (100.0)		9 (52.9)	24 (100.0)	

BI-RADS, breast imaging reporting and data system; DM, digital mammography; DBT, digital breast tomosynthesis; MRI, magnetic resonance imaging.

TABLE 2 | Diagnostic performance of each modality used alone and in combination.

	Cohort	AUC(95%CI)	ACC (95%CI)	SEN	SPE	PPV	NPV
DM alone	Training Cohort	0.727 (0.612–0.842)	0.696 (0.583–0.795)	0.739	0.636	0.739	0.636
	Validation Cohort	0.694 (0.524–0.863)	0.707 (0.545–0.839)	0.750	0.647	0.750	0.647
DBT alone	Training Cohort	0.850 (0.766–0.940)	0.798 (0.692–0.880)	0.804	0.788	0.841	0.743
	Validation Cohort	0.830 (0.698–0.968)	0.781 (0.624–0.894)	0.708	0.882	0.895	0.682
DWI MRI	Training Cohort	0.858 (0.775–0.942)	0.810 (0.706–0.890)	0.913	0.667	0.793	0.846
	Validation Cohort	0.831 (0.696–0.966)	0.781 (0.624–0.894)	0.750	0.824	0.857	0.700
DCE MRI	Training Cohort	0.879 (0.978–0.960)	0.861 (0.765–0.928)	0.957	0.727	0.830	0.923
	Validation Cohort	0.855 (0.727–0.984)	0.829 (0.674–0.929)	0.833	0.824	0.870	0.778
DM plus DBT	Training Cohort	0.909 (0.842–0.976)	0.861 (0765–0.928)	0.826	0.909	0.927	0.790
	Validation Cohort	0.880 (0.779–0.981)	0.805 (0.651–0.912)	0.708	0.941	0.944	0.700
DWI plus DCE	Training Cohort	0.930 (0.877–0.982)	0.873 (0.780–0.938)	0.891	0.849	0.891	0.849
	Validation Cohort	0.885 (0.768–1.000)	0.878 (0.738–0.959)	0.875	0.882	0.913	0.833

DM, digital mammography; DBT, digital breast tomosynthesis; DWI, diffusion-weighted imaging; DCE, dynamic contrast enhanced; AUC, area under the ROC curve; CI, confidence interval; Acc, accuracy; Sen, sensitivity; Spe, specificity; PPV, positive predictive value; NPV, negative predictive value.

the highest AUCs and sensitivities among the four modalities, but had relatively low specificities. The diagnostic performance of DWI plus DCE was significantly higher than DM plus DBT in terms of sensitivity. Besides, the DWI plus DCE yielded the highest positive predictive values (PPV) and the lowest misdiagnosis rates.

Development of the Combined Radiomics Signature and Nomogram

Radiomics features selected from the four modalities were combined and further selected to generate a combined feature set consisting of seven features, three from DBT, two from DCE, and two from DWI. Diagnostic performance of each feature was evaluated and is listed in **Table 3**. The combined radiomics signature (combined Rad score, **Supplementary S3**.) integrating the seven features and their corresponding LASSO coefficients was built and shown as follows:

$$\begin{aligned}\text{Combined Rad score} &= 0.5665 - \text{Wavelet_HHL_glszm_ZonePercentage} \times 2.7374 \\ &+ \text{Wavelet_LHL_firstorder_Skewness} \times 1.4977 \\ &+ \text{Log_sigma_3_0_mm_3D_glrlm_ShortRunLowGrayLevelEmphasis} \\ &\times 2.1381 + \text{Wavelet_HHLglcm_Imcl} \times 1.8133 \\ &+ \text{Original_glcm_ClusterShade} \times 1.4596 \\ &+ \text{Logarithm_glcm_InverseVariance} \times 1.6268 \\ &- \text{Exponential_glcm_MCC} \times 0.7365.\end{aligned}$$

A radiomics nomogram was constructed integrating the combined Rad score with the age and menstruation status (**Figure 2A**). The risk of being a malignant lesion can be read off the scale in the last row by vertically drawing a line from the total points. Calibration curves are shown in **Figures 2B, C**,

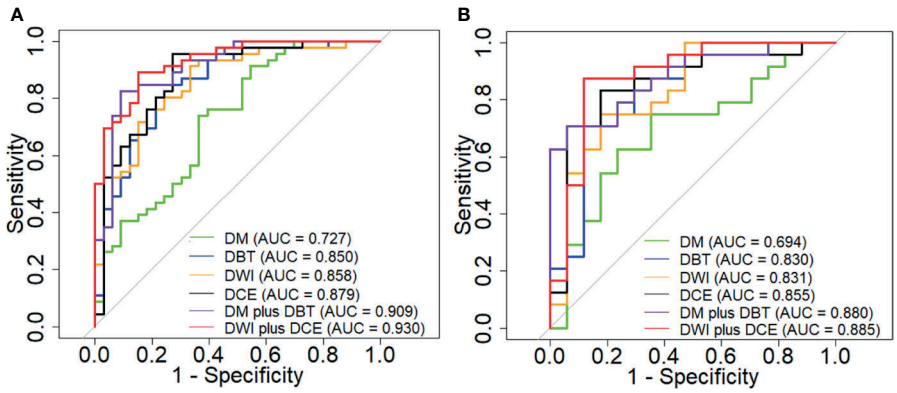


FIGURE 1 | ROC curves of the DM, DBT, DCE MRI and DWI MRI used individually and in comibination in the training **(A)** and validation **(B)** cohort.

TABLE 3 | Diagnostic performance of the selected features for the diagnosis of breast lesions.

Feature	Dataset	Mean ± SD		*P*-value	AUC
		Benign	Malignant		
Wavelet_HHL_glszm_ZonePercentage	Training Cohort	0.006 ± 0.006	0.002 ± 0.002	<0.001	0.772
	Validation Cohort	0.006 ± 0.006	0.002 ± 0.002	0.021	0.716
Wavelet_LHL_firstorder_Skewness	Training Cohort	0.076 ± 0.297	-0.147 ± 0.215	<0.001	0.737
	Validation Cohort	0.040 ± 0.377	-0.176 ± 0.168	0.010	0.740
Log_sigma_3_0_mm_3D_glrlm_ShortRunLowGrayLevelEmphasis	Training Cohort	0.057 ± 0.028	0.037 ± 0.020	<0.001	0.736
	Validation Cohort	0.062 ± 0.062	0.035 ± 0.015	0.181	0.625
Wavelet_HHLglcm_Imcl	Training Cohort	-0.099 ± 0.049	-0.072 ± 0.033	<0.001	0.738
	Validation Cohort	-0.085 ± 0.042	-0.067 ± 0.015	0.181	0.625
Original_glcm_Clus-terShade	Training Cohort	-2,413.833 ± 11,596.710	3,361.392 ± 14,159.810	0.026	0.648
	Validation Cohort	-2,950.967 ± 10,227.370	5,047.669 ± 1,264.26	0.013	0.730
Logarithm_glcm_InverseVariance	Training Cohort	0.161 ± 0.026	0.146 ± 0.022	<0.001	0.667
	Validation Cohort	0.152 ± 0.022	0.151 ± 0.020	<0.001	0.507
Exponential_glcm_MCC	Training Cohort	0.583 ± 0.305	0.776 ± 0.158	0.002	0.710
	Validation Cohort	0.539 ± 0.269	0.761 ± 0.155	0.003	0.772

Glszm, gray level size zone matrix; glrlm, gray level run length matrix; glcm, gray level co-occurrence matrix; SD, standard deviation; AUC, area under the ROC curve.

indicating acceptable agreements between the nomogram-estimated probabilities and actual outcomes of the lesions. The 45-degree blue line and the red dotted line represent an ideal diagnosis and the performance of our nomogram, respectively. As the red dotted line is closer to the blue line represents a better diagnostic performance. **Figures 2D, E** show that the nomogram exhibited better diagnostic capabilities compared with the combined Rad score or the clinical model alone (AUCs in the training cohort, nomogram *vs.* combined Rad score *vs.* clinical model, 0.975 *vs.* 0.964 *vs.* 0.782; AUCs in the validation cohort, nomogram *vs.* combined Rad score *vs.* clinical model, 0.978 *vs.* 0.983 *vs.* 0.690). The diagnostic performance of the combined Rad score, clinical model and nomogram are shown in **Table 4**.

Figure 3 shows the results of the decision curve analysis for each model. The nomogram exhibited a greater net benefit compared with the combined Rad score or the clinical model. When the threshold probability of the patient was between 0.44 and 0.68, or over 0.78, a greater benefit can be obtained by using the nomogram, indicating a good potential in clinical applications.

DISCUSSION

Prior to this study, there have been researches evaluating the diagnostic capabilities of DM (32, 38, 39), DBT (40, 41), MRI (42–44) separately for detecting breast cancer, all based on subjective visual examinations, and lack of direct and quantitative comparisons of different modalities. On the contrary, this study performed comprehensive radiomics analyses to quantitatively assess the diagnostic performance of different modalities separately and in combination. We found that the radiomics signature derived from DM always showed the worst diagnostic performance in terms of AUC, sensitivity, and specificity compared with the other individual modalities. This may be explainable since the DM only obtains one image, which may lead to overlapping glands, and, hence, is not sufficient to analyze the distribution of dense and adipose tissues (52). The result was in accordance with previous studies that also showed the DM-based diagnosis often leads to high false negative and false positive rates due to the fact that the lesions may be

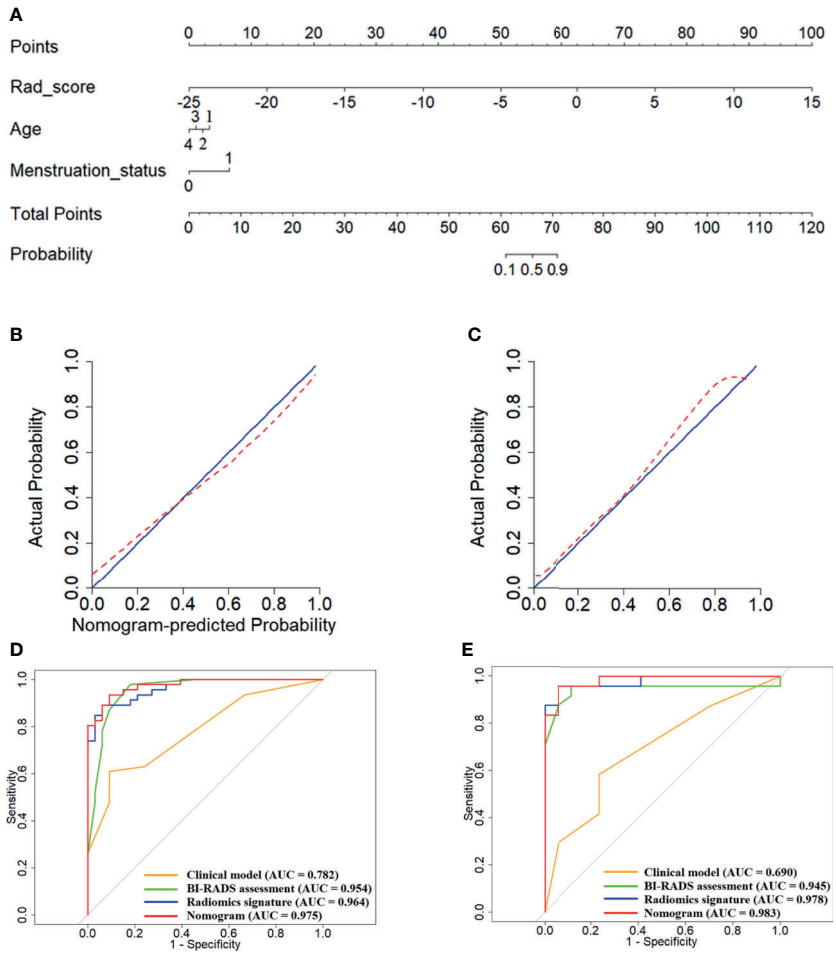


FIGURE 2 | Development and validation of the nomogram model integrating the combined Rad score, age and menstruation status. **(A)** Construction of the nomogram; **(B, C)**, Calibration curves of the nomogram in the training **(B)** and validation **(C)** cohort; **(D, E)**, ROC curves of the nomogram, combined Rad score and clinical model in the training **(D)** and validation **(E)** cohort.

obscured or hidden by the overlapping fibroglandular tissues (5, 53). The addition of DBT to DM can significantly improve the diagnostic AUC, accuracy, specificity, PPV, and NPV, and generate a similar sensitivity compared with the DM alone. This was in line with some previous reports that also indicated that breast DBT can lead to improvements in AUC and specificity by visual assessments (54, 55). This may be because the DBT can improve the lesion visibility by providing thin section tomographic images and reducing the overlap of breast tissues, and, hence, represents a clearer edge, shape, and structure of the lesion. The addition of DBT to DM did not improve the diagnostic sensitivity by visual assessments compared with DM alone as reported in an earlier study (14). The discordance may be because they performed the research with a cancer-only population. The DCE plus DWI yielded higher AUCs and sensitivities, but lower specificities than the DM plus DBT. The result was partially in line with a previous literature that also indicated that the MRI was superior to the X-ray technology in the diagnostic AUC and sensitivity, but weaker in the specificity (14, 56).

The DBT showed a similar diagnostic AUC, slightly increased specificity, and lower sensitivity compared with DCE or DWI, which was in line with a previous research that also demonstrated the inferiority of breast DBT in the sensitivity compared with MRI by visual examinations (14, 53, 57). This may be explained since the DCE can reflect the neoangiogenesis within the tumor that is associated with the growth and progression of the malignant tumor (58). While, the DWI can represent tissue microenvironments and membrane integrities through depicting the diffusivity of the tissues (59). Therefore, the MRI tends to be more sensitive than DBT or DM on tumors with higher malignant degrees. The DCE yielded higher AUC, accuracy, sensitivity, and specificity compared with DWI, which may be due to the higher resolution and the use of a contrast agent in DCE (44). We found that the addition of DBT to MRI (DBT plus DCE plus DWI) can increase the AUC and sensitivity

TABLE 4 | Comparison of the clinical model, BI-RADS assessment, combined Rad score, and nomogram.

	Training cohort								Validation cohort						
	AUC (95% CI)	Acc (95% CI)	Sen	Spe	PPV	NPV	*P*	AUC (95% CI)	Acc (95% CI)	Sen	Spe	PPV	NPV	*P*
M1	0.782 (0.687–0.877)	0.734 (0.623–0.827)	0.609	0.909	0.903	0.625		0.690 (0.531–0.849)	0.659 (0.494–0.799)	0.583	0.765	0.778	0.565	
M2	0.954 (0.908–1.000)	0.912 (0.826–0.964)	0.978	0.818	0.883	0.964		0.945 (0.861–1.000)	0.927 (0.801–0.985)	0.958	0.882	0.920	0.938	
M3	0.964 (0.931–0.997)	0.911 (0.826–0.964)	0.935	0.909	0.935	0.909		0.978 (0.941–1.000)	0.951 (0.835–0.994)	0.958	0.941	0.958	0.941	
M4	0.975 (0.948–1.000)	0.924 (0.842–0.972)	0.913	0.849	0.894	0.875		0.983 (0.955–1.000)	0.951 (0.835–0.994)	0.958	0.941	0.958	0.941	
M1 *vs.* M2							<0.001							<0.001
M2 *vs.* M3							0.741							0.404
M3 *vs.* M4							0.178							0.596
M1 *vs.* M4							0.001							<0.001

M1, Clinical model; M2, BI-RADS assessment; M3, Combined Rad score; M4, Nomogram model; AUC, area under the ROC curve; CI, confidence interval; Acc, accuracy; Sen, sensitivity; Spe, specificity; PPV, positive predictive value; NPV, negative predictive value.

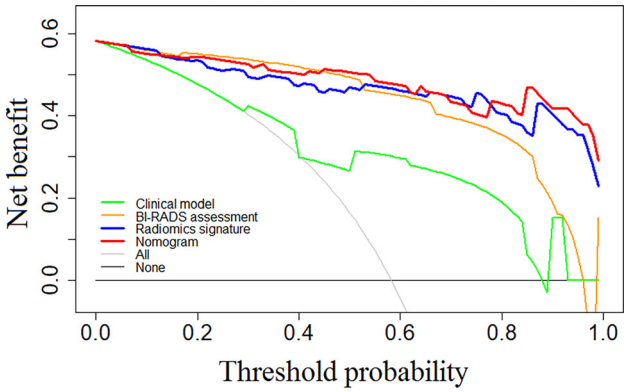


FIGURE 3 | Showed results of the decision curve analysis for each model. The nomogram exhibited greater net benefit compared with the combined Rad score or the clinical model. When the threshold probability of the patient was between 0.44 and 0.68, or over 0.78, greater benefit can be obtained by using the nomogram, indicating good potential in clinical applications.

compared with MRI alone (DCE plus DWI). This indicated that the DBT and MRI are complementary, their combination can significantly improve the predictive capabilities. While, our results were inconsistent with a previous report that showed no improvement in the diagnostic sensitivity by combing DM, DBT, DCE, and ultrasound (60). Since they involved ultrasound, direct comparisons between our study and their work was impossible.

In the clinical practice, although integrating MRI with X-rays allows the radiologists to give judgments more easily, the diagnosis still relies on subjective experiences. We selected a total of seven quantitative features as the most important predictors, three from DBT, two from DCE, and two from DWI. There were one original and six transformed features. The developed combined Rad score integrating these features significantly improved the diagnostic performance compared with any modality alone. The Original_glcm_ClusterShade feature measures the skewness and the uniformity of the gray level co-occurrence matrix within the tumor. A higher value of this feature implies a greater asymmetry about the mean and a greater heterogeneity of the lesion. We found that this feature was bigger in the malignant lesions than in the benign lesions, which suggests that a tumor with more asymmetry and complexity in the tumor texture tends to be malignant. Among the six transformed features, one belonged to the first-order and five belonged to the textural feature class. The first-order feature describes the distribution of voxel intensities in the image region. While, the textural feature quantifies the complexity of a tumor and the thickness of the texture. Our findings suggest that the tumor heterogeneity may be closely related to breast cancer, since textural features in the medical image often reflect tumor heterogeneities. The results were partially in line with previous studies that also highlighted the correlations between the textural features and breast cancer (61, 62). Our findings may explain that the proposed combined Rad score can significantly improve the diagnostic performance with regard to AUC and sensitivity than

visual assessments, since most of the identified features (6 of 7) were derived from the transformed images that were generated by filtering the original images with various filters, and, thus, can hardly be understood by human.

A clinical model was built integrating age and menstruation status, and showed a lower AUC, sensitivity, and specificity than the combined Rad score. The nomogram incorporating the combined Rad score with the age and menstruation status achieved the best overall diagnostic performance compared with the combined Rad score, clinical model, and BI-RADS assessment. Decision curves demonstrated a better clinical usefulness of the nomogram with more net benefits across the majority of the range of threshold probabilities. Therefore, we suggest that our nomogram may be considered as an effective tool that can assist in decision making for the diagnosis of breast cancer. To use our nomogram, radiologists need to manually segment lesions on the DBT and MRI images for each patient, then calculate the probability of being benign or malignant. After that, clinicians can incorporate the nomogram-predicted probabilities with other clinical information to give a comprehensive decision on further examinations and treatments.

This study has limitations. First, this retrospective study had a relatively small sample size, which may cause inherent bias. Second, all data were obtained from a single hospital. Further multi-center trials are warranted to confirm the present findings. Third, our radiomic methods rely on manual segmentations of the ROIs, which were subjective and time-consuming. Future studies are needed to explore deep learning-based automatic segmentation methods on breast data.

CONCLUSIONS

Our results showed that the DBT performed similar to DCE and DWI in terms of AUC and sensitivity, but better in specificity for detecting malignant lesions. The DBT plus DM can provide a lower AUC and sensitivity, but a higher specificity compared with DCE plus DWI. The proposed nomogram achieved the best diagnostic performance, and may help clinicians make precise decisions regarding treatments.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

All analyses of human data conducted in this study were reviewed and approved by the Institutional Review Board of the Cancer Hospital of China Medical University and in accordance with the ethical standards of the institutional and/or national research committee. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

XJ and SN contributed to the study concepts and manuscript preparation. SN and XW contributed to the study design. NZ and GL contributed to the data acquisition. SN and NZ contributed to the quality control of the data and algorithms. XJ and XW contributed to the data analysis and interpretation. YL and E-NC contributed to the statistical analysis. XJ, YD and YK contributed to the manuscript review. All authors contributed to the article and approved the submitted version.

FUNDING

The study was funded by the Key Program of Ministry of Science and Technology of China (2017YFC1309100), Climbing Fund of National Cancer Center (NCC201806B011), Shenyang Municipal Science and Technology Project (F16-206-9-23), National Natural Science Foundation of China (81872363), Support program of youth science and technology innovation talents of Shenyang City (RC180269), Major Technology Plan Project of Shenyang (17-230-9-07), Supporting Fund for Big data in Health Care (HMB201903101), Special foundation for the central government guides the development of local science and technology of Liaoning Province (2018416029), China National Natural Science Foundation (31770147), and Medical-Engineering Joint Fund for Cancer Hospital of China Medical University and Dalian University of technology (LD202029).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.725922/full#supplementary-material

REFERENCES

1. McGuire A, Brown JA, Malone C, McLaughlin R, Kerin MJ. Effects of Age on the Detection and Management of Breast Cancer. *Cancers (Basel)* (2015) 7:908–29. doi: 10.3390/cancers7020815
2. Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid Review: Radiomics and Breast Cancer. *Breast Cancer Res Treat* (2018) 169:217–29. doi: 10.1007/s10549-018-4675-4
3. Xie WY, Li YS, Ma YD. Breast Mass Classification in Digital Mammography Based on Extreme Learning Machine. *Neurocomputing* (2016) 173:930–41. doi: 10.1016/j.neucom.2015.08.048
4. Michell MJ. Breast Screening Review–A Radiologist's Perspective. *Br J Radiol* (2012) 85:845–7. doi: 10.1259/bjr/21332901
5. Fischer U, Baum F, Obenauer S, Luftner-Nagel S, von Heyden D, Vosshenrich R, et al. Comparative Study in Patients With Microcalcifications: Full-Field Digital Mammography vs Screen-Film Mammography. *Eur Radiol* (2002) 12:2679–83. doi: 10.1007/s00330-002-1354-x
6. Niklason LT, Kopans DB, Hamberg LM. Digital Breast Imaging: Tomosynthesis and Digital Subtraction Mammography. *Breast Dis* (1998) 10:151–64. doi: 10.3233/BD-1998-103-415
7. Kopans DB. Digital Breast Tomosynthesis from Concept to Clinical Care. *AJR Am J Roentgenol* (2014) 202:299–308. doi: 10.2214/AJR.13.11520

8. Gilbert FJ, Tucker L, Young KC. Digital Breast Tomosynthesis (DBT): A Review of the Evidence for Use as a Screening Tool. *Clin Radiol* (2016) 71:141–50. doi: 10.1016/j.crad.2015.11.008
9. Orel SG, Schnall MD. MR Imaging of the Breast for the Detection, Diagnosis, and Staging of Breast Cancer. *Radiology* (2001) 220:13–30. doi: 10.1148/radiology.220.1.r01jl3113
10. Fischer U, Kopka L, Grabbe E. Breast Carcinoma: Effect of Preoperative Contrast-Enhanced MR Imaging on the Therapeutic Approach. *Radiology* (1999) 213:881–8. doi: 10.1148/radiology.213.3.r99dc01881
11. Friedewald SM, Rafferty EA, Rose SL, Durand MA, Plecha DM, Greenberg JS, et al. Breast Cancer Screening Using Tomosynthesis in Combination With Digital Mammography. *JAMA* (2014) 311:2499–507. doi: 10.1001/jama.2014.6095
12. Lee CI, Cevik M, Alagoz O, Sprague BL, Tosteson AN, Miglioretti DL, et al. Comparative Effectiveness of Combined Digital Mammography and Tomosynthesis Screening for Women With Dense Breasts. *Radiology* (2015) 274:772–80. doi: 10.1148/radiol.14141237
13. Tang W, Hu FX, Zhu H, Wang QF, Gu YJ, Peng WJ. Digital Breast Tomosynthesis Plus Mammography, Magnetic Resonance Imaging Plus Mammography and Mammography Alone: A Comparison of Diagnostic Performance in Symptomatic Women. *Clin Hemorheol Microcirc* (2017) 66:105–16. doi: 10.3233/CH-16242
14. Kim WH, Chang JM, Moon HG, Yi A, Koo HR, Gweon HM, et al. Comparison of the Diagnostic Performance of Digital Breast Tomosynthesis and Magnetic Resonance Imaging Added to Digital Mammography in Women With Known Breast Cancers. *Eur Radiol* (2016) 26:1556–64. doi: 10.1007/s00330-015-3998-3
15. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006
16. Brem RF. Clinical *Versus* Research Approach to Breast Cancer Detection With CAD: Where Are We Now? *AJR Am J Roentgenol* (2007) 188:234–5. doi: 10.2214/AJR.06.1449
17. Zhou M, Scott J, Chaudhury B. Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches. *AJNR Am J Neuroradiol* (2017) 39:208–16. doi: 10.3174/ajnr.A5391
18. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning Methods for Quantitative Radiomic Biomarkers. *Sci Rep* (2015) 5:13087. doi: 10.1038/srep13087
19. Chai R, Ma H, Xu M, Arefan D, Cui X, Liu Y, et al. Differentiating Axillary Lymph Node Metastasis in Invasive Breast Cancer Patients: A Comparison of Radiomic Signatures From Multiparametric Breast MR Sequences. *J Magn Reson Imaging* (2019) 50:1125–32. doi: 10.1002/jmri.26701
20. Cui X, Wang N, Zhao Y, Chen S, Li S, Xu M, et al. Preoperative Prediction of Axillary Lymph Node Metastasis in Breast Cancer Using Radiomics Features of DCE-MRI. *Sci Rep* (2019) 9:2240. doi: 10.1038/s41598-019-38502-0
21. Dong Y, Feng Q, Yang W, Lu Z, Deng C, Zhang L, et al. Preoperative Prediction of Sentinel Lymph Node Metastasis in Breast Cancer Based on Radiomics of T2-Weighted Fat-Suppression and Diffusion-Weighted MRI. *Eur Radiol* (2018) 28:582–91. doi: 10.1007/s00330-017-5005-7
22. Han L, Zhu Y, Liu Z, Yu T, He C, Jiang W, et al. Radiomic Nomogram for Prediction of Axillary Lymph Node Metastasis in Breast Cancer. *Eur Radiol* (2019) 29:3820–9. doi: 10.1007/s00330-018-5981-2
23. Liu C, Ding J, Spuhler K, Gao Y, Serrano Sosa M, Moriarty M, et al. Preoperative Prediction of Sentinel Lymph Node Metastasis in Breast Cancer by Radiomic Signatures From Dynamic Contrast-Enhanced MRI. *J Magn Reson Imaging* (2019) 49:131–40. doi: 10.1002/jmri.26224
24. Xie T, Zhao Q, Fu C, Bai Q, Zhou X, Li L, et al. Differentiation of Triple-Negative Breast Cancer From Other Subtypes Through Whole-Tumor Histogram Analysis on Multiparametric MR Imaging. *Eur Radiol* (2019) 29:2535–44. doi: 10.1007/s00330-018-5804-5
25. Holli-Helenius K, Salminen A, Rinta-Kiikka I, Koskivuo I, Brück N, Boström P, et al. MRI Texture Analysis in Differentiating Luminal A and Luminal B Breast Cancer Molecular Subtypes - A Feasibility Study. *BMC Med Imag* (2017) 17:69. doi: 10.1186/s12880-017-0239-z
26. Fan M, Li H, Wang S, Zheng B, Zhang J, Li L. Radiomic Analysis Reveals DCE-MRI Features for Prediction of Molecular Subtypes of Breast Cancer. *PloS One* (2017) 12:e0171683. doi: 10.1371/journal.pone.0171683
27. Fan M, Zhang P, Wang Y, Peng W, Wang S, Gao X, et al. Radiomic Analysis of Imaging Heterogeneity in Tumours and the Surrounding Parenchyma Based on Unsupervised Decomposition of DCE-MRI for Predicting Molecular Subtypes of Breast Cancer. *Eur Radiol* (2019) 29:4456–67. doi: 10.1007/s00330-018-5891-3
28. Grimm LJ, Zhang J, Mazurowski MA. Computational Approach to Radiogenomics of Breast Cancer: Luminal A and Luminal B Molecular Subtypes Are Associated With Imaging Features on Routine Breast MRI Extracted Using Computer Vision Algorithms. *J Magn Reson Imaging* (2015) 42:902–7. doi: 10.1002/jmri.24879
29. Huang Y, Cheng Z, Huang X, Liang C, Liang C, Liu Z. Preoperative Evaluation of Histologic Grade in Invasive Breast Cancer With T2W-MRI Based Radiomics Signature. *Zhong Nan Da Xue Xue Bao Yi Xue Ban* (2019) 44:285–9. doi: 10.11817/j.issn.1672-7347.2019.03.009
30. Fan M, Yuan W, Zhao W, Xu M, Wang S, Gao X, et al. Joint Prediction of Breast Cancer Histological Grade and Ki-67 Expression Level Based on DCE-MRI and DWI Radiomics. *IEEE J Biomed Health Inf* (2019) PP(99):1–1. doi: 10.1109/JBHI.2019.2956351
31. Yuan G, Liu Y, Huang W, Hu B. Differentiating Grade in Breast Invasive Ductal Carcinoma Using Texture Analysis of MRI. *Comput Math Methods Med* (2020) 2020:1–14. doi: 10.1155/2020/6913418
32. Ma W, Zhao Y, Ji Y, Guo X, Jian X, Liu P, et al. Breast Cancer Molecular Subtype Prediction by Mammographic Radiomic Features. *Acad Radiol* (2019) 26:196–201. doi: 10.1016/j.acra.2018.01.023
33. Xiong Q, Zhou X, Liu Z, Lei C, Yang C, Yang M, et al. Multiparametric MRI-Based Radiomics Analysis for Prediction of Breast Cancers Insensitive to Neoadjuvant Chemotherapy. *Clin Transl Oncol* (2020) 22:50–9. doi: 10.1007/s12094-019-02109-8
34. Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, et al. Radiomics of Multi-Parametric MRI for Pretreatment Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. *Clin Cancer Res* (2019) 25:3538–47. doi: 10.1158/1078-0432.CCR-18-3190
35. Braman NM, Etesami M, Prasanna P, Dubchuk C, Gilmore H, Tiwari P, et al. Intratumoral and Peritumoral Radiomics for the Pretreatment Prediction of Pathological Complete Response to Neoadjuvant Chemotherapy Based on Breast DCE-MRI. *Breast Cancer Res* (2017) 19:57. doi: 10.1186/s13058-017-0846-1
36. Cain EH, Saha A, Harowicz MR, Marks JR, Marcom PK, Mazurowski MA. Multivariate Machine Learning Models for Prediction of Pathologic Response to Neoadjuvant Therapy in Breast Cancer Using MRI Features: A Study Using an Independent Validation Set. *Breast Cancer Res Treat* (2019) 173:455–63. doi: 10.1007/s10549-018-4990-9
37. Chamming's F, Ueno Y, Ferré R, Kao E, Jannot AS, Chong J, et al. Features From Computerized Texture Analysis of Breast Cancers at Pretreatment MR Imaging Are Associated With Response to Neoadjuvant Chemotherapy. *Radiology* (2018) 286:412–20. doi: 10.1148/radiol.2017170143
38. Mavroforakis M, Georgiou H, Dimitropoulos N, Cavouras D, Theodoridis S. Significance Analysis of Qualitative Mammographic Features, Using Linear Classifiers, Neural Networks and Support Vector Machines. *Eur J Radiol* (2005) 54:80–9. doi: 10.1016/j.ejrad.2004.12.015
39. Verma B, Mcleod P, Klevansky A. Classification of Benign and Malignant Patterns in Digital Mammograms for the Diagnosis of Breast Cancer. *Expert Syst Appl* (2010) 37:3344–51. doi: 10.1016/j.eswa.2009.10.016
40. Tagliafico AS, Valdora F, Mariscotti G, Durando M, Nori J, La Forgia D, et al. An Exploratory Radiomics Analysis on Digital Breast Tomosynthesis in Women With Mammographically Negative Dense Breasts. *Breast* (2018) 40:92–6. doi: 10.1016/j.breast.2018.04.016
41. Sakai A, Onishi Y, Matsui M, Adachi H, Teramoto A, Saito K, et al. A Method for the Automated Classification of Benign and Malignant Masses on Digital Breast Tomosynthesis Images Using Machine Learning and Radiomic Features. *Radiol Phys Technol* (2020) 13:27–36. doi: 10.1007/s12194-019-00543-5
42. Bickelhaupt S, Paech D, Kickingereder P, Steudle F, Lederer W, Daniel H, et al. Prediction of Malignancy by a Radiomic Signature From Contrast Agent-Free Diffusion MRI in Suspicious Breast Lesions Found on Screening

Mammography. *J Magn Reson Imaging* (2017) 46:604–16. doi: 10.1002/jmri.25606
43. Granzier RWY, Verbakel NMH, Ibrahim A, van Timmeren JE, van Nijnatten TJA, Leijenaar RTH, et al. MRI-Based Radiomics in Breast Cancer: Feature Robustness With Respect to Inter-Observer Segmentation Variability. *Sci Rep* (2020) 10:14163. doi: 10.1038/s41598-020-70940-z
44. Hu Q, Whitney HM, Giger ML. Radiomics Methodology for Breast Cancer Diagnosis Using Multiparametric Magnetic Resonance Imaging. *J Med Imaging (Bellingham)* (2020) 7:044502. doi: 10.1117/1.JMI.7.4.044502
45. Zhou H, Dong D, Chen B, Fang M, Cheng Y, Gan Y, et al. Diagnosis of Distant Metastasis of Lung Cancer: Based on Clinical and Radiomic Features. *Transl Oncol* (2017) 11:31–6. doi: 10.1016/j.tranon.2017.10.010
46. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
47. Leijenaar RT, Carvalho S, Velazquez ER, van Elmpt WJ, Parmar C, Hoekstra OS, et al. Stability of FDG-PET Radiomics Features: An Integrated Analysis of Test-Retest and Inter-Observer Variability. *Acta Oncol* (2013) 52:1391–7. doi: 10.3109/0284186X.2013.812798
48. Sauerbrei W, Royston P, Binder H. Selection of Important Variables and Determination of Functional Form for Continuous Predictors in Multivariable Model Building. *Stat Med* (2007) 26:5512–28. doi: 10.1002/sim.3148
49. Tibshirani R. Regression Shrinkage and Selection via the Lasso. *J R Stat Soc Ser B (Methodol)* (1996) 73:273–82. doi: 10.1111/j.2517-6161.1996.tb02080.x
50. Ruopp MD, Perkins NJ, Whitcomb BW, Schisterman EF. Youden Index and Optimal Cut-Point Estimated From Observations Affected by a Lower Limit of Detection. *Biom J* (2008) 50:419–30. doi: 10.1002/bimj.200710415
51. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med Decis Making* (2006) 26:565–74. doi: 10.1177/0272989X06295361
52. Kerlikowske K, Hubbard RA, Miglioretti DL, Geller BM, Yankaskas BC, Lehman CD, et al. Comparative Effectiveness of Digital *Versus* Film-Screen Mammography in Community Practice in the United States: A Cohort Study. *Ann Intern Med* (2011) 155:493–502. doi: 10.7326/0003-4819-155-8-201110180-00005
53. Seely JM, Alhassan T. Screening for Breast Cancer in 2018-What Should We be Doing Today? *Curr Oncol* (2018) 25:S115–24. doi: 10.3747/co.25.3770
54. Haas BM, Kalra V, Geisel J, Raghu M, Durand M, Philpotts LE. Comparison of Tomosynthesis Plus Digital Mammography and Digital Mammography Alone for Breast Cancer Screening. *Radiology* (2013) 269:694–700. doi: 10.1148/radiol.13130307
55. Rafferty EA, Park JM, Philpotts LE, Poplack SP, Sumkin JH, Halpern EF, et al. Assessing Radiologist Performance Using Combined Digital Mammography and Breast Tomosynthesis Compared With Digital Mammography Alone: Results of a Multicenter, Multireader Trial. *Radiology* (2013) 266:104–13. doi: 10.1148/radiol.12120674
56. Girometti R, Tomkova L, Cereser L, Zuiani C. Breast Cancer Staging: Combined Digital Breast Tomosynthesis and Automated Breast Ultrasound *Versus* Magnetic Resonance Imaging. *Eur J Radiol* (2018) 107:188–95. doi: 10.1016/j.ejrad.2018.09.002
57. Berg WA, Zhang Z, Lehrer D, Jong RA, Pisano ED, Barr RG, et al. Detection of Breast Cancer With Addition of Annual Screening Ultrasound or a Single Screening MRI to Mammography in Women With Elevated Breast Cancer Risk. *JAMA* (2012) 307:1394–404. doi: 10.1001/jama.2012.388
58. Clauser P, Carbonaro LA, Pancot M, Girometti R, Bazzocchi M, Zuiani C, et al. Additional Findings at Preoperative Breast MRI: The Value of Second-Look Digital Breast Tomosynthesis. *Eur Radiol* (2015) 25:2830–9. doi: 10.1007/s00330-015-3720-5
59. Kuhl CK, Mielcareck P, Klaschik S, Leutner C, Wardelmann E, Gieseke J, et al. Dynamic Breast MR Imaging: Are Signal Intensity Time Course Data Useful for Differential Diagnosis of Enhancing Lesions? *Radiology* (1999) 211:101–10. doi: 10.1148/radiology.211.1.r99ap38101
60. Mariscotti G, Houssami N, Durando M, Bergamasco L, Campanino PP, Ruggieri C, et al. Accuracy of Mammography, Digital Breast Tomosynthesis, Ultrasound and MR Imaging in Preoperative Assessment of Breast Cancer. *Anticancer Res* (2014) 34:1219–25. doi: 10.1007/s13277-013-1375-x
61. Kontos D, Ikejimba LC, Bakic PR, Troxel AB, Conant EF, Maidment AD. Analysis of Parenchymal Texture With Digital Breast Tomosynthesis: Comparison With Digital Mammography and Implications for Cancer Risk Assessment. *Radiology* (2011) 261:80–91. doi: 10.1148/radiol.11100966
62. Kontos D, Bakic PR, Carton AK, Troxel AB, Conant EF, Maidment AD. Parenchymal Texture Analysis in Digital Breast Tomosynthesis for Breast Cancer Risk Estimation: A Preliminary Study. *Acad Radiol* (2009) 16:283–98. doi: 10.1016/j.acra.2008.08.014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Niu, Wang, Zhao, Liu, Kan, Dong, Cui, Luo, Yu and Jiang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 14 September 2021
doi: 10.3389/fonc.2021.689136

CT Radiomics for the Preoperative Prediction of Ki67 Index in Gastrointestinal Stromal Tumors: A Multi-Center Study

Yilei Zhao[1], Meibao Feng[1], Minhong Wang[2], Liang Zhang[3], Meirong Li[1] and Chencui Huang[4]*

[1] First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, China, [2] First Affiliated Hospital of Wannan Medical College, Wuhu, China, [3] Zhejiang Cancer Hospital, University of Chinese Academy of Sciences, Hangzhou, China, [4] Beijing Deepwise & League of PHD Technology Co., Ltd, Beijing, China

OPEN ACCESS

Edited by:
Xujiong Ye, University of Lincoln, United Kingdom

Reviewed by:
Sikandar Shaikh, Shadan Hospital and Institute of Medical Sciences, India
Changqiang Wu, North Sichuan Medical College, China
Haowen Pang, The Affiliated Hospital of Southwest Medical University, China

***Correspondence:**
Yilei Zhao
1309035@zju.edu.cn

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *31 March 2021*
Accepted: *30 June 2021*
Published: *14 September 2021*

Citation:
Zhao Y, Feng M, Wang M, Zhang L, Li M and Huang C (2021) CT Radiomics for the Preoperative Prediction of Ki67 Index in Gastrointestinal Stromal Tumors: A Multi-Center Study. Front. Oncol. 11:689136. doi: 10.3389/fonc.2021.689136

Purpose: This study established and verified a radiomics model for the preoperative prediction of the Ki67 index of gastrointestinal stromal tumors (GISTs).

Materials and Methods: A total of 344 patients with GISTs from three hospitals were divided into a training set and an external validation set. The tumor region of interest was delineated based on enhanced computed-tomography (CT) images to extract radiomic features. The Boruta algorithm was used for dimensionality reduction of the features, and the random forest algorithm was used to construct the model for radiomics prediction of the Ki67 index. The receiver operating characteristic (ROC) curve was used to evaluate the model's performance and generalization ability.

Results: After dimensionality reduction, a feature subset having 21 radiomics features was generated. The generated radiomics model had an the area under curve (AUC) value of 0.835 (95% confidence interval(CI): 0.761–0.908) in the training set and 0.784 (95% CI: 0.691–0.874) in the external validation cohort.

Conclusion: The radiomics model of this study had the potential to predict the Ki67 index of GISTs preoperatively.

Keywords: gastrointestinal stromal tumors (GIST), ki67, radiomics, computed-tomography (CT), prognostic

INTRODUCTION

Gastrointestinal stromal tumors (GISTs) are the most common mesenchymal-derived tumors of the gastrointestinal tract, with complex biological behaviors. Recurrence and metastasis of GISTs are the main factors associated with patient survival. The pathogenesis of GISTs is mainly due to functional gene mutations of *KIT*, which lead to the continuous activation of the proto-oncogene receptor tyrosine kinase (1). The molecular targeted drug imatinib mesylate blocks tyrosine kinase-

activated pathways and is often used to control postoperative recurrence of high-risk GISTs and in neoadjuvant therapy (2, 3).

The GISTs risk stratification criteria issued by the National Institution of Health is the most commonly used GISTs biological behavior and clinical outcome evaluation criteria (4). The mitosis count of the specimen is a significant factor in these criteria. However, the identification of mitosis count depends on the pathologists' skills and experience, resulting in a problem of reproducibility of findings and affecting the risk stratification accuracy. Ki67 is a nuclear antigen related to proliferating cells and is a marker of cell division and proliferation activity. Its expression level correlates with the degree of tumor cell proliferation. Tumors with active proliferation are susceptible to invasion and distant metastasis (5, 6). The Ki67 index is evaluated by immunohistochemistry. In Ki67 immunohistochemistry, cell nuclei in proliferating cells are stained positively, while the nuclei in nonproliferating cells remain negative. Ki67 immunohistochemistry eliminates the bias in morphological observations and results in a consistent interpretation from different pathologists (7, 8). Second, Ki67 is expressed in all phases of the cell cycle (i.e., G1, S, G2, and M phases), whereas the mitosis is only present in the M phase. Ki67 index is more comprehensive and consistent compared with the mitosis count (9).

In the era of precision therapy for GISTs, the accurate prognosis stratification for patients suffering from GISTs is of great importance. Studies have found that patients with Ki67 index > 8% have a high risk of recurrence, which is an adverse factor for adjuvant imatinib therapy. Ki67 index can also play as an assistant in the stratified diagnosis and treatment among GISTs patients. The CT imaging characteristics, including size, contour, and tumor margin, are correlated with the Ki67 index in GISTs, suggesting that imaging features can be used to predict the expression of Ki67 in GISTs (9). Hence, in this study we constructed a prediction model of the Ki67 index (cutoff value of 8%) for patients with GISTs by extracting the texture and morphological features of tumors from enhanced CT images, and verified the model with independent external data to evaluate the generalization ability.

MATERIALS AND METHODS

Patients

This retrospective study was approved by the Institutional Review Board of the First Affiliated Hospital Zhejiang University School of Medicine (Zhejiang, China). The signed informed consent forms were waived. This study included patients with GISTs in three hospitals from January 1, 2016, to July 1, 2020. The inclusion criteria of this study were as follows (1) GIST was diagnosed pathologically; (2) patient had enhanced CT examination within 15 days before surgery; and (3) the Ki67 index was reported in the pathological results. The exclusion criteria were patients who received neoadjuvant treatment with imatinib or other tyrosine kinase inhibitors before CT examination.

The Ki67 index was evaluated by immunohistochemistry within 7 days after surgical. The patients were divided into two groups based on the cutoff value of 8% for the Ki67 index: high-Ki67 group (Ki67 index >8%) and low-Ki67 group (Ki67 index ≤8%).

According to the above criteria, 344 patients with GISTs were included in this study. Data from the First Affiliated Hospital Zhejiang University School of Medicine (Zhejiang, China) were placed in the training set, and the data from the other two hospitals were placed in the external validation set. Descriptive statistical analyses were performed on the clinical data of the training and validation sets, where a t-test was used for continuous variables and a chi-square test was used for categorical variables. All statistical analyses were performed using R software (version 3.4.1; http://www.rproject.org). $P < 0.05$ (two-sided) was considered statistically significant.

CT Technology

The examination equipment included the Brilliance 64-slice spiral CT and the Brilliance 256-slice spiral CT scanners (Philips Healthcare), and the SOMATOM Definition dual-source CT scanner (Siemens Healthineers). The CT scanning range was from the diaphragm dome to the lower margin of the pubic symphysis and the scanning direction was from the cranial to caudal. The patients were required to hold their breath during the CT. The scanning parameters of the Philips' CT scanners were 64 × 0.625 mm collimation, 250 mm field of view (FOV), 120 kV tube voltage, 250 mA tube current, 5 mm slice thickness, 5 mm slice interval, 0.4 s tube rotation time, 0.891 pitch, standard (B) algorithm, and 512 × 512 matrix. The scanning parameters of the Siemens' CT scanner were 128 × 0.6 mm collimation, 50 min FOV, 120 kV tube voltage, 200 mA tube current, 0.5 s tube rotation time, 0.6 pitch, standard (B) algorithm, and 512 × 512 matrix. A high-pressure syringe was used to inject 70–100 mL of contrast agent into the patient's antecubital vein at a flow rate of 2.5–3.5 mL/s. A plain scan, arterial phase scan at 25–30 s after contrast agent injection, and venous phase scan at 55–60 s after contrast agent injection were performed.

Image Segmentation and Feature Extraction

Tumor segmentation and radiomics feature calculation were based on MATLAB's IBEX software package (10). On the arterial and venous phase CT images before the GIST surgery, the region of interest (ROI) was delineated in layers along the tumor contour's edge (**Figure 1**). ROI segmentation was performed by a radiologist and confirmed by another radiologist. The two radiologists were blinded to the Ki67 index before ROI segmentation.

Before feature extraction, all images were resampled according to the voxel size of 1×1×1 mm^3, and the gray-level value was normalized to 1–64 for preprocessing. The radiomics parameters used in this study included six categories, i.e., histogram parameter (n = 48), 2.5D and 3D gray-level co-occurrence matrix (n = 594), neighborhood gray-tone difference matrix (n = 10), gray-level run-length matrix (n = 34), and shape and size (n = 18). A total of 704 radiomics parameters were extracted from the lesions in each phase, so there were 1,408 radiomics parameters for each patient.

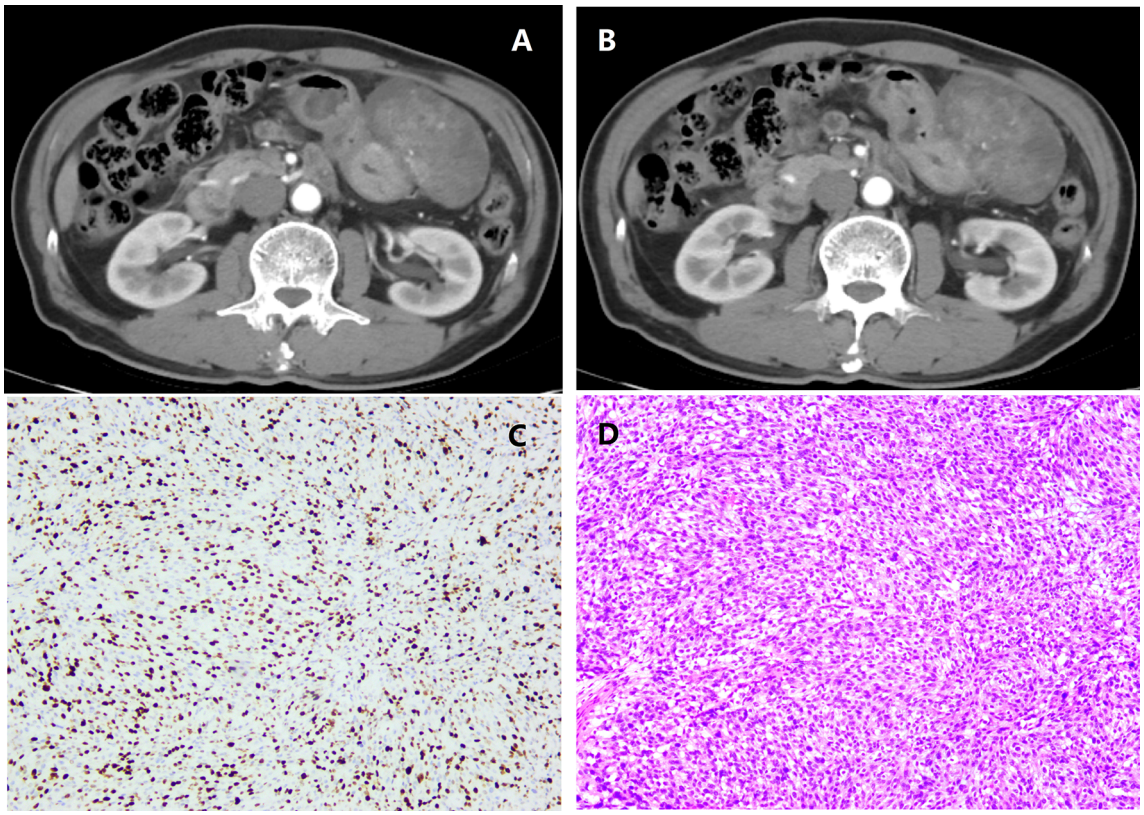


FIGURE 1 | A patient with intestinal gastrointestinal stromal tumor (GIST) belonging to the high malignant potential. **(A)** Arterial phase; **(B)** Venous phase; **(C)** Immunohistochemical photomicrograph showed a high Ki67 index (about 35%, original magnification, ×10); **(D)** Pathology demonstrated a high-risk GIST with a high mitotic rate (> 4/HPF, original magnification, ×20).

Dimensionality Reduction and Modeling

The radiomics features were first subjected to Spearman's correlation analysis, with a correlation coefficient threshold of 0.8. Subsequently, the Boruta algorithm was used to perform dimensionality reduction again. The random forest classifier was used to construct a prediction model, perform 10-fold cross-validation on the training set, and eventually evaluate the model's generalization ability in the external validation set. The performance of the model was evaluated on the basis of receiver operating characteristic (ROC) curves. Delong test was used to evaluate the AUC values of the model in the training set and external validation set. Then, the Brier score was used to quantify overall performance of the model. Brier score is the mean squared difference between the observed and predicted outcome. It is a combination of calibration and differentiation.

RESULTS

The clinical characteristics of the patients with GISTs (**Table 1**) indicated no statistical differences in age, gender, and tumor location between the high-Ki67 group and the low-Ki67 group in the training or validation sets. In addition, no significant difference in the clinical characteristics was found between the training and external validation sets.

After dimensionality reduction, 21 radiomics features were identified, Among these, seven gray-level co-occurrence matrices features (3D), four shape and size features, three gray-level co-occurrence matrices features (2.5D), three neighborhood gray-tone difference matrix features, three histogram features and only one gray level run length matrix were selected. There were 13 features in the venous phase and four in the arterial phase (**Figure 2**). The ROC curve was used to evaluate the performance of the model. The AUC values of the training set and external validation set were 0.835 (95% confidence interval(CI): 0.761–0.908) and 0.784 (95% CI: 0.691–0.874), respectively (**Figure 3**). Brier score of the model was 0.14 in the training set and 0.16 in the external validation set. DeLong test has showed the prediction model had no significantly reduced efficiency in the external validation set.

DISCUSSION

Based on CT images, the radiomics model's performance in preoperative prediction of the Ki67 index was studied. Through independent external validation of multiple centers, this study

TABLE 1 | Patient characteristics in the training and external validation cohorts.

	Training cohort (N=192)			External validation cohort (N=152)		
	Low-Ki67 (N=128)	High-Ki67 (N=64)	P	Low-Ki67 (N=107)	High-Ki67 (N=45)	P
Age (years)	61.8 ± 11.23	63.02 ± 12.82	0.536	55.86 ± 10.74	60.78± 13.29	0.237
Sex (n)			0.609			0.305
Female	61	28		59	27	
Male	67	36		48	18	
Site (n)			0.459			0.286
Gastric	83	26		31	15	
Intestinal	45	38		76	30	

P < 0.05 indicates that difference is statistically significant.

show that the model performed well in predicting whether the Ki67 index was low or high and had good generalization ability.

With the extensive application of adjuvant imatinib therapy, studies have evaluated the correlation between the Ki67 index and adjuvant imatinib therapy. Kaplan-Meier survival analysis and multivariate analysis showed that patients with GISTs and a Ki67 index >8% had a worse prognosis, which was related to imatinib treatment. Particularly, in the high-risk GIST group the rate of recurrence was higher in patients with a Ki67 index >8% (11), suggesting that the Ki67 index >8% was an unfavorable factor for adjuvant imatinib therapy (12, 13). And Sugita et al. have found that 8% cutoff was a good supplementary prognostic factor for NIH risk stratification. Ki67 greater than 8% would indicate poor prognosis in both low - and intermediate-risk groups (14). Raut et al. also concluded that Ki67 was one of the reference markers for imatinib's adjuvant treatment time after surgery (15). High Ki67 expression is associated with poor pathological characteristics and aggressive behaviors of GISTs and can be used as an independent prognostic marker of GISTs (16), But in the existing GIST diagnostic and treatment systems, although it is unrealistic to directly use the Ki67 index to replace the NIH risk stratification of GISTs, the Ki67 index (cutoff value of 8%) can become a supplement to the NIH risk stratification standard for evaluating the prognosis of high-risk GISTs and selecting a suitable population for adjuvant imatinib therapy. CT has the advantages of being noninvasive, objective, and

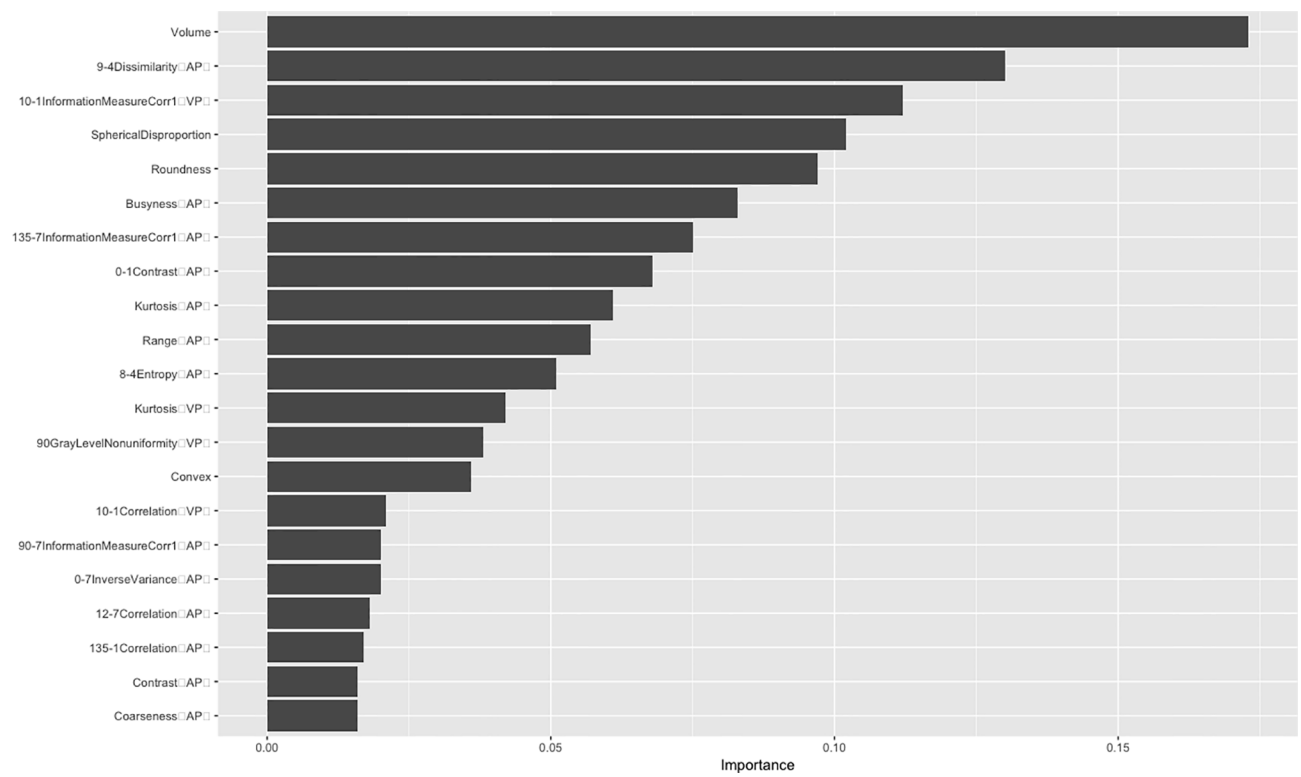


FIGURE 2 | The importance of radiomics features. AP, arterial phase; VP, venous phase.

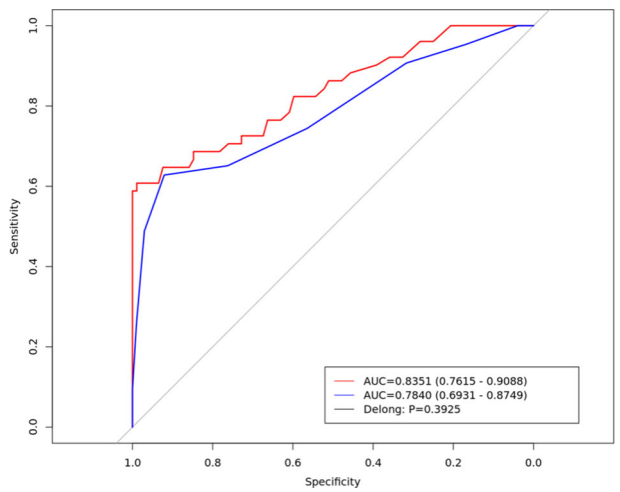


FIGURE 3 | The Receiver operating characteristic curve of the radiomics prediction model in the training set (red line) and external validation set (blue line).

convenient. Preoperative radiomics prediction of the Ki67 index (cutoff value of 8%) provided an additional value for evaluating the prognosis of GISTs.

Numerous studies have demonstrated a close relationship between the imaging characteristics of a tumor and the Ki67 index. The mean, median, and percentage of the apparent diffusion coefficient in functional diffusion imaging of hepatocellular carcinoma are significantly and negatively correlated with the Ki67 level (17). A study by Peng et al. established a multiple linear regression model for preoperative prediction of the Ki67 index of pulmonary nodules with ground-glass opacity based on computed tomography (CT) images (18). Previous studies on GIST imaging for predicting the Ki67 index had two cutoff values. A study by Li et al. used 5% as the cutoff value of the Ki67 index. Through logistic regression analysis, tumor size and ulcers' presence were the most effective features to predict the Ki67 index being ≤5% or >5% in GISTs (19). A study by Zhang et al. set the cutoff value of the Ki67 index to 10% and used radiomics based on enhanced CT images to predict the Ki67 index of patients with GISTs. Their radiomics model performed well, and the AUC of the training set reached 0.801 (20).

NIH risk stratification has showed certain deficiencies with the extensive application of targeted therapies; and it has been removed from the National Comprehensive Cancer Network guidelines since 2015 (21). As a supplementary factor, Ki67 (8%) may make NIH risk stratification better, and its high consistency makes it possible to replace the counting of mitosis count and integrate with other parameters to be a new prognostic stratification (14, 22). Currently, most radiology studies of the GISTs are still focusing on NIH stratification, while those for Ki67 are relatively few. Our results were consistent with those of two previous similar studies, both suggesting that imaging could predict the Ki67 index level, and both found that tumor size played a very important role in Ki67 prediction. Ki67 is related to mitosis in cells, so the higher the index, the more active the tumor proliferation, resulting in a larger tumor size of GISTs (23, 24). Notably, most of the previous radiomics studies of GISTs were mainly single-center studies. The different scanning equipment, parameter settings, and post-processing reconstruction algorithms led to differences in radiomics parameters (25–27). The heterogeneity of single-center imaging data is low, and many prediction models have not been externally validated, so the possibility of overfitting is high (28). Although the prediction model's performance in our study lower in the external validation set, it still maintained at 0.75. Delong test indicating that the model kept a reliable generalization ability.

In the post-NIH stratification era, the Ki67 imaging prediction research has certain clinical significance and great potential. This study has some limitations. First, it is a retrospective study with selection bias; further prospective studies are needed to verify our radiomics model. Second, although this was a multi-center study, we only collected data from three hospitals, and our sample size is small; further studies with a larger sample size with data collected from more hospitals will be necessary. Third, no uniform standard for selecting the cutoff value of the Ki67 index is available, and the cutoff values of the Ki67 index are still in dispute.

CONCLUSION

In summary, the Ki67 index (cutoff value of 8%) can assist NIH risk classification of GISTs, providing supplementary information for the prognostic evaluation of GISTs. As a noninvasive and objective examination, CT radiomics effectively predicted the Ki67 index of GISTs and has the potential for clinical translation. However, verification using multi-center, large sample data will be required before the radiomics model can be truly applied in clinical practices.

DATA AVAILABILITY STATEMENT

The raw/processed data required to reproduce these findings cannot be shared at this time as the data also forms part of an ongoing study. Requests to access the datasets should be directed to gerxyuan@zju.edu.cn.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of the First Affiliated Hospital Zhejiang University School of Medicine. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YZ and ML: conception and design, writing, review, and revision of the manuscript. MW, LZ, CH and MF: analysis data. YZ: study supervision. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Yan L, Zou L, Zhao W, Wang Y, Liu B, Yao H, et al. Clinicopathological Significance of C-KIT Mutation in Gastrointestinal Stromal Tumors: A Systematic Review and Meta-Analysis. *Sci Rep* (2015) 5(1):13718. doi: 10.1038/srep13718
2. Tan Y, Trent JC, Wilky BA, Kerr DA, Rosenberg AE. Current Status of Immunotherapy for Gastrointestinal Stromal Tumor. *Cancer Gene Ther* (2017) 24(3):130–3. doi: 10.1038/cgt.2016.58
3. Corless CL, Barnett CM, Heinrich MC. Gastrointestinal Stromal Tumours: Origin and Molecular Oncology. *Nat Rev Cancer* (2011) 11(12):865–78. doi: 10.1038/nrc3143
4. Joensuu H. Risk Stratification of Patients Diagnosed With Gastrointestinal Stromal Tumor. *Hum Pathol* (2008) 39(10):1411–9. doi: 10.1016/j.humpath.2008.06.025
5. Sozutek D, Yanik S, Akkoca AN, Sözütek A, Ozdemir ZT, Avşar ÇU, et al. Diagnostic and Prognostic Roles of DOG1 and Ki-67, in GIST Patients With Localized or Advanced/Metastatic Disease. *Int J Clin Exp Med* (2014) 7(7):1914–22.
6. Belev B, Brčić I, Prejac J, Golubić ZA, Vrbanec D, Božikov J, et al. Role of Ki-67 as a Prognostic Factor in Gastrointestinal Stromal Tumors. *World J Gastroenterol* (2013) 19(4):523–7. doi: 10.3748/wjg.v19.i4.523
7. Demir L, Ekinci N, Erten C, Kucukzeybek Y, Alacacioglu A, Somali I, et al. Does Immunohistochemistry Provide Additional Prognostic Data in Gastrointestinal Stromal Tumors? *Asian Pac J Cancer Prev Apjcp* (2013) 14 (8):4751–8. doi: 10.7314/APJCP.2013.14.8.4751
8. Hirota S. Differential Diagnosis of Gastrointestinal Stromal Tumor by Histopathology and Immunohistochemistry. *Trans Gastroenterol Hepatol* (2018) 3:27. doi: 10.21037/tgh.2018.04.01
9. Nakamura N, Yamamoto H, Yao T, Oda Y, Nishiyama K, Imamura M, et al. Prognostic Significance of Expressions of Cell-Cycle Regulatory Proteins in Gastrointestinal Stromal Tumor and the Relevance of the Risk Grade. *Hum Pathol* (2005) 36(7):828–37. doi: 10.1016/j.humpath.2005.03.012
10. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE, et al. IBEX: An Open Infrastructure Software Platform to Facilitate Collaborative Work in Radiomics. *Med Phys* (2015) 42(3):1341–53. doi: 10.1118/1.4908210
11. Zhao WY, Xu J, Wang M, Zhang ZZ, Tu L, Wang C-J, et al. Prognostic Value of Ki67 Index in Gastrointestinal Stromal Tumors. *Int J Clin Exp Pathol* (2014) 7(5):2298–304.
12. Heinrich MC, Rankin C, Blanke CD, Demetri GD, Borden EC, Ryan CW, et al. Correlation of Long-Term Results of Imatinib in Advanced Gastrointestinal Stromal Tumors With Next-Generation Sequencing Results: Analysis of Phase 3 SWOG Intergroup Trial S0033. *JAMA Oncol* (2017) 3(7):944–52. doi: 10.1001/jamaoncol.2016.6728
13. Miki Y, Kawamura T, Sugisawa N, Makuuchi R, Nakajima T, Tokunaga M, et al. Risk Classification Using the Ki-67 Labeling Index for Surgically-Treated Gastric Gastrointestinal Stromal Tumors. *Hepatogastroenterology* (2015) 62 (140):919–23.
14. Sugita S, Hirano H, Hatanaka Y, Fujita H, Kubo T, Kikuchi N, et al. Image Analysis is an Excellent Tool for Quantifying Ki-67 to Predict the Prognosis of Gastrointestinal Stromal Tumor Patients. *Pathol Int* (2018) 68(1):7–11. doi: 10.1111/pin.12611
15. Raut CP, Espat NJ, Maki RG, Araujo DM, Trent J, Williams TF, et al. Efficacy and Tolerability of 5-Year Adjuvant Imatinib Treatment for Patients With Resected Intermediate- or High-Risk Primary Gastrointestinal Stromal Tumor: The PERSIST-5 Clinical Trial. *JAMA Oncol* (2018) 4(12):e184060. doi: 10.1001/jamaoncol.2018.4060
16. Jeong SY, Park WW, Kim YS, Park YI, Kim SH, Yoon WJ, et al. Prognostic Significance of Ki-67 Expression in Patients Undergoing Surgical Resection for Gastrointestinal Stromal Tumor. *Korean J Gastroenterol* (2014) 64(2):87–92. doi: 10.4166/kjg.2014.64.2.87
17. Hu XX, Yang ZX, Liang HY, Ding Y, Grimm R', Fu C-X, et al. Whole-Tumor MRI Histogram Analyses of Hepatocellular Carcinoma: Correlations With Ki-67 Labeling Index. *J Magn Reson Imaging* (2017) 46(2):383–92. doi: 10.1002/jmri.25555
18. Peng M, Peng F, Zhang C, Wang Q, Li Z, Hu H, et al. Preoperative Prediction of Ki-67 Labeling Index By Three-Dimensional CT Image Parameters for Differential Diagnosis Of Ground-Glass Opacity (GGO). *PloS One* (2015) 10 (6):e0129206. doi: 10.1371/journal.pone.0129206
19. Li H, Ren G, Cai R, Chen J, Wu X, Zhao J, et al. A Correlation Research of Ki67 Index, CT Features, and Risk Stratification in Gastrointestinal Stromal Tumor. *Cancer Med* (2018) 7(9):4467–74. doi: 10.1002/cam4.1737
20. Zhang QW, Gao YJ, Zhang RY, Zhou XX, Chen SL, Zhang Y, et al. Personalized CT-Based Radiomics Nomogram Preoperative Predicting Ki-67 Expression in Gastrointestinal Stromal Tumors: A Multicenter Development and Validation Cohort. *Clin Trans Med* (2020) 9(1):12. doi: 10.1186/s40169-020-0263-4
21. Judson I, Bulusu R, Seddon B, Dangoor A, Wong N, Mudan S. UK Clinical Practice Guidelines for the Management of Gastrointestinal Stromal Tumours (GIST). *Clin Sarcoma Res* (2017) 7:6. doi: 10.1186/s13569-017-0072-8
22. Uguen A, Conq G, Doucet L, Talagas M, Costa S, De Braekeleer M, et al. Immunostaining of Phospho-Histone H3 and Ki-67 Improves Reproducibility of Recurrence Risk Assessment of Gastrointestinal Stromal Tumors. *Virchows Arch* (2015) 467(1):47–54. doi: 10.1007/s00428-015-1763-2
23. Zhou Y, Hu W, Chen P, Abe M, Shi L, Tan S-Y, et al. Ki67 Is a Biological Marker of Malignant Risk of Gastrointestinal Stromal Tumors: A Systematic Review and Meta-Analysis. *Medicine* (2017) 96(34):e7911. doi: 10.1097/MD.0000000000007911
24. Artigiani Neto R, Logullo AF, Stávale JN, Lourenço LG. Ki-67 Expression Score Correlates to Survival Rate in Gastrointestinal Stromal Tumors (GIST). *Acta Cir Bras* (2012) 27(5):315–21. doi: 10.1590/S0102-86502012000500007
25. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring Computed Tomography Scanner Variability of Radiomics Features. *Invest Radiol* (2015) 50(11):757–65. doi: 10.1097/RLI.0000000000000180
26. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning Methods for Quantitative Radiomic Biomarkers. *Sci Rep* (2015) 5:13087. doi: 10.1038/srep13087
27. Balagurunathan Y, Gu Y, Wang H, Kumar V, Grove O, Hawkins S, et al. Reproducibility and Prognosis of Quantitative Features Extracted From CT Images. *Transl Oncol* (2014) 7(1):72–87. doi: 10.1593/tlo.13844
28. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of A Method to Compensate Multicenter Effects Affecting CT Radiomics. *Radiology* (2019) 291(1):53–9. doi: 10.1148/radiol.2019182023

Conflict of Interest: Author CH was employed by the company Deepwise & League of PHD Technology Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhao, Feng, Wang, Zhang, Li and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

REVIEW
published: 20 September 2021
doi: 10.3389/fonc.2021.698373

Progress of MRI Radiomics in Hepatocellular Carcinoma

Xue-Qin Gong[1], Yun-Yun Tao[1], Yao–Kun Wu[1], Ning Liu[1], Xi Yu[1], Ran Wang[1], Jing Zheng[1], Nian Liu[1], Xiao-Hua Huang[1], Jing-Dong Li[2], Gang Yang[2], Xiao-Qin Wei[3], Lin Yang[1] and Xiao-Ming Zhang[1]*

[1] Medical Imaging Key Laboratory of Sichuan Province, Department of Radiology, Medical Research Center, Affiliated Hospital of North Sichuan Medical College, Nanchong, China, [2] Department of Hepatocellular Surgery, Institute of Hepato-Biliary-Intestinal Disease, Affiliated Hospital of North Sichuan Medical College, Nanchong, China, [3] School of Medical Imaging, North Sichuan Medical College, Nanchong, China

Background: Hepatocellular carcinoma (HCC) is the sixth most common cancer in the world and the third leading cause of cancer-related death. Although the diagnostic scheme of HCC is currently undergoing refinement, the prognosis of HCC is still not satisfactory. In addition to certain factors, such as tumor size and number and vascular invasion displayed on traditional imaging, some histopathological features and gene expression parameters are also important for the prognosis of HCC patients. However, most parameters are based on postoperative pathological examinations, which cannot help with preoperative decision-making. As a new field, radiomics extracts high-throughput imaging data from different types of images to build models and predict clinical outcomes noninvasively before surgery, rendering it a powerful aid for making personalized treatment decisions preoperatively.

Objective: This study reviewed the workflow of radiomics and the research progress on magnetic resonance imaging (MRI) radiomics in the diagnosis and treatment of HCC.

Methods: A literature review was conducted by searching PubMed for search of relevant peer-reviewed articles published from May 2017 to June 2021.The search keywords included HCC, MRI, radiomics, deep learning, artificial intelligence, machine learning, neural network, texture analysis, diagnosis, histopathology, microvascular invasion, surgical resection, radiofrequency, recurrence, relapse, transarterial chemoembolization, targeted therapy, immunotherapy, therapeutic response, and prognosis.

Results: Radiomics features on MRI can be used as biomarkers to determine the differential diagnosis, histological grade, microvascular invasion status, gene expression status, local and systemic therapeutic responses, and prognosis of HCC patients.

Conclusion: Radiomics is a promising new imaging method. MRI radiomics has high application value in the diagnosis and treatment of HCC.

Keywords: hepatocellular carcinoma, magnetic resonance imaging, intravoxel incoherent motion, radiomics, immune checkpoint inhibitors, target therapies, therapeutic response, diagnosis

OPEN ACCESS

Edited by:
Omar Sultan Al-Kadi, The University of Jordan, Jordan

Reviewed by:
Chen Liu, Army Medical University, China
Sangeetha Gupta, Amity University, India

***Correspondence:**
Lin Yang
linyangmd@163.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *21 April 2021*
Accepted: *31 August 2021*
Published: *20 September 2021*

Citation:
Gong X-Q, Tao Y-Y, Wu Y–K, Liu N, Yu X, Wang R, Zheng J, Liu N, Huang X-H, Li J-D, Yang G, Wei X-Q, Yang L and Zhang X-M (2021) Progress of MRI Radiomics in Hepatocellular Carcinoma. Front. Oncol. 11:698373. doi: 10.3389/fonc.2021.698373

INTRODUCTION

Hepatocellular carcinoma (HCC) is the sixth most common cancer and the third leading cause of cancer-related death worldwide (1). Although the diagnostic criteria of HCC continue to improve, its prognosis remains unsatisfactory (2). In addition to certain factors, such as tumor size and number and vascular invasion displayed on traditional imaging, some histopathological features and gene expression parameters are also important in the prognoses of patients with HCC. However, many current staging systems for HCC have not taken into consideration the above-mentioned histopathological features or genetic traits beyond the size and number and vascular invasion of the tumor (3, 4). Most parameters are based on postoperative pathological examinations, which cannot help with preoperative decision-making. To better stratify HCC patients before surgery, make more accurate treatment decisions, and improve the prognoses of patients, there is an urgent need for a noninvasive method that can accurately predict the histopathological features and gene expression parameters before surgery. The rapid development of artificial intelligence has played an important role in personalized precision medicine (5). Radiomics, a new technology, can transform the potential histopathological and physiological information in images into high-dimensional quantitative image features that can be mined (6, 7).The study of radiomics will contribute to the early diagnosis and treatment of HCC and ultimately improve survival (8, 9). In recent years, many studies have confirmed the application values of magnetic resonance imaging (MRI) radiomics in the diagnosis and differentiation (10, 11), histological grading (12, 13), microvascular invasion (MVI) assessment (14, 15), radiogenomics (16, 17),prediction of relapse and prognosis after surgical resection (18–20), response to transarterial chemoembolization(TACE) (21, 22) and systemic treatment efficacy of HCC (23).

To better understand the research hotspots and trends of MRI radiomics in HCC, we used PubMed to identify important recent publications on MRI radiomics in HCC, selected research articles and reviews and used bibliometric method to visually analyze the countries, institution, authors, and keywords of MRI radiomics in HCC. Meanwhile, this study reviews the radiomics workflow from image acquisition and reconstruction, segmentation, feature extraction, feature selection and modeling to model validation, and the research progress of MRI radiomics in HCC.

BIBLIOMETRICS OF MRI RADIOMICS IN HCC

The authors conducted a literature review using PubMed to identify important recent publications and determine the current status of radiomics in HCC. A comprehensive list of MeSH terms and keywords was included in the search: HCC, MRI, radiomics, deep learning(DL), artificial intelligence, machine learning, neural network, texture analysis, diagnosis, histopathology, microvascular invasion, surgical resection, radiofrequency, recurrence, relapse, TACE, targeted therapy, immunotherapy, therapeutic response, and prognosis. The inclusion criteria were as follows: (1) original research articles and review articles published in the English language between May 2017 and June 2021; (2) literature related to MRI radiomics or DL; and (3) literature related to diagnosis and differentiation, histological grading, MVI assessment, radiogenomics, prediction of relapse and prognosis after surgical resection, response to TACE and systemic treatment efficacy of HCC. Articles that were not published in English and those containing irrelevant information on the subject were excluded. We also excluded articles that were published before May 2017 and after June 2021. In total, 129 articles were ultimately retrieved. After screening the titles, abstracts and full texts (if appropriate), only 84 papers met the criteria for inclusion (**Figure 1**). Those 84 papers were then downloaded with the record content of "Full Record and Cited References" and the file format of "Plain Text". As CiteSpace can

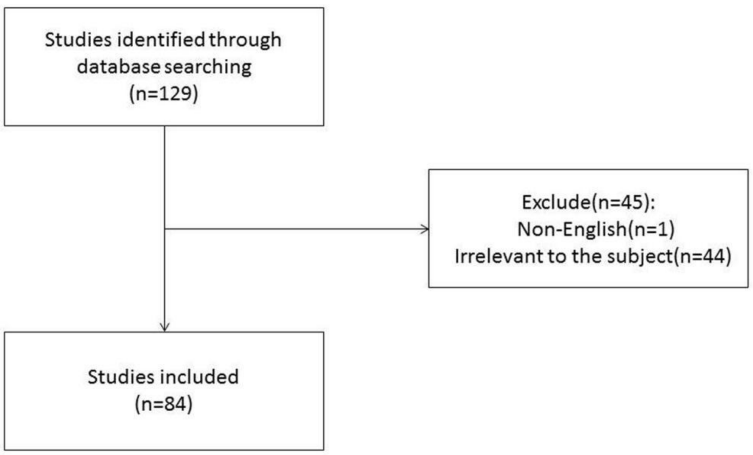


FIGURE 1 | Flow diagram of study selection.

only recognize files named "download *.txt", the files were renamed accordingly. The bibliometric software CiteSpace5.7.R2 (64 bits) was utilized for this study to visually analyze the countries, institution, authors, and keywords draw relevant charts. The articles originated from a total of 12 countries, and the top five countries were China (24), the USA (13), South Korea (5), Germany (3), and France (2). A total of 30 institutions published manuscripts independently or cooperatively. The top five institutions were the Chinese Academy of Sciences (12), Fudan University (10), GE Healthcare (8), Sun Yat-Sen University (7), and Sichuan University (5). Bin Song and Xin Li were the most prolific authors. Meng-Su Zeng, Jie Tian, and Dong-Sheng Gu were also active in this field. "Hepatocellular carcinoma" was the most important term, followed by "radiomics", "recurrence", and "microvascular invasion". According to the link strength of keyword cooccurrence, the network was divided into eight clusters, and the largest cluster was "tumor differentiation (#0)" (**Figure 2**).

RADIOMICS WORKFLOW

Radiomics extracts high-throughput features from images and transforms imaging data into high-resolution mining data spaces through machine learning (25). Quantitative radiological data can therefore be extracted and applied to clinical decision-making (25). The workflow of radiomics usually includes five steps (6), which are described below.

IMAGE ACQUISITION AND RECONSTRUCTION

Imaging techniques that can be used for radiomics include MRI, computed tomography (CT), positron-emission tomography, and ultrasound. Among them, MRI has the advantage of depicting more soft-tissue features. Radiomics is an imaging analysis method; thus, it is vital to standardize high-quality images (26–28). This makes it necessary to preprocess the imaging data; otherwise, a widely promoted standard scanning protocol is needed to reduce the variability in radiomic features and improve the performance of radiomic models (25, 29).

IMAGE SEGMENTATION

Manual, automatic, and semiautomatic segmentation are often used to segment the volume or region of interest in a target tissue (30). Manual segmentation is most reliable, but it involves intraobserver and interobserver variability. Its labor and time cost are high. The segmentation of an image often requires multiple clinicians or the same clinician at multiple times. The intraobserver and interobserver variability can be improved by screening the intraobserver and interobserver consistency. The purpose of automatic segmentation is to mark the regions of interest automatically by a computer. Semiautomatic segmentation involves manual corrections. Automatic segmentation algorithms include image segmentation based on thresholds, image segmentation based on region growing, and image segmentation based on edge detection. Some classical algorithms perform well at delineating liver lesions (31, 32).

IMAGE FEATURE EXTRACTION

Image features include semantic features and nonsemanticfeatures (33). Semantic features include qualitative (shapes, boundaries, etc.) and quantitative features, and their analysis depends on the radiologist's knowledge. Nonsemantic features are quantitative descriptors extracted from tissues of interest, including shape and statistical features (34). The shape features of objects in images include topological features, distances, perimeters, areas, geometric features, and descriptions of shape and orientation. Statistical features can be further divided into first-order, second-order, and high-order features. First-order features are usually called density features, which involve gray-level histogram information simply describing the global distribution of gray levels in an image. Such features cannot describe the local distribution of gray levels in an image or the spatial position of each gray level (35). Second-order features are often called texture features. These reflect the relationships between adjacent voxels. High-order features are usually called filtering features and are generated by wavelet and Laplacian Gaussian filtering, for example, in addition to first-order and second-order features.

FEATURE SELECTION AND MODELING

Many features can be extracted from a high-throughput image, but using all the features to analyze an image will lead to overfitting. The best features can be selected by dimensionality reduction to improve the efficiency of the model. The methods of feature selection can be divided into three categories: filter, wrapper, and embedded (36). The goal of radiomics is to establish a prediction model for clinical outcomes from selected features. The modeling methods include logistic regression, k-nearest neighbor, decision trees, ensemble learning, and support vector machines. It is recommended to test the effectiveness of several forecasting models to select the model with the best performance (37).

MODEL VALIDATION

The prediction model can be validated by internal cross-validation, such that the model can be further optimized and the prediction performance can be maximized. Validation of the

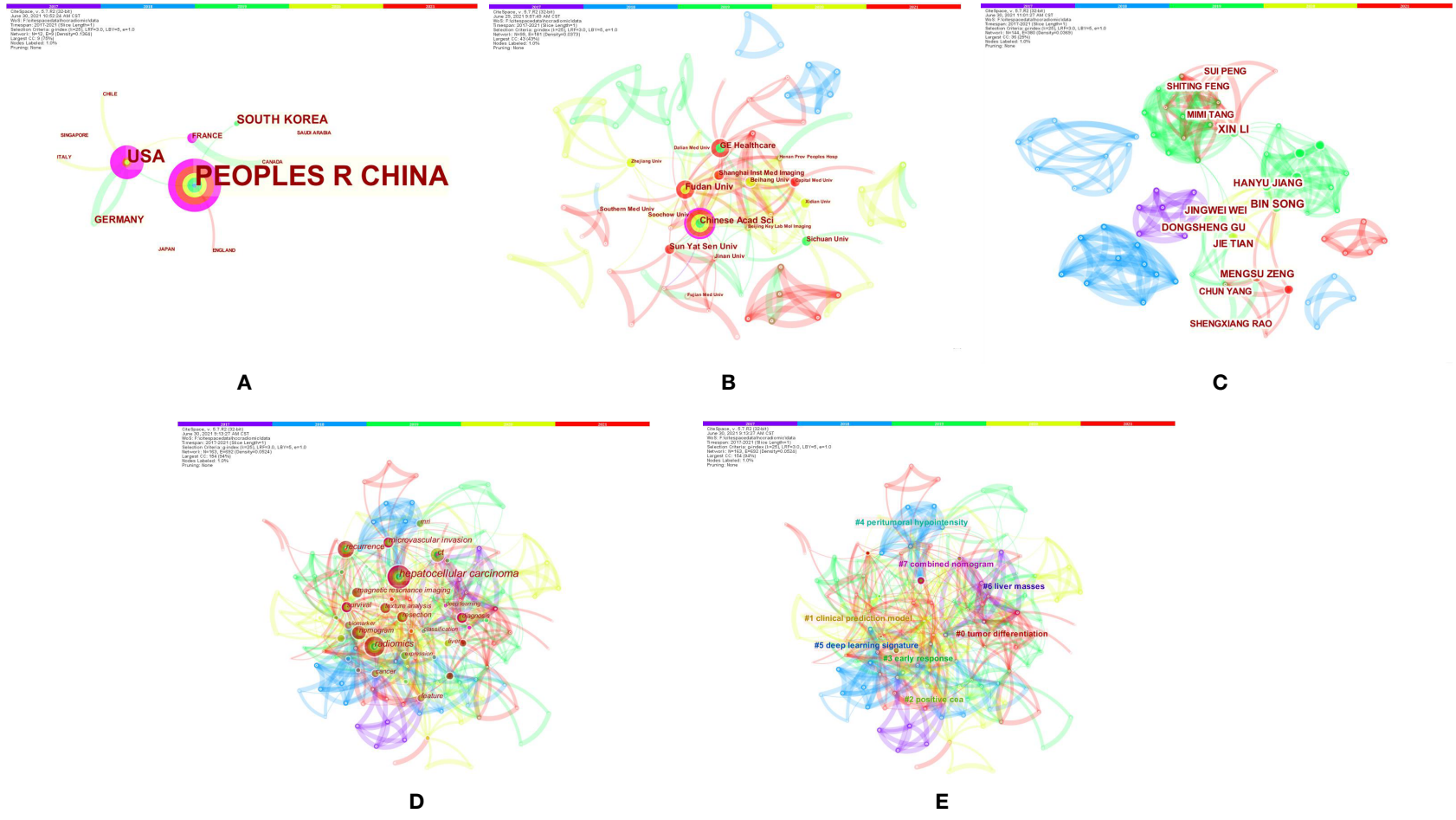


FIGURE 2 | Bibliometrics of magnetic resonance imaging radiomics in hepatocellular carcinoma. **(A)** Co-occurrence map of countries. **(B)** Co-occurrence map of institutions. **(C)** Co-occurrence map of authors. **(D)** Terms in theco-occurrence network. **(E)** Terms in theco-occurrence clusters.

model should be carried out in a separate cohort (37). For differentiation analysis, the receiver operating characteristic (ROC) curve is the most commonly used method to evaluate the performance of the model. The area under the ROC curve (AUC) or the sensitivity and specificity of the model can be used to evaluate whether the model can predict clinical outcomes. For survival analysis, the concordance index (C-index) and the time-related ROC curve are usually used for validation (38).

DIAGNOSIS AND DIFFERENTIATION

At present, the diagnosis of HCC is mainly based on imaging methods such as MRI, CT, and ultrasound. Because HCC has a typical enhancement mode, contrast-enhanced CT and dynamic contrast-enhanced MRI play important roles in the diagnosis of HCC (39–42). The European Association for the Study of the Liver standard (40) and the Liver Imaging Reporting and Data System (43) are widely recognized. However, the evaluation of imaging features may be subjective because radiologists have different experiences and different familiarities with the system (44, 45). Radiomics has important application value in the diagnosis of solid tumors because it uses advanced image processing technology to extract high-throughput data and quantitative analysis of tumor behavior and heterogeneity (6, 46–51).

Radiomics signatures based on conventional precontrast T1-weighted imaging, postcontrast T1-weighted imaging, T2-weighted imaging, diffusion-weighted imaging (DWI), and intravoxel incoherent motion (IVIM), whether alone or in combination with clinical data, are all valuable for HCC differentiation (52–59), and their differentiation efficiency is almost equal to that of experienced radiologists (10-year experience) (52). HCC, intrahepatic cholangiocarcinoma (ICC), and HCC-ICC have common risk factors (60, 61), and their typical qualitative MRI features may overlap (24, 62–64). Therefore, the conventional MRI diagnosis of HCC is still uncertain. According to Liu et al. (54), the imaging features extracted from MR images have great potential to differentiate combined hepatocellular cholangiocarcinoma from cholangiocarcinoma and HCC, showing a maximum AUC of 0.77. Recently, Zhu et al. (56) studied the application value of histogram features on IVIM-DWI in the differential diagnosis of HCC. They found that the histogram parameters of IVIM-DWI could distinguish hepatic hemangiomas, hepatic cysts, and HCC and that the volume of the pseudodiffusion coefficient and perfusion fraction had better diagnostic value than other histogram parameters (56).

In recent years, DL technology has been developed and has achieved excellent performance in the classification of hepatic lesions (65–71). Hamm CA et al. (65) developed a proof-of-concept convolutional neural network (CNN)-based DL system and classified 494 hepatic lesions from six categories on MRI. The system demonstrated 92% accuracy, 92% sensitivity and 98% specificity, and their results showed a 90% sensitivity for classifying HCC compared to 60%/70% for radiologists.

HISTOLOGICAL GRADING

The histological grading of HCC is key to determining the best treatment scheme and prognosis of a patient. High-grade HCC patients have a higher intrahepatic relapse rate than low-grade HCC patients (72, 73), and most high-grade HCC patients need larger safe resection margins and more frequent postoperative follow-up visits (74, 75). The radiomic features of precontrast T1-weighted imaging, postcontrast T1-weighted imaging, and T2-weighted imaging, whether alone or in combination with clinical data (76), are all valuable for identifying poorly differentiated HCC (13, 76–80). In addition, recent studies have shown the application value of functional MRI radiomics based on IVIM-DWI in predicting the pathological grade of HCC (12, 81, 82). Shi et al. (82) performed MRI on 52 HCC patients and extracted histogram indices from IVIM parameter maps. Eighteen IVIM histogram indices showed the capacity to differentiate histopathological grades. By establishing a diagnostic model based on logistic regression and integrating different histogram indices showing significant differences between different subgroups, the maximum diagnostic power for distinguishing HCC histological grades was obtained (AUC=0.917). This study indicated that histogram indices extracted from IVIM parameter maps had great potential in predicting histopathological grade (82). Geng et al. (12) extracted 107 radiomic features from SWI images of 53 HCC patients and Spearman correlation coefficients were used to evaluate the correlation between SWI radiomic features and histopathology. They found that the SWI radiomic features were significantly correlated with histopathological grades.

MVI

MVI is diagnosed depending on postoperative tissue specimens, but detection by conventional imaging is difficult. The presence of MVI indicates that the tumor has strong biological invasiveness, which can increase the relapse rate of HCC more than fourfold (83, 84). Accurate preoperative prediction of MVI of HCC can help doctors adjust treatment strategies in a timely manner (such as expanding the resection range), optimize treatment plans, reduce the risks of postoperative relapse, and improve the prognosis (84, 85). Enhanced MRI is helpful to predict MVI in HCC (83, 86–91). MRI-based radiomics (15, 92–102) and DL systems (103–106) have shown good performance in predicting MVI in HCC. The increase in clinicopathological risk factors and qualitative imaging features can improve the prediction efficiency of the model (14, 98, 107, 108). Li et al. (101) found that tumor volume–based IVIM histogram analysis can be used to predict MVI and that the fifth percentile of the true diffusion coefficient is most beneficial to predict MVI of HCC. Zhang et al. (107) extracted imaging features based on preoperative multimodal MR images and constructed an MVI prediction model (combined model) by combining the clinical features and qualitative imaging features of patients

with HCC. The AUC in the validation cohort of their combined model was 0.858, which was higher than the AUC (0.820) in the validation cohort of the model constructed from individual radiomic features, indicating that the prediction efficiency of the combined model was higher. Song D et al. (104) predicted MVI using radiomics and DL in 601 patients with HCC based on preoperative MRI. Their results showed that the radiomics model achieved an AUC of 0.731, the DL model based only on MRI images achieved an AUC of 0.915, and a DL model combined with clinical parameters achieved an AUC of 0.931. These studies indicated that the model combining radiomics, DL, and clinical parameters showed the best predictive performance.

RADIOGENOMICS

The biological behavior of a tumor is closely related to its gene expression profile. Biopsy is a widely used method to evaluate gene expression before surgery, but biopsy is an invasive examination that may cause bleeding and other complications. Therefore, patients are often unwilling to undergo this examination. In recent years, radiogenomics has gradually become more widely applied in HCC research. The purpose of radiogenomics is to determine the relationship between semantic and quantitative image data and genomic and molecular measurements, thus constructing correlation diagrams related to results or other clinical measurements (33, 109, 110). Segal et al. (111) evaluated the correlation between radiogenomic features and the liver cancer gene phenotype and reported that 78% of liver cancer gene expression profiles could be reconstructed by this combination of features. To date, radiogenomics studies have described the semantic features obtained from MRI (112–115). MRI radiogenomics has the value of predicting gene features with prognostic and therapeutic significance (16, 17, 116–123). Taouli et al. (113) found that there was a strong connection between imaging features, such as the "infiltrative pattern", "mosaic appearance", and "presence of macrovascular invasion", and an aggressive genomic signature determined previously. Shi et al. (82) found that histogram indices extracted from IVIM parameter maps could predict Ki-67 expression. Jun et al. (124) used an immunohistochemical method to detect the expression of programmed cell death-1 (PD-1) and programmed cell death ligand-1 (PD-L1) in 98 ICC patients and extracted radiological features from the arterial phase and portal venous phase of preoperative MR images. The results indicated that the AUCs of the models for predicting PD-1 and PD-L1 expression were 0.897 and 0.897, respectively. The prognoses of PD-1-positive and PD-L1-positive patients were worse than those of PD-1-negative and PD-L1-negative patients, and their 5-year survival rates were 12.5%, 48.3%, 21.9%, and 39.4%, respectively ($P < 0.05$). The results indicated that MRI radiomics could be used as a noninvasive biomarker to evaluate the expression of PD-1 and PD-L1 and the prognosis of ICC patients (**Table 1**).

PREDICTION OF RELAPSE AND PROGNOSIS AFTER SURGICAL RESECTION

Surgical resection is still the main treatment for patients with early HCC (125). However, tumor relapse is still the main cause

TABLE 1 | Summary of radiogenomics studies.

Ref.	Year	Country	Subject number	Key findings
Shi G et al. (82)	2020	China	52	IVIM histogram metrics can predict expression of the cell proliferation marker Ki-67.
Hectors SJ et al. (116)	2020	United States	48	Radiomics features extracted from MR images correlate with quantitative expression of the immune markers CD3, CD68 and CD31and expression of the immunotherapy targets PD-L1 at the protein level, as well as PD1 and CTLA4 at the mRNA level.
Wang W et al. (17)	2020	China	227	The radiomics-based model performs better than the clinico-radiological model for predicting biliary-specific marker CK19 status of HCC.
Gu D et al. (117)	2020	China	293	The MRI-based radiomics signature is significantly related to GPC3positivity (a prognosis factor, was associated with metastasis and recurrence after resection) in patients with HCC.
Ye Z et al. (118)	2019	China	89	Texture analysis on preoperative enhanced MRI can be used to predict the status of the cell proliferation marker Ki-67 after curative resection in patients with HCC.
Fan Y et al. (16)	2021	China	133	Texture analysis based on enhanced MRI can help identify VETC-positive HCC (histological vascular pattern, micrometastases, early recurrence and poor prognosis).
Li Y et al. (119)	2019	China	83	Texture analysis of multiphase MRI images is helpful for predicting expression of the cell proliferation marker Ki-67 in HCC.
Wang HQ et al. (120)	2019	China	86	Texture analysis based on MRI can help identifyCK19-positive HCC(tends to be related to a worse prognosis).
Fan Y et al. (121)	2021	China	151	A combined model including artery phase radiomics score and serum AFP levels based on enhanced MRI can preoperatively predict expression of the cell proliferation marker Ki-67 in HCC.
Huang X et al. (122)	2019	China	100	MRI radiomics features can be used to preoperatively differentiate dual-phenotype HCC from CK7- and CK19 (markers of cholangiocellular carcinoma) -negative HCC.
Chen S et al. (123)	2019	China	207	Radiomics obtained from enhanced MRI can help predict the immunoscore (density of CD3+ and CD8+ T cells) in HCC.

of postoperative death, and the 5-year relapse rate after surgery is close to 70% (126). Improving the ability to preoperatively identify these high-risk patients will guide surgical management, postoperative monitoring, and treatment intervention (127, 128). The radiomic model based on preoperative MRI can be used as a new tool to predict early relapse (18, 19, 129–134), relapse-free survival (135) and overall survival (OS) (136, 137) in patients with HCC after surgery. Hui et al. (130) used preoperative MRI to extract 290 texture parameters to predict the relapse of HCC patients within 730 days after surgical resection. The results showed that the prediction accuracy of texture features based on dynamic contrast-enhanced MRI in the equilibrium phase was 84%. Combining clinical, laboratory, and radiomic data can improve the performance of quantitative models (20, 129, 135, 136, 138). According to Kim et al. (135), the combined clinical and radiomic model had the same performance as the clinicopathological model in predicting early relapse. Zhang et al. evaluated the effectiveness of contrast-enhanced MRI radiomic features in predicting the OS of HCC patients after resection. Their results showed that preoperative clinical features and semantic imaging features were significantly correlated with survival rate; the Barcelona Clinic Liver Cancer stage, uneven tumor margin, and combined rad-score were independently correlated with OS; and the combined model incorporating radiological and radiomic features had a better prediction performance than the clinic-radiological model (136).

PREDICTION OF RESPONSE TO TACE

TACE is recognized as an effective treatment for advanced HCC (125), but its long-term efficacy needs to be further improved (139–141). MRI radiomics can be used to predict the response to TACE treatment and provide a reference for the formulation of individualized treatment plans (21, 22, 142–148). Sun et al. (142) predicted the risk of early postoperative progression based on multiparameter MRI data before TACE. The results showed that the AUC of the model based on DWI features was 0.786 and 0.729 when b=0 and b=500, respectively, followed by the AUC of T2-weighted imaging features (0.729) and the apparent diffusion coefficient (0.714). Compared with any single MRI signal, the MP-MRI signal had a higher AUC, at 0.800. Song et al. (143) revealed that their combined model incorporating radiomic features and clinical radiation risk factors had the best predictive value (C = 0.802).

PREDICTION OF THE SYSTEMIC TREATMENT EFFICACY

The treatment of HCC has been a challenge. Systemic therapies for HCC are current research hotspots. Targeted therapy with sorafenib (149) and lenvatinib (150) and immunotherapy with immune checkpoint inhibitors, especially antibodies against PD-1/PD-L1 pathway members (nivolumab and pembrolizumab), have achieved excellent clinical results (151–158). These results strongly indicated that immune checkpoint inhibitor-based strategies will soon be primary method in the treatment of advanced HCC, and immunotherapy will introduce a new era of HCC therapy. Traditional contrast-enhanced CT and MRI, including functional imaging, are the most commonly used biomarkers for evaluating the therapeutic response in clinical practice (159–172). Research based on contrast-enhanced CT and MR images has shown the value of radiomics and DL in predicting systemic treatment efficacy for advanced HCC (23, 173–175). Mulé et al. (174) analyzed the CT texture features of 92 patients before receiving sorafenib and found that the entropy of portal phase-derived entropy at fine texture scales was an independent predictor of OS, which was confirmed in their validation cohort. Yuan et al. (173) established a radiomics nomogram and measured its ability to evaluate the therapeutic efficacy of anti-PD-1antibodies in the treatment of HCC by combining pretreatment contrast-enhanced CT images and clinical risk factors. The results indicated that the AUCs of the radiomics nomogram were 0.894 and 0.883 in the training and validation cohorts, respectively.

In recent years, MRI radiomics has gradually become more widely applied to systemic treatment evaluation of brain tumors (176, 177). There are still no reports on using MRI radiomics to evaluate the systemic treatments of patients with HCC. We believe that as research progresses, MRI radiomics will play an important role in the evaluation of systemic treatments for HCC in the near future.

CONCLUSION

As a new technology, radiomics can improve the diagnosis and differentiation of HCC, as well as predictions of the stage, histological grade, MVI, gene expression, treatment response, and prognosis of HCC. This is because it allows us to analyze the relationship between high-dimensional quantitative imaging features and clinical and genetic data. Moreover, it is a powerful tool for making personalized treatment decisions before surgery. With the rapid development of targeted therapy and immunotherapy for HCC, radiomics is expected to become a reliable radiological marker for predicting the therapeutic targets and therapeutic responses of HCC patients.

There are still some challenges and limitations in the clinical application of radiomics. First, a key challenge is to ensure that the academic community can obtain high-quality radiological and clinical resources that involve the establishment and promotion of imaging and clinical data acquisition protocols. Second, the analytical methods of radiomics need to be standardized. Third, many radiomics studies are retrospective, whereas a prospective research design is ideal. As technology advances and research progresses, MRI radiomics will play a more important and even irreplaceable role in the diagnosis and treatment of HCC.

AUTHOR CONTRIBUTIONS

X-QG and LY wrote the paper. Y-YT, Y-KW, NinL, XY, RW, JZ, and GY contributed to the literature search and manuscript preparation. NiaL and X-HH revised the paper. X-QW performed the data analysis and created graphs. X-MZ and J-DL designed the research. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Project of Medical Association of Sichuan Province (No. S20070) and the Project of City-University Science and Technology Strategic Cooperation of Nanchong City (North Sichuan Medical College) (No. 20SXQT0324).

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660
2. Hiley C, de Bruin EC, McGranahan N, Swanton C. Deciphering Intratumor Heterogeneity and Temporal Acquisition of Driver Events to Refine Precision Medicine. *Genome Biol* (2014) 15:453. doi: 10.1186/s13059-014-0453-8
3. Hao K, Lamb J, Zhang C, Xie T, Wang K, Zhang B, et al. Clinicopathologic and Gene Expression Parameters Predict Liver Cancer Prognosis. *BMC Cancer* (2011) 11:481. doi: 10.1186/1471-2407-11-481
4. Hao K, Luk JM, Lee NP, Mao M, Zhang C, Ferguson MD, et al. Predicting Prognosis in Hepatocellular Carcinoma After Curative Surgery With Common Clinicopathologic Parameters. *BMC Cancer* (2009) 9:389. doi: 10.1186/1471-2407-9-389
5. Hood L, Friend SH. Predictive, Personalized, Preventive, Participatory (P4) Cancer Medicine. *Nat Rev Clin Oncol* (2011) 8:184–7. doi: 10.1038/nrclinonc.2010.227
6. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
7. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006
8. Sagir KA. Radiomics in Hepatocellular Carcinoma. *J Gastrointest Cancer* (2020) 51:1165–8. doi: 10.1007/s12029-020-00493-x
9. Wei J, Jiang H, Gu D, Niu M, Fu F, Han Y, et al. Radiomics in Liver Diseases: Current Progress and Future Opportunities. *Liver Int* (2020) 40:2050–63. doi: 10.1111/liv.14555
10. Masokano IB, Liu W, Xie S, Marcellin D, Pei Y, Li W. The Application of Texture Quantification in Hepatocellular Carcinoma Using CT and MRI: A Review of Perspectives and Challenges. *Cancer Imag* (2020) 20:67. doi: 10.1186/s40644-020-00341-y
11. Wakabayashi T, Ouhmich F, Gonzalez-Cabrera C, Felli E, Saviano A, Agnus V, et al. Radiomics in Hepatocellular Carcinoma: A Quantitative Review. *Hepatol Int* (2019) 13:546–59. doi: 10.1007/s12072-019-09973-0
12. Geng Z, Zhang Y, Wang S, Li H, Zhang C, Yin S, et al. Radiomics Analysis of Susceptibility Weighted Imaging for Hepatocellular Carcinoma: Exploring the Correlation Between Histopathology and Radiomics Features. *Magn Reson Med Sci* (2020) 20:253–63. doi: 10.2463/mrms.mp.2020-0060
13. Feng M, Zhang M, Liu Y, Jiang N, Meng Q, Wang J, et al. Texture Analysis of MR Images to Identify the Differentiated Degree in Hepatocellular Carcinoma: A Retrospective Study. *BMC Cancer* (2020) 20:611. doi: 10.1186/s12885-020-07094-8
14. Chong HH, Yang L, Sheng RF, Yu YL, Wu DJ, Rao SX, et al. Multi-Scale and Multi-Parametric Radiomics of Gadoxetate Disodium-Enhanced MRI Predicts Microvascular Invasion and Outcome in Patients With Solitary Hepatocellular Carcinoma ≤ 5 Cm. *Eur Radiol* (2021) 31:4824–38. doi: 10.1007/s00330-020-07601-2
15. Chen Y, Xia Y, Tolat PP, Long L, Jiang Z, Huang Z, et al. Comparison of Conventional Gadoxetate Disodium-Enhanced MRI Features and Radiomics Signatures With Machine Learning for Diagnosing Microvascular Invasion. *AJR Am J Roentgenol* (2021) 216:1510–20. doi: 10.2214/AJR.20.23255
16. Fan Y, Yu Y, Wang X, Hu M, Du M, Guo L, et al. Texture Analysis Based on Gd-EOB-DTPA-Enhanced MRI for Identifying Vessels Encapsulating Tumor Clusters (VETC)-Positive Hepatocellular Carcinoma. *J Hepatocell Carcinoma* (2021) 8:349–59. doi: 10.2147/JHC.S293755
17. Wang W, Gu D, Wei J, Ding Y, Yang L, Zhu K, et al. A Radiomics-Based Biomarker for Cytokeratin 19 Status of Hepatocellular Carcinoma With Gadoxetic Acid-Enhanced MRI. *Eur Radiol* (2020) 30:3004–14. doi: 10.1007/s00330-019-06585-y
18. Wen L, Weng S, Yan C, Ye R, Zhu Y, Zhou L, et al. A Radiomics Nomogram for Preoperative Prediction of Early Recurrence of Small Hepatocellular Carcinoma After Surgical Resection or Radiofrequency Ablation. *Front Oncol* (2021) 11:657039. doi: 10.3389/fonc.2021.657039
19. Chong H, Gong Y, Pan X, Liu A, Chen L, Yang C, et al. Peritumoral Dilation Radiomics of Gadoxetate Disodium-Enhanced MRI Excellently Predicts Early Recurrence of Hepatocellular Carcinoma Without Macrovascular Invasion After Hepatectomy. *J Hepatocell Carcinoma* (2021) 8:545–63. doi: 10.2147/JHC.S309570
20. Zhang L, Hu J, Hou J, Jiang X, Guo L, Tian L. Radiomics-Based Model Using Gadoxetic Acid Disodium-Enhanced MR Images: Associations With Recurrence-Free Survival of Patients With Hepatocellular Carcinoma Treated by Surgical Resection. *Abdom Radiol (NY)* (2021) 46:3845–54. doi: 10.1007/s00261-021-03034-7
21. Kuang Y, Li R, Jia P, Ye W, Zhou R, Zhu R, et al. MRI-Based Radiomics: Nomograms Predicting the Short-Term Response After Transcatheter Arterial Chemoembolization (TACE) in Hepatocellular Carcinoma Patients With Diameter Less Than 5 Cm. *Abdom Radiol (NY)* (2021) 46:3772–89. doi: 10.1007/s00261-021-02992-2
22. Zhao Y, Wang N, Wu J, Zhang Q, Lin T, Yao Y, et al. Radiomics Analysis Based on Contrast-Enhanced MRI for Prediction of Therapeutic Response to Transarterial Chemoembolization in Hepatocellular Carcinoma. *Front Oncol* (2021) 11:582788. doi: 10.3389/fonc.2021.582788
23. Zhang L, Xia W, Yan ZP, Sun JH, Zhong BY, Hou ZH, et al. Deep Learning Predicts Overall Survival of Patients With Unresectable Hepatocellular Carcinoma Treated by Transarterial Chemoembolization Plus Sorafenib. *Front Oncol* (2020) 10:593292. doi: 10.3389/fonc.2020.593292
24. Huang B, Wu L, Lu XY, Xu F, Liu CF, Shen WF, et al. Small Intrahepatic Cholangiocarcinoma and Hepatocellular Carcinoma in Cirrhotic Livers May Share Similar Enhancement Patterns at Multiphase Dynamic MR Imaging. *Radiology* (2016) 281:150–7. doi: 10.1148/radiol.2016151205
25. Lambin P, Leijenaar R, Deist TM, Peerlings J, de Jong E, van Timmeren J, et al. Radiomics: The Bridge Between Medical Imaging and Personalized Medicine. *Nat Rev Clin Oncol* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
26. Saha A, Yu X, Sahoo D, Mazurowski MA. Effects of MRI Scanner Parameters on Breast Cancer Radiomics. *Expert Syst Appl* (2017) 87:384–91. doi: 10.1016/j.eswa.2017.06.029
27. Maruyama T, Hayashi N, Sato Y, Hyuga S, Wakayama Y, Watanabe H, et al. Comparison of Medical Image Classification Accuracy Among Three Machine Learning Methods. *J Xray Sci Technol* (2018) 26:885–93. doi: 10.3233/XST-18386
28. Lu L, Ehmke RC, Schwartz LH, Zhao B. Assessing Agreement Between Radiomic Features Computed for Multiple CT Imaging Settings. *PloS One* (2016) 11:e0166550. doi: 10.1371/journal.pone.0166550
29. Ger RB, Zhou S, Chi PM, Lee HJ, Layman RR, Jones AK, et al. Comprehensive Investigation on Controlling for CT Imaging Variabilities in Radiomics Studies. *Sci Rep* (2018) 8:13047. doi: 10.1038/s41598-018-31509-z
30. Polan DF, Brady SL, Kaufman RA. Tissue Segmentation of Computed Tomography Images Using a Random Forest Algorithm: A Feasibility

Study. *Phys Med Biol* (2016) 61:6553–69. doi: 10.1088/0031-9155/61/17/6553
31. Massoptier L, Casciaro S. A New Fully Automatic and Robust Algorithm for Fast Segmentation of Liver Tissue and Tumors From CT Scans. *Eur Radiol* (2008) 18:1658–65. doi: 10.1007/s00330-008-0924-y
32. Häme Y, Pollari M. Semi-Automatic Liver Tumor Segmentation With Hidden Markov Measure Field Model and Non-Parametric Distribution Estimation. *Med Image Anal* (2012) 16:140–9. doi: 10.1016/j.media.2011.06.006
33. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
34. Yang F, Ford JC, Dogan N, Padgett KR, Breto AL, Abramowitz MC, et al. Magnetic Resonance Imaging (MRI)-Based Radiomics for Prostate Cancer Radiotherapy. *Transl Androl Urol* (2018) 7:445–58. doi: 10.21037/tau.2018.06.05
35. Just N. Improving Tumour Heterogeneity MRI Assessment With Histograms. *Br J Cancer* (2014) 111:2205–13. doi: 10.1038/bjc.2014.512
36. Bagherzadeh-Khiabani F, Ramezankhani A, Azizi F, Hadaegh F, Steyerberg EW, Khalili D. A Tutorial on Variable Selection for Clinical Prediction Models: Feature Selection Methods in Data Mining Could Improve the Results. *J Clin Epidemiol* (2016) 71:76–85. doi: 10.1016/j.jclinepi.2015.10.002
37. Lewis S, Hectors S, Taouli B. Radiomics of Hepatocellular Carcinoma. *Abdom Radiol (NY)* (2021) 1:111–23. doi: 10.1007/s00261-019-02378-5
38. Liu Z, Wang S, Dong D, Wei J, Fang C, Zhou X, et al. The Applications of Radiomics in Precision Diagnosis and Treatment of Oncology: Opportunities and Challenges. *Theranostics* (2019) 9:1303–22. doi: 10.7150/thno.30309
39. Marrero JA, Kulik LM, Sirlin CB, Zhu AX, Finn RS, Abecassis MM, et al. Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases. *Hepatology* (2018) 68:723–50. doi: 10.1002/hep.29913
40. EASL Clinical Practice Guidelines. Management of Hepatocellular Carcinoma. *J Hepatol* (2018) 69:182–236. doi: 10.1016/j.jhep.2018.03.019
41. Kokudo N, Hasegawa K, Akahane M, Igaki H, Izumi N, Ichida T, et al. Evidence-Based Clinical Practice Guidelines for Hepatocellular Carcinoma: The Japan Society of Hepatology 2013 Update (3rd JSH-HCC Guidelines). *Hepatol Res* (2015) 45:123–7. doi: 10.1111/hepr.12464
42. Omata M, Cheng AL, Kokudo N, Kudo M, Lee JM, Jia J, et al. Asia-Pacific Clinical Practice Guidelines on the Management of Hepatocellular Carcinoma: A 2017 Update. *Hepatol Int* (2017) 11:317–70. doi: 10.1007/s12072-017-9799-9
43. Chernyak V, Fowler KJ, Kamaya A, Kielar AZ, Elsayes KM, Bashir MR, et al. Liver Imaging Reporting and Data System (LI-RADS) Version 2018: Imaging of Hepatocellular Carcinoma in At-Risk Patients. *Radiology* (2018) 289:816–30. doi: 10.1148/radiol.2018181494
44. Fraum TJ, Tsai R, Rohe E, Ludwig DR, Salter A, Nalbantoglu I, et al. Differentiation of Hepatocellular Carcinoma From Other Hepatic Malignancies in Patients at Risk: Diagnostic Performance of the Liver Imaging Reporting and Data System Version 2014. *Radiology* (2018) 286:158–72. doi: 10.1148/radiol.2017170114
45. Kierans AS, Makkar J, Guniganti P, Cornman-Homonoff J, Lee MJ, Pittman M, et al. Validation of Liver Imaging Reporting and Data System 2017 (LI-RADS) Criteria for Imaging Diagnosis of Hepatocellular Carcinoma. *J Magn Reson Imag* (2019) 49:e205–15. doi: 10.1002/jmri.26329
46. Jian J, Li Y, Pickhardt PJ, Xia W, He Z, Zhang R, et al. MR Image-Based Radiomics to Differentiate Type ɪ and Type ɪɪ Epithelial Ovarian Cancers. *Eur Radiol* (2021) 31:403–10. doi: 10.1007/s00330-020-07091-2
47. Song XL, Ren JL, Zhao D, Wang L, Ren H, Niu J. Radiomics Derived From Dynamic Contrast-Enhanced MRI Pharmacokinetic Protocol Features: The Value of Precision Diagnosis Ovarian Neoplasms. *Eur Radiol* (2021) 31:368–78. doi: 10.1007/s00330-020-07112-0
48. Xia W, Hu B, Li H, Geng C, Wu Q, Yang L, et al. Multiparametric-MRI-Based Radiomics Model for Differentiating Primary Central Nervous System Lymphoma From Glioblastoma: Development and Cross-Vendor Validation. *J Magn Reson Imag* (2021) 53:242–50. doi: 10.1002/jmri.27344
49. Tagliafico AS, Piana M, Schenone D, Lai R, Massone AM, Houssami N. Overview of Radiomics in Breast Cancer Diagnosis and Prognostication. *Breast* (2020) 49:74–80. doi: 10.1016/j.breast.2019.10.018
50. Mokrane FZ, Lu L, Vavasseur A, Otal P, Peron JM, Luk L, et al. Radiomics Machine-Learning Signature for Diagnosis of Hepatocellular Carcinoma in Cirrhotic Patients With Indeterminate Liver Nodules. *Eur Radiol* (2020) 30:558–70. doi: 10.1007/s00330-019-06347-w
51. Jiang H, Liu X, Chen J, Wei Y, Lee JM, Cao L, et al. Man or Machine? Prospective Comparison of the Version 2018 EASL, LI-RADS Criteria and a Radiomics Model to Diagnose Hepatocellular Carcinoma. *Cancer Imag* (2019) 19:84. doi: 10.1186/s40644-019-0266-9
52. Wu J, Liu A, Cui J, Chen A, Song Q, Xie L. Radiomics-Based Classification of Hepatocellular Carcinoma and Hepatic Haemangioma On Precontrast Magnetic Resonance Images. *BMC Med Imag* (2019) 19:23. doi: 10.1186/s12880-019-0321-9
53. Li J, Xue F, Xu X, Wang Q, Zhang X. Dynamic Contrast-Enhanced MRI Differentiates Hepatocellular Carcinoma From Hepatic Metastasis of Rectal Cancer by Extracting Pharmacokinetic Parameters and Radiomic Features. *Exp Ther Med* (2020) 20:3643–52. doi: 10.3892/etm.2020.9115
54. Liu X, Khalvati F, Namdar K, Fischer S, Lewis S, Taouli B, et al. Can Machine Learning Radiomics Provide Pre-Operative Differentiation of Combined Hepatocellular Cholangiocarcinoma From Hepatocellular Carcinoma and Cholangiocarcinoma to Inform Optimal Treatment Planning? *Eur Radiol* (2021) 31:244–55. doi: 10.1007/s00330-020-07119-7
55. Oyama A, Hiraoka Y, Obayashi I, Saikawa Y, Furui S, Shiraishi K, et al. Hepatic Tumor Classification Using Texture and Topology Analysis of Non-Contrast-Enhanced Three-Dimensional T1-Weighted MR Images With a Radiomics Approach. *Sci Rep* (2019) 9:8764. doi: 10.1038/s41598-019-45283-z
56. Ai Z, Han Q, Huang Z, Wu J, Xiang Z. The Value of Multiparametric Histogram Features Based on Intravoxel Incoherent Motion Diffusion-Weighted Imaging (IVIM-DWI) for the Differential Diagnosis of Liver Lesions. *Ann Transl Med* (2020) 8:1128. doi: 10.21037/atm-20-5109
57. Zhong X, Guan T, Tang D, Li J, Lu B, Cui S, et al. Differentiation of Small (≤ 3 Cm) Hepatocellular Carcinomas From Benign Nodules in Cirrhotic Liver: The Added Additive Value of MRI-Based Radiomics Analysis to LI-RADS Version 2018 Algorithm. *BMC Gastroenterol* (2021) 21:155. doi: 10.1186/s12876-021-01710-y
58. Liang W, Shao J, Liu W, Ruan S, Tian W, Zhang X, et al. Differentiating Hepatic Epithelioid Angiomyolipoma From Hepatocellular Carcinoma and Focal Nodular Hyperplasia *via* Radiomics Models. *Front Oncol* (2020) 10:564307. doi: 10.3389/fonc.2020.564307
59. Li Z, Mao Y, Huang W, Li H, Zhu J, Li W, et al. Texture-Based Classification of Different Single Liver Lesion Based on SPAIR T2W MRI Images. *BMC Med Imag* (2017) 17:42. doi: 10.1186/s12880-017-0212-x
60. Nakanuma Y, Xu J, Harada K, Sato Y, Sasaki M, Ikeda H, et al. Pathological Spectrum of Intrahepatic Cholangiocarcinoma Arising in Non-Biliary Chronic Advanced Liver Diseases. *Pathol Int* (2011) 61:298–305. doi: 10.1111/j.1440-1827.2011.02665.x
61. Portolani N, Baiocchi GL, Coniglio A, Piardi T, Grazioli L, Benetti A, et al. Intrahepatic Cholangiocarcinoma and Combined Hepatocellular-Cholangiocarcinoma: A Western Experience. *Ann Surg Oncol* (2008) 15:1880–90. doi: 10.1245/s10434-008-9933-y
62. Fowler KJ, Sheybani A, Parker RR, Doherty S, BE M, Chapman WC, et al. Combined Hepatocellular and Cholangiocarcinoma (Biphenotypic) Tumors: Imaging Features and Diagnostic Accuracy of Contrast-Enhanced CT and MRI. *AJR Am J Roentgenol* (2013) 201:332–9. doi: 10.2214/AJR.12.9488
63. Potretzke TA, Tan BR, Doyle MB, Brunt EM, Heiken JP, Fowler KJ. Imaging Features of Biphenotypic Primary Liver Carcinoma (Hepatocholangiocarcinoma) and the Potential to Mimic Hepatocellular Carcinoma: LI-RADS Analysis of CT and MRI Features in 61 Cases. *AJR Am J Roentgenol* (2016) 207:25–31. doi: 10.2214/AJR.15.14997
64. Sheng RF, Xie YH, Ji Y, Chen CZ, Yang L, Jin KP, et al. MR Comparative Study of Combined Hepatocellular-Cholangiocarcinoma in Normal, Fibrotic, and Cirrhotic Livers. *Abdom Radiol (NY)* (2016) 41:2102–14. doi: 10.1007/s00261-016-0811-y
65. Hamm CA, Wang CJ, Savic LJ, Ferrante M, Schobert I, Schlachter T, et al. Deep Learning for Liver Tumor Diagnosis Part I: Development of a Convolutional Neural Network Classifier for Multi-Phasic MRI. *Eur Radiol* (2019) 29:3338–47. doi: 10.1007/s00330-019-06205-9

66. Azer SA. Deep Learning With Convolutional Neural Networks for Identification of Liver Masses and Hepatocellular Carcinoma: A Systematic Review. *World J Gastrointest Oncol* (2019) 11:1218–30. doi: 10.4251/wjgo.v11.i12.1218
67. Oestmann PM, Wang CJ, Savic LJ, Hamm CA, Stark S, Schobert I, et al. Deep Learning-Assisted Differentiation of Pathologically Proven Atypical and Typical Hepatocellular Carcinoma (HCC) *Versus* Non-HCC on Contrast-Enhanced MRI of the Liver. *Eur Radiol* (2021) 31:4981–90. doi: 10.1007/s00330-020-07559-1
68. Bousabarah K, Letzen B, Tefera J, Savic L, Schobert I, Schlachter T, et al. Automated Detection and Delineation of Hepatocellular Carcinoma on Multiphasic Contrast-Enhanced MRI Using Deep Learning. *Abdom Radiol (NY)* (2021) 46:216–25. doi: 10.1007/s00261-020-02604-5
69. Kim J, Min JH, Kim SK, Shin SY, Lee MW. Detection of Hepatocellular Carcinoma in Contrast-Enhanced Magnetic Resonance Imaging Using Deep Learning Classifier: A Multi-Center Retrospective Study. *Sci Rep* (2020) 10:9458. doi: 10.1038/s41598-020-65875-4
70. Zheng R, Wang L, Wang C, Yu X, Chen W, Li Y, et al. Feasibility of Automatic Detection of Small Hepatocellular Carcinoma (≤2 Cm) in Cirrhotic Liver Based on Pattern Matching and Deep Learning. *Phys Med Biol* (2021) 66:5014. doi: 10.1088/1361-6560/abf2f8
71. Wu Y, White GM, Cornelius T, Gowdar I, Ansari MH, Supanich MP, et al. Deep Learning LI-RADS Grading System Based on Contrast Enhanced Multiphase MRI for Differentiation Between LR-3 and LR-4/LR-5 Liver Tumors. *Ann Transl Med* (2020) 8:701. doi: 10.21037/atm.2019.12.151
72. Sasaki A, Kai S, Iwashita Y, Hirano S, Ohta M, Kitano S. Microsatellite Distribution and Indication for Locoregional Therapy in Small Hepatocellular Carcinoma. *Cancer-Am Cancer Soc* (2005) 103:299–306. doi: 10.1002/cncr.20798
73. Ng IO, Lai EC, Fan ST, Ng MM, So MK. Prognostic Significance of Pathologic Features of Hepatocellular Carcinoma. A Multivariate Analysis of 278 Patients. *Cancer-Am Cancer Soc* (1995) 76:2443–8. doi: 10.1002/1097-0142(19951215)76:12<2443::aid-cncr2820761207>3.0.co;2-f
74. Okusaka T, Okada S, Ueno H, Ikeda M, Shimada K, Yamamoto J, et al. Satellite Lesions in Patients With Small Hepatocellular Carcinoma With Reference to Clinicopathologic Features. *Cancer-Am Cancer Soc* (2002) 95:1931–7. doi: 10.1002/cncr.10892
75. Bruix J, Sherman M. Management of Hepatocellular Carcinoma. *Hepatology* (2005) 42:1208–36. doi: 10.1002/hep.20933
76. Wu M, Tan H, Gao F, Hai J, Ning P, Chen J, et al. Predicting the Grade of Hepatocellular Carcinoma Based on Non-Contrast-Enhanced MRI Radiomics Signature. *Eur Radiol* (2019) 29:2802–11. doi: 10.1007/s00330-018-5787-2
77. Zhou W, Zhang L, Wang K, Chen S, Wang G, Liu Z, et al. Malignancy Characterization of Hepatocellular Carcinomas Based on Texture Analysis Of Contrast-Enhanced MR Images. *J Magn Reson Imag* (2017) 45:1476–84. doi: 10.1002/jmri.25454
78. Yang X, Yuan C, Zhang Y, Wang Z. Magnetic Resonance Radiomics Signatures for Predicting Poorly Differentiated Hepatocellular Carcinoma: A SQUIRE-Compliant Study. *Med (Baltimore)* (2021) 100:e25838. doi: 10.1097/MD.0000000000025838
79. Choi JM, Yu JS, Cho ES, Kim JH, Chung JJ. Texture Analysis of Hepatocellular Carcinoma on Magnetic Resonance Imaging: Assessment for Performance in Predicting Histopathologic Grade. *J Comput Assist Tomogr* (2020) 44:901–10. doi: 10.1097/RCT.0000000000001087
80. Zhou Q, Zhou Z, Chen C, Fan G, Chen G, Heng H, et al. Grading of Hepatocellular Carcinoma Using 3D SE-DenseNet in Dynamic Enhanced MR Images. *Comput Biol Med* (2019) 107:47–57. doi: 10.1016/j.compbiomed.2019.01.026
81. Xu YS, Liu HF, Xi DL, Li JK, Liu Z, Yan RF, et al. Whole-Lesion Histogram Analysis Metrics of the Apparent Diffusion Coefficient: A Correlation Study With Histological Grade of Hepatocellular Carcinoma. *Abdom Radiol (NY)* (2019) 44:3089–98. doi: 10.1007/s00261-019-02109-w
82. Shi G, Han X, Wang Q, Ding Y, Liu H, Zhang Y, et al. Evaluation of Multiple Prognostic Factors of Hepatocellular Carcinoma With Intra-Voxel Incoherent Motions Imaging by Extracting the Histogram Metrics. *Cancer Manag Res* (2020) 12:6019–31. doi: 10.2147/CMAR.S262973
83. Gouw AS, Balabaud C, Kusano H, Todo S, Ichida T, Kojiro M. Markers for Microvascular Invasion in Hepatocellular Carcinoma: Where Do We Stand? *Liver Transpl* (2011) 17 Suppl 2:S72–80. doi: 10.1002/lt.22368
84. Shi M, Guo RP, Lin XJ, Zhang YQ, Chen MS, Zhang CQ, et al. Partial Hepatectomy With Wide *Versus* Narrow Resection Margin for Solitary Hepatocellular Carcinoma: A Prospective Randomized Trial. *Ann Surg* (2007) 245:36–43. doi: 10.1097/01.sla.0000231758.07868.71
85. Sun JJ, Wang K, Zhang CZ, Guo WX, Shi J, Cong WM, et al. Postoperative Adjuvant Transcatheter Arterial Chemoembolization After R0 Hepatectomy Improves Outcomes of Patients Who Have Hepatocellular Carcinoma With Microvascular Invasion. *Ann Surg Oncol* (2016) 23:1344–51. doi: 10.1245/s10434-015-5008-z
86. Renzulli M, Brocchi S, Cucchetti A, Mazzotti F, Mosconi C, Sportoletti C, et al. Can Current Preoperative Imaging Be Used to Detect Microvascular Invasion of Hepatocellular Carcinoma? *Radiology* (2016) 279:432–42. doi: 10.1148/radiol.2015150998
87. Lee S, Kim SH, Lee JE, Sinn DH, Park CK. Preoperative Gadoxetic Acid-Enhanced MRI for Predicting Microvascular Invasion in Patients With Single Hepatocellular Carcinoma. *J Hepatol* (2017) 67:526–34. doi: 10.1016/j.jhep.2017.04.024
88. Ahn SY, Lee JM, Joo I, Lee ES, Lee SJ, Cheon GJ, et al. Prediction of Microvascular Invasion of Hepatocellular Carcinoma Using Gadoxetic Acid-Enhanced MR and (18)F-FDG PET/CT. *Abdom Imag* (2015) 40:843–51. doi: 10.1007/s00261-014-0256-0
89. Kim JY, Kim MJ, Kim KA, Jeong HT, Park YN. Hyperintense HCC on Hepatobiliary Phase Images of Gadoxetic Acid-Enhanced MRI: Correlation With Clinical and Pathological Features. *Eur J Radiol* (2012) 81:3877–82. doi: 10.1016/j.ejrad.2012.07.021
90. Kim KA, Kim MJ, Jeon HM, Kim KS, Choi JS, Ahn SH, et al. Prediction of Microvascular Invasion of Hepatocellular Carcinoma: Usefulness of Peritumoral Hypointensity Seen on Gadoxetate Disodium-Enhanced Hepatobiliary Phase Images. *J Magn Reson Imag* (2012) 35:629–34. doi: 10.1002/jmri.22876
91. Huang M, Liao B, Xu P, Cai H, Huang K, Dong Z, et al. Prediction of Microvascular Invasion in Hepatocellular Carcinoma: Preoperative Gd-EOB-DTPA-Dynamic Enhanced MRI and Histopathological Correlation. *Contrast Media Mol Imag* (2018) 2018:9674565. doi: 10.1155/2018/9674565
92. Meng XP, Wang YC, Zhou JY, Yu Q, Lu CQ, Xia C, et al. Comparison of MRI and CT for the Prediction of Microvascular Invasion in Solitary Hepatocellular Carcinoma Based on a Non-Radiomics and Radiomics Method: Which Imaging Modality Is Better? *J Magn Reson Imag* (2021) 54:526–36. doi: 10.1002/jmri.27575
93. Zhang Y, Shu Z, Ye Q, Chen J, Zhong J, Jiang H, et al. Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma *via* Multi-Parametric MRI Radiomics. *Front Oncol* (2021) 11:633596. doi: 10.3389/fonc.2021.633596
94. Dai H, Lu M, Huang B, Tang M, Pang T, Liao B, et al. Considerable Effects of Imaging Sequences, Feature Extraction, Feature Selection, and Classifiers on Radiomics-Based Prediction of Microvascular Invasion in Hepatocellular Carcinoma Using Magnetic Resonance Imaging. *Quant Imaging Med Surg* (2021) 11:1836–53. doi: 10.21037/qims-20-218
95. Zhang S, Xu G, Duan C, Zhou X, Wang X, Yu H, et al. Radiomics Analysis of MR Imaging With Gd-EOB-DTPA for Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma: Investigation and Comparison of Different Hepatobiliary Phase Delay Times. *BioMed Res Int* (2021) 2021:6685723. doi: 10.1155/2021/6685723
96. Tang M, Zhou Q, Huang M, Sun K, Wu T, Li X, et al. Nomogram Development and Validation to Predict Hepatocellular Carcinoma Tumor Behavior by Preoperative Gadoxetic Acid-Enhanced MRI. *Eur Radiol* (2021). doi: 10.1007/s00330-021-07941-7
97. Wilson GC, Cannella R, Fiorentini G, Shen C, Borhani A, Furlan A, et al. Texture Analysis on Preoperative Contrast-Enhanced Magnetic Resonance Imaging Identifies Microvascular Invasion in Hepatocellular Carcinoma. *HPB (Oxford)* (2020) 22:1622–30. doi: 10.1016/j.hpb.2020.03.001
98. Nebbia G, Zhang Q, Arefan D, Zhao X, Wu S. Pre-Operative Microvascular Invasion Prediction Using Multi-Parametric Liver MRI Radiomics. *J Digit Imag* (2020) 33:1376–86. doi: 10.1007/s10278-020-00353-x
99. Yang L, Gu D, Wei J, Yang C, Rao S, Wang W, et al. A Radiomics Nomogram for Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma. *Liver Cancer* (2019) 8:373–86. doi: 10.1159/000494099

100. Zhu YJ, Feng B, Wang S, Wang LM, Wu JF, Ma XH, et al. Model-Based Three-Dimensional Texture Analysis of Contrast-Enhanced Magnetic Resonance Imaging as a Potential Tool for Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma. *Oncol Lett* (2019) 18:720–32. doi: 10.3892/ol.2019.10378
101. Li H, Zhang J, Zheng Z, Guo Y, Chen M, Xie C, et al. Preoperative Histogram Analysis of Intravoxel Incoherent Motion (IVIM) for Predicting Microvascular Invasion in Patients With Single Hepatocellular Carcinoma. *Eur J Radiol* (2018) 105:65–71. doi: 10.1016/j.ejrad.2018.05.032
102. Peng J, Zhang J, Zhang Q, Xu Y, Zhou J, Liu L. A Radiomics Nomogram for Preoperative Prediction of Microvascular Invasion Risk in Hepatitis B Virus-Related Hepatocellular Carcinoma. *Diagn Interv Radiol* (2018) 24:121–7. doi: 10.5152/dir.2018.17467
103. Song D, Wang Y, Wang W, Wang Y, Cai J, Zhu K, et al. Using Deep Learning to Predict Microvascular Invasion in Hepatocellular Carcinoma Based on Dynamic Contrast-Enhanced MRI Combined With Clinical Parameters. *J Cancer Res Clin Oncol* (2021). doi: 10.1007/s00432-021-03617-3
104. Zhang Y, Lv X, Qiu J, Zhang B, Zhang L, Fang J, et al. Deep Learning With 3d Convolutional Neural Network for Noninvasive Prediction of Microvascular Invasion in Hepatocellular Carcinoma. *J Magn Reson Imag* (2021) 54:134–43. doi: 10.1002/jmri.27538
105. Wei J, Jiang H, Zeng M, Wang M, Niu M, Gu D, et al. Prediction of Microvascular Invasion in Hepatocellular Carcinoma *via* Deep Learning: A Multi-Center and Prospective Validation Study. *Cancers (Basel)* (2021) 13:2368. doi: 10.3390/cancers13102368
106. Zhou W, Jian W, Cen X, Zhang L, Guo H, Liu Z, et al. Prediction of Microvascular Invasion of Hepatocellular Carcinoma Based on Contrast-Enhanced MR and 3D Convolutional Neural Networks. *Front Oncol* (2021) 11:588010. doi: 10.3389/fonc.2021.588010
107. Zhang R, Xu L, Wen X, Zhang J, Yang P, Zhang L, et al. A Nomogram Based on Bi-Regional Radiomics Features From Multimodal Magnetic Resonance Imaging for Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma. *Quant Imaging Med Surg* (2019) 9:1503–15. doi: 10.21037/qims.2019.09.07
108. Feng ST, Jia Y, Liao B, Huang B, Zhou Q, Li X, et al. Preoperative Prediction of Microvascular Invasion in Hepatocellular Cancer: A Radiomics Model Using Gd-EOB-DTPA-Enhanced MRI. *Eur Radiol* (2019) 29:4648–59. doi: 10.1007/s00330-018-5935-8
109. Bai HX, Lee AM, Yang L, Zhang P, Davatzikos C, Maris JM, et al. Imaging Genomics in Cancer Research: Limitations and Promises. *Br J Radiol* (2016) 89:20151030. doi: 10.1259/bjr.20151030
110. Pinker K, Shitano F, Sala E, Do RK, Young RJ, Wibmer AG, et al. Background, Current Role, and Potential Applications of Radiogenomics. *J Magn Reson Imag* (2018) 47:604–20. doi: 10.1002/jmri.25870
111. Segal E, Sirlin CB, Ooi C, Adler AS, Gollub J, Chen X, et al. Decoding Global Gene Expression Programs in Liver Cancer by Noninvasive Imaging. *Nat Biotechnol* (2007) 25:675–80. doi: 10.1038/nbt1306
112. Hectors SJ, Wagner M, Bane O, Besa C, Lewis S, Remark R, et al. Quantification of Hepatocellular Carcinoma Heterogeneity With Multiparametric Magnetic Resonance Imaging. *Sci Rep* (2017) 7:2452. doi: 10.1038/s41598-017-02706-z
113. Taouli B, Hoshida Y, Kakite S, Chen X, Tan PS, Sun X, et al. Imaging-Based Surrogate Markers of Transcriptome Subclasses and Signatures in Hepatocellular Carcinoma: Preliminary Results. *Eur Radiol* (2017) 27:4472–81. doi: 10.1007/s00330-017-4844-6
114. Kitao A, Matsui O, Yoneda N, Kozaka K, Kobayashi S, Sanada J, et al. Hepatocellular Carcinoma With β-Catenin Mutation: Imaging and Pathologic Characteristics. *Radiology* (2015) 275:708–17. doi: 10.1148/radiol.14141315
115. Yoneda N, Matsui O, Kobayashi S, Kitao A, Kozaka K, Inoue D, et al. Current Status of Imaging Biomarkers Predicting the Biological Nature of Hepatocellular Carcinoma. *Jpn J Radiol* (2019) 37:191–208. doi: 10.1007/s11604-019-00817-3
116. Hectors SJ, Lewis S, Besa C, King MJ, Said D, Putra J, et al. MRI Radiomics Features Predict Immuno-Oncological Characteristics of Hepatocellular Carcinoma. *Eur Radiol* (2020) 30:3759–69. doi: 10.1007/s00330-020-06675-2
117. Gu D, Xie Y, Wei J, Li W, Ye Z, Zhu Z, et al. MRI-Based Radiomics Signature: A Potential Biomarker for Identifying Glypican 3-Positive Hepatocellular Carcinoma. *J Magn Reson Imag* (2020) 52:1679–87. doi: 10.1002/jmri.27199
118. Ye Z, Jiang H, Chen J, Liu X, Wei Y, Xia C, et al. Texture Analysis on Gadoxetic Acid Enhanced-MRI for Predicting Ki-67 Status in Hepatocellular Carcinoma: A Prospective Study. *Chin J Cancer Res* (2019) 31:806–17. doi: 10.21147/j.issn.1000-9604.2019.05.10
119. Li Y, Yan C, Weng S, Shi Z, Sun H, Chen J, et al. Texture Analysis of Multi-Phase MRI Images to Detect Expression of Ki67 in Hepatocellular Carcinoma. *Clin Radiol* (2019) 74:813.e19–27. doi: 10.1016/j.crad.2019.06.024
120. Wang HQ, Yang C, Zeng MS, Rao SX, Ji Y, Weng X, et al. Magnetic Resonance Texture Analysis for the Identification of Cytokeratin 19-Positive Hepatocellular Carcinoma. *Eur J Radiol* (2019) 117:164–70. doi: 10.1016/j.ejrad.2019.06.016
121. Fan Y, Yu Y, Wang X, Hu M, Hu C. Radiomic Analysis of Gd-EOB-DTPA-Enhanced MRI Predicts Ki-67 Expression in Hepatocellular Carcinoma. *BMC Med Imag* (2021) 21:100. doi: 10.1186/s12880-021-00633-0
122. Huang X, Long L, Wei J, Li Y, Xia Y, Zuo P, et al. Radiomics for Diagnosis of Dual-Phenotype Hepatocellular Carcinoma Using Gd-EOB-DTPA-Enhanced MRI and Patient Prognosis. *J Cancer Res Clin Oncol* (2019) 145:2995–3003. doi: 10.1007/s00432-019-03062-3
123. Chen S, Feng S, Wei J, Liu F, Li B, Li X, et al. Pretreatment Prediction of Immunoscore in Hepatocellular Cancer: A Radiomics-Based Clinical Model Based on Gd-EOB-DTPA-Enhanced MRI Imaging. *Eur Radiol* (2019) 29:4177–87. doi: 10.1007/s00330-018-5986-x
124. Zhang J, Wu Z, Zhang X, Liu S, Zhao J, Yuan F, et al. Machine Learning: An Approach to Preoperatively Predict PD-1/PD-L1 Expression and Outcome in Intrahepatic Cholangiocarcinoma Using MRI Biomarkers. *ESMO Open* (2020) 5:e000910. doi: 10.1136/esmoopen-2020-000910
125. Bruix J, Reig M, Sherman M. Evidence-Based Diagnosis, Staging, and Treatment of Patients With Hepatocellular Carcinoma. *Gastroenterology* (2016) 150:835–53. doi: 10.1053/j.gastro.2015.12.041
126. Kulik L, El-Serag HB. Epidemiology and Management of Hepatocellular Carcinoma. *Gastroenterology* (2019) 156:477–91.e1. doi: 10.1053/j.gastro.2018.08.065
127. Chen Y, Liu Z, Mo Y, Li B, Zhou Q, Peng S, et al. Prediction of Post-Hepatectomy Liver Failure in Patients With Hepatocellular Carcinoma Based on Radiomics Using Gd-EOB-DTPA-Enhanced MRI: The Liver Failure Model. *Front Oncol* (2021) 11:605296. doi: 10.3389/fonc.2021.605296
128. Zhu WS, Shi SY, Yang ZH, Song C, Shen J. Radiomics Model Based on Preoperative Gadoxetic Acid-Enhanced MRI for Predicting Liver Failure. *World J Gastroenterol* (2020) 26:1208–20. doi: 10.3748/wjg.v26.i11.1208
129. Zhang Z, Jiang H, Chen J, Wei Y, Cao L, Ye Z, et al. Hepatocellular Carcinoma: Radiomics Nomogram on Gadoxetic Acid-Enhanced MR Imaging for Early Postoperative Recurrence Prediction. *Cancer Imag* (2019) 19:22. doi: 10.1186/s40644-019-0209-5
130. Hui T, Chuah TK, Low HM, Tan CH. Predicting Early Recurrence of Hepatocellular Carcinoma With Texture Analysis of Preoperative MRI: A Radiomics Study. *Clin Radiol* (2018) 73:1056.e11–1056.e16. doi: 10.1016/j.crad.2018.07.109
131. Hu C, Song Y, Zhang J, Dai L, Tang C, Li M, et al. Preoperative Gadoxetic Acid-Enhanced MRI Based Nomogram Improves Prediction of Early HCC Recurrence After Ablation Therapy. *Front Oncol* (2021) 11:649682. doi: 10.3389/fonc.2021.649682
132. Zhao Y, Wu J, Zhang Q, Hua Z, Qi W, Wang N, et al. Radiomics Analysis Based on Multiparametric MRI for Predicting Early Recurrence in Hepatocellular Carcinoma After Partial Hepatectomy. *J Magn Reson Imag* (2021) 53:1066–79. doi: 10.1002/jmri.27424
133. Zhang J, Liu X, Zhang H, He X, Liu Y, Zhou J, et al. Texture Analysis Based on Preoperative Magnetic Resonance Imaging (MRI) and Conventional MRI Features for Predicting the Early Recurrence of Single Hepatocellular Carcinoma After Hepatectomy. *Acad Radiol* (2019) 26:1164–73. doi: 10.1016/j.acra.2018.10.011
134. Ahn SJ, Kim JH, Park SJ, Kim ST, Han JK. Hepatocellular Carcinoma: Preoperative Gadoxetic Acid-Enhanced MR Imaging Can Predict Early Recurrence After Curative Resection Using Image Features and Texture Analysis. *Abdom Radiol (NY)* (2019) 44:539–48. doi: 10.1007/s00261-018-1768-9

135. Kim S, Shin J, Kim DY, Choi GH, Kim MJ, Choi JY. Radiomics on Gadoxetic Acid-Enhanced Magnetic Resonance Imaging for Prediction of Postoperative Early and Late Recurrence of Single Hepatocellular Carcinoma. *Clin Cancer Res* (2019) 25:3847–55. doi: 10.1158/1078-0432.CCR-18-2861
136. Zhang Z, Chen J, Jiang H, Wei Y, Zhang X, Cao L, et al. Gadoxetic Acid-Enhanced MRI Radiomics Signature: Prediction of Clinical Outcome in Hepatocellular Carcinoma After Surgical Resection. *Ann Transl Med* (2020) 8:870. doi: 10.21037/atm-20-3041
137. Zhang J, Wang X, Zhang L, Yao L, Xue X, Zhang S, et al. Radiomics Predict Postoperative Survival of Patients With Primary Liver Cancer With Different Pathological Types. *Ann Transl Med* (2020) 8:820. doi: 10.21037/atm-19-4668
138. Wang XH, Long LH, Cui Y, Jia AY, Zhu XG, Wang HZ, et al. MRI-Based Radiomics Model for Preoperative Prediction of 5-Year Survival in Patients With Hepatocellular Carcinoma. *Br J Cancer* (2020) 122:978–85. doi: 10.1038/s41416-019-0706-0
139. Bargellini I, Sacco R, Bozzi E, Bertini M, Ginanni B, Romano A, et al. Transarterial Chemoembolization in Very Early and Early-Stage Hepatocellular Carcinoma Patients Excluded From Curative Treatment: A Prospective Cohort Study. *Eur J Radiol* (2012) 81:1173–8. doi: 10.1016/j.ejrad.2011.03.046
140. Song YG, Shin SW, Cho SK, Choi D, Rhim H, Lee MW, et al. Transarterial Chemoembolization as First-Line Therapy for Hepatocellular Carcinomas Infeasible for Ultrasound-Guided Radiofrequency Ablation: A Retrospective Cohort Study of 116 Patients. *Acta Radiol* (2015) 56:70–7. doi: 10.1177/0284185114520857
141. Terzi E, Golfieri R, Piscaglia F, Galassi M, Dazzi A, Leoni S, et al. Response Rate and Clinical Outcome of HCC After First and Repeated cTACE Performed "on Demand". *J Hepatol* (2012) 57:1258–67. doi: 10.1016/j.jhep.2012.07.025
142. Sun Y, Bai H, Xia W, Wang D, Zhou B, Zhao X, et al. Predicting the Outcome of Transcatheter Arterial Embolization Therapy for Unresectable Hepatocellular Carcinoma Based on Radiomics of Preoperative Multiparameter MRI. *J Magn Reson Imag* (2020) 52:1083–90. doi: 10.1002/jmri.27143
143. Song W, Yu X, Guo D, Liu H, Tang Z, Liu X, et al. MRI-Based Radiomics: Associations With the Recurrence-Free Survival of Patients With Hepatocellular Carcinoma Treated With Conventional Transcatheter Arterial Chemoembolization. *J Magn Reson Imag* (2020) 52:461–73. doi: 10.1002/jmri.26977
144. Yu JY, Zhang HP, Tang ZY, Zhou J, He XJ, Liu YY, et al. Value of Texture Analysis Based on Enhanced MRI for Predicting an Early Therapeutic Response to Transcatheter Arterial Chemoembolisation Combined With High-Intensity Focused Ultrasound Treatment in Hepatocellular Carcinoma. *Clin Radiol* (2018) 73:758.e9–18. doi: 10.1016/j.crad.2018.04.013
145. Weng W, Lü XL, Zhang QQ, Zhao XM, Chen CM, Kong CL, et al. Prediction of Short-Term Prognosis of Hepatocellular Carcinoma After TACE Surgery Based on MRI Texture Analysis Technology. *Zhonghua Yi Xue Za Zhi* (2020) 100:828–32. doi: 10.3760/cma.j.cn112137-20190705-01502
146. Kong C, Zhao Z, Chen W, Lv X, Shu G, Ye M, et al. Prediction of Tumor Response *via* a Pretreatment MRI Radiomics-Based Nomogram in HCC Treated With TACE. *Eur Radiol* (2021). doi: 10.1007/s00330-021-07910-0
147. Abajian A, Murali N, Savic LJ, Laage-Gaupp FM, Nezami N, Duncan JS, et al. Predicting Treatment Response to Intra-Arterial Therapies for Hepatocellular Carcinoma With the Use of Supervised Machine Learning-An Artificial Intelligence Concept. *J Vasc Interv Radiol* (2018) 29:850–7.e1. doi: 10.1016/j.jvir.2018.01.769
148. Liu J, Pei Y, Zhang Y, Wu Y, Liu F, Gu S. Predicting the Prognosis of Hepatocellular Carcinoma With the Treatment of Transcatheter Arterial Chemoembolization Combined With Microwave Ablation Using Pretreatment MR Imaging Texture Features. *Abdom Radiol (NY)* (2021) 46:3748–57. doi: 10.1007/s00261-020-02891-y
149. Llovet JM, Ricci S, Mazzaferro V, Hilgard P, Gane E, Blanc JF, et al. Sorafenib in Advanced Hepatocellular Carcinoma. *N Engl J Med* (2008) 359:378–90. doi: 10.1056/NEJMoa0708857
150. Kudo M, Finn RS, Qin S, Han KH, Ikeda K, Piscaglia F, et al. Lenvatinib *Versus* Sorafenib in First-Line Treatment of Patients With Unresectable Hepatocellular Carcinoma: A Randomised Phase 3 Non-Inferiority Trial. *Lancet* (2018) 391:1163–73. doi: 10.1016/S0140-6736(18)30207-1
151. Bouattour M, Mehta N, He AR, Cohen EI, Nault JC. Systemic Treatment for Advanced Hepatocellular Carcinoma. *Liver Cancer* (2019) 8:341–58. doi: 10.1159/000496439
152. Brown ZJ, Greten TF, Heinrich B. Adjuvant Treatment of Hepatocellular Carcinoma: Prospect of Immunotherapy. *Hepatology* (2019) 70:1437–42. doi: 10.1002/hep.30633
153. Rimassa L, Pressiani T, Merle P. Systemic Treatment Options in Hepatocellular Carcinoma. *Liver Cancer* (2019) 8:427–46. doi: 10.1159/000499765
154. Finn RS, Qin S, Ikeda M, Galle PR, Ducreux M, Kim TY, et al. Atezolizumab Plus Bevacizumab in Unresectable Hepatocellular Carcinoma. *N Engl J Med* (2020) 382:1894–905. doi: 10.1056/NEJMoa1915745
155. Finn RS, Zhu AX. Evolution of Systemic Therapy for Hepatocellular Carcinoma. *Hepatology* (2020) 73:150–7. doi: 10.1002/hep.31306
156. Kelley RK. Atezolizumab Plus Bevacizumab - A Landmark in Liver Cancer. *N Engl J Med* (2020) 382:1953–5. doi: 10.1056/NEJMe2004851
157. Lee MS, Ryoo BY, Hsu CH, Numata K, Stein S, Verret W, et al. Atezolizumab With or Without Bevacizumab in Unresectable Hepatocellular Carcinoma (GO30140): An Open-Label, Multicentre, Phase 1b Study. *Lancet Oncol* (2020) 21:808–20. doi: 10.1016/S1470-2045(20)30156-X
158. Liu Z, Lin Y, Zhang J, Zhang Y, Li Y, Liu Z, et al. Molecular Targeted and Immune Checkpoint Therapy for Advanced Hepatocellular Carcinoma. *J Exp Clin Cancer Res* (2019) 38:447. doi: 10.1186/s13046-019-1412-8
159. Zhao YL, Guo QQ, Yang GR, Wang QD. Early Changes in Apparent Diffusion Coefficient as an Indicator of Response to Sorafenib in Hepatocellular Carcinoma. *J Zhejiang Univ Sci B* (2014) 15:713–9. doi: 10.1631/jzus.B1400010
160. Lee Y, Lee SS, Cheong H, Lee CK, Kim N, Son WC, et al. Intravoxel Incoherent Motion MRI for Monitoring the Therapeutic Response of Hepatocellular Carcinoma to Sorafenib Treatment in Mouse Xenograft Tumor Models. *Acta Radiol* (2017) 58:1045–53. doi: 10.1177/0284185116683576
161. Nakamura Y, Kawaoka T, Higaki T, Fukumoto W, Honda Y, Iida M, et al. Hepatocellular Carcinoma Treated With Sorafenib: Arterial Tumor Perfusion in Dynamic Contrast-Enhanced CT as Early Imaging Biomarkers for Survival. *Eur J Radiol* (2018) 98:41–9. doi: 10.1016/j.ejrad.2017.10.017
162. Kim B, Kim SS, Cho SW, Cheong JY, Huh J, Kim JK, et al. Liver Stiffness in Magnetic Resonance Elastography Is Prognostic for Sorafenib-Treated Advanced Hepatocellular Carcinoma. *Eur Radiol* (2020) 31:2507–17. doi: 10.1007/s00330-020-07357-9
163. Qayyum A, Hwang KP, Stafford J, Verma A, Maru DM, Sandesh S. Immunotherapy Response Evaluation With Magnetic Resonance Elastography (MRE) in Advanced HCC. *J Immunother Cancer* (2019) 7:329. doi: 10.1186/s40425-019-0766-y
164. Yang SH, Lin J, Lu F, Han ZH, Fu CX, Lv P, et al. Evaluation of Antiangiogenic and Antiproliferative Effects of Sorafenib by Sequential Histology and Intravoxel Incoherent Motion Diffusion-Weighted Imaging in An Orthotopic Hepatocellular Carcinoma Xenograft Model. *J Magn Reson Imag* (2017) 45:270–80. doi: 10.1002/jmri.25344
165. Shirota N, Saito K, Sugimoto K, Takara K, Moriyasu F, Tokuuye K. Intravoxel Incoherent Motion MRI as a Biomarker of Sorafenib Treatment for Advanced Hepatocellular Carcinoma: A Pilot Study. *Cancer Imag* (2016) 16:1. doi: 10.1186/s40644-016-0059-3
166. Hsu CY, Shen YC, Yu CW, Hsu C, Hu FC, Hsu CH, et al. Dynamic Contrast-Enhanced Magnetic Resonance Imaging Biomarkers Predict Survival and Response in Hepatocellular Carcinoma Patients Treated With Sorafenib and Metronomic Tegafur/Uracil. *J Hepatol* (2011) 55:858–65. doi: 10.1016/j.jhep.2011.01.032
167. Salvatore V, Baron Toaldo M, Marinelli S, Milazzo M, Palamà C, Venerandi L, et al. Early Prediction of Treatment Response to Sorafenib With Elastosonography in a Mice Xenograft Model of Hepatocellular Carcinoma: A Proof-of-Concept Study. *Ultraschall Med* (2013) 34:541–9. doi: 10.1055/s-0033-1355834

168. Jiang T, Kambadakone A, Kulkarni NM, Zhu AX, Sahani DV. Monitoring Response to Antiangiogenic Treatment and Predicting Outcomes in Advanced Hepatocellular Carcinoma Using Image Biomarkers, CT Perfusion, Tumor Density, and Tumor Size (RECIST). *Invest Radiol* (2012) 47:11–7. doi: 10.1097/RLI.0b013e3182199bb5
169. Jeon MY, Lee HW, Kim BK, Park JY, Kim DY, Ahn SH, et al. Reproducibility of European Association for the Study of the Liver Criteria and Modified Response Evaluation Criteria in Solid Tumors in Patients Treated With Sorafenib. *Liver Int* (2018) 38:1655–63. doi: 10.1111/liv.13731
170. Lencioni R, Montal R, Torres F, Park JW, Decaens T, Raoul JL, et al. Objective Response by mRECIST as a Predictor and Potential Surrogate End-Point of Overall Survival in Advanced HCC. *J Hepatol* (2017) 66:1166–72. doi: 10.1016/j.jhep.2017.01.012
171. Takada J, Hidaka H, Nakazawa T, Kondo M, Numata K, Tanaka K, et al. Modified Response Evaluation Criteria in Solid Tumors Is Superior to Response Evaluation Criteria in Solid Tumors for Assessment of Responses to Sorafenib in Patients With Advanced Hepatocellular Carcinoma. *BMC Res Notes* (2015) 8:609. doi: 10.1186/s13104-015-1565-2
172. Edeline J, Boucher E, Rolland Y, Vauléon E, Pracht M, Perrin C, et al. Comparison of Tumor Response by Response Evaluation Criteria in Solid Tumors (RECIST) and Modified RECIST in Patients Treated With Sorafenib for Hepatocellular Carcinoma. *Cancer-Am Cancer Soc* (2012) 118:147–56. doi: 10.1002/cncr.26255
173. Yuan G, Song Y, Li Q, Hu X, Zang M, Dai W, et al. Development and Validation of a Contrast-Enhanced CT-Based Radiomics Nomogram for Prediction of Therapeutic Efficacy of Anti-PD-1 Antibodies in Advanced HCC Patients. *Front Immunol* (2020) 11:613946. doi: 10.3389/fimmu.2020.613946
174. Mulé S, Thiefin G, Costentin C, Durot C, Rahmouni A, Luciani A, et al. Advanced Hepatocellular Carcinoma: Pretreatment Contrast-Enhanced CT Texture Parameters as Predictive Biomarkers of Survival in Patients Treated With Sorafenib. *Radiology* (2018) 288:445–55. doi: 10.1148/radiol.2018171320
175. Aerts HJ. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. *JAMA Oncol* (2016) 2:1636–42. doi: 10.1001/jamaoncol.2016.2631
176. Galldiks N, Kocher M, Ceccon G, Werner JM, Brunn A, Deckert M, et al. Imaging Challenges of Immunotherapy and Targeted Therapy in Patients With Brain Metastases: Response, Progression, and Pseudoprogression. *Neuro Oncol* (2020) 22:17–30. doi: 10.1093/neuonc/noz147
177. Aslan K, Turco V, Blobner J, Sonner JK, Liuzzi AR, Núñez NG, et al. Heterogeneity of Response to Immune Checkpoint Blockade in Hypermutated Experimental Gliomas. *Nat Commun* (2020) 11:931. doi: 10.1038/s41467-020-14642-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Gong, Tao, Wu, Liu, Yu, Wang, Zheng, Liu, Huang, Li, Yang, Wei, Yang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 24 September 2021
doi: 10.3389/fonc.2021.739933

Prediction of Locoregional Recurrence-Free Survival of Oesophageal Squamous Cell Carcinoma After Chemoradiotherapy Based on an Enhanced CT-Based Radiomics Model

Jie Kong[1], Shuchai Zhu[1], Gaofeng Shi[2], Zhikun Liu[1], Jun Zhang[1] and Jialiang Ren[3]*

[1] Department of Radiation Oncology, The Fourth Hospital of Hebei Medical University, Shijiazhuang, China, [2] Department of Computed Tomography and Magnetic Resonance, The Fourth Hospital of Hebei Medical University, Shijiazhuang, China, [3] Pharmaceutical Diagnosis, GE Healthcare, Beijing, China

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln, United Kingdom

Reviewed by:
Liheng Liu,
Fudan University, China
Bin Song,
Fudan University, China

***Correspondence:**
Shuchai Zhu
sczhu1965@163.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *12 July 2021*
Accepted: *06 September 2021*
Published: *24 September 2021*

Citation:
Kong J, Zhu S, Shi G, Liu Z, Zhang J and Ren J (2021) Prediction of Locoregional Recurrence-Free Survival of Oesophageal Squamous Cell Carcinoma After Chemoradiotherapy Based on an Enhanced CT-Based Radiomics Model. Front. Oncol. 11:739933. doi: 10.3389/fonc.2021.739933

Background and Purpose: Chemoradiotherapy is the standard treatment for moderate and advanced oesophageal cancer. The aim of this study was to establish a predictive model based on enhanced computed tomography examination, and to evaluate its clinical value for detecting locoregional recurrence-free survival (LRFS) in cases of oesophageal squamous cell carcinoma after radiotherapy.

Materials and Methods: In total, 218 patients with pathologically diagnosed oesophageal squamous cell carcinoma who received radical chemoradiotherapy from July 2016 to December 2017 were collected in this study. Patients were randomly divided into either a training group (n=153) or a validation group (n=65) in a 7:3 ratio. Clinical patient information was then recorded. The enhanced computed tomography scan images of the patients were imported into 3D-slicer software (version 4.8.1), and the radiomic features were extracted by the Python programme package. In the training group, the dimensionality reduction of the radiomic features was implemented by Lasso regression, and then a radiological label, the model of predicting LRFS, was established and evaluated. To achieve a better prediction performance, the radiological label was combined with clinical risk factor information to construct a radiomics nomogram. A receiver operating characteristic curve was used to evaluate the efficacy of different models. Calibration curves were used to assess the consistency between the predicted and observed recurrence risk, and the Hosmer-Lemeshow method was used to test model fitness. The C-index evaluated the discriminating ability of the prediction model. Decision curve analysis was used to determine the clinical value of the constructed prediction model.

Results: Of the 218 patients followed up in this study, 44 patients (28.8%) in the training group and 21 patients (32.3%) in the validation group experienced recurrence. There was no difference in LRFS between the two groups (χ^2 = 0.525, P=0.405). Lasso regression

was used in the training group to select six significant radiomic features. The radiological label established using these six features had a satisfactory prediction performance. The C-index was 0.716 (95% CI: 0.645–0.787) in the training group and 0.718 (95% CI: 0.612–0.825) in the validation group. The radiomics nomogram, which included the radiological label and clinical risk factors, achieved a better prediction than the radiological label alone. The C-index was 0.742 (95% CI: 0.674–0.810) in the training group and 0.715 (95% CI: 0.609–0.820) in the validation group. The results of the calibration curve and decision curve analyses indicated that the radiomics nomogram was superior in predicting LRFS of oesophageal carcinoma after radiotherapy.

Conclusions: A radiological label was successfully established to predict the LRFS of oesophageal squamous cell carcinoma after radiotherapy. The radiomics nomogram was complementary to the clinical prognostic features and could improve the prediction of the LRFS after radiotherapy for oesophageal cancer.

Keywords: oesophageal squamous cell carcinoma, chemoradiotherapy, radiomics, enhanced CT, locoregional recurrence-free survival

INTRODUCTION

Oesophageal squamous cell carcinoma (ESCC) is the major histological subtype of oesophageal cancer, especially in high-incidence areas such as China (1–3). Most patients diagnosed with locally advanced oesophageal cancers lose the opportunity for surgery at the time of diagnosis; instead, the standard treatment of concurrent chemoradiotherapy is recommended (4). However, local recurrence is still the main cause of treatment failure (5). Once recurrence occurs, the patient's prognosis is usually poor, with a reported survival time of 3–10 months (6, 7). Reducing the recurrence of oesophageal cancer is an urgent problem to be solved in clinical practice (8). For patients receiving concurrent chemoradiotherapy, the clinical stage is the most important factor for prognosis. However, due to tumour heterogeneity, large differences in recurrence occur even among patients at the same stage (9). Accurate quantification of the inherent heterogeneity of oesophageal cancer and a search for the factors affecting recurrence are crucial to improve local control and prolong survival (10). This will allow clinicians to adopt more active treatments (radiotherapy combined with targeted therapy, immunotherapy, etc.) in high-risk patients.

Radiomics technology is a great innovation in the field of medical image analysis (11). It extracts quantitative features from computed tomography (CT), magnetic resonance, and positron emission tomography images, among other modalities, to locate the changes in the internal characteristics of a tumour. The technology then combines these changes with the biological behaviour of the tumours to reach a precise diagnosis and treatment. Moreover, many studies have focused on predicting therapeutic efficacy (12–15), but few have investigated the prediction of locoregional recurrence-free survival (LRFS) in patients post-radiotherapy.

To explore the relationship between clinical characteristics, radiomic features, and local control after radiotherapy, we performed an enhanced CT scan before and after radiotherapy to select the factors with clinical value. We then constructed a model based on the radiomic characteristics combined with the clinical parameters. The model was used to predict the LRFS of ESCC patients receiving radiotherapy and to provide a theoretical reference for treatment.

MATERIALS AND METHODS

Patients

Patients with pathologically confirmed ESCC and a Karnofsky Performance Status ≥70 who were receiving three-dimensional conformal or intensity-modulated radiotherapy were included in this study. Patients were excluded if they met the following criteria: 1) distant metastasis; 2) low-dose palliative radiotherapy; 3) preoperative or postoperative adjuvant radiotherapy; 4) incomplete clinicopathological information; 5) oesophageal fistula and oesophageal stent implantation; 6) image artifacts or tumour volumes are too small to be recognised on CT images, resulting in poor visualisation quality; or 7) previous malignant tumour history.

Ultimately, 218 patients who received three-dimensional conformal or intensity-modulated radiotherapy at the Fourth Hospital of Hebei Medical University from July 2016 to December 2017 were enrolled. The median age was 67 years (37–84 years) and the median lesion length was 5.0 cm (1.0–12.0 cm). All patients received electronic gastroscopy, oesophageal barium meal contrast, chest enhanced CT scan and abdominal ultrasound or CT examination before treatment, according to the eighth edition of the American Joint Committee on Cancer staging criteria (16).

Radiotherapy

Gross tumour volume included the primary oesophageal tumour and regional lymph nodes. The criteria for determining

oesophageal lesions on CT images were oesophageal wall thickness >5 mm or non-airless oesophagus diameter >10 mm, localised or whole oesophageal wall thickening, and/or local lumen stenosis. The clinical target volume (CTV) was obtained by expanding the GTV to a margin to 2.0-3.0 cm at the long axis and 0.5 cm at the lateral axis. The planning target volume (PTV) was reached by CTV plus a margin of 0.5 cm. The prescription dose for the whole group was 50.0–66.0 Gy, the median dose was 60.0 Gy, and a single dose was 1.8–2.2 Gy.

Chemotherapy

A total of 90 patients received 1–2 cycles of concurrent chemotherapy, with the main regimens of FP (cisplatin, 12.5 mg/m^2×5 days or 25 mg/m^2×3 days; 5-fluorouracil, 450 to 500 mg/m^2×5 days) or TP (paclitaxel, 135 mg/m^2, d1,8 days; cisplatin, 25 mg/m^2, d2, 3, 4 days, 28 days as a cycle, then 1, 5 weeks of administration).

CT Image Acquisition

CT images were collected before and within 1 month after chemoradiotherapy. All patients underwent standard chest contrast-enhanced CT scanning with a CT scanner (SOMATOM Definition Flash CT, SOMATOM Sensation Open CT, Forchheim, Germany). Scan parameters were as follows: tube voltage, 120 kV; tube current, 110 mA; scanning matrix, 512 x 512; conventional scanning layer thickness, 5.0 mm; reconstruction layer thickness, 1.0 mm; mediastinal window width, 350 HU; window position, 40 HU; lung window width, 1200 HU; and window position, -600 HU. In this study, enhanced CT images were used for tumour delineation and feature extraction.

CT Image Segmentation

The ROI profiling process is shown in **Supplementary Figure 1**, Arterial-phase CT images of 218 patients which were retrieved from PACS(Carestream) were imported into the 3D Slicer software (version 4.8.1, http://www.slicer.org), The tumours were manually segmented slice by slice using the software. and an attending physician with more than 5 years of clinical experience independently outlined the region of interest (ROI) of the oesophageal primary tumours. The lesion was considered to be tumours when the oesophageal wall showed focal thickening of ≥5 mm on imaging. Intraluminal air and contrast agent, fatty tissues, tumour necrosis surrounding the lesion, and blood vessels near the gross tumour were removed from the ROI, defined as an area with attenuation values below -50 HU and over 300 HU. The attending physician sketched all tumour ROIs, and the associate chief physician randomly selected 40 cases of sketched tumour ROIs for a consistency test.

Radiomic Feature Extraction and Selection

Radiomic features of the segmented 3D images were extracted using the Python programme package Pyradiomics 1.2.0.(Amsterdam Netherlands). A total of seven categories of imaging features were collected in this study. This included 18 first-order, 14 shape-based histogram, 24 grey level co-occurrence matrix, 16 grey level size zone matrix, 16 grey level run length matrix, 14 grayscale dependence matrix, and 5 neighbourhood grey-tone difference matrix features (17).

The intergroup correlation coefficient (ICC) was used to analyse the consistency of the radiomic features extracted from the ROI of the tumours in the training group. The features with good reproducibility (ICC>0.75) were selected. First, Spearman correlation analysis was performed for any two feature columns. R>0.9 indicated that the two features were highly correlated, and the features with large correlation coefficients with LRFS were retained. Second, the most useful predictive features were selected using the least absolute shrinkage and selection operator (LASSO) Cox regression model, which was applied to reduce high-dimensional data. Ten-fold cross-validation was used in the parameter tuning phase of the LASSO algorithm to extract the effective and predictive features.

Construction of the Radiological Label (Rad-Score) and Radiomics Nomogram

After the imaging features were screened by means of dimensionality reduction, Cox regression was used to calculate the regression coefficient (β). The weighted linear formula was as follows:

$$rad - score = \beta 0 + \beta 1x1 + \beta 2x2 + \beta 3x3 + \ldots\ldots\ldots + \beta nXn\,.$$

For a better prediction effect, a multivariate Cox proportional hazards model was utilised to build a radiomics nomogram. The nomogram was constructed by combining the radiomic signature with the mentioned conventional clinical parameters to determine the model with the optimal predictive performance.

Validation of the Radiological Label (Rad-Score) and Radiomics Nomogram

The efficacy of the radiological label in predicting post-radiotherapy LRFS was determined using the receiver operating characteristic curve. The C-index was used to evaluate the discrimination power, which was defined as the agreement between the predicted and actual RFS probabilities; decision curve analysis was then used to determine the clinical value of the constructed prediction model. Higher clinical utility was observed the farther away the decision curve was from the two extreme curves (treat-all and treat-none). In addition, we used the calibration curve to assess the predictive accuracy and the agreement between the actual and predicted RFS. The Akaike information criterion (AIC) was used to assess the probability of overfitting; a smaller AIC value indicated a better fit of the model.

Moreover, we evaluated the association between the radiological label and LRFS of oesophageal cancer by using Kaplan-Meier analysis in the training group. The optimal cut-off point was calculated to maximise the selection of rank statistics using X-tile software(New Haven USA). Then, the patients with radiological label values above the cut-off point were allocated into a high-risk group, while those with values below the cut-off point were allocated into the low-risk group. The log-rank test was used to measure the difference in survival curves between the two groups. We then performed the same analysis for the radiomics nomogram.

Statistical Analysis

Statistical analysis was performed using R software (version 3.4.4) and SPSS version 25.0 (IBM). Comparisons of patient characteristics were performed using the Mann-Whitney U-test or two-sample t-test. Univariate analysis used the Kaplan-Meier method to calculate and compare the LRFS of different groups of oesophageal cancer patients. A Cox proportional hazards model was utilised to screen the independent influencing factors of LRFS. P<0.05 was statistically significant.

RESULTS

Analysis of the General Clinical Characteristics of Patients

In total, 218 patients (149 males, 69 females) with ESCC were enrolled in the study, with a median age of 67.0 (37.0–84.0) years. Patients were divided into either training or validation groups in a 7:3 ratio. There were 153 patients in the training group, including 106 males and 47 females, with a median age of 67.0 (37.0–81.0) years. The validation group consisted of 65 patients, including 43 males and 22 females, with a median age of 69.0 (46.0–84.0) years. There were no differences in the clinical characteristics between the two groups; P-values ranged from 0.051 to 0.982 (**Table 1**).

TABLE 1 | The distribution of general clinical factors in the training cohort and validation cohort.

Factors	Training cohort (n = 153)	Validation cohort (n = 65)	X^2/t	P
Age (years)	65.98 ± 8.73	67.52 ± 8.41	1.206	0.229
Gender				
Male	105	44	0.018	0.982
Female	48	21		
Location				
Cervical	11	0	5.349	0.148
Upper	41	16		
Middle	77	37		
Lower	24	12		
Length (cm)	5.35 ± 1.95	5.29 ± 2.70	0.211	0.833
Maximum tumour wall thicknesses (Pre-RT) (cm)	1.43 ± 0.47	1.45 ± 0.32	0.377	0.706
Maximum tumour wall thicknesses (Post-RT) (cm)	1.09 ± 0.43	1.45 ± 0.36	1.065	0.288
T stage				
T1-3	118	44	2.126	0.145
T4	35	21		
Tracheal invasion				
No	121	49	0.364	0.546
Yes	32	16		
Prevertebral soft tissue invasion				
No	144	58	1.602	0.206
Yes	9	7		
Aortic invasion				
No	144	56	3.819	0.051
Yes	9	9		
Supraclavicular lymph node metastasis				
Yes	32	14	0.011	0.918
No	121	51		
N stage				
N0	14	7	1.727	0.631
N1	46	24		
N2	63	21		
N3	30	13		
TNM stage				
I-III	97	37	0.808	0.369
IVa	56	28		
Chemotherapy				
Yes	94	38	0.169	0.681
No	59	27		
Chemoradiotherapy				
Yes	67	23	1.330	0.249
No	86	42		
Radiation dose (Gy)				
<60	53	22	1.707	0.426
60	49	16		
>60	51	27		
Short-term clinical effect				
CR	44	18	0.025	0.873
Non-CR	109	47		

Locoregional Recurrence-Free Survival of Patients

All patients with oesophageal cancer were followed up for the full length of the follow-up period. By the end of the last follow-up visit, 44 patients (28.8%) in the training group and 21 patients (32.3%) in the validation group had recurrence. The LRFS of patients in the training group at 1, 2 and 3 years was 77.8%, 71.2%, and 71.2%, and that of patients in the validation group was 75.4%, 66.2%, and 66.2%, respectively. There was no statistically significant difference in LRFS between the two groups (χ^2 = 0.525, P=0.405). Univariate analysis of the data in the training group showed that the length of the lesion, the maximum layer wall thickness of the lesion before and after radiotherapy, aortic invasion, TNM stage, and the short-term efficacy after radiotherapy were correlated with LRFS after radiotherapy (P<0.05). The detailed results are listed in **Table 2**.

Selection of Radiomic Features Associated With Local Recurrence of Oesophageal Carcinoma

Of the 1037 radiomic features extracted from the CT images, 654 had an ICC value>0.75, indicating high reproducibility. Spearman correlation analysis was used to remove features with correlation coefficients greater than 0.9. Basing on the LASSO Cox regression model for LRFS, six radiomic features, including three before and three after radiotherapy, were selected. The correlation coefficients between the screened radiomics features are shown in **Figure 1**.

Construction and Validation of the Radiological Label

A radiological label was established to predict LRFS in patients with oesophageal cancer after radiotherapy. The formula was expressed as follows:

TABLE 2 | Relationship between general clinical characteristics and LRFS in patients with oesophageal cancer.

Factors	Training cohort LRFS (%) (n = 153)			P	Validation cohort LRFS (%) (n = 65)			P
	1-year	2-year	3-year		1-year	2-year	3-year	
Gender								
Male	75.2	69.5	69.5	0.423	71.4	71.4	71.4	0.649
Female	83.3	75.0	75.0		77.3	68.2	63.6	
Location								
Cervical	63.6	63.6	63.6	0.401	–	–	–	0.140
Upper	87.8	78.0	78.0		93.5	87.5	87.5	
Middle	72.7	66.2	66.2		73.2	64.9	59.5	
Lower	83.3	79.2	79.2		58.3	58.3	58.3	
Lesion length								
≤5 cm	85.9	77.6	77.6	0.035	78.4	73.0	73.0	0.235
>5 cm	70.6	63.2	63.2		71.4	60.7	57.1	
Maximum tumour wall thicknesses (Pre-RT)								
≤1.5 cm	84.4	76.7	76.7	0.042	79.5	69.2	64.1	0.724
>1.5 cm	68.3	63.5	63.5		69.2	69.2	69.2	
Maximum tumour wall thicknesses (Post-RT)								
≤1.5 cm	80.7	73.3	73.3	0.048	79.2	71.7	71.7	0.540
>1.5 cm	55.6	55.6	55.6		60.0	60.0	60.0	
T stage								
T1-3	80.5	74.6	74.6	0.099	79.5	72.7	68.2	0.449
T4	68.6	60	60		66.7	61.9	61.9	
Tracheal invasion								
No	79.3	73.6	73.6	0.267	77.6	71.4	67.3	0.533
Yes	71.9	62.5	62.5		68.8	62.5	62.5	
Prevertebral soft tissue invasion								
No	77.8	70.8	70.8	0.691	85.7	71.4	71.4	0.784
Yes	77.8	77.8	77.8		74.1	69	65.5	
Aortic invasion								
No	79.9	72.9	72.9	0.032	76.8	69.6	66.1	0.938
Yes	44.4	44.4	44.4		66.7	66.7	66.7	
Supraclavicular lymph node metastasis								
Yes	74.4	70.1	70.1	0.537	70.8	62.5	58.3	0.036
No	88.6	74.3	74.3		88.2	88.2	88.2	
N stage								
N0	85.7	85.7	85.7	0.135	71.4	42.9	28.6	0.093
N1	80.4	73.9	73.9		70.8	62.5	62.5	
N2	81	73	73		71.4	71.4	71.4	
N3	63.3	56.7	56.7		92.3	84.6	94.6	
TNM stage								
I-III	83.5	77.3	77.3	0.025	78.4	70.3	67.6	0.606
IVa	67.9	60.7	60.7		71.4	67.9	64.3	
Albumin level (Pre-RT) (g/L)								
High (≥40)	72.4	67.1	67.1	0.259	69.7	63.6	60.6	0.333
Low (<40)	82.7	74.7	74.7		80	73.3	73.3	
Chemotherapy								
Yes	79.7	72.9	72.9	0.728	73.7	71.1	68.4	0.668
No	76.6	70.2	70.2		77.8	66.7	63	
Chemoradiotherapy								
Yes	79.1	73.3	73.3	0.573	65.2	60.9	60.9	0.408
No	76.1	68.7	68.7		81	73.8	69	
Radiation dose(Gy)								
<60	77.4	75.5	75.5	0.512	68.2	63.6	59.1	0.401
60	81.6	73.5	73.5		62.5	62.5	62.5	
>60	78.4	64.7	64.7		88.9	77.8	74.1	
Short-term clinical effect								
CR	88.6	86.4	86.4	0.01	88.9	83.3	77.8	0.194
Non-CR	73.4	65.1	65.1		70.2	63.8	61.7	

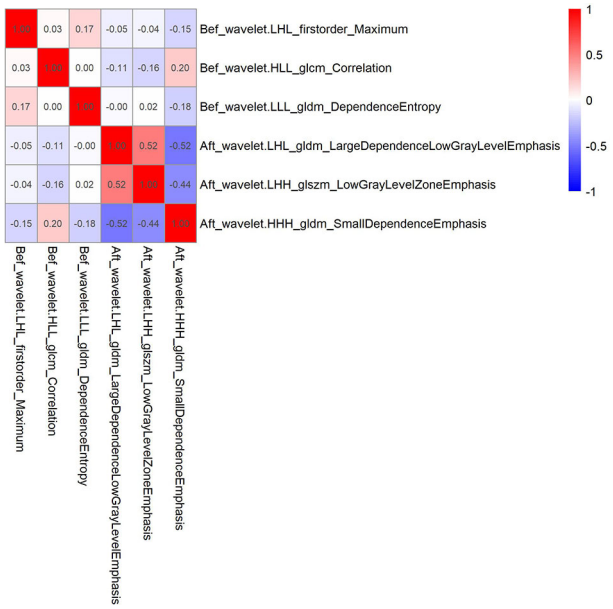


FIGURE 1 | Correlation coefficient between 6 radiomics features screened by Lasso regression in the training cohort.

$$
\begin{aligned}
Rad-score = & -0.506 * Bef_{wavelet} \cdot LHL_{firstorder_{maximum}} \\
& - 0.338 * Bef_{wavelet} \cdot HLL_{glcm_{Correlation}} \\
& + 0.213 * Bef_{wavelet} \cdot LLL_{gldm_{DependenceEntropy}} \\
& - 0.621 * Aft_{wavelet} \cdot LHL_{gldm_{LargeDependenceLowGrayLevel^{Emphasis}}} \\
& - 0.40 * Aft_wavelet \\
& \cdot LHH_glszm_LowGrayLevelZoneEmphasis \\
& - 0.86 * Aft_wavelet \\
& \cdot HHH_gldm_SmallDependenceEmphasis
\end{aligned}
$$

The area under the curve of the radiological label for predicting LRFS after radiotherapy was 0.767 (95% CI: 0.688–0.846) in the training group and 0.728 (95% CI: 0.601–0.856) in the validation group. The C-index of the radiological labels in the training and validation groups were 0.716 (95% CI: 0.645–0.787) and 0.718 (95% CI: 0.612–0.825), respectively. The results of the calibration curves suggest that the radiological labels had a high goodness of fit (P>0.05) (**Figure 2**).

The optimal cut-off point of the radiological label was 0.13, as generated by the X-tile software. Patients were divided into two groups: patients with radiomics signature ≥0.13 were classified as the high-risk group, and those with a score <0.13 were classified as the low-risk group. The distributions of radiomic signature values calculated from the training and validation groups are presented in **Figure 3**.

The 1-, 2- and 3-year LRFS of patients in the low-risk group were 94.36%, 91.55%, and 91.55%, and those in the high-risk group were 65.85%, 53.66%, and 53.66%, respectively. The difference between groups was statistically significant (HR: 6.933 [2.926–16.425], P<0.0001) (**Figure 4A**). Similar results were obtained in the validation group (HR: 2.982 [1.006–8.818], P=0.037) (**Figure 4B**).

Construction and Validation of the Radiomics Nomogram

Cox proportional hazard regression analysis showed that lesion length, TNM stage, and radiological label were independent predictors of LRFS after radiotherapy for oesophageal cancer. Based on the above results, we built a radiomics nomogram with the formula:

$$0.458 * \text{lesion length} + 0.525 * \text{TNM stage}$$

$$+ 0.946 * \text{radiological label}$$

The area under the receiver operating characteristic curve was 0.790 (95% CI: 0.712–0.867) in the training group and 0.727 (95% CI: 0.598–0.856) in the validation group. The calibration curve suggested there was a high goodness of fit between the predictions of the radiomics nomogram and the actual probabilities of 1, 2, and 3-year LRFS (**Figure 5**). The C-index and AIC estimates for the different models are listed in **Table 3**. The C-index values of the radiomics nomogram were 0.742 (95% CI: 0.674–0.810) in the training cohort and 0.715 (95% CI: 0.609–0.820) in the validation cohort. Of all the models, this nomogram had the lowest AIC value in both the training group and the validation group. These results suggested that the predictive power of the radiomics nomogram for LRFS was significantly improved when compared with each feature set alone. Decision curve analysis showed that the radiomics

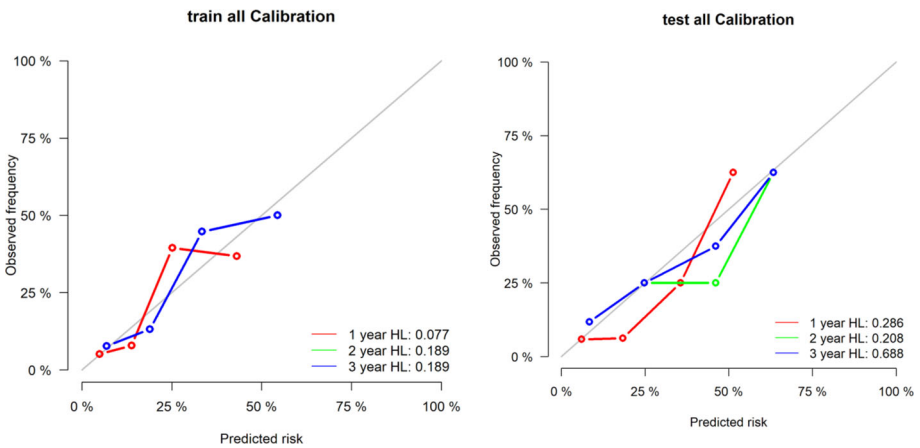


FIGURE 2 | Calibration curves of the radiological label were plotted to assess the agreement between LRFS predicted by the model and the observed LRFS.

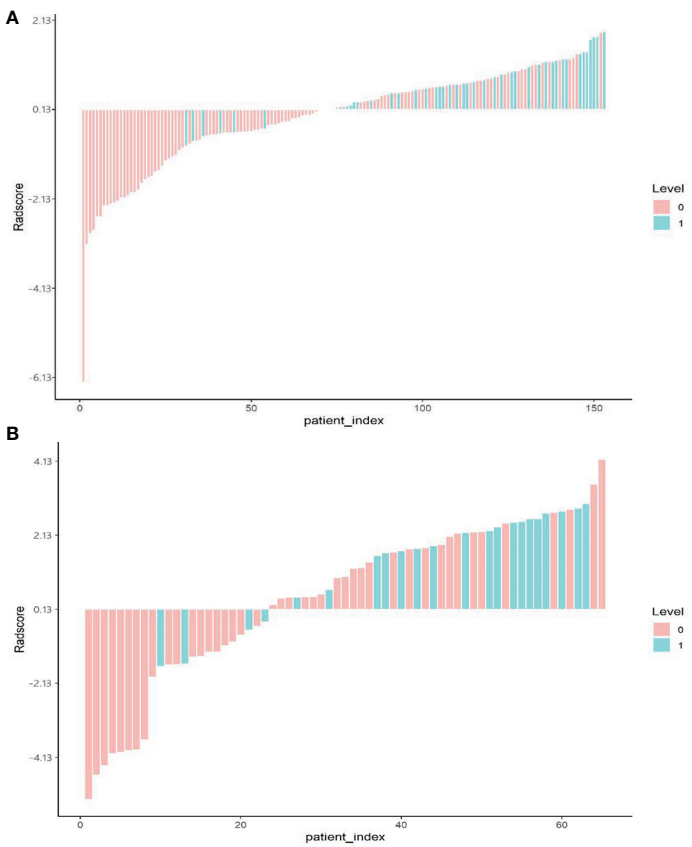


FIGURE 3 | Bar plot of the radiomics signature value for each patient in the training cohort **(A)** and the validation cohort **(B).** 0, good control (red); 1, local uncontrolled or recurrent (green).

nomogram produced a greater net benefit than the other predictive models, as presented in **Figure 5**.

The radiomics nomogram yielded an optimal cut-off point of 0.56. Patients were divided into high-risk (≥0.56) and low-risk (<0.56) groups accordingly. The 1-, 2- and 3-year LRFS of patients in the low-risk group were 90.19%, 85.29%, and 85.29%, and those in the high-risk group were 56.86%, 43.14%, and 43.14%, respectively. The difference was statistically

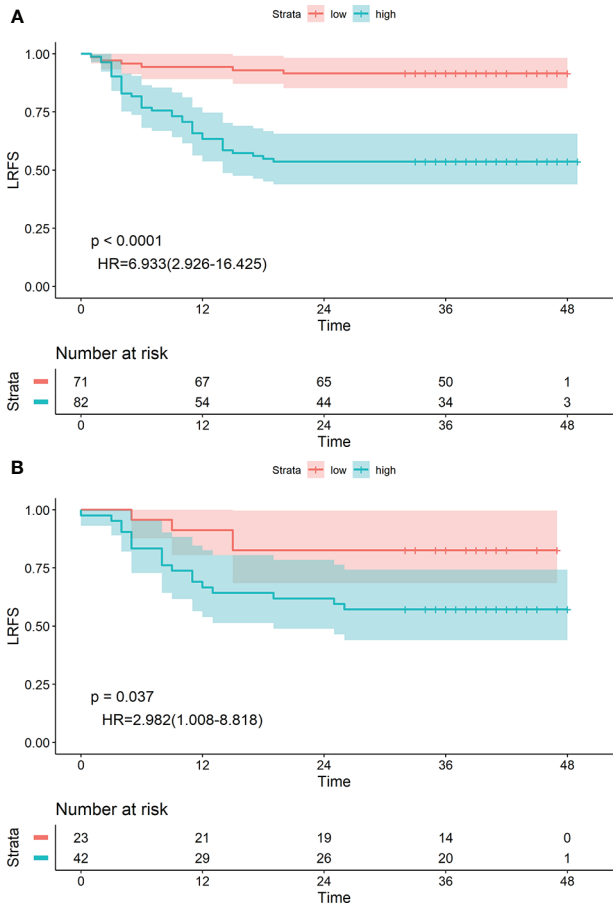


FIGURE 4 | Kaplan-Meier survival analyses of high-risk and low-risk groups divided by radiological label in the training cohort **(A)** and validation cohort **(B)**.

significant [HR: 5.119 (2.735–9.580), P<0.0001] (**Figure 6A**). Similar conclusions were reached in the validation group [HR: 2.829 (1.152–6.947), P=0.018] (**Figure 6B**).

DISCUSSION

As well known, chemoradiotherapy was the standard treatment for the patients of local advanced oesophageal cancer (12). However, a high local recurrence rate still remains (15), and once recurrence was diagnosed, the median OS was 1 year (18), Many studies have suggested that tumour size, degree of differentiation and lymphovascular invasion were the independent factors in predicting the recurrence of oesophageal cancer after treatment (6, 19, 20), but these studies mainly focused on patients who received surgical treatment, and were not applicable to patients receiving radical chemoradiotherapy. Therefore, there is a need to find more effective markers to predict the LRFS of oesophageal cancer after radiotherapy, as this can help us to identify patients with a high recurrence risk to take more active intervention measures.

In clinical practice, CT plays an important role in the diagnosis, staging, efficacy evaluation, and prognosis monitoring of oesophageal cancer (21). However, images obtained in clinical applications cannot capture the underlying tumour characteristics. The value of the our study is that it proposes a novel method for predicting LRFS of patients with oesophageal cancer after chemoradiotherapy. The results of the study showed that radiological label, a non-invasive, quantitative, and low-cost parameter, was an independent factor in predicting LRFS. we concluded that radiological label combined with clinical features can improve the predictive ability. This approach is expected to identify people at high risk of recurrence and support decision-making in clinical treatment for patients with oesophageal cancer.

In this study, we established a radiological label using enhanced CT images before and after radiotherapy for predicting LRFS in patients with oesophageal cancer. The model contains six radiomic features, five of which are textural features, which provide information about tumour heterogeneity for the assessment of the local oesophageal cancer recurrence. Many studies have explored the value of radiomics in clinical practice (22–28). The results of one study, carried out by

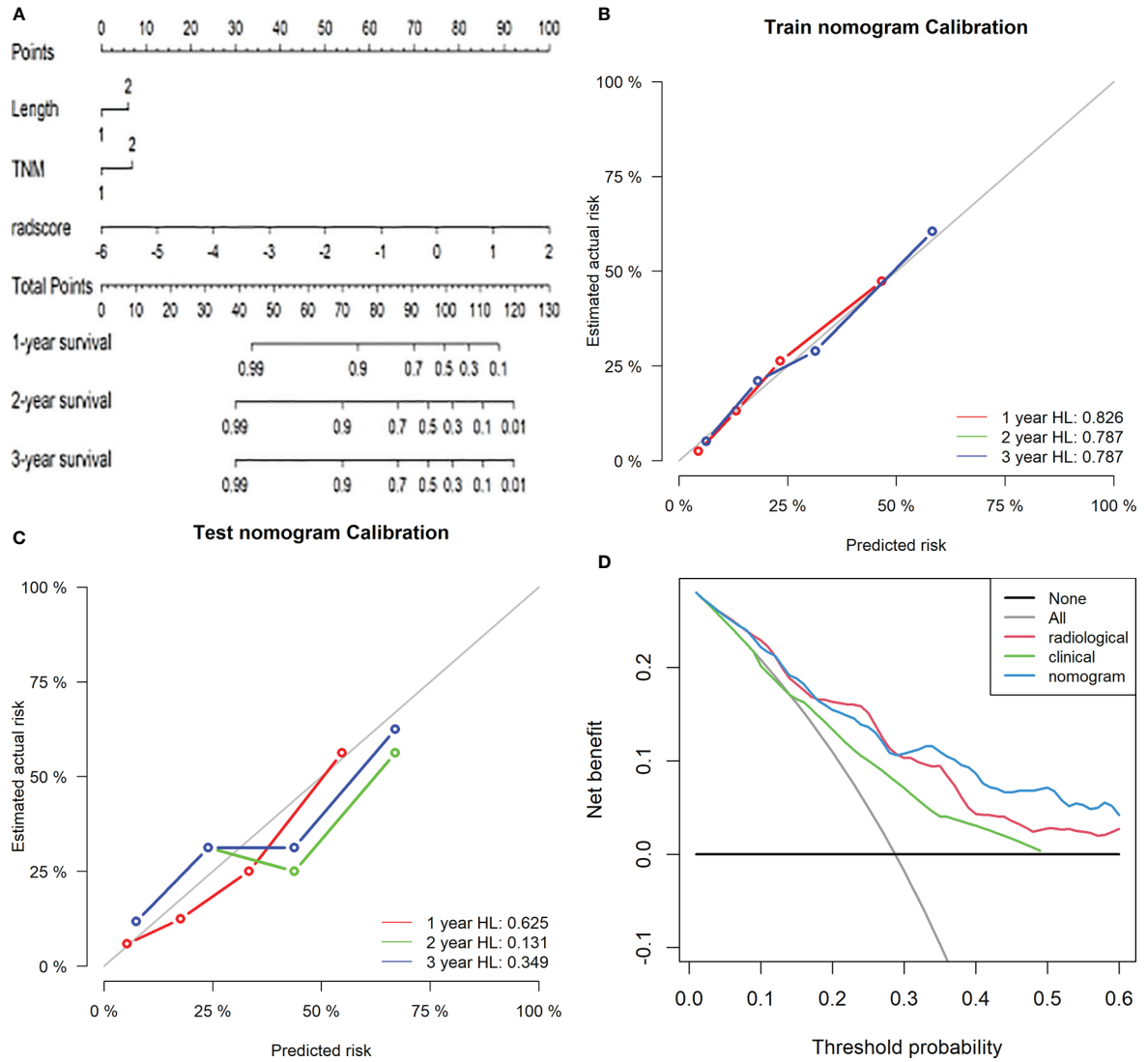


FIGURE 5 | Radiomics nomogram for predicting LRFS of ESCC patients after radiotherapy. Length: 1, length of the oesophageal lesion ≤5 cm; 2, length of lesion >5 cm. TNM: 1, clinical stage I-III of oesophageal cancer; 2, clinical stage IV **(A)**. Calibration curves of the radiomics nomogram in the training **(B)** and validation **(C)** cohorts. Potential incremental values of the radiomics nomogram relative to the radiological label were evaluated by net reclassification improvement (NRI) **(D)**.

TABLE 3 | Discriminating performance of the radiological label and radiomics nomogram.

Model	AIC	Training cohort (n = 153)	Validation cohort (n = 65)
		C-index (95% CI)	C-index (95% CI)
Radiological label	410.78	0.716 (0.645–0.787)	0.718 (0.612–0.825)
Radiomics nomogram	399.28	0.742 (0.674–0.810)	0.715 (0.609–0.820)

Shen et al., suggested that by extracting preoperative CT image features, a radiomic signature can predict the lymph node metastasis status of oesophageal cancer. Thirteen radiomics features were significantly correlated with lymph node metastasis. Another study by Zhang et al. screened 11 imaging features that were significantly associated with the local control of nasopharyngeal carcinoma. Similarly, in our study, the radiological label that was established with six radiomics features obtained satisfactory results in predicting LRFS after radiotherapy for oesophageal cancer in both the training group (C-index, 0.716; 95% CI: 0.645–0.787) and validation group (C-index, 0.718; 95% CI: 0.612–0.825). The patients were stratified into high-risk and low-risk groups according to the radiological label value. The 1-, 2-, and 3-year LRFS values of the low-risk group were significantly higher than those of the high-risk group. The results of our study confirmed that the radiological label was an independent predictor of LRFS.

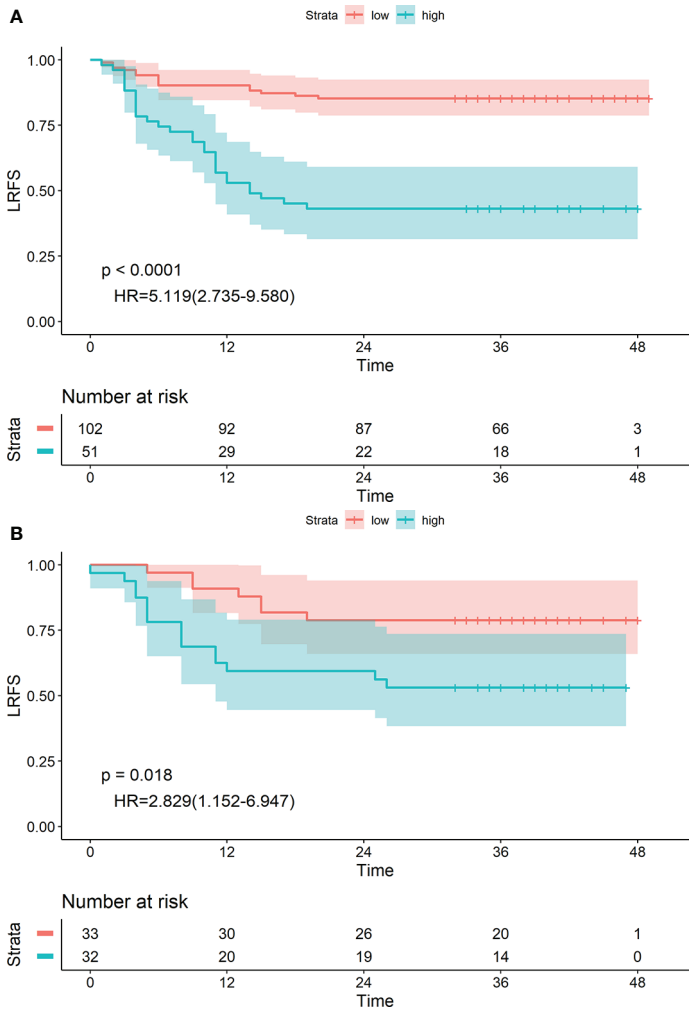


FIGURE 6 | Kaplan-Meier survival analyses of high-risk and low-risk groups divided by the radiomics nomogram in the training **(A)** and validation **(B)** cohorts.

Our radiological label displayed a significant correlation with LRFS. However, clinical characteristics including lesion length, T stage, and N stage were also important influencing factors for local recurrence of oesophageal cancer (29), and these factors can be easily determined during treatment without increasing the burden on patients. Many studies have reported that combining radiological label with clinical risk factors can improve the accuracy of LRFS prediction (30–36). Therefore, we assumed that our model would achieve better performance when combined with these factors, and our results confirmed this hypothesis. The area under the receiver operating characteristic curve for our radiomics nomogram was 0.790 (95% CI: 0.712–0.867) in the training group and 0.727 (95% CI: 0.598–0.856) in the validation group. The C-index was 0.742 (95% CI: 0.674–0.810) in the training group and 0.715 (95% CI: 0.609–0.820) in the validation group. Encouragingly, our study was one of the first clinical studies to explore the use of radiomics in pretherapeutic assessment to predict LRFS after radiotherapy for oesophageal cancer.

Similarly, according to the principle of maximum selected rank statistics, the optimal cut-off point for predicting LRFS by radiomics nomogram was 0.56, and patients were divided into high-risk group and low-risk groups accordingly. There were significant differences in LRFS at 1-, 2-, and 3-year follow-up between the high-and low-risk groups based on this criterion, both in the training and validation sets. Our finding again supports the hypothesis that LRFS was better predicted in patients with oesophageal cancer after radiotherapy when the radiomics label was incorporated with clinical factors.

Admittedly, there were a few limitations in our study. First, patient representation was limited due to the study being conducted in a single-centre; however, as a regional radiotherapy centre, there were sufficient case resources. Furthermore, biases might have been present due to the retrospective nature of the study design; for example, the exact time of relapse of some patients was affected by memory bias over the follow-up period. Genomics was also an important method used to explore the heterogeneity of tumours and to implement individualised

therapy. Therefore, to improve the accuracy of the prediction model, data from a larger multi-centre study is needed to further validate the stability and effectiveness of the model. In future clinical studies, key genetic markers that have been found to be closely related to local recurrence of oesophageal cancer should be added to the prediction model.

In conclusion, the predictive model established by combining enhanced-CT-derived radiomic characteristics with clinical risk factors has great potential and application prospect in predicting the local recurrence of oesophageal cancer after radiotherapy.

DATA AVAILABILITY STATEMENT

The datasets for this manuscript are not publicly available because involving patient confidential information. Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics committee of the Fourth Hospital of Hebei Medical University. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

JK and SZ designed the study. JK and SZ accomplished the manuscript. JK collected the required CT data. ZL and JZ evaluated the curative effect of radiotherapy for oesophageal cancer and determined the local control status of oesophageal cancer. ZL collected and assembled the clinical data. JR performed the statistical analysis and interpretation. JK determined the selection of references and experimental standards. JK and ZL performed the data analysis and interpretation. SZ performed manuscript approval and modification. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (Grant No. 81872456) and the Health Innovation Special Fund of Hebei Province (Grant No. 213777104D).

ACKNOWLEDGMENTS

This is a short text to acknowledge the contributions of specific colleagues, institutions, or agencies that aided the efforts of the authors.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.739933/full#supplementary-material

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018:GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492
2. Di Pardo BJ, Bronson NW, Diggs BS, Thomas CRJr, Hunter JG, Dolan JP. The Global Burden of Esophageal Cancer: A Disability-Adjusted Life-Year Approach. *World J Surg* (2016) 40:395–401. doi: 10.1007/s00268-015-3356-2
3. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global Cancer Statistics, 2012. *CA Cancer J Clin* (2015) 65:87–108. doi: 10.3322/caac.21262
4. Lagergren J, Smyth E, Cunningham D, Lagergren P. Oesophageal Cancer. *Lancet* (2017) 390:2383–96. doi: 10.1016/S0140-6736(17)31462-9
5. Welsh J, Settle SH, Amini A, Xiao L, Suzuki A, Hayashi Y, et al. Failure Patterns in Patients With Esophageal Cancer Treated With Definitive Chemoradiation. *Cancer* (2012) 118:2632–40. doi: 10.1002/cncr.26586
6. Parry K, Visser E, van Rossum PS, Mohammad NH, Ruurda JP, van Hillegersberg R. Prognosis and Treatment After Diagnosis of Recurrent Esophageal Carcinoma Following Esophagectomy With Curative Intent. *Ann Surg Oncol* (2015) 22:1292–300. doi: 10.1245/s10434-015-4840-5
7. Lou F, Sima CS, Adusumilli PS, Bains MS, Sarkaria IS, Rusch VW, et al. Esophageal Cancer Recurrence Patterns and Implications for Surveillance. *J Thorac Oncol* (2013) 8:1558–62. doi: 10.1097/01.JTO.0000437420.38972.fb
8. Zhang W-Y, Chen X-X, Chen W-H, Zhang H, Zou C-L. Nomograms for Predicting Risk of Locoregional Recurrence and Distant Metastases for Esophageal Cancer Patients After Radical Esophagectomy. *BMC Cancer* (2018) 18:879. doi: 10.1186/s12885-018-4796-5
9. Mattonen SA, Palma DA, Haasbeek CJA, Senan S, Ward AD. Early Prediction of Tumor Recurrence Based on CT Texture Changes After Stereotactic Ablative Radiotherapy (SABR) for Lung Cancer. *Med Phys* (2014) 41:033502. doi: 10.1118/1.4866219
10. Dong X, Sun X, Zhao X, Zhu W, Sun L, Huang Y, et al. The Impact of Intratumoral Metabolic Heterogeneity on Postoperative Recurrence and Survival in Resectable Esophageal Squamous Cell Carcinoma. *Oncotarget* (2017) 8:14969–77. doi: 10.18632/oncotarget.14743
11. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
12. Yang Z, He B, Zhuang X, Gao X, Wang D, Li M, et al. CT-Based Radiomic Signatures for Prediction of Pathologic Complete Response in Esophageal Squamous Cell Carcinoma After Neoadjuvant Chemoradiotherapy. *J Radiat Res* (2019) 60:538–45. doi: 10.1093/jrr/rrz027
13. Tan Z, Yang H, Wen J, Luo K, Liu Q, Hu Y, et al. Clinical Predictors of Pathologically Response After Neoadjuvant Chemoradiotherapy for Esophageal Squamous Cell Carcinoma: Long Term Outcomes of a Phase II Study. *J Thorac Dis* (2018) 10:5254–9. doi: 10.21037/jtd.2018.08.88
14. Huang R-W, Chao Y-K, Wen Y-W, Chang H-K, Tseng C-K, Chan S-C, et al. Predictors of Pathological Complete Response to Neoadjuvant Chemoradiotherapy for Esophageal Squamous Cell Carcinoma. *World J Surg Oncol* (2014) 12:170. doi: 10.1186/1477-7819-12-170
15. Hamai Y, Hihara J, Emi M, Furukawa T, Murakami Y, Nishibuchi I, et al. Preoperative Prediction of a Pathologic Complete Response of Esophageal Squamous Cell Carcinoma to Neoadjuvant Chemoradiotherapy. *Surgery* (2018) 164:40–8. doi: 10.1016/j.surg.2018.01.011

16. Rice TW, Ishwaran H, Blackstone EH, Hofstetter WL, Kelsen DP, Apperson-Hansen C. Recommendations for Clinical Staging (Ctnm) of Cancer of the Esophagus and Esophagogastric Junction for the 8th Edition AJCC/UICC Staging Manuals. *Dis Esophagus* (2016) 29:913–9. doi: 10.1111/dote.12540
17. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
18. Ghaly G, Harrison S, Kamel MK, Rahouma M, Nasar A, Port JL, et al. Predictors of Survival After Treatment of Oligometastases After Esophagectomy. *Ann Thorac Surg* (2018) 105:357. Surg Surg Thorac Surg c Surg nn Thorac Su. doi: 10.1016/j.athoracsur.2017.10.039
19. Oppedijk V, van der Gaast A, van Lanschot JJB, van Hagen P, van Os R, van Rij CM, et al. Patterns of Recurrence After Surgery Alone *Versus* Preoperative Chemoradiotherapy and Surgery in the CROSS Trials. *J Clin Oncol* (2014) 32:385–91. doi: 10.1200/JCO.2013.51.2186
20. Weksler B, Kennedy KF, Sullivan JL. Using the National Cancer Database to Create a Scoring System That Identifies Patients With Early-Stage Esophageal Cancer at Risk for Nodal Metastases. *J Thorac Cardiovasc Surg* (2017) 154:1787–93. doi: 10.1016/j.jtcvs.2017.07.036
21. Jayaprakasam VS, Yeh R, Ku GY, Petkovska I, Fuqua JL 3rd, Gollub M, et al. Role of Imaging in Esophageal Cancer Management in 2020: Update for Radiologists. *AJR Am J Roentgenol* (2020) 215:1072nol. doi: 10.2214/AJR.20.22791
22. Ji G-W, Zhang Y-D, Zhang H, Zhu F-P, Wang K, Xia Y-X, et al. Biliary Tract Cancer at CT: A Radiomics-Based Model to Predict Lymph Node Metastasis and Survival Outcomes. *Radiology* (2019) 290:90–8. doi: 10.1148/radiol.2018181408
23. Liang W, Yang P, Huang R, Xu L, Wang J, Liu W, et al. A Combined Nomogram Model to Preoperatively Predict Histologic Grade in Pancreatic Neuroendocrine Tumors. *Clin Cancer Res* (2019) 25:584–94. doi: 10.1158/1078-0432.CCR-18-1305
24. Wu S, Zheng J, Li Y, Wu Z, Shi S, Huang M, et al. Development and Validation of an MRI-Based Radiomics Signature for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer. *EBioMedicine* (2018) 34:76–84. doi: 10.1016/j.ebiom.2018.07.02934
25. Wu S, Zheng J, Li Y, Wu Z, Shi S, Huang M, et al. A Radiomics Nomogram for the Preoperative Prediction of Lymph Node Metastasis in Bladder Cancer. *Clin Cancer Res* (2017) 23:6904–11. doi: 10.1016/j.ebiom.2018.07.029
26. Shen C, Liu Z, Wang Z, Guo J, Zhang H, Wang Y, et al. Building CT Radiomics Based Nomogram for Preoperative Esophageal Cancer Patients Lymph Node Metastasis Prediction. *Transl Oncol* (2018) 11:815–24. doi: 10.1016/j.tranon.2018.04.005
27. Huang Y-Q, Liang C-H, He L, Tian J, Liang C-S, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol* (2016) 34:2157–64. doi: 10.1200/JCO
28. Zhang L-L, Huang M-Y, Li Y, Liang J-H, Gao T-S, Deng B, et al. Pretreatment MRI Radiomics Analysis Allows for Reliable Prediction of Local Recurrence in non-Metastatic T4 Nasopharyngeal Carcinoma. *EBioMedicine* (2019) 42:270–80. doi: 10.1016/j.ebiom.2019.03.050
29. Barbetta A, Sihag S, Nobel T, Hsu M, Tan KS, Bains M, et al. Patterns and Risk of Recurrence in Patients With Esophageal Cancer With a Pathologic Complete Response After Chemoradiotherapy Followed by Surgery. *J Thorac Cardiovasc Surg* (2018) 157:1249–59. doi: 10.1016/j.jtcvs.2018.09
30. Keek S, Sanduleanu S, Wesseling F, de Roest R, van den Brekel M, van der Heijden M, et al. Computed Tomography-Derived Radiomic Signature of Head and Neck Squamous Cell Carcinoma (Peri)Tumoral Tissue for the Prediction of Locoregional Recurrence and Distant Metastasis After Concurrent Chemo-Radiotherapy. *PloS One* (2020) 15:e0237048. doi: 10.1371/journal.pone.0237048
31. Xu X, Wang H, Du P, Zhang F, Li S, Zhang Z, et al. A Predictive Nomogram for Individualized Recurrence Stratification of Bladder Cancer Using Multiparametric MRI and Clinical Risk Factors. *J Magn Reson Imaging* (2019) 50:1893–904. doi: 10.1002/jmri.26749
32. Ji G-W, Zhu F-P, Xu Q, Wang KE, Wu M-Y, Tang W-W, et al. Machine-Learning Analysis of Contrast-Enhanced CT Radiomics Predicts Recurrence of Hepatocellular Carcinoma After Resection:a Multi-Institutional Study. *EBioMedicine* (2019) 50:156–65. doi: 10.1016/j.ebiom.2019.10.057
33. Qiu Q, Duan J, Deng H, Han Z, Gu J, Yue NJ, et al. Development and Validation of a Radiomics Nomogram Model for Predicting Postoperative Recurrence in Patients With Esophageal Squamous Cell Cancer Who Achieved PCR After Neoadjuvant Chemoradiotherapy Followed by Surgery. *Front Oncol* (2020) 10:1398. doi: 10.3389/fonc.2020.01398. eCollection 2020.
34. Fried DV, Mawlawi O, Zhang L, Fave X, Zhou S, Ibbott G, et al. Stage III Non-Small Cell Lung Cancer: Prognostic Value of FDG PET Quantitative Imaging Features Combined With Clinical Prognostic Factors. *Radiology* (2016) 278:214–22. doi: 10.1148/radiol.2015142920
35. Anthony GJ, Cunliffe A, Castillo R, Pham N, Guerrero T, Armato SG 3rd, et al. Incorporation of Pre-Therapy 18F-FDG Uptake Data With CT Texture Features Into a Radiomics Model for Radiation Pneumonitis Diagnosis. *Med Phys (Lancaster)* (2017) 44:3686–94. doi: 10.1002/mp.12282
36. Huang K, Lin Y, Yang L, Wang Y, Cai S, Pang L, et al. A Multipredictor Model to Predict the Conversion of Mild Cognitive Impairment to Alzheimer's Disease by Using a Predictive Nomogram. *Neuropsychopharmacology* (2020) 45:358–66. doi: 10.1038/s41386-019-0551-0

Conflict of Interest: JR was employed by GE Healthcare China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Kong, Zhu, Shi, Liu, Zhang and Ren. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 07 October 2021
doi: 10.3389/fonc.2021.739815

A Novel Radiogenomics Biomarker Based on Hypoxic-Gene Subset: Accurate Survival and Prognostic Prediction of Renal Clear Cell Carcinoma

Jiahao Gao[1†], Fangdie Ye[2,3†], Fang Han[1†], Xiaoshuang Wang[1], Haowen Jiang[2,3,4] and Jiawen Zhang[1*]*

[1] Department of Radiology, Huashan Hospital, Fudan University, Shanghai, China, [2] Department of Urology, Huashan Hospital, Fudan University, Shanghai, China, [3] Fudan Institute of Urology, Huashan Hospital, Fudan University, Shanghai, China, [4] National Clinical Research Center for Aging and Medicine, Huashan Hospital, Fudan University, Shanghai, China

OPEN ACCESS

Edited by:
Giorgio Russo,
National Research Council, Italy

Reviewed by:
Jiong Ni,
Tongji University, China
Zhongxiang Ding,
Zhejiang University, China

***Correspondence:**
Jiawen Zhang
zhangjw2000@126.com
Haowen Jiang
urology_hs@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 12 July 2021
Accepted: 22 September 2021
Published: 07 October 2021

Citation:
Gao J, Ye F, Han F, Wang X, Jiang H and Zhang J (2021) A Novel Radiogenomics Biomarker Based on Hypoxic-Gene Subset: Accurate Survival and Prognostic Prediction of Renal Clear Cell Carcinoma. Front. Oncol. 11:739815. doi: 10.3389/fonc.2021.739815

Purpose: To construct a novel radiogenomics biomarker based on hypoxic-gene subset for the accurate prognostic prediction of clear cell renal cell carcinoma (ccRCC).

Materials and Methods: Initially, we screened for the desired hypoxic-gene subset by analysis using the GSEA database. Through univariate and multivariate cox regression hazard ratio analysis, survival-related hypoxia genes were identified, and a genomics signature was constructed in the TCGA database. Building on this, a hypoxia-gene related radiogenomics biomarker (prediction of hypoxia-genes signature by contrast-enhanced CT radiomics) was constructed in the TCIA-KIRC database by extracting features in the venous phase of contrast-enhanced CT images, selecting features using the mRMR and LASSO algorithms, and building logistic regression models. Finally, we validated the prognostic capability of the new biomarker for patients with ccRCC in an independent validation cohort at Huashan Hospital of Fudan University, Shanghai, China.

Results: The hypoxia-related genomics signature consisting of five genes (IFT57, PABPN1, RNF10, RNF19B and UBE2T) was shown to be significantly associated with survival for patients with ccRCC in the TCGA database, delineated by grouping of the signature expression as either low- or high-risk. In the TCIA database, we constructed a radiogenomics biomarker consisting of 13 radiomics features that were optimal predictors of hypoxia-gene signature expression levels (low- or high-risk) in patients at each institution, that demonstrated AUC values of 0.91 and 0.91 in the training and validation groups, respectively. In the independent validation cohort at Huashan Hospital, our radiogenomics biomarker was significantly associated with prognosis in patients with ccRCC (p=0.0059).

Conclusions: The novel prognostic radiogenomics biomarker that was constructed achieved excellent correlation with prognosis in both the cohort of TCGA/TCIA-KIRC database and the independent validation cohort of Huashan hospital patients with ccRCC. It is anticipated that this work may assist in clinical preferential treatment decisions and promote the process of precision theranostics in the future.

Keywords: clear cell renal cell carcinoma, radiogenomics, contrast-enhanced computed tomography, texture analysis, hypoxia

INTRODUCTION

Recognized as the most common urinary cancer, renal cell carcinoma was responsible for more than 175,000 deaths worldwide in 2020 (1, 2). Clear cell renal cell carcinoma (ccRCC) is the predominant pathological type of kidney cancer, accounting for approximately 75% of these malignancies (3). Surgical resection remains the principal treatment for localized or locally advanced renal cell carcinoma. Although surgically resected localized renal cancers have a 5-year survival ranging from 80% - 95%, non-metastatic renal cancers deemed to be high risk have a probability of recurrence or metastasis as high as 30% - 40%, associated with an extremely high mortality rate (4). In recent years, due to widespread use of abdominal ultrasound and CT scans for vague symptoms, an increasing number of serendipitously discovered, asymptomatic renal cancers have been diagnosed. As noted previously, with rather wide variability in the prognosis of individual patients with ccRCC, there is an urgent need for more precise and readily defined prognostic parameters to group patients according to disease risk, to facilitate individualized clinical oncology treatment options.

Tumors are essentially a phenomenon of mutated gene expression as oncogenes and tumor suppressor genes. Genetic heterogeneity of cancer cells is the underlying cause for differences in drug resistance, disease progression, and patient survival (5–7). In contrast to normal cells, the metabolism of glucose by clear cell renal carcinoma cells, even in the presence of sufficient oxygen, is mainly *via* the glycolytic pathway. This 'pseudo-hypoxia' was found to be associated with the presence of genetic mutations and deletions in the majority of ccRCC. Hypoxia-inducible factor (HIF)-induced gene mutations promote tumor angiogenesis, metabolic reprogramming, and cancer cell proliferation (8–10). Most HIF targeted-genes are involved in hypoxic responses, and the expression status of hypoxia-related genes exerts a significant impact on the prognosis of patients with ccRCC (11, 12). Although a number of genetic mutations have been found in ccRCC (PBRM1, BAP1, SETD2, and KDM5C), effective and definitive genetic prognostic markers for gene expression in kidney cancer patients are lacking. Additionally, small specimens obtained by puncture biopsy do not accurately reflect the heterogeneity of gene expression across the entire tumor. Reliable information on tumor gene expression is usually obtained by genetic sequencing of surgically resected specimens. Highly invasive and costly for patients, they are not currently available on a large scale and lack universal prognostic guidance.

Radiogenomics is a promising technology in cancer-related research. Based on the use of automated, high-throughput feature extraction methods, it can provide insight into the occurrence, development and heterogeneity of tumors by deeply mining the biological nature of medical images and integrating them with genomic data (13). Likewise, this technique is potentially exceptionally promising for linking highly reproducible, non-invasive imaging features with the disease gene expression profile that is distinctly associated with a clinically meaningful prognosis (14). In recent years, radiogenomics has been reported covering a wide range of tumors, encompassing gene sequences, gene expression, molecular subtypes, and tumor heterogeneity (15–17), effectively providing direction for the formulation of clinical treatment plans, particularly devised for the individual patient.

The current study focused on an important prognostic factor in ccRCC: the association between hypoxia-related gene expression and the prognosis of patients with clear cell renal cell carcinoma. The intent was to develop a prognostic radiogenomics biomarker for ccRCC employing the TCIA-KIRC database, then utilizing radiomics to reflect the expression levels of hypoxia-associated gene subsets through contrast-enhanced CT features, and ultimately validating radiogenomics biomarker on a cohort of ccRCC patients from the Huashan Hospital dataset.

MATERIALS AND METHODS

Datasets

The workflow of the study is depicted in **Figure 1**. The study included three cohorts. Cohort 1 contained transcriptome profiles and corresponding clinical data from TCGA-KIRC with the exception of TCIA, and was divided into training and validation groups. It was used to construct a prognostic risk model and determine the prognostic factors that contributed to hypoxia-related gene mutation. This cohort initially included 318 samples; 41 samples were excluded as there were either no clinical reports or insufficient survival data, leaving 277 samples for analysis. For Cohort 2, clinical and pathological information (including gender, age, T stage, M stage, Stage, Grade) as well as imaging and genetic data were obtained from The Cancer Genome Atlas (TCGA) study website. Multi-institutional medical imaging data were retrospectively obtained from The Cancer Imaging Archive (TCIA) renal clear

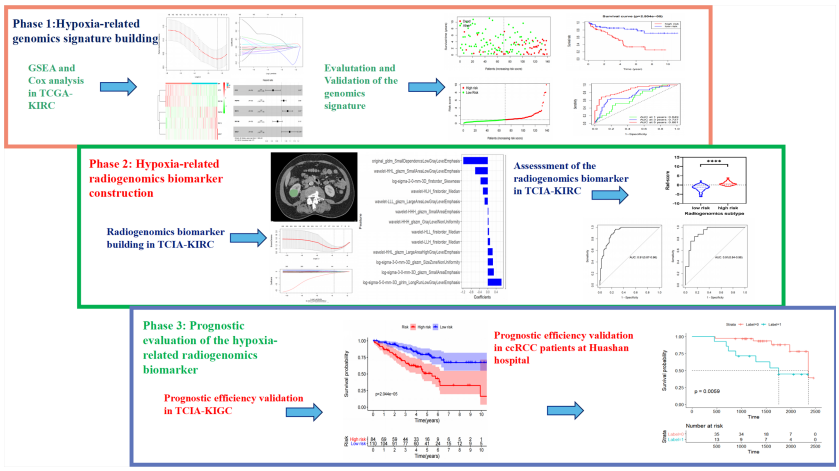


FIGURE 1 | Workflow of the research.

cell carcinoma database. The data were collected with the permission of each institution's ethics review committee and de-identified in accordance with the Health Insurance Portability and Accountability Act (HIPAA). Cohort 3 comprised subjects from Huashan Hospital, Fudan University, and the institutional review board approved the study and waived informed consent. Patients diagnosed with ccRCC who had undergone surgical resection with confirmatory pathology between January 2013 and December 2017 were considered for this retrospective study. The optimized risk prediction model based on genomic data was investigated in Cohort 1, and Cohort 2 was used to explore the relationship between hypoxic-related intratumor heterogeneity and the imaging signature. Cohort 3 was used to assess and validate the performance of the radiogenomics prognostic model.

Imaging Protocol

Cohort 2 initially included 245 patients for whom preoperative baseline abdominal CT or MRI examinations for ccRCC were available. Variations in imaging data in terms of examination modality, machine manufacturer and image acquisition protocol were taken into account. The data were screened according to both inclusion and exclusion criteria. Inclusion criteria included the following: (1) definite postoperative pathology of clear cell renal cell carcinoma with TNM stage; (2) complete imaging data on preoperative enhanced CT or MR scans, with image layer thickness <5 mm, tube voltage 120 KV, and matrix 512*512; and (3) clearly identifiable malignant lesions in the renal parenchymal phase on enhanced CT images. Exclusion criteria included: (1) poor quality enhanced CT images or significant artifacts affecting the region of interest; or (2) inability to successfully extract radiomics features from enhanced CT images. The detailed inclusion procedure is depicted in **Figure S1**.

In Cohort 3, CT examinations of all patients were performed using a 256-row CT system (Brilliance iCT, Philips Medical Systems, The Netherlands). All renal CT images were acquired using a standard three-phase scanning protocol, with parameters as follows: 120 kV; 150-200 mA; rotation time, 0.5-0.75 s; collimation, 128 × 0.625 mm; matrix, 512 × 512; and slice thickness, 1.5 mm. Patients were scanned in the supine position with breath held, inclusive from the top of the diaphragm to the lower edge of the kidney. The abdomen was first scanned, and the enhanced scan utilized a high-pressure syringe injection of a non-ionic contrast agent (1.5 ml/Kg, 3.0 ml/s), with scans at 30 s, 90 s and 300 s following injection to obtain cortical, parenchymal and excretory phase images, respectively. The inclusion and exclusion criteria for Cohort 3 were the same as for Cohort 2.

Cohort 2 initially included 136 samples as a training group from which exclusions included 11 with no obvious tumor, 8 due to poor imaging quality, and 25 with incomplete enhanced CT data. There were 127 males and 67 females with a mean age of 60.45 ± 11.64 years (range, 34 to 88 years). Cohort 3 served as a validation group, derived from Huashan Hospital as an independent prognostic testing dataset. Of 58 initial samples, 8 were excluded due to incomplete enhanced CT imaging data, leaving 50 samples as the final cohort (29 males and 21 females; mean age, 62.9 ± 12.17; range, 33 and 87 years).

Construction of the Hypoxia-Related Genomics Prognostic Model

The transcriptome data and the corresponding clinical survival data (including age, gender, grade, TMN) of the patients with ccRCC were obtained from the TCGA (https://portal.gdc.cancer.gov/). The hypoxic-associated genes were then obtained from GSEA. Univariate Cox proportional hazard regression analysis collected the candidate hypoxia-related genes, which significantly associated with the overall survival by using R-package "survival" ($P < 0.05$). Prognostic hypoxic-related genes were divided into risk related genes (hazard ratio, HR >1) and protective related genes (HR <1). The risk score for each patient

was then calculated using a linear combination of characteristic gene expression, with characteristic gene expression weighted by their regression coefficients = (expression 1× coefficient gene 1) +(expression 2× coefficient gene 2) +… +(expressing 1× coefficient gene). Patients with ccRCC were divided into high-risk and low-risk groups according to the median risk score.

Image Processing and Region of Interest Sketching

In Cohorts 2 and 3, the enhanced CT image data were normalized and resized using the z-score method and the mean normalization method. Parenchymal phase CT images of each patient were used for radiomics feature extraction. For each image sequence, a radiologist (14 years of abdominal imaging experience) segmented the lesion contours on each slice using an open-source software (3D Slicer version 4.11.0; Boston, MA). Radiomic features were extracted for each stage of the 3D volume with a python-based radiomics software (Pyradiomics version 3.0.0; https://github.com/Radiomics/pyradiomics) (18). The extracted features are included in **Supplementary Materials I**.

Intra-class and inter-class correlation coefficients (ICC) were used to assess the reliability of the extracted features. A random selection of 50 patients underwent repeat region of interest (ROI) segmentation performed by the same and an additional radiologist (7 years of abdominal imaging experience) 30 days after the initial segmentation. Relevant clinical and pathological information was blinded to the radiologists.

Radiogenomics Signature Building

In Cohort 2, the extracted radiogenomics features were identified according to the following consecutive steps to construct the radiogenomics model. Initially, features with both intra- and inter- ICC greater than 0.75 were filtered out. Feature dimensionality reduction was further accomplished using the mRMR (minimum Redundancy, Maximum Relevance) method and the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm to select the most optimal and robust features. These features were then combined with their coefficients in the LASSO regression to construct radiogenomics feature labels: Radscores. The Mann-Whitney U test was used to assess the ability of the newly created radiogenomics marker to group patients with different hypoxia gene set expression levels into low- or high-risk groups. We also used ROC curves and area under the curve (AUC) values to evaluate its performance.

We also combined the constructed radiogenomics biomarker with clinical and pathological factors to construct a nomogram which better visualizes the model and increases the reliability and predictive power. First, univariate cox regression analysis was performed for each clinical, pathological variable and radiogenomics biomarker, and then variables with $P<0.05$ in the univariate analysis were included in the multivariate Cox proportional risk model to determine the independent predictors of overall survival (OS). In the multivariate Cox regression, a combined model which consisting radiogenomics biomarker and other useful clinical and pathological factors was built by backward stepwise selection according to the Akaike Information Criterion (AIC), and a nomogram was constructed using R software. Internal validation of the predictive performance of the nomogram was carried out by 1000 resampling with the boots-trap method. ROC curves (1-year, 3-year and 5-year survival) were applied for evaluation. In addition, calibration curves were plotted to ensure the goodness of fit and reliability of the nomogram.

Radiogenomics Signature Validation

For further validation of the prognostic predictive power of our radiogenomics signature for ccRCC, we used Cohort 3, using the same steps as described previously for image acquisition and radiogenomics features extraction to produce the final radiogenomics scores. Kaplan-Meier curves were produced for overall survival of the patients based on this label, and the log-rank test was used to determine whether the new marker was successful in stratifying the prognosis of the patients. Hazard ratios (HR) and their 95% confidence intervals were obtained to assess survival differences between stratified groups. Univariate Cox regression models were used to further identify whether the radiogenomics signature were independently associated with OS.

Statistical Analysis

Cohort 1 comprised a training cohort (n=139) and a validation cohort (n=138) for which risk scores were calculated. Using univariate Cox proportional hazard regression analysis, based on the median risk score, the patients in the training cohort were divided into low- and high-risk groups, and the differences between the groups were evaluated and verified. The same approach was performed in the validation cohort, again separating patients into low- and high-risk groups. The radiogenomic features were then collected from the enhanced CT images to determine the relationship between the imaging report and the genome subcloning, intended to predict the patient survival. Finally, we evaluated the prognostic performance of the radiogenomics signatures on the Cohort 3 dataset with enhanced-CT and matched survival data.

The risk model based on hypoxia-related genes proved to be an independent prognostic factor by univariate and multivariate Cox proportional hazard regression analysis. Kaplan-Meier survival curves were constructed to evaluate differences of overall survival between groups. ROC (receiver operating characteristics) curve analysis was performed to evaluate the prediction performance of the risk model and the combined nomogram. Calibration curves were plotted to assess the accuracy and reliability of the combined nomogram. Two-tailed p-values of less than 0.05 were considered statistically significant. Statistical analyses were all performed on R software (version 3.6.3).

RESULTS

Phase1: Training of Hypoxia-Related Genomic Subclone Model in Cohort 1

In the present study, the risk model was constructed to evaluate hypoxia-related gene status of patients with ccRCC. Initially, a

univariate Cox analysis screened prognostic-related hypoxia genes. Then the LASSO Cox regression algorithm was performed to identify the most valuable prognostic hypoxia-related genes with non-zero regression coefficients (**Figures 2A, B**). This produced 5 gene signatures and the risk score of each patient was calculated by performing multivariate Cox analysis. IFT57 and RNF19B were identified as low-risk prognostic genes whereas PABPN1, RNF10 and UBE2T were considered high-risk prognostic factors (**Figures 2C, D**). The detailed information regarding the selected genes are shown in **Table 1**. The Risk score = (0.0234×expression of RABPN1) + (0.0670×expression of RNF10) + (0.0890×expression of UBE2T) – (0.0924×expression of IFT57) – (0.0432×expression of RNF19B). The risk model could be clearly separated from the PCA (Principal Component Analysis) analysis and the genomic signature established was independently correlated with the survival of ccRCC patients (**Supplementary Materials II**). The distribution of the risk scores as well as the relationship between the risk scores and survival data are illustrated in scatterplots (**Figures 3A, B**). Each of the patients with ccRCC was allocated into either a low- or high-risk group according to the median risk score. To comprehensively evaluate the prognostic value of these five gene signatures in the training group, Kaplan-Meier survival curves demonstrated that the patients in the high-risk group had a significantly shorter overall survival compared to patients in the low-risk group. These results were verified by the K-M survival curves in the validation set (**Figures 3C, D**).

To further determine whether the risk-score constituted an independent prognostic factor that correlated with poor prognosis in patients with ccRCC, we performed both univariate and multivariate Cox proportional hazards regression analyses in both the training and validation sets. These revealed that age, stage, and Risk score were significantly associated with the OS of patients with ccRCC in the training set. Critically important, in the validation set, only the Risk score proved to be significantly associated with OS of patients with ccRCC (**Figure S2**). The correlation between the prognostic genes and pathology is shown in **Figure 4**.

Phase 1: Validation of the Hypoxia-Related Genomics Model in Cohort 2

Based on the hypoxia-related prognostic genes obtained by the above method, we divided Cohort 2 into high- and low-risk groups according to the median value of the Risk score. The low-risk group had a lower death rate and longer survival (**Supplementary Materials IV**). The K-M survival curves

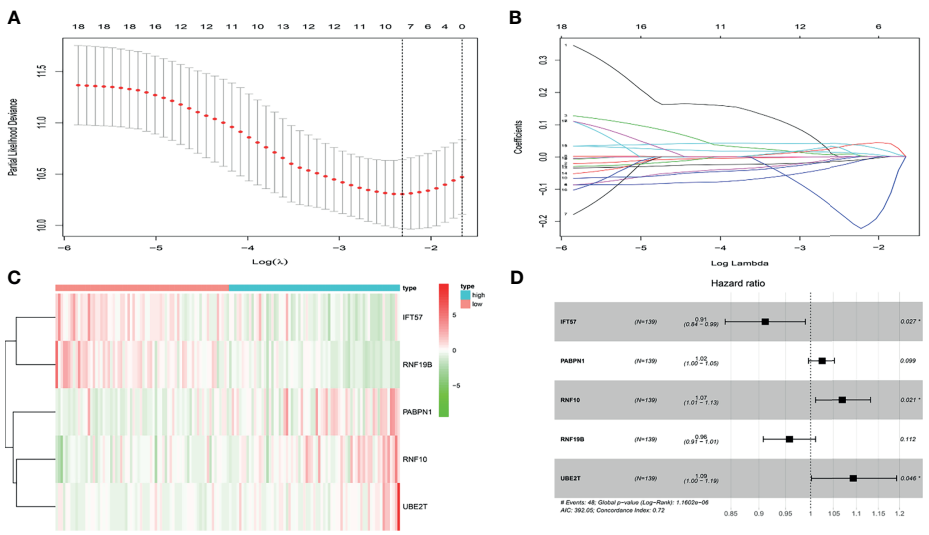


FIGURE 2 | **(A, B)** After univariate logistic regression, the LASSO regression algorithm was performed to further identify the most valuable prognostic hypoxia-related genes with non-zero regression coefficient. **(C)** The heatmap of hypoxia-related prognostic genes expression level. **(D)** The high-risk hypoxia-related prognostic factors selected by multivariate Cox analysis.

TABLE 1 | Identification of prognostic hypoxia-related genes in the multivariate cox regression.

Hypoxia-related genes	Coef	HR	HR.95L	HR.95H	Cox p-value
IFT57	-0.092391582	0.911748056	0.840186788	0.989404414	0.026733824
PABPN1	0.023407623	1.023683732	0.995610475	1.052548572	0.098966027
RNF10	0.066942512	1.069234008	1.010052418	1.131883201	0.021208353
RNF19B	-0.043239752	0.957681757	0.907947788	1.010139965	0.112021756
UBE2T	0.089040247	1.093124651	1.001637183	1.192968395	0.04586362

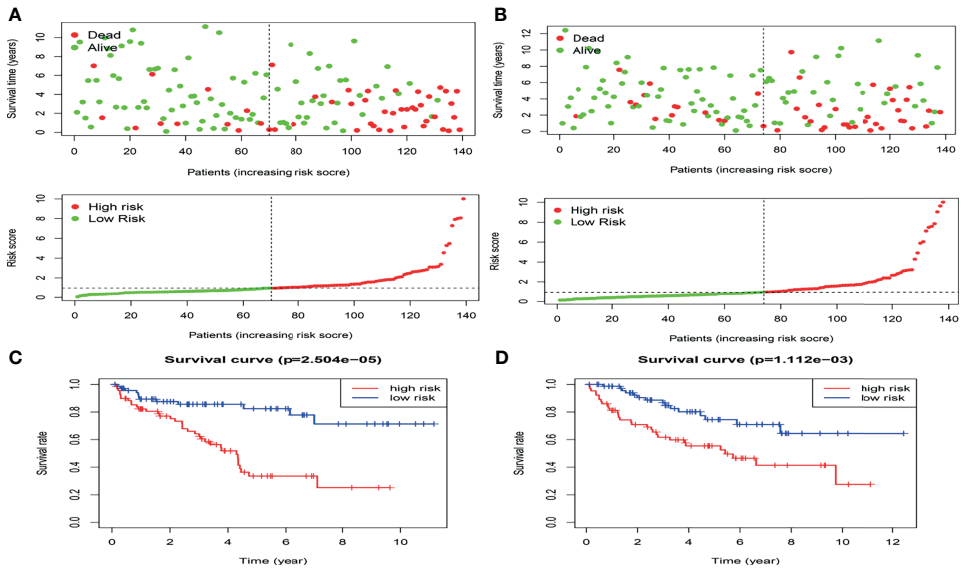


FIGURE 3 | The distribution of the risk scores as well as the relationship between the risk scores and survival status in the training cohort **(A)** and validation cohort **(B)**.The result of K-M analysis for the hypoxia-related risk model in the training cohort **(C)** and validation cohort **(D)**.

indicated that the Risk score was inversely correlated with length of survival (**Figure 7A**). The AUC of the ROC curve of the risk model for overall survival is included in the **Supplementary Materials**.

Phase 2: Radiomics Feature Selection, Construction of the Radiogenomics Biomarker and Combined Nomogram

In Cohort 2, 136 patients were randomized to the training cohort. Enhanced CT images of the renal parenchymal phase from these patients were used to extract radiomics features, starting with 1218 radiomics features. An initial screen of intra- and inter-group ICC > 0.75 reduced the number to 827 features. The mRMR algorithm further downscaled the number of features to 30, following which the most dynamic and relevant features were finally selected for modeling using LASSO regression (**Figure 5A**). The 13 optimal features were combined by multivariate logistic regression for the expression status of the hypoxic gene subgroup, thereby constructing a radiogenomics signature, expressed as radiogenomics score.

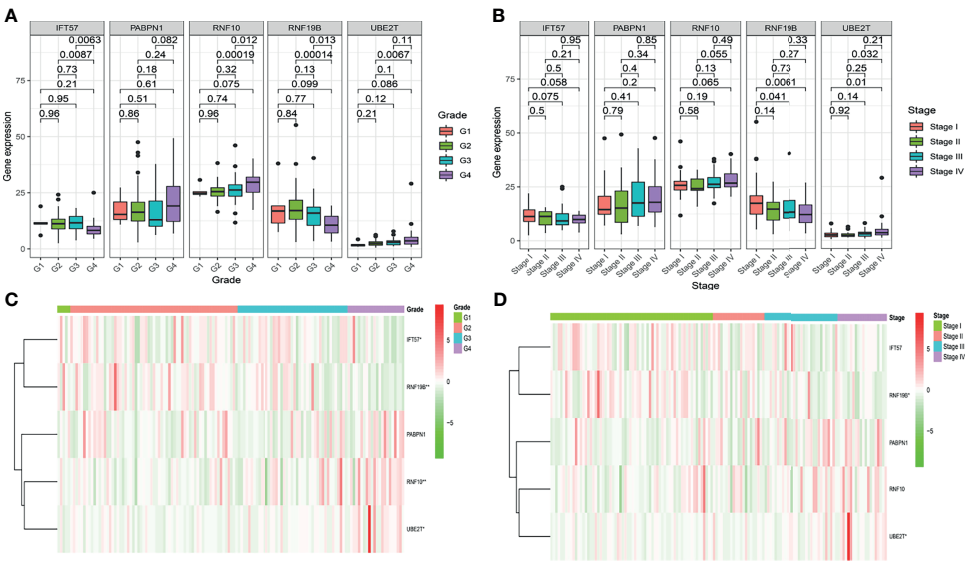


FIGURE 4 | The correlation between the hypoxia-related prognostic genes and clinical pathological, such as Grade **(A, C)** and Stage **(B, D)**, in the training cohort.

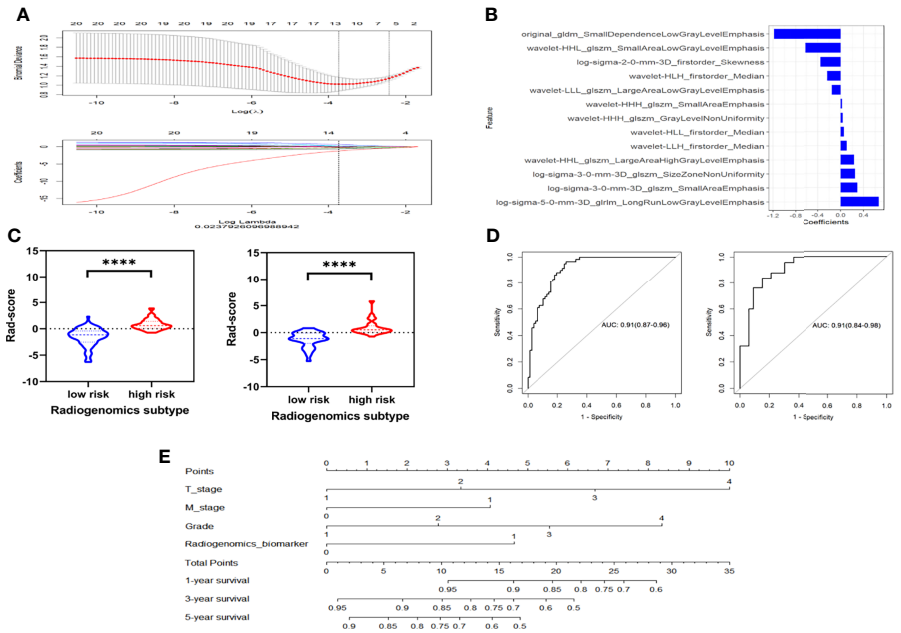


FIGURE 5 | The Lasso regression procedure to select the optimist radiomics features **(A)**. Features selected for hypoxia-related radiogenomics model construction **(B)**. The violin plots of the radiogenomics model in the training cohort and the validation cohort **(C)**. ROC curves shows that the radiogenomics biomarker could exactly distinguish patients into high- an low- hypoxia-related genomics grouping both in the training cohort and validation cohort **(D)**. **(E)** Nomogram of the model incorporating radiogenomics marker and other effective clinical and pathological information through Cox regression. ****p < 0.001.

Details of the included radiomics features are shown in **Figure 5B**. The specific formulae for radiogenomics scores are presented in **Supplementary Materials III**.

We then constructed a combined model incorporating clinical and pathological factors as well as the radiogenomics biomarker. Univariate Cox regression analysis showed that T-stage, M-stage, Grade, Stage and radiogenomics biomarker were associated with survival (P<0.05), and these five factors were included in the multivariate Cox regression. After backward stepwise selection, T stage, M stage, Grade and radiogenomics biomarker formed the final combined nomogram (**Figure 5E**). T stage (HR=2.33, P=0.02) and radiogenomics biomarker (HR=1.81, P=0.03) were shown to be independent predictors of survival (**Table 2**).

Phase 2: Radiogenomics Model and Combined Nomogram Evaluation

The Mann-Whitney U test vividly demonstrates the ability of the radiogenomics signature to group hypoxic gene subsets, as depicted in a violin plot (**Figure 5C**). Shown in **Figure 5D**, this Radiogenomics score in the ROC curves discriminated well between high- and low-risk groups for the subset of hypoxia genes. The AUC values of the ROC curves reached 0.91 in the training group and 0.91 in the validation group.

The ROC curve demonstrated that the combined nomogram incorporating clinical and pathological factors and the radiogenomics biomarker was effective in predicting 1-year, 3-year and 5-year survival rates (AUC=0.789, 0.782, 0.731 respectively) in cohort 2 (**Supplementary Materials V**). It could further improve the predictive effectiveness of the model compared to the Radiogenomics biomarker alone. In addition, the calibration curves of the nomogram for predicting survival are close to the "actual curves" (**Supplementary Materials V**), indicating that the model fits well and is reliable.

In addition, the GSEA demonstrated that linoleic acid/alpha linolenic/glycerophospholipid metabolism and oxidative phosphorylation were significantly enriched in the high-risk model, whereas tight/adherens junction, renal cell carcinoma, ERBB signaling pathway, and the TGF beta signature pathway were enriched in the low-risk model. (**Figure 6**)

Phase 3: Validation of the Prognostic Predictive Performance for the Radiogenomics Biomarker

For the independent validation group, extraction of radiomics features needed to construct the radiogenomics model, and the radiogenomics score were determined for each patient in like manner to the techniques reviewed previously. Based on the best cut-off values of the radiogenomics scores in Cohort 2, 14 patients were included in the predicted-high risk group of hypoxic gene subset, while 36 patients were included in the low-risk group.

The results demonstrated that the radiogenomics signature was associated with overall survival (P <0.01) (**Figure 7B**). Univariable Cox regression analysis revealed that the radiogenomics biomarker proved to be an independent preoperative prognostic factor in patients with ccRCC (HR=1.57, P <0.01).

TABLE 2 | Uni- and multivariable cox regression analysis of predictors of overall survival in cohort 2.

Factors	Univariate Analysis		Multivariate Analysis	
	Hazard Ratio (95%CI)	*P* Value	Hazard Ratio (95%CI)	*P* Value
Age	1.41 (0.84,2.36)	0.19	NA	NA
Gender	1.36 (0.81,2.28)	0.25	NA	NA
T stage	3.97 (2.27,6.94)	<0.01	2.33 (1.21,4.50)	0.02
M stage	3.94 (2.30,6.76)	<0.01	1.67 (0.87,3.24)	0.12
Grade	2.28 (1.57,3.30)	0.03	1.42 (0.93,2.18)	0.10
Stage	3.17 (2.05,4.92)	<0.01	NA	NA
Radiogenomics biomarker	2.46 (1.45,4.19)	<0.01	1.81 (1.04,3.14)	0.03

NA, not available.

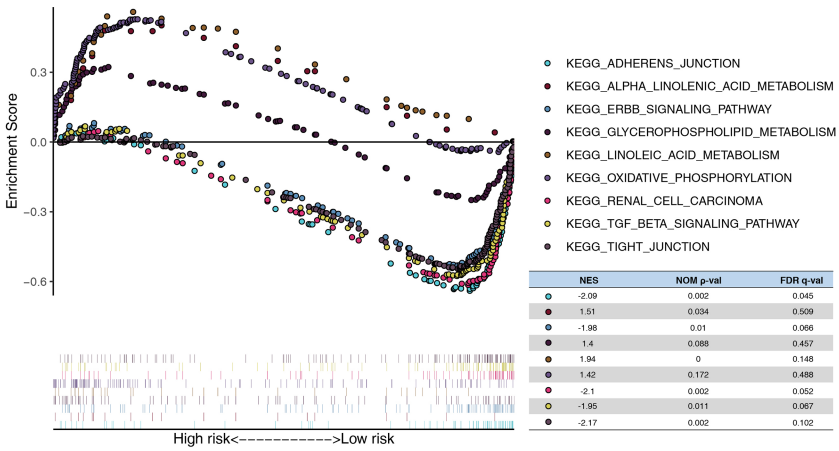


FIGURE 6 | Functional analysis of the risk model. GSEA analysis between high- and low-risk in TCGA. NES, normalized enrichment score; Nom, nominal; FDR, false discovery rate.

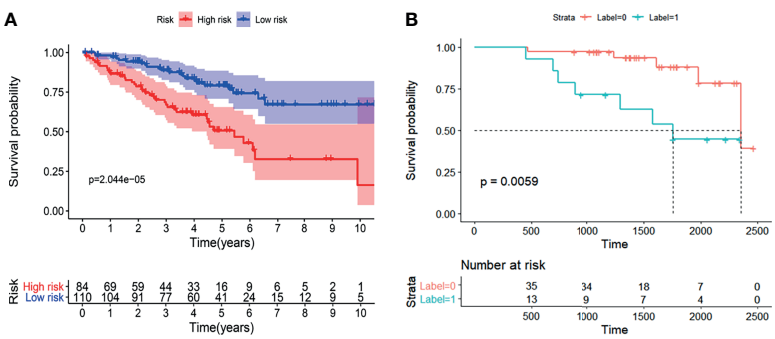


FIGURE 7 | The K-M analysis showed that the established radiogenomics biomarker could successfully divide ccRCC patients in to high- and low-risk, and possessing effective prognostic predictive ability both in the TCIA-KIRC **(A)** database and Huashan validation database **(B)**.

DISCUSSION

In this study, we focused on establishing the relationship between hypoxia gene expression and prognosis of patients with ccRCC. Our data confirmed that high- and low-risk grouping of hypoxia genes were significantly associated with prognosis. Furthermore, a non-invasive, efficient assessment of the expression level of hypoxic gene sets was performed utilizing a radiomics approach. Finally, this association of genomics and radiomics was verified in the validation set of ccRCC patients, and survival statistics established that our radiogenomics biomarker was capable of stratifying the prognosis of ccRCC patients. In addition, we combined A with validated clinical and pathological factors to build an integrated model and visualized it by means of a

nomogram, further improving the reliability and predictive efficacy of the model. The reliability of the radiogenomics biomarker as an independent prognostic marker for ccRCC was further demonstrated in the multivariate COX regression. It would seem appropriate for this biomarker to serve as a non-invasive prognostic marker for ccRCC that could facilitate individualized, potentially targeted treatment of tumors.

A high frequency of VHL gene mutations has been observed in up to 80% of ccRCC. The function of the VHL protein (pVHL) has been well analyzed (9, 10). pVHL contributes to the ubiquitination and degradation of hypoxia-inducible factor (HIF), a transcription factor of VEGF, under hypoxic conditions. In contrast, mutations in the VHL gene lead to reduced degradation of HIF, and may lead to VEGF overexpression and angiogenesis, which in turn lead to the development of ccRCC. Therefore, the expression levels of hypoxia-associated gene sets are significantly associated with the prognosis of patients with ccRCC (19–21). Intratumoral heterogeneity has been cited as a cause of prognostic differences of individual patients, which is reflected by gene expression data. Methods currently in common use to describe tumor heterogeneity include genetic testing and histopathology. Such approaches confirm the heterogeneity of cancer at the genetic level and can yield greater insight into the establishment of effective treatments. However, they are often expensive and invasive. Moreover, the biological characteristics of the excised tumor tissue may differ somewhat from the tumor as a whole, thereby misrepresenting the actual tumor hypoxic gene expression.

Surgery remains the preferred treatment modality for limited kidney cancer as it lacks sensitivity to conventional radiotherapy and chemotherapy. For inoperable patients with intermediate to advanced ccRCC, targeted therapies and immunotherapy are often recommended. However, the heterogeneity of tumors between patients leads to extremely variable treatment responses, and current clinicopathological markers such as Fuhrman grading and TNM staging do not adequately or effectively reflect these biological differences. Clearly, novel independent prognostic biomarkers are needed to predict prognosis of ccRCC (22, 23). Genetic mutations are the principal initiators of tumor cell heterogeneity, and variations in the tumor gene set will continue to accumulate further altering the genetic profile. Therefore, many studies have been conducted to establish genetic biomarkers to predict the prognosis of ccRCC, and these markers have been associated with various physiological changes in the tumor, such as immune infiltration and metabolic changes, and have achieved good predictive results (24–26). Among these, hypoxia-associated genes have been important factors in the development of KIRC as well as predicting prognosis. An alternative method for this type of research is the use of gene chips, but their cost limits clinical application, and there remains a need for efficient, non-invasive and cost-effective clinical biomarkers.

In recent years, with the deep application of artificial intelligence in the field of image processing, radiomics has developed rapidly. Despite tumor imaging being principally limited to morphological features in routine clinical practice, the ability of imaging to provide a comprehensive view of the spatial-temporal heterogeneity of individual tumors is unmatched by other biomarkers and examination modalities. Radiomics can provide non-invasive tools to study tumor biology, capture tumor heterogeneity, and monitor tumor evolution and response to treatment. There have been numerous radiomics studies applied to the differential diagnosis, molecular typing and prognostic prediction of ccRCC (27–29). Unfortunately, neither the computational features extracted by traditional radiomics using machine-learning methods, nor the "black-box" selection of deep learning methods can explain the function of radiomics for the prediction of clinical events. The interpretability of these methods is not strong, and in particular, explanations of the underlying biological and molecular mechanisms are lacking. In contrast, radiogenomics profoundly reflects the nature of tumor heterogeneity and fully explains the latent causes affecting prognostic survival. Both excellent predictive results are achieved through radiogenomics approaches and the biology of this novel biomarker can be summarized in terms of the essence of cancer pathogenesis: relevant gene mutations, thus effectively improving the interpretability of the model and facilitating its future clinical application.

This study has some limitations: (1) the TCIA database contains imaging data from multiple centers with various imaging machines and image acquisition protocols. Although strict inclusion and exclusion criteria were used and the model was validated in an independent center, the results need to be further investigated in future clinical trials in more centers. (2) The region of interest was extracted in the venous phase and segmented manually during the radiomics feature extraction process, which may lack the heterogeneity of tumor features embedded in other phases of CT scans and may have some selection bias. (3) Different combinations of machine learning algorithms other than multivariate logistic regression and COX regression need to be considered in the future to integrate and optimize the best models for more effective feature selection and to improve model prediction performance. (4) More clinical and pathological factors may need to be incorporated in the future to improve the predictive power and reliability of the model.

CONCLUSIONS

In this study, a novel radiogenomics biomarker based on the expression of hypoxic gene subset was developed for the accurate prediction of prognosis in ccRCC. This new biomarker achieved good survival prediction in both the training set of TCIA/TCGA and the independent validation set of Huashan hospital and may assist in clinical preferential treatment decisions for ccRCC in the future and promote the process of precision prognosis and treatment.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Materials**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of Huashan Hospital, Fudan University. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

JG, FY, and FH contributed equally to this work. Study design and concept, JG and FY. Provision of study materials or patients, FH. Data analysis and interpretation, FH and XW. Supervision, JZ and HJ. Manuscript writing: all authors. Final approval of manuscript: all authors. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the National Natural Science Foundation of China (82071877), National Key Research and Development Project in China (No. 2017YFC0113405), and Launching Fund of Huashan Hospital (North Hospital) Affiliated to Fudan University (No. HSBY2020003).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.739815/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* (2021) 71:7–33. doi: 10.3322/caac.21654
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660
3. Gupta K, Miller JD, Li JZ, Russell MW, Charbonneau C. Epidemiologic and Socioeconomic Burden of Metastatic Renal Cell Carcinoma (mRCC): A Literature Review. *Cancer Treat Rev* (2008) 34:193–205. doi: 10.1016/j.ctrv.2007.12.001
4. Lam JS, Shvarts O, Leppert JT, Pantuck AJ, Figlin RA, Belldegrun AS. Postoperative Surveillance Protocol for Patients With Localized and Locally Advanced Renal Cell Carcinoma Based on a Validated Prognostic Nomogram and Risk Group Stratification System. *J Urol* (2005) 174:466–72, 472, 801. doi: 10.1097/01.ju.0000165572.38887.da
5. Moore AL, Batavia AA, Kuipers J, Singer J, Burcklen E, Schraml P, et al. Spatial Distribution of Private Gene Mutations in Clear Cell Renal Cell Carcinoma. *Cancers* (2021) 13:209–49. doi: 10.3390/cancers13092163
6. Marostica E, Barber R, Denize T, Kohane IS, Signoretti S, Golden JA, et al. Development of a Histopathology Informatics Pipeline for Classification and Prediction of Clinical Outcomes in Subtypes of Renal Cell Carcinoma. *Clin Cancer Res* (2021) 27:2868–78. doi: 10.1158/1078-0432.CCR-20-4119
7. Ross-Macdonald P, Walsh AM, Chasalow SD, Ammar R, Papillon-Cavanagh S, Szabo PM, et al. Molecular Correlates of Response to Nivolumab at Baseline and on Treatment in Patients With RCC. *J Immunother Cancer* (2021) 9:1506-20. doi: 10.1136/jitc-2020-001506
8. Akhtar M, Al-Bozom IA, Al Hussain T. Molecular and Metabolic Basis of Clear Cell Carcinoma of the Kidney. *Adv Anat Pathol* (2018) 25:189–96. doi: 10.1097/PAP.0000000000000185
9. Na X, Wu G, Ryan CK, Schoen SR, di'Santagnese PA, Messing EM. Overproduction of Vascular Endothelial Growth Factor Related to Von Hippel-Lindau Tumor Suppressor Gene Mutations and Hypoxia-Inducible Factor-1 Alpha Expression in Renal Cell Carcinomas. *J Urol* (2003) 170:588–92. doi: 10.1097/01.ju.0000074870.54671.98
10. Igarashi H, Esumi M, Ishida H, Okada K. Vascular Endothelial Growth Factor Overexpression is Correlated With Von Hippel-Lindau Tumor Suppressor Gene Inactivation in Patients With Sporadic Renal Cell Carcinoma. *Cancer-Am Cancer Soc* (2002) 95:47–53. doi: 10.1002/cncr.10635
11. Liu Z, Sun T, Piao C, Zhang Z, Kong C. METTL13 Inhibits Progression of Clear Cell Renal Cell Carcinoma With Repression on PI3K/AKT/mTOR/HIF-1α Pathway and C-Myc Expression. *J Transl Med* (2021) 19:209. doi: 10.1186/s12967-021-02879-2
12. Choueiri TK, Bauer TM, Papadopoulos KP, Plimack ER, Merchan JR, McDermott DF, et al. Inhibition of Hypoxia-Inducible Factor-2α in Renal Cell Carcinoma With Belzutifan: A Phase 1 Trial and Biomarker Analysis. *Nat Med* (2021) 27:802–5. doi: 10.1038/s41591-021-01324-7
13. Fan M, Xia P, Clarke R, Wang Y, Li L. Radiogenomic Signatures Reveal Multiscale Intratumour Heterogeneity Associated With Biological Functions and Survival in Breast Cancer. *Nat Commun* (2020) 11:4861. doi: 10.1038/s41467-020-18703-2
14. Lin P, Lin Y, Gao R, Wen R, Qin H, He Y, et al. Radiomic Profiling of Clear Cell Renal Cell Carcinoma Reveals Subtypes With Distinct Prognoses and Molecular Pathways. *Transl Oncol* (2021) 14:101078. doi: 10.1016/j.tranon.2021.101078
15. Calmon R, Dangouloff-Ros V, Varlet P, Deroulers C, Philippe C, Debily M, et al. Radiogenomics of Diffuse Intrinsic Pontine Gliomas (DIPGs): Correlation of Histological and Biological Characteristics With Multimodal MRI Features. *Eur Radiol* (2021) 1-12. doi: 10.1007/s00330-021-07991-x
16. Kirienko M, Sollini M, Corbetta M, Voulaz E, Gozzi N, Interlenghi M, et al. Radiomics and Gene Expression Profile to Characterise the Disease and Predict Outcome in Patients With Lung Cancer. *Eur J Nucl Med Mol Imaging* (2021) 48:3643–55. doi: 10.21203/rs.3.rs-157951/v1
17. Singh G, Manjila S, Sakla N, True A, Wardeh AH, Beig N, et al. Radiomics and Radiogenomics in Gliomas: A Contemporary Update. *Br J Cancer* (2021) 125:641–57. doi: 10.1038/s41416-021-01387-w
18. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
19. Lidgren A, Hedberg Y, Grankvist K, Rasmuson T, Bergh A, Ljungberg B. Hypoxia-Inducible Factor 1alpha Expression in Renal Cell Carcinoma Analyzed by Tissue Microarray. *Eur Urol* (2006) 50:1272–7. doi: 10.1016/j.eururo.2006.05.043
20. Schindl M, Schoppmann SF, Samonigg H, Hausmaninger H, Kwasny W, Gnant M, et al. Overexpression of Hypoxia-Inducible Factor 1alpha is Associated With an Unfavorable Prognosis in Lymph Node-Positive Breast Cancer. *Clinical Cancer Res* (2002) 8:1831–7.
21. Birner P, Schindl M, Obermair A, Plank C, Breitenecker G, Oberhuber G. Overexpression of Hypoxia-Inducible Factor 1alpha is a Marker for an Unfavorable Prognosis in Early-Stage Invasive Cervical Cancer. *Cancer Res* (2000) 60:4693–6.
22. Zhou F, Shen D, Xiong Y, Cheng S, Xu H, Wang G, et al. CTHRC1 Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Kidney Renal Papillary Cell Carcinoma and Kidney Renal Clear Cell Carcinoma. *Front Oncol* (2020) 10:570819. doi: 10.3389/fonc.2020.570819
23. Gkagkalidis K, Kampantais S, Dimitriadis G, Gourvas V, Kapoukranidou D, Mironidou-Tzouveleki M. Expression of HIF-2a in Clear-Cell Renal Cell

Carcinoma Independently Predicts Overall Survival. *Med Mol Morphol* (2020) 53:229–37. doi: 10.1007/s00795-020-00249-3
24. Fan X, Liu B, Wang Z, He D. TACC3 is a Prognostic Biomarker for Kidney Renal Clear Cell Carcinoma and Correlates With Immune Cell Infiltration and T Cell Exhaustion. *Aging* (2021) 13:8541–62. doi: 10.18632/aging.202668
25. Yang F, Zhao G, Ge L, Song Y, Hong K, Zhang H, et al. Identification of a Two-M6a RNA Methylation Regulator Risk Signature as an Independent Prognostic Biomarker in Papillary Renal Cell Carcinoma by Bioinformatic Analysis. *BioMed Res Int* (2021) 2021:4582082. doi: 10.1155/2021/4582082
26. Niu X, Zhu Z, Shao E, Bao J. ACE2 Is a Prognostic Biomarker and Associated With Immune Infiltration in Kidney Renal Clear Cell Carcinoma: Implication for COVID-19. *J Oncol* (2021) 2021:8847307. doi: 10.1155/2021/8847307
27. Raman SP, Chen Y, Schroeder JL, Huang P, Fishman EK. CT Texture Analysis of Renal Masses: Pilot Study Using Random Forest Classification for Prediction of Pathology. *Acad Radiol* (2014) 21:1587–96. doi: 10.1016/j.acra.2014.07.023
28. Chandarana H, Rosenkrantz AB, Mussi TC, Kim S, Ahmad AA, Raj SD, et al. Histogram Analysis of Whole-Lesion Enhancement in Differentiating Clear Cell From Papillary Subtype of Renal Cell Cancer. *Radiology* (2012) 265:790–8. doi: 10.1148/radiol.12111281
29. Juntu J, Sijbers J, De Backer S, Rajan J, Van Dyck D. Machine Learning Study of Several Classifiers Trained With Texture Analysis Features to Differentiate Benign From Malignant Soft-Tissue Tumors in T1-MRI Images. *J Magn Reson Imaging* (2010) 31:680–9. doi: 10.1002/jmri.22095

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Gao, Ye, Han, Wang, Jiang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 07 October 2021
doi: 10.3389/fonc.2021.704607

Exploring CT Texture Parameters as Predictive and Response Imaging Biomarkers of Survival in Patients With Metastatic Melanoma Treated With PD-1 Inhibitor Nivolumab: A Pilot Study Using a Delta-Radiomics Approach

Antonino Guerrisi[1†], Michelangelo Russillo[2†], Emiliano Loi[3], Balaji Ganeshan[4], Sara Ungania[3], Flora Desiderio[1], Vicente Bruzzaniti[3], Italia Falcone[2], Davide Renna[5], Virginia Ferraresi[2], Mauro Caterino[1], Francesco Maria Solivetti[1], Francesco Cognetti[2] and Aldo Morrone[6]*

[1] Radiology and Diagnostic Imaging Unit, Department of Clinical and Dermatological Research, San Gallicano Dermatological Institute IRCCS, Rome, Italy, [2] Medical Oncology Unit 1, Department of Clinical and Cancer Research IRCCS Regina Elena National Cancer Institute, Rome, Italy, [3] Medical Physics and Expert Systems Laboratory, 3 Department of Research and Advanced Technologies, Istituti Fisioterapici Ospitalieri - IRCCS Regina Elena National Cancer Institute, Rome, Italy, [4] Institute of Nuclear Medicine, Imaging Department, University College Hospital, London, United Kingdom, [5] Medical Oncology 1, IRCCS-Regina Elena National Cancer Institute, Rome, Italy, [6] Scientific Director, San Gallicano Dermatological Institute IRCCS, Rome, Italy

In the era of artificial intelligence and precision medicine, the use of quantitative imaging methodological approaches could improve the cancer patient's therapeutic approaches. Specifically, our pilot study aims to explore whether CT texture features on both baseline and first post-treatment contrast-enhanced CT may act as a predictor of overall survival (OS) and progression-free survival (PFS) in metastatic melanoma (MM) patients treated with the PD-1 inhibitor Nivolumab. Ninety-four lesions from 32 patients treated with Nivolumab were analyzed. Manual segmentation was performed using a free-hand polygon approach by drawing a region of interest (ROI) around each target lesion (up to five lesions were selected per patient according to RECIST 1.1). Filtration-histogram-based texture analysis was employed using a commercially available research software called TexRAD (Feedback Medical Ltd, London, UK; https://fbkmed.com/texrad-landing-2/) Percentage changes in texture features were calculated to perform delta-radiomics analysis. Texture feature kurtosis at fine and medium filter scale predicted OS and PFS. A higher kurtosis is correlated with good prognosis; kurtosis values greater than 1.11 for SSF = 2 and 1.20 for SSF = 3 were indicators of higher OS (fine texture: 192 HR = 0.56, 95% CI = 0.32–0.96, p = 0.03; medium texture: HR = 0.54, 95% CI = 0.29–0.99, p = 0.04) and PFS (fine texture:

OPEN ACCESS

Edited by:
Omar Sultan Al-Kadi, The University of Jordan, Jordan

Reviewed by:
Vishwa S. Parekh, Johns Hopkins University, United States
Federica Vernuccio, Azienda Ospedaliera Universitaria Policlinico Paolo Giaccone, Italy
Antonio Napolitano, Bambino Gesù Children Hospital (IRCCS), Italy

***Correspondence:**
Antonino Guerrisi
antonino.guerrisi@ifo.gov.it

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *03 May 2021*
Accepted: *08 September 2021*
Published: *07 October 2021*

Citation:
Guerrisi A, Russillo M, Loi E, Ganeshan B, Ungania S, Desiderio F, Bruzzaniti V, Falcone I, Renna D, Ferraresi V, Caterino M, Solivetti FM, Cognetti F and Morrone A (2021) Exploring CT Texture Parameters as Predictive and Response Imaging Biomarkers of Survival in Patients With Metastatic Melanoma Treated With PD-1 Inhibitor Nivolumab: A Pilot Study Using a Delta-Radiomics Approach. Front. Oncol. 11:704607. doi: 10.3389/fonc.2021.704607

HR = 0.53, 95% CI = 0.29–0.95, p = 0.03; medium texture: HR = 0.49, 209 95% CI = 0.25–0.96, p = 0.03). In delta-radiomics analysis, the entropy percentage variation correlated with OS and PFS. Increasing entropy indicates a worse outcome. An entropy variation greater than 5% was an indicator of bad prognosis. CT delta-texture analysis quantified as entropy predicted OS and PFS. Baseline CT texture quantified as kurtosis also predicted survival baseline. Further studies with larger cohorts are mandatory to confirm these promising exploratory results.

Keywords: melanoma, immunotherapy, x-ray computed tomography, biomarker, precision medicine, image analysis, radiomics

INTRODUCTION

Immunotherapy is changing the landscape of oncology (1, 2). In particular, immune checkpoint inhibitors such as programmed cell death protein 1 (PD-1) inhibitors are demonstrating an increased overall survival and progression-free survival (OS and PFS) in patients with metastatic melanoma (MM) (2–5). However, some patients benefit less than others and many factors are involved in the varied response rates (6). Intra/inter-tumor heterogeneity (ITH) may represent one of the reasons why some patients with MM do not gain real benefits from immunotherapy (7).

Early identification of non-responding patients avoids potential unwanted side effects and reduces the economic burden associated with unnecessary treatments on healthcare providers (8). The development of robust biomarkers for immunotherapy response represents an ongoing important challenge and focus area for research and development (9–11). Although many predictive markers for immunotherapy response have been investigated in MM, validated reliable biomarkers have not yet been identified (12).

Nowadays, imaging criteria (i.e., RECIST 1.1) based on computed tomography (CT) remains the gold standard for evaluating treatment response in clinical trials (13). CT remains indispensable for diagnosis and follow-up as it is reproducible, standardized, and suitable for extracting qualitative–quantitative data (14). However, imaging patterns of the immune mechanism and its associated/atypical response in some patients significantly differ from those seen with the use of more common cytotoxic agents (15). Indeed, treatment response after immunotherapy can be associated with pseudo-progression or hyper-progression (15). To overcome RECIST 1.1 limits, other imaging criteria (i.e., irRECIST, irRC, and iRECIST) have been proposed in clinical trials (16, 17). However, since available evidence for these criteria is still limited in melanoma and may not fully capture all patterns of clinical responses, caution is recommended in the use of these criteria in routine clinical practices (18). Thus, there is an urgent need to look for predictive biomarkers of immunotherapy response in patients with MM in order to provide robust and objective clinical end points.

Radiomics is a promising field of research that uses quantitative imaging methodological approaches that could support the oncologist in treatment decisions (19, 20). An image-based radiomics approach can analyze quantitative information from the whole tumor volume and the multiple lesions located in different anatomical sites within a single examination (21). CT texture analysis (CTTA) provides a panel of quantitative parameters reflecting intra/inter-tumor heterogeneity associated with a high-risk phenotype (22, 23). The delta-radiomics approach allows the evaluation of baseline and post-therapy changes in texture features within target lesions, in order to determine temporal changes in tumor heterogeneity (24).

The purpose of this study was to explore features extracted on baseline CT and post-treatment (delta-radiomics features) as predictors of OS and PFS in patients with MM treated with the PD-1 inhibitor Nivolumab, and its ability to act as a novel imaging biomarker for predicting survival.

MATERIALS AND METHODS

Study Population

Institutional review board approval was obtained for this retrospective study with a waiver of informed consent. At first, 78 MM patients treated with anti-PD-1 therapy, from January 2015 to February 2019, were included in the study. Patient data were extracted from our institutional oncologic digital database and RIS/PACS system (Centricity RIS/PACS, GE-Healthcare). Patients were included in the study only if they had both a contrast-enhanced CT and LDH measurements obtained at baseline/pre-therapy (within 1 month before the start of therapy) and initial post-therapy (about 3 months from the start of therapy). Crucial inclusion criteria of patients in this study required that contrast CT had been performed on the same scanner and with the same institutional protocol for the whole body. Schematic evolution of study population is represented in **Figure 1**.

Survival End Points

OS and PFS were chosen as end points. OS was defined as the time between the beginning of PD-1 inhibitor therapy and the death of patients. PFS was defined as the time between the beginning of PD-1 inhibitor therapy and the evidence of progressive disease (PD) at CT examination, according to

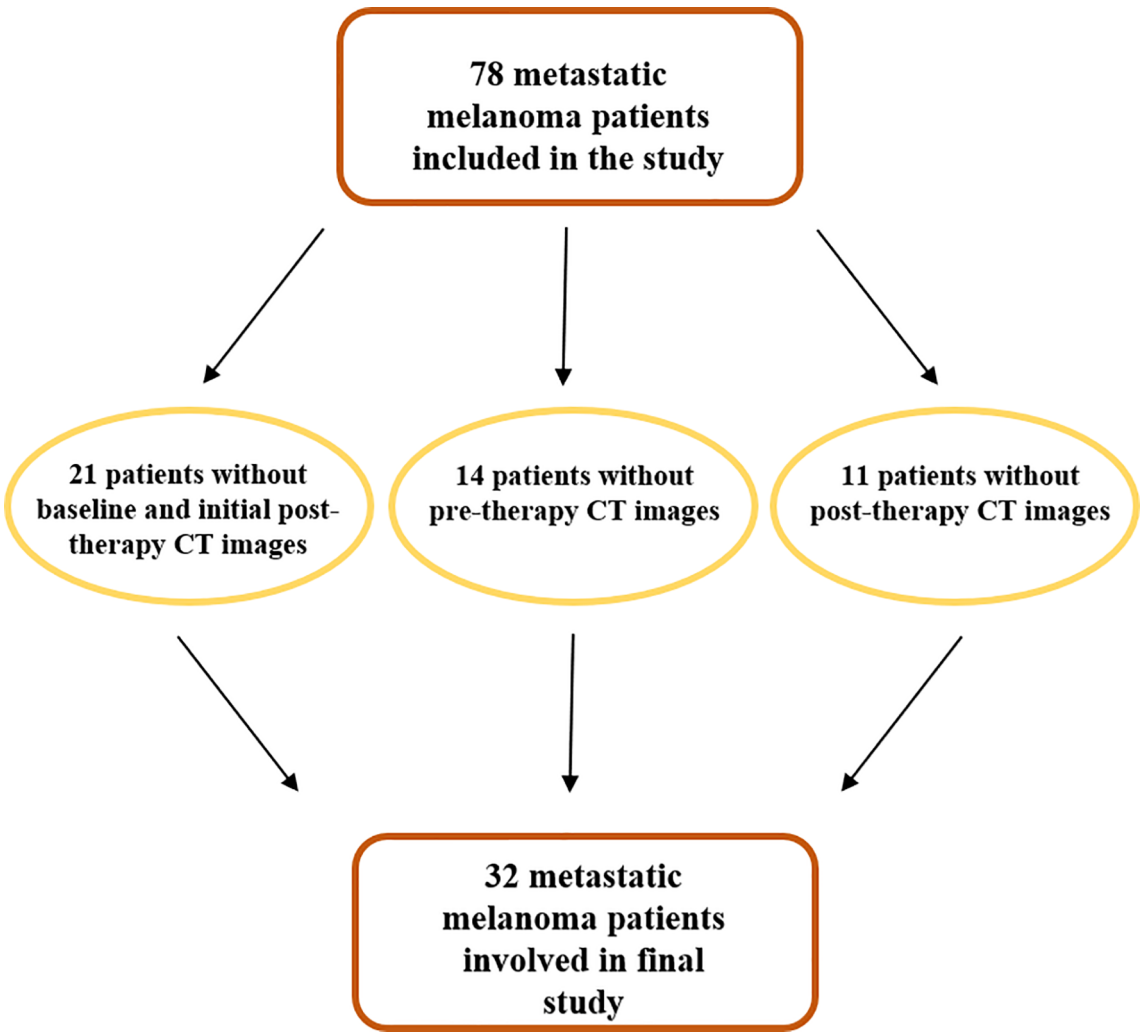


FIGURE 1 | Schematic characteristics of patients enrolled in the study.

RECIST 1.1. Patients, alive or without evidence of PD at the end of the follow-up, were censored to the last follow-up visit.

Clinical Markers

Clinical variables considered for multivariate analysis were serum lactate dehydrogenase (LDH), the percentage change in LDH before and after therapy (PERC-LDH), and the number of metastatic sites involved (4, 25). All clinical variables were dichotomized to the following cutoff points based on the literature: LDH cutoff = 1.5 × upper limit of normal, PERC-LDH cutoff = 20%, and number of metastatic sites cutoff = 3 (11, 25).

CT Examination

CT scans were acquired within a month prior to the start of therapy and, subsequently, every 3 months or in the presence of clinical signs suggesting PD. Images were acquired using a 128 MDCT scanner (Brilliance CT, Philips Healthcare), 64 × 0.625 (128) detector width (mm). The tube voltage was 100 kVp; the tube current product was determined using the automatic tube current modulation (ATCM) technique (Dose Right, Philips Healthcare), with 200 mAs as a reference parameter to minimize radiation exposure. Other scan parameters were as follows: pitch, 0.891; rotation time, 0.4 s; field of view (FOV), 350 mm; slice thickness, 2.5 mm; slice increment, 1.25 mm; matrix 512 × 512; pixel spacing, 0.98 mm × 0.98 mm. After an initial non-contrast scan, all patients received intravenous contrast (Ultravist® 370; Schering AG and Iomeron® 400; Bracco Imaging Italia s.r.l.) at a dose of 1.3 ml/kg and 1.5 ml/kg, respectively, and both at a rate of 3.0 ml/s *via* a 20-gauge catheter placed in the antecubital vein, followed by 40 ml of saline at the same rate. A multiphasic scan was initiated at 15, 55, and 150 s after CT attenuation of the aortic lumen at the level of the celiac trunk, reaching the trigger attenuation threshold of 150 HU, covering the whole body. Raw data of CT scans were

reconstructed with a standard filtered back-projection algorithm in the DICOM format, according to institutional protocol.

Region of Interest Segmentation and CT-Based Radiomic Texture Analysis

Manual segmentation was performed by drawing a region of interest (ROI) around each target lesion. Two radiologists with at least 10 years' experience in oncologic and melanoma imaging, A.G. and F.M.F., evaluated target lesions, applied the RECIST 1.1 criteria, and drew the ROIs in consensus; successively, the same radiologists have evaluated therapy response (up to five lesions were selected per patient). We basically employed what is normally done in routine clinical practice when assessing response to treatment in these ontological patients to be as close and relevant to current practice. No bin width or size was employed, but each individual integer value in the unfiltered and filtered texture map was used (in other words integer binning) in the quantification of texture parameter-based histogram and statistical approach. Lesions with the largest diameter of less than 5 mm were excluded from the analysis. Indeed, smaller lesions will have fewer pixels/distribution of gray-level intensities whereby the statistics may not be optimum. Also, as the filtration-histogram-based texture analysis employed in this study uses a spatial scale filter (SSF) that extracts and enhances features of different sizes corresponding to the SSF value, to reduce the impact of lesion size on the quantification of texture metrics, we recommend that lesions should have a maximum diameter of at least 5 mm to provide a decent number of pixels (statistics)/gray-level intensity variation for extraction of texture features. We can still extract features within lesions <5 mm, but one may not be able to compute the different texture parameters at the higher SSF values. Each ROI was drawn on the slice through the largest diameter of the target lesion around the peripheral margin. Air, streak artifacts, and dense calcifications were excluded from the ROI. Texture feature extraction was performed on the baseline CT and the initial post-therapy contrast enhanced images (in DICOM format) acquired during venous phase. CTTA comprised a filtration-histogram technique. The filtration step, using a Laplacian of Gaussian band-pass filter (similar to a non-orthogonal wavelet), extracted and enhanced texture features of different sizes and intensity variation, corresponding to a SSF that varied from 2 to 6 mm: SSF = 2 corresponds to a fine texture scale, SSF = 3–5 corresponds to a medium texture scale and SSF = 6 corresponds to a coarse texture scale. Quantification of textures at each filter and for unfiltered data (SSF = 0) was undertaken using statistical and histogram-based metrics such as mean intensity (reflects average brightness), standard deviation (SD—reflects deviation from mean), entropy (reflects irregularity of pixel distribution), mean of positive pixels (MPP—only reflects average brightness of positive pixel values), skewness (reflects asymmetry of the histogram distribution), and kurtosis (reflects pointedness/sharpness of the histogram distribution). A detailed description of the above image filtration and quantification is described (26) and a computer modeling study characterized the meaning of filtration-histogram-based texture features in terms of image features and how they relate to the different components (object size, density, and number) of heterogeneity (23). For each patient, the average value of each texture feature from all lesions was calculated and used for the baseline analysis. Percentage changes in texture features on post-treatment from baseline scan were also calculated as part of the delta-radiomics analysis.

Percentage changes in the abovementioned features are indicated with the prefix "perc" and are calculated as follows:

$$\text{perc} - \text{Var} = 100 * \frac{var_{pre} - var_{post}}{var_{pre}}$$

Where var_{pre} is the variable at baseline, var_{post} is the variable at initial post therapy and perc-Var is the percentage change. The whole process of ROI individual lesion segmentation and filtration histogram-based CTTA was undertaken using commercially available proprietary research software called TexRAD (Feedback Medical Ltd). **Figure 2** provides an illustration of the ROI segmentation and filtration-process as part of the CTTA.

Statistical Analysis

The average values of each texture feature from all lesions were calculated for each patient on baseline CT and post-therapy CT and used for the statistical analysis. Average values of the texture metrics were calculated from all the lesions for each patient. This was done because we have the outcome (OS and PFS) per patient. The percentage change was computed for each lesion first and then averaged across all the lesions for each patient. The following steps were undertaken to find independent predictors of OS and PFS from the group of extracted features. Feature selection was performed using LASSO-Cox regression to identify the best predictors of OS and PFS. Indeed, LASSO improves the reliability of regression using a regularization parameter to reduce overfitting and selects optimal predictors. In fact, the LASSO method provides non-zero regression coefficients only for the best predictive features. In this way, it is possible to eliminate overfitting problems. Finally, among these, the feature with the coefficient whose numerical value was greater in absolute value was chosen. In this study, a 10-fold cross-validation was employed for every regression (27). A hazard ratio (HR) and 95% confidence interval (CI) was provided for the best univariate markers of OS and PFS. The best univariate texture predictors were included in the multivariate Cox analysis along with clinical markers, to assess the independence and/or interaction of the significant univariate texture markers in terms of predicting survival. For each feature that resulted as an independent predictor of survival in the multivariate Cox model, Kaplan–Meier (KM) survival curves/analysis differentiated between patients with good prognosis from poor prognosis, based on a median threshold to separate the two prognostic groups. Differences between survival curves were evaluated using a non-parametric log rank test. A two-tailed *p*-value of less than 0.05 indicated a significant difference. All statistical analyses were performed using R-package software (version 3.6.3; R Foundation for Statistical Computing).

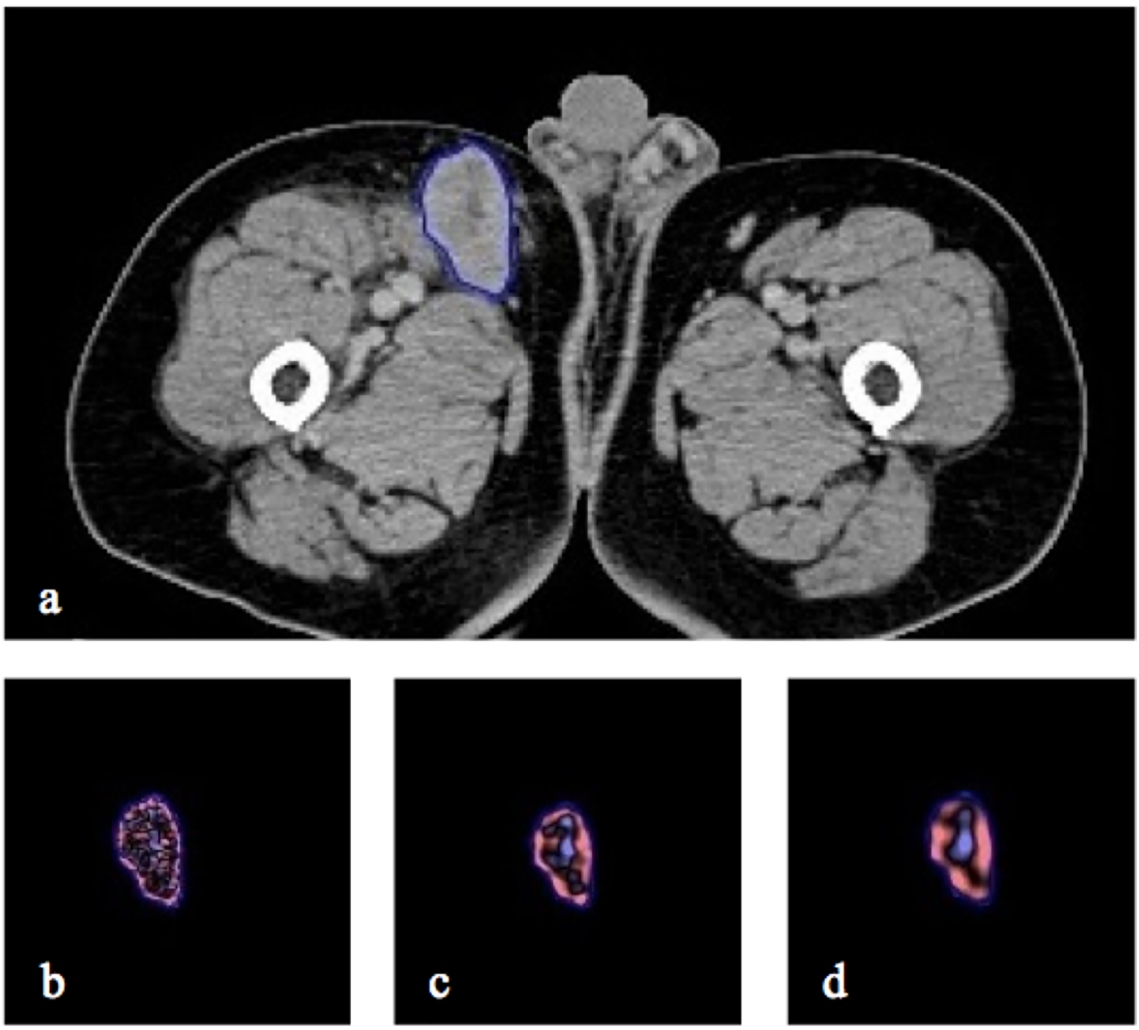


FIGURE 2 | CT axial image with right inguinal lymph node segmented in 2D **(A)** and the resulting. Illustration of different image filtration as part of CTTA at **(B)** fine (SSF = 2mm), **(C)** medium (SSF = 4mm) and **(D)** coarse (SSF = 6mm) texture scales.

TABLE 1 | Main demographic characteristics and relevant clinical data are reported in the table.

Variable	Total	Progression disease	Stable disease	Partial response
No. of patients	32	14	10	8
Age (years)	70 (29, 84)	67,5 (45, 82)	58 (29–79)	74 (54–84)
Sex (F, M)	9 F,23 M	6 F, 8 M	3 F, 7 M	8 M
No. of metastatic sites [≥3 (*N*), <3 (*N*)]	≥ (5), < (27)	≥ (3), < (11)	≥ (2), < (8)	≥ (0), < (8)
LDH pre therapy (UI/L)	389 (238, 827)	450 (313, 827)	331.9 (305–564)	367.5 (238–602)
LDH post therapy (UI/L)	419 (154, 803)	503 (202, 803)	390 (216–552)	287.5 (154–475)
LDH percentage Variation (%)	–3.73 (–50.16, 33.96)	9.28 (–49.6, 23.32)	–5.42 (–34.7–33.9)	–19.1 (–50.2–10.5)
ECOG	1 (0, 2)	1 (1, 2)	1 (0, 2)	0 (0, 1)
Hepatic lesions	Yes (9), No (23)	Yes (3), No (11)	Yes (3), No (7)	Yes (3), No (5)
Lung lesions	Yes (14), No (18)	Yes (9), No (5)	Yes (4), No (6)	Yes (5), No (13)
Mean target lesions size (mm)	30.4 (10, 140)	45.3 (10, 140)	38.7 (10, 88)	28.5 (10–40.3)
OS median/range (months)	31.6/2.73–66.8	20.6/2.73–48.6	38.6/3.5–56.6	54.6/29.3–66.8
PFS median/range (months)	15.1/2.5–52.7	5.3/2.6–24.3	23/2.5–50	46.1/13.2–52.6
Death	Yes (17), No (15)	Yes (12), No (2)	Yes (2), No (8)	Yes (3), No (5)

RESULTS

Patients Characteristics

Of the 78 patients analyzed for the study, only 32 (mean age 60 years; standard deviation 13.9) were really enrolled (see **Figure 2**). **Table 1** presents the demographics, clinical, and follow-up/survival information for all patients and sub-groups based on treatment response.

Overall Survival Analysis

LASSO-Cox regression demonstrated baseline CT kurtosis at fine (SSF = 2) and medium (SSF = 3) texture scales predicting OS (fine texture: LASSO coefficient = −0.3; medium texture: LASSO coefficient = −0.35). A higher baseline kurtosis value was associated with good OS (fine texture: HR = 0.56, 95% CI = 0.32–0.96, p = 0.03; medium texture: HR = 0.54, 95% CI = 0.29–0.99, p = 0.04) (**Tables 2a, b**). LASSO-Cox regression also demonstrated delta-radiomics CT analysis, particularly the percentage change in entropy (Perc-ENTRO) at medium (SSF = 4) and coarse (SSF = 6) scales to predict OS (SSF = 4, LASSO coefficient = 0.02; SSF = 6, LASSO coefficient = 0.03). An increase in Perc-ENTRO was associated with poorer OS (SSF = 4, HR = 1.06, 95% CI = 1.01–1.11, p = 0.05; SSF = 6, HR = 1.06, 95% CI = 1.01–1.11, p = 0.01). A separate multivariate Cox regression analysis, which includes each significant univariate texture marker along with LDH and the number of metastatic sites, are presented in **Tables 2a–c**. Baseline CT kurtosis at fine (**Table 2a**) and medium (**Table 2b**) texture scales were predictors of OS, independent of LDH and the number of metastatic sites. Perc-ENTRO was a predictor at medium (SSF = 4) and coarse scales (SSF = 6) (**Tables 2c, d**).

A Kaplan–Meier survival analysis for significant texture predictors of OS, based on their respective median cutoff, is presented in **Table 3**; Kaplan Meier survival curves are presented in **Figures 3A–D**.

Progression-Free Survival Analysis

LASSO-Cox regression demonstrated baseline CT kurtosis at fine (SSF = 2) and medium (SSF = 3) texture scales predicting PFS (fine texture: LASSO coefficient = −0.35; medium texture: LASSO coefficient = −0.28). **Tables 4a, b** show that a higher baseline kurtosis value was associated with good PFS (fine texture: HR = 0.53, 95% CI = 0.29–0.95, p = 0.03; medium texture: HR = 0.49, 95% CI = 0.25–0.96, p = 0.03). LASSO-Cox regression also demonstrated delta-radiomics CT analysis, particularly Perc-ENTRO at medium (SSF = 4) and coarse (SSF = 6) scales to predict PFS (LASSO coefficient = 0.05 for SSF = 4, 0.04 for SSF = 5 and 0.04 for SSF = 6). An increase in Perc-ENTRO was associated with poorer PFS, for example, SSF = 4 (HR = 1.07, 95% CI = 1.01–1.13, p = 0.009). A separate multivariate Cox regression analysis, including each significant univariate texture marker along with LDH and a number of metastatic sites, is presented in **Tables 4a–c**. Baseline CT kurtosis at fine (**Table 4a**) and medium (**Table 4b**) texture scales were predictors of PFS, independent of LDH and number of metastatic sites. Perc-ENTRO and number of metastatic sites were independent predictors of PFS (**Tables 4c, d**).

Kaplan–Meier survival analysis for significant texture predictors of PFS based on their respective median cutoff are presented in **Table 5**; Kaplan–Meier survival curves are presented in **Figures 4A–D**.

TABLE 2 | (a–d) Multivariate Cox regression analysis including individual significant univariate texture parameters (selected using LASSO) and clinical variables for predicting OS.

	Covariate	HR	95% CI	*p*-value
a.	Kurtosis (SSF = 2)	0.56	0.32–0.96	0.03
	LDH	0.29	0.07–1.12	0.09
	No. of metastatic sites	3.02	0.84–10.9	0.07
b.	Kurtosis (SSF = 3)	0.54	0.29–0.99	0.04
	LDH	0.33	0.08–1.21	0.10
	No. of metastatic sites	3.17	0.87–11.5	0.07
c.	Perc-ENTRO (SSF = 4)	1.06	1.01–1.11	0.05
	No. of met sites	4.02	1.14–14.9	0.04
	Perc-LDH	0.2	0.06–0.7	0.02
d.	Perc-ENTRO (SSF = 6)	1.06	1.01–1.11	0.01
	No. of met sites	4.25	1.14–15.8	0.03
	Perc-LDH	0.2	0.06–0.7	0.01

HR, hazard ratio; CI, confidence interval.

TABLE 3 | Kaplan–Meier analysis for significant texture predictors of OS.

Features	Cutoff	Median OS (months) for patients above the cutoff	Median OS (months) for patients below/equal to the cutoff	*p*-value
Kurtosis (SSF = 2)	1.11	46.6	29.8	0.01
Kurtosis (SSF = 3)	1.20	40.5	30.3	0.02
Perc-ENTRO (SSF = 4)	5	35.7	23.3	0.008
Perc-ENTRO (SSF = 6)	5	35.7	23.3	0.0005

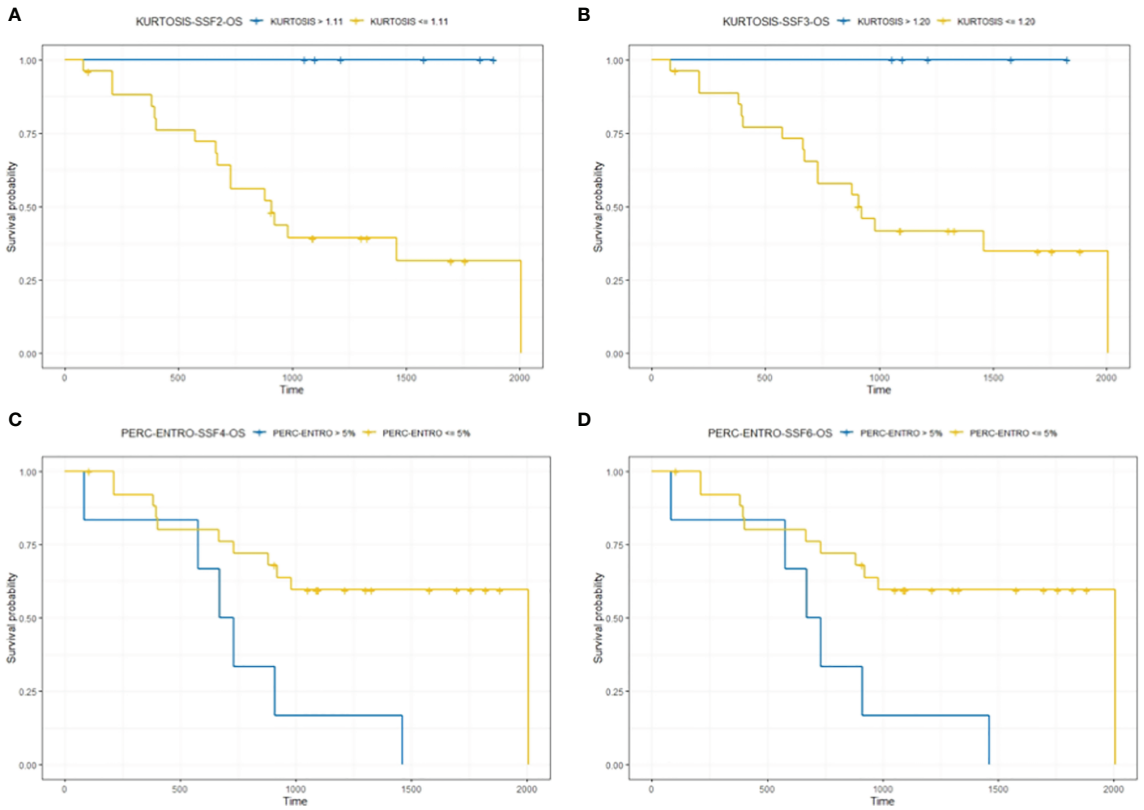


FIGURE 3 | The figure compares survival curve for baseline CT texture parameter kurtosis at fine (SSF = 2) and medium (SSF = 3) texture scales and delta-analysis quantified Perc-ENTRO at medium (SSF = 4) and coarse (SSF = 6) scales for OS. Patients in the good prognostic group, as identified by baseline kurtosis are fine (SSF = 2) and medium (SSF = 3) texture scales, had an improved median survival of around 17 **(A)** and 10 months **(B)** respectively, compared to the poor prognostic group. Furthermore, the good prognostic group, which was defined using Kurtosis, demonstrated zero mortality. Patients in the good prognostic group, as identified by Perc-ENTRO at medium and coarse scales, improved median survival by around 11 months **(C, D)** compared to the poor prognostic group.

DISCUSSION

Our pilot study indicates that baseline and post-therapy contrast-enhanced CT-based radiomics texture features of MM have the potential to act as imaging biomarkers of outcome in terms of OS and PFS in patients treated with Nivolumab. Both kurtosis at baseline CT and percentage change in entropy without filtration, between post-treatment and baseline CT, are the best predictors of outcome and can potentially act as a predictive and response biomarker, respectively. The addition

TABLE 4 | (a–d) Multivariate Cox regression analysis including individual significant univariate texture parameters (selected using LASSO) and clinical variables for predicting PFS.

	Covariate	HR	CI	*p*-value
a.	Kurtosis (SSF = 2)	0.53	0.29–0.95	0.03
	LDH	0.27	0.06–1.26	0.09
	No. of metastatic sites	2.59	0.75–8.9	0.13
b.	Kurtosis (SSF = 3)	0.49	0.25–0.96	0.03
	LDH	0.30	0.06–1.4	0.12
	No. of metastatic sites	2.62	0.76–8.99	0.13
c.	Perc-ENTRO (SSF = 4)	1.07	1.01–1.13	0.009
	No. of met sites	3.98	1.14–13.6	0.03
	Perc-LDH	0.28	0.07–1.1	0.07
d.	Perc-ENTRO (SSF = 6)	1.07	1.02–1.13	0.005
	No. of met sites	4.10	1.19–14.1	0.02
	Perc-LDH	0.27	0.07–1.1	0.06

HR, hazard ratio; CI, confidence interval.

TABLE 5 | Kaplan–Meier analysis for significant texture predictors of PFS.

Features	Cutoff	Median PFS (months) for patients above the cutoff	Median PFS (months) for patients below/equal to the cutoff	*p*-value
Kurtosis (SSF = 2)	1.11	37.8	10.1	0.008
Kurtosis (SSF = 3)	1.20	26	13.2	0.04
Perc-ENTRO (SSF = 4)	5.0	18.7	10.5	0.01
Perc-ENTRO (SSF = 6)	5.0	18.7	10.5	0.0005

of delta-radiomics increases the available quantitative information related to the spatial and temporal intra-tumor heterogeneity (ITH), potentially reflecting tumor phenotypical changes over time that are crucial in assessing response to immunotherapy.

In the era of artificial intelligence and machine learning (ML), quantitative texture-based radiomic features extracted from medical images can provide objective information and, therefore, play an important role in decision support in cancer care pathways and patient management (28, 29). Over the past few years, CTTA has been acknowledged as a promising quantitative imaging tool allowing for measurement of the spatial ITH by analyzing the gray-level distribution of CT images (30). Previous studies explored biological correlates for CTTA, confirming an association between CT heterogeneity and a hypoxic and angiogenic tumor microenvironment (23, 30); at the same time, it is hypothesized that tumor angiogenesis (TA) may have value in predicting patient survival in different cancers (31–37).

In MM, multivariate analysis confirmed that texture variables are significant predictors of survival and suitable to build a prognostic index/composite score along with established important clinical markers, such as LDH and number of metastatic sites (4, 38, 39). We found kurtosis at fine (SSF = 2) and medium (SSF = 3) texture scales in order to best predict the outcome (OS and PFS) of patients treated with Nivolumab

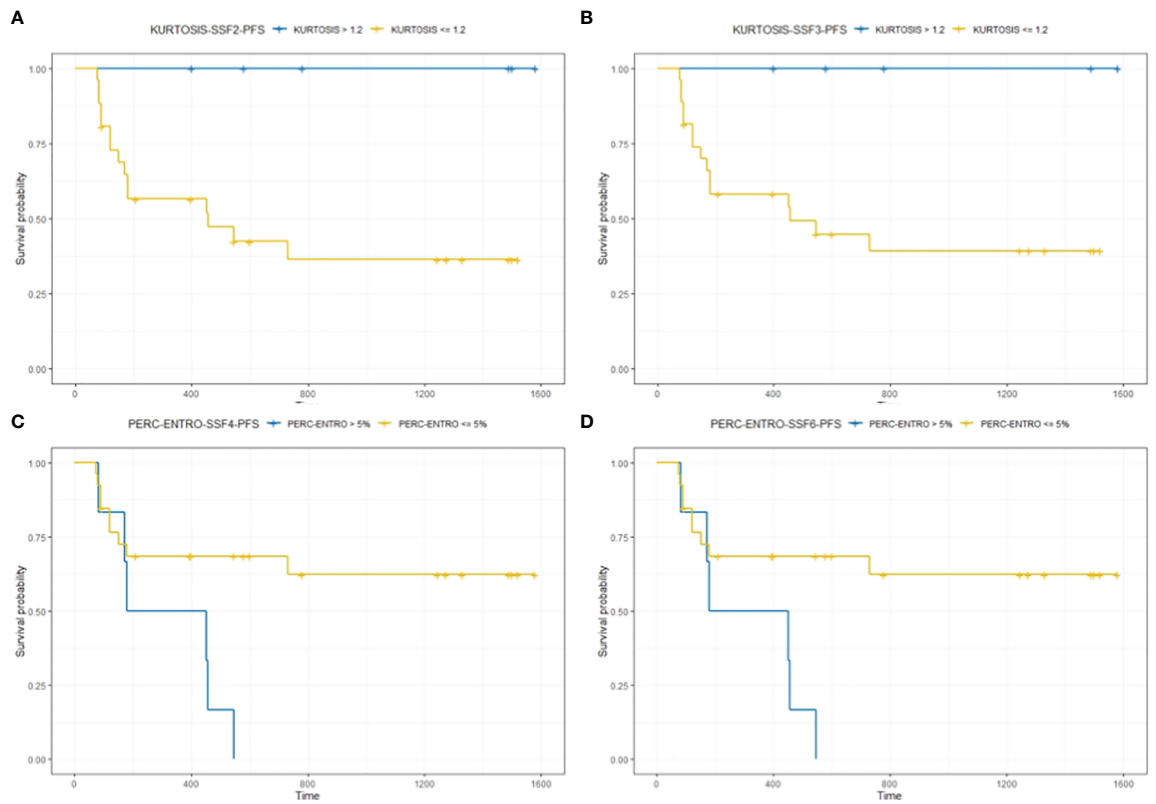


FIGURE 4 | The figure compares survival curves for baseline CT texture parameter kurtosis at fine (SSF = 2) and medium (SSF = 3) texture scales, and delta-analysis quantified Perc-ENTRO at medium and coarse scales for PFS. Patients in the good prognostic group, as identified by baseline kurtosis at fine (SSF = 2) and medium (SSF = 3) texture scales, had an improved median PFS of around 28 **(A)** and 13 months **(B)** respectively, compared to the respective poor prognostic group. Furthermore, the good prognostic group defined using Kurtosis demonstrated zero progression. Patients in the good prognostic group, as identified by Perc-ENTRO, had an improved median PFS of around 8 months **(C, D)** compared to the poor prognostic group.

at baseline CT. Kurtosis is a measure of the sharpness/pointedness of distribution in the histogram of images. A higher value indicates increased tissue contrast, which could be associated with tissue vascularity (23). In our study, hypothesizing kurtosis at fine to medium texture scales may reflect contrast medium changes directly/indirectly associated with small/medium-sized blood vessels (micro-vasculature) within the tumor microenvironment.

In patients treated with immunotherapy, tumor vasculature may impact the host immune response (40). In fact, functional abnormalities of tumor blood vessels, such as architectural defects, can limit lymphocyte recruitment. A correlation between the presence of tumor-infiltrating lymphocytes (TILs) and an improved prognosis has been demonstrated in cancer patients treated with immunotherapy (40, 41). We can assume that a high value of kurtosis may reflect a higher T-cell infiltration within a lesion, resulting in a "T-cell inflamed" phenotype ("hot tumors") associated with lower tumor blood vessel defects and a potential higher responsiveness to PD-1 inhibitor therapy (42). In a recent retrospective study, Schraag et al. have proposed that the kurtosis as an independent predictor of OS in MM patients treated with different immunotherapy regimes, corroborating our results (43). However, some important differences with our pilot study should be emphasized. Indeed, the authors of this study have involved a more heterogeneous population of MM patients (the patients enrolled were treated with CTLA4 and PD1 inhibitors) and have focused their analysis only on the largest target lesion. We believe that this choice may be limiting because it is less representative of the total tumor burden. Texture-based radiomic analysis is dependent on the variation/distribution of the pixel intensities within the ROI enclosing the lesion. Smaller lesions will have fewer pixels/distribution of gray-level intensities whereby the statistics may not be optimum. Also, the lesions <5 mm size may not be clinically relevant. This hypothesis is confirmed by RECIST and other criteria that recommend analysis of up to five target lesions. We basically employed what is normally done in routine clinical practice when assessing response to treatment in these oncological patients (e.g., RECIST 1.1 criteria dictate selecting up to five lesions per patient) to be as close and relevant to current practice. Applying different weightings to different types of lesions or only considering one lesion per patient was not explored as it is not something done routinely and there is no biological rationale "currently" to utilize this. Nevertheless, there are very interesting points to explore in the future, such as certain types (based on anatomy) of metastatic lesions could be weighed differently and could be more robust and sensitive/accurate in early prediction of responders from non-responders.

Contrary to our study that has analyzed the response to Nivolumab, Durot et al. have investigated the role of CTTA in predicting response to immunotherapy in MM patients treated with another anti-PD-1 inhibitor (Pembrolizumab). In this study, the authors reported the role of skewness (i.e., asymmetry of the histogram) as a potential predictor of outcome (31). Pembrolizumab and Nivolumab belong to the same family (anti-PD-1) and are similar (44); it has been suggested that differences observed in clinical data between these two drugs are unlikely to be drug-dependent and are likely to be due to drug-independent differences (44). We can assume that differences in our results are more likely to be due to the small patient population in both studies, which could amplify potential individual patient characteristics, as well as drug administration and imaging protocols. The potential influence of iterative algorithms and contrast administration protocols on radiomic analysis needs to be better investigated (45). In any case, we hypothesize that both kurtosis and skewness may be picking up similar image characteristics. Skewness reflects the preponderance of object brightness/darkness, which could again reflect an aspect of tumor vascularity similar to kurtosis. Therefore, both kurtosis and skewness could have a potential role in reflecting vascular and non-vascular morphological changes within the lesion. These inter-relationships and the above hypotheses need to be investigated in further studies with a larger patient population.

It is well established that greater tumor heterogeneity is an indicator of poor clinical prognosis. Furthermore, tumor-induced angiogenesis contributes to a disorganized micro environment leading to tortuous architecture of the vasculature, eventually resulting in the formation of hypoxic voids and necrosis. Consequently, different sub-clonal cell populations within a tumor arise over time, contributing to different phenotypes (7, 40). It is important to note that phenotypic plasticity can occur due to non-genetic factors and a variation exists at multiple omics levels (46); "phenotypic" delta signatures for predicting survival could better correlate with other molecular markers over time. In the delta-radiomics analysis, the Perc-ENTRO significantly predicted both OS and PFS. Lesions with high entropy usually correlated with higher heterogeneity. In other works, entropy was correlated with the outcome of therapy (34, 37, 47).

The exhibit of a more disorganized microenvironment within the lesion may appear as increased imaging heterogeneity, i.e., higher entropy on the post-treatment scan in comparison to the baseline scan, giving rise to an increased Perc-ENTRO, a response marker indicating the worst outcome (OS and PFS) in our study. We believe that the combination of texture-based radiomic analysis and recent multi-omics approaches can help implement precision medicine and a robust decision-making tool in patients (48–51). This tool/approach is particularly relevant in a MM setting having a complex biology, inter/intra-tumor heterogeneity and is one of the most aggressive cancers (7).

Our study has some limitations. Firstly, the relatively small patient population makes our hypothesized associations speculative and exploratory in nature. Moreover, only a manual 2D analysis on cross-sectional CT images was performed. However, this was undertaken in consensus with two expert radiologists in CT oncological imaging, thereby limiting the potential bias. In any case, both approaches were found to capture heterogeneity and were good predictors of survival (52). Moreover, although our results are promising, of course a validation data set is missing and data have to be confirmed in future studies in order to validate the method.

In fact, this is a pilot study purely for exploratory purposes; the focus of this pilot/proof-of-principle study was to assess the

potential of each texture parameter from the baseline and post-treatment CT scan using filtration-histogram-based technique (a technique that has undergone proper validation as evidenced from the numerous papers using this technique) to predict outcome (OS and PFS) post immunotherapy and compare to existing clinical/imaging markers. Future studies could implement a more complex ML algorithm (incorporating the significant texture features identified in this pilot study) and employ normalization/standardization approaches in a larger cohort, which could further be randomly divided into training and validation sub-cohorts. Furthermore, as the study population in this pilot/proof of principle was very small at 32 patients, the idea was not to employ complex approaches (e.g., ML and numerous radiomic features) as we do not have a dedicated training and testing dataset. Based on this exploratory study, we do propose to undertake a larger prospective study in the future where we will apply the above suggested methodologies (e.g., ML techniques), employing a comprehensive radiomic approach including the filtration-histogram-based technique in addition to higher-order statistics, shape parameters, and training and testing cohort. Another limitation of this pilot study is the lack of the assessment of intra and inter-reader agreement, as only information from segmentation from a consensus reading was performed; but we would like to point out the qualification process undertaken by the filtration-histogram based texture analysis evidenced from numerous other peer-reviewed publications.

However, we would like to point out the quantification process undertaken by the filtration-histogram-based texture analysis evidenced from numerous other peer-reviewed publications: (26, 53, 54). Specifically, the filtration step part of the texture analysis technique reduces the impact of image photon noise, thereby minimizing the impact of image acquisition variation and therefore the quantification of texture features using histogram and statistical approach reflects biologically relevant heterogeneity. The use of the filtration-histogram technique further mitigates the need for the use of larger number of higher-order statistics, which are more abstract in nature, are less reproducible, and increase false discovery rate.

CONCLUSIONS

In conclusion, our study demonstrates the potential role of kurtosis to select MM patients with improved OS and PFS at baseline CT, as an independent predictor of outcome ("predictive-biomarker"). In delta-radiomics analysis, we found Perc-ENTRO to be a good independent predictor for both OS and PFS in MM patients treated with Nivolumab ("response-biomarker"). If this method is validated, we hypothesize that these parameters could potentially improve better patient selection and the response evaluation to immune check point inhibitors and, therefore, be used as an adjunct in decision-making and optimal patient management. On the basis of our promising preliminary results, further studies with a larger MM population treated with PD-1 inhibitors are needed to investigate the usefulness of delta-radiomics based CT texture features in a multi-omics approach.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because we are not yet able to predict whether the necessary internal approvals and permissions and patient consents can be shared, even if anonymously. Requests to access the datasets should be directed to emiliano.loi88@gmail.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by IFO—Istituti Fisioterapici Ospitalieri. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Conceptualization: AG, MR, and AM. Data curation: EL, BG, SU, VB, and DR. Formal analysis: AG, EL, BG, VB, and FS. Investigation: AG and VF. Methodology: AG, EL, VF, and FS. Resources: MR. Supervision: AG, FC, and AM. Validation: BG. Visualization: FD and MC. Writing—original draft: AG and EL. Writing—review and editing: IF, MR, and BG. All authors contributed to the article and approved the submitted version.

FUNDING

BG is affiliated to Institute of Nuclear Medicine, University College London/University College London Hospitals, which receives proportional funding through the UK National Institute of Health Research/Biomedical Research Centre funding scheme. No specific grant number is associated with the presented researcher.

ACKNOWLEDGMENTS

We wish to acknowledge the help provided by Claudia Messina for assistance with literature research. We are particularly grateful for the assistance given by Isabella Sperduti for comments on statistical analyses that greatly improved our manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.704607/full#supplementary-material

REFERENCES

1. Emens LA, Ascierto PA, Darcy PK, Demaria S, Eggermont AMM, Redmond WL, et al. Cancer Immunotherapy: Opportunities and Challenges in the Rapidly Evolving Clinical Landscape. *Eur J Cancer* (2017) 81:116–29. doi: 10.1016/j.ejca.2017.01.035
2. Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Rutkowski P, Lao CD, et al. Five-Year Survival With Combined Nivolumab and Ipilimumab in Advanced Melanoma. *N Engl J Med* (2019) 381(16):1535–46. doi: 10.1056/NEJMoa1910836
3. Shah DJ, Dronca RS. Latest Advances in Chemotherapeutic, Targeted, and Immune Approaches in the Treatment of Metastatic Melanoma. *Mayo Clin Proc* (2014) 89(4):504–19. doi: 10.1016/j.mayocp.2014.02.002
4. Ascierto PA, Long GV, Robert C, Brady B, Dutriaux C, Di Giacomo AM, et al. Survival Outcomes in Patients With Previously Untreated BRAF Wild-Type Advanced Melanoma Treated With Nivolumab Therapy: Three-Year Follow-Up of a Randomized Phase 3 Trial. *JAMA Oncol* (2019) 5(2):187–94. doi: 10.1001/jamaoncol.2018.4514
5. Johnson DB, Peng C, Sosman JA. Nivolumab in Melanoma: Latest Evidence and Clinical Potential. *Ther Adv Med Oncol* (2015) 7(2):97–106. doi: 10.1177/1758834014567469
6. Sambi M, Bagheri L, Szewczuk MR. Current Challenges in Cancer Immunotherapy: Multimodal Approaches to Improve Efficacy and Patient Response Rates. *J Oncol* (2019) 2019:4508794. doi: 10.1155/2019/4508794
7. Grzywa TM, Paskal W, Wlodarski PK. Intratumor and Intertumor Heterogeneity in Melanoma. *Transl Oncol* (2017) 10(6):956–75. doi: 10.1016/j.tranon.2017.09.007
8. Ventola CL. Cancer Immunotherapy, Part 3: Challenges and Future Trends. *P T* (2017) 42(8):514–21.
9. Buder-Bakhaya K, Hassel JC. Biomarkers for Clinical Benefit of Immune Checkpoint Inhibitor Treatment-A Review From the Melanoma Perspective and Beyond. *Front Immunol* (2018) 9:1474. doi: 10.3389/fimmu.2018.01474
10. Kitano S, Nakayama T, Yamashita M. Biomarkers for Immune Checkpoint Inhibitors in Melanoma. *Front Oncol* (2018) 8:270. doi: 10.3389/fonc.2018.00270
11. Wagner NB, Forschner A, Leiter U, Garbe C, Eigentler TK. S100B and LDH as Early Prognostic Markers for Response and Overall Survival in Melanoma Patients Treated With Anti-PD-1 or Combined Anti-PD-1 Plus Anti-CTLA-4 Antibodies. *Br J Cancer* (2018) 119(3):339–46. doi: 10.1038/s41416-018-0167-x
12. Garbe C, Amaral T, Peris K, Hauschild A, Arenberger P, Bastholt L, et al. European Consensus-Based Interdisciplinary Guideline for Melanoma. Part 2: Treatment - Update 2019. *Eur J Cancer* (2020) 126:159–77. doi: 10.1016/j.ejca.2019.11.015
13. Schwartz LH, Litiere S, de Vries E, Ford R, Gwyther S, Mandrekar S, et al. RECIST 1.1-Update and Clarification: From the RECIST Committee. *Eur J Cancer* (2016) 62:132–7. doi: 10.1016/j.ejca.2016.03.081
14. Freeman M, Laks S. Surveillance Imaging for Metastasis in High-Risk Melanoma: Importance in Individualized Patient Care and Survivorship. *Melanoma Manage* (2019) 6(1):MMT12. doi: 10.2217/mmt-2019-0003
15. Borcoman E, Nandikolla A, Long G, Goel S, Le Tourneau C. Patterns of Response and Progression to Immunotherapy. *Am Soc Clin Oncol Educ Book* (2018) 38:169–78. doi: 10.1200/EDBK_200643
16. Oxnard GR, Morris MJ, Hodi FS, Baker LH, Kris MG, Venook AP, et al. When Progressive Disease Does Not Mean Treatment Failure: Reconsidering the Criteria for Progression. *J Natl Cancer Inst* (2012) 104(20):1534–41. doi: 10.1093/jnci/djs353
17. Somarouthu B, Lee SI, Urban T, Sadow CA, Harris GJ, Kambadakone A. Immune-Related Tumour Response Assessment Criteria: A Comprehensive Review. *Br J Radiol* (2018) 91(1084):20170457. doi: 10.1259/bjr.20170457
18. Dromain C, Beigelman C, Pozzessere C, Duran R, Digklia A. Imaging of Tumour Response to Immunotherapy. *Eur Radiol Exp* (2020) 4(1):2. doi: 10.1186/s41747-019-0134-1
19. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, de Jong EEC, van Timmeren J, et al. Radiomics: The Bridge Between Medical Imaging and Personalized Medicine. *Nat Rev Clin Oncol* (2017) 14(12):749–62. doi: 10.1038/nrclinonc.2017.141
20. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
21. Guerrisi A, Loi E, Ungania S, Russillo M, Bruzzaniti V, Elia F, et al. Novel Cancer Therapies for Advanced Cutaneous Melanoma: The Added Value of Radiomics in the Decision Making Process-A Systematic Review. *Cancer Med* (2020) 9(5):1603–12. doi: 10.1002/cam4.2709
22. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006
23. Miles KA, Ganeshan B, Hayball MP. CT Texture Analysis Using the Filtration-Histogram Method: What do the Measurements Mean? *Cancer Imaging* (2013) 13(3):400–6. doi: 10.1102/1470-7330.2013.9045
24. Smith AD, Gray MR, del Campo SM, Shlapak D, Ganeshan B, Zhang X, et al. Predicting Overall Survival in Patients With Metastatic Melanoma on Antiangiogenic Therapy and RECIST Stable Disease on Initial Posttherapy Images Using CT Texture Analysis. *AJR Am J Roentgenol* (2015) 205(3):W283–93. doi: 10.2214/AJR.15.14315
25. Hamid O, Puzanov I, Dummer R, Schachter J, Daud A, Schadendorf D, et al. Final Analysis of a Randomised Trial Comparing Pembrolizumab *Versus* Investigator-Choice Chemotherapy for Ipilimumab-Refractory Advanced Melanoma. *Eur J Cancer* (2017) 86:37–45. doi: 10.1016/j.ejca.2017.07.022
26. Ganeshan B, Goh V, Mandeville HC, Ng QS, Hoskin PJ, Miles KA. Non-Small Cell Lung Cancer: Histopathologic Correlates for Texture Parameters at CT. *Radiology* (2013) 266(1):326–36. doi: 10.1148/radiol.12112428
27. Tibshirani R. The Lasso Method for Variable Selection in the Cox Model. *Stat Med* (1997) 16(4):385–95. doi: 10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3
28. Duffy DJ. Problems, Challenges and Promises: Perspectives on Precision Medicine. *Brief Bioinform* (2016) 17(3):494–504. doi: 10.1093/bib/bbv060
29. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts H. Machine Learning Methods for Quantitative Radiomic Biomarkers. *Sci Rep* (2015) 5:13087. doi: 10.1038/srep13087
30. Lubner MG, Smith AD, Sandrasegaran K, Sahani DV, Pickhardt PJ. CT Texture Analysis: Definitions, Applications, Biologic Correlates, and Challenges. *Radiographics* (2017) 37(5):1483–503. doi: 10.1148/rg.2017170056
31. Durot C, Mule S, Soyer P, Marchal A, Grange F, Hoeffel C. Metastatic Melanoma: Pretreatment Contrast-Enhanced CT Texture Parameters as Predictive Biomarkers of Survival in Patients Treated With Pembrolizumab. *Eur Radiol* (2019) 29(6):3183–91. doi: 10.1007/s00330-018-5933-x
32. Zhang H, Graham CM, Elci O, Griswold ME, Zhang X, Khan MA, et al. Locally Advanced Squamous Cell Carcinoma of the Head and Neck: CT Texture and Histogram Analysis Allow Independent Prediction of Overall Survival in Patients Treated With Induction Chemotherapy. *Radiology* (2013) 269(3):801–9. doi: 10.1148/radiol.13130110
33. Ng F, Ganeshan B, Kozarski R, Miles KA, Goh V. Assessment of Primary Colorectal Cancer Heterogeneity by Using Whole-Tumor Texture Analysis: Contrast-Enhanced CT Texture as a Biomarker of 5-Year Survival. *Radiology* (2013) 266(1):177–84. doi: 10.1148/radiol.12120254
34. Ganeshan B, Skogen K, Pressney I, Coutroubis D, Miles K. Tumour Heterogeneity in Oesophageal Cancer Assessed by CT Texture Analysis: Preliminary Evidence of an Association With Tumour Metabolism, Stage, and Survival. *Clin Radiol* (2012) 67(2):157–64. doi: 10.1016/j.crad.2011.08.012
35. Ganeshan B, Panayiotou E, Burnand K, Dizdarevic S, Miles K. Tumour Heterogeneity in Non-Small Cell Lung Carcinoma Assessed by CT Texture Analysis: A Potential Marker of Survival. *Eur Radiol* (2012) 22(4):796–802. doi: 10.1007/s00330-011-2319-8
36. Alessandrino F, Gujrathi R, Nassar AH, Alzaghal A, Ravi A, McGregor B, et al. Predictive Role of Computed Tomography Texture Analysis in Patients With Metastatic Urothelial Cancer Treated With Programmed Death-1 and Programmed Death-Ligand 1 Inhibitors. *Eur Urol Oncol* (2020) 3(5):680–6. doi: 10.1016/j.euo.2019.02.002
37. Giganti F, Antunes S, Salerno A, Ambrosi A, Marra P, Nicoletti R, et al. Gastric Cancer: Texture Analysis From Multidetector Computed Tomography as a Potential Preoperative Prognostic Biomarker. *Eur Radiol* (2017) 27(5):1831–9. doi: 10.1007/s00330-016-4540-y
38. Wolchok JD, Chiarion-Sileni V, Gonzalez R, Rutkowski P, Grob JJ, Cowey CL, et al. Overall Survival With Combined Nivolumab and Ipilimumab in

Advanced Melanoma. *N Engl J Med* (2017) 377(14):1345–56. doi: 10.1056/NEJMoa1709684
39. Zhang J, Yao YH, Li BG, Yang Q, Zhang PY, Wang HT. Prognostic Value of Pretreatment Serum Lactate Dehydrogenase Level in Patients With Solid Tumors: A Systematic Review and Meta-Analysis. *Sci Rep* (2015) 5:9800. doi: 10.1038/srep09800
40. Georganaki M, van Hooren L, Dimberg A. Vascular Targeting to Increase the Efficiency of Immune Checkpoint Blockade in Cancer. *Front Immunol* (2018) 9:3081. doi: 10.3389/fimmu.2018.03081
41. Hendry SA, Farnsworth RH, Solomon B, Achen MG, Stacker SA, Fox SB. The Role of the Tumor Vasculature in the Host Immune Response: Implications for Therapeutic Strategies Targeting the Tumor Microenvironment. *Front Immunol* (2016) 7:621. doi: 10.3389/fimmu.2016.00621
42. Maleki Vareki S. High and Low Mutational Burden Tumors *Versus* Immunologically Hot and Cold Tumors and Response to Immune Checkpoint Inhibitors. *J Immunother Cancer* (2018) 6(1):157. doi: 10.1186/s40425-018-0479-7
43. Schraag A, Klumpp B, Afat S, Gatidis S, Nikolaou K, Eigentler TK, et al. Baseline Clinical and Imaging Predictors of Treatment Response and Overall Survival of Patients With Metastatic Melanoma Undergoing Immunotherapy. *Eur J Radiol* (2019) 121:108688. doi: 10.1016/j.ejrad.2019.108688
44. Fessas P, Lee H, Ikemizu S, Janowitz T. A Molecular and Preclinical Comparison of the PD-1-Targeted T-Cell Checkpoint Inhibitors Nivolumab and Pembrolizumab. *Semin Oncol* (2017) 44(2):136–40. doi: 10.1053/j.seminoncol.2017.06.002
45. Kim H, Park CM, Lee M, Park SJ, Song YS, Lee JH, et al. Impact of Reconstruction Algorithms on CT Radiomic Features of Pulmonary Tumors: Analysis of Intra- and Inter-Reader Variability and Inter-Reconstruction Algorithm Variability. *PloS One* (2016) 11(10):e0164924. doi: 10.1371/journal.pone.0164924
46. McQuerry JA, Chang JT, Bowtell DDL, Cohen A, Bild AH. Mechanisms and Clinical Implications of Tumor Heterogeneity and Convergence on Recurrent Phenotypes. *J Mol Med (Berl)* (2017) 95(11):1167–78. doi: 10.1007/s00109-017-1587-4
47. Haider MA, Vosough A, Khalvati F, Kiss A, Ganeshan B, Bjarnason GA. CT Texture Analysis: A Potential Tool for Prediction of Survival in Patients With Metastatic Clear Cell Carcinoma Treated With Sunitinib. *Cancer Imaging* (2017) 17(1):4. doi: 10.1186/s40644-017-0106-8
48. Finotello F, Eduati F. Multi-Omics Profiling of the Tumor Microenvironment: Paving the Way to Precision Immuno-Oncology. *Front Oncol* (2018) 8:430. doi: 10.3389/fonc.2018.00430
49. Bodalal Z, Trebeschi S, Nguyen-Kim TDL, Schats W, Beets-Tan R. Radiogenomics: Bridging Imaging and Genomics. *Abdom Radiol (NY)* (2019) 44(6):1960–84. doi: 10.1007/s00261-019-02028-w
50. Gaiser MR, von Bubnoff N, Gebhardt C, Utikal JS. Liquid Biopsy to Monitor Melanoma Patients. *J Dtsch Dermatol Ges* (2018) 16(4):405–14. doi: 10.1111/ddg.13461
51. Valenti F, Falcone I, Ungania S, Desiderio F, Giacomini P, Bazzichetto C, et al. Precision Medicine and Melanoma: Multi-Omics Approaches to Monitoring the Immunotherapy Response. *Int J Mol Sci* (2021) 22(8):3837. doi: 10.3390/ijms22083837
52. Ng F, Kozarski R, Ganeshan B, Goh V. Assessment of Tumor Heterogeneity by CT Texture Analysis: Can the Largest Cross-Sectional Area be Used as an Alternative to Whole Tumor Analysis? *Eur J Radiol* (2013) 82(2):342–8. doi: 10.1016/j.ejrad.2012.10.023
53. Lewis MA, Ganeshan B, Barnes A, Bisdas S, Jaunmuktane Z, Brandner S, et al. Filtration-Histogram Based Magnetic Resonance Texture Analysis (MRTA) for Glioma IDH and 1p19q Genotyping. *Eur J Radiol* (2019) 113:116–23. doi: 10.1016/j.ejrad.2019.02.014
54. Atkinson C, Ganeshan B, Endozo R, Wan S, Aldridge MD, Groves AM, et al. Radiomics-Based Texture Analysis of (68)Ga-DOTATATE Positron Emission Tomography and Computed Tomography Images as a Prognostic Biomarker in Adults With Neuroendocrine Cancers Treated With (177)Lu-DOTATATE. *Front Oncol* (2021) 11:686235. doi: 10.3389/fonc.2021.686235

Conflict of Interest: One of the authors, BG, (who was not a data controller for this study) is the co-founder/co-inventor of TexRAD texture analysis software used in this study and a shareholder (not an employee) of Feedback Plc., a UK based company that owns, develops, and markets the TexRAD texture analysis software.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Guerrisi, Russillo, Loi, Ganeshan, Ungania, Desiderio, Bruzzaniti, Falcone, Renna, Ferraresi, Caterino, Solivetti, Cognetti and Morrone. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 11 October 2021
doi: 10.3389/fonc.2021.743490

CT Radiomics Model for Predicting the Ki-67 Index of Lung Cancer: An Exploratory Study

Qing Fu, Shun li Liu, Da peng Hao, Ya bin Hu, Xue jun Liu, Zaixian Zhang, Wen hong Wang, Xiao yan Tang, Chuan yu Zhang and Shi he Liu**

The Affiliated Hospital of Qingdao University, Qingdao, China

OPEN ACCESS

Edited by:
Omar Sultan Al-Kadi,
The University of Jordan, Jordan

Reviewed by:
Anis Ahmad,
University of Miami Health System, United States
Allegra Conti,
University of Rome Tor Vergata, Italy

****Correspondence:***
Chuan yu Zhang
zhangchuanyu0926@163.com
Shi he Liu
shihe528@126.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *18 July 2021*
Accepted: *15 September 2021*
Published: *11 October 2021*

Citation:
Fu Q, Liu Sl, Hao Dp, Hu Yb, Liu Xj, Zhang Z, Wang Wh, Tang Xy, Zhang Cy and Liu Sh (2021) CT Radiomics Model for Predicting the Ki-67 Index of Lung Cancer : An Exploratory Study. Front. Oncol. 11:743490. doi: 10.3389/fonc.2021.743490

Objective: To establish a radiomics signature and a nomogram model based on enhanced CT images to predict the Ki-67 index of lung cancer.

Methods: From January 2014 to December 2018, 282 patients with lung cancer who had undergone enhanced CT scans and Ki-67 examination within 2 weeks were retrospectively enrolled and analyzed. The clinical data of the patients were collected, such as age, sex, smoking history, maximum tumor diameter and serum tumor markers. Our primary cohort was randomly divided into a training group (n=197) and a validation group (n=85) at a 7:3 ratio. A Ki-67 index ≤ 40% indicated low expression, and a Ki-67 index > 40% indicated high expression. In total, 396 radiomics features were extracted using AK software. Feature reduction and selection were performed using the lasso regression model. Logistic regression analysis was used to establish a multivariate predictive model to identify high and low Ki-67 expression in lung cancer. A nomogram integrating the radiomics score was established based on multiple logistic regression analysis. Area under the curve (AUC) was used to evaluate the prediction efficiency of the radiomics signature and nomogram.

Results: The AUC,sensitivity, specificity and accuracy of the radiomics signature in the training and validation groups were 0.88 (95% CI: 0.82~0.93),79.2%,84.3%,81.2% and 0.86 (95% CI: 0.78~0.94),74.6%,88.1%,79.8%, respectively. A nomogram combining radiomics features and clinical risk factors (smoking history and NSE) was developed. The AUC, sensitivity, specificity and accuracy were 0.87 (95% CI: 0.80~0.95), 75.0%, 90.2% and 83.5% in the validation group, respectively.

Conclusion: The radiomics signature and nomogram based on enhanced CT images provide a way to predict the Ki-67 expression level in lung cancer.

Keywords: lung cancer, Ki-67, tomography, X ray, radiomics model, classification

INTRODUCTION

Lung cancer is one of the most common malignant tumors that endangers human health and life, ranking first in the number of cancer-related deaths (1, 2). The proliferation mode and speed of tumor cells are related to the malignancy and prognosis of lung cancer (3–5). Ki-67 is a nuclear antigen expressed by cells in the proliferation phase that accurately reflect the proliferation activity of cells. Because of its short half-life, Ki-67 is significantly better than those proliferating cell nuclear antigens with a long half-life in evaluating the proliferative activity of tumor (6–9).

Presently, Ki-67 expression in lung cancer must be determined by biopsy or surgical histopathology, but biopsy samples generally represent only a small part of the tumor tissue. Because of the heterogeneous expression of Ki-67 in tumors, the Ki-67 index obtained by needle biopsy samples cannot fully and accurately represent the Ki-67 level of the entire tumor. This situation leads to deviations in results and even misdiagnosis and nonoptimal clinical decision-making (10). As a new research field, radiomics has obvious advantages in assessing tumor heterogeneity. It can discover and analyze different cell phenotypes of tumors (11–13) and provide comprehensive and quantitative tumor measurements, including texture, intensity, heterogeneity and morphological information, enabling a comprehensive analysis of the tumor phenotype (14–16). Zhou B et al. (17) found that twelve CT radiomic features were significantly associated with the Ki-67 of lung cancer, but they did not build a predictive model. Gu Q et al. (18) built a machine learning-based radiomics classifier to predict the Ki-67 index of non-small cell lung cancer, however, their study did not include cases with small cell lung cancer, so the model was not applicable to all patients with lung cancer. Moreover, these past studies have not established a nomogram model that combines radiomic features with clinical parameters, which may have better prediction efficiency. This study aimed to establish a radiomics signature based on enhanced CT images and a nomogram based on radscores and clinical parameters to predict the Ki-67 index of lung cancer.

MATERIALS AND METHODS

Data Cohort

This retrospective study was approved by the Institutional Review Board. The data of 2286 consecutive patients with lung cancer confirmed by surgery between January 2014 and December 2018 were identified for this retrospective study.

The inclusion criteria were as follows: (1) a diagnosis of lung cancer by surgical pathological specimens and immunohistochemical Ki-67 examination and (2) dual-phase enhanced chest CT examination before surgery.

The exclusion criteria were as follows: (1) no Ki-67 immunohistochemistry or enhanced CT examination at our hospital (n=1280); (2) poor image quality or image layer thickness greater than or equal to 5 mm (n=130); (3) incomplete clinical data (n=335); (4) prior neoadjuvant treatment before surgery (n=91); (5) small lesions (long diameter < 1 cm) (n=104); (6) other primary malignancies in the same period (n=64).

Two hundred eighty-two patients (178 men and 104 women with a mean age of 62.0 ± 8.9 years) were enrolled in our study (**Figure 1**), 158 patients with adenocarcinoma (ACC), 69 with squamous cell carcinoma (SCC), and 55 with neuroendocrine carcinoma (NEC) (including 38 patients with small cell lung cancer, 13 patients with large cell lung cancer, and 4 patients with carcinoid cancer). Using a stratified random sampling method, the patients were divided into a training group and a validation group at a ratio of 7:3.

The following information of the enrolled patients was evaluated: smoking history, sex, age, maximum tumor diameter,

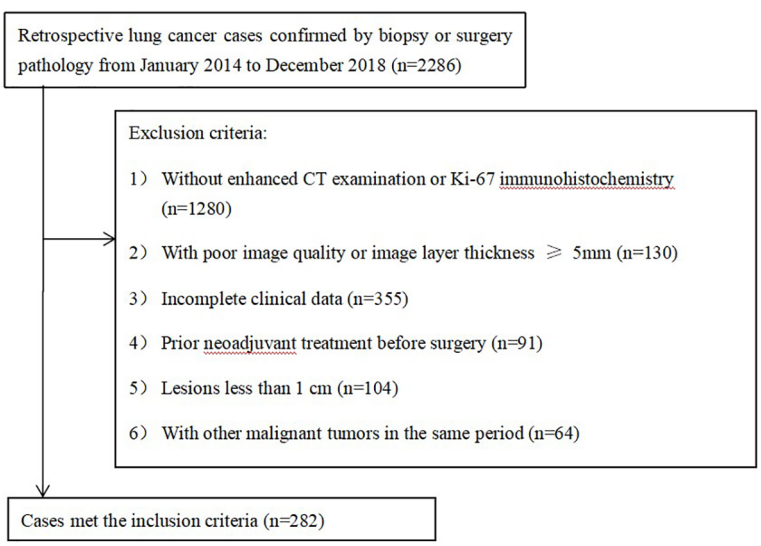


FIGURE 1 | Flow diagram of the patient selection.

pathological type, and serum tumor markers, including neuron-specific enolase (NSE) serum, gastrin-releasing peptide precursor (ProGRP), carcinoembryonic antigen (CEA), squamous cell carcinoma antigen (SCCA) and cytokeratin 19 fragment (cYFRA21-1). Smoking history was defined as smoking for more than one year and smoking more than 20 cigarettes per day on average. Histological classification was based on H&E staining according to the WHO classification of malignant lung tumors.

CT Protocol

All patients were scanned using a SOMATOM (Siemens Medical Systems, Germany) scanner or Brilliance iCT 256 (Philips Healthcare, Netherlands) scanner. The scanning parameters were as follows: tube voltage: 120 kVp; pixel size: 512×512; detector collimation: 64×0.6 and 128×0.625 mm; slice thickness: 5 mm; and reconstructed section thickness: 1 mm. Contrast-enhanced CT images were obtained by intravenous injection of 1.0 ml/kg of contrast material (iohexol injection; 300 mg/ml; Beijing, China) at a rate of 3.0-3.5 ml, followed by a saline flush (20 mL). CT images were acquired at 25 seconds and 70 seconds after the start of contrast medium injection, corresponding to the arterial and venous phases, respectively.

Ki-67 Expression Measurement

Formalin-fixed, paraffin-embedded tissue sections with a thickness of 4 μm were created. The sections were then dried, dewaxed with xylene, rinsed in graded ethanol and rehydrated in double-distilled water. Immunohistochemistry (IHC) staining was performed using a Ki-67 protein antibody (Santa Cruz Biotechnology, California, USA) diluted 1:100. Cells with brown nuclei were considered positive.

The whole specimen was scanned, and positive cells in five areas with the highest positive density were selected, then a percentage of positively labeled cells were determined by counting more than 1000 tumor nuclei at 400 magnification. Because the most active part of tumor proliferation can represent the degree of tumor malignancy and affect the prognosis of patients. So according to previous relevant studies (19–21), the Ki-67 index in this study was the average value of the five areas with the highest percentage of Ki-67-labeled cells, and according to previous studies (22), low Ki-67 expression was defined as ≤ 40% positive staining, while over 40% positive staining was defined as high Ki-67 expression.

Image Normalization and Feature Extraction

The workflow of radiomics implementation is displayed in **Figure 2**. All the images were normalized by z-score transformation, with intensity ranges for each imaging modality across all subjects with a mean of 0 and a standard deviation of 1.ITK-SNAP software (http://www.itksnap.org, version: 3.8.0, USA) was used to outline the lesion on the CT image with the largest diameter of the lesion. All lesion ROI outlining was completed by two radiologists with 12 years (HYB) and 10 years (SLL) of chest CT diagnosis experience, and the intragroup correlation coefficient (ICC) between among the observers was calculated. The ROI was outlined by the HYB once, and the second ROI was performed after a week to assess the observer's ICC. SLL only performed the ROI once to evaluate the ICC between this physician and HYB. ICC>0.75 considers that the consistency is good. Both radiologists were blinded to the patient's clinicopathological information. Commercial software (Analysis Toolkit 1.0.3; GE Healthcare, USA) was used to extract features. In total, 396 quantized features were extracted, such as 9 form factor features, 10 Haralick features, 11 gray level size zone matrix (GLSZM) features, 42 histogram features, 48 gray level cooccurrence matrix (GLCM) features with an offset of 1/4/7, and 60 gray level run-length matrix (GLRLM) features with an offset of 1/4/7.

Development of the Radiomics Signature, Clinical Model, Radiomics Nomogram

To minimize overfitting, the least absolute shrinkage and selection operator (lasso) regression method was used to select the most valuable features from the primary datasets, and then validated in the validation cohort. The linear combination of selected features was used to calculate the radiomic scores (Rad-scores) for each patient. For validation, we evaluated the difference of rad score between the training set and the verification set, and calculated the sample size of the

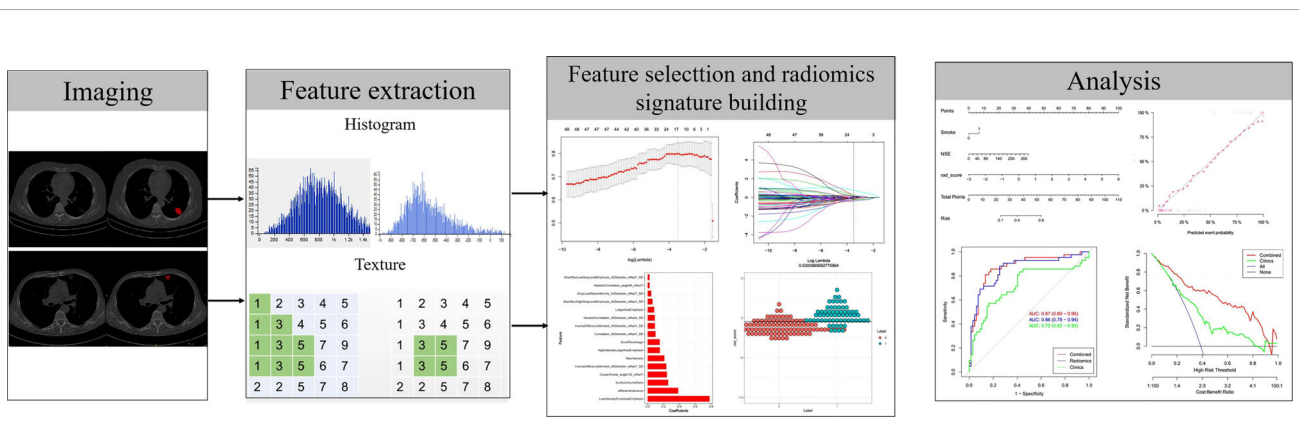


FIGURE 2 | Flowchart of radiomics implementation in this study.

verification set using the method of "comparing the mean between the two groups", which meets the statistical power of more than 0.8. The cut-off value obtained from the training set was used to calculate the metrics of the validation set.The predictive accuracy of the radiomics signature was quantified by the area under the receiver operating characteristic (ROC) curve (AUC) in both the training and validation sets.

Univariate logistic regression was used to select clinical risk factors for high Ki-67 lung cancer. The clinical features and imaging omics features with P<0.05 were used to develop a predictive model to distinguish low Ki-67 lung cancer from high Ki-67 lung cancer using multivariate logistic regression in the primary cohort. In logistic regression, backward stepwise selection was applied using a likelihood ratio test with Akaike's information criterion as the stopping rule.

To provide clinicians with a quantitative tool to predict the Ki-67 level of lung cancer, a radiomics nomogram was established based on multivariate logistic analysis in the primary cohort. The algorithm built by the training set was used to calculate the Rad score in the validation set.

Validation and Assessment of the Radiomics Nomogram

We assessed the value of the radiomics nomograms in training (n=197) and validation (n=85) data sets, including identification, calibration, and clinical value, and quantified the differential performance of AUC. The Hosmer-Lemeshow test was used together with the calibration curve to determine the goodness-of-fit of the nomogram. The validation data set was used to test the internal value of the radiomics nomogram.

Decision curve analysis (DCA) was used to calculate the net benefit of the threshold probability range in the training and validation data sets to estimate whether the nomogram was sufficiently reliable for clinical use. The net benefit was determined by calculating the difference between the true positive rate and weighted false positive rate of different threshold probabilities in the validation set. A "decision curve" was then drawn based on the threshold probability.

Statistical Analysis

R statistical software (http://www.Rproject.org, version 3.4.4) was used for statistical analysis. Lasso regression was performed using the "glmnet" package. The "RMS" package was used to construct multivariate logistic regression, nomogram and calibration charts. DCA was performed using the "DCA. R" function. ROC curves were drawn and analyzed using the "proc" package. The Kolmogorov-Smirnov test was used to test the normality of the quantitative data, and the measurement data conforming to the normal distribution were expressed as $x \pm s$. Counting data was expressed in frequency. Chi-squared test or Fisher's exact test was used to compare the count data between groups, and independent samples t test was used to compare the measurement data. P<0.05 indicated a statistically significant difference.

RESULTS

Comparison of the Clinical Data Results of the Training and Validation Groups and the Low Ki-67 and High Ki-67 Lung Cancer Groups

No significant differences were found in age, sex, the tumor diameter or pathological type between the training and validation groups (P>0.05; **Table 1**). Statistically significant differences were found in sex, age, and the pathological type between the high and low Ki-67 expression groups (P<0.05) (**Table 2**). High Ki-67 expression was more common in men, elderly individuals, and SCC patients.

Extraction/Selection of Radiomics Features and Construction of the Radiomics Signature

First, we performed repeatability evaluation (between and within data sets with a consistency coefficient> 0.75), and then removed highly correlated features (correlation coefficient> 0.6). Finally, we used lasso logistic regression to screen out 16 features (**Figures 3A–C**),including Low Intensity Small Area Emphasis, difference Variance, Surface Volume Ratio, Cluster Shade_ angle135_offset7, Inverse Difference Moment_All Direction_ offset7_SD, Max Intensity, High Intensity Large Area Emphasis, Zone Percentage,Correlation_All Direction_offset7_SD, Inverse Difference Moment_All Direction_offset1_SD, Haralick Correlation_All Direction_offset4_SD, Large Area Emphasis, Short Run High Grey Level Emphasis_All Direction_offset1_SD, Grey Level Non uniformity_All Direction_offset7_SD, Haralick

TABLE 1 | Comparison of the clinical data and pathological staging results of patients in the training and validation groups.

Clinical feature		Training group (n = 197)	Verification group (n = 85)	p value	t value or $\chi 2$ value
Sex	Male	126 (64.0%)	52 (61.2%)	0.66	0.20
	Female	71 (36.0%)	33 (38.8%)		
Age (years)		61.6 ± 8.9	62.4 ± 9.2	0.49	0.69
Tumor maximum diameter (cm)		4.1 ± 2.2	4.0 ± 2.3	0.58	0.55
Smoking	Yes	111 (56.3%)	41 (48.2%)	0.18	1.81
	No	86 (43.7%)	44 (51.8%)		
Pathological type	ACC	110 (55.8%)	48 (56.5%)	0.91	0.19*
	SCC	48 (24.4%)	21 (24.7%)		
	NEC	39 (19.8%)	16 (18.8%)		

*$\chi 2$ value (continuous variables were analyzed by t test, and categorical variables were analyzed by chi-square test).

TABLE 2 | Comparison of the clinical data results between the low Ki-67 and high Ki-67 lung cancer groups.

Clinical feature		low Ki group (n = 175)	high Ki group (n =1 07)	p value	t value or $\chi 2$ value
Age (years)		62.2 ± 8.8	61.8 ± 9.1	0.74	0.33
Sex	Male	93 (53.1%)	85 (79.4%)	<0.01	19.73
	Female	82 (46.9%)	22 (20.6%)		
Smoking	Yes	76 (43.4%)	76 (71.0%)	<0.01	20.36
	No	99 (56.6%)	31 (29.0%)		
Tumor diameter (cm)		3.7 ± 2.0	4.6 ± 2.4	<0.01	-3.46
Pathological type	ACC	135 (77.1%)	23 (21.5%)	<0.01	89.77
	SCC	29 (16.6%)	40 (37.4%)		
	NEC	11 (6.3%)	44 (41.1%)		

*Continuous variables were analyzed by t test, and categorical variables were analyzed by $\chi 2$ test. *u value: the overall variance of the two groups of data was uneven, and the rank-sum test was performed.*

Correlation_angle45_offset7, Short Run Low Grey Level Emphasis_All Direction_offset7_SD.The Rad-scores of each patient in the training and validation sets are shown in **Figures 4A, B**.

Predictive Efficacy of the Imaging Radiomics Signature

The ROC curves of the training and validation groups are shown in **Figures 5A, B**. The AUC, accuracy sensitivity, specificity, positive predictive value, and negative predictive value were 0.88 (95% CI: 0.82~0.93), 81.2%, 79.8%, 84.4%, 88.9%, and 72% in the training group and 0.86 (95% CI: 0.78~0.94), 79.8%, 74.6%, 88.1%, 90.9%, and 68.5% in the validation group, respectively.

Establishment of a Nomogram Combining Radiomics With Clinical Risk Factors

Univariate analysis showed that the clinical factors were significantly related to the classification of low Ki-67 lung cancer and high Ki-67 lung cancer (**Table 3**). They include serum NSE and smoking ($P < 0.05$). The results of multivariate logistic regression analysis suggested that smoking, serum NSE and the rad score were independent predictors of low and high Ki-67 lung cancer classification (**Table 4**). A radiomics nomogram incorporating the predictive factors (including smoking, NSE, and the Rad score) was constructed (**Figure 6**).

The calibration curve showed that the predicted probability of the nomogram was consistent with the pathological findings (**Figure 7**). The results in **Table 5** and **Figure 8** show that the nomogram had better prediction efficiency than the radiomics signature and clinical model. The AUC value of the nomogram in the validation set was 0.87 (95% CI: 0.80-0.95), the accuracy was 0.83 (95% CI: 0.75-0.90), the sensitivity was 75.0%, and the specificity was 90.2%. **Figure 9** shows the DCA of the radiomics nomogram. When the threshold probability is in the range of 0.1–1.0, the radiomics nomogram is superior to the model of "all treatment" and "no treatment" strategies.

DISCUSSION

Ki-67 nuclear protein is a marker of cell cycle and proliferation (9, 23, 24) and is typically used to estimate the population of proliferating cells. In malignant tumors, the percentage of Ki-67-

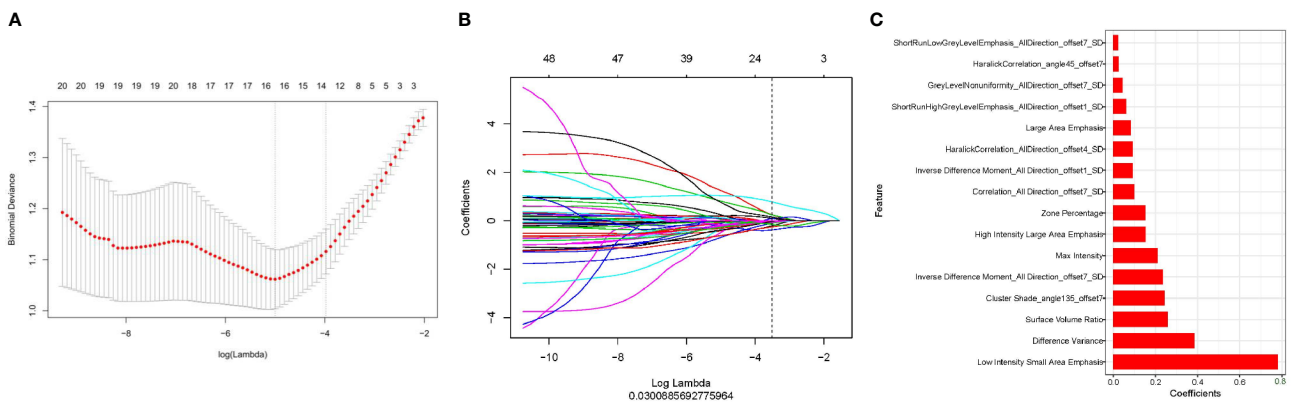


FIGURE 3 | Use of lasso logistic regression to select features. **(A)** Binomial deviation versus parameter λ. Least absolute shrinkage and selection operator (LASSO) regression was used to screen the radiomic features, and cross-validation was used to select the optimal model parameter λ. The vertical axis is the binomial deviation, and the horizontal axis is the log (λ) value. λ, which represents the smallest binomial deviation of the model, is the optimal value (vertical dashed line). **(B)** Graph of the variation of the imaging omics feature coefficient with λ. The number above indicates the number of filtered features. **(C)** Screened 16 radiomics features and their weights.

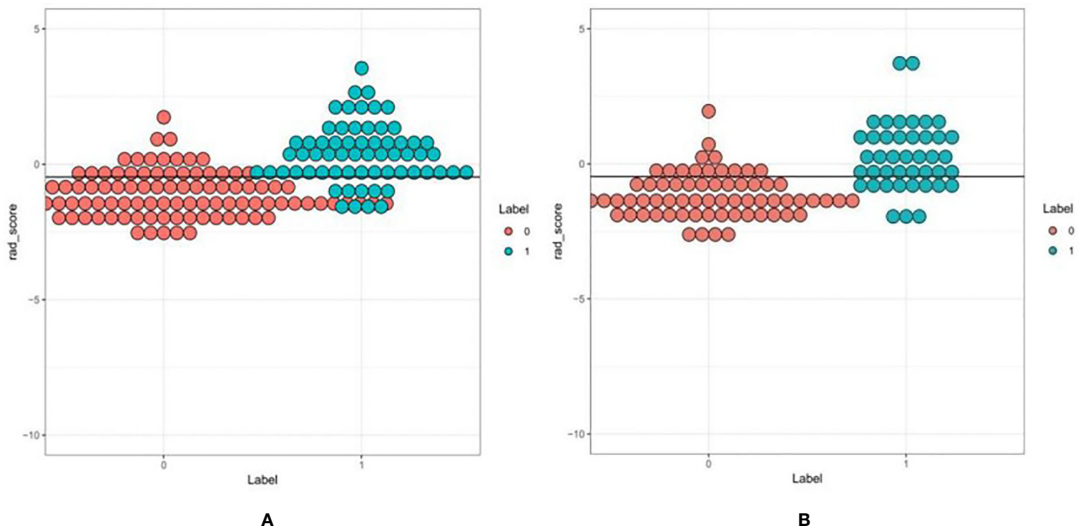


FIGURE 4 | A set of verified rad scores in the training set **(A)** and validation set **(B)**. Red and green represent the true classification: the complete separation of red and green indicates that the radiomics rad-score can be classified well.

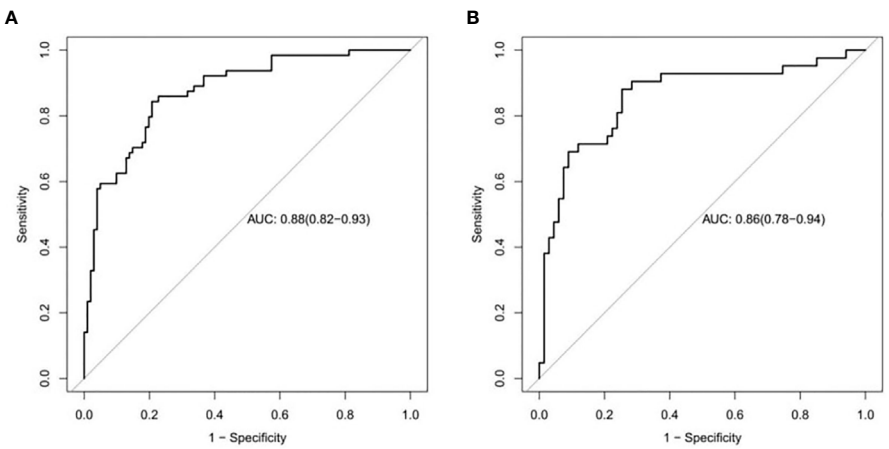


FIGURE 5 | ROC curves to distinguish low Ki-67 lung cancer from high Ki-67 lung cancer based on the CT imaging model prediction model. The AUC in the training set was 0.88. **(A)**, and that in the validation set was 0.86 **(B)**.

TABLE 3 | Positive results of univariate analysis for the classification of low and high Ki-67 lung cancer.

Variable	OR (95% CI)	P-value
Smoking	3.45 (1.78-6.88)	<0.01
Serum	1.00 (1.00-1.00)	0.02
NSE	1.05 (1.02-1.09)	<0.01

TABLE 4 | Positive results of multivariate logistic regression analysis for the classification of low and high Ki-67 lung cancer.

Variable	OR (95% CI)	P-value
(Intercept)	0.48 (0.17-1.30)	0.16
Smoking	2.78 (1.17-6.90)	0.02
NSE	1.02 (0.99-1.06)	0.25
rad_score	5.16 (3.09-9.50)	<0.01

positive cells is related to tumor invasion or tumor progression. Presently, the Ki-67 proliferation index is considered a tumor biomarker that is valuable for tumor diagnosis, treatment and prognosis (4, 24, 25). Previous studies on the CT texture features of non-small cell lung cancer (10, 26, 27) have shown that tumors have internal heterogeneity due to different biological behaviors and metabolic levels, and CT texture analysis can quantify tumor heterogeneity. Lung cancer is a highly heterogeneous tumor, and

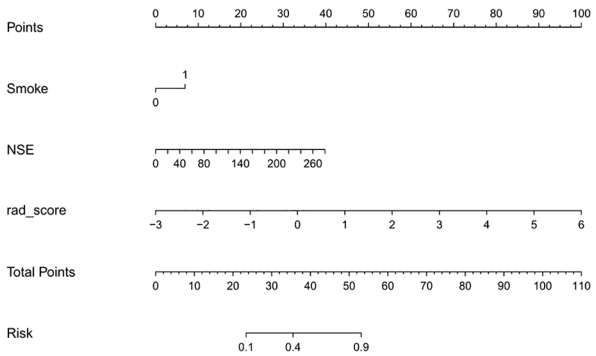


FIGURE 6 | Nomogram used to distinguish between high and low Ki-67 expression levels in lung cancer.

the heterogeneity of Ki-67 expression also exists in lung cancer. The Ki-67 labeling index can range from 1% to 90% in different intratumoral regions (28). High and low Ki-67 expression results in heterogeneity in the tumor cell proliferation rate, cell differentiation and subclonal region composition. Radiomics is a revolution to the traditional visual image features. It obtains high-throughput data and extracts a large number of quantitative features from the image through computer learning software to mine the quantitative information of the shape, texture and heterogeneity of the tumor itself, and screen the most valuable radiomic features to establish a prediction model (18, 29, 30). Radiomics can not only reduce the pain of patients undergoing biopsy, but also improve work efficiency and reduce the cost of patients. Therefore, predicting the expression of Ki-67 by analyzing the CT images of lung cancer is clinically significant.

The present study first used quantitative imaging histology, and then, based on CT images routinely used to diagnose tumors clinically, quantitative image texture analysis was used to estimate Ki-67 expression in lung cancer patients. The radiomics signature was an independent predictor of the expression status of Ki-67 in lung cancer and can distinguish between low Ki-67 lung cancer and high Ki-67 lung cancer well. The AUC of the validation group reached 0.86, and the accuracy, sensitivity, and specificity were 0.80, 0.75, and 0.88, respectively. Radiomics is expected to provide a noninvasive, convenient, and reproducible method to predict the Ki-67 expression status in lung cancer.

In the present study, the training and validation group showed differences in sex, the maximum tumor diameter, smoking status, and pathological type. High Ki-67 expression is more common in men, smokers, and patients with large lesions, SCC and small cell lung cancer. This finding is consistent with previous reports (31, 32). To predict the Ki-67 index, the AUC value of the clinical prediction model established in this study was only 0.72, while that of the radiomics signature was 0.86, much higher than that of the clinical model, indicating that the radiomics signature was significantly better than the simple clinical data prediction model in predicting the Ki-67 index of lung cancer. This study also established a nomogram prediction model combining the radiomics signature and clinically related factors. The data revealed that the AUC value of the validation group was 0.87, which was slightly higher than the predictive power of the radiomics signature alone (AUC=0.86) but significantly higher than the predictive power of the clinical model (AUC=0.72), and the prediction accuracy and specificity of the nomogram were improved. The prediction efficiency of the nomogram was better than that of the clinical model and radiomics signature model. A certain complementarity exists between the radiomics signature model and clinical model, but it is not obvious. The subjects of this study covered all pathological subtypes of lung cancer, so our

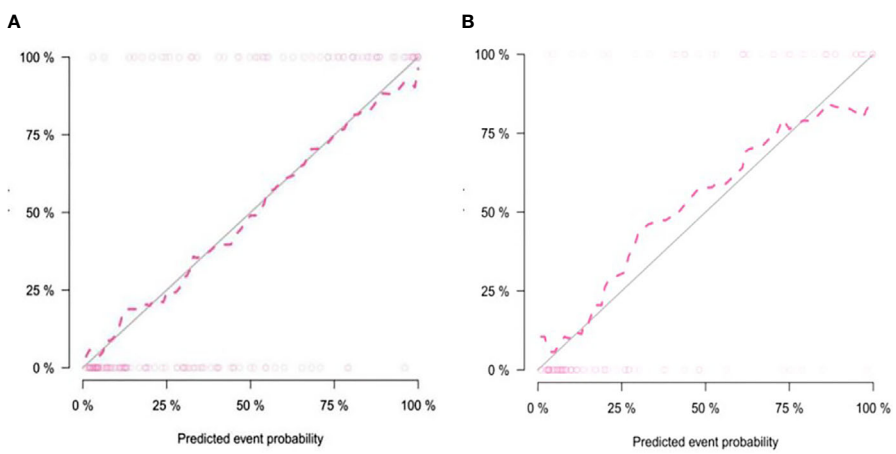


FIGURE 7 | Calibration curve of the nomogram in the training group **(A)** and validation group **(B)**. The solid diagonal line represents the perfect prediction of the ideal model, and the dashed pink line represents the performance of the model. Closer plots of the two lines indicate that the prediction results are in good agreement with the pathological results, and the prediction ability is better.

TABLE 5 | Predictive ability of the radiomics nomogram, radiomics signature, and clinical model for the classification of low and high Ki-67 lung cancer.

Variable		AUC	(95% CI)	Accuracy	Sensitivity	Specificity
Clinical model	Train	0.77	(0.69~0.85)	0.69	0.59	0.82
	Test	0.72	(0.62~0.83)	0.66	0.54	0.80
Radiomics signature	Train	0.88	(0.82~0.93)	0.81	0.79	0.84
	Test	0.86	(0.78~0.94)	0.80	0.75	0.88
Radiomics nomogram	Train	0.91	(0.85~0.98)	0.83	0.79	0.85
	Test	0.87	(0.80~0.95)	0.83	0.75	0.90

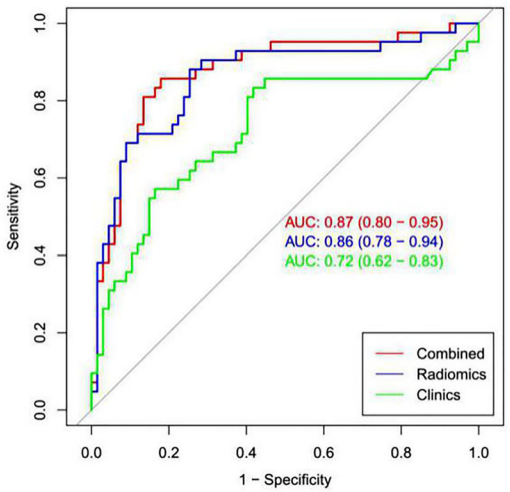


FIGURE 8 | ROC analysis was used to compare the prediction efficiency among the nomogram, radiomics signature and clinical models. The red line shows the nomogram with AUC=0.87, indicating that the radiomics nomogram had better predictive performance than the clinical model (green line; AUC=0.72) or radiomics signature (blue line; AUC= 0.86).

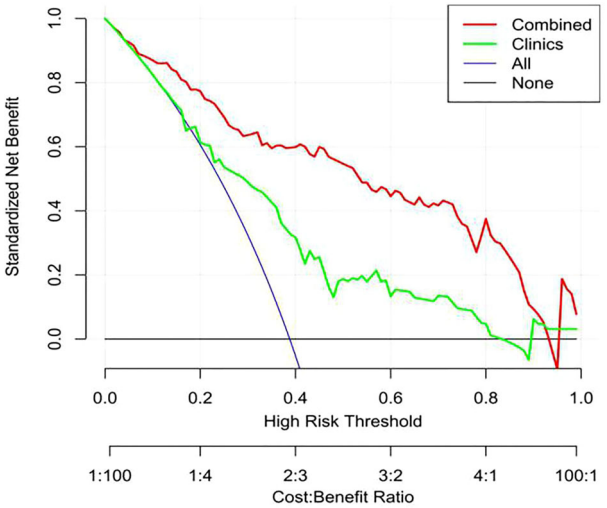


FIGURE 9 | Decision curve analysis (DCA) of the nomogram. The y-axis shows net income, the red line represents the radiomics nomogram, the blue line represents the hypothesis that all patients have high Ki-67 expression, and the black line represents the hypothesis that no patient has high Ki-67 expression. The x-axis shows the threshold probability—that is, the expected benefit of the number of treatments equals the expected benefit of not receiving treatment. The decision curve shows that when the threshold probability is between 0.1 and 1, using a radiomics nomogram to predict Ki-67 expression is more beneficial than treating all patients or not treating patients.

model had better universality. Moreover, the results of this study showed that our model was more effective in predicting the Ki 67 index of lung cancer than the models built in previous studies (17, 18).Our model may become an accurate and noninvasive method to predict the status of Ki-67 in patients with lung cancer.

This study also has some limitations. First, this is a retrospective study with potential selection bias. Second, the sample size of this study was still relatively small, and the predictive ability of radiomics for the Ki-67 index of lung cancer must be further verified in a large sample. Third, although this study included patients with different pathological types of lung cancer, it did not specifically analyze the prediction of the Ki-67 index in a specific pathological type of lung cancer by radiomics. Different cutoff values of the Ki-67 index may need to be established for different pathological types of lung cancer, but the sample size of this study was not sufficiently large to perform this analysis. Therefore, this study is a preliminary exploratory study on the relationship between imaging features and the Ki-67 index of lung cancer. We will expand the sample size and integrate more clinical information to improve the performance and universality of the radiomics model.

In conclusion, we developed and validated the first nomogram model with good diagnostic performance for the classification of low Ki-67 lung cancer and high Ki-67 lung cancer based on the radiomics signature and clinical factors.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Institutional Review Board of

the Affiliated Hospital of Qingdao University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

QF, SHL, and SLL conceived of the project, analyzed the data, and wrote the paper. YBH, ZXZ, WHW and XYT participated in data collection and processing. DPH, XJL and CYZ provided expert guidance, and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Siegel R, Ma J, Zou Z, Jemal A. Cancer Statistics, 2014. *CA Cancer J Clin* (2014) 64:9–29. doi: 10.3322/caac.21208
2. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer Incidence and Mortality Worldwide: Sources, Methods and Major Patterns in GLOBOCAN 2012. *Int J Cancer* (2015) 136(5):E359–86. doi: 10.1002/ijc.29210
3. Tabata K, Tanaka T, Hayashi T, Hori T, Nunomura S, Yonezawa S, et al. Ki-67 Is a Strong Prognostic Marker of Non-Small Cell Lung Cancer When Tissue Heterogeneity Is Considered. *BMC Clin Pathol* (2014) 14(1):1–8. doi: 10.1186/1472-6890-14-23
4. Wen S, Zhou W, Li C, Hu J, Hu X, Chen P, et al. Ki-67 as a Prognostic Marker in Early-Stage Non-Small Cell Lung Cancer in Asian Patients: A Meta-Analysis of Published Studies Involving 32 Studies. *BMC Cancer* (2015) 15 (1):1–13. doi: 10.1186/s12885-015-1524-2
5. Takenaka T, Yano T, Ito K, Morodomi Y, Miura N, Kawano D, et al. Biological Significance of the Maximum Standardized Uptake Values on Positron Emission Tomography in Non-Small Cell Lung Cancer. *J Surg Oncol* (2009) 100(8):688–92. doi: 10.1002/jso.21386
6. Yamashita S, Moroga T, Tokuishi K, Miyawaki M, Chujo M, Yamamoto S, et al. Ki-67 Labeling Index Is Associated With Recurrence After Segmentectomy Under Video-Assisted Thoracoscopic Surgery in Stage I Non-Small Cell Lung Cancer. *Ann Thorac Cardiovasc Surg* (2011) 17 (4):341–6. doi: 10.5761/atcs.oa.10.01573
7. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in Breast Cancer: Prognostic and Predictive Potential. *Lancet Oncol* (2010) 11(2):174–83. doi: 10.1016/S1470-2045(09)70262-1
8. Kim GP, Colangelo LH, Wieand HS, Paik S, Kirsch IR, Wolmark N, et al. Prognostic and Predictive Roles of High-Degree Microsatellite Instability in Colon Cancer: A National Cancer Institute–National Surgical Adjuvant Breast and Bowel Project Collaborative Study. *J Clin Oncol* (2007) 25 (7):767–72. doi: 10.1200/JCO.2006.05.8172
9. Graefe C, Eichhorn L, Wurst P, Kleiner J, Heine A, Panetas I, et al. Optimized Ki-67 Staining in Murine Cells: A Tool to Determine Cell Proliferation. *Mol Biol Rep* (2019) 46(4):4631–43. doi: 10.1007/s11033-019-04851-2
10. Ganeshan B, Abaleke S, Young RC, Chatwin CR, Miles KA. Texture Analysis of Nonsmall Cell Lung Cancer on Unenhanced Computed Tomography: Initial Evidence for a Relationship With Tumour Glucose Metabolism and Stage. *Cancer Imaging* (2010) 10(1):137–43. doi: 10.1102/1470-7330.2010.0021
11. Ganeshan B, Goh V, Mandeville HC, Ng QS, Hoskin PJ, Miles KA. Non–Small Cell Lung Cancer: Histopathologic Correlates for Texture Parameters at CT. *Radiology* (2013) 266:326–36. doi: 10.1148/radiol.12112428
12. Scrivener M, de Jong EEC, van Timmeren JE, Pieters T, Ghaye B, Geets X. Radiomics Applied to Lung Cancer: A Review. *Trans Cancer Res* (2016) 5 (4):398–409. doi: 10.21037/tcr.2016.06.18
13. Wu F, Zhou H, Li F, Wang J, Ai T. Spectral CT Imaging of Lung Cancer: Quantitative Analysis of Spectral Parameters and Their Correlation With Tumor Characteristics. *Acad Radiol* (2018) 25(11):1398–404. doi: 10.1016/j.acra.2018.04.017
14. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5(1):1–9. doi: 10.1038/ncomms5006
15. Grossmann P, Stringfield O, El-Hachem N, Bui MM, Velazquez ER, Parmar C. Defining the Biological Basis of Radiomic Phenotypes in Lung Cancer. *Elife* (2017) 6:e23421. doi: 10.7554/eLife.23421
16. Liu S, Liu S, Zhang C, Yu H, Liu X, Hu Y, et al. Exploratory Study of a CT Radiomics Model for the Classification of Small Cell Lung Cancer and Non-Small-Cell Lung Cancer. *Front Oncol* (2020) 10:1268. doi: 10.3389/fonc.2020.01268
17. Zhou B, Xu J, Tian Y, Yuan S, Li X. Correlation Between Radiomic Features Based on Contrast-Enhanced Computed Tomography Images and Ki-67 Proliferation Index in Lung Cancer: A Preliminary Study. *Thorac Cancer* (2018) 9(10):1235–40. doi: 10.1111/1759-7714.12821
18. Gu Q, Feng Z, Liang Q, Li M, Deng J, Ma M, et al. Machine Learning-Based Radiomics Strategy for Prediction of Cell Proliferation in Non-Small Cell Lung Cancer. *Eur J Radiol* (2019) 118:32–7. doi: 10.1016/j.ejrad.2019.06.025
19. Scholzen T, Gerdes J. The Ki-67 Protein: From the Known and the Unknown. *J Cell Physiol* (2000) 182(3):311–22. doi: 10.1002/(SICI)1097-4652(200003)182:3<311::AID-JCP1>3.0.CO;2-9
20. Alexandrakis MG, Passam FH, Kyriakou DS, Dambaki K, Niniraki M, Stathopoulos E. Ki-67 Proliferation Index: Correlation With Prognostic Parameters and Outcome in Multiple Myeloma. *Am J Clin Oncol* (2004) 27 (1):8–13. doi: 10.1097/01.coc.0000045810.91816.41
21. Higano S, Yun X, Kumabe T, Watanabe M, Mugikura S, Umetsu A, et al. Malignant Astrocytic Tumors: Clinical Importance of Apparent Diffusion Coefficient in Prediction of Grade and Prognosis. *Radiology* (2006) 241 (3):839–46. doi: 10.1148/radiol.2413051276
22. Ahn HK, Jung M, Ha SY, Lee J, Park I, Kim YS, et al. Clinical Significance of Ki-67 and P53 Expression in Curatively Resected Non-Small Cell Lung Cancer. *Tumor Biol* (2014) 35(6):5735–40. doi: 10.1007/s13277-014-1760-0
23. Scholzen T, Gerdes J. The Ki-67 Protein: From the Known and the Unknown. *J Cell Physiol* (2000) 182(3):311–22. doi: 10.1002/(SICI)1097-4652(200003)182:3<311::AID-JCP1>3.0.CO;2-9
24. Antonarakis ES, Keizman D, Zhang Z, Gurel B, Lotan TL, Hicks JL, et al. An Immunohistochemical Signature Comprising PTEN, MYC, and Ki-67 Predicts Progression in Prostate Cancer Patients Receiving Adjuvant Docetaxel After Prostatectomy. *Cancer* (2012) 118(24):6063–71. doi: 10.1002/cncr.27689
25. Jakobsen JN, Sorensen JB. Clinical Impact of Ki-67 Labeling Index in Non-Small Cell Lung Cancer. *Lung Cancer* (2013) 79:1–7. doi: 10.1016/j.lungcan.2012.10.008
26. Yip SS, Aerts HJ. Applications and Limitations of Radiomics. *Phys Med Biol* (2016) 61(13):R150–66. doi: 10.1088/0031-9155/61/13/R150
27. Aerts HJWL, Grossmann P, Tan Y, Oxnard GR, Rizvi N, Schwartz LH, et al. Defining a Radiomic Response Phenotype: A Pilot Study Using Targeted Therapy in NSCLC. *Sci Rep* (2016) 6:33860. doi: 10.1038/srep33860
28. Histed SN, Lindenberg ML, Mena E, Turkbey B, Choyke PL, Kurdziel KA. Review of Functional/Anatomical Imaging in Oncology. *Nucl Med Commun* (2012) 33(4):349–61. doi: 10.1097/MNM.0b013e32834ec8a5
29. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
30. Avanzo M, Stancanello J, El Naqa I. Beyond Imaging: The Promise of Radiomics. *Phys Med* (2017) 38:122–39. doi: 10.1016/j.ejmp.2017.05.071
31. Werynska B, Pula B, Muszczynska-Bernhard B, Piotrowska A, Jethon A, Podhorska-Okolow M, et al. Correlation Between Expression of Metallothionein and Expression of Ki-67 and MCM-2 Proliferation Markers in non-Small Cell Lung Cancer. *Anticancer Res* (2011) 31(9):2833–9. doi: 10.1007/s00520-011-1232-7

32. Okada M, Nishio W, Sakamoto T, Uchino K, Yuki T, Nakagawa A, et al. Effect of Tumor Size on Prognosis in Patients With Non–Small Cell Lung Cancer: The Role of Segmentectomy as a Type of Lesser Resection. *J Thorac Cardiovasc Surg* (2005) 129(1):87–93. doi: 10.1016/j.jtcvs.2004.04.030

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Fu, Liu, Hao, Hu, Liu, Zhang, Wang, Tang, Zhang and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 13 October 2021
doi: 10.3389/fonc.2021.737847

Nomogram Combining Radiomics With the American College of Radiology Thyroid Imaging Reporting and Data System Can Improve Predictive Performance for Malignant Thyroid Nodules

Xingzhi Huang[1†], Zhenghua Wu[1†], Aiyun Zhou[1†], Xiang Min[2], Qi Qi[1], Cheng Zhang[1], Songli Chen[1] and Pan Xu[1]*

[1] Department of Ultrasonography, The First Affiliated Hospital of Nanchang University, Nanchang, China, [2] Department of Head and Neck Otolaryngology, The First Affiliated Hospital of Nanchang University, Nanchang, China

OPEN ACCESS

Edited by:
Giorgio Russo,
Institute of Bioimaging and Molecular Physiology (CNR), Italy

Reviewed by:
Richa Vaish,
Tata Memorial Hospital, India
Chen Liu,
Army Medical University, China
Yanbing Li,
First Affiliated Hospital of Sun Yat-sen University, China

***Correspondence:**
Pan Xu
xupan_1989@126.com

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 07 July 2021
Accepted: 21 September 2021
Published: 13 October 2021

Citation:
Huang X, Wu Z, Zhou A, Min X, Qi Q, Zhang C, Chen S and Xu P (2021) Nomogram Combining Radiomics With the American College of Radiology Thyroid Imaging Reporting and Data System Can Improve Predictive Performance for Malignant Thyroid Nodules. Front. Oncol. 11:737847. doi: 10.3389/fonc.2021.737847

Purpose: To develop and validate a nomogram combining radiomics of B-mode ultrasound (BMUS) images and the American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TI-RADS) for predicting malignant thyroid nodules and improving the performance of the guideline.

Method: A total of 451 thyroid nodules referred for surgery and proven pathologically at an academic referral center from January 2019 to September 2020 were retrospectively collected and randomly assigned to training and validation cohorts (7:3 ratio). A nomogram was developed through combining the BMUS radiomics score (Rad-Score) with ACR TI-RADS score (ACR-Score) in the training cohort; the performance of the nomogram was assessed with respect to discrimination, calibration, and clinical application in the validation and entire cohorts.

Results: The ACR-Rad nomogram showed good calibration and yielded an AUC of 0.877 (95% CI 0.836–0.919) in the training cohort and 0.864 (95% CI 0.799–0.931) in the validation cohort, which were significantly better than the ACR-Score model ($p < 0.001$ and 0.031, respectively). The significantly improved AUC, net reclassification index (NRI), and integrated discriminatory improvement (IDI) of the nomogram were found for both senior and junior radiologists (all $p < 0.001$). Decision curve analysis indicated that the nomogram was clinically useful. When cutoff values for 50% predicted malignancy risk

(ACR-Rad_50%) were applied, the nomogram showed increased specificity, accuracy and positive predictive value (PPV), and decreased unnecessary fine-needle aspiration (FNA) rates in comparison to ACR TI-RADS.

Conclusion: The ACR-Rad nomogram has favorable value in predicting malignant thyroid nodules and improving performance of the ACR TI-RADS for senior and junior radiologists.

Keywords: thyroid nodule, radiomics, nomogram, ultrasound, prediction

INTRODUCTION

With the increasing number of imaging-detected thyroid nodules, overdiagnosis and overtreatment are major clinical challenges in the management of these nodules; therefore, an accurate and practical risk stratification tool is necessary (1). Because B-mode ultrasound (BMUS) is the most accurate imaging modality to assess thyroid nodules, there are a number of risk classification systems based on BMUS images formulated by authoritative associations (2–5). Previous studies have compared different guidelines to find a management guideline that is most beneficial to patients and demonstrated that the 2017 American College of Radiology (ACR) Thyroid Imaging Reporting and Data System (TIRADS) showed accurate diagnostic performance and meaningful reduction in the number of thyroid nodules recommended for biopsy (6–8). However, the relatively low specificity, positive predictive value (PPV), and interobserver variability of the ACR guidelines are impediments to achieving the desired clinical results (7, 9).

Radiomics has the ability to high-throughput mine quantitative image features and discover information reflecting the underlying pathophysiology that cannot be assessed by visual interpretation (10, 11). In recent years, radiomics has been applied to the thyroid, showing that it helps predict malignancy in thyroid nodules and preoperative cervical lymph node staging in papillary thyroid carcinoma (12–14). However, radiomics features are usually analyzed from a single-section image of the target nodule; therefore, radiomics alone might lose some important BMUS information, which makes it impossible to significantly improve the performance of risk stratification systems for all radiologists with different proficiency levels (15).

A nomogram is a graphical tool for a concise and intuitive display of the predicted value of individual outcome events based on multivariate regression analysis. We supposed that a nomogram could adequately combine the visual interpretation and radiomics of BMUS images to achieve better predictive performance. To the best of our knowledge, no published study has investigated the predictive performance of a nomogram combining radiomics with ACR TI-RADS scores for predicting malignant thyroid nodules.

Therefore, the purpose of our study was to develop and validate a nomogram that combines the radiomics score (Rad-Score) and ACR-TIRADS score (ACR-Score) for predicting malignant thyroid nodules and improving performance of the ACR TI-RADS.

MATERIALS AND METHODS

Ethical approval and informed consent were waived because the retrospective study with de-identified data was used, and no protected health information was needed. The study was conducted following guidelines by the Declaration of Helsinki.

Patients

Between January 2019 and September 2020, patients with thyroid nodules (≥10 mm in maximum diameter) in the Head and Neck Otolaryngology Department of our institution were consecutively included. The nodules were enrolled using the following inclusion and exclusion criteria.

The inclusion criteria were as follows: 1) the target nodule had undergone surgical resection; 2) postoperative pathological results were obtained; and 3) BMUS was performed within 2 weeks before the resection. Exclusion criteria are as follows: 1) the pathological result of the nodule was ambiguous, 2) interventional procedures such as fine-needle aspiration (FNA) and radiofrequency ablation were performed before BMUS, and 3) the BMUS image of the target nodule was unclear.

A total of 451 patients (median age, 45 years, range, 20 to 81 years; 93 men and 358 women) were enrolled. If there were multiple nodules in one patient, the nodule with the largest diameter was selected as the target nodule. All nodules were randomly split into a training cohort (n = 315, median age, 45 years, range 20 to 81 years; 68 men and 247 women) and a validation cohort (n = 136, median age, 43 years, range 21 to 70 years; 25 men and 111 women) in a 7:3 ratio.

Clinical and BMUS Information

Clinicopathological data, including age, sex, and nodule pathology, were obtained from medical records. BMUS images were acquired with a Philips iU Elite and Philips EPIQ7 (ultrasound system, Philips Medical System, Bothell, WA, USA) using a 5–12-MHz linear transducer by two radiologists (PX and ZW) with more than 8 years of experience. Images of each target nodule were obtained in transverse and longitudinal planes, and video clips were obtained in at least one plane.

Analysis of the ACR TI-RADS

Two radiologists (AZ and XH, with more than 10 years and 3 years of experience, respectively) who were unaware of the pathological results reviewed the BMUS images of all nodules. The five feature categories in the ACR TI-RADS lexicon

(composition, echogenicity, shape, margin, and echogenic foci) were evaluated, and the ACR-Score of each nodule was calculated (referred to as ACR-Score 1 for AZ, ACR-Score 2 for XH) (16). The Supplement presents the detailed process of calculating the ACR-Score (**Supplementary Tables E1**, **2**). The diameter and location (subcapsular or intrathyroidal) were negotiated to a consensus by the two radiologists.

Analysis of the Radiomics Features

The region of interest (ROI) was delineated manually on the BMUS DICOM image of the target nodule with the largest diameter in sagittal view using open-source software (3D Slicer, version 4.10.2; https://www.slicer.org) (**Supplementary Figure E1**) (17, 18). The reproducibility of the intra- and interobserver agreement for the radiomics features was measured using the first 130 nodules that a radiologist (XH) redelineated twice within 2 weeks. The intraclass correlation coefficient (ICC) was used to evaluate the intra- and interobserver agreement. ICC > 0.75 represented satisfactory agreement. XH delineated the remaining nodules if strong agreement (ICC > 0.90) was achieved. To ensure repeatability of the results, resampling and z-score normalization were performed as preprocessing steps (**Supplementary E1**). Open-source software (Pyradiomics; http://pyradiomics.readthedocs.io/en/latest/index.html) (19) was used to extract a total of 837 texture, intensity, and wavelet features (**Supplementary Table E3**). Then, dimensionality reduction and radiomics feature selection were performed successively by ICC, univariate, least absolute shrinkage and selection operator (LASSO) and linear dependence analyses (20, 21). The methodology used to extract the radiomics features is further described in **Supplementary E2** and **Figure 1A**. The radiomics score (Rad-Score) was generated using a linear combination of the selected features.

Statistical Analysis

Statistical analysis was conducted with IBM SPSS 22.0 software (IBM, New York, USA) and R software (Version 4.0.1, https://www.r-project.org/). The packages of R4.0.1 used are provided in **Supplementary Table E4**. The Shapiro–Wilk test was used to evaluate the normality of the distribution. Continuous data conforming to a normal distribution are expressed as the mean ± standard deviation (SD) and were compared using Student's t-test; nonconforming data are expressed as the median [interquartile range (IQR)] and were compared using the Mann–Whitney U test.

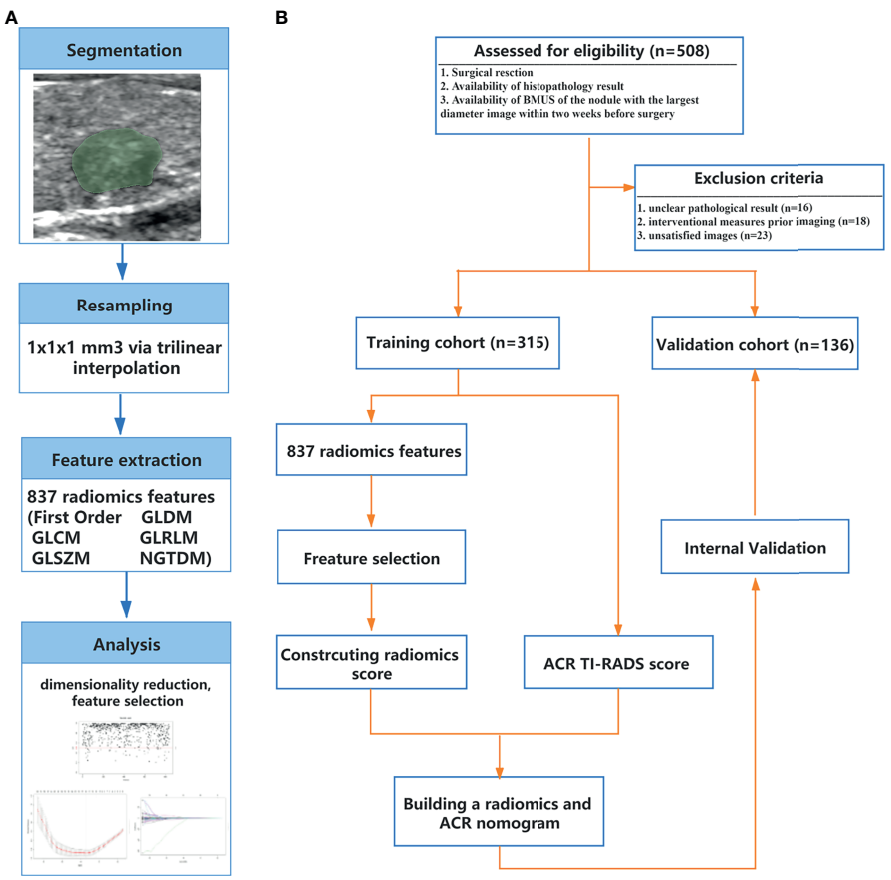


FIGURE 1 | B-model ultrasound radiomics workflow **(A)** and study flowchart **(B)**. ACR, American College of Radiology; BMUS, B-model ultrasound; GLCM, gray-level co-occurrence matrix; GLDM, gray-level dependence matrix; GLRLM, gray-level run length matrix; GLSZM, gray-level size zone matrix; NGTDM, neighboring gray tone difference matrix; TI-RADS, thyroid imaging reporting and data system.

Categorical data are expressed as numbers (%) and were compared using a chi-square test or Fisher's exact test, as appropriate. A $p < 0.05$ represented a statistically significant difference.

Development of the ACR-Rad Nomogram

The ACR-Rad nomogram was developed based on the Rad-Score and the average of ACR-Score 1 and ACR-Score 2. For comparison, the ACR-Score model was built through a univariate logistic equation.

Performance of the ACR-Rad Nomogram

Calibration was evaluated using the Akaike information criterion (AIC), the Bayesian information criterion (BIC), calibration curve, and Hosmer–Lemeshow test (22). Discrimination performance was evaluated using the area under the receiver operator characteristic (ROC) curve (AUC). The Delong test was used to compare AUCs and the LR test to compare the effect across the nested logistic regression models.

Clinical Utility of the ACR-Rad Nomogram

The interobserver agreement for ACR-Score, Rad-Score, and the predicted malignancy risk by the ACR-Rad nomogram was evaluated. The improvement in the predictive accuracy of the nomogram by radiologists with different levels of experience was evaluated by the AUC, index integrated discrimination improvement (IDI), and net reclassification improvement (NRI). Decision curve analysis (DCA) was conducted to determine the clinical usefulness of the nomogram by quantifying the net benefits at different threshold probabilities in the entire cohort.

For clinical management, we compared the performance for biopsy recommended of the cutoff value which was determined in the training cohort with the maximum Youden index (referred to as ACR-Rad_max), and different cutoff values which were determined in the training cohort for prespecified predicted risks of malignancy (20%/30%/40%/50%) (referred to as ACR-Rad_20%/30%/40%/50%, respectively) with ACR TI-RADS in the entire cohort.

RESULTS

Patient Characteristics

The study flowchart is shown in **Figure 1B**. Clinical and pathological characteristics in the training and validation cohorts are summarized in **Tables 1**, **2**. There was no significant difference between the training and validation cohorts for clinicopathological and BMUS characteristics (all $p > 0.05$). The proportions of malignant nodules in the two groups were 62.9% (198/315) and 70.6% (96/136) ($p = 0.114$). Malignant nodules had significantly lower age, diameter, and nodular goiter and significantly higher ACR-Score 1 and ACR-Score 2 than benign nodules in the training and validation cohorts (all $p < 0.05$).

Selecting Radiomics Features and Building the Rad-Score

The rates of intra- and interobserver agreement for the radiomics features reached 94.7% (794/837; mean ICC = 0.920) and 94.0% (787/837; mean ICC = 0.901), respectively (**Supplementary Figure E2**). Seventy-two features were excluded due to unsatisfactory agreement (ICC < 0.75); 192 features were excluded due to insignificant differences based on univariate analysis. Among 14 features selected by LASSO, 4 features were considered to have strong collinearity for the variance inflation factor [VIF] which was more than 10 (**Supplementary Figure E3**). The remaining 10 features were included in the Rad-Score formula (**Supplementary Figure E4**). The Rad-Score of malignant nodules was significantly higher than that of benign nodules in the training [1.265 (0.738–1.900) *vs.* -0.005 (-1.955–0.910), $p < 0.001$] and validation cohorts [1.177 (0.355–1.845) *vs.* -0.320 (-2.182–0.685), $p < 0.001$] (**Table 2**). The Rad-Score yielded an AUC of 0.801 (95% CI 0.750–0.851) in the training cohort and 0.820 (95% CI 0.742–0.898) in the validation cohort (**Figure 2**).

Development and Performance of the ACR-Rad Nomogram

The ACR-Rad nomogram incorporated two predictors: the average ACR-Score [odds ratio (OR) 1.644, 95% CI 1.423–1.928] and Rad-Score (OR 2.269, 95% CI 1.709–3.133) (both $p < 0.001$) (**Figure 3A**). The ACR-Score model was built using the following a univariate logistic regression equation:

$$\text{ACR} - \text{Score model}: \text{logit}\,\pi_{\text{ACR-Score}}$$

$$= -3.791 + 0.603^{*}\text{ACR} - \text{Score}$$

The LR test between the ACR-Rad nomogram and ACR-Score model was $\chi^2 = 4.184$ ($p < 0.001$). The AIC, BIC, calibration curve, and Hosmer–Lemeshow test statistic ($p = 0.640$) showed good calibration of the ACR-Rad nomogram in the training cohort (**Table 3** and **Figure 3B**). An AUC of 0.877 (95% CI 0.836–0.919) also showed good discrimination, which was significantly higher than that of the ACR-Score model (0.833, 95% CI 0.785–0.880) in the training cohort ($p < 0.001$). The favorable calibration of the nomogram was confirmed in the validation cohort, whose Hosmer–Lemeshow test yielded a p value of 0.736 (**Figure 3C**). The AUC (0.864, 95% CI 0.799–0.931) was significantly higher than that of the ACR-Score model (0.802, 95% CI 0.719–0.886) in the validation cohort (**Figures 4A, B**).

Clinical Utility of the ACR-Rad Nomogram

The ICC of ACR-Score (0.677) was considerably lower than that of the Rad-Score and predicted malignancy risk (0.901 and 0.844, respectively). For senior and junior radiologists, the utilization of the ACR-Rad nomogram significantly improved the predictive value for predicting malignant thyroid nodules in terms of AUC, NRI, and IDI compared to the ACR-Score model in entire cohort (all $p < 0.001$) (**Table 4** and **Figure 4C**). Moreover, favorable calibration of the nomogram was confirmed in both radiologists (Hosmer–Lemeshow test $p = 0.715$ and 0.415, respectively) (**Supplementary Figure E5**). The DCA demonstrated that the nomogram had a higher overall net benefit than the ACR-Score model, and was more beneficial than either the treat-all or the treat-none strategy (**Figure 5**).

TABLE 1 | Clinicopathological and ultrasonic characteristics of patients in the training and validation cohorts.

	Entire population (n = 451)	Training cohort (n = 315)	Validation cohort (n = 136)
Sex			
Male	93 (20.6)	68 (21.6)	25 (18.4)
Female	358 (79.4)	247 (78.4)	111 (81.6)
Age, years	45 (34-53)	45 (32-53)	45 (37-53)
Diameter, mm	15.0 (13.0-19.5)	15.0 (13.0-19.5)	15.9 (13.0-19.5)
Nodule pathology			
Benign	157 (34.8)	117 (37.1)	40 (29.4)
Malignant	294 (65.2)	198 (62.9)	96 (70.6)
Tumor location			
Subcapsular thyroid	319 (70.7)	221 (70.2)	98 (72.1)
Intra-thyroidal	132 (29.3)	94 (29.8)	38 (27.9)
Hashimoto thyroiditis			
Positive	162 (35.9)	113 (35.9)	49 (36.0)
Negative	289 (64.1)	202 (64.1)	87 (64.0)
Nodular goiter			
Positive	154 (34.1)	112 (35.6)	42 (30.9)
Negative	297 (65.9)	203 (64.4)	94 (69.1)
ACR-Score 1[b,c]	7 (5-9)	8 (5-9)	7 (5-9)
Composition			
Cystic or spongiform	21 (4.7)	14 (4.4)	7 (5.1)
Cystic and solid	54 (12.0)	41 (13.0)	13 (9.6)
Solid	376 (83.4)	260 (82.5)	116 (85.3)
Echogenicity			
Anechoic	21 (4.7)	14 (4.4)	7 (5.1)
Hyper- or isoechoic	82 (18.2)	65 (20.6)	18 (13.2)
Hypoechoic	277 (61.4)	189 (60.0)	87 (64.0)
Very hypoechoic	71 (15.7)	47 (14.9)	24 (17.6)
Shape			
Taller-than-wide	132 (29.3)	92 (29.2)	40 (29.4)
Not taller-than-wide	319 (70.7)	223 (70.8)	96 (70.6)
Margin			
Smooth or ill defined	163 (36.1)	118 (37.5)	45 (33.1)
Irregular or lobulated	281 (62.3)	193 (61.3)	88 (64.7)
Extrathyroidal extension	7 (1.6)	4 (1.3)	3 (2.2)
Echogenic foci[a]			
No echogenic foci or large comet tail	233 (51.7)	162 (51.6)	71 (51.4)
Macrocalcifications	46 (10.2)	29 (9.2)	17 (12.3)
Peripheral	8 (1.8)	5 (1.6)	3 (2.2)
Punctate	165 (36.6)	118 (37.6)	47 (34.1)
ACR TI-RADS risk level			
TR1	21 (4.7)	14 (4.4)	7 (5.1)
TR2	19 (4.2)	15 (4.8)	5 (3.7)
TR3	31 (6.9)	26 (8.3)	5 (3.7)
TR4	103 (22.8)	69 (21.9)	34 (25.0)
TR5	277 (61.4)	191 (60.6)	85 (62.5)
ACR-Score 2[c]	8 (6-9)	8 (6-9)	8 (6-9)
Rad-Score	0.910 (-0.100-1.550)	0.932 (-0.038-1.694)	0.899 (-0.114-1.473)

Qualitative data were expressed as mean ± standard deviation or number and percentages (%); quantitative data were expressed as median (25%–75% quantiles).
ACR, American College of Radiology; TI-RADS, Thyroid Imaging Reporting and Data System.
[a]*Nodules could have more than one type of echogenic foci.*
[b]*B-model ultrasound findings based on the senior interpretation.*
[c]*ACR-Score 1 was referred for the senior radiologist, ACR-Score 2 for the junior radiologist.*

When applied with the ACR-Rad_max, specificity, accuracy, and PPV significantly increased with unnecessary FNA rates significantly decreasing but at the expense of significantly decreased sensitivity in comparison to ACR TI-RADS for both senior and junior radiologists. With ACR-Rad_20%/30%/40%, the specificity improved insignificantly for the senior radiologist. With ACR-Rad_50%, the significantly increased specificity, accuracy, and PPV and decreased unnecessary FNA rate were observed for the junior radiologist, and the significantly increased specificity was presented for the senior radiologist as well (all $p < 0.05$), with no difference in sensitivity and negative predictive value (NPV) ($p > 0.05$) (**Table 5**).

DISCUSSION

In this study, we proved that BMUS radiomics and the ACR-Rad nomogram based on it and ACR TI-RADS can accurately predict

TABLE 2 | Clinicopathological and ultrasonic characteristics for thyroid nodules in the training and validation cohorts by pathology.

Characteristic	Training cohort			Validation cohort		
	Benign (n = 117)	Malignant (n = 198)	*p* value	Benign (n = 40)	Malignant (n = 96)	*p* value
Sex			0.942			0.511
Male	25 (21.4)	43 (21.7)		6 (24.0)	19 (30.6)	
Female	92 (78.6)	155 (78.3)		34 (76.0)	77 (69.4)	
Age, years	47 (38-55)	42.5 (30-50)	0.001	51.5 (43.5-60)	42.5 (33.5-49)	<0.001
Diameter, mm	17.0 (14.0-32.0)	15.0 (12.0-17.0)	<0.001	19.5 (14.0-34.0)	15.0 (13.0-17.0)	<0.001
Tumor location			0.816			0.236
Subcapsular thyroid	83 (70.9)	138 (69.7)		26 (65.0)	72 (75.0)	
Intra-thyroidal	34 (29.1)	60 (30.3)		14 (35.0)	24 (25.0)	
Hashimoto thyroiditis			0.470			0.344
Positive	39 (33.3)	74 (37.4)		12 (30.0)	37 (38.5)	
Negative	78 (66.7)	124 (62.6)		28 (70.0)	59 (61.5)	
Nodular goiter			0.003			0.021
Positive	54 (46.2)	58 (29.3)		18 (45.0)	24 (25.0)	
Negative	63 (53.8)	140 (70.7)		22 (55.0)	72 (75.0)	
ACR-Score 1[b, c]	4 (3-7)	9 (7-9)	<0.001	5 (2-6)	8 (6-10)	<0.001
Composition			<0.001			<0.001
Cystic or spongiform	14 (12.0)	0 (0)		7 (17.5)	0 (0)	
Cystic and solid	32 (27.4)	9 (4.5)		10 (25.0)	3 (3.1)	
Solid	71 (60.7)	189 (95.5)		23 (57.5)	93 (96.9)	
Echogenicity			<0.001			<0.001
Anechoic	14 (12.0)	0 (0)		7 (17.5)	0 (0)	
Hyper- or Isoechoic	51 (43.6)	14 (7.1)		14 (35.0)	4 (4.2)	
Hypoechoic	35 (29.9)	154 (77.8)		17 (42.5)	70 (72.9)	
Very hypoechoic	17 (14.5)	30 (15.2)		2 (5.0)	22 (22.9)	
Shape			<0.001			0.017
Taller-than-wide	18 (15.4)	74 (37.4)		6 (15.0)	34 (35.4)	
Not taller-than-wide	99 (84.6)	124 (62.6)		34 (85.0)	62 (64.6)	
Margin			<0.001			<0.001
Smooth or ill defined	87 (74.4)	42 (20.1)		29 (72.5)	16 (16.7)	
Irregular or lobulated	29 (24.8)	164 (78.5)		11 (27.5)	77 (80.2)	
Extrathyroidal extension	1 (0.9)	3 (1.4)		0 (0)	3 (3.1)	
Echogenic foci[a]			<0.001			0.028
No echogenic foci or Large comet tail	75 (64.1)	87 (43.5)		25 (59.5)	46 (47.4)	
Macrocalcifications	18 (15.4)	11 (5.5)		8 (19.0)	9 (9.3)	
Peripheral	2 (1.7)	3 (1.5)		2 (4.8)	2 (2.1)	
Punctate	22 (18.8)	99 (49.5)		7 (16.7)	40 (41.2)	
ACR TI-RADS risk level			<0.001			<0.001
TR1	14 (12.0)	0 (0)		7 (17.5)	0 (0)	
TR2	14 (12.0)	1 (0.5)		4 (10.0)	1 (1.0)	
TR3	24 (20.5)	2 (1.0)		5 (12.5)	0 (0)	
TR4	32 (27.4)	37 (18.7)		13 (32.5)	21 (21.9)	
TR5	33 (28.2)	158 (79.8)		11 (27.5)	74 (77.1)	
ACR-Score 2[c]	5 (4-8)	9 (7-10)	<0.001	6 (2-9)	8 (7-10)	<0.001
Rad-score	-0.005 (-1.955-0.910)	1.265 (0.738-1.900)	<0.001	-0.320 (-2.182-0.685)	1.177 (0.355-1.845)	<0.001

Qualitative data were expressed as mean ± standard deviation or number and percentages (%); quantitative data were expressed as median (25%–75% quantiles).
ACR, American College of Radiology; TI-RADS, Thyroid Imaging Reporting and Data System.
[a] *Nodules could have more than one type of echogenic foci.*
[b] *B-model ultrasound findings based on the senior interpretation.*
[c] *ACR-Score 1 was referred for the senior radiologist, ACR-Score 2 for the junior radiologist.*

malignancy in thyroid nodules, and the nomogram showed significantly better discrimination and calibration performance than the guideline alone. Excellent repeatability and clinical application of the nomogram were demonstrated in the entire cohort. With performing with 50% risk cutoff, the nomogram increased the specificity, accuracy, and PPV and decreased the unnecessary FNA rates of ACR TI-RADS for radiologists of different proficiency levels.

Predicting malignant thyroid nodules and reduction in the number of meaningless biopsies are original intentions of many guidelines that the ACR TI-RADS can meet. The ACR guideline assigns points to the five feature categories of BMUS. The sum of the points is used to determine the probability of malignancy and provides recommended management procedures (16). However, the clinical application of the ACR guideline is strongly subjective (6, 13, 14). Hoang et al. (6) found that when the judgment of composition was wrong, malignant nodules would be misclassified. Although the ACR risk stratification system is fault-tolerant, in our study, the interobserver agreement was unsatisfactory (ICC = 0.677) and lower than that of the Rad-

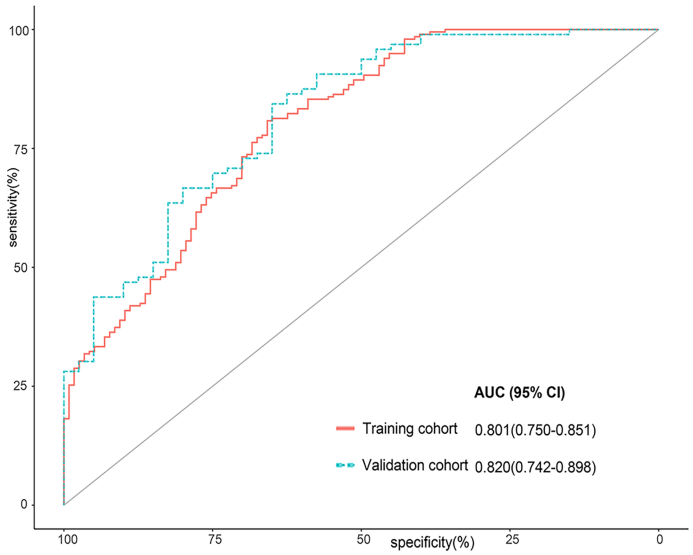


FIGURE 2 | Receiver operator characteristic (ROC) curves for the Rad-Score in the training cohort and the validation cohort.

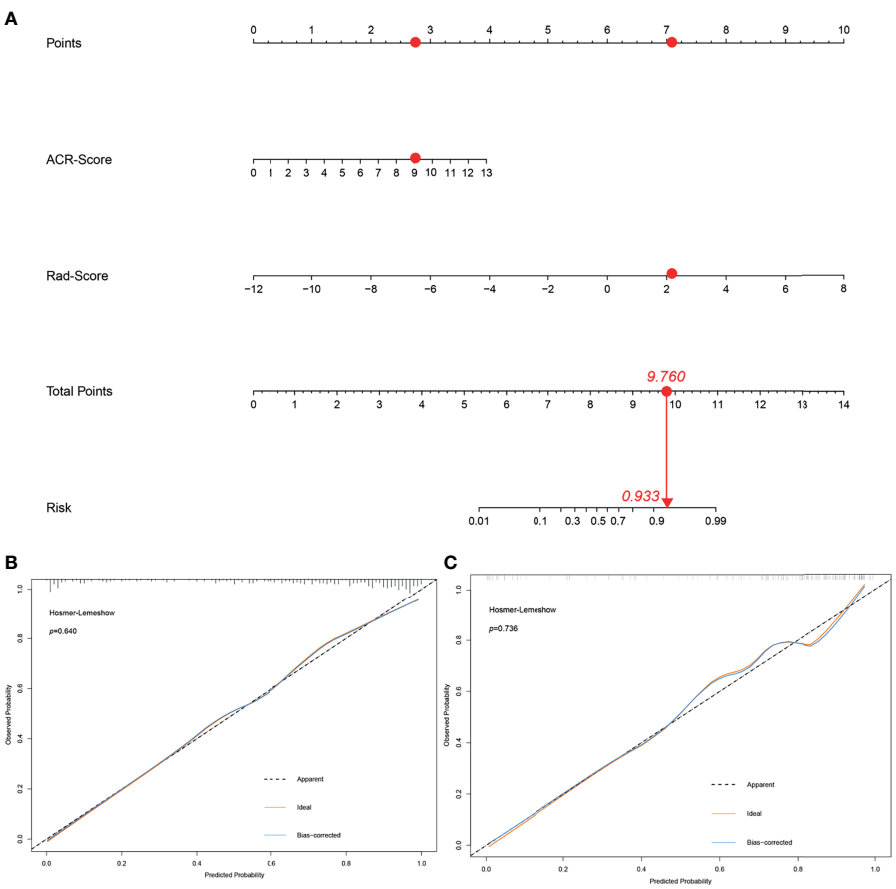


FIGURE 3 | The ACR-Rad nomogram to predict malignancy in thyroid nodules **(A)** and calibration curves of the nomogram in the training **(B)** and validation **(C)** cohort. Red dots described the calculation process of an ACR-Rad nomogram point of a malignant thyroid nodule.

TABLE 3 | Performance of the ACR-Rad nomogram for predicting malignant thyroid nodules in the training and validation cohorts.

		Multivariate analysis		Discrimination		Calibration	Goodness of fit	
		Odds ratio (95% CI)	*p* value	AUC (95% CI)[a]	*p* value	Hosmer–Lemeshow *p* value[a]	AIC	BIC
ACR-Rad nomogram	ACR-Score[b]	1.644 (1.423-1.928)	<0.001	T: 0.877 (0.836-0.919)/V: 0.864 (0.799-0.931)		T: 0.640/V: 0.736	257.52	268.78
	Rad-Score	2.269 (1.709-3.133)	<0.001					
ACR-Score model		1.827 (1.603-2.114)	<0.001	T: 0.833 (0.785-0.880)/V: 0.802 (0.719-0.886)	0.001/0.031	T: 0.415/V: 0.824	299.43	306.93

ACR, American College of Radiology; AIC, Akaike information criterion; AUC, area under the receiver operating characteristic curve; BIC, Bayesian information criterion; LR, likelihood ratio.
[a]*T, training cohort; V, validation cohort.*
[b]*The average of ACR-Score 1 and ACR-Score 2 was applied.*

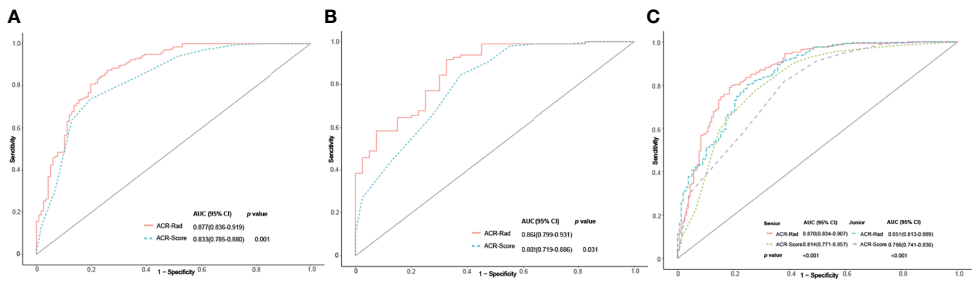


FIGURE 4 | Receiver operating characteristic (ROC) curves of the ACR-Rad nomogram and ACR-Score model in the training **(A)**, validation **(B)**, and entire **(C)** cohorts.

TABLE 4 | Performance of the ACR-Rad nomogram for predicting malignant thyroid nodules with interpretations from the senior and junior radiologists.

		AUC (95% CI)	Categorical NRI (95% CI)	Continuous NRI (95% CI)	IDI (95% CI)
ACR-Rad nomogram *vs.* ACR-Score model	For the senior radiologist	0.870 (0.834-0.907) *vs.* 0.814 (0.771-0.857)	0.181 (0.089-0.273)	0.688 (0.505-0.871)	0.121 (0.086-0.155)
	p value	<0.001	<0.001	<0.001	<0.001
	For the junior radiologist	0.851 (0.813-0.889) *vs.* 0.786 (0.741-0.830)	0.252 (0.157-0.348)	0.721 (0.539-0.903)	0.138 (0.100-0.175)
	p value	<0.001	<0.001	<0.001	<0.001

ACR, American College of Radiology; AUC, area under the receiver operating characteristic curve; IDI, index integrated discrimination improvement; NRI, net reclassification improvement.

Score and risk prediction value of the nomogram (ICC = 0.901 and 0.844, respectively). The reason may be due to weaker judgment of the junior radiologist in scoring spongiform, very hypoechoic and ill defined. Furthermore, the specificity of the ACR TI-RADS is weak (38.85%–57.96% in our study).

Previous studies have reported that combining clinical characteristics (such as age, thyrotropin, or sex) with ultrasound features (such as ACR TI-RADS lexicon, hypoechoic halo, or blood flow) slightly increased the accuracy of these models in discriminating malignant nodules from benign nodules than risk stratification systems (23, 24). However, the abovementioned clinical characteristics in the study of Liang et al. (13) were not significantly different. In our study, there was no significant difference in the gender. Moreover, other subjective ultrasound features might make little contribution to solve current challenges.

With the recent development of radiomics, its application in predicting the malignancy of thyroid nodules has received attention. Previous studies reported that radiomics showed good performance in predicting thyroid cancer, which was even higher than the risk classification guidelines with interpretations from non-experts (13, 25, 26). In our study, both the ACR-Score and Rad-Score were independent predictive factors of malignant nodules, and the Rad-Score had favorable diagnostic performance in magnificent nodules. However, radiomics alone cannot improve the performance of the ACR TI-RADS for senior radiologists who are experienced to evaluate comprehensively ultrasound features correlated with properties of the nodules (15).

Park et al. (14) demonstrated that when combined with a 5% predicted malignancy risk cutoff of radiomics with the ACR or American Thyroid Association (ATA) guidelines, the performance significantly increased and unnecessary FNA rates reduced; in consequence, combining radiomics with ultrasound-based risk stratification systems is a potential approach to predict magnificent thyroid nodules. Luo et al. (27) constructed a

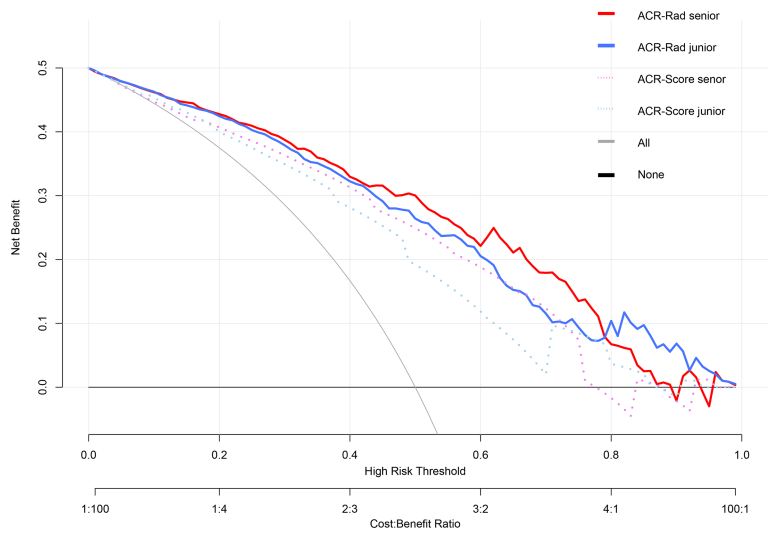


FIGURE 5 | Decision curve analysis (DCA) of the ACR-Rad nomogram in predicting malignancy in thyroid nodules for the senior and junior radiologists. The vertical axis measures standardized net benefit. The horizontal axis shows the corresponding risk threshold. The DCA showed that the ACR-Rad nomogram had a higher overall net benefit than the ACR-Score model for both senior and junior radiologist.

nomogram including the Rad-Score and feature categories of the ACR TI-RADS and determined that a combination model was better than radiomics and the ACR TI-RADS alone for discriminating benign and malignant thyroid nodules. In our study, the ACR-Rad nomogram could be a more convenient tool to combine the ACR-Score and Rad-Score and a better predictive model for thyroid cancer. For senior and junior radiologists, the nomogram had significantly improved predictive performance in comparison with the ACR TI-RADS.

TABLE 5 | Diagnostic performance and unnecessary FNA rates of ACR TI-RADS and risk cutoff values for the ACR-Rad nomogram.

	ACR Senior	ACR-Rad_20%	ACR-Rad_30%	ACR-Rad_40%	ACR-Rad_50%	ACR-Rad_max
Sensitivity	88.44% (84.22%–91.86%)	96.94% (94.27%–98.59%)[a]	94.90% (91.72%–97.12%)[a]	89.12% (84.98%–92.44%)	85.03% (80.43%–88.91%)	79.93% (74.89%–84.36%)[a]
Specificity	57.96% (49.83%–65.78%)	51.59% (43.49%–59.63%)	59.24% (51.12%–67.00%)	66.24% (58.27%–73.59%)	71.97% (64.26%–78.84%)[a]	80.25% (73.16%–86.17%)[a]
Accuracy	77.83% (73.71%–81.58%)	81.15% (77.23%–84.66%)	82.48% (78.65%–85.88%)	81.15% (77.23%–84.66%)	80.49% (76.52%–84.05%)	80.04% (76.05%–83.64%)
PPV	79.75% (76.54%–82.63%)	78.95% (76.11%–81.53%)	81.34% (78.28%–84.06%)	83.17% (79.82%–86.07%)	85.03% (81.49%–88.00%)	88.35% (84.62%–91.26%)[a]
NPV	72.80% (65.51%–79.04%)	90.00% (82.29%–94.57%)[a]	86.11% (78.83%–91.17%)[a]	76.47% (69.70%–82.12%)	71.97% (65.79%–77.43%)	68.11% (62.66%–73.10%)
Unnecessary FNA rate	20.25% (16.02%–25.02%)	21.05% (16.96%–25.63%)	18.66% (14.68%–23.19%)	16.83% (12.86%–21.42%)	14.97% (11.09%–19.57%)	11.65% (8.06%–16.13%)[a]
	Junior					
Sensitivity	90.88% (87.01%–93.90%)	97.62% (95.16%–99.04%)[a]	95.92% (92.98%–97.87%)[a]	92.86% (89.29%–95.52%)	87.07% (82.69%–90.69%)	80.27% (75.26%–84.67%)[a]
Specificity	38.85% (31.19%–46.95%)	46.50% (38.51%–54.62%)	52.87% (44.75%–60.87%)[a]	57.96% (49.83%–65.78%)[a]	64.33% (56.30%–71.81%)[a]	74.52% (66.96%–81.13%)[a]
Accuracy	72.85% (68.50%–76.89%)	79.82% (75.82%–83.43%)[a]	79.21% (76.32%–81.83%)[a]	80.71% (76.76%–84.25%)[a]	79.16% (75.11%–82.82%)[a]	78.27% (74.17%–81.99%)
PPV	73.70% (71.11%–76.14%)	77.36% (74.68%–79.83%)	87.37% (79.58%–92.47%)	80.53% (77.44%–83.29%)[a]	82.05% (78.67%–85.00%)[a]	85.51% (81.78%–88.58%)[a]
NPV	69.32% (60.00%–77.29%)	91.25% (83.11%–95.67%)[a]	80.93% (77.00%–84.45%)[a]	81.25% (73.75%–86.98%)	72.66% (65.90%–78.52%)	66.86% (61.15%–72.11%)
Unnecessary FNA rate	26.45% (21.98%–31.30%)	22.64% (18.48%–27.24%)	20.79% (16.69%–25.38%)	19.47% (15.39%–24.09%)[a]	17.95% (13.85%–22.67%)[a]	14.49% (10.56%–19.21%)[a]

[a]*A statistically significant difference.*

FNA, fine-needle aspiration; NPV, negative predictive value; PPV, positive predictive value.

The radiomics model has not been sufficiently evaluated by prior studies, which were limited to comparisons of discrimination performance or only had senior radiologists assigned to score and delineate nodules (28, 29). Our study evaluated the repeatability, discrimination, and clinical utilization of the ACR-Rad nomogram applied by senior and junior radiologists, proving that there was strong consistency in processing nodule texture information and it significantly increased the predictive performance among radiologists of different proficiency levels, which can compensate for the relatively low repeatability and accuracy of ACR TI-RADS. Moreover, the appropriate cutoff of the ACR-Rad nomogram can significantly reduce unnecessary FNA rates, increase the specificity and PPV, and maintain the high sensitivity of the guideline especially for junior radiologists.

Our study had several limitations. First, this study was a single-center retrospective study; thus, selection bias may be inevitable. The proportion of benign nodules in this study was much lower than that in other studies (13–15), because we chose nodules with postoperative pathology instead of FNA or follow-up. Second, BMUS images were only acquired with Philips ultrasound instruments. We should investigate the influence from images of different ultrasound instruments. Third, on account of the overlap between the shape features and the ACR TI-RADS, we did not analyze them. The predictive performance of shape features should be explored further.

In conclusion, the ACR-Rad nomogram, combined with ACR TI-RADS and BMUS radiomics, has the potential to be a convenient and accurate tool to predict malignancy and improve performance for radiologists at different proficiency levels in thyroid nodules.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The requirement for informed consent was waived. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

XH performed the formal analysis, contributed to the methodology, administered the project, and wrote the original draft of the manuscript. ZW curated the data and collected the resources. AZ conceived the study and wrote and reviewed the manuscript. XM curated the data. QQ curated the data. CZ curated the data. SC curated the data. PX supervised the study and wrote, reviewed, and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by "the key research and development projects of Jiangxi Province" (no. 20181BBG70031).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.737847/full#supplementary-material

REFERENCES

1. Singh Ospina N, Iñiguez-Ariza NM, Castro MR. Thyroid Nodules: Diagnostic Evaluation Based on Thyroid Cancer Risk Assessment. *BMJ (Clinical Res ed)* (2020) 368:l6670. doi: 10.1136/bmj.l6670
2. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, et al. American Association of Clinical Endrocrinologist, American College of Endocrinoly, and Associazione Medici Endocrinologi Medical Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules 2016 Update. *Endocr Pract: Off J Am Coll Endocrinol Am Assoc Clin Endocrinologists* (2016) 22(5):622–39. doi: 10.4158/ep161208.Gl
3. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients With Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid: Off J Am Thyroid Assoc* (2016) 26 (1):1–133. doi: 10.1089/thy.2015.0020
4. Hoang JK, Lee WK, Lee M, Johnson D, Farrell S. US Features of Thyroid Malignancy: Pearls and Pitfalls. *Radiographics: Rev Publ Radiol Soc North Am Inc* (2007) 27(3):847–60. doi: 10.1148/rg.273065038
5. Russ G, Bonnema SJ, Erdogan MF, Durante C, Ngu R, Leenhardt L. European Thyroid Association Guidelines for Ultrasound Malignancy Risk Stratification of Thyroid Nodules in Adults: The EU-TIRADS. *Eur Thyroid J* (2017) 6(5):225–37. doi: 10.1159/000478927
6. Hoang JK, Middleton WD, Farjat AE, Langer JE, Reading CC, Teefey SA, et al. Reduction in Thyroid Nodule Biopsies and Improved Accuracy With American College of Radiology Thyroid Imaging Reporting and Data System. *Radiology* (2018) 287(1):185–93. doi: 10.1148/radiol.2018172572
7. Ruan JL, Yang HY, Liu RB, Liang M, Han P, Xu XL, et al. Fine Needle Aspiration Biopsy Indications for Thyroid Nodules: Compare a Point-Based Risk Stratification System With a Pattern-Based Risk Stratification System. *Eur Radiol* (2019) 29(9):4871–8. doi: 10.1007/s00330-018-5992-z
8. Yoon JH, Lee HS, Kim EK, Moon HJ, Park VY, Kwak JY. Pattern-Based *vs.* Score-Based Guidelines Using Ultrasound Features Have Different Strengths in Risk Stratification of Thyroid Nodules. *Eur Radiol* (2020) 30(7):3793–802. doi: 10.1007/s00330-020-06722-y
9. Ha EJ, Na DG, Baek JH, Sung JY, Kim JH, Kang SY. US Fine-Needle Aspiration Biopsy for Thyroid Malignancy: Diagnostic Performance of Seven Society Guidelines Applied to 2000 Thyroid Nodules. *Radiology* (2018) 287(3):893–900. doi: 10.1148/radiol.2018171074
10. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
11. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical

Images Using Advanced Feature Analysis. *Eur J Cancer (Oxford Engl: 1990)* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
12. Jiang M, Li C, Tang S, Lv W, Yi A, Wang B, et al. Nomogram Based on Shear-Wave Elastography Radiomics Can Improve Preoperative Cervical Lymph Node Staging for Papillary Thyroid Carcinoma. *Thyroid: Off J Am Thyroid Assoc* (2020) 30(6):885–97. doi: 10.1089/thy.2019.0780
13. Liang J, Huang X, Hu H, Liu Y, Zhou Q, Cao Q, et al. Predicting Malignancy in Thyroid Nodules: Radiomics Score *Versus* 2017 American College of Radiology Thyroid Imaging, Reporting and Data System. *Thyroid: Off J Am Thyroid Assoc* (2018) 28(8):1024–33. doi: 10.1089/thy.2017.0525
14. Park VY, Lee E, Lee HS, Kim HJ, Yoon J, Son J, et al. Combining Radiomics With Ultrasound-Based Risk Stratification Systems for Thyroid Nodules: An Approach for Improving Performance. *Eur Radiol* (2021) 31(4):2405–13. doi: 10.1007/s00330-020-07365-9
15. Zhao CK, Ren TT, Yin YF, Shi H, Wang HX, Zhou BY, et al. A Comparative Analysis of Two Machine Learning-Based Diagnostic Patterns With Thyroid Imaging Reporting and Data System for Thyroid Nodules: Diagnostic Performance and Unnecessary Biopsy Rate. *Thyroid: Off J Am Thyroid Assoc* (2021) 31(3):470–81. doi: 10.1089/thy.2020.0305
16. Tessler FN, Middleton WD, Grant EG, Hoang JK, Berland LL, Teefey SA, et al. ACR Thyroid Imaging, Reporting and Data System (TI-RADS): White Paper of the ACR TI-RADS Committee. *J Am Coll Radiol: JACR* (2017) 14(5):587–95. doi: 10.1016/j.jacr.2017.01.046
17. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3d Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magn Reson Imaging* (2012) 30(9):1323–41. doi: 10.1016/j.mri.2012.05.001
18. Yip SSF, Parmar C, Blezek D, Estepar RSJ, Pieper S, Kim J, et al. Application of the 3D Slicer Chest Imaging Platform Segmentation Algorithm for Large Lung Nodule Delineation. *PloS One* (2017) 12(6):e0178944. doi: 10.1371/journal.pone.0178944
19. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.Can-17-0339
20. Huang YQ, Liang CH, He L, Tian J, Liang CS, Chen X, et al. Development and Validation of a Radiomics Nomogram for Preoperative Prediction of Lymph Node Metastasis in Colorectal Cancer. *J Clin Oncol: Off J Am Soc Clin Oncol* (2016) 34(18):2157–64. doi: 10.1200/jco.2015.65.9128
21. Li Y, Hong HG, Ahmed SE, Li Y. Weak Signals in High-Dimension Regression: Detection, Estimation and Prediction. *Appl Stochastic Models business Industry* (2019) 35(2):283–98. doi: 10.1002/asmb.2340
22. Kramer AA, Zimmerman JE. Assessing the Calibration of Mortality Benchmarks in Critical Care: The Hosmer-Lemeshow Test Revisited. *Crit Care Med* (2007) 35(9):2052–6. doi: 10.1097/01.Ccm.0000275267.64078.B0
23. Maia FF, Matos PS, Silva BP, Pallone AT, Pavin EJ, Vassallo J, et al. Role of Ultrasound, Clinical and Scintigraphyc Parameters to Predict Malignancy in Thyroid Nodule. *Head Neck Oncol* (2011) 3:17. doi: 10.1186/1758-3284-3-17
24. Xia J, Chen H, Li Q, Zhou M, Chen L, Cai Z, et al. Ultrasound-Based Differentiation of Malignant and Benign Thyroid Nodules: An Extreme Learning Machine Approach. *Comput Methods Programs Biomed* (2017) 147:37–49. doi: 10.1016/j.cmpb.2017.06.005
25. Bhatia KS, Lam AC, Pang SW, Wang D, Ahuja AT. Feasibility Study of Texture Analysis Using Ultrasound Shear Wave Elastography to Predict Malignancy in Thyroid Nodules. *Ultrasound Med Biol* (2016) 42(7):1671–80. doi: 10.1016/j.ultrasmedbio.2016.01.013
26. Nam SJ, Yoo J, Lee HS, Kim EK, Moon HJ, Yoon JH, et al. Quantitative Evaluation for Differentiating Malignant and Benign Thyroid Nodules Using Histogram Analysis of Grayscale Sonograms. *J Ultrasound Med: Off J Am Inst Ultrasound Med* (2016) 35(4):775–82. doi: 10.7863/ultra.15.05055
27. Luo P, Fang Z, Zhang P, Yang Y, Zhang H, Su L, et al. Radiomics Score Combined With ACR TI-RADS in Discriminating Benign and Malignant Thyroid Nodules Based on Ultrasound Images: A Retrospective Study. *Diagn (Basel Switzerland)* (2021) 11(6):1–12. doi: 10.3390/diagnostics11061011
28. Galimzianova A, Siebert SM, Kamaya A, Rubin DL, Desser TS. Quantitative Framework for Risk Stratification of Thyroid Nodules With Ultrasound: A Step Toward Automated Triage of Thyroid Cancer. *AJR Am J Roentgenology* (2020) 214(4):885–92. doi: 10.2214/ajr.19.21350
29. Song G, Xue F, Zhang C. A Model Using Texture Features to Differentiate the Nature of Thyroid Nodules on Sonography. *J Ultrasound Med: Off J Am Inst Ultrasound Med* (2015) 34(10):1753–60. doi: 10.7863/ultra.15.14.10045

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Huang, Wu, Zhou, Min, Qi, Zhang, Chen and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 13 October 2021
doi: 10.3389/fonc.2021.744756

Radiomics Based on Contrast-Enhanced MRI in Differentiation Between Fat-Poor Angiomyolipoma and Hepatocellular Carcinoma in Noncirrhotic Liver: A Multicenter Analysis

Xiangtian Zhao[1†], Yukun Zhou[2†], Yuan Zhang[1], Lujun Han[3], Li Mao[4], Yizhou Yu[4], Xiuli Li[4], Mengsu Zeng[5], Mingliang Wang[5*] and Zaiyi Liu[1*]*

[1] Department of Radiology, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, [2] Medical Imaging Center, The First Affiliated Hospital of Guangzhou University of Chinese Medicine, Guangzhou, China, [3] Department of Radiology, Sun Yat-Sen University Cancer Center, Guangzhou, China, [4] AI Lab, Deepwise Healthcare, Beijing, China, [5] Department of Radiology, Zhongshan Hospital, Fudan University, Shanghai, China

Objective: This study aims to develop and externally validate a contrast-enhanced magnetic resonance imaging (CE-MRI) radiomics-based model for preoperative differentiation between fat-poor angiomyolipoma (fp-AML) and hepatocellular carcinoma (HCC) in patients with noncirrhotic livers and to compare the diagnostic performance with that of two radiologists.

Methods: This retrospective study was performed with 165 patients with noncirrhotic livers from three medical centers. The dataset was divided into a training cohort (n = 99), a time-independent internal validation cohort (n = 24) from one center, and an external validation cohort (n = 42) from the remaining two centers. The volumes of interest were contoured on the arterial phase (AP) images and then registered to the venous phase (VP) and delayed phase (DP), and a total of 3,396 radiomics features were extracted from the three phases. After the joint mutual information maximization feature selection procedure, four radiomics logistic regression classifiers, including the AP model, VP model, DP model, and combined model, were built. The area under the receiver operating characteristic curve (AUC), diagnostic accuracy, sensitivity, and specificity of each radiomics model and those of two radiologists were evaluated and compared.

Results: The AUCs of the combined model reached 0.789 (95%CI, 0.579–0.999) in the internal validation cohort and 0.730 (95%CI, 0.563–0.896) in the external validation cohort, higher than the AP model (AUCs, 0.711 and 0.638) and significantly higher than the VP model (AUCs, 0.594 and 0.610) and the DP model (AUCs, 0.547 and 0.538). The diagnostic accuracy, sensitivity, and specificity of the combined model were 0.708, 0.625, and 0.750 in the internal validation cohort and 0.619, 0.786, and 0.536 in the external

OPEN ACCESS

Edited by:
Joseph Ross Mitchell, Independent Researcher, Tampa, FL, United States

Reviewed by:
Huaiqiang Sun, Sichuan University, China
Guolin Ma, China–Japan Friendship Hospital, China

***Correspondence:**
Xiangtian Zhao
weybey@163.com
Zaiyi Liu
zyliu@163.com
Mingliang Wang
wang.mingliang@zs-hospital.sh.cn

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *20 July 2021*
Accepted: *21 September 2021*
Published: *13 October 2021*

Citation:
Zhao X, Zhou Y, Zhang Y, Han L, Mao L, Yu Y, Li X, Zeng M, Wang M and Liu Z (2021) Radiomics Based on Contrast-Enhanced MRI in Differentiation Between Fat-Poor Angiomyolipoma and Hepatocellular Carcinoma in Noncirrhotic Liver: A Multicenter Analysis. Front. Oncol. 11:744756. doi: 10.3389/fonc.2021.744756

validation cohort, respectively. The AUCs for the two radiologists were 0.656 and 0.594 in the internal validation cohort and 0.643 and 0.500 in the external validation cohort. The AUCs of the combined model surpassed those of the two radiologists and were significantly higher than that of the junior one in both validation cohorts.

Conclusions: The proposed radiomics model based on triple-phase CE-MRI images was proven to be useful for differentiating between fp-AML and HCC and yielded comparable or better performance than two radiologists in different centers, with different scanners and different scanning parameters.

Keywords: hepatocellular carcinoma, angiomyolipoma, radiomics, magnetic resonance imaging, machine learning

INTRODUCTION

Hepatic angiomyolipoma (AML) is a mesenchymal benign tumor belonging to the perivascular epithelioid cell tumors (PEComas), which is a group of tumors believed to be derived from perivascular epithelioid cells and the co-expression of melanocytic and muscle marker. Histologically, it contains variable proportions of blood vessels, smooth muscle cells, and adipose tissue. Although only a few hundred cases of hepatic AMLs have ever been recorded all over the world, increasing numbers of cases are being reported due to the development of modern imaging techniques in recent years (1). The hepatic AML lesions often grow slowly and do not cause any clinical symptoms. Therefore, once the diagnosis of AML is established, conservative treatment and annual imaging follow-up is recommended in patients without indications for surgical resection (2). Typically, the diagnosis of AML is suggested in the case of a middle-aged woman when a solitary tumor occurs in a noncirrhotic liver and intratumoral macroscopic fat is detected on computed tomography (CT) or magnetic resonance imaging (MRI) (3). However, the amount of fat component in the hepatic AML varies greatly, ranging from 10 to 90% of the tumor volume (4) and, in some instances, cannot be easily identified on imaging (5, 6). In that case, many radiologists tend to misdiagnose these fat-poor AMLs (fp-AMLs) as other common hypervascular liver tumors, particularly hepatocellular carcinoma (HCC), with a frequency of 50% due to the overlapping imaging features (7), especially in areas with a high prevalence of hepatic viral infections like China. This can lead to unsuitable therapeutic schemes such as surgical therapy and liver transplantation. Therefore, it is crucial to accurately distinguish between fp-AML and HCC before surgery.

Unfortunately, correct preoperative diagnosis of fp-AML is currently challenging and mainly depends on histological findings. It is well known that a clinical history of chronic liver disease may be an important clue for the diagnosis of HCC, such as cirrhosis caused by hepatitis B virus (HBV) or hepatitis C virus (HCV) or excessive alcohol use. However, up to 20–30% of HCCs can develop in patients with normal livers (8). Hepatic AML has also been reported to occur in hepatitis B carriers (9). In terms of imaging, it has been proven that it is difficult to differentiate fp-AML from HCC in noncirrhotic liver by the use of only a dynamic enhancement pattern as most of the tumors are seen as a well-defined, hypervascular enhancing mass on arterial phase (AP), followed by a washout pattern on venous phase (VP) or equilibrium phase (6). Besides this, although previous studies pointed out that the presence of early draining vein and absent tumor capsule were useful findings for the differentiation of fp-AML from HCC in noncirrhotic liver (6, 10), these signs were subjective and dependent on the experience of the radiologist (5). In addition, preoperative fine needle aspiration cytology (FNAC) of AML can obtain definite histological evidence to improve the diagnostic accuracy with negligible risk (11). However, FNAC has some limitations because the trabecular growth pattern in hepatic epithelioid AML may mimic the cells of HCC (12).

Radiomics is an emerging field in image analysis, which extracts a large number of high-dimensional quantitative features from the image data and provides information that reflects the underlying pathophysiology (13). Several studies have proven that MRI-based radiomics features have the ability to discriminate different tumor phenotypes (14–17). We assumed that, using radiomics, we could extract and quantify the differences in conventional contrast-enhanced MRI (CE-MRI) images between fp-AML and HCC.

In this study, we aimed to develop a radiomics model based on triple-phase CE-MRI images to differentiate between fp-AML and HCC in the noncirrhotic liver and validate using external data. Moreover, we compared the diagnostic performance of radiomics model and radiologists in distinguishing these two kinds of tumors.

Abbreviations: CE, contrast-enhanced; MRI, magnetic resonance imaging; AML, angiomyolipoma; fp, fat-poor; HCC, hepatocellular carcinoma; VOI, volume of interest; AP, arterial phase; VP, venous phase; DP, delayed phase; AUC, area under the receiver operating characteristic curve; PEComas, perivascular epithelioid cell tumors; CT, computed tomography; HBV, hepatitis B virus; HCV, hepatitis C virus; FNAC, fine needle aspiration cytology; Gd-DTPA, gadolinium-diethylene triamine pentaacetic acid; T1W, T1-weighted; FS, fat-saturation; T2W, T2-weighted; LI-RADS, liver imaging reporting and data system; MITK, medical imaging interaction toolkit; ICC, inter- and intra- class correlation coefficient; LoG, Laplacian of Gaussian; GLCM, gray-level co-occurrence matrix; GLSZM, gray-level size zone matrix; GLRLM, gray-level run length matrix; GLDM, gray-level dependence matrix; JMIM, joint mutual information maximization; LR, logistic regression; ROC, receiver operating characteristic; IDN, inverse difference normalized; LGLZE, low gray-level zone emphasis.

MATERIALS AND METHODS

Patient Population

This multicenter retrospective study was carried out in three centers: Shanghai Zhongshan Hospital (center A), Guangdong

Sun Yat-Sen University Cancer Center (center B), and Guangdong Provincial People's Hospital (center C), approved by the institutional review board of each center, and patient informed consent was waived.

The patient enrollment process for this study is shown in **Figure 1**. First, a thorough search of the electronic medical record system of each center was performed between January 2012 and December 2019 for the diagnosis of hepatic AML. All the patients who both had a histologic diagnosis of AML and had undergone a liver MRI using the contrast agent gadolinium-diethylene triamine pentaacetic acid (Gd-DTPA) within 15 days before their surgery were included. The exclusion criteria were as follows: (1) patients with the presence of macroscopic intralesional fat on unenhanced T1-weighted (T1W) images (lose signal at fat saturation imaging or demonstrate etching artifact at the fat–water interface at chemical shift imaging) (18), (2) patients who received chemotherapy or radiotherapy before surgery, and (3) patients with insufficient CE-MRI image quality or improper timing for dynamic enhancement sequence.

To establish a control group, we subsequently searched the same databases of each hospital for an initial diagnosis of HCC during the same period by applying the same inclusion criteria. The exclusion criteria were as follows: (1) lesions with obvious necrosis, cyst, hemorrhage, or macroscopic fat, (2) lesions with hypo-enhancement on AP, (3) patients who received chemotherapy or radiotherapy before surgery, (4) patients with multiple HCCs, (5) patients with insufficient CE-MRI image quality or improper timing for dynamic enhancement sequence, (6) lesions with intrahepatic vascular invasion or extrahepatic metastases, and (7) patients with morphologic liver cirrhosis (19). Consequently, the patients who had a single and hypervascular HCC without definite evidence of morphologic cirrhosis were identified in each center. In view of the fact that AMLs are much less common than HCCs, we randomly selected some of these patients according to the ratio of 1:2 to alleviate the offset caused by the distribution and improve the statistical power (20), relative to the number of AML patients who were eventually enrolled in each center, using a commercially available random number generator (QuickCalcs, GraphPad).

In total, 165 patients were enrolled in this multicenter study, including 55 fp-AMLs (center A, n = 41; center B, n = 11; and center C, n = 3) and 110 HCCs (center A, n = 82; center B, n = 22; and center C, n = 6). Considering the small sample sizes of center B and center C, we grouped the patients from these two centers into one external validation cohort.

For center A, according to the TRIPOD statement, the patients were divided into training and internal validation cohorts according to the time of receiving surgical treatment and the ratio of 4:1. A total of 99 patients treated between February 2012 and January 2017 constituted the training cohort, whereas 24 patients treated between March 2017 and December 2019 constituted the internal validation cohort.

CE-MRI Image Acquisition

The MRI examinations were performed using 1.5- or 3.0-T systems from various vendors. At each center, the MRI protocols contained unenhanced images and dynamic sequences after an intravenous contrast agent injection, including axial fat saturation (fs) T2-weighted (T2W), T1W in-phase/out-of-phase, unenhanced axial fs T1W and dynamic triple-phase CE-MRI. All patients received 0.2 mmol/kg body weight of Gd-DTPA (Magnevist, Bayer Schering Pharma, Berlin, Germany) *via* a power injector (Spectris Solaris® EP MR, MEDRAD Inc., Indianola, IA, USA) at an infusion rate of 1.5–2 ml/s. After an intravenous contrast agent injection, three-dimensional fs T1W gradient-echo sequence [VIBE (Siemens Healthcare), LAVA (GE Healthcare), and THRIVE (Philips Healthcare)] was used to acquire dynamic enhanced images. The images in AP, VP, and delayed phase (DP) were acquired during suspended respiration at 25–35, 60–75, and 150–180 s, respectively. The detailed parameters of CE-MRI sequences used in each imaging center are reported in **Table 1**.

Radiologists Interpretation of the Enhanced MRI Images

Two abdominal radiologists (XZ and YZ, with 10 and 5 years of experience, respectively) independently reviewed the images of the internal and the external validation cohort. The radiologists

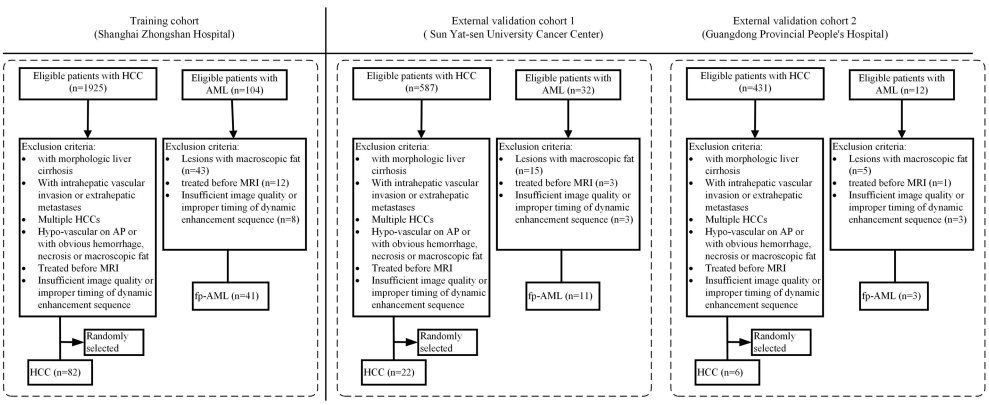


FIGURE 1 | The patient enrollment process for this study.

TABLE 1 | Detailed parameters of contrast-enhanced three-dimensional fs T1W gradient-echo sequences in each center.

Center	Scanner	Vendor	Field strength (t)	Patients	TR/TE (ms)	Matrix	Flip angle
Center A (*n* = 123)	Aera	Siemens	1.5	18	3.51/1.39	260 × 352	10°
	Avanto	Siemens	1.5	21	5.04/2.31	200 × 288	10°
	Ingenia	Philips	3.0	3	4.30/1.65	528 × 528	10°
	UIHMR560	UI	1.5	36	4.4/2.2	320 × 512	10°
	UIHMR770	UI	3.0	31	3.28/1.45	324 × 480	10°
	Verio	Siemens	3.0	14	4.07/1.43	250 × 352	9°
Center B (*n* = 33)	Achieva	Philips	3.0	3	3.12/1.51	480 × 480	10°
	Aera	Siemens	1.5	4	4.63/2.16	460 × 640	10°
	Discovery MR750	GE	3.0	4	4.05/1.64	512 × 512	15°
	Signa HDxt	GE	1.5	5	3.98/1.90	512 × 512	15°
	Trio	Siemens	3.0	10	4.15/1.86	250 × 320	9°
	uMR780	UI	3.0	7	3.3/1.45	336 × 480	10°
Center C (*n* = 9)	Ingenia	Philips	3.0	5	4.01/1.94	384 × 384	10°
	Achieva	Philips	3.0	4	4.02/1.94	384 × 384	10°

FS, fat-suppressed; UI, United Imaging; GE, General Electric; TE, echo time; T1W, T1-weighted; TR, repetition time.

were blinded to clinical information and did not know the exact number of each type of tumor but were aware that the tumors were finally diagnosed with fp-AML or HCC. The two radiologists assessed each specific phase and judged this based on the signal intensity of majority of the tumor. According to the features defined with reference to the definitions and annotations in the Liver Imaging Reporting and Data System (LI-RADS) (21), the main signs that were often used for differential diagnosis between fp-AML and HCC were recorded, including the draining hepatic vein and intra-tumor vessel, the presence of a complete capsule, and the pattern of enhancement (wash in and wash out or prolonged enhancement). When the lesion demonstrated specific MRI features such as intra-tumor vessel, draining hepatic vein, prolonged enhancement, no washout in the VP, and lack of complete capsule, it would be classified as fp-AML; otherwise, it would be classified as HCC (3, 10).

Radiomics Workflow

An overview of our workflow is illustrated in **Figure 2**. Firstly, the enhanced MRI data were collected, including the AP, VP, and DP images. Then, the images of each phase were normalized by the histogram-matching method. The delineation was performed on AP and then registered to the other two phases, and the misalignment was manually corrected. For each phase, the radiomic features were extracted from the tumor region of the original images and the preprocessed images. Finally, the feature selection method was used to select the optimal feature subset. The

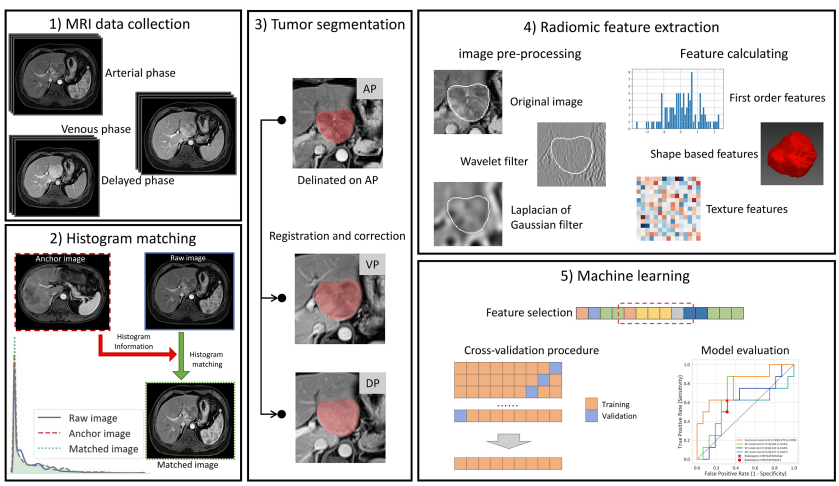


FIGURE 2 | The workflow of our study. (1) The collection of CE-MRI data, including the arterial phase (AP), venous phase, and delayed phase images. (2) Histogram matching: The images of each phase were matched to the corresponding phase of the first patient by histogram matching. (3) Tumor segmentation: The delineation was performed on AP and then registered to the other two phases, and the misalignment was manually corrected. (4) Radiomic feature extraction: For each phase, the radiomic features were extracted from the tumor region of the original images and the preprocessed images. (5) Machine learning. The feature selection method was used to select the optimal feature subset, and then the models were trained by the cross-validation procedure and evaluated on the internal and external validation cohort.

models were trained by the cross-validation procedure and evaluated in the internal and external validation cohort.

The Segmentation of Tumor Images

The tumor volume of interest (VOI) was manually delineated slice by slice using the Medical Imaging Interaction Toolkit (MITK) software (v.2013.12.00, Heidelberg, Germany), referencing to the sagittal and coronal images reconstructed by the software. To reduce the workload of segmentation and increase the accuracy of tumor contouring, differently from many previous studies, the manual VOI delineation was performed only on the AP images in our study, and delineation was registered to the DP and VP images by DEEDs, an efficient 3D discrete deformable alignment algorithm (22), in accordance with the image information of the three phases. It was proved that the DEEDs algorithm outperformed the other common registration algorithm and achieved a dice coefficient of 0.70 for the four large organs (liver, spleen, and kidneys) (23). Then, the misalignment between the image and the registered contour on the other two phases was manually corrected. In this way, when the part or whole of the tumor had a similar signal intensity to the surrounding liver parenchyma and it was difficult to manually outline the contour on a single phase, the VOI of AP could be used for reference, and the tumor contour could be relatively accurately confirmed under the condition of triple-phase image registration.

The inter-observer reliability and intra-observer reproducibility of feature extraction were tested using the inter- and intra-class correlation coefficients (ICCs). After 30 cases of CE-MRI images (10 fp-AMLs and 20 HCCs) were selected randomly, radiologist 1 (XZ) and radiologist 2 (YZ) performed VOI segmentation manually, respectively. Radiologist 2 repeated the VOI segmentation 2 weeks later to assess the intra-observer reproducibility. The feature extraction was considered to represent a good agreement when the ICC was greater than 0.8. The remaining image segmentation was performed by radiologist 2 and reviewed by radiologist 1.

Radiomics Feature Extraction

In the case of MRI, the signal intensity values vary according to the acquisition parameters used, which affect the extracted radiomic features (24). To calibrate the variations due to the scanner manufacturer and magnetic field strength in our cohort, histogram standardization (25) was used to match the input image histogram onto the standard image (in our case, the MRI of the first patient in the training cohort).

Radiomics extraction was performed using Pyradiomics V2.1.0. The images were resampled to a pixel spacing of 1 × 1 × 1 mm to counteract the interference caused by the non-uniform spatial resolution. Then, the original images were preprocessed by the wavelet filters or Laplacian of Gaussian filters with different parameters. For each phase, 1,132 radiomic features were obtained from the original images and the preprocessed images: (1) 234 first-order features, (2) 14 shape-based features, (3) 286 gray-level co-occurrence matrix features (GLCM), (4) 208 gray-level size zone matrix features (GLSZM), (5) 208 gray-level run length matrix features, and (6) 182 gray-level dependence matrix features (GLDM). Finally, a total of 3,396 radiomic features were extracted from triple-phase CE-MRI images for each patient.

Construction and Validation of the Radiomics Signatures

The radiomic features extracted from the AP, VP, and DP images were used to build the AP model, VP model, and DP model, respectively. Then, the combined model was trained on all radiomic features of the images of three phases. The construction strategies of the four models were the same.

The features were normalized by Z-score normalization before the model building. To avoid information disclosure, the mean and standard deviation values were calculated only on the training set, and the entire dataset was normalized by the mean and standard deviation values from the training set. The features with poor consistency (intra-ICC or inter-ICC lower than 0.8) were filtered out. To reduce the redundancy of the features and to avoid overfitting, the joint mutual information maximization (JMIM) method (25), which utilizes mutual information and the maximum–minimum criterion, was used to select the subset of features. Considering the sample size of the training cohort, 10 radiomic features (10% of the sample size of the training cohort) were selected to avoid over-fitting (26). The logistic regression (LR) model was built by a repetitive (five runs) 10-fold cross-validation using the training cohort. After the hyper-parameters were determined by the cross-validation procedure, the LR model with optimal parameters was built on the entire training cohort.

The area under the receiver operator characteristic (ROC) curve was used to evaluate the performance of the radiomic models. After the cutoff value that maximizes the Youden Index was obtained on the cross-validation result, the accuracy, sensitivity, and specificity were also calculated. The output of the prediction was calibrated by the isotonic regression method.

Statistical Analysis

The ROC curves were drawn by using Matplotlib (version 3.1.0), and the area under the ROC curve (AUC), accuracy, sensitivity, and specificity were calculated by the Scikit-learn python package (version 0.20.3). The kappa consistency test was adopted to assess inter-observer agreement between the two radiologists. The level of agreement was interpreted as slight if $\kappa$was 0.01 to 0.20, fair if 0.21 to 0.40, moderate if 0.41 to 0.60, substantial if 0.61 to 0.80, and almost perfect if 0.81 to 1. The DeLong test was used for pairwise comparisons between the combined model and the remaining three models and between the best-performing radiomics model and each radiologist. For the comparison of the sensitivity and specificity between the best-performing radiomics model and the assessment of the radiologists, the McNemar chi-square test was employed. The abovementioned statistical analysis was performed on R software (version 3.6.0; https://www.r-project.org/) environments. A two-sided p <0.05 was considered statistically significant throughout the study.

RESULTS

Patient Demographics

The mean age of patients in the fp-AML group was lower than that of patients in the HCC group (47.1 ± 12.6 *vs.* 55.8 ± 12.0 years,

t = -4.306, $p < 0.001$). Male predominance was observed in the HCC group, while most fp-AML patients were female [87% (48/55) *vs.* 18% (20/110), $p < 0.001$].

There was no patient with tuberous sclerosis in the fp-AML group. In the HCC group, more patients had a preexisting chronic liver disease that was caused by chronic HBV or HCV infection, compared with those with fp-AML [79% (87/110) *vs.* 9% (5/55), $p < 0.001$). None of the HCCs was of the fibrolamellar variant.

Radiomics Analysis

Of the 3,396 radiomics features extracted from AP, VP, and DP images, 2,585 were demonstrated to have a good inter- and intra-observer agreement, including 958 AP features, 823 VP features, and 804 DP features. Then, the JMIM feature selection method selected 10 optimal features for each model. For the combined model, the 10 optimal features included seven features from AP, two features from DP, and one feature from VP.

The detailed diagnostic performance of each model is shown in **Table 2**. The AUCs of the AP model, the VP model, and the DP model reached 0.711 (95%CI, 0.489–0.933), 0.594 (95%CI, 0.339–0.848), and 0.547 (95%CI, 0.257–0.837) in the internal validation cohort and 0.638 (95%CI, 0.466–0.809), 0.61 (95%CI, 0.434–0.786), and 0.538 (95%CI, 0.355–0.722) in the external validation cohort. The combined model reached the AUC of 0.789 (95%CI, 0.579–0.999) in the internal validation cohort, significantly higher than the VP model ($p = 0.015$) and the DP model ($p = 0.004$). In the external validation cohort, the AUC of the combined model was 0.730 (95%CI, 0.563–0.896), also significantly higher than the VP model ($p = 0.035$) and the DP model ($p = 0.008$). The ROC curves are shown in **Figure 3**.

With the cutoff value of 0.6 that maximizes the Youden Index, the accuracy, sensitivity, and specificity of the combined model reached 0.708, 0.625, and 0.75 in the internal validation cohort and 0.619, 0.786, and 0.536 in the external validation cohort, respectively. The accuracy and the specificity of the combined model were higher or comparable than the other three single-phase models in the internal and the external validation cohorts. The sensitivity of the combined model was not lower than the other single-phase models on the internal validation cohort and the external validation cohort, except for the AP model.

The beta coefficients of the combined model were viewed as the importance of the features (illustrated in **Figure 4**), and the formula used to calculate the predicted probability of fp-AML by the combined model is listed in **Supplementary Data S1**. The features that contributed most to the diagnosis of fp-AML were wavelet-LLL_firstorder_RootMeanSquared_ap, wavelet-LLL_firstorder_Mean_ap, and original_firstorder_90Percentile_ap. On the other side, the features that contributed most to the diagnosis of HCC were wavelet-LHL_firstorder_Mean_vp, wavelet-LHH_glszm_LowGrayLevelZon-eEmphasis_ap, and wavelet-HLL_firstorder_RootMeanSquared_vp. The waterfall figure of the calibrated prediction results of each case is shown in **Figure 5**.

Compared With the Interpretation of the Radiologists

The comparison of diagnostic performance between the models and the radiologists is shown in **Figure 3** and **Table 3**. The interobserver agreement (κ) values between the two radiologists were 0.565 in the internal validation cohort and 0.146 in the external validation cohort. The AUCs of radiologist 2 reached 0.594 (95%CI, 0.375–0.813) and 0.500 (95%CI, 0.351–0.649) in the internal and external validation cohorts, respectively, significantly inferior to the combined model ($p = 0.043$ and 0.027). The AUCs of radiologist 1 were 0.656 (95%CI, 0.442–0.871) and 0.643 (95%CI, 0.486–0.799) in the internal and external validation cohorts, respectively, tending to be lower than the combined model, but not significant (both $p > 0.05$).

TABLE 2 | The detailed performance of arterial phase (AP) model, venous phase (VP) model, delayed phase (DP) model, and combined model.

	Model	Area under the receiver operating characteristic curve (95%CI)	Accuracy	Sensitivity	Specificity
Training cohort (n = 99)	AP	0.863 (0.776–0.95)	0.798 (79/99)	0.848 (28/33)	0.773 (51/66)
	VP	0.756 (0.659–0.853)	0.636 (63/99)	0.879 (29/33)	0.515 (34/66)
	DP	0.752 (0.647–0.856)	0.657 (65/99)	0.909 (30/33)	0.53 (35/66)
	Combined	0.866 (0.78–0.953)	0.828 (82/99)	0.758 (25/33)	0.864 (57/66)
Cross-validation (n = 99)	AP	0.826 (0.729–0.923)	0.808 (80/99)	0.818 (27/33)	0.803 (53/66)
	VP	0.708 (0.605–0.811)	0.677 (67/99)	0.818 (27/33)	0.606 (40/66)
	DP	0.6 (0.484–0.715)	0.535 (53/99)	0.788 (26/33)	0.409 (27/66)
	Combined	0.841 (0.747–0.936)	0.848 (84/99)	0.758 (25/33)	0.894 (59/66)
Internal validation cohort (n = 24)	AP	0.711 (0.489–0.933)	0.625 (15/24)	0.875 (7/8)	0.5 (8/16)
	VP	0.594* (0.339–0.848)	0.625 (15/24)	0.625 (5/8)	0.625 (10/16)
	DP	0.547** (0.257–0.837)	0.375 (9/24)	0.625 (5/8)	0.25 (4/16)
	Combined	0.789 (0.579–0.999)	0.708 (17/24)	0.625 (5/8)	0.75 (12/16)
External validation cohort (n = 42)	AP	0.638 (0.466–0.809)	0.524 (22/42)	0.929 (13/14)	0.321 (9/28)
	VP	0.61* (0.434–0.786)	0.595 (25/42)	0.714 (10/14)	0.536 (15/28)
	DP	0.538** (0.355–0.722)	0.405 (17/42)	0.714 (10/14)	0.25 (7/28)
	Combined	0.73 (0.563–0.896)	0.619 (26/42)	0.786 (11/14)	0.536 (15/28)

The p-value was calculated by the De Long's test.
**p < 0.05, **p < 0.01.*
AP, arterial phase; VP, venous phase; DP, delayed phase.

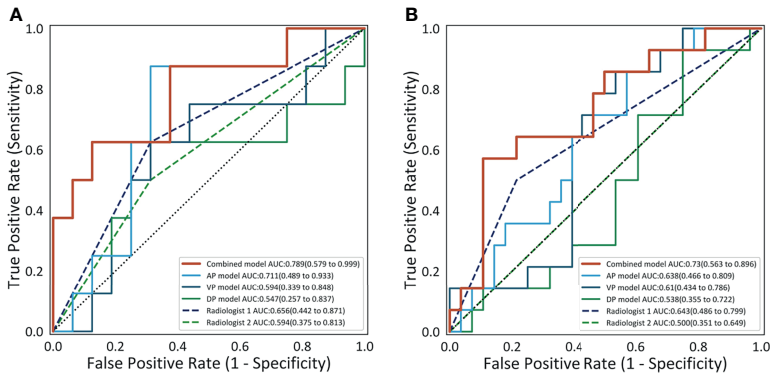


FIGURE 3 | The receiver operating characteristic curves of the four models and the performance of the two radiologists on the internal validation cohort **(A)** and on the external validation cohort **(B)**.

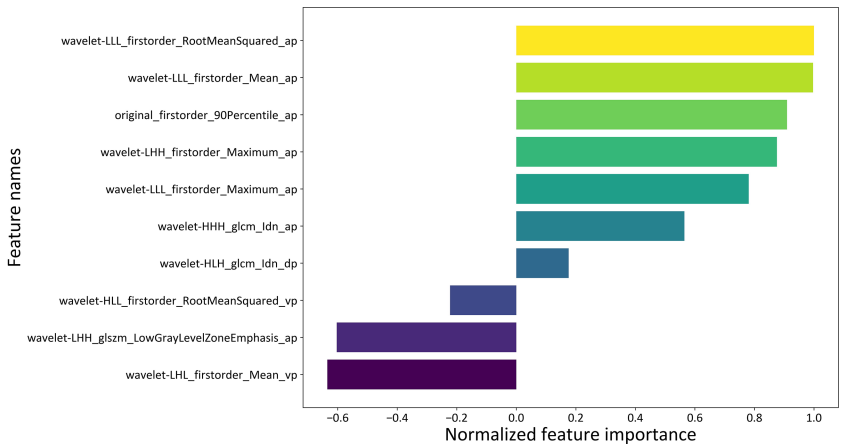


FIGURE 4 | The importance of the features of the combined model.

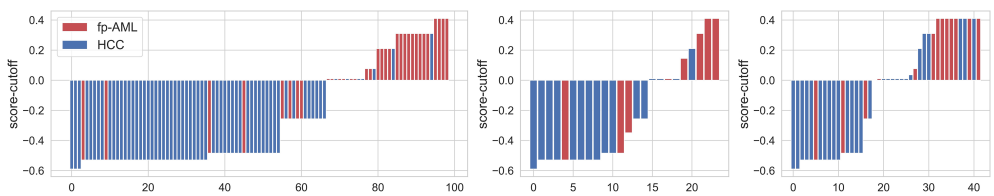


FIGURE 5 | The calibrated radiomics scores for each patient in the training, internal validation, and external validation cohorts. The red bars represent the scores for fat-poor angiomyolipoma patients, while the blue bars represent the scores for the hepatocellular carcinoma patients.

In the internal and external validation cohorts, the differences in accuracy, sensitivity, and specificity between the combined model and each radiologist were not statistically significant (all $p > 0.05$), except that the sensitivity of radiologist 2 was significantly lower than that of the combined model ($p = 0.023$). Representative cases in which diagnoses were corrected using the radiomics approach are shown in **Figure 6**.

DISCUSSION

The present study showed that the combined radiomics model incorporating triple-phase CE-MRI images had a favorable predictive value for differentiating fp-AML from HCC in patients without morphological liver cirrhosis, with the AUCs of 0.866, 0.789, and 0.730, respectively, in the training cohort,

TABLE 3 | The detailed comparison between the performance of two radiologists and the combined model.

	AUC	Accuracy	Sensitivity	Specificity
Internal validation cohort (n = 24)				
Radiologist 1	0.656 (0.442–0.871)	0.667 (16/24)	0.625 (5/8)	0.688 (11/16)
Radiologist 2	0.594 (0.375–0.813)*	0.625 (15/24)	0.500 (4/8)	0.688 (11/16)
Model	0.789 (0.579–0.999)	0.708 (17/24)	0.625 (5/8)	0.750 (12/16)
External validation cohort (n = 42)				
Radiologist 1	0.643 (0.486–0.799)	0.690 (29/42)	0.500 (7/14)	0.786 (22/28)
Radiologist 2	0.500 (0.351–0.649)*	0.571 (24/42)	0.286 (4/14)*	0.714 (20/28)
Model	0.730 (0.563–0.896)	0.619 (26/42)	0.786 (11/14)	0.536 (15/28)

The p-value was calculated by the De Long's test or McNemar chi-square test when appropriate.
**p < 0.05.*
AUC, area under the receiver operating characteristic curve.

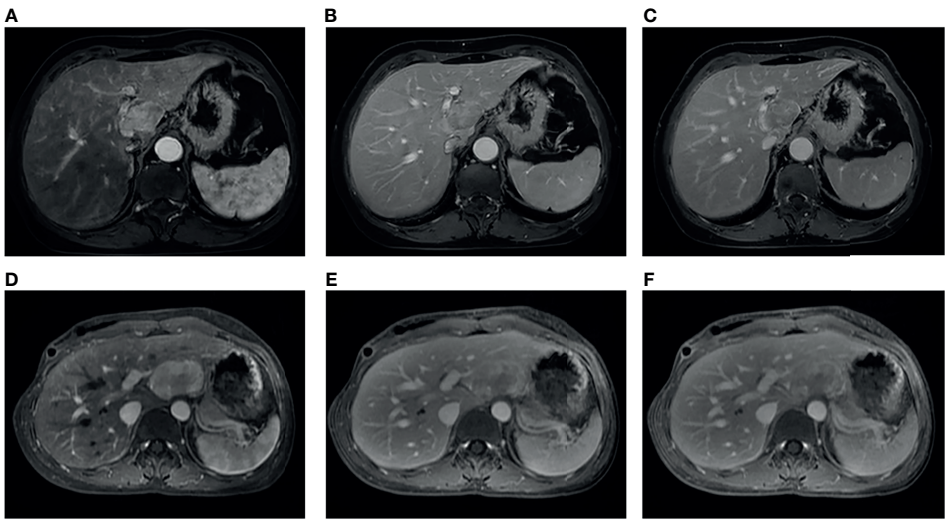


FIGURE 6 | Two representative cases: case 1 **(A–C)**, a 65-year-old female without chronic hepatitis B virus (HBV) infection, and case 2 **(D–F)**, a 36-year-old female with chronic HBV infection. These two cases were both misdiagnosed as hepatocellular carcinoma by two radiologists, whereas the model output was consistent with the correct diagnosis of fat-poor angiomyolipoma.

internal validation cohort, and external validation cohort. The performance of the model was comparable to that of an experienced radiologist with 10 years of experience and better than that of a junior radiologist with 5 years of experience in both the internal validation cohort and the external validation cohort. To the best of our knowledge, this study represents the first multi-center and multi-scanner assessment of the role of multi-phase CE-MRI-based machine learning to differentiate fp-AML from HCC with a large sample size. The performance of this approach in the external validation cohort is encouraging, which suggests its potential to augment the diagnostic performances of radiologists, even in different centers with different scanners or different scanning parameters.

Many previous studies have used various strategies to discriminate between fp-AML and HCC. Due to the rarity of hepatic AML, especially the cases with no or minimal fat, most of these studies enrolled a small number of patients. A study with a relatively large sample size of 30 hepatic epithelioid AML indicated that specific MRI features, such as intra-tumor vessel, draining hepatic vein, prolonged enhancement, and lack of capsule, may contribute to a more confident diagnosis, consistent with the results of some previous reports (7, 27). However, some authors have put forward different views. Kim et al. (5) investigated 12 patients with lipid-poor AML and 27 patients with HCC and analyzed the presence of peripheral capsule and several imaging features related with the vascular components of AMLs on MRI images, including the feeding artery dilatation, multiple aneurysmal arteries, and the early draining veins. They found that none of these imaging features was significantly different between lipid-poor AML and HCC. The authors speculated that this could be explained by the fact that AML and HCC were both hypervascular tumors that frequently shared similar imaging features related to their vascular component and also might be attributed to the weaker arterial enhancement of gadoxetic acid and the lower spatial resolution of MRI compared to CT. In comparison to the

different literatures regarding the frequency of the drainage veins on CT or MRI images in AML patients, the value varied greatly, ranging from 25 to 83.3%, according to previous reports (5, 6, 28). As to the incidence of tumor psudocapsule, this feature was reported to be found in 11.1– 42% of AMLs (3, 5). Those differences, on the one hand, could be attributed to the differences of sample sizes, patient composition, and scanning methods among these studies; on the other hand, this could also mean that the evaluation of these imaging signs is subjective and depends on the experiences of the radiologist—for example, it was pointed out that sometimes it was hard to differentiate the enhanced tumor vessels in the peripheral portion from the tumor capsules in AMLs (28). This may also explain the poor or moderate degree of agreement of the diagnosis results of the two radiologists in our study. Although having been properly trained before on the interpretation of MRI images for an accurate understanding of the useful imaging signs and following clear instructions for diagnosis in our study, radiologist 2 showed a relatively low and unsatisfactory level of sensitivity in the diagnosis in the external validation cohort due to limited diagnostic experience. Besides this, Kim et al. (5) also proposed that lipid-poor AML frequently showed more homogeneous hypointensity than HCC on the hepatobiliary phase of gadoxetic acid-enhanced MRI, and this feature could better differentiate these two diseases. However, gadoxetic acid-enhanced MRI is currently not the first-line examination of focal hepatic lesions in China due to its higher price than the conventional extracellular contrast agents and the longer scan time. Hence, compared to subjective and qualitative analyses, radiomics is objective, quantitative, and reproducible. Moreover, the radiomic analysis based on Gd-DTPA-enhanced conventional MRI images in our study does not require additional scanning time and cost, and it might prove to be a practical tool.

Our study indicated that, compared with the VP and DP models, the AP radiomics model showed a higher AUC. After adding the three phases of images together to form a combined model, the final radiomics signature contained 10 features: seven from the AP, two from the VP, and one from the DP. These results indicated that AP played a major role in distinguishing these two tumors. Although it is well known that AML and HCC usually both demonstrate intense contrast enhancement, these two tumors still seem to be different on AP. It had been proved that tumoral vessels connecting with the early draining vein in AML were more prominent and ectatic than those in HCC, and the latter tends to be faint and negligible (6). Thus, even if not showing obvious differences by visual assessment, conventional AP images might still reflect underlying, invisible, histological differences. Our results suggested that radiomics could detect these microscopic differences between fp-AML and HCC contained in routine AP images. Actually, our results were consistent with a previous study based on the measurement of mean attenuation values with 12 patients who underwent CT (28). In that study, the authors demonstrated that the hepatic AML appeared to have a more intense contrast enhancement and higher mean attenuation values exceeding 120 HU than that of HCCs on AP.

As far as we know, there is only one study based on MRI radiomics to distinguish hepatic AML from HCC. Recently, Liang et al. (7) demonstrated that the radiomics model based on AP images performed well in distinguishing epithelioid AML from HCC and focal nodular hyperplasia, especially for MRI. This was similar to our results; however, the study had not been externally verified, and the performance of the model in other participant data was not clear (26). Furthermore, unlike our study, they only used the AP images and single-layer region of interest. As mentioned above, the accuracy and specificity of the combined model were higher or comparable than the other three single-phase models in the internal and external validation cohorts in our study. Therefore, it is a better choice to combine data from multiple phases. However, whether the VOI analysis is superior to the single-layer analysis is still an unresolved question. Considering that previous studies (29, 30) had confirmed that a whole-tumor analysis had higher inter-observer consistency and better ability to reflect tumor heterogeneity than a two-dimensional analysis, we used VOI analysis in this study.

A previous study has explained the relationships between image features and texture parameters (31), which have different meanings and are expected to be related to histological features that reflect tumor heterogeneity. In our study, fp-AML was positively associated with wavelet-GLCM-inverse difference normalized on DP and AP. Interestingly, this feature reflects the local homogeneity of an image, which may be explained by the lower tissue homogeneity in HCC compared with that in fp-AMLs. Moreover, we found that HCC was significantly associated with the histogram parameters on VP, which reflected the characteristics of earlier washout on VP in HCCs (3). Besides this, GLSZM-low-gray-level-zone-emphasis (LGLZE) on AP was one of the top three ranked parameters for predicting HCC in our study. The GLSZM provides information on the size of the homogeneous zones for each gray level in three dimensions, and LGLZE is the distribution of the low gray level zones. According to a previous study, compared to high gray-level values, gray level runs with low gray-level values in two-dimensional images of the cell nuclei in ovarian cancer patients, indicating a higher probability for strong invasion ability and a poor prognosis (32), which seemed to be in agreement with our results.

There are several limitations in our study. Firstly, compared with HCC, fp-AML is encountered less frequently in clinical practice owing to its rarity. Although we performed a multicentric trial employing a relatively larger sample size, the number of patients with fp-AML was still far less than the patients with HCC. However, our study of 55 patients represents, to our knowledge, the largest cohort of hepatic fp-AML patients analyzed for differential diagnosis of HCC so far. In addition, we followed a fp-AML–HCC ratio of 1:2 to lessen the impact of imbalanced datasets that exist. Secondly, using multicentric CE-MRI datasets for radiomic feature extraction can pose a greater challenge due to the variations resulting from differences in imaging equipment and acquisition parameters. To overcome this problem, we adopted the histogram matching techniques to correct scanner-dependent intensity variations. Besides this, it has been proved that if the spatial resolution of the MRI images used in radiomics analysis is high enough, it can offset the influence of different scan parameters on the results (33).

In our study, all three centers adopted three-dimensional fs T1W gradient-echo sequence for dynamic enhancement imaging, which provided high-slice selective spatial resolution (2 to 3 mm) (34). Thirdly, radiomics signature was constructed using CE-MRI images only in this multicenter study. The reason was that the CE-MRI images were retrospectively collected, so we finally adopted only the enhanced sequence to obtain the largest possible sample size.

In conclusion, this multicenter study indicates the proposed CE-MRI-based radiomics model incorporating triple-phase images that can be useful for differentiating between fp-AML and HCC and yields comparable or better performance than that of two radiologists in both the internal validation cohort and the external validation cohort.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the institutional review board of Zhongshan Hospital, Guangdong Sun Yat-Sen University Cancer Center, Guangdong Provincial People's Hospital. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YKZ and XZ carried out the studies, participated in collecting data, and drafted the manuscript. YZ and MW carried out the studies and participated in collecting data. LH participated in collecting data. LM, YY, and XL performed the statistical analysis. MW, MZ, and ZL participated in its design. All authors contributed to the article and approved the submitted version.

FUNDING

The authors declare that this study received funding from the Key R&D Program of Guangdong Province, China (2021B0101420006), the National Science Fund for Distinguished Young Scholars (81925023), the National Natural Science Foundation of China (82102145), the Shanghai Municipal Key Clinical Specialty (shslczdzk03202) and the Guangdong Medical Research Fund (A2021086). The funders were not involved in the study design, collection, analysis, and interpretation of data, the writing of this article, or the decision to submit it for publication.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.744756/full#supplementary-material

REFERENCES

1. Klompenhouwer AJ, Dwarkasing RS, Doukas M, Pellegrino S, Vilgrain V, Paradis V, et al. Hepatic Angiomyolipoma: An International Multicenter Analysis on Diagnosis, Management and Outcome. *HPB (Oxford)* (2020) 22:622–29. doi: 10.1016/j.hpb.2019.09.004
2. Calame P, Tyrode G, Weil Verhoeven D, Felix S, Klompenhouwer AJ, Di Martino V, et al. Clinical Characteristics and Outcomes of Patients With Hepatic Angiomyolipoma: A Literature Review. *World J Gastroenterol* (2021) 27:2299–311. doi: 10.3748/wjg.v27.i19.2299
3. Lee SJ, Kim SY, Kim KW, Kim JH, Kim HJ, Lee MG, et al. Hepatic Angiomyolipoma Versus Hepatocellular Carcinoma in the Noncirrhotic Liver on Gadoxetic Acid-Enhanced MRI: A Diagnostic Challenge. *AJR Am J Roentgenol* (2016) 207:562–70. doi: 10.2214/AJR.15.15602
4. Wang Z, Xu HX, Xie XY, Xie XH, Kuang M, Xu ZF, et al. Imaging Features of Hepatic Angiomyolipomas on Real-Time Contrast-Enhanced Ultrasound. *Br J Radiol* (2010) 83:411–8. doi: 10.1259/bjr/81174247
5. Kim R, Lee JM, Joo I, Lee DH, Woo S, Han JK, et al. Differentiation of Lipid Poor Angiomyolipoma From Hepatocellular Carcinoma on Gadoxetic Acid-Enhanced Liver MR Imaging. *Abdom Imaging* (2015) 40:531–41. doi: 10.1007/s00261-014-0244-4
6. Jeon TY, Kim SH, Lim HK, Lee WJ. Assessment of Triple-Phase CT Findings for the Differentiation of Fat-Deficient Hepatic Angiomyolipoma From Hepatocellular Carcinoma in non-Cirrhotic Liver. *Eur J Radiol* (2010) 73:601–6. doi: 10.1016/j.ejrad.2009.01.010
7. Liang W, Shao J, Liu W, Ruan S, Tian W, Zhang X, et al. Differentiating Hepatic Epithelioid Angiomyolipoma From Hepatocellular Carcinoma and Focal Nodular Hyperplasia. *Via Radiomics Models Front Oncol* (2020) 10:564307. doi: 10.3389/fonc.2020.564307
8. Sasaki K, Shindoh J, Margonis GA, Nishioka Y, Andreatos N, Sekine A, et al. Effect of Background Liver Cirrhosis on Outcomes of Hepatectomy for Hepatocellular Carcinoma. *JAMA Surgery* (2017) 152:e165059–e59. doi: 10.1001/jamasurg.2016.5059
9. Hwang JY, Lee SW, Baek YH, Kim JH, Kim HY, Bae SH, et al. A Case of Hepatic Angiomyolipoma Which was Misdiagnosed as Hepatocellular Carcinoma in a Hepatitis B Carrier. *Case Rep Hepatol* (2012) 2012:606108–08. doi: 10.1155/2012/606108
10. Liu W, Wang J, Huang Q, Lu Q, Liang W. Comparison of MRI Features of Epithelioid Hepatic Angiomyolipoma and Hepatocellular Carcinoma: Imaging Data From Two Centers. *Front Oncol* (2018) 8:600. doi: 10.3389/fonc.2018.00600
11. Zhou H, Guo M, Gong Y. Challenge of FNA Diagnosis of Angiomyolipoma: A Study of 33 Cases. *Cancer Cytopathol* (2017) 125:257–66. doi: 10.1002/cncy.21824
12. Xie L, Jessurun J, Manivel JC, Pambuccian SE. Hepatic Epithelioid Angiomyolipoma With Trabecular Growth Pattern: A Mimic of Hepatocellular Carcinoma on Fine Needle Aspiration Cytology. *Diagn Cytopathol* (2012) 40:639–50. doi: 10.1002/dc.21703
13. Lewis S, Hectors S, Taouli B. Radiomics of Hepatocellular Carcinoma. *Abdom Radiol (NY)* (2021) 46:111–23. doi: 10.1007/s00261-019-02378-5
14. Li Y, Jian J, Pickhardt PJ, Ma F, Xia W, Li H, et al. MRI-Based Machine Learning for Differentiating Borderline From Malignant Epithelial Ovarian Tumors: A Multicenter Study. *J Magn Reson Imaging* (2020) 52:897–904. doi: 10.1002/jmri.27084

15. Wu J, Liu A, Cui J, Chen A, Song Q, Xie L. Radiomics-Based Classification of Hepatocellular Carcinoma and Hepatic Haemangioma on Precontrast Magnetic Resonance Images. *BMC Med Imaging* (2019) 19:23. doi: 10.1186/s12880-019-0321-9
16. Chen C, Ou X, Wang J, Guo W, Ma X. Radiomics-Based Machine Learning in Differentiation Between Glioblastoma and Metastatic Brain Tumors. *Front Oncol* (2019) 9:806. doi: 10.3389/fonc.2019.00806
17. Xu M, Fang M, Zou J, Yang S, Yu D, Zhong L, et al. Using Biparametric MRI Radiomics Signature to Differentiate Between Benign and Malignant Prostate Lesions. *Eur J Radiol* (2019) 114:38–44. doi: 10.1016/j.ejrad.2019.02.032
18. Pokharel SS, Macura KJ, Kamel IR, Zaheer A. Current MR Imaging Lipid Detection Techniques for Diagnosis of Lesions in the Abdomen and Pelvis. *RadioGraphics* (2013) 33:681–702. doi: 10.1148/rg.333125068
19. Vilgrain V, Lagadec M, Ronot M. Pitfalls in Liver Imaging. *Radiology* (2015) 278:34–51. doi: 10.1148/radiol.2015142576
20. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of Controls in Case-Control Studies. III. Design Options. *Am J Epidemiol* (1992) 135:1042–50. doi: 10.1093/oxfordjournals.aje.a116398
21. Tang A, Bashir MR, Corwin MT, Cruite I, Dietrich CF, Do RKG, et al. Evidence Supporting LI-RADS Major Features for CT- and MR Imaging-Based Diagnosis of Hepatocellular Carcinoma: A Systematic Review. *Radiology* (2018) 286:29–48. doi: 10.1148/radiol.2017170554
22. Heinrich MP, Jenkinson M, Brady M, Schnabel JA. MRF-Based Deformable Registration and Ventilation Estimation of Lung CT. *IEEE Trans Med Imaging* (2013) 32:1239–48. doi: 10.1109/TMI.2013.2246577
23. Xu Z, Lee CP, Heinrich MP, Modat M, Rueckert D, Ourselin S, et al. Evaluation of Six Registration Methods for the Human Abdomen on Clinically Acquired CT. *IEEE Trans BioMed Eng* (2016) 63:1563–72. doi: 10.1109/TBME.2016.2574816
24. Um H, Tixier F, Bermudez D, Deasy JO, Young RJ, Veeraraghavan H. Impact of Image Preprocessing on the Scanner Dependence of Multi-Parametric MRI Radiomic Features and Covariate Shift in Multi-Institutional Glioblastoma Datasets. *Phys Med Biol* (2019) 64:165011. doi: 10.1088/1361-6560/ab2f44
25. Robitaille N, Mouiha A, Crepeault B, Valdivia F, Duchesne S. The Alzheimer's Disease Neuroimaging I. Tissue-Based MRI Intensity Standardization: Application to Multicentric Datasets. *Int J BioMed Imaging* (2012) 2012:347120. doi: 10.1155/2012/347120
26. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): Explanation and Elaboration. *Ann Intern Med* (2015) 162:W1–73. doi: 10.7326/M14-0698
27. Ahmadi T, Itai Y, Takahashi M, Onaya H, Kobayashi T, Tanaka YO, et al. Angiomyolipoma of the Liver: Significance of CT and MR Dynamic Study. *Abdom Imaging* (1998) 23:520–6. doi: 10.1007/s002619900391
28. Cai PQ, Wu YP, Xie CM, Zhang WD, Han R, Wu PH. Hepatic Angiomyolipoma: CT and MR Imaging Findings With Clinical-Pathologic Comparison. *Abdom Imaging* (2013) 38:482–9. doi: 10.1007/s00261-012-9932-0
29. Feng Z, Rong P, Cao P, Zhou Q, Zhu W, Yan Z, et al. Machine Learning-Based Quantitative Texture Analysis of CT Images of Small Renal Masses: Differentiation of Angiomyolipoma Without Visible Fat From Renal Cell Carcinoma. *Eur Radiol* (2018) 28:1625–33. doi: 10.1007/s00330-017-5118-z
30. Hotker AM, Mazaheri Y, Wibmer A, Zheng J, Moskowitz CS, Tickoo SK, et al. Use of DWI in the Differentiation of Renal Cortical Tumors. *AJR Am J Roentgenol* (2016) 206:100–5. doi: 10.2214/AJR.14.13923
31. Miles KA, Ganeshan B, Hayball MP. CT Texture Analysis Using the Filtration-Histogram Method: What do the Measurements Mean? *Cancer Imaging: Off Publ Int Cancer Imaging Society* (2013) 13:400–06. doi: 10.1102/1470-7330.2013.9045
32. Nielsen B, Albregtsen F, Danielsen HE. Statistical Nuclear Texture Analysis in Cancer Research: A Review of Methods and Applications. *Crit Rev Oncog* (2008) 14:89–164. doi: 10.1615/critrevoncog.v14.i2-3.10
33. Mayerhoefer ME, Szomolanyi P, Jirak D, Materka A, Trattnig S. Effects of MRI Acquisition Parameter Variations and Protocol Heterogeneity on the Results of Texture Analysis and Pattern Discrimination: An Application-Oriented Study. *Med Physics* (2009) 36:1236–43. doi: 10.1118/1.3081408
34. AlObaidy M, Ramalho M, Busireddy KK, Liu B, Burke LM, Altun E, et al. High-Resolution 3D-GRE Imaging of the Abdomen Using Controlled Aliasing Acceleration Technique - a Feasibility Study. *Eur Radiol* (2015) 25:3596–605. doi: 10.1007/s00330-015-3780-6

Conflict of Interest: YY, LM, and XL were employed by the company Deepwise AI Lab, Deepwise Inc.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Zhao, Zhou, Zhang, Han, Mao, Yu, Li, Zeng, Wang and Liu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.tpdel

frontiers in Oncology

ORIGINAL RESEARCH
published: 14 October 2021
doi: 10.3389/fonc.2021.700204

Diagnostic Value of CT- and MRI-Based Texture Analysis and Imaging Findings for Grading Cartilaginous Tumors in Long Bones

Xue-Ying Deng[1,2†], *Hai-Yan Chen*[1,2†], *Jie-Ni Yu*[3], *Xiu-Liang Zhu*[3], *Jie-Yu Chen*[1,2], *Guo-Liang Shao*[1,2*] *and Ri-Sheng Yu*[3*]

[1] Department of Radiology, Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou, China, [2] Institue of Cancer and Basic Medicine (ICBM), Chinese Academy of Sciences, Hangzhou, China, [3] Department of Radiology, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln, United Kingdom

Reviewed by:
Bin-Yan Zhong,
The First Affiliated Hospital of Soochow University, China
Zhongqiu Wang,
Affiliated Hospital of Nanjing University of Chinese Medicine, China

***Correspondence:**
Guo-Liang Shao
shaogl@zjcc.org.cn
Ri-Sheng Yu
risheng-yu@zju.edu.cn

[†]These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 25 April 2021
Accepted: 28 September 2021
Published: 14 October 2021

Citation:
Deng X-Y, Chen H-Y, Yu J-N, Zhu X-L, Chen J-Y, Shao G-L and Yu R-S (2021) Diagnostic Value of CT- and MRI-Based Texture Analysis and Imaging Findings for Grading Cartilaginous Tumors in Long Bones. Front. Oncol. 11:700204. doi: 10.3389/fonc.2021.700204

Objective: To confirm the diagnostic performance of computed tomography (CT)-based texture analysis (CTTA) and magnetic resonance imaging (MRI)-based texture analysis for grading cartilaginous tumors in long bones and to compare these findings to radiological features.

Materials and Methods: Twenty-nine patients with enchondromas, 20 with low-grade chondrosarcomas and 16 with high-grade chondrosarcomas were included retrospectively. Clinical and radiological information and 9 histogram features extracted from CT, T1WI, and T2WI were evaluated. Binary logistic regression analysis was performed to determine predictive factors for grading cartilaginous tumors and to establish diagnostic models. Another 26 patients were included to validate each model. Receiver operating characteristic (ROC) curves were generated, and accuracy rate, sensitivity, specificity and positive/negative predictive values (PPV/NPV) were calculated.

Results: On imaging, endosteal scalloping, cortical destruction and calcification shape were predictive for grading cartilaginous tumors. For texture analysis, variance, mean, perc.01%, perc.10%, perc.99% and kurtosis were extracted after multivariate analysis. To differentiate benign cartilaginous tumors from low-grade chondrosarcomas, the imaging features model reached the highest accuracy rate (83.7%) and AUC (0.841), with a sensitivity of 75% and specificity of 93.1%. The CTTA feature model best distinguished low-grade and high-grade chondrosarcomas, with accuracies of 71.9%, and 80% in the training and validation groups, respectively; T1-TA and T2-TA could not distinguish them well. We found that the imaging feature model best differentiated benign and malignant cartilaginous tumors, with an accuracy rate of 89.2%, followed by the T1-TA feature model (80.4%).

Conclusions: The imaging feature model and CTTA- or MRI-based texture analysis have the potential to differentiate cartilaginous tumors in long bones by grade. MRI-based texture analysis failed to grade chondrosarcomas.

Keywords: cartilaginous tumors, texture analysis, chondrosarcoma, enchondroma, radiology

INTRODUCTION

Cartilaginous neoplasms are a heterogeneous group of bone tumors with abundant chondroid matrix and hyaline cartilage differentiation (1). The majority of cartilaginous neoplasms are enchondromas, the second most common benign bone tumors, with an incidence of 2.9% in the knee and 2.1% in the shoulder as detected by routine MR examinations (2). Chondrosarcomas are the second most frequent malignant bone tumors next to osteosarcoma, accounting for nearly 20–27% of bone sarcomas (3), and can be stratified into grades 1 to 3 and dedifferentiated chondrosarcoma based on their histopathological findings (1, 4).

Notably, 30% of chondrosarcomas are grade 1, which are low-grade neoplasms with low recurrence rates, locally aggressive behavior and limited metastatic probability (5, 6). In contrast, high-grade chondrosarcomas include grade 2, grade 3, and dedifferentiated chondrosarcomas, and these have higher recurrence rates, metastatic spread and poor survival outcomes, with a five-year survival rate of 53% (7). Given the indolent clinical course of enchondromas, active surveillance is supported to avoid unnecessary surgeries; meanwhile, surgical excision is of paramount importance for chondrosarcomas (8, 9). On imaging, typical chondrogenic tumors present with lobulated patterns, with hyperintensity on T2-weighted magnetic resonance imaging (MRI) and ring- and arc- or popcorn-like calcifications on computed tomography (CT) (10). However, enchondromas and low-grade chondrosarcomas can both demonstrate typical chondrogenic images. Therefore, the overlap of radiological features and histopathological criteria between benign enchondromas and low-grade chondrosarcomas as well as between low-grade and high-grade chondrosarcomas has led to difficulty in correctly and reliably differentiating and grading cartilaginous tumors; thus, a more accurate method is needed.

As a novel tool for objective quantitative assessment of the heterogeneity of lesions, texture analysis can extract, analyze, and interpret imaging features and has been widely used in differential diagnosis, grading, tumor staging and therapeutic response (11–13). Radiomics nomogram based on non-enhanced MRI showed hopeful performance in distinguishing enchondroma from chondrosarcomas, but their performance in differentiating high-grade chondrosarcomas from low-grade chondrosarcomas has not been sufficiently proven (14). CT has been less commonly used to grade cartilaginous tumors than MRI, and no study based on CT texture analysis has been report (15). However, MRI has been applied to analyze cartilaginous tumors but is less effective for visualizing calcification or bone destruction than CT (16, 17). Thus, we evaluated the radiological characteristics through both CT and MRI and extracted texture features. Therefore, this study aimed to assess and validate the diagnostic performance of CT-based texture analysis (CTTA) and MRI-based texture analysis for grading cartilaginous tumors in long bones and to compare these findings to radiological features.

Abbreviations: CTTA, Computed Tomography Texture Analysis; T1-TA, T1-weighted imaging Texture Analysis; T2-TA, T2-weighted imaging Texture; ROI, Region of Interest; ROC, Receiver Operating Characteristic; PPV, Positive predictive value; NPV, Negative predictive value; CI, Confidence Interval.

MATERIALS AND METHODS

This retrospective study was approved by our institutional review board, and the requirement for informed consent was waived.

Patients

Patients with long bone cartilaginous tumors confirmed by pathology at the Second Affiliated Hospital, Zhejiang University School of Medicine from January 1, 2009, to May 1, 2015 were retrospectively recorded as the training group. The inclusion criteria were as follows: 1) detailed pathological information, especially tumor grade; 2) CT and MRI examinations at a maximum time interval of 2 months before surgery or biopsy; and 3) enchondromas and chondrosarcomas were found in long bones, including the humerus, radius, ulna, femur, tibia, and fibula. The exclusion criteria were as follows: 1) no intergraded imaging data of MRI or CT for evaluation (n=6) and 2) an unclear pathology report (n=3). Finally, 65 patients with cartilaginous tumors were included in our study, including 29 patients with enchondromas, 20 with low-grade chondrosarcomas, and 16 with high-grade chondrosarcomas. Among these patients, all underwent plain CT scans, while 40 underwent contrast-enhanced MRI examinations, including 17 patients with enchondromas, 14 with low-grade chondrosarcomas and 9 with high-grade chondrosarcomas; the other patients had plain MRI scans only. The average time interval between CT/MRI and surgery or biopsy was 7.4/9.2 days, ranging from 0 to 52 days.

Imaging Acquisition

A variety of MRI machines (SIEMENS Aera 1.5T; SIEMENS Avanto 1.5T; GE Healthcare Signa HDxt 1.5T) was used due to our retrospective design, and different coils based on different lesion locations were used. However, standard pulse sequences including fat-suppressed T2-weighted imaging (T2WI; repetition time (TR), 2300–6600 ms; echo time (TE), 25–110 ms), axial T1-weighted imaging (T1WI; TR, 190–600 ms; TE, 1.2–12 ms), and fat-suppressed contrast-enhanced T1WI (TR, 420–630 ms; TE, 7–18 ms) were performed on each patient. The field of view (FOV) varied from 120 x 120 mm to 380 x 380 mm because of the different locations, the matrix size ranged from 288 x 224 to 320 x 224, and the slice thickness was 3 mm or 5 mm. Gadolinium-diethylenetriamine pentaacetic acid (Gd-DTPA, Omniscan, 0.1 mmol/kg body weight, 0.5 mmol/ml) was intravenously administered at a rate of 2 ml/s. Furthermore, longitudinal imaging (coronal or sagittal) was acquired for each patient.

Two kinds of CT scanners (Siemens Emotion 64 and TOSHIBA Aquilion 16) were used. Patients underwent plain CT scans, and the parameters were as follows: voltage, 120 kV; maximum tube current, 250 mAs; slice thickness, 3 mm; reconstructed slice thickness, 1 mm; slice collimation, 0.6 mm; and FOV, 380 X 380 or 120 x 120 mm.

Imaging Analysis

Two musculoskeletal radiologists (with 15 and 8 years of clinical experience in musculoskeletal radiology) reviewed the imaging

features without knowledge of the pathological outcomes. Disagreements were settled by consensus after discussion with a third radiologist with 32 years of experience in radiology.

Demographic information, including age, sex, and symptoms, was included. Imaging features were collected as follows: 1) the largest diameter, which was evaluated by measuring the maximum tumor extent in centimeters in axial scans; 2) the aspect ratio was defined as the length divided by width of the lesion, which accounted for the anisotropy of the shape of the tumor; 3) the location was defined as epiphysis, metaphysis, and diaphysis; 4) the calcified shape was defined as ring or plaque calcification and ground glass calcification in CT; 5) the calcified area was defined as the ratio of calcification to the tumor diameter at the maximum diameter plane and was stratified into<1/3, 1/3~2/3, and>2/3; 6) endosteal scalloping was defined as a local thinning of the osseous cortex by the nearby lesions, which formed a lobular outline (4); 7) periosteal reaction was defined as abnormal thickness or focal proliferation of the periosteal; 8) cortical destruction was defined as a sclerotic or lytic process destroying the continuity of the cortical bone; 9) blurring edge was characterized as an ill-defined tumor margin; 10) Fat replacement, which meant there was a fat-like signal surrounded by tumor tissue; 11) hemorrhage was defined as the presence of a bleeding signal in MRI, which was high signal intensity on T1WI, and low signal on T2WI; 12) peritumoral edema, which was defined as high signal intensity on fat-suppressed T2WI surrounding soft tissue without contrast enhancement; 13) soft tissue mass was defined as normal tissue that was replaced or displaced by a solid, extraosseous mass with contrast enhancement; and 14) the contrast enhanced pattern, which was evaluated on MRI and was divided into three types—type I was defined as a mostly continuous ring- and arc-enhancement of the lesion, type II was defined as a small patchy enhancement with or without ring- and arc-enhancement, and type III was defined as a large patchy, nodule enhancement or with obvious enhancement of the internal septa thickness.

Texture Analysis

MaZda software was used to extract features from CT, T1WI, and T2WI data. First, imaging intensity was normalized and standardized to rectify the effect of different imaging protocols (18). Nine histogram features were extracted in this study, including the mean, variance, skewness, kurtosis, perc.01%, perc.10%, perc.50%, perc.90%, and perc.99% (19). The region of interest (ROI) was outlined in the slice with the maximum diameter of the tumor without any artifacts.

Validation Group

The training and validation group was set at a ratio of 5:2. Thus, another 26 patients with long bone cartilaginous tumors confirmed by pathology at the same center from January 1, 2016 to August 1, 2021 were retrospectively recorded as the validation group. Only significant variables from the training group were evaluated to verify the performance of different models. As shown in **Figure 1**, our study consisted of five steps, including imaging and segmentation, feature extraction, feature selection, model construction and validation.

Statistical Analysis

According to the data distributions, continuous data are presented as the means ± standard deviation or medians (25–75%), while qualitative data are expressed as the frequencies (%). Student's t-tests or Mann-Whitney U tests were used for quantitative data, while Fisher's exact tests or $\chi 2$ tests were used for categorical data. Moreover, binary logistic regression

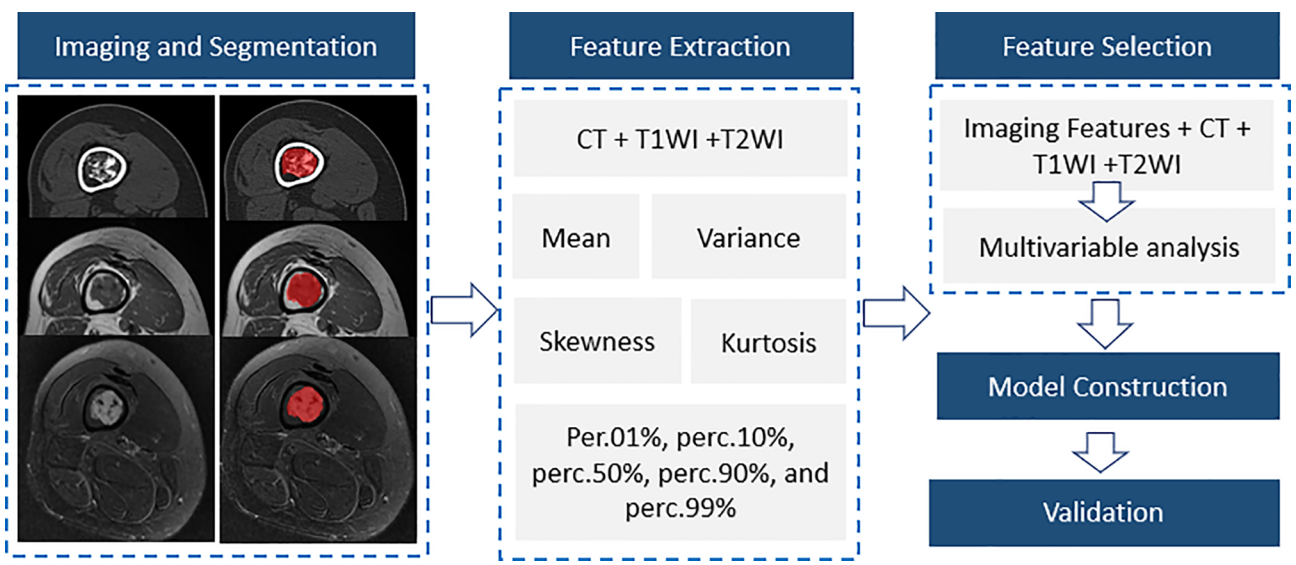


FIGURE 1 | A flow diagram of the whole study including imaging and segmentation, feature extraction, feature selection, model construction and validation.

analysis was performed for variables (p values less than 0.05 in the previous analysis) following the stepwise backwards procedure to differentiate cartilaginous tumors in different models. Then, receiver operating characteristic (ROC) curves were calculated to determine the diagnostic capacity of the model, and the accuracy rate, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and 95% confidence interval (CI) were calculated in both the training and validation groups. ROC analysis was performed using MedCalc software V.18 (Mariakerke, Belgium), while the other analyses used SPSS V.23.0 (IBM company, Chicago, Illinois, USA). ROC curves were drawn using GraphPad Prism version 7.04 (San Diego, California, USA). All tests were two-sided, and a p value less than 0.05 was considered to indicate a significant difference.

RESULTS

Comparison of the Imaging Features Among Groups of Cartilaginous Tumors

The clinical information and imaging features are summarized in **Table 1**, and a comparison of different subgroups of cartilaginous tumors is also presented. Cartilaginous tumors tended to occur in patients in their fifties and were located in the metaphysis of intramedullary long bones; enchondromas were more likely to occur in females, while chondrosarcomas occurred most often in males. Chondrosarcomas more frequently presented with symptoms and larger diameters than enchondromas ($P < 0.05$), while there was no difference between low-grade and high-grade chondrosarcomas ($P > 0.05$). Enchondromas and low-grade chondrosarcomas always included ring and arc or plaque chondroid matrix. The area of calcification decreased with increasing malignancy, while cortical destruction, periosteal reaction and blurry edges increased. Endosteal scalloping of the cortex was more frequent in chondrosarcomas, especially in low-grade chondrosarcomas. Fat replacement was always shown in enchondromas and low-grade chondrosarcomas ($P < 0.05$), while hemorrhage, peritumoral edema and soft tissue mass were only shown in chondrosarcomas. A soft tissue mass always presented with global hypodensity on CT or heterogeneous hyperintensity on T2WI. For the enhanced pattern, enchondromas all presented with continuous ring- and arc-enhancement of the lesion, while chondrosarcomas were more likely to manifest as a small patchy enhancement with or without ring-and arc- enhancement or a large patchy, nodule enhancement or with obvious enhancement of the internal septa thickness (**Figures 2**–**4**).

Different Models to Discriminate the Differentiation of Cartilaginous Tumors

Predictor models based on imaging features, CTTA features, T1WI texture analysis (T1-TA) features and T2WI texture analysis (T2-TA) features are shown in **Table 2**. Regarding imaging features, endosteal scalloping was an independent predictor to differentiate enchondromas from low-grade chondrosarcomas, while cortical destruction could differentiate low-grade chondrosarcomas from high-grade chondrosarcomas. To distinguish benign cartilaginous tumors from malignant cartilaginous tumors, we found that sex, calcified shape and endosteal scalloping might be useful. In terms of texture analysis, we found that the CTTA feature model could differentiate cartilaginous tumors well, while the T1-TA and T2-TA feature models could not differentiate low-grade from high-grade chondrosarcomas. For texture analysis features, variance, mean, perc.01%, perc.10%, perc.99% and kurtosis were extracted through multivariate regression analysis.

Diagnostic Performance of Each Model

The diagnostic performance of the imaging, CTTA, T1-TA, and T2-TA feature models between the training and validation groups are listed in **Table 3**, and the ROC curves are presented in **Figure 5**. For differentiating benign cartilaginous tumors from low-grade chondrosarcomas, we found that the imaging features model reached the highest accuracy rate (83.7%) and AUC (0.841), with a sensitivity value of 75%, a specificity value of 93.1%, a PPV of 88.2% and NPV of 84.4%, followed by the CTTA and T1-TA feature models (accuracy rate: 80%). The validation group also achieved good performance, with an AUC of 0.784. The CTTA feature model was the best method to distinguish low-grade chondrosarcomas from high-grade chondrosarcomas, with an accuracy rate of 71.9% and 80% in the training and validation groups respectively; T1-TA and T2-TA could not distinguish them well. In terms of differentiating benign cartilaginous tumors from malignant cartilaginous tumors, the imaging feature model performed the best, with an accuracy rate of 89.2% and AUC of 0.896, followed by the T1-TA feature model, with an accuracy rate of 80.4% and AUC of 0.804.

DISCUSSION

Accurate grading of cartilaginous tumors is of paramount importance in the management of these lesions, ranging from follow-up for enchondromas to curettage for low-grade chondrosarcomas and amputation or extended resection for high-grade chondrosarcomas (20). In this study, we evaluated the diagnostic performance of CTTA and MRI-based texture analysis for grading cartilaginous tumors in long bones and compared these findings to radiological features both in the training and validation groups. After multivariate analysis, we found that the imaging features model reached the highest accuracy rate both in differentiating enchondromas from low-grade chondrosarcomas and from malignant cartilaginous tumors. Furthermore, the CTTA feature model was the best method to distinguish low-grade chondrosarcomas from high-grade chondrosarcomas, while MRI-based texture analysis could not effectively distinguish low-grade chondrosarcomas from high-grade chondrosarcomas.

In our study, radiological features were evaluated both by MRI and CT, which included calcification information when compared to other studies (17, 21, 22). Imaging features differ when the grade

TABLE 1 | Clinical information and univariate analysis between imaging features and tumor grades.

Variables	Enchondromas (n=29)	Low-grade chondrosarcomas (n=20)	High-grade chondrosarcomas (n=16)	Malignant (n=36)	Benign *vs.* Low grade (P)	Low grade *vs.* High grade (P)	Benign *vs.* Malignant (P)
Age	49 (40–59)	51.5 (39–62)	55.5 (45–62)	54 (44.3–60)	0.662	0.440	0.251
Gender					0.110	0.709	**0.04**
Male	8 (27.6)	10 (50)	9 (56.3)	19 (52.8)			
Female	21 (72.4)	10 (50)	7 (43.8)	17 (47.2)			
Symptomatic	17 (58.6)	19 (95)	12 (75)	31 (86.1)	**0.007**	0.149	**0.012**
Largest diameter/mm	4.1 (2.5–6.5)	6.6 (3.4–18)	7.1 (5.9–10.8)	6.7 (4.5–12.9)	**0.020**	0.440	**<0.001**
Aspect ratio	1.8 (1.5–2.8)	2.5 (1.2–4)	2 (1.2–3.6)	2.2 (1.3–3.5)	0.495	0.962	0.601
Location					0.054	0.103	0.138
Epiphysis	1 (3.4)	4 (20)	3 (18.8)	7 (19.4)			
Metaphysis	21 (72.4)	15 (75)	8 (50)	23 (63.9)			
Diaphysis	7 (24.1)	1 (5)	5 (31.3)	6 (16.7)			
Calcified shape					0.295	0.144	**0.033**
Non calcification	0 (0)	1 (5)	1 (6.3)	2 (5.6)			
Ring and plaque	27 (93.1)	16 (80)	8 (50)	24 (66.7)			
Ground glass	2 (6.9)	3 (15)	7 (43.8)	10 (27.8)			
Calcified area					**0.011**	0.402	**<0.001**
<1/3	6 (20.7)	11 (55)	12 (75)	23 (63.9)			
1/3~2/3	7 (24.1)	6 (30)	2 (12.5)	8 (22.2)			
>2/3	16 (55.2)	3 (15)	2 (12.5)	5 (13.9)			
Endosteal scalloping	2 (6.9)	15 (75)	10 (62.5)	25 (69.4)	**<0.001**	0.418	**<0.001**
Periosteal reaction	0 (0)	8 (40)	11 (68.8)	19 (52.8)	**<0.001**	0.086	**<0.001**
Cortical destruction	0 (0)	9 (45)	13 (81.3)	22 (61.1)	**<0.001**	**0.029**	**<0.001**
Blurry edge	5 (17.2)	10 (50)	11 (68.8)	21 (58.3)	**0.014**	0.257	**0.001**
Fat replacement	20 (69)	8 (40)	0 (0)	8 (22.2)	**0.044**	**0.005**	**<0.001**
Hemorrhage	0 (0)	4 (40)	5 (31.3)	9 (25)	**0.023**	0.470	**0.004**
Peritumoral edema	0 (0)	15 (75)	13 (81.3)	28 (77.8)	**<0.001**	0.709	**<0.001**
Soft tissue mass	0 (0)	6 (30)	10 (62.5)	16 (44.4)	**0.003**	0.051	**<0.001**
Enhanced pattern (n=40)					**<0.001**	0.146	**<0.001**
I type	17 (58.6)	1 (5)	0 (0)	1 (2.8)			
II type	0 (0)	12 (60)	5 (31.3)	17 (47.2)			
III type	0 (0)	1 (5)	4 (25)	5 (13.9)			

The bold values means P < 0.05.

of a cartilaginous tumor changes. We found that sex, calcified shape and endosteal scalloping were independent predictors of the differentiation of benign and malignant cartilaginous tumors. Enchondromas always present with ring and plaque calcification, which indicates mature cartilage matrix differentiation. Douis et al. (23) found that pain related to lesions, tumor length, endosteal scalloping, bone expansion, cortical destruction and soft tissue mass could distinguish enchondromas from low-grade chondrosarcomas, while dynamic contrast-enhanced MRI was not useful in differentiating enchondromas from low-grade chondrosarcomas. Our study also showed that invasive features, such as cortical destruction, blurry edges, periosteal reaction, peritumoral edema and soft tissue mass were more likely to be present in chondrosarcomas, especially in high-grade chondrosarcomas. However, in contrast-enhanced MRI, we found that continuous ring-and-arc enhancement was always shown in enchondromas, while two other types of enhanced patterns were more frequently present in chondrosarcomas. In addition to dynamic contrast-enhanced MRI, diffusion-weighted imaging is another functional sequence used in imaging diagnosis; however, it does not aid in the distinction or grading of cartilaginous tumors (24). Coninck et al. (16) demonstrated that dynamic contrast-enhanced MRI played an important role in distinguishing enchondromas and low-grade chondrosarcomas, with 93.4% accuracy in predicting the diagnosis of chondrosarcomas. When differentiating low-grade and high-grade chondrosarcomas, Sharif et al. (25) found that soft tissue mass, periosteal reaction and bone edema suggested high-grade

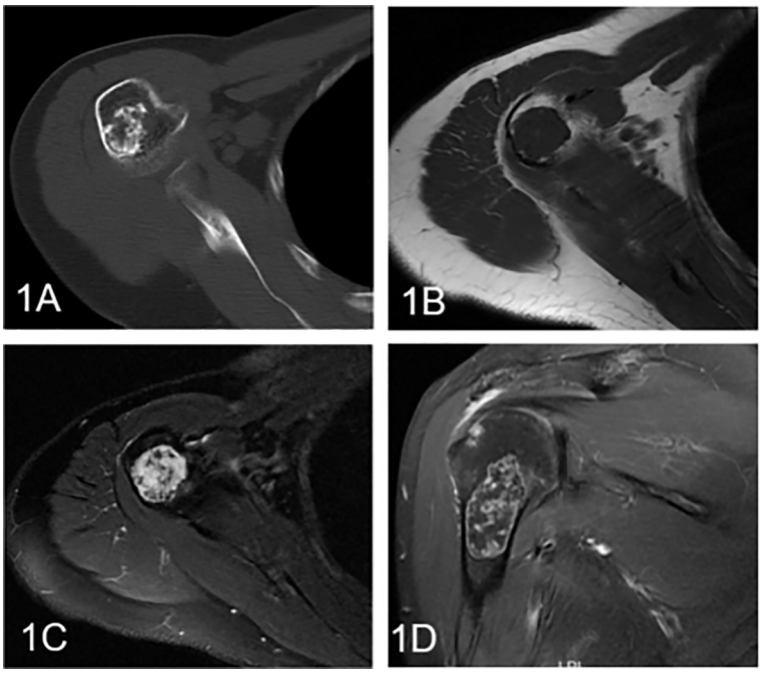


FIGURE 2 | Enchondroma involving the humerus in a 58-year-old female. **(A)** Axial CT shows a well-defined intramedullary osteogenic lesion in the metaphysis of the right humerus, which presents with characteristic chondroid (ring and arc) calcifications. **(B)** Axial T1WI presents a mass with global heterogeneity and a well-defined margin. **(C)** Axial fat-suppressed T2WI shows a heterogeneous hyperintense mass with low signal foci. **(D)** Coronal fat-suppressed contrast-enhanced T1WI demonstrates marginal and septal or ring-and-arc enhancement within the lesion.

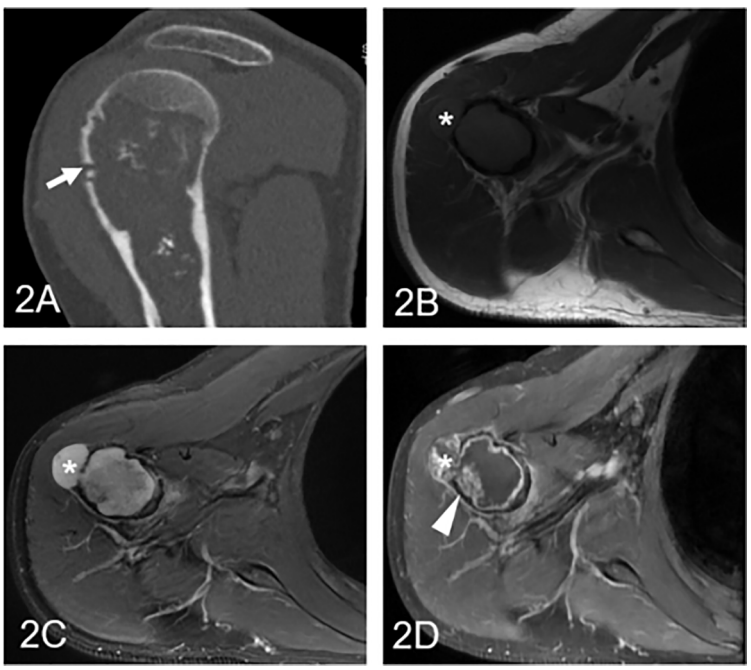


FIGURE 3 | Low-grade chondrosarcoma involving the humerus in a 57-year-old man. **(A)** Coronal CT shows an ill-defined intramedullary osteolytic lesion with several ring and popcorn calcified components, and the arrowhead shows cortical disruption (arrow). **(B)** Axial T1WI presents a homogeneous lesion with an associated soft tissue mass (asterisk). **(C)** Axial fat-suppressed T2WI demonstrates a heterogeneous hyperintense lesion with a soft tissue component (asterisk) and endosteal scalloping of the cortex. **(D)** Coronal fat-suppressed contrast-enhanced T1WI demonstrates patched (arrowhead) and marginal enhancement.

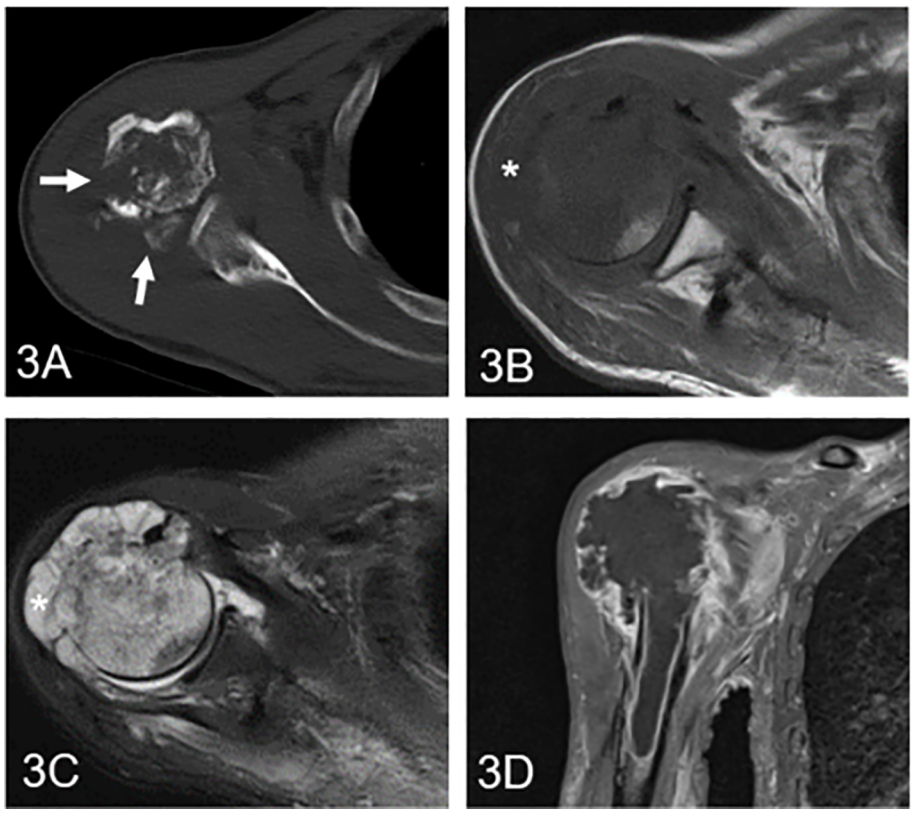


FIGURE 4 | High-grade chondrosarcoma in the proximal humerus in a 61-year-old man. **(A)** Axial CT shows an ill-defined osteolytic lesion with multiple cortical destruction (arrow). **(B)** Axial T1WI presents an irregular soft tissue mass with an aggressive growth pattern (asterisk). **(C)** Axial fat-suppressed T2WI demonstrates a heterogeneous hyperintense mass with low signal components and soft tissue. **(D)** Coronal fat-suppressed contrast-enhanced T1WI shows diffuse and marginal enhancement.

chondrosarcomas, which was consistent with our results. They also found that soft tissue mass and cortical destruction were predictors of high-grade chondrosarcomas in long bones. Soft tissue mass can be found in 28% of chondrosarcomas and can replace marrow fat, indicating a pathognomonic sign of malignancy (21).

The most compelling result is that CTTA could distinguish low-grade from high-grade chondrosarcomas, while MRI-based texture analysis was not able to show the correlation. This is especially beyond the expectation that CT performs better than MRI, which may be because thin-slice CT imaging demonstrates

TABLE 2 | Multivariate analysis among different models.

Model	Compared Tumor Grades	Independent Predictors	Odds Ratio (95%CI)	P
Imaging feature model	Benign *vs.* low-grade	Endosteal scalloping	0.042 (0.004, 0.439)	0.008
	Low-grade *vs.* high-grade	Cortical destruction	5.296 (1.143, 24.548)	0.033
	Benign vs malignant	Gender	0.041 (0.002, 0.813)	0.036
		Calcified shape	205.140 (1.55, 27147.14)	0.033
		Endosteal scalloping	0.007 (0.001, 0.186)	0.003
CTTA feature model	Benign *vs.* low-grade	Variance	1.001 (1.001, 1.002)	0.002
	Low-grade *vs.* high-grade	Mean	1.019 (1.004, 1.034)	0.013
	Benign *vs.* malignant	Variance	1.001 (1.000, 1.001)	0.04
		Perc.01%	0.974 (0.953, 0.995)	0.014
T1-TA feature model	Benign *vs.* low-grade	Kurtosis	1.645 (1.1, 2.459)	0.015
		Perc.10%	0.861 (0.759, 0.997)	0.019
	Benign *vs.* malignant	Kurtosis	1.205 (1.032, 1.407)	0.019
		Perc.10%	0.84 (0.759, 0.930)	0.001
		Perc.99%	1.09 (1.032, 1.152)	0.002
T2-TA feature model	Benign *vs.* low-grade	Variance	1.002 (1.001, 1.003)	0.007
	Benign *vs.* malignant	Variance	1.002 (1.001, 1.003)	0.003

TABLE 3 | Analysis efficiency in different models in the training group and validation group.

Model	Group	Compared Tumor Grades	Accuracy rate	Sensitivity	Specificity	PPV	NPV	AUC (95% CI)
Imaging feature model	Training	Benign *vs*. low-grade	83.7	75	93.1	88.2	84.4	0.841 (0.708, 0.929)
		Low-grade *vs*. high-grade	66.7	81.3	55	59.1	78.6	0.681 (0.505, 0.826)
		Benign *vs*. malignant	89.2	72.2	93.1	92.9	73	0.896 (0.795, 0.958)
	Validation	Benign *vs*. low-grade	78.9	75	81.8	75	81.8	0.784 (0.538, 0.936)
		Low-grade *vs*. high-grade	73.3	85.7	62.5	66.7	83.3	0.741 (0.457, 0.926)
		Benign *vs*. malignant	76.9	86.7	63.6	76.5	77.8	0.788 (0.584, 0.922)
CTTA feature model	Training	Benign *vs*. low-grade	80	81.3	83.3	76.5	87	0.823 (0.670, 0.925)
		Low-grade *vs*. high-grade	71.9	81.3	75	76.5	80	0.785 (0.600, 0.907)
		Benign *vs*. malignant	71.4	79.2	87.5	84.8	82.6	0.854 (0.734, 0.934)
	Validation	Benign *vs*. low-grade	78.9	87.5	72.7	70	88.9	0.801 (0.557, 0.946)
		Low-grade *vs*. high-grade	80	57.1	100	100	72.7	0.786 (0.504, 0.950)
		Benign *vs*. malignant	84.6	100	63.6	78.9	100	0.906 (0.726, 0.985)
T1-TA feature model	Training	Benign *vs*. low-grade	80	100	55.6	73.3	100	0.838 (0.688, 0.935)
		Benign *vs*. malignant	80.4	79.4	68.2	79.4	68.2	0.804 (0.676, 0.898)
	Validation	Benign *vs*. low-grade	84.2	100	72.7	72.7	100	0.949 (0.742, 0.999)
		Benign *vs*. malignant	84.6	73.3	100	100	73.3	0.945 (0.780, 0.996)
T2-TA feature model	Training	Benign *vs*. low-grade	75.6	77.8	73.9	70	81	0.758 (0.600, 0.878)
		Benign *vs*. malignant	79.3	82.9	73.9	82.9	73.9	0.796 (0.670, 0.891)
	Validation	Benign *vs*. low-grade	78.9	75	81.8	75	81.8	0.784 (0.538, 0.936)
		Benign *vs*. malignant	84.2	87.5	81.8	77.8	90	0.847 (0.610, 0.968)

PPV, positive predictive value; NPV, negative predictive value.

calcification well within the lesion. Most of the low-grade chondrosarcomas presented with ring and plaque calcification, while nearly half of the high-grade chondrosarcomas showed ground glass calcification due to an immature cartilage matrix. Lisson et al. (3) included 11 patients with enchondromas and 11 patients with low-grade chondrosarcomas, and only four texture analyses were obtained for each lesion. The authors found that kurtosis in contrast-enhanced T1WI had the greatest power of discrimination, with 86% accuracy. In our study, we found that kurtosis extracted from non-enhanced T1WI was an independent predictor to differentiate benign cartilaginous tumors from chondrosarcomas. Yin et al. (26) used a clinical radiomics nomogram combined with clinical characteristics to successfully predict the early recurrence of pelvic chondrosarcomas and found that a radiomics model based on T1WI outperformed than T2WI or contrast-enhanced radiomics to predict recurrence.

Many texture analysis features based on CT or MRI have been published and suggested depending on histogram analysis or other matrixes (27, 28). We evaluated 9 histogram features and found that variance, mean, perc.01%, perc.10%, perc.99% and kurtosis had

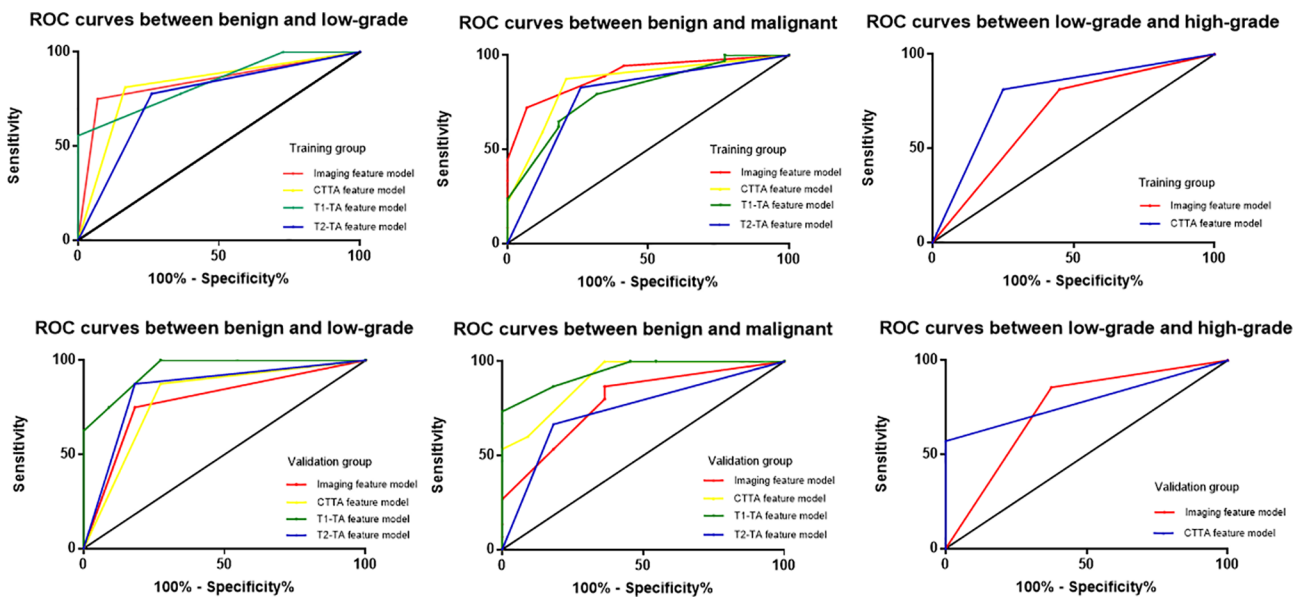


FIGURE 5 | ROC curves among different models to distinguish cartilaginous tumors in both the training group and validation group.

good discriminatory power to grade cartilaginous tumors in different feature models. Hu et al. (29) extracted 22 histogram parameters and integrated them with CT morphological features, which can be used to predict lung metastasis in patients with colorectal cancer. Histogram features may be related to the tumor microenvironment. For example, variance measures the deviation of gray levels from the mean and represents the extent of the histogram, which may reflect on morphologic imaging performance (4). CT and MRI for imaging features should always be initially considered; if there is a need for differentiation of malignant tumor grading, CTTA should be performed.

Our study has several limitations. First, due to our retrospective study design, there may have been inherent selection and unavoidable biases. Second, several MRI and CT machines and protocols have been used, and although we had standardized images before analysis, possible bias may still have been introduced. Third, the number of patients was relatively small, and the number of extracted features was also small; thus, further studies with larger samples or numerous features are needed. Fourth, we did not combine imaging feature models with a texture analysis feature model because both CT and MRI imaging features were incorporated into the imaging feature model; thus, it was impossible to combine an imaging feature model with CT or MRI alone. Finally, we did not use dynamic contrast-enhanced MRI images; on the one hand, not all patients had undergone enhanced MRI, and on the other hand, a recent study found that contrast-enhanced MRI may not distinguish enchondromas from low-grade chondrosarcomas (23).

In conclusion, the imaging feature model and CTTA- or MRI-based texture analysis have the potential to differentiate the grade of cartilaginous tumors in long bones. MRI-based texture analysis failed to grade chondrosarcomas.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding authors.

AUTHOR CONTRIBUTIONS

X-YD and H-YC completed manuscript together. J-NY collected validation group data. J-YC and X-LZ processed the data and the statistics. G-LS and R-SY gave the support of everything we need. All authors did literature researches. All authors contributed to the article and approved the submitted version.

FUNDING

This work was reported by Zhejiang Province Medical, Science and Technology Project, the number is 2021KY091.

REFERENCES

1. Jo VY, Fletcher CD. WHO Classification of Soft Tissue Tumours: An Update Based on the 2013 (4th) Edition. *Pathology* (2014) 46(2):95–104. doi: 10.1097/PAT.0000000000000050
2. Stomp W, Reijnierse M, Kloppenburg M, de Mutsert R, Bovee JV, den Heijer M, et al. Prevalence of Cartilaginous Tumours as an Incidental Finding on MRI of the Knee. *Eur Radiol* (2015) 25(12):3480–7. doi: 10.1007/s00330-015-3764-6
3. Lisson CS, Lisson CG, Flosdorf K, Mayer-Steinacker R, Schultheiss M, von Baer A, et al. Diagnostic Value of MRI-Based 3D Texture Analysis for Tissue Characterisation and Discrimination of Low-Grade Chondrosarcoma From Enchondroma: A Pilot Study. *Eur Radiol* (2018) 28(2):468–77. doi: 10.1007/s00330-017-5014-6
4. Fritz B, Muller DA, Sutter R, Wurnig MC, Wagner MW, Pfirrmann CWA, et al. Magnetic Resonance Imaging-Based Grading of Cartilaginous Bone Tumors: Added Value of Quantitative Texture Analysis. *Invest Radiol* (2018) 53(11):663–72. doi: 10.1097/RLI.0000000000000486
5. Crim J, Schmidt R, Layfield L, Hanrahan C, Manaster BJ. Can Imaging Criteria Distinguish Enchondroma From Grade 1 Chondrosarcoma? *Eur J Radiol* (2015) 84(11):2222–30. doi: 10.1016/j.ejrad.2015.06.033
6. Weinschenk RC, Wang WL, Lewis VO. Chondrosarcoma. *J Am Acad Orthopaedic Surgeons* (2021) 29(13):553–62. doi: 10.5435/jaaos-d-20-01188
7. Houdek MT, Witten BG, Hevesi M, Griffin AM, Salduz A, Wenger DE, et al. Advancing Patient Age Is Associated With Worse Outcomes in Low- and Intermediate-Grade Primary Chondrosarcoma of the Pelvis. *J Surg Oncol* (2020) 121(4):638–44. doi: 10.1002/jso.25854
8. Ahmed S, Jubouri S, Mulligan M. Incidental Long Bone Cartilage Lesions: Is Any Further Imaging Workup Needed? *Skeletal Radiol* (2021) 50(6):1189–96. doi: 10.1007/s00256-020-03664-w
9. Deckers C, Rooy JWJ, Flucke U, Schreuder HWB, Dierselhuis EF, Geest I. Midterm MRI Follow-Up of Untreated Enchondroma and Atypical Cartilaginous Tumors in the Long Bones. *Cancers* (2021) 13(16):4093. doi: 10.3390/cancers13164093
10. Engel H, Herget GW, Füllgraf H, Sutter R, Benndorf M, Bamberg F, et al. Chondrogenic Bone Tumors: The Importance of Imaging Characteristics. *RoFo Fortschr auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin* (2021) 193(3):262–75. doi: 10.1055/a-1288-1209
11. Guo CG, Ren S, Chen X, Wang QD, Xiao WB, Zhang JF, et al. Pancreatic Neuroendocrine Tumor: Prediction of the Tumor Grade Using Magnetic Resonance Imaging Findings and Texture Analysis With 3-T Magnetic Resonance. *Cancer Manage Res* (2019) 11:1933–44. doi: 10.2147/CMAR.S195376
12. Sandrasegaran K, Lin Y, Asare-Sawiri M, Taiyini T, Tann M. CT Texture Analysis of Pancreatic Cancer. *Eur Radiol* (2019) 29(3):1067–73. doi: 10.1007/s00330-018-5662-1
13. Kim J-H, Ko ES, Lim Y, Lee KS, Han B-K, Ko EY, et al. Breast Cancer Heterogeneity: MR Imaging Texture Analysis and Survival Outcomes. *Radiology* (2017) 282(3):665–75. doi: 10.1148/radiol.2016160261
14. Pan J, Zhang K, Le H, Jiang Y, Li W, Geng Y, et al. Radiomics Nomograms Based on Non-Enhanced MRI and Clinical Risk Factors for the Differentiation of Chondrosarcoma From Enchondroma. *J Magnetic Resonance Imaging JMRI* (2021) 10.1002/jmri.27690:1314–23. doi: 10.1002/jmri.27690
15. Miwa S, Yamamoto N, Hayashi K, Takeuchi A, Igarashi K, Tada K, et al. A Radiological Scoring System for Differentiation Between Enchondroma and Chondrosarcoma. *Cancers* (2021) 13(14):3558. doi: 10.3390/cancers13143558
16. De Coninck T, Jans L, Sys G, Huysse W, Verstraeten T, Forsyth R, et al. Dynamic Contrast-Enhanced MR Imaging for Differentiation Between Enchondroma and Chondrosarcoma. *Eur Radiol* (2013) 23(11):3140–52. doi: 10.1007/s00330-013-2913-z
17. Douis H, Singh L, Saifuddin A. MRI Differentiation of Low-Grade From High-Grade Appendicular Chondrosarcoma. *Eur Radiol* (2014) 24(1):232–40. doi: 10.1007/s00330-013-3003-y

18. Dong F, Li Q, Xu D, Xiu W, Zeng Q, Zhu X, et al. Differentiation Between Pilocytic Astrocytoma and Glioblastoma: A Decision Tree Model Using Contrast-Enhanced Magnetic Resonance Imaging-Derived Quantitative Radiomic Features. *Eur Radiol* (2019) 29(8):3968–75. doi: 10.1007/s00330-018-5706-6
19. Szczypiński PM, Strzelecki M, Materka A, Klepaczko A. MaZda—A Software Package for Image Texture Analysis. *Comput Methods Programs Biomed* (2009) 94(1):66–76. doi: 10.1016/j.cmpb.2008.08.005
20. de Camargo OP, Baptista AM, Atanasio MJ, Waisberg DR. Chondrosarcoma of Bone: Lessons From 46 Operated Cases in a Single Institution. *Clin Orthopaedics Related Res* (2010) 468(11):2969–75. doi: 10.1007/s11999-010-1368-7
21. Choi BB, Jee WH, Sunwoo HJ, Cho JH, Kim JY, Chun KA, et al. MR Differentiation of Low-Grade Chondrosarcoma From Enchondroma. *Clin Imaging* (2013) 37(3):542–7. doi: 10.1016/j.clinimag.2012.08.006
22. Jurik AG, Jorgensen PH, Mortensen MM. Whole-Body MRI in Assessing Malignant Transformation in Multiple Hereditary Exostoses and Enchondromatosis: Audit Results and Literature Review. *Skeletal Radiol* (2020) 49(1):115–24. doi: 10.1007/s00256-019-03268-z
23. Douis H, Parry M, Vaiyapuri S, Davies AM. What are the Differentiating Clinical and MRI-Features of Enchondromas From Low-Grade Chondrosarcomas? *Eur Radiol* (2018) 28(1):398–409. doi: 10.1007/s00330-017-4947-0
24. Douis H, Jeys L, Grimer R, Vaiyapuri S, Davies AM. Is There a Role for Diffusion-Weighted MRI (DWI) in the Diagnosis of Central Cartilage Tumors? *Skeletal Radiol* (2015) 44(7):963–9. doi: 10.1007/s00256-015-2123-7
25. Sharif B, Rajakulasingam R, Sharifi S, O'Donnell P, Saifuddin A. MRI Features of Low-Grade and High-Grade Chondrosarcoma in Enchondromatosis. *Skeletal Radiol* (2021) 50(8):1637–46. doi: 10.1007/s00256-021-03718-7
26. Yin P, Mao N, Liu X, Sun C, Wang S, Chen L, et al. Can Clinical Radiomics Nomogram Based on 3D Multiparametric MRI Features and Clinical Characteristics Estimate Early Recurrence of Pelvic Chondrosarcoma? *J Magnetic Resonance Imaging JMRI* (2020) 51(2):435–45. doi: 10.1002/jmri.26834
27. De Robertis R, Maris B, Cardobi N, Tinazzi Martini P, Gobbo S, Capelli P, et al. Can Histogram Analysis of MR Images Predict Aggressiveness in Pancreatic Neuroendocrine Tumors? *Eur Radiol* (2018) 28(6):2582–91. doi: 10.1007/s00330-017-5236-7
28. Leithner D, Horvat JV, Marino MA, Bernard-Davila B, Jochelson MS, Ochoa-Albiztegui RE, et al. Radiomic Signatures With Contrast-Enhanced Magnetic Resonance Imaging for the Assessment of Breast Cancer Receptor Status and Molecular Subtypes: Initial Results. *Breast Cancer Res BCR* (2019) 21(1):106. doi: 10.1186/s13058-019-1187-z
29. Hu T, Wang S, E X, Yuan Y, Huang L, Wang J, et al. CT Morphological Features Integrated With Whole-Lesion Histogram Parameters to Predict Lung Metastasis for Colorectal Cancer Patients With Pulmonary Nodules. *Front Oncol* (2019) 9:1241. doi: 10.3389/fonc.2019.01241

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Deng, Chen, Yu, Zhu, Chen, Shao and Yu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

METHODS
published: 14 October 2021
doi: 10.3389/fonc.2021.738909

Exploring the Value of Radiomics Features Based on B-Mode and Contrast-Enhanced Ultrasound in Discriminating the Nature of Thyroid Nodules

Shi Yan Guo, Ping Zhou, Yan Zhang, Li Qing Jiang and Yong Feng Zhao*

The Department of Ultrasound, Xiangya Third Hospital, Central South University, Changsha, China

Background: With the improvement of ultrasound imaging resolution and the application of various new technologies, the detection rate of thyroid nodules has increased greatly in recent years. However, there are still challenges in accurately diagnosing the nature of thyroid nodules. This study aimed to evaluate the clinical application value of the radiomics features extracted from B-mode ultrasound (B-US) images combined with contrast-enhanced ultrasound (CEUS) images in the differentiation of benign and malignant thyroid nodules by comparing the diagnostic performance of four logistic models.

Methods: We retrospectively collected and ultimately included B-US images and CEUS images of 123 nodules from 123 patients, and then extracted the corresponding radiomics features from these images respectively. Meanwhile, a senior radiologist combined the thyroid imaging reporting and data system (TI-RADS) and the enhancement pattern of the ultrasonography to make a graded diagnosis of the malignancy of these nodules. Next, based on these radiomics features and grades, logistic regression was used to help build the models (B-US radiomics model, CEUS radiomics model, B-US+CEUS radiomics model, and TI-RADS+CEUS model). Finally, the study assessed the diagnostic performance of these radiomics features with a comparison of the area under the curve (AUC) of the receiver operating characteristic curve of four logistic models for predicting the benignity or malignancy of thyroid nodules.

Results: The AUC in the differential diagnosis of the nature of thyroid nodules was 0.791 for the B-US radiomics model, 0.766 for the CEUS radiomics model, 0.861 for the B-US+CEUS radiomics model, and 0.785 for the TI-RADS+CEUS model. Compared to the TI-RADS+CEUS model, there was no statistical significance observed in AUC between the B-US radiomics model, CEUS radiomics model, B-US+CEUS radiomics model, and TI-RADS+CEUS model ($P>0.05$). However, a significant difference was observed between the single B-US radiomics model or CEUS radiomics model and B-US+CEUS radiomics model ($P<0.05$).

Conclusion: In our study, the B-US radiomics model, CEUS radiomics model, and B-US+CEUS radiomics model demonstrated similar performance with the TI-RADS+CEUS

OPEN ACCESS

Edited by:
Joseph Ross Mitchell, Independent researcher, Tampa, FL, United States

Reviewed by:
Sikandar Shaikh, Shadan Hospital and Institute of Medical Sciences, India
Deb Kumar Boruah, Assam Medical College, India

****Correspondence:***
Ping Zhou
zhouping1000@hotmail.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *09 July 2021*
Accepted: *24 September 2021*
Published: *14 October 2021*

Citation:
Guo SY, Zhou P, Zhang Y, Jiang LQ and Zhao YF (2021) Exploring the Value of Radiomics Features Based on B-Mode and Contrast-Enhanced Ultrasound in Discriminating the Nature of Thyroid Nodules. Front. Oncol. 11:738909. doi: 10.3389/fonc.2021.738909

model of senior radiologists in diagnosing the benignity or malignancy of thyroid nodules, while the B-US+CEUS radiomics model showed better diagnostic performance than single B-US radiomics model or CEUS radiomics model. It was proved that B-US radiomics features and CEUS radiomics features are of high clinical value as the combination of the two had better diagnostic performance.

Keywords: thyroid nodules, contrast-enhanced ultrasound, diagnosis performance, radiomics features, clinical value

INTRODUCTION

In recent years, thyroid nodules can be observed in up to 50 - 60% of healthy recipients, but only about 5% are proven to be malignant. As a result, when encountering patients with thyroid nodules today, clinicians are faced with the task of avoiding the overdiagnosis of low-risk cancers without jeopardizing the chances of identifying those rare late-stage or higher-risk tumors that will require timely and appropriate treatment (1, 2). Ultrasound as the principle means for the detection and risk stratification of thyroid nodules provides guidance for their biopsy and nonsurgical treatment (3). However, misdiagnosis and unnecessary operations are still prevalent for certain nodules because of interobserver and intraobserver variability, the overlap in morphological features of certain benign and malignant nodules, and different diagnostic criteria in different regions (4, 5). Fine-needle aspiration biopsy (FNAB) is currently the most reliable and cost-effective examination method for assessing thyroid nodules (6), which can effectively screen out benign thyroid nodules as well as reduce the rate of unnecessary surgeries (2). However, there are still 30-40% of FNAB procedures that generate nondiagnostic results due to various factors, such as the features of the nodules, and the experience of the physician and pathologist (7–9). In addition, it is not widely accepted as an invasive examination. Hence, researchers have been pursuing a precise, uniform, and non-invasive diagnostic method. The notion of radiomics was introduced by Lambin (10) in 2012. The visual image information can be transformed into deep-seated features for quantitative research. It gives us hope to achieve non-invasive precision medicine, which is widely used to evaluate tumors in all parts of the body. Based on a radiomics analysis method, we extracted the corresponding radiomics features from thyroid nodule ultrasound images to build logistic models, and evaluated the clinical application value of these features in differentiating benign and malignant thyroid nodules by comparing the diagnostic performance of the radiomics models with the model based on the grading of nodule malignancy by a senior radiologist.

STUDY SUBJECTS AND METHODS

Study Subjects

This was a retrospective study; the subjects of our study were two-dimensional gray-scale images and contrast-enhanced ultrasound images of patients with thyroid nodules. All of the participants signed an informed consent form prior to the examination, with the approval of the hospital ethics committee. Inclusion criteria: Patients who had received a CEUS examination with complete imaging data in our hospital between September 2020 and August 2021, and had underwent thyroid fine-needle aspiration biopsy or pathological diagnoses within one week after the examination. Exclusion criteria: Patient nodules with an unclear pathological finding, and the proportion of cystic composition over 25% (11), in addition, we also discarded images with poor image quality.

Image Collection

The Siemens X150 Vivi7 color doppler ultrasound diagnostic system was used to perform diagnoses, which was equipped with a 10L4 transducer for B-US and CEUS examination. Each patient lay on the inspection bed in the supine position and was told to fully extend their neck and breathe calmly. By carefully scanning every section of the thyroid nodule, we observed and recorded internal composition, internal echo, the boundary, the margin, aspect ratio, and calcification. Based on a previous study (12), solid nodules with an ill-defined border, irregular margin, hypoechogenicity, an aspect ratio >1, and microcalcification were found to be significantly related to malignancy. Then, the maximum long-axis section of the thyroid nodules was preserved. Next, we adjusted the focus to the lower edge of the thyroid nodules, switching to contrast-enhanced ultrasound mode. This study was conducted using the SonoVue contrast agent (Bracco, Milan, Italy). We continuously monitored the dynamic perfusion process of the nodules in real-time after quickly pushing the contrast agent into the peripheral vein, we then exported the B-US images and CEUS clips in the dicom format (**Figure 1**). These examinations and evaluations were performed by a senior radiologist with over 20 years of experience in diagnostic ultrasound of thyroid disease.

Processing Images

We imported the dynamic contrast-enhanced ultrasound images into the TIC software for the quantitative analysis of CEUS parameters (**Figure 2**). Based on a previous study, the differences of the quantitative parameters of CEUS between benign and malignant nodules, which suggest significant differences of peak contrast perfusion within the nodule, were found between them (13). Therefore, the single frame corresponding to the moment of peak contrast perfusion in the target nodules during contrast ultrasound was selected to represent the whole process for radiomics analysis. In the next step, we normalized the grayscale and voxels of all images. Then, the ITK-SNAP software was used to draw an outline of the area of interest

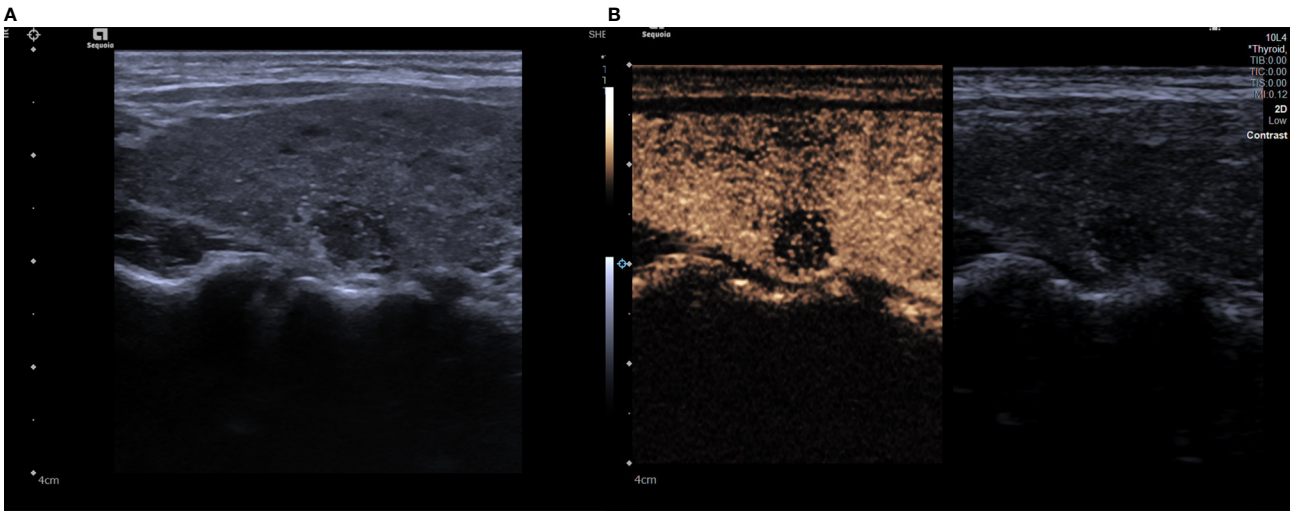


FIGURE 1 | **(A)** A grayscale ultrasound image. We found a 9 x9mm solid hypoechoic nodule in the right lobe of a patient's thyroid, this nodule has faint borders, irregular margin and microcalcifications inside. **(B)** A contrast ultrasound image, the nodule shown has low enhancement. The degree of malignancy risk of the nodule is classified as grade 5.

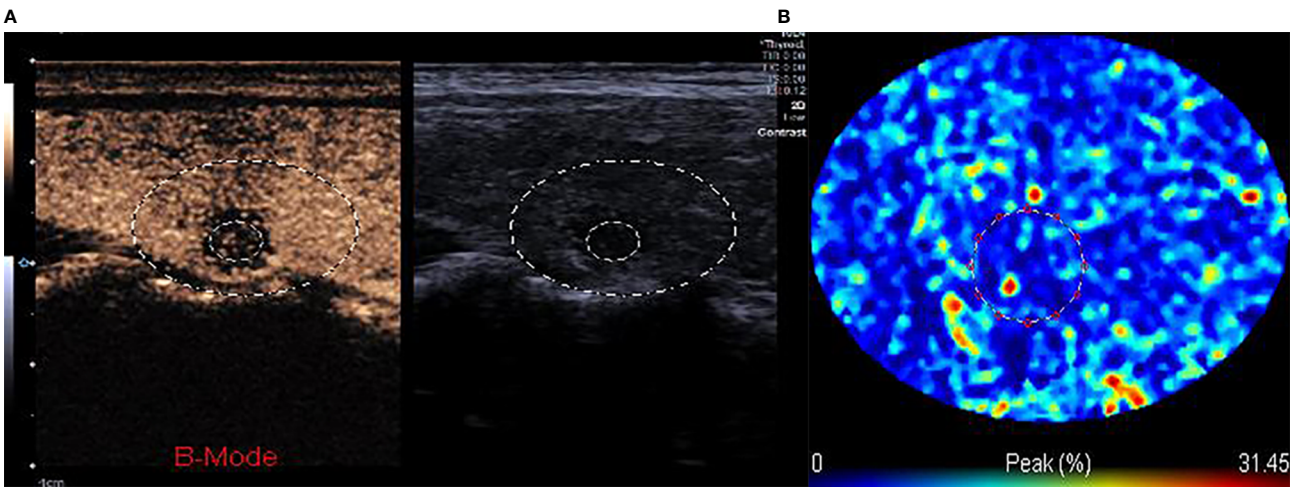


FIGURE 2 | **(A, B)** The TIC analysis curve of the CEUS image. First, we selected the target area with a large circle, second, we drew the ROI inside the nodule with a small circle. According to the TIC curve in the small circle, we selected the frame corresponding to the moment when internal nodule perfusion reached its peak.

(ROI) of the lesion in B-US images and CEUS images respectively (**Figure 3**), which was the target region for radiomics features extraction. The ROI was delineated by a radiologist with over 2 years of experience in diagnostic ultrasound of thyroid disease.

Extraction and Screening of Radiomics Features

The task of feature extraction was completed by 3D-Slicer 4.13.0. In total, 837 radiomics features including first-order statistics, texture features, a grayscale co-occurrence matrix, grayscale tour matrix, grayscale region size matrix, domain grayscale difference matrix, and morphological features were extracted from each B-US image and single frame of CEUS images, respectively. SPSS 23.0 software was used for analyzing the normality of radiomics features. Following the results, the Kruskal-Wallis test was used for preliminary screening. Then, the patients were randomly divided into the training cohort and validation cohort in the ratio of 9:3. The least absolute shrinkage was used to reduce the dimensionality of the radiomics features in R software

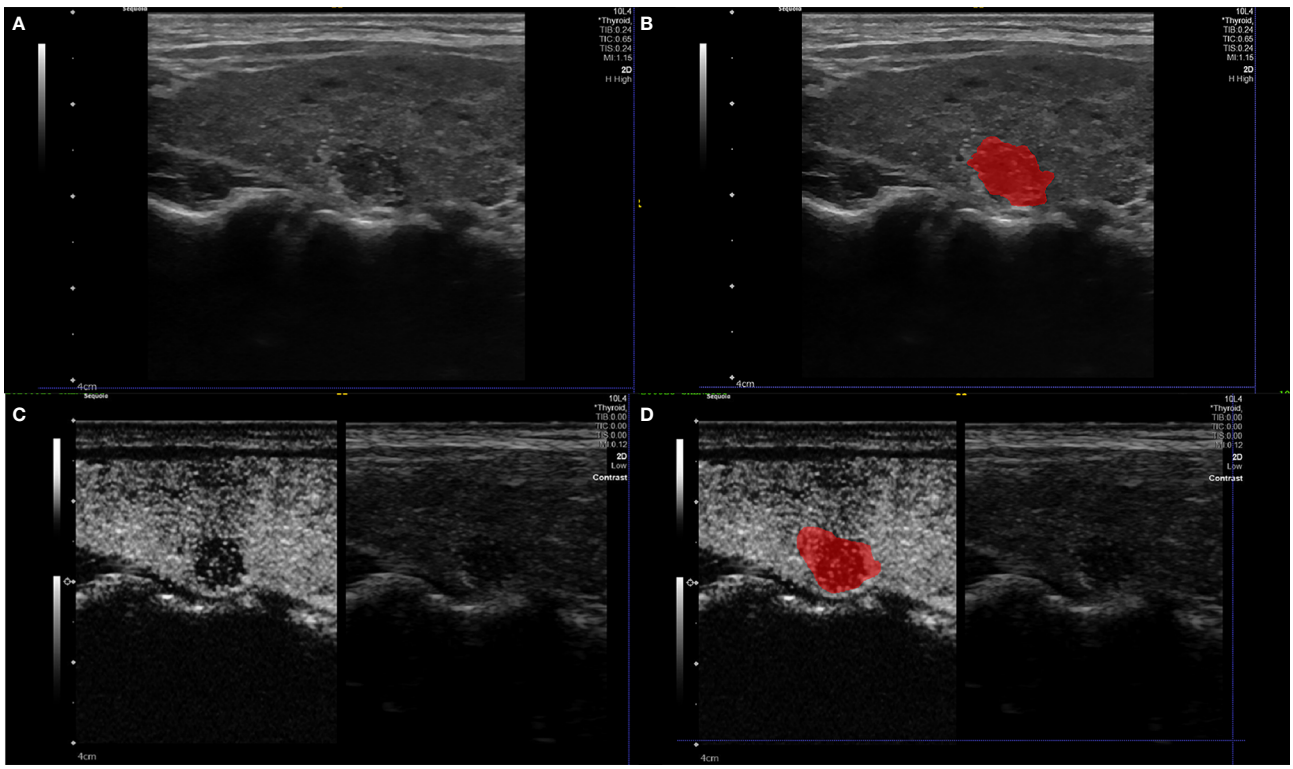


FIGURE 3 | Example **(A, B)** of delineating region of interest (ROI) on a B-US image, we detailed the outline along the nodule boundary on the largest long-axis cross-section of the nodule. Example **(C, D)** of delineating region of interest (ROI) on a CEUS image. The ROI includes the target node and some of its adjacent tissues (22).

(**Figure 4**, **5**), we obtained the final radiomics features for logistic regression analysis to build models for predicting the nature of thyroid nodules. Last, we verified the diagnostic performance of these models in the validation cohort.

Building and Comparison of the Models

TI-RADS + CEUS model: According to the diagnosis criteria of improved TI-RADS and combined with the pattern of contrast-enhanced ultrasound (5), a senior radiologist made risk grading diagnoses of 123 nodules for malignancy in turn. Based on the grades, logistic regression was used to build the TI-RADS + CEUS model.

Radiomics models: Based on the extracted radiomics features from different modality ultrasound images, logistic regression was used to build the B-US radiomics model and CEUS radiomics model. Later, we combined B-US radiomics features and CEUS radiomics features to build the B-US + CEUS radiomics model.

Then, we evaluated the performance of the four types of models for diagnosing the nature of thyroid nodules by comparing the area under the ROC curve.

Statistical Analysis

The statistical analysis and plots were performed using SPSS 23.0 and MedCalc19.6.0. The continuous quantitative data are

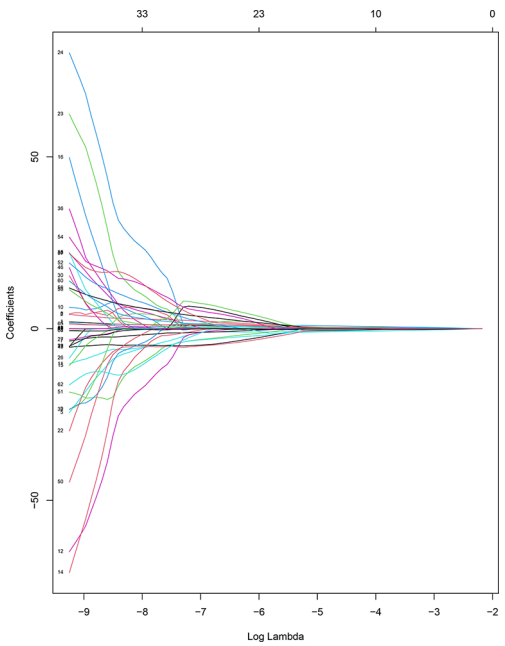


FIGURE 4 | Coefficient convergence graph of B-US radiomics features.

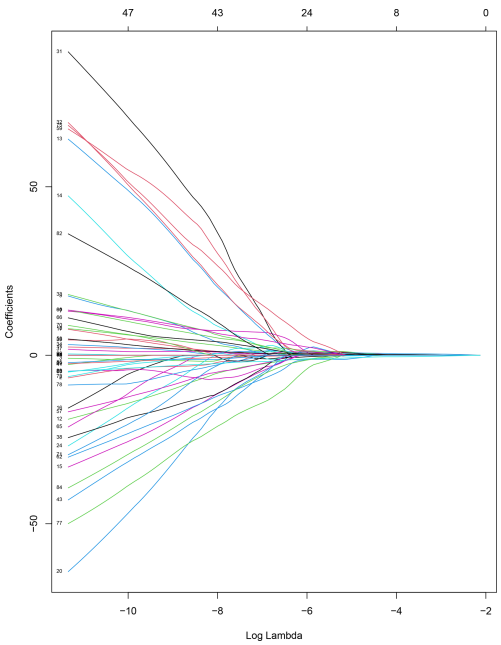


FIGURE 5 | Coefficient convergence graph of CEUS radiomics features.

shown as the mean ± standard deviation. The categorical and rank variables are shown as the number of cases and percentages. The chi-square test, Wilcoxon test, or Fisher`s exact test were applied, as appropriate, to compare the differences between benign and malignant groups. The sensitivity and specificity, positive prediction value (PPV) and negative prediction value (NPV), accuracy, and AUC in the nature of thyroid nodule diagnoses by the B-US radiomics model, CEUS radiomics model, B-US+CEUS radiomics model, and TI-RADS+CEUS model were calculated. The AUC was used to compare the diagnostic performance of the different models. Delong's test was used to test the difference in diagnostic performance between radiomics models and the TI-RADS+CEUS model. Results with P<0.05 meant that the difference was statistically significant.

RESULTS

A total of 123 nodules from 123 patients were enrolled in our study. Of which 95 nodules were excised by resection and FNAB was performed on 28 cases. The pathological types of all nodules are listed as follows: 94 malignant nodules and 29 benign nodules were included in the 123 cases of nodules. Among the benign nodules, 21 were nodular goiters, 3 were adenomas, 2 were adenomatous nodular goiters, and 3 were inflammatory lesions. Among the malignant nodules, 45 malignant nodules were papillary carcinomas, 49 malignant nodules were papillary thyroid microcarcinomas.

The results (**Table 1**) show that gender and age of the patients were not found to differ significantly between benign and malignant groups (P>0.05). Among the sonographic features of thyroid

TABLE 1 | Comparison of the general information between the benign and malignant groups.

Parameter		Malignant group	Benign group	*P value*
Age	Average	43.734 ± 1.189	48.069 ± 2.177	0.941
Gender	Male	23 (24.47%)	4 (13.79%)	0.225
	Female	71 (75.53%)	25 (86.21%)	
Location	Left lobe	45 (47.87%)	14 (48.27%)	1.000
	Gap	7 (7.45%)	2 (6.90%)	
	Right lobe	42 (44.68%)	13 (44.83%)	
Components	Solid	88 (93.62%)	24 (82.76%)	0.156
	Cystic solid mixed	6 (6.38%)	5 (17.24%)	
Echo	Hypoechogenicity	85 (90.42%)	17 (58.62%)	<0.01
	Equalechogenicity	4 (4.26%)	4 (13.79%)	
	Hyperechogenicity	5 (5.32%)	8 (27.59%)	
Border	Well-defined	15 (15.96%)	17 (58.62%)	<0.01
	Ill-defined	79 (84.04%)	12 (41.38%)	
Margin	Regular	39 (41.49%)	22 (75.86%)	<0.01
	Irregular	55 (58.51%)	7 (24.14%)	
Tall than wide Shape	>1	41 (43.62%)	5 (17.24%)	<0.05
	<1	49 (52.13%)	24 (82.76%)	
	=1	4 (4.25%)	0 (0)	
Calcifications	Microcalcifications	64 (68.09%)	6 (20.69%)	<0.01
	No Microcalcifications	30 (31.91%)	23 (79.31%)	
Contrast enhancement pattern	Low enhancement	82 (87.23%)	20 (68.97%)	0.053
	Equal enhancement	9 (9.58%)	7 (24.14%)	
	High enhancement	3 (3.19%)	2 (6.89%)	
Grade of TI-RADS+CEUS	3	0 (0)	7 (24.14%)	<0.01
	4a	16 (17.02%)	12 (41.38%)	
	4b	36 (38.30%)	6 (20.69%)	
	5	42 (44.68%)	4 (13.79%)	

nodules, we found no statistical significance in location, components, and contrast enhancement pattern of nodules (P>0.05). However, a statistic difference in the aspect ratio was observed in the benign group compared with the malignant group (P<0.05). We witnessed the statistically significant difference in the echoes, border, and margin, as well as significant differences in the presence of microcalcifications inside the nodules and grades of nodules in the benign group compared with the malignant group (P<0.01)

In the results, we obtained four radiomics features (**Table 2**) from B-US images to build the logistic regression model. For the differential diagnosis of benign and malignant thyroid nodules, the analysis of ROC curves indicated that the AUCs of the B-US radiomics model were 0.811 and 0.736 (**Figure 6**, **Table 3**) in the training cohort and the validation cohort, respectively. The results demonstrated that the four radiomics features extracted from B-US images had a good ability to distinguish the nature of thyroid nodules.

Finally, we obtained six radiomics features (**Table 4**) from CEUS images to build the logistic regression model. For the differential diagnosis of benign and malignant thyroid nodules, the analysis of ROC curves suggested that the AUCs of the US radiomics model were 0.770 and 0.736 (**Figure 7**, **Table 3**) in the training cohort and the validation cohort, respectively. The results demonstrated that the six radiomics features extracted from CEUS images had a good ability to distinguish the nature of thyroid nodules.

In our study, according to the result (**Tables 5**, **6**, **Figure 8**), we can learn that the sensitivity of the B-US radiomics model, CEUS radiomics model, B-US+CEUS radiomics model, and TI-RADS+CEUS model in the identification of benign and malignant thyroid nodules was 97.87%, 96.81%, 94.68%, and 100%, their specificity was 31.03%, 27.59%, 51.72%, and 24.14%, their accuracy was 82.11%, 80.49%, 84.55%, and 82.11%, their PPV was 82.14%, 81.25%, 86.41%, and 81.03%, their NPV was 81.82%, 72.72%, 75.00%, and 100%, and the AUCs of them were 0.791, 0.766, 0.861, and 0.785, respectively.

Compared to the TI-RADS+CEUS model, Delong's test showed there was no difference in the diagnostic performance between the US radiomics model, CEUS radiomics model, B-US+CEUS radiomics model, and TI-RADS+CEUS model (P>0.05). This indicated that the B-US model and CEUS

TABLE 2 | The B-US radiomics features.

B-US Radiomics Feature
original_ngtdm_Strength
wavelet-LLH_glszm_ZoneEntropy
wavelet-LHL_glszm_GrayLevelVariance
wavelet-LHH_gldm_LargeDependenceHighGrayLevelEmphasis

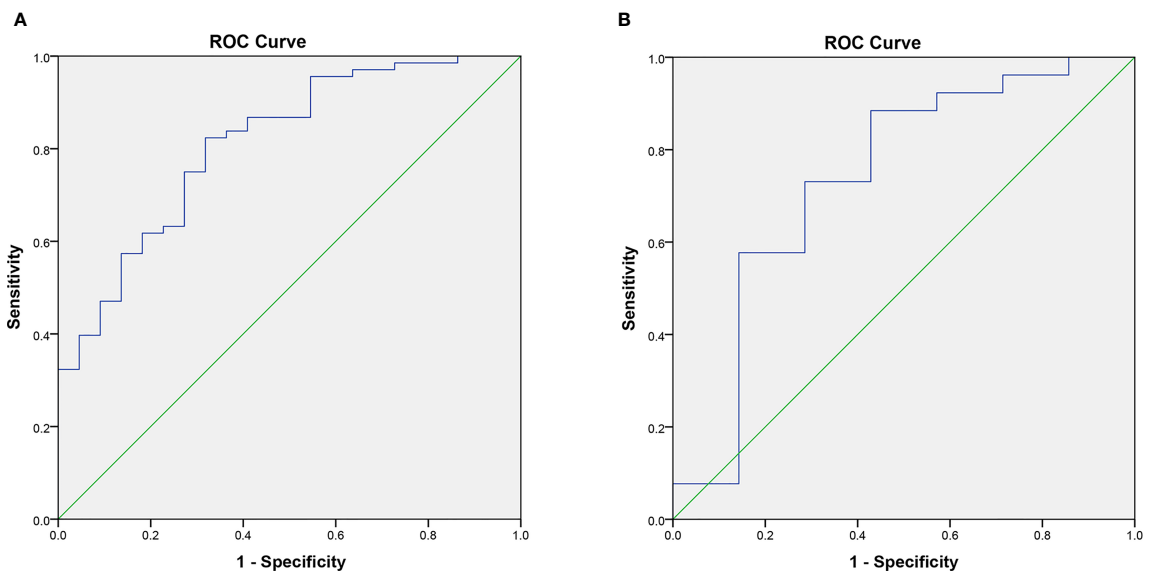


FIGURE 6 | The AUC of the B-US radiomics model in the training cohort **(A)** and validation cohort **(B)**.

TABLE 3 | Comparison of information in the training and validation cohorts.

Cohort	Malignant	Benign	AUC	
			B-US Radiomics	CEUS Radiomics
Training cohort	68	22	0.811	0.770
Validation cohort	26	7	0.736	0.736

TABLE 4 | The CEUS radiomics features.

CEUS radiomics features
wavelet-LHL_glrlm_LongRunLowGrayLevelEmphasis
wavelet-HLL_glcm_MaximumProbability
wavelet-HLL_glrlm_LongRunHighGrayLevelEmphasis
wavelet-HLL_glrlm_RunVariance
wavelet-HLH_gldm_DependenceVariance
wavelet-LLL_ngtdm_Contrast

model and B-US+CEUS radiomics model have comparable diagnostic performance to the TI-RADS+CEUS model. On the other hand, the AUC of the B-US+CEUS radiomics model was the highest among the four models. A significant difference was observed between the single B-US radiomics model or CEUS radiomics model and the B-US+CEUS radiomics model ($P<0.05$). Further analysis found that the specificity, accuracy, and PPV of the B-US+CEUS radiomics model were also the highest. However, its NPV was relatively low compared to the others.

DISCUSSION

The incidence of thyroid cancer has continued to grow rapidly in recent decades. According to statistics, there were 586,202 new cases of thyroid cancer worldwide in 2020, accounting for 3.0% of the total number of cancer cases and ranking 9th in incidence rate (14). Therefore, it is crucial to adopt appropriate treatment by identifying the nature of thyroid nodules accurately and effectively. After high-resolution ultrasound was combined with elastography and imaging technology, it has shown a very valuable performance in diagnosing the nature of thyroid nodules (15–17). However, the nature of certain nodules remains unclear. FNAB also suffers from inadequate sampling rate, and the invasive nature of it would lead to complications (2, 7, 18). Therefore, radiomics analysis can be an effective and promising non-invasive way for predicting the malignancy of thyroid nodules.

Recent studies from numerous researchers have reported that radiomics features extracted from B-US images can be used for the risk prediction of the malignancy in thyroid nodules (19, 20), which was also confirmed in our study, where the diagnostic performance of the B-US radiomics model was similar to the senior radiologist. Meanwhile, the radiomics studies based on multi-modality images is also gradually being developed. A study has proven that the combination of B-mode ultrasound and strain elastography ultrasound, applying radiomics analysis, could further improve the estimate accuracy of lymph node metastasis in patients with papillary thyroid carcinomas (21).

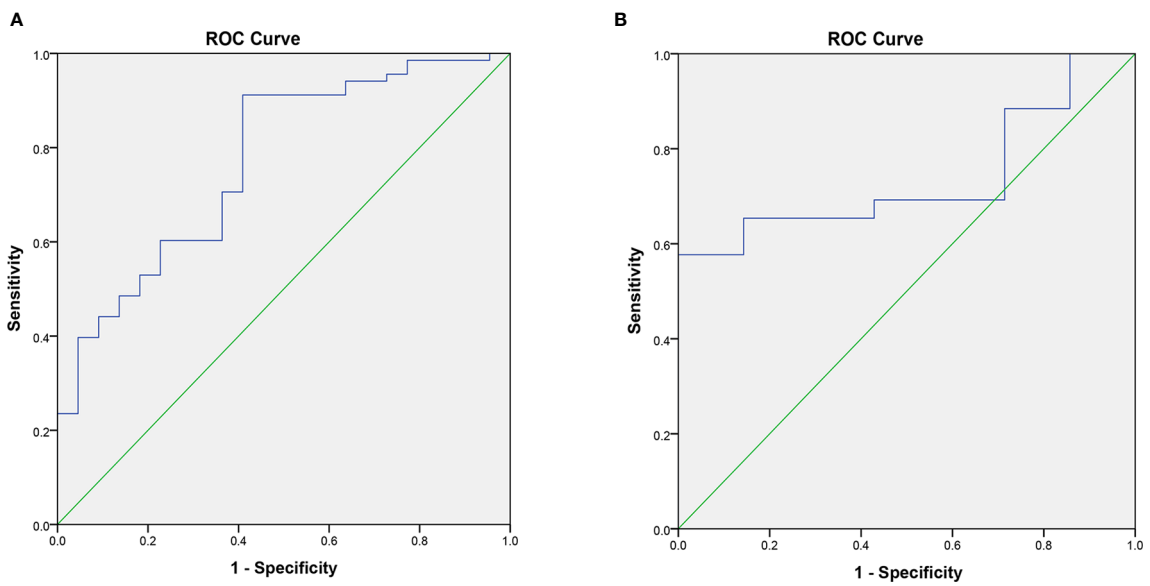


FIGURE 7 | The AUC of the CEUS radiomics model in the training cohort **(A)** and validation cohort **(B)**.

TABLE 5 | Comparison of the diagnostic effectiveness of four types of models.

Models	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Accuracy (%)
B-US radiomics	97.87	31.03	82.14	81.82	82.11
CEUS radiomics	96.81	27.59	81.25	72.72	80.49
B-US+CEUS radiomics	94.68	51.72	86.41	75.00	84.55
TI-RADS+CEUS	100	24.14	81.03	100	82.11

TABLE 6 | Comparison of the AUCs of four types of models.

Models	AUC	95%CI (AUC)		*P* value	
		lower bound	upper bound	Compare to TI-RADS+CEUS	Compare to B-US+CEUS radiomics
B-US radiomics	0.791	0.699	0.883	0.935	0.013
CEUS radiomics	0.766	0.671	0.861	0.743	0.026
B-US+CEUS radiomics	0.861	0.785	0.938	0.217	–
TI-RADS+CEUS	0.785	0.681	0.889	–	0.217

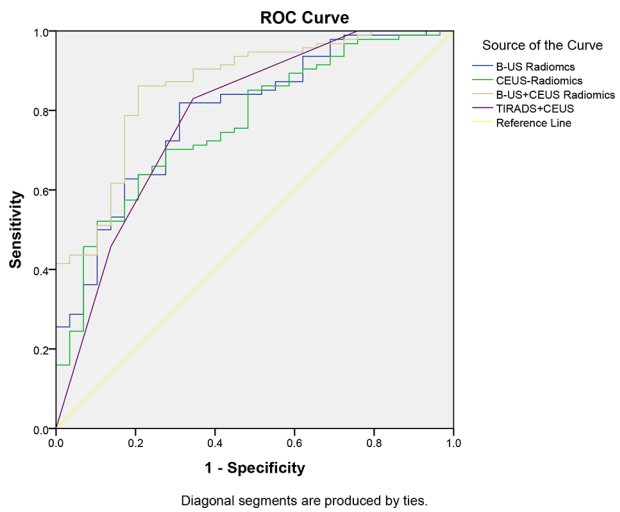


FIGURE 8 | Comparison of the AUCs of the four types of models.

As far as we know, no study has yet determined whether the use of CEUS images for radiomics analysis can provide valuable insights for diagnosis of the nature of thyroid nodules.

Liu et al. (22) demonstrated that a deep learning-based radiomics model, designed for analyzing contrast-enhanced ultrasound, could not merely achieve an accurate preoperative prediction of the progression-free survival for radiofrequency ablation and surgical resection, but could also contribute to the optimized treatment selection between patients with very early or early-stage hepatocellular carcinomas. This demonstrates that the results of radiomics analysis in CEUS images can also provide valuable information for clinical diagnosis and treatment. In our study, we selected single-frame images corresponding to the moment of peak intra-nodal perfusion in thyroid nodules to assess the value of radiomics analysis in CEUS images for diagnosing the nature of thyroid nodules. The results showed that the CEUS radiomics model exhibited equally good diagnostic performance compared to the TI-RADS+CEUS model. Moreover, the B-US+CEUS model showed superior diagnostic performance, highlighted by a specificity of 51.72%, a significant increase compared to the others; its accuracy and PPV were also at the highest. Although the sensitivity and NPV were not very good compared to the TI-RADS+CEUS model and B-US radiomics model, it still attained the best diagnostic performance after balancing all the differences.

The result suggested that the information provided by B-US radiomics features and CEUS radiomics features may complement each other in diagnosis of the nature of thyroid nodules, both of which may have a strong correlation with certain pathological features of thyroid cancer, but this correlation has not yet been revealed. In any case, the combination of both radiomics features have presented a significant improvement in diagnostic performance that is a promising prospect for clinical application.

However, the present study also had several limitations. First of all, the subjects were recruited from a single center, and the sample size was small. Besides, in view of the particularity of this examination, most nodules were grade 4a-4b in this experiment, so these limitations increased the likelihood of bias. As a result, there was no statistical significance in components and contrast enhancement patterns of nodules between the benign and malignant groups and a reduced specificity of all models in this study. Second, due to the limitation of technology, we only selected a single frame of the CEUS images to replace the whole process of perfusion, so other factors that have not been taken into consideration could have affected the diagnostics.

In the future, multi-center, large sample data are desired to further confirm the findings of our study. We expect to search for more advanced and effective technical methods to extract the radiomics features of dynamic images, and carry out more comprehensive and multi-faceted evaluation.

CONCLUSION

In summary, both the B-US radiomics features and CEUS radiomics features have the ability of qualitative diagnosis of thyroid nodules. By combining the two radiomics features we can obtain higher prediction performance, which shows a highly valuable clinical application in the diagnosis of thyroid nodules.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Board of the Third Xiangya

Hospital CSU. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SG did the first draft. PZ and YZ provided instructive advice and useful suggestions for this manuscript. YFZ and LJ provided statistical advice for this manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by grants from the National Natural Science Foundation of China (No. 81871367).

REFERENCES

1. Gharib H, Papini E, Garber JR, Duick DS, Harrell RM, Hegedüs L, et al. American Association of Clinical Endocrinologists, American College of Endocrinology, and Association Medici Endocrinologi Medial Guidelines for Clinical Practice for the Diagnosis and Management of Thyroid Nodules-2016 UPDATE. *Endocrine Prac: Off J Am Coll Endocrinol Am Assoc Clin Endocrinologists* (2016) 22(5):622–39. doi: 10.4158/EP161208.GL
2. Grani G, Sponziello M, Pecce V, Ramundo V, Durante C. Contemporary Thyroid Nodule Evaluation and Management. *J Clin Endocrinol Metab* (2020) 105(9):2869–83. doi: 10.1210/clinem/dgaa322
3. Shin JH, Baek JH, Chung J, Ha EJ, Kim JH, Lee YH, et al. Ultrasonography Diagnosis and Imaging-Based Management of Thyroid Nodules: Revised Korean Society of Thyroid Radiology Consensus Statement and Recommendations. *Korean J Radiol* (2016) 17(3):370–95. doi: 10.3348/kjr.2016.17.3.370
4. Brito JP, Gionfriddo MR, Al Nofal A, Boehmer KR, Leppin AL, Reading C, et al. The Accuracy of Thyroid Nodule Ultrasound to Predict Thyroid Cancer: Systematic Review and Meta-Analysis. *J Clin Endocrinol Metab* (2014) 99 (4):1253–63. doi: 10.1210/jc.2013-2928
5. Zhang Y, Zhou P, Tian SM, Zhao YF, Li JL, Li L. Usefulness of Combined Use of Contrast-Enhanced Ultrasound and TI-RADS Classification for the Differentiation of Benign From Malignant Lesions of Thyroid Nodules. *Eur Radiol* (2017) 27(4):1527–36. doi: 10.1007/s00330-016-4508-y
6. Haugen BR, Alexander EK, Bible KC, Doherty GM, Mandel SJ, Nikiforov YE, et al. 2015 American Thyroid Association Management Guidelines for Adult Patients With Thyroid Nodules and Differentiated Thyroid Cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. *Thyroid: Off J Am Thyroid Assoc* (2016) 26 (1):1–133. doi: 10.1089/thy.2015.0020
7. Shaha AR. Comment on "The Diagnosis and Management of Thyroid Nodules: A Review". *Gland Surg* (2018) 7(5):424–5. doi: 10.21037/gs.2018.06.01
8. Ha EJ, Lim HK, Yoon JH, Baek JH, Do KH, Choi M, et al. Primary Imaging Test and Appropriate Biopsy Methods for Thyroid Nodules: Guidelines by Korean Society of Radiology and National Evidence-Based Healthcare Collaborating Agency. *Korean J Radiol* (2018) 19(4):623–31. doi: 10.3348/kjr.2018.19.4.623
9. Jiang D, Zang Y, Jiang D, Zhang X, Zhao C. Value of Rapid on-Site Evaluation for Ultrasound-Guided Thyroid Fine Needle Aspiration. *J Int Med Res* (2019) 47(2):626–34. doi: 10.1177/0300060518807060
10. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer (Oxford England: 1990)* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
11. Zhan J, Ding H. Application of Contrast-Enhanced Ultrasound for Evaluation of Thyroid Nodules. *Ultrasonography (Seoul Korea)* (2018) 37(4):288–97. doi: 10.14366/usg.18019
12. Kwak JY, Han KH, Yoon JH, Moon HJ, Son EJ, Park SH, et al. Thyroid Imaging Reporting and Data System for US Features of Nodules: A Step in Establishing Better Stratification of Cancer Risk. *Radiology* (2011) 260(3):892–9. doi: 10.1148/radiol.11110206
13. Ma JJ, Sun M, Tian Y, Chen J, Cheng JN, Wang XH, et al. Values of the Quantitative Parameters of Contrast-Enhanced Ultrasound in the Diagnosis of Thyroid Benign and Malignant Nodules. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao Acta Academiae Medicinae Sinicae* (2020) 42(1):80–5. doi: 10.3881/j.issn.1000-503X.11269
14. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
15. Zhang YZ, Xu T, Gong HY, Li CY, Ye XH, Lin HJ, et al. Application of High-Resolution Ultrasound, Real-Time Elastography, and Contrast-Enhanced Ultrasound in Differentiating Solid Thyroid Nodules. *Medicine* (2016) 95 (45):e5329. doi: 10.1097/MD.0000000000005329
16. Pang T, Huang L, Deng Y, Wang T, Chen S, Gong X, et al. Logistic Regression Analysis of Conventional Ultrasonography, Strain Elastosonography, and Contrast-Enhanced Ultrasound Characteristics for the Differentiation of Benign and Malignant Thyroid Nodules. *PloS One* (2017) 12(12):e0188987. doi: 10.1371/journal.pone.0188987
17. Tuan PA, Duc NM, An M, Vien MV, Giang BV. The Role of Shear Wave Elastography in the Discrimination Between Malignant and Benign Thyroid Nodules. *Acta informatica medica: AIM: J Soc Med Inf Bosnia Herzegovina: casopis Drustva za medicinsku informatiku BiH* (2020) 28(4):248–53. doi: 10.5455/aim.2020.28.248-253
18. Cappelli C, Pirola I, Agosti B, Tironi A, Gandossi E, Incardona P, et al. Complications After Fine-Needle Aspiration Cytology: A Retrospective Study of 7449 Consecutive Thyroid Nodules. *Br J Oral Maxillofacial Surg* (2017) 55 (3):266–9. doi: 10.1016/j.bjoms.2016.11.321
19. Liang J, Huang X, Hu H, Liu Y, Zhou Q, Cao Q, et al. Predicting Malignancy in Thyroid Nodules: Radiomics Score Versus 2017 American College of Radiology Thyroid Imaging, Reporting and Data System. *Thyroid: Off J Am Thyroid Assoc* (2018) 28(8):1024–33. doi: 10.1089/thy.2017.0525
20. Zhou H, Jin Y, Dai L, Zhang M, Qiu Y, Wang K, et al. Differential Diagnosis of Benign and Malignant Thyroid Nodules Using Deep Learning Radiomics of Thyroid Ultrasound Images. *Eur J Radiol* (2020) 127:108992. doi: 10.1016/j.ejrad.2020.108992
21. Liu T, Ge X, Yu J, Guo Y, Wang Y, Wang W, et al. Comparison of the Application of B-Mode and Strain Elastography Ultrasound in the Estimation of Lymph Node Metastasis of Papillary Thyroid Carcinoma Based on a Radiomics Approach. *Int J Comput Assist Radiol Surg* (2018) 13(10):1617–27. doi: 10.1007/s11548-018-1796-5
22. Liu F, Liu D, Wang K, Xie X, Su L, Kuang M, et al. Deep Learning Radiomics Based on Contrast-Enhanced Ultrasound Might Optimize Curative Treatments for Very-Early or Early-Stage Hepatocellular Carcinoma Patients. *Liver Cancer* (2020) 9(4):397–413. doi: 10.1159/000505694

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Guo, Zhou, Zhang, Jiang and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 18 October 2021
doi: 10.3389/fonc.2021.729371

CT-Based Peritumoral and Intratumoral Radiomics as Pretreatment Predictors of Atypical Responses to Immune Checkpoint Inhibitor Across Tumor Types: A Preliminary Multicenter Study

Shuai He[1], Yuqing Feng[2], Qi Lin[3], Lihua Wang[1], Lijun Wei[1], Jing Tong[4], Yuwei Zhang[5], Ying Liu[5], Zhaoxiang Ye[5], Yan Guo[6], Tao Yu[1] and Yahong Luo[1]**

[1] Department of Medical Imaging, Cancer Hospital of China Medical University, Liaoning Cancer Hospital and Institute, Shenyang, China, [2] Department of Oncology, The Fifth People's Hospital of Shenyang, Shenyang, China, [3] Department of Radiology, First Affiliated Hospital of Xiamen University, Xiamen, China, [4] Department of Radiology, General Hospital of Northern Theater Command, Shenyang, China, [5] Department of Radiology, Tianjin Medical University Cancer Institute and Hospital, Tianjin, China, [6] Prognostic Diagnosis, GE Healthcare China, Beijing, China

Objective: To develop and validate a new strategy based on radiomics features extracted from intra- and peritumoral regions on CT images for the prediction of atypical responses to the immune checkpoint inhibitor (ICI) in cancer patients.

Methods: In total, 135 patients derived from five hospitals with pathologically confirmed malignancies receiving ICI were included in this retrospective study. Atypical responses including pseudoprogression (PsP) and hyperprogression disease (HPD) were identified as their definitions. A subgroup of standard progression disease (sPD) in 2018 was also involved in this study. Based on pretreatment CT imaging, a total of 107 features were extracted from intra- and peri-tumoral regions, respectively. The least absolute shrinkage and selection operator (Lasso) algorithm was used for feature selection, and multivariate logistic analysis was used to develop radiomics signature (RS). Finally, a total of nine RSs, derived from intra-tumoral, peri-tumoral, and combination of both regions, were built respectively to distinguish PsP *vs*. HPD, PsP *vs*. sPD, and HPD *vs*. sPD. The performance of the RSs was evaluated with discrimination, calibration, and clinical usefulness.

Results: No significant difference was found when compared in terms of clinical characteristics of PsP, HPD, and sPD. RS based on combined regions outperformed those from either intra-tumoral or peri-tumoral alone, yielding an AUC (accuracy) of 0.834 (0.827) for PsP *vs*. HPD, 0.923 (0.868) for PsP *vs*. sPD, and 0.959 (0.894) for HPD *vs*. sPD in the training datasets, and 0.835 (0.794) for PsP *vs*. HPD, 0.919 (0.867) for PsP *vs*. sPD, and 0.933 (0.842) for HPD *vs*. sPD in the testing datasets. The combined RS showed good fitness (Hosmer–Lemeshow test $p > 0.05$) and provided more net benefit than the

OPEN ACCESS

Edited by:
Giorgio Russo, Institute of Bioimaging and Molecular Physiology (CNR), Italy

Reviewed by:
Xiaohu Li, First Affiliated Hospital of Anhui Medical University, China
Shaocheng Zhu, Henan Provincial People's Hospital, China
Xiaopan Xu, Fourth Military Medical University, China

***Correspondence:**
Tao Yu
18900917557@189.cn
Yahong Luo
Luoyahong8888@hotmail.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *23 June 2021*
Accepted: *23 September 2021*
Published: *18 October 2021*

Citation:
He S, Feng Y, Lin Q, Wang L, Wei L, Tong J, Zhang Y, Liu Y, Ye Z, Guo Y, Yu T and Luo Y (2021) CT-Based Peritumoral and Intratumoral Radiomics as Pretreatment Predictors of Atypical Responses to Immune Checkpoint Inhibitor Across Tumor Types: A Preliminary Multicenter Study. Front. Oncol. 11:729371. doi: 10.3389/fonc.2021.729371

treat-none or treat-all scheme by decision curve analysis in both training and testing datasets.

Conclusion: Pretreatment radiomics are helpful to predict atypical responses to ICI across tumor types. The combined RS outperformed those from either intra- or peri-tumoral alone which may provide a more comprehensive characterization of atypical responses to ICI.

Keywords: radiomics, CT, peritumoral, immune checkpoint inhibitor, atypical responses

INTRODUCTION

The novel development of the immune checkpoint inhibitor (ICI) is now approved in a variety of solid tumors, including melanoma, non-small cell lung cancer (NSCLC), and urothelial and microsatellite instability-high (MSI) cancer, represented by programmed cell death-1 (PD-1) and programmed cell death ligand-1 (PD-L1), which became a crucial therapeutic option to improve prognosis (1). Unlike chemotherapy and tyrosine kinase inhibitor (TKI), ICI plays an antitumor role by blocking the immune checkpoint and enhancing the activity of autologous T cells (2). These effects occur through the restart of intrinsic immune actions, and the efficacy of these effects is strictly associated with the appearance of hypoxia, necrosis, and inflammation at the tumor sites (3). Meanwhile, these biological processes can affect the immune system and adjust antitumor responses, giving rise to atypical responses, including pseudoprogression (PsP) and hyperprogression disease (HPD) (4–7).

During tumor assessment, PsP occurs as a shrinkage in tumor burden after increasing in size or the presence of new lesions (8, 9), whereas HPD is presented as an acceleration of tumor growth after the initiation of immunotherapy, as compared to the period before treatment initiation used as a reference (10–12). Most similarly, all PsP, HPD, and standard progression disease (sPD) patients share common imaging with tumor enlargement at initial radiography assessment. Dissimilarly, compared with sPD and HPD, PsP patients have good clinical outcomes with significant longer progression-free survivals and overall survivals (13, 14). Thus, distinguishing PsP from sPD or even HPD will extremely help evaluate the efficacy of ICI and avoid either premature withdrawal of the treatment or prolonging ineffective treatment. Unfortunately, the identification of a reliable predictive biomarker of atypical responses to immunotherapy across various solid tumors remains an unmet need in clinic practice so far.

Routine standard-of-care CT scans are a noninvasive clinical examination tool for tumor diagnosis, staging, and monitoring treatment response. CT imaging-based radiomics can characterize both intra-tumoral and peri-tumoral heterogeneity from digital images to build mathematical formulas that reflect the underlying pathophysiology (15). Nowadays, radiomics has been successfully applied to the prediction of tumor histology, risk of lymph node metastasis, genetic mutation subtypes, and decoding of PD-L1 expression in cancer patients (16–19). However, investigations using radiomics on prediction of atypical responses to cancer immunotherapy are rather rare. Recently, Wang et al. used CT radiomics to identify five PsP cases from 50 melanoma patients treated with anti-cytotoxic T-lymphocyte antigen-4 (CTLA-4) inhibitor (20). Nonetheless, it needs a larger sample to confirm the ability of radiomics to predict PsP before immunotherapy.

In this study, we aim to evaluate the predictive value of radiomics, including intra- and peri-tumoral features that distinguish among PsP, HPD, and sPD patients, which may assist clinicians in the precision management of personalized immunotherapy.

MATERIAL AND METHODS

Patients

This retrospective multicenter study was approved by each participating institutional review board, and the prerequisite for obtaining informed consent was waived. Data were collected from February 2017 to April 2020 in patients with pathologically proven malignant solid tumors who had been treated with ICI, alone or in combination with chemotherapy.

Four hundred and sixty-three consecutive patients from five Chinese Hospitals were identified. The inclusion criteria were as follows: (1) CT scan prior to the initiation of ICI in less than 2 weeks; (2) at least two cycles of ICI; and (3) all solid tumors were pathologically confirmed. The exclusion criteria were as follows: (1) with history of any other concurrent malignancies; (2) no measurable lesion or with obvious artifacts on CT images; and (3) without a previous and/or follow-up CT scan.

Definitions of Pseudoprogression and Hyperprogression

Immune Response Evaluation Criteria in Solid Tumors (iRECIST) were used to evaluate the response of tumors (21). In our datasets of patients, PsP **(Figure 1A)** was defined as immune unconfirmed progressive disease (iUPD) during evaluation and further response classified as immune complete response (iCR), immune partial response (iPR), or immune stable disease (iSD) (22). HPD **(Figure 1B)** criteria, which are as follows in accordance with previous studies: 1) progression at first post-ICI, 2) increase in tumor size over 50%, and 3) over two-fold increase in progression rate (23, 24). We also defined other immune confirmed progressive diseases (iCPDs) except for HPD as sPD by iRECIST criteria. All measurable target lesions (≥10 mm in the longest diameter for non-nodal lesions and ≥15 mm in the short axis for nodal lesions) allow up to two lesions

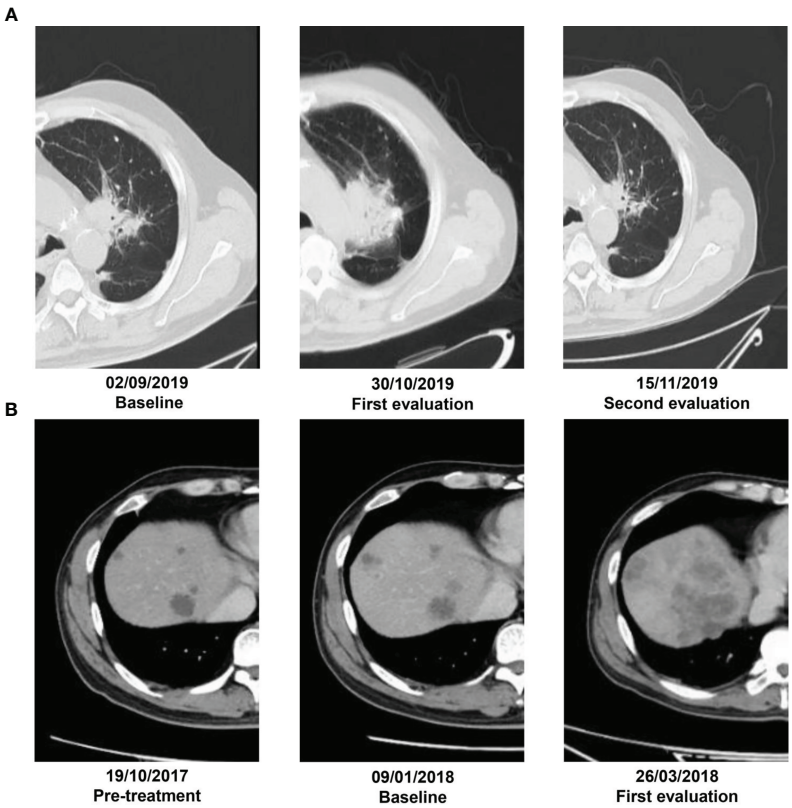


FIGURE 1 | **(A)** A patient with adenocarcinoma by puncture biopsy pathology who was receiving nivolumab therapy. Irregular lesion in the left upper lobe with a diameter of 3.2 cm on the baseline CT scans. By 6 weeks of anti-PD-1 therapy, the lesion increased in diameter of 5.4 cm on the first CT evaluation. At 8 weeks of therapy, it had decreased in size by 2.2 cm. **(B)** Another 64-year-old male was treated with pembrolizumab on January 29, 2018, for liver metastasis from colorectal cancer. It had SD prior to the initiation of immunotherapy but developed rapid tumor growth with appearance of new lesions on the first follow-up and experienced more than two-fold increase from pretreatment tumor growth *versus* treatment.

per organ, and five lesions in total as in iRECIST criteria were used for analysis.

According to the definitions, 34 patients with 42 target lesions undergo PsP, and 43 patients with 67 target lesions experience HPD. A subgroup of sPD (58 out of 220) in 2018 was also included in this study. Finally, the training datasets were recruited from the Liaoning Cancer Hospital while the testing datasets involved patients from The Fifth People's Hospital of Shenyang, First Affiliated Hospital of Xiamen University, General Hospital of Northern Theater Command, and Tianjin Cancer Hospital. The flowchart of patient selection procedure is showed in **Figure 2**.

CT Acquisition

The pretreatment CT scans were acquired on varied datasets of CT scanners **(Supplemental Data)**.

Segmentation and Feature Extraction

Before segmentation, all images were resampled to a common voxel spacing of 1 mm × 1 mm × 1 mm by using the linear interpolation algorithm to construct new data points within the range of discrete datasets of known data points to standardize spacing across all images (25).

Then, the region of interest (ROI) was delineated manually along the tumor contour slice by slice on pretreatment CT images by reader 1 (HS with 9 years of experience) who were blinded to diagnosis and clinical information, using an open-source software (ITK-SNAP, version 3.6.0, http://www.itksnap.org/). The morphologic operation of dilation was then performed to capture the information outside the lesion up to a radial distance of 5 mm; normal tissue or surrounding organs were subsequently excluded from the contours.

Subsequently, a total of 107 radiomics features, which regarded the image biomarker standardization initiative (IBSI) as reference (26), were extracted using A.K. software (Artificial Intelligence Kit, version 2.0.0, GE Healthcare, China) from each region, including the intra- and peri-tumoral ROI.

In order to investigate the reproducibility of the radiomics features obtained by different readers, different times, and different tumor regions (intratumoral and peritumoral), intra- and interclass correlation coefficients (ICCs) were used to assess the reproducibility of the radiomics features extracted from 30 randomly chosen patients. To assess the inter-reader reproducibility, the ROI delineation was performed by two oncologic radiologists (reader 1 and reader 2, HS with 9 years

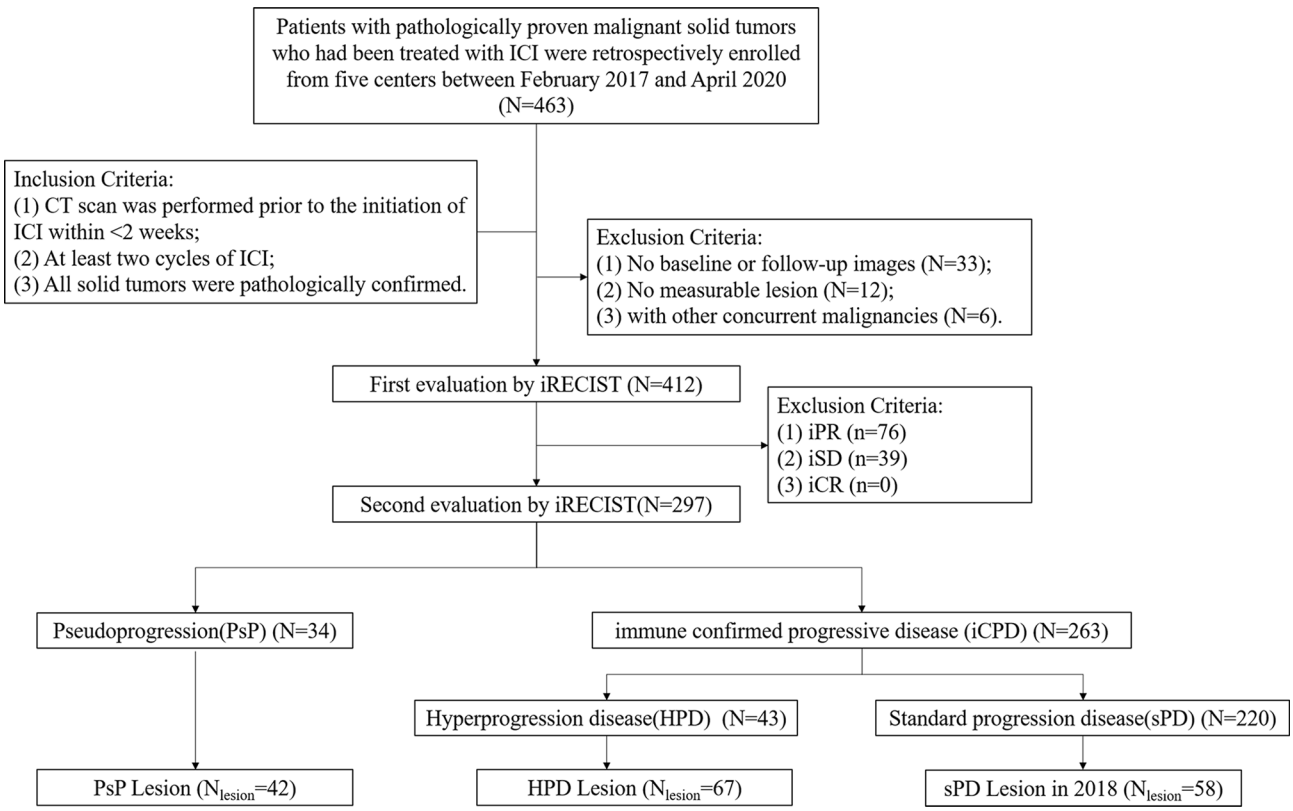


FIGURE 2 | Flowchart of the patient selection procedure.

and WLH with 7 years of experience), respectively. To evaluate the intra-reader reproducibility, reader 1 repeated the ROI delineation at a 1-month interval.

Radiomics Signature Building

To reduce overfitting or selection bias, we adopted a series of methods for dimensionality reduction and feature selection before modeling. In the first step, features of ICCs >0.75 for both inter-reader and intra-reader that were considered a relatively high inter-reader and intra-reader variability in the segmented ROI, were included in subsequent analysis. Subsequently, Spearman correlation analysis was conducted to remove the redundant features which were highly correlated($|r|>0.90$) with other features. Then, the least absolute shrinkage and selection operator (LASSO) regression algorithm with penalty parameter tuning was applied with 10-fold cross validation to select the most useful predictive features with a non-zero coefficient.

Then, a total of nine radiomics signatures (RSs), including intra-tumoral RS, peri-tumoral RS, and combined (intra-plus-peri-tumoral) RS for distinguishing PsP *vs*. HPD, PsP *vs*. sPD, and HPD *vs*. sPD, were built respectively *via* multivariate logistic analysis using the selected optimal feature sets in the training datasets and then tested in the testing datasets. The workflow of the radiomics analysis is shown in **Figure 3**.

Performance Evaluation

The performance of the RSs was evaluated with discrimination, calibration, and clinical application in both training and testing datasets.

Discrimination Degree

Receiver operating characteristic (ROC) curves were plotted, and the area under the ROC curve (AUC) with a 95% confidence interval (CI) was used to assess the diagnostic performance in discriminating PsP *vs*. HPD, PsP *vs*. sPD, and HPD *vs*. sPD in both training and testing datasets. The optimal cutoff of the RSs calculated from the training datasets based on the maximum Youden's index was then applied in the testing datasets. The accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), and negative predictive value (NPV) were calculated in both training and testing datasets.

Calibration Degree

Calibration curves were plotted in both training and testing datasets to explore the agreement between the observed outcome frequencies and predicted probabilities of the RSs. The Hosmer–Lemeshow test was used to determine the goodness of fit of the models, and p values of more than 0.05 were considered well-calibrated.

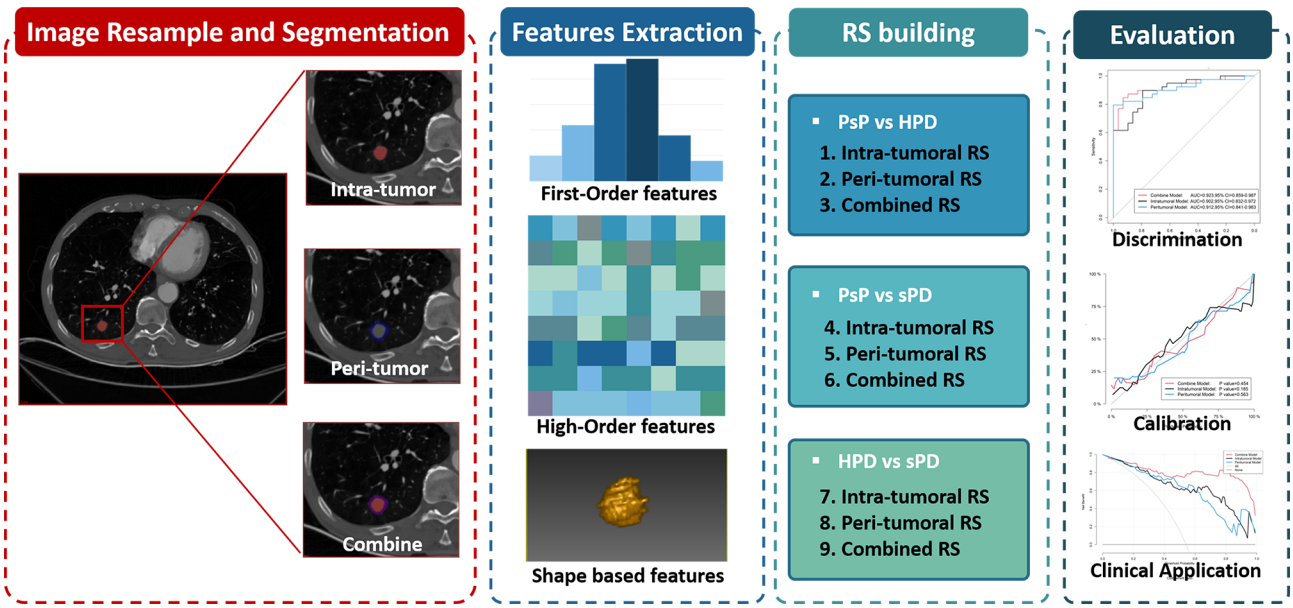


FIGURE 3 | The radiomics workflow including tumor segmentation, feature extraction, radiomics signatures construction, and performance evaluation.

Clinical Application

Decision curve analysis (DCA) was conducted to assess the clinical usefulness by quantifying the net benefits at different threshold probabilities in both training and testing datasets.

Statistical Analysis

All statistical analyses were performed using R language (version 3.5.1, https://www.r-project.org). Categorical variables between two or more groups were compared with the χ^2 test, and continuous variables between groups were compared with either Student's t test or Mann–Whitney U test (for two groups, as appropriate) with Bonferroni correction ($p < 0.017$ indicated significance) or by ANOVA (for three groups). Categorical variables were presented as counts (percentage), and continuous variables were presented as mean (SD) or median (25%, 75%), as appropriate.

ICC was calculated using the "lme4" package. LASSO regression was performed using the "glmnet" package. Multivariate logistic regression was performed using the "rms" package. ROC curves were plotted using the "pROC" package. The calibration curve and Hosmer–Lemeshow test were conducted using the "ModelGood" package. Decision curve analysis was performed using the "dca. R" package.

RESULTS

Patient Dataset

A total of 135 patients including PsP (N = 34), HPD (N = 43), and sPD (N = 58) were analyzed in this study. The most common tumor types included the respiratory system neoplasms (n = 73) and digestive system neoplasms (n = 31). In our population, the incidence rate of HPD (10.44%) is slightly higher than PsP (8.25%) in the whole datasets. More than half of the patients (n = 72) received ICI monotherapy, and 46.67% (n = 63) received combination chemo-immunotherapy.

There was no statistically significant difference in baseline age, gender, pre-chemotherapy or radiotherapy, monotherapy or plus chemotherapy, brain metastasis, bone metastasis, lung metastasis, or hepatic metastasis (all $p > 0.05$) among the three groups. The characteristics of 135 patients are summarized in **Table 1**.

Development and Validation of Radiomics Signatures

RS for Discriminating PsP From HPD

After inter- and intra-reader reproducibility analysis, a total of 210/214 (104/107 from intra-tumoral and 106/107 from peri-tumoral) features showed stability with both intra- and inter-reader ICCs greater than 0.75; the details are shown in **Table 2** and **Supplementary Figure 1**.

The most predictive and strongest features were remained after the process of feature selection (**Table S1**). Then, RSs were calculated for each patient *via* a linear combination of the selected features weighted by respective coefficients (the calculation formulas are shown in **Supplementary Formula**).

The distribution of the combined RSs in the training and testing datasets is shown in **Figure S2**. The RSs derived from intra-tumoral, peri-tumoral, and combined models were all significantly higher in patients with HPD than those with PsP in both training and testing datasets (all $p < 0.017$ after Bonferroni correction). According to the maximum Youden's index, 0.109, 0.386, and −0.298 were respectively set as the optimal cutoff values in the intra-, peri-tumoral, and combined models (**Tables 3-1**). **Table 4**, **Figure S3** and **Figure 4** show the discriminative performance of the models. The combined RS extracted from the intra- and peri-tumoral models yielded the

TABLE 1 | Baseline characteristics of 135 patients.

Characteristics	PsP (N = 34)	HPD (N = 43)	sPD (N = 58)	*p* value
Age, median (range) years	67 (52-81)	62 (45-77)	72 (57-87)	0.368
Gender, No. (%)				0.202
Male	29 (85.3)	32 (74.4)	39 (68.4)	
Female	5 (14.7)	11 (25.6)	18 (31.6)	
Pre- chemotherapy or radiotherapy, No. (%)				0.851
Yes	26 (76.5)	33 (23.3)	46 (80.7)	
No	8 (23.5)	10 (76.7)	11 (19.3)	
Treatment strategy, No. (%)				0.183
Monotherapy	19 (44.1)	17 (39.5)	21 (36.8)	
Combination therapy	15 (44.1)	26 (60.5)	36 (63.2)	
Number of lines of prior systemic cancer therapy, No. (%)				0.337
1	7 (20.6)	9 (20.9)	9 (15.8)	
≥2	27 (79.4)	34 (79.1)	48 (84.2)	
Lung metastasis, No. (%)				0.817
With	18 (52.9)	25 (58.1)	34 (59.6)	
Without	16 (47.1)	18 (41.9)	23 (40.4)	
Brain metastasis, No. (%)				0.124
With	2 (5.9)	7 (16.3)	3 (5.3)	
Without	32 (94.1)	36 (83.7)	54 (94.7)	
Bone metastasis, No. (%)				0.120
With	8 (23.5)	16 (37.2)	11 (19.3)	
Without	26 (76.5)	27 (62.8)	46 (80.7)	
Liver metastasis, No. (%)				0.240
With	5 (14.7)	18 (41.9)	18 (41.9)	
Without	29 (85.3)	29 (85.3)	25 (58.1)	

Psp, pseudoprogression; HPD, hyperprogression disease; sPD, standard progression disease.

highest AUC of 0.834 and an ACC of 0.827 in the training datasets, which were higher than those of the intra-tumoral model (yielded an AUC value of 0.804 and an ACC value of 0.733) alone. However, the RS extracted from the peri-tumoral model yielded the highest AUC of 0.848, along with a lower ACC of 0.773 than those of combined RS. Considering AUC and ACC together, a corresponding result was also found in testing datasets. The combined RS yielded the highest AUC value of 0.835 and an ACC value of 0.794, which were higher than those of intra- (yielded an AUC of 0.769 and ACC of 0.735) and peri-tumoral (yielded an AUC of 0.824 and ACC of 0.765) models alone. Interestingly, the RS extracted from the peri-tumoral model yielded a higher AUC than that of the intra-tumoral model in both training and testing datasets. The calibration curve indicated the good fitness of the conventional model with a p value of the Hosmer–Lemeshow test bigger than 0.05, which is shown in **Figure S4** and **Figure 5**. The clinical usefulness is shown as a decision curve in **Figure S5** and **Figure 6**.

RS for Discriminating PsP From sPD

Similar results were observed in PsP *vs.* sPD. As shown in **Table 4**, **Figure S3** and **Figure 4**, the combined RS yielded the highest AUC of 0.923 and an ACC of 0.868, which were higher than those of the intra- (yielded an AUC of 0.902 and an ACC of 0.838) and peri-tumoral (yielded an AUC of 0.912 and an ACC of 0.868) models alone. The peri-tumoral RS yielded a higher AUC value than that of the intra-tumoral model in both training and testing datasets. The RSs derived from the combined region were all significantly higher in patients with sPD than those with PsP in both training and testing datasets (**Figure S2**).

TABLE 2 | The process of features selection.

Radiomics signatures		Remained feature number			
		Extracted	ICC>0.75	Spearman (|r|<0.90)	LASSO (non-zero)
PsP *VS.* HPD	**Intra-tumoral**	107	104	45	4
	Peri-tumoral	107	106	36	4
	Combined	214	210	79	5
PsP *VS.* sPD	**Intra-tumoral**	107	106	41	7
	Peri-tumoral	107	106	42	8
	Combined	214	212	69	11
HPD *VS.* sPD	**Intra-tumoral**	107	104	47	7
	Peri-tumoral	107	106	43	9
	Combined	214	210	80	12

Psp, pseudoprogression; HPD, hyperprogression disease; sPD, standard progression disease; ICC, inter/intra-class correlation coefficient; LASSO, least absolute shrinkage and selection operator.

TABLE 3-1 | The distribution of the radiomics scores for PsP vs HPD in the training and testing datasets.

RS	Cutoff	Training dataset (N=75)			Testing dataset (N=34)			P value
		PsP	HPD	P value	PsP	HPD	P value	
Intra-tumoral	0.109	-0.56 (-0.78, 0.44)	1.19 (0.13, 2.67)	<0.001*	0.00 (-0.66, 1.08)	1.20 (0.92, 2.04)	0.009*	0.309
Peri-tumoral	0.386	-0.61 (-1.56, 0.43)	1.24 (0.41, 2.52)	<0.001*	-1.66 (-3.80, 0.13)	0.82 (0.07, 2.48)	0.002*	0.278
Combined	-0.298	-0.53 (-0.94, 0.41)	1.34 (0.18, 2.43)	<0.001*	-0.38±0.96	0.88±0.94	0.001*	0.243

TABLE 3-2 | The distribution of the radiomics scores for PsP vs HPD in the training and testing datasets.

RS	Cutoff	Training dataset (N=68)			Testing dataset (N=30)			P value
		PsP	sPD	P value	PsP	sPD	P value	
Intra-tumoral	-0.376	-1.17 (-1.89, -0.46)	2.49 (0.16, 6.15)	<0.001*	-0.69 (-2.27, -0.36)	5.68 (0.26, 7.55)	<0.001*	0.723
Peri-tumoral	0.418	-1.36 (-1.91, 0.02)	3.43 (0.68, 7.28)	<0.001*	-0.49 (-1.45, 0.16)	5.65 (2.87, 8.47)	<0.001*	0.203
Combined	0.541	-1.43 (-2.37, -0.78)	2.56 (0.95, 11.66)	<0.001*	-1.76 (-3.03, -0.59)	3.95 (1.61, 11.87)	<0.001*	0.877

TABLE 3-3 | The distribution of the radiomics scores for HPD vs sPD in the training and testing datasets.

RS	Cutoff	Training dataset (N=85)			Testing dataset (N=38)			P value
		HPD	sPD	P value	HPD	sPD	P value	
Intra-tumoral	0.741	-2.10±2.64	2.20±2.38	<0.001*	-1.25 (-5.95, 0.06)	1.86 (0.63, 3.87)	<0.001*	0.965
Peri-tumoral	-0.154	-1.01 (-15.76, -0.25)	1.37 (0.53, 3.38)	<0.001*	-6.44 (-29.39, 0.00)	1.92 (0.41, 3.28)	<0.001*	0.878
Combined	1.325	-2.84 (-21.49, -0.80)	3.52 (1.75, 5.01)	<0.001*	-18.12 (-26.04, 0.53)	8.14 (3.03, 9.66)	<0.001*	0.345

*RS, radiomics signature; PsP, pseudoprogression; HPD, hyperprogression disease; sPD, standard progression disease; * indicated significant differences; Results for normal and non-normal distributions are means ± standard deviation and quartiles*

The optimal cutoff values in the intra-, peri-tumoral, and combined model are displayed in **Table 3-2**. The calibration curve and decision curve are shown in **Figures S4**, **S5** and **Figures 5**, **6**.

RS for Discriminating HPD From sPD

The RSs derived from the combined region were all significantly higher in patients with sPD than those with HPD in both training and testing datasets (**Figure S2**). According to the maximum Youden's index, 0.741, -0.154, and 1.325 were respectively set as the optimal cutoff values in the intra-, peri-tumoral, and combined models (**Table 3-3**). The combined RS yielded the highest AUC of 0.959 and an ACC of 0.894, which were higher than those of intra- (yielded an AUC of 0.911 and ACC of 0.824) and peri-tumoral (yielded an AUC of 0.894 and ACC of 0.835) models alone. The intra-tumoral RS yielded a higher AUC value than that of peri-tumoral RS in training datasets but lower in testing datasets (**Table 4**, **Figure S3** and **Figure 4**).

DISCUSSION

In this multicenter study, we investigated the ability of pretreatment CT-based RSs extracted from intra- and peri-tumoral regions to predict atypical responses to ICI in multiple solid tumors. Our findings showed that the peri-tumoral regions have additional predictive values relative to the intra-tumoral

TABLE 4 | The discriminative performance of the models in the training and testing datasets.

Radiomics signatures		Training datasets						Testing datasets					
		AUC (95% CI)	ACC	SEN	SPE	PPV	NPV	AUC (95% CI)	ACC	SEN	SPE	PPV	NPV
PsP *VS.* HPD	**Intra-tumoral**	0.804 (0.717, 0.881)	0.733	0.739	0.724	0.810	0.636	0.769 (0.602, 0.913)	0.735	0.857	0.538	0.750	0.700
	Peri-tumoral	0.848 (0.770, 0.918)	0.773	0.783	0.759	0.837	0.688	0.824 (0.688, 0.941)	0.765	0.714	0.846	0.882	0.647
	Combined	0.834 (0.746, 0.914)	0.827	0.935	0.655	0.811	0.864	0.835 (0.704, 0.942)	0.794	0.905	0.615	0.792	0.800
PsP *VS.* sPD	**Intra-tumoral**	0.902 (0.837, 0.957)	0.838	0.872	0.793	0.850	0.821	0.891 (0.788, 0.973)	0.833	0.882	0.769	0.875	0.786
	Peri-tumoral	0.912 (0.846, 0.966)	0.868	0.769	1.000	1.000	0.763	0.900 (0.794, 0.981)	0.833	0.824	0.846	0.833	0.833
	Combined	0.923 (0.865, 0.972)	0.868	0.821	0.931	0.941	0.794	0.919 (0.813, 0.991)	0.867	0.824	0.923	0.933	0.800
HPD *VS.* sPD	**Intra-tumoral**	0.911 (0.857, 0.954)	0.824	0.717	0.949	0.943	0.740	0.891 (0.797, 0.966)	0.763	0.714	0.824	0.833	0.750
	Peri-tumoral	0.894 (0.833, 0.945)	0.835	0.870	0.795	0.833	0.838	0.899 (0.809, 0.969)	0.763	0.810	0.706	0.773	0.700
	Combined	0.959 (0.925, 0.986)	0.894	0.804	1.000	1.000	0.812	0.933 (0.863, 0.985)	0.842	0.857	0.824	0.857	0.824

Psp, pseudo-progression; HPD, hyper-progression disease; sPD, standard progression disease; ROC, receiver operating characteristic; AUC, area under ROC curve; ACC, accuracy; SPE, specificity; SEN, sensitivity; PPV, positive predictive value; NPV, negative predictive value.

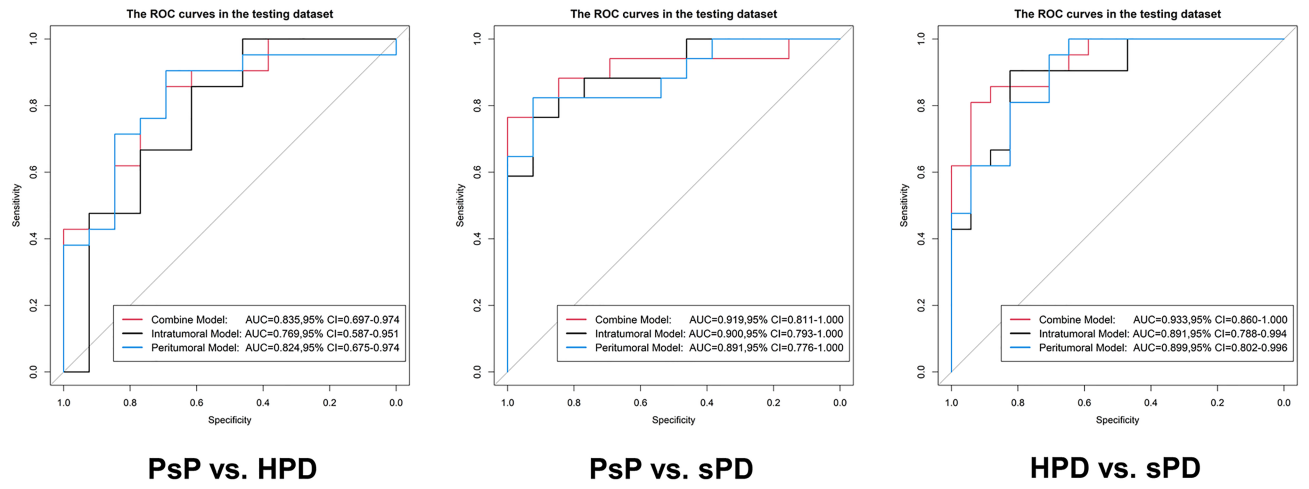


FIGURE 4 | Evaluation of the predictive performance of the radiomics signatures in the testing datasets. In each ROC, the black curve is the ROC of the intratumoral model, the blue curve is the ROC of the peritumoral model, and the red curve is the ROC of the combine model.

regions in different immunotherapy responses, especially for PsP datasets.

As shown above, AUC values ranged from 0.834 to 0.959, whereas ACCs ranged from 0.827 to 0.894 in the combined RSs of the intra- and peri-tumoral regions outperformed than either of them alone. Besides, peri-tumoral RS showed a higher AUC (0.848 *vs*. 0.835) than combined RS in PsP *vs*. HPD, but lower ACC (0.773 *vs*. 0.827). In general, it is clear that the combination of intra- and peri-tumoral regions yielded the overall best classification performance. Noticeably, the most predictive RSs were found to be within an immediate distance of 5 mm from the lesion in predicting PsP when compared with either HPD or sPD, suggesting that features from the peri-tumoral region may have unique power in identifying PsP.

Radiomics is an approach involving a computerized extraction of certain quantitative imaging features, which has shown promise in predicting as well as monitoring treatment response (27). For predicting PsP, our results are in line with the study by Sun et al., who used radiomics from the peri-tumoral region to detect CD8 cells and predict immunotherapy response in multiple datasets (28). Similar results are available in research by Tunali et al.; the authors validated peri-tumoral features highly associated with the tumor-infiltrating lymphocyte (TIL) density on biopsy samples, which may provide a better understanding of the underlying biology (29). More specifically, the immediate surrounding tumor immune microenvironment (TIME) may offer unique information prior to administration of ICI that potentially decodes TILs. However, Shen and his colleagues constructed RS for predicting lymph node

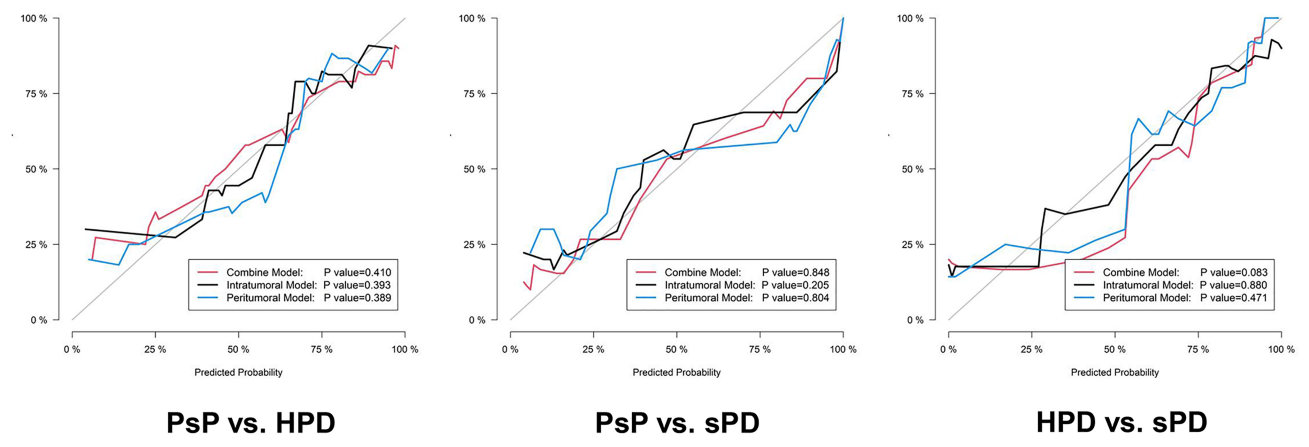


FIGURE 5 | The calibration curves of the proposed radiomics models in the testing datasets. The 45° gray line indicates an ideal prediction. The black, blue, and red lines represent the intratumoral, peritumoral, and combined model predicted results, respectively. The X axis represents the predicted probability, and the Y axis represents true probability. The p value was derived from the Hosmer–Lemeshow test.

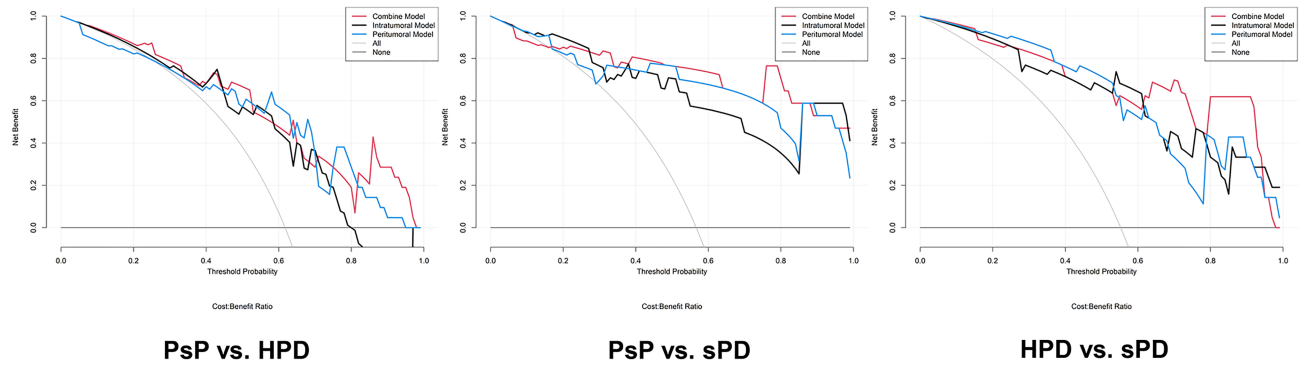


FIGURE 6 | Decision curve analysis for the radiomics signatures in the testing datasets. The result of the decision curve analysis indicated that the prediction of PsP and HPD using the combined RS can give more net benefit than by treating none or all patients in both training and testing datasets.

metastasis of patients with esophageal cancer before surgery from intra- and peri-tumoral areas and found that the former had a better performance in their research (30). In theory, due to the multiregional and microenvironmental heterogeneity in malignant tumors, it is reasonable to speculate that features from the whole tumor and peri-tumoral regions could have a comprehensive understanding of pathophysiology and the best performance in predicting response than those from only single regions (31). Although the biology of PsP and HPD is nevertheless to be understood, many of the present theories hypothesize that various immunoregulatory cells within the TIME could also be liable for this phenomenon (32). Interestingly, there is increasing evidence that TILs could be translated into certain quantitative features from the peri-tumoral region on CT-based radiomics, while further study is still needed (33).

PsP was first described on immunotherapy of the CTLA-4 inhibitor in melanoma, with a patient who experienced an enlargement of a cutaneous lesion after initial treatment, followed by a long-term stability (5). Afterward, PsP was used to describe as clinically improved or stable after a primary disease progression. The incidence of PsP cases observed in our study (8.25%) is consistent with the rate of this phenomenon observed in previous literatures as ranging from 1.1% to 9.1% across multiple solid tumor types (34). However, it should be noted that the incidence of PsP was underestimated in most clinical practices. For instance, PsP can imitate true progression radiographically and may be misclassified as a non-responder then excluded from immunotherapy by an inexperienced physician according to primitive WHO or RECIST criteria. Although iRECIST was proved superior to RECIST1.1 in identifying PsP, it requires an additional 4–8-week reassessment in cases of suspected progression causing an extra cost and a time leg on a potentially ineffective therapy (21). Our results indicate that radiomics can predict PsP who received ICI and may supplement conventional response evaluation criteria.

Actually, PsP is not a real tumor progression. The mechanism behind PsP could be that tumors could have ongoing growth until the activation of effective antitumor immune responses develops (35). Another explanation could be the infiltration of T cells into tumors, leading to a transient increase in tumor burden rather than true proliferation of tumor cells (36). The second hypothesis was later confirmed on tumor biopsies from patients with melanoma experiencing transient progression on a CTLA-4 inhibitor, showing an acute inflammatory reaction with lymphocyte infiltration.

Totally 9 of the 16 radiomics features were extracted from peri-tumoral regions in PsP *vs.* HPD and PsP *vs.* sPD groups that may be capturing data related to the TIME. Of note, the glszm_ZoneEntropy feature has been selected from both intra- and peri-tumoral regions in PsP *vs.* HPD, which was often appeared in tumor grading or staging and differentiation diagnosis (37, 38). Gldm_SmallDependenceEmphasis was used for predicting the genetic mutation status in NSCLC patients (39). Moreover, first-order and shape features quantify the range of gray values in the ROI which reflect the degree of heterogeneity of the tumor (40, 41). In a few clinical trials, PsP has been reported to be more common in younger patients, which may be due to the better reactivity of the immune system and may occur at any time after the start of treatment. However, there was no significant difference in our study, probably due to the limited sample size (42).

Another pattern of atypical responses called HPD was first reported in a case study of NSCLC on nivolumab treatment, with the observation that the patient seems to have an accelerated tumor growth rate after the initiation of ICI (43). The incidence of HPD from 4% to 29% of patients in various cancer types on ICI has been reported in several studies, which is higher than that of PsP (44). Unlike PsP followed subsequently by tumor regression, HPD represents true tumor growth and deserves more attention because patients experiencing HPD have a significantly shorter OS than sPD (3.6 *vs.* 6.2 months), suggesting that HPD has a deleterious effect and that it should be treated as a therapeutic emergency (45). Several mechanisms of HPD such as T cell exhaustion and expansion, aberrant inflammation and oncogenic pathway activation, and modulation of pro-tumorigenic immune subdatasets have been proposed (44). However, some previous reports suggested that being female, advanced age (>65 years), monotherapy, and epidermal growth factor receptor (EGFR) alterations were associated with HPD. However, the reports from different studies sometimes oppose each other (44, 46). We could not find any relationship between clinical variables and HPD in our

study, probably related to the definition of HPD and the heterogeneity of the population enrolled. In contrast to predict PsP, the intra-tumoral RS have a similar performance with peri-tumoral RS, suggesting that the difference in heterogeneity plays a leading role between HPD and sPD, which could be explained by glszm_SizeZoneNonUniformityNormalized and coarseness features (both have higher values indicating more heterogeneity in the imaging). Recently, Tunali et al. reviewed pretreatment contrast-enhanced CT scans and radiomics features of 228 NSCLC patients, and they used parameters derived from both the tumor and tumor border regions to distinguish five HPD patients from non-HPD with an AUC of 0.865 (47). However, it was not capable to be applied because of shortages in cases and model construction.

Nowadays, some approaches are exploratorily used to identify atypical responses to immunotherapy. The first method is liquid biopsies, such as circulating tumor DNA (ctDNA), which was reported to decrease greatly in nine PsP patients with melanoma receiving ICI (47). However, different tumor types or inhibitors are required to further validate the relationship between ctDNA and PsP under immunotherapy. The second method is response evaluation criteria for immunotherapy, such as iRECIST (21). These criteria allow iUPD patients to continue treatment and reevaluate their responses with a time lag (after 4–8 weeks) or unneeded immune-related adverse effects (irAE). The third method is tumor genomic biomarkers, such as MDM2/MDM4 amplifications and EGFR alterations (48). Kato et al. performed next-generation sequencing of four patients with HPD revealed MDM2/MDM4 amplifications in 2 patients and EGFR amplification in one patient. In contrast, Kim et al. found no MDM2/MDM42 amplifications in the 18 patients with HPD (49). Unlike the above noninvasive methods, biopsies of tumors from some patients suspected of PsP have been found to contain dense inflammatory infiltrates or necrosis, instead of increased malignant load. Despite this, biopsies are often limited by a relatively small tissue sample and spatial heterogeneity and carry a procedure risk.

PsP and HPD are both atypical responses to immunotherapy, and oncologists should be conscientious to not confuse them with sPD, so as to avoid changing treatment too early for PsP, or too late in case of (hyper-)progressive disease. To this end, we aimed to use CT imaging, since it is routinely, noninvasive, and informative of the entire tumor burden and can be performed serially. To our knowledge, this is the first study to explore the ability of RS for prediction of atypical responses to ICI. Our findings potentially hold significant clinical applications, because they could provide a clinical framework for the pretreatment identification of atypical responses.

There are several limitations in our study. First is the retrospective nature and relatively small number of atypical responses, which limited the ability to perform stratified analyses, such as unitary tumor. However, we aim to test the true generalization performance of the classifier across multitumor interspecifics. Second, we only explored the radiomics features from pretreatment CT imaging; perhaps different or more information would be obtained with CT evaluations after ICI, although as claimed by the principle of "first, do no harm." Third, lack of LDH, PD-L1 expression, and tumor mutation burden (TMB) in most patients in these retrospective datasets limit the assessment of prediction values. Fourth, the peri-tumoral features were only extracted within an immediate distance of 5 mm from the lesion. It is unclear whether other distances perform a superior prediction ability. Overall, validation of these radiomics biomarkers still needs to be done on larger multisite datasets.

CONCLUSION

In conclusion, atypical responses to ICI are not uncommon phenomena observed with the incidence of 8.25% in PsP and 10.44% in HPD. The present preliminary study suggested that pretreatment CT-based radiomics provided a potential tool to differentiate among PsP, HPD, and sPD, thereby providing possibility for the prediction of atypical responses to ICI. In addition, RS derived from the peri-tumoral outperformed intra-tumoral region in identifying PsP, and the combined RS outperformed those from either intra- or peri-tumoral alone which may provide a more comprehensive characterization of atypical responses to ICI.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

This retrospective study was accepted by our institutional review committee (LNCH-20200611), and the prerequisite for obtaining informed consent was waived. The ethics committee waived the requirement of written informed consent for participation.

AUTHOR CONTRIBUTIONS

SH, TY, and YHL contributed to the conception and design of the study. All authors organized the database. LHW, SH, and YG performed the statistical analysis. YF and QL wrote the first draft of the manuscript. SH, TY, and YHL reviewed the manuscript. All authors contributed to the article and approved the submitted version. SH, YF, and QL have contributed equally to this work. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the Wu Jieping Medical Foundation (No. 320.6750.2020-08-24).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.729371/full#supplementary-material

REFERENCES

1. Carretero-González A, Lora D, Ghanem I, Zugazagoitia J, Castellano D, Sepúlveda JM, et al. Analysis of Response Rate With ANTI PD1/PD-L1 Monoclonal Antibodies in Advanced Solid Tumors: A Meta-Analysis of Randomized Clinical Trials. *Oncotarget* (2018) 9(9):8706–15. doi: 10.18632/oncotarget.24283
2. Arasanz H, Gato-Cañas M, Zuazo M, Ibañez-Vea M, Breckpot K, Kochan G, et al. PD1 Signal Transduction Pathways in T Cells. *Oncotarget* (2017) 8 (31):51936–45. doi: 10.18632/oncotarget.17232
3. Nallasamy P, Chava S, Verma SS, Mishra S, Gorantla S, Coulter DW, et al. PD-L1, Inflammation, non-Coding RNAs, and Neuroblastoma: Immuno-Oncology Perspective. *Semin Cancer Biol* (2018) 52(Pt 2):53–65. doi: 10.1016/j.semcancer.2017.11.009
4. Tazdait M, Mezquita L, Lahmar J, Ferrara R, Bidault F, Ammari S, et al. Patterns of Responses in Metastatic NSCLC During PD-1 or PDL-1 Inhibitor Therapy: Comparison of RECIST 1.1, irRECIST and iRECIST Criteria. *Eur J Cancer* (2018) 88:38–47. doi: 10.1016/j.ejca.2017.10.017
5. Ferrara R, Matos I. Atypical Patterns of Response and Progression in the Era of Immunotherapy Combinations. *Future Oncol* (2020) 16(23):1707–13. doi: 10.2217/fon-2020-0186
6. Lyu N, Kong Y, Li X, Mu L, Deng H, Chen H, et al. Ablation Reboots the Response in Advanced Hepatocellular Carcinoma With Stable or Atypical Response During PD-1 Therapy: A Proof-Of-Concept Study. *Front Oncol* (2020) 10:580241. doi: 10.3389/fonc.2020.580241
7. Queirolo P, Spagnolo F. Atypical Responses in Patients With Advanced Melanoma, Lung Cancer, Renal-Cell Carcinoma and Other Solid Tumors Treated With Anti-PD-1 Drugs: A Systematic Review. *Cancer Treat Rev* (2017) 59:71–8. doi: 10.1016/j.ctrv.2017.07.002
8. Frelaut M, du Rusquec P, de Moura A, Le Tourneau C, Borcoman E. Pseudoprogression and Hyperprogression as New Forms of Response to Immunotherapy. *BioDrugs* (2020) 34(4):463–76. doi: 10.1007/s40259-020-00425-y
9. Ferrara R, Caramella C, Besse B, Champiat S. Pseudoprogression in Non-Small Cell Lung Cancer Upon Immunotherapy: Few Drops in the Ocean? *J Thorac Oncol* (2019) 14(3):328–31. doi: 10.1016/j.jtho.2018.12.011
10. Wang Q, Gao J, Wu X. Pseudoprogression and Hyperprogression After Checkpoint Blockade. *Int Immunopharmacol* (2018) 58:125–35. doi: 10.1016/j.intimp.2018.03.018
11. Frelaut M, Le Tourneau C, Borcoman E. Hyperprogression Under Immunotherapy. *Int J Mol Sci* (2019) 20(11):2674. doi: 10.3390/ijms20112674
12. Adashek JJ, Kato S, Ferrara R, Lo Russo G, Kurzrock R. Hyperprogression and Immune Checkpoint Inhibitors: Hype or Progress? *Oncologist* (2020) 25 (2):94–8. doi: 10.1634/theoncologist.2019-0636
13. Groisberg R, Hong DS, Behrang A, Hess K, Janku F, Piha-Paul S, et al. Characteristics and Outcomes of Patients With Advanced Sarcoma Enrolled in Early Phase Immunotherapy Trials. *J Immunother Cancer* (2017) 5(1):100. doi: 10.1186/s40425-017-0301-y
14. Katz SI, Hammer M, Bagley SJ, Aggarwal C, Bauml JM, Thompson JC, et al. Radiologic Pseudoprogression During Anti-PD-1 Therapy for Advanced Non-Small Cell Lung Cancer. *J Thorac Oncol* (2018) 13(7):978–86. doi: 10.1016/j.jtho.2018.04.010
15. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to Radiomics. *J Nucl Med* (2020) 61(4):488–95. doi: 10.2967/jnumed.118.222893
16. E L, Lu L, Li L, Yang H, Schwartz LH, Zhao B. Radiomics for Classifying Histological Subtypes of Lung Cancer Based on Multiphasic Contrast-Enhanced Computed Tomography. *J Comput Assist Tomogr* (2019) 43 (2):300–6. doi: 10.1097/RCT.0000000000000836
17. Digumarthy SR, Padole AM, Gullo RL, Sequist LV, Kalra MK. Can CT Radiomic Analysis in NSCLC Predict Histology and EGFR Mutation Status? *Medicine (Baltimore)* (2019) 98(1):e13963. doi: 10.1097/MD.0000000000013963
18. Yang X, Pan X, Liu H, Gao D, He J, Liang W, et al. A New Approach to Predict Lymph Node Metastasis in Solid Lung Adenocarcinoma: A Radiomics Nomogram. *J Thorac Dis* (2018) 10(Suppl 7):S807–19. doi: 10.21037/jtd.2018.03.126
19. Sun Z, Hu S, Ge Y, Wang J, Duan S, Song J, et al. Radiomics Study for Predicting the Expression of PD-L1 in non-Small Cell Lung Cancer Based on CT Images and Clinicopathologic Features. *J Xray Sci Technol* (2020) 28 (3):449–59. doi: 10.3233/XST-200642
20. Wang ZL, Mao LL, Zhou ZG, Si L, Zhu HT, Chen X, et al. Pilot Study of CT-Based Radiomics Model for Early Evaluation of Response to Immunotherapy in Patients With Metastatic Melanoma. *Front Oncol* (2020) 10:1524. doi: 10.3389/fonc.2020.01524
21. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, et al. iRECIST: Guidelines for Response Criteria for Use in Trials Testing Immunotherapeutics [Published Correction Appears in Lancet Oncol. 2019 May;20(5):E242]. *Lancet Oncol* (2017) 18(3):e143–52. doi: 10.1016/S1470-2045(17)30074-8
22. Persigehl T, Lennartz S, Schwartz LH. iRECIST: How to do it. *Cancer Imaging* (2020) 20(1):2. doi: 10.1186/s40644-019-0281-x
23. Champiat S, Ferrara R, Massard C, Besse B, Marabelle A, Soria JC, et al. Hyperprogressive Disease: Recognizing a Novel Pattern to Improve Patient Management. *Nat Rev Clin Oncol* (2018) 15(12):748–62. doi: 10.1038/s41571-018-0111-2
24. Gomes da Morais AL, de Miguel M, Cardenas JM, Calvo E. Comparison of Radiological Criteria for Hyperprogressive Disease in Response to Immunotherapy. *Cancer Treat Rev* (2020) 91:102116. doi: 10.1016/j.ctrv.2020.102116
25. Skjelbred HI, Kong J. A Comparison of Linear Interpolation and Spline Interpolation for Turbine Efficiency Curves in Short-Term Hydropower Scheduling Problems. *IOP Conf Ser Earth Environ Sci* (2019) 240:042011. doi: 10.1088/1755-1315/240/4/042011
26. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-Based Phenotyping. *Radiology* (2020) 295(2):328–38. doi: 10.1148/radiol.2020191145
27. Avanzo M, Stancanello J, El Naqa I. Beyond Imaging: The Promise of Radiomics. *Phys Med* (2017) 38:122–39. doi: 10.1016/j.ejmp.2017.05.071
28. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, 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. *Lancet Oncol* (2018) 19(9):1180–91. doi: 10.1016/S1470-2045(18)30413-3
29. Tunali I, Gray JE, Qi J, Abdalah M, Jeong DK, Guvenis A, et al. Novel Clinical and Radiomic Predictors of Rapid Disease Progression Phenotypes Among Lung Cancer Patients Treated With Immunotherapy: An Early Report. *Lung Cancer* (2019) 129:75–9. doi: 10.1016/j.lungcan.2019.01.010
30. Shen C, Liu Z, Wang Z, Guo J, Zhang H, Wang Y, et al. Building CT Radiomics Based Nomogram for Preoperative Esophageal Cancer Patients Lymph Node Metastasis Prediction. *Transl Oncol* (2018) 11(3):815–24. doi: 10.1016/j.tranon.2018.04.005
31. Pak KH, Jo A, Choi HJ, Choi Y, Kim H, Cheong JH. The Different Role of Intratumoral and Peritumoral Lymphangiogenesis in Gastric Cancer Progression and Prognosis. *BMC Cancer* (2015) 15:498. doi: 10.1186/s12885-015-1501-9
32. Belli C, Trapani D, Viale G, D'Amico P, Duso BA, Della Vigna P, et al. Targeting the Microenvironment in Solid Tumors. *Cancer Treat Rev* (2018) 65:22–32. doi: 10.1016/j.ctrv.2018.02.004
33. Devkota L, Starosolski Z, Rivas CH, Stupin I, Annapragada A, Ghaghada KB, et al. Detection of Response to Tumor Microenvironment-Targeted Cellular Immunotherapy Using Nano-Radiomics. *Sci Adv* (2020) 6(28):eaba6156. doi: 10.1126/sciadv.aba6156
34. Michalarea V, Fontana E, Garces AI, Williams A, Smyth EC, Picchia S, et al. Pseudoprogression on Treatment With Immune-Checkpoint Inhibitors in Patients With Gastrointestinal Malignancies: Case Series and Short Literature Review. *Curr Probl Cancer* (2019) 43(5):487–94. doi: 10.1016/j.currproblcancer.2019.01.008
35. Foller S, Oppel-Heuchel H, Grimm MO. Tumorassessment Bei Immun-Checkpoint-Inhibitor-Therapie : Tumoransprechen, Progression Und Pseudoprogression [Tumor Assessment in Immune Checkpoint Inhibitor Therapy : Tumor Response, Progression and Pseudoprogression]. *Urologe A* (2018) 57(11):1316–25. doi: 10.1007/s00120-018-0788-y
36. Wei SC, Levine JH, Cogdill AP, Zhao Y, Anang NAS, Andrews MC, et al. Distinct Cellular Mechanisms Underlie Anti-CTLA-4 and Anti-PD-1 Checkpoint Blockade. *Cell* (2017) 170(6):1120–33.e17. doi: 10.1016/j.cell.2017.07.024
37. Tagliafico AS, Cea M, Rossi F, Valdora F, Bignotti B, Succio G, et al. Differentiating Diffuse From Focal Pattern on Computed Tomography in

Multiple Myeloma: Added Value of a Radiomics Approach. *Eur J Radiol* (2019) 121:108739. doi: 10.1016/j.ejrad.2019.108739
38. Kocak B, Durmaz ES, Ates E, Kaya OK, Kilickesmez O. Unenhanced CT Texture Analysis of Clear Cell Renal Cell Carcinomas: A Machine Learning-Based Study for Predicting Histopathologic Nuclear Grade. *AJR Am J Roentgenol* (2019) 202(6):1077–8. doi: 10.2214/AJR.18.20742
39. Rossi G, Barabino E, Fedeli A, Ficarra G, Coco A, Russo S, et al. Radiomic Detection of EGFR Mutations in NSCLC. *Cancer Res* (2021) 81(3):724–31. doi: 10.1158/0008-5472.CAN-20-0999
40. Ji GW, Zhang YD, Zhang H, Zhu FP, Wang K, Xia YX, et al. Biliary Tract Cancer at CT: A Radiomics-Based Model to Predict Lymph Node Metastasis and Survival Outcomes. *Radiology* (2019) 290(1):90–8. doi: 10.1148/radiol.2018181408
41. Desseroit MC, Tixier F, Weber WA, Siegel BA, Cheze Le Rest C, Visvikis D, et al. Reliability of PET/CT Shape and Heterogeneity Features in Functional and Morphologic Components of Non-Small Cell Lung Cancer Tumors: A Repeatability Analysis in a Prospective Multicenter Cohort. *J Nucl Med* (2017) 58(3):406–11. doi: 10.2967/jnumed.116.180919
42. Martin-Romano P, Castanon E, Ammari S, Champiat S, Hollebecque A, Postel-Vinay S, et al. Evidence of Pseudoprogression in Patients Treated With PD1/PDL1 Antibodies Across Tumor Types. *Cancer Med* (2020) 9(8):2643–52. doi: 10.1002/cam4.2797
43. Petrioli R, Mazzei MA, Giorgi S, Cesqui E, Gentili F, Francini G, et al. Hyperprogressive Disease in Advanced Cancer Patients Treated With Nivolumab: A Case Series Study. *Anticancer Drugs* (2020) 31(2):190–5. doi: 10.1097/CAD.0000000000000864
44. Zhou L, Zhang M, Li R, Xue J, Lu Y. Pseudoprogression and Hyperprogression in Lung Cancer: A Comprehensive Review of Literature. *J Cancer Res Clin Oncol* (2020) 146(12):3269–79. doi: 10.1007/s00432-020-03360-1
45. Ferrara R, Mezquita L, Texier M, Lahmar J, Audigier-Valette C, Tessonnier L, et al. Hyperprogressive Disease in Patients With Advanced Non-Small Cell Lung Cancer Treated With PD-1/PD-L1 Inhibitors or With Single-Agent Chemotherapy. *JAMA Oncol* (2018) 4(11):1543–52. doi: 10.1001/jamaoncol.2018.3676
46. Zang H, Peng J, Zheng H, Fan S. Hyperprogression After Immune-Checkpoint Inhibitor Treatment: Characteristics and Hypotheses. *Front Oncol* (2020) 10:515. doi: 10.3389/fonc.2020.00515
47. Lee JH, Long GV, Menzies AM, Lo S, Guminski A, Whitbourne K, et al. Association Between Circulating Tumor DNA and Pseudoprogression in Patients With Metastatic Melanoma Treated With Anti-Programmed Cell Death 1 Antibodies. *JAMA Oncol* (2018) 4(5):717–21. doi: 10.1001/jamaoncol.2017.5332
48. Adashek JJ, Subbiah IM, Matos I, Garralda E, Menta AK, Ganeshan DM, et al. Hyperprogression and Immunotherapy: Fact, Fiction, or Alternative Fact? *Trends Cancer* (2020) 6(3):181–91. doi: 10.1016/j.trecan.2020.01.005
49. Forschner A, Hilke FJ, Bonzheim I, Gschwind A, Demidov G, Amaral T, et al. MDM2, MDM4 and EGFR Amplifications and Hyperprogression in Metastatic Acral and Mucosal Melanoma. *Cancers (Basel)* (2020) 12(3):540. doi: 10.3390/cancers12030540

Conflict of Interest: Author YG was employed by company GE Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 He, Feng, Lin, Wang, Wei, Tong, Zhang, Liu, Ye, Guo, Yu and Luo. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 02 November 2021
doi: 10.3389/fonc.2021.721318

Identification of Stage IIIC/IV EGFR-Mutated Non-Small Cell Lung Cancer Populations Sensitive to Targeted Therapy Based on a PET/CT Radiomics Risk Model

Dan Shao[1†], Dongyang Du[2†], Haiping Liu[3†], Jieqin Lv[2], You Cheng[1], Hao Zhang[2], Wenbing Lv[2], Shuxia Wang[1] and Lijun Lu[2*]*

[1] Department of Positron Emission Tomography (PET) Center, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, [2] School of Biomedical Engineering and Guangdong Provincial Key Laboratory of Medical Image Processing, Southern Medical University, Guangzhou, China, [3] Department of Positron Emission Tomography/Computed Tomography (PET/CT) Center, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

OPEN ACCESS

Edited by:
Omar Sultan Al-Kadi,
The University of Jordan, Jordan

Reviewed by:
Jinman Kim,
The University of Sydney, Australia
Francesco Bianconi,
University of Perugia, Italy

***Correspondence:**
Shuxia Wang
wang_shuxia@163.com
Lijun Lu
ljlubme@gmail.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *06 June 2021*
Accepted: *11 October 2021*
Published: *02 November 2021*

Citation:
Shao D, Du D, Liu H, Lv J, Cheng Y, Zhang H, Lv W, Wang S and Lu L (2021) Identification of Stage IIIC/IV EGFR-Mutated Non-Small Cell Lung Cancer Populations Sensitive to Targeted Therapy Based on a PET/CT Radiomics Risk Model. Front. Oncol. 11:721318. doi: 10.3389/fonc.2021.721318

Objectives: This project aimed to construct an individualized PET/CT prognostic biomarker to accurately quantify the progression risk of patients with stage IIIC-IV epidermal growth factor receptor (EGFR)-mutated Non-small cell lung cancer (NSCLC) after first-line first and second generation EGFR- tyrosine kinase inhibitor (TKI) drug therapy and identify the first and second generation EGFR-TKI treatment-sensitive population.

Methods: A total of 250 patients with stage IIIC-IV EGFR-mutated NSCLC underwent first-line first and second generation EGFR-TKI drug therapy were included from two institutions (140 patients in training cohort; 60 patients in internal validation cohort, and 50 patients in external validation cohort). 1037 3D radiomics features were extracted to quantify the phenotypic characteristics of the tumor region in PET and CT images, respectively. A four-step feature selection method was performed to enable derivation of stable and effective signature in the training cohort. According to the median value of radiomics signature score (Rad-score), patients were divided into low- and high-risk groups. The progression-free survival (PFS) behaviors of the two subgroups were compared by Kaplan–Meier survival analysis.

Results: Our results shown that higher Rad-scores were significantly associated with worse PFS in the training ($p < 0.0001$), internal validation ($p = 0.0153$), and external validation ($p = 0.0006$) cohorts. Rad-score can effectively identify patients with a high risk of rapid progression. The Kaplan–Meier survival curves of the three cohorts present significant differences in PFS between the stratified slow and rapid progression subgroups.

Conclusion: The PET/CT-derived Rad-score can realize the precise quantitative stratification of progression risk after first-line first and second generation EGFR-TKI drug therapy for NSCLC and identify EGFR-mutated NSCLC populations sensitive to targeted therapy, which might help to provide precise treatment options for NSCLC.

Keywords: lung cancer, PET/CT, radiomics, stratification of progression risk, progression-free survival

INTRODUCTION

Non-small cell lung cancer (NSCLC) is the leading cause of cancer death worldwide (1). The onset of NSCLC is insidious, and most patients present at an advanced stage. For advanced NSCLC, platinum-based combination chemotherapy has reached a bottleneck, with suboptimal efficacy. In recent years, targeted therapy has achieved remarkable results in cancer treatment, showing great promise in the treatment of a variety of advanced tumors. In patients with NSCLC, especially in East Asian populations, the mutation rate of human epidermal growth factor receptor (EGFR) is as high as 40%~55% (2). Therefore, researchers have developed small-molecule tyrosine kinase inhibitor (TKI) drugs that target EGFR in EGFR gene mutation therapy, which have subsequently been found to prolong progression-free survival (PFS) in such patients when compared with chemotherapy in clinical application (3), extend the median PFS of patients to 9~11 months (4).

However, not all patients can benefit from them. In the clinical application, it is found that for a large part of NSCLC patients with Phase IIIC/IV EGFR mutation after the first-line EGFR-TKI treatment the disease progressed rapidly and they could not benefit from the treatment. The third generation of TKI Osimertinib can simultaneously suppress the EGFR-TKI-sensitive gene and T790M, the drug resistance gene of the first and second generation of TKIs, and extend the median PFS of patients. However, not all patients need the treatment of Osimertinib. For some patients, the treatment effect of the first and second generation of TKIs is very good (5). Therefore, how to screen the patients suitable for the treatment of the first and second generation of TKIs is particularly important. The progress of patients receiving the clinical treatment after the treatment of the first and second generation of EGFR-TKIs was accurately quantified and evaluated, thus recognizing the population not sensitive to the treatment, guiding the clinical decision, making the corresponding treatment and follow-up plan, and improving prognosis. It has an important clinical value.

In 2012, Lambin et al. (6) first proposed the concept of radiomics, that is, the high-throughput extraction of image information from medical images and deep-seated mining and analysis of massive image data to reveal the pathophysiological characteristics of diseases (such as tumor heterogeneity) and provide the most accurate decision support for disease diagnosis, prognosis prediction and precision treatment (7–10). In the past, by using the technical means of radiomics in medical imaging, such as computed tomography (CT), magnetic resonance imaging (MRI), ultrasound and positron emission tomography (PET), multiple research teams have revealed the associations between microscopic image information and tumor genotyping, tumor treatment efficacy and prognosis through deep-seated information mining and analysis (11).

^{18}F-fluordeoxyglucos (^{18}F-FDG) PET/CT radiomics analysis can extract additional layers of information on tumor heterogeneity, such as biological metabolism information and anatomical characteristic (12). Tumor cell glucose metabolism produces different genetic variations during growth and treatment, so the glucose metabolism at different sites in the same tumor lesion shows significant heterogeneity (13). ^{18}F-FDG PET, which can reflect the glucose metabolic intensity information of tumor lesions, has played an important role in clinical application for NSCLC (14–17). PET/CT radiomics features can be used as an individualized prognostic biomarker for patients with NSCLC (18–20). Kang F et al. successfully applied the radiomics in the diagnosis and differential diagnosis of lung cancer to reduce the false negative rate (21). Zhang J et al. showed that PET/CT radiomics model could recognize the mutation type of EGFR (22).

In recent years, PET/CT radiomics has been used to quantify the risk of progression of patients with lung cancer after treatment, gastric cancer, and glioblastoma (14–16). However, few studies based on PET/CT radiomics to accurately predict the risk of progression after first and second generation of EGFR-TKI therapy in stage IIIC/IV EGFR-mutated NSCLC patients (23). To address the clinical challenges that the progression risk of patients with stage IIIC-IV EGFR-mutated NSCLC is difficult to accurately quantify and stratify after first-line first and second generation of EGFR-TKI targeted drug therapy, we aimed to construct an individualized prognostic biomarker to accurately quantify the progression risk of patients with stage IIIC-IV EGFR-mutated NSCLC after first-line first and second generation of EGFR-TKI drug therapy and identify the first and second generation of EGFR-TKI treatment-sensitive population, which is expected to aid clinicians in treatment decision-making.

MATERIALS AND METHODS

This retrospective, two-center study was approved by our Institutional Review Board. The requirement for written consent was waived by the board.

Patients

This 2-center retrospective study was conducted jointly by two independent departments. This study was conducted in accordance with the Declaration of Helsinki. Our Institutional

Review Board approved this retrospective study and waived the requirement for informed consent from the patients. Before treatment, all patients successively underwent clinical examination, PET/CT scan, blood tests, and pathological examination. The inclusion criteria were as follows: 1. age 18 and older, stage IIIC-IV NSCLC according to the TNM classification system of the American Joint Committee on Cancer; 2. EGFR mutation; 3. TKI group patients were treated with first-line first and second generation EGFR-TKI according to the criteria established by the National Comprehensive Cancer Network (NCCN) until disease progression, with doses appropriately reduced if severe adverse events occurred; and 4. pretherapy PET/CT was acquired two weeks before the initiation of EGFR-TKI therapy. The exclusion criteria were as follows: 1. patients with a history of a second neoplasm; 2. patients with a history of anticancer therapy or surgical therapy; and 3. HIV-positive patients. The clinicopathological information included epidemiological information and characteristics related to the risk of progression after treatment, including age, sex, and smoking history, pathological type, TNM stage, performance status (PS) score, EGFR gene mutation type, carcinoembryonic antigen (CEA) value, NSCLC-associated antigen value (CYFRA21-1), and the presence or absence of brain/bone/liver/lung/pleural/adrenal metastases. All of the clinicopathological information were complete record for all eligible patients.

In total, 200 patients were recruited from July 2007 to July 2019 from center 1. These patients were randomized into training (N = 140) and internal validation (N = 60) cohorts. Using the same inclusion criteria, 50 patients who initiated TKI-therapy between June 2008 and June 2018 from center 2 was subsequently accrued. This was used as external validation cohort (N = 50). The follow-up interval was 4-6 weeks, and the examinations included clinical physical examination, routine laboratory tests and chest CT or PET/CT. The follow-up duration of this project was 2 years. If no endpoint event (disease progression or death) occurred, the follow-up duration should be at least 3 years. PFS was considered the time from the initiation of EGFR-TKI therapy to the date of confirmed disease progression or death. PFS was censored at the date of death from other causes or the date of the last follow-up visit for progression-free patients.

PET/CT Scans

PET/CT scans of the training and internal validation cohorts were performed in center 1 using a Sensation Biograph Somatom 16 HR PET/CT machine (SIEMENS, Germany). Scans of external validation cohort were acquired in center 2 on eight-section PET/CT scanner (Discovery ST 8; GE Healthcare, Wisconsin, USA). All patients fasted for at least 6 hours before the PET/CT scan. Only patients with blood glucose levels between 72.0 and 144.0 mg/dL (4.0 - 8.0 mmol/L) were subjected to PET/CT scan. The patients were instructed to lie still in a quiet room for 60 ± 5 minutes after they received an intravenous injection of 0.1-0.2 mCi/kg (3.7-7.4 MBq/kg) of ^{18}F-FDG.

Trunk PET scans were performed from the upper thigh to the pharynx nasalis immediately after completion of the CT scans. After a skull CT scan, a 5-minute skull PET scan in one bed position was performed from the foramen magnum to the top of the skull. PET scans in center 1 and center 2 were respectively performed using a three-dimensional model with a matrix of 128×128 voxels and a two-dimensional model with a matrix of 128×128 voxels.

For CT reconstruction, the raw data were converted into images using two-dimensional fast Fourier transform. For three-dimensional reconstruction, CT data were converted by the digital model and prepared for attenuation correction of PET images by using voxel space overlay and interpolation. PET images were reconstructed using the iterative ordered-subset expectation maximization (OSEM) method (4 iteration times and 8 character sets) with scatter correction. Image fusion was completed, producing PET, CT, and PET/CT fusion images in the transverse, sagittal, and coronal planes.

Tumor Segmentation

The contour of primary lung tumor (also called region of interest, ROI) on PET and CT images was first delineated slice-by-slice by a nuclear medicine physician with 10 years of experience in thoracic oncology using ITK-SNAP software (version 3.6.0; www.itksnap.org). To evaluate the reproducibility of feature extraction to different contours of segmentation, the delineated ROIs were perturbed by supervoxel–based contour randomization (24) to produce perturbed ROI in 60 randomly selected patients from training cohort.

Radiomics Feature Extraction

Both PET and CT radiomics features were extracted using the PyRadiomics package (version 3.0; https://github.com/Radiomics/pyradiomics), an open-source platform for easy and reproducible radiomics feature extraction (25). 14 volumetric shape were separately extracted from PET and CT segmentation masks; 18 first-order statistical and 75 texture-matrix radiomics features were extracted from the original and 8 wavelet- and 2 Laplacian of Gaussian (LoG)-filtering-derived PET and CT images, respectively. Ultimately, 2074 = 2 modalities × [14 shape + (18 intensity+ 75 texture) × (1 + 8+2 images)] 3D radiomics features were extracted to quantify the phenotypic characteristics of the tumor region in PET and CT images. The parameter settings of PET/CT image preprocessing and the generation of derived images for customizing the PyRadiomics feature extraction are described in **Supplemental Table S1**. The comprehensive radiomics features list is described in **Supplemental Table S2**. To correct for differences in features caused by the different centers (i.e., different scanners), we used the ComBat compensation method (https://github.com/Jfortin1/ComBatHarmonization), which identified a center-specific transformation to express the feature data in a common space devoid of center effect (26–28). To further ensure that the range of features was relatively uniform, we firstly normalized the training data into z-scores, then the mean and standard deviations of training cohort were used to normalize the feature values of the internal and external validation cohorts.

Radiomics Feature Selection, RAD-Score Building, and Prognostic Modeling

In order to avoid overfitting due to the over-abundance of features relative to sample size, we performed a four-step

feature selection method to enable derivation of stable and effective signature in the training cohort. First, the two-way random effects, absolute agreement, single rater/measurement intra-class correlation coefficient (ICC) (29) was calculated for each feature to evaluate the robustness of feature extraction to different contours of segmentation. Features with ICC greater than 0.8 were considered as robust features and retained (30). Secondly, Pearson correlation analysis was used to assess the correlation between all remaining features. For each feature pair with absolute correlation coefficient > 0.8, the feature that yielded higher absolute column-wise correlation mean was eliminated, which tends to provide redundant information about tumor phenotype. Thirdly, univariate Cox analysis was conducted by performing 10-fold cross validation in the training cohort, and Harrell's concordance index (C-index) was used to measure the prognostic performance. Features were sorted in descending order of the mean validation C-index (in the 10 validation rounds), and the top 10 features were selected. Finally, multivariate Cox analysis that considered all possible combinations of candidate features (in sets of 2 up to 10 features) was conducted by performing 10-fold cross validation in the training cohort, and the feature combination with the highest average validation C-index (in the 10-fold cross validation) and significantly associated with PFS was identified as prognostic signature. The signature was used to refit the final Cox model (Radiomics model) in the whole training cohort, and the radiomics signature score (Rad-score) for each patient was calculated based on the selected signature.

For clinicopathological parameters, univariate Cox analysis was performed to assess the association with PFS, and only significant predictors ($p < 0.05$; p values were corrected for false discovery rate (FDR)) after multiple testing correction with the Benjamini-Hochgerg method (31) in the training cohort, were kept to build multivariable Cox model (Clinicopathological model). To further assess whether the clinicopathological findings can improve the performance of the Rad-score, statistically significant clinicopathological parameters were combined with the Rad-score *via* forward stepwise feature selection using maximum log-likelihood criterion as the stopping rule (32), to build a combined multivariable Cox model (Combined model).

Validation of Prognostic Model for PFS Prediction

Time-dependent receiver operating characteristic (ROC) analyses were performed to estimate the probability of each model in predicting 10-month, one-year and 14-month PFS. C-index and area under the ROC curve (AUC) were quantified to evaluate the prognostic accuracy in the training cohort and another two independent validation cohorts. The prognostic score was generated for multivariable Cox model by using a linear combination of selected features weighted by their respective coefficients. Patients were stratified into slow and rapid-progression subgroups by the median value of prognostic score as computed in training cohort; log-rank test was then used to compare the significant difference between the two Kaplan–Meier curves. The same median value of prognostic score was applied to the two independent validation cohorts to perform risk stratification. All patients who received EGFR-TKI therapy were therefore stratified into slow and rapid-progression subgroups. The PFS behaviors of the two subgroups were compared by Kaplan–Meier survival analysis.

Statistical Analysis

Statistical analysis was performed with MATLAB R2015b, Statistical Program for Social Science (SPSS; version 22.0) and R software (version 3.4.4; http://www.Rproject.org). A two-sided p value < 0.05 was used as the criterion to indicate a statistically significant difference.

RESULTS

Patient Characteristics

The baseline characteristics of the three patient cohorts with stage IIIC-IV EGFR-mutant NSCLC from two institutions are summarized in **Table 1**. 250 patients received first and second generation EGFR-TKI therapy (140 patients, 60 patients, and 50 patients in three cohorts), and 243 of the 250 (97.2%) patients suffered NSCLC progression during the follow-up period. The mean PFS for the training, internal validation, and external validation cohorts was 9.82 ± 6.86 months (range 0.3-35 months), 12.14 ± 8.47 months (range 1-40.5 months), and 15.20 ± 11.49 months (range 2.1-58.5 months), respectively.

PFS Prediction Performance of the RAD-Score and the Combined Model

A three-feature signature was built by four-step feature selection in the training cohort (**Supplemental Tables S3**, **S4**), and it was further used for the construction of multivariable radiomics model (**Supplemental Table S5**), and the Rad-score was calculated for each patient. Rad-score = 0.242 × PET_wavelet_LHL_First-order 90th percentile + 0.214 × PET_wavelet_LHL_GLDM Large Dependence High Gray Level Emphasis + 0.199 × CT_wavelet_HHH_First-order Kurtosis.

The C-index of the radiomics model (Rad-score) for PFS prediction was 0.65 (95% confidence interval (CI): 0.60-0.70) for the training cohort and 0.61 (95% CI: 0.53-0.68) and 0.60 (95% CI: 0.52-0.68) for the internal and external validation cohorts, respectively (**Table 2**). The AUC for the radiomics model (Rad-score) ranged from 0.58 to 0.72, 0.64 to 0.74, and 0.67 to 0.79 for 10-month, one-year, and 14-month PFS probability prediction in the three cohorts, respectively (**Table 2** and **Figure 1A**).

In univariate Cox proportional hazards regression analysis, after FDR correction, only N stage was found to be significantly associated with PFS in the training cohort. The corresponding C-index, p values and hazard ratios (HRs) with 95% CIs are detailed in **Supplemental Table S6**. The C-index of the clinical model (N stage) was 0.59 (95% CI: 0.54-64), 0.56 (95% CI: 0.47-0.64), and 0.56 (95% CI: 0.48-0.64) for training, internal and external validation cohorts, respectively. The AUC of clinical model for 10-month, one-year, 14-month PFS probability prediction ranged from 0.62-0.67, 0.57-0.71. and 0.59-0.62 in the three cohorts, respectively (**Table 2** and **Figure 1B**). The combined model incorporating the Rad-score and N stage yielded a

TABLE 1 | Demographic and clinicopathologic characteristics of the training cohort, internal validation cohort and external validation cohort.

Characteristics	Training set (n = 140)	Internal validation set (n = 60)	*p* value	External validation set (n = 50)
Age (years)			0.642	
Mean ± SD	57.83 ± 11.47	58.57 ± 12.25		61.46 ± 12.99
Gender			0.241	
Male	71 (50.7%)	25 (41.7%)		23 (46.0%)
Female	69 (49.3%)	35 (58.3%)		27 (54.0%)
Location			0.826	
Left	56 (40.0%)	25 (41.7%)		28 (56.0%)
Other	84 (60.0%)	35 (58.3%)		22 (44.0%)
Pathological typing			0.479	
Adenocarcinoma	133 (95.5%)	59 (98.3%)		48 (96.0%)
Other	7 (5.0%)	1 (1.7%)		2 (4.0%)
T category			0.925	
Tl	22 (15.7%)	10 (16.7%)		6 (12.0%)
T2	49 (35.0%)	18 (30.0%)		12 (24.0%)
T3	28 (20.0%)	13 (21.7%)		6 (12.0%)
T4	41 (29.3%)	19 (31.7%)		26 (52.0%)
N category			0.454	
N0	15 (10.7%)	8 (13.3%)		9 (18.0%)
N1	20 (14.3%)	6 (1.0%)		3 (6.0%)
N2	51 (36.4%)	17 (28.3%)		6 (12.0%)
N3	54 (38.6%)	29 (48.3%)		32 (64.0%)
M category			0.231	
M0	5 (3.6%)	0 (0.0%)		2 (4.0%)
M1a	30 (21.4%)	16 (26.7%)		9 (18.0%)
M1b	60 (42.9%)	27 (45.0%)		10 (20.0%)
M1c	45 (32.1%)	17 (28.3%)		29 (58.0%)
Tobacco use			0.311	
Smoker	40 (28.6%)	13 (21.7%)		14 (28.0%)
No smoker	100 (71.4%)	47 (78.3%)		36 (72.0%)
Base PS score			0.770	
< 2	129 (92.1%)	56 (93.3%)		38 (76.0%)
≥ 2	11 (7.9%)	4 (6.7%)		12 (24.0%)
Mutation status			0.319	
EGFR 19Del	64 (45.7%)	30 (50.0%)		–
EGFR 21L858R	53 (37.9%)	25 (41.7%)		–
Other EGFR	23 (16.4%)	5 (8.3%)		–
CEA			0.835	
Mean ± SD	110.64 ± 209.31	130.37 ± 248.97		–
CYFRA21-1			0.987	
Mean ± SD	8.44 ± 11.37	10.19 ± 18.53		–
Brain metastasis			0.064	
Yes	19 (13.6%)	15 (25.0%)		11 (22.0%)
No	121 (86.4%)	45 (75.0%)		39 (78.0%)
Bone metastasis			0.275	
Yes	84 (60.0%)	31 (51.7%)		33 (66.0%)
No	56 (40.0%)	29 (48.3%)		17 (34.0%)
Liver metastasis			0.162	
Yes	21 (15.0%)	4 (6.7%)		2 (4.0%)
No	119 (85.0%)	56 (93.3%)		48 (96.0%)
Lung metastasis			0.828	
Yes	63 (45.0%)	28 (46.7%)		23 (46.0%)
No	77 (55.0%)	32 (53.3%)		27 (54.0%)
Pleural metastasis			0.225	
Yes	48 (34.3%)	26 (43.3%)		17 (34.0%)
No	92 (65.7%)	34 (56.7%)		33 (66.0%)
Adrenal metastasis			0.159	
Yes	25 (17.9%)	6 (10.0%)		10 (20.0%)
No	115 (82.1%)	54 (90.0%)		40 (80.0%)
PFS (months)			0.063	
Mean ± SD	9.82 ± 6.86	12.14 ± 8.47		15.20 ± 11.49

Age, CEA, CYFRA21-1 and PFS are shown as mean ± standard deviation (SD); other data are the number of patients with the percentage in parentheses. Statistical comparison between the training cohort and validation cohort was computed with $\chi 2$ test (categorical variables) or Mann-Whitney U test (continuous variables). PS, performance status; CEA, carcinoembryonic antigen; CYFRA21-1, non-small cell associated antigens; PFS, progression-free survival.

TABLE 2 | Model performance on predicting PFS and time-dependent PFS probability.

Model	Cohorts	C-index (95% CI)	AUC[1] (95% CI)	AUC[2] (95% CI)	AUC[3] (95% CI)
N stage[4]	training	0.59 (0.54-0.64)	0.62 (0.54-0.71)	0.64 (0.55-0.72)	0.67 (0.58-0.74)
	internal validation	0.56 (0.47-0.64)	0.57 (0.44-0.70)	0.60 (0.47-0.73)	0.71 (0.58-0.82)
	external validation	0.56 (0.48-0.64)	0.59 (0.44-0.73)	0.62 (0.47-0.75)	0.59 (0.44-0.72)
Rad-score	training	0.65 (0.60-0.70)	0.72 (0.64-0.80)	0.74 (0.66-0.81)	0.79 (0.71-0.85)
	internal validation	0.61 (0.53-0.68)	0.66 (0.53-0.78)	0.64 (0.50-0.76)	0.67 (0.54-0.79)
	external validation	0.60 (0.52-0.68)	0.58 (0.43-0.72)	0.69 (0.55-0.81)	0.76 (0.62-0.87)
Combined model	training	0.67 (0.62-0.72)	0.75 (0.67-0.82)	0.78 (0.70-0.84)	0.81 (0.74-0.87)
	internal validation	0.61 (0.52-0.69)	0.64 (0.50-0.76)	0.66 (0.52-0.77)	0.72 (0.59-0.83)
	external validation	0.60 (0.51-0.69)	0.60 (0.45-0.73)	0.69 (0.55-0.81)	0.71 (0.56-0.83)

AUC[1], AUCs at 10-month progression-free survival (PFS).
AUC[2], AUCs at one-year progression-free survival (PFS).
AUC[3], AUCs at 14-month progression-free survival (PFS).
N stage[4], Clinicopathological model was built with N stage.
CI, confidence interval; AUC, area under the curve.

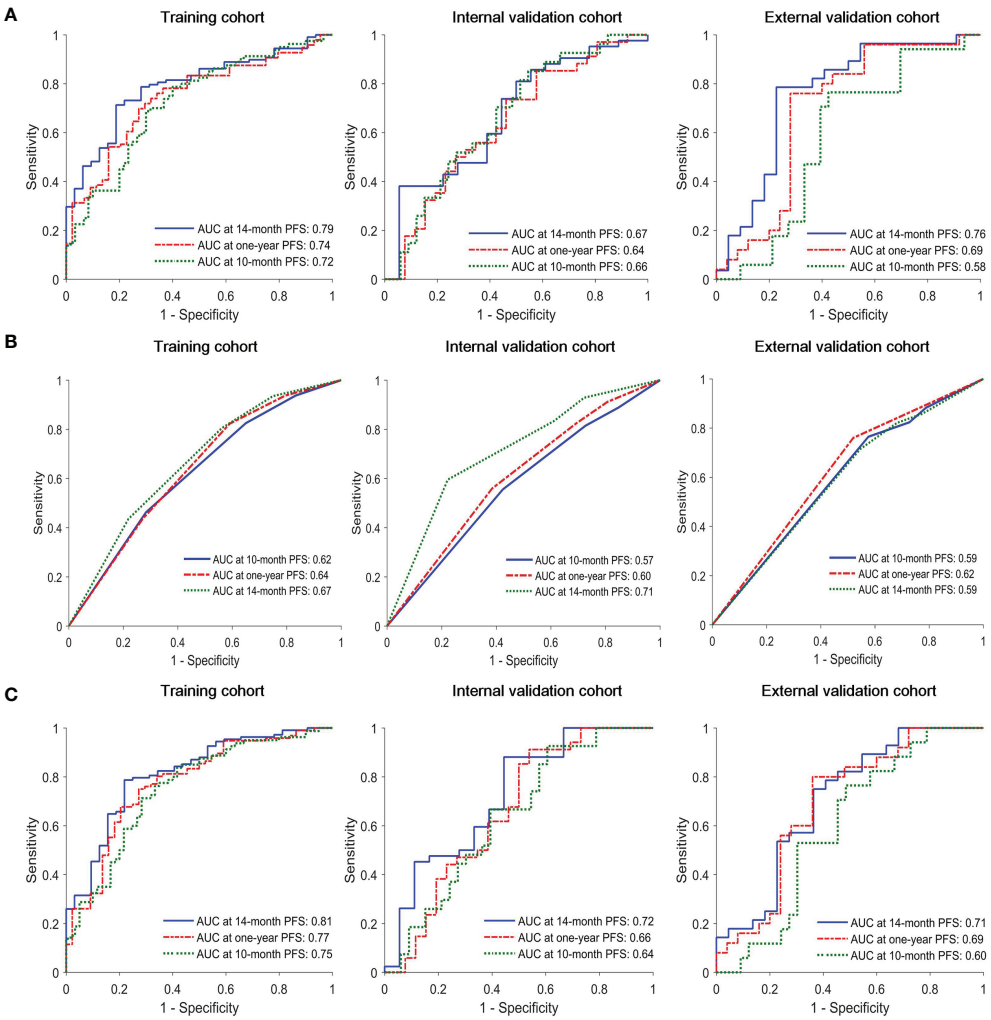


FIGURE 1 | Time-dependent receiver operating characteristic (ROC) curves of **(A)** radiomics model, **(B)** clinical model and **(C)** combined model in the training, interval validation, and external validation cohorts.

C-index of 0.67 (95% CI: 0.62-0.72) in the training cohort, 0.61 (95% CI: 0.52-0.69) in the internal validation cohort, and 0.60 (95% CI: 0.51-0.69) in the external validation cohort; the AUC ranges were 0.60-0.75, 0.66-0.78, and 0.71-0.81 for the 10-month, one-year, and 14-month PFS probability predictions in the three cohorts, respectively (**Table 2** and **Figure 1C**). The statistical comparison of the ROC curves between the combined model and radiomics model (Rad-score) was performed using the DeLong test method (33). The *p* values from the DeLong test are given in **Supplemental Table S7**. No significant differences were found (all *p* values > 0.05); therefore, the combined model showed no performance improvement when compared with the radiomics model (Rad-score) for PFS prediction.

Risk Stratification of EGFR-TKI Therapy and Identification of a Sensitive Population

As shown in **Figures 2A, B** according to the division of median value of Rad-score (median value: -0.0721), each of these three cohorts was divided into slow-progression (blue bars) and rapid-progression (red bars) in the expectation of EGFR-TKI therapy. The Kaplan–Meier survival curves of the three cohorts present significant differences in PFS between the stratified slow and rapid progression subgroups. Higher Rad-scores were significantly associated with worse PFS in the training ($p < 0.0001$), internal validation ($p = 0.0153$), and external validation ($p = 0.0006$) cohorts. The proportion of rapidly progressing patients in each cohort was 50%, 47%, and 50%. Compared with the Rad-score, the Clinicopathological model based on N stage cannot identify the patients with a high risk of rapid progression in the internal ($p = 0.1100$) and external ($p = 0.4637$) validation cohorts (**Figure 2C**). At the same time, the combined model incorporating the Rad-score and N stage cannot successfully achieve risk stratification in the external ($p = 0.1010$) validation cohort (**Figure 2D**).

DISCUSSION

Our 2-center study results showed that the established PET/CT Rad-score had favorable predictive performance for PFS estimation. It can accurately quantify the risk of progression after first-line EGFR-TKI treatment in NSCLC patients and identify populations sensitive to targeted therapy for EGFR-mutant NSCLC to help develop personalized clinical treatment regimens.

At present, few studies based on PET/CT radiomics to predict PFS of NSCLC patients with EGFR mutations. Although, our previous study (23) attempted to investigate the performance of PFS prediction using interim PET/CT (ΔSUVmax and ΔSUVmean) in stage IIIC/IV EGFR-mutant NSCLC patients with EGFR-TKI therapy, this is a single-centre study with a small cohort size (78 patients), the AUCs of the ΔSUVmax and ΔSUVmean were 0.764 and 0.725, respectively. This results has not been validated with independent cohorts from different centers, hence it may be over-fitting. Our results showed that the AUCs of the Rad-score were 0.72, 0.74, and 0.79 for 10-month, one-year, and 14-month PFS probability prediction in the training cohort. This is comparable with previous study. However, the corresponding AUCs were 0.66, 0.64 and 0.67 in the internal validation cohort, and 0.58, 0.69 and 0.76 in the external validation cohort, which reiterates the importance of multi-center independent validation to reflect the actual performance of radiomics model. Furthermore, this results are consistent with study by Kirienko et al. (15), which used PET/CT radiomics signatures to successfully predict disease-free survival (DFS) of 259 patients with NSCLC after surgery; it achieved an AUC of 0.68 *via* using PET/CT signature in the independent validation cohort (90 patients), and an AUC of 0.65 after combining it with clinical predictors. While the C-index of above studies were not reported, so there is no direct comparison of C-index. Importantly, our Rad-score had higher C-index and AUCs than clinical N stage in all three cohorts (**Table 2**), and can effectively identify the patients with low risk of slow progression in internal and external validation cohorts, and for these patients, EGFR-TKI therapy showed great clinical benefits (**Figure 2**).

In this study, a total of 2074 (1037 PET and 1037 CT) radiomics features were extracted to quantify the phenotypic characteristics of the tumor region in PET and CT images, and a 4-step feature selection method was performed to avoid model overfitting. In particular, the robustness of feature extraction to different contours of segmentation was evaluated, which may improve the stability and generalizability of model. Herein, we used supervoxel-based contour randomization (24) to create perturbed ROIs in PET and CT images of 60 randomly selected patients, respectively. Having 60 patients (120 PET and CT ROIs) perturbed would have made it possible to assess the robustness of feature extraction (34–36). Furthermore, due to the perturbed ROIs may deviate from tumor region, even include some non-tumor slices (**Supplementary Figure S1**), which may leads to unnecessary removal of features in robustness assessment, the perturbed results were further checked and adjusted slice by slice, which is also a time-consuming process.

It is worth noting that nearly 90% (1867 of 2074) features were non-robust (309 of 2074 features) and redundant (1558 of 2074 features) (**Table S3**). Furthermore, Chalkidou et al. reported that 10 to 15 observations per predictor variable are minimally required to produce reasonably stable estimates (37). The top 10 features (140 training samples) were therefore selected from the remaining 10% features *via* univariate Cox analysis, and three-feature signature was eventually built *via* exhaustive search, which considered all possible combinations ($1013 \text{ combinations} = \sum_{k=2}^{10} C_{10}^{k}$) in multivariable Cox analysis. The signature consists of two PET features and one CT feature, namely PET_wavelet-LHL_first-order 90th Percentile (a measure to describe the 90th percentile of intensity distribution of wavelet-filtering PET images), PET_wavelet-LHL_gldm Large Dependence High Gray Level Emphasis (a measure to gauge the distribution of large dependence with higher gray-level values in Gray Level Dependence Matrix (GLDM) of wavelet-filtering PET images), and CT_wavelet-HHH_first order Kurtosis (a measure of peakedness in the intensity distribution of wavelet-filtering CT images). Our study showed that the Rad-score built with the three features could predict posttreatment PFS in the training group and the 2 validation groups of EGFR-TKI-treated patients (**Figure 2B**).

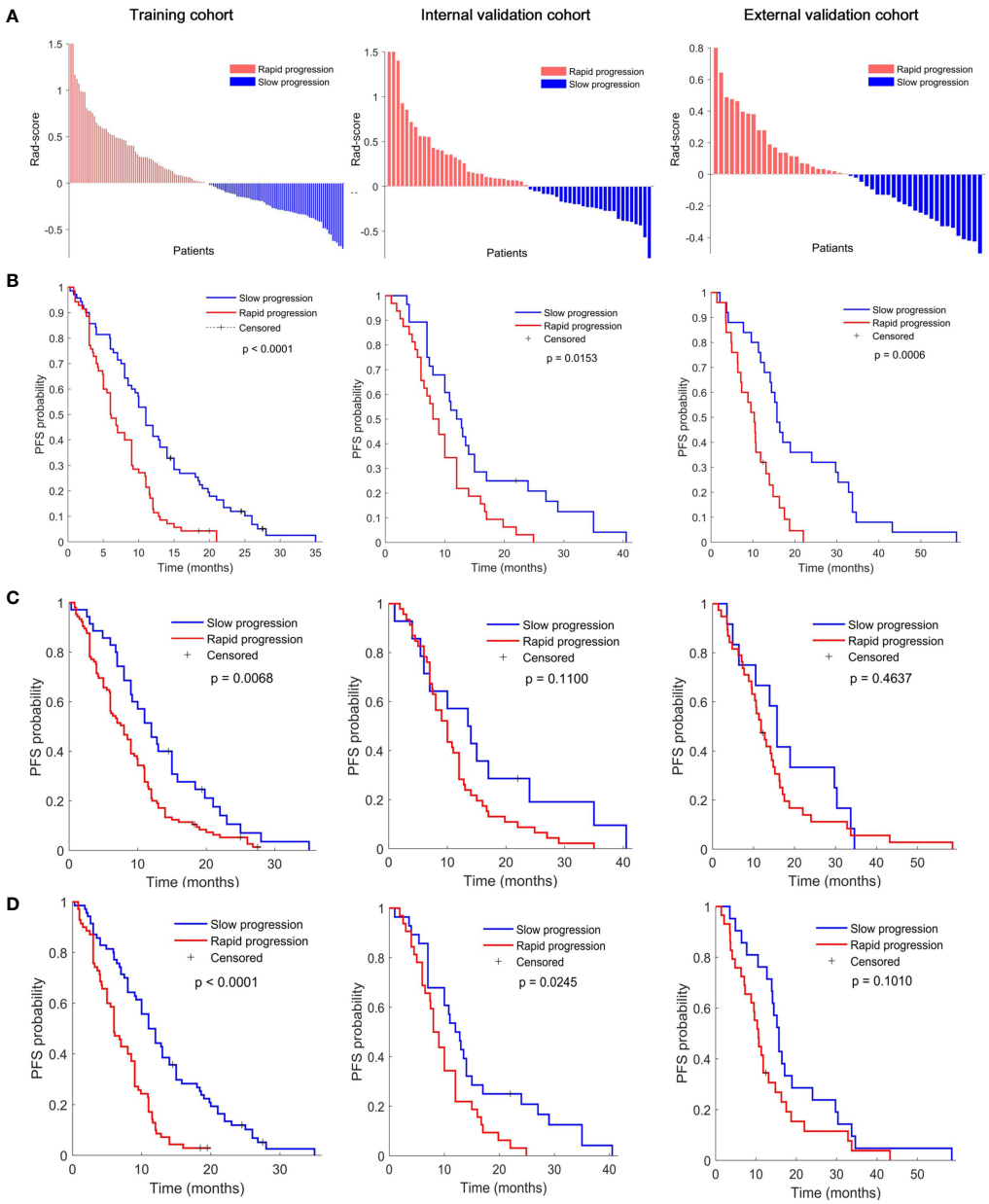


FIGURE 2 | **(A)** Rad-score according to the three-feature signature and Kaplan–Meier survival curves of **(B)** radiomics model, **(C)** clinical model, and **(D)** combined model in the training (left), interval validation (middle), and external validation cohorts (right). All scores have subtracted the cutoff. *P* values were calculated using the log-rank test.

The PET/CT Rad-score identified 123 patients with rapid progression after treatment and 127 patients with slow progression after treatment from 250 patients in the EGFR-TKI treatment group. The PFS of the corresponding two types of patients was 8.56 ± 5.17 months and 14.54 ± 10.19 months. There was a statistically significant difference ($p < 0.0001$) in PFS between the two groups (**Figure 3**). The rapid-progression patients in the EGFR-mutated NSCLC group did not benefit from EGFR-TKI treatment and were a TKI-treatment insensitive population. In this study, the PET/CT Rad-score successfully achieved accurate quantification of the risk of progression after first-line first and second generation EGFR-TKI drug therapy in stage IIIC/IV EGFR-mutated NSCLC patients and identified first and second generation EGFR-TKI treatment-sensitive populations.

Clinicopathological information on the combination of peripheral invasion and metastasis might reflect the biological invasiveness of tumors and have some predictive prognostic value (38). However, our survival analysis results identified only N stage among the clinicopathological factors as a risk factor for predicting PFS in patients receiving EGFR-TKI treatment, while lymph node

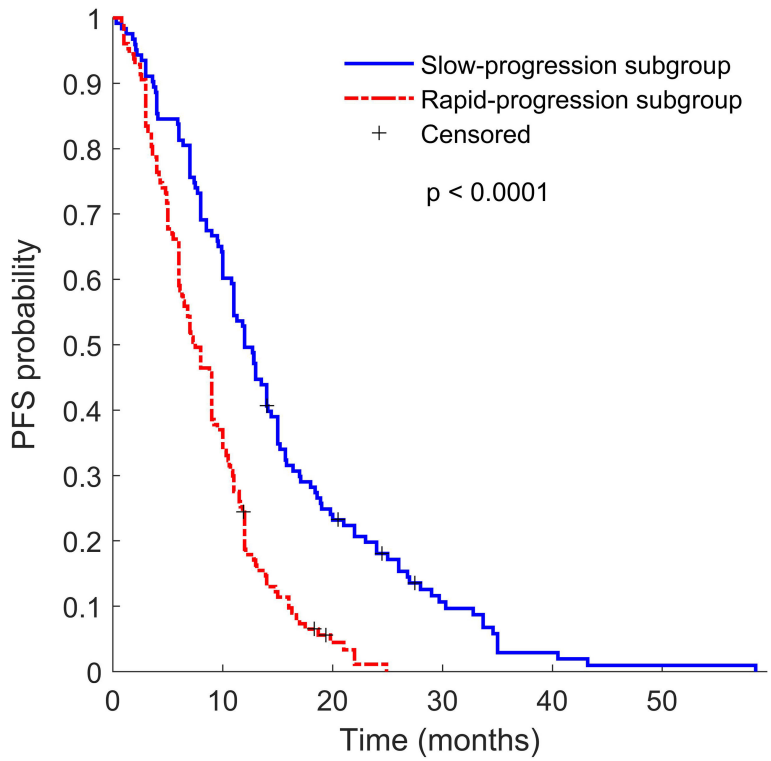


FIGURE 3 | Kaplan–Meier survival curves of slow progression subgroup TKI patients (blue line) and rapid progression subgroup TKI patients (red line). *P* value was calculated using the log-rank test.

metastasis is a well-established prognostic feature for predicting the prognosis of NSCLC (39–42). The predictive value of T stage and distant metastasis varies in different studies (4, 43–45), and our study showed that T stage, M stage, and metastasis at all sites were not prognostic factors. Smoking history is a causative risk factor for lung cancer (46), but whether smoking history is a risk factor affecting the therapeutic effect of lung cancer remains inconclusive (47–49). Pathological type and EGFR mutation (exon 18 deletion, exon 19 deletion, or exon 21 L858R substitution) subtype are still controversial prognostic factors in different trials (5, 39, 50–52). In our study, we found that differences in pathological type as well as three common EGFR mutations did not differ in terms of their benefit for EGFR-TKIs ($p > 0.05$). CEA and NSCLC-associated antigen are important indicators for monitoring the recurrence and progression of NSCLC, but our study did not show them as prognostic factors for predicting PFS. In addition, this analysis did not show that sex had a significant prognostic impact (41).

In the evaluation of the value of the clinicopathological model and Rad-score-clinicopathological combined model, our study showed that the combined model showed no significant PFS prediction performance improvement when compared with the Rad-score, and the AUC did not significantly improve. In addition, compared with the Rad-score, the clinicopathological model based on N stage had poor performance of risk stratification and could not identify sensitive patients with EGFR-TKI therapy in the internal ($p = 0.1100$) and external ($p = 0.4637$) validation cohorts. At the same time, the combined model incorporating the Rad-score and N stage could not successfully achieve risk stratification in the external validation cohort ($p = 0.1010$) and thus could not identify patients with rapid progression who are resistant to EGFR-TKI therapy. The possible reason may be that the clinical factor contains limited and rough prognostic information; moreover, due to tumor heterogeneity, pathological factors cannot reflect the overall tumor differentiation (53, 54). Our findings are in line with those of recent studies that the Rad-score-clinicopathological combined model does not improve the PFS prediction performance (15, 55).

Despite the favorable results of the Rad-score, our study also has some limitations. First, the PET/CT images from two centers were affected by different scanners and protocols. This effect was compensated by the ComBat harmonization method. A more comprehensive method to balance the scanner variance is worth future exploration. Second, we mined 3 prognostic features from PET/CT images and compared their performance with clinicopathological factors. The relationship between the radiomic features and biological level events was not investigated, which limits the broad translation of such radiomics model into clinical application.

Some efforts have been made to introduce biological meaning into radiomics. Tunali et al. performed genomic analysis and

immunohistochemistry analysis for carbonic anhydrase IX, and demonstrated that CT treatment response biomarkers for patients with lung cancer treated with immunotherapy were strongly associated with hypoxia, a prognostic factor (56). Ganeshan et al. took a more detailed look at the correlation of CT texture features from lung cancer tumors to histopathologic markers of angiogenesis and hypoxia, and concluded that CT textures appear to be surrogate measure of tumor hypoxia (57). The interpretability of the radiomics model in terms of the biological properties of tissue is an essential and challenge process (58). Our future studies will plan to pursue biological validation to further explain the radiomics prognostic model.

CONCLUSIONS

In conclusion, a pretreatment PET/CT radiomics biomarker for the prediction of the PFS of NSCLC patients with stage IIIC/IV EGFR mutations was established; it has the potential to realize the precise quantitative stratification of progression risk after first-line first and second generation of EGFR-TKI drug therapy and identify EGFR-mutant NSCLC populations sensitive to targeted therapy. One of the limitation is that the relationship between radiomics biomarker and biological meaning has not been investigated in present study.

DATA AVAILABILITY STATEMENT

The significant feature data supporting the conclusions of this article has been made publicly available. This data can be found here: https://github.com/dudongyangsmu/PFS-of-EGFR-mutated-NSCLC.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics Committee, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences. The ethics committee waived the requirement of written informed consent for participation. Written informed consent was not obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by the National Natural Science Foundation of China under Grant 81871437, the NSFC Incubation Program of GDPH (KY012021162), the Science and Technology Planning Project of Guangdong Province (2019A1515011104, 2018A030313534, 2017A020215110), the Natural Science Foundation of Guangdong Province (2019A1515110377, 2018A030313534), the Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (Lijun Lu, 2018), and the China Postdoctoral Science Foundation funded project under Grant 2020M682792.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.721318/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2019. *CA Cancer J Clin* (2019) 69:7–34. doi: 10.3322/caac.21551
2. Paez JG, Janne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR Mutations in Lung Cancer: Correlation With Clinical Response to Gefitinib Therapy. *Science* (2004) 304:1497–500. doi: 10.1126/science.1099314
3. Sequist LV, Han JY, Ahn MJ, Cho BC, Yu H, Kim SW, et al. Osimertinib Plus Savolitinib in Patients With EGFR Mutation-Positive, MET-Amplified, Non-Small-Cell Lung Cancer After Progression on EGFR Tyrosine Kinase Inhibitors: Interim Results From a Multicentre, Open-Label, Phase 1b Study. *Lancet Oncol* (2020) 21:373–86. doi: 10.1016/S1470-2045(19)30785-5
4. Song J, Shi J, Dong D, Fang M, Zhong W, Wang K, et al. A New Approach to Predict Progression-Free Survival in Stage IV EGFR-Mutant NSCLC Patients With EGFR-TKI Therapy. *Clin Cancer Res* (2018) 24:3583–92. doi: 10.1158/1078-0432.CCR-17-2507
5. Ramalingam SS, Vansteenkiste J, Planchard D, Cho BC, Gray JE, Ohe Y, et al. Overall Survival With Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. *N Engl J Med* (2020) 382:41–50. doi: 10.1056/NEJMoa1913662
6. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
7. Liu Z, Zhang XY, Shi YJ, Wang L, Zhu HT, Tang Z, et al. Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Clin Cancer Res* (2017) 23:7253–62. doi: 10.1158/1078-0432.CCR-17-1038
8. El Naqa I, Ten Haken RK. Can Radiomics Personalise Immunotherapy? *Lancet Oncol* (2018) 19:1138–9. doi: 10.1016/S1470-2045(18)30429-7
9. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, 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. *Lancet Oncol* (2018) 19:1180–91. doi: 10.1016/S1470-2045(18)30413-3
10. Du D, Gu J, Chen X, Lv W, Feng Q, Rahmim A, et al. Integration of PET/CT Radiomics and Semantic Features for Differentiation Between Active Pulmonary Tuberculosis and Lung Cancer. *Mol Imaging Biol* (2021) 23:287–98. doi: 10.1007/s11307-020-01550-4
11. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006
12. Mattonen SA, Davidzon GA, Benson J, Leung ANC, Vasanawala M, Horng G, et al. Bone Marrow and Tumor Radiomics at (18)F-FDG PET/CT: Impact on Outcome Prediction in Non-Small Cell Lung Cancer. *Radiology* (2019) 293:451–9. doi: 10.1148/radiol.2019190357

13. Chen SW, Shen WC, Lin YC, Chen RY, Hsieh TC, Yen KY, et al. Correlation of Pretreatment (18)F-FDG PET Tumor Textural Features With Gene Expression in Pharyngeal Cancer and Implications for Radiotherapy-Based Treatment Outcomes. *Eur J Nucl Med Mol Imaging* (2017) 44:567–80. doi: 10.1007/s00259-016-3580-5
14. Jiang Y, Yuan Q, Lv W, Xi S, Huang W, Sun Z, et al. Radiomic Signature of (18)F Fluorodeoxyglucose PET/CT for Prediction of Gastric Cancer Survival and Chemotherapeutic Benefits. *Theranostics* (2018) 8:5915–28. doi: 10.7150/thno.28018
15. Kirienko M, Cozzi L, Antunovic L, Lozza L, Fogliata A, Voulaz E, et al. Prediction of Disease-Free Survival by the PET/CT Radiomic Signature in Non-Small Cell Lung Cancer Patients Undergoing Surgery. *Eur J Nucl Med Mol Imaging* (2018) 45:207–17. doi: 10.1007/s00259-017-3837-7
16. van Helden EJ, Vacher YJL, van Wieringen WN, van Velden FHP, Verheul HMW, Hoekstra OS, et al. Radiomics Analysis of Pre-Treatment [(18)F]FDG PET/CT for Patients With Metastatic Colorectal Cancer Undergoing Palliative Systemic Treatment. *Eur J Nucl Med Mol Imaging* (2018) 45:2307–17. doi: 10.1007/s00259-018-4100-6
17. Ettinger DS, Wood DE, Aggarwal C, Aisner DL, Akerley W, Bauman JR, et al. NCCN Guidelines Insights: Non-Small Cell Lung Cancer, Version 1.2020. *J Natl Compr Canc Netw* (2019) 17:1464–72. doi: 10.6004/jnccn.2019.0059
18. Bashir U, Siddique MM, McLean E, Goh V, Cook GJ. Imaging Heterogeneity in Lung Cancer: Techniques, Applications, and Challenges. *AJR Am J Roentgenol* (2016) 207:534–43. doi: 10.2214/AJR.15.15864
19. Thawani R, McLane M, Beig N, Ghose S, Prasanna P, Velcheti V, et al. Radiomics and Radiogenomics in Lung Cancer: A Review for the Clinician. *Lung Cancer* (2018) 115:34–41. doi: 10.1016/j.lungcan.2017.10.015
20. Fornacon-Wood I, Faivre-Finn C, O'Connor JPB, Price GJ. Radiomics as a Personalized Medicine Tool in Lung Cancer: Separating the Hope From the Hype. *Lung Cancer* (2020) 146:197–208. doi: 10.1016/j.lungcan.2020.05.028
21. Kang F, Mu W, Gong J, Wang S, Li G, Li G, et al. Integrating Manual Diagnosis Into Radiomics for Reducing the False Positive Rate of (18)F-FDG PET/CT Diagnosis in Patients With Suspected Lung Cancer. *Eur J Nucl Med Mol Imaging* (2019) 46:2770–9. doi: 10.1007/s00259-019-04418-0
22. Zhang J, Zhao X, Zhao Y, Zhang J, Zhang Z, Wang J, et al. Value of Pre-Therapy (18)F-FDG PET/CT Radiomics in Predicting EGFR Mutation Status in Patients With Non-Small Cell Lung Cancer. *Eur J Nucl Med Mol Imaging* (2020) 47:1137–46. doi: 10.1007/s00259-019-04592-1
23. Shao D, Cheng Y, Yuan ZS, Jiang BY, Wang SX. Value of Interim (18)F-FDG PET/CT for Predicting Progression-Free Survival in Stage B/IV EGFR-Mutant Non-Small-Cell Lung Cancer Patients With EGFR-TKI Therapy. *Lung Cancer* (2020) 149:137–43. doi: 10.1016/j.lungcan.2020.09.020
24. Zwanenburg A, Leger S, Agolli L, Pilz K, Troost EGC, Richter C, et al. Assessing Robustness of Radiomic Features by Image Perturbation. *Sci Rep* (2019) 9:614. doi: 10.1038/s41598-018-36938-4
25. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
26. Johnson WE, Li C, Rabinovic A. Adjusting Batch Effects in Microarray Expression Data Using Empirical Bayes Methods. *Biostatistics* (2007) 8:118–27. doi: 10.1093/biostatistics/kxj037
27. Fortin JP, Parker D, Tunc B, Watanabe T, Elliott MA, Ruparel K, et al. Harmonization of Multi-Site Diffusion Tensor Imaging Data. *Neuroimage* (2017) 161:149–70. doi: 10.1016/j.neuroimage.2017.08.047
28. Fortin JP, Cullen N, Sheline YI, Taylor WD, Aselcioglu I, Cook PA, et al. Harmonization of Cortical Thickness Measurements Across Scanners and Sites. *Neuroimage* (2018) 167:104–20. doi: 10.1016/j.neuroimage.2017.11.024
29. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* (2016) 15:155–63. doi: 10.1016/j.jcm.2016.02.012
30. Ligero M, Jordi-Ollero O, Bernatowicz K, Garcia-Ruiz A, Delgado-Munoz E, Leiva D, et al. Minimizing Acquisition-Related Radiomics Variability by Image Resampling and Batch Effect Correction to Allow for Large-Scale Data Analysis. *Eur Radiol* (2021) 31:1460–70. doi: 10.1007/s00330-020-07174-0
31. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Society. Ser B (Methodological)* (1995) 57:289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
32. Lv W, Yuan Q, Wang Q, Ma J, Feng Q, Chen W, et al. Radiomics Analysis of PET and CT Components of PET/CT Imaging Integrated With Clinical Parameters: Application to Prognosis for Nasopharyngeal Carcinoma. *Mol Imaging Biol* (2019) 21:954–64. doi: 10.1007/s11307-018-01304-3
33. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* (1988) 44:837–45. doi: 10.2307/2531595
34. Lu L, Lv W, Jiang J, Ma J, Feng Q, Rahmim A, et al. Robustness of Radiomic Features in [(11)C]Choline and [(18)F]FDG PET/CT Imaging of Nasopharyngeal Carcinoma: Impact of Segmentation and Discretization. *Mol Imaging Biol* (2016) 18:935–45. doi: 10.1007/s11307-016-0973-6
35. Lv W, Yuan Q, Wang Q, Ma J, Jiang J, Yang W, et al. Robustness Versus Disease Differentiation When Varying Parameter Settings in Radiomics Features: Application to Nasopharyngeal PET/CT. *Eur Radiol* (2018) 28:3245–54. doi: 10.1007/s00330-018-5343-0
36. Lv W, Feng H, Du D, Ma J, Lu L. Complementary Value of Intra-and Peri-Tumoral PET/CT Radiomics for Outcome Prediction in Head and Neck Cancer. *IEEE Access* (2021) 9:81818–27. doi: 10.1109/ACCESS.2021.3085601
37. Chalkidou A, O'Doherty MJ, Marsden PK. False Discovery Rates in PET and CT Studies With Texture Features: A Systematic Review. *PloS One* (2015) 10: e0124165. doi: 10.1371/journal.pone.0124165
38. Arshad MA, Thornton A, Lu H, Tam H, Wallitt K, Rodgers N, et al. Discovery of Pre-Therapy 2-Deoxy-2-(18)F-Fluoro-D-Glucose Positron Emission Tomography-Based Radiomics Classifiers of Survival Outcome in Non-Small-Cell Lung Cancer Patients. *Eur J Nucl Med Mol Imaging* (2019) 46:455–66. doi: 10.1007/s00259-018-4139-4
39. Zhou C, Wu YL, Chen G, Feng J, Liu XQ, Wang C, et al. Erlotinib Versus Chemotherapy as First-Line Treatment for Patients With Advanced EGFR Mutation-Positive Non-Small-Cell Lung Cancer (OPTIMAL, CTONG-0802): A Multicentre, Open-Label, Randomised, Phase 3 Study. *Lancet Oncol* (2011) 12:735–42. doi: 10.1016/S1470-2045(11)70184-X
40. Liang W, Zhang L, Jiang G, Wang Q, Liu L, Liu D, et al. Development and Validation of a Nomogram for Predicting Survival in Patients With Resected Non-Small-Cell Lung Cancer. *J Clin Oncol* (2015) 33:861–9. doi: 10.1200/JCO.2014.56.6661
41. Wu YL, Zhou C, Liam CK, Wu G, Liu X, Zhong Z, et al. First-Line Erlotinib Versus Gemcitabine/Cisplatin in Patients With Advanced EGFR Mutation-Positive Non-Small-Cell Lung Cancer: Analyses From the Phase III, Randomized, Open-Label, ENSURE Study. *Ann Oncol* (2015) 26:1883–9. doi: 10.1093/annonc/mdv270
42. Hanna N, Johnson D, Temin S, Baker S Jr, Brahmer J, Ellis PM, et al. Systemic Therapy for Stage IV Non-Small-Cell Lung Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update. *J Clin Oncol* (2017) 35:3484–515. doi: 10.1200/JCO.2017.74.6065
43. Oh JR, Seo JH, Chong A, Min JJ, Song HC, Kim YC, et al. Whole-Body Metabolic Tumour Volume of 18F-FDG PET/CT Improves the Prediction of Prognosis in Small Cell Lung Cancer. *Eur J Nucl Med Mol Imaging* (2012) 39:925–35. doi: 10.1007/s00259-011-2059-7
44. Takeuchi S, Khiewvan B, Fox PS, Swisher SG, Rohren EM, Bassett RL Jr., et al. Impact of Initial PET/CT Staging in Terms of Clinical Stage, Management Plan, and Prognosis in 592 Patients With Non-Small-Cell Lung Cancer. *Eur J Nucl Med Mol Imaging* (2014) 41:906–14. doi: 10.1007/s00259-013-2672-8
45. Seban RD, Mezquita L, Berenbaum A, Dercle L, Botticella A, Le Pechoux C, et al. Baseline Metabolic Tumor Burden on FDG PET/CT Scans Predicts Outcome in Advanced NSCLC Patients Treated With Immune Checkpoint Inhibitors. *Eur J Nucl Med Mol Imaging* (2020) 47:1147–57. doi: 10.1007/s00259-019-04615-x
46. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492
47. Meng X, Loo BWJr., Ma L, Murphy JD, Sun X, Yu J. Molecular Imaging With 11C-PD153035 PET/CT Predicts Survival in Non-Small Cell Lung Cancer Treated With EGFR-TKI: A Pilot Study. *J Nucl Med* (2011) 52:1573–9. doi: 10.2967/jnumed.111.092874
48. Moon SH, Cho SH, Park LC, Ji JH, Sun JM, Ahn JS, et al. Metabolic Response Evaluated by 18F-FDG PET/CT as a Potential Screening Tool in Identifying a Subgroup of Patients With Advanced Non-Small Cell Lung Cancer for Immediate Maintenance Therapy After First-Line Chemotherapy. *Eur J Nucl Med Mol Imaging* (2013) 40:1005–13. doi: 10.1007/s00259-013-2400-4

49. Mu W, Tunali I, Gray JE, Qi J, Schabath MB, Gillies RJ. Radiomics of (18)F-FDG PET/CT Images Predicts Clinical Benefit of Advanced NSCLC Patients to Checkpoint Blockade Immunotherapy. *Eur J Nucl Med Mol Imaging* (2020) 47:1168–82. doi: 10.1007/s00259-019-04625-9
50. Inoue A, Kobayashi K, Maemondo M, Sugawara S, Oizumi S, Isobe H, et al. Updated Overall Survival Results From a Randomized Phase III Trial Comparing Gefitinib With Carboplatin-Paclitaxel for Chemo-Naive Non-Small Cell Lung Cancer With Sensitive EGFR Gene Mutations (NEJ002). *Ann Oncol* (2013) 24:54–9. doi: 10.1093/annonc/mds214
51. Wu YL, Zhou C, Hu CP, Feng J, Lu S, Huang Y, et al. Afatinib Versus Cisplatin Plus Gemcitabine for First-Line Treatment of Asian Patients With Advanced Non-Small-Cell Lung Cancer Harbouring EGFR Mutations (LUX-Lung 6): An Open-Label, Randomised Phase 3 Trial. *Lancet Oncol* (2014) 15:213–22. doi: 10.1016/S1470-2045(13)70604-1
52. Okura N, Nishioka N, Yamada T, Taniguchi H, Tanimura K, Katayama Y, et al. ONO-7475, a Novel AXL Inhibitor, Suppresses the Adaptive Resistance to Initial EGFR-TKI Treatment in EGFR-Mutated Non-Small Lung Cancer. *Clin Cancer Res* (2020) 26:2244–56. doi: 10.1158/1078-0432.CCR-19-2321
53. Sawair FA, Irwin CR, Gordon DJ, Leonard AG, Stephenson M, Napier SS. Invasive Front Grading: Reliability and Usefulness in the Management of Oral Squamous Cell Carcinoma. *J Oral Pathol Med* (2003) 32:1–9. doi: 10.1034/j.1600-0714.2003.00060.x
54. Speight PM, Abram TJ, Floriano PN, James R, Vick J, Thornhill MH, et al. Interobserver Agreement in Dysplasia Grading: Toward an Enhanced Gold Standard for Clinical Pathology Trials. *Oral Surg Oral Med Oral Pathol Oral Radiol* (2015) 120:474–82 e2. doi: 10.1016/j.oooo.2015.05.023
55. Fang J, Zhang B, Wang S, Jin Y, Wang F, Ding Y, et al. Association of MRI-Derived Radiomic Biomarker With Disease-Free Survival in Patients With Early-Stage Cervical Cancer. *Theranostics* (2020) 10:2284–92. doi: 10.7150/thno.37429
56. Tunali I, Tan Y, Gray JE, Katsoulakis E, Eschrich SA, Saller J, et al. Hypoxia-Related Radiomics Predict Immunotherapy Response: A Multi-Cohort Study of NSCLC. *bioRxiv* (2020). doi: 10.1101/2020.04.02.020859
57. Ganeshan B, Goh V, Mandeville HC, Ng QS, Hoskin PJ, Miles KA. Non–small Cell Lung Cancer: Histopathologic Correlates for Texture Parameters at CT. *Radiology* (2013) 266:326–36. doi: 10.1148/radiol.12112428
58. Tomaszewski MR, Gillies RJ. The Biological Meaning of Radiomic Features. *Radiology* (2021) 298:505–16. doi: 10.1148/radiol.2021202553

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Shao, Du, Liu, Lv, Cheng, Zhang, Lv, Wang and Lu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 25 November 2021
doi: 10.3389/fonc.2021.778627

Multiple Survival Outcome Prediction of Glioblastoma Patients Based on Multiparametric MRI

*Bin Wang[1], Shan Zhang[1], Xubin Wu[1], Ying Li[1], Yueming Yan[1], Lili Liu[2], Jie Xiang[1], Dandan Li[1] and Ting Yan[2]**

[1] College of Information and Computer, Taiyuan University of Technology, Taiyuan, China, [2] Department of Pathology & Shanxi Translational Medicine Research Center on Esophageal Cancer, Shanxi Medical University, Taiyuan, China

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln, United Kingdom

Reviewed by:
Johanna Patricia Adevoso Canal,
University of the Philippines Manila, Philippines
Jue Jiang,
Memorial Sloan Kettering Cancer Center, United States

***Correspondence:**
Ting Yan
enntei@hotmail.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *17 September 2021*
Accepted: *01 November 2021*
Published: *25 November 2021*

Citation:
Wang B, Zhang S, Wu X, Li Y, Yan Y, Liu L, Xiang J, Li D and Yan T (2021) Multiple Survival Outcome Prediction of Glioblastoma Patients Based on Multiparametric MRI. Front. Oncol. 11:778627. doi: 10.3389/fonc.2021.778627

Purpose: Construction of radiomics models for the individualized estimation of multiple survival stratification in glioblastoma (GBM) patients using the multiregional information extracted from multiparametric MRI that could facilitate clinical decision-making for GBM patients.

Materials and Methods: A total of 134 eligible GBM patients were selected from The Cancer Genome Atlas. These patients were separated into the long-term and short-term survival groups according to the median of individual survival indicators: overall survival (OS), progression-free survival (PFS), and disease-specific survival (DSS). Then, the patients were divided into a training set and a validation set in a ratio of 2:1. Radiomics features (n = 5,152) were extracted from multiple regions of the GBM using multiparametric MRI. Then, radiomics signatures that are related to the three survival indicators were respectively constructed using the analysis of variance (ANOVA) and the least absolute shrinkage and selection operator (LASSO) regression for each patient in the training set. Based on a Cox proportional hazards model, the radiomics model was further constructed by combining the signature and clinical risk factors.

Results: The constructed radiomics model showed a promising discrimination ability to differentiate in the training set and validation set of GBM patients with survival indicators of OS, PFS, and DSS. Both the four MRI modalities and five tumor subregions have different effects on the three survival indicators of GBM. The favorable calibration and decision curve analysis indicated the clinical decision value of the radiomics model. The performance of models of the three survival indicators was different but excellent; the best model achieved C indexes of 0.725, 0.677, and 0.724, respectively, in the validation set.

Conclusion: Our results show that the proposed radiomics models have favorable predictive accuracy on three survival indicators and can provide individualized probabilities of survival stratification for GBM patients by using multiparametric and multiregional MRI features.

Keywords: multiparametric MRI, multi-survival indicators, glioblastoma, machine learning, radiomics analysis

INTRODUCTION

Glioblastoma (GBM) is the most common primary malignant neoplasm in adults and is nearly uniformly fatal (1), with a median survival time about 12–14 months (2). It is necessary to establish a survival prediction model that is helpful to the treatment decision-making and disease management for GBM patients (3). In clinical studies for GBM, 5-year or 10-year benchmark survival rates are often calculated to convey prognostic information. A recent study pointed out that three survival endpoints including overall survival (OS), progression-free survival (PFS), and disease-specific survival (DSS) can be used in the study of GBM with confidence (4). It is very important to have a sufficiently long follow-up time to capture the events of interest, and the minimum follow-up time needed depends on both the aggressiveness of the type of endpoint (5).

As a noninvasive and preoperative routine examination for GBM (6), the magnetic resonance imaging (MRI) can comprehensively and macroscopically display the whole tumor and provide fine tumor features, including tumor location, shape, size, and heterogeneity (7). Currently, MRI techniques have great potential for predicting the survival of GBM patients (6, 8, 9). More recently, the field of radiomics has been introduced to extract high-throughput quantitative imaging features from MRI, transform the features into minable data, and establish a prediction or prognosis model connecting image features and tumor phenotype (10, 11). In common MRI acquisitions, four image sequences, i.e., T1-weighted gadolinium contrast-enhanced (T1CE), T1-weighted (T1), T2-weighted (T2), and T2-weighted fluid-attenuated inversion recovery (FLAIR) sequences, are recommended for the diagnosis of a brain tumor (12). It is widely believed that multiparametric MRI can improve the diagnostic efficiency and performance of survival stratification (13). Moreover, the heterogeneity of GBM is reflected in the fact that it usually contains different heterogeneous subregions (such as edema, enhanced and non-enhanced core); this inherent heterogeneity is also reflected in its imaging phenotype because its subregions are described by different intensity distributions of multimodal MRI scanning, reflecting the differences in tumor biology, which all contribute to prognosis prediction (8, 9, 14). Although previous studies have explored the survival time of GBM patients, most of them focus on OS, and the research on the heterogeneity within the tumor is insufficient. Therefore, it is necessary and feasible that a multiparametric MRI- and multiregion-based radiomics approach may improve the performance of the multi-survival stratification in GBM patients.

The aim of the present study was to develop and validate radiomics models based on multiparameter and multiregion MRI for the individualized estimation of the multiple survival indicator stratification in GBM patients.

MATERIALS AND METHODS

Patient Population and Study Design

A total of 134 patients (i) with clinical information (such as OS of patients) from The Cancer Genome Atlas (TCGA[1]) GBM Project and (ii) the corresponding MRI data from The Cancer Imaging Archive (TCIA[2]) were retrospectively included, with the following criteria, (i) including all four types of MRI sequences (T1C, T1, T2, FLAIR); (ii) the MRI sequences were acquired prior to surgery or biopsy; (iii) the MRI sequences were acquired without excessive movement or artifacts; (iv) sufficient clinical data on OS, PFS, or DSS. Additional clinical data for all patients, including age, gender, race, and Karnofsky Performance status (KPS), were obtained from TCGA-GBM Project (see **Supplementary Material E2**).

In this paper, OS was defined as the time between the date of pathological diagnosis and the date of death or the date of last clinical follow-up. PFS was calculated from the date of initial diagnosis to the new tumor event whether it was (a) a progression of disease, (b) local recurrence, (c) distant metastasis, or (d) new primary tumors at all sites or the time the patient was last known to be alive. Similarly, DSS was measured from the date of initial diagnosis until death from the GBM or last follow-up examination. Then, these patients were divided into long-term and short-term survival groups respectively according to the median of OS, PFS, and DSS. Next, all patients were randomly separated into a training cohort (OS and PFS: n = 94; DSS: n = 90) and a validation cohort (OS and PFS: n = 40; DSS: n = 37) at a ratio of 2:1. In this retrospective study, the requirement for informed consent was waived because all the patient data in TCGA were deidentified.

The main design idea of this paper is as follows: firstly, the multimodal MRI (multiparameter) data of GBM patients are preprocessed; next, the radiomics features are extracted from different regions of interest (ROIs) (multiregions), and then the analysis of variance (ANOVA) and least absolute shrinkage and selection operator (LASSO) methods are used to analyze and screen the extracted imaging features (i.e., radiomics signature) related to the three kinds of survival endpoints. Finally, combined with clinical risk factors such as age, gender, and

[1] cancergenome.nih.gov
[2] www.cancerimagingarchive.net/

KPS, the radiomics nomogram, which is based on a Cox proportional hazards model, was analyzed. The overall workflow of the proposed method is shown in **Figure 1**.

Imaging Data Acquisition and Preprocessing

All 134 patients underwent four MRI modalities, i.e., T1C, T1, T2, and FLAIR sequence. The range of image acquisition parameters of the four MRI sequences is provided in **Supplementary Material E1**, section S1. The matrix size of all the MRI sequences was 128 × 128. Diverse parameters of different MRI sequences were used during image acquisition, which may have a great influence on three-dimensional (3D) analyses; thus, two-dimensional (2D) preprocessing was performed in this study. First, the four modalities of all subjects are co-registered to the same anatomical template. Next, the planar resolution of each modality was uniformly resampled to 128 × 128. Finally, since the TCGA-GBM database contains multisite data, the scanner model, pixel spacing, slice thickness, and contrast vary within the selected cohort. To consider these differences, all images were resampled to a common voxel resolution of 1 mm^3, and intensities within each volume were normalized to the [0.1] range.

Multiregional Labeling

In order to obtain information about the survival of patients with GBM from multiple tissue types rather than a single tissue type (15), five different heterogeneous regions were drawn (**Figure 2**). Necrosis (NCR) and non-enhancing tumor (NET) region was defined as ROI A, enhancing tumor (ET) core was defined as ROI B, and edema area (ED) was defined as ROI C. Then, ROI D was generated by merging enhancing tumor region and the first ROI. ROI E was generated by adding edema region to ROI D. We also refer to tumor core (TC) as ROI D and whole tumor (WT) as ROI E. Finally, these five regional contours were respectively mapped to each MRI sequence for each patient.

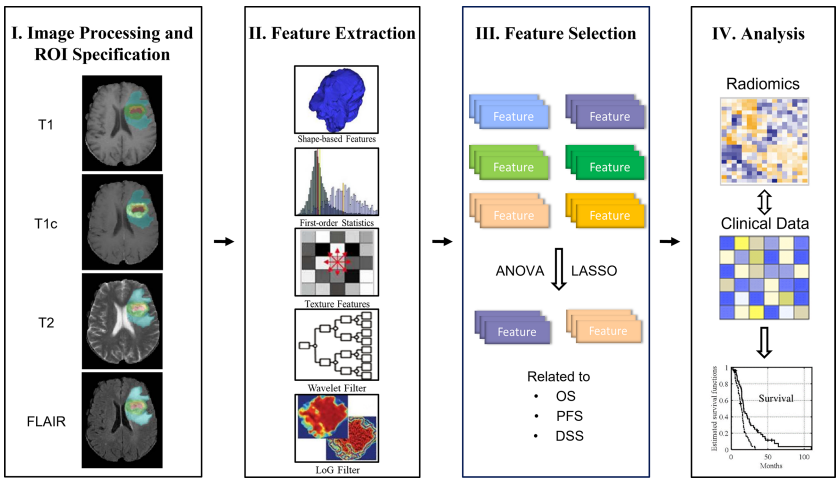


FIGURE 1 | Overall workflow of the study.

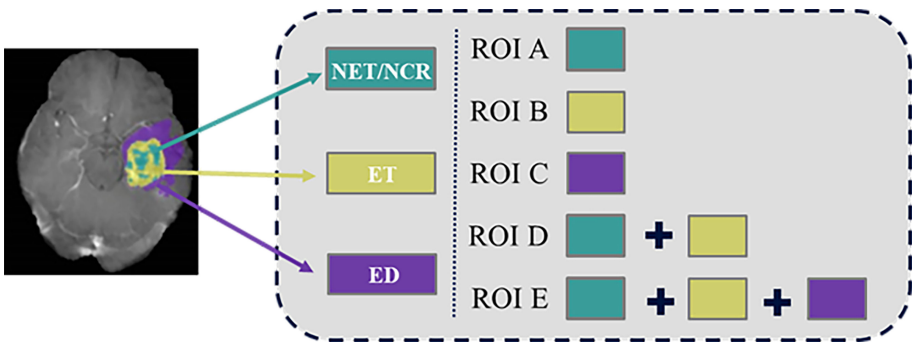


FIGURE 2 | Labeling of the multiple heterogeneous regions.

Radiomics Feature Extraction and Radiomics Signature Construction Related to Survival Indicators

In order to extract high-throughput features, we obtained the original image, eight corresponding Wavelet-filtered images and five corresponding Laplacian of Gaussian (LoG) filtered in each MRI sequence. A total of 5,152 radiomics features were extracted for each patient. A detailed description of the feature extraction is provided in **Supplementary Material E1**, section S2. To reduce the dimension of radiomics features and find out the features that have high evaluation on the prognosis of GBM patients, the ANOVA and the LASSO regression algorithm was adopted to select the survival state-related features among the 5,152 radiomics features in the training cohort.

We use a one-way ANOVA to screen out the characteristics that are separately related to the three survival indicators and have significant differences. Then, LASSO logistic regression method was used to select the characteristics most related to the three survival indicators. LASSO is an effective regression analysis method to constrain the number of independent variables. It can perform feature selection and regularization in high-dimensional data to improve the prediction accuracy through penalty estimation function. The L1 penalty term is added to the ordinary linear model; for ordinary linear model, LASSO estimate is as follows:

$$\hat{\beta}_{Lasso} = \arg \min_{\beta \in R^d} \left(\| Y - X\beta \| ^2 + \lambda \sum_{j=1}^{d} | \beta_j | \right)$$

Where t corresponds to λ one-to-one, which is the adjustment coefficient. If λ is large, it has no effect on the estimated regression parameters, but as λ decreases, the coefficients of most covariates shrink to zero. It makes the model easier to explain: when there are a large number of independent variables, several important independent variables can be found, and the information provided by these variables is the most important in the model.

With the adjustment of λ, the LASSO method can shrink all the coefficients toward zero and set the coefficients of uncorrelated features to zero. Then, 10-fold cross-validation with a maximum area under the curve (AUC) criterion was employed to find an optimal λ, where the final value of λ produces the maximum AUC. The non-zero coefficient is used to construct the regression model, and the corresponding non-zero coefficient is defined as the Rad-score. The fitting formula is generated using a linear combination of selected eigenvalues weighted by their Rad-score. The formula was then used to calculate the radiomics signature of each GBM patient to reflect his or her long-term or short-term survival.

$$\text{Radiomics signature} = \sum_{i=1}^{n} (p_i * v_i)$$

Where p_i (i.e., Rad-score) is the coefficient of the i-th characteristic, and v_i is the i-th characteristic value of patients.

Construction and Assessment of the Radiomics Model With the Training Cohort

The radiomics signature and each clinical factor were first inserted into a Cox proportional hazards model to test whether they were significantly independent prognostic factors for survival stratification in the training cohort. The radiomics signature and significant clinical factors were then utilized to build the Cox proportional hazards model to discriminate the short- and long-term survival outcome of the GBM patients. For comparison, Cox proportional hazards models that used only the significant clinical factors were also established. Finally, based on Cox proportional hazards model, radiomics nomograms of the three survival indicators are respectively constructed, which can directly and individually indicate the probability of survival stratification in the training queue.

The discriminative ability of the radiomics nomogram was quantitatively measured using the C-index, which ranges from 0 to 1. The calibration curves were plotted using observed probabilities and the nomogram-estimated probabilities.

External Validation of the Radiomics Model on the Validation Cohort

The fitting formula that was constructed with the training cohort was applied to all GBM patients in the validation cohort, and the radiomics signature of each patient was calculated. The radiomics nomogram was then validated in this cohort using the radiomics signatures and clinical factors. Finally, the C-index was implemented to evaluate the model results for survival group stratification. Moreover, the calibration curve and the Kaplan–Meier survival curve were also constructed.

Clinical Utility of the Radiomics Model

To estimate the clinical utility of the radiomics nomogram, decision curve analysis (DCA) was performed by calculating the net benefits at different threshold probabilities in the combined training and validation cohorts.

Statistical Analysis

In this study, either Student's *t* tests or Mann–Whitney *U* tests were applied to confirm whether intergroup differences in continuous variables (such as age) existed between the short- and long-term survival groups. Either chi-square tests or Fisher's exact tests were performed on the remaining categorical characteristics to determine whether the constituent ratios differed significantly between the groups. All statistical analyses were performed with R software version 3.6.1 (R Foundation for Statistical Computing; http://www.R-project.org, 2019) using basic statistical functions or additional packages. The following R packages were used: the *glmnet* package was used for the LASSO logistic regression, the *rms* package was used for the nomograms and calibration curves, the *Hmisc* package was used for the comparisons between the C-indices, and the *rmda* package was used to implement the DCA.

Ethics Statement

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and

institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

RESULTS

Clinical Characteristics of the Patients

The cohort of 40 patients ranged in age from 17 to 84 years with a median OS of 359.5 days, PFS of 195 days, and DSS of 376 days. The clinical characteristics and corresponding results of the statistical analyses comparing the long- and short-term survival groups of OS, PFS, and DSS are summarized in **Table 1**.

LASSO Feature Selection and Radiomics Signature Construction

To determine the optimal regulation weight λ for the LASSO algorithm, features with non-zero coefficients for survival stratification were selected by 10-fold cross-validation from the 5,152 radiomics features. The illustration of feature selection using the LASSO algorithm is provided in **Supplementary Material E1**, section S3. A detailed description of the selected non-zero-coefficient features is provided in **Supplementary Material E1**, section S4.

The radiomics signatures for each GBM patient in the training and validation cohorts are presented in **Figure 3**. The prognostic label represented by the x-axis can clearly distinguish the survival status of GBM patients. That is, the patients with long-term survival generally displayed a significantly higher radiomics signature than the patients with short-term survival in both the training and validation cohorts.

Non-Zero-Coefficient Features Analysis

As mentioned above, according to different ROIs and survival endpoints, the radiomics features with non-zero coefficients corresponding to multiregion MRI and three survival periods

TABLE 1 | Characteristics of patients in the training and validation cohorts for OS, PFS, and DSS.

		Training cohort		P	Validation cohort		P
		Short-term	Long-term		Short-term	Long-term	
OS	Patient no.	47	47		15	25	
	Age, mean ± SD	62.13 ± 12.72	54.64 ± 12.69	*0.003*	62.53 ± 17.08	51.54 ± 14.36	*0.045*
	Gender, no. (%)			0.052			0.870
	Male	26 (55.3%)	35 (74.5%)		8 (53.3%)	14 (56.0%)	
	Female	21 (44.7%)	12 (25.5%)		7 (46.7%)	11 (44%)	
	KPS, median (1)	80 (40–100)	80 (60–100)	*0.050*	80 (60–100)	80 (60–100)	0.068
	Race, no. (%)			0.536			0.414
	White	40 (85.1%)	42 (89.4%)		11 (73.3%)	21 (84.0%)	
	Others	7 (14.9%)	5 (10.6%)		4 (26.7%)	4 (16.0%)	
	Hemisphere no. (%)			0.778			0.680
	Unilateral	39 (83.0%)	40 (85.1%)		8 (53.3%)	15 (60.0%)	
	Bilateral	8 (17.0%)	7 (14.9%)		7 (46.7%)	10 (40.0%)	
PFS	Patient no.	47	47		17	23	
	Age, mean ± SD	60.74 ± 13.77	56.15 ± 12.35	0.056	59.64 ± 20.34	54.91 ± 11.47	0.234
	Gender, no. (%)			0.668			0.676
	Male	31 (65.9%)	29 (61.7%)		10 (58.8%)	12 (52.2%)	
	Female	16 (34.1%)	18 (38.3%)		7 (41.2%)	11 (47.8%)	
	KPS, median (1)	80 (40–100)	80 (60–100)	*0.008*	80 (60–100)	80 (60–100)	0.171
	Race, no. (%)			0.536			0.201
	White	40 (85.1%)	42 (89.4%)		12 (70.6%)	20 (87.0%)	
	Others	7 (14.9%)	5 (10.6%)		5 (29.4%)	3 (13.0%)	
	Hemisphere no. (%)			0.102			0.622
	Unilateral	38 (80.9%)	31 (65.9%)		15 (88.2%)	19 (82.6%)	
	Bilateral	9 (19.1%)	16 (34.1%)		2 (11.8%)	4 (17.4%)	
DSS	Patient no.	45	45		15	22	
	Age, mean ± SD	60.58 ± 13.70	55.44 ± 12.34	*0.034*	50.13 ± 15.61	52.18 ± 14.58	0.066
	Gender, no. (%)			*0.028*			0.729
	Male	24 (53.3%)	34 (75.6%)		8 (53.3%)	13 (59.1%)	
	Female	21 (46.7%)	11 (24.4%)		7 (46.7%)	9 (40.9%)	
	KPS, median (1)	80 (60–100)	80 (60–100)	0.055	80 (60-100)	80 (60-100)	*0.035*
	Race, no. (%)			0.725			0.890
	White	40 (88.9%)	41 (91.1%)		12 (80.0%)	18 (36.4%)	
	Others	5 (11.1%)	4 (8.9%)		3 (20.0%)	4 (63.6%)	
	Hemisphere no. (%)			0.694			0.417
	Unilateral	41 (91.1%)	42 (93.3%)		9 (60.0%)	16 (72.7%)	
	Bilateral	4 (8.9%)	3 (6.7%)		6 (40.0%)	6 (27.3%)	

P values <0.05 are shown in italics.

DSS, disease-specific survival; KPS, Karnofsky Performance Scale; OS, overall survival; PFS, progression-free survival.

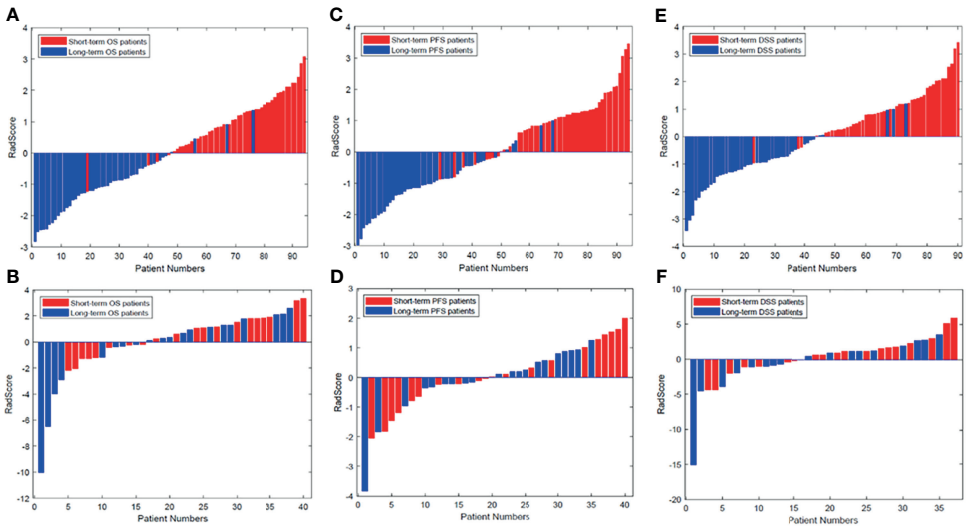


FIGURE 3 | Radiomics signature for each patient in the **(A, C, E)** training cohort and **(B, D, F)** validation cohort. The red bars show the radiomics signature values for the patients with short-term survival (OS, PFS, DFS), and the blue bars show the values for those with long-term survival. DSS, disease-specific survival; OS, overall survival; PFS, progression-free survival.

are selected by LASSO algorithm. The number of different features extracted from each survival indicator and each region is also distinct. According to **Figure 4**, for different survival indicators and different ROIs, the number of non-zero-coefficient radiomics features is inconsistent. As for OS, ROI B has the largest number of features (30 features). For PFS, ROI C has the largest number of features (31 features). Finally, for DSS, ROI A and ROI D have the most features (30 features).

In addition, selected features from different MRI sequences have different coefficients, which can be seen as the importance of one for survival stratification. The results show that T1 sequence and FLAIR sequence account for more than the other sequences. A detailed description of the selected non-zero-coefficient features is provided in **Supplementary Material E1**, section S4.

The Evaluation of the Radiomics Signature and/or Clinical Risk Factors

The radiomics signature, age, gender, and KPS were identified as independent predictors of survival stratification in GBM

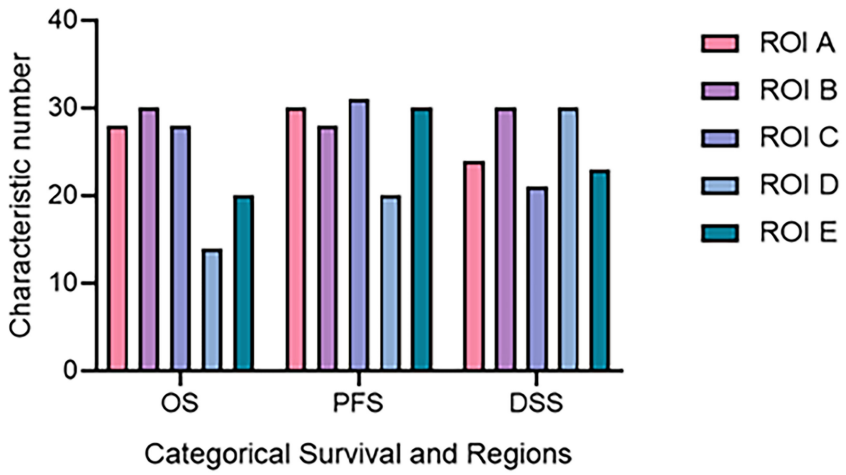


FIGURE 4 | The number of radiomics features corresponding to different ROI and survival time. DSS, disease-specific survival; OS, overall survival; PFS, progression-free survival; ROI, region of interest.

patients. The Cox proportional hazards model was applied based on three clinical predictors and/or the radiomics signature. According to **Table 2**, the C-index that resulted from the combined use of the radiomics signature and three clinical predictors for survival stratification was increased significantly for each item in both training cohort and validation cohort.

In addition, it can be seen from **Table 2** that tumor subregions have different manifestations in the three survival indicators in the validation set. For different heterogenous regions, the C-index of ROI B was 0.725 (95% CI, 0.590–0.859) for OS, higher than those of the other subregions. However, for PFS, the C-index of ROI C was 0.678 (95% CI, 0.540–0.814), and for DSS, the C-index (0.724, 95% CI, 0.594–0.854) of ROI A is the highest. Furthermore, the results on the whole dataset are shown in **Supplementary Material E1**, Section S7, for a more comprehensive assessment.

Radiomics Nomogram Construction and Validation

Based on the multivariate Cox regression, the radiomics nomogram that incorporated the radiomics signature and the three clinical factors was constructed (**Figure 5**). **Figures 5A–C** respectively represent the prediction models with the highest C-index of OS, PFS, and DSS for GBM patients. **Figure 6** illustrates the calibration curve of the proposed nomogram based on the training cohort. Moreover, favorable calibrations (**Figure 6**) and Kaplan–Meier survival curves (**Figure 7**) confirmed the three optimal models (with the highest C-index) of the validation cohort. The Kaplan–Meier survival curves of the model corresponding to other subregions is shown in the **Supplementary Material E1**, section S5.

Clinical Utility of the Radiomics Nomogram

The decision curve for the radiomics nomogram indicates that the use of the radiomics nomogram (**Figure 8**) to stratify the survival of GBM patients was beneficial at all threshold probabilities in our study.

DISCUSSION

This study investigated the prediction of multiple survival of GBM patients based on multiparametric MRI. The radiomics signature involving multiscale texture features, combined with clinical risk factors, could precisely predict the individualized probability of survival stratification for each GBM patient. The performance differs from GBM patients in OS, PFS, and DSS. Furthermore, different tumor subregions and different modalities of MRI play a significant role in the prognosis of GBM patients.

The Radiomics Signature Combined With MultiScale Texture Features Increased the Fit and Accuracy of the Model Significantly

In our study, the incremental value when adding the radiomics signature to the clinical factor-based nomogram was assessed. According to the C-index and Decision curves, the combination of the radiomics signature and clinical predictors demonstrated an enhanced stratification efficacy in both the training and validation cohorts of the three survival indicators. The results suggested that the radiomics signature was more robust than the traditional clinical risk factors, in accordance with many previous studies focusing on radiomics nomogram (16–19).

To develop the radiomics signature, more features are added from original and derived images. More features result from the features based on Wavelet transform and LoG transform and have coefficients of higher importance that influenced the radiomics signature model in terms of survival. Previous studies have shown that multiscale texture analyses of MRI based on feature extraction can automatically predict the survival time (OS and PFS) with a precision and speed beyond the scope of human visual analysis (10, 19, 20). For example, texture features based on gray-level co-occurrence matrix (GLCM) and LoG filter extracted from relative cerebral blood volume maps in contrast-enhancing and non-enhancing regions of GBM tumors were found to predict survival outcomes of GBM patients (21). As demonstrated in this study, more radiomic features need to be added to various survival predictions of GBM (19, 22).

TABLE 2 | The Cox C-index of radiomics signature and clinical factors used for different clinical endpoints of glioblastoma patients in the training cohort and validation cohort.

Cohort	ROI	OS	PFS	DSS
Training	ROI A	0.850 (0.794–0.905)	0.812 (0.756–0.869)	0.845 (0.789–0.901)
	ROI B	0.834 (0.777–0.891)	0.792 (0.731–0.853)	0.844 (0.793–0.894)
	ROI C	0.812 (0.751–0.873)	0.787 (0.751–0.823)	0.784 (0.714–0.855)
	ROI D	0.827 (0.773–0.882)	0.801 (0.744–0.857)	0.851 (0.797–0.904)
	ROI E	0.814 (0.759–0.870)	0.780 (0.722–0.838)	0.828 (0.771–0.885)
	Clinical	0.662 (0.580–0.744)	0.580 (0.489–0.670)	0.631 (0.547–0.715)
Validation	ROI A	0.681 (0.538–0.824)	0.655 (0.516–0.793)	0.724 (0.594–0.854)
	ROI B	0.725 (0.590–0.859)	0.676 (0.540–0.811)	0.710 (0.585–0.836)
	ROI C	0.723 (0.579–0.870)	0.678 (0.540–0.814)	0.710 (0.582–0.838)
	ROI D	0.675 (0.528–0.821)	0.670 (0.533–0.805)	0.700 (0.565–0.833)
	ROI E	0.685 (0.540–0.830)	0.668 (0.528–0.807)	0.706 (0.586–0.825)
	Clinical	0.645 (0.538–0.732)	0.606 (0.527–0.705)	0.611 (0.560–0.692)

The 95% confidence interval is indicated in ().
DSS, disease-specific survival; OS, overall survival; PFS, progression-free survival; ROI, region of interest.

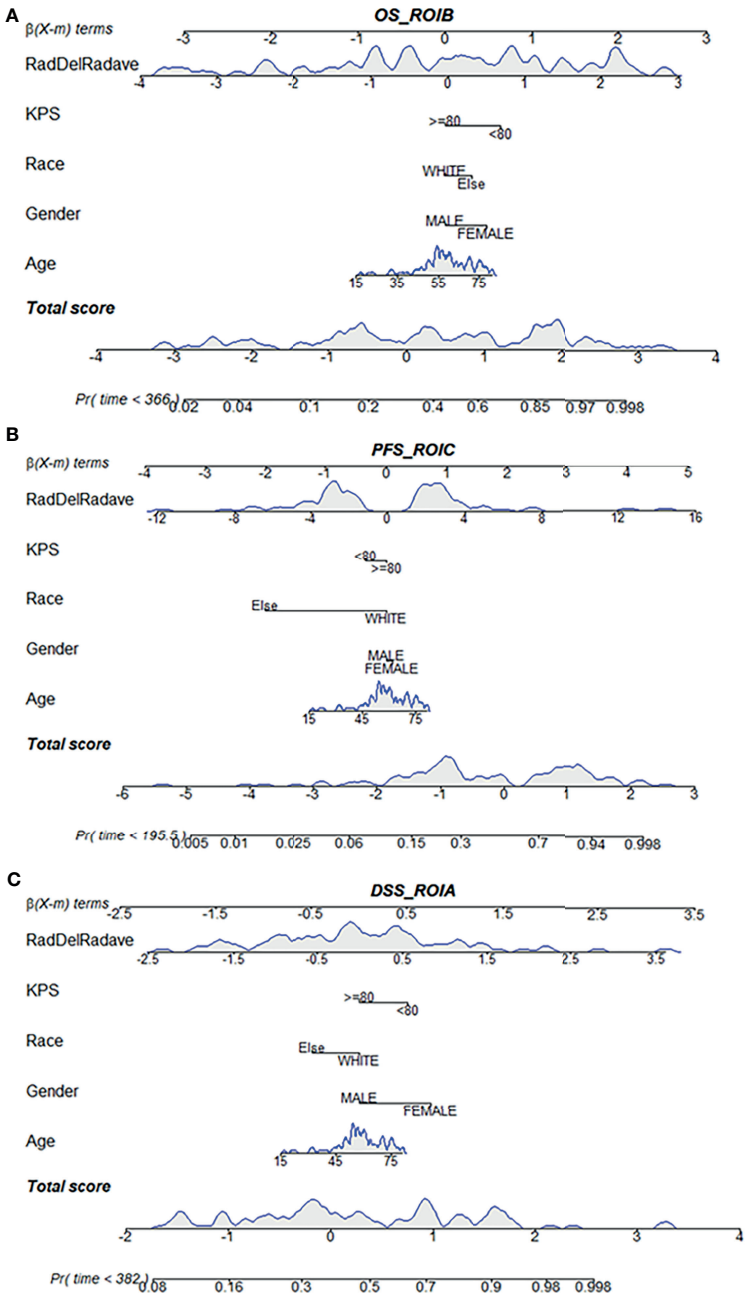


FIGURE 5 | Radiomics nomogram for the three survival indicator stratification of GBM patients. The shaded part indicates the distribution status and probability density of the patients. **(A)** Represents the OS_ROI B model, **(B)** PFS_ROI C model and **(C)** DSS_ROI A model. DSS, disease-specific survival; KPS, Karnofsky Performance Scale; OS, overall survival; PFS, progression-free survival; ROI, region of interest.

The Distinct Performance for the Three Survival Endpoints of Glioblastoma Patients

In order to evaluate and predict the survival status of GBM patients comprehensively and systematically, we examined three different survival indicators: OS, PFS, and DSS, which are the appropriate clinical endpoints for GBM (4). Among the three outcome indicators in this study, OS is an important and commonly used clinical endpoint, with the advantage that it is convenient to record, because it is not difficult to confirm the date of death and there is minimal ambiguity in defining an OS event (23, 24). However, using OS as an endpoint may weaken a clinical study, as deaths because of non-cancer causes do not necessarily reflect tumor biology, aggressiveness, or responsiveness to therapy.

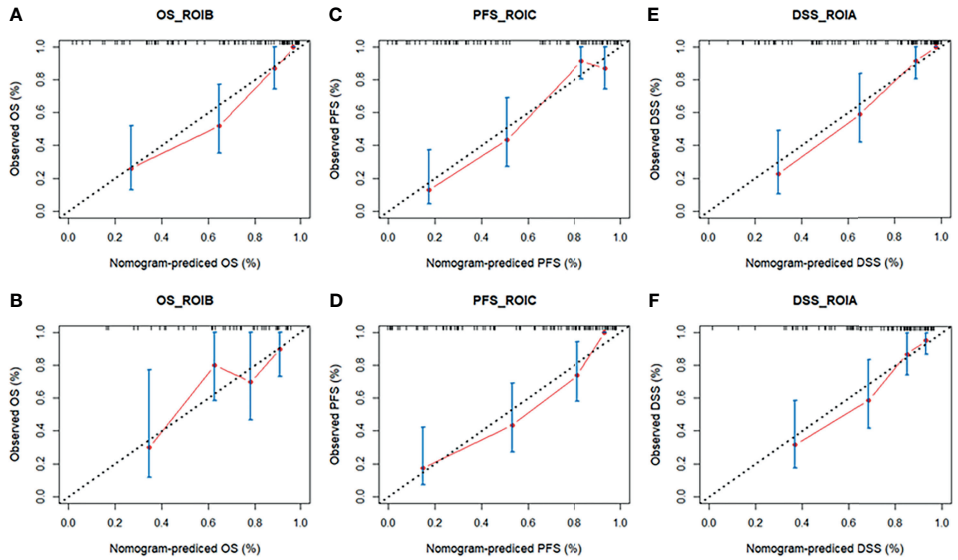


FIGURE 6 | The calibration curves of the radiomics nomogram in the training cohort **(A, C, E)** and validation cohort **(B, D, F)** of the three survival groups. The calibration curves depict the calibration of the nomogram in terms of the agreement between the predicted risk and the observed risk of survival. DSS, disease-specific survival; OS, overall survival; PFS, progression-free survival; ROI, region of interest.

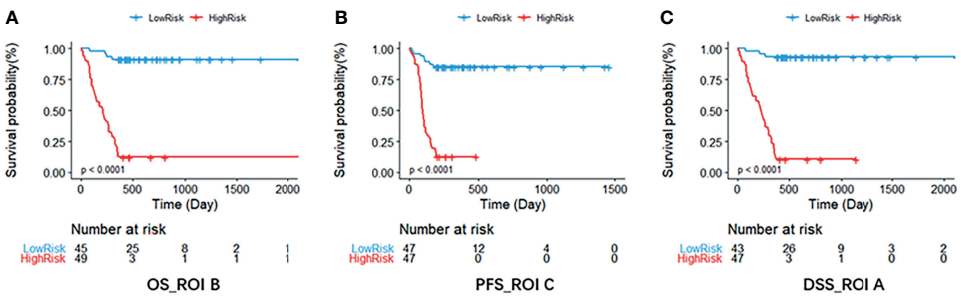


FIGURE 7 | The Kaplan–Meier survival curve shows OS **(A)**, PFS **(B)**, and DSS **(C)** risk stratification for patients with the optimal models in the validation dataset. Patients were classified as low risk and high risk according to radiomics signature. DSS, disease-specific survival; OS, overall survival; PFS, progression-free survival; ROI, region of interest.

DSS can respond to clinical benefits in a targeted manner, and its enhancement can well reflect the clinical benefits of specific diseases, and the deaths caused by specific diseases are reduced or increased. PFS increased node of "deterioration," and "deterioration" is often earlier than death, so the follow-up time of PFS is often shorter than OS and DSS. In view of the relatively short clinical follow-up records of some patients, PFS is generally considered to be a better choice of clinical endpoint than OS and DSS (4). The improvement of PFS includes "no deterioration" and "no death," which indirectly and directly reflect the clinical benefits. Some cancer-related prognostic studies have also shown that OS, PFS, and DSS are important survival indicators, which are closely related to the prediction of clinical benefits (25–29).

Our results showed that the OS and DSS results of GBM patients were relatively consistent, which may be due to the definition of DSS as death caused by a specific disease. And this study focused on glioblastoma multiforme and illustrated that GBM has fewer complications and high mortality. However, in the process of evaluating PFS, it is found that the result of PFS was biased. It is probably due to the large differences between individual patients and the greater changes in imaging characteristics for the progression of tumors. This finding is consistent with another previous study of GBM showing that the stratification of the PFS resulted in worse performance than OS (30).

Tumor Subregions Have Different Manifestations in the Three Survival Indicators

Our approach is based on comprehensive quantitative information from four different MRI sequences and six

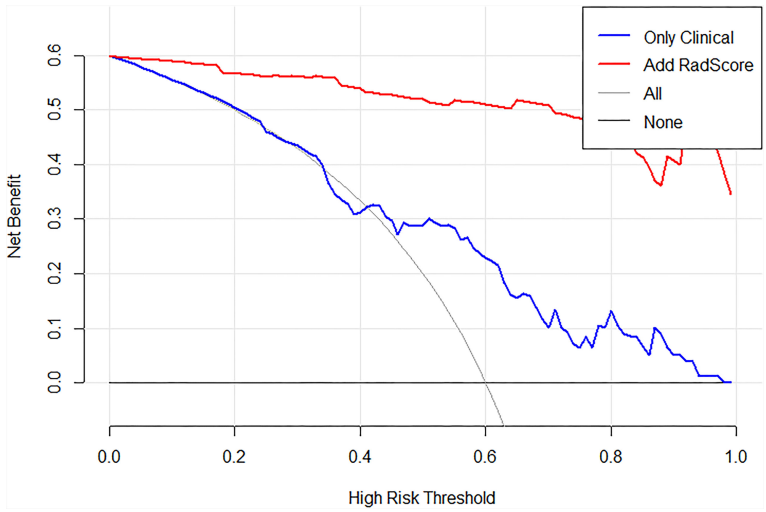


FIGURE 8 | Decision curve analysis (DCA) for the proposed radiomics nomogram, take OS_ROIB as an example. The y-axis represents the net benefit. The x-axis represents the threshold probability. The black line at the bottom represents the hypothesis that no patients had long-term survival. The blue line represents the hypothesis that all patients had long-term survival. The red line represents the net benefit of the radiomics nomogram at different threshold probabilities.

heterogeneous regions that enable a multiparametric three-dimensional characterization of the entire tumor. The selected radiomics features were from different heterogeneous regions and sequences of MRI. According to our research, the absolute values of coefficients obtained by the LASSO algorithm indicate the contribution of specific features for survival stratification.

Concerning different heterogeneous regions, the features from the enhancing tumor core (ROI B) contributed more to the OS stratification than did the features from other subregions. Most GBM prognosis-related studies indicated the association between poor prognosis and radiomics features from contrast-enhanced regions (22, 31). For PFS, our results show that ROI C has the highest C-index, which was the edema area. Some recent studies also revealed the role of features from peritumoral brain edema (3, 32). As for DSS, the C-index of region A is the highest, that is, NCR core and NET region. Previous studies have confirmed that the non-enhancement area of GBM patients is associated with their survival (9). Some recent studies also revealed the role of features from central NCR (17, 33). These results further suggested the role of information contained in non-contrast-enhanced subregions and sequences for GBM prognosis. Our study therefore suggests that subregions of GBM may complement disease stratification of patients with GBM and thereby potentially improve clinical.

Multiparametric MRI Contributed Differently to Predict Survival Stratification of Glioblastoma Patients

In addition, our results show that the selected features based on T1 sequence and FLAIR sequence account for more proportion (see **Supplementary Material E1**, section S4), which is also confirmed in previous studies (30). A comprehensive imaging–genomic analysis of human GBM by using quantitative MRI volumetrics and large-scale genetic and microRNA expression profiles demonstrated the potential for molecular subtyping based on FLAIR (or NER) signal intensity abnormality (34). In another study of high-grade gliomas, Pope et al. (35) analyzed 15 imaging variables obtained from T1-weighted MR images and showed that the presence of non-contrast-enhancing tumor was one of the three variables associated with OS.

In order to more concretely analyze the performance of single-modality and multiparameter MRI, we supplemented the comparison of their C-index of model obtained by each modality on the training set and validation set, respectively (see **Supplementary Material E1**, section S6). The experimental results show that the survival prediction model based on T1 or FLAIR sequence is indeed better than other single modality. However, the performance is not as good as that of multiparameter MRI. When comparing our results with those of older studies, it must be pointed out that radiomics analysis may reveal new insights into the underlying heterogeneity of cancers, driving a valuable prospect to noninvasively delve into GBM heterogeneity (32).

Several Limitations

Some limitations of this study need to be further investigated. First, the number of enrolled subjects is relatively small in our study, and 134 subjects in TCGA-GBM dataset were used, resulting in low robustness results. This is because we have adopted relatively strict exclusion criteria such as patients with incomplete OS, PFS, and DSS data will be excluded. However, to solve this problem, we use cross-validation in the training process of building the model, which makes our result generalizable to the population. Second, the 2D manual

segmentation used in this study may induce bias about tumor slice selection and manual ROI delineation. Furthermore, Hainc et al. (36) have investigated that the variation of slices and ROI delineation method could affect the radiomics features. These findings could be the guidance for our future work. Finally, although with high efficiency and sparsity, LASSO regression method may be less stable when a large number of features were included in the model. Other feature selection methods should be investigated in the future work.

Imaging-related limitations may result from the limited through-plane resolution of the T2 and FLAIR data compared to the higher-resolution T1 data. As a result, the assessment of fine structural details in one of the three spatial dimensions on the FLAIR data was hampered by some degree of blurring.

In conclusion, this study provides reasonable evidence of radiomics based on multiparametric MRI in assessing OS, PFS, and DSS of GBM patients. The features based on diverse regions correlate significantly with GBM survival. Disparate MRI modalities and subregions can provide distinctive but supplemental information. Compared to several survival analysis studies of GBM patients (22, 30, 32), the focus of this study was on the proposed radiomics model, which integrated radiomics signature of heterogeneous regions and three clinical predictors, and can visually and individually estimate the probability of multiple survival stratification for each GBM patient, which suggests its great potential for clinical application. In the future, prognostic research on GBM could consider focusing on the tumor regions mentioned in this paper that have a significant impact on the three survival indicators.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

BW and SZ conceptualized and designed the study, developed the methodology, and analyzed and interpreted the data. SZ, YY and XW performed the data collection or acquisition, statistical analysis. LC participated in discussion and language editing.YL, DL and TY reviewed the manuscript. All authors contributed to the interpretation of the results, drafting the manuscript work or revising it critically for important intellectual content and approval of final version to be published and agreement to be accountable for the integrity and accuracy of all aspects of the work.

FUNDING

This study was supported by the National Natural Science Foundation of China (62176177, 61873178, 61906130) and the National Key R & D Program of China (2018AAA0102604).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.778627/full#supplementary-material

REFERENCES

1. Louis DN, Perry A, Reifenberger G, Deimling AV, Figarella-Branger D, Cavenee WK, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: A Summary. *Acta Neuropathol* (2016) 131(6):803–20. doi: 10.1007/s00401-016-1545-1
2. Therese A, Dolecek T, Jennifer M, Propp J, Nancy E, Stroup N, et al. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2005-2009. *Neuro Oncol* (2012) 14(suppl 5):v1–49. doi: 10.1093/neuonc/nos218
3. Prasanna P, Patel J, Tiwari P, Madabhushi A, Partovi S. Radiomic Features From the Peritumoral Brain Parenchyma on Treatment-Naive Multi-Parametric MR Imaging Predict Long Versus Short-Term Survival in Glioblastoma Multiforme: Preliminary Findings. *Eur Radiol* (2017) 27:4188–97. doi: 10.1007/s00330-016-4637-3
4. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* (2018) 78(13 Supplement):3287. doi: 10.1158/1538-7445.AM2018-3287
5. Tai P, Yu E, Cserni G, Vlastos G, Royce M, Kunkler I, et al. Minimum Follow-Up Time Required for the Estimation of Statistical Cure of Cancer Patients: Verification Using Data From 42 Cancer Sites in the SEER Database. *BMC Cancer* (2005) 5(1):1–9. doi: 10.1186/1471-2407-5-48
6. Gutman DA, Cooper LAD, Hwang SN, Holder CA, Gao J, Aurora TD, et al. MR Imaging Predictors of Molecular Profile and Survival: Multi-Institutional Study of the TCGA Glioblastoma Data Set. *Radiology* (2013) 267(2):560–9. doi: 10.1148/radiol.13120118
7. Aerts JH. The Potential of Radiomic-Based Phenotyping in Precision Medicine: A Review. *JAMA Oncol* (2016) 2(12):1636–42. doi: 10.1001/jamaoncol.2016.2631
8. Itakura H, Achrol AS, Mitchell LA, Loya JJ, Liu T, Westbroek EM, et al. Magnetic Resonance Image Features Identify Glioblastoma Phenotypic Subtypes With Distinct Molecular Pathway Activities. *Sci Trans Med* (2015) 7(303):303ra138–303ra138. doi: 10.1126/scitranslmed.aaa7582
9. Jain R, Poisson LM, Gutman D, Scarpace L, Hwang SN, Holder CA, et al. Outcome Prediction in Patients With Glioblastoma by Using Imaging, Clinical, and Genomic Biomarkers: Focus on the Nonenhancing Component of the Tumor. *Radiology* (2014) 272(2):131691. doi: 10.1148/radiol.14131691
10. Lambin P, Leijen Aa R R, Deist TM, Peerlings J, Jong ED, Timmeren JV, et al. Radiomics: The Bridge Between Medical Imaging and Personalized Medicine. *Nat Rev Clin Oncol* (2017) 14(12):749–62. doi: 10.1038/nrclinonc.2017.141

11. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Ferté C. Promises and Challenges for the Implementation of Computational Medical Imaging (Radiomics) in Oncology. *Ann Oncol* (2017) 28(6):1191–206. doi: 10.1093/annonc/mdx034
12. Fouke SJ, Benzinger T, Gibson D, Olson JJ, Ryken TC, Kalkanis SN. The Role of Imaging in the Management of Adults With Diffuse Low Grade Glioma A Systematic Review and Evidence-Based Clinical Practice Guideline. *J Neurooncol* (2015) 125:457–79. doi: 10.1007/s11060-015-1908-9
13. Shukla G, Alexander GS, Bakas S, Nikam R, Shi W. Advanced Magnetic Resonance Imaging in Glioblastoma: A Review. *JHN Journal* (2018) 13(5):1. doi: 10.29046/JHNJ.013.1.005
14. Wu CX, Lin GS, Lin ZX, Zhang JD, Chen L. Peritumoral Edema on Magnetic Resonance Imaging Predicts a Poor Clinical Outcome in Malignant Glioma. *Oncol Lett* (2015) 10(5):2769–76. doi: 10.3892/ol.2015.3639
15. Zacharaki EI, Wang S, Chawla S, Dong SY, Wolf R, Melhem ER, et al. Classification of Brain Tumor Type and Grade Using MRI Texture and Shape in a Machine Learning Scheme. *Magn Reson Med* (2010) 62(6):1609–18. doi: 10.1002/mrm.22147
16. Cheng W, Zhang C, Ren X, Wang Z, Liu X, Han S, et al. Treatment Strategy and IDH Status Improve Nomogram Validity in Newly Diagnosed GBM Patients. *Neuro Oncol* (2017) 5):5. doi: 10.1093/neuonc/nox012
17. Ellingson BM, Harris RJ, Woodworth DC, Kevin L, Okkar Z, Mason WP, et al. Baseline Pretreatment Contrast Enhancing Tumor Volume Including Central Necrosis is a Prognostic Factor in Recurrent Glioblastoma: Evidence From Single- and Multicenter Trials. *Neuro Oncol* (2017) 1):89. doi: 10.1093/neuonc/now187
18. Haley G, Daniel L, Kattan MW, Arnab C, Gilbert MR, Lassman AB, et al. An Independently Validated Nomogram for Individualized Estimation of Survival Among Patients With Newly Diagnosed Glioblastoma: NRG Oncology RTOG 0525 and 0825. *Int J Radiat Oncol Biol Phys* (2016) 5:S92–2. doi: 10.1016/j.ijrobp.2016.06.231
19. Meldolesi E, Dinapoli N, Gatta R, Damiani A, Valentini V, Farchione A. How Can Radiomics Improve Clinical Choices? In: V Valentini, H-J Schmoll, CJH van de Velde, editors. *Multidisciplinary Management of Rectal Cancer: Questions and Answers*. Cham: Springer International Publishing (2018). p. 135–49. doi: 10.1007/978-3-319-43217-5_18
20. Lao J, Chen Y, Li ZC, Li Q, Ji Z, Jing L, et al. A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. *Sci Rep* (2017) 7:10353. doi: 10.1038/s41598-017-10649-8
21. Lee J, Jain R, Khalil K, Griffith B, Bosca R, Rao G. Texture Feature Ratios From Relative CBV Maps of Perfusion MRI Are Associated With Patient Survival in Glioblastoma. *AJNR Am J Neuroradiol* (2016) 37(1):37–43. doi: 10.3174/ajnr.A4534
22. Chaddad A, Sabri S, Niazi T, Abdulkarim B. Prediction of Survival With Multi-Scale Radiomic Analysis in Glioblastoma Patients. *Med Biol Eng Comput* (2018) 56(12):1–14. doi: 10.1007/s11517-018-1858-4
23. Hudis CA, Barlow WE, Costantino JP, Gray RJ, Pritchard KI, Chapman JAW, et al. Proposal for Standardized Definitions for Efficacy End Points in Adjuvant Breast Cancer Trials: The STEEP System. *J Clin Oncol* (2007) 25 (15):2127–32. doi: 10.1200/JCO.2006.10.3523
24. Punt CJA, Buyse M, Köhne C-H, Hohenberger P, Labianca R, Schmoll HJ, et al. Endpoints in Adjuvant Treatment Trials: A Systematic Review of the Literature in Colon Cancer and Proposed Definitions for Future Trials. *J Natl Cancer Inst* (2007) 99(13):998. doi: 10.1093/jnci/djm024
25. Ahluwalia P, Ahluwalia M, Mondal AK, Sahajpal N, Kolhe R. Prognostic and Therapeutic Implications of Extracellular Matrix Associated Gene Signature in Renal Clear Cell Carcinoma. *Sci Rep* (2021) 11(1):7561. doi: 10.1038/s41598-021-86888-7
26. Li J, Wang J, Chen R, Bai Y, Lu X. The Prognostic Value of Tumor-Infiltrating T Lymphocytes in Ovarian Cancer. *Oncotarget* (2017) 8(9):15621–31. doi: 10.18632/oncotarget.14919
27. Mclendon R, Friedman A, Bigner D, Meir EV, Brat DJ, Mastrogianakis GM, et al. Comprehensive Genomic Characterization Defines Human Glioblastoma Genes and Core Pathways. *Nature* (2008) 455(7216):1061–8. doi: 10.1038/nature07385
28. Pan J. LAYN Is a Prognostic Biomarker and Correlated With Immune Infiltrates in Gastric and Colon Cancers. *Front Immunol* (2019) 10. doi: 10.3389/fimmu.2019.00006
29. Takaya Y, Yoshihide K, Rei U, Noriyuki K, Haruo M, Kazuya T, et al. Stereotactic Body Radiotherapy for Kidney Cancer: A 10-Year Experience From a Single Institute. *J Radiat Res* (2021) 3:3. doi: 10.1093/jrr/rrab031
30. Kickingereder P, Burth S, Wick A, Gotz M, Eidel O, Schlemmer HP, et al. Radiomic Profiling of Glioblastoma: Identifying an Imaging Predictor of Patient Survival With Improved Performance Over Established Clinical and Radiologic Risk Models. *Radiology* (2016) 280(3):880–9. doi: 10.1148/radiol.2016160845
31. Brynolfsson P, Nilsson D, Henriksson R, Hauksson J, Karlsson M, Garpebring A, et al. ADC Texture—An Imaging Biomarker for High-Grade Glioma? *Med Phys* (2014) 41(10):101903. doi: 10.1118/1.4894812
32. Wang K, Wang Y, Fan X, Wang J, Li G, Ma J, et al. Radiological Features Combined With IDH1 Status for Predicting the Survival Outcome of Glioblastoma Patients. *Neuro Oncol* (2016) 18(4):589. doi: 10.1093/neuonc/nov239
33. Liu S, Wang Y, Xu K, Wang Z, Fan X, Zhang C, et al. Relationship Between Necrotic Patterns in Glioblastoma and Patient Survival: Fractal Dimension and Lacunarity Analyses Using Magnetic Resonance Imaging. *Sci Rep* (2017) 7:8302. doi: 10.1038/s41598-017-08862-6
34. Zinn PO, Bhanu M, Pratheesh S, Singh SK, Sadhan M, Jolesz FA, et al. Radiogenomic Mapping of Edema/Cellular Invasion MRI-Phenotypes in Glioblastoma Multiforme. *PloS One* (2011) 6(10):e25451. doi: 10.1371/journal.pone.0025451
35. Pope WB, Sayre J, Perlina A, Villablanca JP, Cloughesy TF. MR Imaging Correlates of Survival in Patients With High-Grade Gliomas. *Ajnr Am J Neuroradiol* (2005) 26(10):2466–74. doi: 10.1080/02841850500216293
36. Hainc N, Stippich C, Stieltjes B, Leu S, Bink A. Experimental Texture Analysis in Glioblastoma. *Invest Radiol* (2017) 52(6):367–73. doi: 10.1097/RLI.0000000000000354

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Wang, Zhang, Wu, Li, Yan, Liu, Xiang, Li and Yan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 01 December 2021
doi: 10.3389/fonc.2021.699812

Multi-Phase CT-Based Radiomics Nomogram for Discrimination Between Pancreatic Serous Cystic Neoplasm From Mucinous Cystic Neoplasm

Jiahao Gao[1,2†], *Fang Han*[1,2†], *Xiaoshuang Wang*[1], *Shaofeng Duan*[3] *and Jiawen Zhang*[1,2*]

[1] Department of Radiology, Huashan Hospital North, Fudan University, Shanghai, China, [2] Department of Radiology, Huashan Hospital, Fudan University, Shanghai, China, [3] Department of Life Sciences, GE Healthcare, Shanghai, China

OPEN ACCESS

Edited by:
Joseph Ross Mitchell, Independent Researcher, Tampa, FL, United States

Reviewed by:
Jingqin Fang, Army Medical University, China
Gerardo Blanco-Fernández, Universidad de Extremadura, Spain

***Correspondence:**
Jiawen Zhang
zhangjw2000@126.com

[†]*These authors have contributed equally to this work*

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *24 April 2021*
Accepted: *15 November 2021*
Published: *01 December 2021*

Citation:
Gao J, Han F, Wang X, Duan S and Zhang J (2021) Multi-Phase CT-Based Radiomics Nomogram for Discrimination Between Pancreatic Serous Cystic Neoplasm From Mucinous Cystic Neoplasm. Front. Oncol. 11:699812. doi: 10.3389/fonc.2021.699812

Purpose: This study aimed to develop and verify a multi-phase (MP) computed tomography (CT)-based radiomics nomogram to differentiate pancreatic serous cystic neoplasms (SCNs) from mucinous cystic neoplasms (MCNs), and to compare the diagnostic efficacy of radiomics models for different phases of CT scans.

Materials and Methods: A total of 170 patients who underwent surgical resection between January 2011 and December 2018, with pathologically confirmed pancreatic cystic neoplasms (SCN=115, MCN=55) were included in this single-center retrospective study. Radiomics features were extracted from plain scan (PS), arterial phase (AP), and venous phase (VP) CT scans. Algorithms were performed to identify the optimal features to build a radiomics signature (Radscore) for each phase. All features from these three phases were analyzed to develop the MP-Radscore. A combined model comprised the MP-Radscore and imaging features from which a nomogram was developed. The accuracy of the nomogram was evaluated using receiver operating characteristic (ROC) curves, calibration tests, and decision curve analysis.

Results: For each scan phase, 1218 features were extracted, and the optimal ones were selected to construct the PS-Radscore (11 features), AP-Radscore (11 features), and VP-Radscore (12 features). The MP-Radscore (14 features) achieved better performance based on ROC curve analysis than any single phase did [area under the curve (AUC), training cohort: MP-Radscore 0.89, PS-Radscore 0.78, AP-Radscore 0.83, VP-Radscore 0.85; validation cohort: MP-Radscore 0.88, PS-Radscore 0.77, AP-Radscore 0.83, VP-Radscore 0.84]. The combination nomogram performance was excellent, surpassing those of all other nomograms in both the training cohort (AUC, 0.91) and validation cohort (AUC, 0.90). The nomogram also performed well in the calibration and decision curve analyses.

Conclusions: Radiomics for arterial and venous single-phase models outperformed the plain scan model. The combination nomogram that incorporated the MP-Radscore,

tumor location, and cystic number had the best discriminatory performance and showed excellent accuracy for differentiating SCN from MCN.

Keywords: pancreatic cystic neoplasm, radiomics, nomogram, contrast-enhanced computed tomography (CECT), texture analysis

INTRODUCTION

Pancreatic cystic neoplasms (PCNs) have been increasingly diagnosed in recent years as a direct result of the extensive use of abdominal cross-sectional imaging. The prevalence of incidentally discovered PCNs in the general population has been reported to range from 2.6 to 19.6% (1, 2). Considerable attention has been focused on serous cystic neoplasms (SCNs) and mucinous cystic neoplasms (MCNs) because of the significant difference in the probability of malignant transformation between the two (3). SCNs have an extremely low incidence of malignancy (4). The current management strategy for SCN is conservative, based on regular surveillance with rare interventions performed only because of symptoms (5, 6). MCNs are diagnosed almost exclusively in middle-aged women, but with a very definite potential for malignant transformation (7–9). In contrast to SCN, surgical resection has been advocated for many, if not most, MCN patients. Recognizing the marked difference in the risk of malignancy and the consequent nearly opposite clinical management strategies between these cystic neoplasms, it is vital to correctly discriminate between the two.

Currently, even the high-quality imaging modalities such as computed tomography (CT) and ultrasound do not provide adequate discrimination between SCN and MCN (10, 11). Clearly, radiological imaging approaches, especially multi-detector computed tomography (MDCT), play a pivotal role in the preoperative diagnosis of PCNs. It has been reported that the discrimination efficacy of CT for SCNs was ranged from 27 to 91% (12, 13). Compared with CT, MRI/magnetic resonance cholangiopancreatography (MRCP) could further improve the diagnostic accuracy of PCNs, with an accuracy of 40–95% (14, 15) providing a better view of the pancreatic duct system and allowing to detect the presence of a solid component or mural nodule. Endoscopic ultrasound with fine-needle aspiration (EUS-FNA) has become a promising tool for classifying specific subtypes of PCNs. Adding EUS-FNA to CT and MRI has improved the diagnostic accuracy by 36% and 54%, respectively (16). Balanced against these potential benefits is the invasive nature of EUS and the variable risk of FNA-associated complications (17). These limitations and the significant cost curtail their application for the routine evaluation of PCNs (18).

Radiomics is an emerging and rapidly developing method for advanced image analysis. Relative to PCNs, radiomics has been successfully applied to the entire spectrum of the disease process, including differential diagnosis, malignant assessment, and prognosis prediction (19–21). Most radiomics studies utilizing MDCT pancreatic scans have been limited to the venous phase (VP) for feature extraction. Clearly, the different phases reflect unique vascular enhancement and texture information. Logically, a radiomics model constructed from a plain scan (PS) or arterial phase (AP) scan would augment diagnostic efficiency. To the best of our knowledge, no prior studies have reported feature extraction of all three phases of contrast-enhanced CT to discriminate PCN subtypes.

Our aim was to compare the predictive efficacy of each single-phase radiomics model, and then to construct a combination nomogram, incorporating a multi-phase (MP) radiomics model with clinical imaging factors that would noninvasively and accurately discriminate SCNs from MCNs.

MATERIALS AND METHODS

Patient Population

The Institutional Review Board of Huashan Hospital of Fudan University approved this retrospective study, and the requirement for informed consent was waived. Patients who were diagnosed with SCNs or MCNs for whom surgical resection was performed in our hospital between January 2011 and December were enrolled in this study. The inclusion criteria were: (1) SCNs or MCNs with surgical pathologic confirmation; (2) contrast-enhanced CT scans (slice thickness: 1.5 mm) performed within one month prior to pancreatic surgery. The exclusion criteria were: (1) CT images with serious artifacts and (2) patients whose radiomics features could not be successfully extracted from the CT images. The details of patient enrollment are shown in **Figure S1** in **Supplementary Materials**. The final study group comprised 115 patients with SCNs and 55 with MCNs. Patients were randomly grouped in a ratio of 7:3, with 120 and 50 patients in the training and validation cohorts, respectively.

Patient demographic and clinical information was collected from the hospital medical record system. Demographic information (age and sex) and eight imaging factors known to be valuable in distinguishing SCNs from MCNs from previous studies were selected as the basis for constructing the clinical model (22, 23). Two radiologists with considerable experience in abdominal imaging (13 and 6 years, respectively) evaluated the features in consensus including: (1) lesion size, (2) tumor location (head, neck, body, and tail), (3) cyst number (single or multiple), (4) calcification (absent or present), (5) septation (absent or present), (6) lesion shape (oval or irregular lobulation), (7) wall enhancement (absent or present), and (8) mural nodules (absent or present). Both radiologists were blinded to the correlative pathological details. Among these, the lesion size was outlined and decided unanimously by two doctors simultaneously, while the other features were assessed by

each doctor, and the results were derived separately. If the two radiologists did not agree on a specific feature in the same patient, a third expert with 23 years' experience in abdominal radiology reviewed the features and helped establish the final decision. The inter-reader agreement of imaging factors was also assessed, as shown in **Supplementary Materials** (**Table S1**). The framework of the study is shown in **Figure 1**.

Image Acquisition

CT examinations of all patients were performed using the same 256-slice CT system (Brilliance iCT, Philips Medical Systems, The Netherlands). All pancreatic CT images were acquired using a standard dual-phase scanning protocol. The CT scan parameters were as follows: 120 kV; 150–200 mAs; rotation time, 0.5-0.75 s; collimation, 128×0.625 mm; matrix, 512×512; and slice thickness, 1.5 mm. An anionic contrast agent (370 mgI/mL, Iopamidol-370, GEhealthcare, Princeton, NJ) was administered at a dose of 1.5 mL/kg, 3.0 mL/s. AP images were obtained 30 s after the injection of contrast agent, and VP images were obtained 45 s after the AP acquisition. All images were downloaded from the hospital archives.

Tumor Segmentation and Single-Phase Radiomics Feature Extraction

PS, AP, and VP CT images in each patient were used for feature extraction. The window width and window level were 300 and 40 HU, respectively. For each phase, one radiologist (13 years' experience in abdominal imaging) segmented the lesion contour on each slice using open-source software (3D Slicer version 4.11.0; Boston, MA). With the technical support of a radiomics software based on Python (Pyradiomics version 3.0.0; https://github.com/Radiomics/pyradiomics) (24), radiomics features were extracted in three-dimensional volume for each phase. The extracted features were classified into six categories: (1) shape features, (2) first order statistics, (3) gray level co-occurrence matrix features, (4) gray-level run length matrix features, (5) gray-level size zone matrix features, and (6) gray-level dependence matrix features. Details of the features are provided in **Supplementary Materials I**.

To estimate both intra- and inter-observer reproducibility of extracted features, 60 patients were randomly chosen for a repeat region of interest (ROI) segmentation at 30 days following the initial segmentation, performed by the same radiologist and an additional one (with 6 years' experience in abdominal imaging). The radiologists were blinded to the associated clinical and pathological information. The intra- and inter-class correlation coefficients (ICCs) were used to evaluate feature reliability (25).

Feature Selection, Single-Phase Radiomics Signature Construction, and Performance Comparison

In the training cohort, a three-step procedure was developed to select the radiomics features extracted in each phase. First, features with both intra- and inter-ICC less than 0.75, were

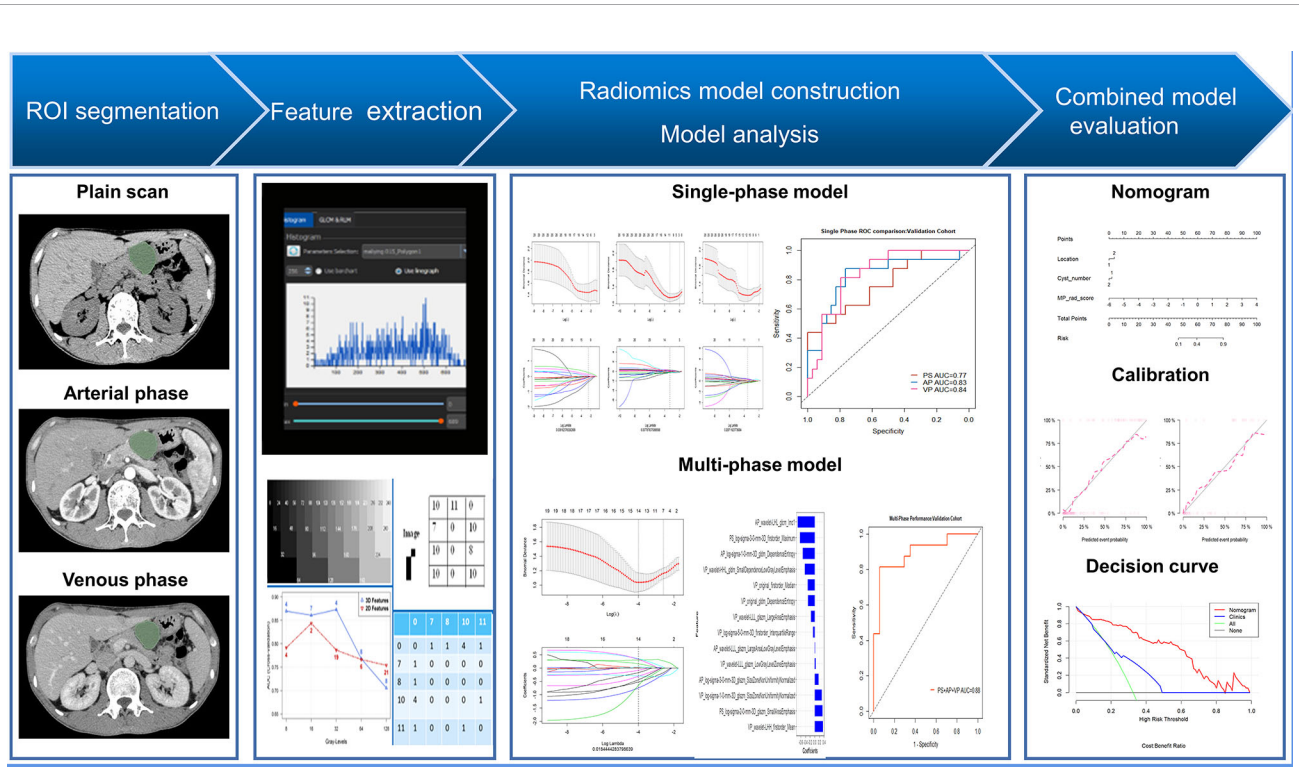


FIGURE 1 | Radiomics workflow.

excluded from this process. The mRMR method and the least absolute shrinkage and selection operator (LASSO) algorithm were used to select the most robust and optimal features to construct the single-phase radiomics model. The selected optimal features were then combined with its coefficient in the LASSO regression to construct the radiomics signature: Radscores (including PS-Radscore, AP-Radscore, VP-Radscore). The Mann-Whitney U test was used to evaluate the discrimination capability of the each-phase Radscore. We also used receiver operating characteristic (ROC) curve analysis and area under the curve (AUC) values to compare the performance of the single-phase radiomics signature. The detailed performance of each of the radiomics signatures is shown in **Figure 2** and **Supplementary Materials II**. We also constructed and evaluated the two-phase combined radiomics model (**Supplementary Materials III**).

Combined Model Building and Nomogram Development

A MP radiomics feature set was developed by integrating all 3654 (1218*3) features of the three phases. We then used the same three-step feature extraction method to obtain an MP radiomics signature, MP-Radscore. The discrimination capability of the MP radiomics model was also evaluated using the Mann-Whitney U test and ROC curve analysis. Univariate analysis was conducted to estimate the differences between SCN and MCN patients for each clinical and imaging feature. In the training cohort, variables with $P < 0.100$, in the univariate regression, were then allocated to a multivariable logistic regression. The clinical model was constructed by incorporating factors with $P < 0.100$ in the multivariate analysis (26). Finally, a combination multivariate logistic model was constructed using MP-Radscores together with selected clinical imaging factors. Variance inflation factor (VIF) analysis was performed on the combination model to further reduce the probability of overfitting. The nomogram was developed to visualize the optimal model, specifically to score each patient and quantify the degree of disease tendency.

Model Validation and Clinical Use Evaluation

The combination model was first evaluated in the training cohort (n = 120) and subsequently confirmed in the validation cohort (n = 50). ROC curves and AUC values were used to evaluate the discriminatory performance of the combined models. Calibration curves and the Hosmer-Lemeshow test were conducted to estimate the consistency between the predictive results of the combination model and expected probabilities. We also used the Delong test to compare the predictive efficiency between the combination model and the venous radiomics approach to confirm the advances of our combination model.

Decision curve analysis (DCA) was performed to determine the clinical value of the nomogram and calculate the net benefits of the models at different threshold probabilities (27).

Statistical Analysis

Continuous variables are presented as means and standard deviations. Student's t-test and chi-square test were employed to evaluate the statistical differences in continuous and discrete variables, respectively. In the ROC test, accuracy, sensitivity, and specificity at the cutoff value were calculated to evaluate the efficiency of the radiomics model, clinical model, and the combination model. The inter-reader agreement of imaging factors was assessed using the kappa test, and the simple kappa coefficient was used as an assessment criterion for consistency. A two-tailed P value less than 0.05, was deemed as statistically significant. All statistical analyses were performed using the R software (version 3.6.3). The R packages and the main code used are included in **Supplementary Materials V**.

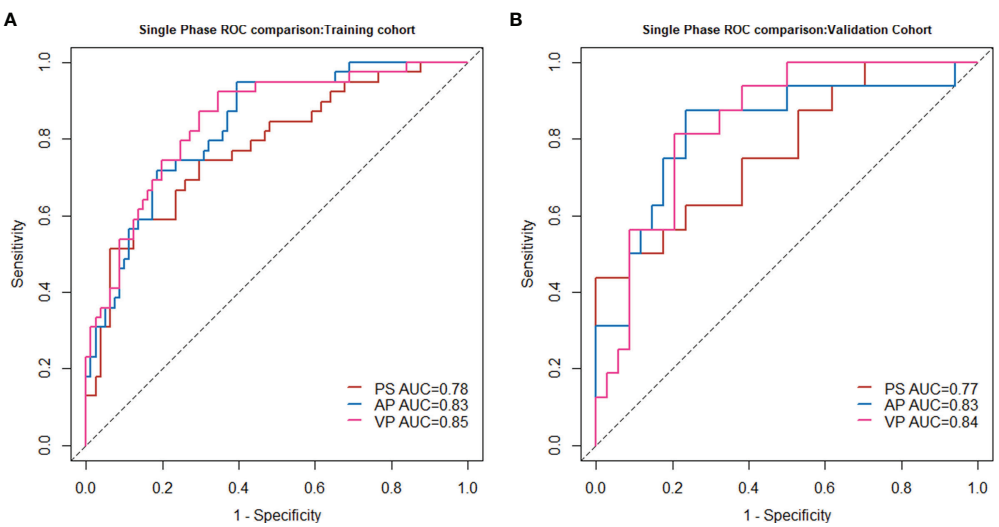


FIGURE 2 | Single-phase radiomics model performance in the training cohort **(A)** and validation cohort **(B)**.

RESULTS

Patients Characteristics

Patient characteristics are summarized in **Table 1**. There were 120 patients (SCN: 81, MCN: 39) in the training cohort and 50 (SCN: 34, MCN: 16) in the validation cohort. No significant differences were found in the clinical and imaging features between the two cohorts.

Feature Selection, Single-Phase Radiomics Signature Construction, and Performance Comparison

From each phase of contrast-enhanced CT scans, 1218 features were extracted. After ICC assessment applying the minimum criteria for intra- or inter-ICC values, 0.75, 765, 922, and 840 features remained from the PS, AP, and VP CT images, respectively. Of these, 30 features from each phase were retained using the mRMR algorithm for LASSO regression. Finally, 11,11, and 12 radiomics features were selected to construct the radiomics signature for the radiomics model of PS, AP, and VP CT scans, respectively. Features were identified using multivariate logistic regression analysis to construct single-phase radiomics signatures (PS-Radscore, VP-Radscore, and AP-Radscore). The formulas for calculating the Radscores and features selected for the single-phase radiomics model are presented in **Supplementary Materials II**.

In all three single-phase Radscores, there was a significant difference between SCN and MCN patients in the training cohort ($P < 0.010$), and importantly, this was confirmed in the independent validation cohort ($P < 0.010$). The PS, AP, and VP radiomics models yielded AUC values of 0.78, 0.83, and 0.85, respectively, for the training cohort, and 0.77, 0.83, and 0.84, respectively, for the validation cohort. The AUC values in the radiomics model in AP and VP were similar and higher than those in the radiomics model of the plain scan. The performance of the single-phase radiomics model is shown in **Figure 2**.

Combined Model Building and Nomogram Development

Using the three-step selection process described above for single-phase radiomics model construction, 14 features (including 2 in PS, 4 in AP, and 8 in VP) were similarly selected from the MP radiomics feature set (**Figure 3**). The MP-Radscore was built to improve the discrimination efficacy of the MP radiomics model. ROC curves showed that the MP radiomics model performed better than the models based on single CT phase (AUC: 0.89 and 0.88 in the training and validation cohorts, respectively). The performance of the MP radiomics model in the Mann-Whitney U test and ROC curves are shown in **Figure 4**. The detailed calculation formulas of the MP-Radscore and combined nomogram are included in **Supplementary Materials IV**.

In the univariate analysis of the clinical model building, only tumor location and cyst number were significantly correlated with pathologic results ($P < 0.100$). Tumor location and cyst number were statistically significant ($P < 0.100$) in the multivariate logistic regression analysis, therefore comprising the clinical model. The results of the univariate and multivariate logistic regression

TABLE 1 | Characteristics of Patients in the Training and Validation Cohorts.

Characteristics	Training Cohort (n = 120)			Validation Cohort (n = 50)		
	SCN (n = 81)	MCN (n=39)	*P*	SCN (n = 34)	MCN (n = 16)	*P*
Age, mean ± SD	51.2 ± 11.5	50.8 ± 14.9	.890	56.4 ± 14	53.9 ± 11.1	.523
Sex, No (%)						
Male	15 (18.5)	3 (7.7)	.199	11 (36.7)	2 (12.5)	.251
Female	66 (81.5)	36 (92.3)		23 (63.3)	14 (87.5)	
Lesion size (cm)	4.0 ± 2.5	4.7 ± 2.1	.113	4.3 ± 2.4	5.5 ± 2.7	.105
Tumor location						
Head and neck	37 (45.7)	9 (23.1)	.029	13 (38.2)	13 (81.2)	.011
Body and tail	44 (54.3)	30 (76.9)		21 (61.8)	3 (18.8)	
Cyst number						
Single	50 (61.7)	33 (93.9)	.019	16 (47.1)	12 (75.0)	.120
Multiple	31 (38.3)	6 (6.1)		18 (52.9)	4 (25.0)	
Calcification						
Absent	53 (65.4)	25 (64.1)	1.00	20 (58.8)	12 (75.0)	.426
Present	28 (34.6)	14 (35.9)		14 (41.2)	4 (25.0)	
Septation						
Absent	40 (49.4)	14 (35.9)	.232	11 (32.4)	8 (50.0)	.375
Present	41 (50.6)	25 (64.1)		23 (67.6)	8 (50.0)	
Lesion shape						
Oval	50 (61.7)	27 (69.2)	.548	20 (58.8)	12 (75.0)	.426
Irregular lobulation	31 (38.3)	12 (30.8)		14 (41.2)	4 (25.0)	
Wall enhancement						
Absent	57 (70.4)	26 (64.1)	1.00	23 (67.6)	11 (68.8)	1.00
Present	24 (29.6)	13 (35.9)		11 (32.4)	5 (31.2)	
Mural nodules						
Absent	64 (79.0)	35 (89.7)	.233	23 (67.6)	15 (93.8)	.096
Present	17 (21.0)	4 (10.3)		11 (32.4)	1 (6.2)	

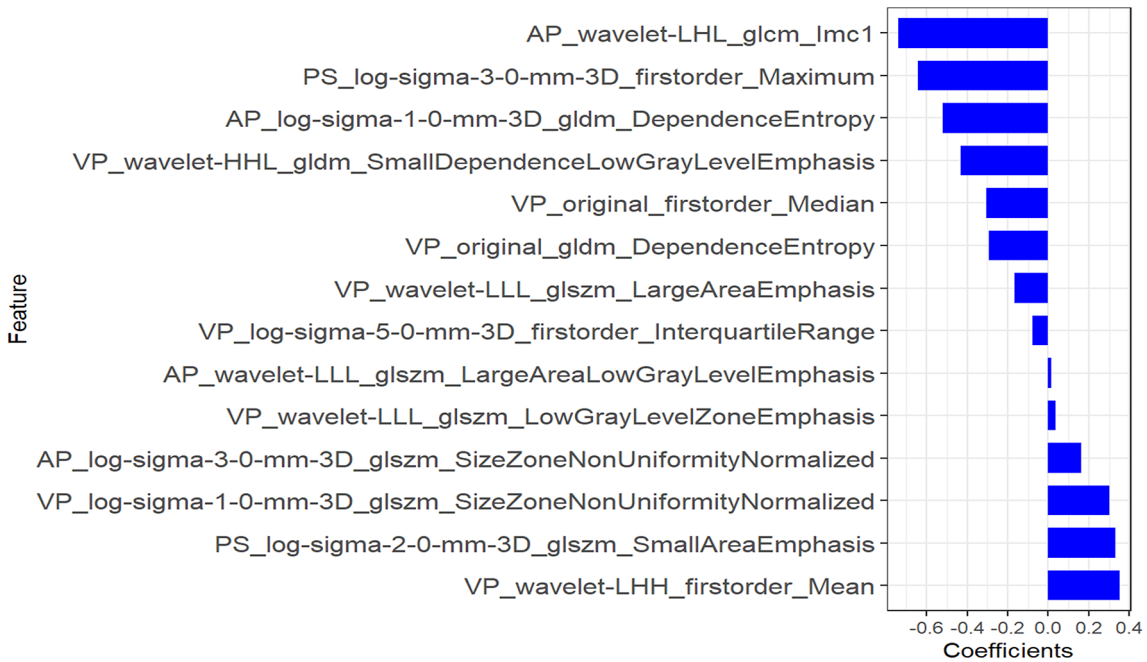


FIGURE 3 | Features selected for multi-phase radiomics model construction.

analyses are shown in **Table 2**. The combination model was constructed by incorporating the MP-Radscore, tumor location, and cyst number. A nomogram was established to visualize the combined model **(Figure 5A).**

Combination Model Validation and Clinical Use Evaluation

The combination nomogram exhibited best predictive performance (AUC: 0.91 and 0.90 in the training and validation cohorts, respectively) for the discrimination between SCNs and MCNs (**Figures 5B, C** and **Table 3**). The Delong test demonstrated statistical differences in AUC values between the combination nomogram and the clinical model ($P < 0.010$). Significant differences were also found in the ROC curves between the combination nomogram and VP model ($Z = 1.962$, $P = 0.0497 < 0.0500$) in the validation cohort. Calibration curves (**Figure S6**) revealed good agreement between the predictive and observation probabilities of our combination nomogram ($P = 0.480$ and 0.582 for the training and validation cohorts, respectively).

The decision curve analysis indicated that the combination nomogram provided a net benefit over either a "treat-all" or "treat-none" strategy, and the clinical model at a threshold probability over 10% (**Figure 6**). The combination nomogram demonstrated excellent clinical practicality.

DISCUSSION

In this retrospective study, we constructed and validated an MP CT-based radiomics nomogram to differentiate SCN from MCN. The combination model, incorporating the MP radiomics model plus clinical imaging factors, exhibited better diagnostic performance than any of the single-phase radiomics models or a clinical model alone did. The decision curve analysis also confirmed that the combination model achieved better discriminatory accuracy than the clinical model did. Relating specifically to the single-phase performance comparison, the radiomics model of the AP and the VP performed better than the PS model in terms of AUC values.

The exact morphologic details of MDCT are crucial to exclude tumor invasion of PCNs. Key imaging morphologic factors (tumor size, location, lesion shape, calcification, segmentation, etc.) derived from pathologic characteristics form the basis for radiologic differentiation of PCN subtypes (28, 29). Nevertheless, the diagnostic accuracy of cross-sectional imaging, such as MDCT, still falls short of ideal discrimination (30). Therefore, intrusive methods, such as EUS-FNA, have been developed to add diagnostic precision for preoperative PCN subtyping. Clearly, achieving this degree of accuracy requires highly skilled endoscopists and cytologists (31, 32).

Apart from the invasive techniques described above, radiomics offers a promising noninvasive technology intended to achieve similar results. We successfully established a combination radiomics model and achieved superior capacity to differentiate SCNs from MCNs. Among numerous clinical imaging factors, only cyst number and tumor location were statistically essential to be included in the combination model. Considering that the clinical and imaging features of PCNs pathologically diagnosed as SCN in the study were not consistent with typical SCN manifestations, we also analyzed

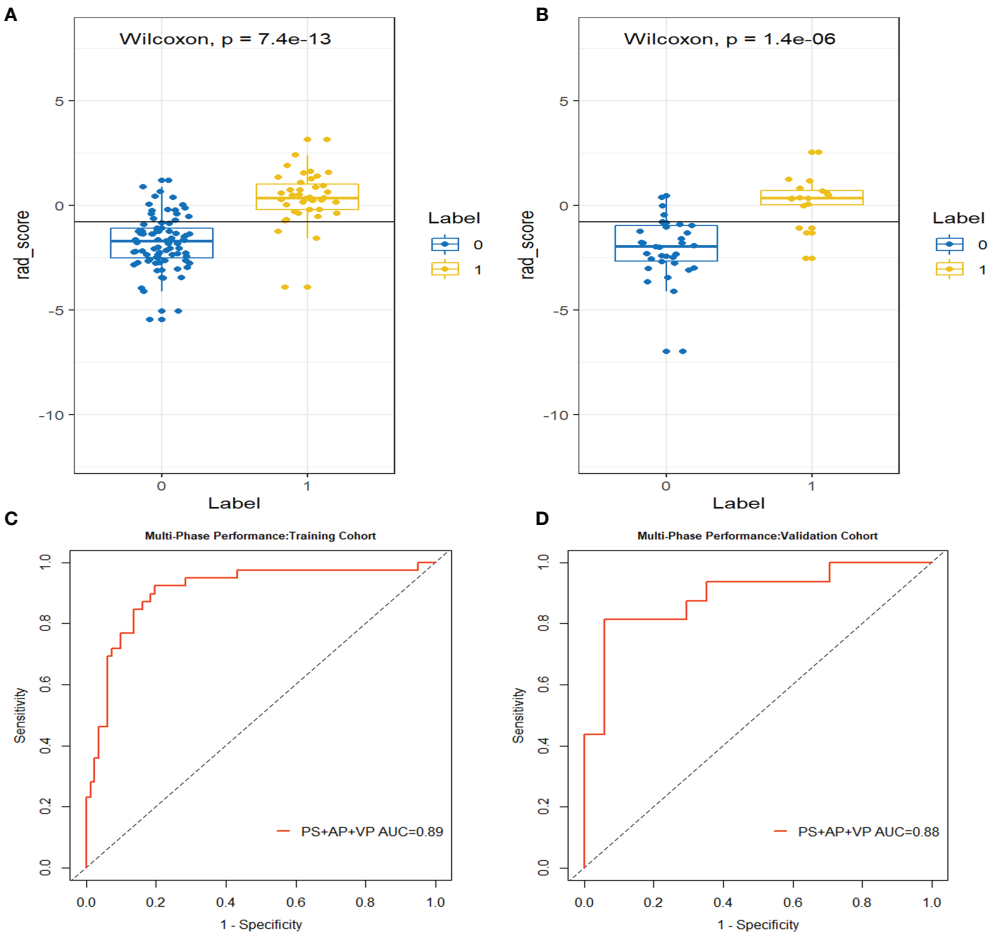


FIGURE 4 | The box-dot plots of the multi-phase radiomics model in the training cohort **(A)** and the validation cohort **(B)**. The orange markers indicate patients with MCN while the blue markers indicate patients with SCN. The ROC curves for the multi-phase radiomics model in the training cohort **(C)** and the validation cohort **(D)**.

and modeled these features. These results are consistent with a number of previous CT imaging studies. The differences in morphology that discriminate between SCNs and MCNs are limited to tumor location, lobular contour, and a large number of cysts (30, 33). The prediction accuracy of the clinical model alone was poor, with AUC values of only 0.69 and 0.63 in the training group and validation group, respectively. Even with high-quality CT scans interpreted by skilled radiologists, the accuracy of PCN subclassification remains disappointing.

A recent study assessed the discriminatory efficacy of conventional CT imaging features in distinguishing SCN from MCN and presented it by building a nomogram based on multivariate logistic regression (34). In contrast, our study not only considered conventional clinic-radiological features, but also incorporated radiomics features that reflected the deeper dimensional information of the images to construct a comprehensive model. The results showed that the combined model demonstrated better predictive ability than the clinical

TABLE 2 | Variables Elected for Combined Model and Clinical Model.

Intercept & Variable	Combined Model (95%CI)		Clinical Model (95%CI)	
	Odds Ratio	P Value	Odds Ratio	P Value
Intercept	0.74 (0.05,10.13)	<0.01	0.45 (0.06, 3.11)	<0.01
Tumor location	1.69 (0.54, 5.42)	0.05	2.65 (1.62, 2.70)	0.03
Cyst number	0.76 (0.23, 2.48)	0.01	0.30 (0.11, 0.75)	<0.01
MP-Radscore	4.25 (2.58, 7.96)	<0.01	NA	NA

NA, not available.

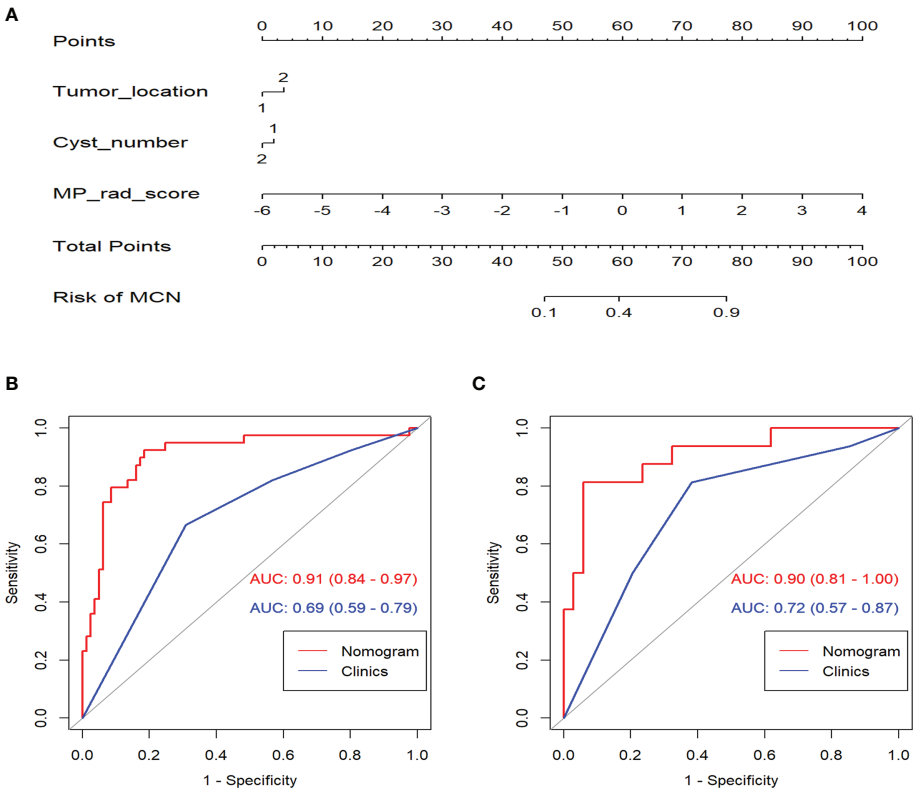


FIGURE 5 | **(A)** The nomogram established for the combined model. ROC curves comparison between the nomogram and clinical model in the training **(B)** and validation cohorts **(C)**.

model alone. Several previous studies have applied radiomics to the differentiation of PCNs and have achieved good results (35, 36). However, further validation is required because of the limited amount of data (N < 80) and the high risk of overfitting. Moreover, a nomogram has not yet been established to visualize the radiomics model. Finally, they performed feature extraction almost uniquely on VP CT images. In the present study, in addition to VP CT scans, we also investigated the radiomics signatures on plain and AP CT images. The VP-Radscore AUC value was the best among the three single-phase radiomics models; similarly, the radiomics models of both the VP and AP had superior AUC values compared to those of the PS model. These results require verification, because they rely heavily on the experience of the radiologist performing manual segmentation (22). Interestingly, several radiomics studies have constructed radiomics models from only PS and have achieved good results in disease prediction (37, 38). The feature composition of the MP radiomics model included 8 features (57.1%) in the VP, 4 in the AP (28.6%), and only 2 (14.3%) in the PS. Our highest quantitative ranking of the VP is consistent with most previously published pancreatic radiomics research, while the PS was used less frequently. Therefore, this study is also significant in that it provides preliminary insight into the effect of contrast-enhanced CT scan phase on the predictive efficacy of imaging histology models and establishes a more comprehensive model to summarize various types of risk factors for prediction.

TABLE 3 | Diagnostic performance of models in the training and validation cohorts.

Models	Training Cohort (n = 120)				Validation Cohort (n = 50)			
	Sensitivity	Specifity	Accuracy (95%CI)	AUC (95%CI)	Sensitivity	Specifity	Accuracy (95%CI)	AUC (95%CI)
Clinical model	0.67	0.69	0.68 (0.59, 0.76)	0.69 (0.59, 0.79)	0.50	0.79	0.70 (0.55, 0.82)	0.72 (0.57, 0.87)
MP-Radscore	0.91	0.80	0.84 (0.76, 0.90)	0.89 (0.82, 0.95)	0.81	0.88	0.82 (0.73, 0.89)	0.88 (0.77, 0.98)
Combined nomogram	0.92	0.81	0.85 (0.77, 0.91)	0.91 (0.84, 0.97)	0.71	0.90	0.78 (0.64, 0.88)	0.90 (0.81, 1.00)

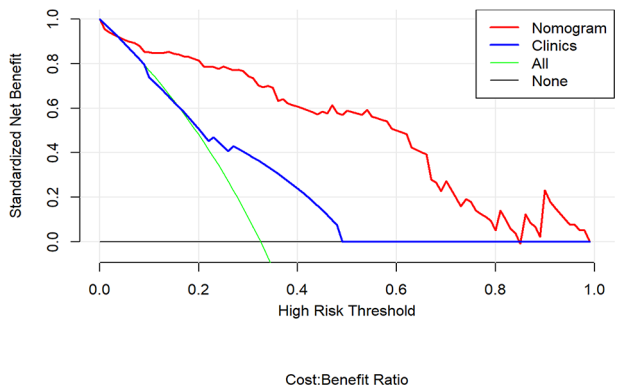


FIGURE 6 | Decision curve analysis for the nomogram compared with clinical model in the validation cohort. It can be concluded that when the threshold probability is over 10% approximately, the nomogram could provide extra profit over the "treat-all" or "treat-none" scheme and the clinical model.

Our study has some limitations. First, this was a retrospective study that was conducted in a single center with a relatively small sample size. Large-scale external validation is needed to further demonstrate the clinical efficacy of the nomogram constructed here. Second, while our method of manual segmentation set the basis for excellent results, this was possible because of our relatively small number of cases. For widespread application of this technique, more research employing automatic or semi-automatic image segmentation is likely to be necessary. Third, the patients included in this study all had SCN or MCN confirmed using surgical pathology, and there may have been a selection bias. IPMN or other pancreatic cystic diseases need to be further studied to broaden the clinical application of this algorithm.

In conclusion, our study has established a novel multi-phase CT-based radiomics nomogram for a noninvasive preoperative differentiation of SCNs from MCNs. The nomogram could provide a reference basis for an accurate diagnosis, thereby avoiding unnecessary surgical resection in clinical practice. We also preliminarily explored the influence of specific feature extraction phases on the predictive efficacy of the radiomics model; the results may be enlightening to subsequent radiomics studies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of Huashan Hospital, Fudan University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JG and FH: made the same effort in study, designed and carried out the experiments. XW: collected and sorted the data. SD: technology support. JZ: guided and modified the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The study was funded by National Key Research and Development Project in China (No. 2017YFC0113405), Launching Fund of Huashan Hospital (North Hospital) Affiliated to Fudan University (No. HSBY2020003).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.699812/full#supplementary-material

REFERENCES

1. Laffan TA, Horton KM, Klein AP, Berlanstein B, Siegelman SS, Kawamoto S, et al. Prevalence of Unsuspected Pancreatic Cysts on MDCT. *AJR Am J Roentgenol* (2008) 191:802–7. doi: 10.2214/AJR.07.3340
2. Zhang X, Mitchell DG, Dohke M, Holland GA, Parker L. Pancreatic Cysts: Depiction on Single-Shot Fast Spin-Echo MR Images. *Radiology* (2002) 223:547–53. doi: 10.1148/radiol.2232010815
3. Gaujoux S, Brennan MF, Gonen M, D'Angelica MI, DeMatteo R, Fong Y, et al. Cystic Lesions of the Pancreas: Changes in the Presentation and Management of 1,424 Patients at a Single Institution Over a 15-Year Time Period. *J Am Coll Surgeons* (2011) 212:590–600, 600-603. doi: 10.1016/j.jamcollsurg.2011.01.016
4. Chatelain D, Hammel P, O'Toole D, Terris B, Vilgrain V, Palazzo L, et al. Macrocystic Form of Serous Pancreatic Cystadenoma. *Am J Of Gastroenterol* (2002) 97:2566–71. doi: 10.1111/j.1572-0241.2002.06024.x
5. Hwang HK, Kim H, Kang CM, Lee WJ. Serous Cyst Adenoma of the Pancreas: Appraisal of Active Surgical Strategy Before It Causes Problems. *Surg Endoscopy* (2012) 26:1560–5. doi: 10.1007/s00464-011-2070-x
6. Strobel O, Z'graggen K, Schmitz-Winnenthal FH, Friess H, Kappeler A, Zimmermann A, et al. Risk of Malignancy in Serous Cystic Neoplasms of the Pancreas. *Digestion* (2003) 68:24–33. doi: 10.1159/000073222
7. Singhi AD, Nikiforova MN, Fasanella KE, McGrath KM, Pai RK, Ohori NP, et al. Preoperative GNAS and KRAS Testing in the Diagnosis of Pancreatic Mucinous Cysts. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2014) 20:4381–9. doi: 10.1158/1078-0432.CCR-14-0513
8. Matthaei H, Wylie D, Lloyd MB, Dal Molin M, Kemppainen J, Mayo SC, et al. MiRNA Biomarkers in Cyst Fluid Augment the Diagnosis and Management of Pancreatic Cysts. *Clin Cancer Res an Off J Am Assoc Cancer Res* (2012) 18:4713–24. doi: 10.1158/1078-0432.CCR-12-0035
9. Nikiforova MN, Khalid A, Fasanella KE, McGrath KM, Brand RE, Chennat JS, et al. Integration of KRAS Testing in the Diagnosis of Pancreatic Cystic

Lesions: A Clinical Experience of 618 Pancreatic Cysts. *Mod Pathol* (2013) 26:1478–87. doi: 10.1038/modpathol.2013.91
10. Correa-Gallego C, Ferrone CR, Thayer SP, Wargo JA, Warshaw AL, Fernández-Del Castillo C. Incidental Pancreatic Cysts: Do We Really Know What We are Watching? *Pancreatol Off J Int Assoc Pancreatol (IAP)* (2010) 10:144–50. doi: 10.1159/000243733
11. Curry CA, Eng J, Horton KM, Urban B, Siegelman S, Kuszyk BS, et al. CT of Primary Cystic Pancreatic Neoplasms: Can CT be Used for Patient Triage and Treatment? *AJR Am J roentgenol* (2000) 175:99–103. doi: 10.2214/ajr.175.1.1750099
12. Katz DS, Friedel DM, Kho D, Georgiou N, Hines JJ. Relative Accuracy of CT and MRI for Characterization of Cystic Pancreatic Masses. *AJR Am J Roentgenol* (2007) 189:657–61. doi: 10.2214/AJR.07.2772
13. Visser BC, Yeh BM, Qayyum A, Way LW, McCulloch CE, Coakley FV. Characterization of Cystic Pancreatic Masses: Relative Accuracy of CT and MRI. *AJR Am J Roentgenol* (2007) 189:648–56. doi: 10.2214/AJR.07.2365
14. Jang DK, Song BJ, Ryu JK, Chung KH, Lee BS, Park JK, et al. Preoperative Diagnosis of Pancreatic Cystic Lesions: The Accuracy of Endoscopic Ultrasound and Cross-Sectional Imaging. *Pancreas* (2015) 44:1329–33. doi: 10.1097/MPA.0000000000000396
15. Lee H, Kim M, Choi J, Hong H, Kim KA. Relative Accuracy of CT and MRI in the Differentiation of Benign From Malignant Pancreatic Cystic Lesions. *Clin Radiol* (2011) 66:315–21. doi: 10.1016/j.crad.2010.06.019
16. Khashab MA, Kim K, Lennon AM, Shin EJ, Tignor AS, Amateau SK, et al. Should We do EUS/FNA on Patients With Pancreatic Cysts? The Incremental Diagnostic Yield of EUS Over CT/MRI for Prediction of Cystic Neoplasms. *Pancreas* (2013) 42:717–21. doi: 10.1097/MPA.0b013e3182883a91
17. Guo J, Sahai AV, Teoh A, Arcidiacono PG, Larghi A, Saftoiu A, et al. An International, Multi-Institution Survey on Performing EUS-FNA and Fine Needle Biopsy. *Endosc Ultrasound* (2020) 9:319–28. doi: 10.4103/eus.eus_56_20
18. Blaszczak AM, Krishna SG. Endoscopic Diagnosis of Pancreatic Cysts. *Curr Opin Gastroen* (2019) 35:448–54. doi: 10.1097/MOG.0000000000000558
19. Carmicheal J, Patel A, Dalal V, Atri P, Dhaliwal AS, Wittel UA, et al. Elevating Pancreatic Cystic Lesion Stratification: Current and Future Pancreatic Cancer Biomarker(s). *Biochim Biophys Acta Rev Cancer* (2020) 1873:188318. doi: 10.1016/j.bbcan.2019.188318
20. Chakraborty J, Midya A, Gazit L, Attiyeh M, Langdon-Embry L, Allen PJ, et al. CT Radiomics to Predict High-Risk Intraductal Papillary Mucinous Neoplasms of the Pancreas. *Med Phys* (2018) 45:5019–29. doi: 10.1002/mp.13159
21. Shen X, Yang F, Yang P, Yang M, Xu L, Zhuo J, et al. A Contrast-Enhanced Computed Tomography Based Radiomics Approach for Preoperative Differentiation of Pancreatic Cystic Neoplasm Subtypes: A Feasibility Study. *Front Oncol* (2020) 10:248. doi: 10.3389/fonc.2020.00248
22. Jais B, Rebours V, Malleo G, Salvia R, Fontana M, Maggino L, et al. Serous Cystic Neoplasm of the Pancreas: A Multinational Study of 2622 Patients Under the Auspices of the International Association of Pancreatology and European Pancreatic Club (European Study Group on Cystic Tumors of the Pancreas). *Gut* (2016) 65:305–12. doi: 10.1136/gutjnl-2015-309638
23. Barthet M, Napoléon B, Palazzo L, Chemali M, Letard J, Laugier R, et al. Management of Cystic Pancreatic Lesions Found Incidentally. *Endoscopy* (2007) 39:926–8. doi: 10.1055/s-2007-966786
24. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
25. Barry WT, Kernagis DN, Dressman HK, Griffis RJ, Hunter JD, Olson JA, et al. Intratumor Heterogeneity and Precision of Microarray-Based Predictors of Breast Cancer Biology and Clinical Outcome. *J Clin Oncol* (2010) 28:2198–206. doi: 10.1200/JCO.2009.26.7245
26. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *BMJ* (2015) 350:g7594. doi: 10.1161/CIRCULATIONAHA.114.014508
27. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. *Eur Urol* (2018) 74:796–804. doi: 10.1016/j.eururo.2018.08.038
28. Sahani DV, Sainani NI, Blake MA, Crippa S, Mino-Kenudson M, del-Castillo CF. Prospective Evaluation of Reader Performance on MDCT in Characterization of Cystic Pancreatic Lesions and Prediction of Cyst Biologic Aggressiveness. *Ajr Am J Roentgenol* (2011) 197:W53–61. doi: 10.2214/AJR.10.5866
29. Brugge WR, Lewandrowski K, Lee-Lewandrowski E, Centeno BA, Szydlo T, Regan S, et al. Diagnosis of Pancreatic Cystic Neoplasms: A Report of the Cooperative Pancreatic Cyst Study. *Gastroenterology* (2004) 126:1330–6. doi: 10.1053/j.gastro.2004.02.013
30. Kim SY, Lee JM, Kim SH, Shin K, Kim YJ, An SK, et al. Macrocystic Neoplasms of the Pancreas: CT Differentiation of Serous Oligocystic Adenoma From Mucinous Cystadenoma and Intraductal Papillary Mucinous Tumor. *AJR Am J Roentgenol* (2006) 187:1192–8. doi: 10.2214/AJR.05.0337
31. Weiss MJ. The Value of Cytology in the Management of Patients With Pancreatic Cysts. *Cancer Cytopathol* (2019) 127:141–2. doi: 10.1002/cncy.22095
32. Chang Y, Tung C, Chang M, Wu C, Chen B, Jan I. Age and Cystic Size are Associated With Clinical Impact of Endoscopic Ultrasonography-Guided Fine-Needle Aspiration on the Management of Pancreatic Cystic Neoplasms. *Scand J Gastroenterol* (2019) 54:506–12. doi: 10.1080/00365521.2019.1601254
33. Goh BKP, Tan Y, Yap W, Cheow P, Chow PKH, Chung YA, et al. Pancreatic Serous Oligocystic Adenomas: Clinicopathologic Features and a Comparison With Serous Microcystic Adenomas and Mucinous Cystic Neoplasms. *World J Surg* (2006) 30:1553–9. doi: 10.1007/s00268-005-0749-7
34. Xie H, Ma S, Guo X, Zhang X, Wang X. Preoperative Differentiation of Pancreatic Mucinous Cystic Neoplasm From Macrocystic Serous Cystic Adenoma Using Radiomics: Preliminary Findings and Comparison With Radiological Model. *Eur J Radiol* (2020) 122:108747. doi: 10.1016/j.ejrad.2019.108747
35. Yang J, Guo X, Ou X, Zhang W, Ma X. Discrimination of Pancreatic Serous Cystadenomas From Mucinous Cystadenomas With CT Textural Features: Based on Machine Learning. *Front Oncol* (2019) 9:494. doi: 10.3389/fonc.2019.00494
36. Kocak B, Durmaz ES, Ates E, Kaya OK, Kilickesmez O. Unenhanced CT Texture Analysis of Clear Cell Renal Cell Carcinomas: A Machine Learning-Based Study for Predicting Histopathologic Nuclear Grade. *AJR Am J Roentgenol* (2019) 10:W1–8. doi: 10.2214/AJR.18.20742
37. Takahashi N, Leng S, Kitajima K, Gomez-Cardona D, Thapa P, Carter RE, et al. Small (< 4 Cm) Renal Masses: Differentiation of Angiomyolipoma Without Visible Fat From Renal Cell Carcinoma Using Unenhanced and Contrast-Enhanced Ct. *AJR Am J Roentgenol* (2015) 205:1194–202. doi: 10.2214/AJR.14.14183
38. Ren S, Zhao R, Zhang J, Guo K, Gu X, Duan S, et al. Diagnostic Accuracy of Unenhanced CT Texture Analysis to Differentiate Mass-Forming Pancreatitis From Pancreatic Ductal Adenocarcinoma. *Abdominal Radiol (New York)* (2020) 45:1524–33. doi: 10.1007/s00261-020-02506-6

Conflict of Interest: Author SD was employed by GE Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Gao, Han, Wang, Duan and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

REVIEW
published: 07 December 2021
doi: 10.3389/fonc.2021.730854

Comprehensive Imaging Characterization of Colorectal Liver Metastases

Drew Maclean[1,2*], *Maria Tsakok*[3], *Fergus Gleeson*[4], *David J. Breen*[1], *Robert Goldin*[5], *John Primrose*[6,7], *Adrian Harris*[4] *and James Franklin*[2]

[1] Department of Radiology, University Hospital Southampton, Southampton, United Kingdom, [2] Department of Medical Imaging, Bournemouth University, Bournemouth, United Kingdom, [3] Department of Radiology, Oxford University Hospitals, Oxford, United Kingdom, [4] Department of Oncology, Oxford University, Oxford, United Kingdom, [5] Department of Metabolism, Digestion and Reproduction, Imperial College London, London, United Kingdom, [6] Department of Surgery, University Hospital Southampton, Southampton, United Kingdom, [7] Academic Unit of Cancer Sciences, University of Southampton, Southampton, United Kingdom

Colorectal liver metastases (CRLM) have heterogenous histopathological and immunohistochemical phenotypes, which are associated with variable responses to treatment and outcomes. However, this information is usually only available after resection, and therefore of limited value in treatment planning. Improved techniques for *in vivo* disease assessment, which can characterise the variable tumour biology, would support further personalization of management strategies. Advanced imaging of CRLM including multiparametric MRI and functional imaging techniques have the potential to provide clinically-actionable phenotypic characterisation. This includes assessment of the tumour-liver interface, internal tumour components and treatment response. Advanced analysis techniques, including radiomics and machine learning now have a growing role in assessment of imaging, providing high-dimensional imaging feature extraction which can be linked to clinical relevant tumour phenotypes, such as a the Consensus Molecular Subtypes (CMS). In this review, we outline how imaging techniques could reproducibly characterize the histopathological features of CRLM, with several matched imaging and histology examples to illustrate these features, and discuss the oncological relevance of these features. Finally, we discuss the future challenges and opportunities of CRLM imaging, with a focus on the potential value of advanced analytics including radiomics and artificial intelligence, to help inform future research in this rapidly moving field.

Keywords: colorectal (colon) cancer, liver, metastasis, radiomic biomarkers, MRI, computed tomography

OPEN ACCESS

Edited by:
Joseph Ross Mitchell, Independent Researcher, Tampa, FL, United States

Reviewed by:
Jianan Chen, University of Toronto, Canada
Kunal Bharat Gala, Tata Memorial Hospital, India

***Correspondence:**
Drew Maclean
dfwmaclean@doctors.org.uk

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *25 June 2021*
Accepted: *15 November 2021*
Published: *07 December 2021*

Citation:
Maclean D, Tsakok M, Gleeson F, Breen DJ, Goldin R, Primrose J, Harris A and Franklin J (2021) Comprehensive Imaging Characterization of Colorectal Liver Metastases. Front. Oncol. 11:730854. doi: 10.3389/fonc.2021.730854

INTRODUCTION

Colorectal cancer (CRC) remains the second leading cause of cancer-related death in the developed world (1). Liver metastases are a major cause of death in patients with CRC and therefore optimising treatment of colorectal liver metastases (CRLM) is an important target for future research. Approximately 15% of patients will have synchronous liver metastases at initial diagnosis, with up to half developing liver metastases during their clinical course (2).

There are two main therapeutic strategies for CRLM. For those patients with polymetastatic disease, palliative systemic therapy is the mainstay of treatment. In patients with 'oligometastatic' disease (3), curative-intent surgical or image-guided treatment can be offered, often following neoadjuvant chemotherapy. Five-year survival rates following hepatic metastasectomy for CRLM are 28-49% (4), with some long-term survivors, in contrast to patients with polymetastatic disease who have a 1-year survival rate of approximately 55%, and 5-year survival rate of 3% (2). Historically, stricter surgical criteria have limited patient numbers proceeding to resection but, as evidence emerges that even patients with extensive disease derive benefit from local treatment (5), more patients are being offered treatment with curative intent. Selecting which patients who may benefit from radical treatment is an important challenge for clinicians, given that a significant proportion of patients undergoing metastasectomy suffer early relapse with incurable disease (4, 6). In the polymetastatic setting, multiple agents, including various cytotoxic regimens, targeted treatments and immunotherapies have transformed the options for patients without a curative option (7). Anticipating which agents will work optimally for each individual patient is critical, especially considering a counterintuitive response of patients to many of these therapies (8), which underlines the need for a more detailed assessment of colorectal metastases prior to commencing therapy.

The pathological literature describes the varied histopathological features of CRLM, both their internal architecture and their interface with the surrounding liver parenchyma (9, 10). Several histopathological and immunohistochemical phenotypes are associated with differential prognostic outcomes (10, 11). Unfortunately, histological phenotypic information is principally only available *after* resection, and is therefore of limited value in pretreatment prognostication, or when planning the neoadjuvant or polymetastatic treatment.

Improved techniques for *in vivo* disease assessment, which can characterise the variable tumour biology, would allow clinicians to plan personalized management strategies. Imaging already plays a central role in assessing the sites and burden of metastatic cancer both before and after treatment (12). Advanced imaging techniques, in particular multiparametric MRI (mpMRI) and functional imaging techniques, combined with novel image analysis techniques, have the potential to improve disease characterisation, and the advantage of being non-invasive, repeatable, and with the potential to assess all tumour sites.

In this narrative review, we outline how imaging techniques could reproducibly characterize the histopathological features of CRLM, with several matched imaging and histology examples to illustrate these features and discuss the oncological relevance of these features. We discuss the future opportunities and challenges of CRLM imaging, with a focus on the potential value of advanced analytics including radiomics and artificial intelligence, to help inform future research in this rapidly moving field. This review was informed by searching PubMed for relevant papers using search terms including 'colorectal', 'liver metastas*', 'MRI', 'CT', 'PET' and 'imaging biomarker' and a search of references.

IMAGING OF CRLM

Contrast enhanced CT (ceCT) is the mainstay of oncological imaging, and is the first line test for staging, surveillance and response assessment. Where there is diagnostic uncertainty, or precision about the number of metastases is crucial, mpMRI is the gold-standard technique for detecting and characterising focal liver lesions. Standard sequences include unenhanced T1- and T2-weighted sequences (including opposed-phase imaging), diffusion weighted imaging (DWI) and multiphase acquisitions following intravenous extracellular gadolinium chelate contrast media or liver-specific contrast agents. For CRLM, multiparametric liver MRI has superior per patient and per lesion sensitivity to CT (13, 14), and provides higher per lesion sensitivity than ^{18}F-fluorodeoxyglucose-positron emission tomography/CT (FDG-PET/CT) (13). Additionally, it provides high quality anatomical information which is invaluable for treatment-planning. It is therefore recommended as part of routine imaging work-up for patients being staged prior to liver-directed therapy (15). 18F-fluorodeoxyglucose-positron emission tomography/CT (FDG-PET/CT) is often used in patients with CRLM to detect extrahepatic disease that would preclude a radical treatment approach (16).

RADIOPATHOLOGICAL FEATURES OF CRLM

Tumour Interface With Normal Liver

Arguably the most clinically-relevant histopathological feature amenable to evaluation by imaging is the interface between normal liver and tumour. Interface features could influence chemotherapy selection, in particular for antiangiogenic agents (17, 18), inform surgical approach (19) and improve risk stratification for recurrence and overall survival (20, 21). These features are also well covered in another recent review (22).

The growth pattern of CRLM has been subdivided into three patterns of interface with the liver parenchyma: 'pushing', 'desmoplastic' and 'replacement' (9). The 'pushing' pattern is characterized by direct abutment of tumour cells on the liver parenchyma, with expansile growth flattening the liver plates. The 'desmoplastic' interface (present in approximately 40%) is differentiated by a rim of desmoplastic, fibrotic stroma incorporating a lymphocytic infiltrate, numerous bile ducts and capillaries (**Figure 1E**). This classification includes the term 'pseudo-encapsulated metastases' and has been linked to improved outcomes compared with non-desmoplastic metastases (17, 18, 20, 21, 23). It has also been suggested a less aggressive surgical approach, with a narrower margin, can be attempted with desmoplastic lesions, thus sparing more normal liver (19). The 'replacement' growth pattern is characterised by intimate contact between tumour cells and hepatocytes with a preserved reticulin pattern within the tumour tissue. Growth of these lesions therefore appears to be *via* vascular co-option, rather than angiogenesis, which is supported by their proven poor response to anti-angiogenic agents (17, 18, 23).

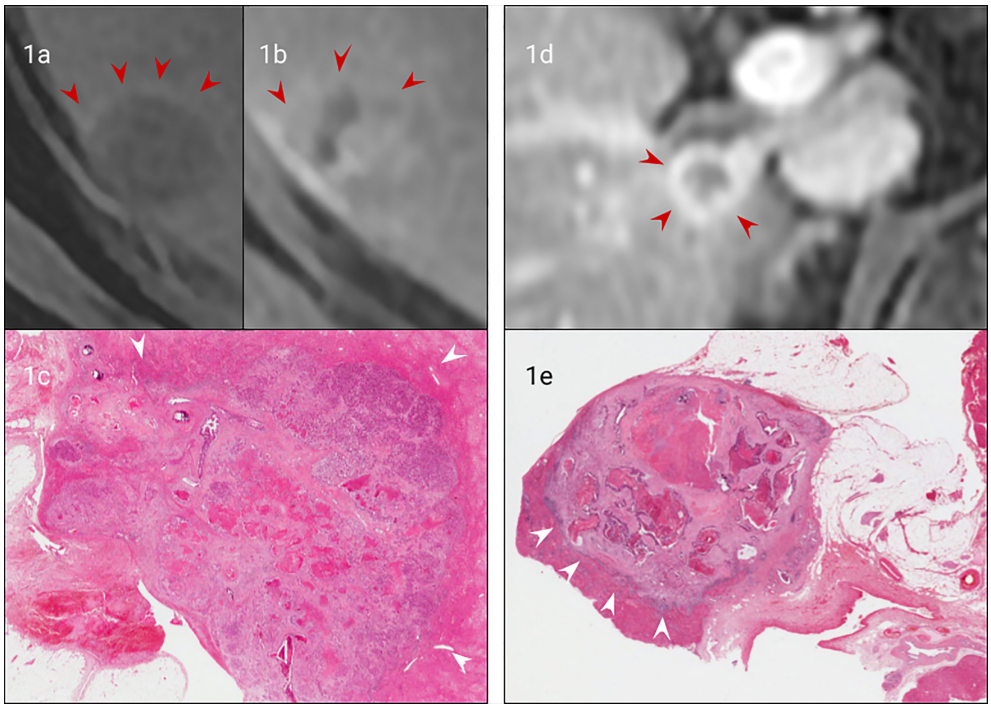


FIGURE 1 | Matched imaging and histology of non-encapsulated **(A–C)** vs capsulated CRLM **(A, D, E)**. Arterial phase gadolinium-enhanced T1 fat saturated MRI of a CRLM with early peripheral enhancement indicating compression of hepatocytes (arrows), **(B)**. Portal venous gadolinium-enhanced T1 fat saturated MRI showing an absence of peripheral enhancement with isointensity to normal liver (arrows), **(C)**. H&E staining (from **(A, B)** confirming no true capsule with peripheral compression of hepatocytes (arrows). **(D)**. Portal venous phase gadolinium-enhanced T1 fat saturated MRI demonstrating clear peripheral enhancement of a fibrotic capsule (arrows), **(E)**. H&E staining of the CRLM (from **(D)** confirming a true fibrotic capsule/desmoplastic interface (arrows).

There are several potential imaging correlates that could predict the presence of these patterns *in vivo*. The presence of a peripheral fibrotic capsule can be indicated by MRI (**Figure 1D**) (24–27), given typical MRI characteristics of fibrous tissue which is typically low T1 and T2 signal, and accumulates contrast on delayed contrast enhanced imaging (e.g. **Figure 1**). These features may reliably distinguish desmoplastic from other tumour types, although further research is required to establish the utility of this as a diagnostic tool.

Hepatocyte-specific contrast agents could also be useful in determining a 'replacement' growth pattern. This is similar to how microvascular invasion (MVI) can be identified with hepatocellular carcinoma (HCC) as reproducible hypoperfusion of liver parenchyma surrounding the tumour due to subtle tumour infiltration (28). However, this feature is currently untested in 'replacement' interface patterns of CRLM.

Vascular, Biliary, Lymphatic and Perineural Invasion

Vascular invasion is a common feature of colorectal cancer and its liver metastases. Intrahepatic portal venous invasion local to resected metastases occurs in approximately a quarter of cases (reported range 10-49%) and hepatic venous invasion in approximately one tenth of cases (reported range 5-24%) (10). Like primary colorectal cancer, venous invasion has been associated with poorer clinical outcomes (29). High resolution pelvic MRI has been shown to reliably identify extramural venous invasion of primary rectal cancer (30) and is now a useful prognostic marker. Similarly, although large vessel invasion is less common in CRLM, adjacent venous invasion can be similarly demonstrated (**Figure 2**).

Lymphatic invasion is a less common feature than venous invasion but has again been shown to be associated with poorer outcomes (31–33). Lymphatic vessels are smaller than the resolution of current imaging techniques, and therefore lymphatic invasion has not been reported to be directly accessible by MRI, although the presence of periportal, retroperitoneal or more distant lymphadenopathy may be a surrogate marker for this pathological feature. Neither biliary nor perineural invasion has been associated with adverse clinical outcomes (10) and there are no imaging studies correlating these features with MRI in CRLM.

Tumour Composition and Markers of Treatment Response

Features of Internal Composition

The internal composition of tumours differs, comprising variable proportions of tumour cells, fibrosis, necrotic material and, in some cases mucin and calcification. It is also influenced by

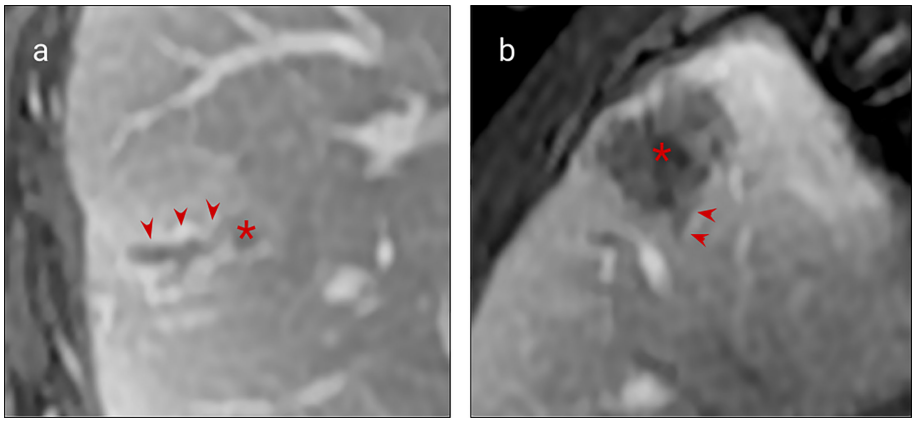


FIGURE 2 | MRI examples of histologically confirmed CRLM with macroscopic vascular invasion. **(A, B)** Portal venous phase gadolinium-enhanced T1 fat saturated MRI shows CRLM lesions (asterisk) with a filling defect within a hepatic vein indicating macroscopic venous invasion (arrows).

adjuvant treatment (34, 35). These tumour components have typical MRI features (as shown in **Table 1** and illustrated in **Figure 3**).

Viable tumour cells contrast with normal liver on T1 and T2 weighted sequences (**Table 1** and **Figure 3A**) and are generally hypoenhancing relative to background liver (**Figures 3A, C, G**) (26, 36–38). Necrosis is commonly found in chemo-naïve CRLMs, manifesting as T1 hyperintensity (differentiating it from other components) with variable enhancement, usually less than viable tumour and fibrosis (**Figures 3A, B, G**) (27, 40), with several histological subtypes recognised. CRLM often exhibit classical garland necrosis: areas of necrotic debris, sometimes referred to as 'dirty', 'usual' or intra-acinar necrosis, surrounded by a rim of viable tumour (**Figures 3A, B**) (27). Fibrosis demonstrates similar T1 and T2 signal characteristics to viable tumour, but typically has delayed gadolinium-enhancement (**Figures 3C, D, H**) (26, 42, 47) which differentiates it from tumour cells. Mucin is identified by pools of homogeneous high T2 and low T1 signal and absent enhancement (**Figures 3E, F**). Calcification characteristically presents signal voids on both T1 and T2-weighted sequences (**Table 1**) (36).

Defining the internal composition of metastases may help to categorise tumour biology and thus influence immunological and surgical strategy recently described Consensus Molecular Subtypes (CMS) of colorectal liver metastases (11), can be characterized by particular histopathological features and could be categorized by imaging. The molecular subtyping of liver metastases may be important even in patients when a resected primary tumour specimen is available, as significant discordance exists between primary and metastatic molecular subtypes (46, 48, 49). Within-patient heterogeneity is implicated as the root cause of a variable immunological response between primary and metastatic lesions (50, 51).

Pitroda et al. (11) have proposed three different CRLM subtypes [rather than the four classical colorectal molecular subtypes (52)]. They have identified a stromal metastasis subtype (with epithelial-mesenchymal transition and angiogenesis molecular signatures plus SMAD3 mutation association, subtype 3) which demonstrates significant baseline intratumoral fibrosis (11), in contrast to an immune-subtype (subtype 2) which demonstrates peritumoral lymphocyte infiltration and minimal intratumoural fibrosis (11). As outlined in **Table 1**, fibrosis is readily identified on MRI (27), whereas the degree of angiogenesis/microvascular density could be evaluated by Diffusion Weighted Imaging (DWI) and DCE-MRI enhancement (53–55). These imaging features are amenable to conventional visual, or radiomic derivation, and could provide a CMS prediction, and support personalised treatment. As yet, there have been no studies linking radiological imaging to CMS subtypes, although deep learning technologies have been applied to histological images to predict CMS subtypes and advanced imaging processing techniques (56), which are discussed in more detail below, have been used to classify other molecular features of CRLM.

Tumour Viability and Response

Chemotherapy and other systemic treatments, such as immunotherapy, are used in the palliative and neoadjuvant settings. Pre-treatment, proportions of viable and necrotic tumour occurs as a result of intralesional hypoxia and ischaemia as tumour growth outstrips angiogenesis or vascular co-option, and has not been found to be prognostically relevant (34, 57). However, markers suggesting reduction of viable tumour following treatment are useful indicators with a strong prognostic value (34, 57). Here imaging has an important advantage over histopathological assessment of CRLM; the ability to evaluate tumour response before and after treatment. Since the features of chemotherapy response, namely fibrosis and necrosis, are present in chemotherapy-naive metastases, evaluation of response without a baseline assessment becomes more challenging.

The method of histological tumour regression grading described by Rubbia-Brandt et al (34) is the most widely used

TABLE 1 | (See corresponding **Figure 3**).

	T2 signal	T1 signal	Enhancement	Diffusion
Tumour cells (**Figures 3A–D**)	High signal relative to liver (36)	Low signal relative to liver (36)	Variable, usually reduced enhancement relative to liver (26, 37)	Reduced ADC value (38, 39)
Necrosis (**Figures 3A, B, G**)	Variable. (36)	Variable (27, 36)	Delayed enhancement, less than fibrosis (26, 27, 36, 40)	Increased ADC value (41)
Fibrosis (**Figures 3C,D,H**)	Low signal relative to liver (27)	Low signal relative to liver (27, 36)	Late enhancement (26, 42, 43)	Increased ADC value (44, 45)
Mucin (**Figures 3E, F**)	High signal (46)	Low signal (46)	No enhancement (46)	Increased ADC value (46)
Calcification	Signal void (36)	Signal void (36)	None	

for determining histological response to chemotherapy; it categorises tumour response according to the balance of fibrosis and residual tumour. This method has been adapted from response grading of other tumours such as rectal and oesophageal tumours (58). Imaging response evaluation using the widely adopted RECIST 1.1 criteria, is based on changes in overall lesion size without taking into account changes in tissue composition (59). This is a limitation of the RECIST 1.1 criteria as tumours can demonstrate a reduction in cellularity without a reduction in size (42) (e.g. **Figures 4**–**6**), and some tumours have a low proportion of viable tumour at the outset, limiting the reliability of RECIST assessment (60–62), and its prognostic value. A combined assessment of the viable tumour volume, using both lesion size and tissue characterization may improve treatment assessment, and is already used in tumours such as HCC (63). Ideally automated lesion segmentation, subsegmentation (64) and feature analysis would integrate this process into the imaging pathway.

There is already evidence to support using morphological features to assess response of CRLM to chemotherapy. Lesion fibrosis, demonstrated as late gadolinium enhancement, which is the principle marker of response on pathological assessment, is a feature strongly linked to improved survival outcomes (**Figure 4**) (26, 42, 65). Similarly, CT/MRI morphological changes (illustrated in **Figures 5** and **6**) are better predictors of survival than RECIST [e.g (66–69)]. Adjuncts and alternatives to RECIST, including DWI and FDG-PET response evaluation have been investigated, but are not routinely deployed in clinical practice.

TOWARDS COMPREHENSIVE IMAGING CHARACTERIZATION OF CRLM: OPPORTUNITIES AND CHALLENGES

The major benefit of imaging is the capacity to perform repeated, *in vivo* assessment of multisite disease, which is a particular advantage in the metastatic setting. In this article, we have described the basis for more complete morphological characterization of CRLM (22). A further strength of imaging is the potential to perform multiparametric imaging, including functional imaging techniques. Functional imaging allows physiological evaluation of tumours, and can aid determination of histological features, and support treatment selection and response. For example, assessment of angiogenesis and microvascular density is possible by diffusion weighted imaging (DWI) (53) and functional vascular assessment by DCE-MRI (54, 55). These techniques could also be used in identifying molecular subtypes of CRLM. Response to antiangiogenic treatment can be monitored with DCE-MRI (54, 70) and this can improve treatment stratification in clinical practice and trials.

Molecular imaging techniques, in particular PET/CT, can provide further functional assessments of CRLM. FDG-PET is well-established in oncological practice, and routinely acquired in staging oligometastatic CRC, principally for detecting disease that would be beyond the scope of local therapy. However, FDG-uptake can also be used as marker of hypoxia, and as a prognostic marker and for response evaluation (71). The potential applications for novel molecular agents to image specific disease features and process are diverse, with hypoxia imaging agents such as 18F-FMISO (72) and antibody-based imaging for CEA (73) are under investigation in the research setting.

However, despite the potential for multiparametric structural, functional and molecular imaging to provide a more comprehensive assessment of CRLM, there are limitations. While these techniques might be valuable, and are often integrated into trials, complex multimodality assessments are challenging in the clinical workflow, and reliably and repeatedly combining information from multiparametric imaging is difficult for human observers. A second major technical challenge to the use of imaging for assessing tissue features of CRLM is the achievable spatial resolution. For MRI, the in-plane spatial resolution is typically in the region of $1mm^2$. Although higher spatial resolutions can be achieved, an important feature of liver imaging is managing respiratory and, to a lesser extent, cardiac motion, which limits acquisition time and spatial resolution achievable in the upper abdomen. This issue is compounded through tissue features which can be substantially smaller than the imaging resolution, which may preclude accurate assessment of some features and place a limit on the achievable performance of imaging.

These two key challenges may be overcome by developments in image acquisition and analysis techniques. MRI acquisition will increasingly use computer-assisted techniques to decrease acquisition time and improve image resolution, for example through the use of 'super-resolution' techniques, which may

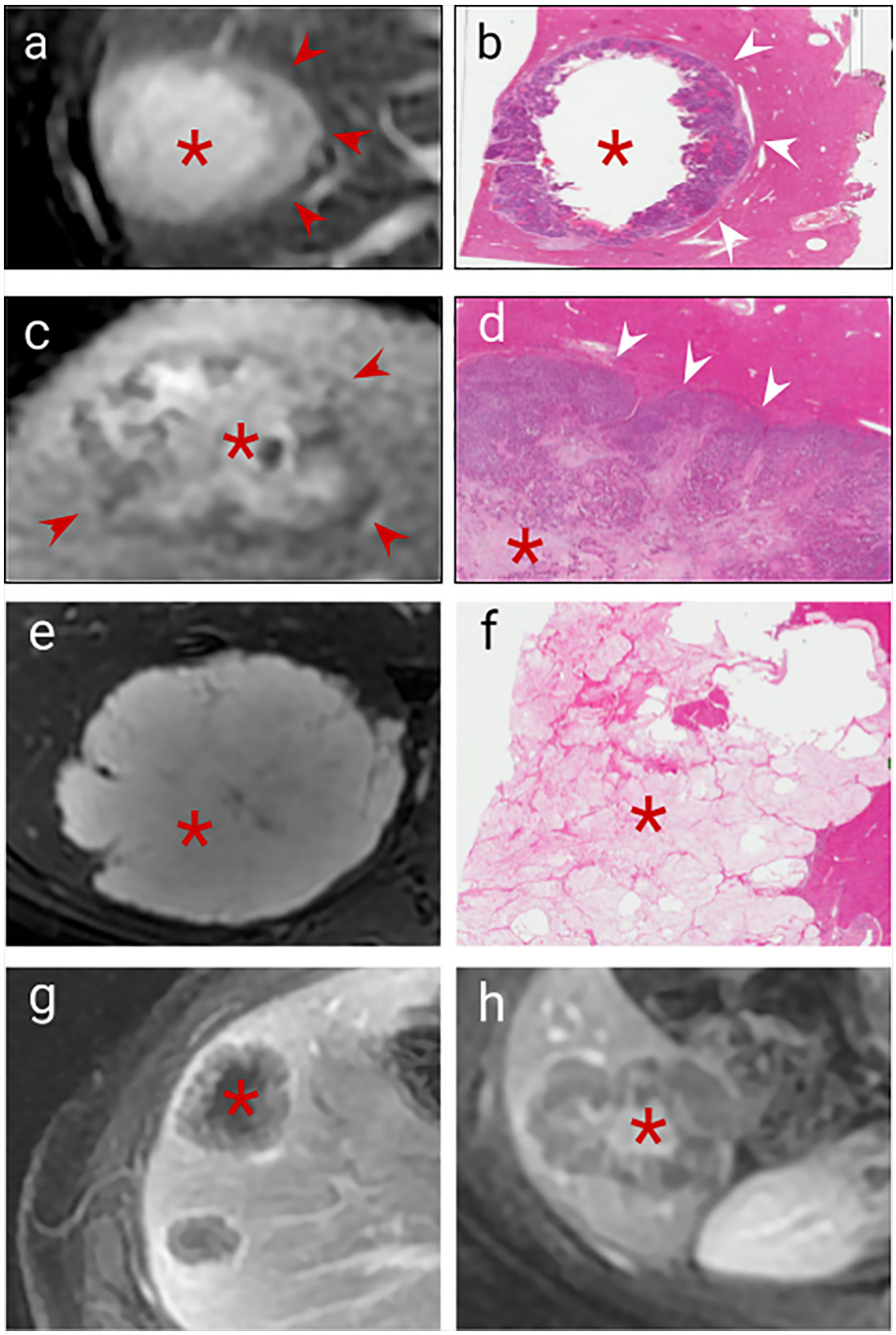


FIGURE 3 | Imaging correlates of internal tumour composition. **(A)**. T2 weighted non-enhanced MRI of a lesion displaying moderately high central T2 signal suggestive of central necrosis (asterisk) and surrounding viable tumour which is higher signal than normal liver (arrows), **(B)**. H&E staining of this CRLM confirms peripheral viable tumour (arrows) with central liquefaction of the metastasis indicating classical Garland necrosis (asterisk), **(C)**. Delayed phase gadolinium-enhanced T1 fat saturated MRI showing avid central delayed enhancement indicating fibrosis (asterisk) with surrounding viable tumour cells which enhance less than normal liver (arrows), **(D)**. H&E staining of this CRLM confirming central fibrosis within the lesion (asteriks) and peripheral viable tumour cells (arrows), **(E)**. T2 weighted sequence demonstrated homogenous high signal mucin (asterisk) **(F)**. H&E staining confirming mucin containing metastasis (asterisk) **(G)**. Delayed phase gadolinium-enhanced T1 fat saturated MRI showing centrally necrotic lesions (no delayed central enhancement, asteriks), **(H)**. Delayed phase Gadolinium enhanced T1 fat saturated MRI showing late central enhancement indicating a centrally fibrotic lesion (asterisk) which contrasts with **(G)** (central necrosis).

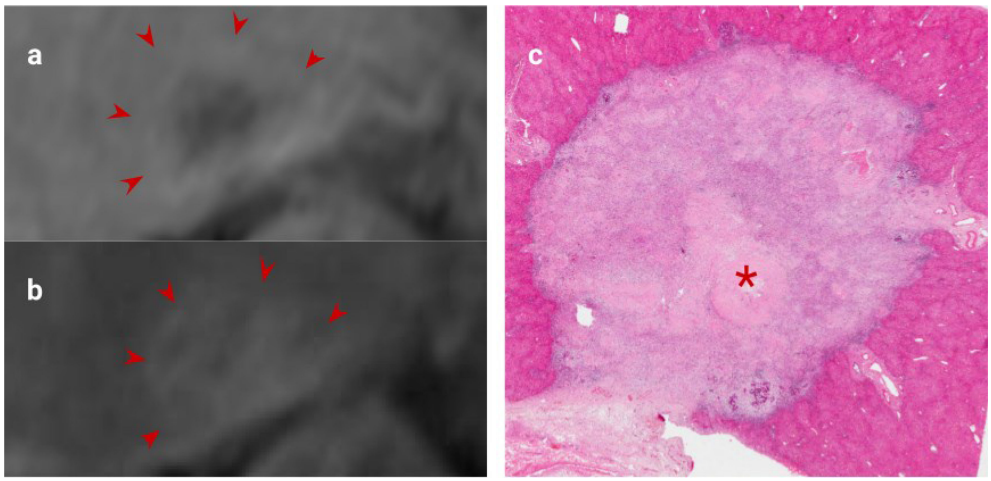


FIGURE 4 | CRLM imaging features following excellent response to Chemotherapy. **(A)**. Portal venous phase gadolinium-enhanced T1 fat saturated MRI showing decreased enhancement relative to liver (arrows), **(B)**. Delayed phase MRI showing increased enhancement relative to liver (arrows), **(C)**. H&E staining of this CRLM confirming almost complete fibrosis of the lesion indicating excellent chemotherapy response. '*' represents the metastasis of interest.

improve the potential for tissue feature assessment by imaging (74). Secondly, improved motion correction and co-registration techniques can help overcome issues with between-acquisition motion (75).

The most promising opportunity within radiology is the incorporation of machine learning in image interpretation. Analysis of tumour features has historically relied on expert radiological assessment of imaging features. However, modern radiomic image analysis can be used to extract high-dimensional data from medical imaging (76), and machine learning techniques can be used for both automated segmentation (77) and feature analysis (78) to produce imaging biomarkers from medical imaging. These analysis strategies can more easily combine multiparametric imaging than a human observer, will remove observer variability, and can become an automated component of the image interpretation pipeline, which would be critical for clinical implementation in patient care.

Several key hurdles must be overcome before radiomics and machine learning becomes robust enough to influence patient care in daily clinical practice. Firstly, many radiomics and machine learning studies on CRLM have been conducted on relatively small datasets (42, 66, 76, 78). Studies conducted on small datasets therefore are at put a Radiomic algorithm at risk of 'overfitting' the data (creating an algorithm too specific to a limited pool of data),

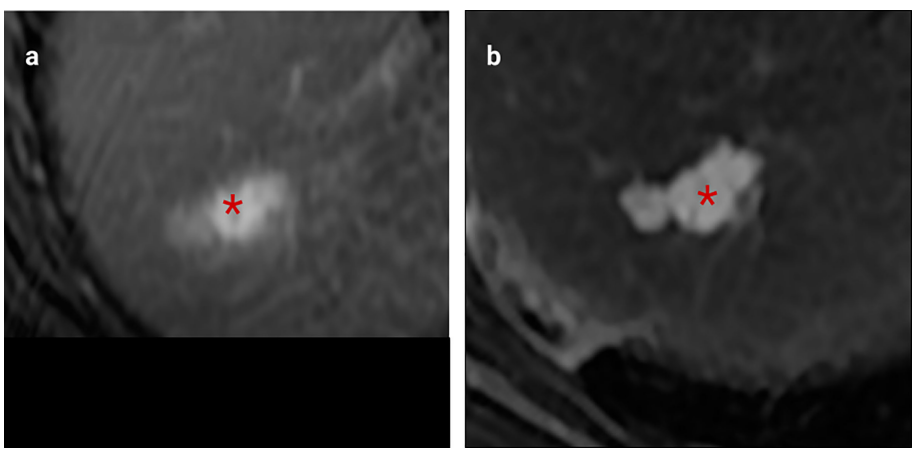


FIGURE 5 | Example of excellent morphological response post-chemotherapy. Both images were obtained using the same MRI scanner performed at 1.5T (GE HDX Twinspeed MR scanner; GE, Milkwaukee, WI) with an 8-channel torso coil. **(A)** T2 weighted MRI showing a poorly defined CRLM with high T2 region of mucin (asterisk). **(B)**. Following neoadjuvant treatment, this lesion demonstrates no change in size but now has a well-defined margin and replacement of intermediate signal cellular tumour with mucin, indicative of a good morphological response (asterisk). There was a complete pathological response at resection.

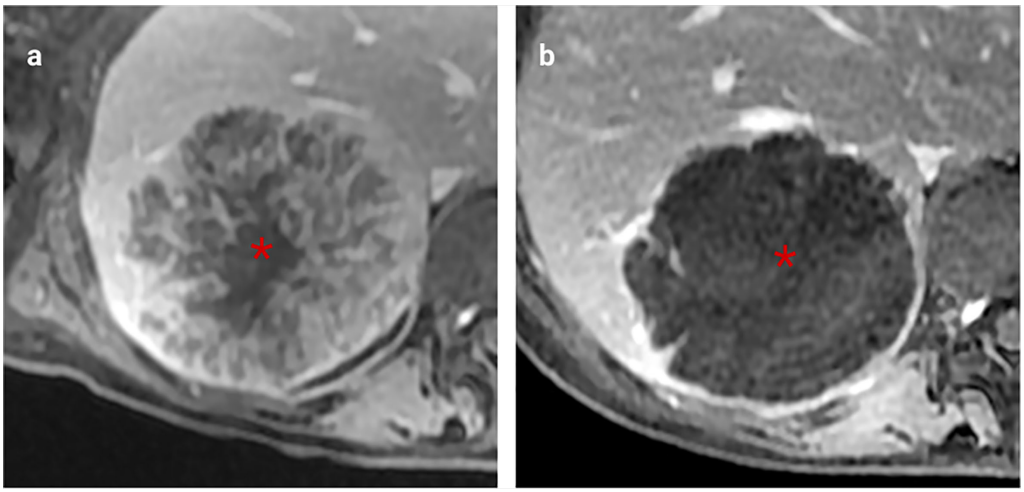


FIGURE 6 | Excellent morphological response to chemotherapy. **(A)**. Heterogenously enhancing CRLM on portal venous phase T1-weighted MRI (asterisk). **(B)**, Post-chemotherapy MRI shows the lesion has become homogenous without a reduction in size, indicating a significant decrease in viable tumour cells.

reducing the generalisability of the study findings (79, 80). The scientific community needs large annotated databases to derive and validate image analysis tools, however the practicality and ethics of sharing scans across multiple institutions, acquired through different and evolving techniques, is an ongoing challenge (79). Aside from acquiring larger, accurately labelled datasets, development of advanced radiomic techniques hold promise for minimising this issue (including various unsupervised clustering methods), but they are unlikely to be the solution without improved sources of data (79, 81, 82).

Even assuming a robust algorithm and analysis platform is developed in the research setting, adapting them into a convenient final product for use in the routine clinical workflow is a further challenge. In addition, many radiomic techniques are time-consuming from an operator perspective with careful lesion contouring required (76, 81). Automated segmentation technologies and other assisted analysis tools will be crucial to ensure workflow ensure clinical adoption.

An important consideration for imaging biomarkers in cancer is the development of alternative techniques for *in vivo* tumour assessment, in particular circulating tumour biomarkers such as circulating tumour DNA (83). However, as these tissues do not allow spatial localization of tumours, it is likely, particularly in the metastatic setting, that imaging can provide complementary information, and the combination of technologies can offer a more comprehensive toolkit to accurately phenotype disease.

CONCLUSION

This review has highlighted the potential for advanced imaging to provide *in vivo* characterization of clinically relevant histopathological features of CRLM. We have outlined the imaging findings of these histopathological features, focusing on the tumour-liver interface, intralesional component analysis, and morphological response assessment following systemic or liver-directed treatment. *In vivo* assessment of the tumour-liver interface has the potential to play and important role in defining the surgical approach and chemotherapy selection. As well as improving our characterisation of response to chemotherapy, imaging analysis of internal tumour components could play an increasingly important role as predictors of CRLM molecular subtypes. However, in the absence of studies providing robust validation of imaging techniques to report these features in practice, our assessment of CRLM by imaging is limited to documenting their size, number and location.

Advanced analysis methods, such as radiomics and machine learning, will be crucial tools in developing and validating novel imaging biomarkers for CRLM. However, these rely on curated and annotated datasets of sufficient size to build reliable algorithms, which is likely to require the cross-institutional collaborations that have been achieved in pathology and molecular biology.

AUTHOR CONTRIBUTIONS

DM was involved with literature review, and writing and editing the manuscript. MT was involved with literature review, and writing and editing the manuscript. FG was involved with writing and editing the manuscript. DB was involved with writing and editing the manuscript. RG was involved with writing and editing the manuscript. JP was involved with writing and editing the manuscript. AH was involved with writing and editing the manuscript. JF was involved with literature review, writing and editing the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Bray F, Jemal A, Grey N, Ferlay J, Forman D. Global Cancer Transitions According to the Human Development Index (2008-2030): A Population-Based Study. *Lancet Oncol* (2012) 13(8):790–801. doi: 10.1016/S1470-2045(12)70211-5
2. Manfredi S, Lepage C, Hatem C, Coatmeur O, Faivre J, Bouvier A-M. Epidemiology and Management of Liver Metastases From Colorectal Cancer. *Ann Surg* (2006) 244(2):254–9. doi: 10.1097/01.sla.0000217629.94941.cf
3. Hellman S, Weichselbaum RR. Oligometastases. *J Clin Oncol* (1995) 13:8–10. doi: 10.1200/jco.1995.13.1.8
4. Kansas GP, Taylor A, Primrose JN, Langeberg W, Kelsh MA, Mowat F, et al. Survival After Liver Resection in Metastatic Colorectal Cancer: Review and Meta-Analysis of Prognostic Factors. *Clin Epidemiol* (2012) 4(1):283–301. doi: 10.2147/clep.s34285
5. Adam R, Wicherts DA, de Haas RJ, Ciacio O, Lévi F, Paule B, et al. Patients With Initially Unresectable Colorectal Liver Metastases: Is There a Possibility of Cure? *J Clin Oncol* (2009) 27:1829–35. doi: 10.1200/jco.2008.19.9273
6. Gomez D, Sangha VK, Morris-Stiff G, Malik HZ, Guthrie AJ, Toogood GJ, et al. Outcomes of Intensive Surveillance After Resection of Hepatic Colorectal Metastases. *Br J Surg* (2010) 97:1552–60. doi: 10.1002/bjs.7136
7. Van Cutsem E, Cervantes A, Adam R, Sobrero A, Van Krieken JH, Aderka D, et al. ESMO Consensus Guidelines for the Management of Patients With Metastatic Colorectal Cancer. *Ann Oncol* (2016) 27(8):1386–422. doi: 10.1093/annonc/mdw235
8. Bridgewater JA, Pugh SA, Maishman T, Eminton Z, Mellor J, Whitehead A, et al. Systemic Chemotherapy With or Without Cetuximab in Patients With Resectable Colorectal Liver Metastasis (New EPOC): Long-Term Results of a Multicentre, Randomised, Controlled, Phase 3 Trial. *Lancet Oncol* (2020) 21(3):398–411. doi: 10.1016/S1470-2045(19)30798-3
9. Vermeulen PB, Colpaert C, Salgado R, Royers R, Hellemans H, Van den Heuvel E, et al. Liver Metastases From Colorectal Adenocarcinomas Grow in Three Patterns With Different Angiogenesis and Desmoplasia [Internet]. *Vol 195 J Pathol* (2001) 195(3):336–42. doi: 10.1002/path.966
10. Knijn N, de Ridder JAM, Punt CJA, de Wilt JHW, Nagtegaal ID. Histopathological Evaluation of Resected Colorectal Cancer Liver Metastases: What Should be Done? *Histopathology* (2013) 63(2):149–56. doi: 10.1111/his.12124
11. Pitroda SP, Khodarev NN, Huang L, Uppal A, Wightman SC, Ganai S, et al. Integrated Molecular Subtyping Defines a Curable Oligometastatic State in Colorectal Liver Metastasis. *Nat Commun* (2018) 9. doi: 10.1038/s41467-018-04278-6
12. Franklin JM, Sharma RA, Harris AL, Gleeson FV. Imaging Oligometastatic Cancer Before Local Treatment. *Lancet Oncol* (2016) 17(9):e406–14. doi: 10.1016/S1470-2045(16)30277-7
13. Niekel MC, Bipat S, Stoker J. Diagnostic Imaging of Colorectal Liver Metastases With CT, MR Imaging, FDG PET, and/or FDG PET/CT: A Meta-Analysis of Prospective Studies Including Patients Who Have Not Previously Undergone Treatment. *Radiology* (2010) 257(3):674–84. doi: 10.1148/radiol.10100729
14. Mainenti PP, Stanzione A, Guarino S, Romeo V, Ugga L, Romano F, et al. Colorectal Cancer: Parametric Evaluation of Morphological, Functional and Molecular Tomographic Imaging. *World J Gastroenterol* (2019) 25(35):5233–56. doi: 10.3748/wjg.v25.i35.5233
15. Adams RB, Aloia TA, Loyer E, Pawlik TM, Taouli B, Vauthey J-N, et al. Selection for Hepatic Resection of Colorectal Liver Metastases: Expert Consensus Statement. *HPB* (2013) 15(2):91–103. doi: 10.1111/j.1477-2574.2012.00557.x
16. Maffione AM, Lopci E, Bluemel C, Giammarile F, Herrmann K, Rubello D. Diagnostic Accuracy and Impact on Management of (18)F-FDG PET and PET/CT in Colorectal Liver Metastasis: A Meta-Analysis and Systematic Review. *Eur J Nucl Med Mol Imaging* (2015) 42(1):152–63. doi: 10.1007/s00259-014-2930-4
17. Lazaris A, Amri A, Petrillo SK, Zoroquiain P, Ibrahim N, Salman A, et al. Vascularization of Colorectal Carcinoma Liver Metastasis: Insight Into Stratification of Patients for Anti-Angiogenic Therapies. *Hip Int* (2018) 4(3):184–92. doi: 10.1002/cjp2.100
18. Frentzas S, Simoneau E, Bridgeman VL, Vermeulen PB, Foo S, Kostaras E, et al. Vessel Co-Option Mediates Resistance to Anti-Angiogenic Therapy in Liver Metastases. *Nat Med* (2016) 22(11):1294–302. doi: 10.1038/nm.4197
19. Oliveira RC, Alexandrino H, Cipriano MA, Tralhão JG. Liver Metastases and Histological Growth Patterns: Biological Behavior and Potential Clinical Implications-Another Path to Individualized Medicine? *J Oncol* (2019) 2019:6280347. doi: 10.1155/2019/6280347
20. Eefsen RL, Vermeulen PB, Christensen IJ, Laerum OD, Mogensen MB, Rolff HC, et al. Growth Pattern of Colorectal Liver Metastasis as a Marker of Recurrence Risk. *Clin Exp Metastasis* (2015) 32(4):369–81. doi: 10.1007/s10585-015-9715-4
21. Pinheiro RS, Herman P, Lupinacci RM, Lai Q, Mello ES, Coelho FF, et al. Tumor Growth Pattern as Predictor of Colorectal Liver Metastasis Recurrence. *Am J Surg* (2014) 207:493–8. doi: 10.1016/j.amjsurg.2013.05.015
22. Paulatto L, Burgio MD, Sartoris R, Beaufrère A, Cauchy F, Paradis V, et al. Colorectal Liver Metastases: Radiopathological Correlation. *Insights Into Imaging* (2020) 11. doi: 10.1186/s13244-020-00904-4
23. Siriwardana PN, Luong TV, Watkins J, Turley H, Ghazaley M, Gatter K, et al. Biological and Prognostic Significance of the Morphological Types and Vascular Patterns in Colorectal Liver Metastases (CRLM). *Medicine* (2016) 95:e2924. doi: 10.1097/md.0000000000002924
24. Farace P, Merigo F, Fiorini S, Nicolato E, Tambalo S, Daducci A, et al. DCE-MRI Using Small-Molecular and Albumin-Binding Contrast Agents in Experimental Carcinomas With Different Stromal Content. *Eur J Radiol* (2011) 78(1):52–9. doi: 10.1016/j.ejrad.2009.04.043
25. Chen P, Wang Y, Fang X, Wang X, Wang G. Prognostic Significance of Peritumoral Fibrosis After Resection of Pancreatic Head Cancer. *Oncol Lett* (2020) 19(2):1235–40. doi: 10.3892/ol.2019.11181
26. Cheung HMC, Karanicolas PJ, Hsieh E, Coburn N, Maraj T, Kim JK, et al. Late Gadolinium Enhancement of Colorectal Liver Metastases Post-Chemotherapy Is Associated With Tumour Fibrosis and Overall Survival Post-Hepatectomy. *Eur Radiol* (2018) 28:3505–12. doi: 10.1007/s00330-018-5331-4
27. Outwater E, Tomaszewski JE, Daly JM, Kressel HY. Hepatic Colorectal Metastases: Correlation of MR Imaging and Pathologic Appearance. *Radiol* (1991) 180(2):327–32. doi: 10.1148/radiology.180.2.2068294
28. Lee S, Kim SH, Lee JE, Sinn DH, Park CK. Preoperative Gadoxetic Acid-Enhanced MRI for Predicting Microvascular Invasion in Patients With Single Hepatocellular Carcinoma. *J Hepatol* (2017) 67(3):526–34. doi: 10.1016/j.jhep.2017.04.024
29. Hayashi M, Inoue Y, Komeda K, Shimizu T, Asakuma M, Hirokawa F, et al. Clinicopathological Analysis of Recurrence Patterns and Prognostic Factors for Survival After Hepatectomy for Colorectal Liver Metastasis. *BMC Surg* (2010) 10:27. doi: 10.1186/1471-2482-10-27
30. Brown G, Radcliffe AG, Newcombe RG, Dallimore NS, Bourne MW, Williams GT. Preoperative Assessment of Prognostic Factors in Rectal Cancer Using High-Resolution Magnetic Resonance Imaging. *Br J Surg* (2003) 90:355–64. doi: 10.1002/bjs.4034
31. Shirabe K, Takenaka K, Gion T, Fujiwara Y, Shimada M, Yanaga K, et al. Analysis of Prognostic Risk Factors in Hepatic Resection for Metastatic Colorectal Carcinoma With Special Reference to the Surgical Margin. *Br J Surg* (1997) 84(8):1077–80. doi: 10.1002/BJS.1800840810
32. Sasaki A, Aramaki M, Kawano K, Yasuda K, Inomata M, Kitano S. Prognostic Significance of Intrahepatic Lymphatic Invasion in Patients With Hepatic Resection Due to Metastases From Colorectal Carcinoma. *Cancer* (2002) 95(1):105–11. doi: 10.1002/cncr.10655
33. Lupinacci RM, Mello ES, Pinheiro RS, Marques G, Coelho FF, Kruger JAP, et al. Intrahepatic Lymphatic Invasion But Not Vascular Invasion is a Major Prognostic Factor After Resection of Colorectal Cancer Liver Metastases. *World J Surg* (2014) 38:2089–96. doi: 10.1007/s00268-014-2511-5
34. Rubbia-Brandt L, Giostra E, Brezault C, Roth AD, Andres A, Audard V, et al. Importance of Histological Tumor Response Assessment in Predicting the Outcome in Patients With Colorectal Liver Metastases Treated With Neo-Adjuvant Chemotherapy Followed by Liver Surgery. *Ann Oncol* (2007) 18(2):299–304. doi: 10.1093/annonc/mdl386
35. Eefsen RL, Engelholm L, Willemoe GL, Van den Eynden GG, Laerum OD, Christensen IJ, et al. Microvessel Density and Endothelial Cell Proliferation Levels in Colorectal Liver Metastases From Patients Given Neo-Adjuvant

Cytotoxic Chemotherapy and Bevacizumab. *Int J Cancer* (2016) 138(7):1777–84. doi: 10.1002/ijc.29904
36. Milot L, Guindi M, Gallinger S, Moulton CA, Brock KK, Dawson LA, et al. MR Imaging Correlates of Intratumoral Tissue Types Within Colorectal Liver Metastases: A High-Spatial-Resolution Fresh *Ex Vivo* Radiologic-Pathologic Correlation Study. *Radiol* (2010) 254(3):747–54. doi: 10.1148/radiol.09090508
37. Colagrande S, Castellani A, Nardi C, Lorini C, Calistri L, Filippone A. The Role of Diffusion-Weighted Imaging in the Detection of Hepatic Metastases From Colorectal Cancer: A Comparison With Unenhanced and Gd-EOB-DTPA Enhanced MRI. *Eur J Radiol* (2016) 85(5):1027–34. doi: 10.1016/j.ejrad.2016.02.011
38. Geschwind J-FH, Artemov D, Abraham S, Omdal D, Huncharek MS, McGee C, et al. Chemoembolization of Liver Tumor in a Rabbit Model: Assessment of Tumor Cell Death With Diffusion-Weighted MR Imaging and Histologic Analysis. *J Vasc Interventional Radiol* (2000) 11:1245–55. doi: 10.1016/s1051-0443(07)61299-8
39. Sica GT, Ji H, Ros PR. CT and MR Imaging of Hepatic Metastases. *Am J Roentgenol* (2000) 174:691–8. doi: 10.2214/ajr.174.3.1740691
40. Koh TS, Thng CH, Hartono S, Dominguez LTM, Lim TKH, Huynh H, et al. Assessment of Tumor Necrotic Fraction by Dynamic Contrast-Enhanced MRI: A Preclinical Study of Human Tumor Xenografts With Histopathologic Correlation. *NMR Biomed* (2014) 27(4):486–94. doi: 10.1002/nbm.3090
41. Chiaradia M, Baranes L, Van Nhieu JT, Vignaud A, Laurent A, Decaens T, et al. Intravoxel Incoherent Motion (IVIM) MR Imaging of Colorectal Liver Metastases: Are We Only Looking at Tumor Necrosis? *J Magnetic Resonance Imaging* (2014) 39:317–25. doi: 10.1002/jmri.24172
42. Shur J, Orton M, Connor A, Fischer S, Moulton C, Gallinger S, et al. A Clinical-Radiomic Model for Improved Prognostication of Surgical Candidates With Colorectal Liver Metastases. *J Surg Oncol* (2020) 121:357–64. doi: 10.1002/jso.25783
43. Murata S, Matsushima S, Sato Y, Yamaura H, Kato M, Hasegawa T, et al. Predicting Chemotherapeutic Response for Colorectal Liver Metastases Using Relative Tumor Enhancement of Gadoxetic Acid Disodium-Enhanced Magnetic Resonance Imaging. *Abdom Radiol (NY)* (2018) 43(12):3301–6. doi: 10.1007/s00261-018-1615-z
44. Wagner M, Doblas S, Daire J-L, Paradis V, Haddad N, Leitão H, et al. Diffusion-Weighted MR Imaging for the Regional Characterization of Liver Tumors. *Radiol* (2012) 264(2):464–72. doi: 10.1148/radiol.12111530
45. Scurr ED, Collins DJ, Temple L, Karanjia N, Leach MO, Koh D-M. Appearances of Colorectal Hepatic Metastases at Diffusion-Weighted MRI Compared With Histopathology: Initial Observations. *Br J Radiol* (2012) 85 (1011):225–30. doi: 10.1259/bjr/11597735
46. Lee JE, Kim SH, Lee S, Choi S-Y, Hwang JA, Woo S-Y. Differentiating Metastatic Mucinous Colorectal Adenocarcinomas From Simple Cysts of the Liver Using Contrast-Enhanced and Diffusion-Weighted MRI. *Br J Radiol* (2018) 91:20180303. doi: 10.1259/bjr.20180303
47. Cheung HMC, Karanicolas PJ, Coburn N, Seth V, Law C, Milot L. Delayed Tumour Enhancement on Gadoxetate-Enhanced MRI Is Associated With Overall Survival in Patients With Colorectal Liver Metastases. *Eur Radiol* (2019) 29(2):1032–8. doi: 10.1007/s00330-018-5618-5
48. Lau YC, Schmeier S, Frizelle F, Purcell R. Consensus Molecular Subtypes of Primary Colon Tumors and Their Hepatic Metastases. *Future Sci OA* (2021) 7 (7):FSO722. doi: 10.2144/fsoa-2021-0021
49. Choi JY, Choi S, Lee M, Park YS, Sung JS, Chang WJ, et al. Clinical Implication of Concordant or Discordant Genomic Profiling Between Primary and Matched Metastatic Tissues in Patients With Colorectal Cancer. *Cancer Res Treat* (2020) 52(3):764–78. doi: 10.4143/crt.2020.044
50. Zou Y, Hu X, Zheng S, Yang A, Li X, Tang H, et al. Discordance of Immunotherapy Response Predictive Biomarkers Between Primary Lesions and Paired Metastases in Tumours: A Systematic Review and Meta-Analysis. *EBioMed* (2021) 63:103137. doi: 10.1016/j.ebiom.2020.103137
51. Stintzing S, Wirapati P, Lenz H-J, Neureiter D, Fischer von Weikersthal L, Decker T, et al. Consensus Molecular Subgroups (CMS) of Colorectal Cancer (CRC) and First-Line Efficacy of FOLFIRI Plus Cetuximab or Bevacizumab in the FIRE3 (AIO KRK-0306) Trial. *Ann Oncol* (2019) 30(11):1796–803. doi: 10.1093/annonc/mdz387
52. Guinney J, Dienstmann R, Wang X, de Reyniès A, Schlicker A, Soneson C, et al. The Consensus Molecular Subtypes of Colorectal Cancer. *Nat Med* (2015) 21(11):1350–6. doi: 10.1038/nm.3967
53. Ullrich RT, Jikeli JF, Diedenhofen M, Böhm-Sturm P, Unruh M, Vollmar S, et al. *In-Vivo* Visualization of Tumor Microvessel Density and Response to Anti-Angiogenic Treatment by High Resolution MRI in Mice. *PloS One* (2011) 6:e19592. doi: 10.1371/journal.pone.0019592
54. Morgan B, Thomas AL, Drevs J, Hennig J, Buchert M, Jivan A, et al. Dynamic Contrast-Enhanced Magnetic Resonance Imaging as a Biomarker for the Pharmacological Response of PTK787/ZK 222584, an Inhibitor of the Vascular Endothelial Growth Factor Receptor Tyrosine Kinases, in Patients With Advanced Colorectal Cancer and Liver Metastases: Results From Two Phase I Studies. *J Clin Oncol* (2003) 21(21):3955–64. doi: 10.1200/JCO.2003.08.092
55. De Bruyne S, Van Damme N, Smeets P, Ferdinande L, Ceelen W, Mertens J, et al. Value of DCE-MRI and FDG-PET/CT in the Prediction of Response to Preoperative Chemotherapy With Bevacizumab for Colorectal Liver Metastases. *Br J Cancer* (2012) 106(12):1926–33. doi: 10.1038/bjc.2012.184
56. Sirinukunwattana K, Domingo E, Richman SD, Redmond KL, Blake A, Verrill C, et al. Image-Based Consensus Molecular Subtype (imCMS) Classification of Colorectal Cancer Using Deep Learning. *Gut* (2021) 70(3):544–54. doi: 10.1136/gutjnl-2019-319866
57. Serrablo A, Paliogiannis P, Pulighe F, Moro SS-M, Borrego-Estella V, Attene F, et al. Impact of Novel Histopathological Factors on the Outcomes of Liver Surgery for Colorectal Cancer Metastases. *Eur J Surg Oncol* (2016) 42 (9):1268–77. doi: 10.1016/j.ejso.2016.02.013
58. Mandard AM, Dalibard F, Mandard JC, Marnay J, Henry-Amar M, Petiot JF, et al. Pathologic Assessment of Tumor Regression After Preoperative Chemoradiotherapy of Esophageal Carcinoma. Clinicopathologic Correlations. *Cancer* (1994) 73(11):2680–6. doi: 10.1002/1097-0142(19940601)73:11<2680::aid-cncr2820731105>3.0.co;2-c
59. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1). *Eur J Cancer* (2009) 45(2):228–47. doi: 10.1016/j.ejca.2008.10.026
60. Lohith G. Radiomic Imaging Analysis-To Renovate RECIST Criteria? *Cancer Ther Oncol Int J* (2017) 6. doi: 10.19080/ctoij.2017.06.555679
61. Villaruz LC, Socinski MA. The Clinical Viewpoint: Definitions, Limitations of RECIST, Practical Considerations of Measurement [Internet]. *Vol 19 Clin Cancer Res* (2013) p:2629–36. doi: 10.1158/1078-0432.ccr-12-2935
62. Barnacle AM, McHugh K. Limitations With the Response Evaluation Criteria in Solid Tumors (RECIST) Guidance in Disseminated Pediatric Malignancy. *Pediatr Blood Cancer* (2006) 46:127–34. doi: 10.1002/pbc.20344
63. Choi J-Y, Lee J-M, Sirlin CB. CT and MR Imaging Diagnosis and Staging of Hepatocellular Carcinoma: Part I. Development, Growth, and Spread: Key Pathologic and Imaging Aspects. *Radiology* (2014) 272:635–54. doi: 10.1148/radiol.14132361
64. Franklin JM, Irving B, Papiez BW, Kallehauge JF, Wang LM, Goldin RD, et al. Tumour Subregion Analysis of Colorectal Liver Metastases Using Semi-Automated Clustering Based on DCE-MRI: Comparison With Histological Subregions and Impact on Pharmacokinetic Parameter Analysis. *Eur J Radiol* (2020) 126:108934. doi: 10.1016/j.ejrad.2020.108934
65. Cheung HMC, Kim JK, Hudson J, Coburn N, Karanicolas PJ, Law C, et al. Late Gadolinium MRI Enhancement of Colorectal Liver Metastases Is Associated With Overall Survival Among Nonsurgical Patients. *Eur Radiol* (2019) 29:3901–7. doi: 10.1007/s00330-019-06177-w
66. Dohan A, Gallix B, Guiu B, Le Malicot K, Reinhold C, Soyer P, et al. Early Evaluation Using a Radiomic Signature of Unresectable Hepatic Metastases to Predict Outcome in Patients With Colorectal Cancer Treated With FOLFIRI and Bevacizumab. *Gut* (2020) 69(3):531–9. doi: 10.1136/gutjnl-2018-316407
67. Mazard T, Boonsirikamchai P, Overman MJ, Asran MA, Choi H, Herron D, et al. Comparison of Early Radiological Predictors of Outcome in Patients With Colorectal Cancer With Unresectable Hepatic Metastases Treated With Bevacizumab. *Gut* (2018) 67(6):1095–102. doi: 10.1136/gutjnl-2017-313786
68. Chun YS. Association of Computed Tomography Morphologic Criteria With Pathologic Response and Survival in Patients Treated With Bevacizumab for Colorectal Liver Metastases. *Jama* (2009) 302:2338. doi: 10.1001/jama.2009.1755
69. Boonsirikamchai P, Asran MA, Maru DM, Vauthey J-N, Kaur H, Kopetz S, et al. CT Findings of Response and Recurrence, Independent of Change in

Tumor Size, in Colorectal Liver Metastasis Treated With Bevacizumab. *Am J Roentgenol* (2011) 197:W1060–6. doi: 10.2214/ajr.11.6459
70. Khan M, Loree JM, Advani SM, Ning J, Li W, Pereira AAL, et al. Prognostic Implications of Mucinous Differentiation in Metastatic Colorectal Carcinoma Can Be Explained by Distinct Molecular and Clinicopathologic Characteristics. *Clin Colorectal Cancer* (2018) 17(4):e699–709. doi: 10.1016/j.clcc.2018.07.005
71. Xia Q, Liu J, Wu C, Song S, Tong L, Huang G, et al. Prognostic Significance of (18)FDG PET/CT in Colorectal Cancer Patients With Liver Metastases: A Meta-Analysis. *Cancer Imaging* (2015) 15:19. doi: 10.1186/s40644-015-0055-z
72. Lee ST, Muralidharan V, Tebbutt N, Wong P, Fang C, Liu Z, et al. Prevalence of Hypoxia and Correlation With Glycolytic Metabolism and Angiogenic Biomarkers in Metastatic Colorectal Carcinoma. *Eur J Nucl Med Mol Imaging* (2020) 48:1585–92. doi: 10.1007/s00259-020-05074-5
73. Hong H, Sun J, Cai W. Radionuclide-Based Cancer Imaging Targeting the Carcinoembryonic Antigen. *Biomark Insights* (2008) 3:435–51. doi: 10.4137/BMI.S1124
74. Yakkundi SV, Subha DP. Convolutional LSTM: A Deep Learning Approach for Dynamic MRI Reconstruction. *2020 4th Int Conf Trends Electron Inf (ICOEI) (48184)* (2020) 48184:1011–5. doi: 10.1109/icoei48184.2020.9142982
75. Papież BW, Franklin JM, Heinrich MP, Gleeson FV, Brady M, Schnabel JA. GIFTed Demons: Deformable Image Registration With Local Structure-Preserving Regularization Using Supervoxels for Liver Applications. *J Med Imaging* (2018) 5:1. doi: 10.1117/1.jmi.5.2.024001
76. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006
77. Vorontsov E, Cerny M, Régnier P, Di Jorio L, Pal CJ, Lapointe R, et al. Deep Learning for Automated Segmentation of Liver Lesions at CT in Patients With Colorectal Cancer Liver Metastases. *Radiology: Artif Intell* (2019) 1:180014. doi: 10.1148/ryai.2019180014
78. Liang M, Cai Z, Zhang H, Huang C, Meng Y, Zhao L, et al. Machine Learning-Based Analysis of Rectal Cancer MRI Radiomics for Prediction of Metachronous Liver Metastasis. *Acad Radiol* (2019) 26(11):1495–504. doi: 10.1016/j.acra.2018.12.019
79. Sala E, Mema E, Himoto Y, Veeraraghavan H, Brenton JD, Snyder A, et al. Unravelling Tumour Heterogeneity Using Next-Generation Imaging: Radiomics, Radiogenomics, and Habitat Imaging. *Clin Radiol* (2017) 72 (1):3–10. doi: 10.1016/j.crad.2016.09.013
80. Dietterich T. Overfitting and Undercomputing in Machine Learning. *ACM Computing Surveys* (1995) 27:326–7. doi: 10.1145/212094.212114
81. Chen J, Milot L, Cheung HMC, Martel AL. Unsupervised Clustering of Quantitative Imaging Phenotypes Using Autoencoder and Gaussian Mixture Model. *Lect Notes Comput Sci* (2019) 1:575–82. doi: 10.1007/978-3-030-32251-9_63
82. Spiegelhalter D. *The Art of Statistics* London, UK: Pelican Books (2019). p. 448.
83. Dasari A, Morris VK, Allegra CJ, Atreya C, Benson AB3rd, Boland P, et al. ctDNA Applications and Integration in Colorectal Cancer: An NCI Colon and Rectal-Anal Task Forces Whitepaper. *Nat Rev Clin Oncol* (2020) 17(12):757–70. doi: 10.1038/s41571-020-0392-0

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Maclean, Tsakok, Gleeson, Breen, Goldin, Primrose, Harris and Franklin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 17 December 2021
doi: 10.3389/fonc.2021.727094

Maximum Standardized Uptake Value of ^{18}F-deoxyglucose PET Imaging Increases the Effectiveness of CT Radiomics in Differentiating Benign and Malignant Pulmonary Ground-Glass Nodules

Rong Niu[1,2], Jianxiong Gao[1,2], Xiaoliang Shao[1,2], Jianfeng Wang[1,2], Zhenxing Jiang[3], Yunmei Shi[1,2], Feifei Zhang[1,2], Yuetao Wang[1,2] and Xiaonan Shao[1,2]**

[1] Department of Nuclear Medicine, The Third Affiliated Hospital of Soochow University, Changzhou, China, [2] Changzhou Key Laboratory of Molecular Imaging, Changzhou, China, [3] Department of Radiology, the Third Affiliated Hospital of Soochow University, Changzhou, China

OPEN ACCESS

Edited by:
Joseph Ross Mitchell,
University of Alberta, Canada

Reviewed by:
Lian-Ming Wu,
Shanghai JiaoTong University, China
Qiqiang Zeng,
Wenzhou Medical University, China

***Correspondence:**
Xiaonan Shao
scorey@sina.com
Yuetao Wang
yuetao-w@163.com

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 18 June 2021
Accepted: 03 December 2021
Published: 17 December 2021

Citation:
Niu R, Gao J, Shao X, Wang J, Jiang Z, Shi Y, Zhang F, Wang Y and Shao X (2021) Maximum Standardized Uptake Value of ^{18}F-deoxyglucose PET Imaging Increases the Effectiveness of CT Radiomics in Differentiating Benign and Malignant Pulmonary Ground-Glass Nodules. Front. Oncol. 11:727094. doi: 10.3389/fonc.2021.727094

To investigate whether the maximum standardized uptake value (SUVmax) of ^{18}F-deoxyglucose (FDG) PET imaging can increase the diagnostic efficiency of CT radiomics-based prediction model in differentiating benign and malignant pulmonary ground-glass nodules (GGNs). We retrospectively collected 190 GGNs from 165 patients who underwent ^{18}F-FDG PET/CT examination from January 2012 to March 2020. Propensity score matching (PSM) was performed to select GGNs with similar baseline characteristics. LIFEx software was used to extract 49 CT radiomic features, and the least absolute shrinkage and selection operator (LASSO) algorithm was used to select parameters and establish the Rad-score. Logistic regression analysis was performed combined with semantic features to construct a CT radiomics model, which was combined with SUVmax to establish the PET + CT radiomics model. Receiver operating characteristic (ROC) was used to compare the diagnostic efficacy of different models. After PSM at 1:4, 190 GGNs were divided into benign group (n = 23) and adenocarcinoma group (n = 92). After texture analysis, the Rad-score with three CT texture features was constructed for each nodule. Compared with the Rad-score and CT radiomics model (AUC: 0.704 (95%CI: 0.562-0.845) and 0.908 (95%CI: 0.842-0.975), respectively), PET + CT radiomics model had the best diagnostic efficiency (AUC: 0.940, 95%CI: 0.889-0.990), and there was significant difference between each two of them (P = 0.001-0.030). SUVmax can effectively improve CT radiomics model performance in the differential diagnosis of benign and malignant GGNs. PET + CT radiomics might become a noninvasive and reliable method for differentiating of GGNs.

Keywords: lung adenocarcinoma, radiomics, standardized uptake value, fluorodeoxyglucose F18, positron emission tomography-computed tomography

INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths worldwide, especially in China (1, 2). The incidence of lung cancer is increasing rapidly. It is predicted that China's lung cancer mortality will increase by about 40% from 2015 to 2030 (1, 2). Early diagnosis and treatment are crucial for improving the prognosis of patients. With the significant increase in the detection of many asymptomatic pulmonary nodules and the change in the epidemiological trend of lung cancer in China, diagnosis, and differentiation of ground-glass pulmonary nodules (GGNs) has become a huge challenge for clinicians (3). It is reported that the probability of malignancy of GGNs is higher than that of solid pulmonary nodules (4), but it can also be caused by benign lesions such as organizing pneumonia and interstitial pneumonia. High-resolution computed tomography (HRCT) is generally recognized as a routine method for differentiating GGNs. However, the radiological features of benign and malignant GGNs are overlapping, and the judgment of the characteristics of the lesion is easily to be affected by subjective factors. Therefore, the diagnostic efficiency of HRCT needs to be improved (5). According to the recommendation of guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images, pulmonary GGNs that cannot be characterized can be further identified by CT follow-up to observe the dynamic changes of GGNs (6). However, some benign GGNs and early lung adenocarcinoma remain stable for a long time, making it difficult to differentiate them (7). Moreover, long-term follow-up often brings panic and anxiety to patients. Pathological examination is the gold standard for the diagnosis of GGNs. However, the cell composition of GGN is relatively small, which requires a highly skillful puncture technique for pathological examination that is difficult to perform. Thus, bronchoscopy and percutaneous lung puncture techniques have limited application value in GGNs. Therefore, it is urgent to develop a reliable and practical noninvasive imaging method to accurately distinguish benign and malignant GGNs to guide the individualized clinical management strategy for GGNs.

Radiomics is a very promising diagnostic method. With the help of mathematical and statistical methods, high-throughput characteristic spatial data can be extracted from the image data of the region of interest, and valuable lesion information that the naked eye may ignore can be effectively captured to improve the accuracy of disease diagnosis (8–10). Radiomics has the advantages of real-time, objective, noninvasive, and reusability. Previous studies have shown that CT texture features are potentially radiological biomarkers in the differential diagnosis of lung cancer (11, 12), the prediction of tumor growth (13), gene expression (14), and the evaluation of therapeutic efficacy (9, 15). Most of these studies are based on solid pulmonary nodules, and there are few reports on the differentiation of benign and malignant GGNs. Theoretically, inflammatory lesions and malignant tumors have completely different biological behaviors, pathological processes, and internal spatial structures. Therefore, radiomics that is based on multi-dimensional characteristics can identify malignant lesions from benign GGNs. Digumarthy et al. (16) differentiated benign and malignant lesions in 108 GGNs obtained from 36 patients and found that only 2/92 radiomic features (cluster shade and surface volume ratio) could be used for model prediction, with AUC=0.624, which is of diagnostic value, but its diagnostic efficacy needs to be improved. It has been reported that the model with the combination of radiomics and semantic parameters can improve the performance of radiomics model alone (17). Besides, PET/CT imaging, as noninvasive dual-modality imaging that reflects tumor heterogeneity, has been recognized for its application in the field of lung cancer. Our previous studies also found that PET metabolic parameters help identify GGNs (18). Thus, we proposed that adding PET metabolic parameters (SUV) on the basis of CT radiomics model will be beneficial in the differentiation of benign and malignant GGNs.

The purpose of this study was to establish a dual-modality comprehensive prediction model based on CT texture parameters, semantic features, and PET metabolic parameters through analyzing the PET/CT images of patients with indeterminate lung GGNs who underwent ^{18}F-FDG PET/CT examination before operation and to investigate whether the SUVmax can increase the diagnostic efficiency of CT radiomics-based prediction model in differentiating benign and malignant GGNs.

Abbreviations: 18F-FDG, 18F-fluorodeoxyglucose; AUC, area under the curve; CI, confidence interval; GGN, ground-glass nodule; HRCT, High-resolution computed tomography; ICC, intraclass correlation coefficients; LASSO, least absolute shrinkage and selection operator; mGGN, mixed ground-glass nodule; PET, positron emission tomography; PSM, propensity score matching; pGGN, pure ground-glass nodule; ROC, receiver operating characteristic; ROI, region of interest; SD, standard deviation; SUVmax, maximum standardized uptake value.

MATERIALS AND METHODS

Research Objects

This was a single-center case-control study. We retrospectively selected patients who received ^{18}F-FDG PET/CT examination in our hospital from January 2012 to March 2020 for indeterminate GGNs. The study was approved by the Institutional Ethics Committee, and no informed consent was required from the patients for retrospective study [approval No.: (2020) Science No. 075]. Inclusion criteria: patients with lung GGN; GGN ≤3 cm; patients who underwent PET/CT scan and breath-holding chest CT scan in our department; the lesions were resected within 1 month after PET/CT examination, and lung adenocarcinoma and benign lesions with complete postoperative pathological data. Exclusion criteria: lesions with poor quality images that affected the measurement; patients who received any anti-tumor treatment; lung cancer patients with stage IB or above; patients with fasting blood glucose level >11.1 mmol/L; patients with impaired liver function (serum alanine aminotransferase or aspartate aminotransferase exceeding five times the upper limit of the normal value).

The total number of GGNs that met the inclusion criteria was 190, belonging to 165 patients, including 53 males and 112 females, aged 60.8 ± 9.2 years (range 31-81 years). All GGNs

were divided into a benign group (n = 23) and an adenocarcinoma group (n = 167) according to postoperative pathology [the pathological classification of adenocarcinoma group was based on the classification of lung adenocarcinoma published by IASLC/ATS/ERS in 2011 (19), and the staging of the lesion was based on the Eighth Edition of the TNM Classification of Lung Cancer published by the Union for International Cancer Control (UICC) in 2017 (20)]. The patient selection process is shown in **Figure 1**.

PET/CT Examination

PET/CT imaging was performed using Siemens biograph mCT (64) PET/CT scanner. The examination procedure was as follows: After fasting for 4-6 h, the height and weight of the patients were measured, and fingertip blood was collected for the blood glucose test. The imaging agent ^{18}F-FDG was injected intravenously into the back of the hand or elbow at a dose of (3.70-5.55) MBq/kg, and the images were collected after the patient rested of (60 ± 5) min. The patient was lying on the examination table supine, with both hands holding the head. Low-dose whole-body CT scanning was performed first from skull base to middle femur, and then whole-body PET scanning was applied with the same range at 2 min/bed. Each patient was scanned for about 6-7 beds. No respiratory gating device was used during image acquisition. Image reconstruction: TrueX + TOF (ultraHD-PET) with 2 iterations and 21 subsets, and Gaussian filtering with a full-width at half maximum of 2.0 mm; matrix (pixels) 200 × 200, zoom 1.00, the image acquisition mode was 3D. The image was evaluated using TrueD software (Siemens). CT data were used for attenuation correction of the PET images, and the corrected PET images were merged with CT images. TrueD software (Siemens) was used to display and analyze the images.

CT Examination and Image Reconstruction

After the whole-body PET/CT scan, the breath-holding chest CT scan was performed immediately, and the GGN site was reconstructed with thin slices. Acquisition and reconstruction conditions: The tube current was automatically adjusted according to the human body's anatomical structure and tissue density. The tube voltage was 140 kV, the rotation time was 0.5 seconds, the pitch was 0.6, the kernel was B70f very sharp, the matrix was 512 × 512, the reconstruction layer thickness was 3.0 mm and 1.0 mm, the window width was 1200 HU, and the window level was -600 HU.

Clinical and Imaging Parameters

The clinical data collected included age, sex, smoking history, and fasting blood glucose level of the patients. The CT parameters included the number of GGN (solitary, multifocal), type of GGN (pGGN, mGGN), location (central, peripheral), shape (round/oval, irregular), margin (smooth, lobulated), abnormal bronchus sign, vacuole sign, pleural indentation sign, vascular convergence sign. PET parameter: SUVmax of nodules.

Texture Analysis and Feature Extraction

The patient's chest CT images were exported from Siemens workstation in DICOM format and uploaded into LIFEx software (version 5.10, http://www.lifexsoft.org). Two experienced nuclear medicine doctors (Niu R and Shao X) manually delineated each target lesion, drew the region of interest (ROI) layer by layer along the lesion's contour, and automatically calculated and extracted texture features for each GGN.

Statistical Analysis

First, the GGNs that met the inclusion criteria were matched with propensity score (PSM) according to benign to malignant ratio at 1:4. The PS variables were age, sex, smoking history, fasting blood glucose level, and GGN number grouping (**Supplementary Table S1**). The continuous variables with normal distribution were expressed as the mean ± standard deviation (SD), continuous variables with non-normal distribution were expressed as the median (Q1-Q3), and the

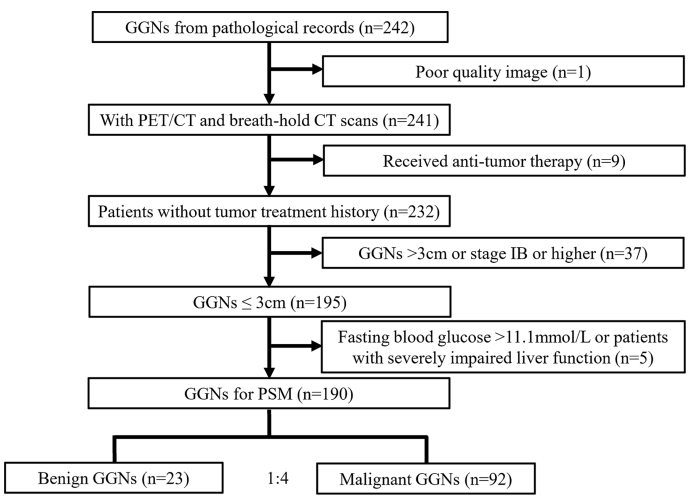


FIGURE 1 | Flow chart of the study. GGN = ground-glass nodule, PSM = propensity score matching.

categorical variables were expressed as frequency (%). T-test or Mann Whitney U test was used to compare continuous variable data between groups, chi-square test or Fisher test was used to compare categorical data between groups.

For texture feature selection, we first removed two parameters with intraclass correlation coefficients (ICCs) lower than 0.75. Next, the Mann Whitney U test was performed to screen the parameters between benign and adenocarcinoma groups (*P*-value was relaxed to 0.15). The best predictive features were selected using the least absolute shrinkage and selection operator (LASSO) algorithm and 10-fold cross-validation, and the Rad-score of each GGN was calculated. Multivariable logistic regression analysis was carried out to construct the joint model using Rad-score, CT conventional morphological parameters (semantic features), and PET parameters (SUVmax) with different CT reconstruction slice thickness. The receiver operating characteristic (ROC) curve was prepared for each model, and the area under the curve (AUC) was calculated. The bootstrap resampling method (times = 500) (21) recommended by the TRIPOD statement was used to internally verify the model and calculate 95% confidence interval (CI) of AUC. Delong test was used to compare whether the differences in effectiveness between the models were statistically significant (P <0.05), and the nomogram of the model was generated. All statistical analyses were performed using R software, version 3.4.3 (http://www.R-project.org; software package: glmnet, pROC, rms, dca.R).

RESULTS

General Data

After PSM at a ratio of 1:4, 190 GGNs were divided into benign and adenocarcinoma groups. There were 23 GGNs in benign group (including 3 organizing pneumonia, 4 fungal infections, 1 interstitial pneumonia, 5 granulomatous inflammation, and 10 other benign lesions), and 92 GGNs in adenocarcinoma group (including 77 invasive lung adenocarcinoma, 8 microinvasive adenocarcinoma, and 7 preinvasive lesions). The general data of GGNs before and after PSM is shown in **Table 1**. Comparing CT semantic features and SUVmax of GGNs after PSM, we found that only pleural indentation and SUVmax were significantly different between the benign and adenocarcinoma groups (P <0.001 and 0.024, respectively). There were no significant differences in nodule type, location, shape, margin, abnormal bronchus sign, vacuole sign, and vascular convergence between the benign group and adenocarcinoma group (all P >0.05) (**Table 2**).

CT Texture Features Analysis

Through texture analysis, 49 features were obtained for each GGN (**Supplementary Table S2**). According to ICC analysis of two different readers, the ICCs of three texture features of 3 mm slice thickness CT images were lower than 0.75, including GLZLM_SZE, GLZLM_SZLGE, and GLZLM_ZP, indicating that these three texture features needed to be eliminated in subsequent analysis. In contrast, the ICCs of GGN texture features of 1 mm slice thickness CT images were all above 0.75 (**Supplementary Table S3**).

Comparison of CT Texture Features of Different Slice Thickness Between Benign Group and Adenocarcinoma Group

Comparing CT texture features of different layer thicknesses of GGNs between benign group and adenocarcinoma group, we found that there were 6 parameters (CONVENTIONAL_HUmin, GLRLM_LGRE, GLRLM_SRLGE, GLRLM_LRLGE, GLZLM_LGZE, and GLZLM_LZLGE) in the texture features of 3 mm slice thickness CT images that passed the primary screening (P = 0.064-0.149). There were 12 parameters in the texture features of 1 mm slice thickness CT images that passed the primary screening (P = 0.018-0.112), including HISTO_Kurtosis, HISTO_ExcessKurtosis, HISTO_Energy, GLCM_Homogeneity, GLRLM_SRE, GLRLM_LRE, GLRLM_LRLGE, GLRLM_RP, NGLDM_Contrast, GLZLM_SZE, GLZLM_LZE, and GLZLM_ZP (**Supplementary Table S4**).

Rad-Scores of CT Images With Different Slice Thickness

LASSO algorithm and 10-fold cross-validation were carried out to extract the best subset of CT radiomics features. For CT

TABLE 1 | General data of GGNs in benign group and adenocarcinoma group before and after PSM.

	Before matching			After matching		
	Benign (n = 23)	Adenocarcinoma (n = 167)	*P*-value	Benign (n = 23)	Adenocarcinoma (n = 92)	*P*-value
Age (years)	55.8 ± 10.5	60.8 ± 8.7	0.013	55.8 ± 10.5	57.4 ± 8.9	0.477
Sex			0.002			0.428
Female	9 (39.1%)	120 (71.9%)		9 (39.1%)	47 (51.1)	
Male	14 (60.9%)	47 (28.1%)		14 (60.9%)	45 (48.9)	
History of smoking			0.039			0.800
No	15 (65.2%)	139 (83.2%)		15 (65.2%)	65 (70.7)	
Yes	8 (34.8%)	28 (16.8%)		8 (34.8%)	27 (29.3)	
Fasting blood glucose (mmol/L)	6.8 ± 1.9	6.7 ± 1.7	0.960	6.8 ± 1.9	6.65 ± 1.69	0.772
GGN number grouping			0.448			1.000
Solitary	15 (65.2%)	95 (56.9%)		15 (65.2%)	58 (63)	
Multifocal	8 (34.8%)	72 (43.1%)		8 (34.8%)	34 (37)	

Results in the table: Mean ± SD/N (%).

TABLE 2 | Comparison of CT semantic features and SUVmax of GGNs between benign group and adenocarcinoma group after PSM.

Features	Benign (n = 23)	Adenocarcinoma (n = 92)	*P*-value
Type			0.625
pGGN	7 (30.4%)	33 (35.9%)	
mGGN	16 (69.6%)	59 (64.1%)	
Location			1.000
Peripheral	22 (95.7%)	88 (95.7%)	
Central	1 (4.3%)	4 (4.3%)	
Shape			0.922
Round/oval	15 (65.2%)	59 (64.1%)	
Irregular	8 (34.8%)	33 (35.9%)	
Margin			0.111
Smooth	16 (69.6%)	47 (51.1%)	
Lobulated	7 (30.4%)	45 (48.9%)	
Abnormal bronchus sign			0.080
No	10 (43.5%)	23 (25.0%)	
Yes	13 (56.5%)	69 (75.0%)	
Vacuole sign			0.904
No	19 (82.6%)	75 (81.5%)	
Yes	4 (17.4%)	17 (18.5%)	
Pleural indentation			<0.001
No	19 (82.6%)	40 (43.5%)	
Yes	4 (17.4%)	52 (56.5%)	
Vascular convergence			0.559
No	2 (8.7%)	5 (5.4%)	
Yes	21 (91.3%)	87 (94.6%)	
SUVmax	2.9 (1.4-6.9)	1.8 (1.1-3.0)	0.024

Results in the table: Median (Q1-Q3)/N (%).

images of 3 mm slice thickness: Rad-score (3 mm) = −0.00198 × CONVENTIONAL_HUmin + 10.16317× GLRLM_LGRE + 66.97979 × GLRLM_SRLGE.

For CT images of 1 mm slice thickness: Rad-score (1 mm) = 11.21344 × GLRLM_LRE + 10.43443 × GLRLM_LRLGE - 0.7472 × NGLDM_Contrast.

ROC curve analysis showed that the AUC values of the Rad-score of the two reconstruction slices with different thickness were 0.634 (95%CI: 0.499-0.768) and 0.704 (95%CI: 0.562-0.845), respectively, with no significant difference between them (Z = 0.702, *P* = 0.483) (**Supplementary Figures S1**, **S2**). The diagnostic efficiency of the Rad-score of the two reconstruction slices is shown in **Table 3**.

Rad-Score in Combination With CT Semantic Features to Construct CT Radiomics Model

Multivariable logistic regression analysis was performed to construct the CT radiomics models to predict the benign and malignant GGNs using semantic characteristic parameters of GGNs and the Rad-scores of CT with different slice thickness individually. The formula of CT radiomics model (3 mm) was as follows: CT radiomics model (3 mm) = −5.20333 + 1.16220 × (abnormal bronchus sign = 1) + 2.12571 × (pleural indentation sign =1) + 2.23666 × Rad-score (3 mm). The AUC value of the 3 mm CT radiomics model was 0.794 (95%CI: 0.704-0.884), which was significantly higher than that of Rad-score (3 mm) (Z = 2.232, *P* = 0.026) (**Figure 2**).

The formula of CT radiomics model (1 mm) was as follows: CT radiomics model (1 mm) = −45.94109 + 3.06092 × (abnormal bronchus sign = 1) + 3.33342 × (pleural indentation sign = 1) + 3.89642 × Rad-score (1 mm). The AUC value of the model was 0.908 (95%CI: 0.842-0.975), which was significantly higher than that of Rad-score (1 mm) (0.704) (Z = 2.769, *P* = 0.006) and CT radiomics model (3 mm) (0.794) (Z = 1.998, *P* = 0.046) (**Figure 2**). The diagnostic efficiency of different models is shown in **Table 3**.

Construction of PET + CT Radiomics Model

Furthermore, based on the CT radiomics model (1 mm) and combined with SUVmax in PET parameters, a dual-modality prediction model (PET + CT radiomics model) was established to

TABLE 3 | Comparison of diagnostic efficiency of different models.

Model	AUC (95%CI)	Best threshold	Sensitivity	Specificity	Accuracy
Rad-score (3 mm)	0.634 (0.499-0.768)	2.162	0.707	0.609	0.687
Rad-score (1 mm)	0.704 (0.562-0.845)	10.920	0.852	0.609	0.779
CT radiomics model (3 mm)	0.794 (0.704-0.884)	0.903	0.794	0.739	0.783
CT radiomics model (1 mm)	0.908 (0.842-0.975)	1.242	0.815	0.956	0.857
PET + CT radiomics model	0.940 (0.889-0.990)	1.931	0.815	1.000	0.870

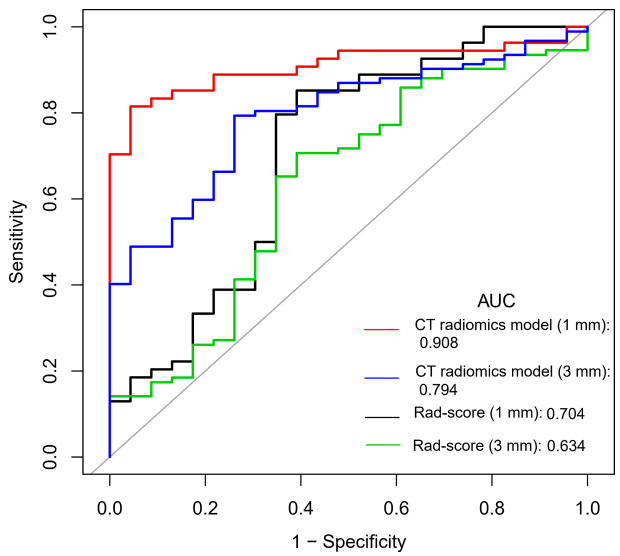


FIGURE 2 | Comparison of ROC curves between Rad-score and radiomics models with different CT reconstruction slice thickness.

predict the benign and malignant GGNs. The formula was as follows: PET + CT radiomics model = −89.87509 + 7.01593 × (abnormal bronchus sign = 1) + 5.03616 × (pleural indentation sign = 1) +7.74753 × Rad-score (1 mm) − 0.84485 × SUVmax. The AUC of this radiomics model was 0.940 (95%CI: 0.889-0.990), and the nomogram of the model as shown in **Figure 3A**. Two examples of how to use the nomogram were showed in **Figures 3B–E**.

We compared the diagnostic efficiency of Rad-score of 1 mm CT, CT radiomics model, and PET + CT radiomics model in differentiating benign and malignant GGNs. The results showed that the AUC values of the three models were significantly different from each other, with Z = 2.174-3.304 and P = 0.001-0.030, and the PET + CT radiomics model had the highest diagnostic efficiency (**Figure 4**). The comparison of diagnostic efficiency of different models is shown in **Table 3**.

DISCUSSION

This study systematically analyzed the semantic features of suspected lung GGNs and 49 radiomics features extracted from each GGN and constructed a CT radiomics model with abnormal bronchus sign, pleural indentation sign, and Rad-score. This CT radiomics model had good differentiation efficiency in benign and malignant GGNs. We further added SUVmax to the CT radiomics model to construct the PET + CT radiomics model, which effectively improved the predictive value of CT radiomics model in differentiating benign and malignant GGNs.

After PSM, the proportion of pleural indentation sign in CT semantic features of adenocarcinoma group was significantly higher than that of benign GGNs. Pleural indentation sign refers to the pleural indentation caused by the traction of subpleural lesions on adjacent pleura, which mostly presents as linear or triangular shadow. Hu et al. reported that this sign is one of the common features of peripheral lung cancer (22). Although some semantic features play an important role in clinical application, they depend on the experience of radiologists or surgeons and their understanding of the signs, and thus they are easily affected by subjective factors. Previous literature also reported that single CT parameters, such as nodule margin characteristics, size, and CT value, have limited values in differentiating benign and malignant GGNs (23). Besides, in univariate comparison, we found that SUVmax of benign GGNs was higher than that of

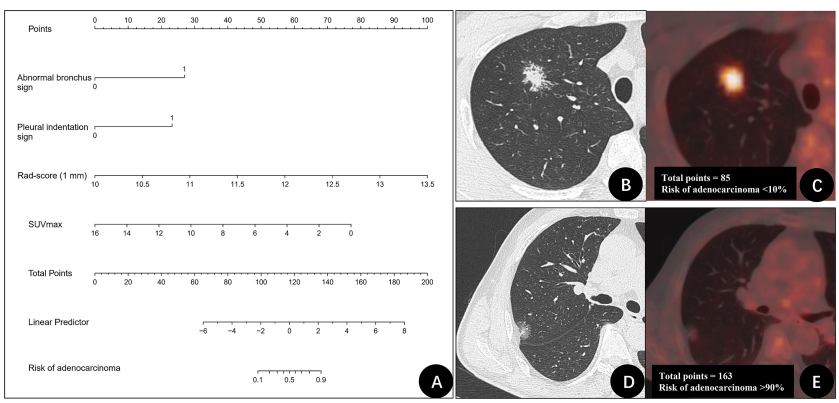


FIGURE 3 | The nomogram of PET + CT radiomics model for differentiating benign and malignant GGNs and two examples. **(A)** Nomogram of PET + CT radiomics model. **(B, C)** A 31-year-old man with a ground-glass nodule (GGN) on the right upper lung lobe. CT image **(B)** and PET/CT fusion image **(C)** show that nodule with abnormal bronchus sign (27 points), and no pleural indentation was identified (0 points). Rad-score (1 mm) was 10.7 (20 points), maximum standardized uptake value (SUVmax) was 8.1 (38 points). The total points were 85 points. The risk of adenocarcinoma for this nodule was < 10%. Postoperative pathologic findings indicated granuloma. **(D, E)** A 61-year-old man with GGN on the right upper lung lobe. CT image **(D)** and PET/CT fusion image **(E)** show that nodule with abnormal bronchus sign (27 points) and pleural indentation sign (23 points). Rad-score (1 mm) was 11.3 (38 points), SUVmax was 1.1 (71 points). The total points were 163 points. The risk of adenocarcinoma for this nodule was > 90%. Postoperative pathologic findings indicated invasive adenocarcinoma.

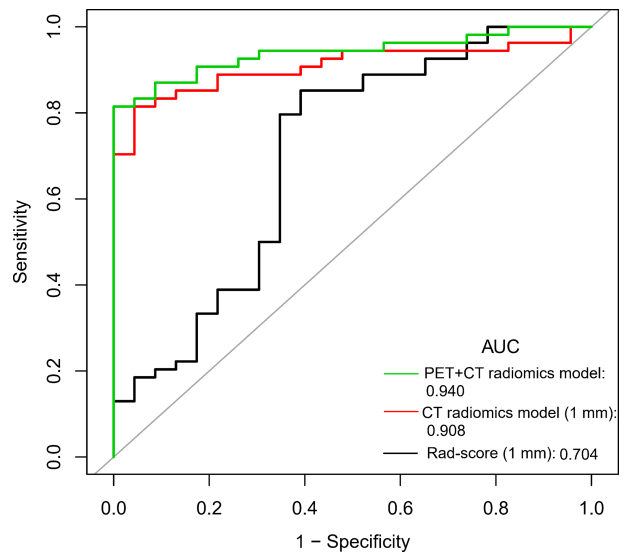


FIGURE 4 | Comparison of ROC curves of Rad-score, CT radiomics model, and PET + CT radiomics model of 1 mm CT.

malignant GGNs, which is consistent with the findings of McDermott et al. (24) and Chun et al. (25). This may be because benign GGN is more common in tuberculosis, organizing pneumonia, and other inflammatory lesions. It is well known that SUV is a semi-quantitative parameter that reflects the uptake of ^{18}F-FDG in the lesion. The increase of SUV value represents the increased uptake of glucose in the lesion, the strong proliferation and growth of cells, and the possibility of malignant tumor. However, ^{18}F-FDG is not a tumor-specific imaging agent. The chemotaxis of inflammatory factors and the accumulation of inflammatory cells may also cause a significant increase in the uptake of FDG (26, 27). On the contrary, early lung adenocarcinoma manifesting as GGN showed lower density, slower growth, less expression of glucose transporter-1 (GLUT-1), less uptake of FDG, and lower SUVmax (28, 29).

This study found that the consistency of CT images 1 mm slice is better, whereas there are three parameters with greater variability in 3 mm CT texture features (including GLZLM_SZE GLZLM_SZLGE, and GLZLM_ZP). In the analysis of more than 100 PET texture features, Leijenaar et al. (30) also found that GLZLM feature-based parameters have the highest variability. In our study, we also found that the number and types of texture features of CT images that were significantly different between benign group and adenocarcinoma group were different with different slice thickness, which may be related to the different spatial information display of image details on CT images with different reconstruction slice thickness. Some scholars reported that applying different reconstruction software, parameter settings, and reconstruction methods can affect radiomics features extraction (31–33).

The Rad-score (3 mm) contains the traditional CT parameters, namely, CONVENTIONAL_HUmin and two GLRLM texture features. HUmin is the minimum CT value of VOI in lesions. The lower the HUmin value is, the higher the possibility of adenocarcinoma is. This may be because malignant GGN is more prone to exhibit vacuole signs than benign lesions. GLRLM was introduced by Galloway (34), which assesses the distribution of discretized grey levels on an image or a stack of images. It describes the roughness or smoothness of the image and reflects the heterogeneity of the tumor. In addition to GLRLM, Rad-score (1 mm) parameter NGLDM_Contrast was also used in the radiomics model. NGLDM_Contrast is the intensity difference between neighboring regions, which provides information about the spatial relationship between an image voxel and its neighboring voxels (35). The lower the NGLDM_Contrast, the higher the possibility of adenocarcinoma. It has been reported that NGLDM_Contrast is also an important prognostic factor for lung cancer, and poor prognosis is associated with low NGLDM_Contrast (36). Comparing the Rad-score between slices with different reconstruction thickness, we found that although the diagnostic efficiency of Rad-score (1 mm) was slightly higher than that of Rad-score (3 mm), there was no statistical difference, and the diagnostic efficiency of both Rad-score (1 mm) and Rad-score (3 mm) was moderate, which is similar to the finding of Digumarthy (15) (AUC = 0.624).

In this study, the diagnostic efficiency of the CT radiomics model was significantly better than that of Rad-score alone, which is consistent with Hyun et al. (37) and Bianconi et al. (38). Moreover, the thin slice CT radiomics model's prediction efficiency is better than that of the 3 mm slice CT radiomics model, which agrees with the previous reports (31). This may be because a thin-slice CT image can improve the image's spatial resolution and facilitate the display of confidential information of lesions. In the thin slice CT radiomics model, the abnormal bronchus sign is also an independent risk factor. Bronchus sign refers to the appearance of air containing bronchus in the lesion. When bronchus is dilated, distorted or cut-off truncated, it often indicates the possibility of malignant lesions. Thus, this sign is called an abnormal bronchus sign (39). Finally, compared with Rad-score and CT radiomics model, PET + CT radiomics model had the best diagnostic efficiency (AUC = 0.940). PET imaging parameters reflect the lesions' functional and metabolic status, provide quantitative information at the molecular level, and complement CT's anatomical images. The establishment of the dual-modality comprehensive PET/CT model helps evaluate the lesions at multiple levels.

There are still some limitations to this study. First of all, patients enrolled in this study were those who received preoperative PET/CT differential diagnosis and staging because of suspicious GGNs, and therefore the number of benign cases is small, which is the reason why we conducted PSM. In addition, there were many types of diseases in the benign group. This heterogeneity may affect the reliability of the model. The radiomics model constructed in this study may be only suitable for the differential diagnosis of GGNs that cannot be determined on CT but not for the screening of GGNs. Secondly, our previous study showed that LIFEx software has requirements for voxels, and it is not suitable for some small or low uptake GGNs. Therefore, we did not perform a PET texture analysis. PET radiomics needs to be further explored. Thirdly, we chose the manual segmentation method for ROI delineation, which is not as stable as the fully/semi-automatic segmentation method.

In a future study, we may try to use the AI-based segmentation method to obtain image information. Fourth, although the PET + CT radiomics model has good internal validation performance, external data remain needed to confirm the robustness and applicability of this radiomics model.

In conclusion, in this study, we successfully constructed CT texture feature-based Rad-score, a CT radiomics model using Rad-score in combination with semantic features (abnormal bronchus sign and pleural indentation sign) and a PET + CT radiomics model using CT radiomics model in combination with SUVmax to differentiate early lung adenocarcinoma from benign lung GGNs and compared the differential diagnostic efficacy of these models. PET + CT radiomics model has the best risk prediction performance and might become a noninvasive and reliable diagnostic tool for differentiating benign and malignant GGNs.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

Our study was approved by the ethics committee of the Third Affiliated Hospital of Soochow University for retrospective analysis and did not require informed consent. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RN and XNS contributed to the study concepts and the study design. RN, JG, YS, FZ and ZJ contributed to data acquisition and reconstructions. XNS, RN, XLS, and ZJ contributed to data analyses and interpretation. XNS contributed to the statistical analysis. XNS, RN, and YW contributed to the manuscript preparation and manuscript editing and reviewing. All authors contributed to the article and approved the submitted version.

FUNDING

Funding from Young Talent Development Plan of Changzhou Health Commission (CZQM2020012) and Key Laboratory of Changzhou High-tech Research Project (Grant No. CM20193010).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.727094/full#supplementary-material

REFERENCES

1. Hong QY, Wu GM, Qian GS, Hu CP, Zhou JY, Chen LA, et al. Prevention and Management of Lung Cancer in China. *Cancer* (2015) 121(Suppl 17):3080–8. doi: 10.1002/cncr.29584
2. Martín-Sánchez JC, Lunet N, González-Marrón A, Lidón-Moyano C, Matilla-Santander N, Clèries R, et al. Projections in Breast and Lung Cancer Mortality Among Women: A Bayesian Analysis of 52 Countries Worldwide. *Cancer Res* (2018) 78(15):4436–42. doi: 10.1158/0008-5472.Can-18-0187
3. Chen KN. Commentary: Pay Attention to Low-Risk Populations for Lung Cancer, But Cautiously Interpret Ground-Glass Nodules Screened by Low-Dose Computed Tomography Scan. *J Thorac Cardiovasc Surg* (2020) 160 (3):833–4. doi: 10.1016/j.jtcvs.2019.10.204
4. Godoy MC, Naidich DP. Overview and Strategic Management of Subsolid Pulmonary Nodules. *J Thorac Imaging* (2012) 27(4):240–8. doi: 10.1097/RTI.0b013e31825d515b
5. Yang W, Sun Y, Fang W, Qian F, Ye J, Chen Q, et al. High-Resolution Computed Tomography Features Distinguishing Benign and Malignant Lesions Manifesting as Persistent Solitary Subsolid Nodules. *Clin Lung Cancer* (2018) 19(1):e75–83. doi: 10.1016/j.cllc.2017.05.023
6. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for Management of Incidental Pulmonary Nodules Detected on CT Images: From the Fleischner Society 2017. *Radiology* (2017) 284(1):228–43. doi: 10.1148/radiol.2017161659
7. Hasegawa M, Sone S, Takashima S, Li F, Yang ZG, Maruyama Y, et al. Growth Rate of Small Lung Cancers Detected on Mass CT Screening. *Br J Radiol* (2000) 73(876):1252–9. doi: 10.1259/bjr.73.876.11205667
8. Chen B, Zhang R, Gan Y, Yang L, Li W. Development and Clinical Application of Radiomics in Lung Cancer. *Radiat Oncol (Lond Engl)* (2017) 12(1):154. doi: 10.1186/s13014-017-0885-x
9. Hassani C, Varghese BA, Nieva J, Duddalwar V. Radiomics in Pulmonary Lesion Imaging. *AJR Am J Roentgenol* (2019) 212(3):497–504. doi: 10.2214/ajr.18.20623
10. Xue B, Wu S, Zheng M, Jiang H, Chen J, Jiang Z, et al. Development and Validation of a Radiomic-Based Model for Prediction of Intrahepatic Cholangiocarcinoma in Patients With Intrahepatic Lithiasis Complicated by Imagologically Diagnosed Mass. *Front Oncol* (2020) 10:598253. doi: 10.3389/fonc.2020.598253
11. Hawkins S, Wang H, Liu Y, Garcia A, Stringfield O, Krewer H, et al. Predicting Malignant Nodules From Screening CT Scans. *J Thorac Oncol* (2016) 11 (12):2120–8. doi: 10.1016/j.jtho.2016.07.002
12. Liu A, Wang Z, Yang Y, Wang J, Dai X, Wang L, et al. Preoperative Diagnosis of Malignant Pulmonary Nodules in Lung Cancer Screening With a Radiomics Nomogram. *Cancer Commun (Lond Engl)* (2020) 40(1):16–24. doi: 10.1002/cac2.12002
13. Lu H, Mu W, Balagurunathan Y, Qi J, Abdalah MA, Garcia AL, et al. Multi-Window CT Based Radiomic Signatures in Differentiating Indolent Versus Aggressive Lung Cancers in the National Lung Screening Trial: A Retrospective Study. *Cancer Imaging* (2019) 19(1):45. doi: 10.1186/s40644-019-0232-6
14. Jia TY, Xiong JF, Li XY, Yu W, Xu ZY, Cai XW, et al. Identifying EGFR Mutations in Lung Adenocarcinoma by Noninvasive Imaging Using Radiomics Features and Random Forest Modeling. *Eur Radiol* (2019) 29 (9):4742–50. doi: 10.1007/s00330-019-06024-y
15. Avanzo M, Stancanello J, Pirrone G, Sartor G. Radiomics and Deep Learning in Lung Cancer. *Strahlenther Onkol* (2020) 196(10):879–87. doi: 10.1007/s00066-020-01625-9
16. Digumarthy SR, Padole AM, Rastogi S, Price M, Mooradian MJ, Sequist LV, et al. Predicting Malignant Potential of Subsolid Nodules: Can Radiomics Preempt Longitudinal Follow Up CT? *Cancer Imaging* (2019) 19(1):36. doi: 10.1186/s40644-019-0223-7

17. Luo T, Xu K, Zhang Z, Zhang L, Wu S. Radiomic Features From Computed Tomography to Differentiate Invasive Pulmonary Adenocarcinomas From non-Invasive Pulmonary Adenocarcinomas Appearing as Part-Solid Ground-Glass Nodules. *Chin J Cancer Res* (2019) 31(2):329–38. doi: 10.21147/j.issn.1000-9604.2019.02.07
18. Niu R, Shao X, Shao X, Wang J, Jiang Z, Wang Y. Lung Adenocarcinoma Manifesting as Ground-Glass Opacity Nodules 3 Cm or Smaller: Evaluation With Combined High-Resolution CT and PET/CT Modality. *AJR Am J Roentgenol* (2019) 213(5):W236–W45. doi: 10.2214/ajr.19.21382
19. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *J Thorac Oncol* (2011) 6(2):244–85. doi: 10.1097/JTO.0b013e318206a221
20. Travis WD, Asamura H, Bankier AA, Beasley MB, Detterbeck F, Flieder DB, et al. The IASLC Lung Cancer Staging Project: Proposals for Coding T Categories for Subsolid Nodules and Assessment of Tumor Size in Part-Solid Tumors in the Forthcoming Eighth Edition of the TNM Classification of Lung Cancer. *J Thorac Oncol* (2016) 11(8):1204–23. doi: 10.1016/j.jtho.2016.03.025
21. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement. *BMJ (Clinical Res Ed)* (2015) 350: g7594. doi: 10.1136/bmj.g7594
22. Hu H, Wang Q, Tang H, Xiong L, Lin Q. Multi-Slice Computed Tomography Characteristics of Solitary Pulmonary Ground-Glass Nodules: Differences Between Malignant and Benign. *Thorac Cancer* (2016) 7(1):80–7. doi: 10.1111/1759-7714.12280
23. Dai J, Yu G, Yu J. Can CT Imaging Features of Ground-Glass Opacity Predict Invasiveness? A Meta-Analysis. *Thorac Cancer* (2018) 9(4):452–8. doi: 10.1111/1759-7714.12604
24. McDermott S, Kilcoyne A, Wang Y, Scott JA, Halpern EF, Ackman JB. Comparison of the (18)F-FDG Avidity at PET of Benign and Malignant Pure Ground-Glass Opacities: A Paradox? *Clin Radiol* (2019) 74(3):187–95. doi: 10.1016/j.crad.2018.12.009
25. Chun EJ, Lee HJ, Kang WJ, Kim KG, Goo JM, Park CM, et al. Differentiation Between Malignancy and Inflammation in Pulmonary Ground-Glass Nodules: The Feasibility of Integrated (18)F-FDG PET/Ct. *Lung Cancer (Amsterdam Netherlands)* (2009) 65(2):180–6. doi: 10.1016/j.lungcan.2008.11.015
26. Prabhu M, Raju S, Chakraborty D, Arora S, Kumar R. Spectrum of 18F-FDG Uptake in Bilateral Lung Parenchymal Diseases on PET/Ct. *Clin Nucl Med* (2020) 45(1):e15–e9. doi: 10.1097/rlu.0000000000002758
27. Qin C, Liu F, Yen TC, Lan X. (18)F-FDG PET/CT Findings of COVID-19: A Series of Four Highly Suspected Cases. *Eur J Nucl Med Mol Imaging* (2020) 47 (5):1281–6. doi: 10.1007/s00259-020-04734-w
28. Shao X, Niu R, Jiang Z, Shao X, Wang Y. Role of PET/CT in Management of Early Lung Adenocarcinoma. *AJR Am J Roentgenol* (2020) 214(2):437–45. doi: 10.2214/ajr.19.21585
29. Scafoglio CR, Villegas B, Abdelhady G, Bailey ST, Liu J, Shirali AS, et al. Sodium-Glucose Transporter 2 Is a Diagnostic and Therapeutic Target for Early-Stage Lung Adenocarcinoma. *Sci Transl Med* (2018) 10(467):eaat5933. doi: 10.1126/scitranslmed.aat5933
30. Leijenaar RT, Carvalho S, Velazquez ER, van Elmpt WJ, Parmar C, Hoekstra OS, et al. Stability of FDG-PET Radiomics Features: An Integrated Analysis of Test-Retest and Inter-Observer Variability. *Acta Oncol (Stockholm Sweden)* (2013) 52(7):1391–7. doi: 10.3109/0284186x.2013.812798
31. Li Y, Lu L, Xiao M, Dercle L, Huang Y, Zhang Z, et al. CT Slice Thickness and Convolution Kernel Affect Performance of a Radiomic Model for Predicting EGFR Status in Non-Small Cell Lung Cancer: A Preliminary Study. *Sci Rep* (2018) 8(1):17913. doi: 10.1038/s41598-018-36421-0
32. He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z. Effects of Contrast-Enhancement, Reconstruction Slice Thickness and Convolution Kernel on the Diagnostic Performance of Radiomics Signature in Solitary Pulmonary Nodule. *Sci Rep* (2016) 6:34921. doi: 10.1038/srep34921
33. Kim H, Park CM, Gwak J, Hwang EJ, Lee SY, Jung J, et al. Effect of CT Reconstruction Algorithm on the Diagnostic Performance of Radiomics Models: A Task-Based Approach for Pulmonary Subsolid Nodules. *AJR Am J Roentgenol* (2019) 212(3):505–12. doi: 10.2214/ajr.18.20018
34. Galloway MM. Texture Analysis Using Gray Level Run Lengths. *Comput Graphics Image Process* (1975) 4(2):172–9. doi: 10.1016/S0146-664X(75)80008-6
35. Mayerhoefer ME, Materka A, Langs G, Häggström I, Szczypiński P, Gibbs P, et al. Introduction to Radiomics. *J Nucl Med* (2020) 61(4):488–95. doi: 10.2967/jnumed.118.222893
36. Koh YW, Park SY, Hyun SH, Lee SJ. Associations Between PET Textural Features and GLUT1 Expression, and the Prognostic Significance of Textural Features in Lung Adenocarcinoma. *Anticancer Res* (2018) 38(2):1067–71. doi: 10.21873/anticanres.12324
37. Hyun SH, Ahn MS, Koh YW, Lee SJ. A Machine-Learning Approach Using PET-Based Radiomics to Predict the Histological Subtypes of Lung Cancer. *Clin Nucl Med* (2019) 44(12):956–60. doi: 10.1097/rlu.0000000000002810
38. Bianconi F, Palumbo I, Fravolini ML, Chiari R, Minestrini M, Brunese L, et al. Texture Analysis on [(18)F] FDG PET/CT in Non-Small-Cell Lung Cancer: Correlations Between PET Features, CT Features, and Histological Types. *Mol Imaging Biol* (2019) 21(6):1200–9. doi: 10.1007/s11307-019-01336-3
39. Zhang Y, Qiang JW, Shen Y, Ye JD, Zhang J, Zhu L. Using Air Bronchograms on Multi-Detector CT to Predict the Invasiveness of Small Lung Adenocarcinoma. *Eur J Radiol* (2016) 85(3):571–7. doi: 10.1016/j.ejrad.2015.12.014

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Niu, Gao, Shao, Wang, Jiang, Shi, Zhang, Wang and Shao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 17 December 2021
doi: 10.3389/fonc.2021.737368

A Comparative Study of Radiomics and Deep-Learning Based Methods for Pulmonary Nodule Malignancy Prediction in Low Dose CT Images

Mehdi Astaraki[1,2*], *Guang Yang*[3,4], *Yousuf Zakko*[5], *Iuliana Toma-Dasu*[2,6], *Örjan Smedby*[1] *and Chunliang Wang*[1]

[1] Department of Biomedical Engineering and Health Systems, KTH Royal Institute of Technology, Huddinge, Sweden, [2] Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden, [3] Cardiovascular Research Centre, Royal Brompton Hospital, London, United Kingdom, [4] National Heart and Lung Institute, Imperial College London, London, United Kingdom, [5] Imaging and Function, Radiology Department, Karolinska University Hospital, Solna, Stockholm, Sweden, [6] Department of Physics, Stockholm University, Stockholm, Sweden

OPEN ACCESS

Edited by:
Xujiong Ye,
University of Lincoln, United Kingdom

Reviewed by:
Yuming Jiang,
Stanford University, United States
Shouliang Qi,
Northeastern University, China

***Correspondence:**
Mehdi Astaraki
mehast@kth.se

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *06 July 2021*
Accepted: *29 November 2021*
Published: *17 December 2021*

Citation:
Astaraki M, Yang G, Zakko Y, Toma-Dasu I, Smedby Ö and Wang C (2021) A Comparative Study of Radiomics and Deep-Learning Based Methods for Pulmonary Nodule Malignancy Prediction in Low Dose CT Images. Front. Oncol. 11:737368. doi: 10.3389/fonc.2021.737368

Objectives: Both radiomics and deep learning methods have shown great promise in predicting lesion malignancy in various image-based oncology studies. However, it is still unclear which method to choose for a specific clinical problem given the access to the same amount of training data. In this study, we try to compare the performance of a series of carefully selected conventional radiomics methods, end-to-end deep learning models, and deep-feature based radiomics pipelines for pulmonary nodule malignancy prediction on an open database that consists of 1297 manually delineated lung nodules.

Methods: Conventional radiomics analysis was conducted by extracting standard handcrafted features from target nodule images. Several end-to-end deep classifier networks, including VGG, ResNet, DenseNet, and EfficientNet were employed to identify lung nodule malignancy as well. In addition to the baseline implementations, we also investigated the importance of feature selection and class balancing, as well as separating the features learned in the nodule target region and the background/context region. By pooling the radiomics and deep features together in a hybrid feature set, we investigated the compatibility of these two sets with respect to malignancy prediction.

Results: The best baseline conventional radiomics model, deep learning model, and deep-feature based radiomics model achieved AUROC values (mean ± standard deviations) of 0.792 ± 0.025, 0.801 ± 0.018, and 0.817 ± 0.032, respectively through 5-fold cross-validation analyses. However, after trying out several optimization techniques, such as feature selection and data balancing, as well as adding context features, the corresponding best radiomics, end-to-end deep learning, and deep-feature based models achieved AUROC values of 0.921 ± 0.010, 0.824 ± 0.021, and 0.936 ± 0.011, respectively. We achieved the best prediction accuracy from the hybrid feature set (AUROC: 0.938 ± 0.010).

Conclusion: The end-to-end deep-learning model outperforms conventional radiomics out of the box without much fine-tuning. On the other hand, fine-tuning the models lead to significant improvements in the prediction performance where the conventional and deep-feature based radiomics models achieved comparable results. The hybrid radiomics method seems to be the most promising model for lung nodule malignancy prediction in this comparative study.

Keywords: lung nodule, benign-malignant classification, lung cancer prediction, radiomics, deep classifier

INTRODUCTION

Lung cancer is notoriously aggressive and accounts for the leading cause of cancer-related death worldwide (1). Early diagnosis of asymptomatic lung cancer plays a vital role in treatment planning that can significantly improve the survival rate of lung cancer patients (2). The National Lung Screening Trial (NLST), a large-scale trial involving more than 50000 individuals, has reported that screening with Low Dose Computed Tomography (LDCT) scans will result in a 20% of reduction in lung cancer mortalities (3). Most lung cancers emerge from small malignant pulmonary nodules that refer to moderately well-marginated round opacities with the largest diameter less than 3cm (4). Although most solitary pulmonary nodules have benign causes, 30%–40% of such nodules are malignant (5). In clinical practice, expert radiologists visually examine the CT volumes on a slice-by-slice basis and subjectively determine the likelihood of nodule malignancy that often yields to relatively high inter/intra-observer variability of the interpretations. Moreover, highly similar visual characteristics shared among benign and malignant pulmonary nodules make this manual assessment task even more challenging (see **Figure 1**). Therefore, it would be beneficial to develop Computer-Aided Diagnosis (CAD) tools to capture latent characteristics of the pulmonary nodules in order to assist the radiologist with the task of benign-malignant lung nodule classification.

In the context of CAD tools, many semi/fully automatic pulmonary nodule classification methods have been proposed in the literature. Among all the recent proposed solutions, radiomics analysis (6–10), and deep learning-based methods (11–15) render the most promising results. Strictly speaking, radiomics analysis aims at building predictive models based on extracting handcrafted features from lung nodules. These radiological image-based features are designed to quantify the latent characteristics of the medical images that are imperceptible to human eyes. On the other hand, deep learning approaches such as Convolutional Neural Networks (CNNs) are trained with an end-to-end scheme to automatically translate the input images into corresponding class labels by adaptively learning deep abstract features in the consecutive convolutional layers.

It has been shown that factors such as nodules' size and heterogeneities in the intensity and textures of nodules are strongly associated with nodule malignancy (16). Radiomic features are often designed to capture such critical features from the nodule structures. The extracted radiomic features from lung nodules were employed to train learning algorithms such as logistic regression (17), linear discriminant analysis (18), random forests (19), and support vector machines (20) for malignancy identification. It has been shown that radiomic textural features are able to quantify the intra-tumor heterogeneities that appeared in CT volumes (21). In this context, several studies have investigated the ability of the textural features such as Gray Level Co-occurrence Matrix (GLCM), Gray Level Run Length Matrix (GLRLM), and Gray Level Zone Length Matrix (GLZLM) to distinguish malignant lung nodules from benign ones (17, 18, 20, 22). In addition, shape-based features were employed to quantify the morphological characteristics of lung nodules with irregular appearance (23, 24). Although each of the radiomics family can capture specific characteristics, their combination could cover different nodule attributes, and several promising results for lung cancer prediction have been reported (25–29).

Gaining from large-scale training image data, CNN models provide a uniform framework for jointly learning the hierarchical representative features extracted directly from the images and classification weights (30). Numerous 2D and 3D CNN networks have been developed for lung nodule classification tasks which were trained with either cropped volumetric patches or 2D slices extracted from multiple views (15, 31, 32). To conquer the challenges of small-scale CT images and the small size of the lung nodules, an Agile-CNN model was proposed based on a hybrid setting of conventional AlexNet and LeNet networks and achieved competitive classification performance (33). To further improve the classification power, recent methods rely on ensemble learning in which multiple different deep learning models are developed, and their outcomes are integrated into a single classification model (34). In this context, Xu et al. (14) employed three shallow 3D networks trained with multi-scale cropped CT volumes to preserve contextual information. By further modifying the training procedure and objective function, they achieved a malignancy prediction score of 0.94 in terms of Area Under the Receiver Operating Characteristic Curve (AUROC) in an unbalanced dataset. To make the learned deep features interpretable, Lei et al. (15) developed a Soft Activation Mapping (SAM) to enable the analysis of fine-grained lung nodule features with a CNN model and then combined the high-level deep features with SAM to improve the classification accuracy to 0.99. Xie et al. (35) developed a semi-supervised adversarial classification model that consists of an unsupervised

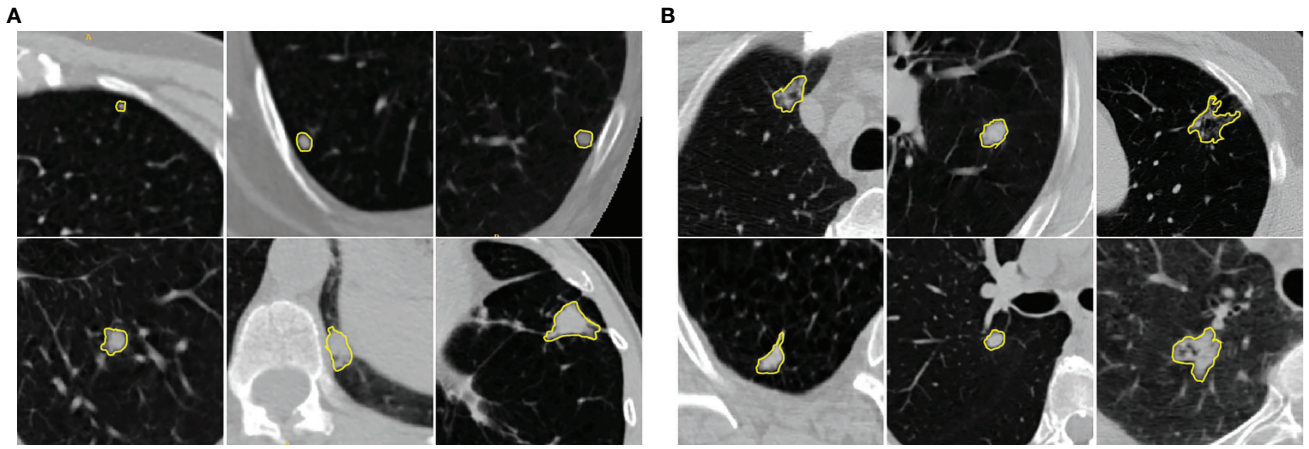


FIGURE 1 | Illustration of **(A)** benign and **(B)** malignant pulmonary nodules in chest LDCT scans. The manually identified nodules are highlighted with yellow contours. The examples show that benign and malignant pulmonary nodules present similar visual characteristics. The cropped patches around the nodules (context images) provide the relative location of the nodules with respect to nearby structures.

adversarial autoencoder network, a supervised classification network, and learnable transition layers to integrate both labeled and unlabeled CT volumes and achieved a classification accuracy of 0.92.

Both deep learning and radiomics have shown great potential to identify lung nodule malignancy in CT volumes and resulted in comparable performance in different datasets. However, there is a lack of conclusive evidence to support one type of method being better than the other. Most of the previous studies were done on relatively small datasets, which may not be considered a fair comparison as it is known that deep-learning based methods require a greater number of training images to achieve optimal performance. Also, many of those studies were performed on private datasets, which does not allow validation by other groups. This study aims to present an objective comparison among a series of carefully selected conventional radiomics methods, end-to-end deep learning models, and deep-feature based radiomics pipelines for pulmonary nodule malignancy prediction on an open database that consists of 1297 manually delineated lung nodules (both source code and annotation labels will be made publicly available upon acceptation). In addition to the prediction models, we also investigated the complementary role of the context region of lung nodules. Finally, a hybrid model was developed by pooling the extracted radiomic features and learned deep features to assess their compatibility with respect to nodule malignancy prediction. Finally, optimization steps such as feature selection methods and balancing the class labels were applied on radiomics, deep features, and their combination to improve the discrimination power.

MATERIALS AND METHODS

The investigated models consist of three major components: a radiomics module, a deep learning module, and a hybrid module. The *radiomics module* incorporates the radiomics analysis starting with feature extraction, followed by feature engineering steps, and finally, building predictive models. The *deep learning module* consists of the development of a dual-pathway CNN model to predict nodule malignancy by simultaneously training nodule target and nodule context images. The deep learning module was investigated thoroughly in our previous research (36) on the same dataset; therefore, in the current study, we adopt the results of already examined models. Finally, the *hybrid module* represents the approach of pooling the radiomic features and deep features together.

Experimental Data

The Kaggle Data Science Bowl 2017 (37) contains a total number of 2101 clinical chest LDCT scans from which 1397, 198, and 506 subjects belong to the training, validation, and test sets, respectively. The objective of this challenge was to automatically predict lung cancer status; for that, each image was labeled as "1" if the patient was diagnosed with lung cancer within one year from the scan and "0" otherwise. The challenge organizer provided only the target labels of the training set at the patient level, and the validation/test labels are not accessible anymore as the challenge platform is closed. It should be noted that additional information, such as nodule segmentation masks, associated clinical data, and laboratory examinations, was not supplied.

In this study, out of 1397 training scans, 968 LDCT volumes were manually inspected by an expert radiologist, which led to the delineation of 1297 pulmonary nodules for further analyses. In addition to the segmentation masks, image patches that best cover each nodule context were cropped and extracted. For each nodule, visual radiological features including "cavitation with thick/thin wall", "attached to the artery/fissure/pleura", "calcification or fat content", "dragging the pleura", "spiculation" as well as "size >3cm" were extracted. **Figure 1** in

Supplementary Material indicates the distribution of the studied nodules.

Image Pre-Processing

Prior to radiomic feature extraction, all the cropped patches were preprocessed in three steps. First, original patches were resampled isotropically to a unified inner plane spacing as $0.2mm^3$ using a bicubic interpolation function. Then the intensity ranges were clamped to [-1000,500] in terms of Hounsfield Units. A further step was applied only for deep learning module analyses in which the image patch was rescaled by zero-padding the original sizes into 128×128×128 voxels followed by intensity normalization in the range of [0,1].

Radiomics Analysis

We adopted the radiomics descriptors to quantify the geometric, intensity, and textural characteristics of the nodules by extracting a total number of 1334 3D descriptors using the standard PyRadiomics package (38). Specifically, 14 geometric, 18 First-Order Statistics (FOS), and 70 Second-Order Statistics (SOS) features were extracted from the target nodules. In addition, FOS and SOS features were also extracted from multi-scale transformed nodules filtered with Wavelet and Laplacian of Gaussians (38) (**Table 1** in **Supplementary Material**). The mentioned features were extracted from both target nodules and context images to capture intra-nodule characteristics as well as nodule context attributes (**Figure 2**- Radiomics Module).

As irrelevant or partially relevant features can adversely impact the classification performance, and in order to minimize the risk of overfitting, filter-based Feature Selection (FS) methods, as well as a wrapper FS method, were employed. In particular, filter-based FS algorithms include (1) Constant (CST): removing constant features; (2) Correlation (Corr): removing linearly related features; (3) Mutual Information (MI): removing nonlinearly related features; (4) RELevance in Estimating Features (RELIEF): estimating the quality of the features based on how well the features can distinguish the subjects that are close to each other; and (5) Least Absolute Shrinkage and Selection Operator (LASSO): applying coefficients to the features and shrink to zero those features which are less predictive (39). On the other hand, Forward Feature Selection (FFS) as a wrapper FS method was adopted to evaluate the performance of the learning algorithms in different combinations of feature subsets. In addition, Principal Component Analysis (PCA) was employed as well to transform the high-dimensional feature set into a lower dimension (26). The prediction power of the selected features was then evaluated with 8 different learning algorithms: Adaptive Boosting (Adab), Decision Tree (DT), Random Forest (RF), K-Nearest Neighbor (KNN), Support Vector Machines (SVM), Linear/Quadratic Discriminant Analysis (LDA/QDA), as well as Naïve Bayesian (**Table 1**).

End-to-End Deep Learning Model

In our previous study on the same dataset (36), we proposed a dual pathway network architecture to train both the nodule target images and nodule context images simultaneously in a unified network. This unified network consists of two convolutional pathways for representation learning, each followed by a few dense layers and a shared final dense layer. In other words, while the nodule target pathway is assumed to mainly learn the association between the intra-nodule representations and class labels, the role of the nodule context pathway is primarily to learn the correlations between the context information and the class labels. Therefore, having concatenated the learned features from each of the pathways in a last shared dense layer, the model is enforced to predict the class labels by adaptively learning the intra- and context-nodules attributes simultaneously. Different supervised models, including VGG (40), ResNet (41), DenseNet (42), EfficientNet (43), and a variational autoencoder (44) as an unsupervised model, were employed for the convolutional backbones. Moreover, the conventional single pathway models were trained with either nodule target or nodule context images separately to extract features from the corresponding regions. More details of the studied method can be found in (36).

Deep Feature Extraction

In addition to end-to-end training of the deep networks, the learned deep features from each network were extracted to train a separate learning algorithm. In specific, for the dual pathway models, 2048 deep features were extracted from the last but one dense layer representing the target and context nodule attributes. Moreover, with the same approach,1024 deep features were extracted from the single pathway networks for each of the nodule target and nodule context images. The extracted deep features were then analyzed by different learning algorithms. Experimentally, random forest was selected as the learning algorithm to train the extracted deep features after feature augmentation, as it consistently led to more stable results than the other classifiers (36).

Nodule Classification

Radiomic features and deep features were used separately to distinguish the benign nodules from malignant ones. In addition, the same task was done by combining the extracted deep features with radiomic features in a hybrid model. In particular, the extracted radiomic features from target nodule images were analyzed both independently and in combination with deep features learned from the target nodule images. Similar analyses were conducted by combining the features extracted and learned from the context nodule images. Finally, to examine how context and target nodule images would complement each other, the prediction power of a combination of target nodule features and context nodule features was investigated as well.

From 1297 studied lung nodules, 876 cases belong to the benign group, and 421 nodules are labeled as malignant. Such an unequal distribution of the class labels leads to a high bias toward the majority class and, in turn, degrades the prediction power of the learning algorithms and will result in poor prediction of the minority class. To tackle this issue, we employed Synthetic Minority Oversampling Technique (SMOTE) (45) to synthesize new samples from the minority class. In specific, SMOTE fits a hypercube among some instances in the feature space of malignant nodules to interpolate new samples.

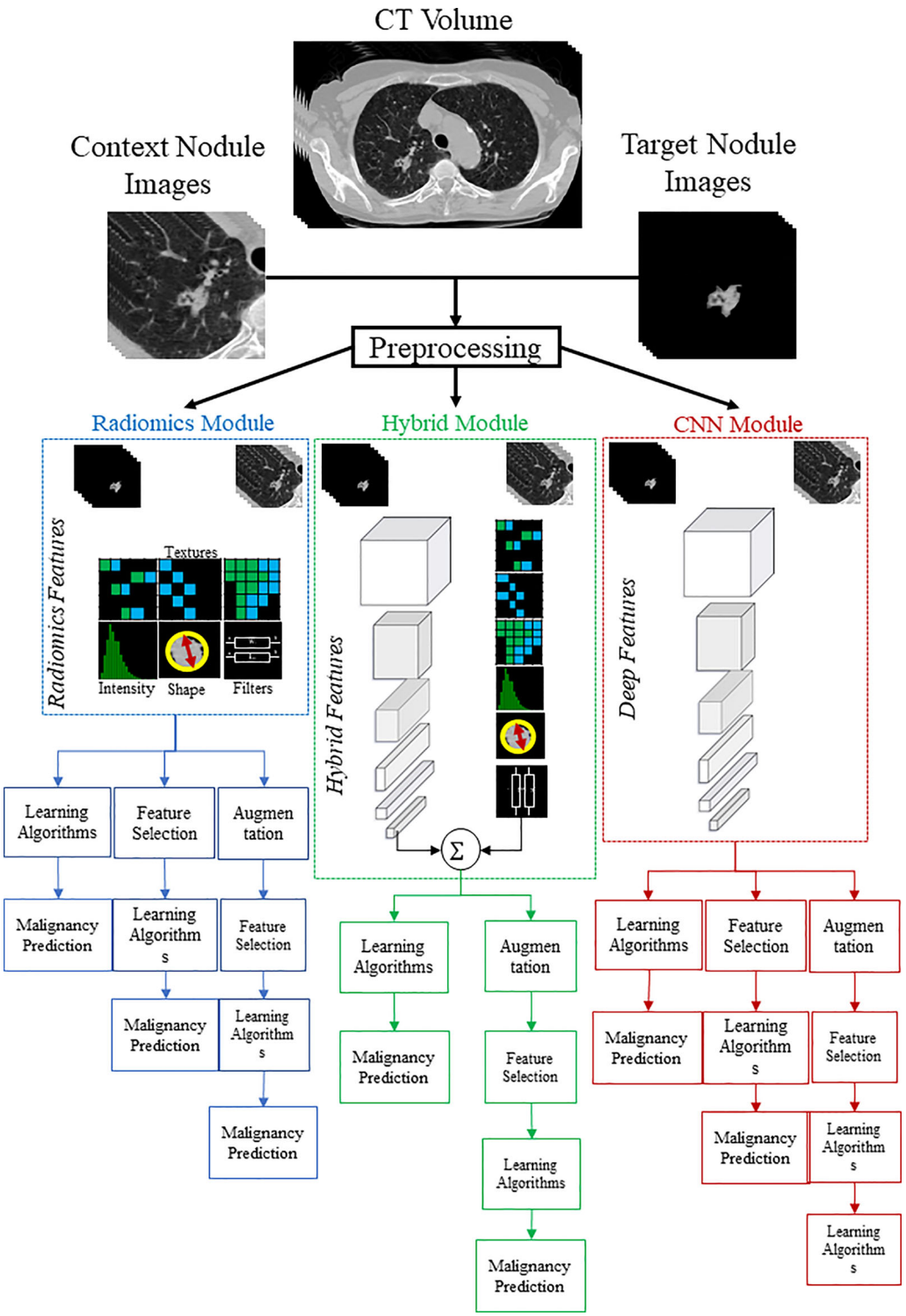


FIGURE 2 | Graphical demonstration of the study pipeline. To predict lung nodule malignancy, three modules were studied. In the CNN module (red color), deep networks were trained with context and target nodule images separately and simultaneously. In the Radiomics module (blue color), handcrafted features were extracted from both target and context nodule images to train the learning algorithms. In the hybrid module (green color), extracted radiomic features were combined with learned deep features to form the hybrid sets.

TABLE 1 | The prediction power of the radiomic features extracted from target nodule images with different learning algorithms and feature selection methods over the balanced dataset.

Learning Algorithm	Target Radiomic Prediction Performance (AUROC)							
	Feature Selection							
	None	***CST***	***Corr***	***LASSO***	***RELIEF***	***MI***	***PCA***	***FFS***
Adab	**0.889 ± 0.016**	**0.863 ± 0.021**	**0.881 ± 0.031**	**0.864 ± 0.026**	0.671 ± 0.019	0.531 ± 0.037	**0.868 ± 0.015**	**0.911 ± 0.016**
DT	0.723 ± 0.011	0.733 ± 0.027	0.703 ± 0.019	0.711 ± 0.028	0.642 ± 0.013	0.523 ± 0.024	0.712 ± 0.026	0.730 ± 0.032
RF	0.871 ± 0.008	0.849 ± 0.025	0.846 ± 0.028	0.856 ± 0.023	**0.765 ± 0.018**	0.517 ± 0.031	0.862 ± 0.026	0.891 ± 0.011
KNN	0.850 ± 0.016	0.846 ± 0.016	0.807 ± 0.036	0.833 ± 0.017	0.735 ± 0.021	0.671 ± 0.089	0.846 ± 0.017	0.870 ± 0.023
SVM	0.777 ± 0.011	0.774 ± 0.029	0.752 ± 0.025	0.775 ± 0.027	0.751 ± 0.020	0.522 ± 0.040	0.775 ± 0.028	0.802 ± 0.008
LDA	0.655 ± 0.045	0.680 ± 0.032	0.785 ± 0.017	0.75 ± 0.027	0.741 ± 0.031	0.735 ± 0.018	0.771 ± 0.028	0.796 ± 0.011
QDA	0.778 ± 0.172	0.696 ± 0.181	0.747 ± 0.016	0.738 ± 0.024	0.753 ± 0.031	**0.840 ± 0.020**	0.752 ± 0.026	0.865 ± 0.006
Naive	0.763 ± 0.006	0.759 ± 0.023	0.742 ± 0.030	0.731 ± 0.022	0.756 ± 0.034	0.583 ± 0.046	0.739 ± 0.024	0.808 ± 0.010

For each feature selection algorithm, the highest value is marked in bold.

Balancing the dataset with SMOTE yields the generation of 455 new instances belong to the malignant nodule class, which increased the total number of studied data to 1752. All the analyses were performed with a 5-fold cross-validation fashion, and the performance of the model was assessed by using the Area Under the Receiver Operating Characteristic Curve (AUROC) metric calculated as averages of corresponding cross-validation folds. Finally, to test statistical significances between different experiments, a pairwise AUROC comparison method proposed by Delong et al. (46) was employed. **Figure 2** shows a graphical illustration to visualize the feature analysis workflow for all the examined modules.

RESULTS

In this section, the performance of the radiomic features for lung nodule malignancy prediction is presented and compared against the prediction power of deep learning models. In addition, the classification power of the hybrid feature pools is quantified as well.

Handcrafted Radiomics

Radiomics analysis was performed by using 1334 3D quantitative descriptors with and without feature selection methods over 8 distinct learning algorithms. The analyses were conducted over the original imbalanced dataset and augmented, balanced set as well. Comparing the results between the balanced (**Tables 1**–**3**) and imbalanced datasets (**Tables 2–4** in **Supplementary Material**), one can observe that synthesizing new samples in the feature space from the minority class resulted in a remarkable improvement in classification performance. For instance, the prediction power of the Adab learning algorithm without any FS method before and after augmenting the data are $AUROC_{unbalanced}$ = 0.779 and $AUROC_{balanced}$ = 0.889, respectively.

In addition, comparing **Tables 1**, **2** reveals the fact that target nodule and context nodule images represent different nodule characteristics that carry distinct prediction powers. However, as can be seen in **Table 3**, combining the target nodule features with context features improved the accuracy and yielded the highest classification power in radiomics analyses. Among the employed learning algorithms, Adab method embedded on decision trees yielded the highest prediction power when feature selection methods were not applied. Comparing the performance of the models after applying different feature selection methods, one can infer that FFS method consistently improved the prediction power and outperformed other feature selections except the two cases of KNN and QDA. Accordingly, the highest predictive value of the target radiomics set was achieved from a subset of features selected with FFS and trained with Adab classifier

TABLE 2 | The prediction power of the radiomic features extracted from context nodule images with different learning algorithms and feature selection methods over the balanced dataset.

Learning Algorithm	Context Radiomic Prediction Performance (AUROC)							
	Feature Selection							
	None	***CST***	***Corr***	***LASSO***	***RELIEF***	***MI***	***PCA***	***FFS***
Adab	**0.895 ± 0.007**	**0.867 ± 0.022**	**0.866 ± 0.020**	**0.871 ± 0.009**	0.580 ± 0.017	0.643 ± 0.078	**0.852 ± 0.009**	**0.916 ± 0.011**
DT	0.718 ± 0.011	0.697 ± 0.025	0.695 ± 0.015	0.702 ± 0.032	0.571 ± 0.021	0.550 ± 0.039	0.697 ± 0.031	0.744 ± 0.027
RF	0.881 ± 0.008	0.843 ± 0.024	0.855 ± 0.009	0.864 ± 0.011	0.645 ± 0.025	0.613 ± 0.044	0.845 ± 0.007	0.901 ± 0.014
KNN	0.852 ± 0.007	0.824 ± 0.019	0.811 ± 0.010	0.843 ± 0.021	0.625 ± 0.015	0.590 ± 0.023	0.827 ± 0.019	0.779 ± 0.029
SVM	0.777 ± 0.012	0.757 ± 0.010	0.689 ± 0.010	0.716 ± 0.008	0.685 ± 0.012	0.571 ± 0.020	0.715 ± 0.009	0.817 ± 0.023
LDA	0.682 ± 0.040	0.727 ± 0.018	0.774 ± 0.014	0.758 ± 0.017	0.743 ± 0.017	0.746 ± 0.022	0.751 ± 0.016	0.842 ± 0.027
QDA	0.841 ± 0.013	0.705 ± 0.033	0.777 ± 0.032	0.751 ± 0.025	**0.770 ± 0.012**	**0.863 ± 0.067**	0.739 ± 0.024	0.872 ± 0.010
Naive	0.767 ± 0.014	0.690 ± 0.006	0.757 ± 0.029	0.745 ± 0.009	0.682 ± 0.010	0.609 ± 0.038	0.728 ± 0.020	0.820 ± 0.013

For each feature selection algorithm, the highest value is marked in bold.

TABLE 3 | The prediction power of the joint context and target radiomic with different learning algorithms and feature selection methods over the balanced dataset.

Learning Algorithm	Combined Radiomic Prediction Performance (AUROC)							
	Feature Selection							
	None	***CST***	***Corr***	***LASSO***	***RELIEF***	***MI***	***PCA***	***FFS***
Adab	**0.908 ± 0.014**	**0.883 ± 0.021**	**0.883 ± 0.016**	**0.888 ± 0.014**	0.570 ± 0.043	0.676 ± 0.005	0.876 ± 0.014	**0.921 ± 0.010**
DT	0.739 ± 0.032	0.699 ± 0.018	0.728 ± 0.011	0.720 ± 0.014	0.568 ± 0.030	0.594 ± 0.020	0.702 ± 0.016	0.772 ± 0.013
RF	0.897 ± 0.016	0.858 ± 0.032	0.877 ± 0.019	0.865 ± 0.028	0.620 ± 0.045	0.621 ± 0.021	**0.880 ± 0.011**	0.910 ± 0.008
KNN	0.872 ± 0.014	0.844 ± 0.014	0.804 ± 0.013	0.860 ± 0.007	0.650 ± 0.018	0.604 ± 0.029	0.848 ± 0.012	0.816 ± 0.028
SVM	0.756 ± 0.023	0.711 ± 0.019	0.709 ± 0.035	0.722 ± 0.022	0.625 ± 0.038	0.574 ± 0.574	0.724 ± 0.022	0.818 ± 0.022
LDA	0.711 ± 0.027	0.726 ± 0.012	0.802 ± 0.023	0.759 ± 0.011	0.730 ± 0.019	0.647 ± 0.032	0.768 ± 0.012	0.827 ± 0.020
QDA	0.862 ± 0.014	0.711 ± 0.015	0.827 ± 0.028	0.766 ± 0.020	**0.736 ± 0.017**	**0.903 ± 0.007**	0.744 ± 0.010	0.887 ± 0.015
Naive	0.783 ± 0.023	0.702 ± 0.020	0.741 ± 0.019	0.731 ± 0.024	0.628 ± 0.020	0.544 ± 0.013	0.736 ± 0.021	0.825 ± 0.021

For each feature selection algorithm, the highest value is marked in bold.

(AUROC = 0.911), which accounts for an improvement of 2.2 percent from the same classifier but without feature selection. Similar behavior was observed in context radiomics where integrating FFS into the Adab classifier improved the discrimination power from 0.895 to 0.916. The classification accuracy was even more improved from 0.908 to 0.921 when the combination of context and target radiomics were analyzed with the same method.

End-to-End Deep Learning Models

Table 4 shows a summary of the best results achieved by the deep learning-based analyses, which were published in our previous study (36). The first column represents the prediction power achieved by end-to-end training of the networks on the imbalanced datasets. In other words, for each of the target nodule, context nodule, and their combined (dual-pathway) images, 5 different networks were trained, and the best performance for each nodule image type is reported in **Table 4**. Comparing the performance of the end-to-end deep networks trained with the target and context nodule images, one can infer that context features were more informative than target nodules. Furthermore, similar to radiomics, integrating the target and context nodule images into a unified network resulted in a slightly higher prediction power. In fact, the AUROC values of dual-pathway models outperformed each of the single pathway models trained with context and target images separately. These enhancements imply that the two distinct image types can become complementary. In other words, the combined features achieved the highest AUROC value by virtue of the joint use of context and target deep features that can adequately complement the intricate characterizations of shape, intensity, and textural heterogeneity of the nodules.

Deep Features

The last column of **Table 4** indicates the results of the deep feature-based radiomics pipeline. Strictly speaking, after training the networks, the imbalanced learned deep features were extracted from one to last dense layer of each model and augmented after applying the SMOTE algorithm. The balanced deep features were then used to train an RF model. More details over the deep feature analyses and the effects of employed different optimization techniques on the prediction powers of deep features were examined and reported in our previous study (36).

Comparing the first column of **Table 4** with **Tables 2–4** in **Supplementary Material** indicates that end-to-end training the deep classifiers with the unbalanced dataset could predict the class labels of the lung nodules more accurately than radiomics analyses over the imbalanced feature set without applying the FS methods. For instance, while integrating the imbalanced target and context radiomic features led to achieving an AUROC of 0.773 with the RF learning algorithm, end-to-end training a dual-pathway model resulted in an AUROC value of 0.824 that accounts for 5.1% of improvement. Besides, significant improvements were achieved when the extracted deep features were employed in a radiomics-based pipeline. Specifically, applying the SMOTE algorithm to the unbalanced extracted deep features could successfully boost the discrimination power of the RF model by up to 12.1% ($AUROC_{context\text{-}imbalanced} = 0.806$ vs. $AUROC_{context\text{-}balanced} = 0.927$). In addition, balancing the deep features seems to be more constructive than balancing the radiomic features in terms of discrimination power. In particular, AUROC values achieved by the RF model on balanced deep features outperformed the performance of the RF model trained with balanced radiomic features with rather

TABLE 4 | The prediction power of the deep learning-based analyses.

Feature Type	Deep Features Prediction Performance (AUROC)	
	End-to-end training	***Feature augmentation***
Target nodule	0.801 [0.777,0.824]	0.906 [0.890,0.921]
Context nodule	0.806 [0.788,0.827]	0.927 [0.912,0.940]
Combined	0.824 [0.798,0.837]	0.936 [0.921,0.950]

The combined model refers to a dual-pathway network that was fed by context and target nodule images simultaneously. Lower and upper limits of confidence interval at 95% level are indicated in square brackets.

significant margins. ($AUROC_{target}$: 0.906 *vs.* 0.891, $AUROC_{context}$: 0.927 *vs.* 0.901; and $AUROC_{combined}$: 0.936 *vs.* 0.910).

Hybrid Feature Analysis

To investigate whether the learned deep features would complement radiomics descriptors, a hybrid model was developed by pooling the radiomics and deep features into a mixture set. Accordingly, for each of the target, context, and combined radiomics pools, corresponding learned features from single-pathways and dual pathway deep networks were merged for both the raw features and augmented features (see **Figure 3**). It should be noted that assessing the capability of hybrid feature pools was performed by employing the Adab learning algorithm integrated with the FFS method because the highest prediction performance of the radiomics was achieved by this method. The results show that incorporating the imbalanced deep features into imbalanced radiomics could slightly improve the performance of imbalanced radiomics alone by up to 1.9% ($AUROC_{radiomics_combined}$: 0.774 *vs.* $AUROC_{hybrid_combined}$: 0.793 *vs.* 0.891). However, this combination was not led to enriching the performance of the end-to-end deep learning methods. On the other hand, noticeable improvements were observed in the performance of the hybrid models after augmenting both the radiomics and deep features. In particular, merging the balanced deep features with balanced radiomics not only successfully improved the prediction power of radiomics alone but also slightly enhanced the performance of deep features as well. These improvements were observed for both target and context nodule images. In addition, the balanced hybrid feature pools of combined target and context features resulted in an AUROC value of 0.929, which is superior to the combined radiomic features alone with an AUROC value of 0.921. Besides, even with the hybrid model, context pools were more representative than the target pools, which is in line with both the radiomics and deep feature pools individually. Last but not least, the best prediction accuracy (AUROC=0.938) among all the analyses was achieved by a hybrid model in which the balanced context radiomics merged with the balanced deep features learned from context images which point to the complementary role of context radiomics and context deep features. **Table 8** in **Supplementary Material** indicates the pairwise statistical comparison between radiomic features and deep features. Finally, **Table 9** in **Supplementary Material** shows the same statistical assessment between target, context, and combined feature sets in the hybrid module.

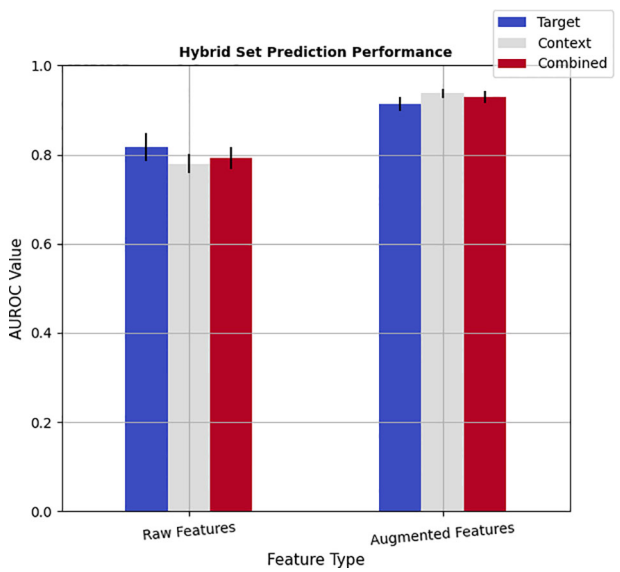


FIGURE 3 | The prediction power of the hybrid model: combinations of deep and radiomic features.

Feature Importance and Class Separability

Figure 2 in **Supplementary Material** demonstrates the top 20 important features that contributed to explaining the class labels when training the RF classifier for different sets of radiomics, and hybrid feature sets. As can be seen, textural features cast the majority of the important features in radiomics analysis in both the target and context nodules. However, the contribution of shape-based features in context radiomics indicates the fact that shape descriptors could be involved in quantifying the context information of the regions covered the nodules. The same figure for the hybrid sets illustrates that most of the important attributes in the hybrid target set derived from radiomics features, while deep features contributed more in the hybrid context set. Moreover, the most informative radiomic features identified from FS methods are reported in **Table 5** in the **Supplementary Material**.

Figure 3 in **Supplementary Material** shows the scatter plots of the class separability of different feature pools. In practice, T-distributed Stochastic Neighbor Embedding (TSNE) (47) statistical method was employed to nonlinearly reduce the dimensionality of the feature space and visualize the distribution of the data points based on their similarity in a 2D space. Because the dimension of the feature sets was larger than 1000, we first applied the PCA to project the feature sets into a 70D space and then apply the TSNE method. Interestingly, the class separability of deep features is more obvious than radiomics. These scatter plots are consistent with the quantitative results reported in **Tables 1**, **2**, **4**, and **Figure 3**.

Finally, to study the impact of the size of training samples on the prediction performance, the feature pools were split randomly into training and test subsets with varying proportions. In specific, the learning algorithm was trained with 25%, 50%, and 70% of the feature pools as training samples, while the rest of 75%, 50%, and 30% were dedicated to test sets. The discrimination scores were then calculated on the test set by applying a 5-fold cross-validation scheme. **Figure 4** shows the results of feature fractioning achieved from the hybrid feature pools, and **Tables 6, 7** in **Supplementary Materials** demonstrate the same evaluations calculated from radiomics and deep features separately. As was expected, increasing the size of the training samples leads to higher prediction powers. Besides, the discrimination powers of different fractions of deep features were higher than those of the radiomic features, while the performance on fractions of hybrid sets and deep features

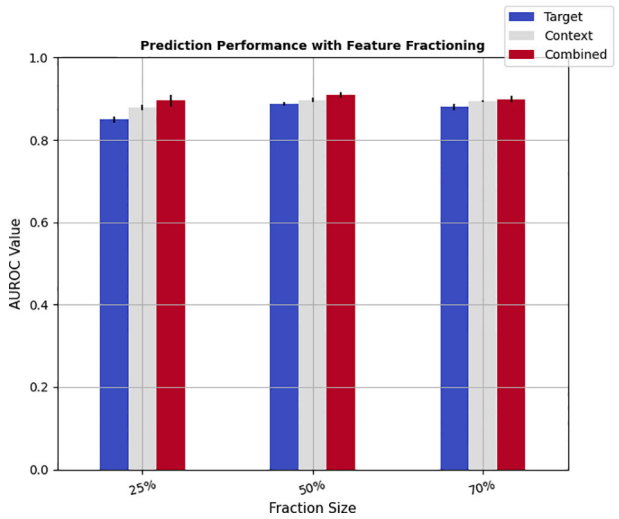


FIGURE 4 | Effect of training size on the prediction power of the hybrid feature sets.

were relatively close. Such results are in agreement with the prediction scores reported in **Figure 3**.

Conventional Radiomics vs. End-to-End Deep Learning Model vs. Deep-Feature-Based Radiomics

In this study, four different strategies, including conventional radiomics analysis, end-to-end deep learning classifiers, deep feature based radiomics as well as the combination of radiomics and deep features, were investigated to predict lung nodule malignancy. Moreover, the effect of fine-tuning steps such as feature augmentation and feature selection on the model performance were examined as well. To better compare the performance of each method, **Table 5** shows a summary of the discrimination powers. In this table, 'fine-tuning' refers to the feature augmentation step, and 'context-info' indicates whether the context nodule images were included or not. Accordingly, the best results achieved by analyzing the radiomic features are presented in 'best conventional radiomics' column, and the best performance achieved by end-to-end training of deep learning models are showed in 'best end-to-end deep learning model' column. In addition, the extracted deep features from the learned networks were analyzed with the radiomics pipeline, i.e., feature selection and feature augmentation steps were applied and followed by training the learning algorithms. The best performance of such analyses is summarized in 'best deep-feature-based radiomics' column. Finally, an overview of the prediction powers achieved by concatenating the radiomic features with the extracted deep features is shown in the last column of the table 'best hybrid model'.

DISCUSSION

The separation of benign from malignant pulmonary nodules on chest LDCT scans is an important step toward the early detection of lung cancers which in return offer the best chance for cure. In clinical practice, this vital step is done manually by expert radiologists on a slice by slice basis. However, the possibility of operator bias on one side, and the presence of highly similar visual characteristics shared between the benign and malignant nodules on the other side, can potentially lessen the accuracy of manual nodule classification. Therefore, many computer-aided models for automatic/semi-automatic classification of pulmonary nodules have been developed as assistant tools to facilitate such a demanding task. In a general view, these models can be categorized into two groups: handcrafted radiomics and end-to-end deep learning models. In this study, we conducted a comprehensive comparison of the performance of radiomics and deep learning models for lung nodule malignancy prediction on a relatively large-scale dataset consists of 1297 manually delineated lung nodules. In addition, we applied several optimization steps on both extracted radiomic features and learned deep features to improve the prediction performance. In this context, to reliably distinguish intra-nodule characteristics from nodule contextual attributes, both radiomics and deep features were extracted from target nodule images and context nodule images. Moreover, in order to efficiently capture the critical nodule characteristics such as shape, intensity, and textural heterogeneities, a hybrid feature set was constructed by pooling deep and radiomic features together.

The high correspondence between intra-nodule heterogeneity and malignancy alludes to the privilege of textural radiomic features. From 1334 radiomic descriptors, almost 73% of them represent textural features from the original and multi-scale filtered images. Interestingly, having performed the radiomics analysis with/without FS methods, textural features contributed

TABLE 5 | Comparing the prediction power of the employed methods.

Prediction Performance Comparison (AUROC)

Fine-tuning	*Context-info*	*Best conventional radiomics*	*Best end-to-end deep learning model*	*Best deep-feature-based radiomics*	*Best hybrid model*
Before	No	0.792 ± 0.025	0.801 [0.777,0.824]	0.753 [0.743,0.775]	0.817 ± 0.032
After	No	0.911 ± 0.016	–	0.906 [0.890,0.921]	0.914 ± 0.015
Before	Yes	0.777 ± 0.017	0.806 [0.788,0.827]	0.761 [0.736,0.779]	0.780 ± 0.022
After	Yes	0.916 ± 0.011	0.824 [0.798,0.837]	0.927 [0.912,0.940]	0.929 ± 0.013

Note that the "best end-to-end deep learning model" column presents the performance of two single pathway models trained with target and context nodule images separately and one dual pathway model trained with both target and context images simultaneously. Lower and upper limits of confidence interval at 95% level are indicated in square brackets.

significantly to classification results. In fact, a majority of the informative features without FS and a majority of the selected features with the FFS method come from textural families, which is in agreement with the reported results by other studies (30, 32, 48). Performing the filter-based FS methods on radiomics has not always yielded the improvement of the prediction power. In other words, statistically less correlated feature subset selected from Corr, LASSO, RELIEF, and MI methods were not necessarily informative with respect to the class labels and therefore lessened the prediction power in some cases. On the other hand, the greedy FFS method reduces the dimensionality of radiomic pools by selecting a combination of feature subsets with the highest prediction performance. In addition, as was expected, combining multiple weak classifiers into a single robust learning algorithm, the Adab method outperformed the rest of the classifiers in terms of prediction power (49). As the underlying training procedures of the RF classifier, to some extent, are close to that of Adab, it achieved the second strong prediction power. Moreover, 35% of the features selected from FFS are repeated between these two classifiers (**Table 5** in **Supplementary Material**). Furthermore, fractioning the radiomics training set to even 25% of all the features resulted in satisfactory prediction power ($AUC_{25\%}$= 0.827 vs. $AUC_{80\%}$= 0.889). While the target nodule images hold intra-nodule attributes such as intensity and textural heterogeneities, shape and morphological features, as well as size-based properties, the context nodule images include the characteristics of surrounding tissues as well. Such differences in the image contents led to slightly higher prediction performance for context images. These results are in agreement with clinical practices in which the expert radiologists not only focus on the visual attributes of the nodules but also closely examine the contextual information around the nodules by inspecting the same nodule in three different orthogonal views in several prior and succeeding slices. The important features which contributed to malignancy prediction in context nodule images consist of several shapes and FOS-based features instead of focusing only on textural radiomics, which implies the fact that the model could capture the changes in the geometry and intensity heterogeneities in the nodule images after including context information. In addition, the observed improvements in the prediction power of combined nodule features point to the complementary role of nodule target and nodule context images. Of course, it should be noticed that applying the FS methods could reduce the risk of overfitting that would occur after combining context and target nodule features. More importantly, employing the SMOTE as an augmentation technique could successfully hinder the classification bias toward the majority class by balancing the class labels, which in return improved the classification accuracy significantly in all the radiomics experiments. In general, our analyses show that radiomic features are capable of quantifying the differences between challenging benign and malignant pulmonary nodules and have a great potential to achieve promising results.

Similar to radiomics, augmented deep context features outperformed the augmented deep target features, and their combination improved the accuracy of malignancy prediction as well. Training end-to-end deep classifiers could perform slightly better than radiomics analyses for each of target nodule images, context nodule images, and their combination. This can be described by the fact that end-to-end training of the networks enforces the models to learn to extract the features with the highest correspondence with respect to the class labels. On the other hand, radiomic features are extracted regardless of the class labels so that they would need further carefully designed processing steps to maximize their discrimination power. Besides that, augmenting the deep features by synthesizing new data points in the feature space to balance the class labels effectuates a terrific improvement in prediction powers. It is noteworthy that employing the conventional strategies such as including an equal number of both class labels in each batch of images during the model training may be helpful to avoid biasing toward the majority class. However, they often would not be able to dramatically boost the prediction performance. In general, although the deep learning models outperformed the radiomics analyses when the mentioned feature selection and augmentation techniques were not applied, we can infer that exerting such fine-tuning steps results in comparable performance.

The classification improvement achieved from integrating deep features with radiomics points to the fact that the hybrid descriptors could successfully cover a wide range of nodule characteristics from quantitative textures to abstract shape features. Such observed improvements are in agreement with other studies performed on lung nodule classification tasks (30, 32). Although radiomic descriptors are designed inspired by radiological quantitative imaging features, CNN models are trained to capture abstract features with high relevancy with respect to the class labels. Nevertheless, fusing these two distinct feature set in order to enhance the prediction power should be conducted carefully. In other words, simply aggregating the two feature sets will dramatically increase the length of the features and potentially increase the risk of overfitting. Such larger feature sets can even exacerbate the problem of class imbalance. Additionally, the careless fusion of these feature sets would increase the number of correlated features, which in return would afflict the performance of the models. The inferior performance of the raw hybrid feature sets can be caused by the described reasons. However, fine-tuning the hybrid feature sets by balancing the classes and reduce the dimensionality of the features by employ the proper FS methods can be considered as beneficial strategies to efficiently gain from the hybrid feature pools that resulted in the highest prediction power.

In this comparative study, we examined the two most conventional approaches of radiomics and deep learning methods to predict pulmonary nodule malignancy in LDCT images. Despite the promising results achieved and compared in this paper, in our future studies, we aim to test the proposed approach on external datasets both as inference models and transfer learning strategies. In addition, we plan to investigate other fusion methods to integrate radiomic descriptors into end-to-end deep learning models efficiently.

CONCLUSION

We conducted a comparative study to distinguish benign pulmonary nodules from malignant nodules in LDCT images. To do so, we performed a comprehensive radiomics analysis by investigating the prediction power of 1334 extracted radiomic features trained with 8 learning algorithms integrated with 7 feature selection methods. We compared the radiomics performance against several deep classifiers trained on the same datasets. We examined the effect of optimization strategies such as synthesizing the feature points to balance the class labels, extracting features from context images, and combine context features with features extracted from target nodule images. Our results suggest that effective integration of radiomics and deep features improves the performance of nodule classification and produces more accurate results.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding author. The implementations can be found at https://github.com/Astarakee/Radiomics_pipeline.

AUTHOR CONTRIBUTIONS

MA: Conceptualization, Methodology, Formal analysis, Writing - original draft, Visualizations. CW: Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration. YZ: data annotation, Writing - review & editing. IT-D: Writing - review & editing, Supervision. GY: Writing - review & editing. ÖS: Writing - review & editing, Supervision, Resources, Funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported in part by the Swedish Childhood Cancer Foundation [grant no. MT2016-0016], in part by the Swedish innovation agency Vinnova [grant no. 2017-01247], and in part by the Swedish Research Council (VR) [grant no. 2018-04375]. Guang Yang was supported in part by the British Heart Foundation [Project Number: TG/18/5/34111, PG/16/78/32402], the European Research Council Innovative Medicines Initiative [DRAGON, H2020-JTI-IMI2 101005122], the AI for Health Imaging Award [CHAIMELEON, H2020-SC1-FA-DTS-2019-1 952172], and the UK Research and Innovation Future Leaders Fellowship [MR/V023799/1].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.737368/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2019. *CA: A Cancer J Clin* (2019) 69:7–34. doi: 10.3322/caac.21551
2. Wu GX, Raz DJ. Lung Cancer Screening. *Lung Cancer*. In: Cham: Springer. (2016). p. 1–23. doi: 10.1007/978-3-319-40389-2_1
3. The National Lung Screening Trial Research Team. Reduced Lung-Cancer Mortality With Low-Dose Computed Tomographic Screening. *N Engl J Med* (2011) 409:365:395. doi: 10.1056/NEJMOA1102873
4. Austin JH, Müller NL, Friedman PJ, Hansell DM, Naidich DP, Remy-Jardin M, et al. Glossary of Terms for CT of the Lungs: Recommendations of the Nomenclature Committee of the Fleischner Society. *Radiology* (1996) 200 (2):327–31. doi: 10.1148/RADIOLOGY.200.2.8685321
5. Erasmus JJ, Connolly JE, McAdams HP, Roggli VL. Solitary Pulmonary Nodules: Part I. Morphologic Evaluation for Differentiation of Benign and Malignant Lesions. *RadioGraphics* (2000) 20:43–58. doi: 10.1148/radiographics.20.1.g00ja0343
6. Aerts HJWL, Velazquez ER, Leijenaar RTH. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5. doi: 10.1038/ncomms5006
7. Chen C-H, Chang C-K, Tu C-Y, Liao W-C, Wu B-R, Chou K-T, et al. Radiomic Features Analysis in Computed Tomography Images of Lung Nodule Classification. *PloS One* (2018) 13:e0192002. doi: 10.1371/journal.pone.0192002
8. Thawani R, McLane M, Beig N, Ghose S, Prasanna P, Velcheti V, et al. Radiomics and Radiogenomics in Lung Cancer: A Review for the Clinician. *Lung Cancer* (2018) 115:34–41. doi: 10.1016/J.LUNGCAN.2017.10.015
9. Wu W, Pierce LA, Zhang Y, Pipavath SNJ, Randolph TW, Lastwika KJ, et al. Comparison of Prediction Models With Radiological Semantic Features and Radiomics in Lung Cancer Diagnosis of the Pulmonary Nodules: A Case-Control Study. *Eur Radiol* (2019) 29:6100–8. doi: 10.1007/s00330-019-06213-9
10. Alahmari SS, Cherezov D, Goldgof DB, Hall LO, Gillies RJ, Schabath MB. Delta Radiomics Improves Pulmonary Nodule Malignancy Prediction in Lung Cancer Screening. *IEEE Access* (2018) 6:77796–806. doi: 10.1109/ACCESS.2018.2884126
11. Mobiny A, Yuan P, Cicalese PA, Moulik SK, Garg N, Wu CC, et al. Memory-Augmented Capsule Network for Adaptable Lung Nodule Classification. *IEEE Trans Med Imaging* (2021) 40(10):2869–79. doi: 10.1109/TMI.2021.3051089
12. Xie Y, Xia Y, Zhang J, Song Y, Feng D, Fulham M, et al. Knowledge-Based Collaborative Deep Learning for Benign-Malignant Lung Nodule Classification on Chest CT. *IEEE Trans Med Imaging* (2019) 38:991–1004. doi: 10.1109/TMI.2018.2876510
13. Liu L, Dou Q, Chen H, Qin J, Heng PA. Multi-Task Deep Model With Margin Ranking Loss for Lung Nodule Analysis. *IEEE Trans Med Imaging* (2020) 39:718–28. doi: 10.1109/TMI.2019.2934577
14. Xu X, Wang C, Guo J, Gan Y, Wang J, Bai H, et al. MSCS-DeepLN: Evaluating Lung Nodule Malignancy Using Multi-Scale Cost-Sensitive Neural Networks. *Med Image Anal* (2020) 65:101772. doi: 10.1016/j.media.2020.101772
15. Lei Y, Tian Y, Shan H, Zhang J, Wang G, Kalra MK. Shape and Margin-Aware Lung Nodule Classification in Low-Dose CT Images *via* Soft Activation Mapping. *Med Image Anal* (2020) 60:101628. doi: 10.1016/J.MEDIA.2019.101628
16. Metz S, Ganter C, Lorenzen S, Van Marwick S, Holzapfel K, Herrmann K, et al. Multiparametric MR and PET Imaging of Intratumoral Biological Heterogeneity in Patients With Metastatic Lung Cancer Using Voxel-by-Voxel Analysis. *PloS One* (2015) 10:132386. doi: 10.1371/journal.pone.0132386
17. Xiong Z, Jiang Y, Che S, Zhao W, Guo Y, Li G, et al. Use of CT Radiomics to Differentiate Minimally Invasive Adenocarcinomas and Invasive

Adenocarcinomas Presenting as Pure Ground-Glass Nodules Larger Than 10 Mm. *Eur J Radiol* (2021) 141:109772. doi: 10.1016/j.ejrad.2021.109772
18. Wei G, Ma H, Qian W, Han F, Jiang H, Qi S, et al. Lung Nodule Classification Using Local Kernel Regression Models With Out-of-Sample Extension. *Biomed Signal Process Control* (2018) 40:1–9. doi: 10.1016/J.BSPC.2017.08.026
19. Buty M, Xu Z, Gao M, Bagci U, Wu A, Mollura DJ. Characterization of Lung Nodule Malignancy Using Hybrid Shape and Appearance Features. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*. Cham: Springer (2016). p. 662–70. doi: 10.1007/978-3-319-46720-7_77
20. Han F, Wang H, Zhang G, Han H, Song B, Li L, et al. Texture Feature Analysis for Computer-Aided Diagnosis on Pulmonary Nodules. *J Digital Imaging* (2015) 28:99–115. doi: 10.1007/s10278-014-9718-8
21. Cherezov D, Goldgof D, Hall L, Gillies R, Schabath M, Müller H, et al. Revealing Tumor Habitats From Texture Heterogeneity Analysis for Classification of Lung Cancer Malignancy and Aggressiveness. *Sci Rep* (2019) 9:1–9. doi: 10.1038/s41598-019-38831-0
22. Savitha G. Jidesh P. A Fully-Automated System for Identification and Classification of Subsolid Nodules in Lung Computed Tomographic Scans. *Biomed Signal Process Control* (2019) 53:101586. doi: 10.1016/j.bspc.2019.101586
23. Yoshimasu T, Kawago M, Hirai Y, Ohashi T, Tanaka Y, Oura S, et al. Fast Fourier Transform Analysis of Pulmonary Nodules on Computed Tomography Images From Patients With Lung Cancer. *Ann Thorac Cardiovasc Surg* (2015) 21:1–7. doi: 10.5761/atcs.oa.13-00318
24. Ciompi F, Jacobs C, Scholten ET, Wille MMW, De Jong PA, Prokop M, et al. Bag-Of-Frequencies: A Descriptor of Pulmonary Nodules in Computed Tomography Images. *IEEE Trans Med Imaging* (2015) 34:962–73. doi: 10.1109/TMI.2014.2371821
25. Kakino R, Nakamura M, Mitsuyoshi T, Shintani T, Hirashima H, Matsuo Y, et al. Comparison of Radiomic Features in Diagnostic CT Images With and Without Contrast Enhancement in the Delayed Phase for NSCLC Patients. *Physica Med* (2020) 69:176–82. doi: 10.1016/j.ejmp.2019.12.019
26. Avanzo M, Stancanello J, El Naqa I. Beyond Imaging: The Promise of Radiomics. *Physica Med* (2017) 38:122–39. doi: 10.1016/j.ejmp.2017.05.071
27. Autrusseau PA, Labani A, De Marini P, Leyendecker P, Hintzpeter C, Ortlieb AC, et al. Radiomics in the Evaluation of Lung Nodules: Intrapatient Concordance Between Full-Dose and Ultra-Low-Dose Chest Computed Tomography. *Diagn Interventional Imaging* (2021) 102:233–9. doi: 10.1016/j.diii.2021.01.010
28. Zhou C, Chan HP, Chughtai A, Hadjiiski LM, Kazerooni EA, Wei J. Pathologic Categorization of Lung Nodules: Radiomic Descriptors of CT Attenuation Distribution Patterns of Solid and Subsolid Nodules in Low-Dose CT. *Eur J Radiol* (2020) 129:109106. doi: 10.1016/j.ejrad.2020.109106
29. Yang X, He J, Wang J, Li W, Liu C, Gao D, et al. CT-Based Radiomics Signature for Differentiating Solitary Granulomatous Nodules From Solid Lung Adenocarcinoma. *Lung Cancer* (2018) 125:109–14. doi: 10.1016/j.lungcan.2018.09.013
30. Xie Y, Zhang J, Xia Y, Fulham M, Zhang Y. Fusing Texture, Shape and Deep Model-Learned Information at Decision Level for Automated Classification of Lung Nodules on Chest CT. *Inf Fusion* (2018) 42:102–10. doi: 10.1016/J.INFFUS.2017.10.005
31. Liu X, Hou F, Qin H, Hao A. Multi-View Multi-Scale CNNs for Lung Nodule Type Classification From CT Images. *Pattern Recognition* (2018) 77:262–75. doi: 10.1016/J.PATCOG.2017.12.022
32. Causey JL, Zhang J, Ma S, Jiang B, Qualls JA, Politte DG, et al. Highly Accurate Model for Prediction of Lung Nodule Malignancy With CT Scans. *Sci Rep* (2018) 8:9286. doi: 10.1038/s41598-018-27569-w
33. Zhao X, Liu L, Qi S, Teng Y, Li J, Qian W. Agile Convolutional Neural Network for Pulmonary Nodule Classification Using CT Images. *Int J Comput Assisted Radiol Surg* (2018) 13:585–95. doi: 10.1007/S11548-017-1696-0
34. Paul R, Schabath M, Gillies R, Hall L, Goldgof D. Convolutional Neural Network Ensembles for Accurate Lung Nodule Malignancy Prediction 2 Years in the Future. *Comput Biol Med* (2020) 122:103882. doi: 10.1016/j.compbiomed.2020.103882
35. Xie Y, Zhang J, Xia Y. Semi-Supervised Adversarial Model for Benign–Malignant Lung Nodule Classification on Chest CT. *Med Image Anal* (2019) 57:237–48. doi: 10.1016/J.MEDIA.2019.07.004
36. Astaraki M, Zakko Y, Toma Dasu I, Smedby Ö, Wang C. Benign-Malignant Pulmonary Nodule Classification in Low-Dose CT With Convolutional Features. *Physica Med* (2021) 83:146–53. doi: 10.1016/j.ejmp.2021.03.013
37. Kaggle, Data Science Bowl 2017 Challenge. Available at: https://www.kaggle.com/c/data-science-bowl-2017.
38. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
39. Zhang B, Tian J, Dong D, Gu D, Dong Y, Zhang L, et al. Radiomics Features of Multiparametric MRI as Novel Prognostic Factors in Advanced Nasopharyngeal Carcinoma. *Clin Cancer Res* (2017) 23:4259–69. doi: 10.1158/1078-0432.CCR-16-2910
40. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. (2015).
41. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE (2016).
42. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely Connected Convolutional Networks. In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. Honolulu, HI, USA: IEEE (2016). p. 2261–9.
43. Tan M, Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In: *36th International Conference on Machine Learning, ICML 2019* Long Beach, California, US: Proceedings of Machine Learning Research (2019). p. 10691–700.
44. Kingma DP, Welling M. Auto-Encoding Variational Bayes. (2014).
45. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-Sampling Technique. *J Artif Intell Res* (2002) 16:321–57. doi: 10.1613/jair.953
46. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics* (1988) 44:837. doi: 10.2307/2531595
47. Van der Maaten L, Hinton G. Visualizing Data Using T-SNE. *J Mach Learn Res* (2008) 9:2579–605.
48. Astaraki M, Wang C, Buizza G, Toma-Dasu I, Lazzeroni M, Smedby Ö. Early Survival Prediction in non-Small Cell Lung Cancer From PET/CT Images Using an Intra-Tumor Partitioning Method. *Physica Med* (2019) 60:58–65. doi: 10.1016/J.EJMP.2019.03.024
49. Nielsen D. Tree Boosting With XGBoost Why Does XGBoost Win Every Machine Learning Competition? (2016). Available at: http://hdl.handle.net/11250/2433761.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Astaraki, Yang, Zakko, Toma-Dasu, Smedby and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 27 January 2022
doi: 10.3389/fonc.2021.742547

Development and Validation of a CT-Based Radiomics Nomogram for Predicting Postoperative Progression-Free Survival in Stage I–III Renal Cell Carcinoma

Haijie Zhang[1,2†], *Fu Yin*[3†], *Menglin Chen*[1], *Liyang Yang*[1], *Anqi Qi*[1], *Weiwei Cui*[1], *Shanshan Yang*[1] *and Ge Wen*[1*]

[1] Department of Imaging, Nanfang Hospital, Southern Medical University, Guangzhou, China, [2] PET/CT Center, Department of Nuclear Medicine, First Affiliated Hospital of Shenzhen University, Shenzhen Second People's Hospital, Shenzhen, China, [3] College of Information Engineering, Shenzhen University, Shenzhen, China

OPEN ACCESS

Edited by:
Giorgio Russo,
Institute of Bioimaging and Molecular Physiology (CNR), Italy

Reviewed by:
Yuanshen Zhao,
Chinese Academy of Sciences (CAS), China
Arie Carneiro,
Albert Einstein Israelite Hospital, Brazil
Dapeng Hao,
The Affiliated Hospital of Qingdao University, China

***Correspondence:**
Ge Wen
m13360022166@163.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 16 July 2021
Accepted: 29 November 2021
Published: 27 January 2022

Citation:
Zhang H, Yin F, Chen M, Yang L, Qi A, Cui W, Yang S and Wen G (2022) Development and Validation of a CT-Based Radiomics Nomogram for Predicting Postoperative Progression-Free Survival in Stage I–III Renal Cell Carcinoma. Front. Oncol. 11:742547. doi: 10.3389/fonc.2021.742547

Background: Many patients experience recurrence of renal cell carcinoma (RCC) after radical and partial nephrectomy. Radiomics nomogram is a newly used noninvasive tool that could predict tumor phenotypes.

Objective: To investigate Radiomics Features (RFs) associated with progression-free survival (PFS) of RCC, assessing its incremental value over clinical factors, and to develop a visual nomogram in order to provide reference for individualized treatment.

Methods: The RFs and clinicopathological data of 175 patients (125 in the training set and 50 in the validation set) with clear cell RCC (ccRCC) were retrospectively analyzed. In the training set, RFs were extracted from multiphase enhanced CT tumor volume and selected using the stability LASSO feature selection algorithm. A radiomics nomogram final model was developed that incorporated the RFs weighted sum and selected clinical predictors based on the multivariate Cox proportional hazard regression. The performances of a clinical variables-only model, RFs-only model, and the final model were compared by receiver operator characteristic (ROC) analysis and DeLong test. Nomogram performance was determined and validated with respect to its discrimination, calibration, reclassification, and clinical usefulness.

Results: The radiomics nomogram included age, clinical stage, KPS score, and RFs weighted sum, which consisted of 6 selected RFs. The final model showed good discrimination, with a C-index of 0.836 and 0.706 in training and validation, and good calibration. In the training set, the C-index of the final model was significantly larger than the clinical-only model (DeLong test, $p = 0.008$). From the clinical variables-only model to the final model, the reclassification of net reclassification improvement was 18.03%, and

the integrated discrimination improvement was 19.08%. Decision curve analysis demonstrated the clinical usefulness of the radiomics nomogram.

Conclusion: The CT-based RF is an improvement factor for clinical variables-only model. The radiomics nomogram provides individualized risk assessment of postoperative PFS for patients with RCC.

Keywords: renal cell carcinoma (RCC), Radiomics, CT, progression-free survival (PFS), predict model, artificial intelligence

INTRODUCTION

Renal cell carcinoma (RCC) is a malignant tumor originating from the proximal tubular epithelial system of renal parenchyma, and accounts for about 85% of all adult renal malignant tumors. It is estimated that, in 2020, there were 431,288 new cases of RCC worldwide, resulting in 179,368 deaths, and accounting for 2.2% and 1.8% of global new cancer morbidity and mortality, respectively. In addition, the incidence of renal cancer is increasing yearly (1). The clear cell renal cell carcinoma (ccRCC) is the most common subtype accounting for about 75% of all RCC (2), and is associated with high invasion and poor prognosis (3, 4). According to the AJCC Tumor Classification Criteria eighth Edition (2017) (5), surgery is the preferred treatment for patients with stage I–III RCC, and is associated with a 5-year survival rate of 71% to 91% (6). However, approximately 20% to 30% of patients will relapse after surgery (7). If we can predict these patients with high risk of recurrence before surgery, and give them targeted treatment and close follow-up, it will be very helpful to improve the prognosis of these patients.

Traditional radiotherapy and chemotherapy have poor efficacy for RCC, and there are no effective adjuvant therapies for RCC. A recent clinical trial showed that a subset of patients with more aggressive disease could benefit from targeted therapy after surgery (8). According to National Comprehensive Cancer Network and European Association of Urology guidelines, adjuvant therapy can reduce the recurrence rate of stage III ccRCC, which is associated with a high recurrence risk. However, about 50% of patients in this high-risk subgroup still do not have postoperative recurrence, and do not need to receive expensive adjuvant targeted therapy. Therefore, there is a need to develop prognostic factors to identify patients who will and will not benefit from adjuvant targeted therapy.

There are currently no markers for a diagnosis of RCC. Tumor stage and pathological nuclear grade are the most important prognostic factors. Nevertheless, distinct outcomes are demonstrated in patients with equivalent tumor-node-metastasis stage and pathological grade; they cannot fully address the issue of individualized treatment for patients with different recurrence risks (9, 10).

The field of artificial intelligence and radiomics has developed rapidly in recent years. Many studies have demonstrated that radiomics can be used to assess the heterogeneity of tumors, thus providing clinicians with more accurate prognostic information to inform treatment decisions (11–13). It has been increasingly reported that radiomics can be used for differentiating benign and malignant renal tumors, as well as discriminating high and low Fuhrman nuclear ccRCC (14, 15). However, to the best of our knowledge, no study has evaluated radiomics for its ability to predict the aggressive potential of ccRCC.

The purpose of this study was to investigate Radiomics Features (RFs) associated with progression-free survival (PFS) of RCC, assessing its incremental value over clinical factors, and to develop a visual nomogram in order to provide reference for individualized treatment and prognosis evaluation of RCC.

MATERIALS AND METHODS

Patients

This retrospective study was approved by the Ethics Committee of Southern Medical University, and because of the retrospective nature of the analysis, the requirement of informed patient consent was waived. The data of patients with RCC who were treated in the Department of Urology, Nanfang Hospital, Southern Medical University, from March 2011 to March 2016 were retrospectively collected. Patients were randomly divided into a training set and validation set in a ratio of approximately 7:3. Clinical, pathological, and surgical data collected included age, sex, symptoms (low back pain, hematuria, emaciation, low-grade fever, cough, abdominal mass, and paraneoplastic syndrome), the interval from diagnosis to treatment, Karnofsky performance status (KPS) score, hemoglobin level, serum calcium level, neutrophil count, platelet count, maximum tumor size, stage, and pathological subtype, WHO/ISUP nuclear grade, overall tumor size, growth pattern, necrosis and calcification, treatment method, and adjuvant treatment. Inclusion criteria for the study were as follows: (1) Histological subtype ccRCC; (2) Clinical stage I–III (stage I–II, T1-2N0M0 and stage III, T1-2N1M0, T3N0-1M0); (3) Received radical nephrectomy or partial nephrectomy; and (4) Complete enhanced computed tomography (CT) imaging data, containing non-contrast phase (NCP), cortico-medullary phase (CMP), nephrographic phase (NP), and excretory phase (EP). Patients with a complete cystic renal tumor, positive tumor margins, and inadequate CT images were excluded.

CT Parameters

A Siemens 64-slice (Somatom Definition CT scanner; Siemens Medical Solutions Company, Malvern, PA, USA) CT scanners were used. A supine CT scan was performed from the diaphragmatic apex to the lower poles of both kidneys with

breath holding using the following parameters: CT tube voltage = 120 kV, tube current = 150–320 mA, layer thickness = 5 mm, layer spacing = 5 mm, field of view = 360 mm, matrix = 512 × 512. After obtaining unenhanced images, Omnipaque (GE Healthcare) was injected into the anterior elbow vein with a high-pressure syringe at a dose of 2 ml/kg and an injection rate of 2.5 ml/s, with a maximum dose of 160 ml. Enhanced CT scanning was started at 25–30 s, 75–80 s, and 180–200 s.

Image Segmentation

Image segmentation was performed by 2 radiologists with 5 and 8 years, respectively, of abdominal imaging diagnosis experience using ITK-snap software. CT images were obtained from the PACS system, with a window width of 300–400 HU, a window level of 45–65 HU, and a slice thickness of 5 mm. The volume of the tumor was selected as the region of interest (ROI), and the edges were kept about 1 mm away from tumor edges to reduce interference from adjacent tissues (such as fat or normal renal tissue). Based on the threshold, areas with CT values less than −55 HU and greater than 350 HU pixels were filtered out. Intra-group and inter-group correlation coefficients (ICCs) were used to ensure stability and repeatability. First, 40 images were randomly segmented by the 2 physicians to assess reproducibility between groups. A week later, Doctor A repeated the same procedure to assess the reproducibility within the group. The results showed that ICC > 0.80 between groups and within groups; it means that the image segmentation was consistent, and the remaining image segmentation was performed by Dr. A.

Radiomics Feature Extraction and Selection

RFs were extracted, preprocessed, and filtered from segmented images using the PyRadiomics computing platform. First, the original CT images and 3D segmented images were imported into the platform for loading. Then, the image was preprocessed based on the Simple ITK software package embedded in the platform to ensure that the isotropic voxels of texture feature and shape feature are equidistant from adjacent bodies in all directions. Then, the preprocessed image was filtered based on the platform built-in filter, including PyWavelets and Simple ITK for wavelet filter and logarithmic filter, and NumPy for the remaining filters. Finally, four custom feature extraction methods based on the platform were used for radiomics feature extraction.

In the training set, feature selection was performed based on the stability LASSO algorithm. It performed through 100 times hierarchical 5-fold cross-validation. The LassoCV method used automatically found the optimal penalty coefficient α value through k-fold cross-verification. The criterion of α selection is to minimize generalization error. RFs were extracted in the training set with 5-fold cross-validations, and RFs with R2 > 0.8 were retained in the 5-fold cross-validation test set. The top 20% RFs of the 100 times 5-fold cross-validation were selected as the final features (**Figure 1**).

Development of Clinical Variable-Only, RFs-Only, and the Final Model

Multivariate Cox proportional hazard regression models were used to estimate the correlation coefficients of selected RFs to PFS, and to calculate the RFs-weighted sum as an independent variable in both training and validation sets. In the training set, univariate and multivariate Cox proportional hazard regression models were used to investigate factors associated with PFS. Independent variables with a value of $p < 0.05$ in univariate results were entered into a multivariate model, and variables that were significant in the multivariate model were considered factors associated with PFS,

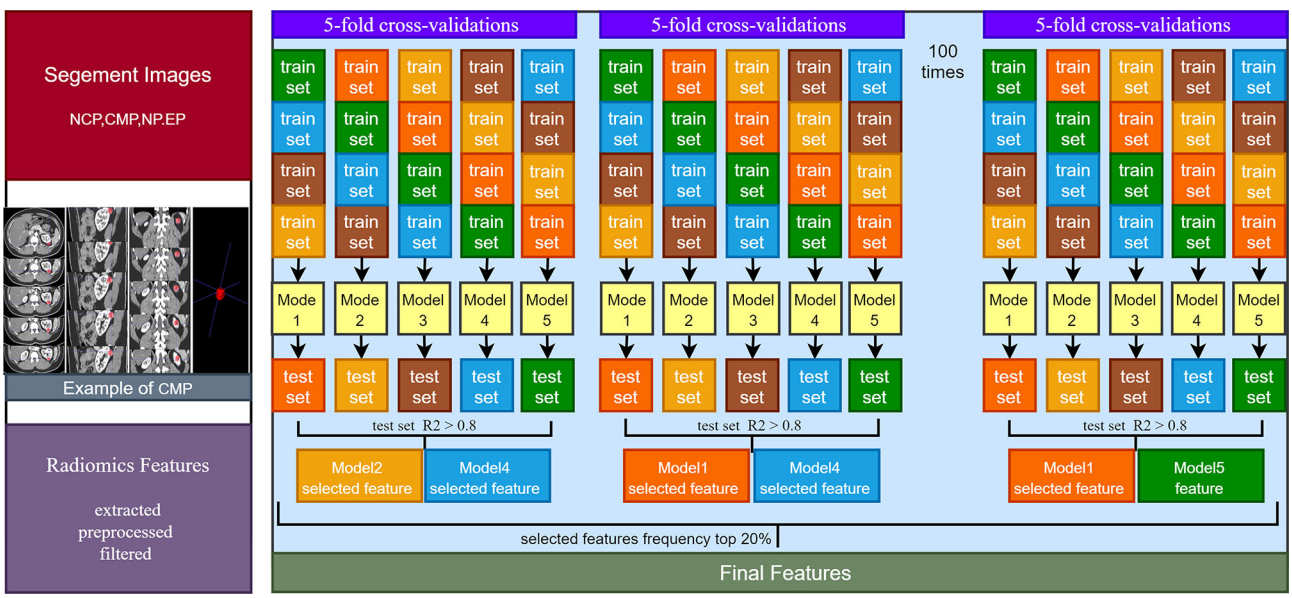


FIGURE 1 | Image segmentation and feature extraction, selection schematic diagram.

and estimated hazard ratios (HRs) were calculated. The final associated factors were used to build a multivariate Cox regression model. The optimal cutoff value of continuous variables was determined by the maximally selected rank statistics from the "maxstat" R package and used to predict PFS. Kaplan-Meier function was used to predict PFS for categorical variables.

Assessment of the Performance of Different Models

The probabilities were used as an independent continuous variable in receiver operating characteristic (ROC) curve analysis, and the C-index (area under curve the ROC curve [AUC]) and likelihood parameters were determined. A cutoff value was determined by maximizing the Youden index and used to predict PFS. ROC analysis was used in the validation set to assess the final model's diagnostic effectiveness.

To compare the performance of a clinical variable-only model, RFs-only model, and the final model, ROC analysis and the DeLong test were used. The Hosmer–Lemeshow test was used to check the calibration. Decision curve analysis was used to observe the net benefits. Finally, index net reclassification improvement (NRI) and integrated discrimination improvement (IDI) were used to calculate the increment from the clinical variables-only model to the final model.

Statistical Analysis

Continuous data were presented as mean ± standard deviation, and categorical data were presented as number and percentage (%). For comparisons of means between groups, Student's independent *t*-test or Mann–Whitney *U* test was used, depending on normality assumption. Categorical data were tested using chi-square test or Fisher's exact text (if an expected value ≤ 5 was found). In all analysis, a 2-tailed value of $p < 0.05$ was considered to indicate statistical significance. Statistical analyses were performed using IBM SPSS version 25 software (IBM Corporation, Somers, New York). A nomogram was established using the associated factors in the training set with R statistical software (version 3.5.2) and the "rms" package. The surv_cutpoint function in the "surviminer" package finds the best cutoff value for a continuous variable. The "Predict ABEL" package is used to calculate the NRI and IDI. The decision curve analysis was also performed with R software and "rmda" package.

RESULTS

Patient Clinical Characteristics and RFs of the Training and Validation Sets

A total of 175 patients were included in the analysis, with 125 in the training set and 50 in the validation set. The mean age of the training set was 52.31 ± 14.51 years, and that of the validation was 52.06 ± 13.19 years. The male:female ratio of the training set was 2.05:1, and that of the validation set was 1.63:1. The mean PFS of the training set was 55.83 ± 22.25 months, and that of the validation set was 61.06 ± 20.29 months. In the training set and

TABLE 1 | Patient's clinical characteristics between training set and validation set.

Parameters	Training set (*n* = 125)	Validation set (*n* = 50)	*p*
Age, years	52.31 ± 14.51	52.06 ± 13.19	0.915
Gender			0.513
Male	84 (67.20%)	31 (62.00%)	
Female	41 (32.80%)	19 (38.00%)	
Symptoms			0.811
No	65 (52.00%)	25 (50.00%)	
Yes	60 (48.00%)	25 (50.00%)	
Interval from diagnosis to treatment			1.000
>1 year	115 (92.00%)	46 (92.00%)	
<1 year	10 (8.00%)	4 (8.00%)	
KPS score			0.064
score ≥80	115 (92.00%)	50 (100.00%)	
score <80	10 (8.00%)	0 (0.00%)	
Hemoglobin			0.451
≥120 g/L	96 (76.80%)	41 (82.00%)	
<120 g/L	29 (23.20%)	9 (18.00%)	
Serum calcium			1.000
<10.2 mg/dl	117 (93.60%)	47 (94.00%)	
>10.2 mg/dl	8 (6.40%)	3 (6.00%)	
Neutrophils			0.367
$\leq 7 \times 10^9$/L	109 (87.20%)	46 (92.00%)	
$>7 \times 10^9$/L	16 (12.80%)	4 (8.00%)	
Platelet			0.240
≤ Normal level	92 (73.60%)	41 (82.00%)	
> Normal level	33 (26.40%)	9 (18.00%)	
Tumor size group			0.737
<40	54 (43.20%)	18 (36.00%)	
40–<70	49 (39.20%)	22 (44.00%)	
70–<100	15 (12.00%)	8 (16.00%)	
≥100	7 (5.60%)	2 (4.00%)	
T stage			0.539
T1	97 (77.60%)	38 (76.00%)	
T2	16 (12.80%)	9 (18.00%)	
T3	12 (9.60%)	3 (6.00%)	
N stage			0.760
N0	115 (92.00%)	47 (94.00%)	
N1	10 (8.00%)	3 (6.00%)	
Clinical stage			0.306
Stage I–I: $T_{1\text{-}2}N_0M_0$	105 (84.00%)	45 (90.00%)	
Stage III: $T_{1\text{-}2}N_1M_0$, $T_3N_{0\text{-}1}M_0$	20 (16.00%)	5 (10.00%)	
WHO/ISUP nuclear grade			0.160
Low	91 (72.80%)	31 (62.00%)	
High	34 (27.20%)	19 (38.00%)	
Grow pattern			0.241
Outside	40 (32.00%)	20 (40.00%)	
Middle	60 (48.00%)	17 (34.00%)	
Inside	25 (20.00%)	13 (26.00%)	
Necrosis			0.956
No	32 (25.60%)	13 (26.00%)	
Yes	93 (74.40%)	37 (74.00%)	
Calcification			0.855
No	101 (80.80%)	41 (82.00%)	
Yes	24 (19.20%)	9 (18.00%)	
Surgery type			0.586
Partial nephrectomy	48 (38.40%)	17 (34.00%)	
Radical nephrectomy	77 (61.60%)	33 (66.00%)	
Adjuvant therapy			0.551
No	107 (85.60%)	41 (82.00%)	
Yes	18 (14.40%)	9 (18.00%)	
PFS, month	55.83 ± 22.25	61.06 ± 20.29	0.152

(Continued)

TABLE 1 | Continued

Parameters	Training set (*n* = 125)	Validation set (*n* = 50)	*p*
Progression			0.252
No	84 (67.20%)	38 (76.00%)	
Yes	41 (32.80%)	12 (24.00%)	
1-year survival	121 (96.80%)	49 (98.00%)	1.000
3-year survival	107 (85.60%)	47 (94.00%)	0.122
5-year survival	98 (78.40%)	44 (88.00%)	0.142
RFs			
X1	160.39 ± 181.67	134.74 ± 149.50	0.377
X2	54.58 ± 54.19	59.69 ± 71.86	0.609
X3	37.42 ± 17.05	36.21 ± 15.84	0.665
X4	60.72 ± 34.13	59.03 ± 26.54	0.753
X5	0.21 ± 0.09	0.22 ± 0.09	0.739
X6	59.40 ± 41.28	66.25 ± 66.52	0.412
RFs weighted sum	−1.41 ± 1.28	−1.44 ± 1.31	0.900

NCP, Non-contrast phase; CMP, cortico-medullary phase; NP, nephrographic phase; EP, excretory phase.

validation set, 32.80% and 24.00% of patients experienced disease progression, respectively. Patient characteristics are summarized in **Table 1**, all clinical characteristics, follow-up results, and RFs of the training set and validation set were comparable (all, $p > 0.05$).

A total of 107 RFs were extracted from the 3D multiphase CT images of each phase of each patient. The RF features were categorized as follows: (1) first-order statistics, (2) shape-based features, and (3) texture features. Of the 107 RFs, there were18 first-order statistics features, 14 shape-based features, 24 gray-level cooccurrence matrix (GLCM) features, 16 gray-level size zone matrix (GLSZM) features, 16 gray-level run length matrix (GLRLM) features, 14 gray-level dependence matrix (GLDM) features, and 5 neighboring gray-tone difference matrix (NGTDM) features. A total of 428 (4×107) RFs were extracted from the 4 phase CT images. There were 6 RFs final selected, including X1, X2, X3, X4, X5, and X6. **Table 2** shows RF's designation, phase, abbreviation, classification, and description. A multivariate Cox regression model for PFS of the training set was established to integrate the RFs indices into a single index, the RFs-weighted sum, using the formula: RFs weighted sum = $5.5056 \times 10^{-5} \times X1 + 0.0013 \times X2 + 0.0028 \times X3 + 0.0053 \times X4 - 10.4363 \times X5 + 0.0048 \times X6$.

Predictive Model of PFS Using the Training Set

The univariate and multivariate Cox regression analyses results of the relations of independent variables to PFS in the training set are shown in **Table 3**. Variables significant in univariate results were entered into the multivariate model. Because of the high correlation between T stage and clinical stage ($r = 0.63$, $p < 0.001$), T stage was excluded in the final model. Thus, the final model for PFS was established using age, clinical stage, KPS score, and RFs-weighted sum (**Table 4**).

The results showed that patients with higher age, higher clinical stage, KPS score < 80, and larger RFs-weighted sum were more likely to have disease progression. The surv_cutpoint functions for age and RFs-weighted sum are shown in **Figures 2A, B**, the cutoff value of RFs-weighted sum and age was 48 and −0.73, and significant differences were found between age ≥ 48 and < 48 ($p = 0.013$) and between RFs-weighted sum ≥ −0.73 and < −0.73 ($p < 0.001$).The Kaplan–Meier survival functions for clinical stage and KPS score are shown in **Figures 3A, B**, and significant differences were found between clinical stage I–II and III ($p < 0.001$), and between KPS score ≥ 80 and < 80 ($p = 0.006$).

ROC Analysis and Nomogram

Table 5 and **Figure 4** show the results of ROC analysis of the final models of the training set and validation set. The C-index of training and validation models was 0.836 and 0.706, respectively. The Hosmer–Lemeshow test indicated that the final models of the training set (chi-square = 15.05, $p = 0.058$) and validation set (chi-square = 13.84, $p = 0.086$) were acceptable. A nomogram of the final model was established for clinical use (**Figure 5**), and included risk estimations of PFS, and 1-, 3-, and 5-year survival.

Comparisons of Clinical Variables-Only, RFs-Only, and Final Models

Three models were compared to investigate the importance of the RFs indices. The models included a clinical variables-only model (age, clinical stage, and KPS score), RFs-only model

TABLE 2 | Patient's selected RF's designation, phase, abbreviation, classification, and description.

Phase	Designation	Abbreviation	Category	Description
Corticomedullary phase	Size-Zone Non-Uniformity	X1	GLSZM	The variability of size zone volumes in the image, with a lower value indicating more homogeneity in size zone volumes.
Corticomedullary phase	Complexity	X2	NGTDM	An image is considered complex when there are many primitive components in the image, i.e., the image is non-uniform and there are many rapid changes in gray-level intensity.
Corticomedullary phase	Least Axis Length	X3	Shape	This feature yield the smallest axis length of the ROI-enclosing ellipsoid and is calculated using the largest principal component λ least.
Excretion period	Maximum 2D Diameter Row	X4	Shape	It is defined as the largest pairwise Euclidean distance between tumor surface mesh vertices in the column-slice (usually the sagittal) plane.
Non-enhanced phase	Surface Volume Ratio	X5	Shape	A lower value indicates a more compact (sphere-like) shape. This feature is not dimensionless, and is therefore (partly) dependent on the volume of the ROI.
Parenchyma phase	Maximum 2D Diameter Slice	X6	Shape	It is defined as the largest pairwise Euclidean distance between tumor surface mesh vertices in the row-column (generally the axial) plane.

TABLE 3 | Univariate and Multivariate Cox-regression results in training set.

Parameters	Univariate		Multivariate	
	HR (95% CI)	*p*	HR (95% CI)	*p*
Age, years	1.03 (1.00 to 1.05)	**0.032**	1.04 (1.01 to 1.07)	**0.012**
Gender				
Male	ref.	–		
Female	1.44 (0.77 to 2.68)	0.251		
Symptoms				
No	ref.	–		
Yes	1.06 (0.58 to 1.96)	0.842		
Interval from diagnosis to treatment				
>1 year	ref.	–		
<1 year	1.60 (0.57 to 4.49)	0.376		
KPS score				
score ≥80	ref.	–	ref.	–
score <80	3.21 (1.34 to 7.69)	**0.009**	2.80 (0.99 to 7.94)	**0.035**
Hemoglobin				
≥120 g/L	ref.	–		
<120 g/L	1.54 (0.78 to 3.01)	0.210		
Serum calcium				
<10.2 mg/dl	ref.	–		
>10.2 mg/dl	1.79 (0.64 to 5.03)	0.271		
Neutrophils				
$\leq 7 \times 10^9$/L	ref.	–	ref.	–
$>7 \times 10^9$/L	2.27 (1.08 to 4.77)	**0.031**	1.62 (0.65 to 4.02)	0.299
Platelet				
≤ Normal level	ref.	–	ref.	–
> Normal level	1.88 (1.00 to 3.52)	0.049	1.85 (0.76 to 4.52)	0.176
Tumor size group		**<0.001**		0.181
<40	ref.	–	ref.	–
40–<70	4.84 (1.94 to 12.10)	<0.001	1.41 (0.42 to 4.69)	0.576
70–<100	12.73 (4.44 to 36.51)	<0.001	5.48 (0.81 to 36.84)	0.080
≥100	19.98 (5.76 to 69.30)	<0.001	7.76 (0.84 to 71.25)	0.070
T stage		**<0.001**		**0.015**
T1	ref.	–	ref.	–
T2	2.62 (1.16 to 5.90)	0.021	0.29 (0.08 to 1.05)	0.059
T3	5.65 (2.59 to 12.32)	<0.001	0.04 (0.00 to 0.40)	0.006
N stage				
N0	ref.	–	ref.	–
N1	6.05 (2.60 to 14.11)	**<0.001**	0.20 (0.03 to 1.30)	0.091
Clinical stage				
Stage I–II: $T_{1\text{-}2}N_0M_0$	ref.	–	ref.	–
Stage III: $T_{1\text{-}2}N_1M_0$, $T_3N_{0\text{-}1}M_0$	5.90 (3.05 to 11.41)	**<0.001**	66.14 (6.38 to 685.77)	**<0.001**
WHO/ISUP nuclear grade				
Low	ref.	–		
High	1.86 (0.99 to 3.49)	0.052		
Grow pattern		0.345		
Outside	ref.	–		
Middle	1.00 (0.48 to 2.09)	0.992		
Inside	1.68 (0.74 to 3.82)	0.216		
Necrosis				
No	ref.	–		
Yes	1.83 (0.81 to 4.14)	0.145		
Calcification				
No	ref.	–	ref.	–
Yes	2.34 (1.20 to 4.56)	**0.013**	1.98 (0.84 to 4.66)	0.117
Surgery type				
Partial nephrectomy	ref.	–		
Radical nephrectomy	1.35 (0.71 to 2.59)	0.359		
Adjuvant therapy				
No	ref.	–		
Yes	1.16 (0.49 to 2.77)	0.732		
RFs weighted sum	2.72 (1.95 to 3.78)	**<0.001**	1.95 (1.00 to 3.80)	**0.049**

P values less than 0.05 are in bold.

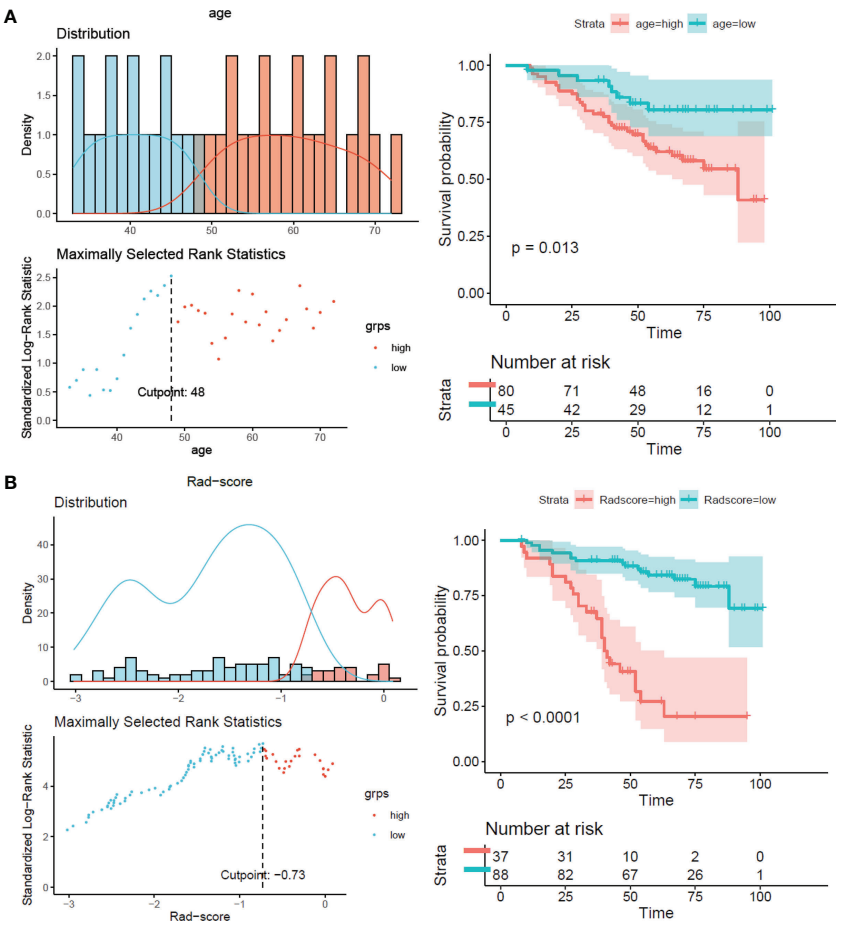


FIGURE 2 | Surv_cutpoint function and survival analysis of PFS in the training set. **(A)** RFs-weighted sum. **(B)** Age.

(RFs-weighted sum), and final model (age, clinical stage, KPS score, and RFs-weighted sum). The ROC analysis of the models is shown in **Figure 6**. It was found that the C-index of the final model was significantly larger than the clinical-only model (DeLong test, $p = 0.008$), but not significant compared to the RFs-only model ($p = 0.402$). However, no significance was found among the paired comparisons of models in the validation set (all $p > 0.05$).

The decision curve analysis is shown in **Figure 7**. The net benefit was higher in the final model than in the clinical variables-only model all the time and also higher than that of the RFs-only model after a 0.4 risk-threshold. The NRI from the clinical variables-only model to the final model was 18.03% ($Z = 1.80$, $p = 0.072$), marginally significant; and the IDI was 19.08% ($Z = 6.39$, $p < 0.001$).

DISCUSSION

In this retrospective analysis, we developed a radiomics nomogram that incorporates three clinical factors and RFs-weighted sum for noninvasive, individualized prediction of PFS in patients with clinical stage I–III ccRCC, which can enable physicians to select reasonable treatment tactics and individualized monitoring protocols to improve clinical outcomes. To the best of our knowledge, this is the first prediction model developed to predict PFS of resectable ccRCC using CT-based radiomics. Through cross-validation and calibration, the RFs selection of this study ensures reliability and avoids over-fitting of the model (16). There were 41 progressive cases in the training set used for modeling, and the nomogram contains 4 factors. The variable selection is consistent with the 10–15 EPV (Event per Variable) criteria proposed by Peduzzi et al. (17), indicating that the model is reliable. The proposed radiomics nomogram demonstrated favorable discrimination in both the training set (C-index, 0.836) and validation set (C-index, 0.706), with high sensitivity, specificity, and accuracy.

Many factors, including clinical, anatomical, pathological, and molecular factors and treatment methods, are related to the prognosis of RCC. Early treatment, being asymptomatic, and a higher KPS score (>80) are very important for prolonging the

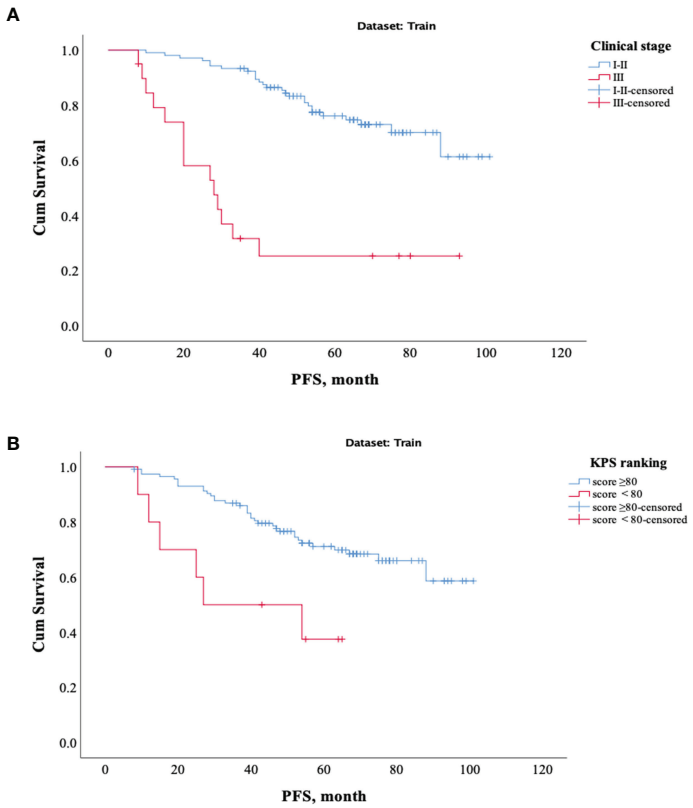


FIGURE 3 | Kaplan–Meier survival analysis of PFS in the training set. **(A)** Clinical stage. **(B)** KPS score.

TABLE 4 | Final model by training set.

Parameters	HR (95% CI)	*p*
Age, years	1.01 (0.99 to 1.04)	0.221
Clinical stage		
Stage I–II: $T_{1\text{-}2}N_0M_0$	ref.	–
Stage III: $T_{1\text{-}2}N_1M_0$, $T_3N_{0\text{-}1}M_0$	3.79 (1.85 to 7.76)	<0.001
KPS ranking		
Score ≥80	ref.	–
Score <80	3.14 (1.26 to 7.86)	0.014
RFs weighted sum	2.48 (1.75 to 3.53)	<0.001

survival of patients with RCC (18). Inflammatory cells (19–21), the TNM classification, and the histological factors included tumor nuclear grade, subtype, sarcomatoid features, microvascular invasion, tumor necrosis, and collection system invasion play a very important role in the tumor prognosis (22, 23). Many molecular factors such as CAIX, VEGF, HIF, Ki67, p53, p21, cell cycle PTEN, E-cadherin, CD44, and CXCR4, as well as other cell cycle and proliferation markers, may be associated with the prognosis of RCC (24–27). Partial nephrectomy is associated with improved survival of early RCC (28). Park et al. (29) reviewed preoperative laboratory data in 747 RCC patients and revealed that clinical information supporting aggressive ccRCC included an older age, larger size, lower hemoglobin, albumin, and calcium, as well as higher platelet and neutrophil. However, few radiologic parameters have been reported as prognostic factors of ccRCC in contrast to pathological markers. We enrolled these variables in this study; univariate Cox regression analysis showed that age, KPS score, neutrophils, tumor size, T stage, N stage, clinical stage, and calcification were associated with PFS, which was consistent with previous studies. However, affected by the RFs weighted sum, the clinical factors of the final model were only age, KPS score, and clinical stage. While RCC can be seen in all age groups, the median age of onset is 64 years, with a high incidence ranging from 50 to 70 years old. The mean age of

TABLE 5 | C-index and diagnostic index of final model in both dataset.

Data set	C-index (95% CI)	*p*	Sensitivity	Specificity	Youden	Accuracy	PPV	NPV	PLR	NLR
Training (*n* = 125)	0.836 (0.763 to 0.909)	<0.001	0.76	0.77	0.53	0.77	0.62	0.87	3.34	0.32
Validation (*n* = 50)	0.706 (0.511 to 0.901)	0.033	0.58	0.74	0.32	0.70	0.41	0.85	2.22	0.57

PPV, Positive predictive value; NPV, negative predictive value; PLR, positive likelihood ratio; NLR, negative likelihood ratio.

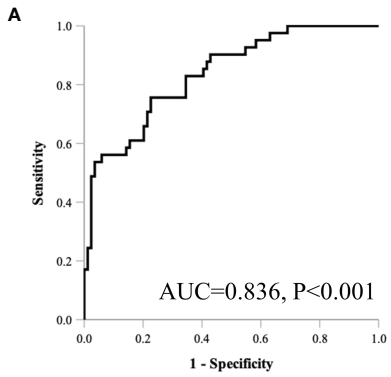


FIGURE 4 | ROC results of the final model of the training set **(A)** and validation set **(B)**.

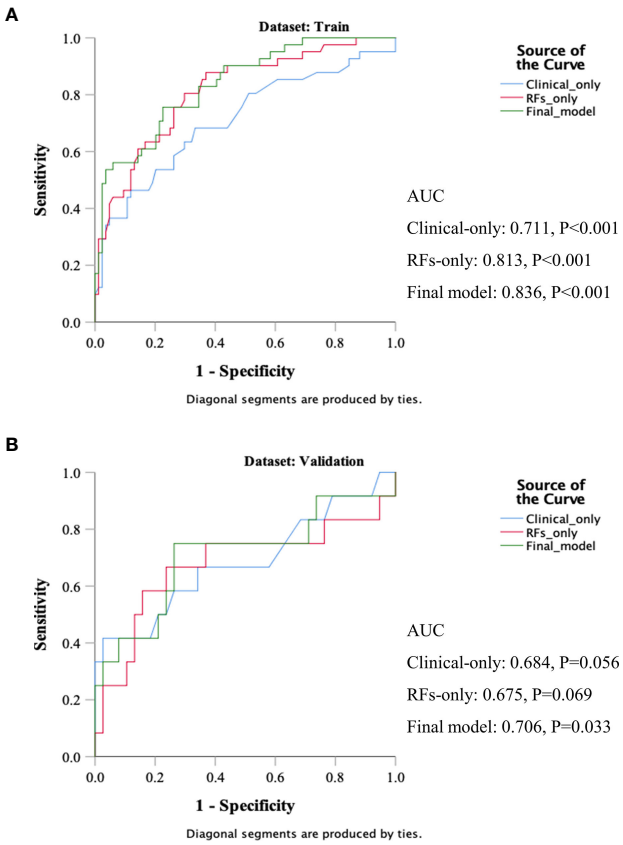


FIGURE 6 | ROC results of the clinical variables-only model, RFs only model, and final model of the training set **(A)** and validation set **(B)**.

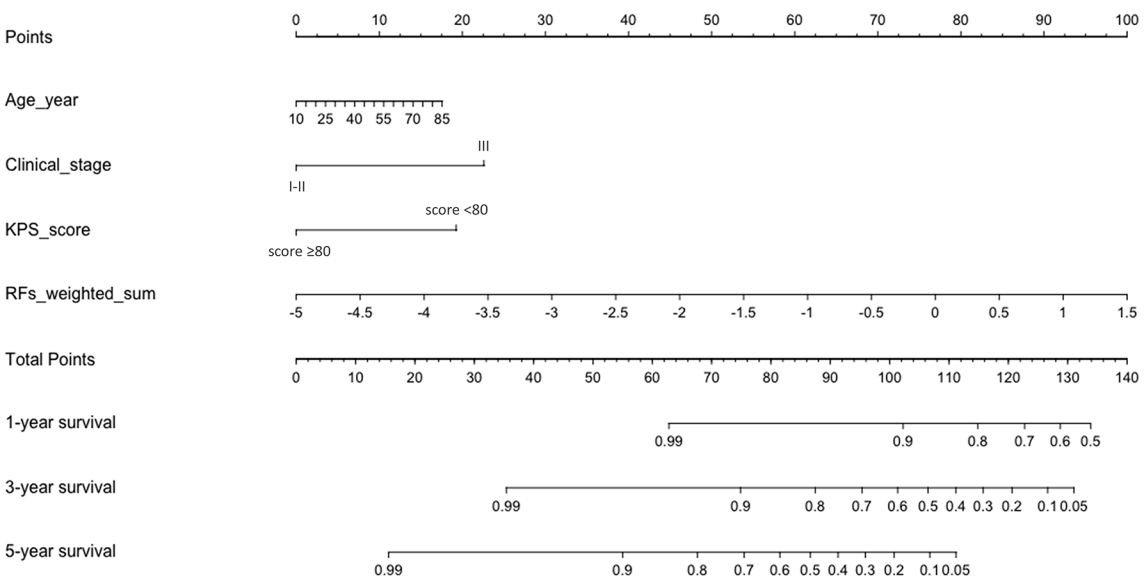


FIGURE 5 | A nomogram for PFS was established that included age, clinical stage, KPS score, and RFs-weighted sum.

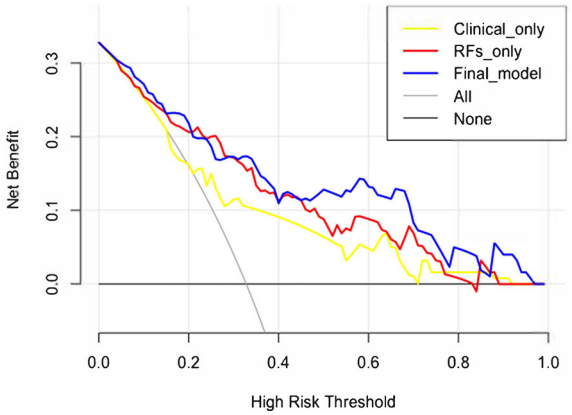


FIGURE 7 | Decision curve analysis results of the clinical variables-only model, RFs only model, and final model.

patients in this study was 52 years, which is lower than that reported in the literature, and this may be related to the small sample size, patient race, and the relatively early disease stage.

Various scoring systems have been developed to predict the risk of postoperative recurrence in patients with RCC. The response evaluation criteria in solid tumors (RECIST) is the most commonly used prognostic evaluation method for tumors (30). However, it is impossible to predict the treatment effect before treatment. Other systems include the University of California, Los Angeles Integrated Staging System (UISS) (31), the Stage Size Grade and Necrosis (SSIGN) model (32), the Leibovich scoring system (33), the Kattan Nomogram (34), and the Karakiewic prognostic model (35). However, some of the parameters used in the model such as tumor necrosis and clinical presentation are subject to inter-observer variability. Different observation end points of different models result in different accuracy results in different studies, and ethnic differences and tumor diversity also limit the use of some systems; in various external validation samples, the results are not consistent (36, 37). Hence, further research and validation are needed.

RFs contain information about tumor heterogeneity and can reflect tumor phenotypes. Our study has filled a gap in the literature on PFS risk of stage I–III RCC in the setting of radiomics. In the recent literature, Radiomics nomogram has demonstrated excellent efficacy in differential diagnosis, nuclear grading, prognosis, and gene expression of RCC (38–43). Among the 6 RFs selected in this study, there were 3 features from the corticomedullary phase, suggesting that the corticomedullary phase may contain more abundant information to predict PFS. The results showed that RFs-weighted sum was an important factor that improved the diagnostic efficiency of the clinical variables-only model. The decision curve analysis revealed that using the radiomics nomogram to predict PFS in patients with stage I–III ccRCC presents more notable benefits than solely relying on clinical variables-only model.

There are several limitations to our study. First, owing to the limitation of the retrospective study and small number of cases, the follow-up time we used was at least 5 years. It would be more interesting to enroll patients without recurrence evidence for more than 5 or 10 years. Second, as a single-center study, the patient population was relatively homogeneous and small. During the 5-year recruiting period, there a large proportion of patients with stage I–II in this study (training set: 84%, verification set: 90%). A large-scale independent prospective multicenter study is needed to evaluate the generalizability of the results, and further work would focus on it.

In conclusion, this study presented a CT-based radiomics nomogram that showed satisfactory performance in predicting PFS in patients with stage I–III ccRCC, as a non-invasive and quantitative method that can be used as an efficient tool to complement individualized treatment.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Southern Medical University. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

HZ, FY, and GW contributed to conception and design of the study. MC, LY, AQ, WC, and SY organized the database. HZ and FY performed the statistical analysis. HZ wrote the first draft of the manuscript. FY wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
2. Lipworth L, Morgans AK, Edwards TL, Barocas DA, Chang SS, Herrell SD, et al. Renal Cell Cancer Histological Subtype Distribution Differs by Race and Sex. *BJU Int* (2016) 117(2):260–5. doi: 10.1111/bju.12950
3. Keegan KA, Schupp CW, Chamie K, Hellenthal NJ, Evans CP, Koppie TM. Histopathology of Surgically Treated Renal Cell Carcinoma: Survival

Differences by Subtype and Stage. *J Urol* (2012) 188(2):391–7. doi: 10.1016/j.juro.2012.04.006
4. Capitanio U, Cloutier V, Zini L, Isbarn H, Jeldres C, Shariat SF, et al. A Critical Assessment of the Prognostic Value of Clear Cell, Papillary and Chromophobe Histological Subtypes in Renal Cell Carcinoma: A Population-Based Study. *Bju Int* (2009) 103(11):1496–500. doi: 10.1111/j.1464-410X.2008.08259.x
5. Paner GP, Stadler WM, Hansel DE, Montironi R, Lin DW, Amin MB. Updates in the Eighth Edition of the Tumor-Node-Metastasis Staging Classification for Urologic Cancers. *Eur Urol* (2018) 73(4):560–9. doi: 10.1016/j.eururo.2017.12.018
6. Tsui KH, Shvarts O, Smith RB, Figlin RA, deKernion JB, Belldegrun A. Prognostic Indicators for Renal Cell Carcinoma: A Multivariate Analysis of 643 Patients Using the Revised 1997 TNM Staging Criteria. *J Urol* (2000) 163 (4):1090–5; quiz 1295. doi: 10.1016/s0022-5347(05)67699-9
7. Stewart SB, Thompson RH, Psutka SP, Cheville JC, Lohse CM, Boorjian SA, et al. Evaluation of the National Comprehensive Cancer Network and American Urological Association Renal Cell Carcinoma Surveillance Guidelines. *J Clin Oncol* (2014) 32(36):4059–65. doi: 10.1200/JCO.2014.56.5416
8. Ravaud A, Motzer RJ, Pandha HS, George DJ, Pantuck AJ, Patel A, et al. Adjuvant Sunitinib in High-Risk Renal-Cell Carcinoma After Nephrectomy. *N Engl J Med* (2016) 375(23):2246–54. doi: 10.1056/NEJMoa1611406
9. Amin MB, Greene FL, Edge SB, Compton CC, Gershenwald JE, Brookland RK, 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 Cancer J Clin* (2017) 67(2):93–9. doi: 10.3322/caac.21388
10. Martínez-Salamanca JI, Huang WC, Millán I, Bertini R, Bianco FJ, Carballido JA, et al. Prognostic Impact of the 2009 UICC/AJCC TNM Staging System for Renal Cell Carcinoma With Venous Extension. *Eur Urol* (2011) 57(1):120–7. doi: 10.1016/j.eururo.2010.10.001
11. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.CAN-17-0339
12. Ganeshan B, Nathan P, Juttla JK, Vinayan A, Miles KA. Assessment of Response to Tyrosine Kinase Inhibitors in Metastatic Renal Cell Cancer: CT Texture as a Predictive Biomarker. *Radiology* (2011) 261(1):165–71. doi: 10.1148/radiol.11110264
13. Lubner MG, Stabo N, Abel EJ, Del Rio AM, Pickhardt PJ. CT Textural Analysis of Large Primary Renal Cell Carcinomas: Pretreatment Tumor Heterogeneity Correlates With Histologic Findings and Clinical Outcomes. *AJR Am J Roentgenol* (2016) 207(1):96–105. doi: 10.2214/AJR.15.15451
14. Bektas CT, Kocak B, Yardimci AH, Turkcanoglu MH, Yucetas U, Koca SB, et al. Clear Cell Renal Cell Carcinoma: Machine Learning-Based Quantitative Computed Tomography Texture Analysis for Prediction of Fuhrman Nuclear Grade. *Eur Radiol* (2019) 29(3):1153–63. doi: 10.1007/s00330-018-5698-2
15. Feng Z, Rong P, Cao P, Zhou Q, Zhu W, Yan Z, et al. Machine Learning-Based Quantitative Texture Analysis of CT Images of Small Renal Masses: Differentiation of Angiomyolipoma Without Visible Fat From Renal Cell Carcinoma. *Eur Radiol* (2018) 28(4):1625–33. doi: 10.1007/s00330-017-5118-z
16. Babyak MA. What You See may Not be What You Get: A Brief, Nontechnical Introduction to Overfitting in Regression-Type Models. *Psychosom Med* (2004) 66(3):411–21. doi: 10.1097/01.psy.0000127692.23278.a9
17. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A Simulation Study of the Number of Events Per Variable in Logistic Regression Analysis. *J Clin Epidemiol* (1996) 49(12):1373–9. doi: 10.1016/s0895-4356(96)00236-3
18. Van Poppel H, Da Pozzo L, Albrecht W, Matveev V, Bono A, Borkowski A, et al. A Prospective, Randomised EORTC Intergroup Phase 3 Study Comparing the Oncologic Outcome of Elective Nephron-Sparing Surgery and Radical Nephrectomy for Low-Stage Renal Cell Carcinoma. *Eur Urol* (2011) 59(4):543–52. doi: 10.1016/j.eururo.2010.12.013
19. Grivennikov SI, Greten FR, Karin M. Immunity, Inflammation, and Cancer. *Cell* (2010) 140(6):883–99. doi: 10.1016/j.cell.2010.01.025
20. Szczerba BM, Castro-Giner F, Vetter M, Krol I, Gkountela S, Landin J, et al. Neutrophils Escort Circulating Tumour Cells to Enable Cell Cycle Progression. *Nature* (2019) 566(7745):553–7. doi: 10.1038/s41586-019-0915-y
21. Bambace NM, Holmes CE. The Platelet Contribution to Cancer Progression. *J Thromb Haemost* (2011) 9(2):237–49. doi: 10.1111/j.1538-7836.2010.04131.x
22. Zhang L, Zha Z, Qu W, Zhao H, Yuan J, Feng Y, et al. Tumor Necrosis as a Prognostic Variable for the Clinical Outcome in Patients With Renal Cell Carcinoma: A Systematic Review and Meta-Analysis. *BMC Cancer* (2018) 18 (1):870. doi: 10.1186/s12885-018-4773-z
23. Fuhrman SA, Lasky LC, Limas C. Prognostic Significance of Morphologic Parameters in Renal Cell Carcinoma. *Am J Surg Pathol* (1982) 6(7):655–63. doi: 10.1097/00000478-198210000-00007
24. Sim SH, Messenger MP, Gregory WM, Wind TC, Vasudev NS, Cartledge J, et al. Prognostic Utility of Pre-Operative Circulating Osteopontin, Carbonic Anhydrase IX and CRP in Renal Cell Carcinoma. *Br J Cancer* (2012) 107 (7):1131–7. doi: 10.1038/bjc.2012.360
25. Choueiri TK, Pal SK, McDermott DF, Morrissey S, Ferguson KC, Holland J, et al. A Phase I Study of Cabozantinib (XL184) in Patients With Renal Cell Cancer. *Ann Oncol* (2014) 25(8):1603–8. doi: 10.1093/annonc/mdu184
26. Choueiri TK, Escudier B, Powles T, Mainwaring PN, Rini BI, Donskov F, et al. Cabozantinib Versus Everolimus in Advanced Renal-Cell Carcinoma. *N Engl J Med* (2015) 373(19):1814–23. doi: 10.1056/NEJMoa1510016
27. Minardi D, Lucarini G, Milanese G, Di Primio R, Montironi R, Muzzonigro G. Loss of Nuclear BAP1 Protein Expression Is a Marker of Poor Prognosis in Patients With Clear Cell Renal Cell Carcinoma. *Urol Oncol* (2016) 34(8):338. e11–8. doi: 10.1016/j.urolonc.2016.03.006
28. Boorjian S. Commentary on "Long-Term Survival Following Partial vs. Radical Nephrectomy Among Older Patients With Early-Stage Kidney Cancer". Tan HJ, Norton EC, Ye Z, Hafez KS, Gore JL, Miller DC. *Urol Oncol* (2012) 31 (1):126–27. doi: 10.1016/j.urolonc.2012.11.004
29. Park YH, Baik KD, Lee YJ, Ku JH, Kim HH, Kwak C. Late Recurrence of Renal Cell Carcinoma >5 Years After Surgery: Clinicopathological Characteristics and Prognosis. *BJU Int* (2012) 110(11 Pt B):E553–8. doi: 10.1111/j.1464-410X.2012.11246.x
30. Hillman SL, An MW, O'Connell MJ, Goldberg RM, Schaefer P, Buckner JC, et al. Evaluation of the Optimal Number of Lesions Needed for Tumor Evaluation Using the Response Evaluation Criteria in Solid Tumors: A North Central Cancer Treatment Group Investigation. *J Clin Oncol* (2009) 27 (19):3205–10. doi: 10.1200/JCO.2008.18.3269
31. Patard JJ, Kim HL, Lam JS, Dorey FJ, Pantuck AJ, Zisman A, et al. Use of the University of California Los Angeles Integrated Staging System to Predict Survival in Renal Cell Carcinoma: An International Multicenter Study. *J Clin Oncol* (2004) 22(16):3316–22. doi: 10.1200/JCO.2004.09.104
32. Sorbellini M, Kattan MW, Snyder ME, Reuter V, Motzer R, Goetzl M, et al. A Postoperative Prognostic Nomogram Predicting Recurrence for Patients With Conventional Clear Cell Renal Cell Carcinoma. *J Urol* (2005) 173(1):48–51. doi: 10.1097/01.ju.0000148261.19532.2c
33. Leibovich BC, Blute ML, Cheville JC, Lohse CM, Frank I, Kwon ED, et al. Prediction of Progression After Radical Nephrectomy for Patients With Clear Cell Renal Cell Carcinoma: A Stratification Tool for Prospective Clinical Trials. *Cancer* (2003) 97(7):1663–71. doi: 10.1002/cncr.11234
34. Kattan MW, Reuter V, Motzer RJ, Katz J, Russo P. A Postoperative Prognostic Nomogram for Renal Cell Carcinoma. *J Urol* (2001) 166(1):63–7. doi: 10.1097/00005392-200107000-00015
35. Karakiewicz PI, Briganti A, Chun FK, Trinh QD, Perrotte P, Ficarra V, et al. Multi-Institutional Validation of a New Renal Cancer-Specific Survival Nomogram. *J Clin Oncol* (2007) 25(11):1316–22. doi: 10.1200/JCO.2006.06.1218
36. Beisland C, Gudbrandsdottir G, Reisæter LA, Bostad L, Wentzel-Larsen T, Hjelle KM. Contemporary External Validation of the Leibovich Model for Prediction of Progression After Radical Surgery for Clear Cell Renal Cell Carcinoma. *Scand J Urol* (2015) 49(3):205–10. doi: 10.3109/21681805.2014.980844
37. Pichler M, Hutterer GC, Chromecki TF, Jesche J, Kampel-Kettner K, Rehak P, et al. External Validation of the Leibovich Prognosis Score for Nonmetastatic Clear Cell Renal Cell Carcinoma at a Single European Center Applying Routine Pathology. *J Urol* (2011) 186(5):1773–7. doi: 10.1016/j.juro.2011.07.034
38. Nie P, Yang G, Wang Z, Yan L, Miao W, Hao D, et al. A CT-Based Radiomics Nomogram for Differentiation of Renal Angiomyolipoma Without Visible Fat

From Homogeneous Clear Cell Renal Cell Carcinoma. *Eur Radiol* (2020) 30 (2):1274–84. doi: 10.1007/s00330-019-06427-x
39. Kang B, Sun C, Gu H, Yang S, Yuan X, Ji C, et al. T1 Stage Clear Cell Renal Cell Carcinoma: A CT-Based Radiomics Nomogram to Estimate the Risk of Recurrence and Metastasis. *Front Oncol* (2020) 10:579619. doi: 10.3389/fonc.2020.579619
40. Zheng Z, Chen Z, Xie Y, Zhong Q, Xie W. Development and Validation of a CT-Based Nomogram for Preoperative Prediction of Clear Cell Renal Cell Carcinoma Grades. *Eur Radiol* (2021) 31(8):6078–86. doi: 10.1007/s00330-020-07667-y
41. Huang Y, Zeng H, Chen L, Luo Y, Ma X, Zhao Y. Exploration of an Integrative Prognostic Model of Radiogenomics Features With Underlying Gene Expression Patterns in Clear Cell Renal Cell Carcinoma. *Front Oncol* (2021) 11:640881. doi: 10.3389/fonc.2021.640881
42. Zhang H, Yin F, Chen M, Qi A, Lai Z, Yang L, et al. A Reliable Prediction Model for Renal Cell Carcinoma Subtype Based on Radiomic Features From 3d Multiphase Enhanced CT Images. *J Oncol* (2021) 2021:6595212. doi: 10.1155/2021/6595212
43. Antunes J, Viswanath S, Rusu M, Valls L, Hoimes C, Avril N, et al. Radiomics Analysis on FLT-PET/MRI for Characterization of Early Treatment Response in Renal Cell Carcinoma: A Proof-Of-Concept Study. *Transl Oncol* (2016) 9 (2):155–62. doi: 10.1016/j.tranon.2016.01.008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Yin, Chen, Yang, Qi, Cui, Yang and Wen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 25 February 2022
doi: 10.3389/fonc.2022.758863

CT-Based Radiomics Showing Generalization to Predict Tumor Regression Grade for Advanced Gastric Cancer Treated With Neoadjuvant Chemotherapy

Yong Chen[1†], Wei Xu[2†], Yan-Ling Li[3†], Wentao Liu[2], Birendra Kumar Sah[2], Lan Wang[1], Zhihan Xu[4], Michael Wels[5], Yanan Zheng[2], Min Yan[2], Huan Zhang[1], Qianchen Ma[6], Zhenggang Zhu[2*] and Chen Li[2*]*

[1] Department of Radiology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2] Department of General Surgery, Shanghai Key Laboratory of Gastric Neoplasms, Shanghai Institute of Digestive Surgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [3] Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiology, Peking University Cancer Hospital and Institute, Beijing, China, [4] Siemens Healthineers Ltd., Shanghai, China, [5] Department of Diagnostic Imaging Computed Tomography Image Analytics, Siemens Healthcare GmbH, Forchheim, Germany, [6] Department of Pathology, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

Objective: The aim of this study was to develop and validate a radiomics model to predict treatment response in patients with advanced gastric cancer (AGC) sensitive to neoadjuvant therapies and verify its generalization among different regimens, including neoadjuvant chemotherapy (NAC) and molecular targeted therapy.

Materials and Methods: A total of 373 patients with AGC receiving neoadjuvant therapies were enrolled from five cohorts. Four cohorts of patients received different regimens of NAC, including three retrospective cohorts (training cohort and internal and external validation cohorts) and a prospective Dragon III cohort (NCT03636893). Another prospective SOXA (apatinib in combination with S-1 and oxaliplatin) cohort received neoadjuvant molecular targeted therapy (ChiCTR-OPC-16010061). All patients underwent computed tomography before treatment, and thereafter, tumor regression grade (TRG) was assessed. The primary tumor was delineated, and 2,452 radiomics features were extracted for each patient. Mutual information and random forest were used for dimensionality reduction and modeling. The performance of the radiomics model to predict TRG under different neoadjuvant therapies was evaluated.

Results: There were 28 radiomics features selected. The radiomics model showed generalization to predict TRG for AGC patients across different NAC regimens, with areas under the curve (AUCs) (95% interval confidence) of 0.82 (0.76~0.90), 0.77 (0.63~0.91), 0.78 (0.66~0.89), and 0.72 (0.66~0.89) in the four cohorts, with no statistical difference observed (all $p > 0.05$). However, the radiomics model showed poor predictive value on

OPEN ACCESS

Edited by:
Joseph Ross Mitchell, University of Alberta, Canada

Reviewed by:
Guolin Ma, China-Japan Friendship Hospital, China
Feng Liu, Central South University, China

***Correspondence:**
Chen Li
leedoctor@sina.com
Zhenggang Zhu
zzg1954@hotmail.com
Qianchen Ma
qianchen_ma@sina.com

†These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *06 October 2021*
Accepted: *25 January 2022*
Published: *25 February 2022*

Citation:
Chen Y, Xu W, Li Y-L, Liu W, Sah BK, Wang L, Xu Z, Wels M, Zheng Y, Yan M, Zhang H, Ma Q, Zhu Z and Li C (2022) CT-Based Radiomics Showing Generalization to Predict Tumor Regression Grade for Advanced Gastric Cancer Treated With Neoadjuvant Chemotherapy. Front. Oncol. 12:758863. doi: 10.3389/fonc.2022.758863

the SOXA cohort [AUC, 0.50 (0.27~0.73)], which was significantly worse than that in the training cohort (p = 0.010).

Conclusion: Radiomics is generalizable to predict TRG for AGC patients receiving NAC treatments, which is beneficial to transform appropriate treatment, especially for those insensitive to NAC.

Keywords: radiomics, gastric cancer, neoadjuvant therapy, tumor regression grade, generalization

INTRODUCTION

Gastric cancer (GC) is a serious health problem in the world, causing an estimated 783,000 deaths in 2018 (1). Despite surgery being the only curative approach, more than half of cases are initially diagnosed as advanced disease, with a limited 5-year survival of 20%–30% (2). Meanwhile, even after R0 gastrectomy, relapse rates remain high, in the range of 40%–60% (2).

Neoadjuvant chemotherapy (NAC) is beneficial to improving R0 resection and prognosis in patients with advanced gastric cancer (AGC) by downstaging the tumor, eradicating micrometastasis, and reducing the risk of recurrence (3). Compared with only about half of patients with a good condition suitable to receive postoperative chemotherapy, NAC (the preoperative part of perioperative chemotherapy) is feasible to most patients, highlighting its importance (4). Although accumulated studies have been made to investigate regimens with more safety and effectiveness since the landmark MAGIC study launched in 2006, a considerable proportion of cases are insensitive to NAC, leaving unnecessary cytotoxicity to those patients (4–9). Even for the newly reported triplet FLOT regimen (docetaxel, oxaliplatin, fluorouracil, and leucovorin), which is under impassioned discussion as the new standard for NAC, completed or subtotal pathological regression was achieved in only 37% of cases (6). Uniform standard of care is an absence in NAC even if great progress has been achieved in this area (10, 11). Besides, the emergence of treatments for precision medicine, such as molecular targeted therapy and immunotherapy, brings vitality to neoadjuvant approaches (12). The choice of appropriate treatment is beset with difficulties where controversial results were observed from different studies (12). It is of urgent need to find an easy-to-use and noninvasive tool to predict tumor sensitivity to different neoadjuvant regimens.

Transformation in artificial intelligence (AI) has provoked a new area of medical image analysis named radiomics, which noninvasively provides insight into tumor heterogeneity by extracting and analyzing high-throughput image features (13, 14). Radiomics extends the scope of biopsy, providing possibilities in dynamic surveillance, prognosis prediction, and treatment decision (15). Especially, compared with conventional imaging metrics, radiomics effectively gauges tumor microenvironment status, providing tailored treatment for individuals (16–18). Radiomics also allows for predicting treatment sensitivity to non–small cell lung cancer treated with different systemic anticancer therapies (19). Accumulated evidence has identified the prognostic value of radiomics in evaluating tumor sensitivity for GC, but there is a lack of clear elucidation whether radiomics is generalizable among different anticancer regimens (20–22).

In this study, we aimed to predict treatment response for GC patients sensitive to neoadjuvant therapies using radiomics and verify the generalization of this AI technology among different regimens, including NAC regimens and molecular targeted therapy.

Abbreviations: AGC, advanced gastric cancer; AI, artificial intelligence; AUC, area under the curve; CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; CECT, contrast-enhanced computed tomography; GC, gastric cancer; IBSI, image biomarker standardization initiative; ICC, intraclass correlation coefficient; NAC, neoadjuvant chemotherapy; NPV, negative predictive value; PPV, positive predictive value; RECIST, Response Evaluation Criteria in Solid Tumors; RF, random forest; ROC, receiver operating characteristic; RQS, radiomics quality score; TRG, tumor regression grade; VOI, volume of interest.

MATERIALS AND METHODS

The main objective was to test whether the radiomics model we built was generalizable to detect AGC tumors sensitive to neoadjuvant approaches. We firstly trained and tested the radiomics model on baseline CT images for patients treated with the EOX regimen (epirubicin, oxaliplatin, and capecitabine). Then, we sought to investigate whether this model was practical on real-world data by applying it to predict therapeutic response on a clinical trial [the Dragon III study (23), ClinicalTrials.gov: NCT03636893] and a retrospective external validation cohort incorporated with three mixture NAC regimens. We finally aimed to validate whether our radiomics model could predict treatment response to molecular targeted therapy from another clinical trial (ChiCTR.gov.cn: ChiCTR-OPC-16010061) (24).

Participants

Patients consecutively enrolled in this study were histopathologically diagnosed as having gastric adenocarcinoma and with advanced disease (cT2-4a/bNxM0) based on contrast-enhanced CT (CECT) images. No prior treatment before NAC was demanded, and all patients underwent CECT scans less than 3 weeks before treatment started. Meanwhile, all patients enrolled were tolerant of NAC and have completed the planned preoperative schedule. The enrollment for all patients is presented in **Figure 1**.

Patients receiving the triplet EOX regimen were retrospectively recruited from May 2009 to September 2017. A total of 200 patients (women, 64; mean age, 59.5 ± 9.7 years) were finally recruited and were divided into a training cohort and an internal validation cohort to construct the radiomics model at a ratio of 2:1. Therefore, there

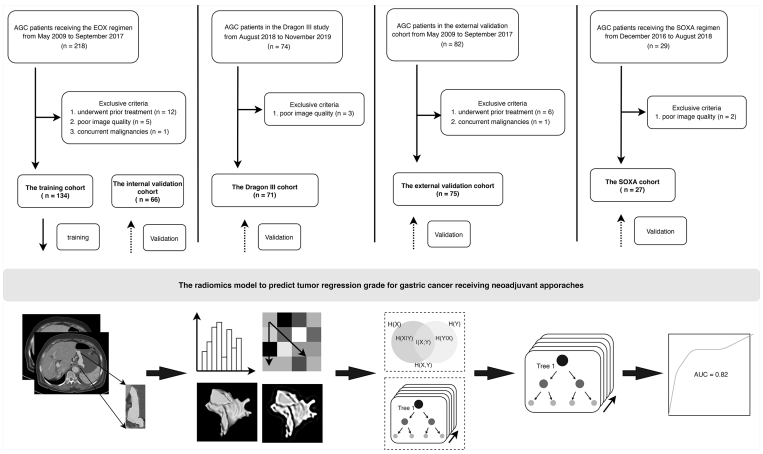


FIGURE 1 | Flowchart and patient enrollment of this study. AGC, advanced gastric cancer. For the regimens, EOX (epirubicin, oxaliplatin, and capecitabine), SOXA (apatinib in combination with S-1 and oxaliplatin).

were 134 and 66 patients in the training and internal validation cohorts, respectively.

The Dragon III study was sought to compare the efficacy between the doublet SOX regimen (S-1 and oxaliplatin) and the triplet FLOT regimen. This clinical trial was conducted from August 2018 to November 2019. Finally, there were 71 patients (women, 24; mean age, 61.3 ± 9.9 years) enrolled in this study, and three patients were excluded due to poor image quality of CECT. Those patients receiving the two regimens were considered as an entity and named as the Dragon III cohort. There were 27 and 44 patients receiving the SOX regimen and the FLOT regimen in the Dragon III cohort, respectively.

Patients in the external validation cohort were from another tertiary referral hospital (Peking University Cancer Hospital and Institute) and were retrospectively recruited from January 2018 to December 2019. In this cohort, three doublet NAC regimens were used, including the SOX regimen, the XELOX regimen (oxaliplatin and capecitabine), and the FOLFOX regimen (oxaliplatin, folinic acid, and fluorouracil). There were 75 AGC patients (women, 24; mean age, 59.4 ± 10.5 years) enrolled in this cohort. There were 42, 30, and 3 patients receiving the SOX, the XELOX, and the FOLFOX regimens, respectively.

Patients in the clinical trial (ChiCTR.gov.cn: ChiCTR-OPC-16010061) were recruited from December 2016 to August 2018. The single-arm, open-label, phase II trial was designed to investigate the added value of apatinib in combination with SOX (the SOXA regimen) to improve the pathologic response of AGC patients. Two patients were excluded due to poor image quality of CECT, and 27 patients (women, 10; mean age, 59.3 ± 8.0 years) were enrolled in this study as the SOXA cohort.

This study was approved by the ethics committee of Ruijin Hospital and Peking University Cancer Hospital and Institute. For patients in the training, internal validation, and external validation cohorts, written informed consent was waived due to the retrospective nature. Baseline clinical data for all patients included gender, age, tumor invasion, lymph node status, tumor location, tumor size (cm), differentiation status, carcinoembryonic antigen (CEA), and carbohydrate antigen 19-9 (CA19-9). Definition for those indices is presented in the **Supplementary Material S1**.

Protocols for Neoadjuvant Regimens

For patients receiving the SOX and FLOT regimens in the DRAGON III study (i.e., the DRAGON III cohort) and patients receiving the SOXA regimen (the SOXA cohort), the protocols are described in the corresponding clinical trials (23, 24). Treatment protocols for patients receiving the EOX regimen and patients in the external validation cohort are presented in **Supplementary Material S2**.

Imaging Protocol

All patients underwent abdomen CECT scanning before treatment. The information and parameters for involved CT scanners are presented in **Supplementary Table S1**. The images on portal-venous and delayed phases were anonymously retrieved for further analysis.

Tumor Regression Grade

Evaluation of tumor response to treatment was in accordance with Ryan criteria (25) after the completion of planned neoadjuvant courses, where completed response represents no viable cancer cells, moderate response represents single cells or small groups of cancer cells, minimal response represents residual cancer outgrown by fibrosis, and poor response represents minimal or no tumor kill (extensive residual cancer). Two pathologists, who had 7 and 10 years of experience, respectively, evaluated tumor response and resolved disputes with consensus. In this study, we considered patients with a completed response or moderate response to neoadjuvant treatment as responders and minimal response or poor response as non-responders.

Tumor Delineation and Feature Extraction

Before segmentation, all images were resampled into a uniform voxel space of 1 * 1 * 1 mm. The volume of interest (VOI) was

performed on the primary tumor by open-source software (3d-slicer, version 4.10.2) on baseline axial images from portal-venous and delayed phases along with tumor boundary slice by slice and omitting the first and last slices to avoid potential partial volume effect. Radiomics features were extracted from an in-house research platform (Syngo Via, Version VB20, Research Frontier, Siemens Healthineer) (26).

A total of 110 original radiomics features (without any preprocessing producers applied) were extracted from each phase. In addition, 1,116 features were calculated after preprocessing for each phase. Therefore, there were 2,452 radiomics features finally extracted for each patient. All features were compliant with the benchmarks of the image biomarker standardization initiative (IBSI) (27). Detailed information is presented in **Supplementary Material S3**.

To guarantee reproducibility of the radiomics features, we performed reliability analysis by repeatedly delineating the tumors by two radiologists (with 10 years and 5 years of experience in abdominal imaging, respectively) on images from portal-venous and delayed phases from 70 randomly selected patients from the training cohort. The two radiologists were aware of GC but unaware of pathological and clinical results. Intraclass correlation coefficient (ICC) analysis was used for reliability analysis, and features with ICC >0.80 were considered robust and remained for further analysis.

Dimensionality Reduction, Radiomics Score Generation, and Modeling

To avoid dimensionality curse and overfitting, mutual information that measures the mutual dependence between features and label was used to select candidate features in the training cohort. Random forest (RF), a widely used ensemble machine learning algorithm, was chosen in our studies for radiomics score generation and model establishment (28,29). RF was generally reported to own the best discriminative and diagnostic power in radiomics studies due to their ability to handle high dimensional features and strong generalization. Grid search with 10-fold cross-validation was implemented for parameter tunings in the training cohort. The contribution of features to modeling was ranked by feature importance calculated by Gini-impurity. Subsequently, the radiomics scores that were generated by the established model for every patient were calculated and included for further analysis. The performance of the radiomics model was then verified in the internal validation cohort.

Generalization Analysis

To investigate the generalization of our radiomics model among different regimens, we further applied the model to predicting TRG for patients in the DRAGON III cohort and the external validation cohort (each included two and three different regimens). Besides, to test whether the model was generalizable to predict treatment response undergoing molecular targeted therapy, we used it to assess TRG on patients in the SOXA cohort.

Model Comparison

For comparison, we evaluated the degree of tumor response on CT images in accordance with the Response Evaluation Criteria in Solid Tumors (RECIST, version 1.1) (30) for patients receiving the EOX regimen and compared its performance with our radiomics model. In addition, the clinical model was built in the training cohort and validated in corresponding cohorts as the radiomics model did by incorporating significant clinical indices into the multivariate logistic regression model after univariate analysis. Furthermore, we also integrated the radiomics scores and significant clinical characteristics into a combined model to investigate whether there was any improvement to predict TRG for AGC patients in different regimens.

Subgroup Analysis

Given the SOX regimen and the FLOT regimen indicating no significant TRG difference in the Dragon III study, which was inconsistent with previous studies, we emphasized the comparison between the SOX regimen and the FLOT regimen in the Dragon III cohort. We were also interested in the added value of apatinib in the SOXA cohort compared with patients receiving the SOX regimen only and compared the difference between the two regimens. For patients receiving the SOX regimen, only those recruited from the randomized controlled trial (RCT) clinical trial (i.e., patients in the Dragon III cohort) were considered to match the single-arm SOXA clinical trial. To avoid the influence of response rate within different cohorts, we further split responders and non-responders in each regimen and independently compared them between the regimens. To simplify the analysis, we analyzed the top 10 features in subgroup analysis according to the importance of ranking.

Statistical Analysis

Continuous variables were compared by independent-samples t-test or Mann–Whitney U test based on their distribution. Categorical variables were compared using χ^2 or Fisher's exact test. For model assessment, discrimination ability included receiver operating characteristic (ROC) curve and area under the curve (AUC), sensitivity, specificity, and accuracy were performed for all models. The optimal threshold was selected based on the Youden index for the radiomics score in the training cohort, and Delong test was implemented for the AUC comparison between models in each cohort. The reliability of those models was evaluated by Cohens' kappa index. An individualized nomogram and a decision curve were implemented to present the clinical utility of our models. The goodness of fit of all models was assessed by calibration curve and Briers score. The radiomics analysis was compliant with published guidance, and we performed a radiomics quality score (RQS) for our study to assess the quality of our study (15). All statistical analyses were performed with software R (version 3.6.0, http://www.r-project.org) and Python Scikit-learn package (version 3.7, Scikit-learn version 0.24.1, https://scikit-learn.org/stable/index.html). Package resources are listed in **Supplementary Table S2**. Statistical significance was considered as a two-sided $p < 0.05$.

RESULTS

Clinical Characteristics

Demographic data in each cohort are presented in **Table 1**. For patients receiving the EOX regimen, there was no response bias

TABLE 1 | Demographic data for all patients in each cohort.

	The training cohort (n = 134)	The internal validation cohort (n = 66)	The Dragon III cohort (n = 71)	The external validation cohort (n = 75)	The SOXA cohort (n = 27)	p value
TRG (response, %)	98 (73.1%)	50 (75.8%)	13 (18.3%)	23 (30.7%)	16 (59.3%)	<0.001
Regimens	EOX	EOX	SOX and FLOT	SOX, XELOX and FOLFOX	apatinib plus SOX	
Age (years)	58.4 ± 10.2	60.3 ± 8.8	61.3 ± 9.9	59.4 ± 10.5	59.3 ± 8.0	0.328
Gender (female, %)	41 (30.6%)	23 (34.8%)	24 (33.8%)	25 (33.3%)	9 (33.3%)	0.977
Tumor invasion						<0.001
cT2	0 (0.0%)	0 (0.0%)	0 (0.0%)	6 (8.0%)	0 (0.0%)	
cT3	0 (0.0%)	0 (0.0%)	0 (0.0%)	23 (28%)	0 (0.0%)	
cT4	134 (100.0%)	66 (100%)	71 (100.0%)	46 (61.4%)	27 (100.0%)	
Lymph node status						<0.001
cN0	0 (0.0%)	0 (0.0%)	4 (5.6%)	3 (4.0%)	0 (0.0%)	
cN1	23 (17.2%)	5 (7.6%)	10 (14.1%)	26 (34.7%)	7 (25.9%)	
cN2	55 (41.0%)	29 (43.9%)	48 (67.6%)	29 (38.7%)	15 (55.6%)	
cN3	56 (41.8%)	32 (48.5%)	9 (12.7%)	17 (22.7%)	5 (18.5%)	
Tumor size (cm)	6.46 ± 2.32	7.16 ± 1.90	5.64 ± 1.96	5.23 ± 1.64	4.83 ± 2.38	<0.001
Tumor location						0.002
Upper	33 (24.6%)	16 (24.2%)	23 (32.4%)	18 (24.0%)	14 (51.9%)	
Middle	38 (28.4%)	18 (27.3%)	4 (5.6%)	15 (20.0%)	3 (11.1%)	
Lower	41 (30.6%)	19 (28.8%)	36 (50.7%)	28 (37.3%)	5 (18.5%)	
Diffuse	22 (16.4%)	13 (19.7%)	9 (11.3%)	14 (18.7%)	5 (18.5%)	
Differentiation status						<0.001
Well differentiated	2 (1.5%)	0 (0%)	0 (0%)	1 (1.3%)	1 (3.7%)	
Moderately differentiated	11 (8.2%)	4 (6.1%)	9 (12.7%)	41 (54.7%)	3 (11.1%)	
Poorly differentiated	78 (58.2%)	35 (53.0%)	57 (80.3%)	33 (44%)	15 (55.6%)	
SRCC	8 (6.0%)	6 (9.1%)	1 (1.4%)	0 (0%)	1 (3.7%)	
Not evaluated	35 (26.1%)	21 (31.8%)	4 (5.6%)	0 (0%)	7 (25.9%)	
CA19-9 (positive, %)	43 (32.1%)	16 (24.2%)	13 (18.3%)	7 (9.3%)	5 (18.5%)	0.005
CEA (positive, %)	39 (29.1%)	24 (36.4%)	18 (25.4%)	12 (16.2%)	5 (18.5%)	0.061

CA19-9, carbohydrate antigen 19-9; CEA, carcinoembryonic antigen; SRCC, signet-ring cell carcinoma; TRG, tumor regression grade. For the regimens, EOX (epirubicin, oxaliplatin, and capecitabine), FOLFOX (oxaliplatin, folinic acid and 5-fluorouracil), FLOT (5-fluorouracil, leucovorin, docetaxel and oxaliplatin), SOX (S-1 and oxaliplatin), SOXA (apatinib in combination with SOX), XELOX (oxaliplatin and capecitabine).

between the training cohort and the internal validation cohort (73.1% vs. 75.8%, p = 0.691). Univariate analysis for differentiation status was not performed, given 28% of cases were not accurately evaluated by a specimen from the endoscopic biopsy. For other clinical indices, only age and tumor size were statistically different between responders and non-responders in the training cohort (p = 0.023 and p = 0.020, respectively). Gender, tumor invasion, lymph node status, tumor location, CA19-9, and CEA showed no statistical difference (all p > 0.05, see **Supplementary Table S3**). Finally, we integrated age and tumor size as clinical risk factors into a clinical model using multivariate logistic regression.

Dimensionality Reduction

The average ICCs for features from the portal-venous and the delayed phase images were 0.75 and 0.80, respectively. There were 668 and 776 radiomics features with a threshold of 0.80 in each phase, and finally, a total of 1,444 radiomics features were considered as robust and remained for further dimensionality reduction.

A total of 741 radiomics features showed non-zero mutual information. Subsequently, 164 radiomics features were selected with a threshold value of 0.05. Finally, 28 features with the highest importance ranking were used for modeling after RF (**Figure 2**). The RQS of our study was 30 (83.3% of ideal score), guaranteeing good quality of the radiomics process (**Supplementary Table S4**).

Modeling

We utilized the 28 features to construct a radiomics model to predict TRG for AGC patients. The radiomics score of the training cohort was 0.600. The model achieved AUCs of 0.82 (95% CI, 0.76~0.90) and 0.77 (95% CI, 0.63~0.91) in the training and internal validation cohorts in patients treated with the EOX regimen. Furthermore, the radiomics model showed generalizability to other NAC regimens because similar performance was observed [for the Dragon III cohort and the external validation cohort, AUCs of 0.78 (95% CI, 0.66~0.89) and 0.72 (95% CI, 0.66~0.89), respectively]. Delong test indicated that no significant difference was found when compared with the training cohort (p = 0.571 and p = 0.216, respectively). However, the radiomics model showed a poor predictive value for TRG in the SOXA cohort [AUC of 0.50 (95% CI, 0.27~0.73) vs. 0.82 in the training cohort, p = 0.010]. Detailed information for the performance of the radiomics model in each cohort is presented in **Table 2**. The radiomics score for patients in each cohort is shown in **Supplementary Figure S1**. The Kappa indices of the radiomics models of the training cohort, the internal validation cohort, the Dragon III cohort, the external validation cohort, and the SOXA cohort were 0.485, 0.365, 0.221, 0.156, and 0.069, respectively.

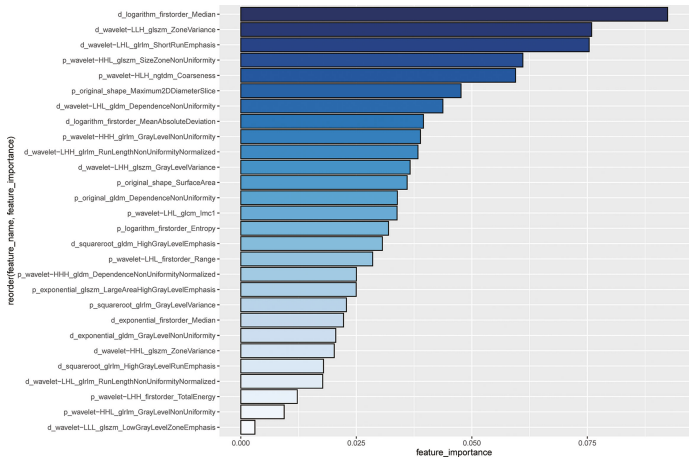


FIGURE 2 | Importance ranking for 28 selected radiomics features using random forest The length of the bin and the depth of the color blue represent the important degree of the radiomics features. The feature name is ordered by the following rule: phase (p or d represents the venous-portal phase or the delayed phase) _ pre-processing_feature category_feature name. For example, for the first feature, i.e., d_logarithm_firstorder_Median indicates that the feature is named as Median from the first-order category, with transformation by logarithm.

Model Comparison

The RECIST 1.1 showed poor performance to predict TRG, which reached AUCs of 0.53 (95% CI, 0.44~0.62) and 0.67 (95% CI, 0.53~0.81) in the training and internal validation cohorts, respectively. The radiomics model in the training cohort was significantly better than the RECIST model ($p < 0.001$), but no statistical difference was found in the internal validation cohort ($p = 0.338$). Detailed information on discriminative metrics for the RECIST models is presented in **Table 2**. The Kappa indices of the RECIST models of the training cohort and the internal validation cohort were 0.057 and 0.316, respectively.

The clinical model, incorporating age and tumor size in multivariate logistic regression, achieved AUCs of 0.69 (95% CI, 0.60~0.79) and 0.59 (95% CI, 0.43~0.74) in the training and internal validation cohorts for patients receiving the EOX regimen. For the Dragon III and external validation cohorts, the performance of the clinical models was 0.60 (95% CI, 0.46~0.74) and 0.68 (95% CI, 0.56~0.81), respectively. For patients receiving NAC regimens, the performance of the radiomics models was better than that of the clinical models to predict TRG in each cohort, although the marginal difference was only observed in the training cohort and the Dragon III cohort ($p = 0.052$ and $p = 0.057$, respectively). On the contrary, the clinical model showed a higher predictive value compared with the radiomics model in the SOXA cohort, with an AUC of 0.57 (95% CI, 0.34~0.81), and no statistical difference was found for the comparison ($p = 0.664$). Detailed information on discriminative metrics for the clinical models is presented in **Table 2**. The Kappa indices of the clinical models of the training cohort, the internal validation cohort, the Dragon III cohort, the external validation cohort, and the SOXA cohort were 0.269, 0.097, 0.145, 0.002, and 0.080, respectively.

The combined models that integrated the radiomics score and clinical risk factors showed limited improvement compared with the radiomics models in each cohort (**Table 2**). The results indicated a similar trend observed for the radiomics models, where a similar predictive power was found for patients receiving NAC regimens, while significantly declined performance was found in the SOXA regimen. When compared with the clinical models, the combined models in the training and internal validation cohorts presented statistically higher predictive value to TRG ($p = 0.003$ and $p = 0.030$, respectively). The ROC curves for all models in each cohort are presented in **Figure 3**. The Kappa indices of the combined models of the training cohort, the internal validation cohort, the Dragon III cohort, the external validation cohort, and the SOXA cohort were 0.521, 0.418, 0.073, 0.002, and 0.119, respectively.

The nomogram indicated that individuals with lower radiomics scores, younger age, and smaller tumor size are prone to respond to NAC treatment; the calibration curve revealed that the radiomics model and the combined model showed better goodness of fit compared with the clinical model because lower Brier scores were observed (0.159 and 0.136 vs. 0.182) (**Supplementary Figure S2**). The DCA curve indicates that the combined model and the radiomics model have higher clinical net benefit than the clinical model (**Figure 4**, **Supplementary Figure S2**).

Subgroup Analysis

We further analyzed the distribution difference of radiomics features in terms of response status between the SOX and FLOT regimens as well as the SOX and SOXA regimens using top 10 importance ranking features. In the Dragon III cohort, response rates for patients receiving the SOX and FLOT regimens were 18.5% and 18.2%, and there were five and eight responders in each regimen, respectively. In the non-responder group, no radiomics features showed a significant difference between the regimens (**Supplementary Figure S3**). However, in the responder group, four radiomics features were statistically different between the two regimens (**Supplementary Figure S4**).

TABLE 2 | Performance of different models in each cohort.

		AUC	Accuracy	Sensitivity	Specificity	PPV	NPV
The training cohort	The radiomics model	0.82 (0.76~0.90)	0.78 (0.70~0.84)	0.79 (0.69~0.86)	0.79 (0.62~0.91)	0.92 (0.83~0.96)	0.56 (0.41~0.70)
	The RECIST model	0.53 (0.44~0.62)	0.60 (0.51~0.68)	0.67 (0.57~0.76)	0.39 (0.24~0.57)	0.75 (0.64~0.83)	0.30 (0.18~0.45)
	The clinical model	0.69 (0.60~0.79)	0.62 (0.53~0.70)	0.55 (0.45~0.65)	0.81 (0.63~0.91)	0.89 (0.77~0.95)	0.40 (0.29~0.52)
	The combined model	0.83 (0.75~0.91)	0.81 (0.74~0.88)	0.88 (0.79~0.93)	0.64 (0.46~0.79)	0.87 (0.78~0.93)	0.66 (0.48~0.80)
The internal validation cohort	The radiomics model	0.77 (0.63~0.91)	0.73 (0.60~0.83)	0.74 (0.59~0.85)	0.69 (0.42~0.88)	0.88 (0.74~0.96)	0.46 (0.26~0.67)
	The RECIST model	0.67 (0.53~0.81)	0.73 (0.60~0.83)	0.78 (0.64~0.88)	0.56 (0.31~0.79)	0.85 (0.71~0.93)	0.45 (0.24~0.68)
	The clinical model	0.59 (0.43~0.74)	0.49 (0.36~0.61)	0.40 (0.27~0.55)	0.75 (0.47~0.92)	0.83 (0.62~0.95)	0.29 (0.16~0.45)
	The combined model	0.78 (0.65~0.92)	0.77 (0.65~0.87)	0.82 (0.68~0.91)	0.63 (0.34~0.84)	0.87 (0.74~0.95)	0.53 (0.30~0.75)
The Dragon III cohort	The radiomics model	0.78 (0.66~0.89)	0.70 (0.58~0.81)	0.54 (0.26 0.80)	0.74 (0.61~0.84)	0.32 (0.15~0.55)	0.88 (0.75~0.95)
	The clinical model	0.60 (0.46~0.74)	0.42 (0.31~0.55)	0.31 (0.10~0.61)	0.45 (0.32~0.58)	0.11 (0.04~0.27)	0.74 (0.56~0.87)
	The combined model	0.71 (0.59~0.84)	0.63 (0.51~0.75)	0.77 (0.46~0.94)	0.60 (0.47~0.73)	0.30 (0.16~0.49)	0.92 (0.78~0.98)
The external validation cohort	The radiomics model	0.72 (0.66~0.89)	0.49 (0.38~0.61)	0.91 (0.71~0.99)	0.31 (0.19~0.45)	0.37 (0.25~0.51)	0.89 (0.64~0.98)
	The clinical model	0.68 (0.56~0.81)	0.60 (0.48~0.71)	0.74 (0.51~0.89)	0.54 (0.40~0.68)	0.42 (0.27~0.58)	0.82 (0.65~0.93)
	The combined model	0.76 (0.64~0.87)	0.40 (0.29~0.52)	0.96 (0.76~0.99)	0.15 (0.07~0.29)	0.33 (0.23~0.46)	0.89 (0.51~0.99)
The SOXA cohort	The radiomics model	0.50 (0.27~0.73)	0.52 (0.32~0.71)	0.44 (0.21~0.69)	0.64 (0.32 0.88)	0.64 (0.32~0.88)	0.44 (0.21~0.69)
	The clinical model	0.57 (0.34~0.81)	0.56 (0.35~0.75)	0.63 (0.36~0.84)	0.46 (0.18~0.75)	0.63 (0.36~0.84)	0.46 (0.18~0.75)
	The combined model	0.51 (0.27~0.72)	0.59 (0.39~0.78)	0.75 (0.47~0.92)	0.36 (0.12~0.68)	0.63 (0.39~0.83)	0.50 (0.17~0.83)

AUC, area under the curve; NPV, negative predictive value; PPV, positive predictive value; RECIST, Response Evaluation Criteria in Solid Tumors. The SOXA cohort is defined as the patients receiving the regimen of apatinib in combination with SOX (S-1 and oxaliplatin).

The response rate for patients receiving the SOXA regimen was 59.3% (16/27). In the non-responder group, 3 of 10 radiomics features showed a significant difference, and they were all from the delayed phase images (**Supplementary Figure S5**). Intriguingly, those features also presented significance between the two regimens in the responder group (**Supplementary Figure S6**). Apart from that, another radiomics features also shared a significantly different distribution.

DISCUSSION

In this study, we developed and validated an AI method using quantitative imaging named radiomics and successfully predicted TRG for AGC patients treated with NAC, which showed generalized power among different NAC regimens by validating its predictive value in the Dragon III cohort and an external validation cohort. More importantly, radiomics noninvasively revealed that NAC regimens and molecular targeted therapy probably shared different pharmacological effects on tumors. This observation was identified by a significantly declined performance of the radiomics model in the SOXA cohort and the results from the subgroup analysis. Our findings are helpful to evaluate the benefit of NAC before treatment and provide the opportunity to timely adjust to an appropriate strategy for patients insensitive to NAC.

Previous studies have demonstrated the predictive value of radiomics in evaluating pathological response after NAC for GC (21, 31, 32). However, few pieces of research reported the generalization of radiomics among different regimens except for a study in non–small cell lung cancer, which has first demonstrated that radiomics could serve as an early indication for different systemic anticancer therapies (19). To verify whether radiomics could exert predictive value to NAC for GC, we first trained and validated our radiomics model in patients treated with the EOX regimen. The radiomics model constructed by selected 28 radiomics features achieved a comparable AUC (0.77) in the internal validation cohort compared with previously reported performance from 0.722 to 0.82 (21, 31, 32). To investigate the generalization of our radiomics model in neoadjuvant agents, we further applied it on the Dragon III cohort and external validation cohort including five different NAC regimens. The radiomics model continued to show its efficacy in the two cohorts, identifying the generalization of our model. For comparison, the commonly used RECIST 1.1 criteria only showed

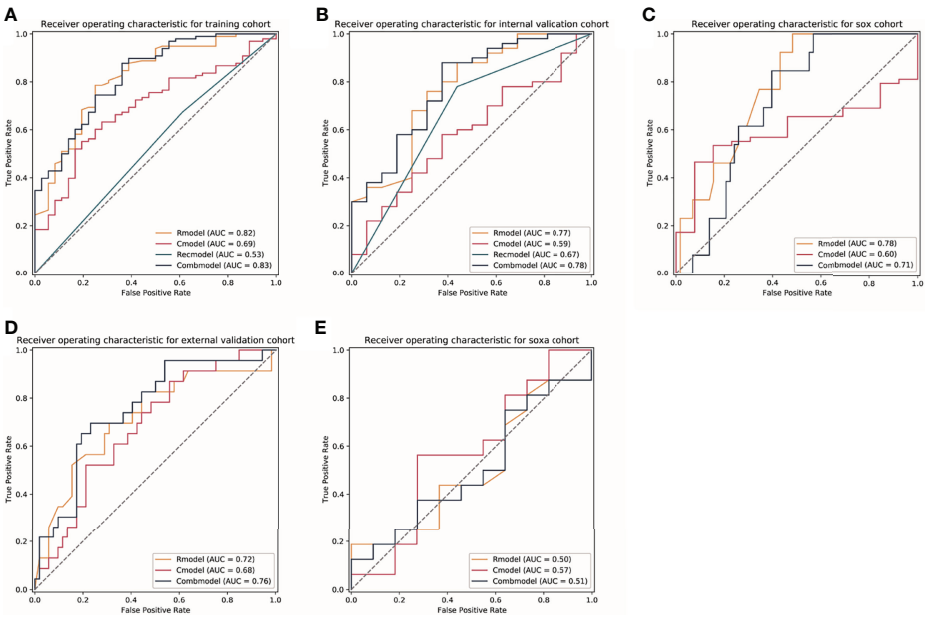


FIGURE 3 | The receiver operator characteristic (ROC) curves for models in each cohort. **(A–E)** Represents ROC curves of models for patients in the training cohort, the internal validation cohort, the Dragon III cohort, the external validation cohort, and the SOXA cohort. Rmodel, the radiomics model; Cmodel, the clinical model; Recmodel, the RECIST model; Combmodel, the combined model. The SOX cohort is defined the patients receiving the regimen of S-1 and oxaliplatin; the SOXA cohort is defined as the patients receiving the regimen of apatinib in combination with SOX (S-1 and oxaliplatin). AUC, area under the curve.

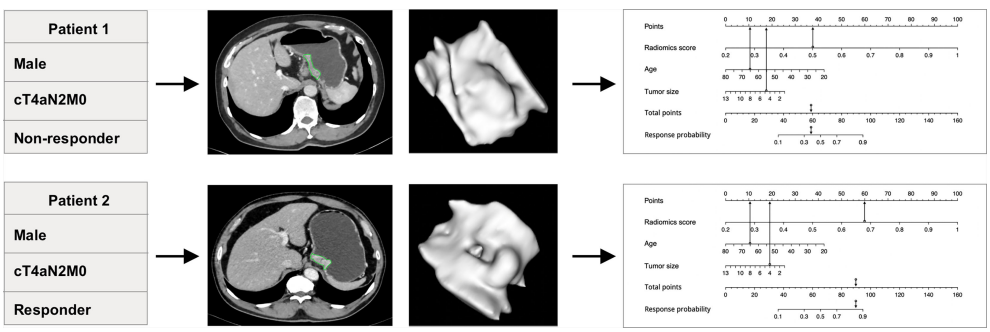


FIGURE 4 | The predictive value of the individualized nomogram to predict tumor regression grade (TRG) for two patients receiving the EOX (epirubicin, oxaliplatin, and capecitabine) regimen. The two patients had similar clinical baseline information but different insensitivity to neoadjuvant chemotherapy treatment (both were men, 65 years old, cT4aN2M0, and similar tumor size). The individualized nomogram integrated radiomics score, age, and tumor size successfully predicted the outcomes of the patients, which mainly relied on the performance of radiomics score.

limited value to predict tumor response to NAC, urging the application of our radiomics model as an early prediction tool for AGC after neoadjuvant treatment. Overall, radiomics shows its potentiality in early prediction of tumor response for GC patients receiving NAC, and more importantly, this efficacy is generalizable among different NAC regimens.

The FLOT regimen is a new candidate NAC approach for AGC patients. Intriguingly, in the Dragon III study, no significantly improved response rate was observed between the FLOT and SOX regimens (23). We also performed a subgroup analysis between the two regimens to investigate whether the radiomics features could provide evidence for the finding in the clinical trial. No features were found statistically different in the non-responders between the two groups, which was consistent with what was observed in the Dragon III study. However, four radiomics features demonstrated a significant difference in the responders between the two groups. Due to the low response rate observed for the two regimens (18.2% and 18.5% for FLOT and SOX, respectively), the sample size in the responder subgroup for each regimen was small, which probably weakened the efficacy of the test. Therefore, whether there was any tumor heterogeneity for patients who were sensitive to the regimens remained to be elucidated.

Unlike indistinctive cytotoxicity from conventional chemotherapy drugs, apatinib was an antiangiogenic agent, which highly selectively targets vascular endothelial growth factor receptor 2 (VEGFR-2) tyrosine kinase, mediating the growth of the tumor by inhibiting angiogenesis (33). Apatinib alone exerts its antitumor effects by inhibiting tumor growth, reducing microvascular density, and enhancing apoptosis (34). This molecular targeted drug was considered as third- or subsequent-line therapy for chemotherapy-refractory advanced or metastatic GC in phase II and III studies, with good tolerance and improved survival (35, 36). The synergistic antineoplastic effects of apatinib combined with chemotherapy agents were observed in preclinical studies (37, 38), which was demonstrated in the clinical practice setting, in combination with preoperative SOX regimen to improve the pathological response of AGC (24). However, the intrinsic changes in tumor heterogeneity initiated by apatinib have not been clarified. In this study, the radiomics model constructed by the NAC regimen performed poorly in the SOXA cohort, revealing the existence of the underlying difference between the two different neoadjuvant approaches. We then performed a subgroup analysis between patients receiving the SOX regimen and the SOXA regimen to investigate the added value of apatinib. Due to the SOXA cohort being a single-arm clinical trial, patients receiving the SOX regimens were from the Dragon III study. Three key radiomics features simultaneously expressed differences in both the responder and non-responder groups, indicating that the discriminative tumor heterogeneity after the addition of apatinib could be detected early by quantitative imaging. Meanwhile, the clinical model maintained a similar performance regardless of which neoadjuvant approach was taken, laterally highlighting the importance of our radiomics model for a discriminative capability for NAC and molecular targeted therapy. Those findings suggested that our radiomics model is helpful to screen patients insensitive to NAC regimens before treatment and thus provide the opportunity for those patients to choose appropriate treatment such as molecular targeted therapy or immunotherapy.

Our study achieved an RQS of 30 (83.3% of ideal score), which is higher than most radiomics studies (reviews report an average RQS percentage ranging from 9.4% to 26.1%) (39–42), guaranteeing the robustness and reproducibility of our study. The calibration curve showed well goodness of fit of the radiomics model, indicating the robustness of the results. The DCA curve revealed that our radiomics model presented higher net benefit in almost the whole period compared with the clinical model, increasing the clinical use of radiomics in providing advice for patients if they should receive NAC treatment.

Our study has several limitations. Except for data from two clinical trials, patients receiving the EOX regimen in the external validation cohort were retrospectively recruited, and bias selection was inevitable in this proof-of-concept study. Also, a potential confounder is that we had limited information in molecular or genetic aspects for those patients, and the diversity among patients may lead to different responses to neoadjuvant therapies. The response rate in each cohort varied, which conformed to the complex condition in the real world. Furthermore, except for the EOX regimen, limited sample size in other NAC regimens hindered the construction of models for each cohort or NAC regimen to investigate the similarity of radiomics features expressed in different NAC regimens. For the same reason, an independent radiomics model for the SOXA cohort was also not realized. Therefore, we could conclude that radiomics was capable of predicting different pharmacological effects on tumors between NAC regimens and molecular targeted therapy, but the nuances that reflected radiomics features between the two different neoadjuvant approaches remained to be further elucidated. Although TRG is reported as a surrogate for prognosis, the predictive value of our radiomics model needs to be elucidated in further investigation with follow-up evidence.

In conclusion, in this study, we demonstrated that our radiomics model is generalizable to predict TRG for AGC patients receiving NAC treatments, which is beneficial to transform appropriate treatment for GC, especially for those insensitive to NAC. We also identified the intrinsic difference between cytotoxic neoadjuvant chemo agents and molecular targeted therapy from a radiomics perspective, which is helpful to construct individualized models for different neoadjuvant approaches.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**. Further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of Ruijin Hospital and Peking University Cancer Hospital and Institute. The ethics committee waived the requirement of written informed consent for participation. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

CL, ZZ and QM contributed to the conception of the study and design of the study. YC, WX and Y-LL performed the data collection and validation. YC wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was funded by Shanghai Science and Technology Commission Science and Technology Innovation Action Clinical Innovation Field (18411953000), the National Natural Science Foundation of China (81771789), and Beijing Natural Science Foundation (Z180001).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.758863/full#supplementary-material

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492
2. Wagner AD, Syn NL, Moehler M, Grothe W, Yong WP, Tai BC, et al. Chemotherapy for Advanced Gastric Cancer. *Cochrane Database Syst Rev* (2017) 8:CD004064. doi: 10.1002/14651858.CD004064.pub4
3. Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H. Gastric Cancer. *Lancet* (2016) 388:2654–64. doi: 10.1016/S0140-6736(16)30354-3
4. Ychou M, Boige V, Pignon JP, Conroy T, Bouche O, Lebreton, et al. Perioperative Chemotherapy Compared With Surgery Alone for Resectable Gastroesophageal Adenocarcinoma: An FNCLCC and FFCD Multicenter Phase III Trial. *J Clin Oncol* (2011) 29:1715–21. doi: 10.1200/JCO.2010.33.0597
5. Cunningham D, Allum WH, Stenning SP, Thompson JN, Van De Velde CJH, Nicolson M, et al. Perioperative Chemotherapy Versus Surgery Alone for Resectable Gastroesophageal Cancer. *N Engl J Med* (2006) 355:11–20. doi: 10.1056/NEJMoa055531
6. Al-Batran S-E, Hofheinz RD, Pauligk C, Kopp H-G, Haag GM, Luley KB, et al. Histopathological Regression After Neoadjuvant Docetaxel, Oxaliplatin, Fluorouracil, and Leucovorin Versus Epirubicin, Cisplatin, and Fluorouracil or Capecitabine in Patients With Resectable Gastric or Gastro-Oesophageal Junction Adenocarcinoma (FLOT4-AIO): Results From the Phase 2 Part of a Multicentre, Open-Label, Randomised Phase 2/3 Trial. *Lancet Oncol* (2016) 17:1697–708. doi: 10.1016/S1470-2045(16)30531-9
7. Alderson D, Cunningham D, Nankivell M, Blazeby JM, Griffin SM, Crellin A, et al. Neoadjuvant Cisplatin and Fluorouracil Versus Epirubicin, Cisplatin, and Capecitabine Followed by Resection in Patients With Oesophageal Adenocarcinoma (UK MRC OE05): An Open-Label, Randomised Phase 3 Trial. *Lancet Oncol* (2017) 18:1249–60. doi: 10.1016/S1470-2045(17)30447-3
8. Eto K, Hiki N, Kumagai K, Shoji Y, Tsuda Y, Kano Y, et al. Prophylactic Effect of Neoadjuvant Chemotherapy in Gastric Cancer Patients With Postoperative Complications. *Gastric Cancer* (2018) 21:703–9. doi: 10.1007/s10120-017-0781-y
9. Al-Batran S-E, Homann N, Pauligk C, Goetze TO, Meiler J, Kasper S, et al. Perioperative Chemotherapy With Fluorouracil Plus Leucovorin, Oxaliplatin, and Docetaxel Versus Fluorouracil or Capecitabine Plus Cisplatin and Epirubicin for Locally Advanced, Resectable Gastric or Gastro-Oesophageal Junction Adenocarcinoma (FLOT4): A Randomised, Phase 2/3 Trial. *Lancet* (2019) 393:1948–57. doi: 10.1016/S0140-6736(18)32557-1
10. Smyth EC, Verheij M, Allum W, Cunningham D, Cervantes A, Arnold D, et al. Gastric Cancer: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann Oncol* (2016) 27:v38–49. doi: 10.1093/annonc/mdw350
11. Wang FH, Shen L, Li J, Zhou ZW, Liang H, Zhang XT, et al. The Chinese Society of Clinical Oncology (CSCO): Clinical Guidelines for the Diagnosis and Treatment of Gastric Cancer. *Cancer Commun (Lond)* (2019) 39:10. doi: 10.1002/cac2.12193
12. Petrillo A, Pompella L, Tirino G, Pappalardo A, Laterza MM, Caterino M, et al. Perioperative Treatment in Resectable Gastric Cancer: Current Perspectives and Future Directions. *Cancers (Basel)* (2019) 11:399. doi: 10.3390/cancers11030399
13. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
15. Lambin P, Leijenaar RTH, Deist TM, Peerlings J, De Jong EEC, Van Timmeren J, et al. Radiomics: The Bridge Between Medical Imaging and Personalized Medicine. *Nat Rev Clin Oncol* (2017) 14:749–62. doi: 10.1038/nrclinonc.2017.141
16. Aslan K, Turco V, Blobner J, Sonner JK, Liuzzi AR, Nunez NG, et al. Heterogeneity of Response to Immune Checkpoint Blockade in Hypermutated Experimental Gliomas. *Nat Commun* (2020) 11:931. doi: 10.1038/s41467-020-14642-0
17. Devkota L, Starosolski Z, Rivas CH, Stupin I, Annapragada A, Ghaghada KB, et al. Detection of Response to Tumor Microenvironment-Targeted Cellular Immunotherapy Using Nano-Radiomics. *Sci Adv* (2020) 6:eaba6156. doi: 10.1126/sciadv.aba6156
18. Jiang Y, Wang H, Wu J, Chen C, Yuan Q, Huang W, et al. Noninvasive Imaging Evaluation of Tumor Immune Microenvironment to Predict Outcomes in Gastric Cancer. *Ann Oncol* (2020) 31:760–8. doi: 10.1016/j.annonc.2020.03.295
19. Dercle L, Fronheiser M, Lu L, Du S, Hayes W, Leung DK, et al. Identification of Non-Small Cell Lung Cancer Sensitive to Systemic Cancer Therapies Using Radiomics. *Clin Cancer Res* (2020) 26:2151–62. doi: 10.1158/1078-0432.CCR-19-2942
20. Jiang Y, Yuan Q, Lv W, Xi S, Huang W, Sun Z, et al. Radiomic Signature of (18)F Fluorodeoxyglucose PET/CT for Prediction of Gastric Cancer Survival and Chemotherapeutic Benefits. *Theranostics* (2018) 8:5915–28. doi: 10.7150/thno.28018
21. Li Z, Zhang D, Dai Y, Dong J, Wu L, Li Y, et al. Computed Tomography-Based Radiomics for Prediction of Neoadjuvant Chemotherapy Outcomes in Locally Advanced Gastric Cancer: A Pilot Study. *Chin J Cancer Res* (2018) 30:406–14. doi: 10.21147/j.issn.1000-9604.2018.04.03
22. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, 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. *Lancet Oncol* (2018) 19:1180–91. doi: 10.1016/S1470-2045(18)30413-3
23. Sah BK, Zhang B, Zhang H, Li J, Yuan F, Ma T, et al. Neoadjuvant FLOT Versus SOX Phase II Randomized Clinical Trial for Patients With Locally Advanced Gastric Cancer. *Nat Commun* (2020) 11:6093. doi: 10.1038/s41467-020-19965-6
24. Zheng Y, Yang X, Yan C, Feng R, Sah BK, Yang Z, et al. Effect of Apatinib Plus Neoadjuvant Chemotherapy Followed by Resection on Pathologic Response in Patients With Locally Advanced Gastric Adenocarcinoma: A Single-Arm, Open-Label, Phase II Trial. *Eur J Cancer* (2020) 130:12–9. doi: 10.1016/j.ejca.2020.02.013
25. Ryan R, Gibbons D, Hyland JM, Treanor D, White A, Mulcahy HE, et al. Pathological Response Following Long-Course Neoadjuvant Chemoradiotherapy for Locally Advanced Rectal Cancer. *Histopathology* (2005) 47:141–6. doi: 10.1111/j.1365-2559.2005.02176.x
26. Wels M, Lades F, M̈hlberg A, Suehling M. General Purpose Radiomics for Multi-Modal Clinical Research. *Proc SPIE* (2019) 1095046. doi: 10.1117/12.2511856
27. Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-Based Phenotyping. *Radiology* (2020) 295:328–38. doi: 10.1148/radiol.2020191145
28. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts H. Machine Learning Methods for Quantitative Radiomic Biomarkers. *Sci Rep* (2015) 5:13087. doi: 10.1038/srep13087
29. Zhang Y, Oikonomou A, Wong A, Haider MA, Khalvati F. Radiomics-Based Prognosis Analysis for Non-Small Cell Lung Cancer. *Sci Rep* (2017) 7:46349. doi: 10.1038/srep46349
30. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1). *Eur J Cancer* (2009) 45:228–47. doi: 10.1016/j.ejca.2008.10.026
31. Sun KY, Hu HT, Chen SL, Ye JN, Li GH, Chen LD, et al. CT-Based Radiomics Scores Predict Response to Neoadjuvant Chemotherapy and Survival in Patients With Gastric Cancer. *BMC Cancer* (2020) 20:468. doi: 10.1186/s12885-020-06970-7
32. Giganti F, Marra P, Ambrosi A, Salerno A, Antunes S, Chiari D, et al. Pre-Treatment MDCT-Based Texture Analysis for Therapy Response Prediction in Gastric Cancer: Comparison With Tumour Regression Grade at Final Histology. *Eur J Radiol* (2017) 90:129–37. doi: 10.1016/j.ejrad.2017.02.043
33. Scott LJ. Apatinib: A Review in Advanced Gastric Cancer and Other Advanced Cancers. *Drugs* (2018) 78:747–58. doi: 10.1007/s40265-018-0903-9
34. Peng QX, Han YW, Zhang YL, Hu J, Fan J, Fu SZ, et al. Apatinib Inhibits VEGFR-2 and Angiogenesis in an *In Vivo* Murine Model of Nasopharyngeal Carcinoma. *Oncotarget* (2017) 8:52813–22. doi: 10.18632/oncotarget.17264

35. Li J, Qin S, Xu J, Guo W, Xiong J, Bai Y, et al. Apatinib for Chemotherapy-Refractory Advanced Metastatic Gastric Cancer: Results From a Randomized, Placebo-Controlled, Parallel-Arm, Phase II Trial. *J Clin Oncol* (2013) 31:3219–25. doi: 10.1200/JCO.2013.48.8585
36. Li J, Qin S, Xu J, Xiong J, Wu C, Bai Y, et al. Randomized, Double-Blind, Placebo-Controlled Phase III Trial of Apatinib in Patients With Chemotherapy-Refractory Advanced or Metastatic Adenocarcinoma of the Stomach or Gastroesophageal Junction. *J Clin Oncol* (2016) 34:1448–54. doi: 10.1200/JCO.2015.63.5995
37. Tian S, Quan H, Xie C, Guo H, Lu F, Xu Y, et al. YN968D1 Is a Novel and Selective Inhibitor of Vascular Endothelial Growth Factor Receptor-2 Tyrosine Kinase With Potent Activity *In Vitro* and *In Vivo*. *Cancer Sci* (2011) 102:1374–80. doi: 10.1111/j.1349-7006.2011.01939.x
38. Li F, Zhu T, Cao B, Wang J, Liang L. Apatinib Enhances Antitumour Activity of EGFR-TKIs in Non-Small Cell Lung Cancer With EGFR-TKI Resistance. *Eur J Cancer* (2017) 84:184–92. doi: 10.1016/j.ejca.2017.07.037
39. Park JE, Kim D, Kim HS, Park SY, Kim JY, Cho SJ, et al. Quality of Science and Reporting of Radiomics in Oncologic Studies: Room for Improvement According to Radiomics Quality Score and TRIPOD Statement. *Eur Radiol* (2020) 30:523–36. doi: 10.1007/s00330-019-06360-z
40. Stanzione A, Gambardella M, Cuocolo R, Ponsiglione A, Romeo V, Imbriaco M. Prostate MRI Radiomics: A Systematic Review and Radiomic Quality Score Assessment. *Eur J Radiol* (2020) 129:109095. doi: 10.1016/j.ejrad.2020.109095
41. Ursprung S, Beer L, Bruining A, Woitek R, Stewart GD, Gallagher, et al. Radiomics of Computed Tomography and Magnetic Resonance Imaging in Renal Cell Carcinoma-A Systematic Review and Meta-Analysis. *Eur Radiol* (2020) 30:3558–66. doi: 10.1007/s00330-020-06666-3
42. Wang H, Zhou Y, Li L, Hou W, Ma X, Tian R. Current Status and Quality of Radiomics Studies in Lymphoma: A Systematic Review. *Eur Radiol* (2020) 30:6228–40. doi: 10.1007/s00330-020-06927-1

Conflict of Interest: ZX is employed by Siemens Healthineers Ltd. MW is employed by Siemens Healthcare GmbH.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Chen, Xu, Li, Liu, Sah, Wang, Xu, Wels, Zheng, Yan, Zhang, Ma, Zhu and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers in Oncology

ORIGINAL RESEARCH
published: 15 March 2022
doi: 10.3389/fonc.2022.787489

Response Prediction to Concurrent Chemoradiotherapy in Esophageal Squamous Cell Carcinoma Using Delta-Radiomics Based on Sequential Whole-Tumor ADC Map

*Dianzheng An[1], Qiang Cao[2], Na Su[3], Wanhu Li[3], Zhe Li[2], Yanxiao Liu[2], Yuxing Zhang[4] and Baosheng Li[2]**

[1] Department of Radiation Oncology, Shandong Cancer Hospital Affiliated to Shandong University, Shandong University, Jinan, China, [2] Department of Radiation Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China, [3] Department of Radiology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China, [4] Department of Imaging, Anyang Tumor Hospital, The Fourth Affiliated Hospital of Henan University of Science and Technology, Anyang, China

OPEN ACCESS

Edited by:
Xujiong Ye, University of Lincoln, United Kingdom

Reviewed by:
Mathieu Hatt, Institut National de la Santé et de la Recherche Médicale (INSERM), France
Kun Qing, City of Hope National Medical Center, United States

***Correspondence:**
Baosheng Li
bsli@sdfmu.edu.cn

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *30 September 2021*
Accepted: *21 February 2022*
Published: *15 March 2022*

Citation:
An D, Cao Q, Su N, Li W, Li Z, Liu Y, Zhang Y and Li B (2022) Response Prediction to Concurrent Chemoradiotherapy in Esophageal Squamous Cell Carcinoma Using Delta-Radiomics Based on Sequential Whole-Tumor ADC Map. Front. Oncol. 12:787489. doi: 10.3389/fonc.2022.787489

Purpose: The purpose of this study was to investigate the association between the radiomics features (RFs) extracted from a whole-tumor ADC map during the early treatment course and response to concurrent chemoradiotherapy (cCRT) in patients with esophageal squamous cell carcinoma (ESCC).

Methods: Patients with ESCC who received concurrent chemoradiotherapy were enrolled in two hospitals. Whole-tumor ADC values and RFs were extracted from sequential ADC maps before treatment, after the 5th radiation, and after the 10th radiation, and the changes of ADC values and RFs were calculated as the relative difference between different time points. RFs were selected and further imported to a support vector machine classifier for building a radiomics signature. Radiomics signatures were obtained from both RFs extracted from pretreatment images and three sets of delta-RFs. Prediction models for different responders based on clinical characteristics and radiomics signatures were built up with logistic regression.

Results: Patients (n=76) from hospital 1 were randomly assigned to training (n=53) and internal testing set (n=23) in a ratio of 7 to 3. In addition, to further test the performance of the model, data from another institute (n=17) were assigned to the external testing set. Neither ADC values nor delta-ADC values were correlated with treatment response in the three sets. It showed a predictive effect to treatment response that the AUC values of the radiomics signature built from delta-RFs over the first 2 weeks were 0.824, 0.744, and 0.742 in the training, the internal testing, and the external testing set, respectively. Compared with the evaluated response, the performance of response prediction in the internal testing set was acceptable (p = 0.048).

Conclusions: The ADC map-based delta-RFs during the early course of treatment were effective to predict the response to cCRT in patients with ESCC.

Keywords: esophageal squamous cell carcinoma, concurrent chemoradiotherapy, diffusion-weighted imaging, radiomics, treatment response

INTRODUCTION

Worldwide, esophageal cancer ranks 7th in cancer incidence and 6th in mortality rate reported by the World Health Organization (1). Esophageal squamous cell carcinoma (ESCC) is one of the top ten characteristic tumors in China with a proportion over 90% in all esophageal cancer patients, while adenocarcinoma is dominant in the West (2). The 5-year survival rate for locally advanced ESCC is limited to 15-30%, and the prognosis varies from patients even with the same treatment (3, 4). Though concurrent chemoradiotherapy (cCRT) has become the prevalent treatment for local advanced ESCC, individual differences for therapeutic sensitivity still exist (5). It is challenging to accurately predict the sensitivity to cCRT and distinguish the discrepancy of ESCC to assist clinical decisions.

Response Evaluation Criteria in Solid Tumors (RECIST, version1.1) is the most widely used tool for assessing solid tumor response to nonsurgical treatment. It relies on imaging techniques, such as X-ray and computed tomography (CT), providing little information about the molecule, cell, histopathology, and biology (6). Compared to these imaging modalities, magnetic resonance imaging (MRI) is able to precisely depict the histopathological layers of the esophageal wall in an *ex vivo* evaluation (7). Notably, diffusion-weighted imaging (DWI) with physiological information reflecting water diffusion properties of tissues is potentially practical to monitor tumor response. The apparent diffusion coefficient (ADC) and the change of ADC (delta-ADC/ΔADC) generated from DW-MRI showed a correlation with tumor response to treatment (8–10). Nevertheless, the utilization of the ADC value as an imaging biomarker is still controversial (11–13).

With the development of computerized image processing technology during the past decades, quantitative imaging analysis is becoming more and more popular in radiology (14). Radiomics, which extracts numeric radiologic data from medical imaging, quantitatively describes the shape, intensity, texture, and other features of target structures (15). It has been reported that various radiomics features (RFs) are associated with tumor genes, pathology, and outcome of treatment in different tumors (16–20). Furthermore, the change of RFs between pretreatment images and images of other time points during or after treatment, referred to as delta-radiomics, has been investigated in lung cancer and rectal cancer as promising prognostic factors (21, 22). Theoretically, delta-RFs (ΔRFs) reflect much detailed information of changes induced by chemotherapy, radiation, and immunotherapy throughout treatment rather than just at pretreatment.

It is assumed that radiomics analysis of the ADC map is helpful to predict early treatment response to cCRT of ESCC patients. Aiming to predict response to concurrent CRT in patients with ESCC, this study developed radiomics models based on RFs and ΔRFs during early treatment. In addition, the prediction value of ADC and ΔADC values were also analyzed.

MATERIALS AND METHODS

Study Population

All protocols of this prospective and observational study were approved by the Institutional Review Board (Project ID: 201404005). Patients from Shandong Cancer Hospital and Institute were randomly assigned to the training set and the internal testing set in a ratio of 7:3 to ensure an adequate sample size. The external testing set was enrolled in Anyang Tumor Hospital.

All methods were carried out following relevant guidelines and regulations. All patients provided written informed consent before enrollment. Inclusion criteria were as follows: (a) ESCC histologically proven by endoscopic biopsy; (b) no contraindications to MR examination; (c) clinical stage T3 or T4 defined on endoscopic ultrasonography, diagnostic CT scans, or ^{18}fluorodeoxyglucose positron emission tomography/computed tomography (^{18}F-FDG PET/CT) scan; (d) no prior anticancer treatment; (e) no distant metastasis except lymph nodes; (f) no other primary tumor. Clinical staging was based on the tumor-node-metastasis classification of malignant tumors (UICC, Version 8th). All patients were treated with concurrent CRT. The daily fractional dose of radiotherapy was 1.8–2.0 Gy, administered 5 days a week, and the total dose was 50.4–66.0 Gy with 6–15 MV X-rays performed by intensity-modulated radiation therapy. Before the early treatment response evaluated, all patients accepted two cycles of chemotherapy concurrently during radiotherapy every 3 weeks. One of the following proposals of chemotherapy would be adopted: 1) an intravenous injection of paclitaxel (150 mg/m^2) on day 1 and cisplatin (75 mg/m^2) on days 1–3; and 2) S-1 administered orally twice daily at 80 mg/m^2 for 2 weeks and cisplatin (75 mg/m^2) on days 1–3.

Response Assessment

The response was assessed according to RECIST 1.1 (6). Upper gastrointestinal endoscopy, chest and abdomen CT, or whole-body ^{18}F-FDG PET/CT was performed 2–3 months after treatment finished to compare with the pretreatment images. The definition of chemoradiotherapy sensitivity was as follows: the sensitive group comprised patients with complete response (CR) and partial response (PR); the resistant group consisted of patients with stable disease (SD) and progression disease (PD).

Sequential MRI Acquisition

MRI examination with DWI and T2-weighted imaging (T2WI) in the axial orientation was sequentially performed at about 1–3 days before treatment, 5th fraction of radiation completed (5f), and 10th fraction of radiation completed (10f) as shown in **Figure 1A**. All patients were scanned using a 3.0 T MRI system (Achieva, Philips Medical Systems, Netherlands; Ingenia, Philips Medical Systems, Netherlands) with body phased-array coil anterior and spine array coil posterior. T2WI was performed by using a two-dimensional fluid-attenuated inversion recovery sequence (repetition time/echo time, 1,277 ms/70 ms; flip angle, 140°; bandwidth, 315 kHz; section thickness, 4 mm; slice gap, 1 mm; field of view, 350 × 350mm; matrix, 384 × 276; no. of slices, 24–36; imaging time, 5 minutes). DWI was performed by using a single-shot echo-planar imaging (repetition time/echo time, 1,244.7 ms/47.9 ms; flip angle, 140°; bandwidth, 260 kHz; section thickness, 4 mm; slice gap, 1 mm; field of view, 360 × 360 to 400 × 400 mm; matrix, 128 × 128; no. of slices, 24–36; imaging time, 5 minutes). The number of signal averaged (NSA) was 1. Diffusion gradients were applied in three orthogonal directions by different diffusion weightings (*b*-values) of 0, 300, and 600 s/mm^2. Especially, b = 600 s/mm^2 was assigned a NSA of 4 to improve the signal-to-noise ratio. Respiratory triggering was used.

The ADC map was generated automatically at the MR workspace (Philips Medical Systems Extended MR workspace, Netherlands) from DWI, which was overlayed and averaged in three directions with the *b*-value of 0 and 600 s/mm^2.

ADC Value and Extraction of Original RFs

The location of the primary tumor was firstly identified on DWI (b = 600 s/mm^2) as areas of high signal (**Figure 2**). Although it was not outlined on DWI, it assisted in accurately finding the corresponding location of the tumor in the ADC map. Regions of interest (ROIs) were manually drawn on the ADC map *via* a segmentation editor in the 3D Slicer software (version 4.10.0, http://www.slicer.org) (23). The definition of ROI was the region of the tumor that was of low signal in the ADC map but excluding the lumen. Every slice containing the tumor was outlined. These ROIs constituted a volume of interest (VOI) with the entire tumor visualized on the ADC map as showed in **Figure 2**. The mean tumor ADC value was calculated by averaging the measured ADC values of the VOI. For whole-tumor radiomics feature, the Slicer Radiomics, which was based

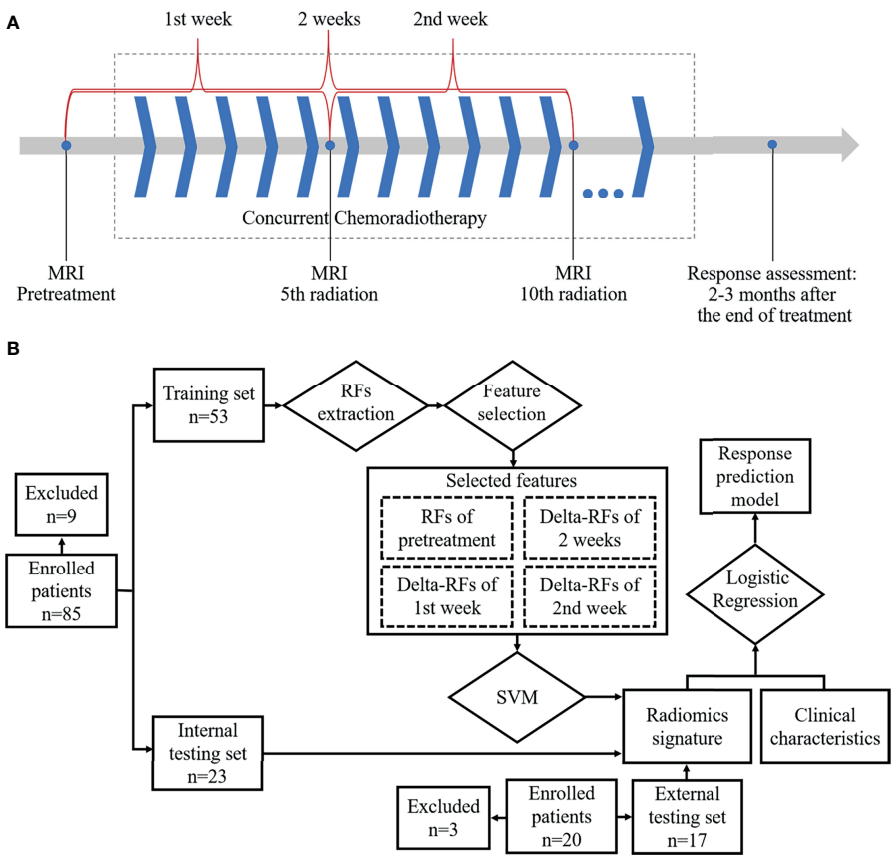


FIGURE 1 | Workflow: **(A)** Magnetic resonance imaging acquisition schema and time ranges of the change of radiomics features. **(B)** Workflow for building radiomics signature-based prediction model.

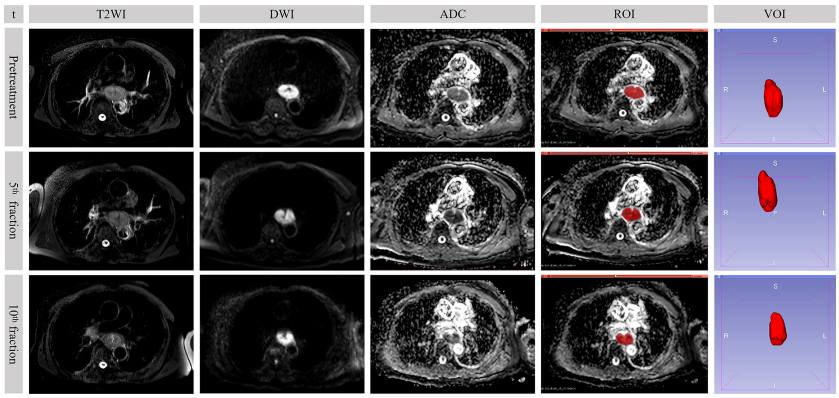


FIGURE 2 | Sequential MR images in a 73-year-old male who underwent concurrent chemoradiotherapy with partial response. The value of *b*-factor of DWI in the figure was 600 s/mm^2. t, time point; DWI, diffusion-weighted imaging; ADC, apparent diffusion coefficient; ROI, region of interest; VOI, volume of interest.

on PyRadiomics program (Revision 2.2.0) in the 3D Slicer software, automatically extracted whole-tumor RFs from the three-dimensional reconstructed VOI (24). In our practice, resampled voxel size, LoG kernel sizes, and bin width were respectively set to 1, 1, and 25. In total, 851 RFs were extracted including 18 first-order statistics features, 14 shape-based (2D and 3D) features, 24 gray-level co-occurrence matrix (GLCM) features, 16 gray-level run length matrix (GLRLM) features, 16 gray-level size zone matrix (GLSZM) features, 5 neighboring gray tone difference matrix (NGTDM) features, 14 gray-level dependence matrix (GLDM) features, and 744 wavelet-based features. Detailed classification of the extracted RFs was summarized in **Supplementary Table 1**. In addition, the RFs mentioned in our study were consistent with the definition of the image biomarker standardization initiative (25).

RF Selection

Selection procedures were performed on the pretreatment images with three steps.

Because DWI is susceptible to field inhomogeneities, large amounts of magnetization artifacts in the chest area, and patient movement, two consecutive DWI acquisitions of the same patient may give images with slightly different imaging characteristics, which may affect radiomics analysis. Thus, the first is to evaluate the repeatability of the radiomics features between two ADC maps in a short time by the intraclass correlation coefficient (ICC) method. Five extra patients with ESCC were enrolled before the first treatment in Shandong Cancer Hospital and Institute (**Supplementary Table 2**) who were not included in the training and internal testing sets. The interval between the two consecutive DWI scans was 5 minutes (**Supplementary Figure 1**). VOIs were delineated in the ADC map by an experienced radiologist (WL, with 20 years of experience in MRI), then the VOI was duplicated to the other ADC map from the same patient. The reliability coefficient of RFs was supposed to be higher than 0.75.

Secondly, to select RFs with high interobserver reproducibility, VOIs were delineated by two experienced radiologists (WL, appointed as reader 1; NS, appointed as reader 2 with 11 years of experience in MRI) who were blind to clinical information, treatment plan, and response about patients. Bland–Altman analysis was used to test the interobserver reproducibility of the RFs from the VOI delineated by different radiologists. The differences between the same parameters extracted from two readers' delineation were plotted against their average and reported as a percentage. The lower and upper reproducibility limits were calculated as ± 1.96 standard deviations. In our case, the mean and standard deviation of the differences were supposed to be less than 5% and 10%, respectively, as stated in related work (26, 27).

Thirdly, to reduce the number of features and select the most significant features correlated with treatment response, the minimum redundancy maximum relevance (mRMR) algorithm was used to identify and rank the top 30 features. Finally, the selected RFs were applied in the training, internal, and external testing sets from the delineation of reader 1.

Change of the Mean of ADC Values and RFs

The changes of the whole-tumor ADC value between images of different time points were respectively calculated as follows:

$$\Delta ADC_{1st\ week} = \frac{ADC_{5f} - ADC_{pretreatment}}{ADC_{pretreatment}} \quad \text{(Eq.1)}$$

$$\Delta ADC_{2nd\ week} = \frac{ADC_{10f} - ADC_{5f}}{ADC_{5f}} \quad \text{(Eq.2)}$$

$$\Delta ADC_{2\ weeks} = \frac{ADC_{10f} - ADC_{pretreatment}}{ADC_{pretreatment}} \quad \text{(Eq.3)}$$

where $\Delta ADC_{1st\ week}$, $\Delta ADC_{2nd\ week}$, and $\Delta ADC_{2\ weeks}$ denoted the change of the ADC value within the 1st week, the 2nd week, and the first two weeks, respectively. Similarly, $ADC_{pretreatment}$, ADC_{5f}, and ADC_{10f} denoted the ADC value of images acquired

before treatment, after the 5th radiation, and after the 10th radiation, respectively.

The changes of the selected RFs were defined as the relative difference between each measurement as follows:

$$\Delta RF_{1st\ week} = \frac{RF_{5f} - RF_{pretreatment}}{RF_{pretreatment}} \quad \text{(Eq.4)}$$

$$\Delta RF_{2nd\ week} = \frac{RF_{10f} - RF_{5f}}{RF_{5f}} \quad \text{(Eq.5)}$$

$$\Delta RF_{2\ weeks} = \frac{RF_{10f} - RF_{pretreatment}}{RF_{pretreatment}} \quad \text{(Eq.6)}$$

where $\Delta RF_{1st\ week}$, $\Delta RF_{2nd\ week}$, and $\Delta RF_{2\ weeks}$ denoted the change of RFs within the 1st week, 2nd week, and first two weeks, and $RF_{pretreatment}$, RF_{5f}, and RF_{10f} denoted the RFs in images acquired before treatment, after the 5th radiation, and after the 10th radiation, respectively.

Construction of Prediction Model Based on Radiomics Signature

As shown in **Figure 1B**, $RF_{pretreatment}$, $\Delta RF_{1st\ week}$, $\Delta RF_{2nd\ week}$, and $\Delta RF_{2\ weeks}$ were used to train a support vector machine (SVM) classifier in the training set, and then a radiomics signature was built by the trained classifier and evaluated in the internal and external testing sets. The cost-based SVM classifier, which used the radial basis function kernel and a cost coefficient of 0.001, was trained by RFs with the 10-fold cross-validation. The SVM classifier transformed the feature space to a high-dimensional space where a separating hyperplane maximized the margin between classes. For each patient, the SVM classifier generated a radiomics signature for evaluating predictions of treatment response. The receiver operating characteristic (ROC) curves of obtained radiomics signatures in the training, internal, and external testing sets were analyzed and then tested by the DeLong test. The chi-square test was used to compare the predicted responses, which were corresponding to radiomics signatures in the internal testing with the signatures' ground truth, i.e., treatment response evaluated by RECIST1.1. Afterward, the eligible radiomics signatures together with clinical characteristics were imported to a multivariate logistic regression-based prediction model to predict the sensitivity to treatment. Furthermore, the performance of radiomics signatures, the ADC value, clinical characteristics, and the regression model was assessed by the ROC curve, respectively. The area under the curve (AUC) was calculated, and the optimal Youden's index was determined.

Statistical Analysis

Statistical analysis was performed using the R software (version 3.6.1, http://www.R-project.org). The ICC method, Bland–Altman method, mRMR approach, SVM classifier, logistic regression analysis, and ROC curve analysis were implemented based on the R packages "irr", "BlandAltmanLeh", "mRMRe", "e1071", and "pROC." Chi-square test or t-test was used to examine the correlations between clinical features and treatment response. The differences in ADC and ΔADC values were compared between the sensitive group and the resistant group by Mann–Whitney U test. *P*-value < 0.05 was considered as an indicator of the statistically significant difference.

RESULTS

Clinical Characteristics

In Shandong Cancer Hospital and Institute, eighty-five consecutive patients with ESCC were enrolled. Nine patients were excluded, given that the DWI of 6 patients were deformed or had high noise and 3 patients' radiotherapy was interrupted. Finally, 76 patients were enrolled in the study during the period of enrollment from June 2014 to September 2019. Fifty-four of the 76 patients experienced clinical PR, 2 with CR and 20 with SD. Patients (n=76) were randomly assigned to training (n=53) and internal testing sets (n=23) in a ratio of 7 to 3. There was no significant difference in clinical characteristics between the training and internal testing sets (sex, age, tumor location, clinical T stage, lymph node status, and radiation dose) and treatment response in both sets (**Table 1**). In Anyang Tumor Hospital, twenty consecutive patients with ESCC were enrolled during the period of enrollment from April 2017 to May 2019. Three patients were excluded, given that 3 patients' radiotherapy was interrupted. Finally, 17 patients were enrolled in the external testing set. Eleven of the 17 patients experienced clinical PR and 6 with SD. Clinical characteristics are presented in **Table 1**.

We measured the voxel volume of the whole-tumor ADC map to substitute the tumor volume and the maximum 3D diameter and their changes between each time point. **Supplementary Tables 3, 4** show that these parameters had no relationship with treatment response.

Association of Sequential ADC Value and Treatment Response

The mean values of $ADC_{pretreatment}$, ADC_{5f}, and ADC_{10f} of all patients are shown in **Table 2**. All of the ADC values and relative changes during the first two weeks showed no significant difference between the sensitive and resistant groups in the training and internal testing sets. In the external testing set, ADC in the 5th radiation showed association with treatment response ($p = 0.048$). We noticed that the ADC value was gradually higher than before treatment as the treatment progressed.

Generation and Validation of the Radiomics Signature

A total of 560 RFs (**Supplementary Table 5**) were selected by the ICC method from two consecutive whole-tumor ADC map acquisitions in the same patient. Afterward, 224 RFs (**Supplementary Table 6**) were selected from pretreatment images in the training set *via* the Bland–Altman method, and 30 RFs (**Supplementary Table 7**) were finally chosen out according to mRMR. Most of these RFs were calculated to measure the local homogeneity of the image. To predict good

TABLE 1 | Patients and tumor characteristics association with treatment response in the training, internal, and, external testing sets.

Characteristics	Training set (n=53)		*p*	Internal testing set (n=23)		*p*	*p* †	External testing set (n=17)		p
	Sensitive group (n=36)	Resistant group (n=17)		Sensitive group (n=18)	Resistant group (n=5)		0.361	Sensitive group (n=11)	Resistant group (n=6)	
Sex			0.821			0.048	0.793			1.000
Male	28	12		16	3			6	3	
Female	8	5		2	2			5	3	
Age (year)			0.412			0.446	0.675			0.660
mean ± sd	63.1 ± 8.14	65.2 ± 10.3		63.9 ± 5.6	59.0 ± 10.7			62.2 ± 8.2	64.5 ± 6.1	
range	43-78	44-80		51-72	39-70			51-72	39-70	
T-stage			0.730			0.576	0.602			0.515
T3	25	11		14	3			10	4	
T4	11	6		4	2			1	2	
LN Status			0.647			1.000	0.834			1.000
N-	7	5		3	1			4	2	
N+	29	12		15	4			7	4	
Location			0.023			0.662	1.000			0.043
Cervival	0	4		1	1			1	3	
Upper thoracic	15	7		7	2			6	2	
Middle thoracic	14	3		6	2			4	0	
Lower thoracic	7	3		4	0			0	1	
Dose (Gy)			0.234			1.000	0.823			1.000
<55	19	6		9	2			9	2	
≥55	17	11		8	3			8	3	

LN, lymph node.

† Difference between the training set and the internal testing set in characteristics of patients and tumor.

**P < 0.05, statistically significant.*

and bad responder, four radiomic signatures based on an SVM classifier were respectively built with $RF_{pretreatment}$, $\Delta RF_{1st\ week}$, $\Delta RF_{2nd\ week}$, and $\Delta RF_{2\ weeks}$. The effectiveness of these 4 radiomics signatures to classify the sensitive group versus the resistant group is shown in **Figure 3**. Only the radiomics signature based on $\Delta RF_{2\ weeks}$, denoted as R-$Signature_{2\ weeks}$, discriminated treatment response with an AUC value higher than 0.5 in the training set [AUC = 0.824, 95% confidence interval (CI): 0.679–0.968], the internal testing set (AUC = 0.744, 95% CI: 0.465–1.0), and the external testing set (AUC = 0.742, 95% CI: 0.478–0.919). No difference was found between the training and internal testing sets according to the results of the DeLong test (p = 0.580), nor between the training and external testing sets (p = 0.526). Compared with the evaluated response, the performance of response prediction in internal testing was acceptable (p = 0.048). The chi-square test of prediction results between R-$Signature_{2\ weeks}$ in the training set and the internal testing set achieved a p-value of 1.000, which implied that the distribution of R-$Signature_{2\ weeks}$ between the two sets was not statistically significantly different.

Combining Model Analysis

As elaborated in **Supplementary Table 8**, in the training set, the outputs of univariate logistics regression showed that tumor location and R-$Signature_{2\ weeks}$ were separately associated with treatment response. We performed the tumor location and R-$signature_{2\ weeks}$ to the prediction model of treatment response by multivariate logistics regression (**Supplementary Table 9**) in the three sets. R-$Signature_{2\ weeks}$ was identified as an independent factor predicting treatment response in multivariate analysis in

TABLE 2 | Association between ADC values or ΔADC values and treatment response in the training, internal, and external testing sets.

Set	Time point	ADC value (10^{-3} mm2/s)		*p*	Time range	ΔADC		p
		Sensitive group	Resistant group			Sensitive group	Resistant group	
Training set	Pre-treatment	1.695 ± 0.687	1.644 ± 0.626	0.746	1st week	0.200 ± 0.415	0.135 ± 0.293	0.391
	5th radiation	1.723 ± 0.612	2.067 ± 0.875	0.253	2nd week	-0.160 ± 0.218	-0.175 ± 0.286	0.819
	10th radiation	2.057 ± 0.720	2.118 ± 0.697	0.381	2 weeks	0.275 ± 0.370	0.255 ± 0.400	0.746
Internal testing set	Pre-treatment	1.556 ± 0.599	1.466 ± 0.946	0.363	1st week	0.177 ± 0.444	0.351 ± 0.396	0.491
	5th radiation	1.847 ± 0.624	1.893 ± 0.990	0.914	2nd week	-0.152 ± 0.239	0.126 ± 0.630	0.538
	10th radiation	2.145 ± 0.831	1.973 ± 0.924	0.691	2 weeks	0.405 ± 0.595	0.538 ± 0.692	0.745
External testing set	Pre-treatment	1.948 ± 0.247	1.818 ± 0.388	0.149	1st week	0.065 ± 0.037	0.042 ± 0.0447	0.216
	5th radiation	2.084 ± 0.237	1.787 ± 0.422	0.048*	2nd week	0.119 ± 0.117	0.145 ± 0.113	0.591
	10th radiation	2.387 ± 0.263	2.078 ± 0.283	0.078	2 weeks	0.177 ± 0.110	0.179 ± 0.127	1.000

ADC, apparent diffusion coefficient.

**P < 0.05, statistically significant.*

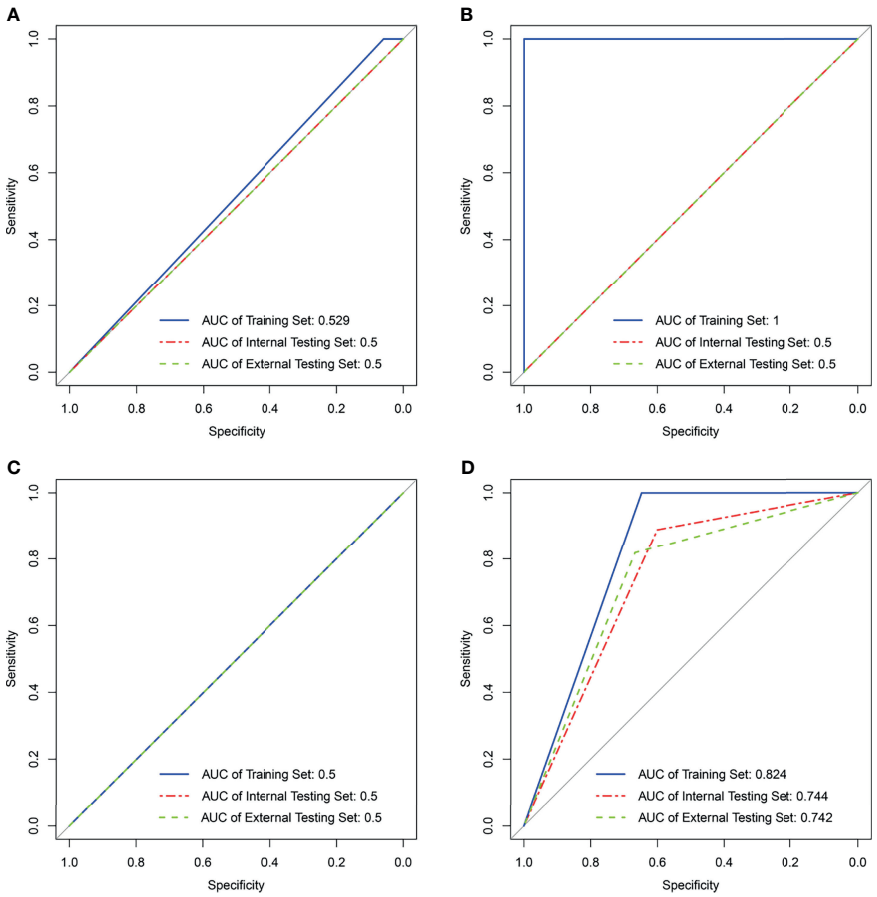


FIGURE 3 | Results of support vector machine classifier in the training, internal, and external testing sets with four kinds of radiomics features. **(A)** Result of SVM generating radiomics signature$_{pretreatment}$. **(B)** Result of SVM generating radiomics signature$_{1st\ week}$. **(C)** Result of SVM generating radiomics signature$_{2nd\ week}$. **(D)** Result of SVM generating radiomics signature$_{2\ weeks}$. NA in **(B, C)** means not applicable.

the three sets. **Figure 4** shows the results of ROC analyses on the tumor location, R-Signature$_{2\ weeks}$, and the regression model in the prediction of treatment response in the training, internal, and external testing sets. **Table 3** shows the effectiveness of our proposed regression model in discriminating the sensitive and resistant groups.

DISCUSSION

In this study, a radiomics signature was developed that incorporated the change of RFs based on the whole-tumor ADC map for predicting treatment response to concurrent CRT in patients with ESCC. Compared with pretreatment, our study showed that the radiomics signature based on $\Delta RF_{2\ weeks}$ was able to predict the treatment response for evaluations in the following 2–3 months.

DWI-MRI was widely and routinely performed in kinds of tumors for diagnosis and treatment assessment. As designed in our study, ADC values were obtained at pretreatment, after the 5th radiation, and after the 10th radiation. However, the mean value of pretreatment ADC was not significantly different between sensitive and resistant groups, which was consistent with the studies of Kozumi in patients with ESCC undergoing concurrent CRT (11). In fact, results of prediction of treatment response using ADC or ΔADC were not consistent among previous studies, or even the opposite (8–10, 12). In our study, ADC_{5f} in the external testing showed association with treatment response, which was different from those in the training and internal testing sets. Since the diffusion movement of water molecules in the tissue was closely related to the distribution of the intratumoral structure, the ADC value tended to be affected by many factors, such as cell density, nuclear area, nuclear–cytoplasmic ratio, tumor angiogenesis, and proteins (28–30). In addition, tissue edema commonly occurred during CRT because of changes in blood perfusion, cell death, and blocked lymphatic drainage induced by radiation and cytotoxicity, which further caused the change of the ADC value (31). We observed that the ADC value increased in both the sensitive group and the resistance group, but only one set of positive results was

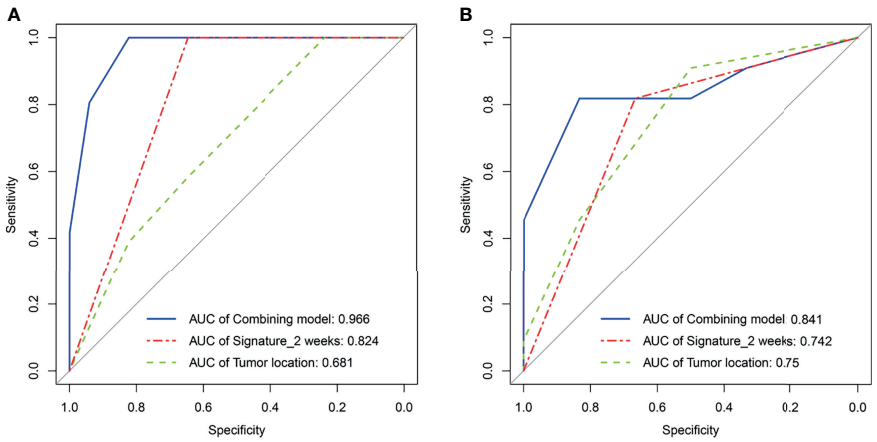


FIGURE 4 | Receiver operating characteristic curves analysis for tumor location, radiomics signature$_{2\ weeks}$, and regression model, respectively, in the training set **(A)** and the external testing set **(B)**.

obtained in the external testing set. Thus, $ADC_{pretreatment}$ and the changes during the first ten radiations were limited in predicting the treatment response.

The mean value of ADC and ΔADC reflected the average level of the Brownian motion of water molecules and a rough change in the whole tumor, while radiomics offered more heterogeneity information involving the above molecular and pathologic characteristics in different regions. The rise of radiomics using quantitative image features of tumors provided an opportunity for the development of predictive biomarkers (32). Over the years, although several studies on the prediction performance of radiomics analysis to treatment response had been reported in patients with esophageal cancer, the predictive radiomics features were hard to reproduce since there were thousands of RFs and diverse selection criteria. For example, in the ^{18}F-FDG PET, different radiomics features were found with discriminatory capability in predicting response to CRT in patients with esophageal cancer (33–35). Many researchers are committed to the standardization and unification of radiomics due to its virtue of convenience to use (36, 37). In our study, the radiomics signature was developed from 30 radiomics features from pretreatment images, and these radiomics features were selected with high interobserver reproducibility, high relevance, and low redundancy. Nevertheless, it was inadequate to predict the response merely using RFs from pretreatment images as what the single-shot image modeling did.

In our study, the selected features were extracted from a four-gray-level matrix, which was the ADC value matrix in the ADC map. Sixteen selected feature names were explained representing homogeneity or heterogeneity. The interpretation of other features was based on their calculation, including the distribution pattern of gray and the neighborhood intensity of the image. Their changes are recorded by ΔRFs. We proposed to train an SVM classifier by the RFs extracted from the images of four time points, including $RF_{pretreatment}$, $\Delta RF_{1st\ week}$, $\Delta RF_{2nd\ week}$, and $\Delta RF_{2\ weeks}$. R-$Signature_{pretreatment}$ showed no correlation with treatment response. Instead, containing the change of whole-tumor heterogeneity over treatment, the

TABLE 3 | Performance of radiomics signatures and models for predicting treatment response in the training, internal, and external testing sets.

Set	Predictor	Sensitivity (%)	Specificity (%)	Accuracy (%)	PPV (%)	NPV (%)	*p*	AUC (95% CI)
Training set	Tumor location	100.0	23.5	75.5	73.5	100.0	0.046	0.681 (0.504-0.773)
	R-Signature$_{2\ weeks}$	100.0	64.7	88.7	85.7	100.0	<0.0001	0.824 (0.694-0.915)
	Tumor location+ R-Signature$_{2\ weeks}$	100.0	82.4	94.3	92.3	100.0	<0.0001	0.966 (0.875-0.996)
Internal testing set*	Tumor location	22.2	100.0	39.1	100.0	26.3	0.256	0.650 (0.425-0.835)
	R-Signature$_{2\ weeks}$	88.9	60.0	82.6	88.9	60.0	0.044	0.744 (0.522-0.901)
	Tumor location+ R-Signature$_{2\ weeks}$	–	–	–	–	–	–	–
External testing set	Tumor location	90.9	50	76.5	75.0	75.0	0.0499	0.674 (0.486-0.924)
	R-Signature$_{2\ weeks}$	81.8	66.7	76.5	81.8	66.7	0.0465	0.742 (0.478-0.919)
	Tumor location+ R-Signature$_{2\ weeks}$	81.8	83.3	82.3	90.0	71.4	0.027	0.841 (0.586-0.970)

**Tumor location was showed a p-value of 0.256, which means the model was not established.*
PPV, positive predictive value; NPV, negative predictive value; AUC, area under curve; CI, confidence interval.
**P < 0.05, statistically significant.*

radiomics signature based on $\Delta RF_{2\ weeks}$ predicted the treatment response with an accuracy of 0.887, 0.826, and 0.765 in the three sets, respectively. In this study, the trend of ADC change during the initial two weeks showed an increase with no correlation with treatment response. This increase was thought to reflect the reduction in cell membrane integrity and changes in tumor cell numbers due to cytotoxic drugs and radiation, resulting in less restriction on the Brownion motion of water molecules (12). As the treatment progressed, tumor cells gradually decomposed and were absorbed, and the blood vessels and fibers in the microenvironment changed together. During this time, the changes of internal structures were not synchronized due to heterogeneity including different cell-cycle, oxygen status, and proliferative potential (38, 39). Our study was designed to record meaningful changes as early as possible (40). By screening radiomics features, effective predictions can be made through statistical methods. The results showed that the larger the time span, the more significant the changes obtained in radiomics features. Delta-radiomics has previously been used for the prediction of response to tumor treatment with ADC maps and CT (41, 42). As radiation and reaction to cytotoxic drugs accumulated, early changes from the tumor interior were recorded and quantified by a ratio that reduced the variability of the data. Moreover, Nasief et al. predicted the treatment response for CRT of pancreatic cancer by combining CT-based delta-radiomics and increasing CA19-9 (43).

In our study, esophageal tumor location was associated with treatment response in the training and external testing sets, while few studies had reported that the location of primary tumors was associated with the sensitivity of CRT in ESCC. One explanation was that biases existed in the clinical characteristic distribution in different responders with treatment due to the small sample size. Such as the result in the internal testing set, although the AUC changed, there was no increase in diagnostic performance. The other one was that the uneven dose distribution reduced the treatment effect since cervical esophagus had physiological curvature affected by the peripheral anatomical structure (44). The association of tumor locations and treatment response was to be further investigated in a larger cohort.

Our study had several limitations. Firstly, although an external testing set was included, the cohort of our study was still small. A multicenter study with a larger patient cohort is required. Secondly, though several selected procedures were used, esophageal tumors were manually delineated, which introduced an uncertain level of observer dependency. Thirdly, compared with RECIST 1.1 in the definitive cCRT, the pathological evaluation of neoadjuvant CRT was more convincing, and a systematic control comparison was supposed to be studied in the future.

In conclusion, we developed a radiomics signature-based model that incorporated the early change of RFs based on the sequential whole-tumor ADC map to predict treatment response to cCRT in patients with ESCC. This model provided a solution to quantify intratumoral changes and utilized them to guide treatment planning at an early date.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of Shandong Cancer Hospital and Institute. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

Study conception and design: DA, BL, and QC. Collection of patient data: DA, YL, NS, and YZ. ROI delineation: WL and NS. Data analysis and statistical analysis: DA, QC, and ZL. Manuscript writing: DA. Manuscript editing: QC, ZL, and BL. All authors contributed to the article and approved the submitted version.

FUNDING

This study has received funding from the National Natural Science Foundation of China (81530060), the National Natural Science Foundation of China (81874224), and the academic promotion program of Shandong First Medical University (2019LJ004).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.787489/full#supplementary-material

REFERENCES

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* (2018) 68:394–424. doi: 10.3322/caac.21492
2. Rustgi AK, El-Serag HB. Esophageal Carcinoma. *N Engl J Med* (2014) 371:2499–509. doi: 10.1056/NEJMra1314530
3. Yuequan J, Shifeng C, Bing Z. Prognostic Factors and Family History for Survival of Esophageal Squamous Cell Carcinoma Patients After Surgery. *Ann Thorac Surg* (2010) 90:908–13. doi: 10.1016/j.athoracsur.2010.05.060
4. van Hagen P, Hulshof MCCM, Van Lanschot JJB, Steyerberg EW, Henegouwen MVB, Wijnhoven BPL, et al. Preoperative Chemoradiotherapy for Esophageal or Junctional Cancer. *N Engl J Med* (2012) 366:2074–84. doi: 10.1046/j.1440-1673.1999.00653.x

5. Cooper JS, Guo MD, Herskovic A, Macdonald JS, Martenson JAJr., Al-Sarraf M, et al. Chemoradiotherapy of Locally Advanced Esophageal Cancer: Long-Term Follow-Up of a Prospective Randomized Trial (RTOG 85-01). *JAMA* (1999) 281:1623–7. doi: 10.1001/jama.281.17.1623
6. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1). *Eur J Cancer* (2009) 45:228–47. doi: 10.1016/S1359-6349(09)70018-7
7. Wei Y, Wu S, Gao F, Sun T, Zheng D, Ning P, et al. Esophageal Carcinoma: *Ex Vivo* Evaluation by High-Spatial-Resolution T2-Mapping MRI Compared With Histopathological Findings at 3.0 T. *J Magn Reson Imaging* (2017) 45:1609–16. doi: 10.1002/jmri.25509
8. De Cobelli F, Giganti F, Orsenigo E, Cellina M, Esposito A, Agostini G, et al. Apparent Diffusion Coefficient Modifications in Assessing Gastro-Oesophageal Cancer Response to Neoadjuvant Treatment: Comparison With Tumour Regression Grade at Histology. *Eur Radiol* (2013) 23:2165–74. doi: 10.1007/s00330-013-2807-0
9. Borggreve AS, Goense L, van Rossum PS, Heethuis SE, van Hillegersberg R, Lagendijk JJ, et al. Preoperative Prediction of Pathologic Response to Neoadjuvant Chemoradiotherapy in Patients With Esophageal Cancer Using ^{18}F-FDG PET/CT and DW-MRI: A Prospective Multicenter Study. *Int J Radiat Oncol* (2020) 106(5):998–1009. doi: 10.1016/j.ijrobp.2019.12.038
10. Jacobs L, Intven M, van Lelyveld N, Philippens M, Burbach M, Seldenrijk K, et al. Diffusion-Weighted MRI for Early Prediction of Treatment Response on Preoperative Chemoradiotherapy for Patients With Locally Advanced Rectal Cancer. *Ann Surg* (2016) 263:522–8. doi: 10.1097/SLA.0000000000001311
11. Kozumi M, Ota H, Yamamoto T, Umezawa R, Matsushita H, Ishikawa Y, et al. Oesophageal Squamous Cell Carcinoma: Histogram-Derived ADC Parameters Are Not Predictive of Tumour Response to Chemoradiotherapy. *Eur Radiol* (2018) 28:4296–305. doi: 10.1007/s00330-018-5439-6
12. Partridge SC, Zhang Z, Newitt DC, Gibbs JE, Chenevert TL, Rosen MA, et al. Diffusion-Weighted MRI Findings Predict Pathologic Response in Neoadjuvant Treatment of Breast Cancer: The ACRIN 6698 Multicenter Trial. *Radiology* (2018) 289:618–27. doi: 10.1148/radiol.2018180273
13. Imanishi S, Shuto K, Aoyagi T, Kono T, Saito H, Matsubara H. Diffusion-Weighted Magnetic Resonance Imaging for Predicting and Detecting the Early Response to Chemoradiotherapy of Advanced Esophageal Squamous Cell Carcinoma. *Dig Surg* (2013) 30:240–8. doi: 10.1159/000351435
14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278:563–77. doi: 10.1148/radiol.2015151169
15. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, Van Stiphout RG, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
16. Li H, Zhu Y, Burnside ES, Drukker K, Hoadley KA, Fan C, 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* (2016) 281:382–91. doi: 10.1148/radiol.2016152110
17. Aerts HJ WL, Velazquez ER, Leijenaar R TH, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5:4006. doi: 10.1038/ncomms5006
18. Liu Y, Zhang Y, Cheng R, Liu S, Qu F, Yin X, et al. Radiomics Analysis of Apparent Diffusion Coefficient in Cervical Cancer: A Preliminary Study on Histological Grade Evaluation. *J Magn Reson Imaging* (2019) 49:280–90. doi: 10.1002/jmri.26192
19. Coroller TP, Agrawal V, Narayan V, Hou Y, Grossmann P, Lee SW, et al. Radiomic Phenotype Features Predict Pathological Response in Non-Small Cell Lung Cancer. *Radiother Oncol* (2016) 119:480–6. doi: 10.1016/j.radonc.2016.04.004
20. Xiong J, Yu W, Ma J, Ren Y, Fu X, Zhao J. The Role of PET-Based Radiomic Features in Predicting Local Control of Esophageal Cancer Treated With Concurrent Chemoradiotherapy. *Sci Rep* (2018) 8:9902. doi: 10.1038/s41598-018-28243-x
21. van Timmeren JE, Leijenaar R TH, van Elmpt W, Reymen B, Oberije C, Monshouwer R, et al. Survival Prediction of Non-Small Cell Lung Cancer Patients Using Radiomics Analyses of Cone-Beam CT Images. *Radiother Oncol* (2017) 123:363–9. doi: 10.1016/j.radonc.2017.04.016
22. Boldrini L, Cusumano D, Chiloiro G, Casà C, Masciocchi C, Lenkowicz J, et al. Delta Radiomics for Rectal Cancer Response Prediction With Hybrid 0.35 T Magnetic Resonance-Guided Radiotherapy (MRgRT): A Hypothesis-Generating Study for an Innovative Personalized Medicine Approach. *Radiol Med* (2019) 124:145–53. doi: 10.1007/s11547-018-0951-y
23. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3d Slicer as an Image Computing Platform for the Quantitative Imaging Network. *Magn Reson Imaging* (2012) 30:1323–41. doi: 10.1016/j.mri.2012.05.001
24. Van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77:e104–7. doi: 10.1158/0008-5472.CAN-17-0339
25. Zwanenburg A, Leger S, Vallières M, Löck S. Image Biomarker Standardisation Initiative. *arXiv preprint* (2016) arXiv:1612.07003. doi: 10.48550/arXiv.1612.07003
26. Bland JM, Altman DG. Comparing Methods of Measurement: Why Plotting Difference Against Standard Method Is Misleading. *Lancet* (1995) 346:1085–7. doi: 10.1016/s0140-6736(95)91748-9
27. Barnhart HX, Barboriak DP. Applications of the Repeatability of Quantitative Imaging Biomarkers: A Review of Statistical Analysis of Repeat Data Sets. *Transl Oncol* (2009) 2:231–5. doi: 10.1593/tlo.09268
28. Driessen JP, Caldas-Magalhaes J, Janssen LM, Pameijer FA, Kooij N, Terhaard CH, et al. Diffusion-Weighted MR Imaging in Laryngeal and Hypopharyngeal Carcinoma: Association Between Apparent Diffusion Coefficient and Histologic Findings. *Radiology* (2014) 272:456–63. doi: 10.1148/radiol.14131173
29. Aoyagi T, Shuto K, Okazumi S, Hayano K, Satoh A, Saitoh H, et al. Apparent Diffusion Coefficient Correlation With Oesophageal Tumour Stroma and Angiogenesis. *Eur Radiol* (2012) 22:1172–7. doi: 10.1007/s00330-011-2359-0
30. Mori N, Ota H, Mugikura S, Takasawa C, Ishida T, Watanabe G, et al. Luminal-Type Breast Cancer: Correlation of Apparent Diffusion Coefficients With the Ki-67 Labeling Index. *Radiology* (2014) 274:66–73. doi: 10.1148/radiol.14140283
31. Lepke RA, Libshitz HI. Radiation-Induced Injury of the Esophagus. *Radiology* (1983) 148:375–8. doi: 10.1148/radiology.148.2.6867327
32. O'Connor JPB, Rose CJ, Waterton JC, Carano RA, Parker GJ, Jackson A. Imaging Intratumor Heterogeneity: Role in Therapy Response, Resistance, and Clinical Outcome. *Clin Cancer Res* (2015) 21:249–57. doi: 10.1158/1078-0432.CCR-14-0990
33. Tixier F, Le Rest CC, Hatt M, Albarghach N, Pradier O, Metges JP, et al. Intratumor Heterogeneity Characterized by Textural Features on Baseline ^{18}F-FDG PET Images Predicts Response to Concomitant Radiochemotherapy in Esophageal Cancer. *J Nucl Med* (2011) 52:369–78. doi: 10.2967/jnumed.110.082404
34. Nakajo M, Jinguji M, Nakabeppu Y, Nakajo M, Higashi R, Fukukura Y, et al. Texture Analysis of ^{18}F-FDG PET/CT to Predict Tumour Response and Prognosis of Patients With Esophageal Cancer Treated by Chemoradiotherapy. *Eur J Nucl Med Mol Imaging* (2017) 44:206–14. doi: 10.1007/s00259-016-3506-2
35. Beukinga RJ, Hulshoff JB, Mul V EM, Noordzij W, Kats-Ugurlu G, Slart RH, et al. Prediction of Response to Neoadjuvant Chemotherapy and Radiation Therapy With Baseline and Restaging ^{18}F-FDG PET Imaging Biomarkers in Patients With Esophageal Cancer. *Radiology* (2018) 287:983–92. doi: 10.1148/radiol.2018172229
36. Depeursinge A, Fageot J, Al-Kadi OS. "Fundamentals of Texture Processing for Biomedical Image Analysis: A General Definition and Problem Formulation". In: *Biomedical Texture Analysis*. London: Academic Press (2017). p. 1–27.
37. Toennies KD. *Guide to Medical Image Analysis*. London: Springer Press (2017). doi: 10.1007/978-1-4471-7320-5
38. Yaes RJ. Tumor Heterogeneity, Tumor Size, and Radioresistance. *Int J Radiat Oncol Biol Phys* (1989) 17(5):993–1005. doi: 10.1016/0360-3016(89)90147-8. Sc D, MD.
39. Sun XX, Yu Q. Intra-Tumor Heterogeneity of Cancer Cells and Its Implications for Cancer Treatment. *Acta Pharmacol Sin* (2015) 36:1–10. doi: 10.1038/aps.2015.92
40. Davnall F, Yip C, Ljungqvist G, Selmi M, Ng F, Sanghera B, et al. Assessment of Tumor Heterogeneity: An Emerging Imaging Tool for Clinical Practice? *Insights Into Imaging* (2012) 3:573–89. doi: 10.1007/s13244-012-0196-6

41. Gao Y, Kalbasi A, Hsu W, Ruan D, Fu J, Shao J, et al. Treatment Effect Prediction for Sarcoma Patients Treated With Preoperative Radiotherapy Using Radiomics Features From Longitudinal Diffusion-Weighted MRIs. *Phys Med Biol* (2020) 65(17):175006. doi: 10.1088/1361-6560/ab9e58
42. Lin P, Yang PF, Chen S, Shao YY, Xu L, Wu Y, et al. A Delta-Radiomics Model for Preoperative Evaluation of Neoadjuvant Chemotherapy Response in High-Grade Osteosarcoma. *Cancer Imaging* (2020) 20(1):1–12. doi: 10.1186/s40644-019-0283-8
43. Nasief H, Hall W, Zheng C, Tsai S, Wang L, Erickson B, et al. Improving Treatment Response Prediction for Chemoradiation Therapy of Pancreatic Cancer Using a Combination of Delta-Radiomics and the Clinical Biomarker CA19-9. *Front Oncol* (2020) 9:1464. doi: 10.3389/fonc.2019.01464
44. Fenkell L, Kaminsky I, Breen S, Huang S, Van Prooijen M, Ringash J. Dosimetric Comparison of IMRT vs. 3D Conformal Radiotherapy in the Treatment of Cancer of the Cervical Esophagus. *Radiother Oncol* (2008) 89:287–91. doi: 10.1016/j.radonc.2008.08.008

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 An, Cao, Su, Li, Li, Liu, Zhang and Li. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 19 April 2022
doi: 10.3389/fonc.2022.807725

Predictive Value of Delta-Radiomics Texture Features in 0.35 Tesla Magnetic Resonance Setup Images Acquired During Stereotactic Ablative Radiotherapy of Pancreatic Cancer

*Garrett Simpson, William Jin, Benjamin Spieler, Lorraine Portelance, Eric Mellon, Deukwoo Kwon, John C. Ford and Nesrin Dogan**

Radiation Oncology, Miller School of Medicine, University of Miami, Miami, FL, United States

OPEN ACCESS

Edited by:
Omar Sultan Al-Kadi, The University of Jordan, Jordan

Reviewed by:
Hanbo Chen, University of Toronto, Canada
Todd A. Aguilera, University of Texas Southwestern Medical Center, United States
Ahmed Elamir, University of Texas Southwestern Medical Center, United States, in collaboration with reviewer TA

***Correspondence:**
Nesrin Dogan
NDogan@med.miami.edu

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: *02 November 2021*
Accepted: *21 March 2022*
Published: *19 April 2022*

Citation:
Simpson G, Jin W, Spieler B, Portelance L, Mellon E, Kwon D, Ford JC and Dogan N (2022) Predictive Value of Delta-Radiomics Texture Features in 0.35 Tesla Magnetic Resonance Setup Images Acquired During Stereotactic Ablative Radiotherapy of Pancreatic Cancer. Front. Oncol. 12:807725. doi: 10.3389/fonc.2022.807725

Purpose: The purpose of this work is to explore delta-radiomics texture features for predicting response using setup images of pancreatic cancer patients treated with magnetic resonance image guided (MRI-guided) stereotactic ablative radiotherapy (SBRT).

Methods: The total biological effective dose (BED) was calculated for 30 patients treated with MRI-guided SBRT that delivered physical doses of 30–60 Gy in three to five fractions. Texture features were then binned into groups based upon BED per fraction by dividing BED by the number of fractions. Delta-radiomics texture features were calculated after delivery of 20 Gy BED (BED20 features) and 40 Gy BED (BED40 features). A random forest (RF) model was constructed using BED20 and then BED40 features to predict binary outcome. During model training, the Gini Index, a measure of a variable's importance for accurate prediction, was calculated for all features, and the two features that ranked the highest were selected for internal validation. The two features selected from each bin were used in a bootstrapped logistic regression model to predict response and performance quantified using the area under the receiver operating characteristic curve (AUC). This process was an internal validation analysis.

Results: After RF model training, the Gini Index was highest for gray-level co-occurrence matrix-based (GLCM) sum average, and neighborhood gray tone difference matrix-based (NGTDM) busyness for BED20 features and gray-level size zone matrix-based (GLSZM) large zones low gray-level emphasis and gray-level run length matrix-based (GLRLM) run percentage was selected from the BED40-based features. The mean AUC obtained using the two BED20 features was AUC = 0.845 with the 2.5 percentile and 97.5 percentile values

ranging from 0.794 to 0.856. Internal validation of the BED40 delta-radiomics features resulted in a mean AUC = 0.567 with a 2.5 and 97.5 percentile range of 0.502–0.675.

Conclusion: Early changes in treatment quantified with the BED20 delta-radiomics texture features in low field images acquired during MRI-guided SBRT demonstrated better performance in internal validation than features calculated later in treatment. Further analysis of delta-radiomics texture analysis in low field MRI is warranted.

Keywords: delta-radiomics, texture analysis, pancreas cancer, MRI, low field (0.35 T)

INTRODUCTION

Radiomic analysis is the extraction of quantitative image descriptors from radiographic images. Radiomic texture features quantify image gray-level spatial relationships that are hypothesized to represent a tumor's underlying microenvironment (1). Tumor heterogeneity presents a major obstacle for personalized medicine and effective treatment, and quantification of the heterogeneity could provide valuable insight (2, 3). Utilization of radiomic features is attractive, as it can provide a non-invasive patient-specific biomarker based on routinely acquired images (4–6). Interest in radiomic texture features intensifies, as they are found to have utility in predictive modeling, support clinical decision-making, and are linked with phenotypic and genetic profiles of various cancers (7–9).

Application of radiomic analysis in pancreatic cancer patients is of particular interest due to the poor treatment outcomes. The incidence of pancreatic cancer continues to increase with a low 5-year overall survival of approximately 9% and is predicted to be one of the top 5 contributors of cancer deaths among men and women in the United States in 2021 (10). Surgery offers patients with pancreatic ductal adenocarcinoma (PDAC) the best chance of long-term survival, but only a minority of patients, 15%–20%, qualify for resection (11). Patients diagnosed as having borderline resectable or unresectable disease generally receive upfront chemotherapy (nCRT) followed by stereotactic body radiation therapy (SBRT) (12–14). This regime can induce changes in the gross disease and allow patients to qualify for surgery. Furthermore, SBRT has improved complete pathological response rates from 2.5% to 7.8% (15). Real-time disease monitoring *via* radiomic analysis may provide valuable personalized information that could lead to beneficial adaptation of the patient care path and improved outcome.

The requirement and desire for better visualization of tumor target and organs at risk locations in patient populations like PDAC patients during therapy have driven the development of new image guidance technology. Magnetic resonance imaging (MRI)-guided radiotherapy units provide an ideal platform for treatment of pancreas SBRT patients because of the ability to acquire high-quality images. MR images have improved soft tissue contrast compared to traditional X-ray guidance modalities and allow for safe dose escalation and avoidance of critical organs in proximity to the target. The daily MRIs acquired prior to each treatment are a by-product of the patient care path but may contain valuable information about patient disease over the course of treatment. These daily images provide an opportunity for constructing models predictive of treatment response for PDAC patients using radiomic analysis. A previous work has indicated that the low field strength MR images provide an adequate base for texture analysis (16). Furthermore, a longitudinal image series may bring the opportunity to monitor tumor microenvironment changes during treatment.

Using the average radiomic feature value throughout the treatment demonstrated possible prognostic utility, but consideration of the time that a prediction is made is important (17). A prediction made too late in the treatment is not ideal because the knowledge provided is unlikely to positively impact a patient's care. Interest and application of radiomic texture features has shifted towards delta-radiomics, which considers changes in radiomic feature values in response during treatment. Delta-radiomics may provide valuable patient-specific outcome prediction with multiple fractions remaining. Indication of disease response prior to the end of radiotherapy treatment could provide valuable information for physicians to consider options for the remainder of the treatment. A limited number of studies have explored delta-radiomic analysis on diagnostic images acquired prior, periodically during, or after treatment (18–21). Nasief et al. used delta-radiomics texture features calculated as the change relative to pre-fraction values extracted from CT images of pancreatic cancer patients to predict treatment outcome (22). High-contrast diagnostic images with better resolution are generally considered ideal for radiomics analysis; however, the limited number of images could limit exploration of delta-radiomics. The increased reliance on imaging in radiation oncology is producing longitudinal image series. The indication of prognostic performance of delta-radiomics features extracted from cone beam computed tomography images was encouraging and has led to further interest in using setup images from other modalities (23). MRI-guided units produce daily setup images for each patient with previous work from our group, indicating that these images may provide a basis for radiomics analysis (24). Radiomics analysis of low field MRIs has gained popularity, but delta-radiomics analysis of images in SBRT for response prediction in PDAC patients is still evolving. The purpose of this study is to use low field strength MR images acquired prior to each daily treatment of PDAC patients treated with MRI-guided SBRT to evaluate the feasibility of using delta-radiomic features for predicting patient response.

MATERIALS AND METHODS

Patient Database

To qualify, patients should have biopsy-confirmed PDAC, including locally advanced or borderline resectable disease. Additionally, each patient should have completed chemotherapy prior to MR-guided SBRT. Patients' individual treatment schedules were used to group fractions with similar amounts of delivered dose together. The total biological effective dose (BED) was calculated for each patient using an alpha/beta ratio = 10 (25). Patients were treated with three to five fractions (1–2 weeks), with a total BED ranging from 54.8 to 132 Gy with a median of 100 Gy (physical dose ranging from 30 to 60 Gy and a median of 50 Gy). Treatments were planned and adapted as needed using an isotoxic approach to escalate the dose to the target until organs at risk dose constraints neared their maximums (26).

Patient treatment response was based upon post-surgery pathology or post-treatment imaging studies. Treatment response for patients who had undergone curative-intent resection following SBRT utilized tumor response grading with the College of American Pathologists (TRG-CAP). Response for patients with imaging studies was determined according to modified response evaluation criteria in solid tumors (mRECIST 1.1) (27, 28). A previously used binary response classification scheme was used to assign each patient to the responder (RS) or non-responder (NR) category (17). Response for patients able to undergo resection was determined using TRG-CAP. TRG-CAP scores ≤2 were considered responders and a score = 3 as NR. For patients who did not qualify for curative resection, follow-up imaging using mRECIST 1.1 criteria was utilized and was determined using dynamic CT, MRI, or PET images acquired within 1–3 months after SBRT.

Daily Setup Images and Volume Delineation

Daily images were acquired on the 0.35-T split-bore MRI-guided radiation treatment unit prior to each daily treatment (MRIdian, ViewRay Inc., Cleveland, OH). Images utilized for this work were acquired using the clinical pulse sequence. The sequence is a true fast imaging with balanced steady-state free procession pulse sequence (bSSFP). bSSFP is a popular choice for MR-guided RT systems because of the short acquisition time needed for volumetric image acquisitions, high signal contrast in soft tissue, insensitivity to motion, and minimal spatial integrity perturbation. The images contained a blend of T1- and T2-weighted visual characteristics (29). Clinical setup images were acquired with $1.5 \times 1.5 \times 3.0$ mm^3 voxel dimensions. Three variations of the clinical bSSFP sequence were included and can be seen in **Table 1**. The inclusion of multiple variations of the same pulse sequence was deemed acceptable because of identical acquisition times, identical voxel sizes, identical flip angles, and similar repetition time (TR) and echo time (TE) (30). Any difference in parameters should have a minimal impact on feature value. Images of patients with arms above their heads were acquired using an MR-compatible board and surface coil. After patients finished treatment, the images were exported to MIM Maestro v6.5.5 (MIM, Software, Cleveland, OH) for contouring. The gross tumor volume (GTV) was contoured on each daily setup image by a radiation oncologist with expertise in PDAC.

Radiomic Texture Feature Extraction

After contouring, the GTVs were exported from MIM to be processed in MATLAB 2020b (The MathWorks, Inc., Natick, MA). Prior to radiomic texture feature calculation, the intensity range of each GTV was normalized by limiting the dynamic range using the methods of Collewet et al., the "$\pm 3\sigma$" method (S4) (31). The GTVs were then quantized to 64 gray levels with the histogram equalization method using a combination of in-house programs and publicly available codes (24, 32). Radiomic texture features belonging to gray-level co-occurrence matrix-based (GLCM, IBSI aggregation code IAZD), gray-level size zone matrix-based (GLSZM, IBSI aggregation code KOBO), gray-level run length matrix-based (GLRLM, IBSI aggregation code IAZD), and neighborhood gray tone difference matrix-based (NGTDM, IBSI aggregation code KOBO) were calculated from each daily setup image (33–39). Details regarding aggregation codes and feature names for the 39 radiomic texture features are displayed in **Table 2**. Five radiomic features were modified in accordance with that of Fave et al. to remove some intrinsic volume dependence (40).

All radiomic texture features were calculated *via* the 3-D definitions using the 26-nearest neighbor voxels. GLCM-based features were calculated using a voxel displacement of one. The GLCM is populated by calculating the probability of gray-level intensities occurring together within a given region. While the GLCM records two voxel intensities, GLSZM and GLRLM record connected isotone zones and regions, respectively. Coordinates in the GLSZM correspond to the probability of a zone of a gray level of a size occurring with the region of interest. Similarly, the GLRLM encodes the probability of a gray level occurring in a connected run of a certain size. The NGTDM

TABLE 1 | The three versions of the bSSFP pulse sequence employed by the MR system for alignment of patients prior to radiotherapy delivery.

Acquisition Parameter	Image Pulse Sequences		
	Clinical A	Clinical B	Clinical C
TR/TE (ms)	3.00/1.27	3.33/1.43	3.36/1.44
Bandwidth (Hz/pixel)	604	537	534
Field of view (mm)	540 × 465 × 432	400 × 400 × 432	350 × 350 × 432
Matrix size	360 × 360 × 144	266 × 266 × 144	234 × 234 × 144

All pulse sequences have identical flip angles and voxel dimensions.

TABLE 2 | Second-order radiomic texture features serving as the basis for the delta-radiomics texture features.

Matrix encoding class (IBSI/aggregation code)	Radiomic texture feature (IBSI code)
GLCM (LFYI/IAZD)	Contrast (ACUI)
	Dissimilarity (8S9J)
	Homogeneity (IB1Z)
	Correlation (NI2N)
	Energy (8ZQL)*
	Variance (UR99)
	Entropy (TU9B)
	Sum average (ZGXS)
GLRLM (TPOI/IAZD)	Short run emphasis (220V)
	Long run emphasis (W4KF)
	Gray-level non-uniformity (R5YN)*
	Run length non-uniformity (W92Y)*
	Run percentage (9ZK5)
	Low gray-level run emphasis (V3SW)
	High gray-level run emphasis (G3QZ)
	Short run low gray-level emphasis (HTZT)
	Short run high gray-level emphasis (GD3A)
	Long run low gray-level emphasis (IVPO)
	Long run high gray-level emphasis (3KUM)
	Gray-level variance (8CE5)
	Run length variance (8CE5)
GLSZM (9SAK/KOBO)	Small zone emphasis (5QRC)
	Large zone emphasis (48P8)
	Gray-level non-uniformity (JNSA)
	Zone-size non-uniformity (4JP3)
	Zone percentage (P30P)
	Low gray-level zone emphasis (XMSY)
	High gray-level zone emphasis (5GN9)
	Small zone low gray-level emphasis (5RAI)
	Small zone high gray-level emphasis (HW1V)
	Large zone low gray-level emphasis (YH51)
	Large zone high gray-level emphasis (J17V)
	Gray-level variance (BYLV)
	Zone-size variance (BYLV)
NGTDM (IPET/KOBO)	Coarseness (QCDE)*
	Contrast (65HE)
	Busyness (NQ30)*
	Complexity (HDEZ)
	Strength (1X9X)

*GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run length matrix; GLSZM, gray-level size zone matrix; NGTDM, neighborhood gray tone difference matrix; IBSI, is the Image Biomarker Standardization Initiative. Features marked with * were IBSI features modified to decrease volume independence (39).*

records the average difference between each gray level and the voxels surrounding it. Each radiomic texture feature sums over the probabilities within the class matrix. The result is a single descriptive number for each feature emphasizing different characteristics of gray-level intensity relationships within the image.

Two sets of delta-radiomics texture features were calculated for each patient using the BED/fraction to group features calculated from different time points into similar delivered dose regions. The total BED was divided by the number of fractions and serves to represent features as a function of time and of dose. Delta-radiomics features were calculated by taking the difference in feature values between pre-radiation images (fraction 1 setup images) and then after a certain amount of dose was delivered. Specifically, for this patient cohort, the following dose bins of 20 and 40 Gy were used and named BED20 and BED40, respectively. BED20 features were calculated after the delivery of 20–30 Gy and BED40 features after 40 Gy.

Radiomic Texture Feature Selection and Analysis

A supervised machine learning approach, random forest (RF), was used to explore the data and select potentially predictive features using a two-step process (41). RF is an ensemble method that creates many weaker learners called decision trees by splitting the training data and combining many decision trees into a forest. Combining the weak learners allows the model to overcome variability in the data and is considered a strong learner (42). The first step of the process was to select potentially prognostic values using the Gini Index that quantifies how important each feature is for accurate predictions during RF model training (43). An RF model was trained using BED20 or BED40 delta-radiomics texture features to predict binary patient outcome. The Gini Index, a measure of variable importance for predictive accuracy, was obtained during model training by calculating the out-of-bag error using bootstrapped datasets from building 500 trees (weak learners) in R (R Foundation for Statistical Computing, Vienna, Austria) (42). The top 2 ranking features were evaluated using an internal validation approach to avoid overfitting due to the limited sample size of patients. A bootstrapped logistic regression model was used to evaluate the pre-selected delta-radiomic features by bootstrapping the data 1,000 times and selecting two-thirds (20 patients at a time) to construct the logistic regression. The resulting model was used to predict the outcome for all patients and to calculate area under the receiver operating characteristic curve (AUC) for each iteration. The numbers reported are internal validation AUCs based upon the 2.5 and 97.5 percentile values, and the mean value of the 1,000 iterations.

RESULTS

Patient Database

A total of 30 patients were included for analysis. The RS group consisted of 11 patients with complete or partial responses. The patients with progressive and stable diseases were included in the NR group and totaled 19. **Table 3** displays each patient's response along with the total BED and delivered BED/fraction.

TABLE 3 | Characteristics of patients included for delta-radiomics texture analysis.

Patient Number	Response	Total BED	BED/Fx	BED20	BED40
1	RS	72	14.4	Fx1–Fx3	Fx1–Fx4
2	NR	59.5	11.9	Fx1–Fx3	Fx1–Fx5
3	NR	72	14.4	Fx1–Fx3	Fx1–Fx4
4	NR	100	20	Fx1–Fx2	Fx1–Fx3
5	NR	59.5	11.9	Fx1–Fx3	Fx1–Fx5
6	RS	54.8	11.0	Fx1–Fx3	Fx1–Fx5
7	RS	72	14.4	Fx1–Fx3	Fx1–Fx4
8	NR	72	14.4	Fx1–Fx3	Fx1–Fx4
9	NR	72	14.4	Fx1–Fx3	Fx1–Fx4
10	RS	59.5	11.9	Fx1 - Fx3	Fx1–Fx5
11	NR	100	20.0	Fx1–Fx2	Fx1–Fx3
12	NR	100	20.0	Fx1–Fx2	Fx1–Fx3
13	RS	100	20.0	Fx1–Fx2	Fx1–Fx3
14	RS	132	26.4	Fx1–Fx2	Fx1–Fx3
15	RS	100	20.0	Fx1–Fx2	Fx1–Fx3
16	RS	100	20.0	Fx1–Fx2	Fx1–Fx3
17	NR	100	20.0	Fx1–Fx2	Fx1–Fx3
18	RS	61.5	12.3	Fx1–Fx3	Fx1–Fx5
19	RS	100	20.0	Fx1–Fx2	Fx1–Fx3
20	RS	100	20.0	Fx1–Fx2	Fx1–Fx3
21	NR	100	20.0	Fx1–Fx2	Fx1–Fx3
22	NR	100	20.0	Fx1–Fx2	Fx1–Fx3
23	NR	100	20.0	Fx1–Fx2	Fx1–Fx3
24	NR	100	20.0	Fx1–Fx2	Fx1–Fx3
25	NR	100	20.0	Fx1–Fx2	Fx1–Fx3
26	NR	100	20.0	Fx1–Fx2	Fx1–Fx3
27	NR	100	20.0	Fx1–Fx2	Fx1–Fx3
28	NR	100	20.0	Fx1–Fx2	Fx1–Fx3
29	NR	100	20.0	Fx1–Fx2	Fx1–Fx3
30	NR	72	14.4	Fx1–Fx3	Fx1–Fx4

The second column contains the binary response classification for each patient, the third column is the total BED for the entire treatment, and the column "BED/Fx" is the total BED divided equally between the number of fractions (in Gy). The final two columns contain the fractions used to calculate the difference corresponding to the binning scheme for delivered dose for each type of delta-radiomics texture features.

Feature Selection and Analysis

The delta-radiomics texture features selected by each model as important for prediction during model training can be seen in **Figure 1**. Based on the Gini Index, GLCM sum average and NGTDM busyness were selected for further analysis during RF model training for BED20-based delta-radiomics texture features, as seen in the top half of **Figure 1**. The higher the value of the Gini Index is, the more important the feature is for model predictive accuracy. The other features included were ranked lower based on the Gini Index, indicating less predictive importance. For the BED40 delta-radiomics texture features (bottom of **Figure 1**), GLSZM large zones low gray-level emphasis and GLRLM run percentage were selected as most important for model prediction accuracy. The Gini Index values for the top 6 features were ranked closer to each other than the BED20-based features. The results of the bootstrapped logistic regression modeling using the respective features are displayed in **Table 4**. Delta-radiomics features extracted early in treatment, BED20-based features, achieved higher AUCs than the BED40 features. The mean internal validation AUC obtained during the bootstrapped logistic regression using the BED20 features with the percentile range from 2.5% to 97.5% was AUC = 0.845 [0.794–0.856]. The mean AUC for the BED40 delta-radiomics was AUC = 0.567 [0.502–0.675].

The ROCs calculated during the logistic regression analysis and the RF model ROC estimates are displayed in **Figure 2**. The values of the top 2 BED20 delta-radiomics texture features are displayed in **Figure 3**, with the Y-axis as the mean radiomic texture feature values and the X-axis containing the bin for delivered dose that the feature values were assigned to for RS and NR patients for the two most predictive delta-radiomic texture features selected at the BED20 timepoint. The features are graphed as the percent change from pre-RT values for each feature and group of patients. The RS group generally had small increases in GLCM sum average values (negative percent change from pre-RT values) from pre-radiation features to BED20, then values return to near pre-treatment levels by the BED40 timepoint. Most of the NR group features decreased at BED20 (positive percent change) and returned to approximately pre-radiation levels also. NGTDM busyness increased from pre-radiation values to BED20 for both RS and NR. While the RS group's features had little change between BED20 and BED40, the NR group's features continued to increase in the same period.

DISCUSSION

Delta-radiomics using patient setup imaging was initially explored by Fave et al. using CBCT images of lung patients (23).

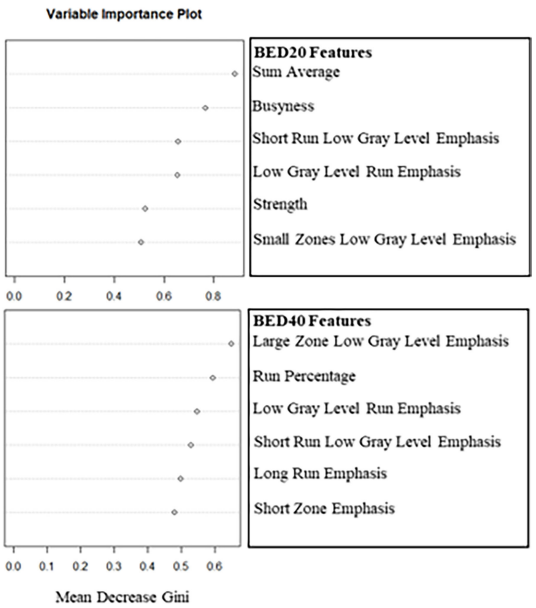


FIGURE 1 | Plots of the mean decrease in the Gini Index used to select the two most important deltaradiomics texture features after training the RF to predict binary outcome. The top plot consists of the BED20-based radiomics texture features and the bottom contains the BED40-based features.

TABLE 4 | The left column contains row labels for the BED20 (middle column) and the BED40 (right column) for relevant radiomic texture features selected by the Gini Index and their internal validation AUCs obtained from bootstrapped logistic regression analysis.

Delta-radiomics feature type:	BED20	BED40
Feature 1	GLCM sum average	Low gray-level run emphasis
Feature 2	Large zones low gray level-emphasis	Run percentage
Mean AUC	0.845	0.567
2.5% AUC value	0.794	0.502
97.5% AUC value	0.856	0.675

The concept of monitoring changes in disease *via* radiomic texture features induced by treatment is beginning to take hold with the increased reliance of imaging in radiation oncology. Jeon et al. calculated delta-radiomics texture features from T2-weighted diagnostic MR images acquired before and after preoperative chemoradiotherapy in locally advanced rectal cancer patients (18). The region of interest was a dynamic range limited using the Collewet method and quantized to 64 intensity levels, as in this work. The study investigated the predictive ability with local recurrence, distant metastasis, and disease-free survival as endpoints. They selected a least absolute shrinkage and selection operator (LASSO) model to preselect delta-radiomics texture features and used a Cox regression analysis to determine the prognostic ability of the features. Delta-radiomic features were predictive of all study outcomes. Interestingly, GLCM sum average and GLSZM large zone low gray-level emphasis were selected in their study as predictive. Only second order texture features were included for this work for a few reasons. The number of features was limited to keep the feature space small for the limited sample size of patients. The texture features included have been a mainstay in radiomics texture analysis studies, since their introduction and previous work conducted by the authors included repeatability analysis of the same features included in this work (24). Lastly, as the feature space begins to include higher order texture features, the interpretability of such features can become very difficult.

Fave's work exposed the possibility of extracting predictive information from low-quality setup images and set the stage for applications in low field strength daily patient setup MR images. Boldrini et al. used 0.35-T setup images of rectal cancer patients acquired at fractions 1, 5, 10, 15, 20, and 25 to predict complete clinical response (16). Delta-radiomics features were defined as the ratio between pre-RT (fraction 1 setup image) subsequent time points. The most predictive timepoint was after the delivery of 22 Gy (physical dose). GLCM energy, GLRLM gray-level nonuniformity, and least axis length were selected as predictive of complete clinical response. Although only 16 patients were included, the selection of GLCM energy as one of the best radiomic predictors is encouraging.

Nasief et al. used CT-based delta-radiomics texture features extracted from pancreatic cancer patients treated with 28 fractions of CT-guided chemoradiation therapy to predict pathological response (22). Delta-radiomics texture features were calculated using the relative change in feature values from fraction 1 to each subsequent fraction. Comparisons of the current study and Nasief et al. should be minimized due to the differences in imaging modalities and fractionation, but Nasief et al. were able to achieve an AUC = 0.94 that is very encouraging of the performance of delta-radiomics texture features. A recent study by Cusumano et al. used 35 pancreatic cancer patients and considered radiomic texture features extracted from time points defined by delivered BEDs of 20, 40, and 60 Gy to predict local control following MR-guided RT (44). The patient library used for their work included patients treated with 5 fractions (17 patients) and 15 fractions (18 patients), and the use of BED was used to group patients to similar points in treatment. In our current study, the BED was used to group changes in features with delivered BED. Cusumano et al. extracted a total of 60 radiomic texture features that adhered to IBSI feature definitions. One feature calculated at the BED 40 Gy time point was predictive of local control at 1 year, with AUC = 0.78 and with 95% confidence interval of 0.61–0.94 but did not overlap with the features selected in this study. The large confidence interval could easily be attributed to the heterogeneity of fractionation of the patient population, and it is possible that different features could be prognostic at different points in time for different study endpoints. The pancreatic patients in this current work are all treated with SBRT, while Cusumano et al. used hypo-fractionated and SBRT patients. The results of this work suggested that changes induced from a baseline pre-RT value

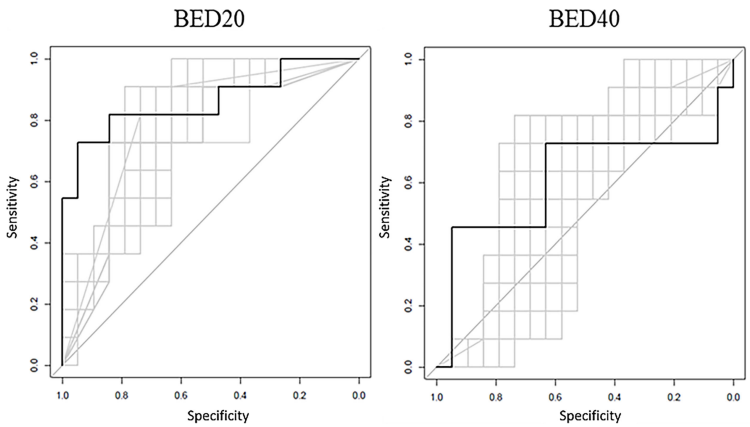


FIGURE 2 | The black ROC curves represent the RF model estimates based on model training. The RF model AUC = 0.876 for BED20 and AUC = 0.601 for the BED40 RF model. The gray curves represent the returned bootstrapped logistic regression produced using the top two features, with an average AUC = 0.845 for BED20 and AUC = 0.567 for BED40-based features.

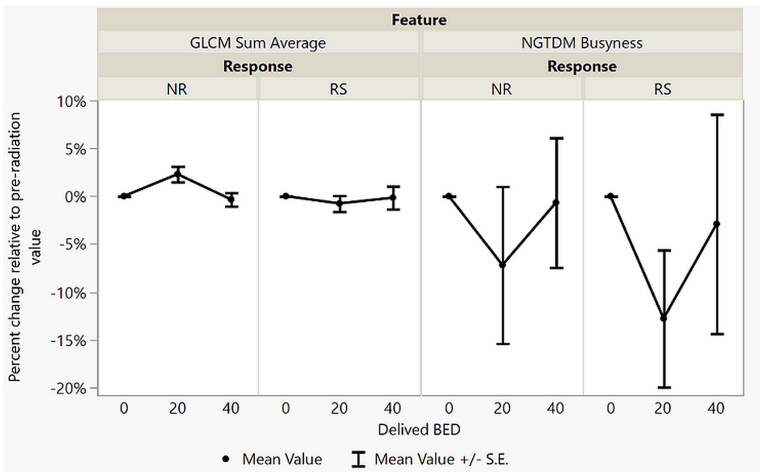


FIGURE 3 | The values of the top two performing texture features in terms of internal validation AUC from the BED20 model. Y-axes are the percent change relative to the pre-RT (0 Gy) values graphed at each point of 20 Gy and 40 Gy. The plots display the mean percent change in each feature value, the points, the ranges represent the range of the standard error from the means.

early in SBRT may have clinical utility. With both studies, the way response has been defined is likely an oversimplification of response, and the underlying biology and the accuracy of delta-radiomics feature analysis could be suppressed.

There are a few shortcomings of this study. Besides possibly oversimplifying response, the use of mRECIST for response determination is less than ideal, as it relies itself on imaging-based evaluation. Expansion of the library to include more patients with pathology-based responses could significantly strengthen results, as the response for 10 patients was determined with pathology (versus 20 using mRECIST) in this work. The development and further validation of more specific response evaluation methods, such as PET response criteria in solid tumors (PERCIST) and/or metabolic imaging, might help increase certainty of classifying patients as RS and NR (45). The relatively small sample size may not represent a fair sample of the true patient population, and a larger patient cohort is being pursued that will allow external validation. The use of the internal validation AUC in this study served its purpose, providing a common measuring stick to evaluate whether delta-radiomics texture analysis is feasible and could contain predictive utility. Although the behavior of the features in **Figure 3** is easy to see when graphed and analyzed, it is disconnected from any underlying changes to the tumor microenvironment. In addition, from **Table 3**, it is apparent that the time and range of dose delivered between the pre-RT (fraction 1 setup scan) and BED40

point compared to the BED20 point is spread out. Without correlation studies and larger patient cohorts, it is difficult to understand that the underpinnings GLCM sum average and NGTDM busyness are quantifying. This is a portion of research that radiomics texture features must bridge prior to wide clinical acceptance. The BED20 delta-radiomics texture features could be capturing and quantifying an initial inflammatory mechanism, a change in tumor signaling, or even just the altering of the physical cellular structure after radiation damage. It is possible that by the BED40 dataset's acquisition, the delta-radiomics has become too heterogenous to extract any predictive features in the small cohort of 30 patients, or these mechanisms have simply stopped. Lastly, but always prevalent in radiomic texture analysis studies, variability in target volume contouring may impact results.

CONCLUSION

This work sought to explore the application and feasibility of delta-radiomics texture analysis in low field strength daily MR setup images. While it remains to be seen if the predictive power of the delta-radiomic texture features selected and the method of analysis utilized in this study will hold in a larger patient cohort, the application of delta-radiomics should be further pursued. The workflow developed in this study could be easily modified and applied to larger data sets as a step in building clinical models and applied to other clinical sites. Changes detected early in treatment could provide valuable information with time left for physicians to consider changes to the treatment regime, including further dose escalation or alternative immunotherapies. Further expansion of the patient library using more patients, external validation, and exploration of a multi-institutional dataset for building a model would appear to be justified.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the raw image data will not be shared due to ongoing data collection. Requests to access the datasets should be directed to Garrett Simpson, GNS24@miami.edu.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by ViewRay Omnibus Patient Registry and Image Analysis Protocol, IRB#: 20160817, University of Miami IRB. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

GNS performed texture analysis and statistical analysis, drafted manuscript, and prepared the figures/tables. WJ gathered patient response data, verified contouring, and made manuscript contributions. BS gathered patient response data, verified contouring, and made manuscript contributions. LP gathered patient response data, verified contouring, and made manuscript contributions. EM verified patient contouring and made manuscript contributions. DK was the statistician who contributed to machine learning work. JCF contributed to texture analysis, methods, figures, tables, and manuscript sections. ND contributed to texture analysis, methods, figures, tables, and manuscript sections. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
2. Almendro V, Marusyk A, Polyak K. Cellular Heterogeneity and Molecular Evolution in Cancer. In: AK Abbas, SJ Galli, PM Howley, editors. *Annual Review of Pathology: Mechanisms of Disease*, vol. 8. Palo Alto: Annual Reviews (2013). p. 277–302.
3. Tu CF, Qi P, Li XY, Mo YZ, Li XL, Xiong W, et al. Tumor Heterogeneity: The Challenge of Precision Medicine. *Prog Biochem Biophys* (2015) 42(10):881–90. doi: 10.3390/ijms19123733
4. Aerts H, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Cavalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5:8. doi: 10.1038/ncomms5006
5. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout R, Granton P, et al. Radiomics: Extracting More Information From Medical Images Using Advanced Feature Analysis. *Eur J Cancer* (2012) 48(4):441–6. doi: 10.1016/j.ejca.2011.11.036
6. Avanzo M, Stancanello J, El Naqa I. Beyond Imaging: The Promise of Radiomics. *Phys Med* (2017) 38:122–39. doi: 10.1016/j.ejmp.2017.05.071
7. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. *Radiology* (2016) 281(3):947–57. doi: 10.1148/radiol.2016152234
8. Li H, Zhu YT, Burnside ES, Huang E, Drukker K, Hoadley KA, et al. Quantitative MRI Radiomics in the Prediction of Molecular Classifications of Breast Cancer Subtypes in the TCGA/TCIA Data Set. *NPJ Breast Cancer* (2016) 2:16012. doi: 10.1038/npjbcancer.2016.12
9. Lucia F, Visvikis D, Desseroit M-C, Miranda O, Malhaire J-P, Robin P, et al. Prediction of Outcome Using Pretreatment 18F-FDG PET/CT and MRI Radiomics in Locally Advanced Cervical Cancer Treated With Chemoradiotherapy. *Eur J Nucl Med Mol Imaging* (2018) 45(5):768–86. doi: 10.1007/s00259-017-3898-7
10. Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2020. *CA Cancer J Clin* (2020) 70(1):7–30. doi: 10.3322/caac.21590
11. Garcea G, Dennison AR, Pattenden CJ, Neal CP, Sutton CD, Berry DP. Survival Following Curative Resection for Pancreatic Ductal Adenocarcinoma. A Systematic Review of the Literature. *JOP* (2008) 9(2):99–132.
12. Chuong MD, Springett GM, Freilich JM, Park CK, Weber JM, Mellon EA, et al. Stereotactic Body Radiation Therapy for Locally Advanced and Borderline Resectable Pancreatic Cancer Is Effective and Well Tolerated. *Int J Radiat Oncol Biol Phys* (2013) 86(3):516–22. doi: 10.1016/j.ijrobp.2013.02.022
13. Herman JM, Chang DT, Goodman KA, Dholakia AS, Raman SP, Hacker-Prietz A, et al. Phase 2 Multi-Institutional Trial Evaluating Gemcitabine and Stereotactic Body Radiotherapy for Patients With Locally Advanced Unresectable Pancreatic Adenocarcinoma. *Cancer* (2015) 121(7):1128–37. doi: 10.1002/cncr.29161
14. Shaib WL, Hawk N, Cassidy RJ, Chen Z, Zhang C, Brutcher E, et al. A Phase 1 Study of Stereotactic Body Radiation Therapy Dose Escalation for Borderline

Resectable Pancreatic Cancer After Modified FOLFIRINOX (Nct01446458). *Int J Radiat Oncol Biol Phys* (2016) 96(2):296–303. doi: 10.1016/j.ijrobp.2016.05.010
15. Mellon EA, Jin WH, Frakes JM, Centeno BA, Strom TJ, Springett GM, et al. Predictors and Survival for Pathologic Tumor Response Grade in Borderline Resectable and Locally Advanced Pancreatic Cancer Treated With Induction Chemotherapy and Neoadjuvant Stereotactic Body Radiotherapy. *Acta Oncol (Stockholm Sweden)* (2017) 56(3):391–7. doi: 10.1080/0284186X.2016.1256497
16. Boldrini L, Cusumano D, Chiloiro G, Casà C, Masciocchi C, Lenkowicz J, et al. Delta Radiomics for Rectal Cancer Response Prediction With Hybrid 0.35 T Magnetic Resonance-Guided Radiotherapy (MRgRT): A Hypothesis-Generating Study for an Innovative Personalized Medicine Approach. *Radiol Med* (2018) 124:145–53. doi: 10.1007/s11547-018-0951-y
17. Simpson G, Spieler B, Dogan N, Portelance L, Mellon EA, Kwon D, et al. Predictive Value of 0.35 T Magnetic Resonance Imaging Radiomic Features in Stereotactic Ablative Body Radiotherapy of Pancreatic Cancer: A Pilot Study. *Med Phys* (2020) 47(8):3682–90. doi: 10.1002/mp.14200
18. Bauer TM, El-Rayes BF, Li XB, Hammad N, Philip PA, Shields AF, et al. Carbohydrate Antigen 19-9 is a Prognostic and Predictive Biomarker in Patients With Advanced Pancreatic Cancer Who Receive Gemcitabine-Containing Chemotherapy. *Cancer* (2013) 119(2):285–92. doi: 10.1002/cncr.27734
19. Plautz TE, Zheng C, Noid G, Li XA. Time Stability of Delta-Radiomics Features and the Impact on Patient Analysis in Longitudinal CT Images. *Med Phys* (2019) 46(4):1663–76. doi: 10.1002/mp.13395
20. Lin P, Yang PF, Chen S, Shao YY, Xu L, Wu Y, et al. A Delta-Radiomics Model for Preoperative Evaluation of Neoadjuvant Chemotherapy Response in High-Grade Osteosarcoma. *Cancer Imaging* (2020) 20(1):7. doi: 10.1186/s40644-019-0283-8
21. Carles M, Fechter T, Grosu AL, Sorensen A, Thomann B, Stoian RG, et al. F-18-FMISO-PET Hypoxia Monitoring for Head-and-Neck Cancer Patients: Radiomics Analyses Predict the Outcome of Chemo-Radiotherapy. *Cancers* (2021) 13(14):3449. doi: 10.3390/cancers13143449
22. Nasief H, Zheng C, Schott D, Hall W, Tsai S, Erickson B, et al. A Machine Learning Based Delta-Radiomics Process for Early Prediction of Treatment Response of Pancreatic Cancer. *NPJ Precis Oncol* (2019) 3(1):25. doi: 10.1038/s41698-019-0096-z
23. Fave X, Zhang LF, Yang JZ, Mackin D, Balter P, Gomez D, et al. Delta-Radiomics Features for the Prediction of Patient Outcomes in Non-Small Cell Lung Cancer. *Sci Rep* (2017) 7:11. doi: 10.1038/s41598-017-00665-z
24. Simpson G, Ford JC, Llorente R, Portelance L, Yang F, Mellon EA, et al. Impact of Quantization Algorithm and Number of Gray Level Intensities on Variability and Repeatability of Low Field Strength Magnetic Resonance Image-Based Radiomics Texture Features. *Phys Med* (2020) 80:209–20. doi: 10.1016/j.ejmp.2020.10.029
25. Fowler JF. The Linear-Quadratic Formula and Progress in Fractionated Radiotherapy. *Br J Radiol* (1989) 62(740):679–94. doi: 10.1259/0007-1285-62-740-679
26. Mihaylov IB, Mellon EA, Yechieli R, Portelance L. Automated Inverse Optimization Facilitates Lower Doses to Normal Tissue in Pancreatic Stereotactic Body Radiotherapy. *PloS One* (2018) 13(1):e0191036. doi: 10.1371/journal.pone.0191036
27. Chatterjee D, Katz MH, Rashid A, Varadhachary GR, Wolff RA, Wang H, et al. Histologic Grading of the Extent of Residual Carcinoma Following Neoadjuvant Chemoradiation in Pancreatic Ductal Adenocarcinoma: A Predictor for Patient Outcome. *Cancer* (2012) 118(12):3182–90. doi: 10.1002/cncr.26651
28. Schwartz LH, Litière S, de Vries E, Ford R, Gwyther S, Mandrekar S, et al. RECIST 1.1-Update and Clarification: From the RECIST Committee. *Eur J Cancer* (2016) 62:132–7. doi: 10.1016/j.ejca.2016.03.081
29. Huang TY, Huang IJ, Chen CY, Scheffler K, Chung HW, Cheng HC. Are TrueFISP Images T2/T1-Weighted? *Magn Reson Med* (2002) 48(4):684–8. doi: 10.1002/mrm.10260
30. Mayerhoefer ME, Szomolanyi P, Jirak D, Materka A, Trattnig S. Effects of MRI Acquisition Parameter Variations and Protocol Heterogeneity on the Results of Texture Analysis and Pattern Discrimination: An Application-Oriented Study. *Med Phys* (2009) 36(4):1236–43. doi: 10.1118/1.3081408
31. Collewet G, Strzelecki M, Mariette F. Influence of MRI Acquisition Protocols and Image Intensity Normalization Methods on Texture Classification. *Magn Reson Imaging* (2004) 22(1):81–91. doi: 10.1016/j.mri.2003.09.001
32. Vallières M. A Radiomics Model From Joint FDG-PET and MRI Texture Features for the Prediction of Lun Metastases in Soft-Tissue Sarcomas of the Extremities. *Phys Med Biol* (2015) 60(14):5471–96. doi: 10.1088/0031-9155/60/14/5471
33. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. *IEEE Trans Syst Man Cybern* (1973) SMC3(6):610–21. doi: 10.1109/TSMC.1973.4309314
34. Thibault G, Fertil B, Navarro C, Pereira S, Lévy N, Sequeira J, et al. *Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification*. Conference Paper at 10th International Conference on Pattern Recognition and Information Processing, PRIP 2009 in Minsk, Belarus (2009).
35. Haralick RM. Statistical and Structural Approaches to Texture. *Proc IEEE* (1979) 67(5):786–804. doi: 10.1109/PROC.1979.11328
36. Amadasun M, King R. Textural Features Corresponding to Textural Properties. *IEEE Trans Syst Man Cybern* (1989) 19(5):1264–74. doi: 10.1109/21.44046
37. Dasarathy BV, Holder EB. Image Characterizations Based on Joint Gray Level—Run Length Distributions. *Pattern Recognit Lett* (1991) 12(8):497–502. doi: 10.1016/0167-8655(91)80014-2
38. Chu A, Sehgal CM, Greenleaf JF. Use of Gray Value Distribution of Run Lengths for Texture Analysis. *Pattern Recognit Lett* (1990) 11(6):415–9. doi: 10.1016/0167-8655(90)90112-F
39. Zwanenburg A, Vallières M, Abdalah M, Aerts H, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative:Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* (2020) 295(2):328–38. doi: 10.1148/radiol.202019 1145
40. Fave X, Zhang L, Yang J. Impact of Image Preprocessing on the Volume Dependence and Prognostic Potential of Radiomics Features in Non-Small Cell Lung Cancer. *Transl Cancer Res* (2016) 5:349. doi: 10.21037/tcr.2016.07.11
41. Kwon D, Reis IM, Breto AL, Tschudi Y, Gautney N, Zavala-Romero O, et al. *Classification of Suspicious Lesions on Prostate Multiparametric MRI Using Machine Learning*. SPIE (2018) p. 1–9.
42. Breiman L. Random Forests. *Mach Learn* (2001) 45(1):5–32. doi: 10.1023/A:1010933404324
43. Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, et al. A Comparison of Random Forest and its Gini Importance With Standard Chemometric Methods for the Feature Selection and Classification of Spectral Data. *BMC Bioinfrm* (2009) 10(1):213. doi: 10.1186/1471-2105-10-213
44. Cusumano D, Boldrini L, Yadav P, Casà C, Lee SL, Romano A, et al. Delta Radiomics Analysis for Local Control Prediction in Pancreatic Cancer Patients Treated Using Magnetic Resonance Guided Radiotherapy. *Diagnostics* (2021) 11(1):72. doi: 10.3390/diagnostics11010072
45. Padhan A, Liu G, Mu-Koh D, Chenevert T, Thoeny H, Takahara T, et al. Diffusion-Weighted Magnetic Resonance Imaging as a Cancer Biomarker: Consensus and Recommendations. *Neoplasia* (2009) 11(2):102–25. doi: 10.1002/mp.14200

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Simpson, Jin, Spieler, Portelance, Mellon, Kwon, Ford and Dogan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 08 July 2022
doi: 10.3389/fonc.2022.757389

The Value of CT-Based Radiomics for Predicting Spread Through Air Spaces in Stage IA Lung Adenocarcinoma

Xiaoyu Han[1,2†], Jun Fan[3†], Yuting Zheng[1,2†], Chengyu Ding[4], Xiaohui Zhang[4], Kailu Zhang[1,2], Na Wang[3], Xi Jia[1,2], Yumin Li[1,2], Jia Liu[1,2], Jinlong Zheng[1,2] and Heshui Shi[1,2*]*

[1] Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2] Hubei Province Key Laboratory of Molecular Imaging, Wuhan, China, [3] Department of Pathology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [4] Clinical Solution, Philips Healthcare, Shanghai, China

OPEN ACCESS

Edited by:
Joseph Ross Mitchell, University of Alberta, Canada

Reviewed by:
Alessandro Gonfiotti, University of Florence, Italy
Jinwei Qiang, Jinshan Hospital, China

***Correspondence:**
Jinlong Zheng
zhengjl40@163.com
Heshui Shi
heshuishi@hust.edu.cn

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Cancer Imaging and Image-directed Interventions, a section of the journal Frontiers in Oncology

Received: 12 August 2021
Accepted: 17 June 2022
Published: 08 July 2022

Citation:
Han X, Fan J, Zheng Y, Ding C, Zhang X, Zhang K, Wang N, Jia X, Li Y, Liu J, Zheng J and Shi H (2022) The Value of CT-Based Radiomics for Predicting Spread Through Air Spaces in Stage IA Lung Adenocarcinoma. Front. Oncol. 12:757389. doi: 10.3389/fonc.2022.757389

Objectives: Spread through air spaces (STAS), a new invasive pattern in lung adenocarcinoma (LUAD), is a risk factor for poor outcome in early-stage LUAD. This study aimed to develop and validate a CT-based radiomics model for predicting STAS in stage IA LUAD.

Methods: A total of 395 patients (169 STAS positive and 226 STAS negative cases, including 316 and 79 patients in the training and test sets, respectively) with stage IA LUAD before surgery were retrospectively included. On all CT images, tumor size, types of nodules (solid, mix ground-glass opacities [mGGO] and pure GGO [pGGO]), and GGO percentage were recorded. Region of interest (ROI) segmentation was performed semi-automatically, and 1,037 radiomics features were extracted from every segmented lesion. Intraclass correlation coefficients (ICCs), Pearson's correlation analysis and least absolute shrinkage and selection operator (LASSO) penalized logistic regression were used to filter unstable (ICC < 0.75) and redundant features (r > 0.8). A temporary model was established by multivariable logistic regression (LR) analysis based on selected radiomics features. Then, seven radiomics features contributing the most were selected for establishing the radiomics model. We then built two predictive models (clinical-CT model and MixModel) based on clinical and CT features only, and the combination of clinical-CT and Rad-score, respectively. The performances of these three models were assessed.

Results: The radiomics model achieved good performance with an area under of curve (AUC) of 0.812 in the training set, versus 0.850 in the test set. Furthermore, compared

with the clinical-CT model, both radiomics model and MixModel showed higher AUC and better net benefit to patients in the training and test cohorts.

Conclusion: The CT-based radiomics model showed satisfying diagnostic performance in early-stage LUAD for preoperatively predicting STAS, with superiority over the clinical-CT model.

Keywords: adenocarcinoma, radiomics, spread through air spaces, lung cancer, stage IA

INTRODUCTION

In 2015, spread through air spaces (STAS), a novel invasive pattern in lung cancer, was recognized by the World Health Organization (WHO) Classification (1). STAS has been reported to occur in 14.8-56.4% of lung adenocarcinomas (LUADs) (2–5). Recent reports have shown that the presence of STAS is an independent risk factor for recurrence and low overall survival in small or early-stage LUAD (4, 5). Moreover, preoperative detection of STAS could help choose an appropriate surgery type (6, 7). Correspondingly, Ren et al (6) reported patients with STAS undergoing sublobar resection have a higher rate of pulmonary metastases than patients with STAS administered a lobectomy (25.8% vs 8.2%). Another study (7) showed the presence of STAS is associated with higher cumulative incidence of recurrence (CIR) and death (CID) in patients with sublobar resection compared with those undergoing lobectomy (5-year CIR, 39% vs. 16%; 5-year CID, 16% vs. 8%). Thus, it is critical to determine the STAS status in LUAD prior to the surgical decision-making.

It has been previously reported that some CT findings, including maximum tumor diameter, nodule type and percentage of the solid component, are related to STAS, with promising diagnostic efficacy (0.64-0.77) (8–11). However, the identification of these CT-based morphological features depends on the radiologist's experience. Furthermore, the use of such qualitative CT features to predict STAS could inevitably lead to inestimable misdiagnosis and overdiagnosis. Radiomics is a characterization algorithm that can extract and analyze a large number of quantitative image features from medical images (12). Numerous studies have revealed the promising potential of radiomics in predicting gene mutations (13, 14), lymph node metastasis (15), therapeutic response (16) and clinical prognosis (17) in lung cancer. Two recent studies performing radiomics analysis of STAS have predicted the existence of STAS by establishing different models (18, 19); the established radiomics models achieved moderate performances for STAS prediction with areas under the curves (AUCs) of 0.63 and 0.754, respectively. However, none of the two studies compared the radiomics signature-based and clinical or CT morphological features-based models. This is of great interest because the introduction of radiomics into routine the clinical workflow is unlikely to be accepted if it does not provide additional predictive value compared to clinical factors or morphological CT features. Consequently, in the present study, in addition to establishing a radiomics model and assessing its capabilities, we simultaneously developed the clinical-CT and mixed models, and compared their predictive values for the STAS status.

MATERIALS AND METHODS

This retrospective study was approved by the Ethics Committee of Wuhan Union Hospital (S377), and the requirement for written informed consent was waived.

Patients and Inclusion Criteria

A total of 1051 patients with Stage IA adenocarcinoma (T1a-cN0M0) confirmed by curative surgery between September 2015 and July 2021 in Wuhan Union Hospital were retrospectively assessed. Then, 126 patients were excluded according to the following exclusion criteria (1): previous chemoradiotherapy (n=45); (2) a history of lung operation (n=17); (3) no thin-section CT before treatment (n=44); or (4) no plain chest CT imaging (n=30, **Figure 1**). As shown in **Figure 1**, the incidence of STAS in our institution was relatively low (169/923,18.3%). To overcome potential data imbalance, we randomly divided the STAS-negative cases by 3:7 into groups and matched them with the STAS-positive group at a nearly 1:1 ratio. Such method for balancing data has been validated in previous studies (11, 18, 20). Due to the fact that the incidence of STAS negative is higher than that of positive cases, we divided the negative data by a ratio of 3:7(226 cases) rather than 2:8(150 cases) to ensure the STAS negative cases were slightly larger than the positive ones. In addition, we had reperformed these models by 2:8 ratio for divide the negative data, and found that this ratio does not affect our main conclusions. In total, 395 patients (226 STAS negative and 169 STAS positive cases) were included, which were randomly assigned into training database (316 patients) and test dataset (79 patients), with a ratio of 0.8:0.2. In the train cohort, 136 patients were presented as STAS positive while 180 were negative. In the test cohort, 33 cases were positive for STAS while 46 were negative for STAS. All included patients had single lung adenocarcinoma.

The patients' clinical characteristics and pathological findings, including age, sex, smoking history, histologic subtypes, surgical margin, lymphatic metastasis, epidermal growth factor receptor (EGFR) mutations and anaplastic large-cell lymphoma kinase (ALK) were recorded. A history of smoking was defined as lifetime exposure to more than 100 cigarettes. TNM staging was performed according to the IASLC 8th TNM Lung Cancer Staging System (21).

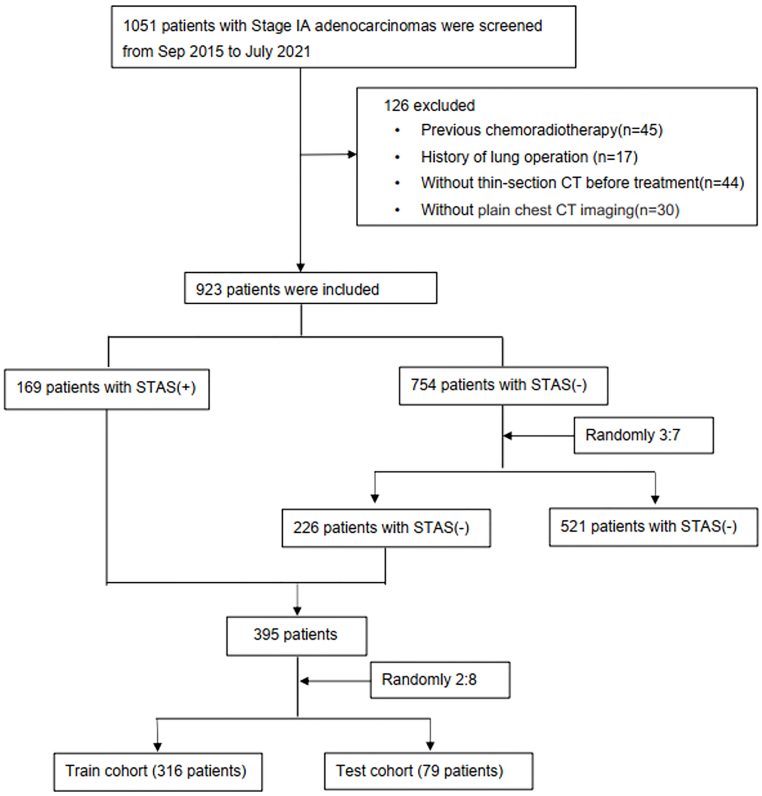


FIGURE 1 | Study flowchart. STAS, spread through air spaces; STAS+, presence of STAS; STAS-, absence of STAS.

Histopathological Analysis

Two pathologists (JF and NW, with 4 and 10 years of experience in thoracic pathology, respectively) blinded to clinical findings re-evaluated the hematoxylin and eosin-stained slides of all included patients using a multiheaded microscope. The predominant subtype of lung adenocarcinoma was assessed based on the International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society multidisciplinary classification of LUAD. STAS positivity was defined according to the WHO definition of STAS as tumor cells were found in the lung air spaces beyond the edge of the primary tumor, which is mainly composed of the following three forms: (1) air spaces filled by micropapillary structure without central fibrovascular cores; (2) air spaces filled by the solid component of the tumor; (3) air spaces filled by multiple discrete and single cells(1). After independent assessments, differences were resolved by consensus.

CT Acquisition

CT was performed on two multislice spiral CT scanners (SOMATOM Definition AS+ and Siemens Healthineers, Germany) at our institution, using the following parameters: detector collimation widths, 64 x 0.6 mm and 128 x 0.6 mm; tube voltages, 120 kV. The tube current was regulated by an automatic exposure control system (CARE Dose 4D). Images were reconstructed at a slice thickness of 1.5 mm or 1 mm and an interval of 1.5 mm or 1 mm. No contrast medium was used. Digital Imaging and Communications in Medicine (DICOM) images from the picture archiving and communication system (PACS) were imported to the 3D-slicer software.

Two senior radiologists (HSS and JLZ, with 31 and 25 years of experience in thoracic radiology, respectively) evaluated the CT images to determine tumor size (longest diameter in MPR images), tumor density type (solid, mix ground-glass opacities [mGGO], and pure GGO [pGGO]), and GGO ratio (GGO diameter/tumor diameter) in consensus on the PACS. These specific CT features were chosen as the most contributing risk factors for STAS in patients with LUAD (8–11). The above two radiologists blinded to clinical and histologic findings assessed CT features on both axial CT and multiplanar reconstruction (MPR) images.

Radiomics Feature Extraction

The regions of interest on CT images were semi-automatically delineated layer by layer by three junior radiologists (XH, YZ and JX with 5, 3 and 2 years of experience in thoracic imaging, respectively). All the three radiologists were aware of tumor presence and location but unaware of the pathological reports and STAS status. The 3D-slicer software was used for segmenting the lesions on each slice of CT scans semi-automatically and independently. Then, the three-dimensional volumes of interest (3D-VOIs) of tumors were automatically reconstructed with the

3D-slicer software (**Figure 2**). Two senior radiologists (HSS and JLZ) were responsible for checking all tumor segmentations, and any deviations were addressed with additional corrections. As reported in a recent publication (22), the semi-automatic segmentation from the 3D-Slicer is a better alternative to manual segmentation, as it can produce more robust and reproducible radiomic features. In addition, we randomly selected 40 cases at a ratio of 1:10 of all sample (396 cases) for estimating intraclass correlation coefficients (ICCs) analysis, which could ensure the precision and assurance of results (23). One observer (XH) repeated the segmentation after one week for intra-observer variability analysis. The other observer (YZ) performed the segmentation on the same image set using the same method, for inter-observer variability analysis.

The Philips Radiomics Tool (Philips Healthcare, China) was used for radiomics feature extraction, and core feature calculation was based on pyRadiomics (24). A B-spline curve interpolation algorithm was used to resample each 3D CT image to a spacing of (0.7, 0.7, 1.0) mm. For each VOI, a total of 1037 3D-radiomic features, including direct, wavelet transformed, logarithm transformed and gradient filtered features, were extracted (types and numbers are shown in **Figure 2**, and details can be also found at pyradiomics.readthedocs.io/en/latest/features.html).

Statistical Analysis

The SPSS software (SPSS, version 21, IBM, Chicago, IL, USA) and R (version 4.0.2; http://www.Rproject.org) were used for all statistical analyses. LASSO binary logistic regression was performed with the 'glmnet' package. Multivariate binary logistic regression was carried out with the 'rms' package. Decision curve analysis was performed with the 'rmda' package. Receiver operating characteristic (ROC) analysis was carried out with the 'pROC' package. Descriptive analysis was performed to describe the distribution of the variables of interest for the training and testing cohorts. All continuous data and categorical variables were expressed as mean ± standard deviation and frequency (percentage), respectively. Independent samples Student's *t* test was applied to compare continuous data in two groups, and the chi-square test was used to compare categorical variables. $P<0.05$ (two-tailed) was considered statistically significant.

All radiomics features were normalized to the z-score. Intraclass correlation coefficient (ICC) determination and Pearson's correlation analysis were performed to exclude redundant and unstable features (r>0.8, ICC>0.75). ICC < 0.5, between 0.5 and 0.75, between 0.75 and 0.9, and > 0.90 indicated poor, moderate, good, and excellent reliability, respectively (25). Least absolute shrinkage and selection operator (LASSO) analysis was performed to determine features for further assessment (26). Five-fold cross-validation and maximum area under the curve (AUC) were used as feature filtering criteria. A classification model based on the selected features was established by a multivariate logistic regression (LR) algorithm. Feature selection was performed on the training set. In this work, we selected no model using the one-standard-error rule as the final model, because a five-fold cross-validation LASSO analysis was performed to filter features. A temporary model was built behind the LASSO analysis as to preserve features that are significant enough for the feature coefficients. More features kept after the LASSO analysis reflect a better performance for the final multivariate logistic regression algorithm (27). Three models were built based on three sets of features: a radiomics model based on the most contributing radiomics features from the temporary model, a clinical and CT model based on clinical and CT features only, and a mixed model (MixModel) combined

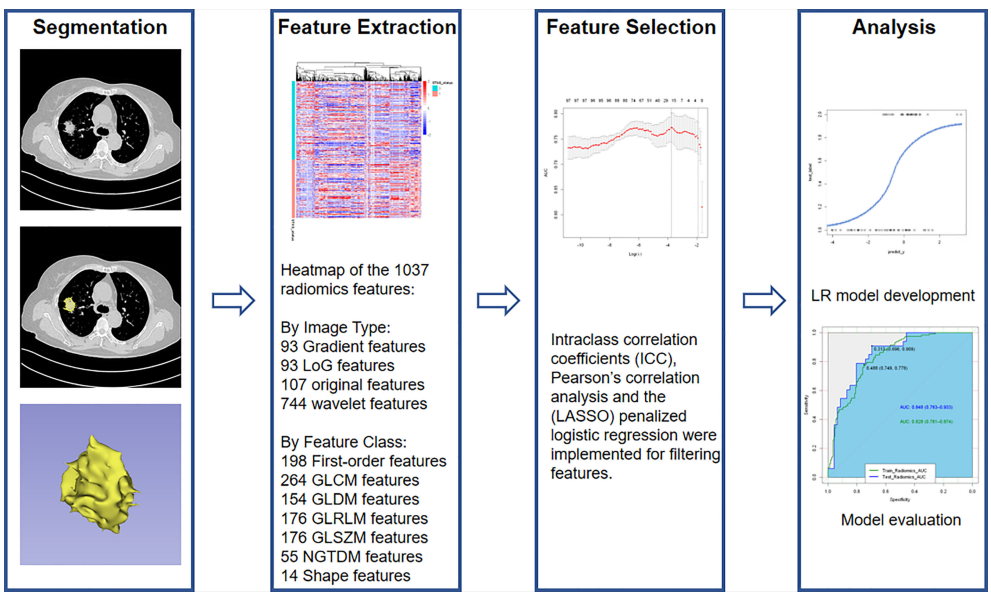


FIGURE 2 | Radiomics workflow.

the Rad-score with clinical and CT characteristics. The performances of these models for predicting STAS were evaluated in the training set, then in the test set by plotting ROC curves and calculating the areas under the curves (AUCs). The accuracy, sensitivity, specificity, negative predictive value (NPV) and positive predictive value (PPV) were then calculated for each model. The predictive ability was also illustrated by the confounder matrix. The DeLong test was conducted to compare diagnostic efficiency among the different models. Furthermore, decision curve analysis was preformed to determine the clinical usefulness of the three models by quantifying the net benefits at different threshold probabilities in the data set.

RESULTS

Clinicopathological and CT Features of STAS

This study included 395 eligible patients (average age, 59 ± 10 years; 207 males) in total. The clinicopathological and main CT features of the final study population are shown in **Table 1**. STAS was found in 169/928 (18.2%) patients. No significant differences were found between STAS-positive and STAS-negative cases in age, gender and smoking history (p=0.268, p=0.232 and p=0.053, respectively). Concerning CT features, tumors with STAS tended to be larger than STAS negative counterparts (21 ± 6.3 mm vs 18 ± 6.7 mm, p<0.001). Tumor densities differed between the STAS and non-STAS groups (p<0.001). The majority of STAS-positive tumors manifested as solid nodules (152/169, 89.9%; **Figure 3**), followed by the mix GGO (13/169, 7.7%) and pure GGO (4/169, 2.4%) groups. Moreover, the less the GGO ratio, the higher the possibility of STAS positivity (0.08 ± 0.21 vs 0.38 ± 0.42, p<0.001). Then, age, gender, smoking history, diameter, density and GGO ratio were selected for predicting STAS in the clinical and CT model (Clinical-CT Model) by the LR algorithm (**Table 2**). Based on clinical and CT features, the model had an AUC of 0.721 (a sensitivity of 69.9% and a specificity of 61.7%) in the training cohort; the AUC was 0.804 (a sensitivity of 72.7% and a specificity of 76.1%) in the test cohort (**Figure 4**). The distribution of the selected clinical and CT features for the patients with STAS and those without STAS in the training and test cohorts are shown in **Table 3**.

TABLE 1 | Associations of spread through air spaces with Clinicopathological features and CT findings.

Factor	Total patients	STAS (+)	STAS (-)	*p*-value
N	395	169	226	
Gender				0.268
Male	207 (52.4%)	94 (55.6%)	113 (50%)	
Female	188 (47.6%)	75 (44.4%)	113 (50%)	
Age, years	59 ± 10	60 ± 10	58 ± 10	0.232
History of smoking	63 (15.9%)	34 (20.1%)	29 (12.8%)	0.053
Diameter, mm	19 ± 6.7	21 ± 6.3	18 ± 6.7	<0.001*
GGO ratio	0.25 ± 0.37	0.08 ± 0.21	0.38 ± 0.42	<0.001*
Density				<0.001*
pGGO	39 (9.9%)	4 (2.4%)	35 (15.5%)	
mGGO	85 (21.5%)	13 (7.7%)	72 (31.9%)	
Solid	271 (68.6%)	152 (89.9%)	119 (52.7%)	
Histologic subtypes				<0.001*
Lepidic predominant	38 (9.6%)	4 (2.4%)	34 (15%)	
Acinar predominant	171 (43.3%)	74 (43.8%)	97 (42.9%)	
Micropapillary	26 (6.6%)	21 (12.4%)	5 (2.2%)	
Papillary predominant	106 (26.8%)	37 (21.9%)	69 (30.5%)	
Solid predominant	44 (11.1%)	30 (17.8%)	14 (6.2%)	
Mucinous predominant	10 (2.5%)	3 (1.8%)	7 (3.1%)	
Resection margin				0.013*
Negative	373 (94.4%)	154 (91.1%)	219 (96.9%)	
Positive	22 (5.6%)	15 (8.9%)	7 (3.1%)	
Pleural invasion				0.001*
Absence	358 (90.6%)	144 (85.2%)	214 (94.7%)	
Present	37(9.4%)	25 (14.8%)	12 (5.3%)	
Perineural invasion				
Absence	379 (95.9%)	160 (94.7%)	219 (96.9%)	0.266
Present	16 (4.1%)	9 (5.3%)	7 (3.1%)	
EGFR				0.039*
Negative	154/261 (59%)	76/115 (59%)	78/146 (53.4%)	
Positive	107/261 (41%)	39/115 (33.9%)	68/146 (46.6%)	
ALK				
Negative	288/303 (95%)	11/124 (8.9%)	4/179 (2.2%)	0.009*
Positive	15/303 (5%)	113/124 (91.1%)	175/179 (97.8%)	

**P<0.05 based on comparisons between the two groups. Data are mean ± SD or n/N (%). STAS, spread through air spaces; STAS+, presence of spread through air spaces; STAS-, absence of spread through air spaces; EGFR, epidermal growth factor receptor; ALK, anaplastic large-cell lymphoma kinase; GGOs, ground-glass opacities; pGGO, pure GGO; mGGO, mix GGO.*

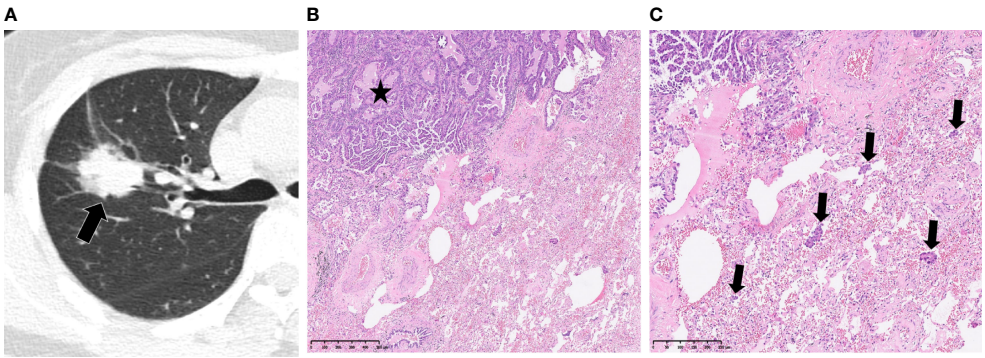


FIGURE 3 | Spread through air spaces in a 52-year-old woman with papillary adenocarcinoma. **(A)** Axial CT image presenting a slightly lobulated, solid tumor in the right upper lobe (arrow). **(B, C)** Photomicrographs showing detached papillary clusters of tumor cells (arrows) in the alveolar space beyond the edge of the main tumor (*). Hematoxylin-eosin staining, magnification x50 **(B)**, x100 **(C)**.

TABLE 2 | Features included in the clinical-CT model and their coefficients.

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	-1.10	1.25	-0.88	0.378
Sex	0.06	0.25	0.24	0.811
Gender	0.01	0.01	-0.18	0.861
Smoking	0.61	0.37	1.68	0.093
Size	0.03	0.02	1.70	0.089
GGO ratio	-1.59	0.99	-1.62	0.106
Solid nodule	0.65	0.99	0.66	0.509
mGGO	-0.25	0.71	-0.35	0.723

GGOs, ground-glass opacities; mGGO, mix GGO.

Pathologically, the histologic subtypes differed between the STAS and non- STAS groups (p<0.001). Specifically, STAS-positive tumors had a lower frequency of lepidic predominant subtypes (2.4% vs 15%), while STAS-negative ones had a lower rate of solid predominant (6.2% vs 17.8%). The positive rate of resection margin was 5.6% (22/395) in this cohort and tended to occur more in STAS-positive tumors (8.9% vs 3.1%, p=0.013). Pleural invasion was observed more frequently in patients with tumors positive for STAS versus the STAS-negative group (14.8% vs 5.3%, p=0.001), whereas perineural invasion showed no significant difference (5.3% vs 3.1%, p=0.266). EGFR and ALK analyses were available in 261 and 303 patients, respectively. STAS positivity was associated with reduced incidence of EGFR (p=0.039) and higher incidence of ALK (p=0.009). Lymphatic metastasis was negative in all patients.

Radiomics Model Building and Validation

After intraclass correlation coefficients (ICC) and Pearson's correlation analysis, 98 radiomics features were selected for predicting STAS. Based on LASSO penalized logistic regression analysis, 18 features showed significant associations between radiomics and STAS (**Figures 5A, B**). Then, the top seven radiomic features with coefficients greater than 0.1 (two first-order and five second order parameters, including GLCM, GLSZM and GLDM features) were identified by the LR model (**Figure 5C**).

Finally, these seven radiomics features were entered into the LR model for building the radiomics model. Features contained in the radiomics model, and their coefficients are shown in **Table 4**. The radiomics model achieved good performance both in the training cohort(AUC: 0.812, sensitivity: 75%, and specificity: 78.3%) and the test cohort(AUC:0.850,; sensitivity:75.8%, and specificity: 76.1%) (**Figure 4**).

The rad-score of each lesion was calculated using the following formula:

$$\begin{aligned}\text{Rad-Score} = {} & 0.585917936 + 0.49 * \text{wavelet . LHH _ firstorder _ TotalEnergy} + 0.21 * \\ & \text{wavelet . HLL _ ngtdm _ Complexity} + 0.46 * \\ & \text{wavelet . LLL _ gldm _ SmallDependenceHighGrayLevelEmphasis} + 1.28 * \\ & \text{log . sigma.6.0.mm.3D _ glcm _ MCC} + 0.78 * \text{gradient _ glcm _ Correlation} + \\ & 0.30 * \text{original _ glszm _ SmallAreaEmphasis} + 0.77 * \text{original _ firstorder _ Minimum}\end{aligned}$$

The rad scores for both the training and test sets are shown in **Figure S1**. STAS-positive tumors had significantly higher rad-scores than STAS-negative tumors in both the training and validation sets (p <0.001).

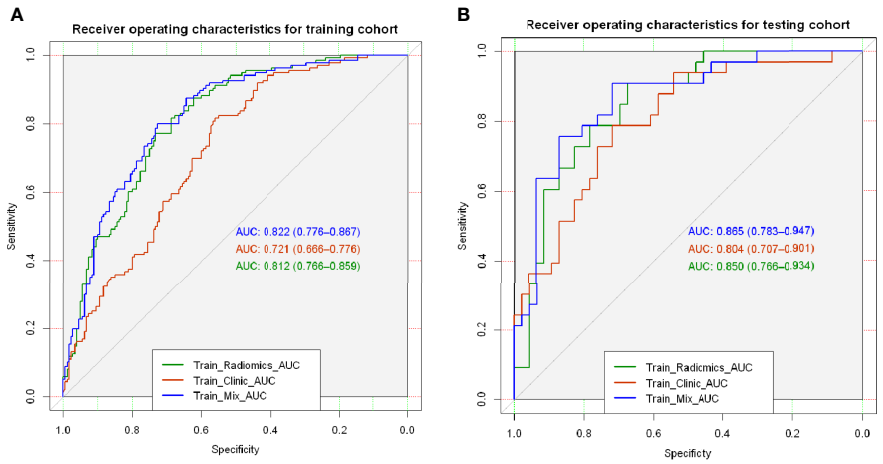


FIGURE 4 | Performances of the three models in the training **(A)** and test **(B)** groups.

TABLE 3 | Comparison of selected clinical and CT features between LUADs with STAS and those without STAS in the training and test cohorts.

	Train cohort			Test cohort		
	STAS(+) N=136	STAS(-) N=180	P value	STAS(+) N=33	STAS(-) N=46	P value
Gender	75 (55.1%)	92 (51.1%)	0.496	18 (54.4%)	20 (43.5%)	0.368
Age	59.4 ± 10.2	58.2 ± 10.3	0.310	60.0 ± 11.1	58.9 ± 9.1	0.660
Smoking	25 (18.4%)	22 (12.2%)	0.151	9 (27.3%)	7 (15.2%)	0.258
Size	21.3 ± 6.3	18 ± 6.7	<0.001	21.7 ± 5.9	18.2 ± 6.9	0.019*
GGO ratio	0.08 ± 0.20	0.38 ± 0.42	<0.001	0.08 ± 0.22	0.45 ± 0.40	<0.001*
Solid nodule	123 (55.6%)	100 (55.6%)	<0.001	29 (87.9%)	19 (41.3%)	<0.001*
mGGO	9 (6.6%)	49 (27.2%)	<0.001	4 (12.1%)	22 (47.8%)	0.001*

**P<0.05 based on comparisons between the two groups. Data are mean ± SD. STAS, spread through air spaces; STAS+, presence of spread through air spaces; STAS-, absence of spread through air spaces,GGOs, ground-glass opacities; mGGO, mix GGO.*

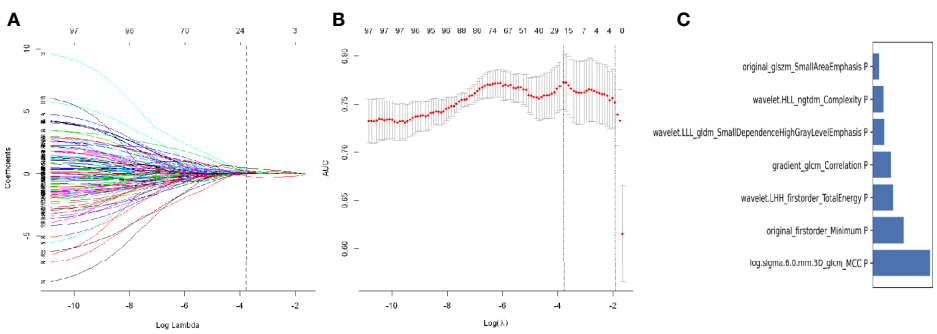


FIGURE 5 | Least absolute shrinkage and selection operator (LASSO) logistic regression of radiomics features **(A)** and the regularization parameter λ **(B)**. **(C)** The feature weights of selected radiomics features.

Performance Comparison Among Different Models

The MixModel of a comprehensive nomogram model was developed with the retain clinical CT characteristics(including age, gender, smoking history, diameter, density and GGO ratio) and rad-score(**Figure 6**), and showed AUCs of 0.822 and 0.865 in the training and test cohorts, respectively. Features contained in Mixmodel and their coefficients are listed in **Table 5**. Subsequently, we separately compared AUCs among MixModel, clinical-CT model and radiomics model (**Figure 4**). Both in the training cohort and the test cohort, the MixModel showed improvement in diagnostic ability compared with the clinical-CT model and

TABLE 4 | Features included in the radiomics model and their coefficients.

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	-0.76	0.20	-3.82	<0.001
wavelet.LHH_firstorder_TotalEnergy	0.49	0.20	2.50	0.012
wavelet.HLL_ngtdm_Complexity	0.21	0.186	1.15	0.249
wavelet.LLL_gldm_SmallDependenceHighGrayLevelEmphasis	0.46	0.28	1.68	0.093
log.sigma.6.0.mm.3D_glcm_MCC	1.28	0.52	2.46	0.014
gradient_glcm_Correlation	0.78	0.18	4.23	<0.001
original_glszm_SmallAreaEmphasis	0.30	0.23	1.30	0.194
original_firstorder_Minimum	0.77	0.20	3.73	<0.001

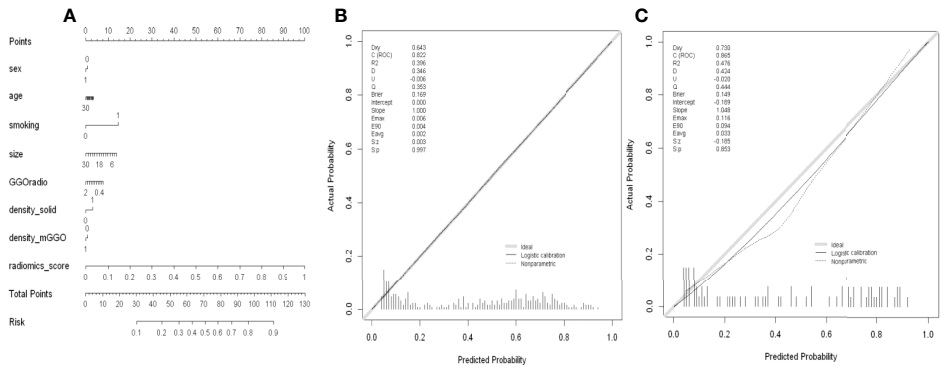


FIGURE 6 | **(A)** Nomogram of MixModel for predicting presence of spread through air spaces(STAS). For each patient, draw a vertical line between the variable value and the corresponding point line, and then assign a score for each variable based on the clinical and imaging characteristics to obtain a total score. The risk of STAS can be predicted according to the total score. **(B)** Calibration curve for the MixModel in training cohort. **(C)** Calibration curve for the MixModel in validation cohort.

TABLE 5 | Features included in Mixmodel and their coefficients.

	Estimate	Std. Error	z value	Pr (>|z|)
(Intercept)	-2.4	1.4	-1.8	0.08
Sex	-0.1	0.3	-0.2	0.8
Age	0.02	0.0	-0.1	0.9
Smoking	0.4	0.4	1.1	0.3
Size	0.0	0.0	-0.8	0.4
GGOradio	-0.4	1.1	-0.4	0.7
density_solid	0.4	1.1	0.4	0.7
density_mGGO	0.1	0.8	0.1	0.9
radiomics_score	5.1	0.7	7.1	0.0

GGOs, ground-glass opacities; mGGO, mix GGO

radiomics model. In addition, the AUC of radiomics model was larger than that of the clinical-CT model in both cohorts (0.812 vs 0.721, $p<0.001$; 0.850 vs 0.804, $p=0.228$).

The results of confounder matrix analysis in the training and test sets in MixModel, and the clinical-CT and radiomics models are summarized in **Table 6**. The Mixmodel model showed the highest accuracy (number of correct predictions divided by that of total predictions) among the three models in the training set and in the test set.

Calibration Analysis and Clinical Use

For the radiomics model, calibration curve analysis showed P values of 0.954 and 0.792 in the training and test sets, respectively, indicating a good degree of fit for the model in both sets (**Figure S2**). Similarly, the clinical-CT model also showed good calibration abilities in both the training and test sets (**Figure S2**). The clinical usefulness of the three predictive models were examined by decision curve analysis (**Figure S3**). Compared with the treat-all and treat-none models, MixModel, and the clinical-CT and radiomics models could bring net benefits to patients, among which the radiomics model had the best benefit, while the clinical-CT model had the lowest.

DISCUSSION

In this study, we developed a CT-based radiomics model with good performance, which was superior to the clinical-CT model

TABLE 6 | Confounder matrix for the training and testing sets in the three models.

Predicted results	Actual results		Accuracy (%)	Sensitivity (%)	Specificity (%)
	STAS (-)	STAS (+)			
Clinical-CT model					
Training data set			65.2	69.9	61.7
STAS(-)	111	41			
STAS(+)	69	95			
Testing data set			74.7	72.7	76.1
STAS(-)	35	9			
STAS(+)	11	24			
Radiomics model					
Training data set			76.9	75	78.3
STAS(-)	141	34			
STAS(+)	39	102			
Testing data set			76	75.8	76.1
STAS(-)	35	8			
STAS(+)	11	25			
MixModel					
Training data set			78.5	75	81.1
STAS(-)	146	34			
STAS(+)	34	102			
Testing data set			79.7	74.3	80.4
STAS(-)	37	7			
STAS(+)	9	26			

Rows correspond to the prediction of the logistic algorithm, and columns to known outcomes. STAS, spread through air spaces; STAS+, presence of spread through air spaces; STAS-, absence of spread through air spaces.

established by demographic characteristics and selected CT features, indicating the discrimination value of radiomics features for STAS positivity in IA-stage LUAD cases before surgery. Furthermore, MixModel also outperformed the clinical-CT model, presenting a higher value in predicting STAS. This report provides a powerful tool for preoperative decision-making in early-stage LUAD patients

As expected, this study showed that STAS-positive tumors tended to be larger, solid tumors with reduced GGO ratio on CT images, which was consistent with previous studies (8–11). In addition, Kim et al. (11) found that a predictive model using the percentage of the solid component could achieve an AUC of 0.77 for STAS detection. In this study, the clinical-CT model also had a similar AUC of 0.721 in the training cohort, and an AUC of 0.804 in the test cohort. The clues behind the associations of imaging features with STAS could be found in pathological findings. Several studies have revealed the connections between STAS and pathological characteristics (28, 29). Similar to previous studies (28), STAS-positive tumors in the present study had a high frequency of micropapillary, papillary or solid pattern growth, which might partially explain the association of STAS with solid tumor predominance on CT images. This study also found that STAS-positive tumors tended to be along positive resection margin and pleural invasion. In addition, the current study further analyzed the correlation between genetic mutations and STAS. We found that STAS positivity was associated with lower incidence of EGFR mutations and higher incidence of ALK mutations. Meanwhile, previous investigations have confirmed that the occurrence of GGOs is significantly associated with EGFR mutations (30) and the presence of solid nodules is one of the vital CT features of ALK rearrangement in LUAD (31). Thus, taken together, we might reasonably consider that STAS could be a potential factor in tumor aggressiveness. The larger size and solid nature on CT scans in this study supported such biological behavior.

In this study, the AUCs of the radiomics model were 0.812 and 0.850 in the training and test sets, respectively. Totally, seven features with coefficients >0.1 were selected for the tumors, including three first-order and seven second order indexes, including GLCM, GLSZM and GLDM features. First-order statistics are defined as the distribution of voxel intensity within the image region delineated by the mask through commonly used and basic metrics, while second order parameters involve the spatial position relationship with voxel intensity. Accordingly, many gray level features inferring intratumor heterogeneity were included in the radiomics model, suggesting that gray level features can contribute to the high diagnostic accuracy observed. Moreover, the present results demonstrated that the firstorder_Minimum feature was closely related to STAS, with the highest estimate coefficient (0.77) among the selected first-order parameters. The firstorder_Minimum feature referred to the lowest gray level intensity within the tumor, and STAS-positive tumors had higher firstorder_Minimum values than STAS-negative tumors. Therefore, these findings suggested that STAS-positive tumors tend to be more heterogeneous and solid components. Similarly, previous reports (18, 19) have also shown correlations between radiomics features representing gray level characteristics and STAS-positive tumors, such as Size-zone non-uniformity and Grey level variance. In addition, as shown in our clinical-CT model, the solid-density type and lower GGO ratio were the most critical features determining STAS risk. Since the automatic extraction of radiomics features by computer is more objective and accurate than subjective and manual measurements, our study confirms the reliability and interpretability of the features extracted by the radiomics analysis.

Two previous studies have explored the relationships between CT-based radiomics features and STAS in lung adenocarcinomas (18, 19). Chen et al. (19) developed a Naïve Bayes model using five radiomics features to predict STAS that achieved AUCs of 0.63 and 0.69 in the internal and external validation sets, respectively. Another report by Jiang et al. (18) built a random forest model using 12 CT-based radiomics features and showed a good AUC of 0.754 for predicting STAS. Our radiomics model established by LR outperformed those in the above two studies for STAS prediction (0.828,0.848). In addition to the differences in modeling methodologies, the discrepancy among the three studies might be related to patient inclusion criteria, sample size, and different data compositions. Chen et al. included both stage IA and IB adenocarcinomas in their study, while Jiang et al. analyzed LUAD patients with no TNM stage restriction. Meanwhile, only stage IA adenocarcinoma (T1a-cN0M0) patients were included in the present study. In addition, the number of STAS-positive patients in this investigation was twice those reported in the above two studies. In addition, STAS-positive tumors accounted for nearly 50% of all cases in this study, while the STAS-positive rates in the above two studies were less than 30%. Therefore, further studies with larger samples and better design are needed to confirm the present results.

There were several limitations in this research. Firstly, since this was a single-center retrospective study, the present radiomics model was not verified by external data. Thus, further multicenter studies are needed to confirm our results. Secondly, our study only included patients who underwent surgery for removing tumors, which might exclude cases with small tumors. However, tumor lesions larger than 3 cm were ruled out. Thirdly, since CT was performed on two different scanners, image acquisition protocols were slightly different, which might lead to some bias. Fourthly, we only included the specific CT features supported by previous reports (8–11), so the results might not represent the total CT morphological characteristics of tumors. However, the associations of other CT findings (such as satellite nodules) in LUAD with STAS remain controversial (32). Moreover, we calculated the mean and standard deviation in the training and testing sets, separately, but it might be better for clinical deployment to determine all model hyperparameters in the training set alone. Finally, since the patients were examined from 2015 to 2021, whose follow-up time was limited, we did not evaluate the effects of STAS on patient outcome.

In conclusion, the CT-based radiomics model showed a satisfying diagnostic performance for preoperatively predicting STAS, which can provide decision-making support for treatment planning in stage-IA LUAD. Besides, this radiomics model outperformed the clinical-CT model, indicating the additional value of radiomics features for predicting STAS positivity in LUAD. However, since this was a single-center retrospective study, these conclusions need to be confirmed in further prospective multicenter studies.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Wuhan Union Hospital. The ethics committee waived the requirement of written informed consent for participation. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

The acquisition, data explanation, and manuscript draft were completed by XH and JF. XH, YZ and KZ were responsible for the analysis of CT images and the delineation of VOIs. CD and XZ were responsible for data analysis and interpretation. YL and JL acquired the clinical information. JF and NW performed the pathological analysis. HS and JZ designed the study and made multiple revisions to the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by the National Natural Science Foundation of China (grant number: 82071921).

ACKNOWLEDGMENTS

We would like to thank all colleagues for help during the current study. We are also very grateful for all selfless volunteers who participated in the study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.757389/full#supplementary-material

Supplementary Figure 1 | The Rad-score of each lesion in the training set **(A)** and test set **(B)**. STAS+, presence of STAS; STAS-, absence of STAS.

Supplementary Figure 2 | Calibration curves for the clinical-CT model **(A, B)** and radiomics model **(C, D)** in the training and test sets, respectively.

Supplementary Figure 3 | Decision curves for the three predictive models.

REFERENCES

1. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol* (2015) 10(9):1243–60. doi: 10.1097/jto.0000000000000630
2. Kadota K, Nitadori JI, Sima CS, Ujiie H, Rizk NP, Jones DR, et al. Tumor Spread Through Air Spaces is an Important Pattern of Invasion and Impacts the Frequency and Location of Recurrences After Limited Resection for Small

Stage I Lung Adenocarcinomas. *J Thorac Oncol* (2015) 10(5):806–14. doi: 10.1097/jto.0000000000000486
3. Onozato ML, Kovach AE, Yeap BY, Morales-Oyarvide V, Klepeis VE, Tammireddy S, et al. Tumor Islands in Resected Early-Stage Lung Adenocarcinomas Are Associated With Unique Clinicopathologic and Molecular Characteristics and Worse Prognosis. *Am J Surg Pathol* (2013) 37(2):287–94. doi: 10.1097/PAS.0b013e31826885fb
4. Shiono S, Yanagawa N. Spread Through Air Spaces is a Predictive Factor of Recurrence and a Prognostic Factor in Stage I Lung Adenocarcinoma. *Interactive Cardiovasc Thorac Surg* (2016) 23(4):567–72. doi: 10.1093/icvts/ivw211
5. Dai C, Xie H, Su H, She Y, Zhu E, Fan Z, et al. Tumor Spread Through Air Spaces Affects the Recurrence and Overall Survival in Patients With Lung Adenocarcinoma >2 to 3 Cm. *J Thorac Oncol* (2017) 12(7):1052–60. doi: 10.1016/j.jtho.2017.03.020
6. Ren Y, Xie H, Dai C, She Y, Su H, Xie D, et al. Prognostic Impact of Tumor Spread Through Air Spaces in Sublobar Resection for 1A Lung Adenocarcinoma Patients. *Ann Surg Oncol* (2019) 26(6):1901–8. doi: 10.1245/s10434-019-07296-w
7. Eguchi T, Kameda K, Lu S, Bott MJ, Tan KS, Montecalvo J, et al. Lobectomy Is Associated With Better Outcomes Than Sublobar Resection in Spread Through Air Spaces (STAS)-Positive T1 Lung Adenocarcinoma: A Propensity Score-Matched Analysis. *J Thorac Oncol* (2019) 14(1):87–98. doi: 10.1016/j.jtho.2018.09.005
8. Qi L, Xue K, Cai Y, Lu J, Li X, Li M. Predictors of CT Morphologic Features to Identify Spread Through Air Spaces Preoperatively in Small-Sized Lung Adenocarcinoma. *Front Oncol* (2020) 10:548430. doi: 10.3389/fonc.2020.548430
9. Zhong Y, Xu Y, Deng J, Wang T, Sun X, Chen D, et al. Prognostic Impact of Tumour Spread Through Air Space in Radiological Subsolid and Pure Solid Lung Adenocarcinoma. *Eur J Cardio-Thoracic Surg* (2021) 59(3):624–32. doi: 10.1093/ejcts/ezaa361
10. Zhang Z, Liu Z, Feng H, Xiao F, Shao W, Liang C, et al. Predictive Value of Radiological Features on Spread Through Air Space in Stage cIA Lung Adenocarcinoma. *J Thorac Dis* (2020) 12(11):6494–504. doi: 10.21037/jtd-20-1820
11. Kim SK, Kim TJ, Chung MJ, Kim TS, Lee KS, Zo JI, et al. Lung Adenocarcinoma: CT Features Associated With Spread Through Air Spaces. *Radiology* (2018) 289(3):831–40. doi: 10.1148/radiol.2018180431
12. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More Than Pictures, They Are Data. *Radiology* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
13. Zhang G, Cao Y, Zhang J, Ren J, Zhao Z, Zhang X, et al. Predicting EGFR Mutation Status in Lung Adenocarcinoma: Development and Validation of a Computed Tomography-Based Radiomics Signature. *Am J Cancer Res* (2021) 11(2):546–60.
14. Song L, Zhu Z, Mao L, Li X, Han W, Du H, et al. Clinical, Conventional CT and Radiomic Feature-Based Machine Learning Models for Predicting ALK Rearrangement Status in Lung Adenocarcinoma Patients. *Front Oncol* (2020) 10:369. doi: 10.3389/fonc.2020.00369
15. Gu Y, She Y, Xie D, Dai C, Ren Y, Fan Z, et al. A Texture Analysis-Based Prediction Model for Lymph Node Metastasis in Stage IA Lung Adenocarcinoma. *Ann Thorac Surg* (2018) 106(1):214–20. doi: 10.1016/j.athoracsur.2018.02.026
16. Coroller TP, Agrawal V, Huynh E, Narayan V, Lee SW, Mak RH, et al. Radiomic-Based Pathological Response Prediction From Primary Tumors and Lymph Nodes in NSCLC. *J Thorac Oncol* (2017) 12(3):467–76. doi: 10.1016/j.jtho.2016.11.2226
17. Lee G, Park H, Sohn I, Lee SH, Song SH, Kim H, et al. Comprehensive Computed Tomography Radiomics Analysis of Lung Adenocarcinoma for Prognostication. *Oncologist* (2018) 23(7):806–13. doi: 10.1634/theoncologist.2017-0538
18. Jiang C, Luo Y, Yuan J, You S, Chen Z, Wu M, et al. CT-Based Radiomics and Machine Learning to Predict Spread Through Air Space in Lung Adenocarcinoma. *Eur Radiol* (2020) 30(7):4050–7. doi: 10.1007/s00330-020-06694-z
19. Chen D, She Y, Wang T, Xie H, Li J, Jiang G, et al. Radiomics-Based Prediction for Tumour Spread Through Air Spaces in Stage I Lung Adenocarcinoma Using Machine Learning. *Eur J Cardio-Thoracic Surg* (2020) 58(1):51–8. doi: 10.1093/ejcts/ezaa011
20. Zhuo Y, Feng M, Yang S, Zhou L, Ge D, Lu S, et al. Radiomics Nomograms of Tumors and Peritumoral Regions for the Preoperative Prediction of Spread Through Air Spaces in Lung Adenocarcinoma. *Trans Oncol* (2020) 13 (10):100820. doi: 10.1016/j.tranon.2020.100820
21. Goldstraw P, Chansky K, Crowley J, Rami-Porta R, Asamura H, Eberhardt WE, et al. The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J Thorac Oncol* 11(1):39–51. doi: 10.1016/j.jtho.2015.09.009
22. Haniff NSM, Abdul Karim MK, Osman NH, Saripan MI, Che Isa IN, Ibahim MJ. Stability and Reproducibility of Radiomic Features Based Various Segmentation Technique on MR Images of Hepatocellular Carcinoma (HCC). *Diagnostics* (2021)30;11(9):1573. doi: 10.3390/diagnostics11091573
23. Zou GY. Sample Size Formulas for Estimating Intraclass Correlation Coefficients With Precision and Assurance. *Stat Med* (2012) 31(29):3972–81. doi: 10.1002/sim.5466
24. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res* (2017) 77(21):e104–7. doi: 10.1158/0008-5472.can-17-0339
25. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropractic Med* (2016) 15(2):155–63. doi: 10.1016/j.jcm.2016.02.012
26. Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics Signature: A Potential Biomarker for the Prediction of Disease-Free Survival in Early-Stage (I or II) Non-Small Cell Lung Cancer. *Radiology* (2016) 281(3):947–57. doi: 10.1148/radiol.2016152234
27. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning With Applications in R.* New York: Springer Science and Business Media (2013). p. 440 pp.
28. Lee JS, Kim EK, Kim M, Shim HS. Genetic and Clinicopathologic Characteristics of Lung Adenocarcinoma With Tumor Spread Through Air Spaces. *Lung Cancer (Amsterdam Netherlands)* (2018) 123:121–6. doi: 10.1016/j.lungcan.2018.07.020
29. Terada Y, Takahashi T, Morita S, Kashiwabara K, Nagayama K, Nitadori JI, et al. Spread Through Air Spaces Is an Independent Predictor of Recurrence in Stage III (N2) Lung Adenocarcinoma. *Interactive Cardiovasc Thorac Surg* (2019) 29(3):442–8. doi: 10.1093/icvts/ivz116
30. Liu Y, Kim J, Qu F, Liu S, Wang H, Balagurunathan Y, et al. CT Features Associated With Epidermal Growth Factor Receptor Mutation Status in Patients With Lung Adenocarcinoma. *Radiology* (2016) 280(1):271–80. doi: 10.1148/radiol.2016151455
31. Zhou JY, Zheng J, Yu ZF, Xiao WB, Zhao J, Sun K, et al. Comparative Analysis of Clinicoradiologic Characteristics of Lung Adenocarcinomas With ALK Rearrangements or EGFR Mutations. *Eur Radiol* (2015) 25(5):1257–66. doi: 10.1007/s00330-014-3516-z
32. Ledda RE, Milanese G, Gnetti L, Borghesi A, Sverzellati N, Silva M. Spread Through Air Spaces in Lung Adenocarcinoma: Is Radiology Reliable Yet? *J Thorac Dis* (2019) 11(Suppl 3):S256–s261. doi: 10.21037/jtd.2019.01.96

Conflict of Interest: Authors, CD and XZ were employed by Philips Healthcare.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Han, Fan, Zheng, Ding, Zhang, Zhang, Wang, Jia, Li, Liu, Zheng and Shi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: frontiersin.org/about/contact

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership