ARTIFICIAL INTELLIGENCE AND MRI: BOOSTING CLINICAL DIAGNOSIS

EDITED BY: Antonio Napolitano, Natalie Julie Serkova, Daniel Rodriguez Gutierrez and Oliver Diaz

PUBLISHED IN: Frontiers in Oncology and Frontiers in Physics

frontiers Research Topics

frontiers

ISSN 1664-8714
ISBN 978-2-88976-719-9
DOI 10.3389/978-2-88976-719-9

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

ARTIFICIAL INTELLIGENCE AND MRI: BOOSTING CLINICAL DIAGNOSIS

Topic Editors:
Antonio Napolitano, Bambino Gesù Children's Hospital (IRCCS), Italy
Natalie Julie Serkova, University of Colorado Anschutz Medical Campus, United States
Daniel Rodriguez Gutierrez, Nottingham University Hospitals NHS Trust, United Kingdom
Oliver Diaz, University of Barcelona, Spain

Citation: Napolitano, A., Serkova, N. J., Gutierrez, D. R., Diaz, O., eds. (2022). Artificial Intelligence and MRI: Boosting Clinical Diagnosis. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-88976-719-9

Table of Contents

frontiers in Oncology

ORIGINAL RESEARCH
published: 26 May 2021
doi: 10.3389/fonc.2021.659969

Comparison of Different Machine Models Based on Contrast-Enhanced Computed Tomography Radiomic Features to Differentiate High From Low Grade Clear Cell Renal Cell Carcinomas

Xu Pei[1], Ping Wang[1], Jia-Liang Ren[2], Xiao-Ping Yin[1,3], Lu-Yao Ma[1], Yun Wang[1], Xi Ma[1] and Bu-Lang Gao[1]*

[1] CT/MRI Room, Affiliated Hospital of Hebei University, Baoding, China, [2] Department of Pharmaceutical Diagnostics, GE Healthcare China (Shanghai) Co Ltd., Shanghai, China, [3] Key Laboratory of Cancer Radiotherapy and Chemotherapy Mechanism and Regulations, Baoding, China

Purpose: This study was to investigate the role of different radiomics models with enhanced computed tomography (CT) scan in differentiating low from high grade renal clear cell carcinomas.

Materials and Methods: CT data of 190 cases with pathologically confirmed renal cell carcinomas were collected and divided into the training set and testing set according to different time periods, with 122 cases in the training set and 68 cases in the testing set. The region of interest (ROI) was delineated layer by layer.

Results: A total of 402 radiomics features were extracted for analysis. Six of the radiomic parameters were deemed very valuable by univariate analysis, rank sum test, LASSO cross validation and correlation analysis. From these six features, multivariate logistic regression model, support vector machine (SVM), and decision tree model were established for analysis. The performance of each model was evaluated by AUC value on the ROC curve and decision curve analysis (DCA). Among the three prediction models, the SVM model showed a high predictive efficiency. The AUC values of the training set and the testing set were 0.84 and 0.83, respectively, which were significantly higher than those of the decision tree model and the multivariate logistic regression model. The DCA revealed a better predictive performance in the SVM model that possessed the highest degree of coincidence.

Conclusion: Radiomics analysis using the SVM radiomics model has highly efficiency in discriminating high- and low-grade clear cell renal cell carcinomas.

Keywords: renal clear cell carcinoma, enhanced computed tomography, imaging histology, logistic regression, radiomics

OPEN ACCESS

Edited by:
Natalie Julie Serkova, University of Colorado, United States

Reviewed by:
S. Gene Kim, Cornell University, United States
Subathra Adithan, Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER), India

***Correspondence:**
Xiao-Ping Yin
browngao@126.com;
yinxiaoping78@sina.com
orcid.org/0000-0002-7386-1069

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

Received: *28 January 2021*
Accepted: *28 April 2021*
Published: *26 May 2021*

Citation:
Pei X, Wang P, Ren J-L, Yin X-P, Ma L-Y, Wang Y, Ma X and Gao B-L (2021) Comparison of Different Machine Models Based on Contrast-Enhanced Computed Tomography Radiomic Features to Differentiate High From Low Grade Clear Cell Renal Cell Carcinomas. Front. Oncol. 11:659969. doi: 10.3389/fonc.2021.659969

INTRODUCTION

Clear cell renal cell carcinoma (ccRCC) accounts for 70% of renal cancers (1). Since the long-term survival of clear cell carcinoma patients correlates negatively to the Fuhrman grading (2–4), it is crucial to accurately grade clear cell carcinoma of the kidney as early as possible. Grading ccRCC through aspiration biopsy is controversial as the operation itself carries risk of metastatic spread (5, 6). Previous studies on RCC were mostly based on analysis of images of conventional computed tomography (CT) (7–9), which was often interfered by human factors and lack of quantification. Through precise quantitative analysis of medical images, radiomics provides researchers an effective way to detect biological characteristic changes caused by tumor microenvironment (10–12). Classic CT information or CT-based radiomics has been applied to establish predictive models for ccRCC grade. In three logistic regression models of radiomics based on non-texture features, texture fraction and non-texture feature combined with texture fraction for identifying high- and low-grade ccRCCs (13), the area under the operating curve (AUC) values in the three models were 0.826, 0.878, and 0.843 for the training set and 0.671, 0.771, and 0.780 for the testing set, respectively. Some image features like tumor size (TS) and permeability surface-area product (PS) were helpful in differentiating high- from low-grade ccRCCs based on conventional CT studies, with the AUC of TS and PS of 0.7 (14). The sensitivity and specificity were 0.8 and 0.6 for TS and 0.7 and 0.8 for PS, respectively. Moreover, gene fragments and radiomics can be combined to establish a two-group model for differentiating ccRCC from non-clear cell RCC (non-ccRCC), with the AUC of the training set and testing set being 0.969 and 0.900, respectively (15). Some studies confirmed that necrosis can independently predict the biological invasiveness of ccRCCs (16, 17). Moreover, only the logistic regression model was utilized in most of these studies lacking comparison between different predictive modeling methods. Therefore, in this study, three models including logistic regression, decision tree and support vector machine (SVM) were established and compared for ccRCC grading performance.

MATERIALS AND METHODS

Patients

This retrospective study was approved by the Ethics Review Committee of Affiliated Hospital of Hebei University with all patients given their signed informed consent. All methods were performed in accordance with the relevant guidelines and regulations. Patients with ccRCC were enrolled between January, 2017 and December, 2018 in our hospital. Inclusion criteria were a single lesion with clear grades of RCC and preoperative enhanced CT images in the cortical phase with fast-in and fast-out enhancement (cortical phase showed the clearest). Exclusion criteria were: (I) carcinomar metastasis, (II) cystic changes in the lesion of carcinoma, (III) necrosis volume >80% of the maximal lesion volume, and (IV) poor image quality. In accord with these criteria, 42 unqualified samples were excluded, and 190 eligible samples were included. In this study, I-II grade ccRCC was defined as low-grade renal clear cell carcinoma, and III-IV grade ccRCC was defined as high-grade renal carcinoma (18) (**Figure 1**). Among the qualified 190 patients with ccRCC, 133 cases were of grade I-II ccRCC and 57 cases were of grade III-IV ccRCC, including 98 males and 92 females with an age range of 27–88 years (mean 58.30 ± 8.70) (**Table 1**). Their maximal diameters of the carcinoma ranged 2-12 cm (mean 5.6 ± 4.4) from post-operative pathological exams.

CT Image Acquisition

Abdominal plain and enhancement CT scans were performed with a 64-row CT scanner (GE Discovery HD 750, GE Health Care, Chicago, IL, USA). Contrast agent was iodophor alcohol, a non-ionic iodine contrast agent. The post-injection scanning time points were 30-35s, 50-60s and 180s, covering the medullary phase and renal pelvis stage. Scanning parameters were as follows: cortical phase, pitch: 0.984:1, layer thickness: 5 mm, field of view: 40 cm×40 cm, matrix: 512×512, tube voltage: 100-120 kV, tube current: 134-409 mA, window width: 250-450 HU, and window position: 30-50 HU.

Volumes of Interest (VOIs) Segmentation

The cortical phase images of enhanced CT from 190 subjects were imported into the ITK-SNAP software (19), and the region of interest (ROI) was delineated by one radiologist with 8 years of working experience and checked by another radiologist with 10 years of working experience.

Radiomics Feature Extraction and Selection

The radiomics features were extracted from the original and filtered images with the AK software (Artificial Intelligence Kit V3.0.0.R, GE Healthcare, China). A total of 402 features were obtained, including 42 histogram features, 144 gray-level co-occurrence matrices features (GLCM), 180 gray-level run length matrices features (GLRM), 11 gray-level zone matrices features (GLSZM), 15 shape-based features, and 10 Haralick features. The feature selection procedure was as follows: Firstly, the data of patients from January 2017 to April 2018 were included in the training set, and the data of patients from April 2018 to December 2018 were included in the testing set, with the data of 122 patients in the training set (with 81 cases of I-II ccRCC and 41 cases of III-IV ccRCC) and 68 patients in the testing set (with 52 cases of I-II ccRCC and 16 cases of III-IV ccRCC). Secondly, the data were preprocessed, including replacing missing values with the median value and standardizing the Z-score of features in all data. Thirdly, the extracted features were analyzed by one-way ANOVA and Wilcox rank-sum test, with the significant P value set at less than 0.05. Then, the least absolute shrinkage and selection operator (LASSO) method, which has been shown to be suitable for high dimensional data analysis (13), were used for further feature screening. The LASSO method selects features using a tuning parameter (Lambda), with some coefficients in the covariance can be shrunk to zero when the cross-validation error

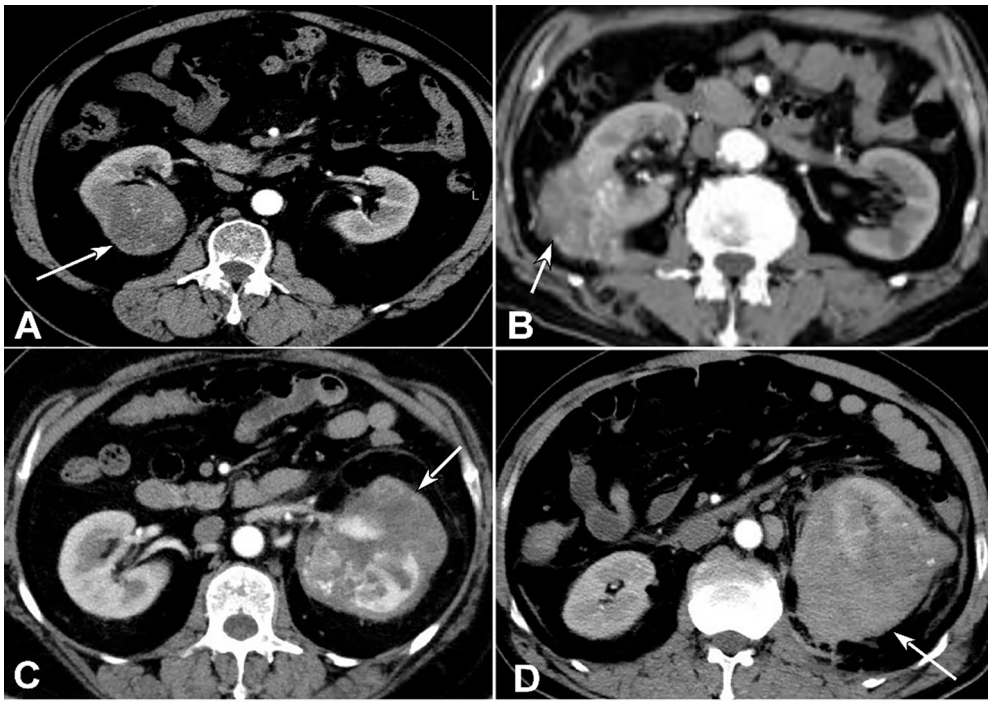


FIGURE 1 | Clear cell renal cell carcinomas (ccRCC) with different grades. **(A)** Grade I ccRCC (arrow) was demonstrated. **(B)** Grade II ccRCC was shown (arrow). **(C)** Grade III ccRCC was revealed (arrow). **(D)** Grade IV ccRCC (arrow) was displayed.

is the smallest. All the feature selection procedure performed on the training set and applied on the testing set. The finally selected features were used to construct models.

RADIOMIC MODEL BUILDING AND VALIDATION

The ROC curves of each model in the training set (data of 122 patients) and testing set (68 patients) were calculated with all available patients and the AUC values were derived (**Figure 3**).

TABLE 1 | Demography of patients in two sets.

Variables	Training set	Testing set
Case no.	122(64%)	68 (36%)
Sex		
Male	66(54%)	33(49%)
Female	56(46%)	35(51%)
Age mean (range, y)	55.6(28-85)	56.1(31-87)
< 60	59(48%)	35(51%)
≧60	33(49%)	63(52%)
Subtype		
Low-grade ccRCC	81(66%)	52(76%)
High-grade ccRCC	41(34%)	16(24%)
Tumor size (cm, mean ± SD)		
Low-grade ccRCC	6.48 ± 3.46	6.57 ± 3.31
High-grade ccRCC	7.21 ± 3.13	8.31 ± 3.31

Low grade, grades I-II; High-grade, grades III-IV; SD, standard deviation.

The predictive performances of three models (logistic regression, decision tree, and SVM) were compared for analysis. The decision curve analysis (DCA) was conducted to evaluate the clinical usefulness of the models for ccRCC prediction. DCA quantified the net benefits at different threshold probabilities in the training and testing set (**Figure 4**).

Statistical Analysis

Statistical analysis was performed with the R software (version: 3.6.3, www.r-project.org). The Chi-square test was used to evaluate the distribution difference in high and low-grade cc RCCs. The LASSO, SVM, and decision tree model were conducted based on 'glmnet', 'e1071', and 'rpart' packages, respectively. The receiver operating characteristics (ROC) curve analysis was performed to determine the AUC, accuracy, specificity and sensitivity for evaluating the performance of the model. The significance was set at $P < 0.05$.

RESULTS

The six most valuable features selected by LASSO for radiomics modelling were GLCMEntropy, GreyLevelNonuniformity, ShortRunEmphasis, LongRunLowGreyLevelEmphasis, ShortRunLowGreyLevelEmphasis, and IntensityVariability. The LASSO regression was shown in **Figure 2**. The specific parameters and feature extraction used in the six most valuable features were demonstrated in **Table 2**. These features were used

to establish three models of logistic regression, decision tree and SVM in the training set with 122 patients. Each model was trained and assessed using the repeated ten-fold cross-validation method in the training set. Performance of differentiating high grade from low grade ccRCC was evaluated with the testing set (68 patients) (**Figure 3**).

The AUC values in the training set and testing sets are respectively 0.63 (95% CI 0.53-0.73) and 0.64 (95% CI 0.48-0.8) with the logistic regression model, 0.84 (95% CI 0.76-0.92) and 0.83 (95% CI 0.69-0.96) with SVM model, and finally, 0.69 (95% CI 0.60-0.78) and 0.72 (95% CI 0.56-0.87) with the decision tree model. The cutoff value of each model was obtained from the Youden index from the ROC curve, with the value being 0.366, 0.38, and 0.276, respectively, in the logistic regression, SVM, and decision tree for the test set. The results presented in **Tables 3** and **4** showed that the SVM model had achieved the best performance.

DCA was conducted to evaluate clinical usefulness of the models in prediction by quantifying the net benefits (relative benefits), at different threshold probabilities in both sets (**Figure 4**). The SVM model had the best performance in prediction of low- and high- grade renal cell carcinoma. In the DCA analysis (**Figure 4**), the SVM model was shown to obtain the highest benefit in the range of 0.34-0.49 which contained the cutoff value 0.38 for the SVM model. The "benefit" was relative and indicated the efficiency of the models in the test set.

The prediction performance of the three models for low and high grade RCC was verified and compared (**Figures 5**–**7**). There was no significant (P=0.054) difference in the high and low-grade distribution of ccRCCs between the training and testing sets.

In verification of the logistic regression model (**Figure 5**), the true negative rate (specificity) for predicting grade I-II ccRCC was 60.5% (49/81) in the training and 53.7% (22/41) in the testing set, and the true positive rate (sensitivity) for predicting grade III-IV ccRCC was 65.4% (34/52) in the training and 75% (12/16) in the testing set. In verification of the SVM model (**Figure 6**), the true negative rate (specificity) was 76.5% (62/81) for predicting grade I-II ccRCC in the training and 85.4% (35/41) in the testing set, and the true positive rate (sensitivity) was 84.6% (44/52) in the training set. The testing set also exhibited a true negative rate of 75% (12/16) for predicting grade III-IV ccRCC. In verification of the decision tree model (**Figure 7**), the true negative rate (specificity) was 77.8% (63/81) for predicting grade I-II ccRCC with the true positive rate (sensitivity) of 55.8% (29/52) in the training set. For the testing set, the true negative rate was 82.9% (34/41) for predicting grade I-II ccRCC and 62.5% (10/16) for predicting grade III-IV ccRCC.

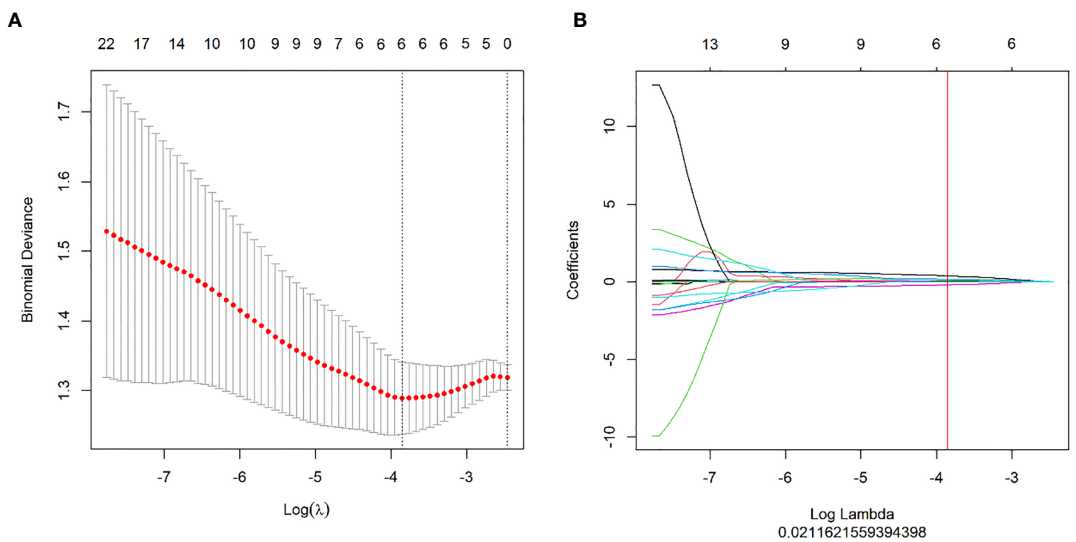


FIGURE 2 | Feature selection with the LASSO method. **(A)** The tuning parameter (λ) changes in the LASSO model. The binomial deviance curve was generated with the log (λ). The minimum criteria for five-fold cross-validation were applied. The best λ = 0.0212 was obtained at the minimal binomial deviance. **(B)** The LASSO coefficient profile plot with different log (λ) was shown. The vertical red line was the best λ with 6 selected radiomic features.

TABLE 2 | Specific parameters and feature extraction in six features.

ID	Class	Type	Offset	Direction
1	GLCM	Entropy	7	Angle90
2	RLM	GreyLevelNonuniformity	7	All (3D)
3	RLM	ShortRunEmphasis	7	Angle0
4	RLM	LongRunLowGreyLevelEmphasis	7	Angle0
5	RLM	ShortRunLowGreyLevelEmphasis	4	All (3D)
6	Histogram	IntensityVariability	–	–

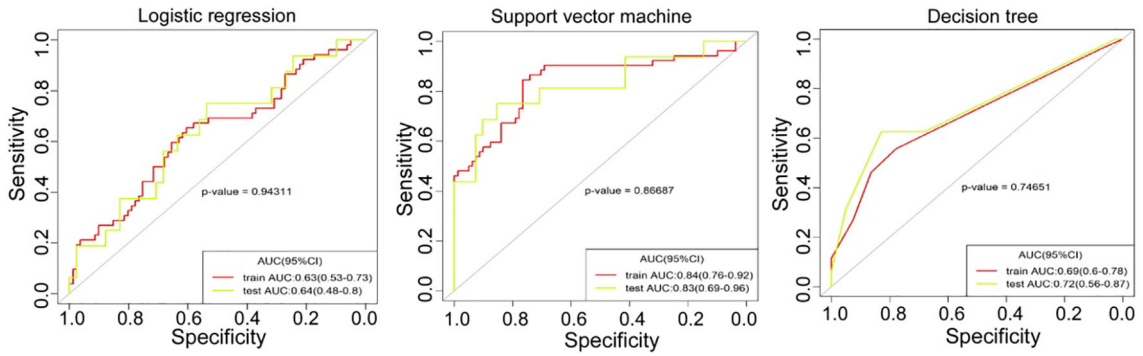


FIGURE 3 | The receiver operating characteristics (ROC) curve analysis was performed for three models of logistic regression, support vector machine and decision tree in the training set and testing set.

The AUC, accuracy, specificity, and sensitivity were used to evaluate the performance of the model (**Tables 1** and **2**). Through comprehensive comparison of the AUC value, specificity, sensitivity of the three models, the best prediction efficiency, observed in the sSVM model, was therefore selected for prediction purpose in this study. The SVM model had the greatest accuracy (0.797 and 0.825), sensitivity (0.846 and 0.825) and specificity (0.742 and 0.750) in both the training and testing set compared with the logistic regression model (0.624 and 0.596, 0.654 and 0.750, 0.605 and 0.537, respectively) and the decision tree model (0.692 and 0.772, 0.558 and 0.625, 0.778 and 0.829, respectively).

DISCUSSION

The present study was aimed at the differentiation of high- from low-grade ccRCCs, because pathological grades highly correlate with ccRCC metastasis and prognosis (20). ccRCC has different clinical prognoses at different grades, and early identification of pathological grade of ccRCCs is valuable for timely clinical treatment and patient health.

Radiomics analysis is to extract a multitude of features form medical images to analyze size, shape, and texture, with useful spatial information on pixel or voxel distribution and modes. The recent advancements in the study of ccRCCs were based on imaging histology except for its grading (21). In the modeling and identification of high- and low-grade ccRCCs, previous studies (22, 23) used *in vivo* diffusion-weighted imaging (DWI) and imaging histology to achieve the AUC value of 0.8, whereas an AUC value of 0.73 was reached by the Renometric score based on CT imaging in identification of high-level RCCs (23). The AUC values for SVM model in the training and testing sets in our study were 0.84 and 0.83, respectively, higher than 0.8 or 0.73 of methods described earlier.

Ding et al. (13) applied radiomics to establish three logistic regression models to identify high and low-grade ccRCCs, achieving the AUC values in the training sets of the three

TABLE 3 | ROC curve analysis of three models in the training set.

Parameter	Logistic (Train)	SVM (Train)	Decision Tree (Train)
AUC	0.632 (CI: 0.533–0.730)	0.840 (CI: 0.653–0.758)	0.688 (CI: 0.601–0.775)
Accuracy	0.624 (CI: 0.530–0.707)	0.797 (CI: 0.719–0.862)	0.692 (CI: 0.606–0.769)
Sensitivity	0.654 (CI: 0.462–0.788)	0.846 (CI: 0.558–0.942)	0.558 (CI: 0.385–0.681)
Specificity	0.605 (CI: 0.272–0.741)	0.742 (CI: 0.284–0.852)	0.778 (CI: 0.575–0.904)

ROC, Receiver operating characteristic; AUC, area under the operating curve; CI, confidence interval.

TABLE 4 | ROC curve analysis of three models in the testing set.

Parameter	Logistic regression	Support vector machine	Decision Tree
AUC	0.639 (CI: 0.476–0.802)	0.826 (CI: 0.688–0.964)	0.717 (CI: 0.564–0.871)
Accuracy	0.596 (CI: 0.458–0.724)	0.825 (CI: 0.701–0.913)	0.772 (CI: 0.642–0.873)
Sensitivity	0.750 (CI: 0.436–0.938)	0.750 (CI: 0.438–0.938)	0.625 (CI: 0.320–0.812)
Specificity	0.537 (CI: 0.195–0.756)	0.854 (CI: 0.341–0.976)	0.829 (CI: 0.400–0.951)

ROC, Receiver operating characteristic; AUC, area under the operating curve; CI, confidence interval.

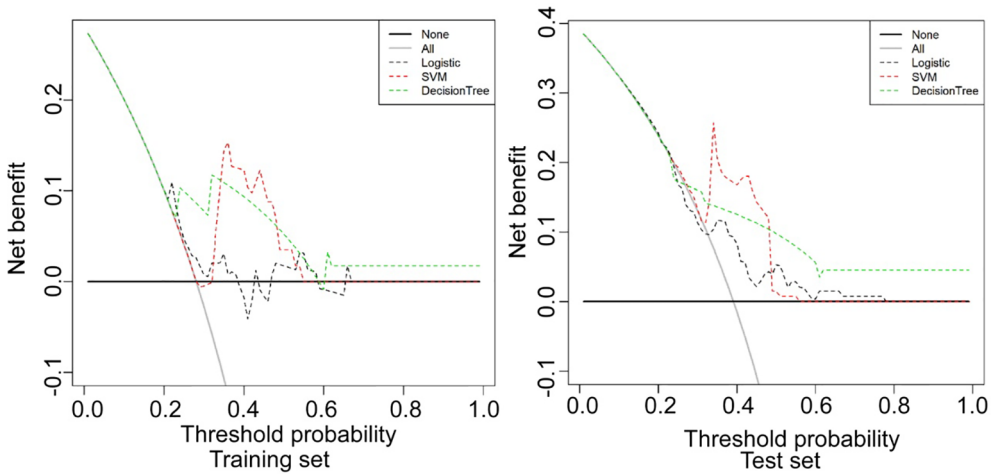


FIGURE 4 | Decision curve analysis (DCA) was conducted to evaluate the clinical usefulness of the models in prediction by quantifying the net benefits at different threshold probabilities in the training and testing set. The SVM model had the best performance in prediction of low- and high-grade renal cell carcinoma. Logistic, logistic regression model; SVM, support vector machine model; Decision tree, decision tree model.

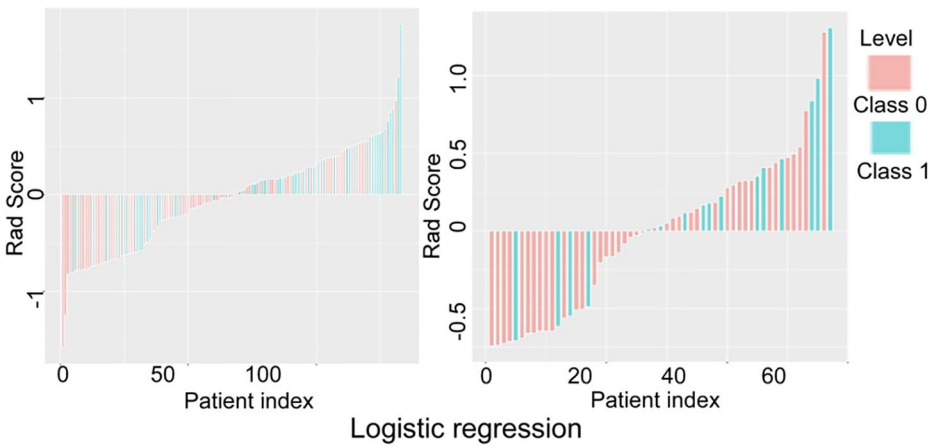


FIGURE 5 | Verification and comparison of the logistic regression model in predicting low and high grade renal cell carcinoma in the training and testing set. In the training set, the true negative rate (specificity) for predicting grade I-II clear cell renal cell carcinoma (ccRCC) was 60.5% (49/81), and the true positive rate (sensitivity) for predicting grade III-IV ccRCC was 65.4% (34/52). In the testing set, the true negative rate was 53.7% (22/41) for predicting grade I-II ccRCC and 75% (12/16) for predicting grade III-IV.

models of 0.826, 0.878, 0.843 compared with the AUC values in the testing sets of 0.671, 0.771 and 0.780, respectively. Although the results in training set were better, the scores in testing set were not as satisfactory probably due to a trend of over-fitting. In addition, Ding et al. extracted the texture features from the maximal diameter level of the mass and collected less heterogeneous information of the mass (13). Compared with the study by Ding et al, our SVM-based model performed better, with our SVM-based AUC in the training and testing set being 0.84 and 0.83, respectively. Shu et al. (24) established three radiomic models based on renal CT enhancement images in the cortical and parenchymal phases, including cortical phase model, parenchymal phase model, and in combination. The corresponding accuracy, AUC value, sensitivity and specificity were 0.719, 0.766, 0.818 and 0.822) for the cortical phase model, 0.738, 0.602, 0.693 and 0.677 for the parenchymal phase model, and 0.777, 0.838, 0.838 and 0.839 for the combined model. Comparing these results to the study with 3D texture analysis based model by Shu et al. (24), our results have better accuracy, AUC value, and sensitivity. Although the model produced by Shu et al. (24) possessed slightly higher specificity with the combined multi-period model outperforming the one-period model, their study used full data to build the model without using independent test data to validate their results.

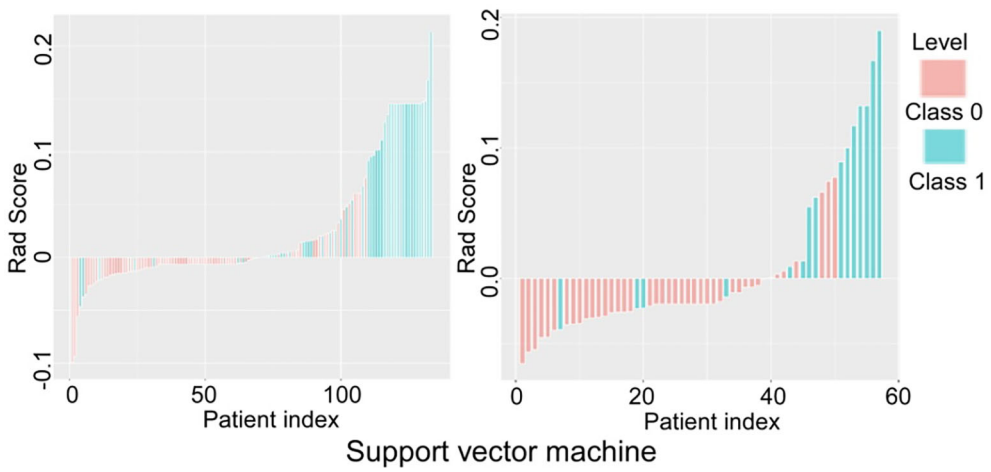


FIGURE 6 | Verification and comparison of the support vector machine model in predicting low and high grade renal cell carcinoma in the training and testing set. In the training set, the true negative rate (specificity) was 76.5% (62/81) for predicting grade I-II clear cell renal cell carcinoma (ccRCC), and the true positive rate (sensitivity) was 84.6% (44/52). In the testing set, the true negative rate was 85.4% (35/41) for predicting grade I-II ccRCC and 75% (12/16) for predicting grade III-IV ccRCC.

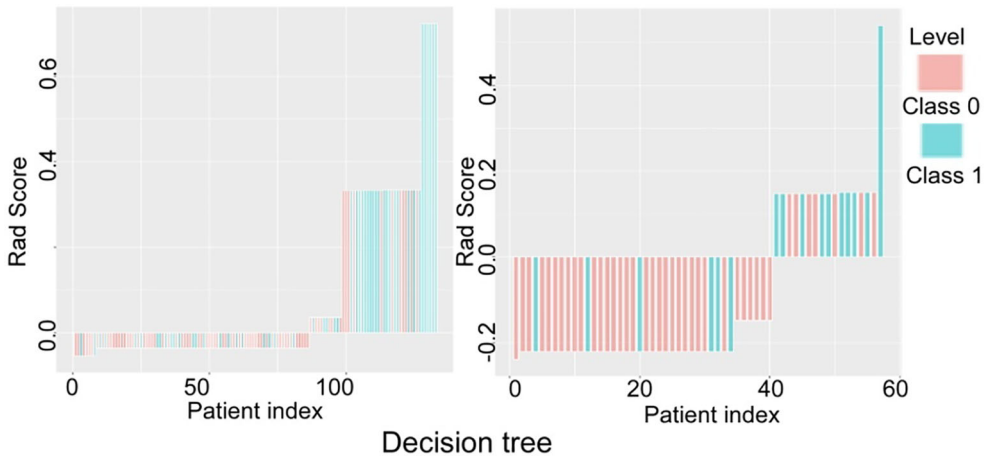


FIGURE 7 | Verification and comparison of the decision tree model in predicting low and high grade renal cell carcinoma in the training and testing set. In the training set, the true negative rate (specificity) was 77.8% (63/81) for predicting grade I-II clear cell renal cell carcinoma (ccRCC), and the true positive rate (sensitivity) was 55.8% (29/52). In the testing set, the true negative rate was 82.9% (34/41) for predicting grade I-II ccRCC and 62.5% (10/16) for predicting grade III-IV ccRCC.

Radiomics-based grading models demonstrated better performance than the model based on conventional CT parameters. Chen et al. (14) pointed out that tumor size (TS) and permeability surface-area products (PS) were helpful in distinguishing the high and low grade clear cell renal cancers, with the AUC of both TS and PS being 0.7 and the sensitivity and specificity being 0.8 and 0.6 for TS and 0.7 and 0.8) for PS. The grading performance in our study was also better than this study (15).

Heterogeneity is an important feature of malignant tumors and is closely related to their biological behavior. CT enhanced imaging can be used to effectively evaluate tumor heterogeneity (25). After studying low enhancement on multiphase contrast-enhanced CT images for predicting presence of high tumor grade of ccRCC (26), Miles et al. found that low tumor enhancement in the cortico-medullary phases was an independent predictor of high tumor grade, which may be useful in clinical care of patients with nonsurgical approaches. It is speculated that the higher the grade of renal clear cell carcinoma, the more abundant the small capillaries (27), which is supported by another study by Li et al (15). In addition, necrosis is highly correlated with heterogeneity of tumors, which is of great significance (28). In this study (28), various processing techniques including voxel normalization and various filtering processes were used to extract a variety of high and low order features, including gray matrix and 3D morphological features. Finally, LASSO cross-processing was used to select the most valuable six histological features.

After looking into a variety of common first-order features that reflect tumor heterogeneity, such as average gray level, kurtosis and entropy, Feng et al. proposed that entropy is an independent and excellent radiomic feature to describe a degree of disorder in images (29). In terms of lesion density distribution, larger entropy values suggest more randomness while smaller entropy values indicate uniformity. Thus, high-grade tumors with relatively large liquefaction necrosis volume have reduced the entropy detectable as a radiomic feature and were consequently excluded from our study. In our study, we only studied the primary renal cell carcinoma rather than metastatic carcinomas from other resources. However, if the renal cancerous lesions of the primary renal cell carcinoma contained large-area necrosis or cystic changes, they would be excluded from the study, because necrosis contained inactive tissue and cystic changes contained liquid materials. Solid mass should be retained as much as possible. The radiomics captured tissues primarily with active and biological behavior, namely solid mass tissues. Cystic degeneration and necrosis are similar in nature, and the doping of these changes in the samples may lower the evaluation efficiency of the results.

In our study, GLCM_entropy, Greylevel_Nonuniformity, and Intensity_Variability of the six features reflect the degree of random gray distribution in ROI, which is usually used to demonstrate the tumor heterogeneity. ShortRun_Emphasis and ShortRunLowGreyLevel_Emphasis are used to show the fine texture of the tumor, whereas LongRunLowGreyLevel_Emphasis is used to reflect the coarse texture within the tumor. The SVM model in our study used the RBF kernel with C value 1 and gamma 0.001. The SVM is a nonlinear model which can get greater and better results than the linear model. The SVM model may be used for machine learning with small samples, for improving generalization and solving higher-dimensional problems as well as for avoiding structural selection in neural networks. There are some limitations in our study. Firstly, the overall sample size was relatively small. Secondly, patient data was not comprehensively collected, with the construction of models having excluded diagnostic elements from biochemistry, immunohistochemistry and genetic studies. Thirdly, when the VOI was delineated, the accuracy of the delineated lesions was reduced, due to unclear margins of some tumor masses or the influences by partial volume effect. Fourthly, the current single-center study lacked independent validation and evaluation from external professionals. Although our scanning parameters and reconstruction methods had been standardized, they should have been fixed with multicenter studies, thus necessitating a unified measurement standard for obtaining necessary information. Lastly, this study was limited to its retrospective nature and involvement with only Chinese ethnicity.

In summary, the current study uses radiomics analysis to differentiate the grade of ccRCC, and the support vector machine-based model exhibits the best performance for differentiating high- and low-grade ccRCC when compared to the logistic regression model and the decision tree model.

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 Affiliated Hospital of Hebei University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Study design: XP and J-LR. Data collection: XP, PW, L-YM, YW, and XM. Data analysis: XP, PW, J-LR, X-PY, and B-LG. Supervision: L-YM. Manuscript writing: XP. Revision: B-LG. All authors contributed to the article and approved the submitted version.

REFERENCES

1. Znaor A, Lortet-Tieulent J, Laversanne M, Jemal A, Bray F. International Variations and Trends in Renal Cell Carcinoma Incidence and Mortality. *Eur Urol* (2015) 67:519–30. doi: 10.1016/j.eururo.2014.10.002
2. Ficarra V, Martignoni G, Maffei N, Brunelli M, Novara G, Zanolla L, et al. Original and Reviewed Nuclear Grading According to the Fuhrman System: A Multivariate Analysis of 388 Patients With Conventional Renal Cell Carcinoma. *Cancer* (2005) 103:68–75. doi: 10.1002/cncr.20749
3. 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) 15:804–34. doi: 10.6004/jnccn.2017.0100
4. Rouprêt M, Babjuk M, Compérat E, Zigeuner R, Sylvester RJ, Cowan NC, et al. European Association of Urology Guidelines on Upper Urinary Tract Urothelial Carcinoma: 2017 Update. *Eur Urol* (2018) 73:111–22. doi: 10.1016/j.eururo.2017.07.036
5. 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
6. Hodgdon T, McInnes MD, Schieda N, Flood TA, Lamb L, Thornhill RE. Can Quantitative Ct Texture Analysis be Used to Differentiate Fat-Poor Renal Angiomyolipoma From Renal Cell Carcinoma on Unenhanced Ct Images? *Radiology* (2015) 276:787–96. doi: 10.1148/radiol.2015142215
7. Pierorazio PM, Hyams ES, Tsai S, Feng Z, Trock BJ, Mullins JK, et al. Multiphasic Enhancement Patterns of Small Renal Masses (≤4 Cm) on Preoperative Computed Tomography: Utility for Distinguishing Subtypes of Renal Cell Carcinoma, Angiomyolipoma, and Oncocytoma. *Urology* (2013) 81:1265–71. doi: 10.1016/j.urology.2012.12.049
8. Rosenkrantz AB, Matza BW, Portnoy E, Melamed J, Taneja SS, Wehrli NE. Impact of Size of Region-of-Interest on Differentiation of Renal Cell Carcinoma and Renal Cysts on Multi-Phase CT: Preliminary Findings. *Eur J Radiol* (2014) 83:239–44. doi: 10.1016/j.ejrad.2013.10.020
9. Teng J, Gao Y, Chen M, Wang K, Cui X, Liu Y, et al. Prognostic Value of Clinical and Pathological Factors for Surgically Treated Localized Clear Cell

Renal Cell Carcinoma. *Chin Med J (Engl)* (2014) 127:1640–4. doi: 10.3760/cma.j.issn.0366-6999.20132831
10. 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
11. 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
12. 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
13. Ding J, Xing Z, Jiang Z, Chen J, Pan L, Qiu J, et al. CT-Based Radiomic Model Predicts High Grade of Clear Cell Renal Cell Carcinoma. *Eur J Radiol* (2018) 103:51–6. doi: 10.1016/j.ejrad.2018.04.013
14. Chen C, Kang Q, Xu B, Guo H, Wei Q, Wang T, et al. Differentiation of Low- and High-Grade Clear Cell Renal Cell Carcinoma: Tumor Size Versus CT Perfusion Parameters. *Clin Imaging* (2017) 46:14–9. doi: 10.1016/j.clinimag.2017.06.010
15. Li ZC, Zhai G, Zhang J, Wang Z, Liu G, Wu GY, et al. Differentiation of Clear Cell and non-Clear Cell Renal Cell Carcinomas by All-Relevant Radiomics Features From Multiphase CT: A VHL Mutation Perspective. *Eur Radiol* (2019) 29:3996–4007. doi: 10.1007/s00330-018-5872-6
16. Hötker AM, Karlo CA, Zheng J, Moskowitz CS, Russo P, Hricak H, et al. Clear Cell Renal Cell Carcinoma: Associations Between CT Features and Patient Survival. *AJR Am J Roentgenol* (2016) 206:1023–30. doi: 10.2214/AJR.15.15369
17. Oh S, Sung DJ, Yang KS, Sim KC, Han NY, Park BJ, et al. Correlation of CT Imaging Features and Tumor Size With Fuhrman Grade of Clear Cell Renal Cell Carcinoma. *Acta Radiol* (2017) 58:376–84. doi: 10.1177/0284185116649795
18. Bektas S, Bahadir B, Kandemir NO, Barut F, Gul AE, Ozdamar SO. Intraobserver and Interobserver Variability of Fuhrman and Modified Fuhrman Grading Systems for Conventional Renal Cell Carcinoma. *Kaohsiung J Med Sci* (2009) 25:596–600. doi: 10.1016/S1607-551X(09)70562-5
19. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability. *Neuroimage* (2006) 31:1116–28. doi: 10.1016/j.neuroimage.2006.01.015
20. Guðmundsson E, Hellborg H, Lundstam S, Erikson S, Ljungberg B. Metastatic Potential in Renal Cell Carcinomas ≤7 Cm: Swedish Kidney Cancer Quality Register Data. *Eur Urol* (2011) 60:975–82. doi: 10.1016/j.eururo.2011.06.029
21. Yan L, Liu Z, Wang G, Huang Y, Liu Y, Yu Y, et al. Angiomyolipoma With Minimal Fat: Differentiation From Clear Cell Renal Cell Carcinoma and Papillary Renal Cell Carcinoma by Texture Analysis on CT Images. *Acad Radiol* (2015) 22:1115–21. doi: 10.1016/j.acra.2015.04.004
22. Maruyama M, Yoshizako T, Uchida K, Araki H, Tamaki Y, Ishikawa N, et al. Comparison of Utility of Tumor Size and Apparent Diffusion Coefficient for Differentiation of Low- and High-Grade Clear-Cell Renal Cell Carcinoma. *Acta Radiol* (2015) 56:250–6. doi: 10.1177/0284185114523268
23. Wang HK, Zhu Y, Yao XD, Zhang SL, Dai B, Zhang HL, et al. External Validation of a Nomogram Using RENAL Nephrometry Score to Predict High Grade Renal Cell Carcinoma. *J Urol* (2012) 187:1555–60. doi: 10.1016/j.juro.2011.12.099
24. Shu J, Tang Y, Cui J, Yang R, Meng X, Cai Z, et al. Clear Cell Renal Cell Carcinoma: CT-Based Radiomics Features for the Prediction of Fuhrman Grade. *Eur J Radiol* (2018) 109:8–12. doi: 10.1016/j.ejrad.2018.10.005
25. Ganeshan B, Miles KA. Quantifying Tumour Heterogeneity With CT. *Cancer Imaging* (2013) 13:140–9. doi: 10.1102/1470-7330.2013.0015
26. Miles KA, Ganeshan B, Hayball MP. 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
27. Zhu YH, Wang X, Zhang J, Chen YH, Kong W, Huang YR. Low Enhancement on Multiphase Contrast-Enhanced CT Images: An Independent Predictor of the Presence of High Tumor Grade of Clear Cell Renal Cell Carcinoma. *AJR Am J Roentgenol* (2014) 203:295–300. doi: 10.2214/AJR.13.12297
28. Pichler M, Hutterer GC, Chromecki TF, Jesche J, Kampel-Kettner K, Rehak P, et al. Histologic Tumor Necrosis is an Independent Prognostic Indicator for Clear Cell and Papillary Renal Cell Carcinoma. *Am J Clin Pathol* (2012) 137:283–9. doi: 10.1309/AJCPLBK9L9KDYQZP
29. 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) 19:6. doi: 10.1186/s40644-019-0195-7

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

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 30 July 2021
doi: 10.3389/fonc.2021.665807

CT-Based Pelvic T_1-Weighted MR Image Synthesis Using UNet, UNet++ and Cycle-Consistent Generative Adversarial Network (Cycle-GAN)

*Reza Kalantar[1], Christina Messiou[1,2], Jessica M. Winfield[1,2], Alexandra Renn[2], Arash Latifoltojar[2], Kate Downey[2], Aslam Sohaib[2], Susan Lalondrelle[3], Dow-Mu Koh[1,2] and Matthew D. Blackledge[1]**

[1] Division of Radiotherapy and Imaging, The Institute of Cancer Research, London, United Kingdom, [2] Department of Radiology, The Royal Marsden Hospital, London, United Kingdom, [3] Gynaecological Unit, The Royal Marsden Hospital, London, United Kingdom

OPEN ACCESS

Edited by:
Oliver Diaz,
University of Barcelona, Spain

Reviewed by:
Guang Yang,
Imperial College London,
United Kingdom
Xiaoran Chen,
ETH Zürich, Switzerland

***Correspondence:**
Matthew D. Blackledge
Matthew.Blackledge@icr.ac.uk

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

Received: 09 February 2021
Accepted: 15 July 2021
Published: 30 July 2021

Citation:
Kalantar R, Messiou C, Winfield JM, Renn A, Latifoltojar A, Downey K, Sohaib A, Lalondrelle S, Koh D-M and Blackledge MD (2021) CT-Based Pelvic T_1-Weighted MR Image Synthesis Using UNet, UNet++ and Cycle-Consistent Generative Adversarial Network (Cycle-GAN). Front. Oncol. 11:665807. doi: 10.3389/fonc.2021.665807

Background: Computed tomography (CT) and magnetic resonance imaging (MRI) are the mainstay imaging modalities in radiotherapy planning. In MR-Linac treatment, manual annotation of organs-at-risk (OARs) and clinical volumes requires a significant clinician interaction and is a major challenge. Currently, there is a lack of available pre-annotated MRI data for training supervised segmentation algorithms. This study aimed to develop a deep learning (DL)-based framework to synthesize pelvic T_1-weighted MRI from a pre-existing repository of clinical planning CTs.

Methods: MRI synthesis was performed using UNet++ and cycle-consistent generative adversarial network (Cycle-GAN), and the predictions were compared qualitatively and quantitatively against a baseline UNet model using pixel-wise and perceptual loss functions. Additionally, the Cycle-GAN predictions were evaluated through qualitative expert testing (4 radiologists), and a pelvic bone segmentation routine based on a UNet architecture was trained on synthetic MRI using CT-propagated contours and subsequently tested on real pelvic T_1 weighted MRI scans.

Results: In our experiments, Cycle-GAN generated sharp images for all pelvic slices whilst UNet and UNet++ predictions suffered from poorer spatial resolution within deformable soft-tissues (e.g. bladder, bowel). Qualitative radiologist assessment showed inter-expert variabilities in the test scores; each of the four radiologists correctly identified images as acquired/synthetic with 67%, 100%, 86% and 94% accuracy. Unsupervised segmentation of pelvic bone on T1-weighted images was successful in a number of test cases

Conclusion: Pelvic MRI synthesis is a challenging task due to the absence of soft-tissue contrast on CT. Our study showed the potential of deep learning models for synthesizing realistic MR images from CT, and transferring cross-domain knowledge which may help to expand training datasets for 21 development of MR-only segmentation models.

Keywords: convolutional neural network (CNN), generative adversarial network (GAN), medical image synthesis, radiotherapy planning, magnetic resonance imaging (MRI), computed tomography (CT)

INTRODUCTION

Computed tomography (CT) is conventionally used for the delineation of the gross tumor volume (GTV) and subsequent clinical/planning target volumes (CTV/PTV), along with organs-at-risk (OARs) in radiotherapy (RT) treatment planning. Resultant contours allow optimization of treatment plans by delivering the required dose to PTVs whilst minimizing radiation exposure of the OARs by ensuring that spatial dose constraints are not exceeded. Magnetic resonance imaging (MRI) offers excellent soft-tissue contrast and is generally used in conjunction with CT to improve visualization of the GTV and surrounding OARs during treatment planning. However, manual definition of these regions is repetitive, cumbersome and may be subject to inter- and/or intra-operator variabilities (1). The recent development of the combined MR-Linac system (2) provides the potential for accurate treatment adaption through online MR-imaging acquired prior to each RT fraction. However, re-definition of contours for each MR-Linac treatment fraction requires approximately 10 minutes of downtime whilst the patient remains on the scanner bed, placing additional capacity pressures on clinicians wishing to adopt this technology.

Deep learning (DL) is a sub-category of artificial intelligence (AI), inspired by the human cognition system. In contrast to traditional machine learning approaches that use hand-engineered image-processing routines, DL is able to learn complex information from large datasets. In recent years, DL-based approaches have shown great promise in medical imaging applications, including image synthesis (3, 4) and automatic segmentation (5–7). There is great promise for DL to drastically accelerate delineation of the GTV and OARs in MR-Linac studies, yet a major hurdle remains the lack of large existing pre-contoured MRI datasets for training supervised segmentation networks. One potential solution is transferring knowledge from pre-existing RT planning repositories on CT to MRI in order to facilitate domain adaptive segmentation (8). Previous studies have reported successful implementation of GANs in generating realistic CT images from MRI (3, 9–11) as well as MRI synthesis from CT in the brain (12). To date, few studies have investigated MRI synthesis in the pelvis. Dong et al. (13) proposed a synthetic MRI-assisted framework for improved multi-organ segmentation on CT. However, although the authors suggested that synthetic MR images improved segmentation results, the quality of synthesis was not investigated in depth. MR image synthesis from CT is a challenging task due to large soft-tissue signal intensity variations. In particular, MRI synthesis in the pelvis offers the considerable difficulty posed by geometrical differences in patient anatomy as well as unpredictable discrepancies in bladder and bowel contents.

In this study, we compare and contrast paired and unpaired generative techniques for synthesizing T_1-weighted (T_1W) MR images from pelvic CT scans as a basis for training algorithms for OARs and tumor delineation on acquired MRI datasets. We include in our analysis the use of state-of-the-art UNet (14) and UNet++ (15) architectures for paired training, testing two different loss functions [L_1 and VGG-19 perceptual loss (16)], and compare our results with a Cycle-Consistent Generative Adversarial Network (Cycle-GAN) (17) for unpaired MR image synthesis. Subsequently, we evaluate our results through blinded assessment of synthetic and acquired images by expert radiologists, and demonstrate our approach for pelvic bone segmentation on acquired T_1W MRI from a framework trained solely on synthetic $_1$W MR images with CT-propagated contours.

MATERIALS AND METHODS

Patient Population and Imaging Protocols

Our cohort consisted of 26 patients with lymphoma who underwent routine PET/CT scanning (Gemini, Philips, Cambridge, United States) and whole-body T_1W MRI (1.5T, Avanto, Siemens Healthcare, Erlangen, Germany) before and after treatment (see **Table 1** for imaging protocols). From this cohort, image series with large axial slice angle mismatch between CT and MR images, and those from patients with metal implants were excluded, leaving 28 paired CT/MRI datasets from 17 patients. The studies involving human participants were reviewed and approved by the Committee for

TABLE 1 | Imaging parameters for acquired CT and T_1W MR images.

CT parameters		T_1W MR parameters	
Peak Voltage Output (kVp)	120	Acquisition Sequence	2D Spoiled Gradient Echo
Acquisition Type	Helical	Echo Time (ms)	4.8
Slice Thickness (mm)	3-6.5	Repetition Time (ms)	386
Matrix Size	512 × 512	Phase Encoding Direction	Anterior-Posterior
Pixel Spacing (mm^2)	0.74 × 0.74-1.17 × 1.17	Acquired Matrix Size (read)	256
Exposure (mAs)	26-80	Reconstructed Matrix Size (read)	512
		Reconstructed Pixel Size (mm^2)	0.74 × 0.74-0.82 × 0.82
		Flip Angle	70°
		Slice Orientation	Axial
		Slice Thickness (mm)	5
		Acceleration	GRAPPA (R=2)
		Bandwidth	Pixel
		(Hz/pixel)	331

Some parameters are shown as the range of values (minimum-maximum) existing in the patient datasets.

Clinical Research at the Royal Marsden Hospital. The patients/participants provided written informed consent to participate in this study.

Model Architectures

We investigated three DL architectures for MR image synthesis: (i) UNet, (ii) UNet++, and (iii) Cycle-GAN. UNet is one of the most popular DL architectures for image-to-image translations, with initial applications in image segmentation (14). In essence, UNet is an auto-encoder with addition of skip connections between encoding and decoding sections to maintain spatial resolution. In this study, a baseline UNet model was designed consisting of 10 consecutive convolutional blocks (5 encoding and 5 decoding blocks), each using batch normalization and ReLU activation for CT-to-MR image generation (**Figure 1A**). Additionally, a UNet++ model with interconnected skip connection pathways, as described in (15), was developed with the same number of encoder-decoder sections and kernel filters as the baseline UNet (**Figure 1B**). UNet++ was reported to enhance performance (15), therefore we deployed this architecture to assess its capabilities for paired image synthesis.

GANs are the state-of-the-art approaches for generating photo-realistic images based on the principles of game theory (18). In image synthesis applications, GANs typically consist of two CNNs, a generator and a discriminator. During training, the generator produces a target synthetic image from an input image with different modality; the discriminator then attempts to classify whether the synthetic image is genuine. Training is successful once the generator is able to synthesize images that the discriminator is unable to differentiate from real examples. Progressive co-training of the generator and discriminator leads to learning of the global conditional probability distribution from input to target domain. In this study, a Cycle-GAN model (17) was developed to facilitate unpaired CT-to-MR and MR-to-CT learning. The baseline UNet model was used as the network generator, and the discriminator composed of 5 blocks containing 2D convolutional layers followed by instance normalization and leaky ReLU activation. This technique offers the advantage that it does not require spatial alignment between training T_1W MR and CT images. The high-level schematic of the Cycle-GAN network is shown in **Figure 2**.

For segmentation, we propose a framework that first generates synthetic T1W MR images from CT, propagates ground-truth CT contours and outputs segmentation contours on acquired T1W MR images. To examine the capability of our fully-automated DL framework for knowledge transfer from CT to MRI, we generate ground-truth contours of the bones using a

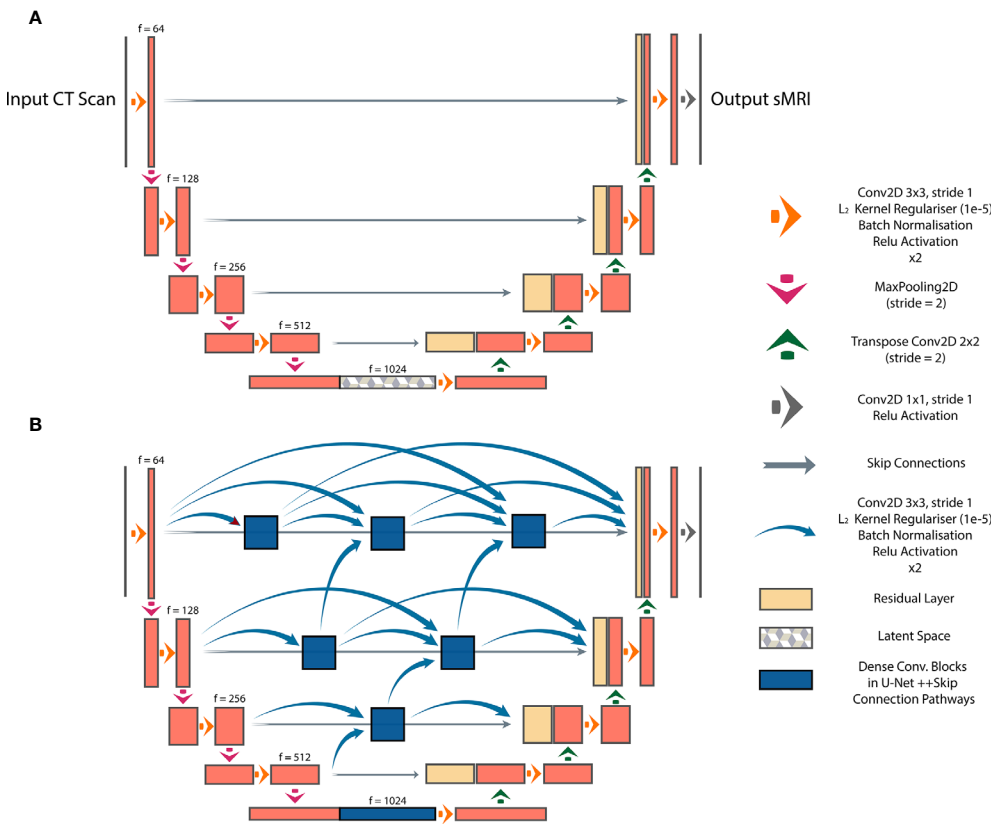


FIGURE 1 | Paired image-to-image networks, **(A)** UNet with symmetrical skip connections between the encoder and decoder, **(B)** UNet++ with interconnected skip connection convolutional pathways.

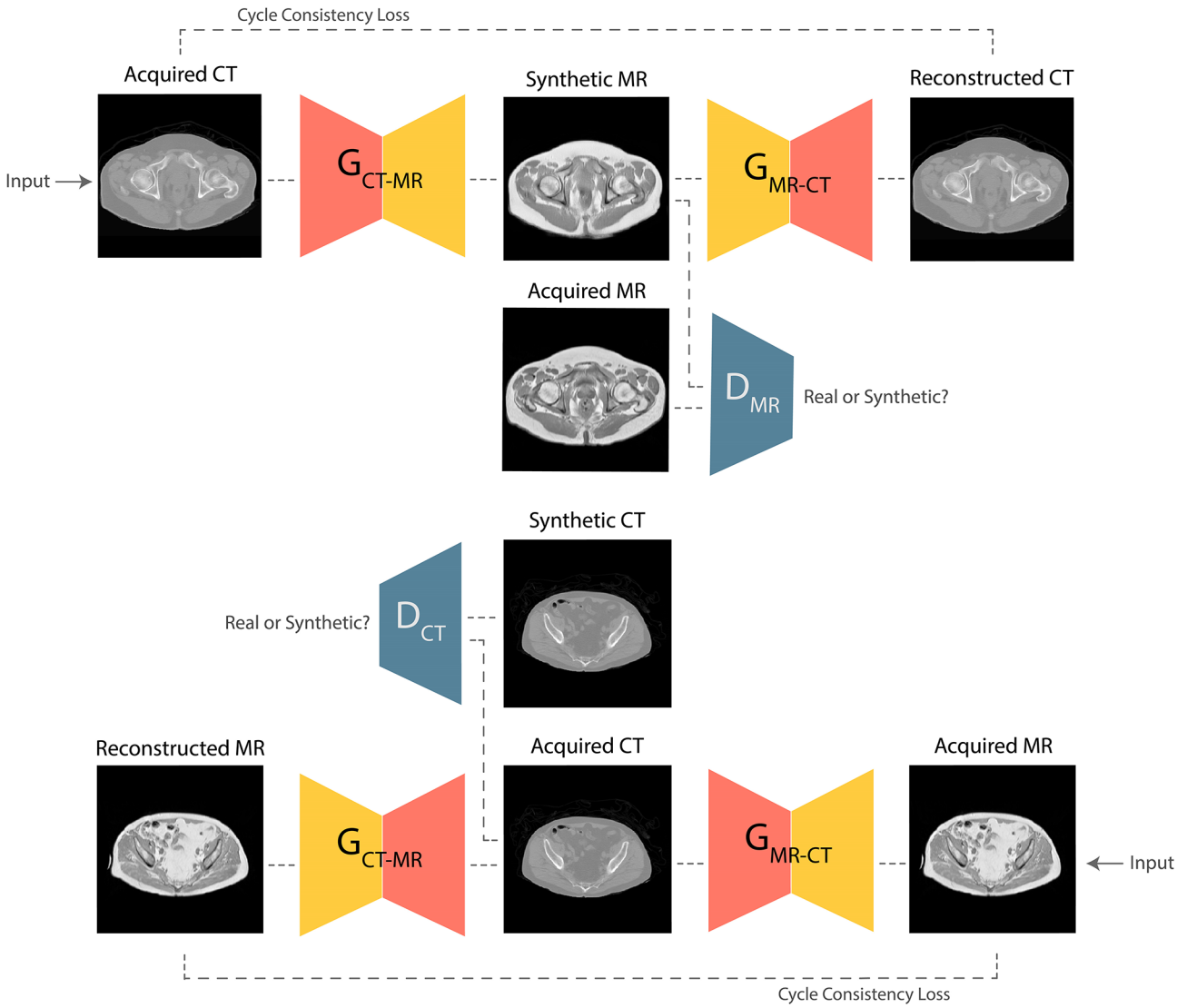


FIGURE 2 | Schematic of the Cycle-GAN network. During training, images from CT domain are translated to MRI domain and reconstructed back to CT domain under the governance of adversarial and cycle consistency loss terms respectively. Co-training of CT-to-MRI and MRI-to-CT models leads to generation of photo-realistic predictions.

Gaussian mixture model proposed by Blackledge etal. (19) and transfer them to synthetic MR images as a basis for our segmentation training. A similar UNet model to the architecture presented in **Figure 1**, with 5 convolutional blocks (convolution-batchnorm-dropout(p=0.5)-ReLU) in the encoding and decoding sections was developed to perform binary bone segmentation from synthetic MR images. The schematic of our proposed synthesis/segmentation framework is illustrated in **Figure 3**.

Image Preprocessing

In preparation for paired training, the corresponding CT and T_1W MR slices from the anatomical pelvic station in each patient were resampled using a 2D affine transformation followed by non-rigid registration using multi-resolution B-Spline free-form deformation (loss = Mattes mutual information, histogram bins = 50, gradient descent line search optimizer parameters: learning rate = 5.0, number of iterations = 50, convergence window size = 10) (20). The resulting co-registered images were visually qualified based on the alignment of rigid pelvic landmarks. In CT images, signal intensities outside of the range -1000 and 1000 HU were truncated to limit the dynamic range. The T_1W MR images were corrected using N4 bias-field correction to reduce inter-patient intensity variations and inhomogeneities (21) and signal intensities above 1500 (corresponding to infrequent high intensity fatty regions) were truncated. Subsequently, the training images were normalized

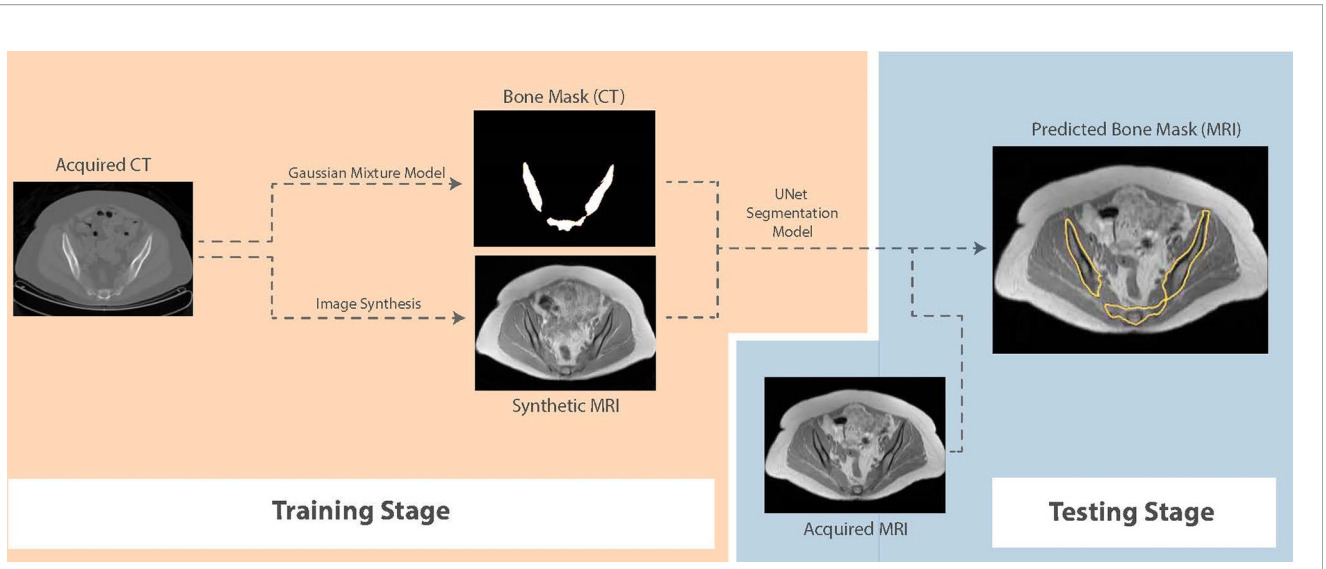


FIGURE 3 | Schematic of the proposed fully-automatic combined synthesis and segmentation framework for knowledge transfer from CT scans to MR images. The intermediate synthesis stage enables segmentation training using CT-based contours and MR signal distributions.

to intensity ranges (0,1) and (-1,1) prior to paired (UNet, UNet++) and unpaired (Cycle-GAN) training respectively.

Objective Functions

Common loss functions in image synthesis are mean absolute error (MAE or L_1) and mean squared error (MSE or L_2) between the target domain and the synthetic output. However, such loss functions ignore complex image features such as texture and shape. Therefore, for UNet/UNet++ models, we compared L_1 loss in the image space with L_1 loss calculated based on the features extracted from a previously-trained object classification network, deriving the "perceptual loss". For this purpose, the VGG-19 classification network was used (16), which is composed of 5 convolutional layers and 19 layers overall, and used features extracted from the "block conv2d" layer. For Cycle-GAN training, the difference between L_1 and the structural similarity index (SSIM) (defined as L_1 – SSIM) was used as the loss to govern the cycle consistency, whilst L_1 and L_2 losses were used for the generator and the discriminator respectively. For segmentation training, the Dice loss (1, 2) was used to perform binary division of bone on MR images.

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \quad (1)$$

$$Dice\ loss = 1 - DSC \quad (2)$$

where A and B denote the generated and ground-truth contours.

Model Training and Evaluation

The dataset was split to 981, 150 and 116 images from 11, 3 and 3 patients for training, validation and testing respectively. All models were trained for 150 epochs using the Adam optimizer (learning rate = 1e-4; UNet and UNet++ models: batch size = 5, Cycle-GAN: batch size=1) on a NVIDIA RTX6000 GPU (Santa Clara, California, United States) (**Table 2**). During paired UNet/UNet++ training, the peak signal-to-noise ratio (PSNR), SSIM, L_1 and L_2 quantitative metrics, as described in (22), were recorded at each epoch for the validation images. The trained weights with the lowest validation loss were used to generate synthetic T_1W MR images from the test CT images. Optimal weights from the Cycle-GAN model were selected based on visual examination of the network predictions of the validation data following each epoch. Subsequently, synthetic images from all models were evaluated against the ground-truth acquired MR images quantitatively using the above-mentioned imaging metrics. An additional qualitative test was designed to obtain unbiased clinical examination of predictions from the Cycle-GAN model. This test consisted of two sections: (i) to blindly classify randomly-selected test images as synthetic or acquired, and outline reasoning for answers (18 synthetic and 18 acquired test MR slices), and (ii) to describe key differences between synthetic and acquired test T_1W MR images when the input CT and ground truth acquired MR images were also provided (10 sets of images from 3 test patients). This test was completed by 4 radiologists (two with <5 years and two with >5 years of

TABLE 2 | Learnable parameters (in millions) of UNet, UNet++ and Cycle-GAN models.

	UNet (L_1)	UNet (VGG)	UNet++ (L_1)	Cycle-GAN
Trainable Parameters (M)	31	31	36	31(G), 11(D)

G and D denote the Cycle-GAN generator and discriminator respectively.

experience). The segmentation network was trained on Cycle-GAN generated synthetic MR images (training: 14, validation: 3 patients) for 600 epochs using the Adam optimizer (learning rate = 1e-4) and batch size of 1. To avoid overfitting, random linear shear and rotation (range:0, π/60) were applied to images during training.

RESULTS

Quantitative assessment of synthetic T_1W MR images from the validation dataset during paired algorithm training suggested that the UNet and UNet++ models with L_1 loss displayed higher PSNR and SSIM, and lower L_1 and L_2 values compared with the generated images from the UNet model with the VGG-19 perceptual loss (**Figure 4**). Quantitative analysis of synthetic images from the test patients revealed a similar trend for UNet and UNet++ model predictions and showed that the Cycle-GAN quantitative values were the lowest in all metrics but the SSIM where it was only higher than UNet (VGG) predictions (**Table 3**). Moreover, qualitative evaluation of predictions from all models revealed a noticeable difference in sharpness (spatial resolution) between the images generated from paired (UNet and UNet++) and unpaired (Cycle-GAN) training. It was observed that despite UNet and UNet++ models generating relatively realistic predictions for pelvic slices consisting of fixed and bony structures (e.g. femoral heads, hip bone, muscles), they yielded blurry and unrealistic patches for deformable and variable pelvic structures such as bowel, bladder and rectum. In contrast, the Cycle-GAN model generated sharp images for all pelvic slices, yet a disparity in contrast was observed for soft-tissues with large variabilities in training patient MRI slices (e.g. bowel content, gas in rectum and bowel, bladder filling) (**Figure 5**).

Our expert radiologist qualitative testing on Cycle-GAN predicted images suggested that there were inter-expert variabilities in scores from section one of the test, highlighting the differences in subjective decisions amongst the experts in a number of test images. Experts 1 and 2 (<5 years of experience) scored 67% and 100% whilst experts 3 and 4 (>5 years of experience) correctly identified 86% and 94% of total 36 test images. Hence, no particular correlation was observed between the percentage scores and the participants' years of experience (**Figure 6A**). Radiologist comments on the synthetic images (following unblinding) are presented in **Figure 6B**.

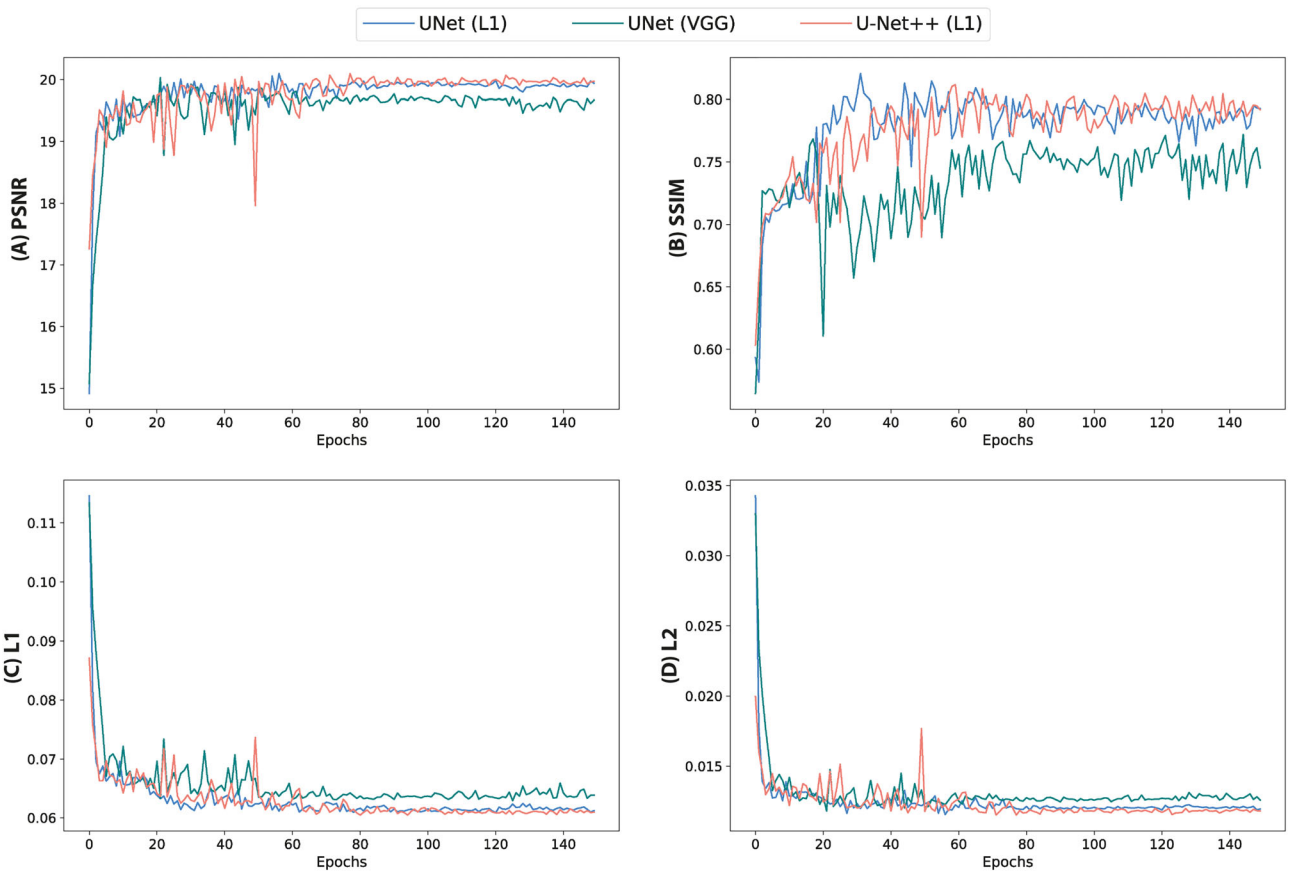


FIGURE 4 | Quantitative metrics calculated from validation images during training of UNet and UNet++ models for 150 epochs. **(A)** PSNR, **(B)** SSIM, **(C)** L_1 loss and **(D)** L_2 loss.

TABLE 3 | Quantitative analysis of predictions from the trained models on test patients.

	UNet (L_1)	UNet (VGG)	UNet++ (L_1)	Cycle-GAN PSNR
PSNR	**20.169 ± 0.196**	19.668 ± 0.189	20.080 ± 0.193	18.279 ± 0.156
SSIM	**0.809 ± 0.003**	0.728 ± 0.003	0.804 ± 0.003	0.783 ± 0.003
MAE	**0.043 ± 0.001**	0.047 ± 0.001	0.044 ± 0.001	0.057 ± 0.001
MSE	**0.011 ± 0.001**	**0.011 ± 0.001**	0.013 ± 0.001	0.016 ± 0.001

The calculated metrics are presented as mean and standard deviation. The pixel intensities outside the body were excluded when deriving these measurements. The best quantitative metrics are shown in bold.

The bone segmentation results using our fully-automated approach showed that our proposed framework successfully performed unsupervised segmentation of the bone from acquired T_1W MR images, without the requirement of any manually annotated regions of interest (ROIs). The outcome from various pelvic slices across 8 patients from our in-house cohort are presented in **Figure 7**. The segmentation results from cases 5 to 8 were from patients not used in the synthesis and segmentation components of our framework. Test case 8 demonstrates the predicted bone contours from a patient with metal hip implant.

DISCUSSION AND CONCLUSION

One major limitation in adaptive RT on the MR-Linac system is the need for manual annotation of OARs and tumors on patient scans for each RT fraction which requires significant clinician interaction. DL-based approaches are promising solutions to automate this task and reduce burden on clinicians. However, the development of these algorithms is hindered by the paucity of pre-annotated MRI datasets for training and validation. In this study, we developed paired and unpaired training for T_1W MR image synthesis from pelvic CT scans as a data generative tool for training of segmentation algorithms for MR-Linac RT treatment planning. Our results suggested that the Cycle-GAN network generated synthetic images with the greatest visual fidelity across all pelvic slices whilst the synthetic images from UNet and UNet++ appeared less sharp, which is likely due to soft-tissue misalignments during the registration process. The observed disparity in contrast in Cycle-GAN images for bladder, bone marrow and bowel loops may be due to large variabilities in our relatively small training dataset. Although the direct impact of these contrast discrepancies on MRI segmentation performance is yet to be evaluated, the Cycle-GAN predictions appeared more suitable for CT contour propagation to synthetic MRI than UNet and UNet++ images due to distinctive soft-tissue boundaries and high-resolution synthesis.

Quantitative analysis of all model predictions indicated that the imaging metrics did not fully conform with the output image visual fidelity and apparent sharpness. This finding was in fact in line previous studies comparing paired and unpaired MRI synthesis (12, 22). CT-to-MR synthesis in the pelvis offers the considerable challenge of generating soft-tissue contrasts absent on acquired CT scans. Although quantitative metrics such as the PSNR, SSIM, L_1 and L_2 differences are useful measures when comparing images, they may not directly correspond to photo-realistic network outcome. This was evident in quantitative evaluation of the images generated from the UNet and UNet++ models trained with L_1 loss in the image space against UNet with VGG-19 perceptual loss and Cycle-GAN predictions. Therefore, expert clinician qualitative assessments may provide a more reliable insight into the performance of medical image generative networks. In this study, our expert evaluation test based on Cycle-GAN predictions suggested that despite a number of suboptimal soft-tissue contrast predictions (e.g. urinary bladder filling, bone marrow, nerves), there were differences in radiologist accuracies for correctly identifying synthetic from acquired MR images. The fact that 3/4 radiologists were unable to accurately identify synthetic images in all cases highlights the capability of our model to generate realistic medical images that may be indistinguishable from acquired MRI.

Our segmentation results demonstrated the capability of our fully-automated framework in segmenting bones on acquired MRI images with no manual MR contouring. Domain adaptation offers a significant clinical value in transferring knowledge from previously-contoured OARs by experts on CT to MR-only treatment planning procedures. Additionally, it potentially enables expanding medical datasets which are essential for training supervised DL models. Such a technique is also highly valuable outside the context of radiotherapy, as body MRI has increasing utility for monitoring patients with secondary bone disease from primary prostate (23) and breast (24) cancers, and multiple myeloma (25). Quantitative assessment of response of these diseases to systemic treatment using MRI is hindered by the lack of automated skeletal delineation algorithms to monitor changes in large volume disease regions (26).

GANs are notoriously difficult to train due to their large degree of application-based hyper-parameter optimization and non-standardized training techniques. However, this study showed that even when trained on relatively small datasets, GANs may have the potential to generate realistic images to overcome the challenge of medical image data shortage. Therefore, fut ure studies will investigate the performance of the proposed framework on larger datasets and alternative pelvic OARs, as well as exploring novel techniques to enforce targeted organ contrast during GAN and segmentation training. Additionally, future research will examine the performance sensitivity on the level of manual MRI contours required for training cross-domain DL algorithms.

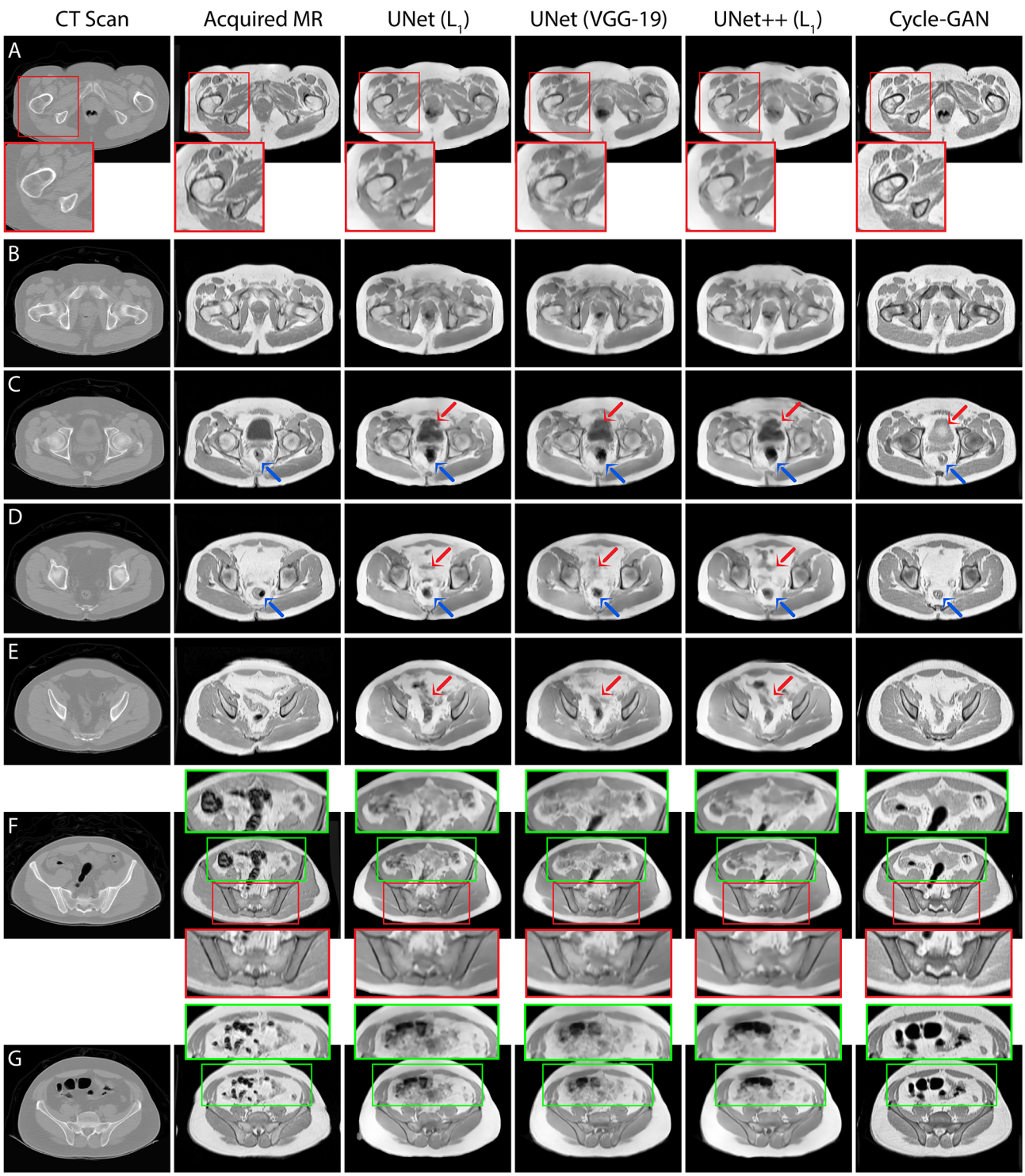
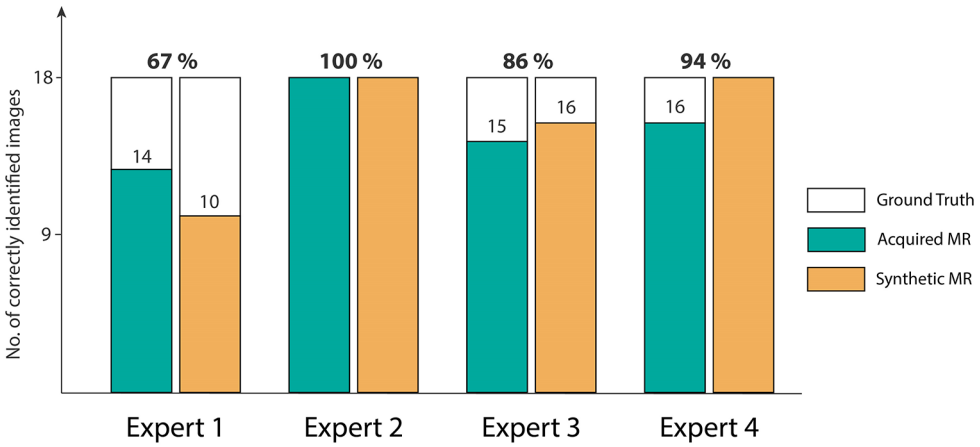


FIGURE 5 | T_1T1W MRI predictions generated from 3 independent test patients using UNet, UNet++ and Cycle-GAN models (panel **A**: patient 1, panels **B–E**, **G**: patient 2, panel **F**: patient 3). Red box: Predictions from pelvic slices with relatively fixed geometries including the bones demonstrate sharp boundaries between anatomical structures, with visually superior results for the Cycle-GAN architecture (panels **A**, **F**). Green box: The superior resolution of the Cycle-GAN architecture is further exemplified in slices with deformable structures such as the bowel loop (panels **F**, **G**). In highly deformable regions, minor contrast disparity in anatomical structures can be observed in the synthetic MRI; examples include prediction of bladder (red arrows in panel **C**), lower gastrointestinal region (red arrows in panels **D**, **E**) and rectum (blue arrows in panels **C**, **D**).

A. Section One



B. Section Two

Expert 1
• Unrealistic bone contrast (e.g. bone marrow, ileum) • Lack of bone heterogeneity, they appear 'too perfect' • Synthetic and acquired MRI indistinguishable in 5 test slices
Expert 2
• Bone marrow signal hypointensity • Muscle appearance (radially-orientated linear hyperintensities) • Tendon/nerves underrepresentation • Urinary bladder hyperintensity on synthetic MR
Expert 3
• Unrealistic bone contrast (e.g. bone marrow) • Cortex appearance (too thick) • Less motion and susceptibility artifacts on synthetic MR (smoothing effect) • Soft-tissue contrast good overall
Expert 4
• Unrealistic bone/muscle boundary contrast • Streak artifact and 'grainy' texture on synthetic MR • Less contrast on bladder and surrounding peritoneal fat • Less contrast between bowel loops and adjacent sidewall structures • Unusual subcutaneous fat appearance

FIGURE 6 | **(A)** Section One: Expert scores for identifying evenly-distributed test patient MRI slices as synthetic or acquired, **(B)** Section Two: Expert comments on Cycle-GAN synthetic MRI when presented along with the ground truth CT and acquired T_1W MRI (Experts 1 and 2 with <5 years of experience, and experts 3 and 4 with >5 years of experience).

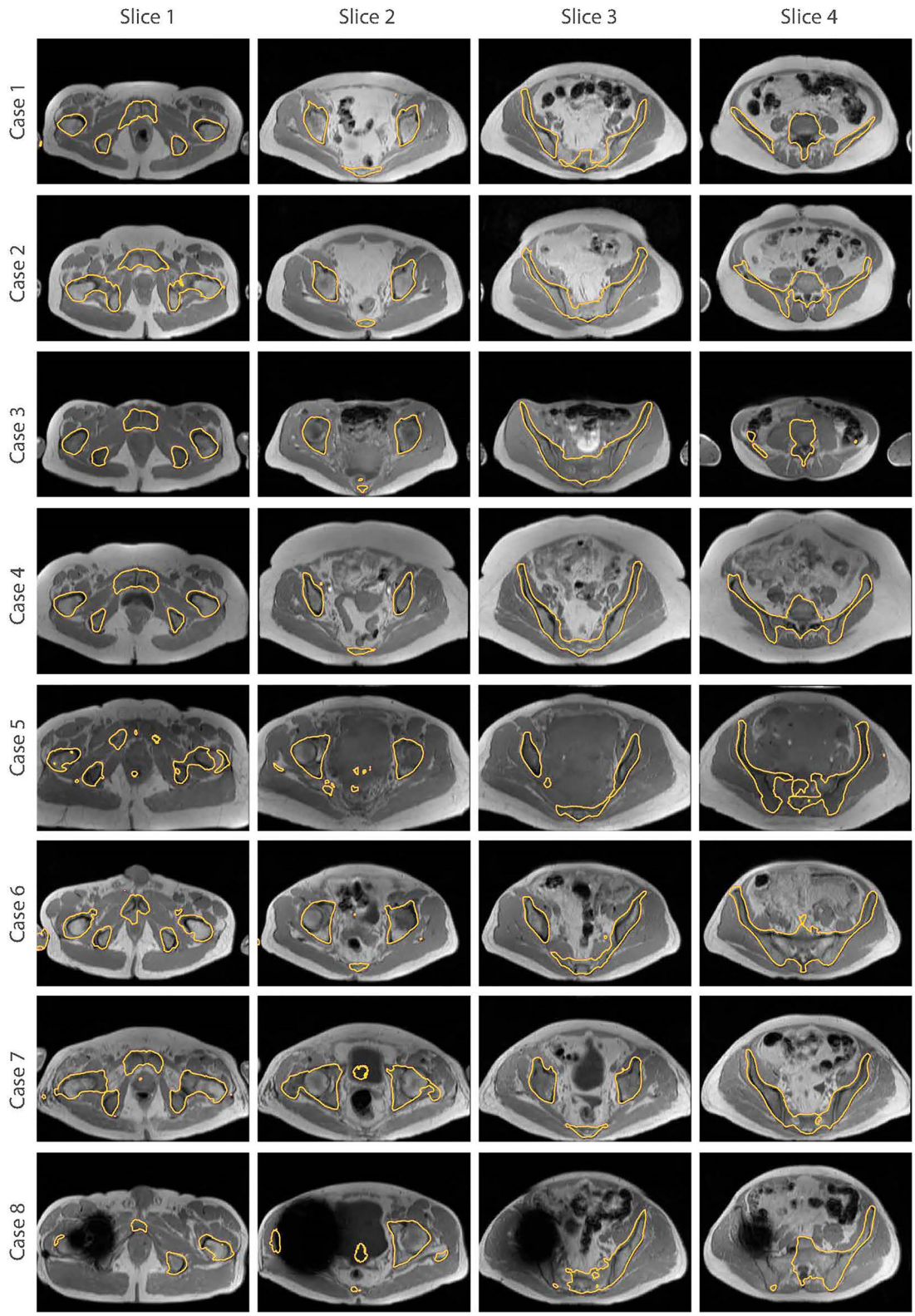


FIGURE 7 | Bone segmentation results from acquired T_1W MRI scans of 8 test patients using the proposed fully-automated framework. The combined synthesis/segmentation network allows transfer of organ- specific encoded spatial information from CT to MRI without the need to manually define ROIs. Cases 5 to 8 were patients not included in the synthesis stage of network training. Case 8 shows bone segmentation results from a patient with metal hip.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The datasets presented in this article are not readily available due to patient confidentiality concerns. Requests to access these datasets should be directed to matthew.Blackledge@icr.ac.uk.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Committee for Clinical Research at the Royal Marsden Hospital. The patients/participants provided written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work, and approved it for publication.

FUNDING

This project represents independent research funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and the Institute of Cancer Research, London, United Kingdom.

REFERENCES

1. Millioni R, Sbrignadello S, Tura A, Iori E, Murphy E, Tessari P. The Inter- and Intra-Operator Variability in Manual Spot Segmentation and its Effect on Spot Quantitation in Two-Dimensional Electrophoresis Analysis. *Electrophoresis* (2010) 31:1739–42. doi: 10.1002/elps.200900674
2. Raaymakers B, Lagendijk J, Overweg J, Kok J, Raaijmakers A, Kerkhof E, et al. Integratinga 1.5 T Mri Scanner With a 6 Mv Accelerator: Proof of Concept. *Phys Med Biol* (2009) 54:N229. doi: 10.1088/0031-9155/54/12/N01
3. Wolterink JM, Dinkla AM, Savenije MH, Seevinck PR, van den Berg CA, Is˙gum I. "Deep Mr to Ct Synthesis Using Unpaired Data". In: *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, Cham (2017). p. 14–23. doi: 10.1007/978-3-319-68127-6_2
4. Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, et al. "Medical Image Synthesis With Context-Aware Generative Adversarial Networks". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham (2017). p. 417–25. doi: 10.1007/978-3-319-66179-7_48
5. Song Y, Hu J, Wu Q, Xu F, Nie S, Zhao Y, et al. Automatic Delineation of the Clinical Target Volume and Organs at Risk by Deep Learning for Rectal Cancer Postoperative Radiotherapy. *Radiother Oncol* (2020) 145:186–92. doi: 10.1016/j.radonc.2020.01.020
6. Boon IS, Au Yong T, Boon CS. Assessing the Role of Artificial Intelligence (Ai) in Clinical Oncology: Utility of Machine Learning in Radiotherapy Target Volume Delineation. *Medicines* (2018) 5:131. doi: 10.3390/medicines5040131
7. Lustberg T, van Soest J, Gooding M, Peressutti D, Aljabar P, van der Stoep J, et al. Clinical Evaluation of Atlas and Deep Learning Based Automatic Contouring for Lung Cancer. *Radiother Oncol* (2018) 126:312–7. doi: 10.1016/j.radonc.2017.11.012
8. Masoudi S, Harmon S, Mehralivand S, Walker S, Ning H, Choyke P, et al. Cross Modality Domain Adaptation to Generate Pelvic Magnetic Resonance Images From Computed Tomography Simulation for More Accurate Prostate Delineation in Radiotherapy Treatment. *Int J Radiat Oncol Biol Phys* (2020) 108:e924. doi: 10.1016/j.ijrobp.2020.07.572
9. Yang H, Sun J, Carass A, Zhao C, Lee J, Xu Z, et al. "Unpaired Brain Mr-to-Ct Synthesis Using a Structure-Constrained Cyclegan". In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer (2018). p. 174–82. doi: 10.1007/978-3-030-00889-5_20
10. Lei Y, Harms J, Wang T, Liu Y, Shu H-K, Jani AB, et al. Mri-Only Based Synthetic Ct Generation Using Dense Cycle Consistent Generative Adversarial Networks. *Med Phys* (2019) 46:3565–81. doi: 10.1002/mp.13617
11. Liang X, Chen L, Nguyen D, Zhou Z, Gu X, Yang M, et al. Generating Synthesized Computed Tomography (Ct) From Cone-Beam Computed Tomography (CBCT) Using Cyclegan for Adaptive Radiation Therapy. *Phys Med Biol* (2019) 64:125002. doi: 10.1088/1361-6560/ab22f9
12. Jin C-B, Kim H, Liu M, Jung W, Joo S, Park E, et al. Deep Ct to Mr Synthesis Using Paired and Unpaired Data. *Sensors* (2019) 19:2361. doi: 10.3390/s19102361
13. Dong X, Lei Y, Tian S, Wang T, Patel P, Curran WJ, et al. Synthetic Mri-Aided Multi-Organ Segmentation on Male Pelvic Ct Using Cycle Consistent Deep Attention Network. *Radiother Oncol* (2019) 141:192–9. doi: 10.1016/j.radonc.2019.09.028
14. Ronneberger O, Fischer P, Brox T. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, Cham (2015). p. 234–41. doi: 10.1007/978-3-319-24574-4_28
15. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Trans Med Imaging* (2019) 39:1856–67. doi: 10.1109/TMI.2019.2959609
16. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. (2014) arXiv preprint arXiv:1409.1556.
17. Zhu J-Y, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. In: *Proceedings of the IEEE International Conference on Computer Vision* (2017) p. 2223–32. doi: 10.1109/ICCV.2017.244
18. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems* (2014). p. 2672–80.
19. Blackledge MD, Collins DJ, Koh D-M, Leach MO. Rapid Development of Image Analysis Research Tools: Bridging the Gap Between Researcher and Clinician With Pyosirix. *Comput Biol Med* (2016) 69:203–12. doi: 10.1016/j.compbiomed.2015.12.002
20. Keszei AP, Berkels B, Deserno TM. Survey of Non-Rigid Registration Tools in Medicine. *J Digital Imaging* (2017) 30:102–16. doi: 10.1007/s10278-016-9915-8
21. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. "N4itk: Improved N3 Bias Correction". In: *IEEE Transactions on Medical Imaging*. (2010) p. 29:1310–20. doi: 10.1109/TMI.2010.2046908
22. Li W, Li Y, Qin W, Liang X, Xu J, Xiong J, et al. Magnetic Resonance Image (Mri) Synthesis From Brain Computed Tomography (Ct) Images Based on Deep Learning Methods for Magnetic Resonance (Mr)-Guided Radiotherapy. *Quant Imaging Med Surg* (2020) 10:1223. doi: 10.21037/qims-19-885
23. Padhani AR, Lecouvet FE, Tunariu N, Koh D-M, De Keyzer F, Collins DJ, et al. Metastasis Reporting and Data System for Prostate Cancer: Practical Guidelines for Acquisition, Interpretation, and Reporting of Whole-Body Magnetic Resonance Imaging-Based Evaluations of Multiorgan Involvement in Advanced Prostate Cancer. *Eur Urol* (2017) 71:81–92. doi: 10.1016/j.eururo.2016.05.033
24. Padhani AR, Koh D-M, Collins DJ. Whole-Body Diffusion-Weighted Mr Imaging in Cancer: Current Status and Research Directions. *Radiology* (2011) 261:700–18. doi: 10.1148/radiol.11110474
25. Messiou C, Hillengass J, Delorme S, Lecouvet FE, Moulopoulos LA, Collins DJ, et al. Guidelines for Acquisition, Interpretation, and Reporting of Whole-Body Mri in Myeloma: Myeloma Response Assessment and Diagnosis System (My-Rads). *Radiology* (2019) 291:5–13. doi: 10.1148/radiol.2019181949

26. Blackledge MD, Collins DJ, Tunariu N, Orton MR, Padhani AR, Leach MO, et al. Assessment of Treatment Response by Total Tumor Volume and Global Apparent Diffusion Coefficient Using Diffusion-Weighted MRI in Patients With Metastatic Bone Disease: A Feasibility Study. *PLoS One* (2014) 9:e91779. doi: 10.1371/journal.pone.0091779

Author Disclaimer: The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 20 August 2021
doi: 10.3389/fonc.2021.699789

Machine Learning-Based Analysis of Magnetic Resonance Radiomics for the Classification of Gliosarcoma and Glioblastoma

Zenghui Qian[1†], Lingling Zhang[2†], Jie Hu[1], Shuguang Chen[3], Hongyan Chen[2], Huicong Shen[2], Fei Zheng[2], Yuying Zang[2] and Xuzhu Chen[2]*

[1] Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China, [2] Department of Radiology, Beijing Tiantan Hospital, Capital Medical University, Beijing, China, [3] School of Mathematical Sciences, Nankai University, Tianjin, China

OPEN ACCESS

Edited by:
Natalie Julie Serkova,
University of Colorado, United States

Reviewed by:
Amrita Guha,
Tata Memorial Hospital, India
Chirag Kamal Ahuja,
Post Graduate Institute of Medical Education and Research (PGIMER), India

***Correspondence:**
Xuzhu Chen
radiology888@aliyun.com

[†]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: 24 April 2021
Accepted: 04 August 2021
Published: 20 August 2021

Citation:
Qian Z, Zhang L, Hu J, Chen S, Chen H, Shen H, Zheng F, Zang Y and Chen X (2021) Machine Learning-Based Analysis of Magnetic Resonance Radiomics for the Classification of Gliosarcoma and Glioblastoma. Front. Oncol. 11:699789. doi: 10.3389/fonc.2021.699789

Objective: To identify optimal machine-learning methods for the radiomics-based differentiation of gliosarcoma (GSM) from glioblastoma (GBM).

Materials and Methods: This retrospective study analyzed cerebral magnetic resonance imaging (MRI) data of 83 patients with pathologically diagnosed GSM (58 men, 25 women; mean age, 50.5 ± 12.9 years; range, 16-77 years) and 100 patients with GBM (58 men, 42 women; mean age, 53.4 ± 14.1 years; range, 12-77 years) and divided them into a training and validation set randomly. Radiomics features were extracted from the tumor mass and peritumoral edema. Three feature selection and classification methods were evaluated in terms of their performance in distinguishing GSM and GBM: the least absolute shrinkage and selection operator (LASSO), Relief, and Random Forest (RF); and adaboost classifier (Ada), support vector machine (SVM), and RF; respectively. The area under the receiver operating characteristic curve (AUC) and accuracy (ACC) of each method were analyzed.

Results: Based on tumor mass features, the selection method LASSO + classifier SVM was found to feature the highest AUC (0.85) and ACC (0.77) in the validation set, followed by Relief + RF (AUC = 0.84, ACC = 0.72) and LASSO + RF (AUC = 0.82, ACC = 0.75). Based on peritumoral edema features, Relief + SVM was found to have the highest AUC (0.78) and ACC (0.73) in the validation set. Regardless of the method, tumor mass features significantly outperformed peritumoral edema features in the differentiation of GSM from GBM ($P < 0.05$). Furthermore, the sensitivity, specificity, and accuracy of the best radiomics model were superior to those obtained by the neuroradiologists.

Conclusion: Our radiomics study identified the selection method LASSO combined with the classifier SVM as the optimal method for differentiating GSM from GBM based on tumor mass features.

Keywords: gliosarcoma, glioblastoma, machine learning, radiomics, differentiation

INTRODUCTION

Gliosarcoma (GSM), a variant of glioblastoma (GBM), differs from GBM in many respects (1). GSM is associated with lower ratios of epidermal growth factor receptor (EGFR) and O^6-methylguanine-DNA methyltransferase (MGMT) promoter methylation without isocitrate dehydrogenase (IDH) mutations as well as the expression of the v-raf murine sarcoma viral oncogene homolog B1(BRAF) gene at codon 600 (BRAF V600E) (2–6). Clinically, GSM is associated with a higher ratio of extracranial metastasis (7, 8) and a poorer prognosis (3, 9–11). These molecular, genetic, and clinical differences between GSM and GBM indicate that the former may be treated as a unique entity.

While the similarity in the clinical presentation of the two types of tumors underscores the importance of their radiological differentiation, most of the radiological signs of the two tumors overlap (2, 4). Prior imaging research has therefore sought to find a method by which to reliably distinguish the two types of tumors: peritumoral edema seen on routine magnetic resonance imaging (MRI) is more severe in patients with GSM (1, 2), and other imaging modalities, including diffusion weighted imaging (DWI), perfusion weighted imaging (PWI), and magnetic resonance spectroscopy (MRS), have also proven to be helpful in the identification of the tumors (7, 12). However, these imaging methods have not been substantive enough to guide clinical practice due to some limitations. First, qualitative radiological features are susceptible to intra and interobserver variability and lacking reproducibility among evaluators. Second, these radiological modalities only focus on the tumor masses of GSM and GBM when peritumoral edema also requires attention.

Radiomics, a new method for imaging data analysis, has been successfully used for the differentiation of central nervous system tumors: e.g., differentiation between primary central nervous system lymphoma and atypical GBM (13), between GBM and metastasis (14–16), and between GBM and anaplastic oligodendroglioma (17). Like any high-throughput data-mining field, the curse of dimensionality presents a challenge for radiomics analysis. Feature selection is the process of removing irrelevant features that are most conducive to reducing the difficulty of learning task and minimizing the risk of overfitting. This study extracted a large panel of radiomic features from the tumor masses and peritumoral edema of GSM and GBM to inform an optimal machine learning-based algorithm for differentiating GSM from GBM.

MATERIALS AND METHODS

Patient Enrollment

The ethics committee of our hospital approved this retrospective study. This study enrolled 83 patients with GSM (58 men, 25 women; mean age, 50.5 ± 12.9 years; range, 16-77 years) between July 2009 and August 2018 and 100 consecutive patients with GBM (58 men; 42 women; mean age, 53.4 ± 14.1 years; range, 12-77 years) between December 2016 and February 2017. The inclusion criteria for this study were as follows: (I) pathologically confirmed GBM or GSM, as defined by the World Health Organization (WHO) criteria; (II) available preoperative multi-parametric MRI data, including T2-weighted imaging (T2WI) and contrast enhanced (CE) data; (III) patients with no history of preoperative treatment for the tumor before receiving MR; and (IV) available clinical data. Patients were excluded if (I) preoperative MR images were not available in our institute; (II) the images were inadequate for image analysis (for example, they featured obvious artifacts); (III) the lesion showed no enhancement on post-contrast images; or (IV) the lesion was recurrent or had received previous treatment. The clinical and imaging characteristics of all patients were retrospectively assessed, including age, gender, tumor location, and the identification of intra-tumoral necrosis and cystic changes and peritumoral edema. The flowchart of 83 patients with GSM and 100 patients with GBM is presented as **Supplementary Figure 1**. The patients were randomly assigned to either the training (n = 93) or validation groups (n = 90).

MRI Data Acquisition and Region of Interest Segmentation

MRI data included pre- and post-contrast scanning. The detailed scanning parameters are shown in **Supplementary Table 1**. The presence of intra-tumoral necrosis and cystic changes and peritumoral edema were determined for each case. The intra-tumoral necrosis and cystic changes were defined as low signal intensity without enhancement on post-contrast images and high signal on T2WI. The peritumoral edema was defined as low signal intensity around enhanced tumors and high signal on T2WI. The identification of intra-tumoral necrosis, cystic changes, and peritumoral edema were performed by two of the co-authors; conflicting opinions were resolved with discussion.

Several postprocessing steps following the acquisition of MR images were performed to reduce data heterogeneity bias. The adjustment of image resolution was first conducted to resample all voxel size to $3.00 \times 3.00 \times 3.00$ mm^3 without gaps between consecutive slices for each MRI image. Image intensity normalization transformed MR imaging intensity into standardized ranges (0–1). The contour of the tumor on axial images in the CE sequence and the high signal around the tumor in the T2 sequence (the tumor itself and peritumoral edema) were manually segmented into region of interest (ROI) on multiple slices with the opensource software MRIcro (http://www.mccauslandcenter.sc.edu/mricro/). The ROI of the peritumoral edema on CE images was generated by the voxel-wise subtraction of the contrast enhancement in CE sequence from high signals on T2WI using FSL (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL).

Radiomic Feature Extraction and Stability Evaluation

PyRadiomics (http://readthedocs.org/projects/pyradiomics/) computed a total of nine feature categories, including first-order statistics, shape descriptors, texture classes (gray level

co-occurrence matrix, GLCM), gray level run length matrix (GLRLM), and gray level size zone matrix (GLSZM), and six built-in filters (wavelet, Laplacian of Gaussian (LoG), square, square root, logarithm, and exponential), resulting in a total of 1,303 radiomic features (13 shape features, 18 first-order intensity statistics features, 68 texture features, 86 square features, 86 square root features, 86 logarithm features, 86 exponential features, 172 LoG features, and 688 wavelet features). First-order features are intensity-based statistical features describing the distribution of voxel intensities. Shape features describe the size and shape of the ROIs. GLCM, GLRLM and GLSZM features are all texture-related features defined by different computations based on the gray level of the image. All of the features were defined in compliance with the Imaging Biomarker Standardization Initiative (IBSI). All the radiomics features were listed in the **Supplementary Table 2**.

Feature Selection and Classification

A total of three feature selection methods based on statistical approaches were applied in this study: least absolute shrinkage and selection operator (LASSO), Relief and Random Forest (RF). While LASSO and RF are embedded methods, Relief is a filter method. The embedded methods (LASSO and RF) and filter method (Relief) are commonly and effectively used feature selection methods. From the performance of the final model, the wrapped feature selection is better than the filtered feature selection, but the model needs to be trained multiple times, so the computational cost is relatively large. We chose these methods mainly because of their efficiency and popularity among previous studies. In the LASSO algorithm, the shrinkage parameter lambda was identified when the misclassification error was smallest in 10-fold cross-validation. The LASSO, Relief, and RF curve analysis were conducted based on the "glmnet", "vsurf", and "CORElearn" packages by R software (version 3.4.0, R Foundation for Statistical Computing), respectively. Then, three machine-learning classifiers were then applied for feature classification: adaboost classifier (Ada), support vector machine (SVM), and RF. These classifiers are widely used pattern recognition tools and imported from the Python (version 3.6.4) machine learning library named scikit-learn (version 19.0).

Differentiation Performance of the Radiomics Models

The three subsets of selected features were then used as an input to each of the three machine-learning classifiers, which generated nine (3×3 = 9) radiomics models. We applied 5-fold cross-validation as the criteria for each of the nine radiomics models in the training cohort. The differentiation performance was evaluated in the validation cohort. The area under the curve (AUC) and accuracy (ACC) from the receiver operating characteristic curve analysis were calculated to evaluate the differentiation performances of the radiomics models. The optimal thresholds of the AUCs were determined by maximizing the sum of the sensitivity and specificity values calculated for the differentiation of GBM from GSM.

To compare the differentiation performances of the radiomics models and neuroradiologists in differentiating GBM from GSM, we employed the two aforementioned neuroradiologists, who were blinded to the clinical and pathological data, to manually differentiate the GBM from GSM according to all of the sequences (T1WI, T2WI, and CET1WI) showing on the Picture Archiving and Communication Systems (PACS), just as the daily radiological diagnosis workflow before ROI segmentation. They were allowed to see the full MRI images used in this study for the first time. The results of inter-observer variation and concordance with final histopathology statistics between the two neuroradiologists are shown in **Supplementary Table 3**. The chi-square test was performed to compare the proportion of predicted GBM/GSM between the neuroradiologists and the best radiomics model. The entire analysis process is shown in **Figure 1**.

Statistical Analysis

Differences in the clinical and MRI characteristics between GBM and GSM were evaluated using the *t*-test and chi-square test. *P*-values of less than 0.05 were considered to indicate statistical significance. The statistical analysis and figure plots were performed using R (version 3.0.1; http://www.R-project.org) and SPSS (SPSS Inc.).

RESULTS

Clinical and Routine MRI Characteristics

GBM and GSM showed no difference in patient age and gender (P=0.151; χ^2 = 2.758, P=0.097). The ratio of intra-tumoral necrosis and cystic changes was 98.8% (82/83) and 95.0% (95/100) among patients with GSM and GBM, respectively. This difference was non-significant (**Table 1**). The prevalence of peritumoral edema was 94.0% (78/83) and 83.0% (83/100) among patients with GSM and GBM, respectively. The difference was significant (χ^2 = 5.166, *P*=0.023).

Selection of Stable Features

We calculated intraclass correlation coefficient (ICC) to select for the robustness of radiomic features in tumor mass and peritumoral edema. For the tumor mass, 918 of the 1,303 (70.5%) extracted radiomic features showed high stability, including 13 shape features, 18 first-order intensity statistics features, 70 texture features, 84 square features, 81 square root features, 80 logarithm features, 89 exponential features, 179 LoG features, and 304 wavelet features. For the peritumoral edema, 815 of the 1,303 (62.5%) extracted radiomic features showed high stability, including 13 shape features, 18 first-order intensity statistics features, 64 texture features, 70 square features, 89 square root features, 65 logarithm features, 80 exponential features, 162 LoG features, and 254 wavelet features.

Unsupervised clustering of these stable features was conducted and presented as a heat map to yield two imaging subtypes (**Figure 2**). However, the association between the imaging and histology subtypes was not obvious.

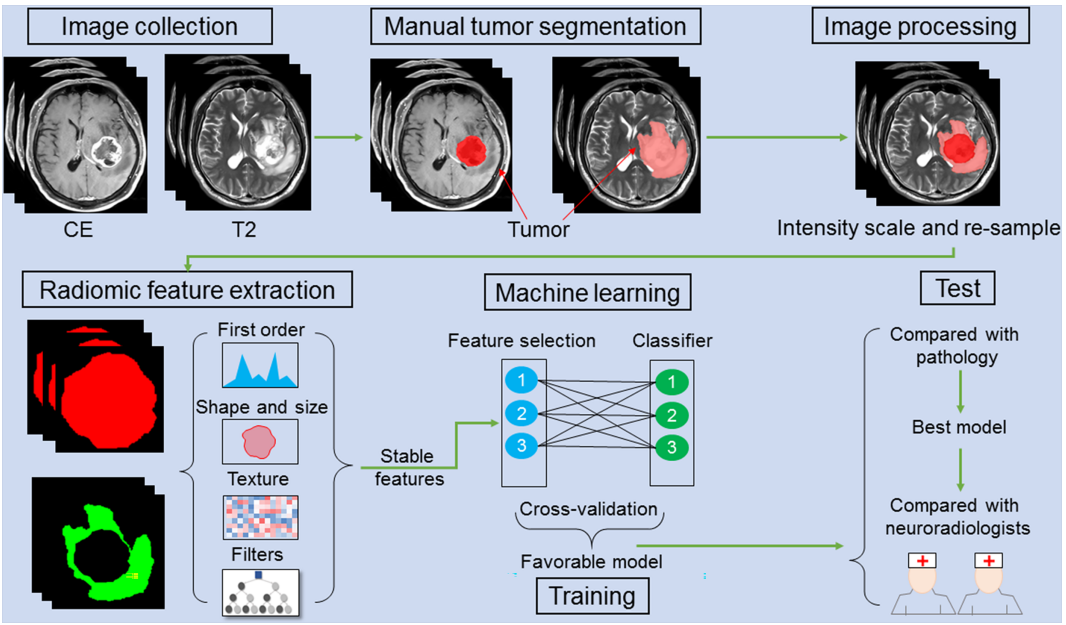


FIGURE 1 | A schematic figure shows the radiomic analysis process. After feature extraction, stable features are selected. Three feature selection and classification methods are combined with favorable models selected and cross-validated in the training cohort. In an independent validation cohort, the optimal model is identified by comparing with pathology. The performance of the optimal model is compared with that of the two neuroradiologists.

TABLE 1 | Clinical and MRI characteristics of patients with GSM and GBM.

	Training cohort			Validaion cohort		
	GSM (n=43)	GBM (n=50)	*P* value	GSM (n=40)	GBM (n=50)	*P* value
Age (years)	51.1	51.6	0.884†	49.8	55.2	0.044†
Sex						
Female	9	23	0.011*	16	19	0.847*
Male	34	27		24	31	
Localization						
Supratentorial	43	47	0.296*	39	50	0.444*
Infratentorial	0	3		1	0	
Necrosis						
Yes	42	47	0.720*	40	48	0.501*
No	1	3		0	2	
Edema						
Yes	39	42	0.337*	39	41	0.047*
No	4	8		1	9	

**Chi-square test, †Student's t-test. GBM, glioblastoma; GSM, gliosarcoma; MRI, magnetic resonance imaging.*

Feature Selection and Radiomics Model Construction

Based on tumor mass features in the training set, the selection method LASSO + classifier SVM was found to feature the highest AUC (0.96) and ACC (0.85), followed by those of Relief + RF (AUC = 0.94, ACC = 0.81), LASSO + RF (AUC = 0.91, ACC = 0.84), and LASSO + Ada (AUC = 0.91, ACC = 0.81; **Tables 2–4** and **Figures 3**, **4**). A similar result was found using the tumor mass features in the validation set: the selection method LASSO + classifier SVM featured the highest AUC (0.85) and ACC (0.77), followed by those of Relief + RF (AUC = 0.84, ACC = 0.72) and LASSO + RF (AUC = 0.82, ACC = 0.75). In both the training and validation set, regardless of the method, tumor mass features significantly outperformed those of the peritumoral edema in the differentiation of GSM from GBM ($P< 0.05$). The illustration of the 5-fold cross-validated ROC curve of the LASSO + SVM radiomics model in the training cohort and ROC curve of the LASSO + SVM radiomics model in the validation set are shown in **Figure 5**.

To avoid biases and confirm the efficacy of the radiomics model, we compared the performance of the selection method LASSO + classifier SVM in 90 validation cases with that of experienced and inexperienced raters. As shown in **Table 4**, the clinical performance of the LASSO + SVM radiomics model was superior to that of the neuroradiologists in terms of sensitivity, specificity, and accuracy.

DISCUSSION

This retrospective study developed and validated a favorable predictive model with radiomics features extracted from tumor mass and peritumoral edema to distinguish GSM from GBM. Importantly, the trend of the diagnostic performance of this machine-learning radiomics model was similar in the training set, validation set, and cross-validation analysis. In our study, two neuroradiologists independently rendered diagnosis of the

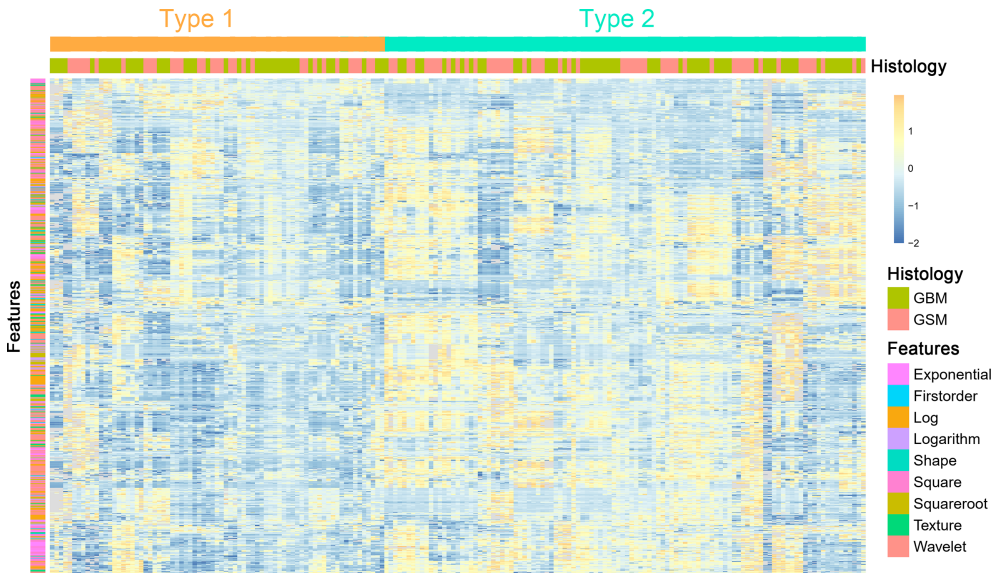


FIGURE 2 | A heat map shows the stable radiomic features. Each column and row correspond to one patient and z-score normalized radiomic feature, respectively.

TABLE 2 | The AUC of the cross-combination methods.

AUC	Ada	RF	SVM
TMF			
LASSO	0.91 (0.81)	0.89 (0.82)	0.96 (0.85)
Relief	0.85 (0.79)	0.91 (0.84)	0.94 (0.81)
RF	0.87 (0.81)	0.84 (0.77)	0.82 (0.79)
PEF			
LASSO	0.84 (0.75)	0.79 (0.71)	0.81 (0.77)
Relief	0.78 (0.76)	0.84 (0.77)	0.84 (0.78)
RF	0.81 (0.69)	0.80 (0.73)	0.76 (0.68)

The AUC of the cross-combination methods based on tumor mass and peritumoral edema features is showed in the training set (no brackets) and the validation set (in brackets). Ada, adaboost; AUC, area under the receiver-operating characteristic curve; LASSO, least absolute shrinkage and selection operator; PEF, peritumoral edema feature; RF, random forest; SVM, support vector machine; TMF, tumor mass feature.

TABLE 3 | The ACC of the cross-combination methods.

ACC	Ada	RF	SVM
TMF			
LASSO	0.83(0.74)	0.81(0.75)	0.87(0.77)
Relief	0.77(0.70)	0.80(0.72)	0.84(0.75)
RF	0.77(0.71)	0.76(0.70)	0.71(0.65)
PEF			
LASSO	0.73(0.68)	0.69(0.63)	0.71(0.67)
Relief	0.72(0.64)	0.75(0.70)	0.79(0.73)
RF	0.74(0.63)	0.71(0.68)	0.71(0.63)

The ACC of the cross-combination methods based on tumor mass and peritumoral edema features are showed in the training set (no brackets) and the validation set (in brackets). ACC, accuracy; ACC, accuracy; Ada, adaboost; LASSO, least absolute shrinkage and selection operator; PEF, peritumoral edema feature; RF, random forest; SVM, support vector machine; TMF, tumor mass feature.

TABLE 4 | Comparison of predictive performance between radiomic model and neuroradiologists in the validation set.

	Sensitivity, *P*	Specificity, *P*	Accuracy, *P*
Neuroradiologist with 3 years of experiences	0.40, <0.001*	0.44, <0.001*	0.42, <0.001*
Neuroradiologist with 10 years of experiences	0.70, 0.015*	0.34, <0.001*	0.50, <0.001*
LASSO_SVM	0.78, —	0.76, —	0.77, —

**Chi-square test. LASSO, least absolute shrinkage and selection operator; SVM, support vector machine.*

two kinds of tumors based on the routine MRI; their accuracy was less than 50.0%, lower than the accuracy of the radiomics analysis, suggesting the superiority of radiomics relative to human analysis in distinguishing GSM from GBM.

In agreement with previous research (18, 19), our study indicated that GSM usually showed enhancement on the solid component with peritumoral edema on routine MRI. These findings, however, are insufficient to inform the distinction of GSM from GBM. Some advanced imaging modalities, such as DWI, PWI, and MRS (7, 12, 20), have therefore been used to better identify the characteristics of GSM. On DWI, the thicker or more solid components of GSM show a restricted diffusion ratio of as high as 72.7% (8/11) (7); on PWI, the tumor featured high perfusion (7); on MRS, GSM shows a lactate peak indicating local necrosis and hypoxia of the tumor and a higher lipid-choline ratio than do GBM (12, 20). These indices obtained from the advanced MR modalities were all derived from analysis of the solid part of the tumor. However, due to the fact that GSM and GBM usually evince necrosis and cystic changes, a comprehensive differentiation between the two tumors should

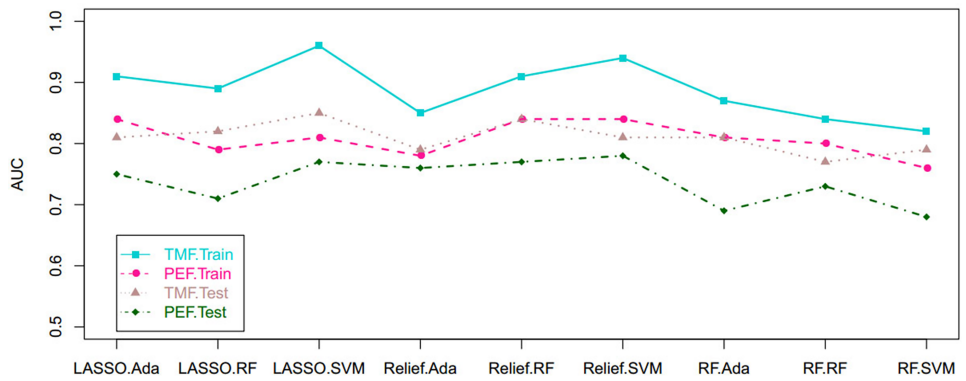


FIGURE 3 | Scatterplots depict the AUC of the cross-combination methods based on the features derived from the tumor and peritumoral edema, respectively. AUC, area under the curve.

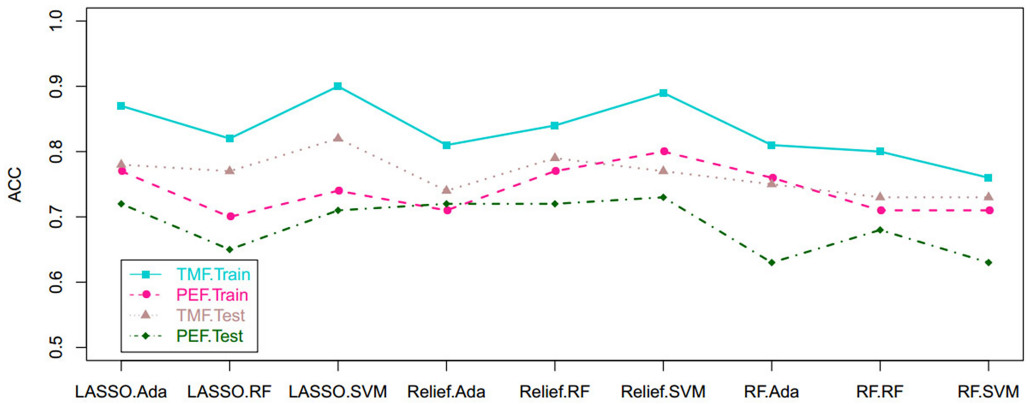


FIGURE 4 | Scatterplots show the ACC of the cross-combination methods based on the features derived from the tumor and peritumoral edema, respectively. ACC, accuracy.

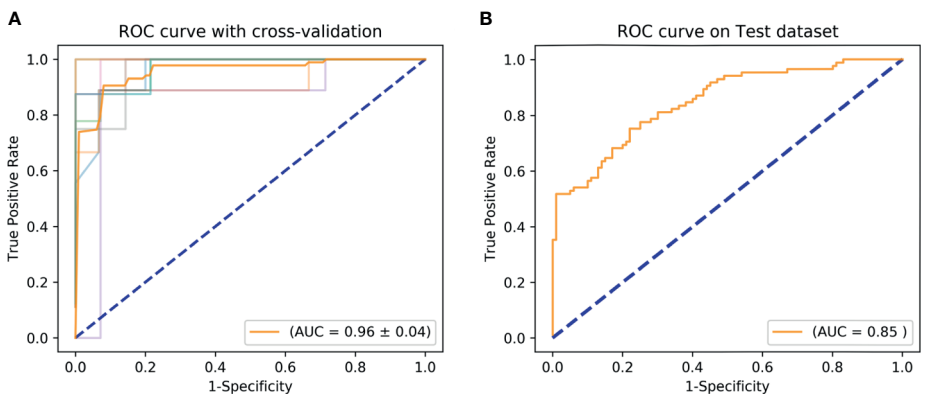


FIGURE 5 | ROC curve shows the optimal classifier for differentiating GSM from GBM. **(A)** The AUC of 5-fold cross-validated ROC is 0.96 in the training set. **(B)** The AUC of 5-fold cross-validated ROC is 0.85 in the validation set. AUC, area under the curve; GBM, glioblastoma; GSM, gliosarcoma; ROC, receiver operating characteristic.

simultaneously involve the solid part and non-solid components. The peritumoral region, which usually shows as edema, is also neglected during differentiation.

In our study, the differentiation between GSM and GBM not only included the whole part of the lesion but also the peritumoral edema outside of the lesion. Our investigation revealed that, based on the peritumoral edema region, the two tumors can be differentiated with the radiomics method of Relief + SVM (AUC, 0.78; ACC, 0.73). Showing as high signal intensity on T2WI, this region included both vasogenic edema and the infiltration of tumor cells (21–23). However, compared with this region, analysis of the tumor mass itself allowed for the more efficient differentiation between tumor types. This can be explained by the fact that there are far more tumor cells in the region of tumor mass than in the peritumoral region. Moreover, the whole region of the tumor mass, including necrosis, cystic changes, and other non-enhanced components, was analyzed for its capacity to inform differentiation. As previous studies that employed PWI, DWI, and MRS (7, 12), only focused on the solid part of the two kinds of tumors, our analysis is more factual and practicable.

Radiomics is an emerging non-invasive method that extracts high-dimensional sets of imaging features to build appropriate models for survival prediction (24), distant metastasis prediction (25), and molecular characteristics classification (26). However, dimensionality is a critical challenge in radiomics analysis and limits the potential of the radiomics model. Hence, this study compared three feature selection methods and classification methods for improving the stability and classification performance of the radiomics model. After performing nine cross-combinations comparisons, we found the LASSO selection method and the classifier SVM to best differentiate of GSM from GBM. The LASSO is a regularization technique used to minimize the number of non-zero elements and make the solution unique (27). It is therefore often used to solve the problem of large sets of radiomics features derived from a relatively small sample size. The SVM is a powerful classification algorithm that can estimate the classification probabilities and control complexity. These properties account for its effective application in the fields of neuroimaging and molecular biology (16, 28) and its superb pairing with the LASSO selection method in our radiomics analysis.

Our study has several limitations. First, it may be subjective to selective bias as a retrospective study. Second, the scanning parameters were not uniform, requiring the preprocessing of the data. Third, compared with the large radiomic features dataset, the sample size was relatively small. Therefore, our results may be caused by overfitting. Fourth, only T2WI and axial post contrast T1WI were used in our radiomic analysis, multi-model imaging data (such as DWI, PWI, MRS) needs to be integrated into our model in the future, to improve its performance. Finally, being a single center study, our study is lack of external independent validation.

In conclusion, this retrospective study presents the machine learning-based MR radiomics model as a non-invasive tool for preoperatively differentiating GSM from GBM with favorable predictive accuracy and stability. Prospective studies are needed to further validate its classification ability.

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 Beijing Tiantan Hospital. Written informed consent for participation was not provided by the participants' legal guardians/next of kin because: As a retrospective study, it was approved by our institute committee without the informed consent of the patients.

AUTHOR CONTRIBUTIONS

ZQ and LZ performed study design, information collection, statistical analysis, and manuscript editing. HC, HS, and XC guided study design, reviewed images, and revised manuscript. JH and SC provided technical support. FZ and YZ collected images and clinical information. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by the National Natural Science Foundation of China under grant numbers 81772005 and 82001897 and Collaborative innovative major special project supported by Beijing Municipal Science & Technology Commission under grant number Z191100006619088.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.699789/full#supplementary-material

REFERENCES

1. Lowder L, Hauenstein J, Woods A, Chen HR, Rupji M, Kowalski J, et al. Gliosarcoma: Distinct Molecular Pathways and Genomic Alterations Identified by DNA Copy Number/SNP Microarray Analysis. *J Neurooncol* (2019) 143:381–92. doi: 10.1007/s11060-019-03184-1
2. Han SJ, Yang I, Tihan T, Prados MD, Parsa AT. Primary Gliosarcoma: Key Clinical and Pathologic Distinctions From Glioblastoma With Implications as a Unique Oncologic Entity. *J Neurooncol* (2010) 96:313–20. doi: 10.1007/s11060-009-9973-6
3. Zhang G, Huang S, Zhang J, Wu Z, Lin S, Wang Y. Clinical Outcome of Gliosarcoma Compared With Glioblastoma Multiforme: A Clinical Study in

Chinese Patients. *J Neurooncol* (2016) 127:355–62. doi: 10.1007/s11060-015-2046-0
4. Peckham ME, Osborn AG, Palmer CA, Tsai A, Salzman KL. Gliosarcoma: Neuroimaging and Immunohistochemical Findings. *J Neuroimaging* (2019) 29:126–32. doi: 10.1111/jon.12565
5. Wang L, Sun J, Li Z, Chen L, Fu Y, Zhao L, et al. Gliosarcomas With the BRAF V600E Mutation: A Report of Two Cases and Review of the Literature. *J Clin Pathol* (2017) 70:1079–83. doi: 10.1136/jclinpath-2017-204620
6. Behling F, Barrantes-Freer A, Skardelly M, Nieser M, Christians A, Stockhammer F, et al. Frequency of BRAF V600E Mutations in 969 Central Nervous System Neoplasms. *Diagn Pathol* (2016) 11:55. doi: 10.1186/s13000-016-0506-2
7. Sampaio L, Linhares P, Fonseca J. Detailed Magnetic Resonance Imaging Features of a Case Series of Primary Gliosarcoma. *Neuroradiol J* (2017) 30:546–53. doi: 10.1177/1971400917715879
8. Ramos R, Morais N, Silva AI, Almeida R. Gliosarcoma With Neuroaxis Metastases. *BMJ Case Rep* (2015) 2015:bcr2015212970. doi: 10.1136/bcr-2015-212970
9. Hashmi FA, Salim A, Shamim MS, Bari ME. Biological Characteristics and Outcomes of Gliosarcoma. *J Pak Med Assoc* (2018) 68:1273–5.
10. Saadeh F, El IS, Najjar M, Assi HI. Prognosis and Management of Gliosarcoma Patients: A Review of Literature. *Clin Neurol Neurosurg* (2019) 182:98–103. doi: 10.1016/j.clineuro.2019.05.008
11. Castelli J, Feuvret L, Haoming QC, Biau J, Jouglar E, Berger A, et al. Prognostic and Therapeutic Factors of Gliosarcoma From a Multi-Institutional Series. *J Neurooncol* (2016) 129:85–92. doi: 10.1007/s11060-016-2142-9
12. Buhl R, Stark AM, Hugo HH, Rohr A, Mehdorn HM. Gliosarcoma: Clinical Experiences and Additional Information With MR Spectroscopy. *Neurol Res* (2009) 31:873–7. doi: 10.1179/174313209X395490
13. Suh HB, Choi YS, Bae S, Ahn SS, Chang JH, Kang SG, et al. Primary Central Nervous System Lymphoma and Atypical Glioblastoma: Differentiation Using Radiomics Approach. *Eur Radiol* (2018) 28:3832–9. doi: 10.1007/s00330-018-5368-4
14. Artzi M, Bressler I, Ben BD. Differentiation Between Glioblastoma, Brain Metastasis and Subtypes Using Radiomics Analysis. *J Magn Reson Imaging* (2019) 50:519–28. doi: 10.1002/jmri.26643
15. 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
16. 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
17. 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
18. Yi X, Cao H, Tang H, Gong G, Hu Z, Liao W, et al. Gliosarcoma: A Clinical and Radiological Analysis of 48 Cases. *Eur Radiol* (2019) 29:429–38. doi: 10.1007/s00330-018-5398-y
19. Han L, Zhang X, Qiu S, Li X, Xiong W, Zhang Y, et al. Magnetic Resonance Imaging of Primary Cerebral Gliosarcoma: A Report of 15 Cases. *Acta Radiol* (2008) 49:1058–67. doi: 10.1080/02841850802314796
20. Raab P, Pilatus U, Hattingen E, Franz K, Hermann E, Zanella FE, et al. Spectroscopic Characterization of Gliosarcomas-Do They Differ From Glioblastomas and Metastases? *J Comput Assist Tomogr* (2016) 40:815–9. doi: 10.1097/RCT.0000000000000419
21. Farshidfar Z, Faeghi F, Mohseni M, Seddighi A, Kharrazi HH, Abdolmohammadi J. Diffusion Tensor Tractography in the Presurgical Assessment of Cerebral Gliomas. *Neuroradiol J* (2014) 27:75–84. doi: 10.15274/NRJ-2014-10008
22. Chen XZ, Yin XM, Ai L, Chen Q, Li SW, Dai JP. Differentiation Between Brain Glioblastoma Multiforme and Solitary Metastasis: Qualitative and Quantitative Analysis Based on Routine MR Imaging. *AJNR Am J Neuroradiol* (2012) 33:1907–12. doi: 10.3174/ajnr.A3106
23. Wang S, Zhou J. Diffusion Tensor Magnetic Resonance Imaging of Rat Glioma Models: A Correlation Study of MR Imaging and Histology. *J Comput Assist Tomogr* (2012) 36:739–44. doi: 10.1097/RCT.0b013e3182685436
24. Qian Z, Li Y, Sun Z, Fan X, Xu K, Wang K, et al. Radiogenomics of Lower-Grade Gliomas: A Radiomic Signature as a Biological Surrogate for Survival Prediction. *Aging (Albany NY)* (2018) 10:2884–99. doi: 10.18632/aging.101594
25. Coroller TP, Grossmann P, Hou Y, Rios VE, Leijenaar RT, Hermann G, et al. CT-Based Radiomic Signature Predicts Distant Metastasis in Lung Adenocarcinoma. *Radiother Oncol* (2015) 114:345–50. doi: 10.1016/j.radonc.2015.02.015
26. Li Y, Qian Z, Xu K, Wang K, Fan X, Li S, et al. MRI Features Predict P53 Status in Lower-Grade Gliomas via a Machine-Learning Approach. *NeuroImage Clin* (2018) 17:306–11. doi: 10.1016/j.nicl.2017.10.030
27. Gui J, Li H. Penalized Cox Regression Analysis in the High-Dimensional and Low-Sample Size Settings, With Applications to Microarray Gene Expression Data. *Bioinformatics* (2005) 21:3001–8. doi: 10.1093/bioinformatics/bti422
28. Han H, Jiang X. Overcome Support Vector Machine Diagnosis Overfitting. *Cancer Inform* (2014) 13:145–58. doi: 10.4137/CIN.S13875

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 02 September 2021
doi: 10.3389/fonc.2021.684996

A Radiomics Model for Predicting Early Recurrence in Grade II Gliomas Based on Preoperative Multiparametric Magnetic Resonance Imaging

Zhen-hua Wang, Xin-Lan Xiao, Zhao-Tao Zhang, Keng He and Feng Hu*

Department of Radiology, The Second Affiliated Hospital of Nanchang University, Nanchang, China

OPEN ACCESS

Edited by:
Antonio Napolitano, Bambino Gesù Children Hospital (IRCCS), Italy

Reviewed by:
S Senthil Kumaran, All India Institute of Medical Sciences, India
Kajari Bhattacharya, Tata Memorial Hospital, India

***Correspondence:**
Xin-Lan Xiao
jx_xiaoxinlan@sina.com

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

Received: *24 March 2021*
Accepted: *12 August 2021*
Published: *02 September 2021*

Citation:
Wang ZH, Xiao XL, Zhang ZT, He K and Hu F (2021) A Radiomics Model for Predicting Early Recurrence in Grade II Gliomas Based on Preoperative Multiparametric Magnetic Resonance Imaging. Front. Oncol. 11:684996. doi: 10.3389/fonc.2021.684996

Objective: This study aimed to develop a radiomics model to predict early recurrence (<1 year) in grade II glioma after the first resection.

Methods: The pathological, clinical, and magnetic resonance imaging (MRI) data of patients diagnosed with grade II glioma who underwent surgery and had a recurrence between 2017 and 2020 in our hospital were retrospectively analyzed. After a rigorous selection, 64 patients were eligible and enrolled in the study. Twenty-two cases had a pathologically confirmed recurrent glioma. The cases were randomly assigned using a ratio of 7:3 to either the training set or validation set. T1-weighted image (T1WI), T2-weighted image (T2WI), and contrast-enhanced T1-weighted image (T1CE) were acquired. The minimum-redundancy-maximum-relevancy (mRMR) method alone or in combination with univariate logistic analysis were used to identify the most optimal predictive feature from the three image sequences. Multivariate logistic regression analysis was then used to develop a predictive model using the screened features. The performance of each model in both training and validation datasets was assessed using a receiver operating characteristic (ROC) curve, calibration curve, and decision curve analysis (DCA).

Results: A total of 396 radiomics features were initially extracted from each image sequence. After running the mRMR and univariate logistic analysis, nine predictive features were identified and used to build the multiparametric radiomics model. The model had a higher AUC when compared with the univariate models in both training and validation data sets with an AUC of 0.966 (95% confidence interval: 0.949–0.99) and 0.930 (95% confidence interval: 0.905–0.973), respectively. The calibration curves indicated a good agreement between the predictable and the actual probability of

developing recurrence. The DCA demonstrated that the predictive value of the model improved when combining the three MRI sequences.

Conclusion: Our multiparametric radiomics model could be used as an efficient and accurate tool for predicting the recurrence of grade II glioma.

Keywords: radiomics, grade II gliomas, MRI, multiparametric, recurrence

INTRODUCTION

Glioma is a brain tumor originating from central glial cells with a high mortality rate (1–3). According to the World Health Organization (WHO), grade I and grade II tumors are classified as low-grade gliomas (LGG). LGGs are generally benign, with a recurrence rate of about 36% (4). Nevertheless, the clinical course of LGG may be unpredictable, as some of these tumors recur soon after primary treatment and/or undergo malignant transformation (5–7). A previous report indicated that low-grade gliomas (WHO II grade) have a 5-year survival rate of as high as 50% (8). Surgical resection followed by chemoradiation is the standard treatment option for gliomas. However, the risk and timing of recurrence following treatment in LGG are still difficult to predict accurately (9–12). Therefore, there is a need to identify accurate indicators for early detection and recurrence to provide timely, optimal treatment and improve survival.

Although histological analysis of surgical specimens is still considered the gold standard to grade gliomas, it may not always provide an accurate result (13) as the small sample obtained during the biopsy may not always reflect the grading heterogeneity within the entire tumor (14, 15). A substantial assessment would require the acquisition of samples from multiple regions within the tumor currently not widely accepted in clinical practice. Furthermore, a biopsy is an invasive procedure and also carries some risk. The acquisition of repeated biopsies is not always considered to be ethical as it may aggravate patient suffering.

The factors leading to poor OS post-surgery in LGG are still not well understood. Previous studies identified age, the extent of the tumor resection, and the expression of specific genes, including Ki-67 and the isocitrate dehydrogenase 1 (IDH1), as indicators for OS (16). Yet, to our knowledge, there is no accurate quantitative tool that could be used to predict at an early stage the risk of recurrence following the first tumor resection, highlighting the need to develop predictive models.

An alternative method that can be used to assess tumor recurrence post-surgery is magnetic resonance imaging (MRI). Previous studies have shown that radiomics could be used to quantitatively extract and assess numerous imaging features to effectively differentiate between high and low-grade gliomas (17, 18) and differentiate tumor recurrence from radiation necrosis (19). When combined with clinical data, imaging features could be used to assess the OS and hence optimize the treatment for the patient. Therefore, this study aimed to create a radiomics model based on clinical and imaging features to predict the risk of developing recurrence in grade II glioma after the first resection.

MATERIALS AND METHODS

Participants

Retrospective analyses were performed on the follow-up medical records of 103 adult patients with histologically confirmed supratentorial grade II gliomas (according to WHO 2016 classification). All patients who had their first extensive glioma resection between May 2017 and November 2019 were included in the study. All patients had a MRI T1-contrast enhanced (T1CE) examination within 72 h after surgery to exclude the presence of a conspicuous residual tumor after surgery and received the same adjuvant chemoradiation treatment using a radiotherapy dose of 50.4 Gy in 28 fractions and 75 mg/m^2 of temozolomide orally (20). Patients below 18 years with poor MRI images and tumor hemorrhage were excluded from the study (**Figure 1**). A total of 64 patients were ultimately included in the study.

Data Collection

After being discharged, the patients were regularly followed up by the neurosurgery group of the hospital. A periodical MRI examination was performed after treatment, and any tumor progression was noted in the patient's medical records according to the neuro-oncology (RANO) criteria (21). A biopsy was performed in those patients who had an obvious tumor progression noted on the MRI to further confirm the findings. The age, sex, progression-free survival (PFS), Ki-67, and IDH1 mutations were obtained from the patients' medical records. Three magnetic resonance imaging (MRI) sequences, including T1-weighted (T1W1), T2-weighted (T2WI), and T1-contrast enhanced (T1CE), were acquired.

MRI Parameters

All the patients underwent multi-sequence imaging protocol on a 3.0 Tesla MRI system (Discovery 750; GE Healthcare, Milwaukee, WI, USA), with an eight-channel head coil (GE Healthcare, Chicago, IL, USA). For the T1-weighted image (T1WI) acquisition, the repetition time/echo time (TR/TE), matrix size, field-of-view (FOV), slice thickness, slice gap, and acquisition time were 1,750/25.4 ms; 512 × 512, 220 × 220, 5 mm, 1.5 mm, and 89 s, respectively. For the T2WI acquisition, the (TR/TE), matrix size, FOV, layer thickness, layer spacing, and the number of layers were 4600/102 ms, 224 × 320, 220 × 220, 6 mm, 1, and 18, respectively.

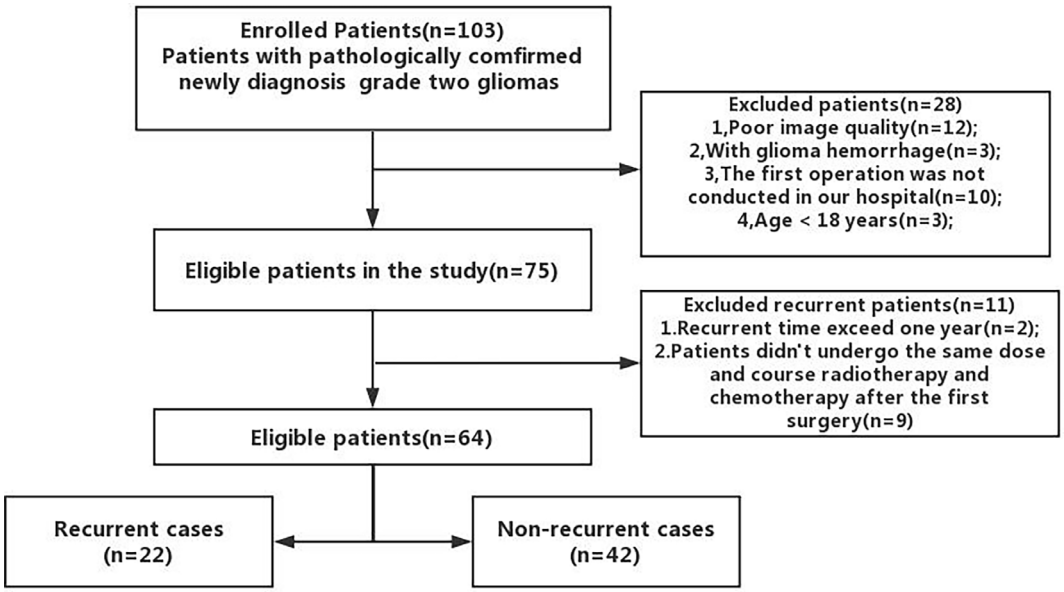


FIGURE 1 | Flow diagram illustrating the patient selection process.

The axial T1CE sequence was acquired by repeating the T1WI described above after a bolus injection of 0.1 mmol/kg of gadodiamide (Omniscan, GE Healthcare, Cork, Ireland).

Description of the Region of Interest and Assessment of the MRI Sequences

The ITK-snap software (www.itk-snap.org) was used to analyze the MRIs. A region of interest (ROI) was blindly delineated by two senior radiologists with more than 10 years of work experience. The boundaries of most low-grade tumors without contrast enhancement were determined on the T2WI images as these images are widely accepted in the identification of hyperintense signals representing the tumor regions (22). Then, the contours of the tumor delineated on the T2WI were transferred to the T1WI and T1CE images. In tumors with contrast enhancement, the tumor boundaries were delineated on the T1CE images by selecting the enhanced region. The delineated region was transferred onto the T1WI and T2WI images.

After the delineation of the ROI, all the patients were divided into the recurrent group (RG) and non-recurrent group (NRG) based on the RANO criteria (indicated in **Table 1**) and biopsy findings by two radiologists. In case of any disagreement, a consensus was reached through discussion, especially when there was a discrepancy between the two readers, as illustrated in **Figure 2**.

Feature Extraction

Radiomic features were extracted using the AK software (Artificial Intelligence Kit V3.0.0.R, GE Healthcare). A total of

TABLE 1 | RANO criteria used to evaluate treatment response in low-grade gliomas.

Criterion	Complete remission	Partial remission	Stable disease	Progress disease
T1CE	Not seen	Decrease ≥50%	Increase or decrease in the range of -25% ~ +25%	Increase ≥ 25%*
T2WI/FLAIR	Stable or diminished	Stable or diminished	Stable or diminished	Increase*
New lesion	None (apart from those consistent with radiation effects, and no new or increased enhancement)	None (apart from those consistent with radiation effects, and no new or increased enhancement)	None (apart from those consistent with radiation effects, and no new or increased enhancement)	Present*
Corticosteroids	None	Stable or diminished	Stable or diminished	Not apply
Clinical status	Stable or improved	Stable or improved	Stable or improved	Deteriorative*(not attributable to other causes apart from the tumor, or decrease in corticosteroid dose)
Requirement for response	All	All	All	Any

CE, contrast-enhanced; FLAIR, fluid-attenuated inversion recovery.
**Progress is determined by anyone project.*

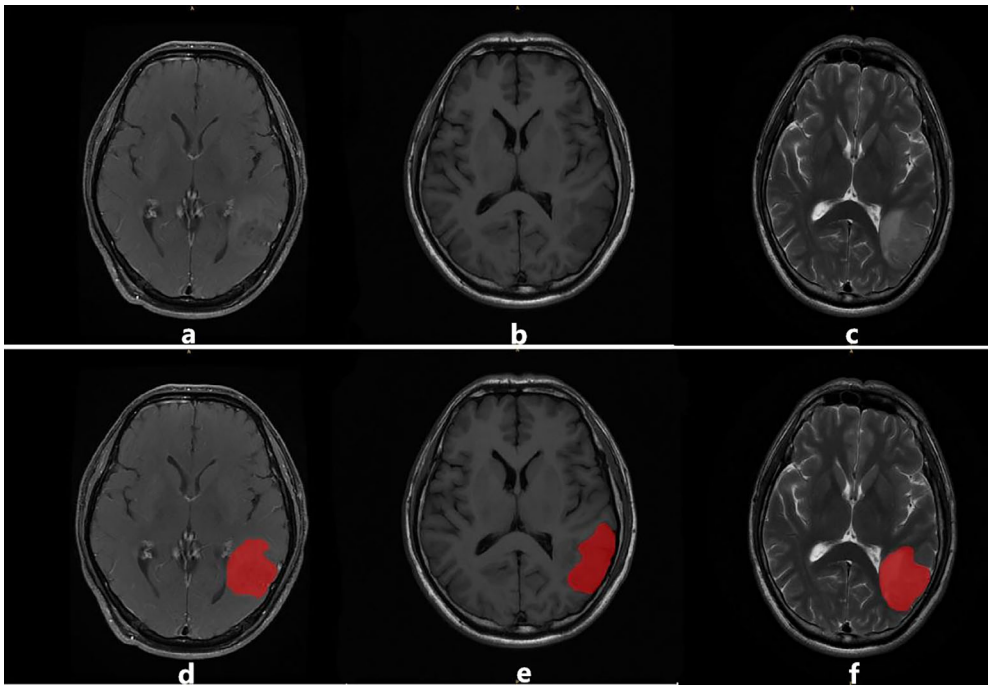


FIGURE 2 | An example of image segmentation: **(A–C)** illustrate T1CE, T1WI, and T2WI sequences, respectively. Images **(D–F)** illustrate the region of interest (ROI) in red delineated by the radiologists for feature extraction.

396 features were extracted from each MRI sequence, including the Laplacian of Gaussian (LoG), rotation invariant local binary patterns (RILBP), the gray level co-occurrence matrix (GLCM), intensity-based features (IBF), directional Gabor texture features (DGTF), and rotation invariant circular Gabor features (RICGF), etc. These features were then used to construct the multiparametric model.

Data Preprocessing and Feature Screening

The dataset was randomly categorized into the training or validation set using a ratio of 7:3. All cases in the training set were used to train the predictive model, while cases in the validation set were used to evaluate the model's performance independently. Variables with zero variance were excluded from the analysis. The missing values were substituted with the median value. Finally, the z-score was used to standardize the data (23). Feature screening was performed by using the minimum redundancy-maximum relevance (mRMR) (24) method alone or in combination with univariate logistic analysis. A *p*-value below 0.05 was deemed statistically significant.

Development and Validation of Models

Logistic regression analysis was used to construct predictive models based on the extracted optimal feature subsets of the training dataset. A receiver operator curve (ROC) was used to assess the performance of the radiomics models, and the sensitivity, specificity, and area under the curve (AUC) were calculated using five-fold cross validation. Calibration curves and decision curve analyses (DCA) were used to assess the clinical predictive performance of the models. The models were constructed using the R software (version 4.0.2), and a two-tailed *p*-value below 0.05 was deemed statistically significant.

Statistical Analysis

According to the normality of samples based on the Shapiro-Wilk test, the independent samples t-test, the chi-square (x^2) test, Fisher's exact test and the Mann-Whitney U-test were used to identify any differences in age, gender, and other baseline characteristics between the training set and validation set. This data was analyzed using the statistical package for the social sciences (SPSS) version 22.0 software.

Ethical Considerations

Ethical approval was obtained from our hospital ethics committee. The need to obtain informed consent from patients was waived due to the retrospective nature of the study.

RESULTS

Patient Characteristics

The characteristics of the tumors and patients are summarized in **Table 2**. A total of 64 patients were included in the analysis. Following the first surgical resection, 64 patients were confirmed as grade II gliomas. According to the RANO criteria, 29 patients

TABLE 2 | Baseline demographics and clinical characteristics of patients in the training and validation datasets.

Clinicopathological Variable		Training set (n=44)			Validation set (n=20)		
		NRG	RG	*p*-value	NRG	RG	*p*-value
Numbers of cases		30	14		12	8	
Age		40.60 ± 12.20	48.36 ± 9.74	0.047	39.77 ± 14.31	51.25 ± 8.12	0.053
Gender, n(%)	Female	15 (50)	6 (42.9)	0.659	4 (30)	4 (50)	0.648
	Male	15 (50)	8 (57.1)		8 (70)	4 (50)	
IDH1-mutation, n(%)	Wild-type	6 (20)	5 (35.7)	0.287	5 (41.7)	3 (37.5)	1.00
	Mutation-type	24 (80)	9 (64.3)		7 (58.3)	5 (62.5)	
Tumor crossing the midline, n(%)	Non	25 (83.3)	10 (71.4)	0.610	9 (75)	6 (75)	1.00
	Yes	5 (16.7)	4 (28.6)		3 (25)	2 (25)	
Ki-67 [median (IQR)]		5.0 (2.0-8.0)	5.5 (3.0-10.0)	0.533	5.0 (2.25-8.0)	7.0 (5.0-14.75)	0.238

NRG, Non-recurrent group; RG, recurrent group; IQR, interquartile range.

were thought to have a tumor recurrence and underwent a biopsy. The biopsy confirmed the recurrence in 22 patients, while the other 7 patients were diagnosed with pseudo-response.

Clinicopathological Characteristics

Among the 64 patients included in the study, 22 had a pathologically confirmed recurrent tumor, and the rest did not have any recurrence. The patients were randomly divided into training and validation datasets using a ratio of 7:3. The baseline characteristics of the subjects are summarized in **Table 2**. There was no significant difference in the age (p = 0.251), gender (p = 0.475), frequency of glioma recurrence (p = 0.845), Ki-67 (p = 0.486), and IDH1 (p = 0.885) mutation status and tumors crossing the midline (p = 0.307) between the training and validation group. There was a statistically significant difference ($p < 0.05$) in age between the RG and NRG in the training set. All other clinicopathological features did not differ significantly between the two groups.

Performance of the Radiomics Models

We extracted 396 features from the ROIs of every sequence. After running the mRMR algorithm, six features were selected from the T1WI images, five features from the T2WI images, and four features from the T1CE images. These three sequences were subsequently combined to identify the most important predictive features of the multiparametric model. Based on the univariate logistic analysis and mRMR, nine predictive features were eventually identified, and their correlation coefficients are illustrated in **Figure 3**. The low correlation coefficient between the nine features indicates little redundancy among every feature cluster.

The features screened from the T1WI, T2WI, T1CE, and multiparametric sequences are summarized in **Table 3**. Four radiomics models were established for predicting tumor recurrence based on the screened optimal predictive features and their contributing predictive weight for each image sequence, as illustrated in **Table 3**. In the T1WI sequence, six predictive features were included in the model, eventually resulting in an AUC of 0.842 and 0.79 in the training and validation datasets, respectively. In the T2WI sequence, five predictive features were used to construct the models, resulting in an AUC of 0.785 in the training set and 0.790 in the validation set. In the T1CE sequence, four predictive features were used to develop the predictive model, which resulted in an AUC of 0.784 in the training set and 0.803 in the validation set. The multiparametric MRI model included nine predictive features from the T1WI, T2WI, and T1CE sequence, resulting in the best overall performance with an AUC of 0.966 and 0.930 for the training and validation datasets, respectively (**Table 4** and **Figure 4**). The calibration curves of the model also indicated a good agreement between the predicted probability and actual tumor recurrence both in the training set and validation set, indicating that the model was well-calibrated (**Figure 5**).

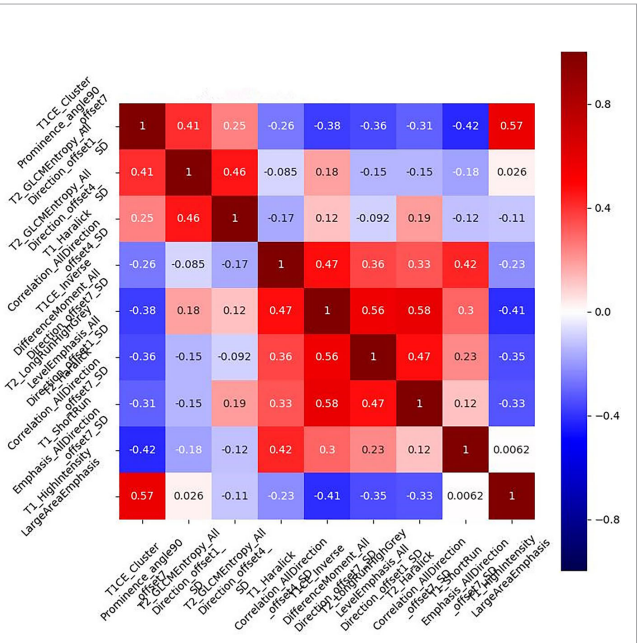


FIGURE 3 | Correlation coefficient of the combined multiparametric models. The correlation coefficients of the selected nine features were low, indicating that the nine feature clusters were not redundant. The magnitude of the correlation is illustrated in the color bar on the right.

The DCA for the individual T1WI, T2WI, T1CE, and these combined multiparametric models are illustrated in **Figure 6**. The net benefit of the model constructed based on the three sequences was higher than the one based on the individual

TABLE 3 | The screened features and their coefficients in the models for the different imaging sequences.

Modality	Variables	Coefficient.	Std.Err.	Z	*P* > |z|	[0.025	0.975]
T1CE	Intercept	-0.3825	0.5451	-0.7017	0.4828	-1.4508	0.6858
	GLCMEntropy_AllDirection_offset4_SD	0.5540	0.5905	0.9382	0.3481	-0.6034	1.7114
	Compactness2	0.6443	0.6101	1.0560	0.2910	-0.5515	1.8401
	ShortRunEmphasis_AllDirection_offset7_SD	-1.1118	0.9869	-1.1265	0.2600	-3.0461	0.8226
	LongRunEmphasis_angle0_offset1	-1.2584	0.7682	-1.6382	0.1014	-2.7640	0.2472
T1WI	intercept	0.0196	0.3750	0.0523	0.9583	-0.7154	0.7546
	GLCMEntropy_AllDirection_offset7_SD	0.9189	1.1705	0.7850	0.4324	-1.3753	3.2132
	LowGreyLevelRunEmphasis_AllDirection_offset1_SD	-3.9480	3.5082	-1.1254	0.2604	-10.8240	2.9279
	RunLengthNonuniformity_AllDirection_offset4_SD	0.3930	0.4920	0.7987	0.4245	-0.5714	1.3573
	ShortRunEmphasis_AllDirection_offset7_SD	-0.3207	0.4130	-0.7765	0.4375	-1.1302	0.4888
	ShortRunEmphasis_angle90_offset4	1.0657	2.3549	0.4525	0.6509	-3.5498	5.6812
	Variance	1.7391	0.7825	2.2225	0.0262	0.2054	3.2728
T2WI	intercept	-0.1973	0.4298	-0.4591	0.6461	-1.0397	0.6451
	ClusterShade_angle45_offset7	0.0706	0.3401	0.2075	0.8356	-0.5960	0.7372
	Correlation_AllDirection_offset1_SD	-0.4181	0.4751	-0.8800	0.3788	-1.3494	0.5131
	Sphericity	-0.6831	0.4032	-1.6941	0.0902	-1.4734	0.1072
	HighIntensityLargeAreaEmphasis	-0.0726	0.4370	-0.1662	0.8680	-0.9291	0.7838
	LongRunEmphasis_angle90_offset1	-1.4411	0.9134	-1.5777	0.1146	-3.2313	0.3491
Multiparametric	intercept	-0.1338	2.3732	-0.0564	0.9550	-4.7852	4.5176
	T1CE_ClusterProminence_angle90_offset7	-2.4287	2.7045	-0.8980	0.3692	-7.7295	2.8721
	T1CE_InverseDifferenceMoment_AllDirection_offset7_SD	2.3638	3.4494	0.6853	0.4932	-4.3970	9.1245
	T2_GLCMEntropy_AllDirection_offset1_SD	2.0994	2.7421	0.7656	0.4439	-3.2750	7.4739
	T2_GLCMEntropy_AllDirection_offset4_SD	-1.2254	0.0745	1.683	0.0924	-0.8635	1.5517
	T2_LongRunHighGrayLevelEmphasis_AllDirection_offset1_SD	0.9696	1.2682	0.7646	0.4445	-1.5161	3.4553
	T2_HaralickCorrelation_AllDirection_offset7_SD	-10.1476	6.6254	-1.5316	0.1256	-23.1332	2.8380
	T1_HaralickCorrelation_AllDirection_offset4_SD	-1.8642	1.6618	-1.1218	0.2620	-5.1213	1.3929
	T1_ShortRunEmphasis_AllDirection_offset7_SD	-1.9502	1.4283	-1.3654	0.1721	-4.7497	0.8493
	T1_HighIntensityLargeAreaEmphasis	2.6325	1.6886	1.5590	0.1190	-0.6770	5.9420

TABLE 4 | The performance of the models for predicting tumor recurrence in the training and validation datasets.

Modality	Features screening method	Number of features after screened	Cohort	AUC (95% CI)	Sensitivity	Specificity	Accuracy
T1WI	mRMR	6	Training	0.842 (0.674–0.905)	0.7	0.7	0.75
			Validation	0.79 (0.687–0.902)	0.778	0.778	0.78
T2WI	mRMR	5	Training	0.785 (0.697–0.912)	0.727	0.682	0.705
			Validation	0.79 (0.679–0.92)	0.8	0.5	0.65
T1CE	mRMR	4	Training	0.784 (0.665–0.913)	0.889	0.556	0.722
			Validation	0.802 (0.693–0.911)	0.78	0.778	0.8
Multi-modalities	ULA + mRMR	9	Training	0.966 (0.949–0.99)	0.905	0.952	0.929
			Validation	0.93 (0.905–0.973)	1	0.8	0.90

ULA, univariate logistic analysis.

imaging sequence, to which it was superior across nearly the entire range of clinically useful threshold risks.

DISCUSSION

Surgery followed by chemoradiation is the main treatment option for patients diagnosed with LGG. Tumor recurrence post-treatment is one of the factors leading to poor OS. Surgical resection is one of the treatment options for patients diagnosed with recurrent LGG. Still, guidelines issued by several professional bodies state that there is limited high-level clinical evidence on the effectiveness of a secondary invasive resection on survival. A study by Patrizz et al. (25) indicated that histopathologic findings following chemoradiation do not always correlate with clinical outcomes in patients diagnosed with recurrence post-surgery. First of all, the pathological specimens may not always reflect the nature of the whole tumor. Furthermore, several studies found that other clinical factors that may have an impact on survival including age, radiotherapy dose, and the extent of tumor resection (16). Multiparametric MRI has played an important role in distinguishing between LGG and HGG as well as recurrence from radiation-induced necrosis. However, to our knowledge, currently, there is no suitable clinical and image-based predictive model to assess the risk of recurrence post-surgery in LGG patients. Therefore in this study, we made use of the imaging data of 64 LGG patients to develop a model that could be used to predict recurrence in these patients and hence enable clinicians to identify the patients that are most likely to benefit from additional surgery.

In our study, there was no difference in the baseline characteristics between the RG and NRG except for age. Consistent with the retrospective study by Li et al. (16), age

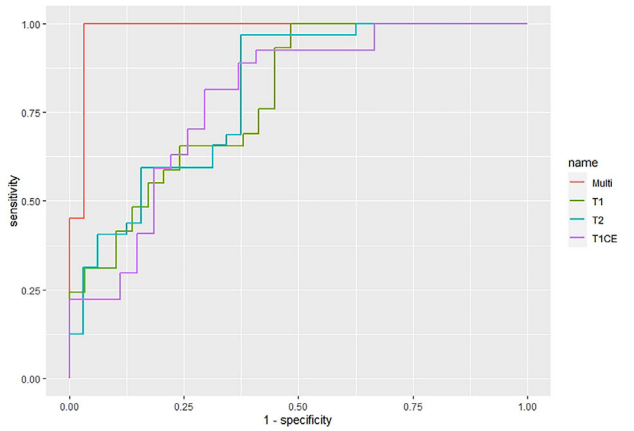


FIGURE 4 | The ROC curves of the four imaging prediction models whereby the green curve represents the T1WI model, the blue curve represents the T2WI, the purple curve represents T1CE, and the red curve represents the multiparametric MRI model.

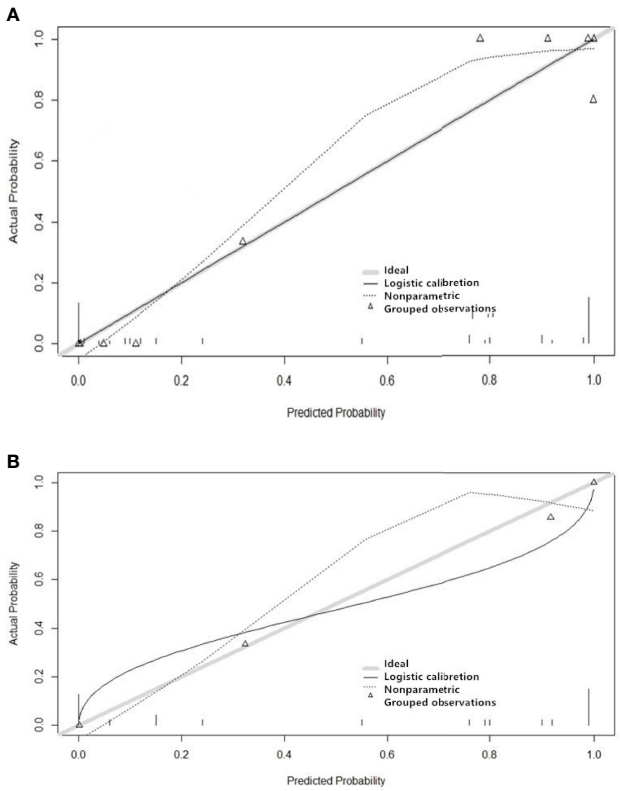


FIGURE 5 | Performance of the four imaging models for predicting the recurrence of grade II gliomas. The y-axis represents the actual probability, and the x-axis represents the predicted probability. Figure **(A)** shows the model's calibration of the training set, and Figure **(B)** shows the validation set. A calibration curve describes the consistency between the predicted and actual tumor recurrence rate. The 45° gray heavy lines represent the ideal prediction performance, the non-45dotted lines represent the prediction performance of the model, and non-45° solid lines represent the corrected prediction performance of the model. The closer the solid line is to the ideal gray line, the better the prediction accuracy of the model.

was found to be an important risk factor for recurrence in grade II gliomas following the first surgery. Jansen et al. (26) conducted a long-term follow-up of 110 patients with LGG (WHO Grade II) after resection. Their results demonstrated that the initial extent of the resection influenced the progression-free survival, time to malignant transformation, and overall survival. Moreover, Patrizz et al. (25) indicated that the radiotherapy dose after surgery has a significant impact on survival in LGG patients. In our study, all patients had an extensive tumor resection and received the same radiation dose. Therefore, the effects of these variables on tumor recurrence could not be assessed.

Studies have shown a high correlation between certain genetic alterations, recurrence, and prognosis in grade II and III gliomas. Mutations of the isocitrate dehydrogenase (IDH)1/2 genes are common events in gliomas (27), especially among grade II gliomas, where IDH1 mutations are observed in about 70% to 80% of cases (27, 28). Some studies indicated that IDH1 mutation status could improve OS and PFS in grade II and III glioma (19, 29). Although the IDH1 mutation has been identified as an independent positive prognostic biomarker for survival in patients with glioma (26, 30), the association between the IDH mutant status and the risk of developing recurrence is still not clear. In the present study, the proportion of IDH mutation cases was noticeably higher in NRG compared with RG [31/42(73.8%) *vs* 14/22(63.6%)]; however, the statistic results showed that there was not a significant difference between NRG and RG (**Table 2**), which indicated that there might not be a link between the IDHI mutation and tumor recurrence; nevertheless, due to the limitation of our relatively small sample size, it still needs a big sample for further verification.

The RANO criteria are still widely used to assess the tumor response post-treatment and the need for additional treatment (31, 32). Despite being used extensively, the accuracy rate of the RANO criteria in distinguishing between tumor recurrence and pseudo-response (32, 33) in our study was only 75.86%. The multi-parameters radiomics model developed in our study resulted in higher prediction accuracy in both testing and validation datasets.

In order to develop our radiomics model, numerous features were extracted from each of the three MRI sequences. It is important to acknowledge that the sample size in our study was relatively small, potentially over-fitting the model (34). In order to reduce this risk, mRMR was used for feature dimensionality reduction. This technique has been widely used in several studies and involves selecting features from the mutually correlated distance or similarity score hence facilitating the data screening process (35, 36).

Numerous studies evaluated the use of radiomics models in predicting recurrence in glioma after radiotherapy. Wang et al. (37) proposed a radiomics model based on MRI and PET images to discriminate between tumor recurrence from radiation necrosis. The model performed well in both training and validation datasets with an AUC of 0.988 and 0.914,

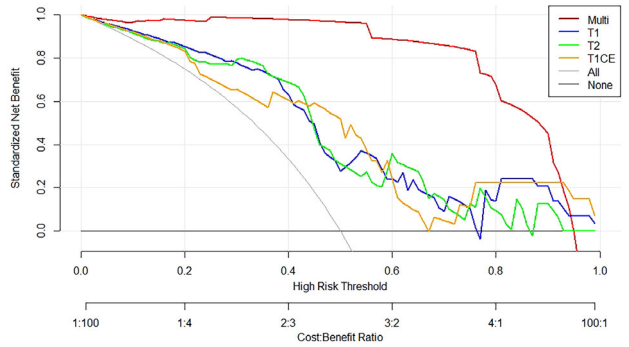


FIGURE 6 | Decision curve analysis constructed using the radiomics features extracted from T1WI, T2WI, and T1CE sequences. The y-axis measures the net benefit of the T1WI (blue curve), T2WI (green curve), T1CE (yellow curve), and the multiparametric (red curve) images. The gray curve represents the assumption that all patients were treated, and the straight black line at the bottom of the figure represents the assumption that none of the patients were treated.

respectively. A similar model based on 51 glioma patients developed by Quan Zhang et al. (19) achieved outstanding performance with an AUC of 0.962 following validation. However, to the best of our knowledge, this is the first multiparametric model developed to predict recurrence in LGG before surgery. Our model also achieved an excellent performance, with an AUC of 0.966 and 0.930 in the testing and validation dataset, respectively.

In the study, a total of nine optimal features were selected for the construction of the multiparametric radiomics model. Among these features, there were three gray level run length matrix (GLRLM) features (T2_LongRunHighGrayLevel Emphasis_AllDirection_offset1_SD, T1_ShortRunEmphasis_ AllDirection_offset7_SD, and T1_ShortRunEmphasis_ AllDirection_offset7_SD), one gray level size zone matrix (GLSZM) feature (T1_HighIntensityLargeAreaEmphasis), and the rest were gray level co-occurrence matrix (GLCM) features (**Table 3**). The above results indicate that GLCM features played the most important role in the model. In some previous radiomics studies, the GLCM features also played an important role in predicting the IDH mutation status. Checkout et al. developed a new approach to predict IDH mutation status that outperformed competing methods (38), while Park et al. (39) found that GLCM was one of the strongest IDH status prediction factors. Furthermore, in a study by Chaddad et al. (40), GLCM had a significant role in predicting survival in patients with glioblastoma. Combined with these previous studies, we can reasonably infer that GLCM may convey information that could potentially be used to predict recurrence.

Both calibration and discrimination are valuable aspects of a prediction model (41). AUC is a common evaluation index of discrimination, while calibration reflects the level of agreement between the actual observed outcomes and the model's predicted outcomes (42). However, the AUC focuses merely on the predictive accuracy of the signature. As such, it does not tell us whether the model is worth using at all. DCA is a statistical method that incorporate consequences and, thus, can inform the decision of whether to use this model (43). Therefore to further complement the AUC findings, a DCA was also performed to evaluate the clinical value of the models (44). In our study, both the AUC and calibration curve (**Figure 5**) showed that our model has a high prediction accuracy. Furthermore, the DCA curves showed that within a relatively large threshold range, our proposed radiomics models could be used to improve the treatment decision-making process. However, the DCA showed that multiparametric MRI models had a significantly higher performance when compared with models based on a single MRI sequence across nearly the entire range of clinically useful threshold risks (**Figure 6**).

This study has some limitations that have to be acknowledged. The majority of the patients with recurrent LGG at our institution generally prefer to be treated with radiotherapy and chemotherapy as opposed to surgery. This limited the sample size in our study and hence limited the number of clinical, pathological, molecular, and imaging features that could be used to train the model. In order to improve the robustness and generalizability of the model, further studies with a larger sample from multiple institutions with a longer follow-up are warranted. A larger sample will also allow us to apply different machine learning strategies to improve the prediction performance of the model. Further research is also recommended to illustrate the relationship between specific imaging features and pathology. Finally, additional studies are also recommended to evaluate the impact of early recurrence prediction on the provision of timely interventions and ultimately survival.

CONCLUSION

The application of our radiomics model-based features extracted from multiparametric MRI could be used to predict the risk of early recurrence of grade II gliomas after the first surgical resection. This model could be used to guide the clinicians' decision on the need for further invasive treatment such as biopsy and surgery in LGG patients.

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 Second Affiliated Hospital of Nanchang University Medical Research Ethical Committee. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Conceptualization: Z-hW, X-LX, and KH. Data curation: Z-hW. Formal analysis: Z-hW, X-LX, and KH. Funding acquisition: Z-hW and X-LX. Investigation: Z-hW, X-LX, and FH. Methodology: Z-hW, X-LX, and Z-TZ. Project administration: Z-hW, X-LX, and KH. Resources: Z-hW, X-LX, and KH. Software: Z-hW, X-LX, and FH. Supervision: Z-hW, X-LX, and KH. Validation: Z-hW, X-LX, and Z-TZ. Visualization: Z-hW, X-LX, and Z-TZ. Writing—original draft: Z-hW and X-LX. Writing—review and editing: Z-hW and X-LX.

FUNDING

This work was supported and funded by the Key research and development projects in Jiangxi Province, China (NO.20171ACG70002).

ACKNOWLEDGMENTS

We acknowledge the Department of Radiology of The Second Affiliated Hospital of Nanchang University.

REFERENCES

1. Ganau L, Paris M, Ligarotti G, Ganau M. Management of Gliomas: Overview of the Latest Technological Advancements and Related Behavioral Drawbacks. *Behav Neurol* (2015) 2015:862634. doi: 10.1155/2015/862634
2. Tully PA, Gogos AJ, Love C, Liew D, Drummond KJ, Morokoff AP. Reoperation for Recurrent Glioblastoma and Its Association With Survival Benefit. *Neurosurgery* (2016) 79:678–89. doi: 10.1227/NEU.0000000000001338
3. Ostrom Q, Gittleman H, Fulop J, Liu M, Blanda R, Kromer C, et al. CBTRUS Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2008-2012. *Neuro Oncol* (2015) 17(Suppl 4):iv1–iv62. doi: 10.1093/neuonc/nov189
4. Fukuya Y, Ikuta S, Maruyama T, Nitta M, Saito T, Tsuzuki S, et al. Tumor Recurrence Patterns After Surgical Resection of Intracranial Low-Grade Gliomas. *J Neurooncol* (2019) 144:519–28. doi: 10.1007/s11060-019-03250-8
5. Daniels TB, Brown PD, Felten SJ, Wu W, Buckner JC, Arusell RM, et al. Validation of EORTC Prognostic Factors for Adults With Low-Grade Glioma: A Report Using Intergroup 86-72-51. *Int J Radiat Oncol Biol Phys* (2011) 81:218–24. doi: 10.1016/j.ijrobp.2010.05.003
6. Okita Y, Narita Y, Miyakita Y, Ohno M, Matsushita Y, Fukushima S, et al. IDH1/2 Mutation Is a Prognostic Marker for Survival and Predicts Response to Chemotherapy for Grade II Gliomas Concomitantly Treated With Radiation Therapy. *Int J Oncol* (2012) 41:1325–36. doi: 10.3892/ijo.2012.1564
7. Komori T, Muragaki Y, Chernov MF. Pathology and Genetics of Gliomas. *Prog Neurol Surg* (2018) 31:1–37. doi: 10.1159/000466835
8. Goodenberger ML, Jenkins RB. Genetics of Adult Glioma. *Cancer Genet* (2012) 205(12):613–21. doi: 10.1016/j.cancergen.2012.10.009
9. Westphal M, Giese A. Local Control of Gliomas: The Next Best Step—A Good Step? *Front Radiat Ther Oncol* (1999) 33:214–26. doi: 10.1159/000061237
10. De Fazio S, Russo E, Ammendola M, Donato Di Paola E, De Sarro G. Efficacy and Safety of Bevacizumab in Glioblastomas. *Curr Med Chem* (2012) 19:972–81. doi: 10.2174/092986712799320646
11. Tugcu B, Postalci L, Gunaldi O, Tanriverdi O, Akdemir H. Efficacy of Clinical Prognostic Factors on Survival in Patients With Glioblastoma. *Turk Neurosurg* (2010) 20:117–25. doi: 10.5137/1019-5149.JTN.2461-09.4
12. Thompson G, Lawrie TA, Kernohan A, Jenkinson MD. Interval Brain Imaging for Adults With Cerebral Glioma. *Cochrane Database Syst Rev* (2019) 12(12):CD013137. doi: 10.1002/14651858.CD013137.pub2
13. Kessler J, Hohmann T, Güttler A, Petrenko M, Ostheimer C, Hohmann U, et al. Radiosensitization and a Less Aggressive Phenotype of Human Malignant Glioma Cells Expressing Isocitrate Dehydrogenase 1 (IDH1) Mutant Protein: Dissecting the Mechanisms. *Cancers* (2019) 11(6):889. doi: 10.3390/cancers11060889
14. Kani Y, Cecere TE, Lahmers K, LeRoith T, Zimmerman KL, Isom S, et al. Diagnostic Accuracy of Stereotactic Brain Biopsy for Intracranial Neoplasia in Dogs: Comparison of Biopsy, Surgical Resection, and Necropsy Specimens. *J Vet Intern Med* (2019) 33:1384–91. doi: 10.1111/jvim.15500
15. Reardon DA, Wen PY. Glioma in 2014: Unravelling Tumour Heterogeneityimplications for Therapy. *Nat Rev Clin Oncol* (2015) 12:69–70. doi: 10.1038/nrclinonc.2014.223
16. Li J, Niu X, Gan Y, Yang Y, Wang T, Zhang H, et al. Clinical and Pathologic Features and Prognostic Factors for Recurrent Gliomas. *World Neurosurg* (2019) 128:e21–30. doi: 10.1016/j.wneu.2019.02.210
17. Tian Q, Yan LF, Zhang X, Zhang X, Hu YC, Han Y, et al. Radiomics Strategy for Glioma Grading Using Texture Features From Multiparametric MRI. *J Magn Reson Imaging* (2018) 48(6):1518–28. doi: 10.1002/jmri.26010
18. Wang Q, Li Q, Mi R, Ye H, Zhang H, Chen B, et al. Radiomics Nomogram Building From Multiparametric MRI to Predict Grade in Patients With Glioma: A Cohort Study. *J Magn Reson Imaging: JMRI* (2019) 49(3):825–33. doi: 10.1002/jmri.26265
19. Zhang Q, Cao J, Zhang J, Bu J, Yu Y, Tan Y, et al. Differentiation of Recurrence From Radiation Necrosis in Gliomas Based on the Radiomics of Combinational Features and Multimodality MRI Images. *Comput Math Methods Med* (2019) 2019:2893043. doi: 10.1155/2019/2893043
20. Shaw E, Arusell R, Scheithauer B, O'Fallon J, O'Neill B, Dinapoli R, et al. Prospective Randomized Trial of Low-*Versus* High-Dose Radiation Therapy in Adults With Supratentorial Low-Grade Glioma: Initial Report of a North Central Cancer Treatment Group/Radiation Therapy Oncology Group/ Eastern Cooperative Oncology Group Study. *J Clin Oncol* (2002) 20 (9):2267–76. doi: 10.1200/JCO.2002.09.126
21. Yamanaka T, Shirota S, Sawai Y, Murai T, Fujita N, Hosoi H. Osteoporosis as a Risk Factor for the Recurrence of Benign Paroxysmal Positional Vertigo. *Laryngoscope* (2013) 123(11):2813–6. doi: 10.1002/lary.24099
22. Li Y, Liu X, Xu K, Qian Z, Wang K, Fan X, et al. MRI Features can Predict EGFR Expression in Lower Grade Gliomas: A Voxel-Based Radiomic Analysis. *Eur Radiol* (2018) 28:356–62. doi: 10.1007/s00330-017-4964-z
23. Kranzusch R, Aus dem Siepen F, Wiesemann S, Zange L, Jeuthe S, Ferreira da Silva T, et al. Z-Score Mapping for Standardized Analysis and Reporting of Cardiovascular Magnetic Resonance Modified Look-Locker Inversion Recovery (MOLLI) T1 Data: Normal Behavior and Validation in Patients With Amyloidosis. *J Cardiovasc Magn Reson* (2020) 22(1):6. doi: 10.1186/s12968-019-0595-7
24. 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
25. Patrizz A, Dono A, Zhu P, Tandon N, Ballester LY, Esquenazi Y. Tumor Recurrence or Treatment-Related Changes Following Chemoradiation in Patients With Glioblastoma: Does Pathology Predict Outcomes? *J Neuro-Oncol* (2021) 152(1):163–72. doi: 10.1007/s11060-020-03690-7
26. Jansen E, Hamisch C, Ruess D, Heiland DH, Goldbrunner R, Ruge M, et al. Observation After Surgery for Low Grade Glioma:Long-Term Outcome in the Light of the 2016 WHO Classification. *J Neurooncol* (2019) 145(3):501–7. doi: 10.1007/s11060-019-03316-7
27. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W, et al. IDH1 and IDH2 Mutations in Gliomas. *N Engl J Med* (2009) 360:765–73. doi: 10.1056/NEJMoa0808710
28. Balss J, Meyer J, Mueller W, Korshunov A, Hartmann C, von Deimling A. Analysis of the IDH1 Codon 132 Mutation in Brain Tumors. *Acta Neuropathol* (2008) 116:597–602. doi: 10.1007/s00401-008-0455-2
29. Jiang H, Cui Y, Wang J, Lin S. Impact of Epidemiological Characteristics of Supratentorial Gliomas in Adults Brought About by the 2016 World Health Organization Classification of Tumors of the Central Nervous System. *Oncotarget* (2017) 8(12):20354–61. doi: 10.18632/oncotarget.13555
30. Nobusawa S, Watanabe T, Kleihues P, Ohgaki H. IDH1 Mutations as Molecular Signature and Predictive Factor of Secondary Glioblastomas. *Clin Cancer Res* (2009) 15:6002–7. doi: 10.1158/1078-0432.CCR-09-0715

31. Li T, Chen L, Gan M. Quality Control of Imbalanced Mass Spectra From Isotopic Labeling Experiments. *BMC Bioinf* (2019) 20(1):549. doi: 10.1186/s12859-019-3170-1
32. Wick W, Chinot OL, Bendszus M, Mason W, Henriksson R, Saran F, et al. Evaluation of Pseudoprogression Rates and Tumor Progression Patterns in a Phase III Trial of Bevacizumab Plus Radiotherapy/Temozolomide for Newly Diagnosed Glioblastoma. *Neuro Oncol* (2016) 18(10):1434–41. doi: 10.1093/neuonc/now091
33. Ellingson BM, Wen PY, Cloughesy TF. Modified Criteria for Radiographic Response Assessment in Glioblastoma Clinical Trials. *Neurotherapeutics* (2017) 14(2):307–20. doi: 10.1007/s13311-016-0507-6
34. Balagurunathan Y, Kumar V, Gu Y, Kim J, Wang H, Liu Y, et al. Test-Retest Reproducibility Analysis of Lung CT Image Features. *J Digit Imaging* (2014) 27:805–23. doi: 10.1007/s10278-014-9716-x
35. Cai Y, Huang T, Hu L, Shi X, Xie L, Li Y. Prediction of Lysine Ubiquitination With mRMR Feature Selection and Analysis. *Amino Acids* (2012) 42(4):1387–95. doi: 10.1007/s00726-011-0835-0
36. Alshamlan H, Badr G, Alohali Y. mRMR-ABC: A Hybrid Gene Selection Algorithm for Cancer Classification Using Micro-Array Gene Expression Profiling. *BioMed Res Int* (2015) 2015:604910. doi: 10.1155/2015/604910
37. Wang K, Qiao Z, Zhao X, Li X, Wang X, Wu T, et al. Individualized Discrimination of Tumor Recurrence From Radiation Necrosis in Glioma Patients Using an Integrated Radiomics-Based Model. *Eur J Nucl Med Mol Imaging* (2020) 47(6):1400–11. doi: 10.1007/s00259-019-04604-0
38. Chekouo T, Mohammed S, Rao A. A Bayesian 2D Functional Linear Model for Gray-Level Co-Occurrence Matrices in Texture Analysis of Lower Grade Gliomas. *NeuroImage Clin* (2020) 28:102437. doi: 10.1016/j.nicl.2020.102437
39. Park CJ, Choi YS, Park YW, Ahn SS, Kang SG, Chang JH, et al. Diffusion Tensor Imaging Radiomics in Lower-Grade Glioma: Improving Subtyping of Isocitrate Dehydrogenase Mutation Status. *Neuroradiology* (2020) 62(3):319–26. doi: 10.1007/s00234-019-02312-y
40. Chaddad A, Daniel P, Desrosiers C, Toews M, Abdulkarim B. Novel Radiomic Features Based on Joint Intensity Matrices for Predicting Glioblastoma Patient Survival Time. *IEEE J BioMed Health Inform* (2019) 23:795–804. doi: 10.1109/JBHI.2018.2825027
41. Lagendijk J, Steyerberg EW, Daalderop LA, Been JV, Steegers E, Posthumus AG. Validation of a Prognostic Model for Adverse Perinatal Health Outcomes. *Sci Rep* (2020) 10(1):11243. doi: 10.1038/s41598-020-68101-3
42. Gerry S, Bonnici T, Birks J, Kirtley S, Virdee PS, Watkinson PJ, et al. Early Warning Scores for Detecting Deterioration in Adult Hospital Patients: Systematic Review and Critical Appraisal of Methodology. *BMJ (Clin Res Ed.)* (2020) 369:m1501. doi: 10.1136/bmj.m1501
43. Van Calster B, Wynants L, Verbeek J, 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
44. Vickers AJ, Elkin EB. Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Med Decision Making: An Int J Soc Med Decision Making* (2006) 26(6):565–74. doi: 10.1177/0272989X06295361

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 08 September 2021
doi: 10.3389/fphy.2021.684184

Comparison of Complex *k*-Space Data and Magnitude-Only for Training of Deep Learning–Based Artifact Suppression for Real-Time Cine MRI

Hassan Haji-Valizadeh[†‡], *Rui Guo*[‡], *Selcuk Kucukseymen, Yankama Tuyen, Jennifer Rodriguez, Amanda Paskavitz, Patrick Pierce, Beth Goddu, Long H. Ngo and Reza Nezafat* *

Department of Medicine, Cardiovascular Division, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, United States

Propose: The purpose of this study was to compare the performance of deep learning networks trained with complex-valued and magnitude images in suppressing the aliasing artifact for highly accelerated real-time cine MRI.

Methods: Two 3D U-net models (Complex-Valued-Net and Magnitude-Net) were implemented to suppress aliasing artifacts in real-time cine images. ECG-segmented cine images (n = 503) generated from both complex k-space data and magnitude-only DICOM were used to synthetize radial real-time cine MRI. Complex-Valued-Net and Magnitude-Net were trained with fully sampled and synthetized radial real-time cine pairs generated from highly undersampled (12-fold) complex *k*-space and DICOM images, respectively. Real-time cine was prospectively acquired in 29 patients with 12-fold accelerated free-breathing tiny golden-angle radial sequence and reconstructed with both Complex-Valued-Net and Magnitude-Net. Cardiac function, left-ventricular (LV) structure, and subjective image quality [1(non-diagnostic)-5(excellent)] were calculated from Complex-Valued-Net– and Magnitude-Net–reconstructed real-time cine datasets and compared to those of ECG-segmented cine (reference).

Results: Free-breathing real-time cine reconstructed by both networks had high correlation (all $R^2 > 0.7$) and good agreement (all $p > 0.05$) with standard clinical ECG-segmented cine with respect to LV function and structural parameters. Real-time cine reconstructed by Complex-Valued-Net had superior image quality compared to images from Magnitude-Net in terms of myocardial edge sharpness (Complex-Valued-Net = 3.5 ± 0.5; Magnitude-Net = 2.6 ± 0.5), temporal fidelity (Complex-Valued-Net = 3.1 ± 0.4; Magnitude-Net = 2.1 ± 0.4), and artifact suppression (Complex-Valued-Net = 3.1 ± 0.5; Magnitude-Net = 2.0 ± 0.0), which were all inferior to those of ECG-segmented cine (4.1 ± 1.4, 3.9 ± 1.0, and 4.0 ± 1.1).

Conclusion: Compared to Magnitude-Net, Complex-Valued-Net produced improved subjective image quality for reconstructed real-time cine images and did not show any difference in quantitative measures of LV function and structure.

Keywords: real-time cine, artifact suppression, deep learning, complex, magnitude

OPEN ACCESS

Edited by:
Oliver Diaz,
University of Barcelona, Spain

Reviewed by:
Zahra Raisi-Estabragh,
Queen Mary University of London, United Kingdom
Angelica Atehortua,
University of Barcelona, Spain

***Correspondence:**
Reza Nezafat
rnezafat@bidmc.harvard.edu

***Present Address:**
Hassan Haji-Valizadeh,
Canon Medical Research USA Inc., Mayfield Village, OH, United States

These authors have contributed equally to this work

Specialty section:
This article was submitted to Medical Physics and Imaging, a section of the journal Frontiers in Physics

Received: *22 March 2021*
Accepted: *09 August 2021*
Published: *08 September 2021*

Citation:
Haji-Valizadeh H, Guo R, Kucukseymen S, Tuyen Y, Rodriguez J, Paskavitz A, Pierce P, Goddu B, Ngo LH and Nezafat R (2021) Comparison of Complex k-Space Data and Magnitude-Only for Training of Deep Learning–Based Artifact Suppression for Real-Time Cine MRI. Front. Phys. 9:684184. doi: 10.3389/fphy.2021.684184

INTRODUCTION

Cardiovascular MR (CMR) is the clinical gold-standard imaging modality for evaluation of cardiac function and structure. Breath-hold ECG-segmented cine imaging using balanced steady-state free-procession readout (bSSFP) allows for accurate and reproducible measurement of left-ventricular (LV) and right-ventricular (RV) function and volume [1–3]. In this technique, *k*-space is divided into different segments collected over consecutive cardiac cycles within a single breath-hold scan. However, ECG-segmented cine acquisition has limited spatial and temporal resolution, is sensitive to changes in heart rate, and requires repeated breath-holds [4–6]. Alternatively, free-breathing real-time cine has been proposed and pursued using rapid real-time imaging or multiple averaging with or without motion correction [7–12]. Using free-breathing real-time cine is advantageous because it does not require multiple breath-holds and is insensitive to heart rate variations. However, real-time cine has lower temporal and spatial resolution than ECG-segmented cine [10, 11]. Therefore, there is a need to further accelerate data collection for real-time cine MRI.

Over the past three decades, there has been considerable progress in the development of accelerated real-time cine imaging including parallel imaging and compressed sensing [13–18]. Parallel imaging is almost always used in cine imaging for both real-time and ECG-segmented acquisition with robust and highly reliable image quality [13]. However, the acceleration rate of parallel imaging cannot be more than three without compromising image quality [19–21]. Compressed sensing has recently been integrated into applications by vendors enabling higher acceleration rates than parallel imaging; however, reconstruction time is long, and acceleration rates beyond four can result in degradation of image quality [17]. Alternative techniques that exploit spatial–temporal correlation and sparsity of cine data have also been explored [22–26]; however, these approaches can suffer from temporal data filtering, often removing information that is crucial to cardiac cine evaluation. Therefore, despite considerable interest from the image reconstruction community, these techniques are rarely clinically used.

Deep learning–based reconstruction has been recently proposed to enable rapid reconstruction of accelerated cine MRI. Hauptmann et al. [27] showed that a 3D U-net was capable of reconstructing accelerated (acceleration rate = 13) real-time cine MRI. Schlemper et al. [28] showed that a trained cascade network was able to rapidly reconstruct accelerated (acceleration rate = 11) cine MRI. Kustner et al. [29] showed that (3 + 1)-dimensional complex-valued spatio-temporal convolutions and multi-coil data processing (CINENet) could reconstruct accelerated ($9 \leq$ acceleration rate ≤ 15) 3D ECG-segmented cine. El-Rewaidy et al. [30] reconstructed accelerated radial cine MRI (acceleration rate = 14) using a complex-valued network (MD-CNN) designed to process MR data in both k-space and image space. Daming et al. [31] used a complex U-net with a combined mean-squared error and perceptual loss (PCNN) to reconstruct real-time cine MRI (acceleration rate = 15).

While promising, popular deep learning–based reconstructions methods [27–32] for cine MRI rely on supervised learning and, as such, require training with large and diverse patient datasets. However, prospectively acquiring large patient datasets within a clinical setting can be difficult due to long scanning times, respiratory/cardiac motion, or contrast washout. To overcome these limitations, Hauptmann et al. proposed training a deep learning network using synthetic data generated from DICOMs (Digital Imaging and Communications in Medicine) [27]. The use of DICOM imaging is advantageous because it is readily available in large numbers at centers with cardiac MR expertise. While promising, DICOM usage during training is theoretically non-optimal given that DICOM images are magnitude images, which lack phase and multi-coil information; furthermore, vendors often apply different filtering techniques to improve image quality in the DICOM creation process. The effect of using DICOM images for training on the performance of a deep learning model has not yet been rigorously studied.

In this study, we sought to investigate differences in performance between two deep learning–based models trained to suppress artifacts in 12-fold accelerated real-time cine. Paired complex-valued k-space data and DICOM images of ECG-segmented cine (n = 503) were used to synthetize highly undersampled radial real-time cine data. Both artifact suppression models were made using 3D U-net architectures. One model was trained with synthetic radial real-time cine images generated from complex *k*-space data (Complex-Valued-Net), while the other model was trained with synthetic radial real-time cine images generated from DICOM images (Magnitude-Net). The performance of the two models was evaluated against prospectively collected free-breathing real-time cine CMR with radial acquisition.

METHODS

Figure 1 summarizes our study which was designed to compare the performance of deep learning–based networks trained to suppress aliasing artifacts in highly accelerated real-time cine using complex-valued images (derived from k-space data) and magnitude-only images (derived from DICOM images). We prepared a dataset containing both complex *k*-space data and corresponding magnitude images (i.e., DICOM) scanned by breath-holding ECG-segmented cine using a Cartesian trajectory to synthesize radial real-time cine data (**Figure 1A**) [27]. Two 3D U-net models [33], Complex-Valued-Net and Magnitude-Net, were developed to remove aliasing artifacts in complex-valued and magnitude images of highly accelerated radial real-time cine, respectively. Complex-Valued-Net and Magnitude-Net were trained using synthetized radial real-time cine with aliasing artifacts generated from complex-valued *k*-space and magnitude-only images, respectively. "Artifact-free" images used to produce synthetized radial cine were used as the ground truth (**Figure 1B**). Finally, the performance of both networks was compared using prospectively acquired free-

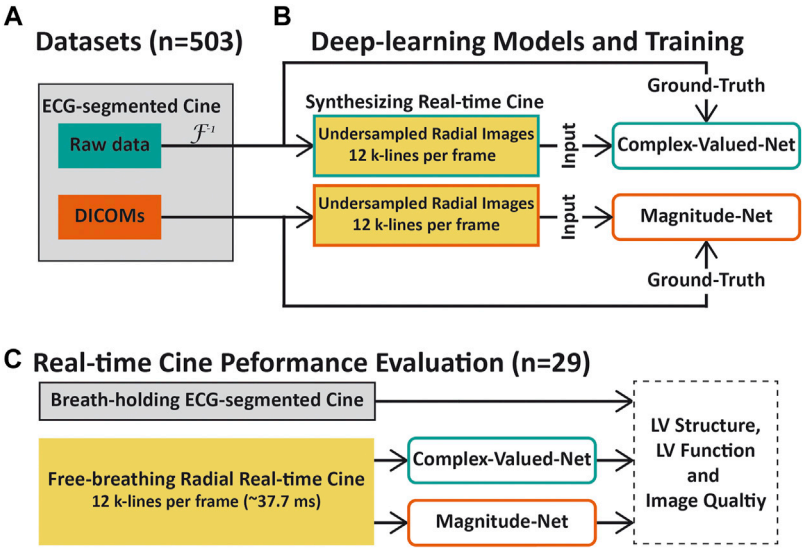


FIGURE 1 | Overview of this study. **(A)** Cine images of 503 patients with both raw *k*-space data and DICOMs were collected. These images were scanned using a breath-holding cine sequence with a Cartesian trajectory. **(B)** Raw *k*-space data and DICOMs of ECG-segmented cine were used to synthesize highly accelerated radial real-time cine datasets for training Complex-Valued-Net and Magnitude-Net, respectively. **(C)** Performance comparison between the two neural networks. Real-time radial cine and corresponding ECG-segmented cine images were collected from 29 patients. The left-ventricular function, structural parameters, and subjective image scores were used to compare the performance of both deep learning models with respect to aliasing artifact suppression. For quantitative and qualitative evaluation, Magnitude-Net reconstruction, Complex-Valued-Net reconstruction, and ECG-segmented cine were compared to one another in pairs.

breathing highly accelerated (12*x*) radial real-time cine in 29 patients. Quantitative functional and structural parameters of the LV and qualitative visual assessments of the LV were compared against reference values derived from ECG-segmented cine images (**Figure 1C**).

Training Datasets

We retrospectively collected short-axis (SAX) cine data from 503 patients (286 males, 55.4 ± 15.8 years) who underwent clinical scans at BIDMC from October 2018 to May 2020. Imaging was performed on a 3T MR scanner (MAGNETOM Vida Siemens Healthineers, Erlangen, Germany) using a breath-hold ECG-segmented sequence with the following parameters: bSSFP readout, FOV = 355 × 370 mm^2, in-plane resolution = 1.7 × 1.4 mm^2, slice thickness = 8 mm, TE/TR = 1.41/3.12 ms, flip angle = 42°, GRAPPA acceleration rate = 2–3, ~18 cardiac phases at a temporal resolution of ~55.3 ms, receiver bandwidth = 1,502 Hz/pixel, Cartesian sampling pattern, and slices per volume = 11 ± 1 (from 9 to 17). Cine's paired raw *k*-space data and DICOM images were used in this study. This study protocol was approved by the institutional review board, and written consent was waived. Patient information was handled in compliance with the Health Insurance Portability and Accountability Act.

Synthesizing Real-Time Cine Training Data

Supplementary Figure S1 shows the data preparation workflow for producing synthetic accelerated radial real-time cine datasets from ECG-segmented cine data acquired using the Cartesian trajectory. The complex-valued multi-coli *k*-space data with an acceleration rate of 2–3 were first reconstructed by GRAPPA [21] offline. Offline GRAPPA reconstruction was implemented with the code made available by Dr. Chiew (https://users.fmrib.ox.ac.uk/~mchiew/Teaching.html).

Then, GRAPPA-reconstructed images and the original DICOM images exported from the scanner were interpolated to achieve 2 × 2 mm^2 in-plane resolution with a temporal resolution of 37.7 ms. We chose these interpolated spatial and temporal resolutions to match the temporal and spatial resolutions used during prospective real-time cine scanning (see below). These GRAPPA-reconstructed or DICOM images were also used as the ground truth in training of two neural networks, respectively. Subsequently, backward non-uniform fast Fourier transform (NUFFT) [34] was applied to GRAPPA-reconstructed and DICOM images to produce complex-valued radial *k*-space. Twelve lines per frame, which were distributed over the whole *k*-space with a tiny golden-angle rotation (32.049°) [35, 36], were chosen to simulate highly accelerated radial *k*-space of real-time cine.

For both Complex-Valued-Net and Magnitude-Net, simulated highly accelerated radial *k*-space data were transformed into image space using forward NUFFT. Specifically, for complex-valued multi-coil *k*-space, the above procedures were performed on a coil-by-coil basis. Finally, a coil-combined image was generated using sensitivity-encoding coil combination [37]. An auto-calibrated sensitivity profile for each coil was produced as previously described [38]. Note that a GPU-based implementation of NUFFT (https://cai2r.net/resources/gpunufft-an-open-source-gpu-library-for-3d-gridding-with-direct-matlab-interface/) was used for synthetic MRI generation.

Deep Learning Models and Training

Supplementary Figure S2 presents an in-depth description of the 3D residual U-net architecture used for Complex-Valued-Net and Magnitude-Net. The U-net architecture of both networks comprised five million kernels and two max-pooling layers/up-convolutional layers. Each convolutional processing layer consisted of 3 × 3 × 3 kernels, batch normalization, and rectified linear activation function (ReLU) [33].

The input/output of each network consisted of paired artifact-free ground truth images and their corresponding undersampled, artifact-contaminated images (size: M × N × T = 144 × 144 × 20). Specifically, for Complex-Valued-Net, we concatenated real and imaginary components of complex-valued input/output pairs to enable real-valued deep learning model processing of complex-valued data (size: 2M × N × T = 288 × 144 × 20) [39]. For Magnitude-Net, a ReLU operator was positioned at the final layer to force the output to be non-negative [27]. L_2 loss function was used to train both networks.

Both networks were implemented using PyTorch (Facebook, Menlo Park, California) and trained on a DGX-1 workstation (NVIDIA Santa Clara, California, United States) equipped with 88 Intel Xeon central processing units (2.20 GHz), eight NVIDIA Tesla V100 graphics processing units (GPUs), and 504 GB RAM. Each GPU has 32 GB memory and 5120 Tensor cores. Each network was trained with 2,900 iterations using an ADAM optimizer and with a 15% drop-out rate. Each iteration randomly chose cine images of 16 LV slices from different patients (batch size). For synthetized real-time cine with ≥20 frames, the starting frame was randomly selected to achieve 20 consecutive frames. For <20 timeframes, the dynamic series was circularly padded to 20. Both input and output images were normalized by the 95th percentile magnitude pixel intensity within the central region (i.e., 48 × 48) across 20 frames. The initial learning rate was 0.001, which decreased by 5% after every 100 iterations. The cost function and optimizer were selected to match parameters proposed by Hauptmann et al. [27] for neural network training using DICOM-derived simulated real-time cine.

Real-Time Cine Performance Evaluation

Twenty-nine patients (16 males, 58 ± 16 years) were prospectively recruited. Free-breathing radial real-time cine research sequences in addition to clinically indicated CMR sequences were collected from each patient. Written informed consent was obtained from each patient prior to CMR imaging. Clinical indications and characteristics of these patients are listed in **Supplementary Table S1**. Breath-hold ECG-segmented cine was performed using the same imaging parameters as those detailed in *Training Datasets*. Free-breathing radial real-time cine was collected with the following parameters: bSSFP readout, FOV = 288 × 288 mm^2, resolution = 2 × 2 mm^2, slice thickness = 8 mm, TE/TR = 1.3/3.2 ms, flip angle = 43°, receiver bandwidth = 1,085 Hz/pixel, radial lines per phase = 12, and temporal resolution = 37.7 ms. The rotating angle of the radial line was 32.049° [36]. Both sequences imaged a stack of 14 SAX slices covering the entire LV. Breath-holding ECG-segmented cine was reconstructed by the scanner. For free-breathing real-time cine, NUFFT first transformed radial *k*-space data into complex-valued and magnitude images. Subsequently, two neural networks were used to remove aliasing artifacts.

Data Analysis

We used both quantitative imaging parameters and qualitative assessments of image quality to compare the performance of both deep learning reconstructions. ECG-segmented cine images collected using the standard clinical protocol were used as a reference. For each patient in our independent validation dataset, one reader (HH), trained by a clinical reader (SK) with 5 years of experience, calculated the following cardiac function and structural parameters: LV ejection fraction (LVEF), LV end-diastolic volume (LVEDV), LV end-systolic volume (LVESV), LV stroke volume (LVSV), and LV mass (LVMass). All quantifications were performed using CVI42 (v5.9.3, Cardiovascular Imaging, Calgary, Canada). Linear regression and Bland–Altman analysis were performed to evaluate correlation and agreement between real-time cine and ECG-segmented cine. A paired Student's t-test was conducted to compare the difference between two approaches in measures of LV function and structural parameters. $p < 0.05$ was considered statistically significant. Three pairwise group comparisons were assessed using the t-test with Bonferroni correction, with p less than 0.0167 considered significant.

Subjective image quality was graded by one reader (SK) with 5 years of CMR experience. Cine images of all patients obtained from the three methods were randomized and de-identified. For each method, whole LV cine images from each subject were scored with respect to conspicuity of endocardial borders (1: non-diagnostic, 2: poor, 3: adequate, 4: good, 5: excellent), temporal fidelity of wall motion (1: non-diagnostic, 2: poor, 3: adequate, 4: good, 5: excellent), and artifact level on the myocardium (1: non-diagnostic, 2: severe, 3: moderate, 4: mild, 5: minimal). **Supplementary Figure S3** shows representative graded images. The z-test was used to compare image quality between every two methods, and a p-value < 0.05 was considered significant. SAS version 9.4 (SAS Institute, Cary, NC, United States) was utilized for all above analyses. Note that we elected not to quantitatively and qualitatively analyze real-time cine reconstructed with gridding because gridding alone did not produce diagnostic image quality.

RESULTS

Figures 2A,B show images obtained from the basal, mid, and apical cavities of one subject at end-systole and -diastole by ECG-segmented cine and free-breathing real-time cine *via* gridding, Complex-Valued-Net, and Magnitude-Net reconstruction. **Supplementary Videos S1–S4** show the corresponding movies for dynamic display. We also show representative end-systolic images for three patients in **Supplementary Figure S4**. In both **Figure 2** and **Supplementary Figure S4**, free-breathing real-time

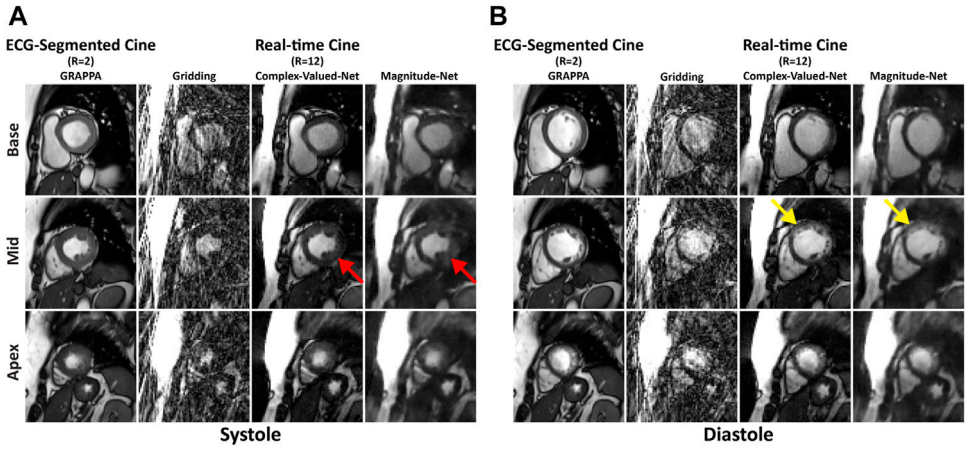


FIGURE 2 | Images at end-systolic **(A)** and end-diastolic **(B)** phases for three short-axis slices (base, mid, apex) in one patient. Magnitude-Net exhibits more image artifact (red arrow) and greater blurring (yellow arrow) at the myocardial wall than Complex-Valued-Net. Gridding reconstruction produces non-diagnostic image quality.

cine reconstructed by Magnitude-Net had more artifacts in the myocardial wall and greater blurring than ECG-segmented cine and real-time cine by Complex-Valued-Net.

Supplementary Table S2 summarizes LV structure and cardiac function values from ECG-segmented cine and free-breathing real-time cine in 29 patients. The mean difference and 95% CI between every two methods are listed in **Supplementary Table S3**. According to Bland–Altman analysis (**Figures 3A–C**), mean differences between ECG-segmented cine and real-time cine by Complex-Valued-Net reconstruction were −0.9 ± 6.5% (p = 0.48) for LVEF, 0.9 ± 13.6 ml (p = 0.73) for LVEDV, and 2.2 ± 12.5 ml (p = 0.34) for LVESV. Correspondingly, mean differences between real-time cine by Magnitude-Net and ECG-segmented cine images were −2.3 ± 5.1% (p = 0.02), −0.5 ± 15.5 ml (p = 0.85), and 3.7 ± 9.8 ml (p = 0.05) for LVEF, LVEDV, and LVESV, respectively (**Figures 3D–F**). **Supplementary Figure S5** compares real-time cine and ECG-segmented cine according to LVSV and LVMass using Bland–Altman analysis. For real-time cine images reconstructed by Complex-Valued-Net, the mean difference was −1.4 ± 16.3 ml (p = 0.65) for LVSV and 2.2 ± 15.2 g (p = 0.43) for LVMass. For Magnitude-Net real-time cine, the mean difference was −4.2 ± 15.4 ml (p = 0.15) and 1.6 ± 18.0 g (p = 0.64) for LVSV and LVMass, respectively. Free-breathing real-time cine reconstructed by both Complex-Valued-Net and Magnitude-Net had high correlation with ECG-gated segmented cine on quantification of LV function and structure (all $R^2 \geq 0.74$ and all slope ≥ 0.88) (**Figures 3G–I** and **Supplementary Figures S5C, F**). The difference between real-time cine images reconstructed by Complex-Valued-Net and Magnitude-Net in quantification of LV function and structure was 1.4 ± 5.1% (p = 0.15) for LVEF, 1.4 ± 8.1 ml (p = 0.36) for LVEDV, −1.4 ± 8.7 ml (p = 0.39) for LVESV, 2.8 ± 12.1 ml (p = 0.22) for LVSV, and 0.7 ± 10.8 g (p = 0.74) for LVMass.

Figure 4 shows the mean/standard deviation and distribution of image quality scores across all patients. **Supplementary Table S4** lists the percentages as two grades (1–3 and 4–5) of image quality scores across all patients by each method. The corresponding differences in the percentage of two grade groups (1–3 and 4–5) among three methods are listed in **Table 1**. The table shows that 79% of ECG-segment cine images had good or excellent scores (>3) for myocardial edge (4.1 ± 1.4) and temporal fidelity (3.9 ± 1.0). In contrast, 50% of real-time cine images reconstructed by both Complex-Valued-Net and Magnitude-Net scored less than or equal to 3 (myocardial edge: 3.5 ± 0.5 vs 2.6 ± 0.5; temporal fidelity: 3.1 ± 0.4 vs 2.1 ± 0.4), suggesting poor image quality. ECG-segment cine had less artifact (4.0 ± 1.1) than real-time cine (Complex-Valued-Net: 3.1 ± 0.5; Magnitude-Net: 2.0 ± 0.0). All z-tests were found to be significant ($p < 0.05$).

DISCUSSION

This study compares the performance of deep learning approaches for reconstruction of highly accelerated real-time cine using synthetized training data generated from complex-valued multi-coil k-space data (Complex-Valued-Net) and real-valued DICOMs (Magnitude-Net). Our subjective assessment of image quality demonstrates that Complex-Valued-Net yields better image quality than Magnitude-Net. However, the clinically relevant parameters of LV function and structure extracted from real-time cine reconstructed by both Complex-Valued-Net and Magnitude-Net were highly correlated and had excellent agreement with those of clinical breath-holding ECG-segmented cine.

There is a growing body of literature in deep learning, beyond CMR, in which magnitude images are used for training a variety of deep learning techniques [27, 40–42]. However, there is also concern regarding the impact that discarded phase information may have on the clinical interpretation of reconstructed images [27, 29, 43–45]. Our study demonstrates that availability of complex k-space data improves overall image quality; however, these improvements in image quality do not necessary impact

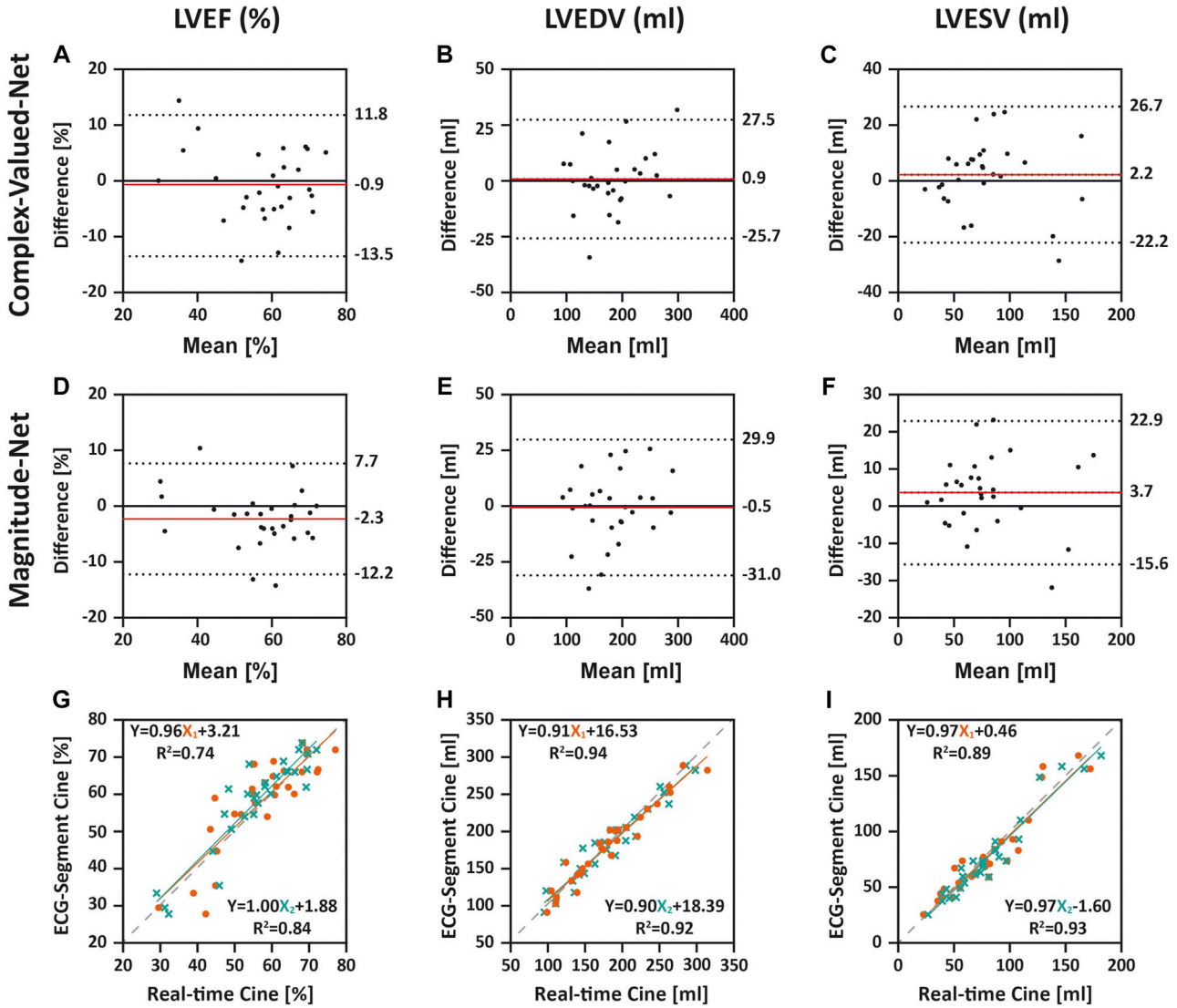


FIGURE 3 | Comparison between ECG-segmented cine and real-time cine for quantifying left-ventricular ejection fraction (LVEF), left-ventricular end-diastolic volume (LVEDV), and left-ventricular end-systolic volume (LVESV) using Bland–Altman analysis **(A–F)** and linear regression **(G–I)**. In Bland–Altman, dotted lines indicate upper and lower 95% limits of agreement and the red line represents the mean difference. The difference was calculated as real-time cine (Complex-Valued-Net and Magnitude-Net) minus ECG-segmented cine. In linear regression, X_1 and X_2 indicate real-time cine reconstructed by Complex-Valued-Net and Magnitude-Net, respectively. The dashed line shows a reference line with a slope of 1. All three quantifications from real-time cine using both Complex-Valued-Net and Magnitude-Net had good agreement and high correlation with quantifications by ECG-segmented cine (all $p > 0.0167$).

clinical interpretation and quantification. This observation is not unique, and it is often debated whether "prettier" images necessarily lead to better diagnostic information. While the resulting data do not show clinically meaningful differences in LV function and structural parameters, an improvement in overall image quality may still be clinically relevant. For example, we often rely on wall motion abnormality to assess the presence of ischemia, which can be visually assessed by reviewing cine images [46]. One can envision that improved image quality may still be clinically relevant and provide additional confidence in image assessments. Further studies in patients with different imaging indications are warranted.

In cine imaging, voxel-values are not meaningful; however, in quantitative CMR imaging (e.g., T_1/T_2 mapping, quantitative perfusion, or phase-contrast), voxel-values represent a tissue-specific meaning [47]. While qualitative imaging such as cine imaging is more forgiving in terms of artifact and inaccuracy during image reconstruction, quantitative CMR imaging is very sensitive to image artifacts. In addition, complex *k*-space data carry crucial information in quantitative imaging and cannot simply be discarded. Therefore, complex *k*-space data will still be needed for quantitative CMR image reconstruction with deep learning, despite our findings showing that

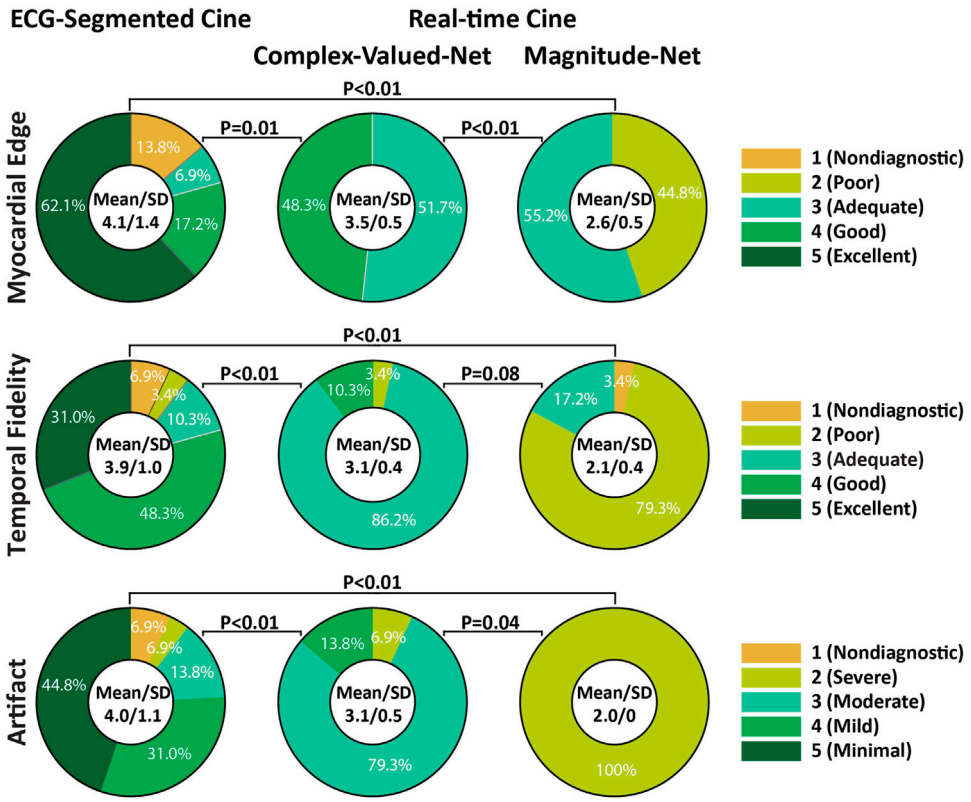


FIGURE 4 | Distribution and average image quality scores across all cine images of 29 patients by three methods: ECG-segmented cine, Complex-Valued-Net real-time cine, and Magnitude-Net real-time cine. The P-values of z-tests between every two methods regarding each criterion are labeled. Real-time cine by Complex-Valued-Net reconstruction yielded superior subjective scores for all three criteria compared to those by Magnitude-Net. ECG-segmented breath-hold cine had the highest score across all three criteria.

TABLE 1 | Differences in percentage of two grades (1–3 and 4–5) of image quality scores between three methods.

	Complex-Valued-Net vs ECG-segmented cine		Magnitude-Net vs ECG-segmented cine		Complex-Valued-Net vs Magnitude-Net	
	Rate difference (95% CI)	P	Rate difference (95% CI)	P	Rate difference (95% CI)	P
Myocardial edge	−0.31 (−0.54, −0.08)	0.01	−0.79 (−0.94, −0.65)	<0.01	0.48 (0.30, 0.66)	<0.01
Temporal fidelity	−0.69 (−0.87, −0.51)	<0.01	−0.79 (−0.94, −0.65)	<0.01	0.10 (−0.01, 0.21)	0.08
Artifact	−0.62 (−0.82, −0.42)	<0.01	−0.76 (−0.91, −0.60)	<0.01	0.14 (0.01, 0.26)	0.04

**For myocardial edge and temporal fidelity, 1: non-diagnostic; 2: poor; 3: adequate; 4: good; 5: excellent. For artifact, 1: non-diagnostic; 2: severe; 3: moderate; 4: mild; 5: minimal.*

magnitude-only images may be sufficient in real-time cine imaging. Further studies are needed to rigorously study other imaging sequences.

For this study, our goal was not necessarily to study or develop a new architecture but was instead motivated by Hauptmann et al. and their important contribution of using readily available DICOMs for network training [27]. Raw complex *k*-space data will still be needed for deep learning models that integrate complex *k*-space data for image reconstruction. However, limited availability of complex *k*-space data will remain a major challenge for training such networks on different applications, diseases, scanner vendors, field strengths, and number of coils. On the contrary, if one can train the model using only DICOM images, there are vast amounts of available data for different organs, sequences, diseases, and vendors that could greatly impact the adoption of deep learning artifact reconstruction techniques.

This study has several limitations. Our training data were not collected using prospectively acquired datasets using radial *k*-space filling, but instead training data were synthesized in a similar manner as proposed by Hauptmann et al. [27]. We used ECG-gated cine images with Cartesian sampling to extract reference values for different LV functional and structural parameters for comparison with real-time radial imaging [27]. There may be differences between the two approaches

due to the *k*-space sampling scheme. Additionally, ECG-segmented data were collected with breath-holding, while real-time data were collected during free-breathing. The evaluation of deep learning reconstruction methodologies was limited to image quality assessment and quantification of left-ventricular functional and structural parameters (i.e., EF, LVEDV, LVESV, LVSV, and LVMass). We chose these metrics because of their clinical importance. That said, further studies are warranted to evaluate the capacity of the presented methods (Magnitude-Net and Complex-Valued-Net) for diagnosis of cardiovascular diseases. Real-time cine reconstructed with gridding was not quantitatively or qualitatively analyzed because gridding alone produced non-diagnostic image quality. Subjective image assessment was performed by a single reader, and there may be differences in image perception by different reviewers. Both Magnitude-Net and Complex-Valued-Net suffer from reduced temporal fidelity compared to ECG-gated segmented cine. Such a loss of temporal fidelity can be especially problematic during systolic phases and may be a source of error during qualitative and quantitative evaluation. All patients in our testing cohort were in sinus rhythm. Only a single neural network architecture (i.e., 3D U-net) was used to compare the performance of magnitude vs complex-valued synthetic training data. We chose this network architecture because, to the best of our knowledge, it is the only architecture shown to be capable of reconstructing radial real-time cine MRI acquired with bSSFP readout [27, 31]. Other state-of-the-art approaches such as cascade networks [28, 29] have yet to be investigated for radial real-time cine reconstruction. Future collaborations are warranted to first extend other state-of-the-art methods to radial real-time cine reconstruction and then compare the performance of different synthetic training data (i.e., magnitude vs. complex-valued) using these methods. ECG-segmented cine images used for training were gathered from one cardiac MR center. As such, trained networks could contain bias which can prevent generalization. Although we used a relatively large number of patients for training, our testing cohort with real-time radial imaging was relatively small, and images were acquired at a single clinical center. Future studies with more patients and imaging from different centers are required to evaluate proposed deep learning methodologies for real-time cine reconstruction.

CONCLUSION

Despite improved subjective image quality in real-time cine images reconstructed using a deep learning model trained with complex *k*-space data when compared to magnitude-only data, there were no differences with respect to quantitative measures of LV function and structural parameters.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

This study was approved by the BIDMC Institutional Review Board (IRB) and was Health Insurance Portability and Accountability Act (HIPPA)-compliant. This study was performed under two IRB approved protocols, including one allowing use of retrospective data collected as part of a clinical exam for machine learning research; informed consent was waived for use of previously collected data. In addition, we prospectively recruited subjects for this study, and written-informed consent was obtained from all prospective participants.

AUTHOR CONTRIBUTIONS

HH-V and RN contributed to study design and validation of deep learning models. HH-V contributed to training of deep learning models. RG, SK, YT, and LN performed data analysis, and RG and RN prepared the manuscript. JR, AP, PP, and BG contributed to data collection. RN contributed to data interpretation. All authors critically revised the paper and read and approved the final manuscript.

FUNDING

RN was funded by the NIH (5R01HL127015-02, 1R01HL129157-01A1, 5R01HL129185, and 1R01HL154744) and the American Heart Association (15EIA22710040). HH-V was supported by the NIH (5T32HL007374-41).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2021.684184/full#supplementary-material

REFERENCES

1. Carr JC, Simonetti O, Bundy J, Li D, Pereles S, and Finn JP. Cine MR Angiography of the Heart with Segmented True Fast Imaging with Steady-State Precession. *Radiology* (2001) 219(3):828–34. doi:10.1148/radiology.219.3.r01jn44828
2. Hundley WG, Bluemke DA, Finn JP, Flamm SD, Fogel MA, Friedrich MG, et al. ACCF/ACR/AHA/NASCI/SCMR 2010 Expert Consensus Document on Cardiovascular Magnetic Resonance. *J Am Coll Cardiol* (2010) 55(23): 2614–62. doi:10.1016/j.jacc.2009.11.011
3. Leiner T, Bogaert J, Friedrich MG, Mohiaddin R, Muthurangu V, Myerson S, et al. SCMR Position Paper (2020) on Clinical Indications for Cardiovascular Magnetic Resonance. *J Cardiovasc Magn Reson* (2020) 22(1):76. doi:10.1186/s12968-020-00682-4
4. Atkinson DJ, and Edelman RR. Cineangiography of the Heart in a Single Breath Hold with a Segmented TurboFLASH Sequence. *Radiology* (1991) 178(2):357–60. doi:10.1148/radiology.178.2.1987592

5. James CC, Orlando S, Jeff B, Debiao L, Scott P, and Finn JP. Cine MR Angiography of the Heart with Segmented True Fast Imaging with Steady-State Precession1. *Radiology* (2001) 219(3):828–34. doi:10.1148/radiology.219.3.r01jn44828
6. Finn JP, Nael K, Deshpande V, Ratib O, and Laub G. Cardiac MR Imaging: State of the Technology. *Radiology* (2006) 241(2):338–54. doi:10.1148/radiol.2412041866
7. Goebel J, Nensa F, Bomas B, Schemuth HP, Maderwald S, Gratz M, et al. Real-Time SPARSE-SENSE Cardiac Cine MR Imaging: Optimization of Image Reconstruction and Sequence Validation. *Eur Radiol* (2016) 26(12):4482–9. doi:10.1007/s00330-016-4301-y
8. Setser RM, Fischer SE, and Lorenz CH. Quantification of Left Ventricular Function with Magnetic Resonance Images Acquired in Real Time. *J Magn Reson Imaging* (2000) 12(3):430–8. doi:10.1002/1522-2586(200009)12:3<430::aid-jmri8>3.0.co;2-v
9. Plein S, Smith WHT, Ridgway JP, Kassner A, Beacock DJ, Bloomer TN, et al. Qualitative and Quantitative Analysis of Regional Left Ventricular wall Dynamics Using Real-Time Magnetic Resonance Imaging: Comparison with Conventional Breath-Hold Gradient echo Acquisition in Volunteers and Patients. *J Magn Reson Imaging* (2001) 14(1):23–30. doi:10.1002/jmri.1146
10. Kellman P, Chefd'hotel C, Lorenz CH, Mancini C, Arai AE, and McVeigh ER. Fully Automatic, Retrospective Enhancement of Real-Time Acquired Cardiac Cine MR Images Using Image-Based Navigators and Respiratory Motion-Corrected Averaging. *Magn Reson Med* (2008) 59(4):771–8. doi:10.1002/mrm.21509
11. Kellman P, Chefd'hotel C, Lorenz CH, Mancini C, Arai AE, and McVeigh ER. High Spatial and Temporal Resolution Cardiac Cine MRI from Retrospective Reconstruction of Data Acquired in Real Time Using Motion Correction and Resorting. *Magn Reson Med* (2009) 62(6):1557–64. doi:10.1002/mrm.22153
12. Kaji S, Yang PC, Kerr AB, Tang WHW, Meyer CH, Macovski A, et al. Rapid Evaluation of Left Ventricular Volume and Mass without Breath-Holding Using Real-Time Interactive Cardiac Magnetic Resonance Imaging System. *J Am Coll Cardiol* (2001) 38(2):527–33. doi:10.1016/s0735-1097(01)01399-7
13. Feng L, Srichai MB, Lim RP, Harrison A, King W, Adluru G, et al. Highly Accelerated Real-Time Cardiac Cine MRI Usingk-tSPARSE-SENSE. *Magn Reson Med* (2013) 70(1):64–74. doi:10.1002/mrm.24440
14. Tsao J, Boesiger P, and Pruessmann KP. K-T BLAST Andk-T SENSE: Dynamic MRI with High Frame Rate Exploiting Spatiotemporal Correlations. *Magn Reson Med* (2003) 50(5):1031–42. doi:10.1002/mrm.10611
15. Tsao J, Kozerke S, Boesiger P, and Pruessmann KP. Optimizing Spatiotemporal Sampling fork-t BLAST Andk-T SENSE: Application to High-Resolution Real-Time Cardiac Steady-State Free Precession. *Magn Reson Med* (2005) 53(6):1372–82. doi:10.1002/mrm.20483
16. Jung H, Park J, Yoo J, and Ye JC. Radial K-T FOCUSS for High-Resolution Cardiac Cine MRI. *Magn Reson Med* (2010) 63(1):68–78. doi:10.1002/mrm.22172
17. Otazo R, Kim D, Axel L, and Sodickson DK. Combination of Compressed Sensing and Parallel Imaging for Highly Accelerated First-Pass Cardiac Perfusion MRI. *Magn Reson Med* (2010) 64(3):767–76. doi:10.1002/mrm.22463
18. Ding Y, Chung Y-C, Jekic M, and Simonetti OP. A New Approach to Autocalibrated Dynamic Parallel Imaging Based on the Karhunen-Loeve Transform: KL-TSENSE and KL-TGRAPPA. *Magn Reson Med* (2011) 65(6):1786–92. doi:10.1002/mrm.22766
19. Deshmane A, Gulani V, Griswold MA, and Seiberlich N. Parallel MR Imaging. *J Magn Reson Imaging* (2012) 36(1):55–72. doi:10.1002/jmri.23639
20. Blaimer M, Breuer F, Mueller M, Heidemann RM, Griswold MA, and Jakob PM. SMASH, SENSE, PILS, GRAPPA: How to Choose the Optimal Method. *Top Magn Reson Imaging* (2004) 15(4):223–36. doi:10.1097/01.rmr.0000136558.09801.dd
21. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, et al. Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA). *Magn Reson Med* (2002) 47(6):1202–10. doi:10.1002/mrm.10171
22. Otazo R, Candès E, and Sodickson DK. Low-Rank Plus Sparse Matrix Decomposition for Accelerated Dynamic MRI with Separation of Background and Dynamic Components. *Magn Reson Med* (2015) 73(3):1125–36. doi:10.1002/mrm.25240
23. Lingala SG, Hu Y, DiBella E, and Jacob M. Accelerated Dynamic MRI Exploiting Sparsity and Low-Rank Structure: K-T SLR. *IEEE Trans Med Imaging* (2011) 30(5):1042–54. doi:10.1109/tmi.2010.2100850
24. Zhao B, Haldar JP, Christodoulou AG, and Liang ZP. Image Reconstruction from Highly Undersampled (K,t)-Space Data with Joint Partial Separability and Sparsity Constraints. *IEEE Trans Med Imaging* (2012) 31(9):1809–20. doi:10.1109/TMI.2012.2203921
25. Haji-Valizadeh H, Rahsepar AA, Collins JD, Bassett E, Isakova T, Block T, et al. Validation of Highly Accelerated Real-Time Cardiac Cine MRI with Radial K-Space Sampling and Compressed Sensing in Patients at 1.5T and 3T. *Magn Reson Med* (2018) 79(5):2745–51. doi:10.1002/mrm.26918
26. Uecker M, Zhang S, Voit D, Karaus A, Merboldt K-D, and Frahm J. Real-Time MRI at a Resolution of 20 Ms. *NMR Biomed* (2010) 23(8):986–94. doi:10.1002/nbm.1585
27. Hauptmann A, Arridge S, Lucka F, Muthurangu V, and Steeden JA. Real-Time Cardiovascular MR with Spatio-Temporal Artifact Suppression Using Deep Learning-Proof of Concept in Congenital Heart Disease. *Magn Reson Med* (2019) 81(2):1143–56. doi:10.1002/mrm.27480
28. Schlemper J, Caballero J, Hajnal JV, Price AN, and Rueckert D. A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction. *IEEE Trans Med Imaging* (2018) 37(2):491–503. doi:10.1109/tmi.2017.2760978
29. Küstner T, Fuin N, Hammernik K, Bustin A, Qi H, Hajhosseiny R, et al. CINENet: Deep Learning-Based 3D Cardiac CINE MRI Reconstruction with Multi-Coil Complex-Valued 4D Spatio-Temporal Convolutions. *Sci Rep* (2020) 10(1):13710. doi:10.1038/s41598-020-70551-8
30. El-Rewaidy H, Fahmy AS, Pashakhanloo F, Cai X, Kucukseymen S, Csecs I, et al. Multi-Domain Convolutional Neural Network (MD-CNN) for Radial Reconstruction of Dynamic Cardiac MRI. *Magn Reson Med* (2021) 85(3):1195–208. doi:10.1002/mrm.28485
31. Shen D, Ghosh S, Haji-Valizadeh H, Pathrose A, Schiffers F, Lee DC, et al. Rapid Reconstruction of Highly Undersampled, Non-Cartesian Real-Time Cine K-Space Data Using a Perceptual Complex Neural Network (PCNN). *NMR Biomed* (2021) 34(1):e4405. doi:10.1002/nbm.4405
32. Qin C, Schlemper J, Caballero J, Price AN, Hajnal JV, and Rueckert D. Convolutional Recurrent Neural Networks for Dynamic MR Image Reconstruction. *IEEE Trans Med Imaging* (2019) 38(1):280–90. doi:10.1109/tmi.2018.2863670
33. Ronneberger O, Fischer P, and Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: N Navab, J Hornegger, WM Wells, and AF Frangi, editors. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th international conference; October 5-9, 2015; Munich, Germany. Cham: Springer International Publishing (2015).
34. Matej S, Fessler JA, and Kazantsev IG. Iterative Tomographic Image Reconstruction Using Fourier-Based Forward and Back-Projectors. *IEEE Trans Med Imaging* (2004) 23(4):401–12. doi:10.1109/tmi.2004.824233
35. Wundrak S, Paul J, Ulrici J, Hell E, and Rasche V. A Small Surrogate for the Golden Angle in Time-Resolved Radial MRI Based on Generalized Fibonacci Sequences. *IEEE Trans Med Imaging* (2015) 34(6):1262–9. doi:10.1109/tmi.2014.2382572
36. Wundrak S, Paul J, Ulrici J, Hell E, Geibel M-A, Bernhardt P, et al. Golden Ratio Sparse MRI Using Tiny golden Angles. *Magn Reson Med* (2016) 75(6):2372–8. doi:10.1002/mrm.25831
37. Pruessmann KP, Weiger M, Scheidegger MB, and Boesiger P. SENSE: Sensitivity Encoding for Fast MRI. *Magn Reson Med* (1999) 42(5):952–62. doi:10.1002/(sici)1522-2594(199911)42:5<952::aid-mrm16>3.0.co;2-s
38. Walsh DO, Gmitro AF, and Marcellin MW. Adaptive Reconstruction of Phased Array MR Imagery. *Magn Reson Med* (2000) 43(5):682–90. doi:10.1002/(sici)1522-2594(200005)43:5<682::aid-mrm10>3.0.co;2-g
39. Eo T, Jun Y, Kim T, Jang J, Lee HJ, and Hwang D. KIKI -Net: Cross-Domain Convolutional Neural Networks for Reconstructing Undersampled Magnetic Resonance Images. *Magn Reson Med* (2018) 80(5):2188–201. doi:10.1002/mrm.27201
40. Terpstra ML, Maspero M, d'Agata F, Stemkens B, Intven MPW, Lagendijk JJW, et al. Deep Learning-Based Image Reconstruction and Motion Estimation from Undersampled Radial K-Space for Real-Time MRI-Guided Radiotherapy. *Phys Med Biol* (2020) 65(15):155015. doi:10.1088/1361-6560/ab9358

41. Fahmy AS, Neisius U, Chan RH, Rowin EJ, Manning WJ, Maron MS, et al. Three-Dimensional Deep Convolutional Neural Networks for Automated Myocardial Scar Quantification in Hypertrophic Cardiomyopathy: A Multicenter Multivendor Study. *Radiology* (2019) 294(1):52–60. doi:10.1148/radiol.2019190737
42. Fahmy AS, El-Rewaidy H, Nezafat M, Nakamori S, and Nezafat R. Automated Analysis of Cardiovascular Magnetic Resonance Myocardial Native T1 Mapping Images Using Fully Convolutional Neural Networks. *J Cardiovasc Magn Reson* (2019) 21(1):7. doi:10.1186/s12968-018-0516-1
43. El-Rewaidy H, Neisius U, Mancio J, Kucukseymen S, Rodriguez J, Paskavitz A, et al. Deep Complex Convolutional Network for Fast Reconstruction of 3D Late Gadolinium Enhancement Cardiac MRI. *NMR Biomed* (2020) 33(7): e4312. doi:10.1002/nbm.4312
44. Han Y, Sunwoo L, and Ye JC. ${k}$ -Space Deep Learning for Accelerated MRI. *IEEE Trans Med Imaging* (2020) 39(2):377–86. doi:10.1109/tmi.2019.2927101
45. Lee D, Yoo J, Tak S, and Ye JC. Deep Residual Learning for Accelerated MRI Using Magnitude and Phase Networks. *IEEE Trans Biomed Eng* (2018) 65(9): 1985–95. doi:10.1109/tbme.2018.2821699
46. Hundley WG, Hamilton CA, Thomas MS, Herrington DM, Salido TB, Kitzman DW, et al. Utility of Fast Cine Magnetic Resonance Imaging and Display for the Detection of Myocardial Ischemia in Patients Not Well Suited for Second Harmonic Stress Echocardiography. *Circulation* (1999) 100(16): 1697–702. doi:10.1161/01.cir.100.16.1697
47. Messroghli DR, Moon JC, Ferreira VM, Grosse-Wortmann L, He T, Kellman P, et al. Clinical Recommendations for Cardiovascular Magnetic Resonance Mapping of T1, T2, T2* and Extracellular Volume: A Consensus Statement by the Society for Cardiovascular Magnetic Resonance (SCMR) Endorsed by the European Association for Cardiovascular Imaging (EACVI). *J Cardiovasc Magn Reson* (2017) 19(1):75. doi:10.1186/s12968-017-0389-8

Conflict of Interest: HH-V is currently an employee of Canon Medical Research USA Inc. The work presented in this study was performed during his post-doctoral training fellowship at Beth Israel Deaconess Medical Center.

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 09 September 2021
doi: 10.3389/fonc.2021.697721

Calculation of Apparent Diffusion Coefficients in Prostate Cancer Using Deep Learning Algorithms: A Pilot Study

*Lei Hu[1], Da Wei Zhou[2], Cai Xia Fu[3], Thomas Benkert[4], Yun Feng Xiao[1], Li Ming Wei[1] and Jun Gong Zhao[1]**

[1] Department of Diagnostic and Interventional Radiology, Shanghai Jiao Tong University Affiliated Sixth People's Hospital, Shanghai, China, [2] State Key Laboratory of Integrated Services Networks, School of Telecommunications Engineering, Xidian University, Xi'an, China, [3] Magnetic Resonance (MR) Application Development, Siemens Shenzhen Magnetic Resonance Ltd., Shenzhen, China, [4] MR Application Predevelopment, Siemens Healthcare GmbH, Erlangen, Germany

Background: Apparent diffusion coefficients (ADCs) obtained with diffusion-weighted imaging (DWI) are highly valuable for the detection and staging of prostate cancer and for assessing the response to treatment. However, DWI suffers from significant anatomic distortions and susceptibility artifacts, resulting in reduced accuracy and reproducibility of the ADC calculations. The current methods for improving the DWI quality are heavily dependent on software, hardware, and additional scan time. Therefore, their clinical application is limited. An accelerated ADC generation method that maintains calculation accuracy and repeatability without heavy dependence on magnetic resonance imaging scanners is of great clinical value.

Objectives: We aimed to establish and evaluate a supervised learning framework for synthesizing ADC images using generative adversarial networks.

Methods: This prospective study included 200 patients with suspected prostate cancer (training set: 150 patients; test set #1: 50 patients) and 10 healthy volunteers (test set #2) who underwent both full field-of-view (FOV) diffusion-weighted imaging (f-DWI) and zoomed-FOV DWI (z-DWI) with *b*-values of 50, 1,000, and 1,500 s/mm^2. ADC values based on f-DWI and z-DWI (f-ADC and z-ADC) were calculated. Herein we propose an ADC synthesis method based on generative adversarial networks that uses f-DWI with a single *b*-value to generate synthesized ADC (s-ADC) values using z-ADC as a reference. The image quality of the s-ADC sets was evaluated using the peak signal-to-noise ratio (PSNR), root mean squared error (RMSE), structural similarity (SSIM), and feature similarity (FSIM). The distortions of each ADC set were evaluated using the T2-weighted image reference. The calculation reproducibility of the different ADC sets was compared using the intraclass correlation coefficient. The tumor detection and classification abilities of each ADC set were evaluated using a receiver operating characteristic curve analysis and a Spearman correlation coefficient.

OPEN ACCESS

Edited by:
Oliver Diaz,
University of Barcelona, Spain

Reviewed by:
Huaiqiang Sun,
Sichuan University, China
Wenwu Ling,
Sichuan University, China

****Correspondence:***
Jun Gong Zhao
zhaojungongradio@hotmail.com

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

Received: *20 April 2021*
Accepted: *11 August 2021*
Published: *09 September 2021*

Citation:
Hu L, Zhou DW, Fu CX, Benkert T, Xiao YF, Wei LM and Zhao JG (2021) Calculation of Apparent Diffusion Coefficients in Prostate Cancer Using Deep Learning Algorithms: A Pilot Study. Front. Oncol. 11:697721. doi: 10.3389/fonc.2021.697721

Results: The s-ADC_{b1000} had a significantly lower RMSE score and higher PSNR, SSIM, and FSIM scores than the s-ADC_{b50} and s-ADC_{b1500} (all $P < 0.001$). Both z-ADC and s-ADC_{b1000} had less distortion and better quantitative ADC value reproducibility for all the evaluated tissues, and they demonstrated better tumor detection and classification performance than f-ADC.

Conclusion: The deep learning algorithm might be a feasible method for generating ADC maps, as an alternative to z-ADC maps, without depending on hardware systems and additional scan time requirements.

Keywords: apparent diffusion coefficient, diffusion magnetic resonance imaging, deep learning, prostatic neoplasms, supervised machine learning

INTRODUCTION

Diffusion-weighted imaging (DWI) currently constitutes an integral part of multiparametric magnetic resonance imaging (MRI) examinations of the prostate. Apparent diffusion coefficients (ADCs) obtained with DWI are highly valuable for detecting and staging prostate cancer, evaluating cancer aggressiveness (1, 2), guiding targeted biopsies, and assessing the response to treatment (3–10). Clinically, the accuracy of the ADC measurement depends on the quality of the DWI image.

Single-shot echo-planar imaging (SS-EPI)-based sequences are preferred for DWI because of its ability to acquire the images rapidly and the robustness of the technique against motion artifacts. However, because of its high sensitivity to chemical shifts and magnetic susceptibilities (11), conventional SS-EPI DWI suffers from significant anatomic distortions (12) and susceptibility artifacts, resulting in reduced ADC calculation accuracy and reproducibility (12–14). Another limitation is the low signal-to-noise ratios observed during DWI, which result in noise-induced signal intensity biases (15, 16) and inaccurate ADC maps. These drawbacks may lead to an error in judgment regarding the condition of a patient and a potential misdiagnosis of malignant lesions or over-treatment of benign lesions. Zoomed field-of-view (FOV) DWI (z-DWI) is an appealing attempt to address these limitations. This method reduces the scanning time as well as artifacts, distortions, and blurring of images, and it also has improved spatial resolution (17, 18). Additionally, z-DWI can effectively improve the ADC map accuracy (17, 18); however, the technique depends on radio frequency design and software platforms (17–19), which can make it unaffordable for many small- and medium-sized hospitals and their patients. Moreover, a reduced FOV may prevent the visualization of lymph nodes (3). Therefore, the clinical application of z-DWI is limited. A method that can consistently generate high-quality ADC images with reduced equipment costs will be of more benefit to patients in clinical practice.

Recently, the advent of generative adversarial networks (GANs) (20) has shown promise for optimizing medical image quality without relying on software and equipment conditions (21). As a generative model, the objective of a GAN is to learn the underlying training data distributions to generate realistic images that are indistinguishable from the input datasets (21). With their ability to mimic data distributions, GANs have been used to translate low-quality images into high-quality counterparts. Previous studies have successfully used GANs to improve computed tomography (CT) or MRI quality in terms of de-noising (22), increased resolution (23), artifact reduction (24), and motion correction (25). Inspired by these image optimization solutions, we hypothesized that deep learning algorithms based on GANs might be promising for generating ADC maps with good image quality and improved ADC calculation accuracy. The purpose of this study was to establish and evaluate a supervised learning framework based on a GAN to synthesize realistic zoomed FOV ADC images using conventional full FOV SS-EPI DWI images with a single *b*-value.

MATERIALS AND METHODS

Patients and Healthy Volunteers

This prospective study was approved by the local ethics committee, and informed consent was obtained from each participant. All the procedures involving human participants were performed in accordance with the 1964 Helsinki Declaration and its later amendments. A total of 200 consecutive patients underwent preoperative MRI examinations and subsequent MRI fusion ultrasound-guided biopsies for suspected prostate cancer (PCa) between December 2018 and May 2020. The inclusion criteria were as follows: patients with (1) at least one prostate lesion visible on DWI and ADC maps and (2) complete clinical information and pathologic examination information, including biopsy reports. Ten healthy volunteers were also recruited for the study. The study included four steps: (1) MRI examinations, (2) model training, (3) image quality assessments, and (4) ADC assessments (**Figure 1**).

MRI Examinations and Datasets

All the patients and volunteers underwent multiparametric MRI examinations of the prostate using a 3T MRI scanner (MAGNETOM Skyra, Siemens Healthcare, Erlangen, Germany) equipped with a phased-array 18-channel body coil and an integrated 32-channel spine coil. Both a transversal single-shot full FOV-EPI DWI (f-DWI) and a prototypic non-parallel transmission zoomed EPI DWI (z-DWI) with *b*-values of 50, 1,000, and 1,500 s/mm^2 were performed with the ADC reconstruction maps (f-ADC and z-ADC) using a standard

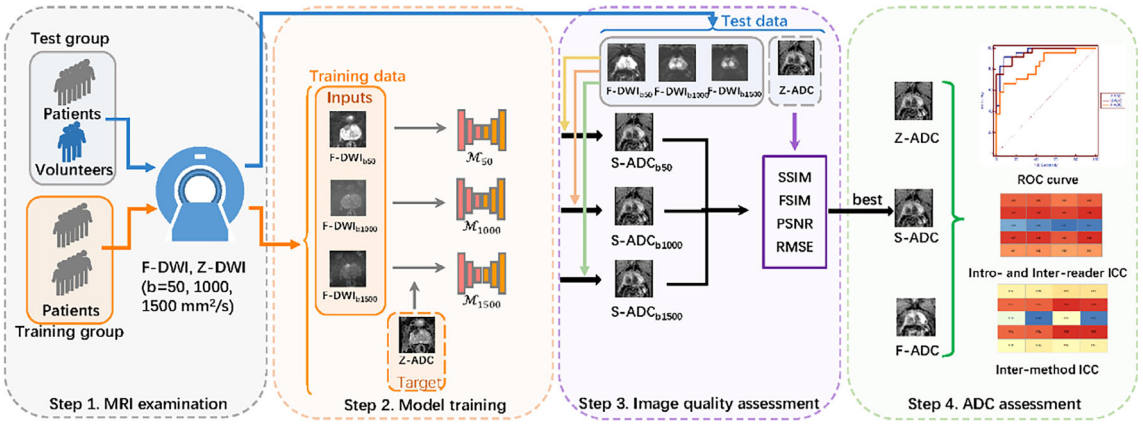


FIGURE 1 | Overall study flow diagram. Step 1: All the patients and healthy volunteers underwent multiparametric magnetic resonance imaging examinations of the prostate, including full field-of-view (FOV) diffusion-weighted imaging (f-DWI) and zoomed FOV diffusion-weighted imaging with *b*-values of 50, 1,000, and 1,500 s/mm^2. Step 2: The models that used full f-DWI with different *b*-values (f-DWI$_{b50}$, f-DWI$_{b1000}$, and f-DWI$_{b1500}$) to synthesize the apparent diffusion coefficient (s-ADC) maps (s-ADC$_{b50}$, s-ADC$_{b1000}$, and s-ADC$_{b1500}$) were trained. Step 3: The image quality of s-ADC$_{b50}$, s-ADC$_{b1000}$, and s-ADC$_{b1500}$ were evaluated using the peak signal-to-noise ratio, root mean square error, structural similarity, and feature similarity. Step 4: An ADC assessment was performed to determine reproducibility, tumor detection, and classification.

mono-exponential with all the acquired *b*-values (14). Axial T2-weighted images were obtained from all the participants, and the total examination time was approximately 7 min and 40 s. The detailed scan parameters are shown in **Table 1**.

Patient images were randomly divided into two groups (training set: 150 patients, test set #1: 50 patients). The training set was used to build the framework and train different models to synthesize the ADC maps (s-ADCs). Test set #1 was used to test the reproducibility of the s-ADC prostate lesion measurements, along with tumor detection. The images of the healthy volunteers were regarded as test set #2, which was used to test the reproducibility and consistency of the normal prostate tissue s-ADC calculations, including the peripheral zone (PZ) and the transitional zone (TZ).

Data Pre-Processing

Before the model training could occur, image selection, cropping, and registration were performed on f-DWI with *b*-values of 50, 1,000, and 1,500 s/mm^2 and the z-ADC images. The first and last slices that did not cover the prostate were removed manually. The images with severe distortion and artifacts were also removed. Ultimately, there were between five and 20 DWI images selected for each person. Finally, there were 2,250 images from each set for the 150 patients in the training set, 750 images from each set for the 50 patients in test set #1, and 145 images from each set for the 10 healthy volunteers in test set #2.

Due to hardware limitations of the graphics cards and the CPU memory, we used only axial slices of the cropped data to train the two-dimensional generation models. The f-DWI data had an original voxel size of 2.13 × 2.13 × 3.3 mm^3 and a matrix size of 178 ×132, whereas the z-ADC data had a voxel size of 0.95 × 0.95 × 3.3 mm^3 and a matrix size of 112 × 200. The f-DWI data were first resampled to a voxel size of 0.95 × 0.95 × 3.3 mm^3 with a matrix size of 360 × 267, and both modalities were cropped at the center to extract the relevant prostate region. The f-DWI data were then aligned to the z-ADC data using the affine transformation implemented by the Advanced

TABLE 1 | The magnetic resonance imaging sequence parameters.

Parameter	T2-weighted imaging	F-DWI	Z-DWI
Field-of-view, FOV (mm^2)	200 × 200	380 × 280	190 × 106
Imaging matrix	320 × 320	132 × 178	112 × 200
Thickness (mm)	3.5	3	3
Distance fact	0	10%	10%
B-value (s/mm^2)	n.a.	50, 1,000, 1500	50, 1,000, 1500
Echo time (ms)	101	73	76
Time to repeat (ms)	6,000	4,200	3,800
Bandwidth (Hz/pixel)	200	1,872	1,612
Scan time (min)	2:08	3:05	2:27

f-ADC, mean apparent diffusion coefficient (ADC) map derived from full FOV diffusion-weighted imaging with all available b-values (b = 50, 1,000, and 1,500 mm^2/s); z-ADC, ADC map derived from zoomed FOV diffusion-weighted imaging with all available b-values (b = 50, 1,000, and 1,500 mm^2/s), n.a., no available.

Normalization Tools (https://github.com/ANTsX/ANTs). To facilitate the model training, all the two-dimensional axial slices were scaled to a unified resolution of 224 × 224 pixels.

To select a suitable *b*-value for ADC synthesis, we first used 2,250 paired f-DWI images with *b*-values of 50 s/mm^2 and the ground truth z-ADC maps from the training set as inputs and references, respectively, to train our framework-based model M_{50} to synthesize ADC maps (s-ADC_{b50}). Similarly, the M_{1000} and M_{1500} models based on the f-DWI images with *b*-values of 1,000 and 1,500 s/mm^2 were trained to synthesize ADC maps (s-ADC_{b1000} and s-ADC_{b1500}).

Model Training

We have proposed a GAN-based framework to generate realistic z-ADC maps from f-DWI maps (**Figure 2**).

The generator *G* translates the input image (f-DWI) *X* into the synthesized ADC (s-ADC) *S* with a quality similar to the ground truth z-ADC, which is used as reference *Y*. The discriminator *D* takes either the *S* or the *Y* as input and determines whether the input is a real z-ADC map. Specifically, we used a deep residual network structure with skip connections to construct the generator *G* and a full convolution network to construct the discriminator *D*. The adversarial loss of the discriminator *D* is formulated as follows:

$$\mathcal{L}_D = \mathbb{E}_{Y\sim P_Y}[(D(Y)-1)^2] + E_{S\sim P_S}[(D(S))^2]$$

The adversarial loss of the generator *G* is formulated as follows:

$$\mathcal{L}_G^{adv} = \mathbb{E}_{S\sim P_S}[(D(S)-1)^2]$$

Considering that the standard GAN might not adequately preserve the tumors/lesions during image-to-image translation (26), we introduced a multi-level verification (MLV) mechanism, including a pre-trained recognition model *C*. This mechanism promotes the generator *G* to better retain the features, which helps in the diagnosis. Using *C*, the proposed MLV mechanism provides more details about the tumor/lesion features when they are extracted from the input images. *G* represents the tumor/lesion texture, making it better and more robust against changes in appearance and geometric transformations (27).

We first obtained a recognition model *C*, which was pre-trained on a VGG-19-based network using the processed images from the patients and healthy volunteers with a benign or malignant label (28). Subsequently, the multiple layers of model *C* extracted the multi-level features from the fake synthetic ADC map *S* and the ground truth ADC map. The sum of the mean square errors of the features in each level layer was used as the multi-level feature loss to supervise the generator *G*.

Inspired by the current work (29) and considering the use of multi-level features, we selected the features in the 0, 1st, 3rd, and 5th level layers. The loss of the multi-level verification mechanism is formulated as follows:

$$\mathcal{L}_G^{mlv} = \Sigma_{i=0,1,3,5}\theta_i \cdot || C^i(S) - C^i(Y) ||_2^2$$

where $\theta_i \in (0, 1)$ denotes the weight parameter for the loss ($\mathcal{L}_{mlf}^i$) at different levels, and it is optimized in each epoch to cause a faster decrease in the loss of the larger items. The θ_i in the *j - th* epoch can be computed as follows:

$$\theta_i^j = \Sigma_{n=0,1,3,5}\frac{|| C_i^{j-1}(S) - C_i^{j-1}(Y) ||_2^2}{|| C_n^{j-1}(S) - C_n^{j-1}(Y) ||_2^2}$$

where $C_i^{j-1}(\cdot)$ denotes the feature of the *i - th* layer in the *(j-1)-th* epoch, and $C_n^{j-1}(\cdot)$ indicates the feature of the *n - th* layer in the *(j-1)-th* epoch. We initialized θ_i to 1/4. The objective function of generator *G* is formulated as follows:

$$\mathcal{L}_G = \mathcal{L}_G^{adv} + \lambda_1\mathcal{L}_G^{mlv}$$

with γ_1 set to 10^{-1}.

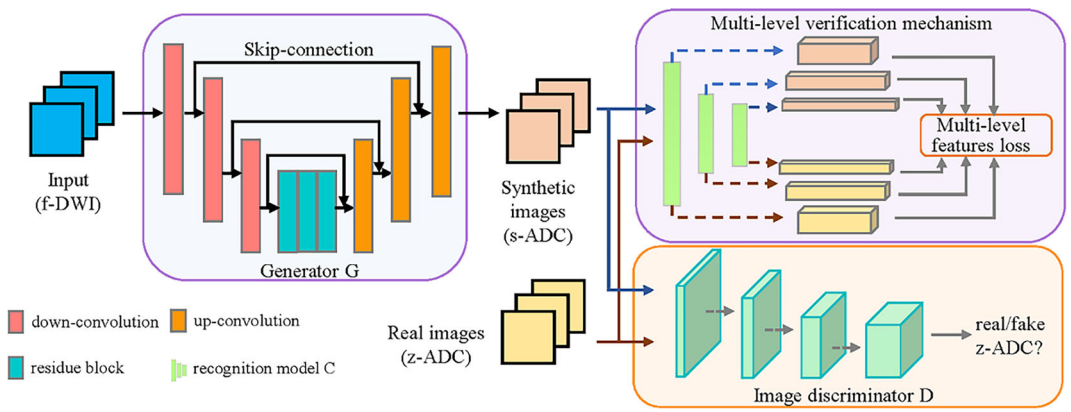


FIGURE 2 | Illustration of our framework. The proposed framework consists of a generator (*G*), which was constructed using a deep convolution network with skip connections, and an image discriminator (*D*) constructed using a full convolution network. The *G* transforms the f-DWI into a synthesized apparent diffusion coefficient (s-ADC) using zoomed field-of-view diffusion-weighted imaging (z-ADC) as a reference. The *D* takes either s-ADC or z-ADC as the input and determines whether the input is a real z-ADC. In addition, to promote *G* in an effort to retain better features for diagnosis, we introduced a multi-level verification mechanism, including a pre-trained recognition model (*C*), to extract the multi-level features from the s-ADC and the z-ADC.

Experimental Settings

The generator consists of three convolution layers, followed by five residual blocks and three deconvolution layers. Each convolution or deconvolution layer is followed by an instance-normalization layer and a ReLu activation layer. The discriminator consists of five convolution layers. The learning rate was set to 0.001 for both the generator and the discriminator. The batch size was set to 5, and the epoch was set to 50. The details of the generator and discriminator can be found at https://github.com/huxiaolie/ADC_generation. All the algorithms were implemented using Python 3.6 (https://www.python.org/downloads/release/python-362/) and Pytorch 1.6.0 (https://pytorch.org/get-started/previous-versions/) on an Ubuntu 16.04 system with an NVIDIA TITAN XP GPU.

Image Quality Assessment

The s-ADC sets were synthesized using each model with inputs from the f -DWI images with *b*-values of 50, 1,000, and 1,500 s/mm^2 for test set #1 (50 patients) and test set #2 (10 healthy volunteers), and they were compared using peak signal-to-noise ratios (PSNRs), root mean square errors (RMSEs), structural similarities (SSIMs), and feature similarities (FSIMs) (30).

A radiologist with 6 years of experience with prostate MRIs measured the anterior–posterior (AP) and left–right (LR) diameters of each prostate on the ADC set on the slice on which the prostate showed the greatest cross-sectional area. The differences in the measured AP and LR diameters of the prostate relative to the T2-weighted image (T2WI) were computed for f-ADC, z-ADC, and s-ADC, with the best performance from the above-mentioned quantitative evaluation.

ADC Measurement Assessment

For the patient study, two radiologists with 5 and 10 years of experience with prostate MRIs and who were unaware of the clinical, surgical, and histologic findings independently drew a circular region of interest (ROI) with an area of approximately 0.5–0.8 cm^2 in the center of the lesion, excluding its edges. For the healthy volunteer study, the readers drew circular ROIs with an area of approximately 0.5 cm^2 in the peripheral and transitional zones on the ADC maps using axial T2-weighted images as the anatomical reference. The mean ADC values for each ROI were recorded.

The ADC sets of all the patients and healthy volunteers were measured twice using Image J (NIH Image, Bethesda, MD) in a different order, with an interval of 2 weeks. The first measurement given by the two readers showed the consistency of the ADC measurements for each ADC set. The second measurement showed the repeatability of the ADC values for each ADC set.

Tumor Detection Assessment

The s-ADC set with the best image quality and ADC measurement assessment among the three s-ADC sets was selected for tumor detection assessments. The selected s-ADC was compared with the f-ADC and z-ADC in terms of the ability to differentiate benign from malignant lesions. The correlation between the ADC values in the different ADC sets and tumor grades was also evaluated.

Statistical Analyses

Analyses of the baseline characteristics between the training group and the test group were conducted. An independent *t*-test was used to assess normally distributed continuous variables. The Mann–Whitney *U*-test was used to assess non-normally distributed continuous variables.

To assess differences in the image quality metrics (PSNR, RMSE, SSIM, and FSIM) between any two s-ADC sets, a paired Student's *t*-test was applied. The intraclass correlation coefficient (ICC) was used to assess the inter-and intra-reader repeatability of the ADC measurements for each tissue (malignant lesion, benign lesion, peripheral zone, and transitional zone) in each ADC set (f-ADC, z-ADC, and s-ADC). The ICC was also used to evaluate the inter-method reliability of the ADC values for each tissue between the synthesized image (s-ADC) and the reference image (z-ADC). A receiver operating characteristic (ROC) curve analysis was performed to assess the ability to discriminate between benign and malignant prostate lesions based on the ADC values. The differences in the area under the curve (AUC) values were tested using DeLong tests. The statistical analyses were performed using MedCalc software. Two-tailed tests were used to calculate all the *P*-values. Statistical significance was set at $P < 0.05$.

RESULTS

Demographic Characteristics

The patient characteristics are summarized in **Table 2**. There were no significant differences in the mean ages between the patients with and without PCa ($P = 0.557$). The mean prostate-specific antigen (PSA) level was significantly higher in patients with PCa compared to those without PCa ($P < 0.001$).

There were no significant differences in mean ages and mean PSA between the training set and test set #1 (mean ages: 68 ± 10 *vs.* 68 ± 12 years, $P = 0.974$; PSA: 29.872 ± 69.461 *vs.* 39.296 ± 92.604, $P = 0.154$). The mean age of test set #2 (healthy volunteers, 24 ± 3 years) is significantly lower than that of the training set and test set #1 ($P < 0.001$).

Image Quality Assessment

Visual comparisons of the s-ADC values generated with different *b*-value inputs are shown in **Figure 3**. We observed that the s-ADC_{b50} displayed blurred images of the prostate, bladder, rectum, pelvic floor muscles, and pubic symphysis in both the patients and the volunteers. Compared with s-ADC_{b50}, s-ADC_{b1000} and s-ADC_{b1500} could delineate normal tissues and lesions more clearly and sharply, which was in line with the

TABLE 2 | The clinical characteristics of the patient cohort.

Characteristics	Patients without cancer (*n* = 106)	Patients with cancer (*n* = 94)	*P*-value
Mean age (y) [range]	70 (52–87)	71 (48–88)	0.675
total PSA (ng/ml)	11.079 ± 9.013	57.002 ± 125.88	<0.001
Position, no.			
Peripheral zone	44	63	<0.001
Transitional zone	62	31	
Gleason score (*n*, %)			
6	— —	8	
7	— —	46	
8	— —	24	
9	— —	16	

The data are mean ± standard deviation, unless otherwise indicated.
PSA, prostate-specific antigen.

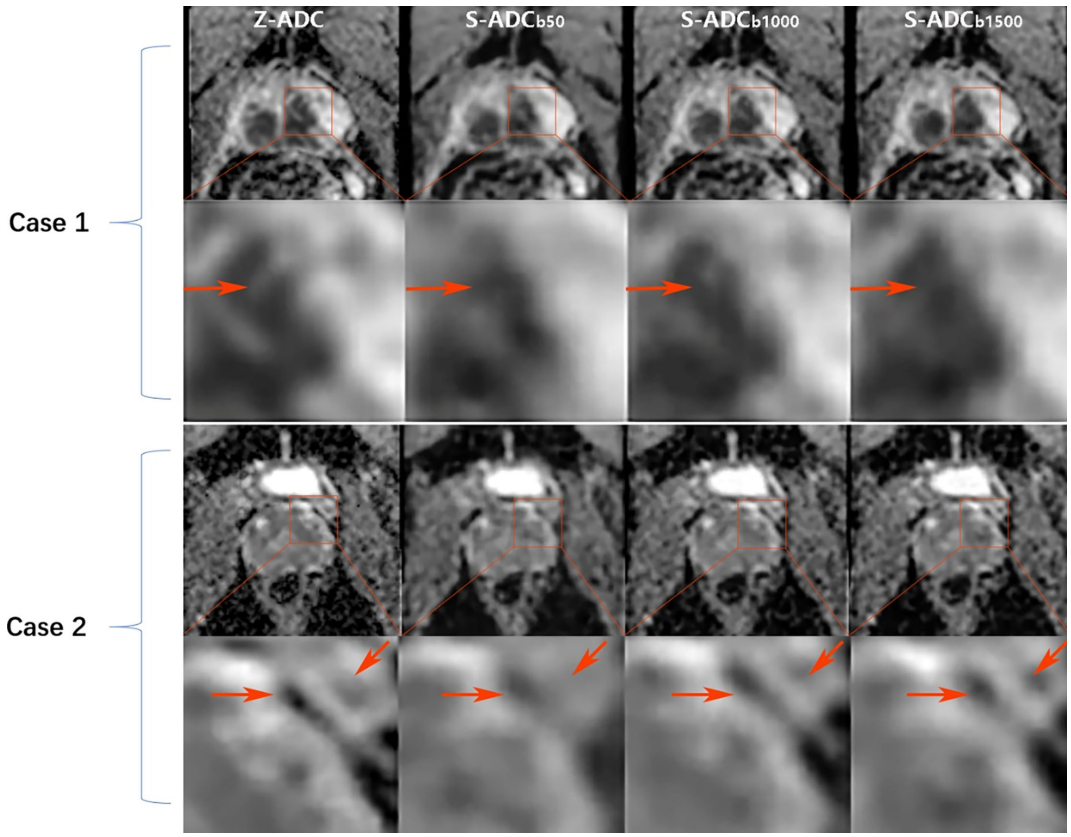


FIGURE 3 | Comparison of the synthesized apparent diffusion coefficient (s-ADC) maps. Case 1: An 82-year-old man with prostate cancer from test set 1 who had an initial prostate-specific antigen level of 13.04 ng/ml. Case 2: A 27-year-old healthy man from test set 2. For these two cases, both the s-ADC_{b1000} and s-ADC_{b1500} performed well in displaying the prostate, pelvic floor muscles, pubic symphysis, and the entire cancer lesion. However, the s-ADC_{b50} images of these structures are fuzzy. According to the local enlargement of the images (the images in the second and fourth lines), the s-ADC_{b1000} is more similar to the z-ADC than to the s-ADC_{b1500}, and it retains more details of the z-ADC (shown as red arrows).

ground truth. According to the magnified images of the local tissue structures, s-ADC_{b1000} provided more details than s-ADC_{b1500} with reference to z-ADC.

As shown in the violin plots (**Figure 4**), s-ADC_{b1000} performed better than the other two s-ADC sets in terms of the distribution, median, and inter-quartile ranges of the RMSE, SSIM, FSIM, and PSNR scores. The mean RMSE scores of s-ADC_{b50}, s-ADC_{b1000}, and s-ADC_{b1500} were 4.1×10^{-3}, 2.5×10^{-3}, and 3.1×10^{-3}, respectively. The mean PSNR scores of s-ADC_{b50}, s-ADC_{b1000}, and s-ADC_{b1500} were 48.0, 53.4, and 51.0, respectively. The mean SSIM scores of s-ADC_{b50}, s-ADC_{b1000}, and s-ADC_{b1500} were 0.972, 0.986, and

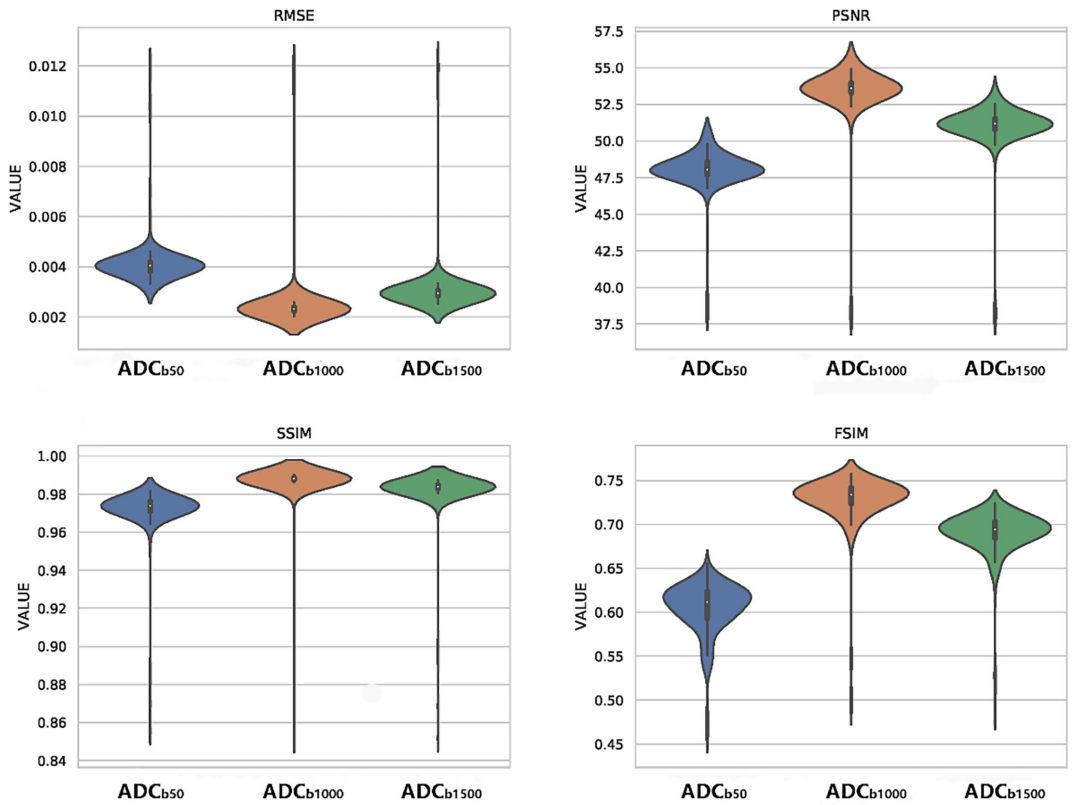


FIGURE 4 | Violin plots of the quantitative metric distributions of the s-ADC sets.

0.982, respectively. The mean FSIM scores of s-ADC_{b50}. s-ADC_{b1000}, and s-ADC_{b1500} were 0.604, 0.728, and 0.690, respectively. s-ADC_{b1000} had a significantly lower RMSE score and higher PSNR, SSIM, and FSIM scores than s-ADC_{b50} and s-ADC_{b1500} (all $P < 0.05$).

To evaluate the distortion in the ADC maps, the differences in the AP and LR diameters of the prostate relative to T2WI were both significantly lower for s-ADC (AP, 2.734; LR, 3.204) and z-ADC (AP, 2.755; LR, 3.073) than for f-DWI (AP, 5.916; LR, 5.053) (all $P < 0.001$).

ADC Measurement Assessment

The measurements of all the ADC sets (f-ADC, z-ADC, s-ADC_{b50}, ADC_{b1000}, and s-ADC_{b1500}) on various tissues from both readers are presented in **Table 3**. For all the ADC sets, the ADC values of the TZ are significantly lower than those of the PZ, while the ADC values of the malignant lesions are significantly lower than those of the benign lesions (all $P < 0.05$).

Figure 5 presents the results of the intra-reader reproducibility (**Figures 5A, B**) and inter-reader consistency (**Figure 5C**) analyses for each ADC set calculation. Both readers reported that the reproducibility of the ADC measurements for f-ADC, z-ADC, s-ADCb1000, and s-ADCb1500 was excellent for all the tissues, while the reliability of the ADC measurements for s-ADC_{b50} was good. The inter-reader consistency of all the ADC set measurements was excellent for all the tissues. **Table 4** shows the consistency of the ADC values between the z-ADC and s-ADC sets. The consistency of the ADC values in the transitional zone between z-DWI and s-DWI_{b50} was good, and the consistency of the ADC values between z-ADC and s-ADC_{b50} for the remaining tissues was excellent. For the s-ADC_{b1000} and s-ADC_{b1500} values, the consistency of the ADC values for z-ADC for all the tissues was excellent.

Tumor Detection Assessment

Among the three s-ADC sets, s-ADC_{b1000} performed the best in the image quality assessment and ADC evaluation. Therefore, it was selected for further comparisons with f-ADC and z-ADC in terms of tumor detection and classification (**Figure 6**). The ADC values for patients with malignant lesions and those with benign lesions measured by the two readers were used to compute the ROC curves (**Figure 7**). The comparisons of AUCs for both readers based on the f-ADC, z-ADC, and s-ADC sets are summarized in **Table 5**. Both the z-ADC and s-ADC sets showed significantly better predictive capabilities than the f-ADC set ($P \leq 0.027$). The differences in AUCs between s-ADC and z-ADC were not statistically significant (reader 1: $z = 0.134$, $P = 0.893$; reader 2: $z = 0.094$, $P = 0.925$).

TABLE 3 | The mean apparent diffusion coefficient (ADC) values ($\times 10^{-3}$ mm^2/s) of the different ADC sets.

Parameter	ADC value ($\times 10^{-3}$ mm^2/s)			
	Peripheral zone (n = 10)	Transitional zone (n = 10)	Benign lesions (n = 26)	Malignant lesions (n = 24)
Reader 1				
f-ADC	1.90 ± 0.11	1.41 ± 0.13	1.40 ± 0.28	1.06 ± 0.25
z-ADC	1.43 ± 0.17	1.20 ± 0.16	0.98 ± 0.18	0.61 ± 0.11
s-ADC$_{b50}$	1.43 ± 0.25	1.20 ± 0.18	1.09 ± 0.23	0.68 ± 0.13
s-ADC$_{b1000}$	1.43 ± 0.16	1.20 ± 0.16	0.99 ± 0.18	0.61 ± 0.17
s-ADC$_{b1500}$	1.46 ± 0.18	1.26 ± 0.16	1.01 ± 0.17	0.67 ± 0.18
Reader 2				
f-ADC	1.94 ± 0.14	1.39 ± 0.19	1.42 ± 0.29	1.06 ± 0.25
z-ADC	1.49 ± 0.16	1.22 ± 0.14	0.98 ± 0.18	0.61 ± 0.11
s-ADC$_{b50}$	1.44 ± 0.13	1.18 ± 0.14	1.02 ± 0.24	0.69 ± 0.13
s-ADC$_{b1000}$	1.48 ± 0.21	1.18 ± 0.13	0.99 ± 0.16	0.61 ± 0.15
s-ADC$_{b1500}$	1.45 ± 0.12	1.18 ± 0.09	1.00 ± 0.16	0.70 ± 0.10

The ADC values of the lesions were calculated using images from the patients in test set 1. The ADC values of the normal prostate tissues in the peripheral and transitional zones were calculated using images from the healthy volunteers in test set 2.

f-ADC, ADC map derived from full field-of-view (FOV) diffusion-weighted imaging (f-DWI) with all available b-values (b =50, 1,000, and 1,500 s/mm^2); z-ADC, ADC map derived from the zoomed FOV diffusion-weighted imaging and all available b-values (b = 50, 1,000, and 1,500 s/mm^2); s-ADC$_{b50}$, ADC map synthesized using our proposed deep learning framework with input from the f-DWI (b = s/mm^2); s-ADC$_{b1000}$, ADC map synthesized using our proposed deep learning framework with input from the f-DWI (b =1,000 s/mm^2); s-ADC$_{b1500}$, ADC map synthesized using our proposed deep learning framework with input from the f-DWI (b =1,500 s/mm^2).

DISCUSSION

The main contribution of our study to the literature is the proposed GAN-based ADC synthesis method that can be used to generate s-ADC maps using single *b*-value DWIs with better image quality and stronger ADC calculation accuracy and reproducibility than a full FOV ADC, but without dependence on software, hardware, and additional scanning time that zoomed FOV ADC technology requires. A shorter scan time will lead to better patient comfort and fewer motion artifacts due to involuntary or autonomous motions. The high reproducibility and accuracy of the ADC calculations may effectively reduce the risk of delayed treatment or unnecessary overtreatment due to the misdiagnoses of benign and malignant lesions. Therefore, the GAN-based ADC synthesis method can increase the clinical benefits to patients, reduce treatment times, and lower the costs incurred by patients and hospitals.

In previous studies (31–33), GANs have been used successfully for image-to-image transformations, such as in generating MRI or PET images using CT images or synthesizing CT images from MRI images, and they have performed well in terms of the traditional pixel-wise metrics. However, GAN-generated images do not have a physical meaning, and they can often lead to spurious images (21). As a result, it is difficult for GANs and their extensions to win the trust of clinicians. Therefore, we not only compared traditional pixel-wise metrics, including the RMSE, SSIM, FSIM, and PSNR scores, between the s-ADC sets and reference images, but we also compared the s-ADC and ADC values generated by traditional methods to evaluate the clinical value of GAN-generated images.

In the present study, we evaluated s-ADC maps that were based on DWI inputs with different *b*-values and found that the choice of *b*-values influenced the s-ADC values. Based on a subjective visual evaluation, the s-ADC$_{b1000}$ maps delineated normal tissues and lesions more clearly than the s-ADC$_{b50}$ maps, and they provided more details for targeted images than the s-ADC$_{b1500}$ set. The quantitative evaluation results are also consistent with the visual evaluation results. Among the three s-ADC sets, the s-ADC$_{b1000}$ set achieved a lower RMSE score and higher SSIM, FSIM, and PSNR scores than the s-ADC$_{b50}$ and s-ADC$_{b1500}$ sets, indicating that the s-ADC$_{b1000}$ set is more similar to the realistic z-ADC in terms of noise distribution, image structure, and features. Additionally, the s-ADC$_{b1000}$ set showed better intra-reader repeatability and inter-reader consistency than the s-ADC$_{b50}$ and s-ADC$_{b1500}$ sets. Moreover, the s-ADC$_{b1000}$ set showed the best ADC value inter-method consistency with the z-ADC set, suggesting that a DWI with a *b*-value of 1,000 s/mm^2 might be more suitable for synthesizing ADC maps than one with a *b*-value of 50 or 1,500 s/mm^2. The similarity between the target image z-DWI and s-DWI strongly depends on how much useful information the input f-DWI can provide to the generator for the extraction of meaningful features to begin the mapping between f-DWI and z-ADC. Low-*b*-value DWIs suffer from T2 shine-through or black-through effects, whereas high-*b*-value DWIs might be affected by diffusion kurtosis effects (34). These effects have a negative influence on image quality and lesion information, causing a relatively lower similarity between the s-ADC$_{b50}$ and s-ADC$_{b1500}$ sets and the z-ADC set compared to the s-ADC$_{b1000}$ set (3).

In our study, both the z-ADC and s-ADC sets showed less distortion and better reproducibility of the quantitative ADC values for all the evaluated tissues; they also showed better tumor detection and classification capacity than the f-ADC sets. The ADC values are generated for most of the current clinical implementations by calculating the signal intensity decay using two or more DWI sets with different *b*-values (1–5, 9–11, 13, 14). The reproducibility and accuracy of the calculated ADC values are affected by the choice of *b*-values (3, 4, 34) and the DWI image quality (14). The application of a significant number of

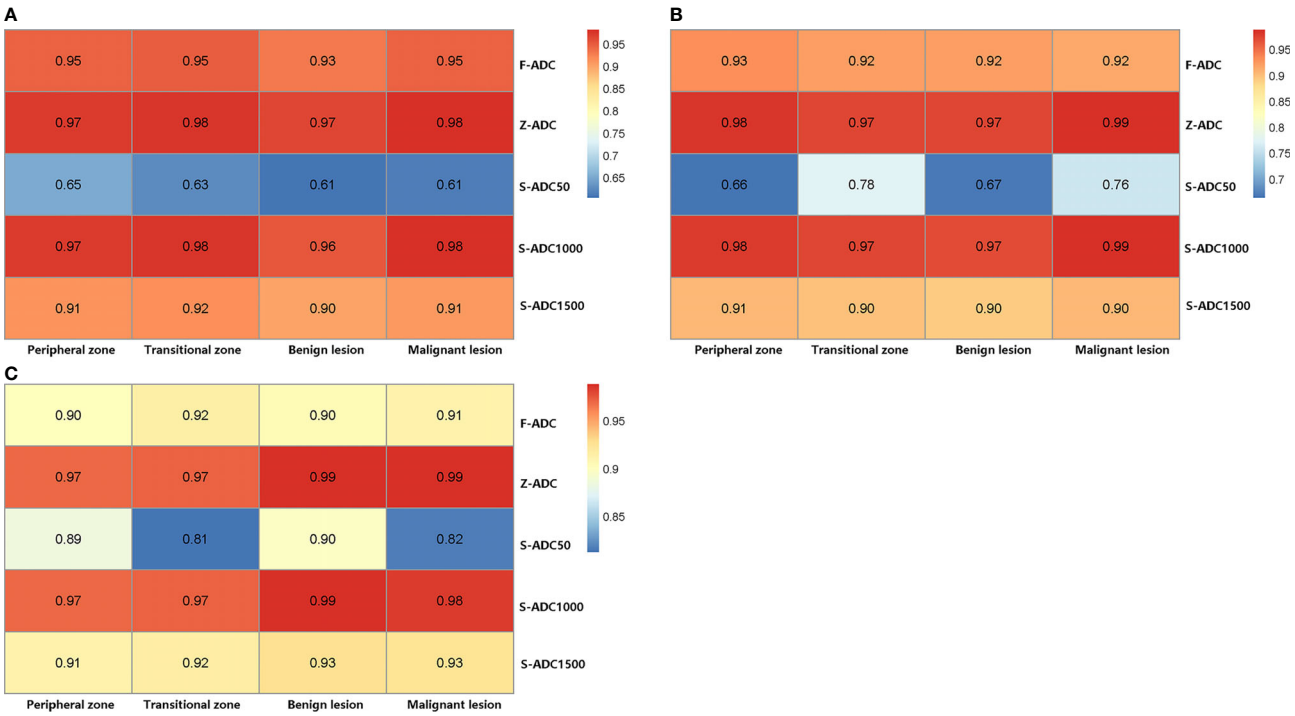


FIGURE 5 | The reproducibility of the apparent diffusion coefficient (ADC) measurements as evaluated by the intraclass correlation coefficient. **(A)** The ADC measurement repeatability of reader 1 and **(B)** reader 2, and **(C)** the consistency of the ADC measurements between readers 1 and 2.

TABLE 4 | Comparison between the inter-method intraclass correlation coefficients from the z-DWI and s-DWI sets.

Parameter	Inter-method intraclass correlation coefficient		
	s-apparent diffusion coefficient $(ADC)_{b50}$ *vs.* z-ADC	s-ADC_{b1000} *vs.* z-ADC	s-ADC_{b1500} *vs.* z-ADC
Reader 1			
Peripheral zone (*n* = 10)	0.87 (0.76–0.98)	0.99 (0.99–1.00)	0.99 (0.94–1.00)
Transitional zone (*n* = 10)	0.78 (0.58–0.98)	0.98 (0.87–1.00)	0.95 (0.73–0.99)
Benign lesion (*n* = 50)	0.86 (0.74–0.99)	0.98 (0.94–1.00)	0.98 (0.95–0.99)
Malignant lesion (*n* = 50)	0.89 (0.76–0.95)	0.90 (0.88–0.98)	0.88 (0.74–0.95)
Reader 2			
Peripheral zone (*n* = 10)	0.81 (0.61–0.99)	0.99 (0.97–1.00)	0.98 (0.88–1.00)
Transitional zone (*n* = 10)	0.78 (0.58–0.98)	0.99 (0.93–1.00)	0.97 (0.76–1.00)
Benign lesion (*n* = 50)	0.86 (0.73–0.99)	0.98 (0.95–0.99)	0.97 (0.93–0.99)
Malignant lesion (*n* = 50)	0.82 (0.70–0.94)	0.88 (0.72–0.95)	0.88 (0.72–0.95)

z-ADC, ADC map derived from zoomed field-of view (FOV) diffusion-weighted imaging and all the available b-values (b = 50, 1,000, and 1,500 s/mm^2); s-ADC_{b50}, ADC map synthesized using our proposed deep learning framework with input from full FOV diffusion-weighted imaging (f-DWI) (b = 50 s/mm^2); s-ADC_{b1000}, ADC map synthesized using our proposed deep learning framework with input from f-DWI (b = 1,000 s/mm^2); s-ADC_{b1500}, ADC map synthesized using our proposed deep learning framework with input from f-DWI (b = 1500 s/mm^2).

b-values improves the reproducibility and accuracy of the calculated ADC values, although it also increases the scanning time (3, 35). In contrast to traditional ADC calculation methods, our proposed method takes advantage of the ability of GAN to simulate data distribution and synthesize ADC maps that are highly similar to real zoomed FOV ADC maps that use a full FOV DWI with a single *b*-value. Considering the excellent image quality consistency and similar tumor detection and classification abilities between the s-ADC and z-ADC maps, we believe that the deep learning algorithm might be a feasible method for generating ADC maps as an alternative to z-ADC maps without requiring a strong dependence on software, hardware, and additional scan time (36).

Our study has several limitations. First, the s-ADC_{b1000} set showed the best image quality among the s-ADC sets; however, it remains unknown whether a DWI set with a *b*-value of 1,000

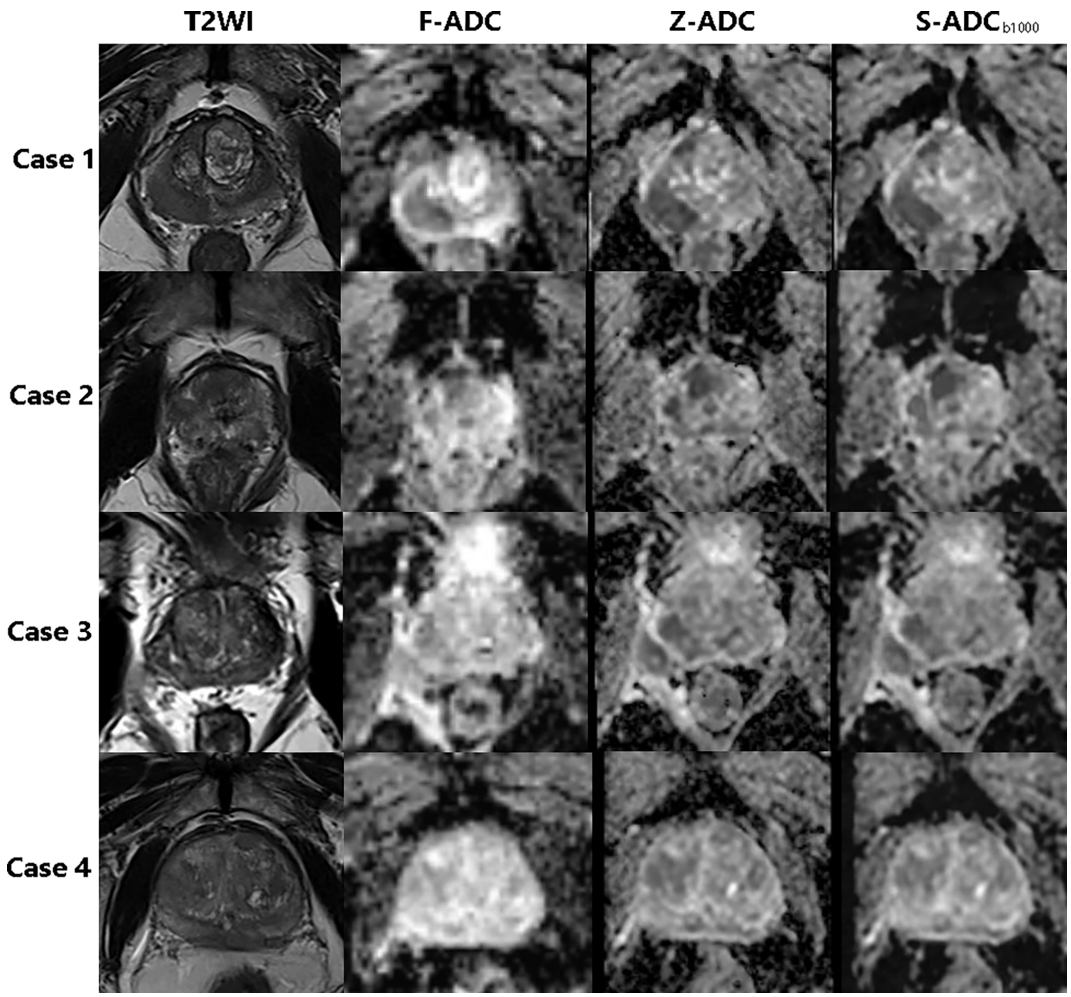


FIGURE 6 | The T2-weighted image, f-ADC, z-ADC, and s-ADC$_{b1000}$ of four different patients. Case 1: An 85-year-old man with prostate cancer in the right peripheral zone and an initial prostate-specific antigen level of 0.157 ng/ml. Case 2: An 85-year-old man with prostate cancer in the central zone and an initial prostate-specific antigen level of 21.44 ng/ml. Case 3: A 67-year-old man with an inflammatory nodule in the right peripheral zone and an initial prostate-specific antigen level of 14.37 ng/ml. Case 4: A 77-year-old man with prostate cancer in the central zone and an initial prostate-specific antigen level of 56.62 ng/ml.

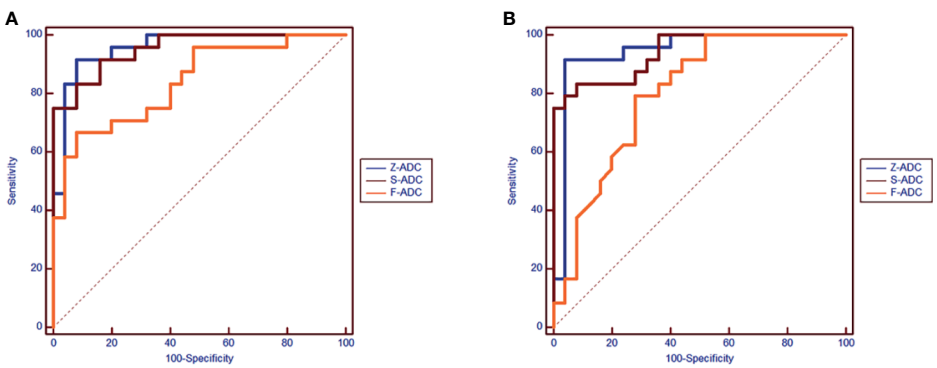


FIGURE 7 | The receiver operating characteristic comparison of the diagnostic accuracy of the f-ADC, z-ADC, and s-ADCb1000 sets (**A**: reader 1, **B**: reader 2).

TABLE 5 | The comparison of areas under the curve (AUCs) based on the f-apparent diffusion coefficient (ADC), z-ADC, and s-ADC sets for both readers.

		z-ADC *vs.* s-ADC$_{b1000}$	z-ADC *vs.* f-ADC	s-ADC $_{b1000}$ *vs.* f-ADC
Reader-1	AUC	0.96 *vs*. 0.95	0.96 *vs*. 0.84	0.95 *vs*. 0.84
	z value	0.134	2.445	2.207
	P-value	0.893	0.015	0.027
Reader-2	AUC	0.94 *vs*. 0.94	0.94 *vs*. 0.80	0.94 *vs*. 0.80
	z value	0.094	2.652	2.29
	P-value	0.925	0.008	0.022

f-ADC, ADC map derived from full field-of-view (FOV) diffusion-weighted imaging (f-DWI) and all the available b-values (b = 50, 1,000, and 1,500 s/mm^2); z-ADC, ADC map derived from zoomed FOV diffusion-weighted imaging and all the available b-values (b = 50, 1,000, and 1,500 s/mm^2); s-ADC$_{b1000}$, ADC map synthesized using our proposed deep learning framework with input from f-DWI (b = 1,000 s/mm^2).

s/mm^2 is the most appropriate for ADC map synthesis. In future studies, s-ADC sets generated using DWI sets with more potential *b*-values should be compared. Second, as ADC values vary across vendors, the generalizability of our model across MRI scanners from different vendors requires multi-center verification.

In conclusion, the GAN-based ADC synthesis method can generate s-ADC maps using a single *b*-value DWI with good image quality and high reproducibility and ADC calculation accuracy.

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 Shanghai Jiao Tong University Affiliated Sixth People's Hospital. 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

JZ is the guarantor of the integrity of the entire study (study concepts/study design, data acquisition, and data analysis/interpretation). LH participated in statistical analysis, manuscript drafting and manuscript revision for important intellectual content. LH and JZ edited the manuscript. LH and DZ participated in literature research. LH, DZ, YX, and LW participated in clinical studies. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the National Natural Science Foundation of China (nos. 81901845 and 81671791), the Science Foundation of Shanghai Jiao Tong University Affiliated Sixth People's Hospital (no. 201818), and the Shanghai Key Discipline of Medical Imaging (no. 2017ZZ02005).

REFERENCES

1. Turkbey B, Shah VP, Pang Y, Bernardo M, Xu S, Kruecker J, et al. Is Apparent Diffusion Coefficient Associated With Clinical Risk Scores for Prostate Cancers That are Visible on 3-T MR Images? *Radiology* (2011) 258(2):488–95. doi: 10.1148/radiol.10100667
2. Hambrock T, Somford DM, Huisman HJ, van Oort IM, Witjes JA, Hulsbergen-van de Kaa CA, et al. Relationship Between Apparent Diffusion Coefficients at 3.0-T MR Imaging and Gleason Grade in Peripheral Zone Prostate Cancer. *Radiology* (2011) 259(2):453–61. doi: 10.1148/radiol.11091409
3. Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, et al. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2. *Eur Urol* (2019) 76 (3):340–51. doi: 10.1016/j.eururo.2019.02.033
4. Surov A, Meyer HJ, Wienke A. Correlations Between Apparent Diffusion Coefficient and Gleason Score in Prostate Cancer: A Systematic Review. *Eur Urol Oncol* (2020) 3(4):489–97. doi: 10.1016/j.euo.2018.12.006
5. Bai K, Sun Y, Li W, Zhang L. Apparent Diffusion Coefficient in Extraprostatic Extension of Prostate Cancer: A Systematic Review and Diagnostic Meta-Analysis. *Cancer Manag Res* (2019) 11:3125–37. doi: 10.2147/CMAR.S191738
6. Nguyen TB, Ushinsky A, Yang A, Nguyentat M, Fardin S, Uchio E, et al. Utility of Quantitative Apparent Diffusion Coefficient Measurements and Normalized Apparent Diffusion Coefficient Ratios in the Diagnosis of Clinically Significant Peripheral Zone Prostate Cancer. *Br J Radiol* (2018) 91(1088):20180091. doi: 10.1259/bjr.20180091
7. Kato S, Kamijima S, Inaoka T, Kamiya N, Sasai D, Terada H, et al. Quantitative Evaluation of the Relative Apparent Diffusion Coefficient Values on Multiparametric Magnetic Resonance Imaging to Predict Higher Gleason Score Prostate Cancer. *Scand J Urol* (2018) 52(3):180–5. doi: 10.1080/21681805.2018.1481143
8. Nougaret S, Robertson N, Golia Pernicka J, Molinari N, Hötker AM, Ehdaie B, et al. The Performance of PI-RADSv2 and Quantitative Apparent Diffusion Coefficient for Predicting Confirmatory Prostate Biopsy Findings in Patients Considered for Active Surveillance of Prostate Cancer. *Abdom Radiol (NY)* (2017) 42(7):1968–74. doi: 10.1007/s00261-017-1086-7
9. Gupta RT, Kauffman CR, Garcia-Reyes K, Palmeri ML, Madden JF, Polascik TJ, et al. Apparent Diffusion Coefficient Values of the Benign Central Zone of the Prostate: Comparison With Low- and High-Grade Prostate Cancer. *AJR Am J Roentgenol* (2015) 205(2):331–6. doi: 10.2214/AJR.14.14221
10. Boesen L, Chabanova E, Løgager V, Balslev I, Thomsen HS. Apparent Diffusion Coefficient Ratio Correlates Significantly With Prostate Cancer

Gleason Score at Final Pathology. *J Magn Reson Imaging* (2015) 42(2):446–53. doi: 10.1002/jmri.24801
11. Babourina-Brooks B, Cowin GJ, Wang D. Diffusion-Weighted Imaging in the Prostate: An Apparent Diffusion Coefficient Comparison of Half-Fourier Acquisition Single-Shot Turbo Spin-Echo and Echo Planar Imaging. *Magn Reson Imaging* (2012) 30(2):189–94. doi: 10.1016/j.mri.2011.09.024
12. Nketiah G, Selnaes KM, Sandsmark E, Teruel JR, Krüger-Stokke B, Bertilsson H, et al. Geometric Distortion Correction in Prostate Diffusion-Weighted MRI and its Effect on Quantitative Apparent Diffusion Coefficient Analysis. *Magn Reson Med* (2018) 79(5):2524–32. doi: 10.1002/mrm.26899
13. Mazaheri Y, Vargas HA, Nyman G, Shukla-Dave A, Akin O, Hricak H. Diffusion-Weighted MRI of the Prostate at 3.0 T: Comparison of Endorectal Coil (ERC) MRI and Phased-Array Coil (PAC) MRI-The Impact of SNR on ADC Measurement. *Eur J Radiol* (2013) 82(10):e515–20. doi: 10.1016/j.ejrad.2013.04.041
14. Gibbs P, Pickles MD, Turnbull LW. Repeatability of Echo-Planar-Based Diffusion Measurements of the Human Prostate at 3 T. *Magn Reson Imaging* (2007) 25(10):1423–9. doi: 10.1016/j.mri.2007.03.030
15. Henkelman RM. Measurement of Signal Intensities in the Presence of Noise in MR Images. *Med Phys* (1985) 12(2):232–33. doi: 10.1118/1.595711
16. Gudbjartsson H, Patz S. The Rician Distribution of Noisy MRI Data. *Magn Reson Med* (1995) 34(6):910–14. doi: 10.1002/mrm.1910340618
17. Riffel P, Michaely HJ, Morelli JN, Pfeuffer J, Attenberger UI, Schoenberg SO, et al. Zoomed EPI-DWI of the Pancreas Using Two-Dimensional Spatially Selective Radiofrequency Excitation Pulses. *PloS One* (2014) 9(3):e89468. doi: 10.1371/journal.pone.0089468
18. Rosenkrantz AB, Chandarana H, Pfeuffer J, Triolo MJ, Shaikh MB, Mossa DJ, et al. Zoomed Echo-Planar Imaging Using Parallel Transmission: Impact on Image Quality of Diffusion-Weighted Imaging of the Prostate at 3T. *Abdom Imaging* (2015) 40(1):120–26. doi: 10.1007/s00261-014-0181-2
19. Thierfelder KM, Scherr MK, Notohamiprodjo M, Weiß J, Dietrich O, Mueller-Lisse UG, et al. Diffusion-Weighted MRI of the Prostate: Advantages of Zoomed EPI With Parallel-Transmit-Accelerated 2D-Selective Excitation Imaging. *Eur Radiol* (2014) 24(12):3233–41. doi: 10.1007/s00330-014-3347-y
20. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Nets. In: *Advances in Neural Information Processing Systems*. The MIT Press (2014)p. 2672–80.
21. Armanious K, Jiang C, Fischer M, Kustner T, Hepp T, Nikolaou K, et al. MedGAN: Medical Image Translation Using GANs. *Comput Med Imaging Graph* (2020) 79:101684. doi: 10.1016/j.compmedimag.2019.101684
22. Wolterink JM, Leiner T, Viergever MA, Isgum I. Generative Adversarial Networks for Noise Reduction in Low-Dose CT. *IEEE Trans Med Imaging* (2017) 36(12):2536–45. doi: 10.1109/TMI.2017.2708987
23. Kim KH, Do WJ, Park SH. Improving Resolution of MR Images With an Adversarial Network Incorporating Images With Different Contrast. *Med Phys* (2018) 45(7):3120–31. doi: 10.1002/mp.12945
24. Wang J, Zhao Y, Noble JH, Dawant BM. Conditional Generative Adversarial Networks for Metal Artifact Reduction in CT Images of the Ear. *Med Image Comput Dawna Assist Interv* (2018) 11070:3–11. doi: 10.1007/978-3-030-00928-1_1
25. Johnson PM, Drangova M. Conditional Generative Adversarial Network for 3D Rigid-Body Motion Correction in MRI. *Magn Reson Med* (2019) 82 (3):901–10. doi: 10.1002/mrm.27772
26. Cohen JP, Luck M, Honari S. (2018). Distribution Matching Losses can Hallucinate Features in Medical Image Translation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Medical Image Computing and Computer Assisted Intervention – MICCAI 2018, Lecture Notes in Computer Science, pp. 529–36.
27. Xu H, Huang C, Huang X, Xu C, Huang M. Combining Convolutional Neural Network Markov Random Field Semantic Image Retrieval. *Adv Multimedia* (2018) 2018:1–11.
28. Simonyan K, Zisserman A. *Very Deep Convolutional Networks for Large-Scale Image Recognition*. The 3rd International Conference on Learning Representations (2014).
29. Chen C, Liu W, Tan X, Wong KK. Semi-Supervised Learning for Face Sketch Synthesis in the Wild. In: C Jawahar, H Li, G Mori, K Schindler, editors. *Computer Vision – ACCV 2018. Lecture Notes in Computer Science*, vol. 11361. Cham: Springer (2018). doi: 10.1007/978-3-030-20887-5_14
30. Zhou F, Yao R, Liu B, Qiu G. Visual Quality Assessment for Super-Resolved Images: Database and Method. *IEEE Trans Image Process* (2019) 28(7):3528–41. doi: 10.1109/TIP.2019.2898638
31. Jin CB, Kim H, Liu M, Jung W, Joo S, Park E, et al. Deep CT to MR Synthesis Using Paired and Unpaired Data. *Sensors (Basel)* (2019) 19(10):2361. doi: 10.3390/s19102361
32. Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, et al. Medical Image Synthesis With Context-Aware Generative Adversarial Networks. *Med Image Comput Jinn Assist Interv* (2017) 10435:417–25. doi: 10.1007/978-3-319-66179-7_48
33. Bentaieb A, Hamarneh GJI. ToMI. *Adversarial Stain Transf Histopathol Image Anal* (2017) PP(99):1–1.
34. Peng Y, Jiang Y, Antic T, Sethi I, Schmid-Tannwald C, Eggener S, et al. Apparent Diffusion Coefficient for Prostate Cancer Imaging: Impact of B Values. *AJR Am J Roentgenol* (2014) 202(3):W247–53. doi: 10.2214/AJR.13.10917
35. Park SY, Kim CK, Park BK, Kwon GY. Comparison of Apparent Diffusion Coefficient Calculation Between Two-Point and Multipoint B Value Analyses in Prostate Cancer and Benign Prostate Tissue at 3 T: Preliminary Experience. *AJR Am J Roentgenol* (2014) 203(3):W287–94. doi: 10.2214/AJR.13.11818
36. Hu L, Zhou DW, Fu CX, Benkert T, Jiang CY, Li RT, et al. Advanced Zoomed Diffusion-Weighted Imaging *vs*. Full-Field-of-View Diffusion-Weighted Imaging in Prostate Cancer Detection: A Radiomic Features Study. *Eur Radiol* (2021) 31(3):1760–9. doi: 10.1007/s00330-020-07227-4

Conflict of Interest: CF was employed by Siemens Shenzhen Magnetic Resonance, Ltd. TB was employed by MR Application Predevelopment, Siemens Healthcare GmbH, Erlangen, Germany.

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.

frontiers in Oncology

METHODS
published: 17 September 2021
doi: 10.3389/fonc.2021.704850

Multi-Task Deep Supervision on Attention R2U-Net for Brain Tumor Segmentation

Shiqiang Ma[1], *Jijun Tang*[1,2,3] *and Fei Guo*[4]*

[1] School of Computer Science and Technology, College of Intelligence and Computing, Tianjin University, Tianjin, China, [2] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, [3] Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, United States, [4] School of Computer Science and Engineering, Central South University, Changsha, China

Accurate automatic medical image segmentation technology plays an important role for the diagnosis and treatment of brain tumor. However, simple deep learning models are difficult to locate the tumor area and obtain accurate segmentation boundaries. In order to solve the problems above, we propose a 2D end-to-end model of attention R2U-Net with multi-task deep supervision (MTDS). MTDS can extract rich semantic information from images, obtain accurate segmentation boundaries, and prevent overfitting problems in deep learning. Furthermore, we propose the attention pre-activation residual module (APR), which is an attention mechanism based on multi-scale fusion methods. APR is suitable for a deep learning model to help the network locate the tumor area accurately. Finally, we evaluate our proposed model on the public BraTS 2020 validation dataset which consists of 125 cases, and got a competitive brain tumor segmentation result. Compared with the state-of-the-art brain tumor segmentation methods, our method has the characteristics of a small parameter and low computational cost.

Keywords: brain tumor segmentation, attention mechanism, multi-task learning, semi-supervised learning, multi-scale feature fusion, deep supervision

OPEN ACCESS

Edited by:
Natalie Julie Serkova,
University of Colorado, United States

Reviewed by:
Shengfeng He,
South China University of Technology, China
Rupal Kapdi,
Nirma University, India

***Correspondence:**
Fei Guo
guofeieileen@163.com

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

Received: *04 May 2021*
Accepted: *26 August 2021*
Published: *17 September 2021*

Citation:
Ma S, Tang J and Guo F (2021) Multi-Task Deep Supervision on Attention R2U-Net for Brain Tumor Segmentation. Front. Oncol. 11:704850. doi: 10.3389/fonc.2021.704850

1. INTRODUCTION

Brain tumors are the most common primary malignant tumors of the brain caused by the canceration of glial cells in the brain and spinal cord. Brain tumors have the characteristics of high morbidity and mortality. Automatic segmentation technology of brain tumor can assist professional doctors to diagnose brain lesions and provide imaging technical support for the diagnosis and treatment of brain tumor patients. With the development of convolutional neural networks, the brain tumor automatic segmentation technology based on deep learning had achieved a high segmentation accuracy. However, the location of brain tumor regions and accurate segmentation of tumor edges have always been the difficulties of deep learning methods. In order to obtain accurate segmentation results, deep learning methods usually require a numerous parameters and a long calculation time, which leads to extremely high demands on the hardware. Therefore, it is of great significance to develop a simple and efficient network architecture.

Since 2015, a variety of Convolutional Neural Networks (CNN) architectures for brain tumor segmentation have been proposed. Havaei et al. proposed the InputCascadeCNN model (1), which

used cascaded CNN to segment brain tumor regions. After the network obtained a small feature map, it used two CNN branches with different convolution kernel sizes to further extract local feature and global information, and fused multi-scale information. Dvorak et al. proposed a 6-layer CNN, the brain image was cropped into multiple patches, and these patches were clustered using k-means to obtain N clustering results and formed a dictionary as the input of network (2). Pereira et al. used a 3X3 convolution kernel to extract the segmentation features (3), like VGG (4). When the receptive field of the same size was obtained, a smaller convolution kernel could effectively reduce the amount of network parameters and enabled the network to be designed deeper. At the same time, the author used intensity normalization in the data preprocessing process. Kamnitsas et al. proposed DeepMedic (5), using residual block (6) in the CNN architecture. DeepMedic used images of different resolutions as the input of two branch networks to obtain multi-scale information and fused the multi-scale information. Randhawa et al. (7) used a classification network to classify each input pixel. Kamnitsas et al. proposed EMMA (8), which merged the outputs of multiple independent networks through an average confidence.

Although a variety of network structures have been proposed, the location of tumor regions and accurate segmentation of tumor boundaries have always been the difficulties of brain tumor segmentation. The traditional deep learning method usually used the fully connected layer as the last layer of the network, but one-dimensional probability information will lose the spatial structure information of the image, which is not suitable for image segmentation. Fully convolutional neural networks (FCN) (9) and U-Net (10) used a fully convolutional layer as the last layer of network, and used an up-sampling operation that is symmetrical to down-sampling to keep the size of the feature map consistent with the input size of the network. This method effectively improves the ability of neural network to locate the region of interest (ROI). However, the shape and pixel intensity of brain tumor data are affected by differences between patients and data collection agencies, which makes it difficult for traditional U-Net and FCN to obtain accurate location and segmentation accuracy when the number of parameters is small.

In order to further improve the performance of the U-Net architecture, a variety of improved U-Net architectures have been proposed. DCSNN (11) extends the architecture of U-Net with a residual module by adding a symmetric mask in multiple layers. Isensee et al. proposed an improved U-Net architecture (12), which used the pre-activation residual block (13) as the basic unit of network. At the same time, the leaky rectified linear unit (leaky ReLU) was used to prevent the gradient from disappearing, and batch normalization (14) was replaced with instance normalization (15), which improved the stability of the network for a feature extraction of small batches. nnU-Net (16) used 2D U-Net, 3D U-Net, and cascaded 3D U-Net to adaptively segment inputs of different resolutions. Although most of the improved u-net methods improve the segmentation accuracy, they also increase the depth, parameters, and computing time of deep learning network.

The depth of the network and the size of the parameters will directly affect the ability of feature extraction, usually a deeper network structure and larger parameters will improve the segmentation accuracy. However, the increase of parameters will lead to an over fitting problem and reduce the robustness of the network. Too deep network structure will lead to the problem of vanishing gradient and exploding gradient in network training. In order to solve the vanishing gradient problem and exploding gradient problem of the deep network, deep supervision methods were introduced (17–19). In theory, when the size of convolution kernel remains the same, as the number of network layers becomes deeper, the network gained a stronger nonlinear expression capability. However, with the deepening of the network, backpropagation becomes difficult, resulting in a decrease in network performance. Chen et al. proposed VoxResNet, which was used in brain segmentation. In order to solve the problem of automatic segmentation caused by the difference in the shape of 3D image slices, the author merged the deep supervision results containing multi-level context information as the final output of network (20). Zeng et al. used a multi-level deep supervision of 3D U-Net to alleviate the potential gradient vanishing problem in a Proximal femur segmentation (21). Zhang et al. used deep supervision in a retinal vessel segmentation to learn a better semantically representation and help convergence (22). Zeng et al. proposed a multi-scale deep supervision method in infant brain MR image segmentation, which addresses that the final loss cannot supervise a shallow fracture extraction (23).

Similarly, a deep supervision method was also used in the brain tumor segmentation (12). Deep supervision usually used the same label to perform a single task, mainly focusing on solving the problem of gradient vanishing. When Resnet was proposed, the problem of gradient vanishing was effectively improved. Andriy Myronenko proposed a multi-task learning method (24), which used U-Net to perform brain tumor segmentation tasks and used another decoder branch for image reconstruction. This method was similar to a deep supervision, replacing the label of a decoder branch with a reconstruction label, thereby preventing the problem of network overfitting. Similarly, Chen et al. proposed the Multi-task Attention-based Semi-Supervised Learning (MASSL) framework, which used soft segmentation to obtain pseudo-labels of tumor and non-tumor regions, and used pseudo-labels to supervise the reconstruction branch (25). They proposed that multi-task learning could improve the capture of segmentation features in the encoder part. Jiang et al. used two decoder branches with different up-sampling structures to help the encoder part to collect more abundant brain tumor regional features (26). Weninger et al. used the three tasks of segmentation, classification, and reconstruction to jointly train the shared encoder part (27). The methods above used other related tasks as labels for deep supervision, and obtained accurate brain tumor segmentation results. It showed that the deep supervision method could not only improve the vanishing gradient problem of deep network, but also enabled the network to learn a richer visual representation and prevented overfitting.

In the brain MRI image of the patient, the brain tumor area is small, so the brain tumor segmentation has a problem of class imbalance. In order to focus on the brain tumor area, the visual attention mechanism was introduced into the medical image segmentation network. Hu et al. used the global max-pooling layer to adaptively calculate the weight of each channel, and feed the weight back to the feature channel (28). On this basis, Li et al. designed a dynamic selection mechanism for the convolution kernel based on the working principle of visual neuron, and adaptively adjusted the receptive field size obtained by the convolution kernel through multi-scale information, and used softmax to Features of different sizes are merged (29). Woo et al. used the channel attention module and spatial attention module to adaptively select the beneficial channel features and spatial features, and used element-wise summation and sigmoid activation function to fuse the two features (30).

In this paper, we proposed a new end-to-end brain tumor segmentation network. We made partial modifications to the Attention U-Net (31) framework and design MTDS and APR module. Our work aims to enhance the ability of network to capture the features of brain tumor and reduce the impact of class imbalance, and improve the accuracy of brain tumor segmentation.

2. METHODS

The detailed description of our proposed automatic brain tumor segmentation method will be given in this section. The proposed deep learning model architecture is presented, including the UNet-like basic network, APR module, and MTDS.

2.1 Basic Network

The design of the model needs to consider the distribution characteristics of the dataset. Compared with natural images, medical images are symmetrical and have a simpler semantic information and a more fixed image structure. However, medical images often contain noise and artifacts, and the boundary information is blurred. In the view of a single structure and the fuzzy boundary of medical images, the autoencoder structure with skip connection has become the benchmark for brain tumor segmentation. The structure of convolutional autoencoder can reduce the amount of network parameters while obtaining high-level semantic features, saving computing resources. Skip connection combines low-level and high-level features to help the network reconstruct the detailed information of ROI. Our basic network is similar to Attention U-Net. In order to obtain a higher tumor segmentation accuracy, we adjusted the structure of the network.

The model structure is shown in **Figure 1**, similar to LinkNet (32), we combined the U-Net structure and the ResNet structure. According to the statement in (33), the skip connection of U-Net cannot eliminate the vanishing gradient problem, but the shortcut of ResNet can prevent the vanishing gradient problem. In addition, the skip connection of U-Net helps to increase the convergence speed the same as the shortcut of ResNet. The main structure includes encoder, decoder, and deep supervision. Encoder consists of 3 down-sampling, 4 APR module, and 4 Squeeze-and-excitation (SE) modules. For the first Residual Units of the encoder part, the number of convolution kernel is 32, and doubles with each next residual unit. Decoder includes 3 up-sampling, 3 pre-activation convolution blocks, 3 SE modules, 1 convolutional layer (1x1), and 1 sigmoid. In the SE module, some channels are considered to have no important contribution to the segmentation task, and their weights are very small, which leads to overfitting and vanishing gradients problem. Therefore, we added the dropout layer to prevent the network from overfitting and improve the robustness of the deep learning network. The random change of channel weight helps the network learn the visual expression of different channel features in brain tumor segmentation. The experimental results also prove this conclusion. The SE module is shown in **Figure 2**, and **Table 1** reports the results of comparative experiments with or without dropout in the SE module.

2.2 Multi-Task Deep Supervision

In the brain tumor automatic segmentation model, we use the MTDS method to optimize the training process of deep learning network and extract richer visual features. In the process of back propagation, the deep network converges slowly or even hard to converge due to the problem of vanishing gradient. Deep supervision techniques are used to alleviate the training difficulty of deep networks. However, unreasonable network design affects the hierarchical feature expression ability of the network, and even disrupt the network optimization goal. Usually, the shallow layers of the network extract low-level features in the image, such as boundary information. The deep layers of the network can extract high-level features, in other words, the semantic information of an image. When deep supervision is designed in the front of the network, it forces the network to change the normal learning process, resulting in an inconsistent loss of optimization goals and affecting the segmentation accuracy. This impact became more serious in many deep networks (34).

Based on the problems above, we use the ground truth of multiple segmentation tasks as the label for deep supervision, and optimize the training process through multiple associated sub-segmentation tasks. While solving the vanishing gradient problem, the ability of the network to extract segmentation features of a sub-tumor region is improved. The comparison between our proposed deep supervision method and other methods is shown in **Figure 3**. The sub-segmentation task is used as the regularization item of the network to improve the generalization ability of the model and prevent overfitting. Normally, whole tumors consist of the peritumoral edema, enhancing tumor, and the necrotic and the non-enhancing tumor. The area of enhancing tumor is smaller than the area of peritumoral edema and the necrotic and the non-enhancing tumor. High-level semantic information is not conducive to capturing the features of the enhancing tumor area, while low-level boundary information can better express the detailed features of the enhancing tumor. In our method, the enhancing tumor ground truth is used as

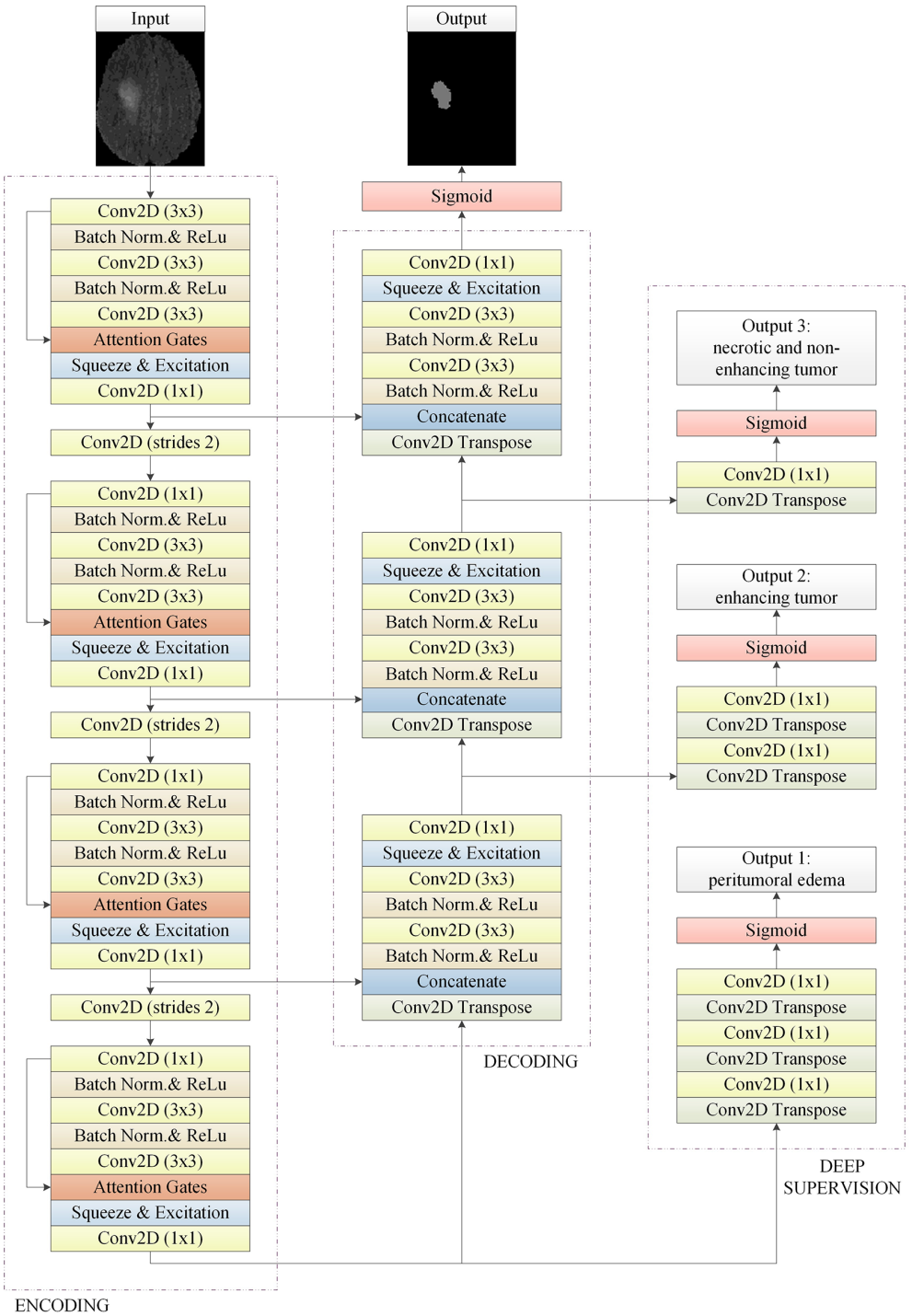


FIGURE 1 | The basic 2D convolutional neural network for brain tumor segmentation. It consists of encoding, decoding, and deep supervision. Our approach is an end-to-end network, the input of the network is a 2D image composed of four modes, and the output is the whole brain tumor prediction result of each 2D image. Output1, output2, and output3 are the subregions of the brain tumors, which are the peritumoral edema, enhancing tumor, and the necrotic and the non-enhancing tumor, respectively. Multi-task deep supervision with progressive relationships can help our method accurately extract the visual features of each stage.

the label of first deep supervision, and the shallow layers of network can better capture the boundary details of the enhancing tumor area. Segmentation of the necrotic area and segmentation of the peritumoral edema area are respectively used as the other two deep supervision tasks, and the final output of the network is the segmentation of the whole tumor area. The optimization objective of whole brain tumor segmentation and multi-task auxiliary segmentation can be expressed as follows:

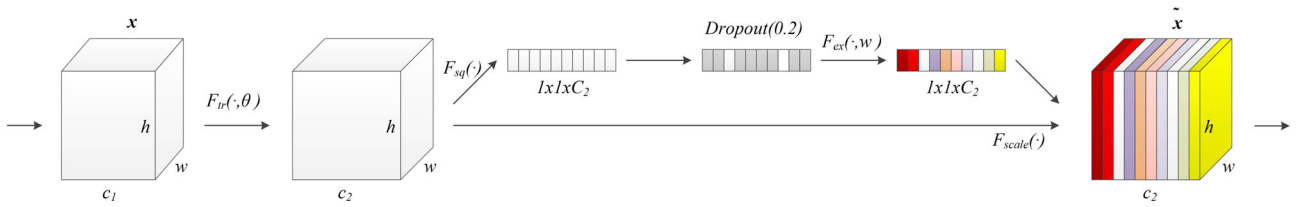


FIGURE 2 | Our proposed SE module with the dropout layer. Adding the dropout layer can prevent overfitting and improve the robustness of the deep learning network. The SE module assigns different weights to the feature channels to help the network obtain the most effective features of the brain tumor regions.

$$\arg\max_{\omega_m,\omega_\alpha} L_m(\omega_m; D) + L_\alpha(\omega_\alpha; D) \tag{1}$$

where D is the brain tumor datasets with annotation, ω_m is the learnable weight matrices of whole brain tumor segmentation network, and ω_a correspond to the learnable weight matrices of multi-task auxiliary segmentation network. L_m denotes the total loss function of whole brain tumor segmentation, and $L\alpha$ is the loss function of multi-task auxiliary segmentation.

2.3 Attention Pre-Activation Residual Module

In addition to the function of identity mapping, residual module is a simple multi-scale feature fusion method (36). Multi scale feature representation is very important for image segmentation. Except to the pixel intensity, the morphological features of the tumor region are of great importance for brain tumor segmentation. Learning the difference between the morphological features of brain tumor and the surrounding normal brain tissue by deep convolution network is helpful to the accurate segmentation of the brain tumor region. The combination of the boundary information of the tumor region and its high-level semantic information can make the deep convolution network accurately locate ROI (31). Based on the residual module, the improved multi-scale information fusion of deep convolution network is beneficial to the classification, segmentation, and detection of visual tasks.

Therefore, Res2Net (37) and other network structures are proposed. Res2Net designed a residual structure, which can significantly increase the multi-scale information of the residual module. However, the feature fusion of Res2Net is simple, so that it is difficult to make full use of the multi-scale information. On this basis, we propose an APR module, which is used to improve the attention of the deep network to ROI. This structure combines the pre-activation residual units (13) and attention gates (AGs) (31). The APR module can be seen in **Figure 4**. Thanks to the excellent performance of the pre-activation residual units in the field of medical image segmentation (24, 26, 33, 38), we use the pre-activation residual units as the basic module of the segmentation network. Pre-activation residual units can help information propagation, which include 2 batch normalization, 2 rectified linear unit (ReLU), and 2 weight layers. The output x_{l+1} of the pre-activation residual units can be expressed as follows:

$$x_{l+1} = x_l + F(x_l, \omega_l) \tag{2}$$

$$F(x_l, \omega_l) = F_r(F_r(x_l, \omega_l)) \tag{3}$$

$$F_r(x_l, \omega_l) = W_x(\sigma_1(W_b(x_l, \omega_l))) \tag{4}$$

where x_l is the input of the pre-activation residual units, w_l is the learnable weight matrices. $F(x_l,w_1)$ denotes the pre-activation residual function, $F(x_l,w_1)$ consists of two cascaded subunits $F_r(x_l,w_1)$. An element-wise addition is used to combine the feature map of x_l and $F(x_l,w_1)$. Each $F_r(x_l,w_1)$ includes a batch normalization ω_b, a ReLU σ_1, and a 3X3 convolutional layer W_x. The 3X3 convolution layer enables the pre-activation residual function to obtain a larger receptive field than the input, which provides multi-scale visual information for the feature fusion of the attention gates.

Attention gates, which is like the shortcut-only gating and 1x1 convolutional shortcut (13), have a stronger visual representational ability. Attention gates consists of a ReLU, 1x1 convolutional layer, and a sigmoid activation function. ROI is selected by analyzing both the activations and contextual information. The output y_{l+1} of attention gates can be expressed as follows:

$$y_{l+1} = y_l \dot{o} F_\alpha(y_l) \tag{5}$$

$$F_\alpha(y_l) = \sigma_2(W_y(\sigma_1(y_l))) \tag{6}$$

where y_l is the input of attention gates, which is the output of the pre-activation residual units ($y_l = x_{l+1}$). $\dot{o}$ is the element-wise

TABLE 1 | The results of comparative experiments with or without dropout in the SE module on the BraTS 2020.

Method	DSC (%)	Sensitivity (%)	Specificity (%)	Hausdroff95
without dropout	88.59	88.52	99.86	7.74
with dropout	**89.18**	**89.24**	**99.91**	**5.77**

The bold values indicate the best results.

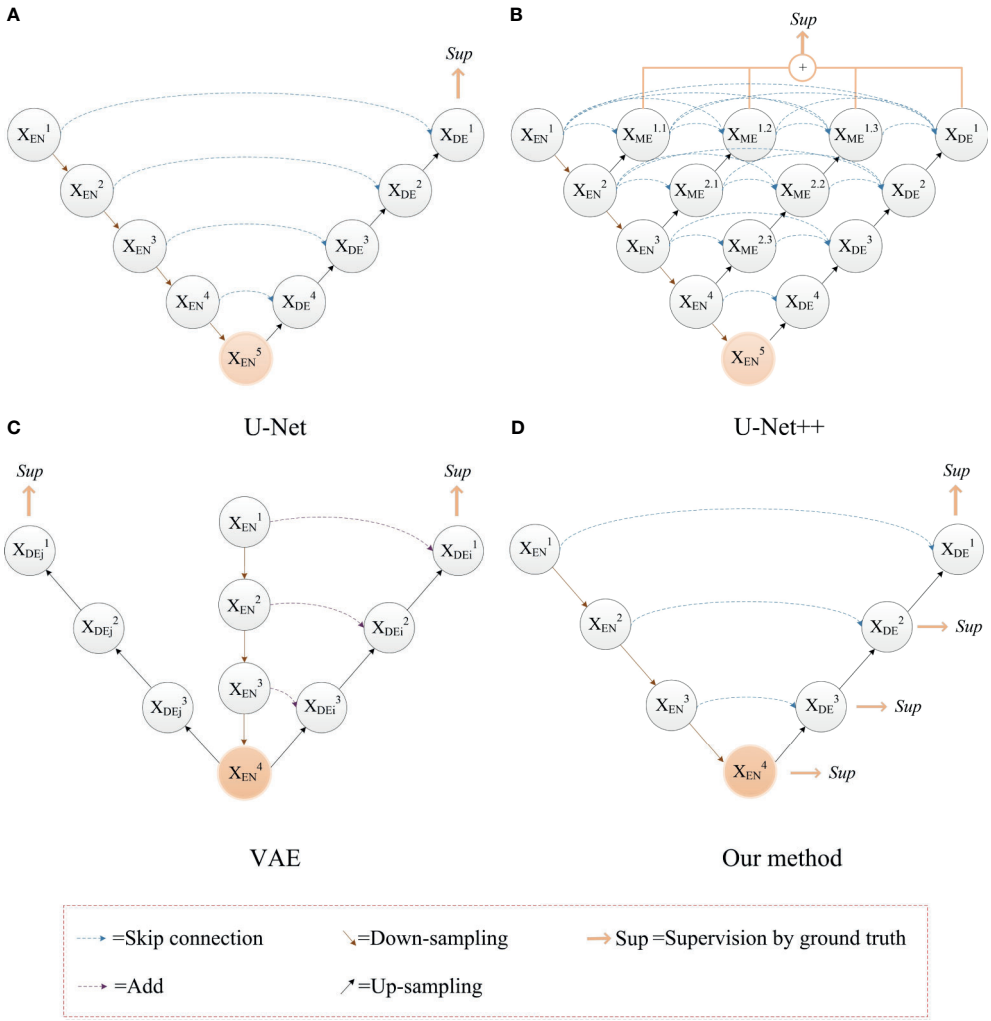


FIGURE 3 | The comparison between our proposed deep supervision method and other methods. **(A)** The U-Net model; **(B)** Use of multiple shortcuts and skip connections: this method adds a deep supervision method to each level of sub network, which affect the hierarchical feature expression ability of network (35); **(C)** Use of image reconstruction task as deep supervision to prevent the network from overfitting (24). **(D)** Our method with deep supervision.

multiplication. $F\alpha(y_1)$ denotes the attention gates function. W_y is a 1x1 convolutional layer used to compute linear transformation. $\sigma_2 = \frac{1}{(1+exp(-yl))}$ is a sigmoid activation function. ReLU and sigmoid can improve the nonlinear expression ability of the attention gates. In addition, sigmoid can make attention gates parameters have a better convergence.

We combine the pre-activation residual unit and attention gates, and obtain the APR module as follows:

$$y_{l+1} = (x_l + F_r(F_r(x_l, \omega_l))) \dot{o} \sigma_2(W_y(\sigma_1(x_l + F_r(F_r(x_l, \omega_l))))) \quad (7)$$

APR module is a multi-scale feature fusion method based on the residual unit. This method obtains multi-scale information from the residual units and generates a gating signal to control the importance of features in different spatial regions, to suppress the feature response of irrelevant background regions.

3 EXPERIMENTS AND RESULTS

In this section, the brain tumor datasets and the pre-processing methods are introduced. And then, we provide the training details of network, including the loss function and optimizer. Post-processing methods for brain tumor segmentation are also introduced. Finally, we introduce the evaluation criteria for the brain tumor segmentation task, and report the results consisting of the ablation experiment and comparison with the state-of-the-art methods.

3.1 Brain Tumor Dataset and Pre-Processing

In this section, we present the details of experimental data, it includes brain tumor datasets, data preprocessing and data augmentation.

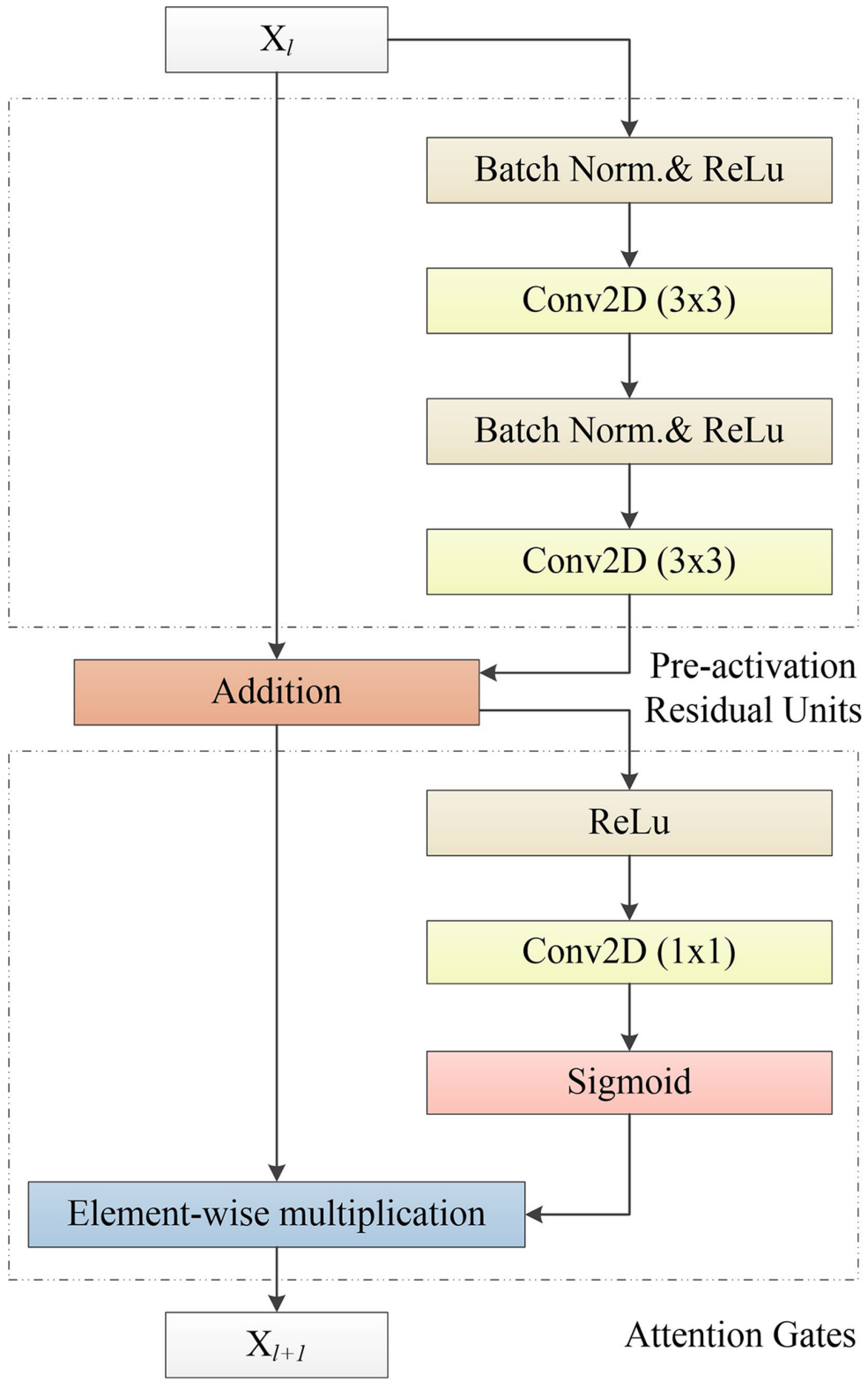


FIGURE 4 | Our proposed APR module, which consists of the Pre-activation Residual Units and Attention Gates. Pre-activation Residual Units obtain feature maps of low-level and high-level scales. Attention Gates obtains the weighted feature map of the 2D image by performing nonlinear processing on the output result of the Pre-activation Residual Units.

3.1.1 Brain Tumor Datasets

The brain image dataset is provided by MICCAI Multimodal Brain Tumor Segmentation Challenge (BraTS) (39, 40). Each sample of the patient includes four modalities. The brain tumor datasets were collected from 19 institutions with the same resolution of 1 mm^3, and were unified to the same anatomical template. The size of each modality was 240x240x155. All BraTS multimodal datasets include four modals, which are native (T1), post-contrast T1 weighted (T1Gd), T2 weighted (T2), and T2 fluid attenuation inversion recovery (T2-FLAIR). **Table 2** summarizes the dataset of BraTS 2017-2020. The training datasets of BraTS 2018-2020 are used to train our network.

3.1.2 Pre-Processing

Due to different data collection agencies, there are differences in the pixel intensity. In order to make the deep learning network learn more uniform and the segmentation features more accurate, it is necessary to use image pre-processing methods to standardize the data.

In the dataset provided by BraTS 2020, the brain area occupies less than 50% of the total area. A large background area increases the proportion of negative samples, making it difficult for deep learning networks to effectively learn brain tumor features (16). In addition, more tumor pixels are incorrectly classified as background. Different from (41, 42), which crops images into small patches, we crop each image to a size of 144x176 to preserve as much brain region information as possible and reduce the interference of background regions. Specifically, we keep the center area of each image and cropped the edge area. Maximizing the preservation of brain information in non-tumor areas is beneficial for the network to better learn to distinguish the difference between tumor and normal brain tissue. After cropping the image, we use min-max normalization (43) to process the image to reduce the difference between the data collected by different institutions. Specifically, we calculated the maximum and minimum pixel intensity of the 3D brain data of each brain tumor patient in a single modality, and normalized the value range of each pixel to 0 and 1 through min-max normalization between. Performing min-max normalization on a single modality of each sample can not only reduce the difference between scans from various institutions, but also avoid the difference of various scans from the same institution. In addition, normalizing the pixel value between 0 and 1 facilitates the back propagation of gradient during the training process.

3.1.3. Data Augmentation

In order to solve the problem of less training data, we also carried out data augmentation operations. Data augmentation can effectively increase the sample size and prevent the model from overfitting. Commonly used data augmentation methods include flipping (44), transposing, and rotating (45). In order to ensure that the pixel intensity of data does not change significantly and to make the network robust to the shape of tumor, we use the data augmentation strategy of flipping. This strategy can enable the deep learning network to learn the shape characteristics of brain tumors, and use the shape information of brain tumors and non-tumor regions to help the network distinguish tumor regions with similar pixel intensity from normal brain tissue regions.

3.2 Loss Function

In the brain tumor images, the proportion of the lesion area is small, in other words, the foreground area is much smaller than the background area. Class imbalance makes it difficult for some commonly used segmentation loss functions to train network parameters effectively. In order to reduce the impact of class imbalance on network training, the network is trained with a combination of dice loss (42) and cross-entropy loss. The joint loss combining dice loss and cross-entropy loss is proven to have an excellent performance in medical image segmentation tasks (46).

Dice loss is a similarity measure method, which is widely used in medical image segmentation, and its value range is [0, 1]. Dice loss can be expressed as follows:

$$L_{dice} = \frac{2\Sigma_{i=1}^{Z} p_i q_i}{\Sigma_{i=1}^{Z} p_i^2 + \Sigma_{i=1}^{Z} \quad q_1^2}, \tag{8}$$

where Z denotes the sums of voxels, $p_i \in P$ is the predicted binary segmentation volume, and $q_i \in Q$ is the ground truth of segmentation volume.

Dice loss focuses on the segmentation results of the foreground regions, so it can improve the impact of class imbalance. But when the foreground area in the image is too small, the predicted segmentation result has a greater impact on the calculation result of loss function, making the training unstable. Therefore, we combine dice loss and cross-entropy to improve the training stability. The loss function of brain tumor segmentation network without deep supervision can be expressed as follows:

$$L_m(\omega_m; D) = 1 - \frac{1}{N}\Sigma_{i=1}^{N} \log f_m(\omega_m; x_i)^{(yi)} - \frac{2\Sigma_{i=1}^{N} y_i \widehat{y}_i}{\Sigma_{i=1}^{N} y_i^2 + \Sigma_{i=1}^{N} \widehat{y}_i^2}, \tag{9}$$

where the brain tumor dataset D including N examples, x_i is the i^{th} image of brain MRI scans, and y_i is the ground truth

TABLE 2 | Summary of the BraTS challenge dataset from 2017 to 2020.

Dataset	Training	Validation	Testing
BraTS 2017	285	46	146
BraTS 2018	285	66	191
BraTS 2019	335	125	166
BraTS 2020	369	125	166

corresponding to x_i. $\hat{y}_i$ denotes the predicted binary segmentation result corresponding to x_i.

3.3 Implementation Details

Our framework was constructed using the TensorFlow2 (47) libraries. The GPU used in the experiment is a virtualized NVIDIA Tesla V100 with only 16 GB of memory. Its computing performance is a quarter of that of a physical GPU. For the training of our method, the total number of epochs is set to 50 and the batch size is set to 32. Adam optimizer (48) is used to optimize the training for all experiments. Adam optimizer, combining the advantages of the AdaGrad and RMSProp optimization algorithms, comprehensively considers the first moment estimation (First Moment Estimation, the mean value of gradient) and the second moment estimation (Second Moment Estimation, the uncentered variance of gradient), and calculate the update step size. The update of parameters of the Adam optimizer is not affected by the scaling transformation of the gradient. It is suitable for the unstable objective function and problems with sparse gradients or very noisy gradients. In our method, the initial learning rate of the Adam optimizer is $1e^{-4}$, the algorithm of learning rate decay is like as (24).

3.4 Post-Processing

In order to further improve the accuracy of the brain tumor segmentation results, we performed post-processing operations on the output of the network. Commonly used post-processing methods for image segmentation include thresholding, erosion, dilation, open operations, close operations, and CRF. For brain tumor segmentation tasks, the pixel intensity and the morphology features of some brain tissues in the brain image are similar to the tumor area, it is easy to interfere with the segmentation of the tumor area, resulting in false positives segmentation results. Through observation, the normal area that is misclassified as a tumor is usually small. In order to reduce the influence of false positives on the segmentation accuracy, we concatenate all the 2D segmentation results of each patient into 3D voxels. And then, we calculate the volume of each independent predicted brain tumor area in each 3D voxel and eliminate the smaller predicted tumor. We keep the largest predicted tumor in each patient and use its volume as the baseline. Then, we compare the volume of other predicted tumors with the baseline. When the volume of other predicted tumors is less than one-tenth of the baseline, we determine that these predicted tumors are false positives.

3.5 Evaluation Metrics

In order to evaluate the segmentation performance of brain tumors more comprehensively, dice similarity coefficient (DSC), sensitivity, specificity, and hausdorff distance (HD) are used as evaluation metrics. All evaluation metrics can be expressed as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN}, \tag{10}$$

$$\text{Specificity} = \frac{TN}{TN + FP}, \tag{11}$$

$$\text{DSC} = \frac{2|U \cap V|}{|U| \cup |V|}, \tag{12}$$

$$\text{Hausdorff} = \max\left\{ \max_{u \in U} \min_{v \in V}(u, v), \max_{v \in V} \min_{u \in U}(u, v) \right\}, \tag{13}$$

where true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*) are usually used to calculate the evaluation metrics in the segmentation methods. Higher values of sensitivity indicate that the larger tumor area is segmented correctly. Higher values of specificity indicate that the larger non-tumor area is segmented correctly. U and V indicate the ground truth of the lesion area and the prediction of network, respectively. Higher values of DSC indicate that the segmentation of the lesion area is more accurate. u and v indicate the set of points on the boundary of ground truth U and the set of points on the boundary of prediction V, respectively. Lower values of Hausdorff distance indicate that the segmentation of the lesion area is more accurate. In this paper, we use Hausdorff95, which is based on the calculation of the 95th percentile of distances between the boundary points in the ground truth and prediction. Due to the presence of outliers in the boundary area, hausdorff95 can avoid the interference of outliers on the segmentation performance.

3.6 Evaluation on Model Architecture

We present a detailed study of the proposed network on the MICCAI Multimodal Brain Tumor Segmentation Challenge 2020 in this section. The training dataset provided by BraTS 2020 is used to train the network. In order to evaluate the segmentation performance of our method more objectively, we upload the predicted results of the validation dataset to the Image Processing Portal (IPP) of CBICA's.

Similar to the training dataset, the validation dataset also includes four modal brain MRI scans. The validation dataset consists of a total of 125 brain data of patients, and for the axial axis, each brain MRI scans of the patient consisted of 155 images with a size of 240x240. The validation dataset contains mixed glioblastoma (GBM/HGG) and lower grade glioma (LGG). In order to match the trained network input, we use the same cropping method as the training dataset to reduce the image size of each validation dataset to 144x176. After obtaining the prediction results, we restore each image to its original size and submit it to the online evaluation system.

3.6.1 Study of Attention Pre-Activation Residual Module

APR module is modular so that it can be easily added to the segmentation structures. In our proposed model, the APR module is used in the encoder part to improve the ability of extracting tumor features. Three structures are designed to compare with the APR module. The first structure does not use the shortcut and attention gates. The second structure adds

the shortcut, but there are no attention gates. The third structure uses the shortcut, and the use of the attention gates is consistent with (31), in other words, combine attention gates with the skip connections.

In **Table 3**, we report the results of the comparative experiment. The results on whole brain tumor predictions demonstrate that the APR module has achieved the first place in three evaluation metrics of dice similarity coefficient, sensitivity, and Hausdorff distance. Due to the large proportion of negative samples, the specificity scores of the four structures are very similar. In addition, the structure of the Attention U-Net has a better segmentation performance for brain tumors, which also proves that the attention gates are helpful for the fusion of multi-scale features. However, for brain tumor segmentation tasks, too large feature scale differences cannot make attention gates accurately weight ROI. This result proves that the APR module contributes to brain tumor segmentation tasks.

3.6.2 Study of Multi-Task Deep Supervision

MTDS is used to extract richer visual features. It can be applied to multi-label segmentation tasks similar to brain tumor segmentation. We design three comparative structures. The first structure does not use deep supervision. The second structure adds deep supervision, but only uses the whole brain tumor mask as the label for all branches. The third structure uses MTDS, and uses enhancing tumor, the necrotic and the non-enhancing tumor, and peritumoral edema as the labels of the three branches, respectively.

Table 4 shows the comparison experiment results of MTDS and the other two structures. The structure with the MTDS strategy has achieved the top rank in all evaluation metrics. Through the comparative experiments, we can find an interesting phenomenon. The segmentation results of structure without deep supervision are better than the structure with single-task deep supervision in the evaluation metrics of DSC, Sensitivity, and Hausdorff95. Although deep supervision techniques can alleviate the difficulty of optimization arising from gradient flow, it interferes with the hierarchical representation generation process. Due to the inconsistency of optimization objectives, the positive optimization effect on the shared shallow parameters is small, which reduces the accuracy of brain tumor segmentation.

3.7 Comparison with State-of-the-Art Methods

Our proposed model is evaluated on the public BraTS 2020 validation dataset to compare its performance with the state-of-the-art methods which are on the BraTS2017, BraTS2018, and BraTS2019 leader board. The results of our method comparison with the state-of-the-art methods are reported in **Table 5**.

Most state-of-the-art methods ensemble the segmentation results of multiple models, and the segmentation results of ensemble of multiple models is usually better than a single one. In order to show the performance of our proposed method more visual, we did not use the ensemble of multiple models, but only used the proposed single model to compare with other methods. For the whole brain tumor segmentation task, the Dice score of whole tumors reached 0.86-0.90, the Sensitivity score of whole tumors reached 0.85-0.92. Specificity scores of all methods are very high, almost over 0.99. The Hausdorff distance is basically between 4 and 7. The experimental results show that our method has a strong competitiveness.

In order to make the comparison result more objective, we retrain several state-of-the-art segmentation models to the brain tumor dataset and evaluated them on the BraTS2020 dataset. It can be seen from **Table 6** that our method has achieved the first place in the DSC, Sensitivity, Specificity, and Hausdorff distance. At the same time, our method has the least number of parameters. **Figure 5** shows a more intuitive comparison between the segmentation results of our method and state-of-the-art methods.

4. DISCUSSION AND CONCLUSION

Brain tumor is a disease that threatens human health. Manual segmentation is time-consuming and subjective. The difficulties

TABLE 3 | Ablation experiment of the APR module without multi-task deep supervision on the BraTS 2020.

Method	DSC (%)	Sensitivity (%)	Specificity (%)	Hausdroff95
without shortcut & AGs	88.57	88.56	99.89	7.03
without AGs	88.69	88.01	**99.90**	6.99
Attention UNet	88.89	89.25	99.88	6.78
APR module	**88.95**	**89.56**	99.89	**6.51**

The bold values indicate the best results.

TABLE 4 | Ablation experiment of deep supervision on the BraTS 2020.

Method	DSC (%)	Sensitivity (%)	Specificity (%)	Hausdroff95
without deep supervision	88.95	89.56	99.89	6.51
single-task deep supervision	88.74	87.95	99.90	6.84
multi-task deep supervision	**89.18**	**89.24**	**99.91**	**5.77**

The bold values indicate the best results.

TABLE 5 | The results of comparison between our proposed method and state-of-the-art methods.

Method	Dataset	DSC (%)	Sensitivity (%)	Specificity (%)	Hausdroff95
BCVUniandes (49)	2017	86.8	84.2	99.5	18.456
BRATZZ27 (50)	2017	88.0	85.6	99.6	5.72
CISA (50)	2017	87.3	85.4	99.4	5.18
CMR (50)	2017	85.6	81.1	99.6	5.87
MIC_DKFZ (50)	2017	90.2	90.1	99.5	6.77
Zhouch (50)	2017	**90.3**	90.3	99.5	**4.74**
RadCNN (51)	2017	89.0	89.1	99.5	6.53
Radiomics-miu (52)	2018	87.6	86.2	99.5	4.90
GBMNet (50)	2018	88.3	**93.4**	98.9	5.46
Mmonteiro2 (50)	2018	87.0	87.4	99.3	5.79
UNetImage (50)	2018	89.9	91.0	99.4	5.10
RA-UNet (50)	2018	89.1	89.4	99.3	5.87
Voxel-GAN (53)	2018	84.0	86.0	99.0	6.41
S3D-Unet (41)	2018	88.7	90.1	99.4	5.51
3D Dense U-Nets (54)	2018	88.9	88.0	98.0	7.27
3D Attention UNet (55)	2019	89.8	90.0	99.4	6.29
MECU-Net (56)	2019	90.2	90.8	99.5	5.41
Multi-step cascaded network (57)	2019	88.6	92.1	99.2	6.23
3D U-Net (58)	2019	89.4	89.7	99.5	5.68
Our method	2020	89.2	89.2	**99.9**	5.77

The bold values indicate the best results.

of the automatic brain tumor segmentation technology include the sensitivity of the algorithm to tumor regions and the suppression of response to non-tumor regions. In order to improve the ability of the convolutional neural networks to locate ROI, we propose the APR module. This module uses the residual units and attention gates to construct a multi-scale feature fusion method. The simple fusion of low-level feature and the high-level feature of residual unit pass the features of non-interest region to the deeper layers of network. It interferes with the extraction of important information about brain tumors from the encoder part. The attention gate added in the residual unit focus attention on the tumor area, reduced the response of non-interest areas, thereby improving the ability of the convolutional neural network to locate the area of interest. This method has proved its superiority in brain tumor segmentation experiments.

In order to improve the utilization of multi-modal information in brain tumor segmentation tasks, we propose a MTDS method. Different modalities have different sensitivities to the tumor area. In order to fully explore the potential information of multimodal data, we have designed multiple branches in the network, and each branch is used to complete a specific task. In order to avoid the chaotic design from interfering with the ability of the network to extract tumor features, we designed a MTDS method for the characteristics of different tumor regions. In addition, MTDS helps the network to extract richer semantic features and alleviate the problem of network overfitting. We also tested its performance on the brain tumor segmentation task, and the results of experiment proved our hypothesis. The experimental results show that our model has a generalization ability and extension possibilities.

In this paper, we focus on the segmentation accuracy and robustness of a single network to the target region. We hope to design a simple and easy-to-use 2D segmentation method to reduce the dependence of network training on the hardware and reduce training time. Due to the few network parameters, our proposed method is not as good as some segmentation results that integrate multiple 3D networks. In future work, we will continue to focus on the improvement of the current method to make it smaller and more flexible, and at the same time have a higher segmentation accuracy. In order to achieve this goal, we will improve the currently proposed attention mechanism to enable it to integrate richer multi-scale features. In addition, we will make the architecture much more general to other medical image segmentation datasets.

TABLE 6 | The results of comparison between our proposed method and state-of-the-art methods on the BraTS 2020.

Method	DSC (%)	Sensitivity (%)	Specificity (%)	Hausdroff95	Parameter
U-Net (10)	87.59	87.04	99.89	8.97	34.5M
ResU-Net (59)	87.06	86.63	99.80	9.16	8.2M
ResU-Net++ (60)	88.48	87.98	99.90	7.42	42.2M
DeepLabV3+ (61)	82.99	84.16	99.82	11.05	39.4M
PSPNet (62)	83.74	82.27	99.87	5.99	35.0M
Attention UNet (31)	87.58	87.26	99.82	8.55	9.3M
Our method	**89.18**	**89.24**	**99.91**	**5.77**	**3.3M**

The bold values indicate the best results.

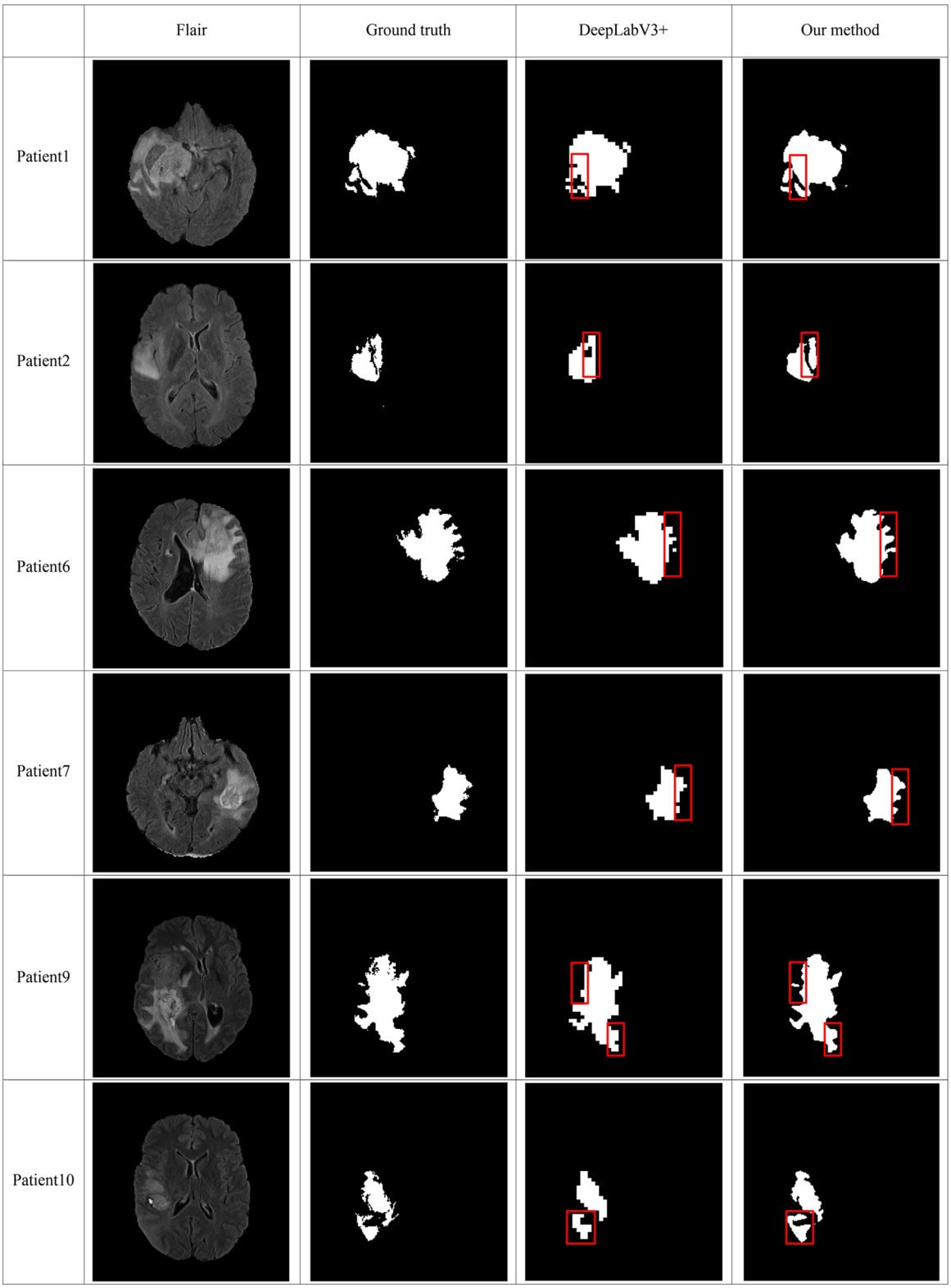


FIGURE 5 | Comparison of brain tumor segmentation results between our method and DeepLabV3+. The differences between the segmentation results of the two methods are marked by the red boxes.

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

SM and FG contributed to the conception of the study. SM performed the experiment. FG contributed significantly to analysis and manuscript preparation. SM performed the data analyses and wrote the manuscript. JT and FG helped perform the analysis with constructive discussions. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by a grant from National Key R&D Program of China (2020YFA0908400) and National Natural Science Foundation of China (NSFC 61772362, 61972280).

REFERENCES

1. Dutil F, Havaei M, Pal C, Larochelle H, Jodoin P. A Convolutional Neural Network Approach to Brain Lesion Segmentation. *Ischemic Stroke Lesion Segment* (2015) 9556:51–6. doi: 10.1007/978-3-319-30858-6_17
2. Dvorak P, Menze B. Structured Prediction With Convolutional Neural Networks for Multimodal Brain Tumor Segmentation. In: *Proceeding of the Multimodal Brain Tumor Image Segmentation Challenge* Springer (2015). p. 13–24.
3. Pereira S, Pinto A, Alves V, Silva CA. Brain Tumor Segmentation Using Convolutional Neural Networks in Mri Images. In: *IEEE Transactions on Medical Imaging* IEEE, vol. 35. (2016). p. 1240–51.
4. Simonyan K, Zisserman A. *Very Deep Convolutional Networks for Large-Scale Image Recognition* Vol. 1409.1556. arXiv preprint arXiv (2014).
5. Kamnitsas K, Ferrante E, Parisot S, Ledig C, AV N, Criminisi A, et al. Deepmedic for Brain Tumor Segmentation. In: *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. German: Springer (2016). p. 138–49.
6. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* IEEE (2016a). p. 770–8.
7. Randhawa RS, Modi A, Jain P, Warier P. Improving Boundary Classification for Brain Tumor Segmentation and Longitudinal Disease Progression. In: *International Workshop on Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. German: Springer (2016). p. 65–74.
8. Kamnitsas K, Bai W, Ferrante E, McDonagh S, Sinclair M, Pawlowski N, et al. Ensembles of Multiple Models and Architectures for Robust Brain Tumour Segmentation. In: *International Miccai Brainlesion Workshop*. German: Springer (2017). p. 450–62.
9. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* IEEE (2015). p. 3431–40.
10. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. German: Springer (2015). p. 234–41.
11. Chen H, Qin Z, Ding Y, Tian L, Qin Z. Brain Tumor Segmentation With Deep Convolutional Symmetric Neural Network. *Neurocomputing* (2020) 392:305–13. doi: 10.1016/j.neucom.2019.01.111
12. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH. Brain Tumor Segmentation and Radiomics Survival Prediction: Contribution to the Brats 2017 Challenge. In: *International Miccai Brainlesion Workshop*. German: Springer (2017). p. 287–97.
13. He K, Zhang X, Ren S, Sun J. Identity Mappings in Deep Residual Networks. In: *European Conference on Computer Vision*. German:Springer (2016). p. 630–45.
14. Ioffe S, Szegedy C. *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift* Vol. 1502.03167. arXiv preprint arXiv (2015).
15. Ulyanov D, Vedaldi A, Lempitsky V. *Instance Normalization: The Missing Ingredient for Fast Stylization* Vol. 1607.08022. arXiv preprint arXiv (2016).
16. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH. Nnu-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nat Methods* (2021) 18:203–11. doi: 10.1038/s41592-020-01008-z
17. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-Net: Learning Dense Volumetric Segmentation From Sparse Annotation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. German: Springer (2016). p. 424–32.
18. Dou Q, Yu L, Chen H, Jin Y, Yang X, Qin J, et al. 3D Deeply Supervised Network for Automated Segmentation of Volumetric Medical Images. *Med Image Anal* (2017) 41:40–54. doi: 10.1016/j.media.2017.05.001
19. Lee CY, Xie S, Gallagher P, Zhang Z, Tu Z. Deeply-Supervised Nets. *Artif Intell Stat* (2015) 38:562–70.
20. Chen H, Dou Q, Yu L, Heng PA. *Voxresnet: Deep Voxelwise Residual Networks for Volumetric Brain Segmentation*. arXiv preprint arXiv (2016). 1608.05895.
21. Zeng G, Yang X, Li J, Yu L, Heng PA, Zheng G. 3D U-Net With Multi-Level Deep Supervision: Fully Automatic Segmentation of Proximal Femur in 3D Mr Images. In: *International Workshop on Machine Learning in Medical Imaging*. German: Springer (2017). p. 274–82.
22. Zhang Y, Chung AC. Deep Supervision With Additional Labels for Retinal Vessel Segmentation Task. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. German: Springer (2018). p. 83–91.
23. Zeng G, Zheng G. Multi-Stream 3D Fcn With Multi-Scale Deep Supervision for Multi-Modality Isointense Infant Brain Mr Image Segmentation. In: *IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. USA: IEEE (2018). p. 136–40.
24. Myronenko A. 3D Mri Brain Tumor Segmentation Using Autoencoder Regularization. In: *International Miccai Brainlesion Workshop*. German: Springer (2018). p. 311–20.
25. Chen S, Bortsova G, Juárez AGU, van Tulder G, de Bruijne M. Multi-Task Attention-Based Semi-Supervised Learning for Medical Image Segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. German: Springer (2019). p. 457–65.
26. Jiang Z, Ding C, Liu M, Tao D. Two-Stage Cascaded U-Net: 1st Place Solution to Brats Challenge 2019 Segmentation Task. In: *International Miccai Brainlesion Workshop*. German: Springer (2019). p. 231–41.
27. Weninger L, Liu Q, Merhof D. Multi-Task Learning for Brain Tumor Segmentation. In: *International Miccai Brainlesion Workshop*. German: Springer (2019). p. 327–37.
28. Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* IEEE (2018).
29. Li X, Wang W, Hu X, Yang J. Selective Kernel Networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* IEEE (2019a). p. 510–9.
30. Woo S, Park J, JY L, So Kweon I. Cbam: Convolutional Block Attention Module. In: *Proceedings of the European Conference on Computer Vision (ECCV)* Springer (2018). p. 3–19.
31. Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. *Attention U-Net: Learning Where to Look for the Pancreas*. arXiv preprint arXiv (2018). 1804.03999.
32. Chaurasia A, Culurciello E. Linknet: Exploiting Encoder Representations for Efficient Semantic Segmentation. In: *2017 IEEE Visual Communications and Image Processing (VCIP)*. USA: IEEE (2017). p. 1–4.

33. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The Importance of Skip Connections in Biomedical Image Segmentation. In: *Deep Learning and Data Labeling for Medical Applications*. German: Springer (2016). p. 179–87.
34. Li D, Chen Q. Dynamic Hierarchical Mimicking Towards Consistent Optimization Objectives. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* IEEE (2020). p. 7642–51.
35. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: A Nested U-Net Architecture for Medical Image Segmentation. In: *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. German: Springer (2018). p. 3–11.
36. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated Residual Transformations for Deep Neural Networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* IEEE (2017). p. 1492–500.
37. Gao S, Cheng MM, Zhao K, Zhang XY, Yang MH, Torr PH. Res2net: A New Multi-Scale Backbone Architecture. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* IEEE (2019).
38. Ibtehaz N, Rahman MS. Multiresunet: Rethinking the U-Net Architecture for Multimodal Biomedical Image Segmentation. *Neural Networks* (2020) 121:74–87. doi: 10.1016/j.neunet.2019.08.025
39. 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. *Sci Data* (2017) 4:170117. doi: 10.1038/sdata.2017.117
40. 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 preprint arXiv (2018). 1811.02629.
41. Chen W, Liu B, Peng S, Sun J, Qiao X. S3d-Unet: Separable 3d U-Net for Brain Tumor Segmentation. In: *International Miccai Brainlesion Workshop*. German: Springer (2018a). p. 358–68.
42. Milletari F, Navab N, Ahmadi SA. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV)*. USA: IEEE (2016). p. 565–71.
43. Nuechterlein N, Mehta S. 3d-Espnet With Pyramidal Refinement for Volumetric Brain Tumor Image Segmentation. In: *International Miccai Brainlesion Workshop*. German:Springer (2018). p. 245–53.
44. Benson E, Pound MP, French AP, Jackson AS, Pridmore TP. Deep Hourglass for Brain Tumor Segmentation. In: *International Miccai Brainlesion Workshop*. German: Springer (2018). p. 419–28.
45. Hv V. Pre and Post Processing Techniques for Brain Tumor Segmentation. In: *Proc. Pre-Conf. 7th MICCAI Brats Challenge* Springer (2018). p. 213–21.
46. Ma S, Li X, Tang J, Guo F. A Zero-Shot Method for 3d Medical Image Segmentation. In: *2021 IEEE International Conference on Multimedia and Expo (ICME)*. USA: IEEE (2021). p. 1–6.
47. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, et al. *Tensorflow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. arXiv preprint arXiv (2016). 1603.04467.
48. Kingma DP, Ba J. *Adam: A Method for Stochastic Optimization*. arXiv preprint arXiv (2014). 1412.6980.
49. González C, Escobar M, Daza L, Torres F, Triana G, Arbeláez P. Simba: Specific Identity Markers for Bone Age Assessment. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. German: Springer (2020). p. 753–63.
50. Jin Q, Meng Z, Sun C, Cui H, Su R. Ra-Unet: A Hybrid Deep Attention-Aware Network to Extract Liver and Tumor in Ct Scans. *Front Bioeng Biotechnol* (2020) 8:1471. doi: 10.3389/fbioe.2020.605132
51. Karnawat A, Prasanna P, Madabushi A, Tiwari P. Radiomics-Based Convolutional Neural Network (Radcnn) for Brain Tumor Segmentation on Multi-Parametric Mri. In: *Proceedings of MICCAI-Brats Conference, Canada* Springer (2017).
52. Banerjee S, Mitra S. Novel Volumetric Sub-Region Segmentation in Brain Tumors. *Front Comput Neurosci* (2020) 14:3. doi: 10.3389/fncom.2020.00003
53. Rezaei M, Yang H, Meinel C. Generative Adversarial Framework for Learning Multiple Clinical Tasks. In: *2018 Digital Image Computing: Techniques and Applications (DICTA)*. USA: IEEE (2018). p. 1–8.
54. Zhang X, Jian W, Cheng K. *3D Dense U-Nets for Brain Tumor Segmentation*. German: Springer (2018) p. 562–70.
55. Islam M, Vibashan V, Jose VJM, Wijethilake N, Utkarsh U, Ren H. Brain Tumor Segmentation and Survival Prediction Using 3D Attention Unet. In: *International Miccai Brainlesion Workshop*. German: Springer (2019). p. 262–72.
56. Cheng X, Jiang Z, Sun Q, Zhang J. Memory-Efficient Cascade 3d U-Net for Brain Tumor Segmentation. In: *International Miccai Brainlesion Workshop*. German: Springer (2019). p. 242–53.
57. Li X, Luo G, Wang K. Multi-Step Cascaded Networks for Brain Tumor Segmentation. In: *International Miccai Brainlesion Workshop*. German: Springer (2019b). p. 163–73.
58. Wang F, Jiang R, Zheng L, Meng C, Biswal B. 3d U-Net Based Brain Tumor Segmentation and Survival Days Prediction. In: *International Miccai Brainlesion Workshop*. German: Springer (2019). p. 131–41.
59. Liu Z, Feng R, Wang L, Zhong Y, Cao L. D-Resunet: Resunet and Dilated Convolution for High Resolution Satellite Imagery Road Extraction. In: *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. USA: IEEE (2019). p. 3927–30.
60. Jha D, Smedsrud PH, Riegler MA, Johansen D, De Lange T, Halvorsen P, et al. Resunet++: An Advanced Architecture for Medical Image Segmentation. In: *2019 IEEE International Symposium on Multimedia (ISM)*. USA: IEEE (2019). p. 225–2255.
61. Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H. Encoder-Decoder With Atrous Separable Convolution for Semantic Image Segmentation. In: *Proceedings of the European Conference on Computer Vision (ECCV)* Springer (2018). p. 801–18.
62. Zhao H, Shi J, Qi X, Wang X, Jia J. Pyramid Scene Parsing Network. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* IEEE (2017). p. 2881–90.

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 01 October 2021
doi: 10.3389/fonc.2021.708655

MRI-Based Grading of Clear Cell Renal Cell Carcinoma Using a Machine Learning Classifier

Xin-Yuan Chen[1†], Yu Zhang[2,3†], Yu-Xing Chen[4†], Zi-Qiang Huang[2], Xiao-Yue Xia[2], Yi-Xin Yan[4], Mo-Ping Xu[4], Wen Chen[4], Xian-long Wang[4] and Qun-Lin Chen[2*]*

[1] Department of Rehabilitation Medicine, The First Affiliated Hospital of Fujian Medical University, Fuzhou, China, [2] Department of Radiology, The First Affiliated Hospital of Fujian Medical University, Fuzhou, China, [3] Department of Radiology, Xiangyang Central Hospital, Affiliated Hospital of Hubei University of Arts and Science, Xiangyang, China, [4] Department of Bioinformatics, School of Basic Medical Sciences, School of Medical Technology and Engineering, Key Laboratory of Medical Bioinformatics, Key Laboratory of Ministry of Education for Gastrointestinal Cancer, Fujian Medical University, Fuzhou, China

OPEN ACCESS

Edited by:
Natalie Julie Serkova, University of Colorado, United States

Reviewed by:
Arnaldo Stanzione, University of Naples Federico II, Italy
Subathra Adithan, Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER), India

***Correspondence:**
Qun-Lin Chen
fychenqunlin@126.com
Xian-long Wang
xwang@fjmu.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 May 2021*
Accepted: *15 September 2021*
Published: *01 October 2021*

Citation:
Chen X-Y, Zhang Y, Chen Y-X, Huang Z-Q, Xia X-Y, Yan Y-X, Xu M-P, Chen W, Wang X-l and Chen Q-L (2021) MRI-Based Grading of Clear Cell Renal Cell Carcinoma Using a Machine Learning Classifier. Front. Oncol. 11:708655. doi: 10.3389/fonc.2021.708655

Objective: To develop a machine learning (ML)-based classifier for discriminating between low-grade (ISUP I-II) and high-grade (ISUP III-IV) clear cell renal cell carcinomas (ccRCCs) using MRI textures.

Materials and Methods: We retrospectively evaluated a total of 99 patients (with 61 low-grade and 38 high-grade ccRCCs), who were randomly divided into a training set (n = 70) and a validation set (n = 29). Regions of interest (ROIs) of all tumors were manually drawn three times by a radiologist at the maximum lesion level of the cross-sectional CMP sequence images. The quantitative texture analysis software, MaZda, was used to extract texture features, including histograms, co-occurrence matrixes, run-length matrixes, gradient models, and autoregressive models. Reproducibility of the texture features was assessed with the intra-class correlation coefficient (ICC). Features were chosen based on their importance coefficients in a random forest model, while the multi-layer perceptron algorithm was used to build a classifier on the training set, which was later evaluated with the validation set.

Results: The ICCs of 257 texture features were equal to or higher than 0.80 (0.828–0.998. Six features, namely Kurtosis, 135dr_RLNonUni, HorzI_GLevNonU, 135dr_GLevNonU, S(4,4)Entropy, and S(0,5)SumEntrp, were chosen to develop the multi-layer perceptron classifier. A three-layer perceptron model, which has 229 nodes in the hidden layer, was trained on the training set. The accuracy of the model was 95.7% with the training set and 86.2% with the validation set. The areas under the receiver operating curves were 0.997 and 0.758 for the training and validation sets, respectively.

Conclusions: A machine learning-based grading model was developed that can aid in the clinical diagnosis of clear cell renal cell carcinoma using MRI images.

Keywords: machine learning, magnetic resonance imaging, texture analysis, clear cell renal cell carcinoma, multi-layer perceptron algorithm

INTRODUCTION

Renal cell carcinoma (RCC) is the most common malignant kidney tumor, and the most common pathological type, accounting for 70–90%, is clear cell renal cell carcinoma (ccRCC) (1). The latest World Health Organization (WHO)/ International Society of Urological Pathology (ISUP) grading system divides ccRCC into four grades, in which grades I and II are low-grade tumors with good prognosis while grades III and IV are high-grade tumors with poor prognosis (2, 3). Current studies have shown a relationship between the different nuclear grades of RCCs and the choice of surgical methods and prognosis (4, 5). Therefore, preoperative determination of the nuclear grade of ccRCC is valuable.

The pathological features of renal masses are frequently evaluated by preoperative percutaneous renal biopsy, but this invasive technique still suffers from low accuracy. This has prompted a search for non-invasive methods that can grade the tumors and aid clinicians in selecting optimal therapeutic regimens. Several studies have proposed the use of images generated by computed tomography (CT) or magnetic resonance imaging (MRI) for identification of potential biomarkers for tumor grading (6, 7). MRI images have the advantage of being free from ionizing radiation exposure and are capable of evaluating both the tumor morphology and the tumor microenvironment (8), but MRI itself is incapable of providing sufficient information for differentiating the grades of ccRCC by most radiologists. However, artificial intelligence can play an important role in interpreting MRI information in comprehensive ways by texture analysis. In this way, MRI images can provide quantitative statistical parameters by identifying subtle texture information not readily observable with the human eye (8). These parameters, rather than the original images, can be then used as the input features for machine learning algorithms to improve the sensitivity of medical imaging diagnosis, and they can also circumvent the requirement for large sample sizes in image-based deep-learning algorithms. At present, the use of magnetic resonance texture analysis (MRTA) to predict ccRCC grades is seldom reported (9–11). The purpose of this study was to explore the value of using MRI textures and machine learning algorithms for predicting the grade of ccRCCs before operations.

MATERIALS AND METHODS

Clinical Data

This retrospective study was approved by our Hospital Authority Review Committee. The requirement for informed consent was waived because of the study's retrospective nature. The analysis included patients who met the following standards hospitalized from July 2016 to January 2020 at the First Affiliated Hospital of Fujian Medical University.

The inclusion criteria were: (i) patients surgically confirmed with ccRCCs; (ii) patients who had undergone preoperative contrast-enhanced MRI (corticomedullary phase, nephrographic phase, and delayed phase) in our hospital within one week before operations; and (iii) patients with single lesions with short diameters of more than 1 centimeter measured on axial T2 weighted imaging.

The exclusion criteria were: (i) patients with MRI images with artifacts, such as respiratory movement or magnetic sensitivity; (ii) patients with long lesion diameters ≤ 1cm; (iii) patients with tumors presenting as obvious cystic degeneration (cystic degeneration portion >75%); and (iv) patients with preoperative puncture biopsy, interventions, or other treatments.

We enrolled 99 patients with histologically verified ccRCC. These patients included 61 cases with low-grade disease (4 grade I cases and 57 grade II cases) and 38 cases with high-grade disease (32 grade III cases and 6 grade IV cases). The low-grade group included 42 males and 19 females, while the high-grade group included 25 males and 13 females. All MRI images were exported from the Picture Archiving and Communication System (PACS) of our hospital.

MRI Examination

All patients underwent a preoperative 3.0 Tesla MR (MAGNETOM Verio, Siemens, Germany) examination with the standard protocol using a phased-array body coil. Image acquisition sequences and parameters were as follows: (a) coronal half-Fourier acquisition single-shot turbo spin-echo (HAST) sequences (repetition time msec/echo time msec, 1400/91; field of view, 340×340 mm; matrix, 224×320; section thickness, 5mm; intersection gap, 1mm); (b) transverse T2-weighted single-shot fast spin-echo sequences (repetition time msec/echo time msec, 2000/91; field of view, 340×340 mm; matrix 224×320; section thickness, 3 mm; intersection gap, 0.8mm); (c) axial diffusion weighted imaging sequences (repetition time msec/echo time msec, 6000/73; field of view, 340×340; section thickness, 4 mm; intersection gap, 0.8mm; and two sets of b values: 50 and 800 sec/mm^2); (d) transverse three-dimensional fat suppressed T1-weighted interpolated spoiled gradient echo (volumetric interpolated breath-hold examination, VIBE) sequences (repetition time msec/echo time msec, 3.92/1.39; field of view, 250×380 mm; matrix, 224×320; section thickness, 3mm; intersection gap, 0.6 mm). The VIBE sequences were performed prior to and three times after intravenous injection of gadopentetate dimeglumine (MultiHance, Bracco Sine, Shanghai, China; 0.1 mmol per kilogram of body weight) at a rate of 2 mL/sec with a power injector (Medrad, Warrendale, USA), followed by a 20 mL saline flush. Corticomedullary phase images were obtained approximately 40–50 seconds after administration of the contrast material using timing, nephrographic phase images were obtained at 80–100 seconds, and excretory phase images were obtained 3 minutes later.

Placement of ROIs

All data were stored anonymously in the Digital Imaging and Communications in Medicine (DICOM) format. The largest cross-section of the tumor on the axial CMP images was first determined, and then images of the selected layer were imported into MaZda (version 4.6, http://www.eletel.p.lodz.pl/mazda/). The two-dimensional region of interest (ROI) was then delineated manually by an experienced radiologist (Ziqiang

Huang), who was engaged in urogenital system imaging diagnosis and blinded to the nuclear grade of the ccRCCs. Note that the edge of the lesion segmentation had shrunk inward by 1–2 mm.

Feature Extraction

The differences in image brightness and contrast were reduced by standardizing the gray scale of the images before texture extraction, so that the image gray scale was within the range of [μ−3σ, μ+3σ], where μ and σ represent the mean gray value and the standard deviation, respectively. The MaZda quantitative texture analysis software package was used to extract texture features, including the gray-scale histogram, co-occurrence matrix, run-length matrix, and gradient and autoregressive models. All 257 radiomics features were extracted from each ROI for each patient. Three feature data sets (Data 1, Data 2, and Data 3) were obtained from the same tumor images by segmenting the data three times.

Reproducibility of Texture Features

The reproducibility of the texture features was evaluated by calculating the intra-class correlation coefficients (ICCs) of 257 texture features among the three feature datasets. Only features with an ICC value equal to or higher than 0.80, indicating excellent reproducibility, were included in further feature selection processes.

Feature Standardization

Before model development, various features were first standardized to make them comparable using the 'robustscale' method in the Python package of scikit-learn (ver. 0.23.2, https://scikit-learn.org/) (12). The specific formula is as follows:

$$z_i = \frac{V_i - M}{\mathrm{IQR}}$$

where V_i is the original feature value, M is the median of the feature, and IQR is the interquartile range (the difference between the third quartile and the first quartile). The following logistic transformation was then performed to minimize the adverse effects of outliers on the stability of the classifier:

$$y_i = 1/\left(1 + e^{-4z_i}\right).$$

Feature Selection

A random forest model was used to select features for model development using the Random Forest Classifier function provided by scikit-learn. A grid search algorithm was then executed to determine a set of hyperparameters using the "GridSearchCV" function provided by scikit-learn. The random forest parameters were the following: 'class_weight' = 'balanced'; 'max_features' = 'log2'; and the rest were default values. A random forest model was then fitted to the training set, and the model then assigned each feature an importance coefficient that represents the information gain for the specific feature, where a larger value indicates a greater importance of the feature. The number of features was determined by repeated iterations based on the accuracy of the model on the validation set, while keeping the number of features as small as possible. Finally, 6 features, namely Kurtosis, 135dr_RLNonUni, Horzl_GLevNonU, 135dr_GLevNonU, S(4,4)Entropy, and S(0,5)SumEntrp, with the largest importance coefficients were selected.

Model Development

We used a multilayer perceptron algorithm (the MLP Classifier function in scikit-learn) to develop the classification model. The model parameters were the following: 'activation' = 'relu', and 'solver' = 'lbfgs', 'learning_rate' = 'constant' and 'hidden_layer_sizes' = '(229),'. The most important parameter was 'hidden_layer_sizes', which determines the number of hidden layers and the number of nodes in each hidden layer. In this work, we included only one hidden layer, which consisted of 229 nodes. The number of nodes was optimized by repeated iterations to achieve optimal accuracy on the validation set.

Statistical Analysis

Univariate analyses were performed with SPSS version 22 (SPSS Inc.). In the training set, the continuous variables (age, tumor size) between low-grade and high-grade groups were analyzed with the Student's *t* test or the Mann–Whitney *U* test. The Chi-squared test was used to analyze the categorical variables (gender) between the two groups. A p value less than 0.05 was considered statistically significant.

RESULTS

Demographic Analysis

The baseline characteristics of the training and validation sets are presented in **Table 1**. The training set consisted of 70 patients with pathologically proven ccRCC lesions (low-grade ccRCCs: 3 grade I lesions and 40 grade II lesions; high-grade ccRCCs: 23 grade III lesions and 4 grade IV lesions). The validation set consisted of 1 grade I lesion, 17 grade II lesions, 9 grade III lesions, and 2 grade IV lesions. In the training set, the mean ages ± standard deviations of the low-grade and high-grade subgroups were 53.5 ± 11.5 years and 57.1 ± 10.9 years, respectively. No statistically significant differences were found for gender and age distribution between the low-grade and high-grade ccRCC groups (p = 0.751 and 0.124, respectively). The average tumor sizes were 4.0 cm and 6.1 cm, respectively, in the low-grade and high-grade subgroups, and the difference was statistically significant (p<0.001).

MRI Texture Analysis and Feature Selection

The MRI images of 99 ccRCC tumors were used to extract 257 texture features with the MaZda software package. The features included 7 histogram features, 220 gray co-occurrence matrix features, 20 run-length matrix features, 5 gradient features, and 5 autoregressive model features.

The ICC ranges of the histogram features, gray level co-occurrence matrix features, run-length matrix features, gradient

TABLE 1 | Analysis of baseline data from patients with ccRCCs.

Characteristic	Low-grade group	High-grade group	Statistics	*P* value
Patients (*n*)	61	38	–	–
Age (mean ± SD, years)	53.5 ± 11.5	57.1 ± 10.9	-1.522	0.124
Gender			0.1	0.751
Male (*n*)	42	25		
Female (*n*)	19	13		
Tumor size (mean ± SD, cm)	4.0 ± 2.1	6.1 ± 2.9	-3.869	<0.001

features, and autoregressive model features were (0.968, 0.998), (0.828, 0.996), (0.880, 0.997), (0.934, 0.986), and (0.863, 0.984), respectively.

Upon obtaining stable texture features, we applied the RobustScale method to standardize the feature values in the training set, and we then carried out a logistic transformation on them to minimize the negative impact of outliers on the model development (see *Materials and Methods* for details). The same formulas were archived and later applied to the validation set. The 257 processed features were input into a random forest model and fitted on the training dataset, while the hyperparameters of the random forest model were optimized with the grid search method.

The model assigned an importance coefficient to each feature. The value of the coefficient represents the importance of the feature. A set of top-ranked probes was selected to develop the MLP model and to optimize the hyperparameters to achieve the highest accuracy in the validation set. With the optimized hyperparameters, the number of features was updated with the new fitting model. This iteration was repeated manually to obtain a minimal set of features without appreciably sacrificing the accuracy. In the final model, 6 texture features were selected for modeling. The heatmap of the 6 selected features is shown in **Figure 1** for the training set. **Figure 1** also shows that the low-grade and high-grade ccRCCs are approximately clustered into two separate groups, demonstrating the rationality of the selected features.

Model Development

A multi-layer perceptron algorithm was used for developing the prediction classifier. The final model has a three-layer structure: an input layer, a hidden layer, and an output layer (see **Figure 2**). The input layer consists of 6 nodes, corresponding to the 6 texture features, and the output layer consists of 2 nodes, corresponding to the low-grade and high-grade groups. The hidden layer has 229 nodes in the final model.

Model Validation

The optimized model was evaluated in the validation set. The predictive indicators of the model in the training set and the validation set are shown in **Table 2**. The accuracy was 95.7% and 86.2%, respectively, in the training set and the validation set. The AUC values were 0.997 and 0.758, respectively, in the two sets (**Figure 3**). In the training set, two low-grade tumors were predicted as high-grade, and one high-grade tumor was incorrectly classified (**Figure 1**). In the validation set, only one

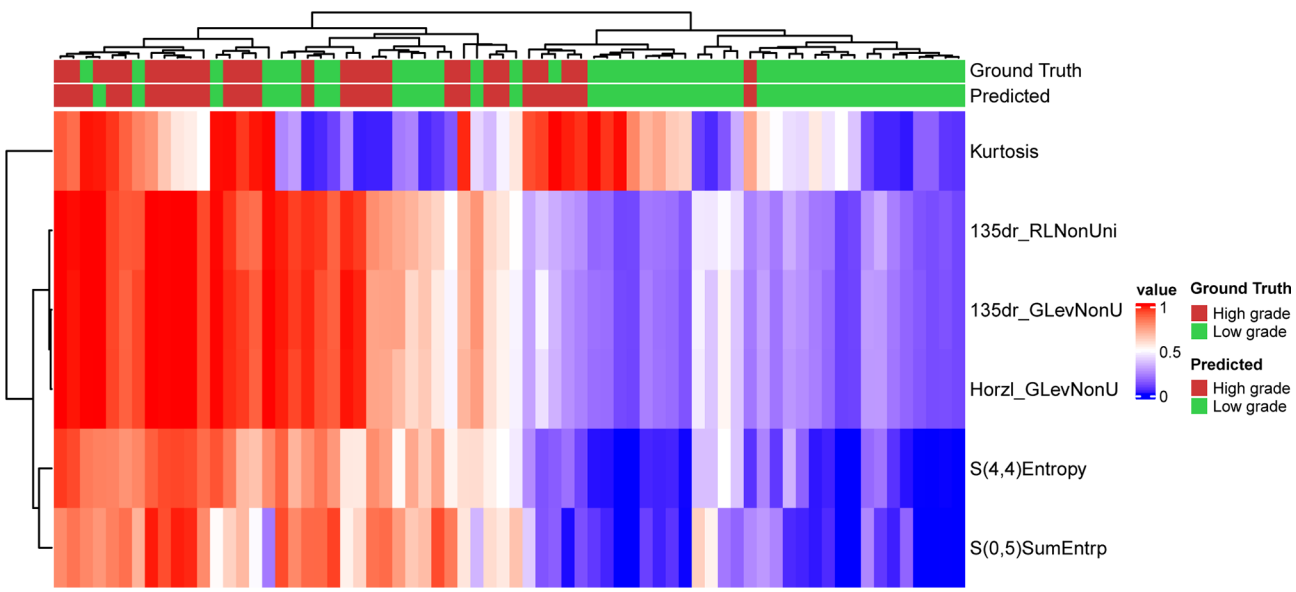


FIGURE 1 | Prediction result and the heatmap of the selected 6 features in the training set along the clustering results of the samples and the features.

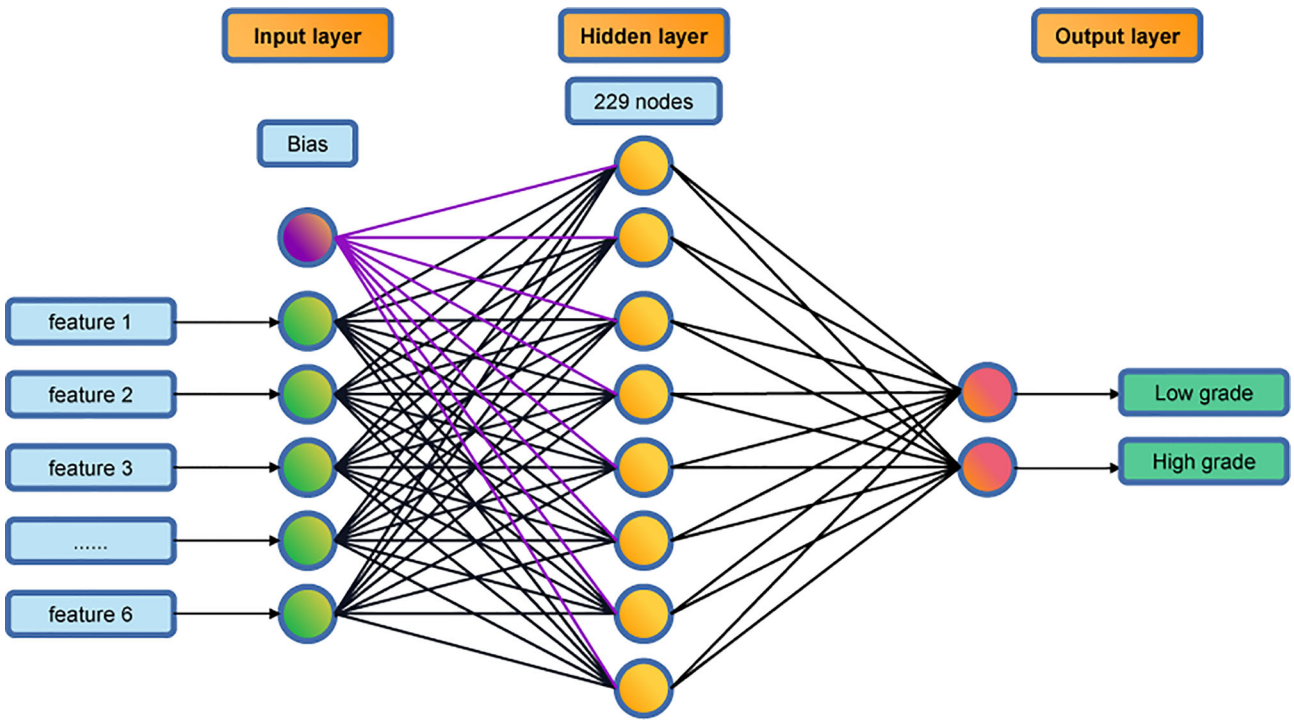


FIGURE 2 | The topological structure of the 3-layer perceptron classifier.

TABLE 2 | Performance of the MLP classifier in the training and validation sets.

	AUC	ACC	SEN	SPE	PPV	NPV
Training set	0.997	0.957	96.30%	95.30%	92.90%	97.60%
Validation set	0.758	0.862	72.70%	94.40%	88.90%	85.00%

low-grade and three high-grade tumors were misclassified (**Figure 4**). **Figure 5** shows that the misclassified low-grade tumors show higher similarities with the high-grade tumors and that, similarly, the misclassified high-grade tumors also show a higher similarity with the low-grade tumors. This result suggests that the selected texture features might be inadequate for discriminating these samples.

Figure 5 shows the distribution of the probabilities predicted by the MLP model in the validation set. The prediction probability of the model's prediction results for the 24/29 samples of the validation set is greater than 0.9, which indicates that the model is highly confident in the prediction result and is relatively robust.

DISCUSSION

In this study, we evaluated the applicability of a machine learning method based on MRI textures for the grade classification of ccRCCs. A three-layer MLP classifier using 6 features from MRI texture analysis exhibited satisfactory, reproducible, and reliable performance in discriminating the high-grade ccRCCs from the low-grade ones, and it outperformed classifiers presented in previous studies (8, 10, 13).

We adopted the latest WHO/ISUP grading system for renal cell carcinoma as the classifying criterion. However, most of the previous studies on the prediction of nuclear grading of ccRCC by texture analysis have been based on the Fuhrman classification system, which has some inevitable inadequacies, such as interpretation difficulties and poor reproducibility in clinical applications (10, 14). Besides the high application value for ccRCC, the WHO/ISUP nuclear grading is also a reliable prognosis indicator of patients with ccRCC (15).

In this study, we attempted to predict the nuclear classification of ccRCCs by quantitative analysis based on the texture features of MRI images. However, in current clinical practice, radiologists estimate the degree of aggressiveness of renal carcinoma based mainly on radiological findings (16, 17). For example, Pedrosa et al. found that some MRI features, which include both qualitative and semiquantitative parameters, can differentiate low-grade and high-grade ccRCCs (18). However, the classification is subjective and depends on the radiologist's

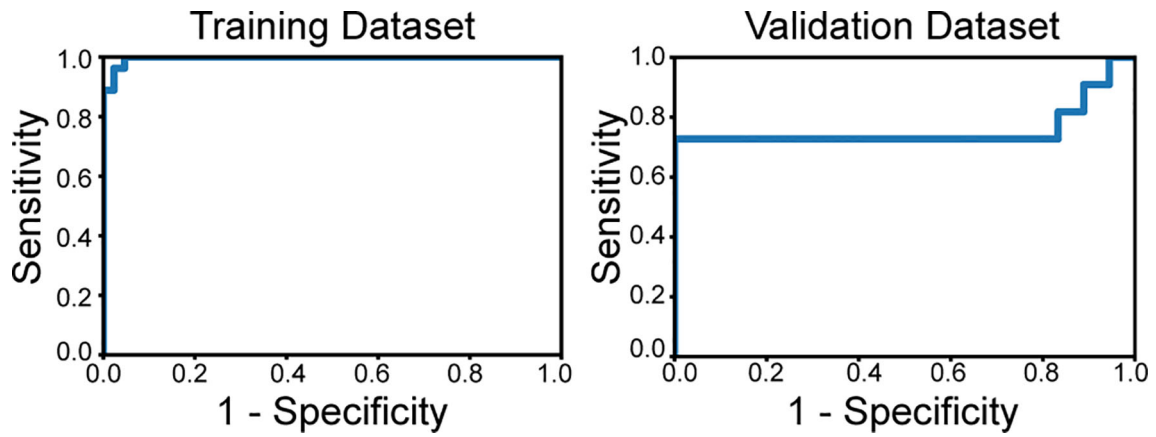


FIGURE 3 | The receiver operating characteristics curves of the classifier applied to the training set and validation set. The area under curve is 0.997 in the training dataset and 0.758 in the validation set.

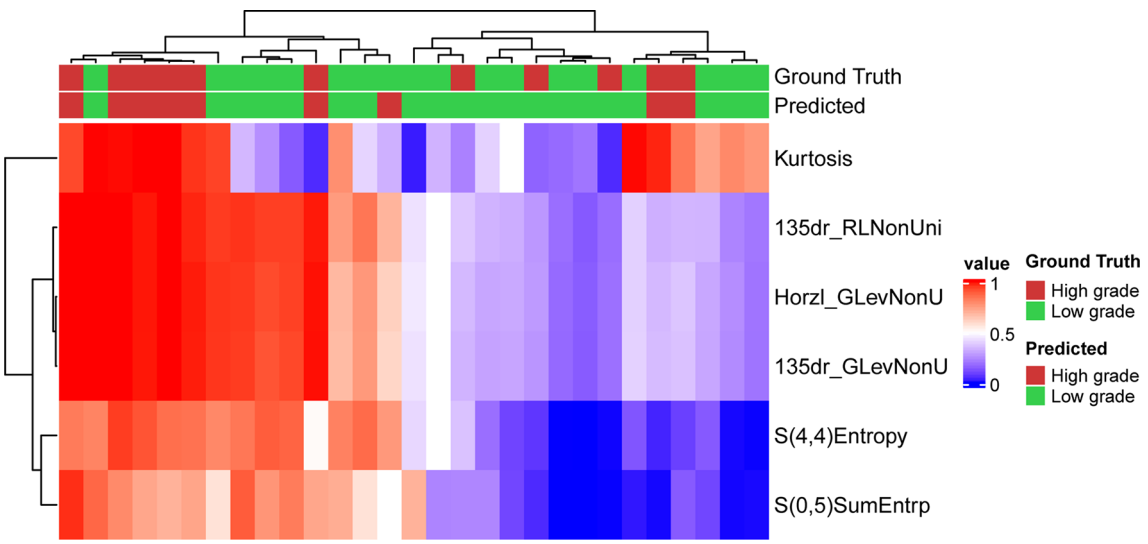


FIGURE 4 | Prediction result and the heatmap of the selected 6 features in the validation set along the clustering results of the samples and the features.

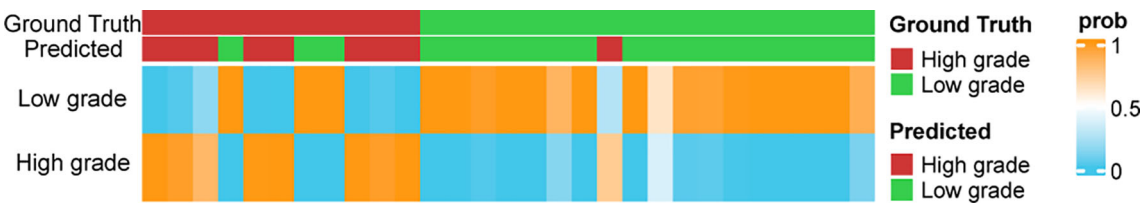


FIGURE 5 | The prediction probabilities assigned by the classifier in the validation set. The top annotation labels show the ground truth and the predicted result.

experience. Quantitative MRI texture analysis is now playing an increasingly important role in the clinical diagnosis and treatment of tumors, and it can be used to distinguish the pathological types and grades of tumors, to evaluate prognosis, and to predict the therapeutic response of tumors (19–21). Compared with CT examinations, MRI has multiple advantages, including multi-parameter imaging, high soft tissue resolution, high signal-to-noise ratio, and freedom from ionizing radiation. The texture features of multiple sequence images can be obtained with MRI, and this provides more feature space for developing imaging markers for tumors. Therefore, MRI texture analysis is a useful and promising method for non-invasive prediction of the ISUP nuclear grade of ccRCCs.

Multiple machine learning models have been successfully constructed to classify low-grade and high-grade ccRCCs (10, 13, 22). After comparing the performance of different models, we obtained an optimal prediction result with MLP. The AUC value of the classifier is 0.997 in the training set, indicating a good performance of the MLP model. Bektas et al. developed machine learning models to predict the Fuhrman nuclear grade of ccRCC based on quantitative CT texture analysis (22). They achieved the best prediction result using an MLP model with an AUC of 0.86. We further validated the application value of our model by creating a validation set to assess the accuracy and stability of the model. Satisfactory results were obtained, with an AUC of 0.758 in the validation set.

Most studies on machine learning-based CT or MRI texture analysis have not validated the developed models for predicting the nuclear grade of ccRCC. A comprehensive review of the radiomics literature on renal mass characterization in terms of validation strategies did not reveal any validation performed in 19 (46%) of the 41 papers reviewed (23). In other words, only slightly more than half of the studies described at least one validation method, and these were predominantly internal validation techniques. The wide clinical use of radiomics will require proper validation strategies for developing machine learning models. Compared with previous studies (24, 25), an independent and prospective test set is needed for further validation of our model in the future.

The current study has some inevitable limitations. One is that it is a single-center and retrospective study, so selectivity bias may exist. Another is the small sample size, which may lead to overfitting and low repeatability of the prediction results. Therefore, further expansion of the sample size and cross-verification of the model at multiple centers are needed. There is a slight imbalance in our dataset where the number of low-grade patients is larger than the number of high-grade patients. This issue could be addressed by the SMOTE algorithm (26). However, due the limited sample size we did not employ the method. Furthermore, the ratio between the sample size of the low-grade and that of the high-grade is approximately 3:2 where the class imbalance problem is not critical to the model performance. A third limitation is that the texture features extracted in this study are based on the two-dimensional ROI of MR images at the maximum level of the tumor, which may be biased by layer selection. Ideally, three-dimensional radiomic features of the whole lesion should be extracted to obtain comprehensive tumor features. A fourth limitation is that manual segmentation of MR images may be affected by the consistency between observers; however, this method is still widely used in texture analysis and remains the "gold standard" (27). Here, the stability of the texture features was evaluated by segmenting the lesions of all patients three times. Tumor size is generally associated with the tumor grade and therefore an important factor in tumor grading system. This information has been encoded in the feature "dr135RLNonUni" (Spearman's $\rho = 0.945$ between the tumor diameter and dr135RLNonUni) and therefore it was implicitly used in the final model. The value of clinical factors other than radiomics signatures will also be investigated in predicting the grades in future study.

CONCLUSIONS

An MLP model was successfully developed to classify the grades of clear-cell renal cell carcinomas, thereby demonstrating that ML-based MRI texture classifiers can be used preoperatively as a complementary tool to predict the ISUP grade of ccRCCs. This model can make a potential contribution to personalized treatment for patients with ccRCCs.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

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 work reported in the above for publications has been done by all authors. X-YC, YZ, and Y-XC contributed to data analysis, manuscript editing and model development and manuscript Preparation. Z-QH and X-YX collected the data of patients. Y-XY, M-PX, and WC helped in images analysis. X-lW and Q-LC helped in manuscript preparation and contributed to the supervision of the whole process. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by Fujian Medical University (Grant No. XRCZX2017001); the Natural Science Foundation of Fujian Province (Grant No. 2019J01294); Fujian Youth Education Research Project (Grant No. JT180193).

REFERENCES

1. Warren AY, Harrison D. WHO/ISUP Classification, Grading and Pathological Staging of Renal Cell Carcinoma: Standards and Controversies. *World J Urol* (2018) 36(12):1913–26. doi: 10.1007/s00345-018-2447-8
2. Evelönn EA, Degerman S, Köhn L, Landfors M, Ljungberg B, Roos G. DNA Methylation Status Defines Clinicopathological Parameters Including Survival for Patients With Clear Cell Renal Cell Carcinoma (ccRCC). *Tumour Biol* (2016) 37(8):10219–28. doi: 10.1007/s13277-016-4893-5
3. Delahunt B. Advances and Controversies in Grading and Staging of Renal Cell Carcinoma. *Mod Pathol* (2009) 22:S24–36. doi: 10.1038/modpathol.2008.183
4. Lam JS, Klatte T, Breda A. Staging of Renal Cell Carcinoma: Current Concepts. *Indian J Urol* (2009) 25(4):446–54. doi: 10.4103/0970-1591.57906
5. Ljungberg B. Prognostic Factors in Renal Cell Carcinoma. *Urologe A* (2004) 43 Suppl 3:S119–20. doi: 10.1007/s00120−004−0594−6
6. Oh J, Lee JM, Park J, Joo I, Yoon JH, Lee DH, et al. Hepatocellular Carcinoma: Texture Analysis of Preoperative Computed Tomography Images can Provide Markers of Tumor Grade and Disease-Free Survival. *Korean J Radiol* (2019) 20(4):569–79. doi: 10.3348/kjr.2018.0501
7. Zhou H, Vallières M, Bai HX, Su C, Tang H, Oldridge D, et al. MRI Features Predict Survival and Molecular Markers in Diffuse Lower-Grade Gliomas. *Neuro Oncol* (2017) 19(6):862–70. doi: 10.1093/neuonc/now256
8. Goyal A, Razik A, Kandasamy D, Seth A, Das P, Ganeshan B, et al. Role of MR Texture Analysis in Histological Subtyping and Grading of Renal Cell Carcinoma: A Preliminary Study. *Abdom Radiol* (2019) 44(10):3336–49. doi: 10.1007/s00261-019-02122-z
9. Rosenkrantz AB, Niver BE, Fitzgerald EF, Babb JS, Chandarana H, Melamed J. Utility of the Apparent Diffusion Coefficient for Distinguishing Clear Cell Renal Cell Carcinoma of Low and High Nuclear Grade. *AJR Am J Roentgenol* (2010) 195(5):W344–51. doi: 10.2214/AJR.10.4688
10. Cui E, Li Z, Ma C, Li Q, Lei Y, Lan Y, et al. Predicting the ISUP Grade of Clear Cell Renal Cell Carcinoma With Multiparametric MR and Multiphase CT Radiomics. *Eur Radiol* (2020) 30(5):2912–21. doi: 10.1007/s00330-019-06601-1
11. Stanzione A, Ricciardi C, Cuocolo R, Romeo V, Petrone J, Sarnataro M, et al. MRI Radiomics for the Prediction of Fuhrman Grade in Clear Cell Renal Cell Carcinoma: A Machine Learning Exploratory Study. *J Digit Imaging* (2020) 33(4):879–87. doi: 10.1007/s10278-020-00336-y
12. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-Learn: Machine Learning in Python. *J Mach Learn Res* (2011) 12:2825–30.
13. 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) 212(6):W1–8. doi: 10.2214/AJR.18.20742
14. Moch H, Cubilla AL, Humphrey PA, Reuter VE, Ulbright TM. The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part a: Renal, Penile, and Testicular Tumours. *Eur Urol* (2016) 70(1):93–105. doi: 10.1016/j.eururo.2016.02.029
15. 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
16. Cornelis F, Tricaud E, Lasserre AS, Petitpierre F, Bernhard JC, Le Bras Y, et al. Multiparametric Magnetic Resonance Imaging for the Differentiation of Low and High Grade Clear Cell Renal Carcinoma. *Eur Radiol* (2015) 25(1):24–31. doi: 10.1007/s00330-014-3380-x
17. Parada Villavicencio C, Mc Carthy RJ, Miller Frank H. Can Diffusion-Weighted Magnetic Resonance Imaging of Clear Cell Renal Carcinoma Predict Low From High Nuclear Grade Tumors. *Abdom Radiol (NY)* (2017) 42(4):1241–49. doi: 10.1007/s00261-016-0981-7
18. Pedrosa I, Chou MT, Ngo L, H Baroni R, Genega EM, Galaburda L, et al. MR Classification of Renal Masses With Pathologic Correlation. *Eur Radiol* (2008) 18(2):365–75. doi: 10.1007/s00330-007-0757-0
19. Hoang UN, Mojdeh Mirmomen S, Meirelles O, Yao J, Merino M, Metwalli A, et al. Assessment of Multiphasic Contrast-Enhanced MR Textures in Differentiating Small Renal Mass Subtypes. *Abdom Radiol (NY)* (2018) 43(12):3400–09. doi: 10.1007/s00261-018-1625-x
20. Nketiah G, Elschot M, Kim E, Teruel JR, Scheenen TW, Bathen TF, et al. T2-Weighted MRI-Derived Textural Features Reflect Prostate Cancer Aggressiveness: Preliminary Results. *Eur Radiol* (2017) 27(7):3050–59. doi: 10.1007/s00330-016-4663-1
21. Park H, Kim KA, Jung JH, Rhie J, Choi SY. MRI Features and Texture Analysis for the Early Prediction of Therapeutic Response to Neoadjuvant Chemoradiotherapy and Tumor Recurrence of Locally Advanced Rectal Cancer. *Eur Radiol* (2020) 30(8):4201–11. doi: 10.1007/s00330-020-06835-4
22. 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
23. Kocak B, Durmaz ES, Erdim C, Ates E, Kaya OK, Kilickesmez O. Radiomics of Renal Masses: Systematic Review of Reproducibility and Validation Strategies. *AJR Am J Roentgenol* (2020) 214(1):129–36. doi: 10.2214/AJR.19.21709
24. Purkayastha S, Zhao Y, Wu J, Hu R, McGirr A, Singh S, et al. Diferentiation of Low and High Grade Renal Cell Carcinoma on Routine MRI With an Externally Validated Automatic Machine Learning Algorithm. *Sci Rep* (2020) 10:19503. doi: 10.1038/s41598-020-76132-z
25. Zhao Y, Chang M, Wang R, Xi IL, Chang K, Huang RY, et al. Deep Learning Based on MRI for Differentiation of Low- and High-Grade in Low-Stage Renal Cell Carcinoma. *J Magn Reson Imaging* (2020) 52(5):1542–49. doi: 10.1002/jmri.27153
26. 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
27. Kocak B, Ates E, Durmaz ES, Ulusan MB, Kilickesmez O. Influence of Segmentation Margin on Machine Learning-Based High-Dimensional Quantitative CT Texture Analysis: A Reproducibility Study on Renal Clear Cell Carcinomas. *Eur Radiol* (2019) 29(9):4765–75. doi: 10.1007/s00330-019-6003-8

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 27 October 2021
doi: 10.3389/fonc.2021.758036

The Value of MRI Findings Combined With Texture Analysis in the Differential Diagnosis of Primary Ovarian Granulosa Cell Tumors and Ovarian Thecoma–Fibrothecoma

Nai-yu Li[1], Bin Shi[1], Yu-lan Chen[1], Pei-pei Wang[1], Chuan-bin Wang[1], Yao Chen[1], Ya-qiong Ge[2], Jiang-ning Dong[1] and Chao Wei[1]**

[1] The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, China, [2] Department of the Healthcare, GE of China, Shanghai, China

OPEN ACCESS

Edited by:
Antonio Napolitano, Bambino Gesù Children Hospital (IRCCS), Italy

Reviewed by:
Gaurav Sharma, University of Texas Southwestern Medical Center, United States
Zheng Han, Johns Hopkins University, United States

***Correspondence:**
Chao Wei
weichaoai121@163.com
Jiang-ning Dong
Dongjn@163.com

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

Received: *13 August 2021*
Accepted: *05 October 2021*
Published: *27 October 2021*

Citation:
Li N-y, Shi B, Chen Y-l, Wang P-p, Wang C-b, Chen Y, Ge Y-q, Dong J-n and Wei C (2021) The Value of MRI Findings Combined With Texture Analysis in the Differential Diagnosis of Primary Ovarian Granulosa Cell Tumors and Ovarian Thecoma–Fibrothecoma. Front. Oncol. 11:758036. doi: 10.3389/fonc.2021.758036

Objective: This study aims to explore the value of magnetic resonance imaging (MRI) and texture analysis (TA) in the differential diagnosis of ovarian granulosa cell tumors (OGCTs) and thecoma-fibrothecoma (OTCA–FTCA).

Methods: The preoperative MRI data of 32 patients with OTCA–FTCA and 14 patients with OGCTs, confirmed by pathological examination between June 2013 and August 2020, were retrospectively analyzed. The texture data of three-dimensional MRI scans based on T2-weighted imaging and clinical and conventional MRI features were analyzed and compared between tumor types. The Mann–Whitney *U*-test, χ^2 test/Fisher exact test, and multivariate logistic regression analysis were used to identify differences between the OTCA–FTCA and OGCTs groups. A regression model was established by using binary logistic regression analysis, and receiver operating characteristic curve analysis was carried out to evaluate diagnostic efficiency.

Results: A multivariate analysis of the imaging-based features combined with TA revealed that intratumoral hemorrhage (OR = 0.037), log-sigma-20mm-3D_glszm_SmallAreaEmphasis (OR = 4.40), and log-sigma-2-0mm-3D_glszm_SmallAreaHighGrayLevelEmphasis (OR = 1.034) were independent features for discriminating between OGCTs and OTCA–FTCA ($P < 0.05$). An imaging-based diagnosis model, TA-based model, and combination model were established. The areas under the curve of the three models in predicting OGCTs and OTCA–FTCA were 0.935, 0.944, and 0.969, respectively; the sensitivities were 93.75, 93.75, and 96.87%, respectively; and the specificities were 85.71, 92.86, and 92.86%, respectively. The DeLong test indicated that the combination model had the highest predictive efficiency ($P < 0.05$), with no significant difference among the three models in differentiating between OGCTs and OTCA–FTCA ($P > 0.05$).

Conclusions: Compared with OTCA–FTCA, intratumoral hemorrhage may be characteristic MR imaging features with OGCTs. Texture features can reflect the microheterogeneity of OGCTs and OTCA–FTCA. MRI signs and texture features can

help differentiate between OGCTs and OTCA–FTCA and provide a more comprehensive and accurate basis for clinical treatment.

Keywords: granulosa cell tumor, fibrothecoma, thecoma, sex cord stromal tumors, magnetic resonance imaging, texture analysis

INTRODUCTION

Ovarian sex cord stromal tumors are rare tumors that account for approximately 7% of all ovarian tumors. According to the 2014 World Health Organization (WHO) ovarian tumor histological classification, these tumors are divided into pure stromal tumors, pure sex cord tumors, luteinized thecoma associated with sclerosing peritonitis, and mixed sex cord stromal tumors. Pure stromal tumors include three subtypes: fibroma, cellular fibroma, and thecoma; these are mainly distinguished based on whether they comprise theca cells, lutein cells, fibroblasts, and fibrocytes. This group of tumors has overlapping features in multidirectional differentiation through histology, which makes it difficult to obtain a pathological diagnosis in some cases. Therefore, these tumors are traditionally named ovarian thecoma-fibrothecoma (OTCA–FTCA) (1, 2).

OTCA–FTCA and ovarian granulosa cell tumors (OGCTs) are the most common sex cord stromal tumors and have a low incidence relative to other ovarian tumors. These tumors are usually discovered by chance during gynecological examinations or routine physical examinations as the symptoms are nonspecific. OTCA–FTCA accounts for 0.5–1.0% of ovarian tumors, is generally benign, and has an excellent prognosis after resection, but a small proportion of these tumors (1.0–5.0%) are malignant (3, 4). OGCTs are rare sex cord stromal tumors with a low malignant potential and account for only 5% of all malignant ovarian tumors, with adult and juvenile forms of subtypes. OGCTs have a low degree of malignancy, show growth patterns of benign tumors, and have potentially malignant behaviors, including local invasion, recurrence, and metastasis (5, 6). OTCA–FTCA is mainly found in menopausal women, and less than 10% occur before age 30 (7, 8); however, OGCTs are more common in postmenopausal women, and the juvenile type is rare and typically occurs before 30 years of age. Sometimes these tumors share similar clinical manifestations (such as elevated estrogen levels leading to endometrial hyperplasia and irregular vaginal bleeding). These tumors can have similar imaging findings, such as combined with cystic degeneration, edema, and hemorrhage, which may cause misdiagnosis in radiography and inappropriate choice of treatment of clinicians (9–11). Therefore, the preoperative diagnosis of OTCA–FTCA and OGCTs is particularly important.

Magnetic resonance imaging (MRI) has high resolution in soft tissues that clearly reveal the lesion characteristics, relationship between the tumor and surrounding tissues, and the status of lymph node disease (9). In particular, the semiquantitative parameters deprived from diffusion-weighted imaging (DWI) have gradually become one of the important tools for evaluating ovarian tumors (12). Texture analysis (TA) has been widely adopted in the differential diagnosis of tumors in recent years and is considered to be an effective means to assess tumor heterogeneity. Not only MRI-based texture analysis but also CT texture-based analysis of the whole tumor has demonstrated high sensitivity and specificity for the characterization of ovarian tumors and may assist in characterizing the differences in ovarian tumor patients. The application of MRI-based texture features combined with conventional MRI features may assist in improving the differentiation of ovarian tumors. These findings, in turn, may guide diagnostic protocols for future patients and can help radiologists make appropriate follow-up decisions (3, 4, 7).

Therefore, the purpose of this study was to identify the best features for distinguishing between OTCA–FTCA and OGCTs through conventional MRI, TA, and the combination of the two diagnostic methods to improve the accuracy of preoperative imaging-based diagnoses and help clinicians choose appropriate treatment methods.

MATERIALS AND METHODS

Clinical Information

This retrospective study was approved by the institutional review board of The First Affiliated Hospital of University of Science and Technology of China (USTC), and the requirement of written informed consent was waived. Between June 2013 and August 2020, 1,586 patients with clinically suspected adnexal disease who underwent 3.0-T MR examinations were reviewed through the picture archiving and communication system at the First Affiliated Hospital of the USTC. A total of 46 patients with histologically proven OGCTs (n = 14, 15–71 years of age) and OTCA–FTCA (n = 32, 24–94 years of age) were included in this study. The inclusion criteria were as follows: (1) surgically diagnosed tumor with a known pathological type (according to the 2014 WHO classification of ovarian tumors), (2) no intervention before the MRI examination, (3) lesion that could be measured and segmented on MRI, and (4) signed informed consent form provided before the examination.

MRI Examination

MRI was performed using a 3.0-T system (Signa HDxT, GE Healthcare) with an eight-channel phased array coil. The routine MRI protocols used to assess the pelvic masses included axial T1-weighted imaging (T1WI), axial/sagittal T2-weighted imaging (T2WI), axial fat-suppressed T2WI (FS T2WI), DWI (b value = 0, 1,000 s/mm^2), and multiple phases of contrast-enhanced (LAVA-FLEX) MRI. For the axial images, the transverse plane was perpendicular to the long axis of the uterine body; for the

sagittal images, the longitudinal plane was parallel to the main body of the uterus. If contraindications were excluded, the patients were often given an intramuscular injection of 20 mg scopolamine 15 min before the examination to suppress MRI motion artefacts caused by peristalsis. Contrast-enhanced pelvic imaging was acquired at the arterial, venous, and delayed phases of contrast medium enhancement in axial planes, which were acquired at 25, 60, and 120 s after the intravenous injection of 0.1 mmol/kg gadodiamide (Omniscan, GE Healthcare) using an Ulrich power injector. Some of the scanning sequences and parameters are shown in **Table 1**.

Radiological Evaluation

Two radiologists (YuC and BS, with 10 and 7 years of experience in gynecological imaging, respectively) who were blinded to the histological results independently analyzed the MRI data of each participant, and discrepancies were resolved by consensus. The following MRI features were recorded and analyzed for the two groups: size (the maximum diameter of the tumor and the shortest perpendicular diameter measured on T2WI, the maximum upper and lower diameter of the tumor measured on sagittal T2WI, and the average size of the aboved diameters), endometrial hyperplasia (endometrium thickness greater than 5 mm after menopause and greater than 16 mm in premenopausal women) (12), apparent diffusion coefficient (ADC) value (10^3 mm^2/s) [mean value obtained from three measurements of a region of interest (ROI) manually placed in the solid components of the tumors and myometrium, and the calculated ratio; the ROI was drawn using GE AW4.5 workstation Functool-MADC software, and attempts were made to avoid tumor necrosis and cystic areas], enhancement degree, T2WI signal, and DWI signal of the solid component of the tumors (hypointense, isointense, or hyperintense compared with the myometrium at the same level), location (left or right), degree of cystic components (graded as 0–4°; grade 0 = no cystic change; grade 1 = area with cystic changes was ≤25%; 25% < grade 2 ≤ 50%; 50% < grade 3 ≤ 75%; and 75% < grade 4), cystic form (no cyst, mainly small sac, mainly large and mixed; small sac ≤1.0 cm, large sac >1.0 cm, or a mix of both), intratumoral hemorrhage (present or absent), and age (years, mean ± standard deviation, SD).

Texture Feature Extraction

The images of OTCA–FTCA and OGCTs were manually segmented, and volumes were extracted using *ITK* Snap software (3.8.0, http://www.itksnap.org). ROIs were delineated around the tumor boundary for each section by two radiologists (YuC and BS). After tumor segmentation, AK software (Analysis Kit Version: 3.2.0; GE Healthcare) was used for texture feature extraction, and 1,316 features, such as the mean, entropy, energy, skewness, kurtosis, and standard deviation, were obtained in this study.

Statistical Analysis

Continuous and categorical variables were compared using the *t*-test and χ^2/Fisher's exact test, respectively. Continuous variables are expressed as the mean ± standard deviation, and categorical variables are expressed as the frequency and percentage (%). Continuous variables were first tested for normality to understand the data distribution, and the variables were tested as follows: (1) an independent-sample *t*-test was used to compare variables both conforming to a normal distribution, and (2) the Mann–Whitney *U*-test was used to compare variables conforming to a skewed distribution and variables conforming to a skewed distribution with those conforming to a normal distribution. Continuous and categorical variables showing significant differences were analyzed by multivariate logistic regression analysis with the forward step method to screen for independent discriminant features, which were used to construct the discriminating model. Receiver operating characteristic (ROC) curve and area under the curve (AUC) analyses were performed with MedCalc (version 19.5.3, https://www.medcalc.org/) to determine the overall diagnostic performance of the radiographic model, texture model, and combined model. SPSS 26.0 software (version 20.0, IBM, Armonk, NY, USA) was used for statistical analysis, and P <0.05 was considered statistically significant. The intraclass correlation coefficient (ICC) was used to evaluate the consistency between evaluator 1 and evaluator 2, and an ICC between 0.81 and 1.00 indicated good agreement.

RESULTS

Pathological, Clinical, and Imaging Findings

The pathological diagnoses of all OTCA–FTCA and OGCTs were made by a pathologist (YuC, with 8 years of experience in gynecological tumors) according to the 2014 WHO ovarian sex cord stromal tumor histological classification. Finally, a total of 32 patients with OTCA–FTCA (mean age, 52.93 ± 12.39 years) and 14 patients with OGCTs (mean age, 49.93 ± 19.19 years) were enrolled in this study. The 14 patients with OGCTs included 12 adult and two juvenile patients, with eight patients with tumors in the right ovary,

TABLE 1 | Partial list of MRI parameters.

SEQUENCE	TE (ms)	TR (ms)	Freq × phase	Nex	FOV	Slice thickness	Interval	Flip angle
FS T2WI	72.5	5,000	320 × 256	2	24 × 24	6	2	90°
T2WI	72.5	4,600	320 × 256	2	24 × 24	6	2	90°
Osag T2WI	72	4,500	320 × 320	2	28 × 28	4	1	90°
T1WI	7.5	500	352 × 192	2	32 × 32	6	2	90°
DWI (b = 0, 1,000 s/mm^2)	/	5,000	96 × 130	6	32 × 32	6	2	90°
Oax LAVA-FLEX	1.4	5.8	320 × 224	1	34 × 31	4	0	15°
Osag LAVA-FLEX	1.3	6.8	268 × 224	1	28 × 25	4	0	15°

six patients with tumors in the left ovary, and three patients with endometrial hyperplasia (21%). There were 11 patients with cystic–solid changes (two small sacs, one large sac, and eight mixed types) (**Figure 1**) and 11 patients with hemorrhage signal (76%) in the tumor (**Figure 2**); the enhancement degree of the solid components of the tumor was mainly mild to moderate (eight patients with mild enhancement and six patients with moderate enhancement), with no patients with marked enhancement. Among the 32 patients with OTCA–FTCA, 23 patients had tumors in the right ovary, nine patients had tumors in the left ovary, and three patients had endometrial hyperplasia (9%); 18 solid masses (**Figure 3**) were observed, with 14 showing mainly cystic–solid changes (one small cyst, 10 large cysts, and three mixed). There were five patients (16%) with hemorrhage signals in the tumor. As shown in **Table 2**, the following four MRI-based characteristics were significantly different between the OGCTs and OTCA–FTCA groups: (1) the mean ADC value of the solid component (z = -1.982, P = 0.047), (2) the degree of enhancement of the solid component (χ^2/9.084, P = 0.003), (3) the cystic form (Fisher/0.006, P = 0.008), and (4) the presence of intratumoral hemorrhage (Fisher/0.000, P=0.000).

Diagnostic Performance of the Texture Features

Least absolute shrinkage and selection operator (Lasso) regression was performed in R (3.6.1, http://www.r-project.org) to reduce the dimensionality of the features and filter and remove redundancy parameters ($|r| > 0.8$) to reduce the impact of data overfitting. First, the Mann–Whitney U-test was applied to the features to explore whether the features were significantly different between the two groups, and 123 features with p <0.05 were retained. Second, univariate logistic regression was applied to explore whether the features were discriminative between the two groups, and 78 features with p <0.05 were retained. Third, minimum redundancy and maximum correlation were applied to eliminate the redundant features and retain the features that were highly correlated with the label, and 10 features were retained. Then, the retained features were enrolled in backward stepwise multivariate logistic regression, and the final model was constructed. The explanation of the texture analysis features is shown in **Table 3**.

The ICC was used to evaluate the consistency between radiologist 1 and radiologist 2 and was 0.81–1.00 (P < 0.001), indicating good consistency. Finally, the average of the two sets of data was used as the new texture data for statistical analysis. As shown in **Table 4**, the following six texture features were significantly different between the OGCTs and OTCA–FTCA groups: (1) log-sigma-2-0-mm-3D_glszm_SmallAreaEmphasis (SAE) (z = -4.201, P = 0.000), (2) log-sigma-2-0-mm-3D_glszm_SmallAreaHighGrayLevelEmphasis (z = -3.187, P = 0.340), (3) log-sigma-3-0-mm-3D_glcm_InverseVariance

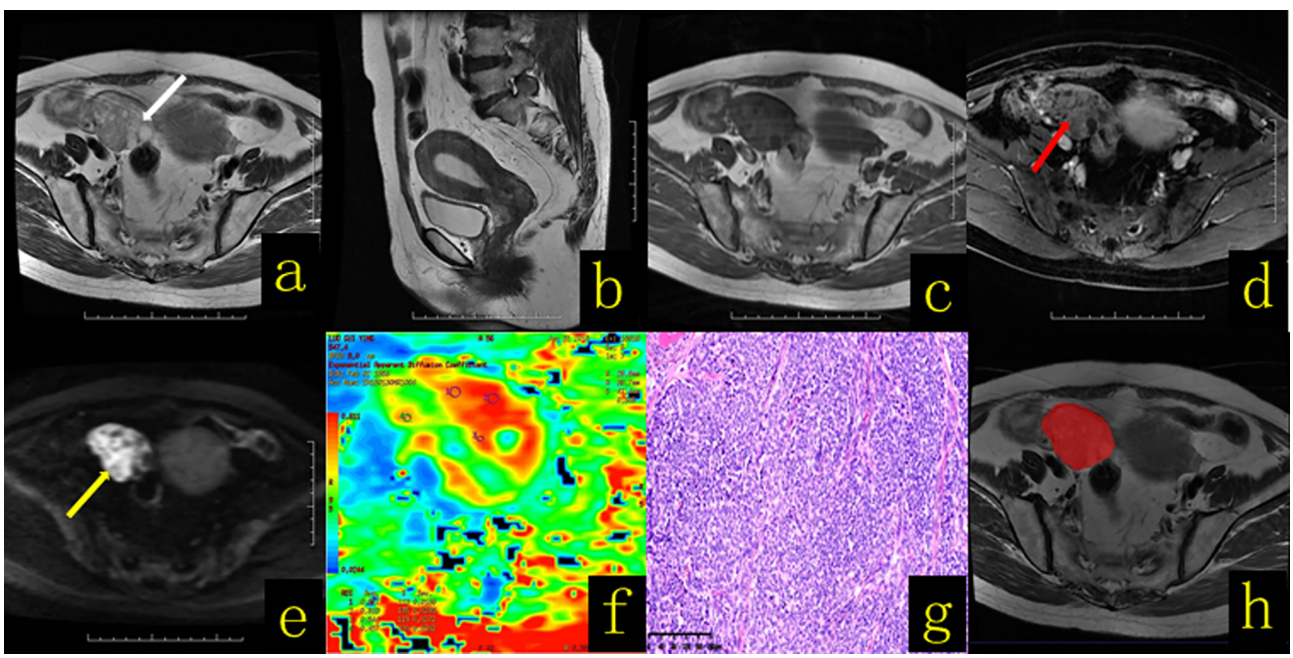


FIGURE 1 | A 61-year-old female patient with an ovarian granulosa cell tumor. **(A)** Axial T2WI revealed a cystic solid mass in the right adnexal region that manifested with a "spongy" or "honeycomb" change (white arrow). **(B)** Sagittal T2WI showed thickening of the endometrium to a thickness of approximately 1.9 cm. **(C)** Axial T1WI revealed a cystic solid mass with a hypo–isointense signal. **(D)** On contrast-enhanced fat-suppressed T1WI, the solid components (red arrow) of the lesion showed mild and moderate enhancement, with a region resembling the myometrium. **(E)** On DWI-MRI (b = 1,000 s/mm^2), the solid part of the lesion appeared hyperintense (yellow arrow), and the cystic part appeared hypointense. **(F)** The apparent diffusion coefficient (ADC) map showed that the average ADC value of the diffuse high-signal area was approximately 0.7 × 10^{-3} mm^2/s. **(G)** Hematoxylin and eosin (H&E) staining (×100) showed that the tumor cells appeared as large islands, diffusely distributed in nests and rich in interstitial separation and blood vessels. **(H)** The texture analysis target area was delineated throughout the whole tumor layer by layer.

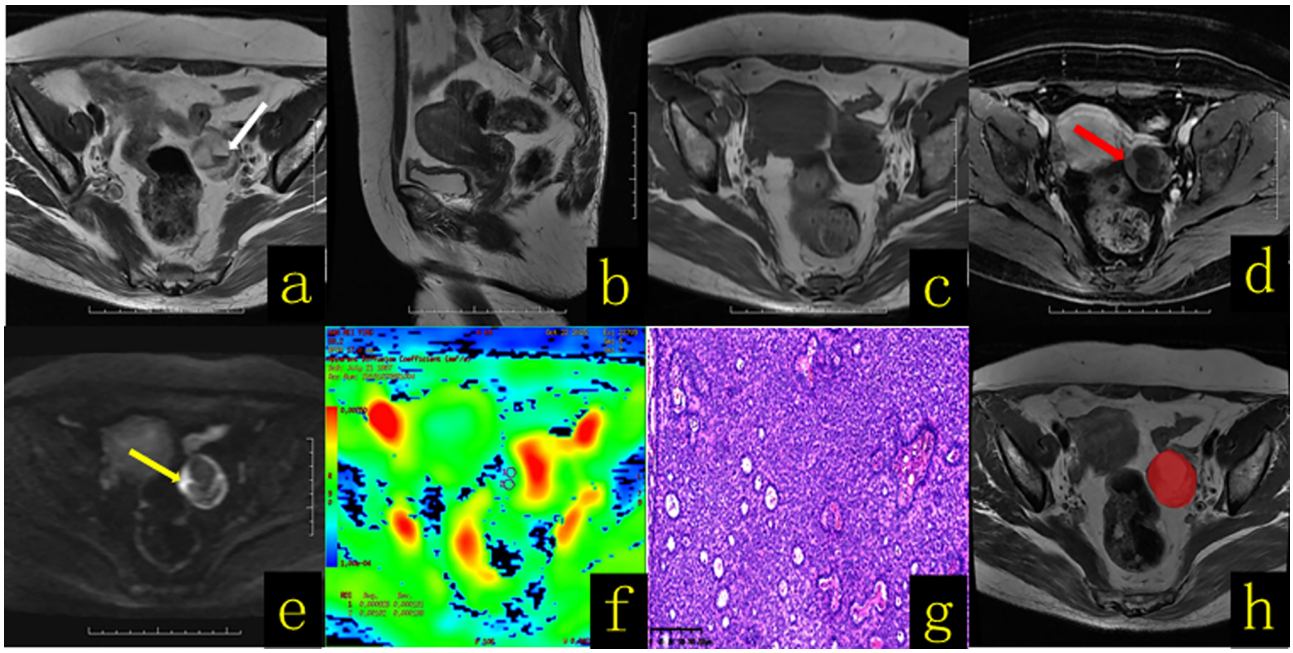


FIGURE 2 | A 58-year-old female patient with an ovarian granulosa cell tumor. **(A)** Axial T2WI revealed a well-defined cystic solid mass in the left adnexal region, with fluid–fluid levels (hemorrhagic content, white arrow). **(B)** Sagittal T2WI showed no thickening of the endometrium. **(C)** Axial T1WI revealed a cystic solid mass with a hypo–isointense signal. **(D)** On contrast-enhanced fat-suppressed T1WI, the solid components (red arrow) of the lesion showed mild enhancement. **(E)** On DWI-MRI (b = 1,000 s/mm^2), the solid part of the lesion (yellow arrow) appeared hyperintense. **(F)** The apparent diffusion coefficient (ADC) map showed that the average ADC value of the diffuse high-signal area was approximately 1.1 × 10^{-3} mm^2/s. **(G)** Hematoxylin and eosin (H&E) staining (×100) showed that the tumor cells were solid tubular structures, and the tubules were composed of uniform cells containing Call–Exner bodies. **(H)** The texture analysis target area was delineated throughout the whole tumor layer by layer.

(z = -3.342, P = 0.001), (4) wavelet-LLH_glcm_MCC (z = -4.106, P = 0.001), (5) wavelet-HLH_glszm_SmallAreaHighGrayLevelEmphasis (z = -2.984, P = 0.003), and (6) wavelet-HLL_glszm_LowGrayLevelZoneEmphasis (z = -3.103, P = 0.002).

Diagnostic Performance of the Predictive Models Based on MRI Characteristics, Texture Features, and Combined Features

The variables with significant differences in the univariate analysis were included in the multivariate logistic regression analysis for screening. As shown in **Table 5**, the overall imaging-based diagnosis (IBD) and overall TA prediction models based on MRI characteristics and texture features were established, respectively: (Y-IBD) = -10.04 + 6.67 × ADC (average) + 4.67 × enhancement degree (solid) (mild = 0, moderate = 1, marked = 3) - 4.63 × intratumoral hemorrhage (present = 0, absent = 1), and (Y-TA) = -11.39 + 33.18 × log-sigma-2-3D_glszm_SmallAreaEmphasis ($x \pm s$) - 0.03 × log- sigma-2-0mm-3D_glszm_SmallAreaHighGrayLevelEmphasis ($x \pm s$). Three IBD and two TA predictive factors were simultaneously included in the multivariate logistic regression analysis, and the combined prediction model was established: (Y-Combine) = -12.33 + 30.76 × log-sigma-20mm-3D_glszm_SmallAreaEmphasis ($x \pm s$) - 0.03 × log-sigma-2-0mm-3D_glszm_SmallAreaHighGrayLevelEmphasis ($x \pm s$) + 3.31 × intratumoral hemorrhage (present = 0, absent = 1). The three prediction models established in this study could accurately predict OGCTs and OTCA–FTCA (P < 0.05). The results of the DeLong test showed that the efficacies of Y-IBD, Y-TA, and Y-Combine were not significantly different (P > 0.05; **Figures 4**, **5** and **Tables 5**, **6**).

DISCUSSION

OTCA–FTCA and OGCTs are the most common sex cord stromal tumors and have a low incidence relative to other ovarian tumors. The radiological knowledge of those rare ovarian tumors is still limited in the reported literature; furthermore, the imaging findings of the two entities are similar. Herein we performed a retrospective review of the MRI findings of 32 patients with OTCA–FTCA and 14 patients with OGCTs in this study at our single institution within 7 years. To the best of our knowledge, this is the first study to describe the detailed MRI sign and TA characteristic in this samples.

In our study, the clinical characteristics [age (years), size (maximum), size (average), menopausal status, presence of endometrial hyperplasia, and location] were compared, and there were no significant differences between the two tumors, indicating that they have similar clinical characteristics, as shown in **Table 2**. Combined with literature reports, we found the following: (1) The incidence of intratumoral hemorrhage in this group of OGCTs was

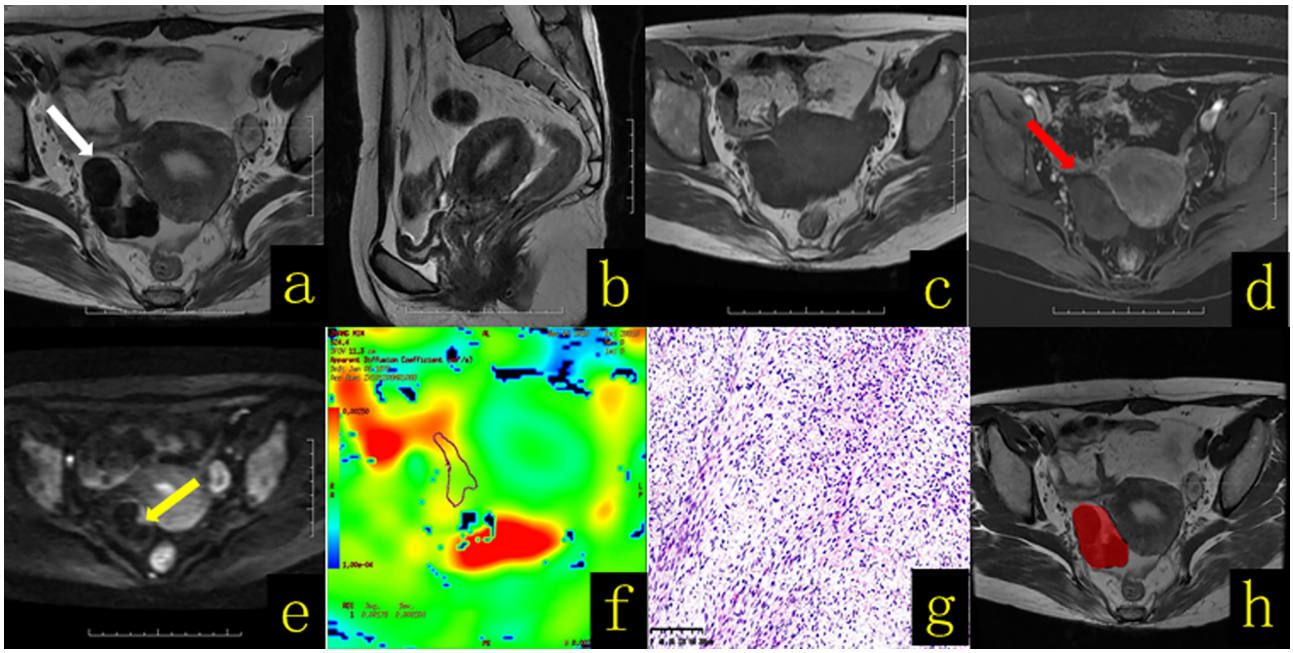


FIGURE 3 | A 65-year-old female patient with right ovarian thecoma–fibrothecoma. **(A)** Axial T2WI revealed a solid mass in the right adnexal region (white arrow), showing mainly a low-signal mass with a semiarc shape and high signal at the left front edge. **(B)** Sagittal T2WI showed thickening of the endometrium to a thickness of approximately 1.2 cm. **(C)** Axial T1WI revealed a solid mass with hypo–isointense signal (white arrow). **(D)** On contrast-enhanced fat-suppressed T1WI, the solid components (red arrow) of the lesion showed mild enhancement. **(E)** On DWI-MRI (b = 1,000 s/mm^2), the solid part of the lesion of the left front edge appeared hyperintense (yellow arrow). **(F)** The apparent diffusion coefficient (ADC) map showed that the average ADC value of the diffuse high-signal area was approximately 1.78 × 10^{-3} mm^2/s. **(G)** Hematoxylin and eosin (H&E) staining (×100) showed that the tumor was composed of spindle cells and collagen fibers arranged in a mat-like pattern with interwoven bundles, and hyaline degeneration of fibrous tissue bands and intercellular edema were observed to varying degrees. The tumor cell nucleus was fusiform to oval, with sparse cytoplasm and containing a small amount of lipids; the mitotic index was <3/10 HPF. **(H)** The texture analysis target area was delineated throughout the whole tumor layer by layer.

as high as 76%, which is higher than that reported in the literature (13). Intratumoral hemorrhage mainly manifested as high signal on T1WI and high signal or low signal on T2WI, and the fluid–fluid level due to hemorrhage could be seen in some lesions. In comparison, the incidence of intratumoral hemorrhage in OTCA–FTCA was only 16% (5/32). The multivariate logistic regression analysis found that the presence of intratumoral hemorrhage could help diagnose OGCTs (OR = 0.12, 95% CI: 0.001–0.284), which is consistent with previous reports that intratumoral hemorrhage is a typical feature of these tumors (14); (2) OTCA–FTCA is composed of theca cells, lutein cells, and fibroblasts. This group of tumors is prone to secondary degenerative changes, such as tumor stromal edema and mucinous degeneration, which may lead to high ADC values (1.50 ± 0.32 × 10^3 mm^2/s). In contrast, OGCTs are low-grade malignant tumors that histologically show diffuse, island, beam, follicular, and sarcoma-like growth patterns. These patterns often exist mixed, and the relatively tight arrangement results in more restricted water molecule diffusion with lower ADC values (1.27 ± 0.37 × 10^3 mm^2/s) than that of OTCA–FTCA. Therefore, the average ADC value was significantly different between the two tumors (Z = -1.982, P = 0.047) (15). When the ADC value was ≤1.34 × 10^3 mm^2/s, its sensitivity for diagnosing OGCTs was 71.34%, and the specificity was 65.62% (AUC = 0.685, 95% CI: 0.532 to 0.814, P = 0.048); (3) In our group, 94% (30/32) of the OTCA–FTCA tumors were mildly enhanced, 6% (2/32) were moderately enhanced, and none showed marked enhancement. In comparison, 57% (8/14) of the OGCTs were mildly enhanced, 43% were moderately enhanced, and none showed marked enhancement. There was a significant difference in the degree of enhancement between the two tumor types (OR = 0.89, 95% CI: 0.015–0.527). It is possible that OTCA–FTCA contains fibrous components, resulting in a lower blood supply and lower enhancement than OGCTs. This is also consistent with previous reports that OTCA–FTCA tumors have a low blood supply, resulting in mild enhancement on MRI (16–18); (4) OGCTs are mostly solid or cystic–solid, and it has been reported in the literature that a "honeycomb" or "sponge" cyst is the characteristic imaging manifestation (19). OTCA–FTCA is often prone to secondary cystic transformation when the tumor volume is large. Some scholars have reported that the cystic transformation rate is 76% (19/25) (20), so the tumor often appears as a cystic–solid or cystic mass, which may be preoperatively misdiagnosed as OGCTs or other ovarian tumor. Other scholars have divided these tumors into solid, cystic, and cystic–solid masses according to the degree of the cystic component. Cystic–solid masses are divided into intratumoral cysts and extratumoral cysts according to whether the cysts are located in the tumor. Intratumoral cysts are divided into peripheral, central,

TABLE 2 | Details of the clinical and MR imaging-based characteristics of 14 histologically proven OGCTs and OTCA–FTCA in 32 patients.

Characteristics	Category	OGCTs (n = 14)	OTCA–FTCA (n = 32)	^{2}/Fisher/z value	P-value
Age (years)		49.93 ± 19.19	52.93 ± 12.39	z/-0.478	0.632
Size (maximum)	/	6.65 ± 4.60	8.08 ± 5.33	z/-0.967	0.333
Size (average)	/	6.47 ± 4.74	7.96 ± 5.18	z/-1.146	0.252
Mean ADC (10^3 s/mm^2)	/	1.27 ± 0.37	1.50 ± 0.32	z/-1.982	0.047
ADC (10^3 s/mm^2, ratio)	/	0.93 ± 0.24	1.05 ± 0.27	z/-1.695	0.090
Menopause	Postmenopausal	10 (71%)	21 (66%)	χ^2/0.149	0.699
	Premenopausal	4 (29%)	11 (34%)		
Endometrial hyperplasia	Present	3 (21%)	3 (9%)	Fisher/0.350	0.264
	Absent	11 (79%)	29 (91%)		
T2WI intensity (solid)	Hypointense	2 (14%)	12 (38%)	Fisher/0.102	0.084
	Isointense	6 (43%)	11 (34%)		
	Hyperintense	5 (36%)	3 (9%)		
	Mixed signal	1 (7%)	6 (19%)		
Location	Right	8 (57%)	23 (72%)	Fisher/0.495	0.327
	Left	6 (43%)	9 (28%)		
DWI intensity (solid)	Isointense	1 (7%)	5 (16%)	Fisher/0.175	0.149
	Hyperintense	1 (7%)	9 (28%)		
	Mixed	12 (86%)	18 (56%)		
Enhancement degree (solid)	Mild	8 (57%)	30 (94%)	χ^2/9.084	0.003
	Moderate	6 (43%)	2 (6%)		
	Marked	0 (0%)	0 (0%)		
Degree of cystic components	None	3 (21%)	18 (56%)	Fisher/0.149	0.229
	<25%	4 (29%)	4 (13%)		
	25–50%	1 (7%)	3 (9%)		
	50%~75%	1 (7%)	1 (3%)		
	>75%	5 (36%)	6 (19%)		
Cystic form	Small cyst	2 (14%)	1 (3%)	Fisher/0.006	0.008
	Large cyst	1 (7%)	10 (31%)		
	Mixed	8 (57%)	3 (9%)		
Intratumoral hemorrhage	Present	11 (76%)	5 (16%)	Fisher/0.000	0.000
	Absent	3 (24%)	27 (84%)		

and diffuse types according to their location. The study showed that peritumoral cysts are a characteristic MRI sign (21). In our study, the types of cysts were divided into five degrees according to the degree of cystic degeneration (no cyst: 0°, 0–25%: 1°, 25–50%: 2°, 50–75%: 3°, and greater than 75%: 4°), and the forms of cystic transformation were divided into four forms (no cystic transformation, small cyst, large cyst, and mixed). Between the two tumor types, there was no significant difference in the degree of cystic transformation (Fisher = 0.149, P = 0.229), but there was a significant difference in the form of cystic transformation (Fisher = 0.006, P = 0.008), indicating that OGCTs mainly demonstrated mixed cystic changes, while OTCA–FTCA predominantly exhibited macrocystic changes. In this study, a weak correlation existed between tumor size and the degree of cystic transformation in the OGCTs group (Kendall's tau-b = 0.618, P < 0.001), and no correlation was observed in the OTCA–FTCA group (Kendall's tau-b = -0.025, P = 0.857). It is inconsistent with related reports (17) and may be caused by the small sample size.

In the multivariate logistic regression analysis, the IBD model established had an AUC of 0.935, and its sensitivity, specificity, and Youden index were 85.71%, 93.75%, and 0.794 (95% CI: 0.822 to 0.987, P < 0.0001), respectively, so the significant features, such as the mean ADC value, enhancement degree, and presence of intratumoral hemorrhage, were important predictors to distinguish between OGCTs and OTCA–FTCA.

TA is different from traditional empirical image analysis based on observations with the naked eye. TA can provide a large amount of imaging information that cannot be recognized by the naked eye by quantitatively analyzing the grayscale information of medical images, realizing the conversion from images to data, and constructing labels to describe the details of the lesion features. Thus, this information could be of value in helping clinicians develop reasonable treatment strategies (22). In recent years, TA has been regarded as an effective means to assess tumor heterogeneity. This method can be used to evaluate the gray-level intensity and position of the pixels within an image to derive texture features that provide a measure of intralesional heterogeneity. TA data are easy to obtain, and no additional imaging is required. In addition, TA plays a relatively important role in evaluating clinical curative effects and predicting prognosis. Many researchers have conducted excellent research, especially with radiomics, in predicting the development trends of tumor lesions (23, 24). The TA in the present study is based on the T2WI sequence because conventional T2WI can reveal the rich histopathological characteristics of tumors, for example, by determining the water content, degree of fibrotic change, necrosis, and hemorrhage (15).

As shown in **Table 4**, the univariate analysis demonstrated that six texture features were significantly different between the OGCTs and OTCA–FTCA groups (P < 0.05). Among the six features, the log-sigma-2-0-mm-3D_glszm_SmallAreaEmphasis, log-sigma-2-0-mm-3D_glszm_SmallAreaHighGrayLevel Emphasis, and log-sigma-3-0-mm-3D_glcm_InverseVariance were derived from the image transform type of Laplacian of Gaussian. The wavelet-LLH_glcm_MCC, wavelet-

TABLE 3 | Explanation of the texture analysis features.

Image type	Features	Feature explanation
log-sigma-2-0-mm-3D	glszm_SmallAreaEmphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}\frac{p(i,j)}{j^2}}{N_z}$	Small area emphasis (SAE): SAE is a measure of the distribution of small size zones, with a greater value indicative of much smaller size zones and more fine textures
	glszm_SizeZoneNonUniformityNormalized $\frac{\sum_{j=1}^{N_s}(\sum_{i=1}^{N_g}p(i,j)^2)}{N_z}$	SZNN measures the variability of size zone volumes throughout the image, with a lower value indicating more homogeneity among zone size volumes in the image. This is the normalized version of the SZN formula
	glszm_SmallAreaHighGrayLevelEmphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}\frac{p(i,j)i^2}{j^2}}{N_z}$	SAHGLE measures the proportion in the image of the joint distribution of smaller size zones with higher gray-level values
log-sigma-3-0-mm-3D	glcm_InverseVariance $\sum_{k=1}^{N_g-1}\frac{p_{x-y}(k)}{k^2}$	Reflects the local variation of the image texture; so, if more uniformity was found in the different regions of the image texture, this indicates that the change is slower, the value will be larger, and vice versa
wavelet-LLH	glcm_MCC $\sqrt{\sum_{k=0}^{N_g}\frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}}$	Maximal correlation coefficient (MCC). The maximal correlation coefficient is a measure of complexity of the texture and 0 ≤ MCC ≤ 1. In case of a flat region, each GLCM matrix has shape (1, 1), resulting in just 1 eigenvalue.
wavelet-HLH_	glszm_SmallAreaHighGrayLevelEmphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}\frac{p(i,j)i^2}{j^2}}{N_z}$	Measures the proportion in the image of the joint distribution of smaller size zones with higher gray-level values
wavelet-HLL	glszm_LowGrayLevelZoneEmphasis $\frac{\sum_{i=1}^{N_g}\sum_{j=1}^{N_s}\frac{p(i,j)}{i^2}}{N_z}$	Measures the distribution of lower gray-level size zones, with a higher value indicating a greater proportion of lower gray-level values and size zones in the image
lbp-3D-k	glszm_ZonePercentage $\frac{N_z}{N_p}$	Measures the coarseness of the texture by taking the ratio of the number of zones and number of voxels in the region of interest (ROI). Values are in the range 1Np ≤ ZP ≤ 1, with higher values indicating a larger portion of the ROI consisting of small zones (indicates a finer texture)
lbp-3D-k	firstorder_Kurtosis $\frac{\mu_4}{\sigma^4}=\frac{\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^4}{(\frac{1}{N_p}\sum_{i=1}^{N_p}(X(i)-\bar{X})^2)^2}$	Kurtosis is a measure of the "peakedness" of the distribution of values in the image region of interest. A higher kurtosis implies that the mass of the distribution is concentrated towards the tail (s) rather than towards the mean. A lower kurtosis implies the reverse: that the mass of the distribution is concentrated towards a spike near the mean value
original_shape_Sphericity	Sphericity $\frac{\sqrt[3]{36\pi V^2}}{A}$	Sphericity is a measure of the roundness of the shape of the tumor region relative to a sphere. It is a dimensionless measure, independent of scale and orientation. The value range is 0 < sphericity ≤ 10 <sphericity ≤ 1, where a value of 1 indicates a perfect sphere (a sphere has the smallest possible surface area for a given volume, compared to other solids)

Reference: https://pyradiomics.readthedocs.io/en/latest/features.html.

HLH_glszm_SmallAreaHighGrayLevelEmphasis, and wavelet-HLL_glszm_LowGrayLevelZoneEmphasis were derived from the image transform type of wavelet. The features belong to the gray level co-occurrence matrix (GLCM), and the gray level size zone matrix can assess the second-order joint probability function and quantify gray level zones in the image (25). A gray level zone is defined as the number of connected voxels that share the same gray level intensity (26). In the multivariate logistic regression analysis with the forward step method, we found that two features from the image transform type of Laplacian of Gaussian—log-sigma-2-0-mm-3D_glszm_SmallAreaEmphasis and Log-sigma-20mm-3D_glszm_SmallAreaHighGrayLevelEmphasis—are independent risk predictors for distinguishing between OGCTs and OTCA–FTCA ($P < 0.05$). The Laplacian operator can highlight areas in the image where the intensity changes rapidly. The log-sigma-2-0-mm-3D_glszm_SmallAreaEmphasis and log-sigma-20mm-3D_glszm_SmallAreaHighGrayLevelEmphasis describe the distribution of small size zones and the proportion of the joint distribution of smaller size zones with higher gray level values, respectively (27). In our study, the log-sigma-2-0-mm-3D_glszm_SmallAreaEmphasis value of OTCA–FTCA was

TABLE 4 | Results of the univariate analysis of texture features that were significantly different between the OGCTs and OTCA–FTCA groups.

Features	OGCTs	OTCA–FTCA	Mann–Whitney *U*	*Z*-value	*P*-value
log-sigma-2-0-mm-3D_glszm_SmallAreaEmphasis	0.38 ± 0.094	0.70 ± 0.26	50.000	-4.201	0.000
log-sigma-2-0-mm-3D_glszm_SizeZoneNonUniformityNormalized	0.16 ± 0.059	397.89 ± 676.32	184.000	-0.955	0.340
log-sigma-2-0-mm-3D_glszm_SmallAreaHighGrayLevelEmphasis	92.85 ± 87.99	39.93 ± 73.47	92.000	-3.187	0.001
log-sigma-3-0-mm-3D_glcm_InverseVariance	0.33 ± 0.053	1.55 ± 1.35	84.000	-3.342	0.001
wavelet-LLH_glcm_MCC	0.64 ± 0.12	604.17 ± 873.98	52.000	-4.106	0.000
wavelet-HLH_glszm_SmallAreaHighGrayLevelEmphasis	52.32 ± 29.84	56.13 ± 203.67	99.000	-2.984	0.003
wavelet-HLL_glszm_LowGrayLevelZoneEmphasis	0.04 ± 0.06	3.93 ± 6.73	94.000	-3.103	0.002
lbp-3D-k_glszm_ZonePercentage	0.009 ± 0.003	99.68 ± 195.12	202.000	-0.525	0.599
lbp-3D-k_firstorder_Kurtosis	8.73 ± 4.06	99.65 ± 163.28	206.000	-0.430	0.667
original_shape_Sphericity	0.75 ± 0.04	24.14 ± 34.74	212.000	-0.286	0.775

TABLE 5 | Multivariate logistic regression and receiver operating characteristic curve analysis for the overall IBD, overall TA, and combined IBD with TA models.

Features	Multivariate logistic regression analysis				Receiver operating characteristic analysis		
	B	*P*-value	Odds ratio	95% CI	AUC	Specificity	Sensitivity
Overall IBD							
Mean ADC (10^3 s/mm^2)	6.67	0.015	0.001	0.000 to 0.232	0.685	71.43	65.62
Presence of intratumoral hemorrhage	-4.63	0.020	0.012	0.001 to 0.284	0.815	78.57	84.37
Enhancement degree (solid)	4.67	0.004	102.596	2.055 to 5,121.212	0.683	42.86	93.75
Pre model					0.935	85.71	93.75
Overall TA							
Log-sigma-20mm-3D_glszm_SmallAreaEmphasis	33.18	0.009	3.91	6.540 to 0.0002	0.885	85.71	84.37
Log-sigma-20mm-3D_glszm_SmallAreaHighGrayLevelEmphasis	-0.03	0.036	1.032	1.002 to 1.062	0.795	100.00	71.87
Pre model					0.944	92.86	93.75
Combined IBD and TA							
Presence of intratumoral hemorrhage	3.31	0.030	0.037	0.002 to 0.721	0.815	78.57	84.37
Log-sigma-20mm-3D_glszm_SmallAreaEmphasis	30.76	0.024	4.40	1.089 to 0.018	0.885	85.71	84.37
Log-sigma-20mm-3D_glszm_SmallAreaHighGrayLevelEmphasis	-0.03	0.047	1.034	1.000 to 1.068	0.795	100.00	71.87
Combined model					0.969	92.86	96.87

IBD, imaging-based diagnosis; TA, texture analysis; Pre, prediction.

significantly lower than that of OGCTs, which means that smaller size zones and fine textures were observed in the solid lesions of OTCA–FTCA composed of similar theca cells, lutein cells, and fibroblasts (18). OTCA–FTCA also had significantly less intratumoral hemorrhage than OGCTs in the present study ($P < 0.05$). For the log-sigma-20mm-3D_glszm_SmallAreaHighGrayLevelEmphasis value, the OGCTs obviously contained a greater proportion of the joint distribution of smaller size zones with higher gray level values on T2WI scans than OTCA–FTCA ($P < 0.05$). The log-sigma-20mm-3D_glszm_SmallAreaHighGrayLevelEmphasis is a quantitative index used to compensate for the shortage of MRI findings on T2WI based on solid or cystic components that can be compared. Then, the TA-based predictive model was obtained and had a diagnostic performance/AUC, specificity, and sensitivity of 0.944, 92.86%, and 93.75%, respectively ($P < 0.05$).

The AUC of the IBD and TA combined prediction model to distinguish between OGCTs and OTCA–FTCA was 0.969. When compared with MRI features or TA parameters alone, the combined model showed no significant difference, even though the sensitivity and specificity of the combination were improved to some extent. Therefore, it is believed that the diagnostic performance of the combination model was similar to that of MRI-IBD or T2WI-TA features alone. Furthermore, the diagnostic performance of T2WI-TA parameters was similar to that of MRI-IBD features in helping to distinguish between OGCTs and OTCA–FTCA, which may be less strongly associated with the sample size. However, TA can provide another method to identify OGCTs and OTCA–FTCA.

The present study has several limitations. First of these is the limited study sample size (14 patients with OGCTs and 32 patients with OTCA–FTCA) due to the low incidence of these tumors relative to other ovarian tumors. It might have influence on the final results, such as the rigor of ROC curve analysis. Second, there was an inherent selection bias because the retrospective study was conducted in one institution. We urge the clarification of the imaging findings in larger population-based studies. Third, the ROIs of the ADC and TA were

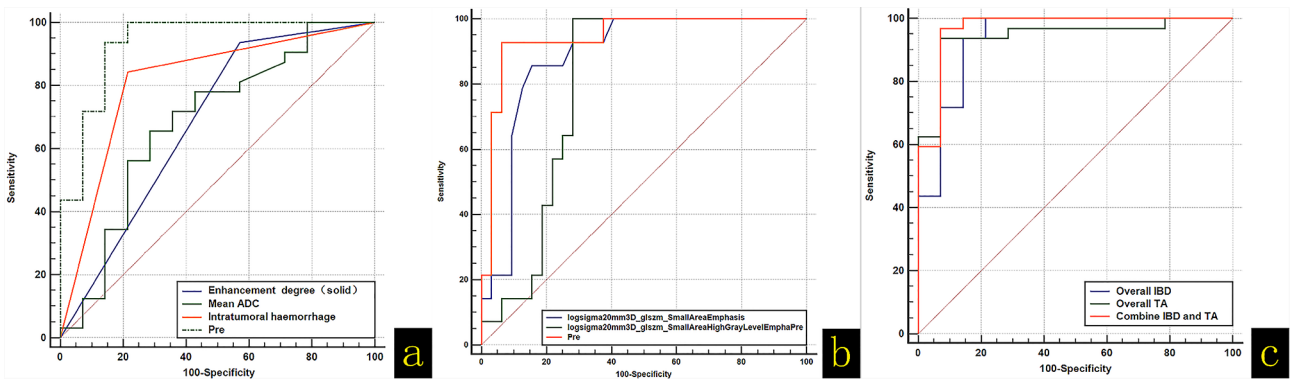


FIGURE 4 | **(A)** (ROC) curve analysis of the diagnostic abilities of apparent diffusion coefficient values (average), enhancement degree, presence of intratumoral hemorrhage, and the prediction models. **(B)** ROC curve analysis of Log-sigma-2-0mm-3D_glszm_SmallAreaEmphasis, Log-sigma-2-0mm-3D_glszm_SmallAreaHighGrayLevelEmphasis, and the prediction models. **(C)** ROC curve analysis of the overall imaging-based diagnosis (IBD), overall texture analysis (TA), and combined IBD with TA models.

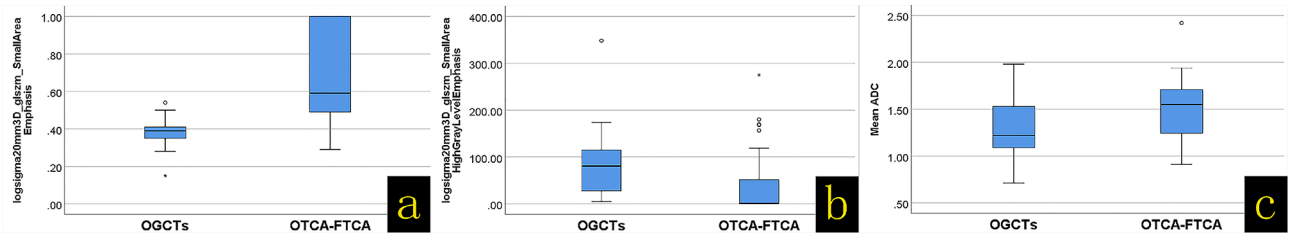


FIGURE 5 | **(A)** Box-and-whisker plot of the Log-sigma-2-0mm 3D_glszm_SmallAreaEmphasis difference between OGCTs and OTCA–FTCA. **(B)** Box-and-whisker plot of the Log-sigma-2-0mm 3D_glszm_SmallAreaHighGrayLevelEmphasis difference between OGCTs and OTCA–FTCA. **(C)** Box-and-whisker plot of the Mean ADC difference between OGCTs and OTCA–FTCA.

TABLE 6 | Receiver operating characteristic analysis of the overall imaging-based diagnosis (IBD), overall texture analysis (TA), and combined IBD with TA models.

Variables	Difference between areas	Standard error	95% confidence interval	*z* statistic	Significance level (*P*)
Overall_IBD~ Overall_TA	0.009	0.04	-0.067 to 0.085	0.23	0.82
Overall_IBD ~ Combine_IBD_and_TA	0.034	0.03	-0.017 to 0.084	1.30	0.19
Overall_TA ~ Combine_IBD_and_TA	0.025	0.02	-0.020 to 0.067	1.09	0.27

performed manually by radiologists based on individual habits, which may also have influence on the final results. In addition, we did not use MRI images other than T2-weighted images for TA in the present study.

In summary, compared with OTCA–FTCA, OGCTs more commonly exhibit intratumoral hemorrhage, mixed cystic degeneration, moderate enhancement, and low ADC values. Particularly, intratumoral hemorrhage may be a common and characteristic MR finding of OGCTs. When it is difficult to distinguish between OGCTs and OTCA–FTCA, TA described here may serve as a supplementary means, although this will require further large sample size validation before widespread implementation in clinical practice.

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 ethics committee of the First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

N-YL, J-ND, and C-BW contributed to conception and design. N-YL organized the database, performed the statistical analysis, and wrote the first draft of the manuscript. BS, P-PW, Y-LC, C-BW, and YC contributed to the collection and arrangement of data. Y-LC, Y-QG and P-PW contributed to data analysis. BS, P-PW, and C-BW wrote sections of the manuscript. All authors contributed to manuscript revision and read and approved the submitted version.

FUNDING

This work was supported by the Anhui Provincial Natural Science Foundation for Youths, China (1908085QH364).

ACKNOWLEDGMENTS

We thank Y-QG from GE Healthcare China for technical assistance.

REFERENCES

1. Kurman RJ, Carcangiu ML, Herrington CS, Young RH. Classification of Tumors of the Ovary. In: *WHO Classification of Tumors, 4th*. Lyon: IARC (2014). p. 44–56.
2. Haroon S, Zia A, Idrees R, Memon A, Fatima S, Kayani N. Clinicopathological Spectrum of Ovarian Sex Cord-Stromal Tumors; 20 Years' Retrospective Study in a Developing Country. *J Ovarian Res* (2013) 6(1):87. doi: 10.1186/1757-2215-6-87
3. Wei C, Chen YL, Li XX, Li NY, Wu YY, Lin TT, et al. Diagnostic Performance of MR Imaging-Based Features and Texture Analysis in the Differential Diagnosis of Ovarian Thecomas/Fibrothecomas and Uterine Fibroids in the Adnexal Area. *AcadRadiol* (2020) 27(10):1406–15. doi: 10.1016/j.acra.2019.12.025
4. Nocito AL, Sarancone S, Bacchi C, Tellez T. Ovarian Thecoma: Clinicopathological Analysis of 50 Cases. *Ann Diagn Pathol* (2008) 12 (1):12–6. doi: 10.1016/j.anndiagpath.2007.01.011

5. Murkey B, Nadkarni T, Bhalerao S, Jassawalla MJ. Delayed Menopause Due to Granulosa Cell Tumor of the Ovary. *J Midlife Health* (2011) 2(2):86–8. doi: 10.4103/0976-7800.92536
6. Kavuri S, Kulkarni R, Reid-Nicholson M. Granulosa Cell Tumor of the Ovary: Cytologic Findings. *Acta Cytol* (2010) 54(4):551–9. doi: 10.1159/000325176
7. Li X, Zhang W, Zhu G, Sun C, Liu Q, Shen Y. Imaging Features and Pathologic Characteristics of Ovarian Thecoma. *J Comput Assist Tomogr* (2012) 36(1):46–53. doi: 10.1097/RCT.0b013e31823f6186
8. Templeman CL, Fallat ME. Benign Ovarian Masses. *Semin Pediatr Surg* (2005) 14(2):93–9. doi: 10.1053/j.sempedsurg.2005.01.004
9. Zhang H, Zhang GF, Wang TP, Zhang H. Value of 3.0 T Diffusion-Weighted Imaging in Discriminating Thecoma and Fibrothecoma From Other Adnexal Solid Masses. *J Ovarian Res* (2013) 6(1):58. doi: 10.1186/1757-2215-6-58
10. Bremmer F, Behnes CL, Radzun HJ, Bettstetter M, Schweyer S. Tumoren Des Gonadenstromas [Sex Cord Gonadal Stromal Tumors]. *Pathologe* (2014) 35 (3):245–51. doi: 10.1007/s00292-014-1901-7
11. Jung SE, Rha SE, Lee JM, Park SY, Oh SN, Cho KS, et al. CT and MRI Findings of Sex Cord-Stromal Tumor of the Ovary. *AJR Am J Roentgenol* (2005) 185 (1):207–15. doi: 10.2214/ajr.185.1.01850207
12. Tanaka YO, Tsunoda H, Kitagawa Y, Ueno T, Yoshikawa H, Saida Y. Functioning Ovarian Tumors: Direct and Indirect Findings at MR Imaging. *Radiographics* (2004) 24 Suppl 1:S147–66. doi: 10.1148/rg.24si045501
13. Fang M, Dong J, Zhong Q, Fang X, Chen Y, Wang C, et al. Value of Diffusion-Weighted Imaging Combined With Conventional Magnetic Resonance Imaging in the Diagnosis of Thecomas and Their Differential Diagnosis With Adult Granulosa Cell Tumors. *Acta Radiol* (2019) 60(11):1532–42. doi: 10.1177/0284185119830280
14. Kim SH, Kim SH. Granulosa Cell Tumor of the Ovary: Common Findings and Unusual Appearances on CT and MR. *J Comput Assist Tomogr* (2002) 26 (5):756–61. doi: 10.1097/00004728-200209000-00016
15. Zhang H, Zhang H, Gu S, Zhang Y, Liu X, Zhang G. MR Findings of Primary Ovarian Granulosa Cell Tumor With Focus on the Differentiation With Other Ovarian Sex Cord-Stromal Tumors. *J Ovarian Res* (2018) 11(1):46. doi: 10.1186/s13048-018-0416-x
16. Yin B, Li W, Cui Y, Chu C, Ding M, Chen J, et al. Value of Diffusion-Weighted Imaging Combined With Conventional Magnetic Resonance Imaging in the Diagnosis of Thecomas/Fibrothecomas and Their Differential Diagnosis With Malignant Pelvic Solid Tumors. *World J Surg Oncol* (2016) 14(1):5. doi: 10.1186/s12957-015-0760-x
17. Chung BM, Park SB, Lee JB, Park HJ, Kim YS, Oh YJ. Magnetic Resonance Imaging Features of Ovarian Fibroma, Fibrothecoma, and Thecoma. *Abdom Imaging* (2015) 40(5):1263–72. doi: 10.1007/s00261-014-0257-z
18. ShinagareAB, MeylaertsLJ, LauryAR, MorteleKJ. MRI Features of Ovarianfibroma and Fibrothecoma With Histopathologic Correlation. *AJR Am J Roentgenol* (2012) 198(3):W296–303. doi: 10.2214/AJR.11.7221
19. Millet I, Rathat G, Perrochia H, Hoa D, Mérigeaud S, Curros-Doyon F, et al. Aspect Enimagerie Des Tumeurs De La Granulosa De L'ovaire: À Propos De Troiscas [Imaging Features of Granulosa Cell Tumors of the Ovary: About Three Cases]. *J Radiol* (2011) 92(3):236–42. doi: 10.1016/j.jradio.2011.02.002
20. Chen J, Wang J, Chen X, Wang Y, Wang Z, Li D. Computed Tomography and Magnetic Resonance Imaging Features of Ovarian Fibrothecoma. *Oncol Lett* (2017) 14(1):1172–8. doi: 10.3892/ol.2017.6228
21. Kato H, Kanematsu M, Ono H, Yano R, Furui T, Morishige K, et al. Ovarian Fibromas: MR Imaging Findings With Emphasis on Intratumoral Cyst Formation. *Eur J Radiol* (2013) 82(9):e417–21. doi: 10.1016/j.ejrad.2013.04.010
22. 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
23. 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
24. 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
25. Haralick RM, Shanmugam K, Dinstein I. Textural Features For Image Classification. *IEEE Syst Man Cy* (1973) SMC 3(6):610–21. doi: 10.1109/TSMC.1973.4309314
26. She Y, Zhang L, Zhu H, Dai C, Xie D, Xie H, et al. The Predictive Value of CT-Based Radiomics in Differentiating Indolent From Invasive Lung Adenocarcinoma in Patients With Pulmonary Nodules. *Eur Radiol* (2018) 28(12):5121–8. doi: 10.1007/s00330-018-5509-9
27. Zwanenburg A, Leger S, Vallières M, Löck S. *Image Biomarker Standardisation Initiative* (2019). Available at: https://arxiv.org/pdf/1612.07003.pdf.

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 17 November 2021
doi: 10.3389/fonc.2021.774248

Diagnosis of Breast Cancer Using Radiomics Models Built Based on Dynamic Contrast Enhanced MRI Combined With Mammography

You-Fan Zhao[1†], Zhongwei Chen[1†], Yang Zhang[2], Jiejie Zhou[1], Jeon-Hor Chen[2,3], Kyoung Eun Lee[4], Freddie J. Combs[2], Ritesh Parajuli[5], Rita S. Mehta[5], Meihao Wang[1] and Min-Ying Su[2,6]*

[1] Department of Radiology, First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China, [2] Department of Radiological Sciences, University of California, Irvine, Irvine, CA, United States, [3] Department of Radiology, E-Da Hospital and I-Shou University, Kaohsiung, Taiwan, [4] Department of Radiology, Inje University Seoul Paik Hospital, Inje University, Seoul, South Korea, [5] Department of Medicine, University of California, Irvine, Irvine, CA, United States, [6] Department of Medical Imaging and Radiological Sciences, Kaohsiung Medical University, Kaohsiung, Taiwan

OPEN ACCESS

Edited by:
Daniel Rodriguez Gutierrez, Nottingham University Hospitals NHS Trust, United Kingdom

Reviewed by:
Kim Brewer, Dalhousie University, Canada
Subathra Adithan, Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER), India

***Correspondence:**
Meihao Wang
wzwmh@wmu.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: *11 September 2021*
Accepted: *29 October 2021*
Published: *17 November 2021*

Citation:
Zhao Y-F, Chen Z, Zhang Y, Zhou J, Chen J-H, Lee KE, Combs FJ, Parajuli R, Mehta RS, Wang M and Su M-Y (2021) Diagnosis of Breast Cancer Using Radiomics Models Built Based on Dynamic Contrast Enhanced MRI Combined With Mammography. Front. Oncol. 11:774248. doi: 10.3389/fonc.2021.774248

Objective: To build radiomics models using features extracted from DCE-MRI and mammography for diagnosis of breast cancer.

Materials and Methods: 266 patients receiving MRI and mammography, who had well-enhanced lesions on MRI and histologically confirmed diagnosis were analyzed. Training dataset had 146 malignant and 56 benign, and testing dataset had 48 malignant and 18 benign lesions. Fuzzy-C-means clustering algorithm was used to segment the enhanced lesion on subtraction MRI maps. Two radiologists manually outlined the corresponding lesion on mammography by consensus, with the guidance of MRI maximum intensity projection. Features were extracted using PyRadiomics from three DCE-MRI parametric maps, and from the lesion and a 2-cm bandshell margin on mammography. The support vector machine (SVM) was applied for feature selection and model building, using 5 datasets: DCE-MRI, mammography lesion-ROI, mammography margin-ROI, mammography lesion+margin, and all combined.

Results: In the training dataset evaluated using 10-fold cross-validation, the diagnostic accuracy of the individual model was 83.2% for DCE-MRI, 75.7% for mammography lesion, 64.4% for mammography margin, and 77.2% for lesion+margin. When all features were combined, the accuracy was improved to 89.6%. By adding mammography features to MRI, the specificity was significantly improved from 69.6% (39/56) to 82.1% (46/56), $p<0.01$. When the developed models were applied to the independent testing dataset, the accuracy was 78.8% for DCE-MRI and 83.3% for combined MRI+Mammography.

Conclusion: The radiomics model built from the combined MRI and mammography has the potential to provide a machine learning-based diagnostic tool and decrease the false positive diagnosis of contrast-enhanced benign lesions on MRI.

Keywords: breast neoplasms, diagnosis, radiomics, machine learning, magnetic resonance imaging, mammography

INTRODUCTION

Breast cancer is the most common cancer in women, and one main cause of cancer deaths (1, 2). Mammography, ultrasound, and magnetic resonance imaging (MRI) are well-established diagnostic modalities, which are known to reveal different aspects of underlying abnormalities and provide complementary information for diagnosis (3, 4). Dynamic contrast-enhanced MRI (DCE-MRI) can assess angiogenesis (5, 6), which is essential for cancer development and progression (7, 8). The high spatial resolution and 3D imaging capability of MRI allow for detecting early small cancers, and for evaluating the extent of the disease for pre-operative staging and treatment planning. However, some benign diseases may show strong contrast enhancements and lead to a false positive diagnosis (9).

Mammography can detect breast cancer based on the presence of mass, microcalcifications, architectural distortion, or asymmetric density. It is a widely used imaging modality for screening and diagnosis, and crucial for detecting breast cancer at an early, curable, stage to decrease mortality (10). However, mammography is limited by breast density, which may compromise the detection sensitivity. For women with a high-risk of developing breast cancer, the screening is recommended to start from a young age, and to mitigate the problem of high density in mammography MRI is commonly used as a supplementary modality. Since different imaging can evaluate different pathological characteristics of the abnormal tissue, combining them may improve the diagnostic accuracy (3). MRI is also commonly used for problem-solving when other imaging shows equivocal findings. For example, in patients with category 4 mammographic microcalcifications, MRI can decrease false positive findings and unnecessary biopsy (11).

Breast Imaging Reporting and Data System (BI-RADS) (12) is used to indicate the level of suspicion in detected abnormality. However, subjective reading using the BI-RADS lexicon only achieved moderate levels of inter-reader agreement (13). For MRI, intra-/inter-observer agreement was particularly worse for non-mass enhancement compared to mass lesions (14, 15). To circumvent this problem, computer-aided diagnosis (CAD) systems have been proposed to develop quantitative models that are not subject to high variations to serve as potential diagnostic tools (16, 17).

Artificial intelligence (AI) based radiomics study has been widely applied for medical applications. The method allows for high-throughput extraction of quantitative features from radiographic images (18), and it has been shown as a feasible approach for diagnosis of breast cancer using mammography (19–22) and MRI (23–25). However, the combined model using different imaging modalities was rarely reported. Features from corresponding lesions on each modality can be extracted, and then combined in the selection process to develop better models based on their complementary information.

The purpose of this study was to evaluate the diagnostic performance of radiomics models built based on DCE-MRI and mammography. The motivation was coming from the high false positive diagnosis of contrast-enhanced benign lesions commonly seen on MRI. It is anticipated that the complementary information provided by the radiomics analysis of the lesion on mammography may help to improve the diagnostic accuracy. In mammography, features extracted from the lesion and the margin were used to build separate models. The complementary role of MRI and mammography was first evaluated by the selected features, and then by comparing the performance of final models built using each modality alone and in combination.

Abbreviations: AI, artificial intelligence; AUC, the area under the curve; BI-RADS, Breast Imaging Report and Data System; CAD, computer-aided diagnosis; CC, cranio-caudal; DCE, dynamic contrast enhanced; DCIS, ductal carcinoma *in-situ*; GLCM, gray-level co-occurrence matrix; GLDM, gray-level dependence matrix; GLRLM, gray-level run length matrix; GLSZM, gray-level size zone matrix; IDC, invasive ductal cancer; MIP, maximum intensity projection; MLO, medio-lateral oblique; MRI, magnetic resonance imaging; NGTDM, neighboring gray tone difference matrix; NPV, negative predicting value; PPV, positive predicting value; ROC, receiver operating characteristic; ROI, region of interest; SE, signal enhancement; SVM, support vector machine.

MATERIAL AND METHODS

Study Population

This retrospective study was approved by Institutional Review Board and written informed consent was waived. Earlier patients who received DCE-MRI and mammography for diagnosis between July 2017 and August 2019 and had confirmed pathology were retrospectively identified as the training set. Later patients from September 2019 to July 2020 were used as the independent testing set. The exclusion criteria were: (1) no pathology result; (2) not visible on MRI or mammography; (3) having prior surgery, chemotherapy, or other treatment; (4) the interval between the two examinations longer than one month; (5) poor image quality. Finally, a total of 268 lesions were included, 202 lesions (146 malignant and 56 benign) in the training set, and 66 lesions (48 malignant and 18 benign) in the testing set. The BI-RADS scores of MRI and mammography were obtained from the radiology reports, classified into 2, 3, 4A, 4B, 4C, and 5. In our institution, BI-RADS 4 MRI cases were routinely subdivided to 4A, 4B, and 4C, as validated in Strigel et al. (26).

Image Acquisition

Mammography was performed using Fujifilm Amulet Innovality Digital Mammography System with a resolution of 5828×4728 pixels, including craniocaudal (CC) and mediolateral oblique (MLO) view. MRI was performed on a 3.0T scanner (GE SIGNA HDx) using a dedicated 8-channel bilateral breast coil. The imaging protocol included axial and sagittal T2- and T1-weighted sequences, and the DCE acquisition performed using the volume imaging for breast assessment (VIBRANT) sequence. The parameters were: repetition time= 5msec, echo time= 2msec, flip angle= 10°, slice thickness= 1.2mm, field of view= 34×34cm^2, matrix size= 416×416, temporal resolution= 90sec, and total scan time= 9min. The DCE series consisted of 6 frames: one pre-contrast and 5 post-contrast. The contrast agent, 0.1 mmol/kg body weight of gadopentetate dimeglumine

(Magnevist; Bayer Schering Pharma), was injected after the pre-contrast images were acquired, with a flow rate of 2 mL/s followed by a flush of 20 mL saline.

Tumor Segmentation

For MRI, the tumor region of interest (ROI) segmentation was done using computer algorithms, according to the location and the range of slices. The fuzzy-C-means clustering algorithm was applied to perform segmentation on each DCE slice containing the lesion. The automatic segmentation results were evaluated by two radiologists separately, and adjusted if necessary. Then, the ROIs from all slices were combined, and the 3D connected-component labeling and the hole-filling algorithms were applied to generate the final 3D mask (27, 28). For the corresponding mammography, two radiologists manually outlined the lesion on craniocaudal (CC) or mediolateral oblique (MLO) view by consensus using ITK-SNAP software (version 3.8, www.itksnap.org), with the guidance of the lesion shown on the maximum intensity projection (MIP) of MRI, projected from different angles. The choice of CC or MLO was determined according to the lesion visibility, and only one view was used.

MRI and Mammography Radiomics Feature Extraction

The analysis flowchart is demonstrated in **Figure 1**. For DCE-MRI, three heuristic DCE parametric maps were generated according to: the early wash-in signal enhancement (SE) ratio ((F2-F1)/F1); the maximum SE ratio = ((F3-F1)/F1); the wash-out slope ((F6-F3)/F3) (25), as illustrated in case examples in **Figures 2–5**. The intensity was normalized to mean=0 and standard deviation=1. In the segmented 3D ROI, pixels were transformed into isotropic 0.82×0.82×0.82 mm by B-spline interpolation. The radiomics analysis was performed using the PyRadiomics, an open-source radiomics library written in Python (29). On each parametric map, a total of 107 features were extracted, including 14 shape, 18 first-order, 24 gray-level co-occurrence matrix (GLCM), 14 gray-level dependence matrix (GLDM), 16 gray-level run length matrix (GLRLM), 16 gray-level size zone matrix (GLSZM), and 5 neighboring gray tone difference matrix (NGTDM) features, so there was a total of 321 parameters from 3 maps. Only 268 features showing intra-class coefficient (ICC) ≥0.8 were included in the final analysis, which was determined using two sets of separately segmented tumor ROI to evaluate the reproducibility of extracted radiomics features (30).

For mammography, two different feature sets were analyzed. Considering that the ROI was manually drawn by tracing the visible lesion area based on density, it might not reveal the margin information. To specifically focus on the margin, a 2-cm bandshell was created, by shrinking and expanding the manually-drawn tumor boundary by 1 cm, as shown in **Figure 1**. Because the margin could not be well defined on mammography, shrinking the boundary followed by region growing has been shown as a feasible segmentation method (31), and the method was adopted here to generate the bandshell for analysis of margin features. Similarly, the intensity was normalized to mean=0 and standard deviation=1, and a total of 107 PyRadiomics features were extracted from the outlined lesion mask and also from the bandshell on mammography. The radiomics model was first performed using lesion features alone, margin features alone, and then a combined model was built by considering all lesion and margin features.

Feature Selection and Model Building in Training Set

The procedures are also shown in **Figure 1**. In addition to the normalization on images, each feature extracted from all cases was normalized to mean=0 and standard deviation=1 before training. To evaluate the importance of these features in

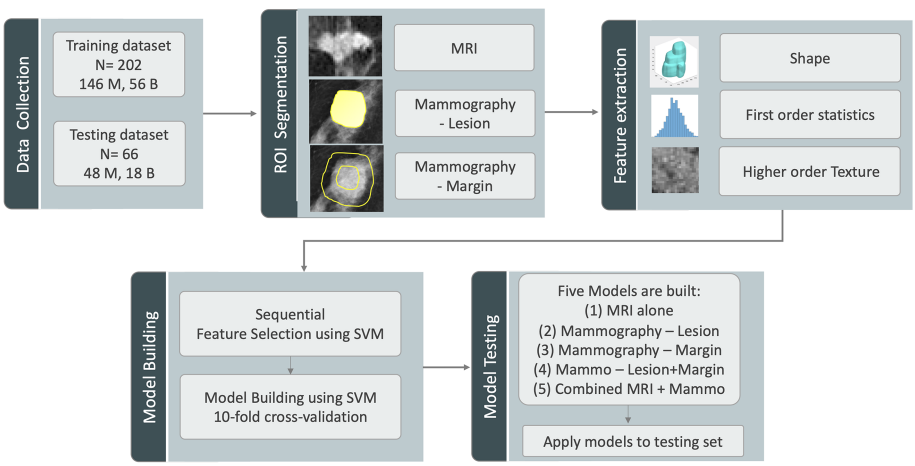


FIGURE 1 | The analysis flowchart. The training and testing sets are assembled according to the time of case enrollment. The analysis starts with ROI segmentation, followed by radiomics feature extraction using Pyradiomics, feature selection and model building in the training set using SVM with cross-validation, and lastly, the testing of the 5 developed models in the testing set.

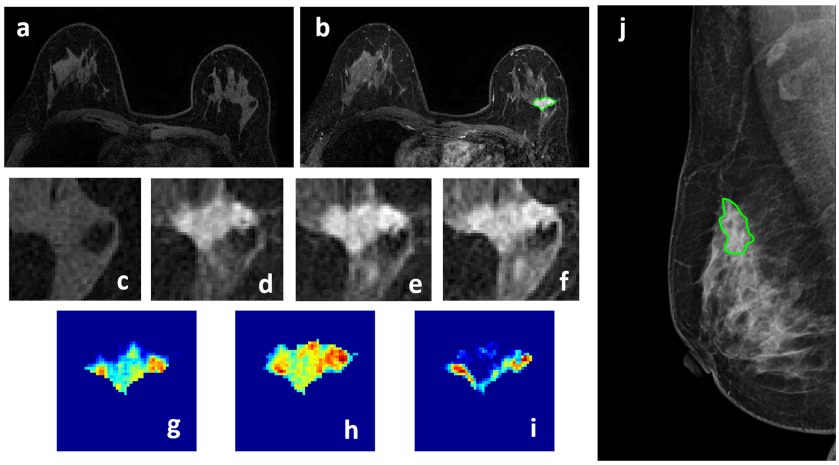


FIGURE 2 | A 50-year-old patient with invasive ductal cancer, showing a strongly enhanced 1.8 x 1.0 cm lesion, with MRI BI-RADS score of 5. **(A)** F1 Pre-contrast image. **(B)** F2 post-contrast image. **(C–I)** Magnified images to demonstrate the margin and internal enhancements within the lesion. **(C)** F1 pre-contrast, **(D)** F2 post-contrast, **(E)** F3 post-contrast, **(F)** The last F6 post-contrast image. **(G)** The wash-in signal enhancement map F2-F1, **(H)** The F3-F1 signal enhancement map, **(I)** The wash-out F6-F3 map. **(J)** A mass lesion with spiculation is clearly noted on mammography as BI-RADS 4C, and manually outlined by a radiologist. The radiomics malignancy probability predicted by MRI, mammography, and combined models were: 0.83, 0.77, 0.88, respectively, true positive.

diagnosis, a sequential forward feature selection method using the support vector machine (SVM) was applied (32, 33). In this process, we used SVM with Gaussian kernel as the objective function to test the performance of models built with a subset of features. In the beginning, an empty candidate set was presented, and features were sequentially added. The 10-fold cross-validation was applied to test the model performance. In each iteration, the training process was repeated 1,000 times to explore the robustness of each feature. After each iteration, the feature which led to the best performance was added to the candidate set. When the addition of features no longer met the criterion, the selection process stopped. Here, we used 10*e^-6 as termination tolerance for the objective function value.

The selected features were used to build the SVM classification model with Gaussian kernel to classify the benign and malignant groups. The diagnostic performance was tested using 10-fold cross-validation. Each case had only one chance to be included in the validation set. The probability of all cases in

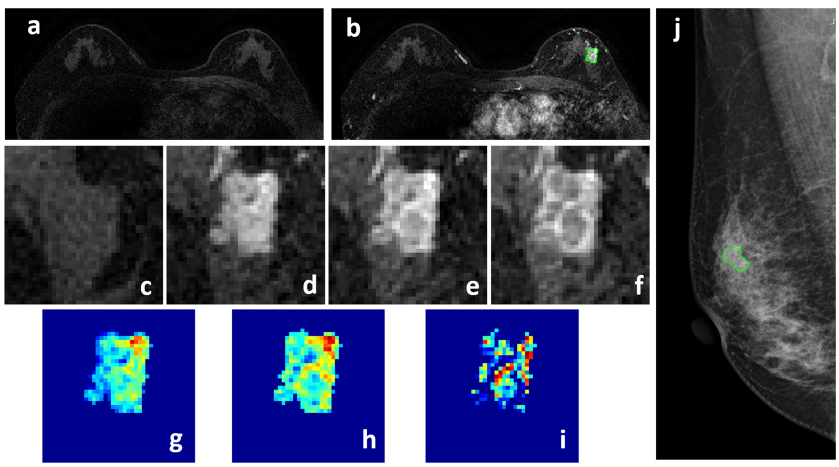


FIGURE 3 | A 58-year-old patient with ductal carcinoma *in situ*, showing a strongly enhanced heterogeneous 1.4 x 0.9 cm lesion, with MRI BI-RADS score of 5. **(A)** F1 Pre-contrast image. **(B)** F2 post-contrast image. **(C–I)** Magnified images to demonstrate the margin and internal enhancements within the lesion. **(C)** F1 pre-contrast, **(D)** F2 post-contrast, **(E)** F3 post-contrast, **(F)** The last F6 post-contrast image. **(G)** The wash-in signal enhancement map F2-F1, **(H)** The F3-F1 signal enhancement map, **(I)** The wash-out F6-F3 map. **(J)** A suspicious BI-RADS 4A mass is seen on mammography. The lesion ROI is outlined with the guidance of MRI. The probability predicted by MRI, mammography, and combined radiomics models were: 0.53, 0.49, 0.62, respectively, true positive.

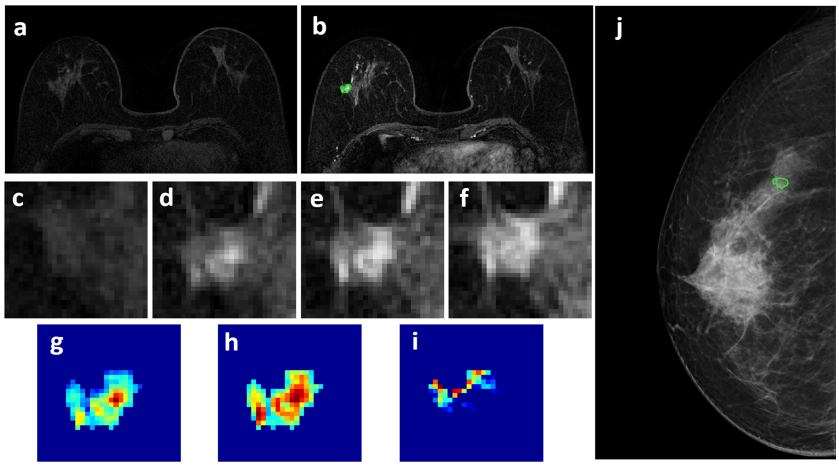


FIGURE 4 | A 63-year-old patient with a 0.7 x 0.7 cm benign adenosis, showing a persistent DCE-MRI enhancement kinetics and determined as BI-RADS 3 on MRI. **(A)** F1 Pre-contrast image. **(B)** F2 post-contrast image. **(C–I)** Magnified images to demonstrate the margin and internal enhancements within the lesion. **(C)** F1 pre-contrast, **(D)** F2 post-contrast, **(E)** F3 post-contrast, **(F)** The last F6 post-contrast image. **(G)** The wash-in signal enhancement map F2-F1, **(H)** The F3-F1 signal enhancement map, **(I)** The wash-out F6-F3 map. **(J)** The lesion is not seen on mammography, determined as BI-RADS 2, and an area is outlined with the guidance of MRI. The probability predicted by MRI, mammography, and combined radiomics models were: 0.42, 0.44, 0.15, respectively, true negative.

the validation set was combined to perform the receiver operating characteristic curve (ROC) analysis, and the area under the curve (AUC) was calculated. Five models were built using features extracted from: 1) DCE-MRI; 2) mammography – lesion ROI; 3) mammography – margin ROI, i.e., the bandshell; 4) mammography lesion+margin; and 5) all combined. The developed model gave a radiomics score, i.e., the malignancy probability, for each case.

Applying the Trained Models to the Testing Set

The developed models from the training set were applied to test their performances in the testing set. The model gave each lesion a radiomics score, and they were used to generate the ROC curves. The sensitivity, specificity, positive predicting value (PPV), negative predicting value (NPV), and overall accuracy of each model were calculated using the threshold of probability ≥0.5 as malignant.

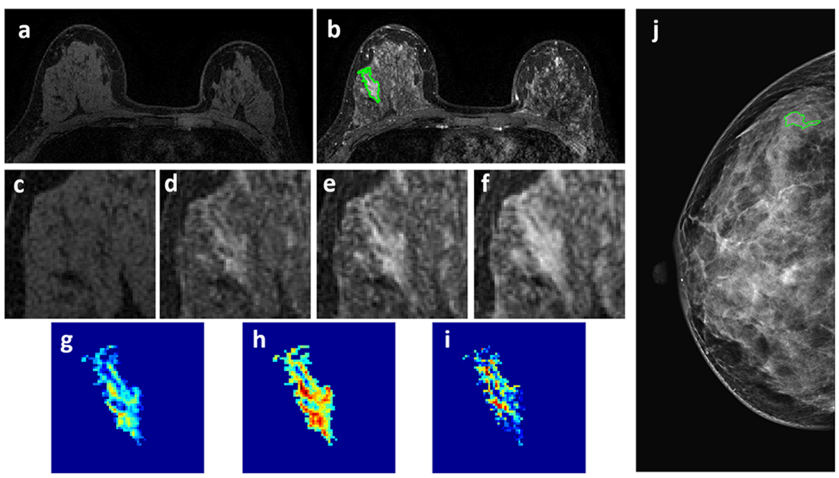


FIGURE 5 | A 46-year-old patient with a 2.7 x 1.3 cm benign adenosis. This is a young woman with extremely dense breasts showing substantial parenchymal enhancements. The lesion shows a persistent DCE-MRI pattern and determined as BI-RADS 4A on MRI. **(A)** F1 Pre-contrast image. **(B)** F2 post-contrast image. **(C–I)** Magnified images to demonstrate the margin and internal enhancements within the lesion. **(C)** F1 pre-contrast, **(D)** F2 post-contrast, **(E)** F3 post-contrast, **(F)** The last F6 post-contrast image. **(G)** The wash-in signal enhancement map F2-F1, **(H)** The F3-F1 signal enhancement map, **(I)** The wash-out F6-F3 map. **(J)** The lesion is not seen on mammography, determined as BI-RADS 2, and an area is outlined with the guidance of MRI. The probability predicted by MRI, mammography, and combined radiomics models were: 0.3, 0.41, 0.11, respectively, true negative.

The Delong test was used to compare the difference between paired ROC curves. The difference in proportions between malignant and benign groups was compared by using the Chi-square (χ^2) test or Fisher's Exact Test.

RESULTS

Patients' Characteristics and BI-RADS Scores

In the training set, the mean age was 50.0 ± 9.6 in the malignant, and 46.6 ± 9.7 in the benign groups. The 1-D longest dimension tumor size measured on MRI was 2.4 ± 1.4 cm (median 2.0 cm) in the malignant, and 2.0 ± 2.3 cm (median 1.5 cm) in the benign groups. In the testing set, the mean age was 51.8 ± 11.2 in the malignant, and 43.5 ± 10.8 in the benign groups. The 1-D longest dimension tumor size measured on MRI was 3.2 ± 1.9 cm (median 2.8 cm) in the malignant, and 2.0 ± 1.4 cm (median 1.5 cm) in the benign groups. The pathological types and BI-RADS distributions in both datasets are listed in **Table 1**. In the training set, the majority of malignant lesions had BI-RADS scores of 4B, 4C, 5 on MRI (132/146 = 90.4%) and mammography (120/146 = 82.2%). In the benign group, a substantial number of patients also had high BI-RADS ≥ 4B diagnosed by MRI (20/56 = 35.7%) and mammography (16/56 = 28.6%). Although the number of patients with BI-RADS ≥ 4B lesions was significantly smaller in the benign compared to the malignant groups ($p < 0.001$), these cases would be recommended for biopsy and led to false positive diagnosis. Similar BI-RADS distributions were also noted in the testing set.

TABLE 1 | Pathological types and BI-RADS scores of lesions in training and testing datasets.

Characteristics	Training (N = 202)	Testing (N = 66)
Benign	56	18
Fibroadenoma	13 (23.2%)	5 (27.8%)
Adenosis	25 (44.6%)	10 (55.6%)
Intraductal papilloma	10 (17.9%)	1 (5.6%)
Inflammation	2 (3.6%)	0 (0.0%)
Others	6 (10.7%)	2 (11.1%)
MRI BI-RADS		
2	9 (16.1%)	1 (5.6%)
3	13 (23.2%)	3 (16.7)
4A	14 (25%)	9 (50%)
4B	14 (25%)	4 (22.2%)
4C	5 (8.9%)	1 (5.6%)
5	1 (1.8%)	0 (0.0%)
Mammography BI-RADS		
2	13 (23.2%)	5 (27.8)
3	16 (28.6%)	6 (33.3%)
4A	11 (19.6%)	4 (22.2%)
4B	12 (21.4%)	3 (16.7%)
4C	4 (7.1%)	0 (0.0%)
5	0 (0%)	0 (0.0%)
Malignant	146	48
Invasive ductal cancer	113 (77.4%)	39 (81.3%)
Ductal carcinoma *in-situ*	23 (15.8%)	3 (6.3%)
Intraductal papillary carcinoma	4 (2.7%)	0 (0.0%)
Mucinous carcinoma	3 (2.1%)	1 (2.1%)
Others	3 (2.1%)	5 (10.4%)
MRI BI-RADS		
3	1 (0.7%)	0 (0.0%)
4A	13 (8.9%)	1 (2.1%)
4B	17 (11.6%)	4 (8.3%)
4C	39 (26.7%)	18 (37.5%)
5	76 (52.1%)	25 (52.1%)
Mammography BI-RADS		
2	0	4 (8.3%)
3	9 (6.2%)	1 (2.1%)
4A	17 (11.6%)	2 (4.2%)
4B	32 (21.9%)	11 (22.9%)
4C	48 (32.9%)	21 (43.8%)
5	40 (27.4%)	9 (18.8%)

BI-RADS, Breast Imaging Report and Data System.

Radiomics Diagnostic Models in Training Set

The selected radiomics features for each model are listed in **Table 2**. The diagnostic sensitivity, specificity, PPV, NPV, accuracy, and AUC obtained from the cross-validation results are summarized in **Table 3**. The overall accuracy was 83.2% for DCE-MRI. In mammography, the accuracy was 75.7% for lesion-ROI, 64.4% for margin-ROI, and when combining both of them it was improved to 77.2%. When all MRI and mammography features were combined to build a model, the accuracy was improved to 89.6%, which was significantly better than the mammography model (77.2%, p=0.001). The combined model was also better than the MRI model (83.2%, p=0.059), but not reaching significance. By adding mammography features to MRI, the specificity was significantly improved from 69.6% (39/56) to 82.1% (46/56) ($p<0.01$), while sensitivity was also improved from 88.4% (129/146) to 92.5% (135/146). **Figure 6** plots the malignant probability predicted by the combined MRI+ Mammography radiomics model in the training set of 146 malignant and 56 benign lesions. Using the threshold of 0.5 as the cut-off, there are 135 true positive, 46 true negative, 11 false negative, and 10 false positive cases, with an overall accuracy of 181/202 = 89.6%.

Four case examples are shown. **Figure 2** is an IDC with BI-RADS 5 MRI and BI-RADS 4C mammography, and the malignancy probability predicted by MRI, mammography, and combined models are: 0.83, 0.77, 0.88, respectively; thus, true positive. **Figure 3** is a DCIS, also with BI-RADS 5 MRI and a lower BI-RADS 4A mammography, and the combined radiomics probability is 0.62, true positive. **Figure 4** is a very small 0.7 cm benign adenosis with BI-RADS 3 MRI and BI-RADS 2 mammography, and the combined radiomics probability is 0.15, true negative. **Figure 5** is another adenosis in a younger woman with BI-RADS 4A MRI and BI-RADS 2 mammography, and the combined radiomics probability is 0.11, true negative. These cases demonstrate that the malignancy probability predicted by radiomics models was consistent with BI-RADS reading, and elaborate how the model may help to improve the diagnostic confidence.

Performance of the Trained Models in Testing Set

The developed models were then applied to cases in the independent testing set to test the performance. The results are

TABLE 2 | Selected radiomics features for modeling using MRI, mammography, and both combined.

Models	Selected Radiomic Features	Numbers
DCE-MRI	**Maximum signal enhancement ratio:** entropy, GLCM sum average, GLCM IMC1, GLDM high gray level emphasis, skewness **Wash-in ratio:** GLRLM RLN **Wash-out ratio:** GLRLM small area emphasis, GLCM sum entropy	8
Mammography (Lesion)	90% value, entropy, GLCM maximum probability, GLDM high gray level emphasis	4
Mammography (Margin)	10% value, GLSZM zone entropy, GLCM IDN	3
Combination of DCE-MRI and mammography	**Maximum signal enhancement ratio:** kurtosis, GLCM IMC1 **Wash-in ratio:** skewness, GLRLM RLN, NGTDM complexity **Wash-out ratio:** GLCM IMC1, GLCM sum entropy **Mammography lesion:** GLCM maximum probability, GLCM IDN	9

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; IMC, informational measure of correlation; IDN, inverse difference normalized; RLN, run length non-uniformity.

listed in **Table 3**. In general, the performance of these 5 models was consistent with the validation results in the training set. The accuracy was 78.8% for DCE-MRI, 69.7% for mammography, and improved to 83.3% when using the combined MRI and mammography model.

Performance of the Combined Model in Each BI-RADS Category

In order to further evaluate the performance of the model in each BI-RADS category, the results from the training and testing sets are combined and listed in **Table 4**. The cases with BI-RADS score of 2, 3, 4A, 4B, 4C, and 5 based on MRI and mammography were separately tabulated. It can be seen clearly that malignant lesions have higher BI-RADS scores compared to benign lesions, but many benign lesions also have ≥4B scores. First, in the malignant group, if we used 2, 3, and 4A as more likely benign, 15 MRI and 33 mammography cases would be diagnosed as benign. The results showed that the model could reach 14/15 = 93.3% accuracy for MRI and 31/33 = 93.9% for mammography lesions, still with a high sensitivity. On the other hand, in the benign group, if we used 4B,

TABLE 3 | The diagnostic performance of developed radiomics models in training and testing datasets.

Models	Sensitivity	Specificity	PPV	NPV	Accuracy	AUC
Training Dataset						
DCE-MRI	88.4% (129/146)	69.6% (39/56)	88.4% (129/146)	69.6% (39/56)	83.2%	0.77
Mammography (Lesion)	84.9% (124/146)	51.8% (29/56)	82.1% (124/151)	56.9% (29/51)	75.7%	0.69
Mammography (Margin)	73.3% (107/146)	41.1% (23/56)	76.4% (107/140)	37.1% (23/62)	64.4%	0.62
Mammography (Lesion+Margin)	84.9% (124/146)	57.1% (32/56)	83.8% (124/148)	59.3% (32/54)	77.2%	0.70
All Combination	92.5% (135/146)	82.1% (46/56)	93.1% (135/145)	80.7% (46/57)	89.6%	0.83
Testing Dataset						
DCE-MRI	87.5% (42/48)	55.6% (10/18)	84% (42/50)	62.5% (10/16)	78.8%	0.80
Mammography (Lesion)	81.3% (39/48)	38.9% (7/18)	78% (39/50)	43.8% (7/16)	69.7%	0.65
Mammography (Margin)	66.7% (32/48)	33.3% (6/18)	59.3% (32/54)	27.3% (6/22)	57.6%	0.53
Mammography (Lesion+Margin)	81.3% (39/48)	38.9% (7/18)	78% (39/50)	43.8% (7/16)	69.7%	0.64
All Combination	91.7% (44/48)	61.1% (11/18)	86.3% (44/51)	73.3% (11/15)	83.3%	0.81

PPV, positive predicting value; NPV, negative predicting value; AUC, the area under the curve; DCE, dynamic contrast enhanced.

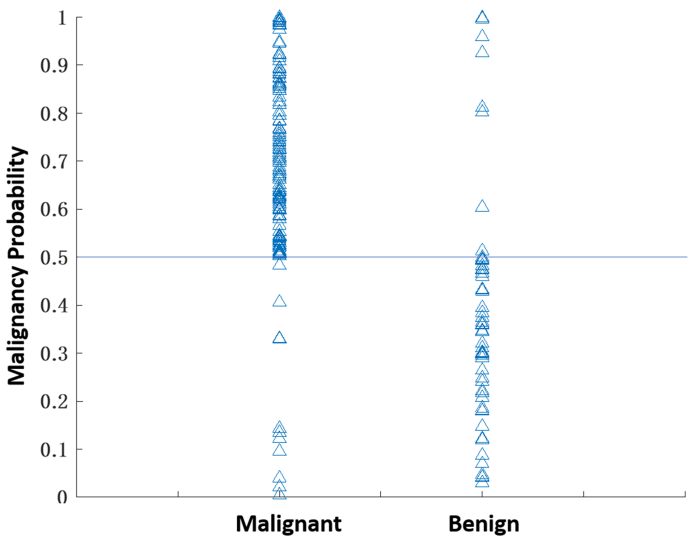


FIGURE 6 | The malignant probability predicted by the combined MRI+Mammography radiomics model in 202 lesions, 146 malignant and 56 benign, in the training set. Using the threshold of 0.5 as the cut-off, there are 135 true positive, 11 false negative, 46 true negative, and 10 false positive cases, with an overall accuracy of 181/202 = 89.6%.

4C and 5 as possibly malignant, 25 MRI and 19 mammography cases would be diagnosed as malignant. The model could achieve 18/25 = 72% accuracy for MRI and 15/19 = 78.9% for mammography lesions. The correct benign diagnosis for these cases may help to avoid unnecessary biopsy.

DISCUSSION

In this study, we developed the radiomics models for diagnosis of breast cancer using DCE-MRI alone, mammography alone, and the combined MRI and mammography. While quite a few studies have reported the radiomics models developed using MRI (23, 24, 34) or mammography (19–22), the combined analysis was rarely reported (35). We further investigated the complementary role of MRI and mammography features in diagnostic sensitivity and specificity. In the training set, the combined model (89.6%) had a higher accuracy than individual ones (83.2% for mammography, 77.2% for mammography). When mammography features were added to MRI features, it could significantly improve specificity from 69.6% (39/56) to 82.1% (46/56); and thus, have the potential to decrease unnecessary biopsy. Interestingly, the sensitivity was also improved, so the higher specificity was not at the expense of compromised sensitivity. Similar findings were seen in the testing set, with slightly lower overall accuracy from 89.6% to 83.3%.

For mammography, we further separated the analysis using features extracted from the lesion-ROI alone, and from the margin-ROI alone by using a bandshell. The results showed that the accuracy was much better for the lesion model than the margin model, but the margin information could help to improve the accuracy. The results were consistent with the knowledge that margin plays an important role in characterization of a lesion for diagnosis.

Since MRI is more expensive than mammography, the most established clinical indication is for pre-operative staging and high-risk screening. It is not always included in the standard diagnostic workup. It has been shown that in the mammography 4 category, particularly in non-palpable lesions presenting only with microcalcifications, MRI can be used to reduce false positives and avoid unnecessary biopsy (11, 36, 37). On the other hand, benign lesions may show enhancements on MRI, and the

TABLE 4 | The number of correctly diagnosed cases made by the combined radiomics model in each BI-RADS category.

BI-RADS Score	Malignant Cases (N = 194)		Benign Cases (N = 74)	
	MRI	Mammography	MRI	Mammography
2	0	3/4	8/10	15/18
3	1/1	9/10	11/16	18/22
4A	13/14	19/19	20/23	9/15
4B	19/21	38/43	13/18	11/15
4C	51/57	67/69	4/6	4/4
5	95/101	43/49	1/1	0

BI-RADS, Breast Imaging Report and Data System.

information from mammography may help to rule out malignancy (38). As in the case examples shown in **Figures 4** and **5**, the benign lesions might be inconspicuous on mammography and had low BI-RADS score of 2, and we had to use MIP generated from MRI as a reference to locate them. Since MRI and mammography evaluate different aspects of the underlying pathology, they should be reviewed together to determine which information needs to be weighted more.

Radiomics is becoming an active research field in breast cancer diagnosis. Due to the large number of images acquired using different MR sequences, radiomics provides an efficient analysis method to extract information. Therefore, more MRI radiomics studies were reported than ultrasound, mammography, and 18F FDG PET/CT (34). MRI radiomics was shown to provide better discrimination than conventional parameters for the diagnosis of breast cancer (23, 24). Mammography radiomics analysis has also been performed in several diagnostic studies (19–22). However, since the patient cohort is different, the diagnostic accuracy will be highly dependent on the inclusion/exclusion criteria, and not directly comparable among studies. Mao et al. (19) used four modeling algorithms, including SVM, naive Bayes classifier, k-NN classifier, and logistic regression to differentiate between benign and malignant cases, and showed a high vibration of 0.629-0.978 in the obtained accuracy. The radiologists' reading accuracy was 0.772. Lei et al. (20) applied radiomics to diagnose patients showing BI-RADS 4 calcifications on mammography, and achieved AUC of 0.80 in the validation cohort. For characterizing microcalcifications, since the lesion area was not well-defined, the ROI drawing will affect the extracted features, and thus, the diagnostic results. Huang et al. (21) applied mammography radiomics for distinguishing male malignant and benign lesions, and reported an AUC of 0.82 – a very unique study in rarely reported male patients. Another study by Niu et al. (22) also analyzed patients showing abnormal lesions on mammography and MRI, close to our patient cohort, but their goal was to evaluate the combined effect of mammography and digital breast tomosynthesis (DBT), as well as the combined effect of DCE and diffusion weighted MRI. The reported accuracy based on the mammography was close to ours, around 0.70. Multi-modal radiomics combining different imaging modalities are rarely reported. In a study by Chen et al., the multimodal classifier achieved a better diagnostic performance than any single modality (35). Since each imaging modality is unique in its acquisition method and parameter setting, the extracted features from a lesion may be different and provide complementary information to improve diagnostic accuracy.

In this study, the cases were identified from the MRI database first, and then only those with mammography performed within one month were further selected for analysis. All lesions showed strong enhancements on MRI, and the information was used to determine a corresponding ROI on mammography. Co-registration of MRI and mammography to ensure that the traced ROI is indeed coming from the same suspicious tissue is not a trivial task. We used maximum intensity projection of MRI as guidance, and it could be projected from different angles to simulate CC view and MLO view to guide the tracing of the suspicious tissues on mammography. Some computer techniques have been proposed for registration between MRI and mammography, e.g., using finite element methods by Hopp et al. (39) and Mertzanidou et al. (40), and the thin-plate spline method by Yang et al. (3). These registration techniques can be considered in future multi-modality radiomics studies. However, since the mammography was acquired using heavily compressed breast tissues in a different body position, it might be difficult to find the precise correspondence. Therefore, in this study we only analyzed the CC or MLO view that had more clear presentation of the lesion.

There were several limitations in this study. First, the models were developed using a dataset from a single institution. The earlier cases were used for training, and the performance was evaluated using 10-fold cross-validation. We assembled an independent testing set using later cases according to time of enrollment, so the developed models from training can be independently tested. Another limitation is that the sample size was relatively small. In our dataset, all benign lesions had to show visible enhancements on MRI and were histologically confirmed, which were very strict criteria and limited the number of eligible cases. However, since the major goal of this study is to investigate whether and how much the addition of mammography radiomics features can complement MRI, using a strict rule to identify eligible cases with histologically confirmed lesions is needed. Third, while all lesions showed enhancements on MRI, lesions not visible with the MRI-guidance on mammography were not included in this study. Since the boundary of these lesions could not be clearly defined, the radiomics features might not be reliably extracted.

In conclusion, the radiomics models built based on combined MRI and mammography had better diagnostic accuracy than models built using single modality alone. The combined model could reach the accuracy of 89.6% in the training and 83.3% in the testing sets. The motivation of this study is to use the complementary information extracted from radiomics analysis of the lesion shown on mammogram to decrease the false positive diagnosis of contrast-enhanced benign lesions on MRI. In the western countries, breast MRI is recommended as a clinical modality for screening of women with a high risk of developing breast cancer, and the false positive diagnosis in a screening population will lead to many unnecessary procedures including biopsy, and patient anxiety. Our study may provide a helpful computer-aided diagnostic tool for such clinical indications. The multimodality radiomics analysis by combining mammography and MRI features has the potential to improve the specificity and reduce unnecessary biopsies, while maintaining a high sensitivity for diagnosis of breast cancer.

DATA AVAILABILITY STATEMENT

The datasets used and analyzed in this study will be made available by the corresponding author on a reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of The First Affiliated Hospital of Wenzhou 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

Study concept and design: Y-FZ and M-YS. Acquisition of data: ZC, JZ, and MW. Analysis of data: YZ, ZC and KL. Drafting of the manuscript: Y-FZ, J-HC, FC, RP, and RM. Critical revision: MW and M-Y S. Statistical analysis: YZ and ZC. Study supervision: MW and M-Y S. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported in part by Foundation of Wenzhou Science & Technology Bureau (No. Y20180185), Medical Health Science and Technology Project of Zhejiang Province Health Commission (No. 2019KY102), Research Incubation Project of First Affiliated Hospital of Wenzhou Medical University (No. FHY2019085), the National Cancer Institute of the National Institutes of Health under award number P30 CA062203, R01 CA127927, R21 CA208938 and the UC Irvine Comprehensive Cancer Center using UCI Anti-Cancer Challenge funds.

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 (6):394–424. doi: 10.3322/caac.21492
2. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654
3. Yang S-N, Li F-J, Liao Y-H, Chen Y-S, Shen W-C, Huang T-C. Identification of Breast Cancer Using Integrated Information From MRI and Mammography. *PloS One* (2015) 10(6):e0128404. doi: 10.1371/journal.pone.0128404
4. 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 (2):105–16. doi: 10.3233/CH-16242
5. Buadu LD, Murakami J, Murayama S, Hashiguchi N, Sakai S, Masuda K, et al. Breast Lesions: Correlation of Contrast Medium Enhancement Patterns on MR Images With Histopathologic Findings and Tumor Angiogenesis. *Radiology* (1996) 200(3):639–49. doi: 10.1148/radiology.200.3.8756909
6. Ikeda O, Nishimura R, Miyayama H, Yasunaga T, Ozaki Y, Tuji A, et al. Evaluation of Tumor Angiogenesis Using Dynamic Enhanced Magnetic Resonance Imaging: Comparison of Plasma Vascular Endothelial Growth Factor, Hemodynamic, and Pharmacokinetic Parameters. *Acta Radiol* (2004) 45(4):446–52. doi: 10.1080/02841850410005372
7. Izrailit J, Reedijk M. Developmental Pathways in Breast Cancer and Breast Tumor-Initiating Cells: Therapeutic Implications. *Cancer Lett* (2012) 317 (2):115–26. doi: 10.1016/j.canlet.2011.11.028
8. Toi M, Inada K, Suzuki H, Tominaga T. Tumor Angiogenesis in Breast Cancer: Its Importance as a Prognostic Indicator and the Association With Vascular Endothelial Growth Factor Expression. *Breast Cancer Res Treat* (1995) 36(2):193–204. doi: 10.1007/BF00666040
9. Parsian S, Rahbar H, Allison KH, Demartini WB, Olson ML, Lehman CD, et al. Nonmalignant Breast Lesions: ADCs of Benign and High-Risk Subtypes Assessed as False-Positive at Dynamic Enhanced MR Imaging. *Radiology* (2012) 265(3):696–706. doi: 10.1148/radiol.12112672
10. Tabár L, Vitak B, Chen HH, Yen MF, Duffy SW, Smith RA. Beyond Randomized Controlled Trials: Organized Mammographic Screening Substantially Reduces Breast Carcinoma Mortality. *Cancer* (2001) 91(9):1724–31. doi: 10.1002/1097-0142(20010501)91:9<1724::aid-cncr1190>3.0.co;2-v
11. Eun NL, Son EJ, Gweon HM, Youk JH, Kim JA. The Value of Breast MRI for BI-RADS Category 4B Mammographic Microcalcification: Based on the 5th Edition of BI-RADS. *Clin Radiol* (2018) 73(8):750–5. doi: 10.1016/j.crad.2018.04.014
12. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA. *ACR BI-RADS® Atlas, Breast Imaging Reporting and Data System*. Reston, VA: American College of Radiology (2013).
13. Rawashdeh M, Lewis S, Zaitoun M, Brennan P. Breast Lesion Shape and Margin Evaluation: BI-RADS Based Metrics Understate Radiologists' Actual Levels of Agreement. *Comput Biol Med* (2018) 96:294–8. doi: 10.1016/j.compbiomed.2018.04.005
14. Ohta T, Nakata N, Nishioka M, Igarashi T, Fukuda K. Quantitative Differentiation of Benign and Malignant Mammographic Circumscribed Masses Using Intensity Histograms. *Jpn J Radiol* (2015) 33(9):559–65. doi: 10.1007/s11604-015-0456-8
15. Li H, Mendel KR, Lan L, Sheth D, Giger ML. Digital Mammography in Breast Cancer: Additive Value of Radiomics of Breast Parenchyma. *Radiology* (2019) 291(1):15–20. doi: 10.1148/radiol.2019181113
16. Grimm LJ, Anderson AL, Baker JA, Johnson KS, Walsh R, Yoon SC, et al. Interobserver Variability Between Breast Imagers Using the Fifth Edition of the BI-RADS MRI Lexicon. *AJR Am J Roentgenol* (2015) 204(5):1120–4. doi: 10.2214/AJR.14.13047
17. Lunkiewicz M, Forte S, Freiwald B, Singer G, Leo C, Kubik-Huch RA. Interobserver Variability and Likelihood of Malignancy for Fifth Edition BI-RADS MRI Descriptors in Non-Mass Breast Lesions. *Eur Radiol* (2020) 30 (1):77–86. doi: 10.1007/s00330-019-06312-7
18. 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
19. Mao N, Yin P, Wang Q, Liu M, Dong J, Zhang X, et al. Added Value of Radiomics on Mammography for Breast Cancer Diagnosis: A Feasibility Study. *J Am Coll Radiol* (2019) 16(4 Pt A):485–91. doi: 10.1016/j.jacr.2018.09.041
20. Lei C, Wei W, Liu Z, Xiong Q, Yang C, Yang M, et al. Mammography-Based Radiomic Analysis for Predicting Benign BI-RADS Category 4 Calcifications. *Eur J Radiol* (2019) 121:108711. doi: 10.1016/j.ejrad.2019.108711
21. Huang Y, Xiao Q, Sun Y, Wang Z, Li Q, Wang H, et al. An Approach Based on Mammographic Imaging and Radiomics for Distinguishing Male Benign and Malignant Lesions: A Preliminary Study. *Front Oncol* (2021) 10:607235. doi: 10.3389/fonc.2020.607235
22. Niu S, Wang X, Zhao N, Liu G, Kan Y, Dong Y, et al. Radiomic Evaluations of the Diagnostic Performance of DM, DBT, DCE MRI, DWI, and Their Combination for the Diagnosis of Breast Cancer. *Front Oncol* (2021) 11:725922. doi: 10.3389/fonc.2021.725922
23. Ye DM, Wang HT, Yu T. The Application of Radiomics in Breast MRI: A Review. *Technol Cancer Res Treat* (2020) 19:1533033820916191. doi: 10.1177/1533033820916191
24. Holli K, Lääperi AL, Harrison L, Luukkaala T, Toivonen T, Ryymin P, et al. Characterization of Breast Cancer Types by Texture Analysis of Magnetic Resonance Images. *Acad Radiol* (2010) 17(2):135–41. doi: 10.1016/j.acra.2009.08.012
25. Zhou J, Zhang Y, Chang KT, Lee KE, Wang O, Li J, et al. Diagnosis of Benign and Malignant Breast Lesions on DCE-MRI by Using Radiomics and Deep Learning With Consideration of Peritumor Tissue. *J Magn Reson Imaging* (2020) 51(3):798–809. doi: 10.1002/jmri.26981
26. Strigel RM, Burnside ES, Elezaby M, Fowler AM, Kelcz F, Salkowski LR, et al. Utility of BI-RADS Assessment Category 4 Subdivisions for Screening Breast MRI. *AJR Am J Roentgenol* (2017) 208(6):1392–9. doi: 10.2214/ajr.16.16730
27. Nie K, Chen JH, Yu HJ, Chu Y, Nalcioglu O, Su MY. Quantitative Analysis of Lesion Morphology and Texture Features for Diagnostic Prediction in Breast MRI. *Acad Radiol* (2008) 15(12):1513–25. doi: 10.1016/j.acra.2008.06.005
28. Newell D, Nie K, Chen JH, Hsu CC, Yu HJ, Nalcioglu O, et al. Selection of Diagnostic Features on Breast MRI to Differentiate Between Malignant and Benign Lesions Using Computer-Aided Diagnosis: Differences in Lesions

Presenting as Mass and Non-Mass-Like Enhancement. *Eur Radiol* (2010) 20 (4):771–81. doi: 10.1007/s00330-009-1616-y
29. 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
30. Commenges D, Jacqmin H. The Intraclass Correlation Coefficient: Distribution-Free Definition and Test. *Biometrics* (1994) 50(2):517–26. doi: 10.2307/2533395
31. Shi J, Sahiner B, Chan HP, Ge J, Hadjiiski L, Helvie MA, et al. Characterization of Mammographic Masses Based on Level Set Segmentation With New Image Features and Patient Information. *Med Phys* (2008) 35(1):280–90. doi: 10.1118/1.2820630
32. Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V. Support Vector Regression Machines. *Adv Neural Inf Process Syst* (1997) 9:155–61
33. Yu MK, Ma J, Fisher J, Kreisberg JF, Raphael BJ, Ideker T. Visible Machine Learning for Biomedicine. *Cell* (2018) 173(7):1562–5. doi: 10.1016/j.cell.2018.05.056
34. Valdora F, Houssami N, Rossi F, Calabrese M, Tagliafico AS. Rapid Review: Radiomics and Breast Cancer. *Breast Cancer Res Treat* (2018) 169(2):217–29. doi: 10.1007/s10549-018-4675-4
35. Chen S, Guan X, Shu Z, Li Y, Cao W, Dong F, et al. A New Application of Multimodality Radiomics Improves Diagnostic Accuracy of Nonpalpable Breast Lesions in Patients With Microcalcifications-Only in Mammography. *Med Sci Monit* (2019) 25:9786–93. doi: 10.12659/MSM.918721
36. Bennani-Baiti B, Dietzel M, Baltzer PA. MRI for the Assessment of Malignancy in BI-RADS 4 Mammographic Microcalcifications. *PloS One* (2017) 12(11):e0188679. doi: 10.1371/journal.pone.0188679
37. Bennani-Baiti B, Baltzer PA. MR Imaging for Diagnosis of Malignancy in Mammographic Microcalcifications: A Systematic Review and Meta-Analysis. *Radiology* (2017) 283(3):692–701. doi: 10.1148/radiol.2016161106
38. Brnic D, Brnic D, Simundic I, Vanjaka Rogosic L, Tadic T. MRI and Comparison Mammography: A Worthy Diagnostic Alliance for Breast Microcalcifications? *Acta Radiol* (2016) 57(4):413–21. doi: 10.1177/0284185115585036
39. Hopp T, Dietzel M, Baltzer PA, Kreisel P, Kaiser WA, Gemmeke H, et al. Automatic Multimodal 2D/3D Breast Image Registration Using Biomechanical FEM Models and Intensity-Based Optimization. *Med Image Anal* (2013) 17(2):209–18. doi: 10.1016/j.media.2012.10.003
40. Mertzanidou T, Hipwell J, Johnsen S, Han L, Eiben B, Taylor Z, et al. MRI to X-Ray Mammography Intensity-Based Registration With Simultaneous Optimisation of Pose and Biomechanical Transformation Parameters. *Med Image Anal* (2014) 18(4):674–83. doi: 10.1016/j.media.2014.03.003

Author Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Chao Family Comprehensive Cancer Center.

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 23 November 2021
doi: 10.3389/fonc.2021.601425

AI and High-Grade Glioma for Diagnosis and Outcome Prediction: Do All Machine Learning Models Perform Equally Well?

Luca Pasquini[1,2], Antonio Napolitano[3], Martina Lucignani[3], Emanuela Tagliente[3], Francesco Dellepiane[2], Maria Camilla Rossi-Espagnet[2,4], Matteo Ritrovato[5], Antonello Vidiri[6], Veronica Villani[7], Giulio Ranazzi[8], Antonella Stoppacciaro[8], Andrea Romano[2], Alberto Di Napoli[2,9] and Alessandro Bozzao[2]*

[1] Neuroradiology Service, Department of Radiology, Memorial Sloan Kettering Cancer Center, New York, NY, United States, [2] Neuroradiology Unit, Neuroscience, Mental Health and Sensory Organs (NESMOS) Department, Sant'Andrea Hospital, La Sapienza University, Rome, Italy, [3] Medical Physics Department, Bambino Gesù Children's Hospital, Scientific Institute for Research, Hospitalization and Healthcare (IRCCS), Rome, Italy, [4] Neuroradiology Unit, Imaging Department, Bambino Gesù Children's Hospital, Scientific Institute for Research, Hospitalization and Healthcare (IRCCS), Rome, Italy, [5] Unit of Health Technology Assessment (HTA), Biomedical Technology Risk Manager, Bambino Gesù Children's Hospital, Scientific Institute for Research, Hospitalization and Healthcare (IRCCS), Rome, Italy, [6] Radiology and Diagnostic Imaging Department, Regina Elena National Cancer Institute, Scientific Institute for Research, Hospitalization and Healthcare (IRCCS), Rome, Italy, [7] Neuro-Oncology Unit, Regina Elena National Cancer Institute, Scientific Institute for Research, Hospitalization and Healthcare (IRCCS), Rome, Italy, [8] Department of Clinical and Molecular Medicine, Surgical Pathology Units, Sant'Andrea Hospital, La Sapienza University, Rome, Italy, [9] Radiology Department, Castelli Romani Hospital, Rome, Italy

OPEN ACCESS

Edited by:
Marco Rengo,
Sapienza University of Rome, Italy

Reviewed by:
Lorenzo Faggioni,
University of Pisa, Italy
Shun Yao,
Sun Yat-sen University, China

***Correspondence:**
Luca Pasquini
pasquinl@mskcc.org

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

Received: 01 September 2020
Accepted: 02 November 2021
Published: 23 November 2021

Citation:
Pasquini L, Napolitano A, Lucignani M, Tagliente E, Dellepiane F, Rossi-Espagnet MC, Ritrovato M, Vidiri A, Villani V, Ranazzi G, Stoppacciaro A, Romano A, Di Napoli A and Bozzao A (2021) AI and High-Grade Glioma for Diagnosis and Outcome Prediction: Do All Machine Learning Models Perform Equally Well? Front. Oncol. 11:601425. doi: 10.3389/fonc.2021.601425

Radiomic models outperform clinical data for outcome prediction in high-grade gliomas (HGG). However, lack of parameter standardization limits clinical applications. Many machine learning (ML) radiomic models employ single classifiers rather than ensemble learning, which is known to boost performance, and comparative analyses are lacking in the literature. We aimed to compare ML classifiers to predict clinically relevant tasks for HGG: overall survival (OS), isocitrate dehydrogenase (IDH) mutation, O-6-methylguanine-DNA-methyltransferase (MGMT) promoter methylation, epidermal growth factor receptor vIII (EGFR) amplification, and Ki-67 expression, based on radiomic features from conventional and advanced magnetic resonance imaging (MRI). Our objective was to identify the best algorithm for each task. One hundred fifty-six adult patients with pathologic diagnosis of HGG were included. Three tumoral regions were manually segmented: contrast-enhancing tumor, necrosis, and non-enhancing tumor. Radiomic features were extracted with a custom version of Pyradiomics and selected through Boruta algorithm. A Grid Search algorithm was applied when computing ten times K-fold cross-validation (K=10) to get the highest mean and lowest spread of accuracy. Model performance was assessed as AUC-ROC curve mean values with 95% confidence intervals (CI). Extreme Gradient Boosting (xGB) obtained highest accuracy for OS (74,5%), Adaboost (AB) for IDH mutation (87.5%), MGMT methylation (70,8%), Ki-67 expression (86%), and EGFR amplification (81%). Ensemble classifiers showed the best performance across tasks. High-scoring radiomic features shed light on possible correlations between MRI and tumor histology.

Keywords: glioblastoma, machine learning, radiomics, survival, high-grade glioma (HGG), genetics

INTRODUCTION

High-grade gliomas (HGG) are considered the most frequent and lethal primary malignant brain tumors of the adult (1). Glioblastoma multiforme is a type of HGG with an estimated incidence rate of 3.19 per 100,000 persons in the United States, a median age of 64 years, and a dismally poor overall survival (OS) despite combined radio-chemotherapy, ranging approximately between 15 and 17 months (1, 2). Although less frequent, the outcome of HGG is similarly poor in the pediatric population (3). Genetic alterations may influence patient outcome, with effects on survival, disease progression, and treatment response (2, 4). These considerations inspired the cIMPACT recommendations for classification of diffused gliomas and the last revision of the World Health Organization (WHO) classification for central nervous system (CNS) tumors, which suggested considering isocitrate dehydrogenase (IDH)-mutant and IDH-wild-type cancers as two separate entities due to the importance of IDH mutation for patient survival (5, 6).

Artificial intelligence (AI) is the term used to describe the use of computers and technology to simulate intelligent behavior and critical thinking comparable to a human being. Specifically, machine learning (ML) is a subfield of AI, defined as a set of methods that can automatically detect a pattern of data, with the ability of using uncovered patterns to predict future data or perform other kinds of decision-making under uncertainty (7). The learning process can be classified as supervised and unsupervised. Unsupervised learning models identify the pattern class information heuristically, providing clusters without a ground-truth knowledge. On the contrary, the supervised learning approach (explored in this article) identifies a pattern that connects the inputs X to the outputs Y, given a labeled set of input-output pairs. In recent years, AI applications in medicine have grown exponentially, involving almost every medical specialty (8). In the field of radiology, the conversion of biomedical images [such as magnetic resonance imaging (MRI), Computerized Tomography (CT), X-Ray, etc.] to mineable data, and their analysis with AI techniques is defined as "radiomics" (9). Thanks to these new developments, it is possible to extract multiple features from radiological images reflecting tissue characteristics, and use them as input for ML models. For example, graytone distribution and mutual dependencies reflect tissue heterogeneity (10). One of the most interesting applications of ML to radiology is the creation of predictive models to estimate clinically relevant variables. Biomedical images intrinsic parameters (represented by radiomic features) contain information about tissue structure, molecular data, and patient outcome, providing important information for patient care through quantitative image analyses (9, 11). AI-powered analyses may aid diagnosis and prognostication, with practical applications in multiple clinical settings, including emergency care (12).

In brain tumors, radiomic research can identify features that describe the tumor microenvironment (13) and build predictive models for tumor variables and patient outcome. Radiomic models have been shown to outperform clinical models based on patient age, Karnofsky performance scale, surgical resection, genetic alterations, in glioblastoma (GBM) outcome prediction (14, 15). Recent studies proposed several high-performance radiomic models for predicting OS, progression-free survival, molecular subtypes of HGG, as well as genetic alterations critical for clinical practice (16–20). Despite these promising results, clinical implementation is extremely limited due to wide variations of model performances (21–23) and controversial findings. For example, a recent study on 152 patients with GBM concluded that MRI features were not adequate for providing reliable and clinically meaningful predictions through ML classification models (24). A recent review calls for improved standardization and clinical application feasibility (25).

Variability in model performance may depend on parameters optimization. Radiomic workflows comprehend multiple steps requiring parameter choice: tumor segmentation on radiologic images to identify regions of interest (ROIs), feature extraction and selection, training, testing and validation of the AI model, performance evaluation (26, 27). The lack of radiomic parameters standardization might limit results generalizability across studies. A possible solution for this limitation is to compare multiple ML algorithms in the same population for different tasks. In fact, the classification method was shown to be the dominant source of performance variation in radiomic analyses (28). Furthermore, most of radiomic models presented for outcome prediction in HGG employ classic ML algorithms, such as logistic regression, support vector machine, and decisional trees (21, 22). Non-ensemble learners showed inferior performance for small or imbalanced datasets when compared to the ensemble counterpart. Few studies have indeed shown comparative results of single learners *vs* ensemble models (29–31). This is not unexpected considering that single classifier approaches try to learn a single hypothesis from the training set, whereas ensemble learning tries to construct a set of hypotheses and combine them in the best way possible (32). In fact, ensemble methods are used to obtain better predictive performance by reducing both the bias (representational problem) and the variance (computational problem) of learning algorithms (33).

In this study, we chose well-established ML classifiers from previous literature in the field and compared their performance to predict outcome variables of HGG: OS, IDH mutation, O-6-methylguanine-DNA-methyltransferase (MGMT) promoter methylation, epidermal growth factor receptor vIII (EGFR) amplification, and Ki-67 expression, based on features extracted from conventional and advanced MRI. Our objectives were (1) to assess the best algorithm for each prediction task, providing a benchmark for future clinical applications. Particularly, we wanted to compare classic and ensemble learners among ML classifiers to provide a comprehensive view on model performance; (2) to evaluate highly predictive radiomic features extracted from different tumor regions, highlighting possible correlations between MR parameters and tumor molecular/genetic characteristics.

MATERIALS AND METHODS

Subjects

This retrospective observational study was conducted in accordance to the Helsinki declaration. Approval from the

institutional review board (IRB) was obtained with protocol number: 19 SA_2020. Consecutive patients with pathologically proven diagnosis of HGG were recruited from March 2005 to May 2019. Data were collected from two institutions: Sant'Andrea Hospital La Sapienza University of Rome (Institution 1) on a 1.5T scanner (Magnetom Sonata, Siemens, Erlangen, Germany), and Regina Elena Institute of Rome (Institution 2) on a 3T system (Discovery MR 750w, GE Healthcare, Milwaukee, WI, USA). We enrolled patients fulfilling the following inclusion criteria: histopathological diagnosis of HGG, presurgical MRI with at least one sequence among structural T1 or T2-weighted images, diffusion or perfusion-weighted images. Exclusion criteria were causes of suboptimal images (for example motion artifacts) and loss of patients' information during follow-up.

All patients received standard treatment after surgery with the same protocol, including focal radiotherapy (RT) and concomitant temozolomide (TMZ), followed by adjuvant TMZ therapy. RT consisted of fractionated focal irradiation (60 Gy) started within 4 weeks after surgery. The radiation dose was delivered in 30 fractions of 2 Gy over 6 weeks. Chemotherapy with TMZ was administered in a dose of 75 mg/m2, 7 days/week. Adjuvant TMZ started 4 weeks after radiation with the following protocol: 150 mg/m2 for the first cycle, increased to 200 mg/m2 for the second cycle; administered 5 days every 28 days up to 12 cycles.

Prediction labels were associated with survival at 12 months after diagnosis (SURV12), MGMT promoter methylation, IDH mutation, Ki-67 expression, and EGFR amplification. These labels were chosen as they usually provide important prognostic information in HGG. Survival cutoff at 12 months was set based on previous studies (34–36).

Histopathological Analysis

Each tumor specimen was fixed in formaldehyde (10%) and embedded in paraffin. Thin sections (2 μm) were mounted and stained with hematoxylin and eosin. The histopathological examination, including tumor grading, was performed taking into account at least three of the following: cellular atypias, number of mitotes, microvascular proliferation, and/or presence of necrosis. The histopathological examination was performed according to the 2016 edition of the WHO classification of CNS tumors.

Immunohistochemistry

A Dako Envision Flex system was employed for the immunohistochemical analysis. The immunostaining patterns of EGFR were evaluated considering both cellular and tissue distribution. The number of immunopositive cells in 10 high-power (40×) areas were counted, and the percentage of immunopositive cells were estimated. The ratio of positive cells/total number of cells was calculated for each field. The mean value of the 10 fields obtained from a section was considered as the estimated percentage of immunoreactivity assigned to the tumor sample. For IDH-1 mutation analysis, we performed a test with IDH-1 R132H antibody. A positive result was defined when a focal or diffuse immunopositivity was detected, while a negative result was when no immunopositive tumor cells were found. Negative cases were further analyzed for IDH-1/2 mutations as previously shown (37). All sequence reactions were carried out using the GenomeLab DTCS quick-start kit (Beckman Coulter, Fullerton, CA, USA). The reactions were carried out in an automated DNA analyzer (CEQ 8000; Beckman Coulter). All sections were immunostained with Ki-67 antibody. The positivity for Ki67 was determined by counting at least 1,000 tumor cells in a homogeneously stained area and then expressed in percentage.

MGMT Methylation Testing

We used EntroGen's MGMT Methylation Detection Kit (MSPCR, Cat. No. MGMT-RT44), a semiquantitative real-time PCR-based essay for detection of MGMT promoter methylation within the DMR2 locus, distinguishing between methylated and non-methylated cytosines. Its target region starts at chr10:131265513 (hg19 genome build) in the MGMT promoter region and covers CpG sites 75–86. The detection of the amplification product was done by using fluorescent hydrolysis fraction. The procedure involves the following steps: (1) isolation of DNA from tumor biopsies, paraffin-embedded sections; (2) bisulfite treatment of the isolated DNA using the EZ DNA methylation-Lightning Kit (Zymo Research, CATD5030); (3) amplification of treated DNA using the provided reagents in the MGMT Promoter methylation Detection kit; (4) data analysis and interpretation using the real-time PCR software.

MRI Acquisition

MRI sequences were acquired with the same protocol including magnetization-prepared rapid acquisition with gradient echo (MPRAGE), fluid-attenuated inversion recovery (FLAIR), T1-weighted, T2-weigthed, diffusion weighted images (DWI), with apparent diffusion coefficient (ADC) map reconstruction, and perfusion weighted images (PWI) with dynamic susceptibility contrast (DSC) technique. Perfusion parametric maps were obtained through a dedicated software package OleaSphere software version 3.0 (Olea Medical, La Ciotat, France). A relative cerebral blood volume (rCBV) map was generated by using an established tracer kinetic model applied to the first-pass data (38). As previously shown (39), we applied a mathematical correction to the dynamic curves to reduce contrast agent leakage effects. Detailed acquisition parameters can be found in the **Supplementary Material**.

Image Processing and Radiomic Feature Extraction

The radiomic workflow of our analysis was developed following the white paper of the Image Biomarker Standardization Initiative (IBSI) (40) and is summarized in **Figure 1**. For every patient, we automatically co-registered MR data to the MPRAGE sequence using FMRIB Linear Image Registration Tool of FSL (https://fsl.fmrib.ox.ac.uk) (41, 42). Tumors were manually segmented by a neuroradiologist, with three ROIs drawn on MPRAGE and FLAIR images using 3D-Slicer (LP, with 7 years of experience in radiology) (https://www.slicer.org/) (43). Doubtful cases were solved as for consensus with a senior neuroradiologist

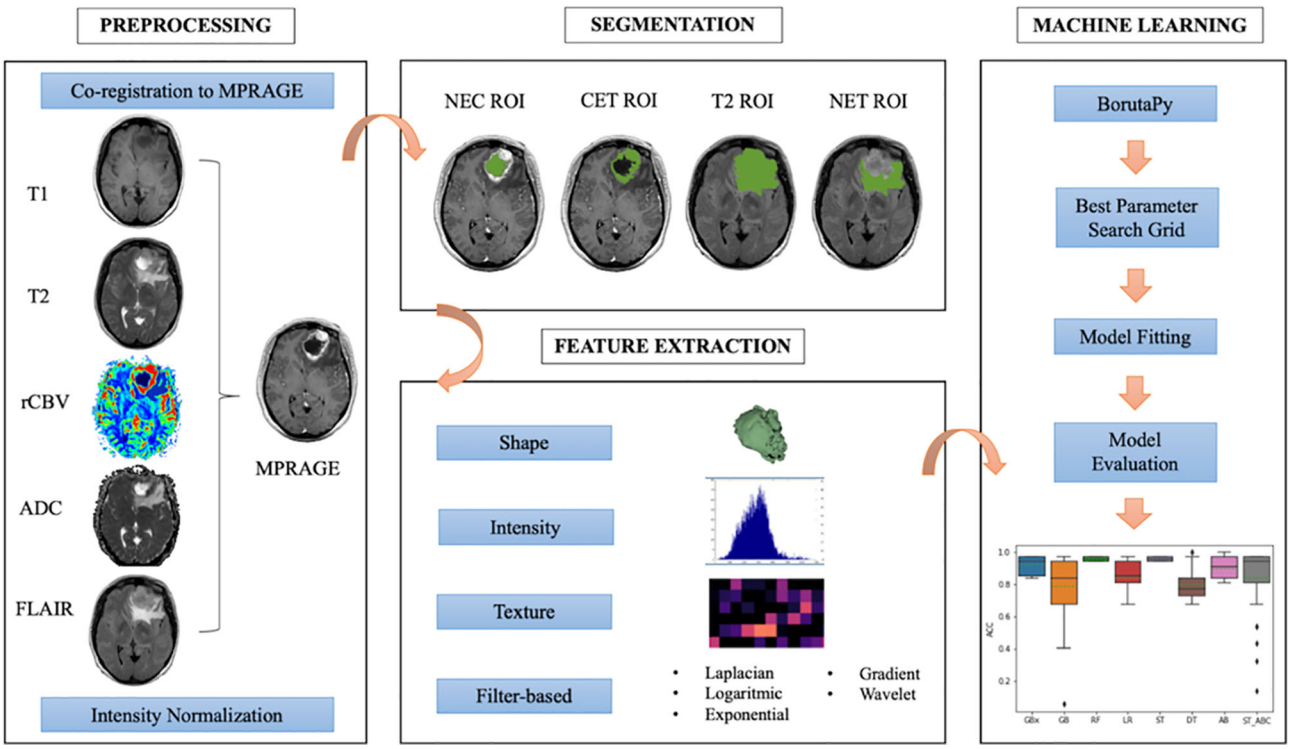


FIGURE 1 | Radiomic workflow followed in the present study.

(AB, with 25 years of experience in radiology). The ROIs were whole tumor (T2), contrast-enhancing tumor (CET), necrosis (NEC). A further non-enhancing tumor (NET) ROI was obtained from the other ROIs as it follows: T2 – (CET+NEC). Based on recent findings (44), we performed intensity non-standardness correction on our multi-institutional data by scaling each image with respect to its mean value within specific brain structure (i.e., NET ROI) using MATLAB R2017a environment (MATLAB 2017, Natick, MA, USA: The MathWorks Inc). The intensity range between 0 and 255 was not rescaled to prevent information loss due to image down-sampling.

We extracted a set of 1,871 radiomic features for each patient from the combination of tumor ROIs (NET, CET, and NEC) and multiparametric MR data (ADC, FLAIR, MPRAGE, rCBV, T1-weigthed, and T2-weighted images). The process was carried out through Pyradiomics package on Python 2.7 (45). Each radiomic set included 14 shape features, 18 intensity features, and 75 texture features [gray-level co-occurrence matrix (GLCM), gray-level difference matrix (GLDM), gray-level size zone matrix (GLSZM), gray-level run length matrix (GLRLM), neighborhood gray tone difference matrix (NGTDM)] from original and filtered images (wavelet decomposition, Laplacian of Gaussian, exponential, logarithmic, and gradient). Additionally, three *ad-hoc* fractal features were computed: box counting two dimensions (2D), box counting three dimensions (3D), and differential box counting, which were integrated in the code of the Pyradiomics pipeline (46). Patients' age at the time of diagnosis was considered a feature in our model for survival prediction only.

Feature Selection and Classification

The pipeline was written in Python and was implemented on Google Colab (47). Prior to any further analysis, each extracted feature distribution was standardized by taking out outliers, removing the mean and scaling it to unit variance with Python Standard Scaler package. Feature selection was then performed in order to identify an ensemble of the most predictive features for each ROI-sequence combination. To this purpose, we used the Boruta algorithm, a powerful and recently introduced feature selector method, that trained a Random Forest Classifier on a duplicated dataset (composed by original and shadow features) and marked a feature as important comparing its Z-scores with that of the duplicate (48). The implementation we used in this work was boruta_py module, freely accessible from github repository (49). Due to the retrospective nature of this study, some MRI sequences were not acquired for all the patients, and some patients lacked full genetic testing, leading to class imbalance issues. In order to overcome this limitation in binary classification, we used Synthetic Minority Over-sampling Technique (SMOTE) approach, which oversamples data of the minority class, creating new synthesized samples from the existing ones (24, 50).

To find the best parameter setting, an optimization search grid algorithm was applied on nine ML classifiers including ensemble and non-ensemble learners (**Figure 2**): AdaBoost (AB),

Extreme Gradient Boosting (xGB), Gradient Boosting (GB), Decision Tree (DT) and Random Forest (RF), Logistic Regressor (LR), two types of Stacking classifiers: stacking (ST) and stacking with AdaBoost (ST_ABC), and KNeighbors (KN). AB, xGB, and GB use a set of weak learners and try to boost them into strong learners. The GB classifier appears in classification studies (24), as it works well with categorical and numerical data; we decided to compare GB performance with xGB, that is the

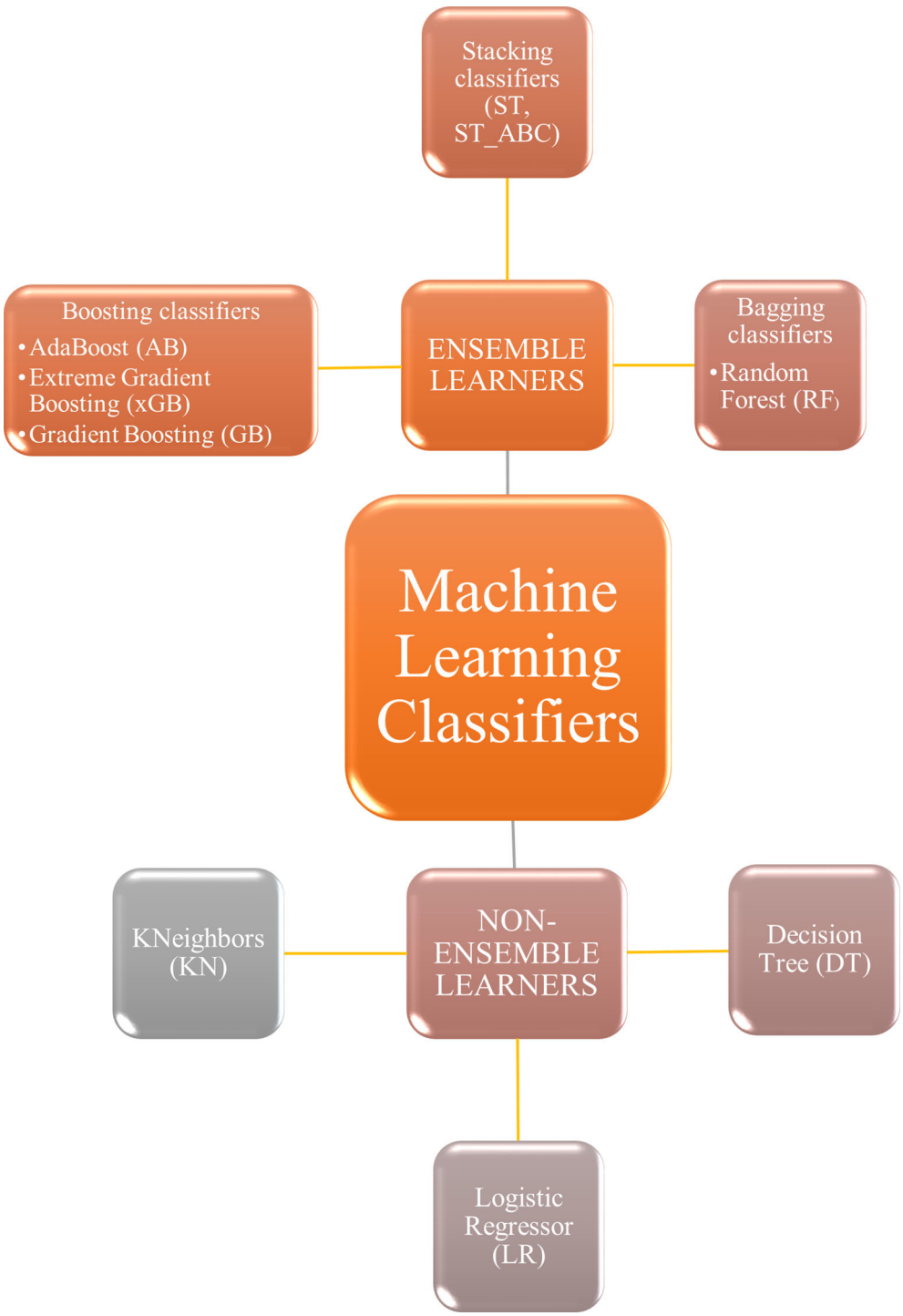


FIGURE 2 | Machine learning classifiers tested in the present study. Non-ensemble learners included KNeighbors, logistic regressor, and decision tree. Ensemble learners included boosting, stacking, and bagging classifiers.

fastest implementation of gradient boosted trees (24, 51). The AB was also often used for brain tumor classification (52, 53), as it works to create a powerful algorithm where instances are reweighted rather than resampled. A Decision Tree algorithm was used in AB as a weak learner. Decision Tree (DT) and Random Forest (RF) are both based upon decision tree algorithms. RF is actually a collection of DTs attempting to classify a new object based on its attributes (54). The RF classifier was already used in brain tumor segmentation problems (55), for the MGMT promoter prediction model (56), for the IDH status prediction (57), and for the survival prediction (58). Logistic Regressor (LR) is one of the most used linear classifiers to disentangle linear relationship between the data (24). The stacked generalization is an ensemble ML algorithm that learns how to best combine the predictions from multiple well-performing ML models. In our case, one classifier was set on the best parameters from GB, RF, and LR (ST), whereas the second was set on best parameters from GB, RF, and AB (ST_ABC) (59). KN relies on distance in data space and is one of the simplest of all the supervised ML algorithms (31). Apart from the extreme gradient boosting classifier which was implemented in xgboost package (60), all classifiers were part of Scikit-learn package (61). Algorithms were chosen based on their known performance and extensive use in the literature.

In order to achieve the most performant and robust model, the Grid Search algorithm, as implemented in Scikit-learn package, was applied when computing 10 times K-fold cross-validation (K=10) and setting the same test split. Given the unbalanced condition for all molecular predictors and in order to reach the same number of trials as for SURV12, an iterative way of K-fold cross-validation was applied. This method made sure that among the possible combinations of data splitting, only those one having the number of minority class subjects at least equal to half of the number of majority class were included among the eligible reshuffles. The Grid Search algorithm was set to look for the highest mean along with the lowest spread of accuracy. The accuracy mean and standard deviation were evaluated on 100 different splitting of training and test data. Once optimal parameters were identified, model performances were also assessed in terms of AUC-ROC curve with 95% CI (28, 62). AUC-ROC curves were also useful when comparing classifiers as they show the trade-off between false positive and true positive rates in the classification (63).

RESULTS

Subjects

The study included 156 adult patients (mean age = 62 y, range = 35–83 y) with confirmed diagnosis of HGG: 121 patients were acquired at Institution 1 and 35 patients at Institution 2. Descriptive statistics performed on genetic variables revealed an odds ratio of 0.607, 1.186, 0.911, and 5.6 for Ki-67, MGMT, IDH, and EGFR respectively, evaluated with reference to SURV12.

Machine Learning Analysis

The distribution of our data is summarized in **Table 1**. For those labels suffering from class imbalance issues, SMOTE was always used. Feature selection produced multiple radiomic signatures composed by 20 features, ordered by importance for the predicted label. The best 15 features for every signature are displayed in the **Supplementary Material**. Nine ML classifiers were compared in the present study. We identified the best classifier and the best ROI-sequence combination in terms of prediction accuracy for each task (SURV12, MGMT, IDH, KI67, and EGFR).

Prediction Performance

Regarding SURV12 prediction, the best performance was achieved by AB and xGB classifiers on ADC radiomic features from NET ROI and T2 radiomic features from NEC ROI (**Table 2**). AB classifier demonstrated accuracy of 73.6% and AUC-ROC mean value of 73.6% (95% CI 71.6–75.3) based on ADC features from NET ROI (**Figure 3A**). xGB classifier achieved accuracy of 74.5% and AUC-ROC mean value of 74.2% (95% CI 71.9–76.3) with T2 radiomic features from NEC ROI (**Figure 3B**). Similarly, xGB classifier provided good accuracy based on FLAIR features from NET ROI (Acc=72.1%; AUC-ROC=72.4%; 95% CI 69.6–75) (**Figure 3C**).

Best results for MGMT prediction (**Table 3**) were obtained from CET ROI on FLAIR images by using AB classifier (Acc=70.8%; AUC-ROC=68.8%; 95% CI 65.9–71.7) (**Figure 4**). High-scoring features mainly included texture parameters (**Figure S4**).

IDH prediction task showed the best performance in our dataset (**Table 4**). Highest accuracy was achieved by AB classifier with rCBV features from NET ROI (Acc= 87.5%; AUC-ROC=86.7%; 95% CI 84.3–89) (**Figure 5A**). Similarly, AB classifier provided good results with T2-based features from CET ROI (Acc=85.9%; AUC-ROC=85.8%; 95% CI 80–84.6) (**Figure 5B**) and NEC ROI (Acc=80.8%; AUC-ROC=80.5%; 95% CI 78.4–82.6) (**Figure 5C**). Good results were also achieved by ST classifier based on T1 features from NET ROI (Acc=84.2%; AUC-ROC=83%; 95% CI 80–85.9) (**Figure 5D**).

The prediction of Ki-67 expression provided excellent results from ADC sequence and CET ROI (**Table 5**). AB classifier provided the highest accuracy (86%) and AUC-ROC value (70%; 95% CI 65.3–72.9) (**Figure 6**).

TABLE 1 | Number of patients and label distributions for label-sequence combination.

	ADC	FLAIR	MPRAGE	rCBV	T1	T2
SURV12 (0/1)	134 (65/69)	140 (68/72)	138 (66/72)	93 (45/48)	122 (61/61)	122 (60/62)
MGMT (0/1)	110 (41/69)	115 (43/72)	114 (42/72)	80 (33/47)	100 (39/61)	102 (39/63)
IDH (0/1)	86 (71/15)	89 (74/15)	89 (74/15)	60 (51/9)	77 (63/14)	78 (65/13)
KI67 (0/1)	100 (18/82)	106 (21/85)	103 (22/81)	77 (16/61)	97 (17/80)	94 (16/78)
EGFR (0/1)	65 (21/44)	69 (23/46)	66 (23/43)	49 (16/33)	65 (22/43)	62 (20/42)

TABLE 2 | Surv12 best results (reported as mean ± standard deviation).

ROI	SEQ		xGB	GB	RF	LR	ST	KN	DT	AB	ST_ABC
NET	ADC	Acc%	71,8 ± 10	68,8 ± 11,4	67,9 ± 6,5	46,3 ± 5,4	71 ± 9	61,2 ± 12,3	59,2 ± 11,7	73,6 ± 9,3	64,2 ± 12,6
NET	ADC	Roc %	71,8 ± 9,7	69,1 ± 11,1	67,9 ± 6,5	46,3 ± 5,4	71 ± 9	61,2 ± 12,3	59,2 ± 11,7	73,6 ± 9,3	64,2 ± 12,6
NET	FLAIR	Acc %	72,1 ± 13,7	67,4 ± 9,9	71,6 ± 8,4	62 ± 13,6	69 ± 12	54,3 ± 15	59 ± 13,7	68,9 ± 7	62,3 ± 14
NET	FLAIR	Roc %	72,4 ± 14	67 ± 11	72,1 ± 7,6	62,3 ± 13,7	69 ± 12,2	53,9 ± 14,8	58,8 ± 13	69,5 ± 7,7	59 ± 13
NEC	T2	Acc %	74,5 ± 11	65,8 ± 12,6	67 ± 16,7	58,7 ± 14,3	73,6 ± 9	52,3 ± 15,2	60,7 ± 11,4	72,7 ± 9,5	58,1 ± 13,9
NEC	T2	Roc %	74,2 ± 10,9	65 ± 11,2	66,4 ± 17	58,8 ± 14,4	73 ± 9,4	52 ± 14,9	59 ± 11	72,5 ± 9,6	56,3 ± 14,3

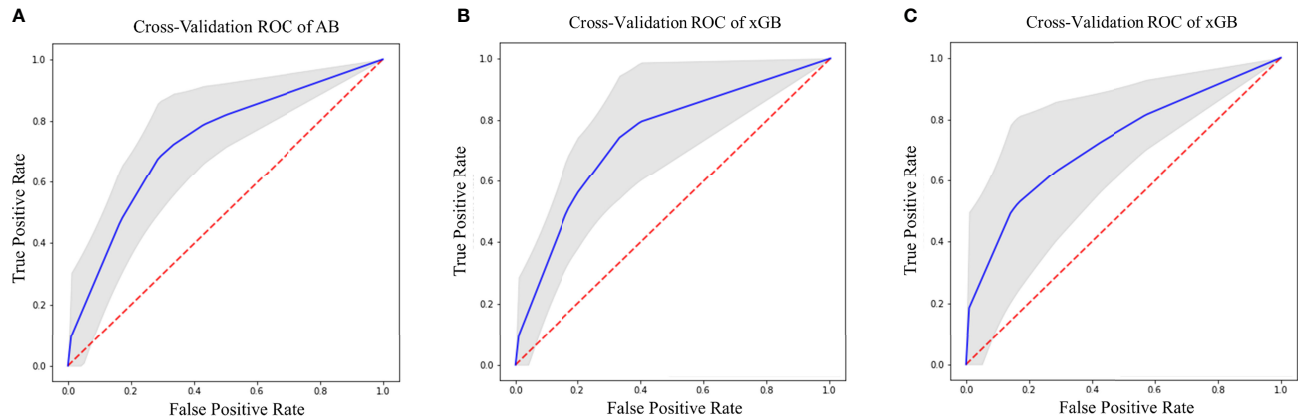


FIGURE 3 | Best ROC curves for Surv12 prediction: **(A)** AB classifier with ADC sequence on NET ROI; **(B)** xGB classifier with T2 sequence on NEC ROI; **(C)** xGB classifier with FLAIR sequence on NET ROI.

TABLE 3 | MGMT best results (reported as mean ± standard deviation).

ROI	SEQ		xGB	GB	RF	LR	ST	KN	DT	AB	ST_ABC
CET	FLAIR	Acc %	63,3 ± 11,3	68,1 ± 13,4	70.7 ± 9,3	65,5 ± 11,4,4	67,9 ± 15,7	52,2 ± 12,7	59,4 ± 14,4	70,8 ± 14,1	64,5 ± 15,7
CET	FLAIR	Roc %	62,8 ± 11,7	66,8 ± 13,4	63,4 ± 12,2	59 ± 10,6	67 ± 16,8	51,4 ± 13,3	55,5 ± 12,1	68,8 ± 14,6	62 ± 14,2

EGFR amplification was correctly predicted by radiomic features extracted from rCBV and T2 images within CET ROI, in both cases with AB classifier (**Table 6**). Particularly, rCBV demonstrated the highest performance (Acc=81%; AUC-ROC=74.3%; 95% CI 70.8–77.8) (**Figure 7A**), while T2 sequence achieved accuracy of 77.8% and AUC-ROC equal to 74.1% (95% CI 70.6–77.6) (**Figure 7B**).

Box-plots figures comparing the best results for each classifier and tables with high-scoring radiomic features are provided in the **Supplementary Material** (**Figures S1**–**S10**).

DISCUSSION

AI has proven to be an accurate tool in predicting survival and molecular profile of gliomas. However, high variability in results across studies and lack of standardization are limiting its use in clinical practice. We studied the best ROI-sequence combination for prediction of clinically relevant variables in HGG, by comparing multiple ML classifiers including classic and ensemble learners. Ensemble classifiers achieved the best performance in every task. The AB was the best classifier overall, with accuracy of 73.6, 70.8, 87.5, 86, and 81% for SURV12, MGMT, IDH, Ki-67, and EGFR respectively, while the LR and KN classifiers always produced suboptimal prediction performances.

These results are in line with previous literature comparing boosting and logistic regression-based classifiers (64). Ensemble models showed high classification performance in different fields. Similar results were observed by Wang et al. using four single classifiers combined with three different algorithms (bagging boosting and stacking) to create ensemble learners for credit scoring (59). All ensemble types yielded a significant improvement compared to base learners (59). In line with our findings, Lu et al. reported higher performances for AdaBoost compared to bagging ensemble algorithms for cancer classification with gene expression data. The idea behind this better performance is that AdaBoost is based on a linear combination of single learners weighted by their own performance, being able to filter out redundant training data attributes and focusing on the important features (65).

Other studies compared ML classifiers in HGG, although with different methodologies and results. Samara et al. conducted a study comparing base models (LR, KN, DT, linear support vector

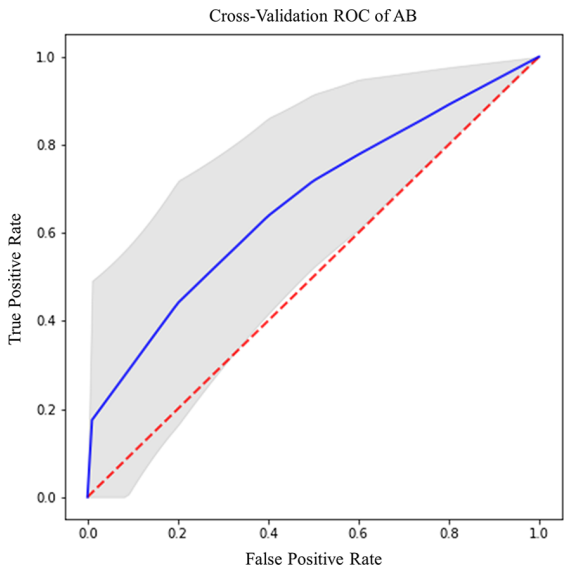


FIGURE 4 | Best ROC curve for MGMT prediction: AB classifier with FLAIR sequence on CET ROI.

machine) and ensemble algorithms (Bootstrap aggregating, AB, RF, and Voting classifier) in a GBM prognostication model based on clinical data (30). In the study, ensemble classifiers attained the highest AUC for every dataset, especially when trained on statistically determined sets or union sets. Osman attempted GBM patients' survival stratification based on conventional MRI sequences with several classifiers. Combining nine selected radiomic features with clinical factors (e.g., age and resection status), even the best prediction accuracy of the ensemble learning classifier appeared low (less than 60%), possibly due to the multi-institutional nature of the study (31). In our approach, we made use of advanced sequences and a larger number of features. Among them we also included fractal dimension-based features which have rarely been implemented in previous studies and may help boosting up the accuracy of our results. Further and important difference regards the use of Boruta algorithm to reduce the features and select only those having higher importance for the model. Also, Kickingereder et al. proposed to evaluate the association of multiparametric MRI features with molecular characteristics (e.g., global DNA methylation subgroup, MGMT, EGFR) in GBM patients, training different models (e.g., stochastic GB, RF, and penalized LR). The authors found associations between established MRI features and molecular characteristics (prediction accuracy of 63% for EGFR with penalized LR). However, the link between them was not strong enough to enable generation of ML classification models for reliable and clinically meaningful predictions (24). In addition to a different set of predicted outcomes, this result might be due to the type and amount of imaging features used for prediction: Kickingereder et al. used 31 imaging parameters for molecular characteristic prediction, while this study extracted 1,871 radiomic features from each image.

A closer look on best performing features and ROI-sequence combinations from our results may unravel interesting associations between MRI parameters and pathologic features of HGG. The best survival prediction was achieved by AB using ADC maps from NET ROI. Also, xGB classifiers showed high performance using T2 images from NEC ROI or FLAIR images from NET ROI, but with higher spread of accuracy (**Table 2**). Previous studies showed heterogeneous results on the same matter (17, 31, 66), depending on size and source of datasets, type and number of extracted features, and model parameters. NET is a common finding in HGG and is considered a combination of infiltrating tumor cells and vasogenic edema (67), whose extension correlates with poor prognosis (68). After surgical resection, recurrence occurs more frequently along the resection margins, due to populations of malignant cells interspersed in the NET (69). Recent research demonstrated that peritumoral MRI textural features from FLAIR and T2 images were predictive of survival as compared to features from enhancing tumor, necrotic regions, and known clinical factors (70, 71). Higher performance of ADC features from NET is coherent with studies demonstrating the inverse correlation between ADC values and tissue cellularity (72–75). In fact, tissue cellularity as measured by ADC can differentiate between vasogenic edema and malignant tumoral tissue within the NET, possibly recognizing patients at higher risk for recurrence (76). Good survival predictivity on NEC ROI is also supported by previous literature. Chaddad et al. reported that shape features, particularly those extracted from necrotic regions, can be used to effectively predict OS of GBM patients (77). Furthermore, our best performing feature for survival prediction on NEC was related to fractal dimension (**Figure S2C**), a measure of shape complexity that has rarely been employed in radiomic studies but demonstrated interesting correlations with patient survival (35).

TABLE 4 | IDH best results (reported as mean ± standard deviation).

ROI	SEQ		xGB	GB	RF	LR	ST	KN	DT	AB	ST_ABC
NET	rCBV	Acc %	83,5 ± 12,8	82,8 ± 12	76,2 ± 16,2	77,3 ± 14,4	86,7 ± 11,8	69,2 ± 17,5	78,7 ± 14,5	87,5 ± 11,9	82,8 ± 12,4
NET	rCBV	Roc %	83,2 ± 12,8	82 ± 13,5	78,3 ± 15,5	78 ± 14,7	85,8 ± 12,3	69 ± 18,3	78,3 ± 15	86,7 ± 12	82 ± 12,4
NET	T1	Acc %	80,2 ± 14	81 ± 13,8	80 ± 12,5	68,7 ± 12	84,2 ± 15	66 ± 21	75,2 ± 13,7	85,9 ± 14	80,9 ± 12
NET	T1	Roc %	79,4 ± 15	80,7 ± 15	78,2 ± 12,3	67,9 ± 11,4	83 ± 14,7	66,7 ± 21,2	76,3 ± 14,5	85,8 ± 14,9	80 ± 13
CET	T2	Acc %	80,2 ± 14	81 ± 13,8	80 ± 12,5	68,7 ± 12	84,2 ± 15	66 ± 21	75,2 ± 13,7	85,9 ± 14	80,9 ± 12
CET	T2	Roc %	79,4 ± 15	80,7 ± 15	78,2 ± 12,3	67,9 ± 11,4	83 ± 14,7	66,7 ± 21,2	76,3 ± 14,5	85,8 ± 14,9	80 ± 13
NEC	T2	Acc %	77,4 ± 9,8	77,9 ± 11	79 ± 11	70,3 ± 12,5	79,2 ± 10,7	69,3 ± 14,3	75,8 ± 12,6	80,8 ± 10,2	79,5 ± 9,5
NEC	T2	Roc %	76,6 ± 10	77 ± 10	78 ± 11,2	70,7 ± 12,6	78,9 ± 9,7	70 ± 14,9	77,5 ± 12,9	80,5 ± 10,6	78,4 ± 9

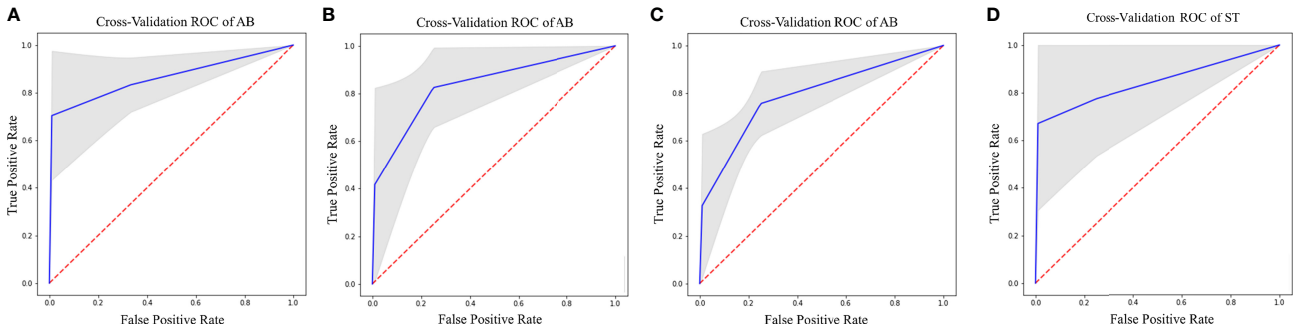


FIGURE 5 | Best ROC curves for IDH prediction: **(A)** AB classifier with rCBV sequence on NET ROI; **(B)** AB classifier with T2 sequence on CET ROI; **(C)** AB classifier with T2 sequence on NEC ROI; **(D)** ST classifier with T1 sequence on NET ROI.

TABLE 5 | KI67 best results (reported as mean ± standard deviation).

ROI	SEQ		xGB	GB	RF	LR	ST	KN	DT	AB	ST_ABC
CET	ADC	Acc %	82,3 ± 8,4	81,6 ± 9,7	83,9 ± 9,8	63,7 ± 13,6	82,6 ± 10,5	67,5 ± 10	76,5 ± 12	86 ± 10,6	83 ± 8,2
CET	ADC	Roc %	64,6 ± 15	64,5 ± 17,3	67,5 ± 18,9	50,8 ± 17,5	63,2 ± 17,8	60 ± 15,7	60 ± 19	70 ± 20	64,4 ± 17

Preoperative prediction of MGMT promoter methylation and IDH mutation represents a crucial objective for radiomic studies due to their pivotal role in patient outcome (2, 4). On conventional and advanced MRI, MGMT methylated HGG may show mixed nodular enhancement, limited edema, lower rCBV, increased Ktrans, and higher ADC minimum values (78, 79). IDH mutant tumors usually show less enhancement, less blood flow on perfusion weighted images, higher mean diffusion values, smaller size, and frontal lobe location (21). Many studies tried to correlate these characteristics with MGMT and IDH status, reporting conflicting results (78). Textural features demonstrated higher accuracy for MGMT promoter methylation prediction, achieving best performance with FLAIR features from CET (70.8%, AB classifier) (**Figures S3** and **S4**). These results are coherent with other reports (80) and confirm that textural features outperform morphological and intensity features in MGMT status prediction (16). Another recent study from Sasaki et al. reported accuracy of 67% for MGMT prediction with textural features (81). A possible explanation for the performance discrepancy is the choice of the classification algorithm: prediction accuracy has great variability depending on the selected model (**Table 3**), with higher performance for ensemble learners. Regarding IDH mutation, our AB classifier achieved an accuracy of 87.5% with rCBV-derived first-order features (median, skewness) from NET (**Figure S6A**), outperforming most of previous models (21, 22). Besides correlating with patient survival (82), perfusion-based features were highly predictive of IDH status in another recent study from our group based on deep-learning (37). Kieckegereder et al. demonstrated that IDH mutation status is associated with a specific hypoxia/angiogenesis transcriptome signature predictable through perfusion MRI (83). Our results seem to confirm a role for perfusion-based analysis in discriminating IDH mutation, reflecting the known correlation with hypoxia inducible factor (HIF) and neoangiogenesis (84). Also, textural features achieved optimal results in the prediction of IDH mutation based on T1 images from NET (84.2%, ST classifier) and T2 images from CET (85.9%, AB classifier). The accumulation of D-2HG derived from IDH mutation induces epigenetic changes that lead to abnormal gene expression and impaired cellular differentiation, possibly contributing to intratumoral heterogeneity. Hsieh et al. demonstrated that textural features can differentiate IDH mutation with 85%

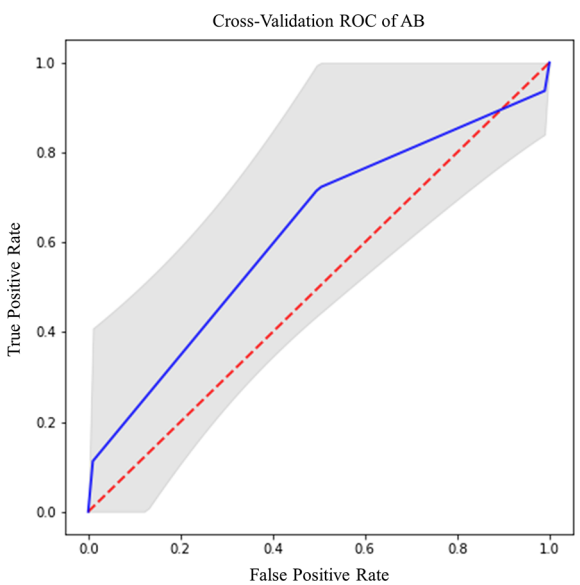


FIGURE 6 | Best ROC curve for KI67 prediction: AB classifier with ADC sequence on CET ROI.

TABLE 6 | EGFR best results (reported as mean ± standard deviation).

ROI	SEQ		xGB	GB	RF	LR	ST	KN	DT	AB	ST_ABC
CET	rCBV	Acc %	69,8 ± 15,1	75,4 ± 15	73,1 ± 16	64,3 ± 16,3	72,9 ± 14,3	61,3 ± 21,4	66,7 ± 19,4	81 ± 13,8	66,5 ± 18,7
CET	rCBV	Roc %	63,9 ± 19,5	64,6 ± 18,5	64,7 ± 20	62,2 ± 21,8	65,7 ± 18,9	63,4 ± 23,3	59,4 ± 23,2	74,3 ± 17,3	62,6 ± 20
CET	T2	Acc %	76,4 ± 15,2	74,7 ± 15	76,4 ± 16	60,8 ± 18,8	76 ± 17,8	59,7 ± 20,4	61,3 ± 18,7	77,8 ± 13,8	71,5 ± 16
CET	T2	Roc %	70,4 ± 22,7	69,7 ± 19,8	76,3 ± 17	65,4 ± 15,7	69,8 ± 22,8	60,2 ± 19,5	55,7 ± 20,4	74,1 ± 17,6	65,6 ± 20,6

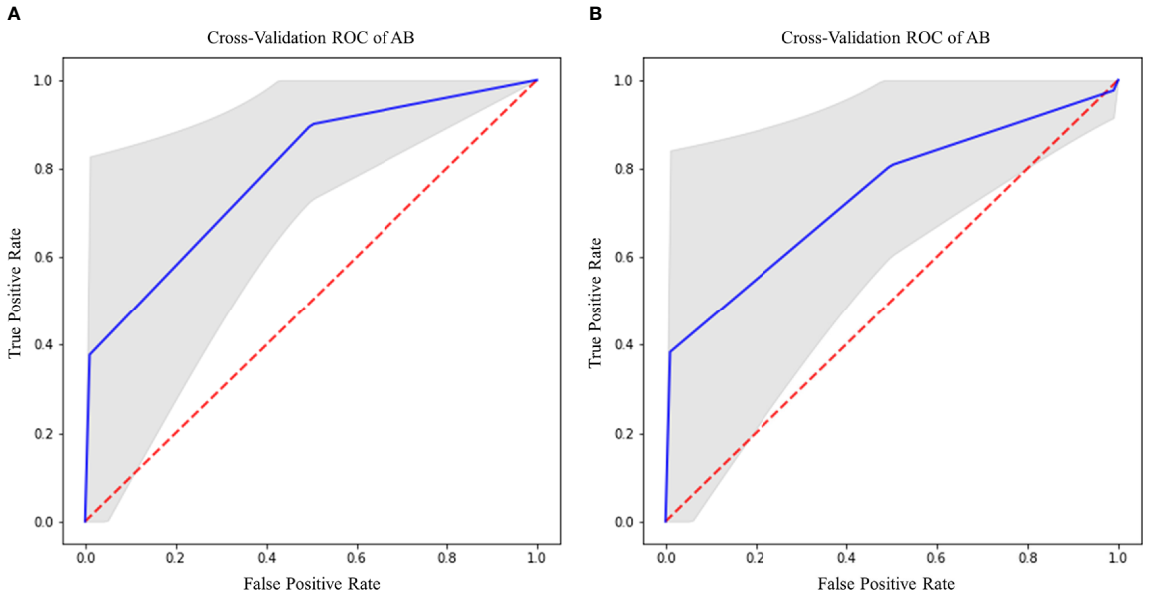


FIGURE 7 | Best ROC curves for EGFR prediction: **(A)** AB classifier with rCBV sequence on CET ROI; **(B)** AB classifier with T2 sequence on CET ROI.

accuracy in 39 patients with GBM. The Authors performed tailored biopsies demonstrating an agreement between prediction results and biopsy-proven pathology of 0.60 (85). Shape features of tumor necrosis demonstrated good accuracy for IDH mutation prediction in our model (**Figure S6D**). Such result may partly explain the relation between necrosis shape and survival as previously discussed (35, 77).

Ki-67 is a nuclear protein expressed by cells entering the mitotic cycle. In gliomas, the expression of Ki-67 is roughly proportional to the histologic grade, representing a proliferative index with prognostic correlation (86). Radiomic models predictive of Ki-67 expression have not been investigated before in the literature. In our analysis we achieved an accuracy of 86% for predicting Ki-67 expression through the AB. Intriguingly, best performing features were texture-based parameters extracted from the solid tumor (CET ROI) on ADC maps (**Figure S8**). These results perfectly agree with the role of Ki-67 as proliferative index in HGG, being ADC an MRI surrogate of cellularity (72, 73).

EGFR is a transmembrane tyrosine-kinase receptor for different growth factors, whose activation leads to DNA synthesis and cellular proliferation (87). Amplification of EGFR (especially EGFRvIII) is a common somatic mutation in HGG (4), with high relevance for the definition of GBM in the recent classification (6). Despite failure of initial attempts of targeting EGFR for therapy, the receptor remains of value for possible future treatments (87). In our results, EGFR showed best prediction performance with ST and AB classifiers. Particularly, rCBV features achieved a performance of 81% with AB classifier and T2 features achieved a performance of 77.8% with AB classifier on CET ROI. Highest scoring features were median intensity values for rCBV and textural features for T2 (**Figures S10A, B**). These results are supported by previous evidence. Hu et al. demonstrated a link between EGFR amplification and rCBV textural features, with correlation to microvessel volume and angiogenesis on tumor biopsies (88). Similarly, T2 textural features were shown to correlate to EGFR amplification (88).

Our study had some limitations. Firstly, even though ML studies in HGG often rely on limited populations (18, 19, 34, 36, 62, 77, 85, 88, 89), our sample size (156 patients) could be considered small. Nevertheless, our dataset includes clinical/genetic information (e.g., survival, MGMT, IDH, EGFR, and KI67), together with radiomic data from different MRI sequences (e.g., MPRAGE, FLAIR, ADC, rCBV, T1-wiethed, and T2-weighted), thus allowing us to combine information from different sources to better predict clinical and genetic variables. Due to the retrospective nature of the study, some sequences were not acquired for all the patients (**Table 1**). For this reason,

prediction accuracy for each label was evaluated separately on each sequence, thus limiting performance bias. Moreover, some labels were not available for all the patients; consequently, the number of subjects split in train and test groups changed for each label-sequence combination. We tried to overcome this limitation by employing two well-known and effective techniques with the aim of balancing the asymmetric labels. Although undersampling of the majority class was considered a more effective approach in respect to an oversampling method (90), we decided to use SMOTE for unbalancing issues. As demonstrated in other SMOTE-based studies (24, 91), it could represent a suitable solution for our purposes. In order to overcome main SMOTE drawbacks (92, 93) we perform ML analysis with a significant number of cross-validations. Since we only split subjects into train and test groups, the lack of an additional validation cohort could represent a limitation of this study. To overcome this issue, we decided to report range of performance obtained applying four times stratified K-fold cross-validation. This approach provides a full accuracy range, which includes the results that an eventual validation test would produce.

CONCLUSIONS

In the present study we were able to predict patient OS and highly relevant molecular features of HGG from preoperative MRI, comparing different ML classifiers. Ensemble classifiers (AB, ST, GB, and xGB) showed optimal performance in prediction tasks for all the studied variables. In particular, AB and xGB obtained maximum accuracy for survival, AB for IDH mutation, MGMT promotor methylation status and Ki-67 expression, and EGFR amplification. Ensemble learning outperformed classic ML algorithms in all tests, in agreement with previous literature. Best performing features from our analysis shed light on possible correlations between MRI and tumor histology, as well as molecular profiles and patient outcome in HGG. Our results may set a path for ML analysis standardization and clinical application. Future developments may include the evaluation of other genetic abnormalities, prediction of recurrence, and response to therapy.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available upon reasonable request to the authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Sant'Andrea Hospital, *via* Grottarossa 1035, 00189, Rome, Italy. Protocol Number: 19 SA_2020. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LP and AN made substantial contributions to the conception and design of the work. LP, AN, ADN, FD, AV, VV, GR, and AS contributed to data acquisition and supervision. AN, ML, ET, and MR contributed to data analysis. LP, AN, ADN, MCR-E, AR, and AB contributed to data interpretation. LP, ML, and EM drafted the manuscript. All authors substantially revised the manuscript. All authors approved the submitted version.

FUNDING

This study was supported by the grant "Progetti di Ateneo 2020" from La Sapienza University (Protocol ID: RP120172B9E252BD). Funding sources did not influence any phase of the present study.

ACKNOWLEDGMENTS

We thank Cassa Galeno for the support provided to the project through the "Eleonora Cantamessa" gold medal award, 2019 edition. We thank Dr. Matteo Nicolai and Giulia Moltoni for the support in reviewing the data for the present study.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.601425/full#supplementary-material

Supplementary Figure 1 | Best results box-plots for Surv12 prediction among all sequences and ROI combination for all classifiers.

Supplementary Figure 2 | First 15 significative features extracted from NET ROI on sequence ADC **(A)** and FLAIR **(B)** and from NEC ROI on sequence T2 **(C)** that provided best Surv12 predictions.

Supplementary Figure 3 | Best results box-plots for MGMT prediction among all sequences and ROI combination for all classifiers.

Supplementary Figure 4 | First 15 significative features extracted from CET ROI on sequence FLAIR that provided best MGMT predictions.

Supplementary Figure 5 | Best results box-plots for IDH prediction among all sequences and ROI combination for all classifiers.

Supplementary Figure 6 | First 15 significative features extracted from NET ROI on sequence rCBV **(A)** and T1 **(B)**, from CET ROI on sequence T2 **(C)**, and within NEC ROI on sequence T2 **(D)**, that provided best IDH predictions.

Supplementary Figure 7 | Best results box-plots for KI-67 prediction among all sequences and ROI combination for all classifiers.

Supplementary Figure 8 | First 15 significative features extracted from CET ROI on ADC sequence, that provided best KI67 predictions.

Supplementary Figure 9 | Best results box-plots for EGFR prediction among all sequences and ROI combination for all classifiers.

Supplementary Figure 10 | First 15 significative features extracted from CET ROI on rCBV **(A)** and T2 sequences, that provided best EGFR predictions.

REFERENCES

1. Tamimi AF, Juweid M. Epidemiology and Outcome of Glioblastoma. In: de Vleeschouwer S, editor. *Glioblastoma*. Brisbane (AU): Codon Publications (2017). doi: 10.15586/codon.glioblastoma.2017.ch8
2. Molinaro AM, Taylor JW, Wiencke JK, Wrensch MR. Genetic and Molecular Epidemiology of Adult Diffuse Glioma. *Nat Rev Neurol* (2019) 15:405–17. doi: 10.1038/s41582-019-0220-2
3. Braunstein S, Raleigh D, Bindra R, Mueller S, Haas-Kogan D. Pediatric High-Grade Glioma: Current Molecular Landscape and Therapeutic Approaches. *J Neuro Oncol* (2017) 134:541–9. doi: 10.1007/s11060-017-2393-0
4. Wang J, Bettegowda C. Genomic Discoveries in Adult Astrocytoma. *Curr Opin Genet Dev* (2015) 30:17–24. doi: 10.1016/j.gde.2014.12.002
5. Louis DN, Wesseling P, Aldape K, Brat DJ, Capper D, Cree IA, et al. cIMPACT-NOW Update 6: New Entity and Diagnostic Principle Recommendations of the cIMPACT-Utrecht Meeting on Future CNS Tumor Classification and Grading. *Brain Pathol* (2020) 30:844–56. doi: 10.1111/bpa.12832
6. Rushing EJ. WHO Classification of Tumors of the Nervous System: Preview of the Upcoming. *5th edition. Memo* (2020) 14:188–91. doi: 10.1007/s12254-021-00680-x
7. Murphy KP. *Machine Learning*. Cambridge, Massachusetts London, England: The MIT Press (1988). doi: 10.1111/j.1468-0394.1988.tb00341.x.
8. Briganti G, le Moine O. Artificial Intelligence in Medicine: Today and Tomorrow. *Front Med* (2020) 7:27. doi: 10.3389/fmed.2020.00027
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. Haralick RM, Dinstein I, Shanmugam K. Textural Features for Image Classification. *IEEE Trans Systems Man Cybernetics* (1973) SMC-3:610–21. doi: 10.1109/TSMC.1973.4309314
11. Barajas RF, Phillips JJ, Parvataneni R, Molinaro A, Essock-burns E, Bourne G, et al. Regional Variation in Histopathologic Features. *Neuro-Oncology* (2012) 14:942–54. doi: 10.1093/neuonc/nos128
12. Bottino F, Tagliente E, Pasquini L, Napoli A, Lucignani M, Talamanca LF, et al. COVID Mortality Prediction With Machine Learning Methods : A Systematic Review and Critical Appraisal. *J Personalized Med* (2021) 11:893. doi: 10.3390/jpm11090893
13. Rudie JD, Rauschecker AM, Bryan RN, Davatzikos C, Mohan S. Emerging Applications of Artificial Intelligence in Neuro-Oncology. *Radiology* (2019) 00:1–12. doi: 10.1148/radiol.2018181928
14. Bae S, Choi YS, Ahn SS, Chang JH, Kang SG, Kim EH, et al. Radiomic MRI Phenotyping of Glioblastoma: Improving Survival Prediction. *Radiology* (2018) 289:797–806. doi: 10.1148/radiol.2018180200
15. Kickingereder P, Neuberger U, Bonekamp D, Piechotta PL, Götz M, Wick A, et al. Radiomic Subtyping Improves Disease Stratification Beyond Key Molecular, Clinical, and Standard Imaging Characteristics in Patients With Glioblastoma. *Neuro-Oncology* (2018) 20:848–57. doi: 10.1093/neuonc/nox188
16. Li ZC, Bai H, Sun Q, Li Q, Liu L, Zou Y, et al. Multiregional Radiomics Features From Multiparametric MRI for Prediction of MGMT Methylation Status in Glioblastoma Multiforme: A Multicentre Study. *Eur Radiol* (2018) 28:3640–50. doi: 10.1007/s00330-017-5302-1
17. Sanghani P, Ang BT, King NKK, Ren H. Overall Survival Prediction in Glioblastoma Multiforme Patients From Volumetric, Shape and Texture Features Using Machine Learning. *Surg Oncol* (2018) 27:709–14. doi: 10.1016/j.suronc.2018.09.002
18. Zhang B, Chang K, Ramkissoon S, Tanguturi S, Bi WL, Reardon DA, et al. Multimodal MRI Features Predict Isocitrate Dehydrogenase Genotype in High-Grade Gliomas. *Neuro-Oncology* (2017) 19:109–17. doi: 10.1093/neuonc/now121
19. Macyszyn L, Akbari H, Pisapia JM, Da X, Attiah M, Pigrish V, et al. Imaging Patterns Predict Patient Survival and Molecular Subtype in Glioblastoma *via* Machine Learning Techniques. *Neuro-Oncology* (2016) 18:417–25. doi: 10.1093/neuonc/nov127
20. Lao J, Chen Y, Li ZC, Li Q, Zhang J, Liu J, et al. A Deep Learning-Based Radiomics Model for Prediction of Survival in Glioblastoma Multiforme. *Sci Rep* (2017) 7:1–8. doi: 10.1038/s41598-017-10649-8
21. Chow D, Chang P, Weinberg BD, Bota DA, Grinband J, Filippi CG. Imaging Genetic Heterogeneity in Glioblastoma and Other Glial Tumors: Review of Current Methods and Future Directions. *Am J Roentgenol* (2018) 210:30–8. doi: 10.2214/AJR.17.18754
22. Fathi Kazerooni A, Bakas S, Saligheh Rad H, Davatzikos C. Imaging Signatures of Glioblastoma Molecular Characteristics: A Radiogenomics Review. *J Magnetic Resonance Imaging* (2020) 52:54–69. doi: 10.1002/jmri.26907
23. Kawaguchi RK, Takahashi M, Miyake M, Kinoshita M, Takahashi S, Ichimura K, et al. Assessing Versatile Machine Learning Models for Glioma Radiogenomic Studies Across Hospitals. *Cancers* (2021) 13:3611. doi: 10.3390/cancers13143611
24. Kickingereder P, Bonekamp D, Nowosielski M, Kratz A, Sill M, Burth S, et al. Radiogenomics of Glioblastoma : Machine Learning – Based Classification of Molecular Characteristics by Using Multiregional Imaging Features. *Radiology* (2017) 000:1–12. doi: 10.1148/radiol.2016161382
25. Sotoudeh H, Shafaat O, Bernstock JD, Brooks MD, Elsayed GA, Chen JA, et al. Artificial Intelligence in the Management of Glioma: Era of Personalized Medicine. *Front Oncol* (2019) 9:768. doi: 10.3389/fonc.2019.00768
26. Larue RTHM, Defraene G, de Ruysscher D, Lambin P, Elmpt WV. Quantitative Radiomics Studies for Tissue Characterization: A Review of Technology and Methodological Procedures. *Br J Radiol* (2017) 90:20160665. doi: 10.1259/bjr.20160665
27. Li Q, Bai H, Chen Y, Sun Q, Liu L, Zhou S, et al. A Fully-Automatic Multiparametric Radiomics Model: Towards Reproducible and Prognostic Imaging Signature for Prediction of Overall Survival in Glioblastoma Multiforme. *Sci Rep* (2017) 7:1–9. doi: 10.1038/s41598-017-14753-7
28. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts HJWL. Machine Learning Methods for Quantitative Radiomic Biomarkers. *Sci Rep* (2015) 5:1–11. doi: 10.1038/srep13087
29. Brunese L, Mercaldo F, Reginelli A, Santone A. An Ensemble Learning Approach for Brain Cancer Detection Exploiting Radiomic Features. *Comput Methods Programs Biomed* (2020) 185:105134. doi: 10.1016/j.cmpb.2019.105134
30. Samara KA, Al Aghbari Z, Abusafia A. GLIMPSE: A Glioblastoma Prognostication Model Using Ensemble Learning—A Surveillance, Epidemiology, and End Results Study. *Health Inf Sci Syst* (2021) 9:5. doi: 10.1007/s13755-020-00134-4
31. Osman AFI. A Multi-Parametric MRI-Based Radiomics Signature and a Practical ML Model for Stratifying Glioblastoma Patients Based on Survival Toward Precision Oncology. *Front Comput Neurosci* (2019) 13:58. doi: 10.3389/fncom.2019.00058
32. Zhou ZH. Ensemble Learning. In: Li SZ, Jain A, editors. *Encyclopedia of Biometrics*. Boston, MA: Springer (2015).
33. Dietterich TG. Ensemble Learning. In: *The Handbook of Brain Theory and Neural Networks*. Cambridge, MA: MIT Press (1998).
34. Lee J, Jain R, Khalil K, Griffith B, Bosca R, Rao G, et al. Texture Feature Ratios From Relative CBV Maps of Perfusion MRI Are Associated With Patient Survival in Glioblastoma. *Am J Neuroradiol* (2016) 37:37–43. doi: 10.3174/ajnr.A4534
35. 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) 7:1–7. doi: 10.1038/s41598-017-08862-6
36. Yang D, Rao G, Martinez J, Veeraraghavan A, Rao A. Evaluation of Tumor-Derived MRI-Texture Features for Discrimination of Molecular Subtypes and Prediction of 12-Month Survival Status in Glioblastoma. *Med Phys* (2015) 42:6725–35. doi: 10.1118/1.4934373
37. Pasquini L, Napolitano A, Tagliente E, Dellepiane F, Lucignani M, Vidiri A, et al. Deep Learning can Differentiate IDH-Mutant From IDH-Wild Type GBM. *J Personalized Med* (2021) 1–12. doi: 10.3390/jpm11040290
38. Ostergaard L, Weisskoff RM, Chesler DA, Gyldensted C, Rosen BR. High Resolution Measurement of Cerebral Blood Flow Using Intravascular Tracer Bolus Passages. Part I: Mathematical Approach and Statistical Analysis. *Magnetic Resonance Med* (1996) 36:715–25. doi: 10.1002/mrm.1910360510
39. Boxerman JL, Schmainda KM, Weisskoff RM. Relative Cerebral Blood Volume Maps Corrected for Contrast Agent Extravasation Significantly Correlate With Glioma Tumor Grade, Whereas Uncorrected Maps Do Not. *Ajnr Am J Neuroradiol* (2006) 27:859–67.

40. 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) 295:328–38. doi: 10.1148/radiol.2020191145
41. Jenkinson M, Bannister P, Brady M, Smith S. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. *NeuroImage* (2002) 17:825–41. doi: 10.1006/nimg.2002.1132
42. Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, et al. Advances in Functional and Structural MR Image Analysis and Implementation as FSL. *NeuroImage* (2004) 23:208–19. doi: 10.1016/j.neuroimage.2004.07.051
43. 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. *Magnetic Resonance Imaging* (2012) 30:1323–41. doi: 10.1016/j.mri.2012.05.001
44. 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
45. Available at: http://www.radiomics.io/pyradiomics.html (Accessed December 2, 2020).
46. MatLab. Available at: https://www.mathworks.com/matlabcentral/fileexchange/13063-boxcount (Accessed December 2, 2020).
47. Bisong E. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners. 1st ed*. Apress (2019).
48. Kursa MB, Rudnicki WR. Feature Selection With the Boruta Package. *J Stat Softw* (2010) 36:1–13. doi: 10.18637/jss.v036.i11
49. Boruta_py. Available at: https://github.com/scikit-learn-contrib/boruta_py (Accessed December 2, 2020).
50. 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
51. Revanuru K, Shah N. Fully Automatic Brain Tumour Segmentation Using Random Forest and Patient Survival Prediction Using XGBoost. In: *Proceedings of the 6th MICCAI-BRATS Challenge*. p. 239–43.
52. Sonavane R, Sonar P. Classification and Segmentation of Brain Tumor Using Adaboost Classifier. In: *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication ICGTSPICC* (2016). p. 396–403. doi: 10.1109/ICGTSPICC.2016.7955334
53. Usman K, Rajpoot K. Brain Tumor Classification From Multi-Modality MRI Using Wavelets and Machine Learning. *Pattern Anal Appl* (2017) 20:871–81. doi: 10.1007/s10044-017-0597-8
54. Naik J, Patel PS. Tumor Detection and Classification Using Decision Tree in Brain MRI. *Int J of Eng Dev and Res* (2013) 49–53.
55. Maier O, Wilms M, Handels H. Image Features for Brain Lesion Segmentation Using Random Forests. In: Crimi A, Maier O, Menze B, Handels H, editors. *LNCS Brainlesion Glioma, MS, Stroke Trauma. Brain Inj. - First Int. BrainLes Work. MICCAI 2015*. Berlin Heidelberg: Springer (2016).
56. Kanas VG, Zacharaki EI, Thomas GA, Zinn PO, Megalooikonomou V, Colen RR. Learning MRI-Based Classification Models for MGMT Methylation Status Prediction in Glioblastoma. *Comput Methods Programs Biomed* (2017) 140:249–57. doi: 10.1016/j.cmpb.2016.12.018
57. de Looze C, Beausang A, Cryan J, Loftus T, Buckley PG, Farrell M, et al. Machine Learning: A Useful Radiological Adjunct in Determination of a Newly Diagnosed Glioma's Grade and IDH Status. *J Neuro Oncol* (2018) 139:491–9. doi: 10.1007/s11060-018-2895-4
58. Shboul ZA, Vidyaratne L, Alam M, Iftekharuddin KM. Glioblastoma and Survival Prediction. In: *Lecture Notes in Computer Science*.
59. Wang G, Hao J, Ma J, Jiang H. A Comparative Assessment of Ensemble Learning for Credit Scoring. *Expert Syst Appl* (2011) 38:223–30. doi: 10.1016/j.eswa.2010.06.048
60. *Xgboost*. Available at: https://github.com/dmlc/xgboost (Accessed December 2, 2020).
61. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Machine Learning in Python. *J Mach Learn Res* (2011) 12:2825–30. doi: 10.4018/978-1-5225-9902-9.ch008
62. Xi Yb, Guo F, Xu ZL, Li C, Wei W, Tian P, et al. Radiomics Signature: A Potential Biomarker for the Prediction of MGMT Promoter Methylation in Glioblastoma. *J Magnetic Resonance Imaging* (2018) 47:1380–7. doi: 10.1002/jmri.25860
63. Hand DJ, Anagnostopoulos C. When is the Area Under the Receiver Operating Characteristic Curve an Appropriate Measure of Classifier Performance? *Pattern Recognition Lett* (2013) 34:492–5. doi: 10.1016/j.patrec.2012.12.004
64. Stollhoff R, Sauerbrei W, Schumacher M. An Experimental Evaluation of Boosting Methods for Classification. *Methods Inf Med* (2010) 49:219–29. doi: 10.3414/ME0543
65. Lu H, Gao H, Ye M, Wang X. A Hybrid Ensemble Algorithm Combining AdaBoost and Genetic Algorithm for Cancer Classification With Gene Expression Data. *IEEE/ACM Trans Comput Biol Bioinf* (2019) 5963:1–1. doi: 10.1109/tcbb.2019.2952102
66. Chen X, Fang M, Dong D, Jiang X, Qin L, Liu Z. Development and Validation of a MRI-Based Radiomics Prognostic Classifier in Patients With Primary Glioblastoma Multiforme. *Acad Radiol* (2019) 26:1292–300. doi: 10.1016/j.acra.2018.12.016
67. Chang EL, Akyurek S, Avalos T, Rebueno N, Spicer C, Garcia J, et al. Evaluation of Peritumoral Edema in the Delineation of Radiotherapy Clinical Target Volumes for Glioblastoma. *Int J Radiat Oncol Biol Phys* (2007) 68:144–50. doi: 10.1016/j.ijrobp.2006.12.009
68. Schoenegger K, Oberndorfer S, Wuschitz B, Struhal W, Hainfellner J, Prayer D, et al. Peritumoral Edema on MRI at Initial Diagnosis: An Independent Prognostic Factor for Glioblastoma? *Eur J Neurol* (2009) 16:874–8. doi: 10.1111/j.1468-1331.2009.02613.x
69. Ruiz-Ontañon P, Orgaz JL, Aldaz B, Elosegui-Artola A, Martino J, Berciano MT, et al. Cellular Plasticity Confers Migratory and Invasive Advantages to a Population of Glioblastoma-Initiating Cells That Infiltrate Peritumoral Tissue. *Stem Cells* (2013) 31:1075–85. doi: 10.1002/stem.1349
70. Prasanna P, Patel J, Partovi S, Madabhushi A, Tiwari P. Radiomic Features From the Peritumoral Brain Parenchyma on Treatment-Naïve 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
71. Choi Y, Ahn KJ, Nam Y, Jang J, Shin NY, Choi HS, et al. Analysis of Peritumoral Hyperintensity on Pre-Operative T2-Weighted MR Images in Glioblastoma: Additive Prognostic Value of Minkowski Functionals. *PloS One* (2019) 14:1–13. doi: 10.1371/journal.pone.0217785
72. Chang PD, Malone HR, Bowden SG, Chow DS, Gill BJA, Ung TH, et al. A Multiparametric Model for Mapping Cellularity in Glioblastoma Using Radiographically Localized Biopsies. *Am J Neuroradiol* (2017) 38:890–8. doi: 10.3174/ajnr.A5112
73. la Violette PS, Mickevicius NJ, Cochran EJ, Rand SD, Connelly J, Bovi JA, et al. Precise Ex Vivo Histological Validation of Heightened Cellularity and Diffusion-Restricted Necrosis in Regions of Dark Apparent Diffusion Coefficient in 7 Cases of High-Grade Glioma. *Neuro-Oncology* (2014) 16:1599–606. doi: 10.1093/neuonc/nou142
74. Gadda D, Mazzoni LN, Pasquini L, Busoni S, Simonelli P, Giordano GP. Relationship Between Apparent Diffusion Coefficients and MR Spectroscopy Findings in High-Grade Gliomas. *J Neuroimaging* (2017) 27:128–34. doi: 10.1111/jon.12350
75. Pasquini L, di Napoli A, Napolitano A, Lucignani M, Dellepiane F, Vidiri A, et al. Glioblastoma Radiomics to Predict Survival: Diffusion Characteristics of Surrounding non-Enhancing Tissue to Select Patients for Extensive Resection. *J Neuroimaging* (2021) 31:1192–200. doi: 10.1111/jon.12903
76. Lemée JM, Clavreul A, Menei P. Intratumoral Heterogeneity in Glioblastoma: Don't Forget the Peritumoral Brain Zone. *Neuro-Oncology* (2015) 17:1322–32. doi: 10.1093/neuonc/nov119
77. Chaddad A, Desrosiers C, Hassan L, Tanougast C. A Quantitative Study of Shape Descriptors From Glioblastoma Multiforme Phenotypes for Predicting Survival Outcome. *Br J Radiol* (2016) 89:20160575. doi: 10.1259/bjr.20160575
78. Smits M, van den Bent MJ. Imaging Correlates of Adult Glioma Genotypes. *Radiology* (2017) 284:316–31. doi: 10.1148/radiol.2017151930
79. Romano A, Calabria LF, Tavanti F, Minniti G, Rossi-Espagnet MC, Coppola V, et al. Apparent Diffusion Coefficient Obtained by Magnetic Resonance Imaging as a Prognostic Marker in Glioblastomas: Correlation With MGMT Promoter Methylation Status. *Eur Radiol* (2013) 23:513–20. doi: 10.1007/s00330-012-2601-4

80. Korfiatis P, Kline TL, Coufalova L, Lachance DH, Parney IF, Carter RE, et al. MRI Texture Features as Biomarkers to Predict MGMT Methylation Status in Glioblastomas. *Med Phys* (2016) 43:2835–44. doi: 10.1118/1.4948668
81. Sasaki T, Kinoshita M, Fujita K, Fukai J, Hayashi N, Uematsu Y, et al. Radiomics and MGMT Promoter Methylation for Prognostication of Newly Diagnosed Glioblastoma. *Sci Rep* (2019) 9:1–9. doi: 10.1038/s41598-019-50849-y
82. Romano A, Pasquini L, di Napoli A, Tavanti F, Boellis A, Rossi Espagnet MC, et al. Prediction of Survival in Patients Affected by Glioblastoma: Histogram Analysis of Perfusion MRI. *J Neuro Oncol* (2018) 139:455–60. doi: 10.1007/s11060-018-2887-4
83. Kickingereder P, Sahm F, Radbruch A, Wick W, Heiland S, von Deimling A, et al. IDH Mutation Status Is Associated With a Distinct Hypoxia/Angiogenesis Transcriptome Signature Which Is Non-Invasively Predictable With rCBV Imaging in Human Glioma. *Sci Rep* (2015) 5:1–9. doi: 10.1038/srep16238
84. Yalaza C, Ak H, Cagli MS, Ozgiray E, Atay S, Aydin HH. R132H Mutation in IDH1 Gene Is Associated With Increased Tumor HIF1-Alpha and Serum VEGF Levels in Primary Glioblastoma Multiforme. *Ann Clin Lab Sci* (2017) 47:362–4.
85. Hsieh KLC, Chen CY, Lo CM. Radiomic Model for Predicting Mutations in the Isocitrate Dehydrogenase Gene in Glioblastomas. *Oncotarget* (2017) 8:45888–97. doi: 10.18632/oncotarget.17585
86. Wong E, Nahar N, Hau E, Varikatt W, Gebski V, Ng T, et al. Cut-Point for Ki-67 Proliferation Index as a Prognostic Marker for Glioblastoma. *Asia Pacific J Clin Oncol* (2019) 15:5–9. doi: 10.1111/ajco.12826
87. Saadeh FS, Mahfouz R, Assi HI. Egfr as a Clinical Marker in Glioblastomas and Other Gliomas. *Int J Biol Markers* (2018) 33:22–32. doi: 10.5301/ijbm.5000301
88. Hu LS, Ning S, Eschbacher JM, Baxter LC, Gaw N, Ranjbar S, et al. Radiogenomics to Characterize Regional Genetic Heterogeneity in Glioblastoma. *Neuro-Oncology* (2017) 19:128–37. doi: 10.1093/neuonc/now135
89. Liu Y, Xu X, Yin L, Zhang X, Li L, Lu H. Relationship Between Glioblastoma Heterogeneity and Survival Time: An MR Imaging Texture Analysis. *Am J Neuroradiol* (2017) 38:1695–701. doi: 10.3174/ajnr.A5279
90. Wallace BC, Small K, Brodley CE, Trikalinos TA. Class Imbalance, Redux. In: *IEEE 11th International Conference on Data Mining*. p. 754–63.
91. Liu R, Hall LO, Bowyer KW, Goldgof DB, Gatenby R, Ahmed KB. Synthetic Minority Image Over-Sampling Technique: How to Improve AUC for Glioblastoma Patient Survival Prediction. In: *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017* (2017). p. 1357–62. doi: 10.1109/SMC.2017.8122802
92. Fernández A, García S, Herrera F, Chawla NV. SMOTE for Learning From Imbalanced Data: Progress and Challenges, Marking the 15-Year Anniversary. *J Artif Intell Res* (2018) 61:863–905. doi: 10.1613/jair.1.11192
93. Cheng K, Zhang C, Yu H, Yang X, Zou H, Gao S. Grouped SMOTE With Noise Filtering Mechanism for Classifying Imbalanced Data. *IEEE Access* (2019) 7:170668–81. doi: 10.1109/ACCESS.2019.2955086

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.

The handling editor declared a shared affiliation with several of the authors, LP, FD, MCR-E, GR, AS, AR, ADN, AB, at time of review.

frontiers in Oncology

ORIGINAL RESEARCH
published: 29 November 2021
doi: 10.3389/fonc.2021.773299

Detection and Segmentation of Pelvic Bones Metastases in MRI Images for Patients With Prostate Cancer Based on Deep Learning

Xiang Liu[1], *Chao Han*[1], *Yingpu Cui*[1], *Tingting Xie*[2], *Xiaodong Zhang*[1] *and Xiaoying Wang*[1]*

[1] Department of Radiology, Peking University First Hospital, Beijing, China, [2] Department of Radiology, Peking University Shenzhen Hospital, Shenzhen, China

OPEN ACCESS

Edited by:
Oliver Diaz,
University of Barcelona, Spain

Reviewed by:
Kavita Singh,
Massachusetts General Hospital and Harvard Medical School, United States
Wei Wei,
Xi'an Polytechnic University, China

***Correspondence:**
Xiaoying Wang
wangxiaoying@bjmu.edu.cn

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

Received: *09 September 2021*
Accepted: *08 November 2021*
Published: *29 November 2021*

Citation:
Liu X, Han C, Cui Y, Xie T, Zhang X and Wang X (2021) Detection and Segmentation of Pelvic Bones Metastases in MRI Images for Patients With Prostate Cancer Based on Deep Learning. Front. Oncol. 11:773299. doi: 10.3389/fonc.2021.773299

Objective: To establish and evaluate the 3D U-Net model for automated segmentation and detection of pelvic bone metastases in patients with prostate cancer (PCa) using diffusion-weighted imaging (DWI) and T1 weighted imaging (T1WI) images.

Methods: The model consisted of two 3D U-Net algorithms. A total of 859 patients with clinically suspected or confirmed PCa between January 2017 and December 2020 were enrolled for the first 3D U-Net development of pelvic bony structure segmentation. Then, 334 PCa patients were selected for the model development of bone metastases segmentation. Additionally, 63 patients from January to May 2021 were recruited for the external evaluation of the network. The network was developed using DWI and T1WI images as input. Dice similarity coefficient (DSC), volumetric similarity (VS), and Hausdorff distance (HD) were used to evaluate the segmentation performance. Sensitivity, specificity, and area under the curve (AUC) were used to evaluate the detection performance at the patient level; recall, precision, and F1-score were assessed at the lesion level.

Results: The pelvic bony structures segmentation on DWI and T1WI images had mean DSC and VS values above 0.85, and the HD values were <15 mm. In the testing set, the AUC of the metastases detection at the patient level were 0.85 and 0.80 on DWI and T1WI images. At the lesion level, the F1-score achieved 87.6% and 87.8% concerning metastases detection on DWI and T1WI images, respectively. In the external dataset, the AUC of the model for M-staging was 0.94 and 0.89 on DWI and T1WI images.

Conclusion: The deep learning-based 3D U-Net network yields accurate detection and segmentation of pelvic bone metastases for PCa patients on DWI and T1WI images, which lays a foundation for the whole-body skeletal metastases assessment.

Keywords: pelvic bones, metastases, prostate cancer, deep learning, magnetic resonance imaging

INTRODUCTION

The nature of bone marrow makes it a favorite fertile soil into which prostate tumors incline to colonize and grow (1, 2); up to 84% of patients with advanced prostate cancer (PCa) experience bone metastases (3), and more than 80% PCa patients developed relapse in the bone following treatment of the primary site (4). The mortality of PCa is 6.6-fold for those with bone metastases compared to those without bone metastases (5). Accurate detection and assessment of metastatic burden in bone are of fundamental importance for radiologists.

Bone scintigraphy (BS) and computed tomography (CT) scans were endorsed as the standard imaging method in the staging and follow-up of metastatic PCa (6), while it is gradually clear that the reduced accuracy of BS and CT in the detection and therapeutic response evaluation of bone metastases reduces their effectiveness in therapy management (7). Multiparametric magnetic resonance imaging (mpMRI) is emerging as a powerful alternative for metastatic PCa. One of the main strengths of mpMRI is to achieve a precise evaluation of bone metastasis *via* the incorporation of anatomic [e.g., T1 weighted imaging (T1WI)] and functional imaging sequences [e.g., diffusion-weighted imaging (DWI)] (7, 8). The value of volumetric measurements for assessing treatment response has been increasingly discussed, and the measurements of lesion volume on mpMRI should be undertaken on high-quality T1WI images according to the METastasis Reporting and Data System (MET-RADS) for PCa (9). Additionally, the volume of bone metastasis assessed with DWI was reported to show a correlation with established prognostic biomarkers and is associated with overall survival in metastatic castration-resistant PCa (10). In short, the detection and delineation of metastases and evaluation of volume change concerning disease progression or therapy on DWI and T1WI images are key tasks as part of optimal patient management.

Heavy workload of mpMRI images evaluation can be tiresome for radiologists, hence bearing the risk of missed diagnosis for lesions and leading to decreased sensitivity. The measurements of all the metastatic lesions are time consuming, in particular, if multiple metastases are present. In this context, automated and accurate segmentation of bone metastases would be highly beneficial.

Driven by the rapid growth in computer science, the performance of deep learning is on par with or even outperforms radiologists in visual identification, which can perform automated data-oriented feature extraction and thus learning directly the most relevant feature representation from the input images (11, 12). The U-Net algorithm is one of the most commonly used deep learning-based convolutional neural networks (CNNs) (13), which shows potential in detection, segmentation, and classification of metastatic lesions on MRI images such as brain metastases (14, 15) and liver metastases (16). Concerning the automated bone metastasis analysis using the deep learning technique, the research trend is mainly on BS (17, 18) and single-photon emission computerized tomography (SPECT) images (19, 20); less attention has been paid to the diagnosis of mpMRI (21, 22). To this end, we intend to apply the 3D U-Net (23) algorithm for the segmentation of bone metastases on mpMRI images. For proof-of-concept, we focused on the detection and segmentation of bone metastases in the pelvic area.

MATERIALS AND METHODS

This retrospective single-center study was approved by the institutional review board, and written informed consent was waived.

Patient Cohort

A cohort of 955 consecutive patients who had undergone pelvic mpMRI for either clinically suspected or confirmed PCa between January 2017 and December 2020 was reviewed using our institutional image archiving system. The exclusion criteria were as follows: (1) poor image quality (significant motion artifact or chemical shift artifact), (2) uncomplete MR image set, (3) obvious destruction of bone structure, and (4) patients with a history of pelvic fracture or surgery. Finally, the images from 859 patients were included for the 3D U-Net model development of pelvic bony structures segmentation, including a dataset of patients with PI-RADS score of 1–2 or biopsy-proven benign prostate hyperplasia (dataset 1, n = 349), a dataset of biopsy-proven PCa patients without bone metastases (dataset 2, n = 280), and a dataset of biopsy-proven PCa patients with bone metastases (dataset 3, n = 230).

All three datasets were used to develop a pelvic bony structure segmentation model. Then, a 3D U-Net model for bone metastases segmentation was developed using datasets 2 and 3. The patients with primary malignant bone tumors (such as osteosarcoma and myeloma) or definite benign findings (hemangiomas, bone island) on pelvic bones (n = 27) and patients who underwent PCa treatment (endocrine therapy, chemotherapy, or radiotherapy, n = 149) were excluded. In total, 334 patients were enrolled for the model development, including 168 PCa patients with bone metastases and 166 PCa patients without bone metastases.

Additionally, 77 patients with biopsy-proven PCa who performed pelvic mpMRI scanning from January 2021 and May 2021 were acquired; according to the above excluding criteria, 63 patients were finally recruited for the external evaluation of the 3D U-Net model including a dataset of 31 PCa patients with bone metastases (dataset 4) and a dataset of 32 PCa patients without bone metastases (dataset 5). The workflow of data enrollment is shown in **Figure 1**.

Image Acquisition

The pelvic mpMRI acquisitions were performed on three 3.0 T MR units (Achieva, Philips Healthcare; Discovery, GE Healthcare; Interia, Philips Healthcare). The standard pelvic mpMRI protocol at our institution included a T1/T2-weighted sequence, DWI with b-values of 0, 800, or 1,000 s/mm^2 along with reconstructed ADC images, T1W images obtained using the 2-point Dixon technique with in-phase (T1WI-IP) and out-phase (T1WI-OP), and dynamic contrast-enhanced imaging. DWI images with high b-values (b = 800 or 1,000 s/mm^2) and T1WI-IP images were selected for PCa bone metastases analyses in this study. Detailed MR imaging parameters of DWI and T1WI-IP sequence are shown in **Table 1**.

Manual Annotation

The manual annotations were performed with an image segmentation software (ITK-SNAP 3.6; Penn Image Computing and Science Laboratory, Philadelphia, PA). Under the supervision of a board-certified radiology expert (with more than 15 years of

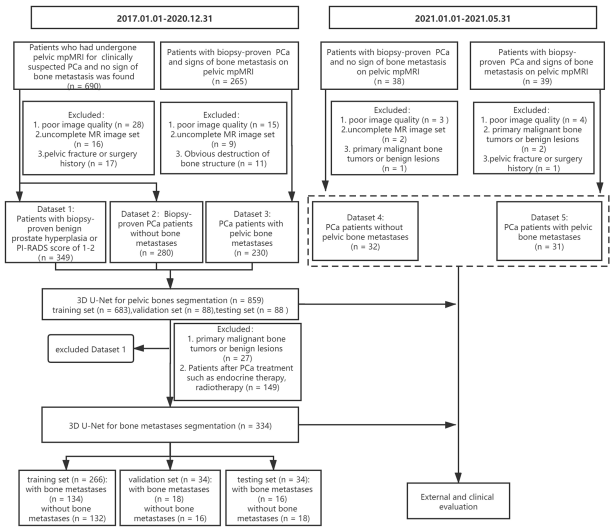


FIGURE 1 | The workflow of data enrollment.

reading experience), a radiology resident with 3 years of reading experience evaluated all mpMRI examinations and, section by section, manually annotated eight pelvic bony structures (lumbar vertebra, sacrococcyx, ilium, acetabulum, femoral head, femoral neck, ischium, and pubis) on DWI images and T1WI-IP images. The bone metastases were included in the annotations, which were recognized as bone tissue in this bony structure segmentation model. The manual annotations of the pelvic bony structures were regarded as the reference standard for the 3D U-Net model evaluation.

To establish the reference standard of bone metastases, the radiology resident and expert radiologist conducted a review of the original radiology report and double reviewed the included MR imaging scans and prior/follow-up imaging before annotation. A bone lesion was considered as a metastasis if it showed an MR imaging correlated with adequate image contrast (positive image contrast on DWI images and negative image contrast on those obtained with the T1WI-IP images). The radiology resident performed manual annotations of the metastatic lesions on DWI and T1WI-IP images in a voxel-wise manner (indicated as A1.1). Then, the expert radiologist modified the annotations of A1.1 and the annotations after modification were indicated as A2.1. Both the resident and expert radiologist repeated the annotations and modifications at least 3 weeks later (indicated as A1.2 and A2.2, respectively). The inter- and intraobserver agreement between the manual annotations (A1.1 vs. A2.1; A1.1 vs. A1.2; A2.1 vs. A2.2; and A1.2 vs. A2.2) were estimated using Dice similarity coefficient (DSC).

The bony metastatic lesions in the 31 PCa patients of the external dataset were manually annotated by the resident radiology under the supervision of the expert radiologist, which was taken as the reference standard for external evaluation of the model.

Model Development

A two-step method for the bone metastases segmentation was proposed using the 3D U-Net model: the first step with a 3D U-Net algorithm for pelvic bone segmentation followed by a second step with a 3D U-Net for bone lesion segmentation within the segmented pelvic bony structures. Both the CNNs were coded by Python3.6, Pytorch 0.4.1, Opencv, Numpy, and SimpleITK, and trained on the GPU NVIDIA Tesla P100 16G.

Model Development for Pelvic Bones Segmentation

The model of the pelvic bony structure segmentation takes the combination of DWI images and T1WI-IP images as input, and each image sequence is used as an independent input data (**Figure 2**). The 859 patients were randomly divided into either training (n = 683), validation (n = 88), or testing (n = 88) sets with a ratio of 8:1:1. During the image preprocessing, the pixel values in images were scaled between 0 and 65,535. Then, the images were resized to 64 × 224 × 224 (z, y, x) by resampling to maintain the optimal image features, and z-score intensity normalization was applied to all images. Skewing (angel: 0–5), shearing (angel: 0–5), and translation (scale: −0.1,0.1) of the images were applied for data

TABLE 1 | MR imaging parameters of DWI and T1WI-IP sequence.

Sequences		3.0 T Discovery	3.0 T Intera	3.0 T Achieva
DWI	b-value (s/mm^2)	800	1000	800
	Imaging matrix	256 × 256	240 × 240	156 × 180
	Echo time (ms)	60	78	54
	Repetition time (ms)	4,000	4,959	3,300
	Field of view (mm^2)	450 × 366	480 × 360	512 × 356
	Section thickness (mm)	8	7	7
	Number of slices	25	28	24
T1WI-IP	Imaging matrix	288 × 192	320 × 200	280 × 180
	Echo time (ms)	2.0	2.4	2.4
	Repetition time (ms)	3.9	7.5	6.7
	Field of view	450 × 360	450 × 350	400 × 400
	Section thickness	4	5	2
	Number of slices	112	112	120
	Bandwidth	166.67	300	450
	Flip angle(°)	13	10	10

T1WI-IP, T1W images obtained using the Dixon technique with in-phase.

augmentation. To remove small spurious segmentation, the two largest connected components of each bone were selected as the final segmentation. A total of 300 epochs of training were performed until validation loss failed to rise. The Adam optimizer was employed to minimize loss with a learning rate of 0.0001, a batch size of 2, and a Dice loss function. Other hyperparameters (such as weight initialization and dropout for regularization) were randomly searched and automatically executed in the validation set during model development.

Model Development for Bone Metastases Segmentation

The volume of interest predicted by the model of pelvic bony structure segmentation was used as the mask for the bone metastases segmentation (**Figure 2**). The network configurations were set as follows: training epoch, 250; learning rate, 0.01; batch size, 5; optimizer, Adam optimizer; and loss function, Dice loss.

For post-processing, automatically detected metastases of <0.2 cm^3 during inference of testing set were regarded as image noise and discarded. The threshold was based on the resolution of T1WI-IP sequences and is determined by referring to the smallest annotated metastases (0.356 cm^3).

Model Evaluation

Model Evaluation for Pelvic Bony Structure Segmentation

The performance of the network was evaluated by comparing the segmentations generated by the 3D U-Net based on image data from the testing set to the corresponding reference standard represented by the manual segmentations on DWI and T1WI-IP images quantitatively. The evaluation metrics used for the bony structures segmentation include the overlap-based metric (DSC), the volume-based metric [volumetric similarity (VS)], and the spatial distance-based metric [Hausdorff distance (HD)] (24).

Model Evaluation for Bone Metastases Segmentation

The performance of the bone metastases segmentation model was evaluated both on detection and segmentation. Detection is defined as the network's ability to detect a metastasis annotated by the radiologist. One bone metastasis was considered detected when the manual annotation and the predicted segmentation had an overlap >0. Segmentation is defined as its ability to provide a contour identical to that of the radiologist.

The detection performance of the network was quantified at the patient and lesion levels. The sensitivity, specificity, accuracy, positive predictive value (PPV), negative predictive value (NPV), and area under the receiver operating characteristic curve (AUC) were used to assess the performance of the model to discriminate between patients with bone metastases and patients without bone metastases. To determine the detection accuracy of the metastases at the lesion level, we compared the lesions obtained with model predictions and manual annotations to determine the true-positive (TP), false-negative (FN), and false-positive (FP) findings. The recall (correctly detected metastases divided by all metastases contained in reference standard), precision (correctly detected metastases divided by all the detected metastases), and F1 score (harmonic mean of precision and recall) were calculated to assess the detection performance of the model on a lesion-by-lesion basis. In addition, we determined the number of distinct metastatic lesions in each case in the testing set and then divided the data into groups with (a) 1, (b) 2–3, (c) 4–5, and (d) >5 lesions to facilitate subgroup analysis of metastases detection at lesion level.

The metastases segmentation performance of the network was assessed using the metrics of DSC, VS, and HD by comparing the CNN-predicted segmentation and manual segmentation. Besides, the volume of the bone metastases in manual annotations and automated segmentations was calculated to further quantitatively estimate the segmentation efficacy of the U-Net algorithm.

Model Evaluation on an External Dataset

The external dataset was used to further assess the efficiency of the model on bone metastases evaluation in the clinical setting. Given the new mpMRI data of PCa patients, the 3D U-Net was supposed to determine the existence of bone metastases (M0 or

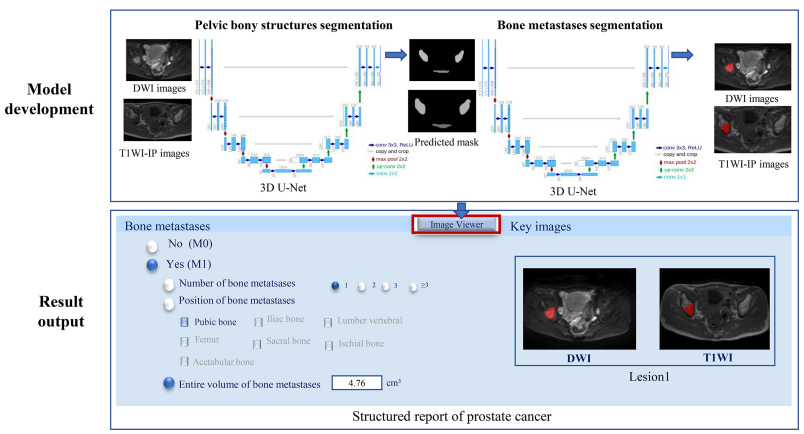


FIGURE 2 | The two-step 3D U-Net for bone metastases segmentation on DWI and T1WI-IP images. T1WI-IP, T1W images obtained using the Dixon technique with in-phase.

M1) and output the number, location, and volume of the bone metastases with corresponding segmented masks (**Figure 2**). A bone lesion was considered as being detected if it was segmented on at least one of the two MR imaging sequences (DWI/T1WI-IP). The accuracy of the M-staging of the model was assessed using the receiver operating characteristic curve analysis, and the segmentation performance (DSC, VS, HD) and quantitative measurements (volume) were assessed by comparison with manual annotations.

Statistical Analysis

MedCalc (version 14.8; MedCalc Software, Ostend, Belgium) and SPSS (version 22.0, IBM Corp., Armonk, NY, USA) were used for the statistical analyses. Numerical data of patients' age were reported as the mean ± SD (standard deviation), and prostate-specific antigen (PSA) levels were reported as (median, quartile). One-way analysis of variance (ANOVA) was used to compare the characteristics of patients (age, PSA level) among training, validation, and testing sets. The segmentation performance of the algorithm (DSC, VS, and HD) between DWI and T1WI-IP images were compared by paired *t*-test. The McNemar's test was used to compare the detection performance (sensitivity, specificity, PPV, NPV, recall, and precision) between the two sequences. Bland–Altman analyses were performed to compare manual versus automated bone metastases volume. $p < 0.05$ was considered indicative of a statistically significant difference.

RESULTS

Characteristics of Patients

The characteristics of patients are shown in **Tables 2**, **3**. The age and PSA level showed no significant difference among the training, validation, and testing sets on both models (all with $p > 0.05$). The average volume of metastases in the external dataset was 7.39 cm^3, and no difference was found between the external dataset and model development dataset ($p = 0.645$). Of the 16 PCa patients with bone metastases in the testing set, 2 patients (12.50%) had one metastasis, 5 patients (31.25%) had two to three metastases, 4 patients (25.00%) had four to five metastases, and 5 patients (31.25%) had more than five metastases.

Assessment of Pelvic Bony Structures Segmentation

As shown in **Table 4** and **Figure 3**, in the testing set of pelvic bone segmentation model, the DSC and VS values of eight pelvic bony structures between model prediction and manual annotation are all above 0.85 on both DWI and T1WI-IP images, while the mean DSC and VS values on T1WI-IP images are significantly higher than those on DWI images (all with $p < 0.05$), and the HD is significantly lower. This may be explained by the higher spatial resolution of the T1WI-IP images. Additionally, as detailed in the Supplementary materials (**Supplementary Tables S1–S4**), no significant difference was found among the patients from different datasets (dataset 1 vs. dataset 2 vs. dataset 3) and different scanners (3.0 T Discovery vs. 3.0 T Achieva vs. 3.0 T Intera) on both DWI and T1WI-IP images.

The Inter- and Intraobserver Agreement of Bone Metastases Annotations

The interobserver agreement of the manual annotations of bone metastases was assessed by calculating the DSC values between A1.1 and A2.1, and A1.2 and A2.2. The intraobserver agreement was assessed by A1.1 vs. A1.2 and A2.1 vs. A2.2. The DSC values on DWI images were as follows: A1.1 vs. A2.1, 0.90 ± 0.08; A1.1 vs. A2.1, 0.91 ± 0.09; A2.1 vs. A2.2, 0.94 ± 0.05; A1.2 vs. A2.2, 0.91 ± 0.08. In T1WI-IP images, the DCS values were as follows: A1.1 vs. A2.1, 0.89 ± 0.09; A1.1 vs. A2.1, 0.90 ± 0.09; A2.1 vs. A2.2, 0.97 ± 0.04; and A1.2 vs. A2.2, 0.92 ± 0.08. The high DSC values between A2.1 vs. A2.2 confirmed the reliability of the manual annotations. A2.2 was regarded as the reference standard for the lesion segmentation model evaluation.

TABLE 2 | Characteristics of patients for the pelvic bony structure segmentation model.

Characteristics	Model development (dataset 1 + dataset 2 + dataset 3)				External dataset	*p*-value
	Training set	Validation set	Testing set	*p*-value		
Age (mean ± SD)	68.3 ± 10.5	67.6 ± 10.9	67.8 ± 11.7	0.756	70.7 ± 8.1	0.062
No. of patients	683	88	88	–	68	–
No. of patients with bone metastases	184	23	23	–	34	–
No. of patients without bone metastases	224	16	18	–	34	–
PSA (median, quartile, ng/ml)						
T-PSA	10.49 (7.11, 15.99)	9.73 (8.46, 12.68)	11.19 (7.44, 15.75)	0.556	12.25 (8.89, 26.92)	0.199
F-PSA	1.95 (1.04, 6.71)	2.07 (1.08, 5.78)	2.19 (1.02, 5.46)	0.266	1.65 (1.05, 5.26)	0.112
F/T-PSA	0.12 (0.09, 0.17)	0.10 (0.07, 0.20)	0.10 (0.08, 0.18)	0.587	0.12 (0.09, 0.18)	0.399
Scanners						
3.0 T Discovery	417	56	57	–	31	–
3.0T Achieva	133	17	15	–	13	–
3.0 T Intera	134	15	16	–	24	–

PSA, prostate-specific antigen; T-PSA, total PSA; F-PSA, free PSA; SD, standard deviation.

TABLE 3 | Characteristics of patients for the bone metastases model.

Characteristics	Model development (from dataset 2 and dataset 3)				External dataset	*p*-value
	Training set	Validation set	Testing set	*P* value		
Age (mean ± SD)	69.6 ± 10.4	65.9 ± 11.2	68.7 ± 8.9	0.548	70.7 ± 8.1	0.268
No. of patients	266	34	34	–	63	–
No. of patients with bone metastases	134	18	16	–	31	–
No. of patients without bone metastases	132	16	18	–	32	–
PSA (median, quartile, ng/ml)						
T-PSA	13.04 (9.10, 20.1)	12.65 (10.13,18.50)	13.95 (12.95, 23.5)	0.305	12.25 (8.89, 26.92)	0.941
F-PSA	1.29 (1.01,5.41)	1.36 (1.08, 4.38)	1.48 (1.07, 4.73)	0.993	1.65 (1.05, 5.26)	0.091
F/T-PSA	0.07 (0.09, 0.18)	0.09 (0.07, 0.16)	0.09 (0.04, 0.11)	0.356	0.12 (0.09, 0.18)	0.573
Average volume of metastases (median, quartile, cm^3)	7.50 (5.47, 31.60)	7.98 (2.72, 31.75)	8.05 (2.93, 31.03)	0.945	7.39 (1.23, 28.23)	0.645
No. of metastatic lesions						
1	30 (22.39%)	4 (22.22%)	2 (12.50%)	–	5 (16.13%)	–
2-3	36 (26.86%)	6 (33.33%)	5 (31.25%)	–	6 (19.35%)	–
4-5	24 (17.91%)	4 (22.22%)	4 (25.00%)	–	8 (25.81%)	–
>5	44 (32.84%)	4 (22.22%)	5 (31.25%)	–	12 (38.71%)	–
Total lesions	664	86	89	–	144	–
Scanners						
3.0 T Discovery	172	17	20	–	31	–
3.0T Achieva	71	10	8	–	13	–
3.0 T Intera	23	7	6	–	24	–

PSA, prostate-specific antigen; T-PSA, total PSA; F-PSA, free PSA; SD, standard deviation.

TABLE 4 | Segmentation performance of pelvic bony structures.

Bony structures	DSC		*p*-value	VS		*p*-value	HD (mm)		*p*-value
	DWI	T1WI-IP		DWI	T1WI-IP		DWI	T1WI-IP	
Lumbar vertebra	0.89 ± 0.05	0.93 ± 0.03	0.001	0.94 ± 0.06	0.96 ± 0.06	0.034	11.45 ± 3.54	10.63 ± 4.66	0.258
Sacrococcyx	0.88 ± 0.04	0.93 ± 0.02	0.001	0.96 ± 0.03	0.98 ± 0.02	0.001	13.36 ± 4.79	9.56 ± 4.53	0.001
Ilium	0.88 ± 0.03	0.94 ± 0.02	0.001	0.97 ± 0.02	0.99 ± 0.02	0.001	13.34 ± 4.15	8.50 ± 3.30	0.001
Acetabulum	0.85 ± 0.04	0.90 ± 0.03	0.001	0.94 ± 0.05	0.96 ± 0.04	0.017	14.95 ± 6.04	10.17 ± 5.60	0.001
Femoral head	0.90 ± 0.04	0.94 ± 0.03	0.001	0.95 ± 0.04	0.97 ± 0.02	0.001	9.00 ± 2.90	4.77 ± 1.51	0.001
Femoral neck	0.88 ± 0.04	0.95 ± 0.03	0.001	0.96 ± 0.04	0.98 ± 0.05	0.015	12.39 ± 4.40	8.50 ± 5.51	0.001
Ischium	0.86 ± 0.04	0.90 ± 0.03	0.001	0.93 ± 0.05	0.96 ± 0.04	0.001	14.88 ± 6.92	14.62 ± 6.27	0.295
Pubis	0.86 ± 0.05	0.88 ± 0.04	0.022	0.92 ± 0.06	0.94 ± 0.05	0.074	14.72 ± 7.08	10.60 ± 4.58	0.001

DSC, Dice similarity coefficient; HD, Hausdorff distance; T1WI-IP, T1W images obtained using the Dixon technique with in-phase; VS, volumetric similarity.

The Detection Accuracy of Bone Metastases

The detection performance of the CNN on DWI and T1WI-IP images at the patient and lesion levels are shown in **Table 5**. The detection performance of the model on DWI images was better than on T1WI-IP images concerning the values of the evaluation metrics, while no significant difference was found between the two sequences (all with $p > 0.05$). The results of the subgroup analysis of detection accuracy at lesion level in the testing set showed the highest recall and precision values in patients with single metastases, and both the recall and precision were above 80% for few metastases (≤5 metastases) and multiple metastases (>5 metastases).

The Segmentation Accuracy of Bone Metastases

The mean DSC, VS, and HD for the automatic metastases segmentation are 0.79 ± 0.05, 0.84 ± 0.09, and 15.05 ± 3.61 mm on DWI images and 0.80 ± 0.06, 0.85 ± 0.08, and 13.39 ± 3.20 mm on T1WI images (**Figure 4A**), which showed no significant difference between the two sequences (p = 0.627, 0.741, and 0.175, respectively).

The volume differences between manual annotation and model prediction of bone metastases on DWI and T1WI-IP images are shown in **Figures 4B, C**. The limit of agreement (LOA) between the automated and manual segmentation on DWI images was −8.4–6.6 cm^3 and −4.4–4.4 cm^3 on T1WI-IP images. Most of the difference values were within the LOA, which showed that the volume of overall metastatic lesions in each patient between manual and automated segmentations agreed closely. Example results of the automatic bone metastases segmentation are shown in **Figure 5**.

Detection and Segmentation Accuracy on the External Dataset

The sensitivity, specificity, and AUC values of the model in determining the M-staging (M0 or M1) were 93.6% (29/31; 95% CI, 78.6%–99.2%), 93.8% (30/32; 95%CI, 79.6%–99.2%), and 0.94

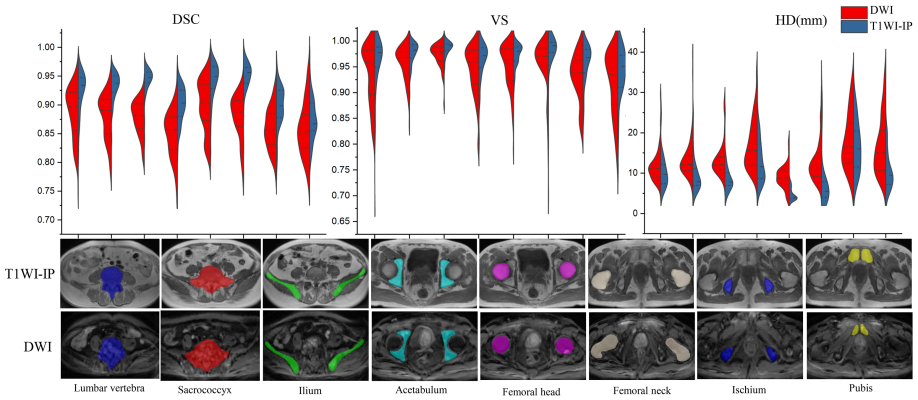


FIGURE 3 | Split violin plots of DSC, VS, and HD (mm) for pelvic bony structures segmentation. DSC, Dice similarity coefficient; HD, Hausdorff distance; T1WI-IP, T1W images obtained using the Dixon technique with in-phase; VS, volumetric similarity.

TABLE 5 | Detection accuracy of bone metastases at patient and lesion levels.

Level	Metrics	DWI	T1WI-IP	*p*-value
Patient-level	Sensitivity (%)	87.5 (61.7–98.4)	81.3 (54.4–96.0)	0.847
	Specificity (%)	83.3 (58.6–96.4)	77.8 (52.4–93.6)	0.852
	Accuracy (%)	85.3 (68.9–95.1)	79.4 (62.1–89.9)	0.789
	PPV (%)	82.4 (56.6–96.2)	76.5 (50.1–93.2)	0.847
	NPV (%)	88.2 (63.6–98.6)	82.4 (56.6–96.2)	0.852
	AUC	0.85 (0.69–0.95)	0.80 (0.62–0.91)	0.442
Lesion-level	Recall (%)	91.01 (81/89)	88.76 (79/89)	0.874
	Precision (%)	84.38 (81/96)	86.81 (79/91)	0.857
	F1-score (%)	87.6	87.8	–
Subgroup analysis				
1	Recall (%)	100 (2/2)	100 (2/2)	
	Precision (%)	100 (2/2)	100 (2/2)	
2–3	Recall (%)	92.9 (13/14)	85.7 (12/14)	
	Precision (%)	86.7 (13/15)	85.7 (12/14)	
4–5	Recall (%)	94.7 (18/19)	84.2 (16/19)	
	Precision (%)	85.7 (18/21)	88.9 (16/18)	
>5	Recall (%)	88.9 (48/54)	90.7 (49/54)	
	Precision (%)	82.8 (48/58)	85.9 (49/57)	

AUC, area under the receiver operating characteristic curve; NPV, negative predictive value; PPV, positive predictive value; T1WI-IP, T1W images obtained using the Dixon technique with in-phase.

(95%CI, 0.85–0.98) on DWI images and 87.1% (27/31; 95%CI, 70.2%–96.4%), 90.6% (29/32; 95%CI, 75.0%–98.0%), and 0.89 (95% CI, 0.85–0.98) on T1WI-IP images. The AUC values between the two sequences showed no significant difference ($p = 0.368$).

At lesion level, the segmentation accuracy of the model for bone metastases achieved average DSC, VS, and HD values of 0.79 ± 0.06, 0.83 ± 0.08, and 16.03 ± 9.74 mm on DWI images, 0.81 ± 0.06, 0.82 ± 0.07, and 17.20 ± 6.73 mm on T1WI-IP images (**Figure 6A**). The mean volumes of manual annotation and model prediction were 15.35 and 14.10 cm^3 on DWI images and 15.68 and 14.40 cm^3 on T1WI-IP images. The volume difference is shown in **Figures 6B, C**.

DISCUSSION

In this work, we developed a two-step deep learning-based 3D CNN for automated detection and segmentation of bone metastases in PCa patients using whole 3D MR images (DWI and T1WI-IP images), in which the first 3D U-Net focuses on the segmentation of pelvic bony structures and the second one on bone metastases segmentation. On heterogeneous scanner data, the first CNN performed excellent segmentation of pelvic bony structures on both DWI and T1WI-IP images (all with DSC > 0.85), which provides a reliable foundation for the subsequent bone metastases segmentation. Furthermore, our result showed that the proposed CNN provided an AUC of 0.854 and 0.795 on DWI and T1WI-IP images for bone metastases detection at the patient level, and high overlap between automated and manual metastases segmentations was observed (DSC = 0.79 and 0.80 on DWI and T1WI-IP images, respectively). Additionally, by testing on an external dataset, this work demonstrates the CNN's potential ability of M-staging in clinical practice (with AUC of 0.936 and 0.889 on DWI and T1WI-IP images).

mpMRI has been identified as an essential and crucial imaging modality in PCa diagnosis and metastases evaluation (25, 26). The importance of DWI and T1WI in the detection and

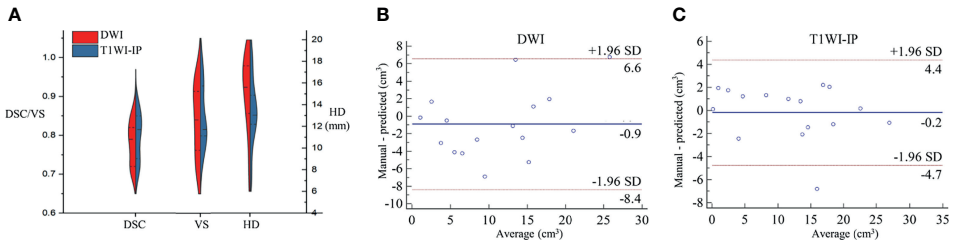


FIGURE 4 | The segmentation accuracy of bone metastases in the testing set. **(A)** Split violin plot of DSC, VS, and HD of the bone metastases on DWI and T1WI-IP images. **(B)** The Bland–Altman plot of the volume difference between manual annotation and model prediction on DWI images. **(C)** The Bland–Altman plot of the volume difference between manual annotation and model prediction on T1WI-IP images. DSC, Dice similarity coefficient; HD, Hausdorff distance; T1WI-IP, T1W images obtained using the Dixon technique with in-phase; VS, volumetric similarity.

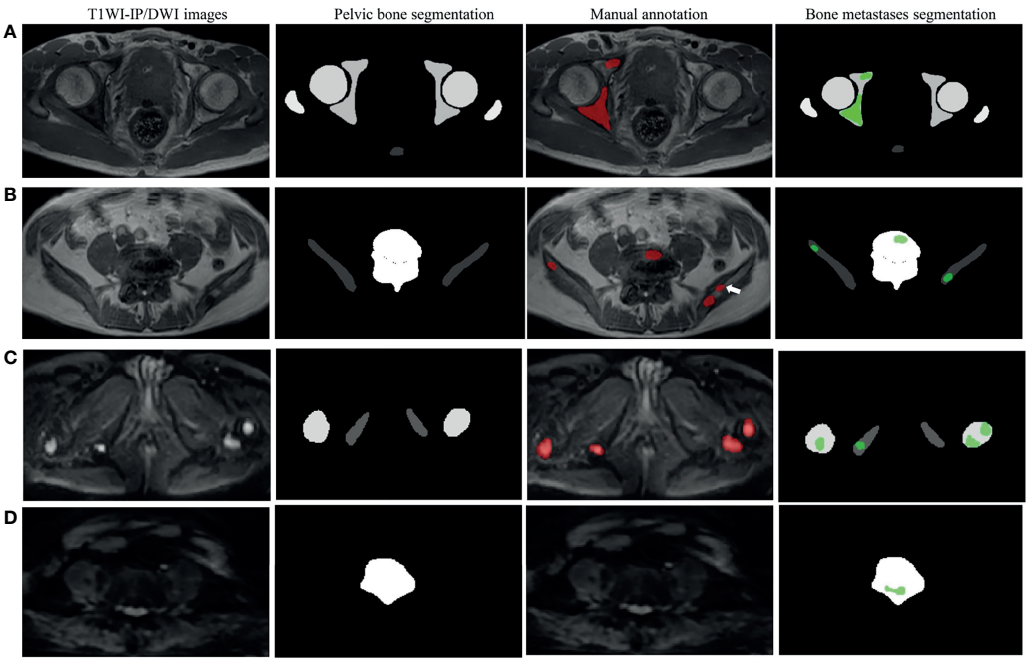


FIGURE 5 | Examples of pelvic bony structure and bone metastases segmentations. **(A)** Two metastases of acetabulum annotated by radiologists were corrected segmented by the model on T1WI-IP images (true positive). **(B)** Four of five metastases annotated by the radiologists were corrected segmented by model on T1WI-IP images; one metastasis on the right ilium was missed (the white arrow pointed, false negative). **(C)** All the four metastases of femoral head and ischium annotated by radiologists were correctly segmented by the model on DWI images (true positive). **(D)** One metastasis of lumbar vertebra was segmented by the model by error, which was not annotated by the radiologists (false positive). T1WI-IP, T1W images obtained using the Dixon technique with in-phase.

quantification of osseous metastasis in patients with PCa has been widely recognized (9, 27). In this study, to avoid the limitation of the application of the CNN if one of these sequences is unavailable, we trained the two-step 3D U-Net CNN using DWI and T1WI-IP images as independent input data. The enrolled participants performed the mpMRI examinations on one of the three different 3.0-T MR scanners with different protocols, and the b-values of the DWI images were different (b = 0, 800 or 0, 1,000 s/mm^2). In a previous publication (28), we proposed a deep learning-based approach for the segmentation of normal pelvic bony structures. It was the proof-of-concept study for the possibility to detect skeletal metastases located on the pelvic bones. In this study, we used two 3D U-Nets in cascade. The first model was trained to segment the pelvic bony structures. Taking the areas predicted by the first model as the mask, the second model was trained to segment the metastatic lesions on the pelvic bones. The combination of the two 3D U-Nets offers the potential for efficient bone metastases location and quantification. It is important to note that the two-step deep learning model has been widely used to improve the accuracy and stability of the system, such as lymph node detection (29) and PCa segmentation (30).

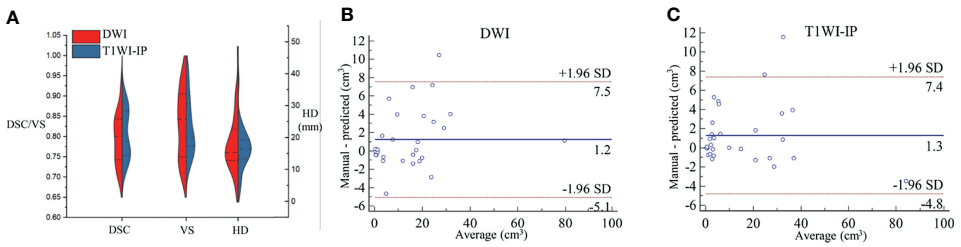


FIGURE 6 | The segmentation accuracy of bone metastases on an external dataset. **(A)** Split violin plot of DSC, VS, and HD of the bone metastases on DWI and T1WI-IP images. **(B)** The Bland–Altman plot of the volume difference between manual annotation and model prediction on DWI images. **(C)** The Bland–Altman plot of the volume difference between manual annotation and model prediction on T1WI-IP images. DSC, Dice similarity coefficient; HD, Hausdorff distance; T1WI-IP, T1W images obtained using the Dixon technique with in-phase; VS, volumetric similarity.

The high number of FP lesions poses a common drawback in automated detection of metastatic lesions, which has been reported to be approximately seven to eight per scan for brain metastases (31, 32). In the present study, by providing high-quality pelvic bone segmentation masks on DWI and T1WI-IP images, the FP interference from other tissues within the pelvic region (such as metastatic lymph nodes, colon, bladder, etc.) can be effectively eliminated. Moreover, a simple post-processing step was added to avoid FP findings by rejecting all structures with a volume <0.2 cm^3, which was smaller than the smallest annotated metastases.

Our CNN not only detects almost all metastases but also incorrectly marks other objects as metastases. Most of these FPs were caused by objects that showed a similar radiological appearance to metastatic lesions on DWI and T1WI-IP images. As shown in **Figure 5D**, the high-intensity spinal cord on DWI images within the mask of the lumbar vertebra was detected as metastases by mistake. In addition, the objects that were not or scarcely represented in the training set and thus had an appearance unknown to the network could result in FP as well. These unknown appearances could be other lesions or conditions such as incidental cysts. An inspection of the 15 FP findings on DWI images showed that nine of the FPs were the spinal cord and nerve root structure, and six of the FPs were benign lesions: four cysts and two hemangiomas. The 12 FP objects on T1WI-IP images included eight spinal cord and nerve root structures, three cysts, and one blood vessel structure.

The FN metastases missed by the CNN networks were the small ones, as can be seen in **Figure 5A**, which might be due to the few occupied voxels compared with large metastases. Additionally, on a subgroup analysis, our results suggest that the networks perform well on patients with few metastases (≤5 metastases) and multiple metastases (>5 metastases) in terms of recall and precision, which boosts the clinical utility of the CNN.

Automated segmentation can help radiologists in dealing with an increased number of image interpretations while maintaining high diagnostic accuracy and, simultaneously, may also assist in evaluating treatment response during oncological follow-up. Volumetric assessment proves to be a promising tool for quantification of tumor burden and treatment response evaluation, which is superior to user-dependent conventional linear measurements because metastatic lesions are irregular (33). Compared with manual segmentation, our proposed CNN achieved a high volumetric correlation on both the testing set and the external dataset, which is crucial to help treatment decision-making and potentially improve patient care.

TNM is considered to be one of the most pivotal factors in evaluating the prognosis of PCa, and the existence of bone metastases is a decisive index for the M-staging (34). Concerning M-staging, on the external dataset, our model achieved an AUC of 0.936 (95%CI, 0.845–0.982) on DWI images and 0.889 (95%CI, 0.845–0.982) on T1WI-IP images, which demonstrated that the two-step 3D U-Net algorithm could be used in a clinical context. Besides, the output of the automated segmentation result to the structure report essentially combines visualization, quantification, and segmentation into one step, producing results that can be directly displayed to the radiologists.

U-Net has been proven to possess the potential for bone metastases segmentation. Lin et al. (19) built two deep learning networks based on U-Net and Mask R-CNN to segment hotspots in bone SPECT images for automatic assessment of metastasis. Their results showed that the U-Net-based model achieved better segmentation performance with a precision and recall value of 0.76 and 0.67 than the Mask R-CNN model (precision, 0.72; recall, 0.65). In addition, Chang et al. (35) demonstrated the capability of U-Net in segmenting spinal sclerotic bone metastases on CT images with a Dice score of 0.83. In this study, we explored the feasibility of the 3D U-Net network for pelvic bone metastases segmentation on DWI and T1WI-IP images, and our results further confirmed the segmentation accuracy of the U-Net for bone metastases. However, the comparisons among a couple of other architectures may be helpful to choose an optimal model for metastases segmentation and detection. In the future, we should further explore the performance of other models.

While this study shows high accuracy and performance using CNNs for bone metastases segmentation, several potential study limitations exist. First, the study has a typical drawback of retrospective setting. Testing of the network performance on prospective multicenter data remains a key step towards understanding its clinical value. Second, the relatively small number of patients needs to be noted. Only patients with PCa

were included here, which potentially limits the transferability of our CNN to a broad range of bone metastases of other primary tumors (rectal cancer, bladder cancer, etc.). In this context, future studies are needed to evaluate the feasibility of the CNN for bone metastases segmentation of other tumors. Third, in clinical practice, the detection of the lesion by the radiologist is usually done by simultaneous review of anatomical and functional MR images. Besides the Dixon T1WI-IP and DWI images, the Fat or Water images from the Dixon sequence and the short time inversion recovery sequence may also be helpful for the bone metastases evaluation (36, 37). Last, the choice of pelvic examinations as the anatomic target to detect bone metastases and assess the positive–negative status of the patients in terms of metastases is insufficient in clinical practice. The axial and probably whole skeleton, at least from skull to thighs, is necessary, as metastases affect the red marrow-containing areas. Future research is needed to allow for the whole-body bone metastases assessment.

CONCLUSION

In summary, our study shows that the deep learning-based 3D U-Net network can automatically detect and segment bone metastases on DWI and T1WI-IP images with high accuracy and thus illustrates the potential use of this technique in a clinically relevant setting.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the datasets are privately owned by Peking University First Hospital and are not made public. Requests to access the datasets should be directed to wangxiaoying@bjmu.edu.cn.

ETHICS STATEMENT

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

AUTHOR CONTRIBUTIONS

XL and XW contributed to the study concept and design. TX and CH contributed to acquisition of data. XL and XW annotated the images data. YC and XZ designed the model and implemented the main algorithm. XL and CH contributed to drafting of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study was supported by Capital's Funds for Health Improvement and Research (2020-2-40710).

ACKNOWLEDGMENTS

The authors gratefully acknowledge the technical support of Yaofeng Zhang, Xiangpeng Wang, and Jiahao Huang from Beijing Smart Tree Medical Technology Co. Ltd.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2021.773299/full#supplementary-material

REFERENCES

1. Buenrostro D, Mulcrone PL, Owens P, Sterling JA. The Bone Microenvironment: A Fertile Soil for Tumor Growth. *Curr Osteoporos Rep* (2016) 14:151–8. doi: 10.1007/s11914-016-0315-2
2. Park SH, Eber MR, Shiozawa Y. Models of Prostate Cancer Bone Metastasis. *Methods Mol Biol* (2019) 1914:295–308. doi: 10.1007/978-1-4939-8997-3_16
3. Hensel J, Thalmann GN. Biology of Bone Metastases in Prostate Cancer. *Urology* (2016) 92:6–13. doi: 10.1016/j.urology.2015.12.039
4. Clamp A, Danson S, Nguyen H, Cole D, Clemons M. Assessment of Therapeutic Response in Patients With Metastatic Bone Disease. *Lancet Oncol* (2004) 5:607–16. doi: 10.1016/s1470-2045(04)01596-7
5. Sathiakumar N, Delzell E, Morrisey MA, Falkson C, Yong M, Chia V, et al. Mortality Following Bone Metastasis and Skeletal-Related Events Among Men With Prostate Cancer: A Population-Based Analysis of US Medicare Beneficiaries, 1999-2006. *Prostate Cancer Prostatic Dis* (2011) 14:177–83. doi: 10.1038/pcan.2011.7
6. Mottet N, Bellmunt J, Bolla M, Briers E, Cumberbatch MG, De Santis M, et al. EAU-ESTRO-SIOG Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Local Treatment With Curative Intent. *Eur Urol* (2017) 71:618–29. doi: 10.1016/j.eururo.2016.08.003
7. Padhani AR, Lecouvet FE, Tunariu N, Koh DM, De Keyzer F, Collins DJ, et al. Rationale for Modernising Imaging in Advanced Prostate Cancer. *Eur Urol Focus* (2017) 3:223–39. doi: 10.1016/j.euf.2016.06.018
8. Scher HI, Morris MJ, Stadler WM, Higano C, Basch E, Fizazi K, et al. Trial Design and Objectives for Castration-Resistant Prostate Cancer: Updated Recommendations From the Prostate Cancer Clinical Trials Working Group 3. *J Clin Oncol* (2016) 34:1402–18. doi: 10.1200/jco.2015.64.2702
9. Padhani AR, Tunariu N. Metastasis Reporting and Data System for Prostate Cancer in Practice. *Magn Reson Imaging Clin N Am* (2018) 26:527–42. doi: 10.1016/j.mric.2018.06.004
10. Perez-Lopez R, Lorente D, Blackledge MD, Collins DJ, Mateo J, Bianchini D, et al. Volume of Bone Metastasis Assessed With Whole-Body Diffusion-Weighted Imaging Is Associated With Overall Survival in Metastatic Castration-Resistant Prostate Cancer. *Radiology* (2016) 280:151–60. doi: 10.1148/radiol.2015150799
11. Chan HP, Samala RK, Hadjiiski LM, Zhou C. Deep Learning in Medical Image Analysis. *Adv Exp Med Biol* (2020) 1213:3–21. doi: 10.1007/978-3-030-33128-3_1
12. Kriegeskorte N, Golan T. Neural Network Models and Deep Learning. *Curr Biol* (2019) 29:R231–r236. doi: 10.1016/j.cub.2019.02.034
13. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation, *18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* Munich, GERMANY (2015) 9351:234–41. doi: 10.1007/978-3-319-24574-4_28
14. Bousabarah K, Ruge M, Brand JS, Hoevels M, Rueß D, Borggrefe J, et al. Deep Convolutional Neural Networks for Automated Segmentation of Brain

Metastases Trained on Clinical Data. *Radiat Oncol* (2020) 15:87. doi: 10.1186/s13014-020-01514-6
15. Park YW, Jun Y, Lee Y, Han K, An C, Ahn SS, et al. Robust Performance of Deep Learning for Automatic Detection and Segmentation of Brain Metastases Using Three-Dimensional Black-Blood and Three-Dimensional Gradient Echo Imaging. *Eur Radiol* (2021) 31:6686–95. doi: 10.1007/s00330-021-07783-3
16. Goehler A, Harry Hsu TM, Lacson R, Gujrathi I, Hashemi R, Chlebus G, et al. Three-Dimensional Neural Network to Automatically Assess Liver Tumor Burden Change on Consecutive Liver MRIs. *J Am Coll Radiol* (2020) 17:1475–84. doi: 10.1016/j.jacr.2020.06.033
17. Aoki Y, Nakayama M, Nomura K, Tomita Y, Nakajima K, Yamashina M, et al. The Utility of a Deep Learning-Based Algorithm for Bone Scintigraphy in Patient With Prostate Cancer. *Ann Nucl Med* (2020) 34:926–31. doi: 10.1007/s12149-020-01524-0
18. Wuestemann J, Hupfeld S, Kupitz D, Genseke P, Schenke S, Pech M, et al. Analysis of Bone Scans in Various Tumor Entities Using a Deep-Learning-Based Artificial Neural Network Algorithm-Evaluation of Diagnostic Performance. *Cancers (Basel)* (2020) 12:2654. doi: 10.3390/cancers12092654
19. Lin Q, Luo M, Gao R, Li T, Man Z, Cao Y, et al. Deep Learning Based Automatic Segmentation of Metastasis Hotspots in Thorax Bone SPECT Images. *PloS One* (2020) 15:e0243253. doi: 10.1371/journal.pone.0243253
20. Lin Q, Li T, Cao C, Cao Y, Man Z, Wang H. Deep Learning Based Automated Diagnosis of Bone Metastases With SPECT Thoracic Bone Images. *Sci Rep* (2021) 11:4223. doi: 10.1038/s41598-021-83083-6
21. Colombo A, Saia G, Azzena AA, Rossi A, Zugni F, Pricolo P, et al. Semi-Automated Segmentation of Bone Metastases From Whole-Body MRI: Reproducibility of Apparent Diffusion Coefficient Measurements. *Diagnostics (Basel)* (2021) 11:499. doi: 10.3390/diagnostics11030499
22. Heredia V, Gonzalez CA, Azevedo RM, Semelka RC. Bone Metastases: Evaluation of Acuity of Lesions using Dynamic Gadolinium-chelate Enhancement, Preliminary Results. *J Magn Reson Imaging* (2011) 34:120–7. doi: 10.1002/jmri.22495
23. Cicek O, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3d U-Net: Learning Dense Volumetric Segmentation From Sparse Annotation. *18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* Munich, GERMANY (2016) 9901:24–32. doi: 10.1007/978-3-319-46723-8_49
24. Taha AA, Hanbury A. Metrics for Evaluating 3D Medical Image Segmentation: Analysis, Selection, and Tool. *BMC Med Imaging* (2015) 15:29. doi: 10.1186/s12880-015-0068-x
25. Stabile A, Giganti F, Rosenkrantz AB, Taneja SS, Villeirs G, Gill IS, et al. Multiparametric MRI for Prostate Cancer Diagnosis: Current Status and Future Directions. *Nat Rev Urol* (2020) 17:41–61. doi: 10.1038/s41585-019-0212-4
26. Patel P, Wang S, Siddiqui MM. The Use of Multiparametric Magnetic Resonance Imaging (mpMRI) in the Detection, Evaluation, and Surveillance of Clinically Significant Prostate Cancer (csPCa). *Curr Urol Rep* (2019) 20:60. doi: 10.1007/s11934-019-0926-0
27. Padhani AR, Lecouvet FE, Tunariu N, Koh DM, De Keyzer F, Collins DJ, et al. METastasis Reporting and Data System for Prostate Cancer: Practical Guidelines for Acquisition, Interpretation, and Reporting of Whole-Body Magnetic Resonance Imaging-Based Evaluations of Multiorgan Involvement in Advanced Prostate Cancer. *Eur Urol* (2017) 71:81–92. doi: 10.1016/j.eururo.2016.05.033
28. Liu X, Han C, Wang H, Wu J, Cui Y, Zhang X, et al. Fully Automated Pelvic Bone Segmentation in Multiparameteric MRI Using a 3D Convolutional Neural Network. *Insights Imaging* (2021) 12:93. doi: 10.1186/s13244-021-01044-z
29. Hu Y, Su F, Dong K, Wang X, Zhao X, Jiang Y, et al. Deep Learning System for Lymph Node Quantification and Metastatic Cancer Identification From Whole-Slide Pathology Images. *Gastric Cancer* (2021) 24:868–77. doi: 10.1007/s10120-021-01158-9
30. Zhu Y, Wei R, Gao G, Ding L, Zhang X, Wang X, et al. Fully Automatic Segmentation on Prostate MR Images Based on Cascaded Fully Convolution Network. *J Magn Reson Imaging* (2019) 49:1149–56. doi: 10.1002/jmri.26337
31. Charron O, Lallement A, Jarnet D, Noblet V, Clavier JB, Meyer P. Automatic Detection and Segmentation of Brain Metastases on Multimodal MR Images With a Deep Convolutional Neural Network. *Comput Biol Med* (2018) 95:43–54. doi: 10.1016/j.compbiomed.2018.02.004
32. Grøvik E, Yi D, Iv M, Tong E, Rubin D, Zaharchuk G. Deep Learning Enables Automatic Detection and Segmentation of Brain Metastases on Multisequence MRI. *J Magn Reson Imaging* (2020) 51:175–82. doi: 10.1002/jmri.26766
33. Chang V, Narang J, Schultz L, Issawi A, Jain R, Rock J, et al. Computer-Aided Volumetric Analysis as a Sensitive Tool for the Management of Incidental Meningiomas. *Acta Neurochir (Wien)* (2012) 154:589–97. doi: 10.1007/s00701-012-1273-9
34. 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
35. Chang CY, Buckless C, Yeh KJ, Torriani M. Automated Detection and Segmentation of Sclerotic Spinal Lesions on Body CTs Using a Deep Convolutional Neural Network. *Skeletal Radiol* (2021). doi: 10.1007/s00256-021-03873-x
36. Costelloe CM, Kundra V, Ma J, Chasen BA, Rohren EM, Bassett RL, et al. Fast Dixon Whole-Body MRI for Detecting Distant Cancer Metastasis: A Preliminary Clinical Study. *J Magn Reson Imaging* (2012) 35(2):399–408. doi: 10.1002/jmri.22815
37. Lecouvet FE, Pasoglou V, Van Nieuwenhove S, Van Haver T, de Broqueville Q, Denolin V, et al. Shortening the Acquisition Time of Whole-Body MRI: 3d T1 Gradient Echo Dixon vs Fast Spin Echo for Metastatic Screening in Prostate Cancer. *Eur Radiol* (2020) 30:3083–93. doi: 10.1007/s00330-019-06515-y

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.

frontiers in Oncology

REVIEW
published: 07 January 2022
doi: 10.3389/fonc.2021.689509

Radiomics and Radiogenomics in Evaluation of Colorectal Cancer Liver Metastasis

Yun Wang, Lu-Yao Ma, Xiao-Ping Yin and Bu-Lang Gao*

CT-MRI Room, Affiliated Hospital of Hebei University, Baoding, China

Colorectal cancer is one common digestive malignancy, and the most common approach of blood metastasis of colorectal cancer is through the portal vein system to the liver. Early detection and treatment of liver metastasis is the key to improving the prognosis of the patients. Radiomics and radiogenomics use non-invasive methods to evaluate the biological properties of tumors by deeply mining the texture features of images and quantifying the heterogeneity of metastatic tumors. Radiomics and radiogenomics have been applied widely in the detection, treatment, and prognostic evaluation of colorectal cancer liver metastases. Based on the imaging features of the liver, this paper reviews the current application of radiomics and radiogenomics in the diagnosis, treatment, monitor of disease progression, and prognosis of patients with colorectal cancer liver metastases.

Keywords: colorectal cancer, liver metastasis, radiomics, gene, treatment, prognosis

OPEN ACCESS

Edited by:
Oliver Diaz,
University of Barcelona, Spain

Reviewed by:
Zhongxiang Ding,
Zhejiang University, China
Fan Xia,
Fudan University, China

****Correspondence:***
Xiao-Ping Yin
yinxiaoping78@sina.com

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

Received: *01 April 2021*
Accepted: *03 December 2021*
Published: *07 January 2022*

Citation:
Wang Y, Ma L-Y, Yin X-P and Gao B-L (2022) Radiomics and Radiogenomics in Evaluation of Colorectal Cancer Liver Metastasis. Front. Oncol. 11:689509. doi: 10.3389/fonc.2021.689509

1 INTRODUCTION

Colorectal cancer (CRC) is the third most prevalent malignancy and the second commonest cause of cancer-related deaths throughout the world (1), with the incidence and mortality still on the rise in recent years (2). Because of the hepatic unique blood circulation characteristics, the liver has become the most common organ for blood metastasis of cancers, accounting for 25% of all cancer metastasis (3) and approximately 35%–55% of CRC (4, 5). The liver has uniquely favorable conditions for stagnation and growth of cancerous cells, with double blood supply from the visceral and portal vascular systems and natural spaces among adjacent endothelial sinusoidal cells that are deficient of a typical basement membrane for covering (3, 6, 7). Hepatic metastasis is a critical indicator of prognosis for patients with primary cancers, and the life expectancy of patients with hepatic metastases from gastrointestinal cancers is only 6 months without appropriate treatment (8). Accurate prediction and differentiation of liver metastases from CRC is critical to making an appropriate therapeutic plan and improving the prognosis of the patients. Ultrasound, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) have been routinely applied to detect and assess liver lesions, including metastases of cancer (9, 10). Some liver metastatic lesions from primary cancers of different systems may have common characteristics, including hyperechoic lesions surrounded by a hypoechoic halo (targeted ring sign) in primary gastrointestinal and vascular carcinomas on ultrasound imaging and presence of calcification in CRC or ovarian carcinomas (7, 11, 12). Metastatic lesions with typical imaging features may be easily identified from specific primary carcinomas; however, this kind of lesion accounts for only a small proportion of metastatic lesions, with most of the metastatic lesions being

atypical on imaging, whose specific origin cannot be identified easily. Thus, thorough laboratory and physical examinations, molecular genetic test, and tissue biopsy have been applied to assess the primary origin of liver metastases even though these tests and examinations are costly, invasive, or time-consuming (13, 14).

With the development of great-volume computing capability, it is currently feasible to quickly extract countless quantitative characteristics from three-dimensional imaging data of MRI, CT, ultrasound, and PET for evaluation of the nature of different lesions, because digital medical images contain considerable information that reflects potential pathophysiology. This technology of transforming digital medical imaging data into high-dimensional data for assessment and decision support is referred to as radiomics (15). The framework of radiomics application is shown in **Figure 1**. The radiomics technology has been motivated by the notion that biomedical images comprise information that mirrors and can be used to reveal basic pathophysiology through quantitative analysis. It has been applied in many conditions, but the most developed field of application is in oncology. Quantitative features of imaging are based on imaging shape, intensity, volume, size, and texture, which provide detailed information on tumor microenvironment and phenotype distinct from that offered by laboratory results, clinical reports, and genomic or proteomic analyses. Combined with other clinical information, these features can be used for correlation analysis with clinical results and decision-making, and radiomics can thus provide countless imaging biomarkers to potentially help cancer diagnosis, detection, prognosis evaluation, prediction of treatment response, and monitoring of disease progression. Radiomics is a young field of study and will undergo a slow progress because of technical complexity, datum overfitting, deficiency of standards for outcome validation, incomplete presentation of outcomes, and unrecognizable confounding factors in the databases.

Radiogenomics refers to the exploring of radiomics data to find correlations with genomic modes and has aroused considerable interest in the research field of oncology (15). Here, in this paper, radiogenomics only indicates the combination of genomic information and radiomic features to enable decision support rather than whole-genome analysis to determine the genetic causes of radiosensitive variations in the scope of radiation oncology. Radiogenomics is important because not all patients have had their cancerous diseases genomically profiled even though they may undergo imaging examinations during the course of disease. Radiogenomic data can provide gene expression or mutation information to increase diagnostic, predictive, and prognostic capability and to enable precision therapy because these radiomics data are originated from the complete tumor lesion rather than a small sample of tissue.

In patients with CRC, one factor significantly affecting the prognosis is the proper management of colorectal cancer liver metastases (CRLM), and surgical treatment stands for the only opportunity of long-term survival. The 5-year survival rate of CRC patients with complete resection of liver metastases has been reported to be approximately 30% higher than that without appropriate treatment of the liver metastases (16). Therefore, one of the keys to improving the prognosis of CRC patients is to detect liver metastases for initiating appropriate treatment as soon as possible. Currently, few studies have been performed on radiomics or radiogenomics of CRLM, and this review focused on the radiomics and radiogenomic features of CRLM, trying to facilitate early detection and appropriate treatment of CRLM besides evaluation of its genetic factors and response to treatment for improving the prognosis. The flow chart of the content of this paper is shown in **Figure 2**.

2 RADIOMICS PROGRESS IN THE DIAGNOSIS AND TREATMENT OF CRLM

In recent years, the field of medical image analysis has developed rapidly, and the development of pattern recognition tools has promoted fast progress of quantitative feature extraction. By extracting a great deal of quantitative features from medical imaging data, radiomics can be used to analyze image

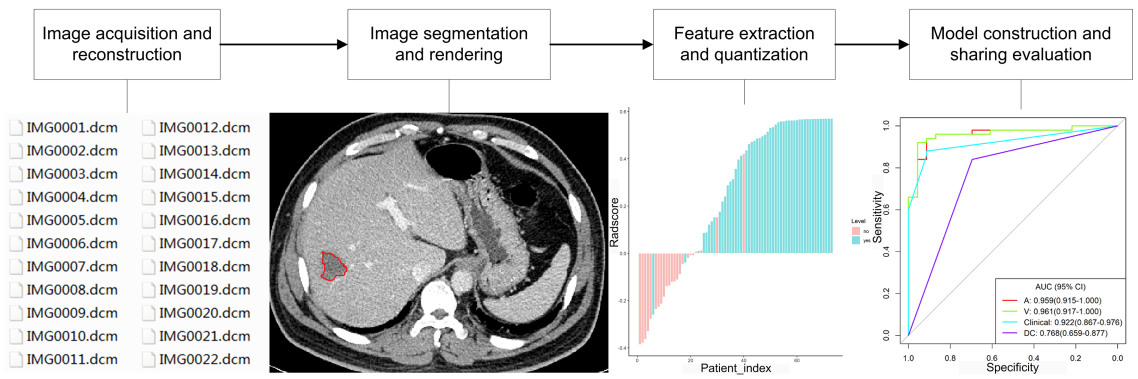


FIGURE 1 | Framework of radiomics application.

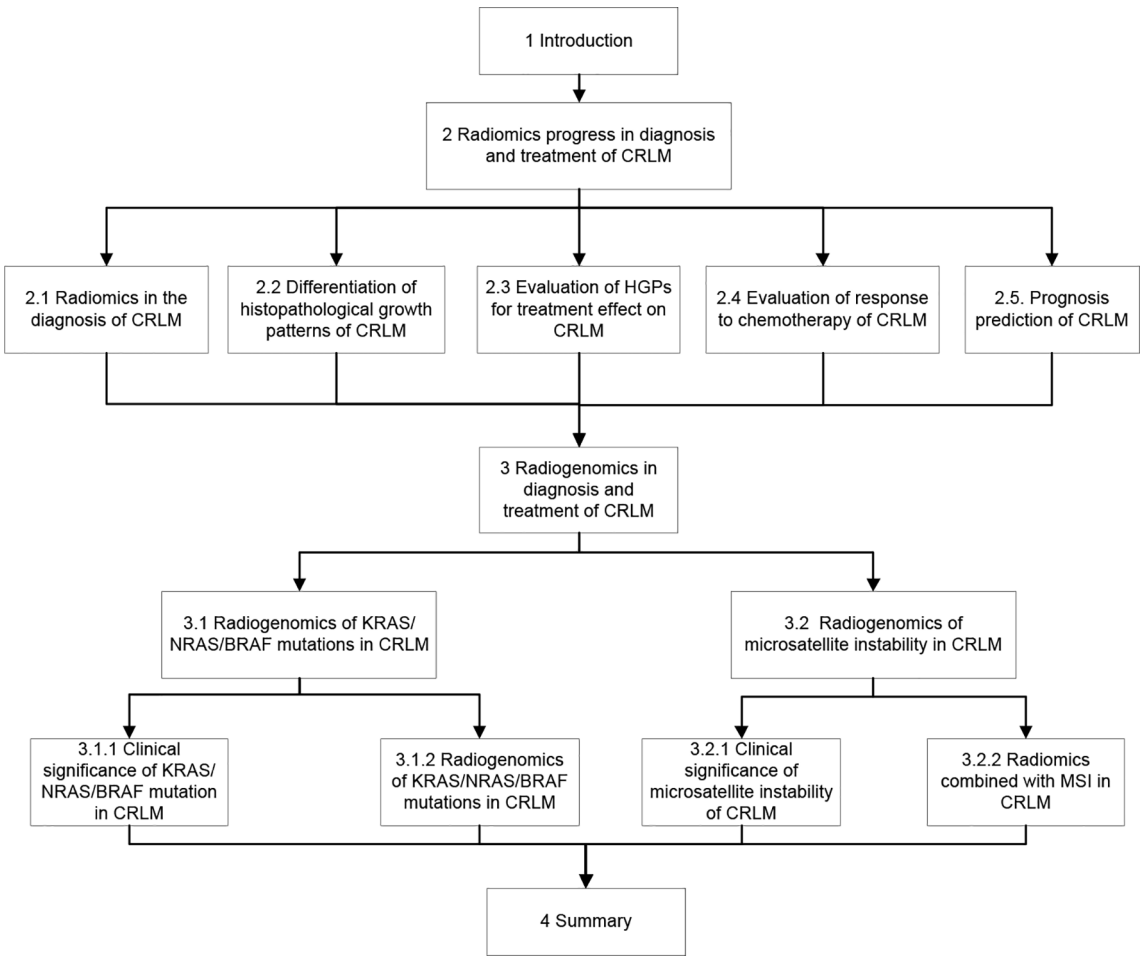


FIGURE 2 | Flow chart of this paper.

information in detail. Compared with traditional approaches of imaging diagnosis, it can significantly improve tumor diagnosis, grading and staging, evaluation of responses to chemotherapy, and prognosis prediction (15, 17, 18), providing professional guidance for treatment planning.

2.1 Radiomics in the Diagnosis of CRLM

With the progress of imaging technology, conventional imaging approaches can effectively detect large and typical CRLM. However, due to the complexity of hepatic hemodynamics and differences of liver parenchymal background on imaging among patients, different imaging modalities perform differently in diagnosis of atypical or tiny liver metastases. It is hard to detect tiny or occult metastases by using the existing imaging approaches; however, identification of these lesions is crucial to early management and improved prognoses. Radiomics features, including entropy, texture and texture ratio, uniformity, and convolutional neural networks (CNNs), have been effectively applied for diagnosis of CRLM. In assessing the capability of whole-liver CT imaging texture analyses of hepatic parenchyma in distinguishing CRC patients with simultaneous hepatic metastasis (n = 10), heterochronous hepatic metastasis within 18 months after initial staging (n = 4), or no hepatic metastasis (n = 15), Rao et al. (19) found that the mean entropy of the whole liver was significantly ($p < 0.05$) higher in patients with synchronous metastases than those without hepatic metastases, whereas the mean uniformity of the whole liver was significantly ($p < 0.05$) lower in patients with synchronous metastases than those without liver metastases. This study indicated that texture evaluation of seemingly disease-free liver is promising to distinguish CRC patients with or without hepatic metastases. After analyzing the texture in non-enhanced CT imaging in seemingly non-diseased regions of liver for impact of hepatic texture by presence of malignant tumors in patients with CRC, Ganeshan et al. (20) found that the fine to medium texture ratio after imaging filtration was significantly ($p < 0.05$) different in seemingly non-diseased hepatic areas in patients with hepatic metastasis compared with those without liver metastasis (entropy, p = 0.0257) or those with extra-liver disease (uniformity, p = 0.0143). Imaging textures of entropy and uniformity have been found to be more advantageous to other features in the diagnosis of CRLM.

CNNs are able to generate useful characteristics from imaging data and have been proven to have high values in predicting oncological outcomes (5, 21–23). Lee et al. used CNNs to generate imaging features from the liver parenchyma in 2019 patients with stage I–III CRC for predicting metachronous liver metastasis based on preoperative abdominal CT imaging (5). They found that the radiomics model combining clinical variables with the top principal components of imaging had the greatest performance (mean AUC = 0.747) to predict 5-year metachronous liver metastasis compared with the model using clinical features only. Even though no hepatic metastasis was found during the initial colectomy, the radiomics features using the CNNs could be used to predict possible metachronous liver metastasis.

2.2 Differentiation of Histopathological Growth Patterns of CRLM

The heterogeneities of genetic, phenotypic, epigenetic, and morphological features inside and outside the CRLM lesion result in different responses to systemic treatment (24, 25). The histopathological growth pattern (HGP) is one such heterogeneity with corresponding microvasculatures. Based on the interface of cancerous cells with adjacent hepatic texture, CRLM has two primary kinds of HGPs: replacement and desmoplastic, with other uncommon kinds of mixed and pushing HGPs (26). The desmoplastic HGP is characterized by separation of the cancerous cells from the hepatic texture by a fibrous band with lymphocytic infiltration and sprouting angiogenesis in the microvasculature. In this pattern, the cancerous cells initiate a reaction similar to the healing of wounds: scar tissues are created with presence of inflammation and new blood vessels. In the replacement HGP, the cancerous cells constitute cellular plates that are in continuity with the hepatocytic plate, allowing the cancerous cells to displace hepatocytes and co-opt the sinusoidal blood vessels at the cancer–liver interface, without disturbing the hepatic stromal architecture or inducing sprouting angiogenesis (25, 27, 28). Desmoplastic metastases are frequently well or moderately differentiated, whereas replacement liver metastases are of poor differentiation, lacking immune reaction and secondary glandular structures (27, 29). The pushing HGP is less common with the hepatocyte plate being compressed and pushed away by the metastatic cancer cells, with no desmoplastic rim around the cancerous cells or direct contact of the cancerous cells with the hepatocytes.

The HGPs of CRLM can be effectively differentiated using multidetector CT-based radiomics and MRI-based radiomics (multi-habitat and multi-sequence) (25, 30). After studying 126 patients with CRLM lesions who had undergone abdominal contract-enhanced CT imaging followed by partial hepatectomy with histopathologically confirmed HGPs including desmoplastic HGP in 68 patients and replacement HGP in 58, Cheng et al. (30) found that the fused radiomics signature had the best predictive performance in differentiating replacement from desmoplastic HGPs (AUC of 0.926 and 0.939, respectively, in the training and validating set), with good discrimination demonstrated in the clinical-radiomics combined model (C-indices of 0.941 and 0.833, respectively, in the training and validating set). Han et al. (25) investigated MRI data of 182 resected CRLM lesions in chemotherapy-free patients including desmoplastic HGPs in 59 patients and replacement HGPs in 123, with the decision tree algorithm being used for radiomics modeling, fused radiomics model being reconstructed from combination of radiomics signatures of all sequences, and clinical and combined models being constructed *via* multivariate logistic regression analysis. They found that the fused radiomics model of tumor zone and the radiomics model of tumor–hepatic interface zone exhibited superior performance to any single sequence or the clinical model and that the radiomics model of tumor–liver interface zone was better than that of the tumor zone (AUC of 0.912 vs. 0.879). The combined model had good discriminating capability, with the AUC of nomogram being 0.971, 0.909, and 0.905, respectively, in the training, internal validating, and external validating set. Their study (25) revealed that MRI-based radiomics is capable of predicting the predominant CRLM HGPs as a potential biomarker for therapeutic strategy. Through analysis of the above studies, it was found that the combination model of radiomics and clinical information can show better discrimination ability than the single radiomics model.

2.3 Evaluation of HGPs for Treatment Effect on CRLM

Metastases are the major death cause in most patients with solid malignancies, and hepatic metastasis is the critical factor for survival of patients with advanced malignant tumors (27, 28). Histological presentations of liver metastases are heterogeneous and reflected by different HGPs that affect clinical outcomes. The desmoplastic HGP is a positive prognostic biomarker while the replacement HGP is a negative one (27). A retrospective study enrolling 732 patients found that the exclusive desmoplastic growth serves as a positive prognostic marker for patients with CRLM, which is not matched by any other factors evaluated (31). In this study, 19% of patients without chemotherapy (n = 367) had desmoplastic growth in the whole tumor–hepatic interface and were independently associated with 50% 5-year survival rate without progression (hazard ratio or HR: 0.54, p = 0.001) and 78% 5-year overall survival (HR: 0.39, $p < 0.001$). CRLM lesions with this kind of HGP are more suitable for regional metastases-directed treatment. On the contrary, replacement HGP is linked to poor pathological responses, with the presence of a large proportion of cancerous cells after chemotherapy, and bad imaging reaction on CT in patients with primary chemotherapy and anti-angiogenesis therapy before surgery for CRLM (32). This type of HGP occurs more often in new hepatic metastatic lesions that grow even during systemic therapy. The replacement HGP indicates not only worse overall and progression-free survival (31, 33, 34), but also resistance to systemic therapy in patients with CRLM (32). A possible reason for the resistance to systemic therapy of the replacement type of HGP is vessel co-option, which serves as an approach of continuous blood supply when the vascular

endothelial growth factor is inhibited by treatment (35). Moreover, different HGPs have varied immune phenotypes that contribute to varied responses to immune therapy. Evidence has indicated that tumors with limited numbers of infiltrated T cells are in essence frequently resistant to immune therapy (36). Vascular co-opting hepatic lesions of metastasis usually have low infiltration of immune cells or inflammatory cells as demonstrated in lesions with the replacement type of HGP in contrast to those with desmoplastic HGP which are frequently surrounded by a lot of lymphocytes in the dense rime (29, 37). Thus, the types of HGPs differentiated using the multidetector CT-based radiomics and MRI-based radiomics (25, 30) may indicate the prognosis of patients with relevant types of HGP in CRLM lesions. In studying the HGP types of CRLM using MRI-based radiomics in comparison with the histopathological types, Han et al. (25) found that more tissue types were presented in the desmoplastic HGP lesion of CRLM, including inflammatory, fibrosis, tumor, and hepatic cells, indicating greater heterogeneity than lesions of replacement HGP. Replacement and desmoplastic HGPs may be able to predict responses to bevacizumab and long-term prognosis. Galjart et al. have convincingly demonstrated that patients with CRLM and an exclusive desmoplastic HGP (100% of the tumor–hepatic interface) undergoing partial hepatectomy have outstandingly good outcomes (31).

2.4 Evaluation of Response to Chemotherapy of CRLM

In CRLM patients, less than 30% were initially resectable (38). In some patients, the metastatic foci, which could not be removed, might disappear on imaging after appropriate therapy, but some metastases could still be detected during radical surgery. Because radiomics can explore subtle changes of tumor and liver texture before and after treatment, it can be used to evaluate the response of CRLM lesions to chemotherapy (39–48). The CRLM lesion uniformity, entropy, homogeneity (variance and angular second moment), gray-tone difference, matrix contrast and shape, skewness, narrowed standard deviation, mean attenuation, density of major hepatic lesion, and histogram parameters for apparent diffusion coefficient maps have all be used to predict responses to chemotherapy. Good responses have been associated with decreased entropy, increased uniformity, higher variance, lower angular second moment, lower baseline skewness value, narrowed standard deviation, high mean attenuation, mean values of histogram parameters for apparent diffusion coefficient maps, and high baseline density of dominant hepatic lesions.

The entropy of CRLM lesions had been reported to decrease in patients with good responses while the uniformity increased after chemotherapy (entropy: −5.13 in good responding patients and +1.27 in non-responding patients, OR = 1.34; uniformity: +30.84 vs. −0.44, respectively, OR = 0.95) (45). However, a higher entropy had also been associated slightly with therapeutic success (6.65 ± 0.26 in patients with good responses vs. 6.51 ± 0.34 in non-responding patients, $P = 0.08$) (41), and a low baseline uniformity was related to a good response (cutoff ≥ 0.42; OR = 20, 95%CI = 1.85–217.4) (46). Two measures for homogeneity of lower angular second moment and a higher variance had been demonstrated to associate with good responding CRLM lesions rather than non-responding lesions on T2 MRI imaging, with the variance of 446.07 ± 329.60 in patients with good responses vs. 210.23 ± 183.39 in non-responding patients ($p < 0.001$) and the angular second moment of 0.96 ± 0.02 vs. 0.98 ± 0.01, respectively ($p < 0.001$).

After investigating therapeutic radiomics features for predicting tumor sensitivity in 667 patients with CRLM to 5-fluorouracil, irinotecan, and folinic acid alone or combined with cetuximab, Dercle et al. (42) found that the radiomics response signature outperformed known biomarkers of the KRAS mutation status and tumor contraction rate in the early prediction of therapeutic sensitivity and for guiding decisions of cetuximab therapy. In evaluating the significance of pre-treatment CT texture analyses for predicting treatment responses in 82 patients with CRLM after combined targeting chemotherapy, Zhang et al. (49) found significant ($p < 0.05$) differences in Entropy, Energy, Variance, Standard deviation, Quantile 95, and sumEntropy between the response and non-response groups in pre-treatment lesions. Lesions with higher Entropy, lower Energy, higher Variance, higher Standard Deviation, and higher sumEntropy seemed to indicate a better therapeutic response. Good diagnostic efficiency was obtained when sumEntropy > 0.867, with a sensitivity of 60.5% and a specificity of 79.5%. Radiomics texture indexes originating from basic CT imaging data of CRLM lesions had the potential capability of imaging biomarkers for predicting cancer response to targeted chemotherapy. By comparing the image features before and after diagnosis and treatment, we found statistically significant radiomics features, such as Entropy, Energy, Variance, Standard deviation, and Quantile, which can all be used to evaluate the remission effect of drugs on CRLM lesions. In the future, these radiomics features can be used clinically as a relatively cheap and noninvasive monitoring means for patients with CRLM or other malignancies.

Most of the reported studies on radiomics are based on CT images, and radiomics features from MRI images can also be used to predict the treatment effect on liver metastases. In order to determine the predictive value of pre-treated MR texture features of CRLM lesions for therapeutic response to chemotherapy, Zhang et al. (48) extracted five histogram features (variance, mean, kurtosis, skewness, and entropy) and five co-occurrence matrix features of gray level (GLCM; entropy), angular second moment, correlation, inverse difference moment, and contrast) from whole liver MRI T2WI data of 26 patients with CRLM before chemotherapy. After careful evaluation, a higher variance, contrast, entropy, entropy, a lower angular second moment, correlation, and inverse difference moment were revealed to significantly ($p < 0.05$) independently associate with good responses to chemotherapy (AUCs 0.602–0.784). Multivariable logistic regression demonstrated that variance ($p < 0.001$) and angular second moment ($p = 0.001$) remained predictive parameters to distinguish responding from non-responding tumors, with the highest AUC of 0.814 (48).

2.5 Prognosis Prediction of CRLM

After active surgical resection, radiofrequency, and chemotherapeutic targeted therapy, some patients with CRLM can achieve a high-quality survival of up to 10 years, whereas others only obtain a short tumor-free survival. Individual differences make the application of personalized treatment strategy particularly important, and identifying risk factors allows clinicians to develop surveillance strategies for patients who are at a higher risk of recurrence. Researchers all over the world have proposed many scoring systems for grading and predicting prognosis of CRLM patients with different tumor loads (17–19), but the ultimate effects on prognosis may be quite different. The radiomics features of heterogeneity, homogeneity, uniformity, Graytone difference matrix contrast, spatial heterogeneity, entropy, texture, and gray level size zone matrix have been used to evaluate the prognoses of patients with CRLM.

Radiomics features have been used to predict the survival of patients with CRLM who have undergone chemotherapy or hepatic surgery because radiomics can assess subtle liver texture differences on different images (40–43, 46, 47, 50–53). An association had been revealed between CRLM heterogeneity/homogeneity and survival. Patients with a greater uniformity of CRLM on CT imaging (cutoff value ≥ 0.42 with a relative risk of 6.94 for overall survival and a relative risk of 5.05 for progression-free survival) had been reported to have poor overall survival and progression-free survival (46). A shorter overall survival had also been demonstrated to associate with metastatic homogeneity on CT imaging (HR: 1.5×10^{20}–1.3×10^{49}) (40). After comparing with before chemotherapy, a radiomic signature based on two heterogeneity features, Graytone Difference Matrix contrast and spatial heterogeneity, had been related to overall survival (HR = 44.3 for patients with superior image quality; HR = 6.5 for patients with conventional image quality) (42), with the radiomic signature having a better value in predicting survival than the 8-week tumor shrinkage or KRAS-mutational status assessed in accordance with the RECIST criteria (AUC 0.80 vs. 0.67 for KRAS and 0.75 for RECIST, $p < 0.001$) in the validation setting. The CRLM heterogeneity at 18F-FDG PET/CT was also confirmed to be a predictor of shorter overall survival (HR 4.29) at multivariant analysis (51), and a model constructed with numbers of metastases, histogram uniformity, and metabolic cancer volume was constructed to predict shorter event-free survival (HR 3.20, $p < 0.001$) (51).

Entropy of the metastatic lesions had been associated with the prognosis of patients with CRLM (40, 41, 50). It had been reported that the overall survival was in a positive correlation with the entropy of CRLM [HR: 0.16–0.63 (40), and HR = 0.65, 95% CI = 0.44–0.95 (50)]. The value of entropy ratio between CRLM and liver texture had also been demonstrated to relate to the prognosis, with a negative correlation between the value and overall survival (HR 1.9) (41). After studying the tumor and liver texture on CT portal venous-phase images in 230 patients with CRLM (120 in the training and 110 in the validation group) before and 2 months after chemotherapy, Dohan et al. (43) established a predictive model of efficacy after 6 months of chemotherapy, which is as effective as the RECIST1.1 evaluation criteria for solid tumors. The radiomic signature with the combination of decreases in sum of target liver lesions, density, and texture analyses of dominant liver lesion at baseline and 2-month CT imaging data could predict the overall survival and detect tumors with good responses better than the RECIST1.1 criteria for CRLM treated by bevacizumab and FOLFIRI as first-line medicines.

Other radiomics features have also been related to the survival. The combination of CRLM correlation and contrast into a single texture parameter had been reported to associate with overall survival (HR 2.35) (53). One texture analysis score combining three features of high baseline density of dominant hepatic lesion, reduction in kurtosis, and decrease in the sum of target hepatic lesions assessed 2 months after chemotherapy had been demonstrated to strongly associate with overall survival (SPECTRA score >0.02 vs. ≤0.02, with the HR of 2.82 in the training set and 2.07 in the validating set) (43). Radiomic evaluation score 2 months later had the same prediction value of prognosis as the RECIST criteria following chemotherapy for 6 months. In the gray level size zone matrix, the small area emphasis (positive parameter of prognosis, HR 0.62) and the minimal pixel value (negative parameter, HR 1.66) had been revealed to be related to progression-free survival (52).

In addition to the above mentioned radiomics features, CRLM density on CT imaging (46), ShapeSI4 (in a radiomic signature) (42), standard deviation (40), future hepatic residual energy and entropy combined as a single linear predictor (53), and AUC of volume histograms at PET-CT (47) have also been reported to associate with the overall survival.

3 RADIOGENOMICS IN DIAGNOSIS AND TREATMENT OF CRLM

Radiogenomics can be used to discover the radiomics features that reflect gene expression or polymorphism for further understanding the occurrence and development of diseases (54). Radiogenomics promises to understand gene expression of diseases through noninvasive and conventional imaging methods. With continuous progress of the technology, radiogenomics has been widely studied in systemic diseases in recent years. Many scholars have reported a correlation between radiomics features and EGFR (epidermal growth factor receptor) mutation (55–57) or ALK (anaplastic lymphoma kinase) rearrangement of lung cancer (58, 59). In detection and management of breast cancers, many researchers have found that breast cancer is associated with radiomics features at the gene sequence level (60), gene expression level (61), and molecular subtype level (62). Marigliano et al. (63) analyzed multiphase CT images (arterial phase, portal-venous phase, and urinary tract phase) of 20 patients with clear cell renal cancer and found that the radiogenomics data derived from these images were well correlated with expressions of some microRNAs (miR-185-5p, miR-21-5p, miR-210-3p, miR-221-3p, and miR-145-5p), especially between entropy and miR-21-5p. Similarly, progress has also been made in radiogenomics for prostate cancer (64).

Currently, there are only limited studies on radiogenomics of tumors involving the liver. Segal et al. were the first in 2007 to explore the correlation of gene expression pattern of a hepatocellular carcinoma with the imaging features, identifying 32 image characteristics from enhanced CT imaging of three phases to be correlated to the expression degrees of 116 genetic biomarkers among 6,732 genes confirmed by microarray analysis (65). However, only three imaging features on average were required to catch expression variations of any genetic marker, and the use of 28 image features combined could explain variations of all 116 genetic markers (65). Moreover, it was found that the genes in some particular molecular profiles had common physiological function, including cellular proliferation and hepatic enzyme syntheses, which could correlate to specific imaging characteristics. Thus, two image features, presence of arteries and absence of low-density halos, were found to correlate with "venous invasion signatures", which are image patterns to predict microscopic venous invasion and OS (65). Kuo et al. (66) also conducted radiogenomic analysis to identify imaging traits in hepatocellular carcinomas, which were related to a genetic expression profile of 61 genes to detect tumor responses to doxorubicin. The enhanced CT imaging data of 30 hepatocellular carcinomas had been studied for six image features, which were found to correlate with the microarray of 18,000 genes.

CRC is a heterogeneous tumor, and its occurrence and development are affected by a variety of factors. Lifestyle habits such as high-fat diet are important risk factors to increase the incidence of CRC (67). Besides external factors, intrinsic genetic factors also affect the occurrence and development of CRC (68). Knowing the status of gene mutation in CRC can effectively provide guidelines for clinical treatment and prognosis evaluation, thus formulating a recurrence surveillance strategy for patients (69).

3.1 Radiogenomics of KRAS/NRAS/BRAF Mutations in CRLM

3.1.1 Clinical Significance of KRAS/NRAS/BRAF Mutation in CRLM

The RAS/RAF/MEK/extracellular signal-regulated kinase signaling cascade is referred to as the pathway of mitogen-activated protein kinase (MAPK), which controls cellular differentiation, proliferation, angiogenesis, migration, and survival. Dysregulation of the pathway constitutes the bases for tumorigenesis (70). This pathway consists of RAS small guanidine triphosphatases (GTPase) and can activate the family proteins of RAF (ARAF, CRAF, and BRAF). Abnormal activation or signaling of the MAPK pathway had been demonstrated in many tumors, including CRC, through some distinctive mechanisms, like mutations in BRAF and RAS (70), which most frequently occur in human neoplasms.

KRAS, NRAS, and HRAS are the RAS oncogenes to encode a family of GTP-adjusted switches and can repeatedly mutate in human cancers (71). Once activated, these genes will cause pleiotropic effects in cells, leading to cellular differentiation, proliferation, and survival. KRAS mutations take up approximately 85% of mutations in the RAS gene in human malignancies, NRAS accounts for approximately 15%, and HRAS accounts for below 1% (72). In CRC, RAS mutations primarily take place in the KRAS gene, and approximately 45% of metastatic CRCs contain activated KRAS mutations (73). NRAS mutation happens in 2%–7% patients with metastatic CRC (71). KRAS gene mutations are related to right-sided colonic cancers, but NRAS gene mutation is related to left-sided primary malignancies and female gender, indicating distinctive biology for NRAS and KRAS mutant molecule subsets of metastatic CRC (74).

KRAS gene is related to the pathogenesis and progression of CRC, and mutation of this gene may cause resistance to EGFR inhibitors and poor tumor response to molecular targeted drugs (75, 76). De Macedo et al. (77) studied the DNA of primary tumor and metastatic tissue in 102 cases of CRLM and found that the KRAS gene was highly homogeneous across the primary CRC cancer areas and consistent in the original cancer lesion with the metastatic tissues in the same person. KRAS mutation is an independent risk factor for the prognosis of patients with CRC (78). Therefore, understanding the KRAS mutation rate in patients with CRC will help treatment planning and prognosis evaluation.

NRAS defines a group of molecules with different clinical features from KRAS-mutant and wild-type metastatic CRC (71). NRAS gene mutation can cause disordered malignant proliferation and promote metastasis (71), thus associating with worse survival and outcomes than KRAS-mutant or wild-type metastatic CRC. Activating mutations in NRAS take place in 30% of cases with skin melanoma, and BRAF mutation happens at a high incidence in these malignancies (74). BRAF or NRAS gene mutation is related to poor survival of metastatic melanoma patients. However, BRAF mutation is reciprocally exclusive with melanoma NRAS mutation and with CRC KRAS mutation.

BRAF mutations take place in 7% of cancers, and approximately 8%–12% of metastatic CRC cases contain BRAF mutations (79). BRAF gene mutation can cause poor drug effect and worse prognosis, and reduce the effect of cancer cell apoptosis, thus aggravating the condition of patients with cancers. Some studies (80) found that the mutation rate of the BRAF gene is higher in patients with lower tumor differentiation.

3.1.2 Radiogenomics of KRAS/NRAS/BRAF Mutations in CRLM

Yang et al. (81) studied 346 radiomic features extracted from portal venous-phase CT imaging data of primary tumors and KRAS/NRAS/BRAF gene mutation in 117 patients with CRC, including 61 cases in the training and 56 in the verification group before treatment. The support vector machine methods and RELIEFF were constructed to choose important features and establish the radiomic features. It was found that the radiomic signature was significantly associated with the KRAS/NRAS/BRAF mutation ($p < 0.001$), with the AUC, sensitivity, and specificity for predicting KRAS/NRAS/BRAF mutation as 0.869, 0.757, and 0.833 in the primary group, and 0.829, 0.686, and 0.857 in the validation group, respectively.

Lubner et al. (50) investigated tumor texture analysis on single CRLM lesion on contrast-enhanced CT imaging in 77 patients before treatment. It was found that entropy (spatial scaling factor or SSF 4, $p = 0.007$), mean positive pixels (SSF 3, $p = 0.002$), and standard deviation (SSF 3, $p = 0.004$) of medium filtration were significantly associated with the tumor stage. Skewness was found to negatively associate with KRAS mutations ($p = 0.02$), whereas the coarse filtration entropy was significantly ($p = 0.03$) associated with survival (HR for death 0.65). Therefore, radiogenomics is expected to understand the gene expression profile of the disease through noninvasive and routine imaging examination and may be a breakthrough in the diagnosis, treatment, disease monitor, and prognosis evaluation of CRC and CRLM.

3.2 Radiogenomics of Microsatellite Instability in CRLM

3.2.1 Clinical Significance of Microsatellite Instability of CRLM

Some kinds of genomic instability are able to drive the initiation and development of CRC. The most common type is chromosomal instability, which is found in 85% of CRC, and another is microsatellite instability (MSI) which occurs in 15% patients with CRC. MSI tumors are a subset of CRC characterized by malfunction of mismatch repair genes (MMR), which can cause failure to repair errors in short tandem repetitive DNA sequences known as microsatellites (82, 83). In the microsatellite sequences, the DNA replication stability is poor and is prone to mismatches. MSI is caused by lack of DNA mismatch repair (MMR) system, arising from germline mutations in the MMR gene, which is prone to the Lynch syndrome, or from epigenetic inactivation of MLH1 in sporadic malignancies. Approximately 5% metastatic CRCs showed MSI or deficient MMR, and sporadic CRC patients with MSI were often related to BRAFV600E mutation *via* its association with CpG methylator phenotype (84).

High-frequency MSI (MSI-H) refers to the occurrence of MSI at two or more sites; low-frequency MSI (MSI-L) is MSI occurring only at one site; microsatellite stability (MSS) indicates MSI, which does not occur at any site (85, 86). MSI has a guiding role in predicting the malignant degree and pathogenesis of tumor, and can also provide direction for clinical selection of treatment plan and prognosis evaluation. Studies have shown that MSI-H can be used as a biomarker to guide clinical immunotherapy for CRLM patients (83, 84, 87). Through transformation therapy of immune drugs, it is possible to remove the metastatic foci so as to further improve survival and quality of life for cancer patients.

3.2.2 Radiomics Combined With MSI in CRLM

Understanding the MSI status is necessary because CRC tissues with MSI have specific biological behavior and may indicate better prognoses and benefit from immunotherapy or resistance to fluorouracil treatment (88). However, the approaches for evaluating MSI status using polymerase chain reaction and immunohistochemistry are performed on pathological tissues from invasive biopsies or surgeries and have not been extensively applied. It is therefore necessary to develop non-invasive and cost-effective methods to predict the MSI status and guide further therapeutic strategies. By extracting 254 radiomics features of intensity from CT imaging of the CRC cancer region in combination with clinical features in 198 patients including 134 patients with microsatellite stable tumors and 64with MSI tumors, Golia Pernicka et al. (89) were able to develop three prediction models with clinical features only, radiomic features only, and combination of radiomic and clinical features. The combined radiomics model outperformed the other two models in predicting MSI, with the AUC of 0.80 and 0.79 for the training and testing set, respectively (specificity 96.8% and 92.5%, respectively).

Fan et al. studied 119 patients with stage II CRC confirmed pathologically, known MSI status, and preoperative enhanced CT images for extracting radiomics features (90). In their study, the radiomics features were obtained from the portal-vein phase CT imaging data of segmented tissues of each complete primary cancer lesion with the Matrix Laboratory software while the radiomic signatures were constructed using the selection operator logistic regression and least absolute shrinkage model. Six radiomics and 11 clinical features were chosen for predicting the MSI status. The model combining both radiomic and clinical features achieved the overall best performance in predicting the MSI status than either the radiomics or clinical feature model alone, yielding the AUC, sensitivity, and specificity of 0.752, 0.663, and 0.841 for the combined model, 0.598, 0.371, and 0.825 for clinical model alone, and 0.688, 0.517, and 0.858 for radiomics model alone, respectively. Combined analyses of radiomic and clinical features improved the predictive efficacy and helped selecting appropriate patients for personalized therapy.

In exploring the value of radiomics analysis derived from dual-energy CT imaging to preoperatively evaluate the MSI status in CRC, Wu et al. (88) investigated 102 CRC patients with pathologically confirmed MSI status and selected nine top features to constitute the radiomic model. They found that radiomic analyses of iodine-based material decomposition imaging data with dual-energy CT has a great capability to predict the MSI status in patients with CRC, with the AUC, accuracy, sensitivity, and specificity of 0.961, 0.875, 1.000, and 0.812 in the training set, and 0.875, 0.788, 0.909, and 0.727 in the testing set, respectively. Good clinical application and calibration were demonstrated with the decision curve and calibration analyses, respectively.

Although there is consistency between CRC MSI status and liver metastasis, there were currently no correlation studies between MSI status and radiomics of liver metastasis.

4 SUMMARY

In the diagnosis, treatment, monitor of disease progression, and prognosis of CRLM, thousands of radiomics features can be extracted, such as image intensity features, high-order features, texture features, and shape features. Due to the lack of unified standards at present, different research teams choose different

radiomics features in the selection of features. Through review of published studies in the literature, it is found that the most widely used radiomics features include entropy, uniformity, variance, and skewness. At present, the unity of the results is relatively poor, but all these results show the feasibility and significance of the application of radiomics and radiogenomics in the diagnosis, treatment, monitor of disease progression, and prognosis of CRLM.

Radiomics and radiogenomics can be widely used in clinical medicine research with noninvasiveness and low cost. However, as a new field, it is still in its infancy, with many limitations. For example, the research data for radiomics mostly come from small samples and single centers, whereas some big data from multicenters are different because of use of different scanning equipment and scanning conditions. Moreover, imaging delineation segmentation approaches may also differ from center to center or from study to study. Future development and research in radiomics and radiogenomics will have to solve these issues for better outcomes.

As an innovative arena in medical imaging, radiomics and radiogenomics can be used to identify pathological process, reveal the underlying pathophysiological mechanisms through medical imaging and clinical data, and identify hidden imaging patterns that can be used to predict tumor biological behavior and patients' prognoses, providing efficient prediction of responses to chemotherapy and survival in addition to accurate and early prediction compared to standard biomarkers. Continuous surveillance of the radiomics and radiogenomics biomarkers will provide adequate information to monitor cancer recurrence and individualized treatment to the constantly changing genome of cancer. The current research results in radiomics and radiogenomics of CRLM warrant further exploration into wider application in other fields.

AUTHOR CONTRIBUTIONS

Study design: X-PY and B-LG. Data collection: YW and L-YM. Supervision: L-YM. Writing of the original version of article: YW. Revision of the original version: B-LG. All authors contributed to the article and approved the submitted version.

FUNDING

This study is funded by Hebei Natural Science Foundation Project (H20212017), Project of Hebei Provincial Department of finance (361007), and Medical Discipline Cultivation Project of Hebei University (2020b05).

REFERENCES

1. Li M, Li X, Guo Y, Miao Z, Liu X, Guo S, et al. Development and Assessment of an Individualized Nomogram to Predict Colorectal Cancer Liver Metastases. *Quant Imaging Med Surg* (2020) 10:397–414. doi: 10.21037/qims.2019.12.16
2. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer Statistics in China, 2015. *CA Cancer J Clin* (2016) 66:115–32. doi: 10.3322/caac.21338
3. Qin H, Wu YQ, Lin P, Gao RZ, Li X, Wang XR, et al. Ultrasound Image-Based Radiomics: An Innovative Method to Identify Primary Tumorous Sources of Liver Metastases. *J Ultrasound Med* (2020) 40(6):1229–44. doi: 10.1002/jum.15506
4. Akgul O, Cetinkaya E, Ersoz S, Tez M. Role of Surgery in Colorectal Cancer Liver Metastases. *World J Gastroenterol* (2014) 20:6113–22. doi: 10.3748/wjg.v20.i20.6113
5. Lee S, Choe EK, Kim SY, Kim HS, Park KJ, Kim D. Liver Imaging Features by Convolutional Neural Network to Predict the Metachronous Liver Metastasis in Stage I-III Colorectal Cancer Patients Based on Preoperative Abdominal Ct Scan. *BMC Bioinformatics* (2020) 21:382. doi: 10.1186/s12859-020-03686-0
6. Harvey CJ, Albrecht T. Ultrasound of Focal Liver Lesions. *Eur Radiol* (2001) 11:1578–93. doi: 10.1007/s003300101002
7. Paley MR, Ros PR. Hepatic Metastases. *Radiol Clin North Am* (1998) 36:349–63. doi: 10.1016/S0033-8389(05)70027-0
8. Brenner H, Kloor M, Pox CP. Colorectal Cancer. *Lancet* (2014) 383:1490–502. doi: 10.1016/S0140-6736(13)61649-9
9. Kirchner J, Sawicki LM, Deuschl C, Gruneisen J, Beiderwellen K, Lauenstein TC, et al. 18 F-Fdg Pet/Mr Imaging in Patients With Suspected Liver Lesions: Value of Liver-Specific Contrast Agent Gadobenate Dimeglumine. *PloS One* (2017) 12:e0180349. doi: 10.1371/journal.pone.0180349
10. Montoya JC, Eckel LJ, DeLone DR, Kotsenas AL, Diehn FE, Yu L, et al. Low-Dose Ct for Craniosynostosis: Preserving Diagnostic Benefit With Substantial Radiation Dose Reduction. *AJNR Am J Neuroradiol* (2017) 38:672–7. doi: 10.3174/ajnr.A5063
11. Katragadda CS, Goldstein HM, Green B. Gray Scale Ultrasonography of Calcified Liver Metastases. *AJR Am J Roentgenol* (1977) 129:591–3. doi: 10.2214/ajr.129.4.591
12. Virmani J, Kumar V, Kalra N, Khandelwal N. Characterization of Primary and Secondary Malignant Liver Lesions From B-Mode Ultrasound. *J Digit Imaging* (2013) 26:1058–70. doi: 10.1007/s10278-013-9578-7
13. Hainsworth JD, Rubin MS, Spigel DR, Boccia RV, Raby S, Quinn R, et al. Molecular Gene Expression Profiling to Predict the Tissue of Origin and Direct Site-Specific Therapy in Patients With Carcinoma of Unknown Primary Site: A Prospective Trial of the Sarah Cannon Research Institute. *J Clin Oncol* (2013) 31:217–23. doi: 10.1200/JCO.2012.43.3755
14. Weiss LM, Chu P, Schroeder BE, Singh V, Zhang Y, Erlander MG, et al. Blinded Comparator Study of Immunohistochemical Analysis Versus a 92-Gene Cancer Classifier in the Diagnosis of the Primary Site in Metastatic Tumors. *J Mol Diagn* (2013) 15:263–9. doi: 10.1016/j.jmoldx.2012.10.001
15. 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
16. House MG, Ito H, Gonen M, Fong Y, Allen PJ, DeMatteo RP, et al. Survival After Hepatic Resection for Metastatic Colorectal Cancer: Trends in Outcomes for 1,600 Patients During Two Decades at a Single Institution. *J Am Coll Surg* (2010) 210:744–752, 752-745. doi: 10.1016/j.jamcollsurg.2009.12.040
17. Bibault JE, Xing L, Giraud P, El Ayachy R, Giraud N, Decazes P, et al. Radiomics: A Primer for the Radiation Oncologist. *Cancer Radiother* (2020) 24:403–10. doi: 10.1016/j.canrad.2020.01.011
18. 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
19. Rao SX, Lambregts DM, Schnerr RS, van Ommen W, van Nijnatten TJ, Martens MH, et al. Whole-Liver Ct Texture Analysis in Colorectal Cancer: Does the Presence of Liver Metastases Affect the Texture of the Remaining Liver? *United Eur Gastroenterol J* (2014) 2:530–8. doi: 10.1177/2050640614552463
20. Ganeshan B, Miles KA, Young RC, Chatwin CR. Texture Analysis in Non-Contrast Enhanced Ct: Impact of Malignancy on Texture in Apparently Disease-Free Areas of the Liver. *Eur J Radiol* (2009) 70:101–10. doi: 10.1016/j.ejrad.2007.12.005

21. Li Q, Kim J, Balagurunathan Y, Liu Y, Latifi K, Stringfield O, et al. Imaging Features From Pretreatment Ct Scans Are Associated With Clinical Outcomes in Nonsmall-Cell Lung Cancer Patients Treated With Stereotactic Body Radiotherapy. *Med Phys* (2017) 44:4341–9. doi: 10.1002/mp.12309
22. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A Survey on Deep Learning in Medical Image Analysis. *Med Image Anal* (2017) 42:60–88. doi: 10.1016/j.media.2017.07.005
23. 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
24. Goey KKH, Sorbye H, Glimelius B, Adams RA, Andre T, Arnold D, et al. Consensus Statement on Essential Patient Characteristics in Systemic Treatment Trials for Metastatic Colorectal Cancer: Supported by the Arcad Group. *Eur J Cancer* (2018) 100:35–45. doi: 10.1016/j.ejca.2018.05.010
25. Han Y, Chai F, Wei J, Yue Y, Cheng J, Gu D, et al. Identification of Predominant Histopathological Growth Patterns of Colorectal Liver Metastasis by Multi-Habitat and Multi-Sequence Based Radiomics Analysis. *Front Oncol* (2020) 10:1363. doi: 10.3389/fonc.2020.01363
26. 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. *J Pathol* (2001) 195:336–42. doi: 10.1002/path.966
27. Latacz E, van Dam PJ, Vanhove C, Llado L, Descamps B, Ruiz N, et al. Can Medical Imaging Identify the Histopathological Growth Patterns of Liver Metastases? *Semin Cancer Biol* (2021) 71:33–41. doi: 10.1016/j.semcancer.2020.07.002
28. van Dam PJ, van der Stok EP, Teuwen LA, Van den Eynden GG, Illemann M, Frentzas S, et al. International Consensus Guidelines for Scoring the Histopathological Growth Patterns of Liver Metastasis. *Br J Cancer* (2017) 117:1427–41. doi: 10.1038/bjc.2017.334
29. van Dam PJ, Daelemans S, Ross E, Waumans Y, Van Laere S, Latacz E, et al. Histopathological Growth Patterns as a Candidate Biomarker for Immunomodulatory Therapy. *Semin Cancer Biol* (2018) 52:86–93. doi: 10.1016/j.semcancer.2018.01.009
30. Cheng J WJ, Tong T, Sheng W, Zhang Y, Han Y, Gu D, et al. Prediction of Histopathologic Growth Patterns of Colorectal Liver Metastases With a Noninvasive Imaging Method. *Ann Surg Oncol* (2019) 26:4587–98. doi: 10.1245/s10434-019-07910-x
31. Galjart B, Nierop PMH, van der Stok EP, van den Braak R, Hoppener DJ, Daelemans S, et al. Angiogenic Desmoplastic Histopathological Growth Pattern as a Prognostic Marker of Good Outcome in Patients With Colorectal Liver Metastases. *Angiogenesis* (2019) 22:355–68. doi: 10.1007/s10456-019-09661-5
32. 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) 22:1294–302. doi: 10.1038/nm.4197
33. Barnhill R, van Dam PJ, Vermeulen P, Champenois G, Nicolas A, Rawson RV, et al. Replacement and Desmoplastic Histopathological Growth Patterns in Cutaneous Melanoma Liver Metastases: Frequency, Characteristics, and Robust Prognostic Value. *J Pathol Clin Res* (2020) 6:195–206. doi: 10.1002/cjp2.161
34. Barnhill R, Vermeulen P, Daelemans S, van Dam PJ, Roman-Roman S, Servois V, et al. Replacement and Desmoplastic Histopathological Growth Patterns: A Pilot Study of Prediction of Outcome in Patients With Uveal Melanoma Liver Metastases. *J Pathol Clin Res* (2018) 4:227–40. doi: 10.1002/cjp2.105
35. Donnem T, Reynolds AR, Kuczynski EA, Gatter K, Vermeulen PB, Kerbel RS, et al. Non-Angiogenic Tumours and Their Influence on Cancer Biology. *Nat Rev Cancer* (2018) 18:323–36. doi: 10.1038/nrc.2018.14
36. Chen DS, Mellman I. Elements of Cancer Immunity and the Cancer-Immune Set Point. *Nature* (2017) 541:321–30. doi: 10.1038/nature21349
37. Hoppener DJ, Nierop PMH, Hof J, Sideras K, Zhou G, Visser L, et al. Enrichment of the Tumour Immune Microenvironment in Patients With Desmoplastic Colorectal Liver Metastasis. *Br J Cancer* (2020) 123:196–206. doi: 10.1038/s41416-020-0881-z
38. Adams RB, Aloia TA, Loyer E, Pawlik TM, Taouli B, Vauthey JN, et al. Selection for Hepatic Resection of Colorectal Liver Metastases: Expert Consensus Statement. *HPB (Oxford)* (2013) 15:91–103. doi: 10.1111/j.1477-2574.2012.00557.x
39. 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
40. Andersen IR, Thorup K, Andersen MB, Olesen R, Mortensen FV, Nielsen DT, et al. Texture in the Monitoring of Regorafenib Therapy in Patients With Colorectal Liver Metastases. *Acta Radiol* (2019) 60:1084–93. doi: 10.1177/0284185118817940
41. Beckers RCJ, Trebeschi S, Maas M, Schnerr RS, Sijmons JML, Beets GL, et al. Ct Texture Analysis in Colorectal Liver Metastases and the Surrounding Liver Parenchyma and Its Potential as an Imaging Biomarker of Disease Aggressiveness, Response and Survival. *Eur J Radiol* (2018) 102:15–21. doi: 10.1016/j.ejrad.2018.02.031
42. Dercle L, Lu L, Schwartz LH, Qian M, Tejpar S, Eggleton P, et al. Radiomics Response Signature for Identification of Metastatic Colorectal Cancer Sensitive to Therapies Targeting Egfr Pathway. *J Natl Cancer Inst* (2020) 112:902–12. doi: 10.1093/jnci/djaa017
43. 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) 69:531–9. doi: 10.1136/gutjnl-2018-316407
44. Liang HY, Huang YQ, Yang ZX, Ying D, Zeng MS, Rao SX. Potential of Mr Histogram Analyses for Prediction of Response to Chemotherapy in Patients With Colorectal Hepatic Metastases. *Eur Radiol* (2016) 26:2009–18. doi: 10.1007/s00330-015-4043-2
45. Rao SX, Lambregts DM, Schnerr RS, Beckers RC, Maas M, Albarello F, et al. Ct Texture Analysis in Colorectal Liver Metastases: A Better Way Than Size and Volume Measurements to Assess Response to Chemotherapy? *United Eur Gastroenterol J* (2016) 4:257–63. doi: 10.1177/2050640615601603
46. Ravanelli M, Agazzi GM, Tononcelli E, Roca E, Cabassa P, Baiocchi G, et al. Texture Features of Colorectal Liver Metastases on Pretreatment Contrast-Enhanced Ct May Predict Response and Prognosis in Patients Treated With Bevacizumab-Containing Chemotherapy: A Pilot Study Including Comparison With Standard Chemotherapy. *Radiol Med* (2019) 124:877–86. doi: 10.1007/s11547-019-01046-4
47. 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
48. Zhang H, Li W, Hu F, Sun Y, Hu T, Tong T. Mr Texture Analysis: Potential Imaging Biomarker for Predicting the Chemotherapeutic Response of Patients With Colorectal Liver Metastases. *Abdom Radiol (NY)* (2019) 44:65–71. doi: 10.1007/s00261-018-1682-1
49. Zhang J, Zhou Y, Qiu M, Wu B. Prediction of the Therapeutic Response After Target-Combined Chemotherapy Treatment for Patients With Liver Metastasis From Colorectal Cancer Using Computed Tomography Texture Analysis. *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi* (2018) 35:914–20. doi: 10.7507/1001-5515.201801062
50. 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
51. Rahmim A, Bak-Fredslund KP, Ashrafinia S, Lu L, Schmidtlein CR, Subramaniam RM, et al. Prognostic Modeling for Patients With Colorectal Liver Metastases Incorporating Fdg Pet Radiomic Features. *Eur J Radiol* (2019) 113:101–9. doi: 10.1016/j.ejrad.2019.02.006
52. Shur J, Orton M, Connor A, Fischer S, Moulton CA, Gallinger S, et al. A Clinical-Radiomic Model for Improved Prognostication of Surgical Candidates With Colorectal Liver Metastases. *J Surg Oncol* (2019) 121:357–364. doi: 10.1002/jso.25783
53. Simpson AL, Doussot A, Creasy JM, Adams LB, Allen PJ, DeMatteo RP, et al. Computed Tomography Image Texture: A Noninvasive Prognostic Marker of Hepatic Recurrence After Hepatectomy for Metastatic Colorectal Cancer. *Ann Surg Oncol* (2017) 24:2482–90. doi: 10.1245/s10434-017-5896-1
54. Bodalal Z, Trebeschi S, Nguyen-Kim TDL, Schats W, Beets-Tan R. Radiogenomics: Bridging Imaging and Genomics. *Abdom Radiol (NY)* (2019) 44:1960–84. doi: 10.1007/s00261-019-02028-w

55. 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) 29:4742–50. doi: 10.1007/s00330-019-06024-y
56. Koo HJ, Kim MY, Park S, Lee HN, Kim HJ, Lee JC, et al. Non-Small Cell Lung Cancer With Resistance to Egfr-Tki Therapy: Ct Characteristics of T790m Mutation-Positive Cancer. *Radiology* (2018) 289:227–37. doi: 10.1148/radiol.2018180070
57. Zhang L, Chen B, Liu X, Song J, Fang M, Hu C, et al. Quantitative Biomarkers for Prediction of Epidermal Growth Factor Receptor Mutation in Non-Small Cell Lung Cancer. *Transl Oncol* (2018) 11:94–101. doi: 10.1016/j.tranon.2017.10.012
58. Mendoza DP, Stowell J, Muzikansky A, Shepard JO, Shaw AT, Digumarthy SR. Computed Tomography Imaging Characteristics of Non-Small-Cell Lung Cancer With Anaplastic Lymphoma Kinase Rearrangements: A Systematic Review and Meta-Analysis. *Clin Lung Cancer* (2019) 20:339–49. doi: 10.1016/j.cllc.2019.05.006
59. Wang H, Schabath MB, Liu Y, Han Y, Li Q, Gillies RJ, et al. Clinical and Ct Characteristics of Surgically Resected Lung Adenocarcinomas Harboring Alk Rearrangements or Egfr Mutations. *Eur J Radiol* (2016) 85:1934–40. doi: 10.1016/j.ejrad.2016.08.023
60. Yamamoto S, Maki DD, Korn RL, Kuo MD. Radiogenomic Analysis of Breast Cancer Using Mri: A Preliminary Study to Define the Landscape. *AJR Am J Roentgenol* (2012) 199:654–63. doi: 10.2214/AJR.11.7824
61. Zhu Y, Li H, Guo W, Drukker K, Lan L, Giger ML, et al. Deciphering Genomic Underpinnings of Quantitative Mri-Based Radiomic Phenotypes of Invasive Breast Carcinoma. *Sci Rep* (2015) 5:17787. doi: 10.1038/srep17787
62. Mazurowski MA, Zhang J, Grimm LJ, Yoon SC, Silber JI. Radiogenomic Analysis of Breast Cancer: Luminal B Molecular Subtype Is Associated With Enhancement Dynamics at Mr Imaging. *Radiology* (2014) 273:365–72. doi: 10.1148/radiol.14132641
63. Marigliano C, Badia S, Bellini D, Rengo M, Caruso D, Tito C, et al. Radiogenomics in Clear Cell Renal Cell Carcinoma: Correlations Between Advanced Ct Imaging (Texture Analysis) and MicroRNAs Expression. *Technol Cancer Res Treat* (2019) 18:1533033819878458. doi: 10.1177/1533033819878458
64. Wibmer AG, Robertson NL, Hricak H, Zheng J, Capanu M, Stone S, et al. Extracapsular Extension on Mri Indicates a More Aggressive Cell Cycle Progression Genotype of Prostate Cancer. *Abdom Radiol (NY)* (2019) 44:2864–73. doi: 10.1007/s00261-019-02023-1
65. 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
66. Kuo MD, Gollub J, Sirlin CB, Ooi C, Chen X. Radiogenomic Analysis to Identify Imaging Phenotypes Associated With Drug Response Gene Expression Programs in Hepatocellular Carcinoma. *J Vasc Interv Radiol* (2007) 18:821–31. doi: 10.1016/j.jvir.2007.04.031
67. O'Keefe SJ, Li JV, Lahti L, Ou J, Carbonero F, Mohammed K, et al. Fat, Fibre and Cancer Risk in African Americans and Rural Africans. *Nat Commun* (2015) 6:6342. doi: 10.1038/ncomms7342
68. Giannakis M, Mu XJ, Shukla SA, Qian ZR, Cohen O, Nishihara R, et al. Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Rep* (2016) 15:857–65. doi: 10.1016/j.celrep.2016.03.075
69. Wang W, Xie G, Ren Z, Xie T, Li J. Gene Selection for the Discrimination of Colorectal Cancer. *Curr Mol Med* (2020) 20:415–28. doi: 10.2174/1566524019666191119105209
70. Sanz-Garcia E, Argiles G, Elez E, Tabernero J. Braf Mutant Colorectal Cancer: Prognosis, Treatment, and New Perspectives. *Ann Oncol* (2017) 28:2648–57. doi: 10.1093/annonc/mdx401
71. Cercek A, Braghiroli MI, Chou JF, Hechtman JF, Kemeny N, Saltz L, et al. Clinical Features and Outcomes of Patients With Colorectal Cancers Harboring Nras Mutations. *Clin Cancer Res* (2017) 23:4753–60. doi: 10.1158/1078-0432.CCR-17-0400
72. Downward J. Targeting Ras Signalling Pathways in Cancer Therapy. *Nat Rev Cancer* (2003) 3:11–22. doi: 10.1038/nrc969
73. Janakiraman M, Vakiani E, Zeng Z, Pratilas CA, Taylor BS, Chitale D, et al. Genomic and Biological Characterization of Exon 4 Kras Mutations in Human Cancer. *Cancer Res* (2010) 70:5901–11. doi: 10.1158/0008-5472.CAN-10-0192
74. Irahara N, Baba Y, Nosho K, Shima K, Yan L, Dias-Santagata D, et al. Nras Mutations Are Rare in Colorectal Cancer. *Diagn Mol Pathol* (2010) 19:157–63. doi: 10.1097/PDM.0b013e3181c93fd1
75. Taniguchi H, Yamazaki K, Yoshino T, Muro K, Yatabe Y, Watanabe T, et al. Japanese Society of Medical Oncology Clinical Guidelines: Ras (Kras/Nras) Mutation Testing in Colorectal Cancer Patients. *Cancer Sci* (2015) 106:324–7. doi: 10.1111/cas.12595
76. Zeinalian M, Emami MH, Naimi A, Salehi R, Hashemzadeh-Chaleshtori M. Immunohistochemical Analysis of Mismatch Repair Proteins in Iranian Colorectal Cancer Patients at Risk for Lynch Syndrome. *Iran J Cancer Prev* (2015) 8:11–7.
77. Petaccia de Macedo M, Melo FM, Ribeiro HSC, Marques MC, Kagohara LT, Begnami MD, et al. Kras Mutation Status Is Highly Homogeneous Between Areas of the Primary Tumor and the Corresponding Metastasis of Colorectal Adenocarcinomas: One Less Problem in Patient Care. *Am J Cancer Res* (2017) 7:1978–89.
78. Chen Y, Wei C, Ning S, Wei w, Liu H, Li Q. Tension Chart. Mutations of KRAS and BRAF Genes in Colorectal Cancer and Their Clinical Significance. *Shandong Med* (2017) 57(16):77–80. doi: 10.3969/j.issn.1002-266x.2017.16.026
79. Lea IA, Jackson MA, Li X, Bailey S, Peddada SD, Dunnick JK. Genetic Pathways and Mutation Profiles of Human Cancers: Site- and Exposure-Specific Patterns. *Carcinogenesis* (2007) 28:1851–8. doi: 10.1093/carcin/bgm176
80. Ye JX, Liu Y, Qin Y, Zhong HH, Yi WN, Shi XY. Kras and Braf Gene Mutations and DNA Mismatch Repair Status in Chinese Colorectal Carcinoma Patients. *World J Gastroenterol* (2015) 21:1595–605. doi: 10.3748/wjg.v21.i5.1595
81. Yang L, Dong D, Fang M, Zhu Y, Zang Y, Liu Z, et al. Can Ct-Based Radiomics Signature Predict Kras/Nras/Braf Mutations in Colorectal Cancer? *Eur Radiol* (2018) 28:2058–67. doi: 10.1007/s00330-017-5146-8
82. Boland CR, Goel A. Microsatellite Instability in Colorectal Cancer. *Gastroenterology* (2010) 138:2073–2087 e2073. doi: 10.1053/j.gastro.2009.12.064
83. Gupta R, Sinha S, Paul RN. The Impact of Microsatellite Stability Status in Colorectal Cancer. *Curr Probl Cancer* (2018) 42:548–59. doi: 10.1016/j.currproblcancer.2018.06.010
84. Cohen R, Pellat A, Boussion H, Svrcek M, Lopez-Trabada D, Trouilloud I, et al. Immunotherapy and Metastatic Colorectal Cancers With Microsatellite Instability or Mismatch Repair Deficiency. *Bull Cancer* (2019) 106:137–42. doi: 10.1016/j.bulcan.2018.09.004
85. Birkenkamp-Demtroder K, Mansilla F, Sorensen FB, Kruhoffer M, Cabezon T, Christensen LL, et al. Phosphoprotein Keratin 23 Accumulates in Mss But Not Msi Colon Cancers *In Vivo* and Impacts Viability and Proliferation *In Vitro*. *Mol Oncol* (2007) 1:181–95. doi: 10.1016/j.molonc.2007.05.005
86. Kok M, Chalabi M, Haanen J. How I Treat Msi Cancers With Advanced Disease. *ESMO Open* (2019) 4:e000511. doi: 10.1136/esmoopen-2019-000511
87. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. Pd-1 Blockade in Tumors With Mismatch-Repair Deficiency. *N Engl J Med* (2015) 372:2509–20. doi: 10.1200/jco.2015.33.15_suppl.lba100
88. 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
89. 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) 44:3755–63. doi: 10.1007/s00261-019-02117-w
90. 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) 26:1633–40. doi: 10.1016/j.acra.2019.02.009

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.

frontiers in Physics

ORIGINAL RESEARCH
published: 12 January 2022
doi: 10.3389/fphy.2021.754248

Deep Learning-Aided Automatic Contouring of Clinical Target Volumes for Radiotherapy in Breast Cancer After Modified Radical Mastectomy

*Jinqiang You[1†], Qingxin Wang[1†], Ruoxi Wang[2], Qin An[2], Jing Wang[1], Zhiyong Yuan[1], Jun Wang[1], Haibin Chen[2], Ziye Yan[2], Jun Wei[2] and Wei Wang[1]**

[1]Department of Radiation Oncology, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Tianjin's Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin, China, [2]Perception Vision Medical Technologies Co. Ltd, Guangzhou, China

OPEN ACCESS

Edited by:
Daniel Rodriguez Gutierrez, Nottingham University Hospitals NHS Trust, United Kingdom

Reviewed by:
Ramy Abdlaty, McMaster University, Canada
Guokai Zhang, Tongji University, China

***Correspondence:**
Wei Wang
weiwang_2@126.com

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

Specialty section:
This article was submitted to Medical Physics and Imaging, a section of the journal Frontiers in Physics

Received: 06 August 2021
Accepted: 08 December 2021
Published: 12 January 2022

Citation:
You J, Wang Q, Wang R, An Q, Wang J, Yuan Z, Wang J, Chen H, Yan Z, Wei J and Wang W (2022) Deep Learning-Aided Automatic Contouring of Clinical Target Volumes for Radiotherapy in Breast Cancer After Modified Radical Mastectomy. Front. Phys. 9:754248. doi: 10.3389/fphy.2021.754248

Purpose: The aim of this study is to develop a practicable automatic clinical target volume (CTV) delineation method for radiotherapy of breast cancer after modified radical mastectomy.

Methods: Unlike breast conserving surgery, the radiotherapy CTV for modified radical mastectomy involves several regions, including CTV in the chest wall (CTV_{cw}), supra- and infra-clavicular region (CTV_{sc}), and internal mammary lymphatic region (CTV_{im}). For accurate and efficient segmentation of the CTVs in radiotherapy of breast cancer after modified radical mastectomy, a multi-scale convolutional neural network with an orientation attention mechanism is proposed to capture the corresponding features in different perception fields. A channel-specific local Dice loss, alongside several data augmentation methods, is also designed specifically to stabilize the model training and improve the generalization performance of the model. The segmentation performance is quantitatively evaluated by statistical metrics and qualitatively evaluated by clinicians in terms of consistency and time efficiency.

Results: The proposed method is trained and evaluated on the self-collected dataset, which contains 110 computed tomography scans from patients with breast cancer who underwent modified mastectomy. The experimental results show that the proposed segmentation method achieved superior performance in terms of Dice similarity coefficient (DSC), Hausdorff distance (HD) and Average symmetric surface distance (ASSD) compared with baseline approaches.

Conclusion: Both quantitative and qualitative evaluation results demonstrated that the specifically designed method is practical and effective in automatic contouring of CTVs for radiotherapy of breast cancer after modified radical mastectomy. Clinicians can significantly save time on manual delineation while obtaining contouring results with high consistency by employing this method.

Keywords: modified radical mastectomy breast cancer surgery, auto-contouring, deep learning, clinical target volume, radiotherapy

1 INTRODUCTION

According to a report from the World Health Organization, breast cancer has overtaken lung cancer as the most prevalent cancer worldwide [1]. Different stages of tumor progression require different types of surgical treatment, including breast-conserving surgery (BCS) and Radical Mastectomy (RM). Modified radical mastectomy (MRM) is widely used in clinical practice for the treatment of breast cancer to ensure surgical efficacy while reducing surgical damage and improving the patient's quality of life [2]. Specifically, MRM has become a cornerstone of breast cancer treatment in China. It involves excising only the mammary gland and clearing the axillary lymph nodes, while preserving the pectoralis major and minor muscles, thereby ensuring postoperative mobility and appearance.

Although MRM is beneficial to patients, it presents a challenge to clinicians in contouring the clinical target volume (CTV) for postoperative radiotherapy because the corresponding CTVs involve several target areas with relatively complex anatomic structures compared with their counterparts in BCS and HS. There are three targets in the CTV delineation for radiotherapy of breast cancer after MRM: CTV in the chest wall (CTV_{cw}), supraclavicular region (CTV_{sc}), and internal mammary lymphatic region (CTV_{im}), among which the position and volume vary significantly. The significant variation between patients and the inter-intra-observation variability [3, 4] also results in highly demanding and time-consuming work for clinicians. Conversely, research has demonstrated that the incidental doses to regions, such as the contralateral breast and thyroid caused by contouring errors can affect patients' quality of life [5–7]. Therefore, there is an urgent need to develop an automatic CTV delineation method for radiotherapy of breast cancer after MRM to reduce the burden on clinicians while improving work efficiency and accuracy.

Currently, most automatic contouring methods are developed for radiotherapy after breast-conserving surgery because they only segment the breast with the mammary gland. For example, atlas-based methods are successful in breast [8] segmentation under the condition that the amount of data and the inter-data variation are small. As the volume of data grows, deep-learning-based approaches have achieved significant development toward remedying the cases with large deformation and other considerable variations and have been adopted by an increasing number of institutes and clinicians.

To the best of our knowledge, this is the first study whose aim is to develop a deep learning-based automatic CTV delineation algorithm for radiotherapy of breast cancer after MRM. In this study, we propose a specifically designed multi-objective segmentation method for automatic CTV delineation for radiotherapy of breast cancer after MRM. An orientation attention mechanism is proposed to tackle the misrecognition of a similar structure between the breast and back sides caused by modified radical surgery. To enable the model to segment the targets correctly with significantly different volumes, an inception block-based multi-scale convolution architecture is constructed to obtain different perception fields and capture the corresponding features. In addition, the model is trained by local dice loss to handle the imbalance between segmentation categories and stabilize the training. Furthermore, three particular data augmentation strategies, namely, attention position variance, deformation simulation, and breast implant simulation, are designed to cope with the problem of data scarcity and differentiation.

The remainder of this paper is organized as follows. 2 introduces related research on automatic breast CTV delineation. 3 Materials and methods describe the specifically designed methods. 4 The experimental results show the quantitative and qualitative results. 5 Discussion and 6 Conclusion and future work.

2 RELATED WORKS

For the past few decades, traditional methods, particularly Atlas-based methods, have been the preferred solution for automatic CTV delineation. Atlas-based approaches perform deformable image registration to match the target and ground truth. Patients are segmented based on an atlas library, and the most anatomically similar will be selected as the target to be transformed into the same coordinate space as the input data. Anders et al. [9] and Velker et al. [10] collected 9 and 124 cases to build a library for breast cancer. The method proposed by Velker achieved good performance on structured CTVs, such as breast and chest wall, with Dice similarity coefficient (DSC) values of 0.87 and 0.89 for left- and right-side breast, respectively.

Atlas-based solutions have been widely utilized in cancer sites, such as the head and neck [11], breast [12], and lungs [13]. However, the performance of these approaches is limited by the degree of deformation, image registration quality, and additional corrections. For instance, for highly variable structures, such as internal mammary nodes, Velker's method achieved poor performance with a DSC of 0.3. In this case, several deep-learning-based approaches have been proposed and have made significant progress in terms of accuracy and consistency [14].

Deep learning methods have demonstrated excellent performance in several fields. Convolutional neural networks (CNNs) have become increasingly irreplaceable in the field of image processing and analysis, producing results by extracting and learning the features from well-organized training data. Deep learning-based semantic segmentation is a suitable solution for automatic CTV delineation. Min et al. [15] proposed a deep learning-based breast segmentation algorithm (a 3D fully convolutional DesnseNet) and compared its performance with the aforementioned atlas-based segmentation methods. The comparison results demonstrated that the deep learning method performed more consistently and robustly on the majority of structures. In addition to the segmentation accuracy, clinicians are concerned with the inference speed of the algorithms because the produced segmentation results still require manual correction. To this end, Jan et al. [16] proposed BibNet, a novel neural network built by U-Net [17] with a multi-resolution level processing structure and residual connections, alongside a full-image processing strategy to increase the

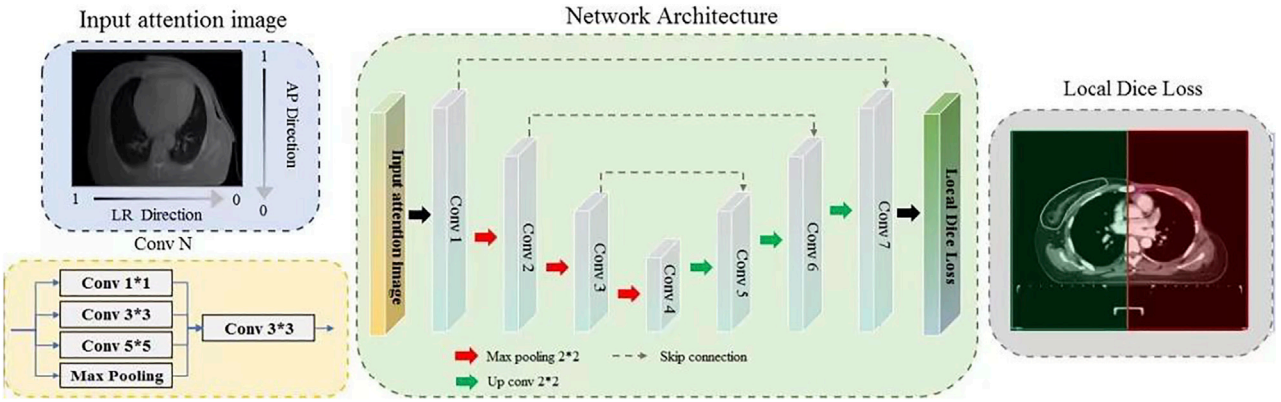


FIGURE 1 | Illustration of the specifically designed deep-learning based multi-objective segmentation method for the automatic delineation of CTVs for Radiotherapy after Modified Radical Mastectomy. Input attention images are obtained by overlapping an anterior-posterior (AP) direction attention map on to input images.

inference speed while improving the segmentation quality. Kuo et al. [18] proposed a deep dilated residual network (DD-ResNet) for auto-segmentation of the clinical target volume for breast cancer radiotherapy, which outperformed deep dilated convolutional neural network (DDCNN) and deep deconvolutional neural network (DDNN). Compared with those references, we use optimizer U-Net to help doctors contouring the region of breast cancer.

3 MATERIALS AND METHODS

3.1 Data Acquisition

The data supporting this study comprised 110 CT scans of patients who underwent modified mastectomy surgery collected from Tianjin Medical University Cancer Institute and Hospital. These patients received adjuvant radiotherapy on the chest wall, supra- and infra-clavicular, and internal mammary lymphatic regions after lumpectomy. Therefore, the CTVs delineated for radiotherapy by an experienced clinician according to the RTOG criteria were set as the ground truth for model training [19]. The CTVs on both the left and right sides were delineated to stabilize model training. Patients with breast implants were also collected in our dataset and extended using the breast implant simulation data augmentation method. The two-dimensional size and thickness of the reconstructed CT images were 512*512 and 5 mm, respectively. The dataset was randomly split into a training set and testing set with 82 cases and 28 cases, respectively. For the sake of splitting our dataset for training and test purpose, the ratio of training and test set about 3:1, which is slightly higher than the 4:1 for most commonly used, was adopted, accommodating the limited overall sample size, resulting in an adequately sized test set.

3.2 Architecture and Strategies

The architecture of the proposed network is illustrated in **Figure 1**. The input images are preprocessed using a specific orientation attention method before being fed into the network. Each convolution block in the network comprises a inception module, followed by an activation layer and a batch normalization layer. The red arrows symbolize max pooling, whereas the green arrows symbolize transpose convolution. Black arrows indicate the inputs and outputs of the model. Local dice loss is employed to train the model for multi-objective segmentation, followed by a sigmoid activation function to generate the output mask. In this study, we focused on the specific characteristics of CTVs after MRM and designed corresponding solutions to accomplish an automatic contouring task.

The breast on the affected side is excised in MRM with only the pectoralis major and minor muscles preserved, resulting in a flat structure that is similar to the back. In addition, the collected data contained patients with left breast cancer and right breast cancer, and even on both sides; therefore, the model should be encouraged to focus more on the affected side and perform delicate segmentation. To this end, an orientation attention mechanism was designed for preprocessing. Specifically, a direction attention map is calculated based on the formula $AP_i = 1 - i/H$ and $LR_i = 1 - i/W$, where i and H/W are the row/column index and image resolution along the anterior–posterior (AP) and left-right (LR) directions, respectively. The input of the model is the product of the AP and LR direction attention map and the normalized CT image with a range of [−1, 1]. The values on the breast and affected sides in the attention map were set to near 1, whereas the opposite side was set to near 0, thereby assigning higher importance to the breast and affected sides. This can be observed in **Figure 1**; the input attention image has a gray gradient along the vertical and horizontal directions. The darker side is emphasized, thus implicitly promoting breast segmentation.

The segmentation targets of the model contained CTV in the chest wall (CTV_{cw}), supra-clavicular region (CTV_{sc}), and internal mammary lymphatic region (CTV_{im}), which vary greatly in volume. CTV_{cw} and CTV_{sc} have thin and long shapes, whereas CTV_{im} only occupies a small region. This imbalance may confuse the model and reduce segmentation performance, especially for

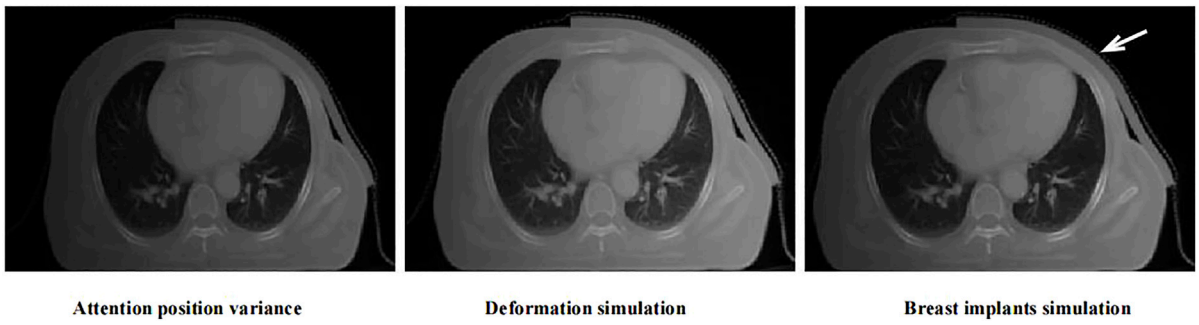


FIGURE 2 | The examples of the proposed data augmentation strategies. The red arrow indicates the position of implanted breast implant.

small targets. Therefore, to enable the model to extract features with different perception fields, thereby performing delicate segmentation of targets with different scales, a network with a multi-scale convolution structure is constructed. This is done by utilizing a refined inception block [20] as a basic convolution element, which can improve the perception field while maintaining minimal pooling operations. Specifically, the input to each convolution block is fed into 1*1, 3*3, and 5*5 convolution layers and a max pooling layer to obtain different perception fields, and the extracted multi-scale features are then fused to model higher-level semantic information. In addition, to overcome the problems of incomplete labels, a novel local loss is introduced for network optimization, where a local mask is calculated based on the label. If parts of the targets are not annotated, the local mask will be initialized by zeros, thereby avoiding optimization of the model with the segmentation error outside the local regions. Benefiting from the larger variation in the breast cancer dataset, this local loss performed excellently in this study. Moreover, the sigmoid activation function is employed in the output layer to produce the probability of the categories of each pixel in the case of overlap among labels.

To cope with individual variations, such as various deformations and cases with breast implants, we designed several targeted data augmentation methods. Three specific data augmentation approaches are exploited to improve data diversity: Attention position variance, deformation simulation, and breast implant simulation. The CT scan center may vary significantly for different patients. Furthermore, the attention map is calculated based on the body center, which may be affected by the coach and other similar materials in the image. Thus, we adjusted the body cancer with limited variation and generated the corresponding input image for training. Breast cancer is a deformable organ, and small deformation is common in breast cancer radiotherapy. Thus, a random elastic deformation vector field was applied to the CT images for deformation augmentation. In particular, a breast implant simulation method was designed for data augmentation. Patients who have undergone breast reconstruction have completely different anatomical structures compared with other patients, which may confuse the model in the training process. In this case, we simulated breast implants in the breast region via morphological processing and density simulations. In the study, We collected CT images from 110 patients with breast cancer for model training and testing. They received radiotherapy from June 6, 2016 to January 31, 2020, at Tianjin Medical University Cancer Hospital. The contouring of target areas have been examined and modified by senior radiotherapy doctors. In order to reduce the influence of individual differences, these CT images are processed by the above data enhancement methods. From **Figure 2**, it can be seen that the simulated images have a relatively similar appearance to the real data. These approaches increase the amount of data, reduce overfitting, and improve the generalization performance of the model.

3.3 Evaluation Metrics

To evaluate this method, the DSC was employed as the quantitative metric, which is defined as the overlap between the segmented mask and the manually labeled mask, witch labled by experienced radiologists. The DSC formula is shown in **Eq. 1**, where A denotes the ground truth, and B denotes the predicted results. Therefore, a higher DSC indicates a more precise segmentation performance.

$$DSC = \frac{2|A \cap B|}{|A| \cup |B|} \tag{1}$$

In some cases, more attention should be paid to segmentation boundaries. Therefore, the Hausdorff distance (HD) and average symmetric surface distance (ASSD) were calculated to evaluate the segmentation performance on boundaries. HD measures the surface distance between two point sets X and Y, as defined by **Eq. 2**. ASSD is the average of all the distances from points on the boundary of the predicted results to the boundary of the ground truth, which is calculated by **Eq. 3**.

$$HD = max\left\{max_{x\in X} min_{y\in Y} d(x,y), max_{y\in Y} min_{x\in X} d(x,y)\right\} \tag{2}$$

$$ASSD = \frac{\sum_{x\in X} min_{y\in Y} d(x,y) + \sum_{y\in Y} min_{x\in X} d(y,x)}{len(X) + len(Y)} \tag{3}$$

where $len(X)$ and $len(Y)$ represent the total number of pixels in the boundary X and boundary Y respectively.

Although the above metrics could provide a scientific assessment of the proposed segmentation method, they are not reliable enough to evaluate the significance of clinical practice

TABLE 1 | Quantitative evaluation of the proposed method and U-Net on CTV_{cw}, CTV_{im}, and CTV_{sc} in terms of DSC, HD and ASSD. The *p* value smaller than 0.05 indicates that there are significant differences between the two approaches.

	DSC		HD		ASSD	
Structures	**U-net**	**Proposed**	**U-net**	**Proposed**	**U-net**	**Proposed**
CTV_{cw}	0.79 ± 0.12	0.92 ± 0.04	13.97 ± 13.33	5.36 ± 3.98	4.7 ± 6.07	1.98 ± 3.15
CTV_{im}	0.66 ± 0.12	0.74 ± 0.09	6.24 ± 5.86	3.86 ± 2.60	1.39 ± 1.38	0.80 ± 0.60
CTV_{sc}	0.60 ± 0.18	0.76 ± 0.10	14.76 ± 11.35	5.67 ± 5.47	3.36 ± 5.36	1.10 ± 0.64
Mean	0.69 ± 0.14	0.81 ± 0.08	11.66 ± 11.18	4.96 ± 3.95	3.15 ± 4.53	1.29 ± 1.41
p value	0.0001		0.0019		0.0015	

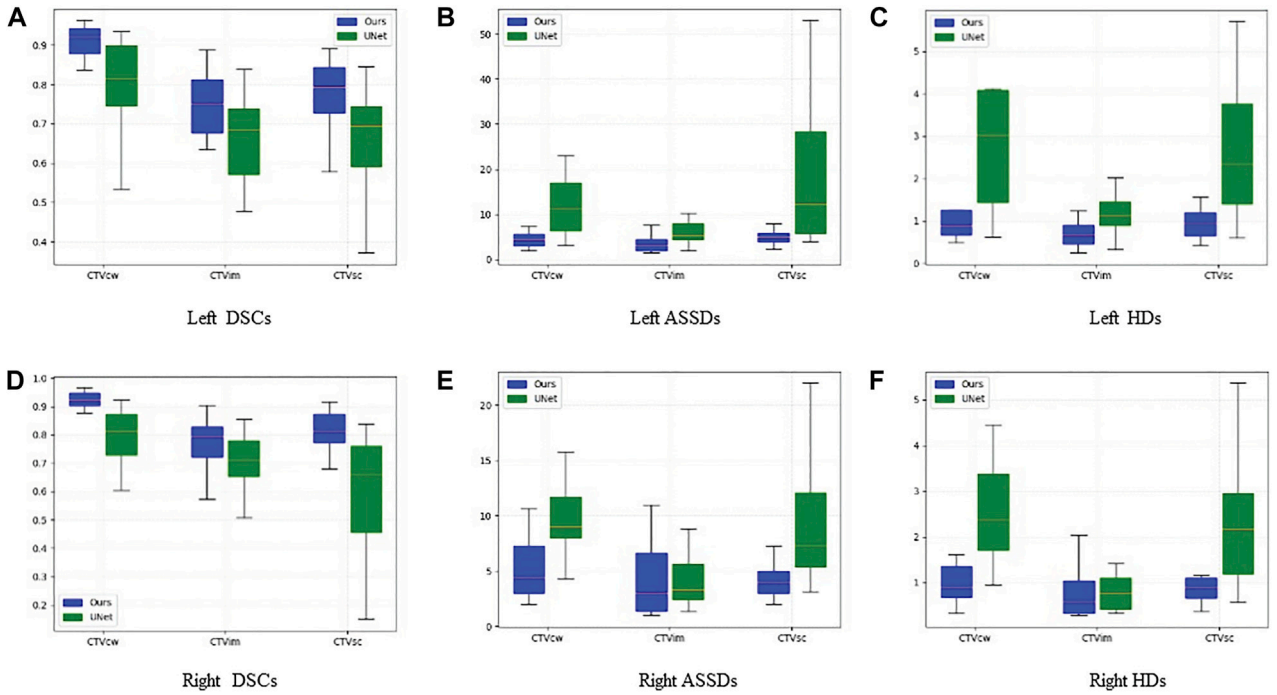


FIGURE 3 | Box-plots of DSC, HD and ASSD in left CTVs and right CTVs on the test set using the gold standard as reference. Blue means our method result. Green means U-Net's result. By comparison, we can see that the effect of blue is much better than that of green. Details are given below.

[21]. To this end, we conducted a user study to obtain a practical assessment by three experienced radiologists.

3.4 Statistical Analysis

A paired *t*-test was conducted to verify the statistical difference between the quantitative evaluation results of the proposed method and other approaches. The test was also performed on the clinicians' scores. A *p value* of less than 0.05 can be regarded as a significant difference between the proposed method and baseline approaches.

4 RESULTS

4.1 Segmentation Performance

Table 1 presents the quantitative evaluation results of the proposed method and the baseline (U-Net) in terms of DSC, HD, and ASSD. It is observed that the proposed method achieved a mean DSC of 0.92 with standard deviation of 0.04 for CTV_{cw}, a mean DSC of 0.74 with standard deviation of 0.09 for CTV_{im}, and a mean DSC of 0.76 with standard deviation of 0.10 for CTV_{sc}. The average DSC over all categories of the proposed method is 0.81, which outperformed the baseline significantly. The *p* value of 0.0001 also demonstrated the significant difference between the two methods. **Figures 3A,B** show the proposed method has larger inter-subject variations in the left CTVs.

The HD and ASSD evaluations illustrated that the proposed method produced smaller surface discrepancies compared with U-Net in all the CTVs. **Figures 3B,C,E,F** revealed that the proposed method tends to generate segmentation results with quite small inter-subject diversity compared with U-Net, thereby demonstrating the inference quality and the robustness of the proposed method.

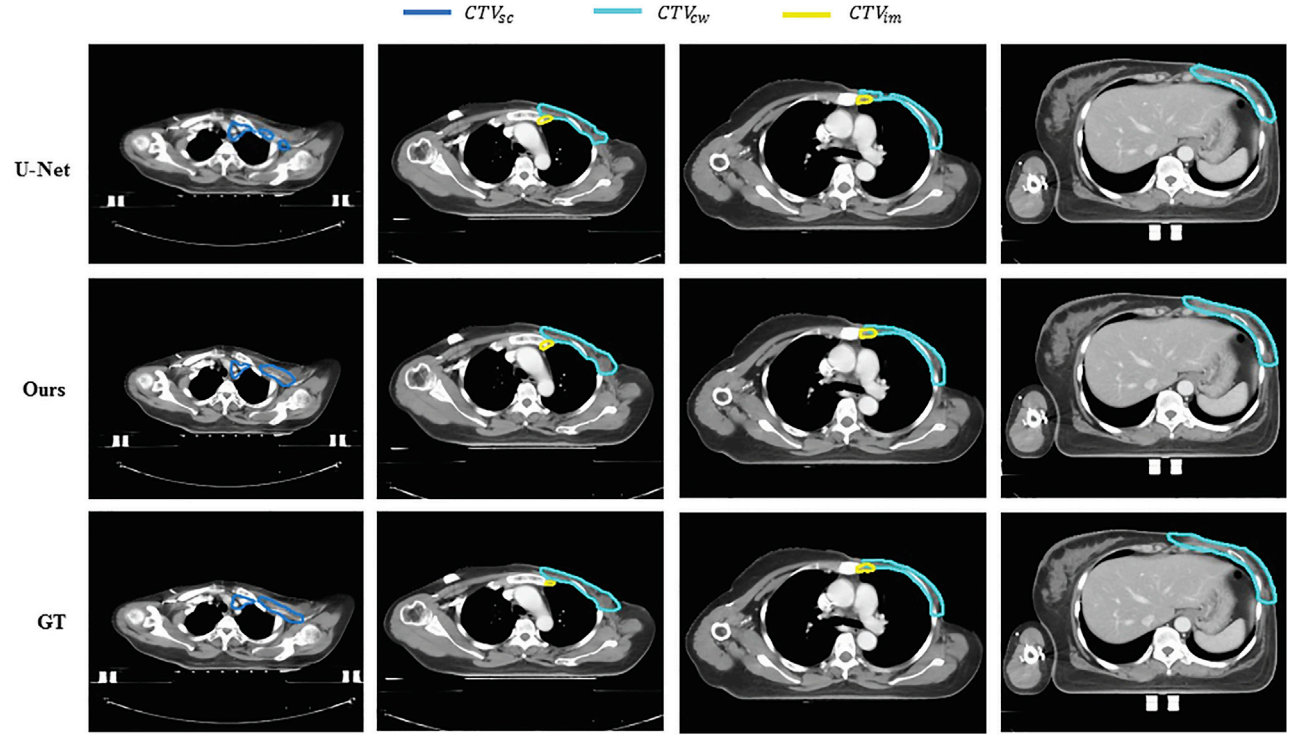


FIGURE 4 | Examples of segmentation results of U-Net and the proposed method against gold standard for the affected side. Different colors represent different segmentation targets. The first row is the result of U-Net,the second row is the result of our method, the third row is the groundtruth of the images. And the different colors represent dfferent segmentation targets. Blue meas the supra-clavicular region, yellow means internal mammary lymphatic region (CTV im), another means CTV in the chest wall (CTV cw).

Specifically, our method can produce significantly better result with small inter-subject diversity compared with U-Net on CTV_{cw} and CTV_{sc}, because the multi-scale convolution module enables the model to extract sufficient features to segment targets with complex structure, such as CTV_{cw} and CTV_{sc}. As for targets with small volume like CTV_{im}, the proposed method can also produce precise results by utilizing receptive fields with different scale.

Figures 4, **5** compare the segmentation results of U-Net, the proposed method with the manual segmentation on the cancer affected side and the contralateral side. The CTV in the chest wall (CTV_{cw}) has an anatomically different structure on the affected side and the contralateral side because the mammary gland is removed. The results produced by U-Net suffer from a moderate degree of under-segmentation and holes in targets, which is not acceptable clinically. It can be seen that our proposed method achieved closer results to the gold standard in terms of shape, location, and volume than those of the counterpart of U-Net.

4.2 Ablation Study

In this section, we explored the importance and effectiveness of the orientation attention mechanism and breast implant simulation.

4.2.1 Importance of Orientation Attention

The input orientation attention strategy is expected to encourage the model to distinguish the breast region from the back region in the transverse CT slices and perform segmentation. To verify the effectiveness of this strategy, we conducted an ablation experiment by removing the input orientation attention mechanism and compared the segmentation performance. **Figure 6** shows the segmentation results for a test case generated by models with and without input orientation attention preprocessing. The model trained without the orientation attention mechanism incorrectly performs segmentation on the back region, whereas the targets are correctly segmented by the model trained with the orientation attention strategy.

4.2.2 Importance of Breast Implants Simulation

`Only six patients with breast implants were included in the training data, which was extremely imbalanced for training. The different anatomical structures between patients with and without breast implants can confuse the model during the training process. Thus, we expect that the proposed breast implant simulation can handle this problem by increasing the amount of data with breast implants. We investigated the

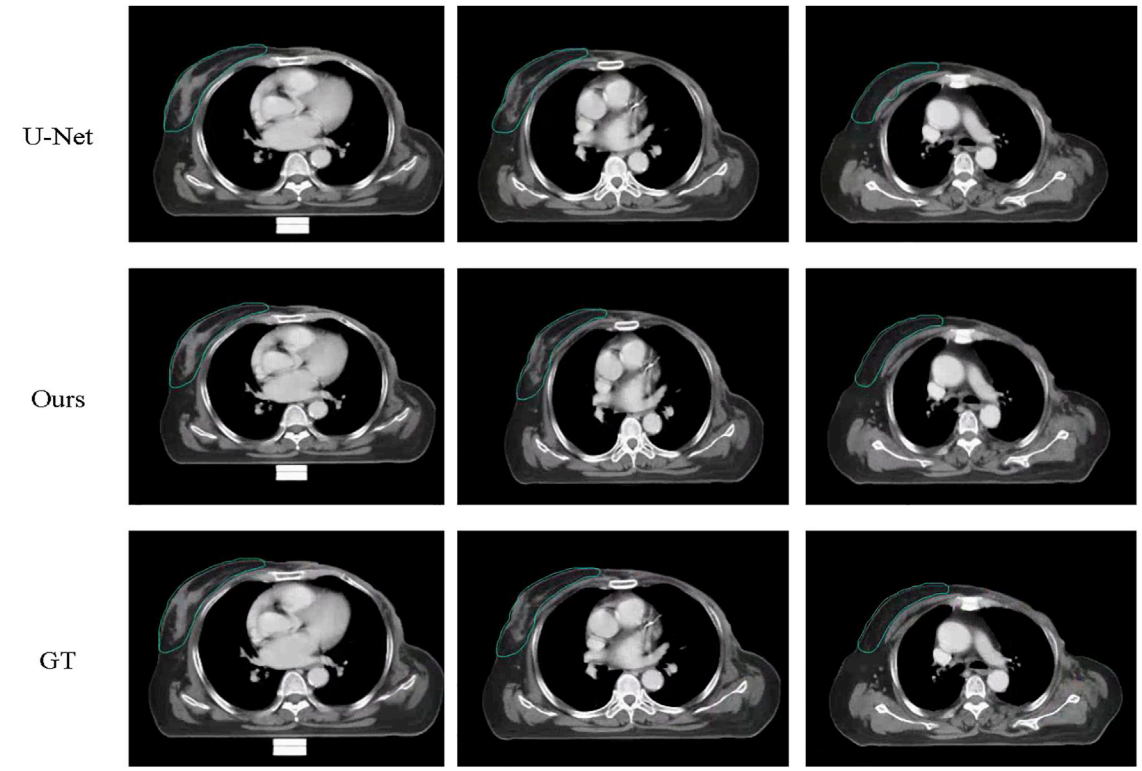


FIGURE 5 | Examples of segmentation results of U-Net and the proposed method against gold standard on the contralateral side. The first image is the result of U-Net, the seconf is the ground truth while the third is our proposed method.

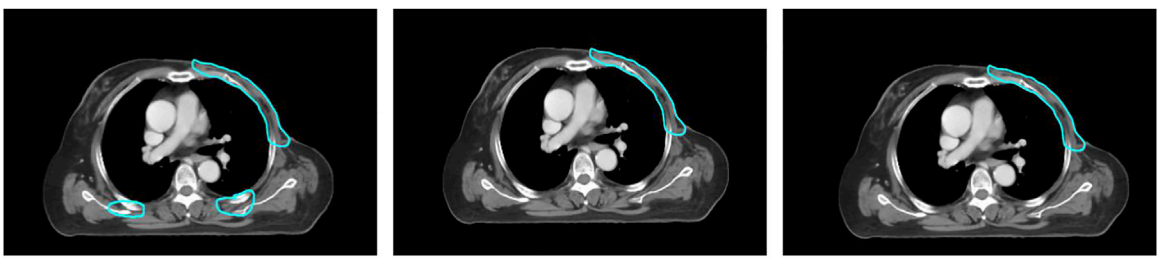


FIGURE 6 | The illustration of the usefulness of the proposed method in recognizing the breast side correctly. The U-Net model incorrectly segments dorsally structurally similar regions as target CTVs, while the proposed method successfully identifies the breast side and segments the target CTVs.

importance of breast implant simulation by training the model with only the original data. **Figure 7** presents the segmentation results for the case of breast implants. It was found that the trained model without specific data augmentation was confused by processing cases with breast implants, resulting in poor segmentation results. The proposed method is well suited to cases with breast implants, whereas U-Net performs poorly.

4.3 Timing Performance

The time required to train the proposed model on two GTX 1080 GPUs was approximately 24 h. By utilizing the automatic segmentation method, the time required to delineate a breast CTV of a patient is drastically reduced from approximately 40 min (manual delineating) to several seconds. Even if some special cases need doctors correct the delineating result maunally, the completion of a breast CTV contouring can be

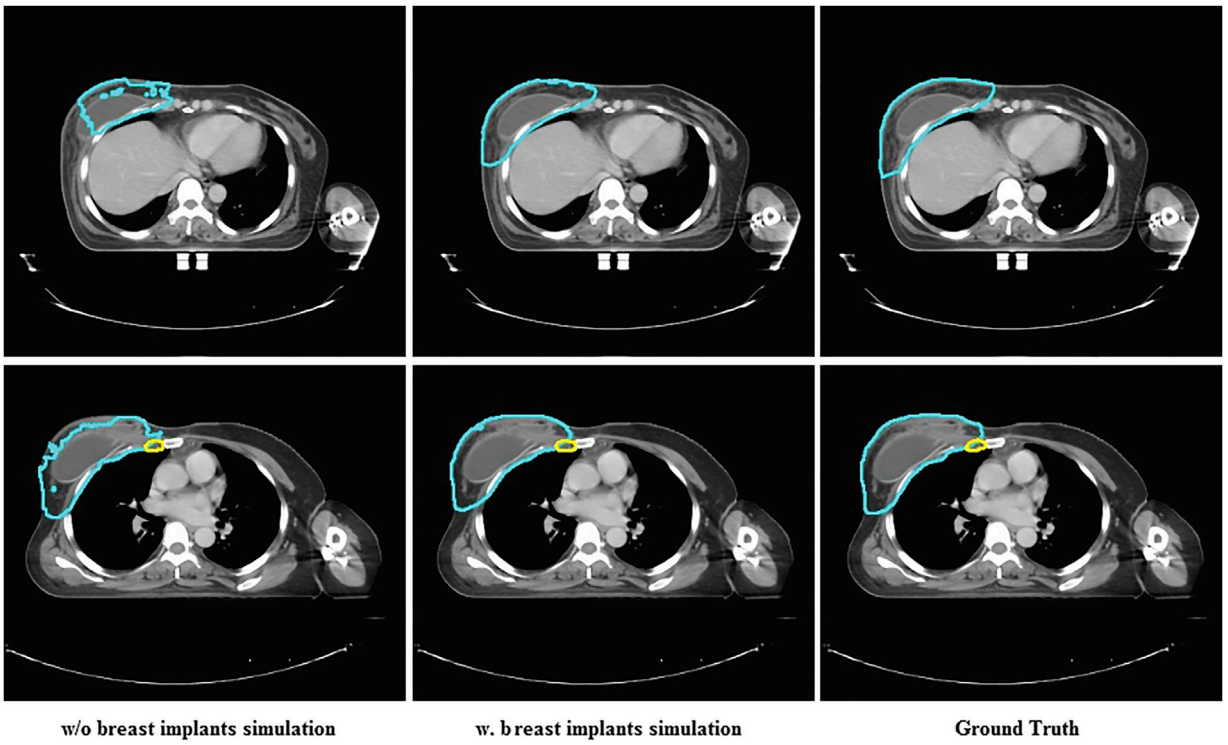


FIGURE 7 | The comparison between the segmentation results of U-Net and the proposed method and ground truth on the case with breast implants.

controlled within 10 min with the manual correction time, demonstrating the feasibility and effectiveness of the proposed approach.

5 DISCUSSION

In this study, we proposed a specifically-designed deep learning-based framework for automatic contouring of 10 targets in CT scans for modified mastectomy RT. The experiment results indicate that our method performed well, exhibiting excellent agreement with the CTVs that were manually delineated by clinicians. In detail, both quantitative and qualitative evaluations demonstrated the feasibility of the proposed methods in contouring CTVs for modified mastectomy RT. The orientation attention provides reliable supervision for the model to recognize the breast and affect sides in CT images. Different from simply applying a deep learning-based segmentation network for automatic CTVs contouring, we conducted statistic analysis of the CTVs in modified mastectomy surgery-based radiotherapy and designed the network according to the statistical characteristics. The multi-scale convolutional structure constructed by refined inception module increases both the width of the network and the adaptability of the network to scales, thereby producing delicate segmentation results of targets with different volume. Besides, the local loss drives the optimization for all of the targets even in the cases with labels missing.

Considering the scarcity of data volume and the variability among data, we designed three data enhancement methods for data expansion to improve the generalization performance of the model while avoiding overfitting. Data augmentation is particularly essential for medical-related researches, since it takes long and a lot to collect medical data. Apart from the attention position and general deformation simulation, we particularly designed the breast implants simulation method to increase the number of cases with breast implants. The breast anatomical structure of patients with breast implants is completely different from the patients without. So a small amount of data with breast implants can affect the model training, resulting in the model not converge. Through the breast implants simulation, the problem of category imbalance is alleviated and the model is able to generate more accurate segmentation results for patients with breast implants.

Although deep learning solutions performs well in producing contouring results for RT (RT is a file that stores the coordinates of the region of interest), the nature of deep learning makes it sort of disputable [22] because it learns how to segment only based on the ground truth delineated by one clinician. Radiotherapy requires clinical input and creativity in terms of science and art [19]. The delineation results of the same case can vary between clinicians, and it is sometimes difficult to determine which one is optimal. Therefore, the ground truth used for training the deep learning model also should have diversity. The reinforcement learning provides a potential way to enable the DL model to learn how to optimally segment targets.

Manual delineation of OARs and CTVs for RT is a laborious task for clinicians, which requires not only experience but also physical exertion. Repetitive work for long periods can lead to reduced productivity and even errors on the part of clinicians [2]. In this case, automatic segmentation algorithms serve as a useful tool for reducing the workload of clinicians and producing highly consistent results. A previous study illustrated that atlas-based automatic segmentation (ABAS) for loco-regional RT of breast cancer reduced the time needed for manual delineation by 93% (before correction) and 32% (after correction) [23]. Our method reduced the time required for contouring from 40 min (manual) to 10 min (automatic) on average. With the assistance of deep learning-based auto-segmentation, radiation oncologists can work more efficiently.

To evaluate the segmentation results more carefully and efficiently, and to explore the detailed gap between the deep learning-based automatic contouring algorithm and manual contouring, we used both HD and ASSD to evaluate the performance of the contouring results on the edges and surfaces. In this case, we further proved the level of advancement of the proposed method on 3D level rather than the 2D level only. **Table 1** and **Figure 3** illustrate that the proposed method can produce segmentation results with better agreement with the manually delineated structures in terms of region and surface.

This study has several limitations. First, we conducted this research in a single center with limited sample size and diversity, which will impose a challenge on the generalization power of the proposed model. The well-performing model may produce unacceptable segmentation results when applied to other centers owing to the variance between the data. Therefore, we plan to validate the proposed method using data from other institutions. Second, the accuracy and pattern of the segmentation results depend heavily on the manual annotations used for training, which can be both advantageous and disadvantageous. As aforementioned, the model can be trained using a homogeneous gold standard created by a single clinician. However, there is no 100% gold standard in clinical settings, as inter-intra-observer variations always exist. Thus, further studies should be conducted to evaluate the generalization of the gold standard created by multiple clinicians. Additionally, it may be more favorable if the OARs are segmented simultaneously. By extracting corresponding features and segment-related organs and tissues, the model can obtain a better perception of the target region. Specifically, the OARs that are most helpful for segmenting target CTVs in the breast region still need to be considered. For instance, the importance of coronary vessels has been increasingly acknowledged.

6 CONCLUSION

Auto-contouring of the CTVs can relieve clinicians from tedious contouring work while improve the consistency and reliability of radiotherapy. In this study, a specifically designed deep learning-based segmentation method was developed to delineate CTVs for modified mastectomy radiotherapy. Qualitative and quantitative evaluations demonstrated the outstanding performance of the proposed method. The method can also handle cases with breast implants and large shape variability. The user study also suggests that the proposed method is practical and beneficial to clinical work by significantly saving time and improving the consistency of decisions.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because data security requirement of our hospital. Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Medical Ethics Committee of Tianjin Medical University Cancer Institute and Hospital. The ethics committee waived the requirement of written informed consent for participation. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. 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 (Grant No. 81872472).

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 A Cancer J Clin* (2021) 71: 209–49. doi:10.3322/caac.21660
2. Abdlaty R, Doerwald L, Hayward J, Fang Q Radiation-Therapy-Induced Erythema: Comparison of Spectroscopic Diffuse Reflectance Measurements and Visual Assessment. In: Medical Imaging 2019: Image Perception, Observer Performance, and Technology Assessment. Bellingham, WA: International Society for Optics and Photonics (2019) 10952: 109520H.
3. Li XA, Tai A, Arthur DW, Buchholz TA, Macdonald S, Marks LB, et al. Variability of Target and normal Structure Delineation for Breast Cancer

Radiotherapy: An Rtog Multi-Institutional and Multiobserver Study. *Int J Radiat Oncol Biol Phys* (2009) 73:944–51. doi:10.1016/j.ijrobp.2008.10.034
4. Louie AV, Rodrigues G, Olsthoorn J, Palma D, Yu E, Yaremko B, et al. Inter-observer and Intra-observer Reliability for Lung Cancer Target Volume Delineation in the 4d-Ct Era. *Radiother Oncol* (2010) 95:166–71. doi:10.1016/j.radonc.2009.12.028
5. Gross JP, Lynch CM, Flores AM, Jordan SW, Helenowski IB, Gopalakrishnan M, et al. Determining the Organ at Risk for Lymphedema after Regional Nodal Irradiation in Breast Cancer. *Int J Radiat Oncol Biol Phys* (2019) 105:649–58. doi:10.1016/j.ijrobp.2019.06.2509
6. Stovall M, Smith SA, Langholz BM, Boice JD, Shore RE, Andersson M, et al. Dose to the Contralateral Breast from Radiotherapy and Risk of Second Primary Breast Cancer in the Wecare Study. *Int J Radiat Oncol Biol Phys* (2008) 72:1021–30. doi:10.1016/j.ijrobp.2008.02.040
7. Yaney A, Ayan A, Pan X, Jhawar S, Healy E, Beyer S, et al. Dosimetric Parameters Associated with Radiation-Induced Esophagitis in Breast Cancer Patients Undergoing Regional Nodal Irradiation. *Radiol Oncol* (2021) 155: 167–73. doi:10.1016/j.radonc.2020.10.042
8. Bell LR, Dowling JA, Pogson EM, Metcalfe P, Holloway L Atlas-based Segmentation Technique Incorporating Inter-observer Delineation Uncertainty for Whole Breast. *J Phys Conf Ser* (2017) 777:012002. doi:10.1088/1742-6596/777/1/012002
9. Anders LC, Stieler F, Siebenlist K, Schäfer J, Lohr F, Wenz F Performance of an Atlas-Based Autosegmentation Software for Delineation of Target Volumes for Radiotherapy of Breast and Anorectal Cancer. *Radiother Oncol* (2012) 102: 68–73. doi:10.1016/j.radonc.2011.08.043
10. Velker VM, Rodrigues GB, Dinniwell R, Hwee J, Louie AV Creation of Rtog Compliant Patient Ct-Atlases for Automated Atlas Based Contouring of Local Regional Breast and High-Risk Prostate Cancers. *Radiat Oncol* (2013) 8:188. doi:10.1186/1748-717x-8-188
11. Lee H, Lee E, Kim N, Kim Jh., Park K, Lee H, et al. Clinical Evaluation of Commercial Atlas-Based Auto-Segmentation in the Head and Neck Region. *Front Oncol* (2019) 9:239. doi:10.3389/fonc.2019.00239
12. Reed VK, Woodward WA, Zhang L, Strom EA, Perkins GH, Tereffe W, et al. Automatic Segmentation of Whole Breast Using Atlas Approach and Deformable Image Registration. *Int J Radiat Oncol Biol Phys* (2009) 73: 1493–500. doi:10.1016/j.ijrobp.2008.07.001
13. Pirozzi S, Horvat M, Piper J, Nelson A SU-E-J-106: Atlas-Based Segmentation: Evaluation of a Multi-Atlas Approach for Lung Cancer. *Med Phys* (2012) 39: 3677. doi:10.1118/1.4734942
14. Hoffman R, Jain AK Segmentation and Classification of Range Images. *IEEE Trans Pattern Anal Mach Intell* (1987) 9:608–20. doi:10.1109/TPAMI.1987.4767955
15. Min SC, Byeong SC, Seung YC, Nalee K, Jaeh ee C, Yong BK, et al. Clinical Evaluation of Atlas- and Deep Learning-Based Automatic Segmentation of Multiple Organs and Clinical Target Volumes for Breast Cancer. *Radiother Oncol* (2020) 677.
16. Schreier J, Attanasi F, Laaksonen H A Full-Image Deep Segmenter for Ct Images in Breast Cancer Radiotherapy Treatment. *Front Oncol* (2019) 9:677. doi:10.3389/fonc.2019.00677
17. Ronneberger O, Fischer P, Brox T U-net: Convolutional Networks for Biomedical Image Segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham, switzerland: Springer (2015). p. 234–41. doi:10.1007/978-3-319-24574-4_28
18. Men K, Zhang T, Chen X, Chen B, Tang Y, Wang S, et al. Fully Automatic and Robust Segmentation of the Clinical Target Volume for Radiotherapy of Breast Cancer Using Big Data and Deep Learning. *Physica Med* (2018) 50:13–9. doi:10.1016/j.ejmp.2018.05.006
19. Abdlaty R, Doerwald-Munoz L, Madooei A, Sahli S, Yeh S, Zerubia J, et al. Hyperspectral Imaging and Classification for Grading Skin Erythema. *Front Phys* (2018) 6. doi:10.3389/fphy.2018.00072
20. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper with Convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; Boston, MA (2015). p. 1–9. doi:10.1109/cvpr.2015.7298594
21. Bertels J, Eelbode T, Berman M, Vandermeulen D, Maes F, Bisschops R, et al. Optimizing the Dice Score and Jaccard index for Medical Image Segmentation: Theory and Practice. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Cham, Switzerland: Springer (2019) 92–100. doi:10.1007/978-3-030-32245-8_11
22. Wong J, Fong A, Mcvicar N, Smith SL, Alexander AS Comparing Deep Learning-Based Auto-Segmentation of Organs at Risk and Clinical Target Volumes to Expert Inter-observer Variability in Radiotherapy Planning. *Int J Radiat OncologyBiologyPhysics* (2019) 105:S70–S71. doi:10.1016/j.ijrobp.2019.06.523
23. Eldesoky AR, Yates ES, Nyeng TB, Thomsen MS, Nielsen HM, Poortmans P, et al. Internal and External Validation of an ESTRO Delineation Guideline - Dependent Automated Segmentation Tool for Loco-Regional Radiation Therapy of Early Breast Cancer. *Radiother Oncol* (2016) 121:424–30. doi:10.1016/j.radonc.2016.09.005

Conflict of Interest: Authors RW, QA, HC, ZY and JW were employed by the company Guangzhou Perception Vision Medical Technologies 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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 21 January 2022
doi: 10.3389/fonc.2022.811767

A Magnetic Resonance Imaging-Based Radiomic Model for the Noninvasive Preoperative Differentiation Between Transitional and Atypical Meningiomas

Jing Zhang[1†], Guojin Zhang[2†], Yuntai Cao[3†], Jialiang Ren[4†], Zhiyong Zhao[5], Tao Han[5], Kuntao Chen[1] and Junlin Zhou[5*]*

[1] Department of Radiology, The Fifth Affiliated Hospital of Zunyi Medical University, Zhuhai, China, [2] Department of Radiology, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, China, [3] Department of Radiology, Affiliated Hospital of Qinghai University, Xining, China, [4] Department of Pharmaceuticals Diagnosis, GE Healthcare, Beijing, China, [5] Department of Radiology, Lanzhou University Second Hospital, Lanzhou, China; Key Laboratory of Medical Imaging of Gansu Province, Lanzhou, China

Preoperative distinction between transitional meningioma and atypical meningioma would aid the selection of appropriate surgical techniques, as well as the prognosis prediction. Here, we aimed to differentiate between these two tumors using radiomic signatures based on preoperative, contrast-enhanced T1-weighted and T2-weighted magnetic resonance imaging. A total of 141 transitional meningioma and 101 atypical meningioma cases between January 2014 and December 2018 with a histopathologically confirmed diagnosis were retrospectively reviewed. All patients underwent magnetic resonance imaging before surgery. For each patient, 1227 radiomic features were extracted from contrast-enhanced T1-weighted and T2-weighted images each. Least absolute shrinkage and selection operator regression analysis was performed to select the most informative features of different modalities. Subsequently, stepwise multivariate logistic regression was chosen to further select strongly correlated features and build classification models that can distinguish transitional from atypical meningioma. The diagnostic abilities were evaluated by receiver operating characteristic analysis. Furthermore, a nomogram was built by incorporating clinical characteristics, radiological features, and radiomic signatures, and decision curve analysis was used to validate the clinical usefulness of the nomogram. Sex, tumor shape, brain invasion, and four radiomic features differed significantly between transitional meningioma and atypical meningioma. The clinicoradiomic model derived by fusing the above features resulted in the best discrimination ability, with areas under the curves of 0.809 (95% confidence interval, 0.743-0.874) and 0.795 (95% confidence interval, 0.692-0.899) and sensitivity values of 74.0% and 71.4% in the training and validation cohorts, respectively. The clinicoradiomic model demonstrated good performance for the differentiation between transitional and atypical meningioma. It is a

OPEN ACCESS

Edited by:
Antonio Napolitano, Bambino Gesù Children's Hospital (IRCCS), Italy

Reviewed by:
Subhas K. Konar, National Institute of Mental Health and Neurosciences (NIMHANS), India
Rosario Sarabia, Hospital Universitario Río Hortega, Spain
Santiago Cepeda, Rio Hortega University Hospital, Spain in collaboration with reviewer RS
Yang Zhang, Rutgers Cancer Institute of New Jersey, United States

***Correspondence:**
Junlin Zhou
ery_zhoujl@lzu.edu.cn
Kuntao Chen
chenkunt2021@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: *09 November 2021*
Accepted: *04 January 2022*
Published: *21 January 2022*

Citation:
Zhang J, Zhang G, Cao Y, Ren J, Zhao Z, Han T, Chen K and Zhou J (2022) A Magnetic Resonance Imaging-Based Radiomic Model for the Noninvasive Preoperative Differentiation Between Transitional and Atypical Meningiomas. Front. Oncol. 12:811767. doi: 10.3389/fonc.2022.811767

quantitative tool that can potentially aid the selection of surgical techniques and the prognosis prediction and can thus be applied in patients with these two meningioma subtypes.

Keywords: meningioma, clinical decision-making, neoplasm grading, radiomics, retrospective studies

INTRODUCTION

Meningiomas are the most common primary intracranial tumors in adults, accounting for 36.7% of all intracranial tumors (1). According to the latest 2016 edition of the World Health Organization (WHO) classification of central nervous system tumors (2), meningiomas have been classified into 3 grades and 15 different subtypes. Among these different subtypes, transitional meningioma (TM) is a common benign meningioma (WHO grade I), whereas atypical meningioma (AM) is an uncommon tumor of intermediate grade between benign and malignant forms (WHO grade II). Pathologically, TM is characterized by the transitional morphological manifestation between endothelial meningiomas and fibrous meningiomas (3). AM is defined as a tumor with increased mitotic activity (≥4 mitoses per 10 high-power fields), brain invasion, and at least three of the following minor criteria: increased cellularity, high nucleus-to-cytoplasm ratio, prominent nucleoli, sheet-like architecture, and spontaneous necrosis foci (4, 5).

According to the European Association of Neuro-Oncology (EANO) guidelines, magnetic resonance imaging (MRI) is the main method used in the provisional diagnosis of meningiomas (6). At present, several studies have explored imaging features to assess the tumor grade, and some imaging features (such as tumor heterogeneity, shape, and tumor-brain interface) may be used as predictive factors to discriminate between tumors of different grades (7–9). Zhang et al. used MRI features to distinguish some subtypes of WHO grade I meningiomas (angiomatous, meningothelial, fibroblastic, and psammomatous meningiomas) and found that angiomatous and meningothelial meningiomas were the most easily identifiable subtypes (10). However, current image-guided evaluation depends on the experience of radiologists, which is non-specific and highly subjective. Recently, our previous study (11) has shown that among meningiomas, WHO grade I TM and WHO grade II AM are more aggressive than other subtypes because the frequency of brain invasion in these two tumors was much higher than in other subtypes. This study showed that TM was more aggressive than other subtypes of WHO grade 1 meningioma, and its biological behavior is close to that of atypical meningioma. Another study observed that several imaging characteristics, such as irregular tumor shape, heterogeneous contrast enhancement, and peritumoral edema were identified as predictors of brain invasion (12). The above research suggests that TM and AM may be similar in their imaging presentation, although the reported data on TM remain scarce, especially regarding its imaging characteristics. However, the clinical treatment plan and prognosis of these two tumors are significantly different due to their different grades. According to EANO guidelines for the treatment of meningiomas, the diagnosis of WHO grade II meningioma (such as AM) implies an increased risk of recurrence, requiring shorter control intervals (every 6 months instead of annually) than in WHO grade I TM (6). Han et al. reported that TMs can be treated with either surgery or external beam radiation, AMs often require a combination of the two modalities (13). The choice of surgical technique may be different. Because AM is more prone to invasive growth and recurrence. Whether to expand the scope of surgical resection, application of intraoperative navigation and preoperative blood preparation, this is closely related to the size of the tumor (AM tends to be slightly larger than TM) and tumor surrounding tissues Moreover, it has been established that higher tumor grades are associated with worse prognosis; higher grades indicate reduced survival and higher rates of tumor recurrence (14). Therefore, precise distinction between TM and AM before surgery is desirable.

Given the above reasons, it is necessary to explore the imaging differences between AM and TM. Radiomic analysis is a reliable tool that can quantify high-dimensional tumor features that cannot be observed with the naked eye, such as intensity, texture, and shape features (15, 16). In recent years, radiomic analysis has rapidly transformed the field of medical imaging analysis, since it provides more stable results and is an objective rather than a subjective assessment. Several studies have demonstrated the applications of radiomics in meningiomas, such as the characterization of the grade and histological subtype, the prediction of brain invasion and recurrence-free survival, and the identification of differential diagnoses in meningioma (11, 17–20). These studies show that the MRI-based radiomics may also be a method for discriminations between AM and TM.

To the best of our knowledge, this is the first study to differentiate TM from AM based on texture feature or radiomic analysis. Therefore, our study aimed first to identify MR and radiomic features that are associated with these two tumors from two MRI modalities [T2-weighted (T2) and T1-weighted post-contrast (T1C)]; second, to combine these two modalities generating a radiomic signature; and third, to build a nomogram fusing clinical factors, MR features, and radiomic signatures to differentiate TM from AM in MRIs of patients with suspected meningioma.

MATERIALS AND METHODS

Study Population and Semantic Features

For this retrospective analysis, ethical approval was obtained from the Institutional Review Board of Lanzhou University

Second Hospital, and the requirement for informed consent was waived. In this study, all patients with TM and AM who underwent surgery in our institute between January 2015 and December 2019 were enrolled according to the following inclusion and exclusion criterias. The inclusion criteria were: (a) histological diagnosis of AM or TM, and (b) MRI, including T1C and T2 sequences, performed within 1 week before surgical tumor resection. The exclusion criteria were: (a) cases with motion artefacts that impacted the assessment; (b) incomplete MRI sequences; and (c) treatment such as radiotherapy, chemoradiotherapy, or surgery before surgical tumor resection.

All tumors were resected with the aid of a microscope. Patients with TM and AM were diagnosed according to the pathological findings. Finally, a total of 242 patients (TM: 19 men, 122 women, mean age 52.3 ± 9.2 years; AM: 46 men, 55 women, mean age 51.5 ± 10.3 years) were enrolled. All patients were randomly divided into a training cohort and a validation cohort in a 7:3 ratio. The patient recruitment flowchart is shown in **Figure 1**.

Two radiologists (reader 1 JZ and reader 2 YTC, with 12 and 15 years of experience in brain MRI interpretation, respectively) independently analyzed the MRI characteristics (including tumor location, maximum diameter, tumor shape, tumor border, dural tail sign, peritumoral edema, T2 signal, enhanced features, bone invasion, sinus invasion, and brain invasion). The image analysis was based on clinical experience. Both readers were blinded to all personal information and the histopathological results before analysis. For qualitative data, agreements were reached after discussion between the two in cases of difference of opinions. When the two readers were unsure, reader 3 (ZYZ) with 19 years of experience confirmed the results. For quantitative data, reader 1 measured the maximum diameter three times on the maximum level of the tumor, and calculated the average of three measurements. Reader 2 performed the data measurement in the same way. The final result was the average of the measurement values of two readers to minimize the deviation of the measurement results. Among MR features evaluation, peritumoral edema was evaluated on T2 images according to the standardized visually accessible Rembrandt Images (VASARI; https://wiki.nci.nih.gov/display/CIP/VASARI) feature set. Brain invasion was diagnosed by pathology. Bone invasion assessments were performed by pathology and surgeon assessment intraoperatively. Sinus invasion was evaluated by an intraoperative neurosurgeon as a diagnostic standard.

Image Acquisition, Segmentation, and Normalization

The MRIs were obtained at our institution with 3.0-T scanners (Siemens Verio or Philips Achieva). The MR sequences included T1C and T2 images, and the detailed parameters of each scanner are shown in **Supplementary Table S1**.

Two radiologists (TH and JZ) without prior knowledge of the pathological records manually segmented MR images using the open-source software ITK-SNAP (version: 3.8.0, www.itksnap.org). On both axial T1C and T2 images, the regions of interest (ROIs) of images were manually delineated on each slice of the entire tumor including hemorrhagic, necrotic lesions and without the surrounding brain tissue, and oedema. T2 images were segmented with reference to T1C images for visual

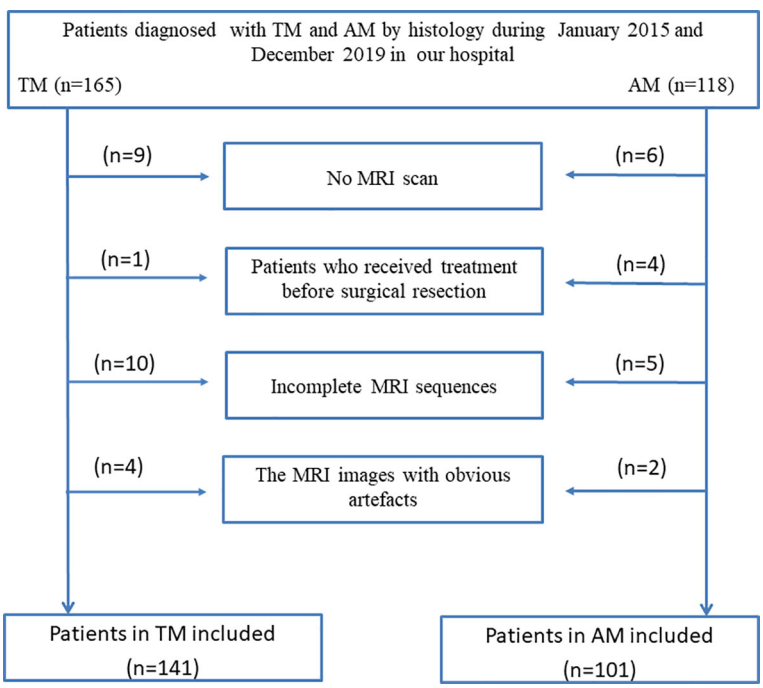


FIGURE 1 | Inclusion and exclusion criteria.

guidance. The segmented tissues on each slice were fused together to generate the volume of interest, as shown in **Supplementary Figure 1**.

To obtain a standard normal distribution of the image intensities, T1C and T2 images were standardized using z-score normalization and resampling after manual segmentation. MR scanners and image segmentation of two additional sets (i.e. two MR scanners set and the re-segmentation set) are described in Supplementary Material.

Feature Extraction and Selection

The PyRadiomics platform was used to extract standardized radiomic features from the T1C and T2 imaging data (21). In this study, feature extraction followed the Image Biomarker Standardization Initiative (IBSI) guideline (22). T1C features were extracted from the volume of interest (VOI) of T1C images, whereas T2 features were extracted from the VOI of T2 images. Finally, a total of 2454 radiomic features were extracted from the VOI of two modalities of the MR images.

For both T1C and T2 features, the least absolute shrinkage and selection operator (LASSO) regression with five-fold cross-validation was conducted to select the radiomic features highly correlated with discrimination of TM and AM (**Supplementary Figure 2**). Features with a *P*-value of less than 0.05 were selected. For clinical factors and MRI features, the correlation between these two factors and discrimination of AM and TM were tested *via* Student's t-test and the chi-square test with the *P*-value set to 0.05. Then, stepwise multivariate logistic regression further selected the most informative features and deleted irrelevant features. Features with a *P*-value of less than 0.05 and preoperative factors were included in the model. Spearman correlation analysis was conducted to examine the correlation between the selected radiomic features and clinicoradiological features to determine whether these features are correlated with each other.

Fusion of Modalities and Radiomic Signature Building

T1C represents the blood supply and the integrity of the blood-brain barrier, whereas T2 is sensitive to peritumoral edema, thus mainly reflecting tissue edema. Therefore, these two modalities were fused by combining the selected radiomic features to increase the performance of the radiomic model. After fusing the modalities, we used stepwise multivariate logistic regression to build a radiomic model discrimination of TM and AM based on the selected radiomic features. The T1C model was built based on T1C features (two features), and the T2 model was built based on T2 features (two features), whereas the fusion model was built based on T1C and T2 fusion features (all four radiomic features). The clinical model was built based on a clinical factor (sex) and MRI features (tumor shape and brain invasion). Thus, the clinicoradiomic model was built by incorporating the clinical factor, MRI features, and the radiomic signature. In the training cohort, the maximum area under the receiver operating characteristic curve (AUC) with three-fold cross-validation determined the final regularization parameter.

Nomogram Building and Validation

Integrated discrimination improvement (IDI) (23) was used to quantify performance improvements. The *P*-values indicated whether the improvement in reclassification was statistically significant after the inclusion of a new factor in the model. In addition, we used the DeLong test to compare the AUC estimates of the performance between different models.

Afterward, a nomogram for clinical usefulness incorporating the radiomic signature and the correlated clinicoradiological features was constructed in the training cohort and validated in the validation cohort. The calibration curves assessed the discrimination ability of the nomogram for the training and test cohorts, and the Hosmer-Lemeshow test evaluated the agreement between the discrimination of TM from AM and the observed outcomes. Then, we used decision curve analysis (DCA) to quantify the net benefits at different threshold probabilities to evaluate the clinical efficacy of the nomogram (24).

Statistical Analysis

In this study, all statistical analyses were performed with R software (version 3.6.4, http://www.Rproject.org). R was also used to assess the prediction models. PyRadiomics was used to extract and select the radiomic features, as well as to build the prediction models. The Spearman correlation test was used to explore differences between clinicoradiological features and radiomic features. Student's t-test and the chi-square test were used to compare continuous and categorical variables, respectively. Generally, two-sided *P*-values less than 0.05 were considered statistically significant. The intra-/inter-class correlation coefficients (ICCs) were used to assess the agreement of the two MR scanners and the extracted features by two radiologists.

RESULTS

Clinical Factors and MR Features

The clinical factors and MR features of the patients are shown in **Table 1**. For clinical factors, sex was found to be significantly different ($P < 0.001$) between the TM and AM groups, whereas age did not differ significantly ($P > 0.05$). For MR features, the parameters maximum diameter, tumor shape, peritumoral edema, enhanced features, bone invasion, and brain invasion were significantly different (all $P < 0.05$) in the univariate analysis. Among them, tumor shape and brain invasion were highly correlated with discrimination of TM from AM and can be used as independent predictive factors according to the multivariate logistic regression analysis. By contrast, tumor location, tumor border, dural tail sign, T2 signal, and sinus invasion were not significantly different (all $P > 0.05$) between the TM and AM groups.

Radiomic Features Correlated With TM and AM

The ICCs were calculated to evaluate the agreement of the two MR scanners and the features extracted by two radiologists, respectively. All values exceeded 0.75, reflecting good agreement. In total, 2454 radiomic features were extracted from each patient.

TABLE 1 | Clinical factors of the patients and magnetic resonance imaging features in the training and validation cohorts.

Characteristics	AM (n = 101)	TM (n = 141)	Univariate analysis (*p* value)	Multivariate analysis (*p* value)
Clinical factors				
Age (years)	51.5 ± 10.3	52.3 ± 9.2	0.543	N/A
Sex			<0.001*	<0.001*
Female	55 (54.5%)	122 (86.5%)		
Male	46 (45.5%)	19 (13.5%)		
Imaging features				
Tumor location			0.442	N/A
Parasinus and parasial	56 (55.4%)	69 (48.9%)		
Skull base	28 (27.8%)	50 (35.5%)		
Convexity	17 (16.8%)	22 (15.6%)		
Maximum diameter (mm)	47.96 ± 15.89	37.36 ± 15.18	<0.001*	N/A
Tumour shape			<0.001*	<0.001*
Circular or quasi- circular	38 (37.6%)	89 (63.1%)		
Irregular	63 (62.4%)	52 (36.9%)		
Tumour border			0.106	N/A
Clear	81 (80.2%)	124 (87.9%)		
Blur	20 (19.8%)	17 (12.1%)		
Dural tail sign			0.175	N/A
Yes	41 (40.6%)	45 (31.9%)		
None	60 (59.4%)	96 (68.1%)		
Peritumoural oedema			<0.001*	N/A
None (0%)	27 (27.3%)	83 (58.9%)		
≤5%	23 (23.2%)	19 (13.5%)		
6-33%	19 (19.2%)	14 (9.9%)		
34-67%	17 (17.2%)	19 (13.5%)		
68-95%	13 (13.1%)	6 (4.3%)		
MRI signal				
T2WI			0.056	N/A
Slightly high signal	50 (49.5%)	59 (41.8%)		
Iso signal	31 (30.7%)	64 (45.4%)		
Mixed signal	20 (19.8%)	18 (12.8%)		
Enhanced features			<0.001*	N/A
Uniform	42 (41.6%)	93 (66.0%)		
Uneven enhancement	59 (58.4%)	48 (34.0%)		
Bone invasion			0.022*	N/A
Yes	38 (37.6%)	33 (23.4%)		
No	63 (62.4%)	108 (76.6%)		
Sinus invasion			0.469	N/A
Yes	25 (24.8%)	41 (29.1%)		
No	76 (75.2%)	100 (70.9%)		
Brain invasion			0.014*	0.011*
Yes	21 (20.8%)	13 (9.2%)		
No	80 (79.2%)	128 (90.8%)		

*Among peritumoural oedema, percentage represents the proportion of peritumoural oedema in the entire abnormality, and the entire abnormality may be comprised of the entire tumour and oedema component.T2 signal is defined by comparing the signal of the gray matter of the brain. A Student's t-test was used to compare the difference in age and maximum diameter, while the chi-square test was used to compare the difference in other features. *P < 0.05. SD, standard deviation. N/A, not available.*

Among them, two T1C features and two T2 features were selected, and all four radiomic features (T1C_WaveletGLSZMwavelet.HHL_GraylevelNonUniformity, T1C_SquareRootGLSZM_squareroot_zoneEntropy, T2_WaveletGLCMwavelet.LLL_JointEnergy, and T2_SquareRootGLDM_squareroot_DependenceEntropy) were significantly different between the TM and AM groups (all $P < 0.05$; **Figure 2**). Their odds ratios are shown in **Supplementary Figure 3**. The weights of each selected radiomics features are shown in **Supplementary Table 2**.

According to the Spearman correlation test, these four features extracted by algorithms from MR images were consistent with some clinicoradiological features evaluated by the radiologists (**Supplementary Table 3**). For example, shape was correlated with the parameters T1C_WaveletGLSZMwavelet.HHL_GraylevelNonUniformity and T1C_SquareRootGLSZM_squareroot_zoneEntropy in both training and validation cohorts (**Figure 3**).

Fusion of Modalities and Model Building

T1C and T2 radiomic features may correspond to different information. Though fusing the selected radiomic features, the radiomics signature can reflect the discrimination factors of TM and AM from different perspectives. Stepwise multivariate logistic regression analysis showed that sex, tumor shape, and brain invasion were significantly different between TM and AM groups (all $P < 0.001$). Thus, the radiomic signature, sex, tumor shape, and brain invasion were selected for the clinicoradiomic model building. The radiomics scores in the AM group were

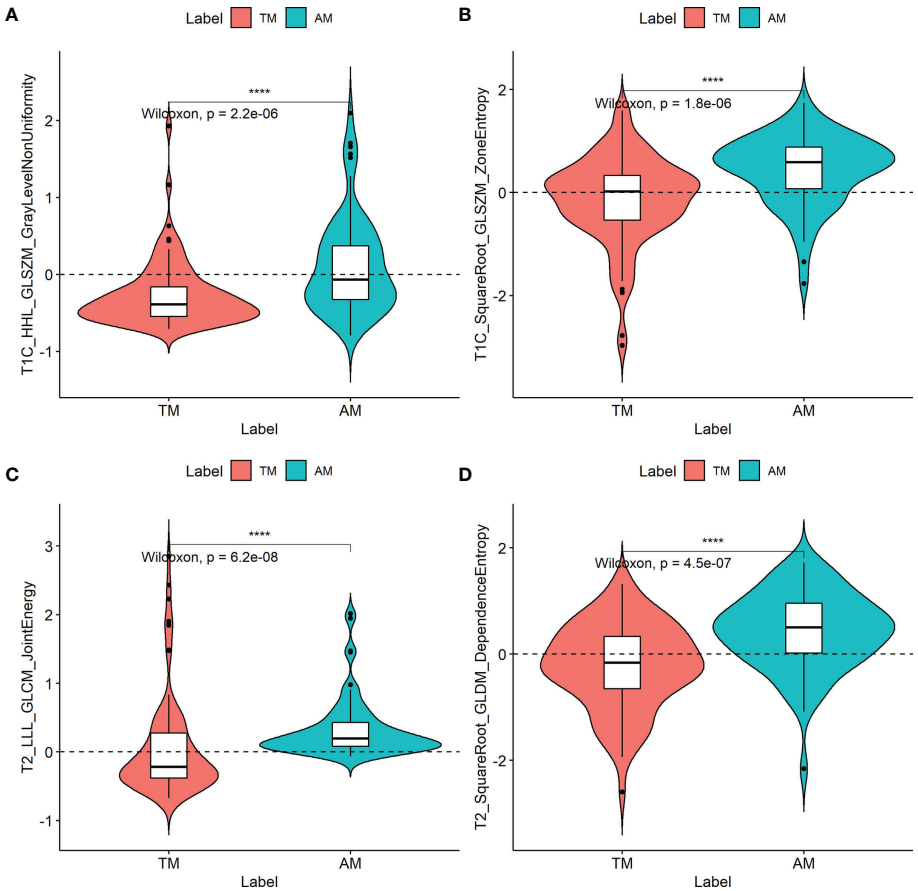


FIGURE 2 | Boxplots of the four radiomic features **(A–D)** with significant differences between transitional meningioma (TM) and atypical meningioma (AM) groups in the training cohort. The symbol **** represents $p < 0.001$.

significantly higher than those in the TM group in the different models, as shown in the **Figure 4**.

The performance of these models was evaluated in the training cohort and then validated in the validation cohort. The discrimination ability of T1C, T2, the radiomic (fusion of T1C and T2), clinical, and the clinicoradiomic models are shown in **Table 2**. The ROC curves for T1C, T2, the radiomic, clinical, and clinicoradiomic models are plotted in **Figure 5**. The clinicoradiomic model (nomogram) demonstrated the best discrimination ability, resulting in AUCs of 0.809 (95% CI, 0.743-0.874) and 0.795 (95% CI, 0.692-0.899) with sensitivity values of 74.0% (95% CI, 49.3%-83.6%) and 71.4% (95% CI, 42.9%-89.3%) for the differentiation of TM from AM in the training and validation cohorts, respectively. The formula for calculating the clinicoradiomic model and the fusion radiomic signature is described respectively in the **Supplementary Results**.

Model Comparison

The IDI index was calculated to assess the predictive usefulness of the different models. The clinicoradiomic model improved the integrated discrimination by 5.75% ($P = 0.002$) and 9.96% ($P < 0.001$) compared to the radiomic model in the training and validation cohorts, respectively. The comparisons between different models are shown in **Table 3**. In addition, Delong test showed that compared with Clinical and T2 models, the discrimination ability of clinicoradiomic model has been significantly improved in the training cohort, P value was 0.014 and 0.004 respectively, and there is no statistical significance in the validation cohort.

Assessment of the Clinicoradiomic Nomogram Performance

The clinicoradiomic model demonstrated the best discrimination ability and was used to construct the nomogram (**Figure 6A**). The calibration curve together with the Hosmer-Lemeshow test were used to measure the consistency between the probability of TM or AM being diagnosed by the clinicoradiomic model and the actual pathological diagnosis. The actual pathological diagnosis was consistent with the predicted probability of TM and AM in both the training and validation cohorts, with P-values of 0.361 and 0.472, respectively, as shown in **Figures 6B, C**.

The DCA assessed the discrimination ability of the clinicoradiomic model based on clinical applications. The clinicoradiomic model provided a net benefit in the DCA at a

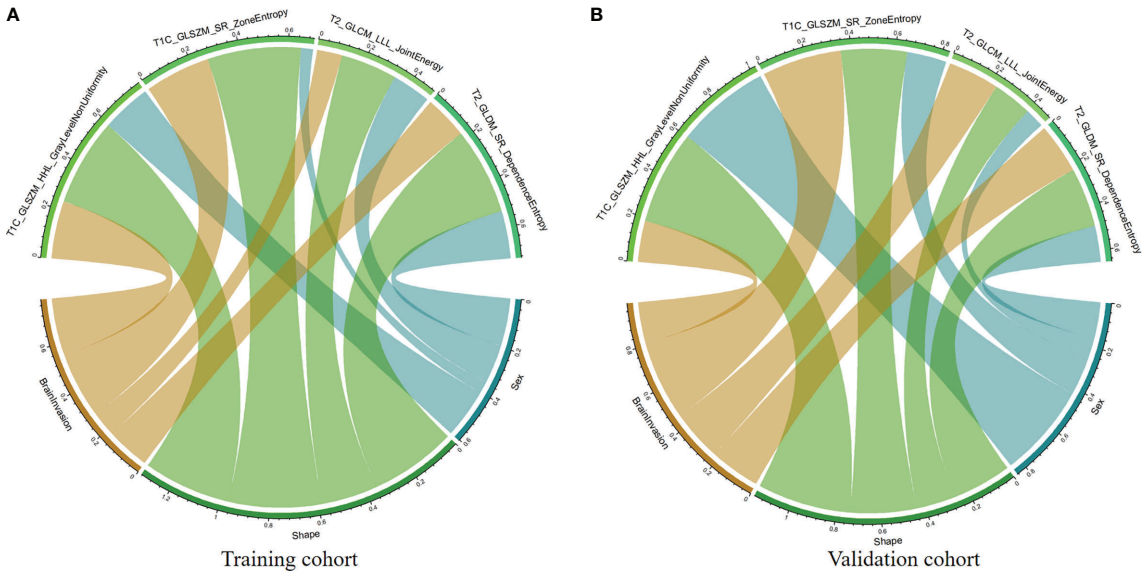


FIGURE 3 | Chord diagram of the correlation between clinicoradiological and selected radiomic features. Correlation analysis of clinicoradiological and selected radiomic features in the training **(A)** and validation **(B)** cohorts. The Spearman correlation test confirms that each link is significantly correlated ($P < 0.05$). The width of a link represents the strength of the correlation. For example, the T1C_SquareRootGLSZM_squareroot_zoneEntropy feature (gray) is highly correlated with tumor shape in both training and validation cohorts.

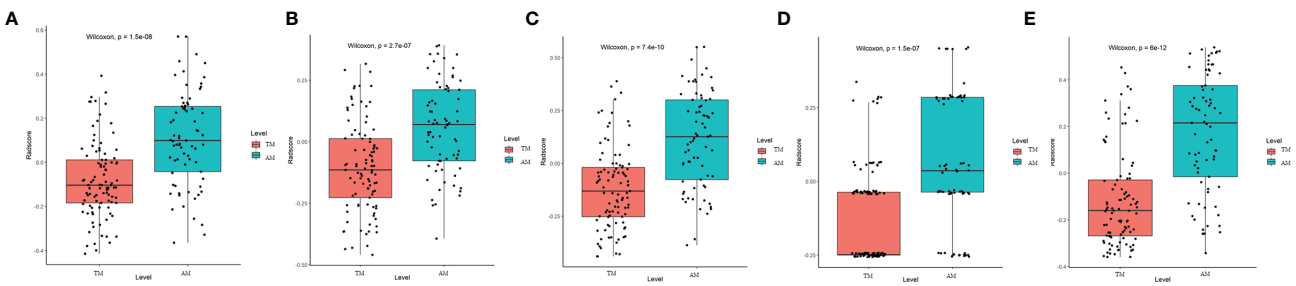


FIGURE 4 | Plots **(A–E)** show the boxplots of the corresponding radiomics score in the T1C, T2, combination of T1C and T2, clinical and clinicoradiomics models, respectively.

threshold probability of above 20% (**Figures 6D, E**). This result indicated that the clinicoradiomic data were clinically useful.

DISCUSSION

This is a preliminary study to develop a clinicoradiomic model that discriminates TM from AM based on MRI. The discrimination ability of this fusion model was validated *via* DCA, discrimination, and calibration curves in an internal validation cohort. One clinical factor, two radiological features, and four radiomic features indicated a high correlation with the ability of a model to discriminate between TM and AM. A multimodality (fusion of T1C and T2) model of radiomics showed good discrimination ability in both training (AUC: 0.776, Sensitivity: 0.685) and validation (AUC: 0.734, Sensitivity: 0.679) cohorts. Moreover, the nomogram incorporating clinicoradiological and radiomic features demonstrated the best performance in both training (AUC: 0.809, Sensitivity: 0.740) and validation (AUC: 0.795, Sensitivity: 0.714) cohorts.

Among clinical factors, sex was only the parameter that was significantly different between TM and AM. Females (86.5%) were prone to TM, whereas the male-to-female ratio was balanced in AM (females 54.5%), which is consistent with the results of other studies (3, 25). Among MR features, tumor shape and brain invasion were significantly different between TM and AM, and based on the stepwise multivariate logistic regression analysis, they can be used as independent discrimination factors. AMs are more irregular, and TMs are mostly circular or quasi-circular, which may be related to the grade of the tumor and the

TABLE 2 | Performance of the sequence models.

Cohort	Model	AUC	ACC	SEN	SPE	PPV	NPV
Training set	T1C	0.754 (0.679-0.829)	0.729 (0.656-0.795)	0.726 (0.589-0.822)	0.732 (0.464-0.866)	0.671 (0.623-0.698)	0.780 (0.692-0.808)
	T2	0.731 (0.655-0.806)	0.694 (0.619-0.762)	0.644 (0.466-0.767)	0.732 (0.556-0.825)	0.644 (0.567-0.683)	0.732 (0.675-0.755)
	Radiomics	0.776 (0.705-0.847)	0.753 (0.681-0.816)	0.685 (0.507-0.795)	0.804 (0.464-0.887)	0.725 (0.661-0.753)	0.772 (0.662-0.789)
	Clinical	0.726 (0.651-0.801)	0.688 (0.613-0.757)	0.534 (0.380-0.639)	0.804 (0.685-0.900)	0.672 (0.593-0.711)	0.696 (0.661-0.720)
	Nomogram	0.809 (0.743-0.874)	0.771 (0.700-0.831)	0.740 (0.493-0.836)	0.794 (0.567-0.866)	0.730 (0.643-0.753)	0.802 (0.743-0.816)
Test set	T1C	0.717 (0.597-0.836)	0.694 (0.575-0.798)	0.750 (0.392-0.893)	0.659 (0.295-0.818)	0.583 (0.423-0.625)	0.806 (0.650-0.837)
	T2	0.670 (0.541-0.798)	0.611 (0.489-0.724)	0.607 (0.392-0.858)	0.614 (0.432-0.886)	0.500 (0.392-0.586)	0.711 (0.633-0.780)
	Radiomics	0.734 (0.616-0.851)	0.722 (0.604-0.821)	0.679 (0.285-0.857)	0.750 (0.431-0.864)	0.633 (0.420-0.686)	0.786 (0.678-0.809)
	Clinical	0.765 (0.653-0.877)	0.736 (0.619-0.833)	0.714 (0.372-0.879)	0.750 (0.499-0.864)	0.645 (0.486-0.691)	0.805 (0.733-0.826)
	Nomogram	0.795 (0.692-0.899)	0.750 (0.634-0.845)	0.714 (0.429-0.893)	0.773 (0.477-0.910)	0.667 (0.545-0.714)	0.810 (0.724-0.833)

T1C, contrast-enhanced T1-weighted imaging; T2WI, T2-weighted imaging; Radiomics, combination of T1C and T2; Clinical, fusion of sex, tumour shape and brain invasion; AUC, area under receiver operating characteristic curve; ACC, balanced accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value.

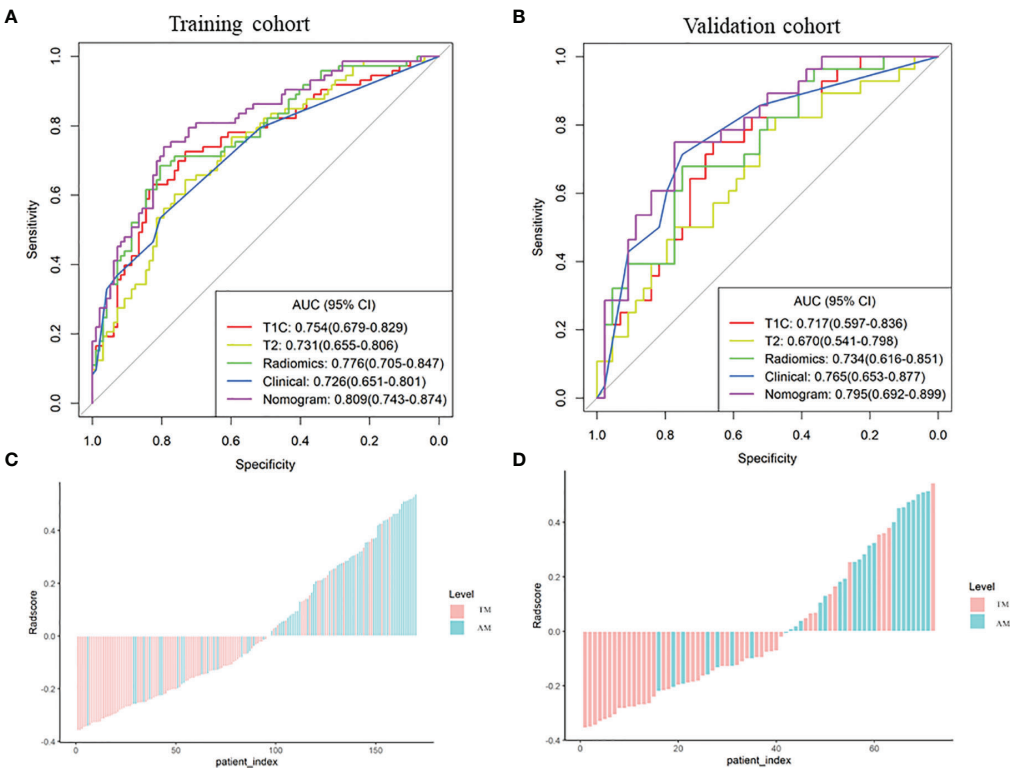


FIGURE 5 | Comparison of the receiver operating characteristic (ROC) curves of the different models. **(A, B)** ROC curves of the different models in the training and validation cohorts. The clinicoradiomic model demonstrates the best discrimination ability among these models, with area under the curve (AUC) values of 0.809 and 0.795 in the training and validation cohorts, respectively. **(C, D)** Radiomic signature histogram of the training and validation cohorts. The red bar shows the sample with transitional meningioma (TM), and the blue bar shows the sample with atypical meningioma (AM).

increased brain volume due to the peritumoral edema (26). Similarly, in most studies, irregular or lobulated tumor growth was associated with high-grade histology in both uni- and multivariate analyses (7, 8, 27, 28), presumably showing a parenchymal reaction of the brain tissue to the extensive tumors growth and the aggressiveness of the meningioma (27). Some authors also found that irregular or lobulated meningiomas were more likely to recur than regular-shaped ones (7). Zhang et al. have reported that the frequencies of brain invasion in TM and AM were much higher than those in other meningioma subtypes. In our study, the incidence of brain invasion in AM (20.8%) was higher than that in TM (9.2%), which is consistent with a previous study (4%-19% in all WHO grade meningiomas) (11, 12, 29). This indicates that WHO grade II AMs are more aggressive than WHO grade I TMs. Moreover, the maximum diameter, peritumoral brain edema, heterogeneous enhancement, and bone invasion were also different in these two tumors according to the univariate

TABLE 3 | Comparison of the different models in the validation cohort.

Initial model	Model introducing new factor	Performance improvement (IDI)	
		Training cohort	Validation cohort
Clinical	Clinicoradiomic	11.1% *P* = 0.00033	7.86% *P* = 0.0396
Combination of T1C and T2	Clinicoradiomic	5.75% *P* = 0.00195	9.96% *P* = 0.00089
T1C	Combination of T1C and T2	4.37% *P* = 0.00538	3.06% *P* = 0.3425
T2	Combination of T1C and T2	7.04% *P* = 0.00029	6.91% *P* = 0.02442

Compared with the T1C and T2 models, the performance of combination of T1C and T2 model improved by 7.04% and 4.37% in discrimination ability, respectively. Compared with combination of T1C and T2 model, the performance of clinicoradiomic model improved by 5.75% in discrimination ability. IDI: Integrated discrimination improvement; Clinicoradiomic, fusion of sex information, tumour shape, brain invasion and radiomic signature.

analysis, in agreement with published reports (30). For example, larger tumor size and tumor volume were more likely to be observed in high-grade meningiomas, and AM is a WHO grade II tumor. Heterogeneous enhancement reflects intratumoral hemorrhage, ischemic necrosis, cystic change, or calcification and is associated with heterogeneous distribution of tumor cells. Previous studies have reported that AMs have significantly more intratumoral necrosis and cystic changes than benign meningiomas (1, 31).

At present, MR radiomics can reproducibly extract objective and quantitative data from different sequences (T2, T1, T1C, and fluid attenuated inversion recovery [FLAIR], among others) to diagnostically discriminate meningiomas from other tumor forms, such as craniopharyngioma from meningioma in the sellar/parasellar area (32) or malignant hemangiopericytoma from angiomatous meningioma (20, 33). Radiomics can use visually imperceptible information about the tumor. Given this background, the radiomics model is a convenient, noninvasive method that does not require tissue biopsy or gene sequencing and may be a valuable approach to differentiate TM from AM since the radiomics model (AUC: 0.776 in the training cohort) outperformed the clinical model (AUC: 0.726). Additionally, we developed and validated a clinicoradiomics model to discriminate between TM and AM. Of the 2454 radiomic features, four were highly correlated with the discrimination between these two tumors. These features were textural image features indicating microscopic descriptions of the tumor including cellularity and tumor-induced compression of normal brain tissue. Textural features can neither be identified by the human visual system nor be easily interpreted to understand their specific meaning (7, 34, 35). We analyzed the four identified radiomic features and found that two gray-level size zone matrix (GLSZM) features, one gray-level dependence matrix (GLDM) feature, and one gray-level co-occurrence matrix (GLCM) feature were significantly associated with the discrimination of TM from AM. According to the definitions of these texture features (36), GLSZM quantifies gray-level zones in an image. The GLDM feature measures the difference between adjacent voxels based on their voxel value, and this feature was most relevant to the discrimination between these two tumors. The GLDM features selected by LASSO include entropy features, where a larger entropy value indicates greater heterogeneity of the tumor (37). The GLCM feature describes the distance and angle of each pixel, which includes energy, correlation, entropy, inertia, and inverse difference (37). Compared to TM, the values of these features were higher in AM. This indicates that these features may reflect microscopic heterogeneity within the tumors. Thus, as a new tool, the radiomic feature could distinguish TM from AM.

We also analyzed the correlation between clinicoradiological factors and radiomic features using Spearman's correlation analysis. We found that some clinicoradiological factors were consistent with some radiomic features extracted from MR images in both training and validation cohorts. This revealed that clinicoradiological features also reflected some radiomic MRI features. For example, shape was correlated with T1C_WaveletGLSZMwavelet.HHL_GraylevelNonUniformity and T1C_SquareRootGLSZM_squareroot_zoneEntropy in both training and validation cohorts. This indicated that some radiomic features corresponded to tumor shape in MR images, such as entropy in GLSZM. The correlation between these features revealed that some specific feature combinations may be explained by some clinicoradiological features to some extent. Besides, the included four features were extracted from different sequences, and the combination of T1C and T2 models of radiomic features indicated a better discrimination ability than the T1C or T2 models alone. These results indicated that different sequences provided distinct information, and multiple sequences could show more information about tumors and increase the discrimination ability of the model. After Delong test, we found that the discrimination ability of clinicoradiomic model is better than that of Clinical and T2 models. However, there is no statistical significance in the validation cohort, which may be related to the small sample size of the validation cohort. This result indicated that we should include more data in future study to improve the performance and stability of the model. At present, the clinical application of radiomics is still in the exploratory research stage, and it needs time to accumulate. Radiomics-derived data, when combined with other pertinent data sources (including clinically obtained, treatment-related or genomic data), can produce accurate robust evidence-based clinical-decision support systems (38).

The nomogram incorporating clinicoradiological factors and radiomic features showed the best discrimination ability compared to radiomic models based on T1C, T2, T1C/T2 images, and clinicoradiological factors. The results indicated that fusing clinicoradiological factors and radiomic features significantly improved classification performance. Combining qualitative and quantitative imaging analyses provided an additive effect because the information contained in these

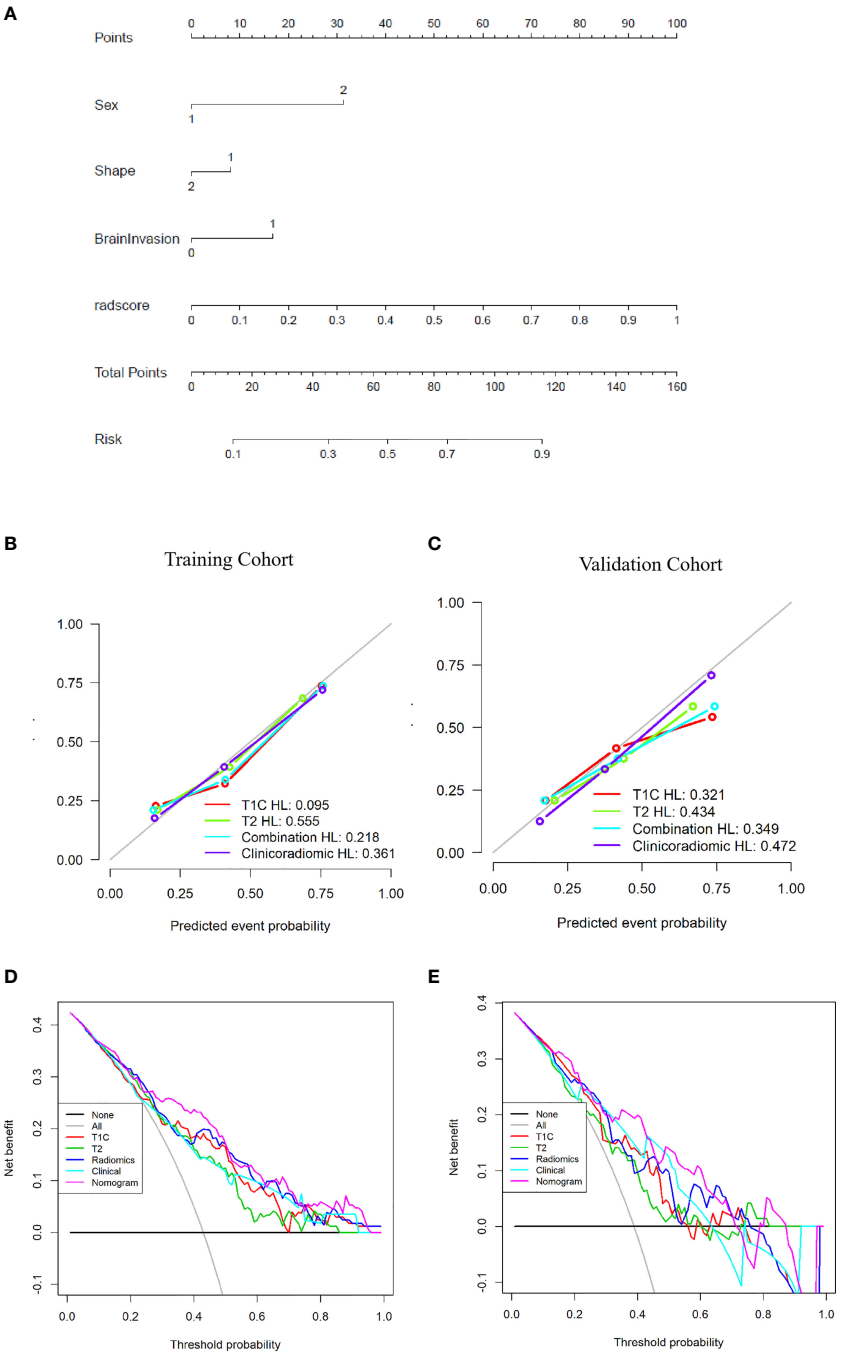


FIGURE 6 | Establishment and performance of the clinicoradiomic model. **(A)** The clinicoradiomic model is used to develop a nomogram. **(B, C)** Calibration curves of the clinicoradiomic nomogram for the training and validation cohorts. The x-axis represents the probability of atypical meningioma (AM) and transitional meningioma (TM) as measured by the clinicoradiomic model, and the y-axis represents the actual rate of AM and TM. The solid line represents the discrimination ability of the nomogram, and the diagonal dotted line represents the ideal evaluation by a perfect model. The *P*-values in the Hosmer-Lemeshow test are 0.361 and 0.472 in the training and validation cohorts, respectively. A closer fit to the diagonal dotted line represents a better evaluation. **(D, E)** Decision curve analysis for the clinicoradiomic model. The x-axis shows the threshold probability, and the y-axis measures the net benefit. The gray line represents all patients with AM, whereas the black line represents all patients with TM. The pink line represents the clinicoradiomic model.

features was complementary (39). Preoperative risk factors were extracted without postoperative factors to build the nomogram. This nomogram may be helpful for both clinicians and radiologists to preoperatively distinguish TM from AM. The nomogram was better than the radiomics model and could be applied in clinical practice for meningioma patients undergoing MRI scans.

Our study has several limitations. First, the analysis of MRI characteristics was independently performed by two radiologists, and imaging assessments are subjective. Second, we selected the T1C and T2 sequences for this study. On T2 images, the tumor had unclear boundaries in some cases. Although we referred to the T1C sequences for visual guidance to delineate tumor borders, there were still deviations. Third, our study was a single-center study, and a multi-center external validation is needed to test the generalizability and robustness of the model in the future. In addition, two neuroradiologists spent plenty of time to manually delineate two independent VOIs of tumors for each MRI sequence in this study, thus, efficient automatic segmentation and co-registration was available for meningiomas in the future research. In future, multimodal studies such as DWI and FLAIR sequences could be combined to improve accuracy.

CONCLUSION

Preoperative identification of TM and AM would aid the clinical decision-making and prognosis prediction. In the radiomic analysis, four radiomic features were highly correlated with the discrimination between these two tumors. After the fusion of T1C and T2 features, the MRI-based radiomics signature effectively identified TM and AM on MR images. The clinicoradiomic model that combined the radiomic signatures and clinicoradiological factors showed the best discrimination ability and may be used in patients with TM and AM.

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 For this retrospective analysis, ethical approval was obtained from the Institutional Review Board of Lanzhou University Second Hospital, and the requirement for informed consent was waived. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JZ, GZ, JLZ, KC, and YC contributed to conception and design of the study. TH organized the database. JR performed the statistical analysis. JZ wrote the first draft of the manuscript. ZZ, JR, YC, and GZ wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This study was supported by grants of National Natural Science Foundation of China (82071872), Doctoral Research Startup Fund of Zunyi Medical University (BS2021-03), Technology Plan Project of Guizhou Province [Qiankehe Support (2020) No. 4Y179], and Science and Technology Fund Project of Guizhou Provincial Health Commission (gzwkj2021-375).

ACKNOWLEDGMENTS

We would like to thank Ning Liu and Guangyou Li (from Zunyi Medical University) for the data interpretation.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.811767/full#supplementary-material

REFERENCES

1. Park YW, Oh J, You SC, Han K, Ahn SS, Choi YS, et al. Radiomics and Machine Learning may Accurately Predict the Grade and Histological Subtype in Meningiomas Using Conventional and Diffusion Tensor Imaging. *Eur Radiol* (2019) 29(8):4068–76. doi: 10.1007/s00330-018-5830-3
2. Louis DN, Perry A, Reifenberger G, von Deimling A, 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
3. Ma X-J, Zhang G-J, Wang W, Li D, Wu Z, Zhang J-T. Proposed Treatment for Intracranial Transitional Meningioma: A Single-Center Series of 298 Cases. *World Neurosurgery* (2019) 127:e280–7. doi: 10.1016/j.wneu.2019.03.104
4. Barresi V, Lionti S, Caliri S, Caffo M. Caffo: Histopathological Features to Define Atypical Meningioma: What Does Really Matter for Prognosis? *Brain Tumor Pathol* (2018) 35(3):168–80. doi: 10.1007/s10014-018-0318-z
5. Dehghan Harati M, Yu A, Magaki SD, Perez-Rosendahl M, Im K, Park YK, et al. Clinicopathologic Features and Pathogenesis of Melanocytic Colonization in Atypical Meningioma. *Neuropathology* (2018) 38(1):54–61. doi: 10.1111/neup.12409
6. Goldbrunner R, Minniti G, Preusser M, Jenkinson MD, Sallabanda K, Houdart E, et al. EANO Guidelines for the Diagnosis and Treatment of Meningiomas. *Lancet Oncol* (2016) 17(9):e383–91. doi: 10.1016/s1470-2045(16)30321-7
7. Yan PF, Yan L, Hu TT, Xiao DD, Zhang Z, Zhao HY, et al. The Potential Value of Preoperative MRI Texture and Shape Analysis in Grading Meningiomas: A Preliminary Investigation. *Transl Oncol* (2017) 10(4):570–7. doi: 10.1016/j.tranon.2017.04.006

8. Lin BJ, Chou KN, Kao HW, Lin C, Tsai WC, Feng SW, et al. Correlation Between Magnetic Resonance Imaging Grading and Pathological Grading in Meningioma. *J Neurosurg* (2014) 121(5):1201–8. doi: 10.3171/2014.7.jns132359
9. Kawahara Y, Nakada M, Hayashi Y, Kai Y, Hayashi Y, Uchiyama N, et al. Prediction of High-Grade Meningioma by Preoperative MRI Assessment. *J Neurooncol* (2012) 108(1):147–52. doi: 10.1007/s11060-012-0809-4
10. Zhang T, Yu J-M, Wang Y-Q, Yin D-D, Fang L-J. WHO Grade I Meningioma Subtypes: MRI Features and Pathological Analysis. *Life Sci* (2018) 213:50–6. doi: 10.1016/j.lfs.2018.08.061
11. Zhang J, Yao K, Liu P, Liu Z, Han T, Zhao Z, et al. 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
12. Adeli A, Hess K, Mawrin C, Streckert EMS, Stummer W, Paulus W, et al. Prediction of Brain Invasion in Patients With Meningiomas Using Preoperative Magnetic Resonance Imaging. *Oncotarget* (2018) 9(89):35974–82. doi: 10.18632/oncotarget.26313
13. 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
14. Rogers L, Barani I, Chamberlain M, Kaley TJ, McDermott M, Raizer J, et al. Meningiomas: Knowledge Base, Treatment Outcomes, and Uncertainties. A RANO Review. *J Neurosurg* (2015) 122(1):4–23. doi: 10.3171/2014.7.JNS131644
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. 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
17. Niu L, Zhou X, Duan C, Zhao J, Sui Q, Liu X, et al. Differentiation Researches on the Meningioma Subtypes by Radiomics From Contrast-Enhanced Magnetic Resonance Imaging: A Preliminary Study. *World Neurosurg* (2019) 126:e646–52. doi: 10.1016/j.wneu.2019.02.109
18. Zhu Y, Man C, Gong L, Dong D, Yu X, Wang S, et al. A Deep Learning Radiomics Model for Preoperative Grading in Meningioma. *Eur J Radiol* (2019) 116:128–34. doi: 10.1016/j.ejrad.2019.04.022
19. Olar A, Goodman LD, Wani KM, Boehling NS, Sharma DS, Mody RR, et al. A Gene Expression Signature Predicts Recurrence-Free Survival in Meningioma. *Oncotarget* (2018) 9(22):16087–98. doi: 10.18632/oncotarget.24498
20. Li X, Lu Y, Xiong J, Wang D, She D, Kuai X, et al. Presurgical Differentiation Between Malignant Haemangiopericytoma and Angiomatous Meningioma by a Radiomics Approach Based on Texture Analysis. *J Neuroradiol* (2019) 46 (5):281–7. doi: 10.1016/j.neurad.2019.05.013
21. 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
22. Zwanenburg A, Vallières 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(2):328–38. doi: 10.1148/radiol.2020191145
23. Miller TD, Askew JW. Askew: Net Reclassification Improvement and Integrated Discrimination Improvement: New Standards for Evaluating the Incremental Value of Stress Imaging for Risk Assessment. *Circ Cardiovasc Imaging* (2013) 6(4):496–8. doi: 10.1161/CIRCIMAGING.113.000797
24. Fitzgerald M, Saville BR, Lewis RJ. Decision Curve Analysis. *JAMA* (2015) 313 (4):409–10. doi: 10.1001/jama.2015.37
25. Barthélemy E, Loewenstern J, Konuthula N, Pain M, Hall J, Govindaraj S, et al. Primary Management of Atypical Meningioma: Treatment Patterns and Survival Outcomes by Patient Age. *J Cancer Res Clin Oncol* (2018) 144 (5):969–78. doi: 10.1007/s00432-018-2618-4
26. Thenier-Villa JL, Alejandro Galarraga Campoverde R, Ramon DELALZA, Conde Alonso C. Predictors of Morbidity and Cleavage Plane in Surgical Resection of Pure Convexity Meningiomas Using Cerebrospinal Fluid Sensitive Image Subtraction Magnetic Resonance Imaging. *Neurol Med Chir (Tokyo)* (2017) 57(1):35–43. doi: 10.2176/nmc.oa.2016-0169
27. Spille DC, Sporns PB, Hess K, Stummer W, Brokinkel B. Prediction of High-Grade Histology and Recurrence in Meningiomas Using Routine Preoperative Magnetic Resonance Imaging: A Systematic Review. *World Neurosurg* (2019) 128:174–81. doi: 10.1016/j.wneu.2019.05.017
28. Liu H, Zhou J, Li W, Liu G. Comparative Analysis of the Magnetic Resonance Imaging Features Between Anaplastic Meningioma and Atypical Meningioma. *J Craniofac Surg* (2016) 27(3):e229–33. doi: 10.1097/SCS.0000000000002361
29. Spille DC, Hess K, Sauerland C, Sanai N, Stummer W, Paulus W, et al. Brain Invasion in Meningiomas: Incidence and Correlations With Clinical Variables and Prognosis. *World Neurosurg* (2016) 93:346–54. doi: 10.1016/j.wneu.2016.06.055
30. Salah F, Tabbarah A, ALArab Y, Asmar K, Tamim H, et al. Can CT and MRI Features Differentiate Benign From Malignant Meningiomas? *Clin Radiol* (2019) 74(11):898.e15–.e23. doi: 10.1016/j.crad.2019.07.020
31. Hsu C-C, Pai C-Y, Kao H-W, Hsueh C-J, Hsu W-L, Lo C-P. Do Aggressive Imaging Features Correlate With Advanced Histopathological Grade in Meningiomas? *J Clin Neurosci* (2010) 17(5):584–7. doi: 10.1016/j.jocn.2009.09.018
32. Tian Z, Chen C, Zhang Y, Fan Y, Feng R, Xu J. Radiomic Analysis of Craniopharyngioma and Meningioma in the Sellar/Parasellar Area With MR Images Features and Texture Features: A Feasible Study. *Contrast Media Mol Imaging* (2020) 2020:4837156. doi: 10.1155/2020/4837156
33. Kanazawa T, Minami Y, Jinzaki M, Toda M, Yoshida K, Sasaki H. Preoperative Prediction of Solitary Fibrous Tumor/Hemangiopericytoma and Angiomatous Meningioma Using Magnetic Resonance Imaging Texture Analysis. *World Neurosurg* (2018) 120:e1208–16. doi: 10.1016/j.wneu.2018.09.044
34. Zhang Y, Chen JH, Chen TY, Lim SW, Wu TC, Kuo YT, et al. Radiomics Approach for Prediction of Recurrence in Skull Base Meningiomas. *Neuroradiology* (2019) 61(12):1355–64. doi: 10.1007/s00234-019-02259-0
35. Wu Q, Yao K, Liu Z, Li L, Zhao X, Wang S, et al. Radiomics Analysis of Placenta on T2WI Facilitates Prediction of Postpartum Haemorrhage: A Multicentre Study. *EBioMedicine* (2019) 50:355–65. doi: 10.1016/j.ebiom.2019.11.010
36. Laukamp KR, Shakirin G, Baessler B, Thiele F, Zopfs D, Hokamp NG, et al. Accuracy of Radiomics-Based Feature Analysis on Multiparametric Magnetic Resonance Images for Noninvasive Meningioma Grading. *World Neurosurg* (2019) 132:e366–90. doi: 10.1016/j.wneu.2019.08.148
37. Chu H, Lin X, He J, Pang P, Fan B, Lei P, et al. Value of MRI Radiomics Based on Enhanced T1WI Images in Prediction of Meningiomas Grade. *Acad Radiol* (2020) 28(5):687–93. doi: 10.1016/j.acra.2020.03.034
38. 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
39. Pallud J, Coroller TP, Bi WL, Huynh E, Abedalthagafi M, Aizer AA, et al. Radiographic Prediction of Meningioma Grade by Semantic and Radiomic Features. *PloS One* (2017) 12(11):e0187908. doi: 10.1371/journal.pone.0187908

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 28 January 2022
doi: 10.3389/fonc.2021.733260

An MRI-Based Radiomics Model for Predicting the Benignity and Malignancy of BI-RADS 4 Breast Lesions

Renzhi Zhang[1†], Wei Wei[2†], Rang Li[3,4], Jing Li[1], Zhuhuang Zhou[4], Menghang Ma[2], Rui Zhao[2] and Xinming Zhao[1]*

[1] Department of Diagnostic Radiology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China, [2] School of Electronics and Information, Xi'an Polytechnic University, Xi'an, China, [3] College of Engineering, Boston University, Boston, MA, United States, [4] Faculty of Environment and Life, Beijing University of Technology, Beijing, China

OPEN ACCESS

Edited by:
Natalie Julie Serkova, University of Colorado, United States

Reviewed by:
Kun Zheng, Peking Union Medical College Hospital (CAMS), China
Stefano Marrone, University of Naples Federico II, Italy
Vishwa S. Parekh, Johns Hopkins University, United States

***Correspondence:**
Xinming Zhao
zhaoxinming@cicams.ac.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 June 2021*
Accepted: *08 December 2021*
Published: *28 January 2022*

Citation:
Zhang R, Wei W, Li R, Li J, Zhou Z, Ma M, Zhao R and Zhao X (2022) An MRI-Based Radiomics Model for Predicting the Benignity and Malignancy of BI-RADS 4 Breast Lesions. Front. Oncol. 11:733260. doi: 10.3389/fonc.2021.733260

Objectives: The probability of Breast Imaging Reporting and Data Systems (BI-RADS) 4 lesions being malignant is 2%–95%, which shows the difficulty to make a diagnosis. Radiomics models based on magnetic resonance imaging (MRI) can replace clinicopathological diagnosis with high performance. In the present study, we developed and tested a radiomics model based on MRI images that can predict the malignancy of BI-RADS 4 breast lesions.

Methods: We retrospective enrolled a total of 216 BI-RADS 4 patients MRI and clinical information. We extracted 3,474 radiomics features from dynamic contrast-enhanced (DCE), T_2-weighted images (T_2WI), and diffusion-weighted imaging (DWI) MRI images. Least absolute shrinkage and selection operator (LASSO) and logistic regression were used to select features and build radiomics models based on different sequence combinations. We built eight radiomics models which were based on DCE, DWI, T_2WI, DCE+DWI, DCE+T_2WI, DWI+T_2WI, and DCE+DWI+T_2WI and a clinical predictive model built based on the visual assessment of radiologists. A nomogram was constructed with the best radiomics signature combined with patient characteristics. The calibration curves for the radiomics signature and nomogram were conducted, combined with the Hosmer-Lemeshow test.

Results: Pearson's correlation was used to eliminate 3,329 irrelevant features, and then LASSO and logistic regression were used to screen the remaining feature coefficients for each model we built. Finally, 12 related features were obtained in the model which had the best performance. These 12 features were used to build a radiomics model in combination with the actual clinical diagnosis of benign or malignant lesion labels we have obtained. The best model built by 12 features from the 3 sequences has an AUC value of 0.939 (95% CI, 0.884-0.994) and an accuracy of 0.931 in the testing cohort. The sensitivity, specificity, precision and Matthews correlation coefficient (MCC) of testing

cohort are 0.932, 0.923, 0.982, and 0.791, respectively. The nomogram has also been verified to have calibration curves with good overlap.

Conclusions: Radiomics is beneficial in the malignancy prediction of BI-RADS 4 breast lesions. The radiomics predictive model built by the combination of DCE, DWI, and T_2WI sequences has great application potential.

Keywords: LASSO, BI-RADS 4, breast lesion, magnetic resonance imaging, radiomics

INTRODUCTION

The 2020 Global Cancer Report released by the International Agency for Research on Cancer (IARC) shows that female breast cancer has replaced lung cancer as the most common cancer in the world with an increase of approximately 2.3 million new cases (11.7%) throughout the year (1). Breast cancer is also estimated to top the list of new morbidity and mortality among all types of cancers in women. During treatment, the diagnosis of benign or malignant breast lesions has become the most basic and important step in the treatment of breast diseases.

According to the breast cancer screening guidelines of the National Comprehensive Cancer Network (NCCN) and BI-RADS, suspicious lesions can be classified into 6 categories (2). The fourth category of breast disease is defined as a type of breast lesions with suspicious malignancy and uncertain pathological types. The probability of being malignant is 2%–95%, although this type of breast disease has a further classification of 4a, 4b, and 4c, because of the large range of the possibility of the existence of malignant lesion, all patients with BI-RADS 4 breast diseases are recommended to undergo biopsy of suspicious areas to clarify their pathological properties (3). Because of the blurry qualitative characteristics of BI-RADS 4 breast diseases, we can see that most patients with BI-RADS 4 of breast diseases are overdiagnosed and treated with the puncture case analysis, which requires a certain degree of trauma to the body. Moreover, the clinical diagnosis inevitably has a certain false-positive rate and missed diagnosis rate (4, 5). Therefore, we propose a hypothesis to establish a predictive radiomics model based on the patients' preoperative imaging information, thereby avoiding patients with benign breast lesions from undergoing invasive pathological testing.

Computer-aided diagnosis methods based on medical imaging have been increasing in clinical application value in recent years. This method only needs to mark the abnormal signs, and then perform common image processing on this basis to get the diagnosis result. Therefore, the concept of radiomics came into being (6). Radiomics is a research method that extracts high-throughput image features from medical images and conducts quantitative research. DCE, DWI, and T_2WI are three routine methods for the diagnosis and observation of breast diseases. Breast DCE imaging has high sensitivity in breast cancer screening for women who have accumulated breast cancer risk for more than 20%–25% (7). However, its specificity depends on a variety of external factors, such as the professional skills of the reader or the method of using quantitative techniques. DWI can characterize the three-dimensional fluidity of water in the body and indirectly detect and visualize the microstructure (8). DWI and apparent diffusion coefficient (ADC) have been successfully applied to the clinical diagnosis and screening of breast cancer. Compared with the average specificity of 80% in DCE for breast cancer diagnosis, the average specificity of the combined diagnosis of DCE and DWI can reach 89.2% (9). Therefore, radiologists mainly use the combination of these two technologies in the diagnosis of breast cancer. T_2WI is usually used to exclude cysts, intramammary lymph nodes, and other benign breast lesions (10). One of the most advantageous characteristics of T_2WI is that signal strength is directly related to the underlying disease state for most breast cancer lesions mainly showing uneven or slightly high signal on T_2WI MRI, while the surrounding tissues show low or medium signal (11). As existing research shows that the T_2WI MRI image has a strong ability to interpret the pathology and diagnosis of breast diseases (12). Research combining T_2WI MRI, DCE MRI, and DWI MRI to increase the actual diagnosis efficiency also picture a wider application of combining multiple MRI features in the field of breast diagnosis (13, 14).

It can be seen from the above results that the combined application of different MRI images can play a greater practical role in the diagnosis of breast diseases. This study aims to establish an auxiliary diagnosis prediction model that can be used to predict benign and malignant breast lesions by combining MRI images with three sequences of DCE, DWI, and T_2WI.

MATERIALS AND METHODS

This study had been reviewed and approved by the Ethics Committee of Cancer Hospital, Chinese Academy of Medical Sciences, and had been in line with the Declaration of Helsinki. All patients participating in this study waived the requirement of informed consent. The image processing methods had met the terms and conditions mentioned in the Transparent Reporting of a Multivariable Model for Individual Prognosis or Diagnosis (TRIPOD), Image Biomarker Standardization Initiative (IBSI), Checklist for Artificial Intelligence in Medical Imaging (CLAIM).

Patients

The patient information for this research study comes from the picture archiving and communication system (PACS) of the Cancer Hospital of the Chinese Academy of Medical Sciences. Due to the following criteria, we continuously enrolled a total of

216 research subjects from 230 patients who received diagnosis at our hospital from September 2018 to December 2019. Each of the patients was classified with the fifth version of BI-RADS guidelines. The criteria for inclusion in the group were as follows: (i) Patients have been diagnosed with a BI-RADS type IV breast lesion. (ii) The pathological diagnoses results were confirmed by puncture pathology diagnoses. (iii) Patients have complete MRI images in PACS with axial DCE, DWI, and T_2WI sequences obtained before patients underwent biopsy.

All the in-group patients were separated into a training cohort of 144 patients and a testing cohort of 72 patients divided by scan time from early to late according to the ratio of 2:1. We use leave-one-out cross-validation (LOOCV), and the training cohort is used to train and validate the model. Testing cohort is used to test the performance of the model.

Semiautomatic Image Segmentation and Feature Extraction

We collected image information of the 216 enrolled patients, which were composed of three kinds of MRI images: DCE, DWI, and T_2WI. All patients' MRI scans were completed before they underwent biopsy. Patients in training and testing cohorts were scanned using the same equipment. The DCE sequences were obtained by a higher axial resolution T1-weighted DCE imaging with a temporal resolution of 90 s. The DWI sequences were obtained by a DWI sequence with two *b*-values (0–1,000 s/mm^2). The T_2WI sequences were obtained from a higher axial resolution T_2WI turbo spin-echo sequence. All scans were done at a magnetic field strength of 3.0 T.

These MRI images were reviewed by 2 radiologists who were not aware of the real pathological diagnoses. They calibrated and refined the segmentation results from Radiomics (www.radiomics.net.cn) of the breast lesion regions of interest (ROI) of the patients' pretreatment images in the DCE, DWI, and T_2WI views. This software is a computerized semiautomatic image segmentation software with high accuracy trained by using a deep learning model. It uses the most recognized nnU-Net model framework, which is a generalized U-net framework and can obtain better training results through a careful preprocessing process and diverse network training schemes. An example of Radiomics software combined with manual fine-tuning for semiautomatic image segmentation is shown in **Figure 1**.

The specific operations of feature extraction were all done through Python (3.7.9), where the "pyradiomics" package (pyradiomics.readthedocs.io) was used in conjunction. The hyperparameters of feature extractor set as follows: Laplacian of Gaussian filter: sigma: [2.0, 3.0, 4.0, 5.0]; normalize: True; normalize scale: 100; resampled pixel spacing (1, 1, 1):; interpolator: sitkBSpline; binWidth: 5; and voxelArrayShift: 300. The other settings are default. A total of 3,474 quantitative original image features including first-order statistics features (18 features), 3D shape-based features (14 features), gray-level cooccurrence matrix features (24 features), gray-level run length matrix features (16 features), gray-level size zone matrix features (16 features), and gray-level dependence matrix features (14 features) were extracted from the corresponding regions of interest in the original images in the DCE, DWI, and T_2WI views (15). The original images were later added with wavelet filtering and Laplace of Gaussian (LoG) filtering respectively to extract the above features. Wavelet filtering was aimed to yield 8 decompositions per level of the original images. LoG filtering was used as an edge enhancement filter which emphasized areas of gray level changes, where sigma defines how the emphasized textures are supposed to be (16, 17).

Radiomics Signature Building

Due to the same coarse-to-fine strategy, we filtered the features to prevent the model from overfitting (18–20). First, a mono-factor analysis was performed on all features, and all features were ranked in order from smallest to largest *p*-value, and the top 5% of features were filtered out. Second, the LASSO algorithm made the image features most relevant to the BI-RADS 4 breast lesions to be filtered out by compressing the correlation coefficients of some of the features and zeroing out another part of the coefficients. Finally, the LOOCV was performed to select a model. After the above steps are completed, the coefficients of most radiomics features were compressed to zero. Then, a radiomics signature was established based on a linearly weighted combination of features with nonzero coefficients (21). The above content is the entire process of our training model. The feature extraction and selection process were implemented in Python (3.7.9).

The ability of the model to recognize benign and malignant lesions is evaluated by the drawn ROC. ROC can directly show the sensitivity (true-positive rate) and false-positive rate of the predicted results of the model. Meanwhile, the AUC value and the accuracy value are two main indicators that can be indirectly obtained from the curve. In addition, we also calculated precision and MCC metrics to evaluate the model.

Development and Performance of the Models Built With Sequences Combinations

Based on the above modeling steps, we performed prediction model building using different combinations of the three sequences, respectively. First, we modeled the three sequences individually based on the eigenvalue data extracted from each sequence. Second, the feature data of each two sequences were combined and modeled based on this data. Finally, we modeled the data using the combined data of the three sequences. These seven different models were used to compare performance in the same computational way.

Development and Performance of the Radiomics Nomogram

To verify that radiomics signature combined with clinical factor has a greater predictive ability, we built the radiomics nomogram (21). Through the progress of organizing the patients' information and univariate logistic regression analysis, we selected age as the clinical factor used for radiomics nomogram building. We constructed a calibration curve for the nomogram to demonstrate its predictive efficacy.

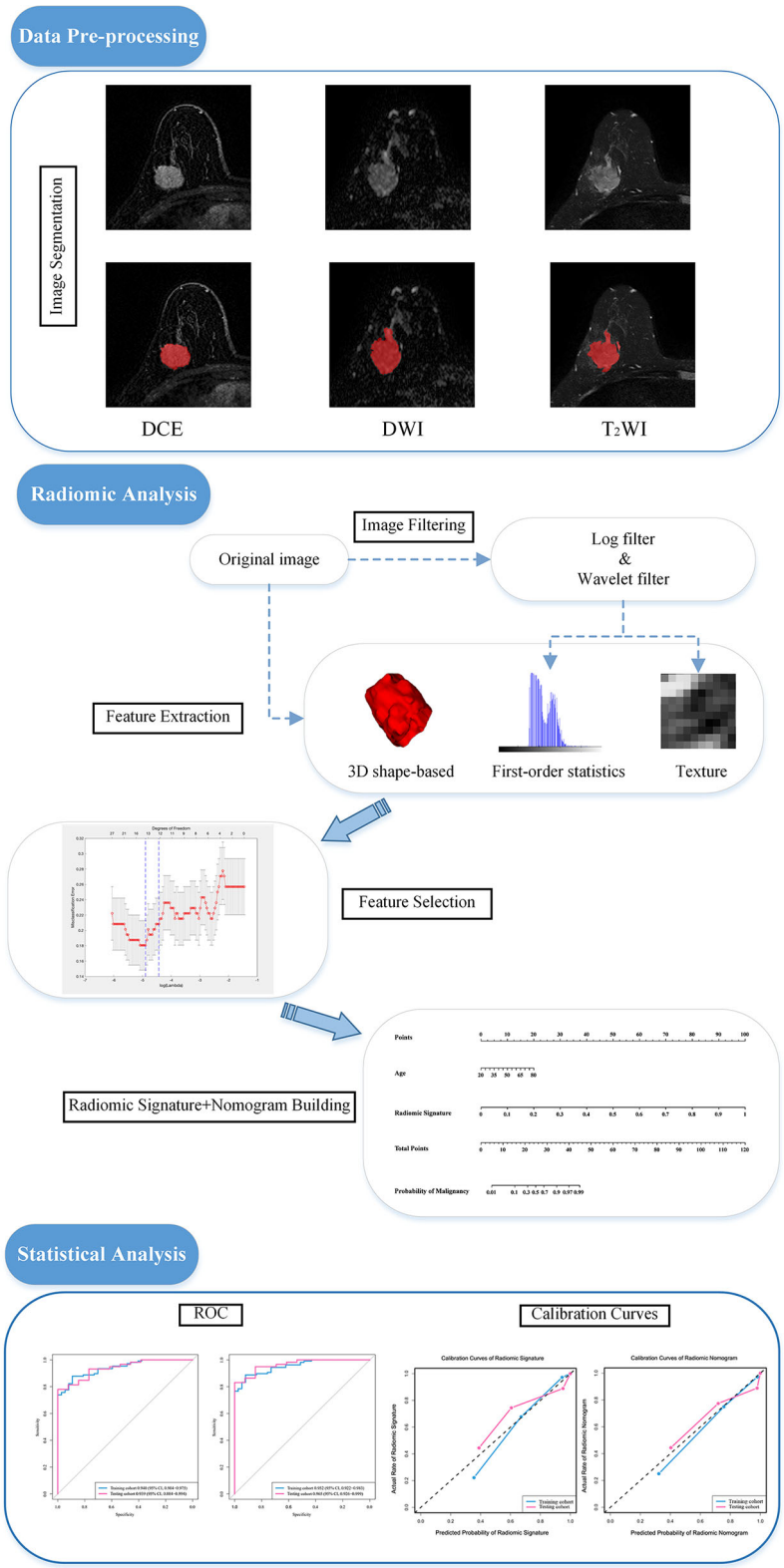


FIGURE 1 | Image segmentation and the procedure of developing a predictive model. The experiment is divided into three main parts: image pre-processing, image radiomic analysis and statistical analysis. Radiomic analysis includes image feature extraction and feature filtering. DCE, dynamic contrast enhanced imaging; DWI, diffusion weighted imaging; T_2WI, $_2$ weighted imaging; ROC, receiver operating characteristic curve.

Imaging Diagnosis by the Radiologists

In order to further verify the application value of the prediction model, we invited two radiologists to participate in our model verification step. Both radiologists come from the Cancer Hospital of the Chinese Academy of Medical Sciences and have more than 10 years of clinical diagnosis experience. The experiment requires the two experts to combine DCE, DWI, and T_2WI MRI images of each patient without contacting the patient and completely ignorant of the clinical information and actual diagnosis results to jointly make their pathological diagnosis of each case of whether benign or malignant breast lesion judgments. During the actual experiment, two radiologists conducted pathological identification of all patients through visual observation of images. By comparing the prediction results of our model with the diagnosis results of these two experts, we can evaluate whether this model has value in clinical application.

After obtaining the diagnosis results of the two radiologists, we used the same calculation method to calculate the accuracy, sensitivity, specificity, precision, and MCC of the radiologist's diagnoses and compare them with the corresponding indicators of the radiomics model.

Data and Statistical Analysis

To further evaluate the results, we plotted nomograms from the clinical information and plotted calibration curves combined with the Hosmer-Lemeshow test for the nomograms and the ROC curves of the radiomics model with the strongest performance. The p-value was calculated by a two samples *t*-test to evaluate the degree of group differentiation of the data. The above statistical evaluation work was done through R version 4.0.5 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

Clinical Characteristics of the Patients

The clinical characteristics of the patients in the training cohort and the testing cohort are shown in **Table 1**. The benign rate of breast lesion in the training cohort is 25.69% and that of the testing cohort is 18.05%. The differences in imaging devices, age, and other clinical characteristics were not statistically significant between the two cohorts.

Feature Selection, Radiomics Signature Development, and Validation

Twelve features were derived from DCE, DWI, and T_2WI MRI images with the principle of coarse-to-fine. The selected features and corresponding coefficients are listed in **Table 2**.

Development, Performance, and Evaluation of the Prediction Models

By comparing the ROC curves obtained after building radiomics prediction models for single sequences separately and for every two sequences combined with the ROC curve of radiomics prediction models built by combining the three sequences, the radiomics models built by combining the DCE, DWI, and T_2WI sequences obtained a significant performance improvement. Also, the AUC value of the model based on three sequences combined is the highest among all AUC values we have got. The performance of our joint prediction model is significantly improved, especially in terms of the specificity of the model prediction. The results are shown in **Table 3**. The ROC curves are shown in **Figure 2**.

Assessment of the Radiomics Nomogram

We conducted the radiomics nomogram by combining the patients' ages and the radiomics signature. The nomogram has achieved an AUC value of 0.965 (95% CI, 0.926–0.999) and an accuracy of 0.912 in the testing cohort. The radiomics nomogram is shown in **Figure 3**. The calibration curves of the radiomics nomogram and the radiomics model based on DCE, DWI, and T_2WI are shown in **Figure 4**.

Comparison of the Prediction Results Between Radiologists and the Radiomics Prediction Model

Two radiologists who have reviewed the in-group patients' images reached a good consensus of their visual diagnosis

TABLE 1 | Basic clinical information of enrolled patients.

Characteristics	Training (N = 144)	Testing (N = 72)	Total (N = 216)	*P*-value
Age at surgery (years), median (range)	45 (22-72)	45 (23-78)	45 (22-78)	0.052[a]
Benign (%)	37 (25.69)	13 (18.06)	50 (23.15)	0.867[b]
Adenomatosis	9 (6.25)	3 (4.17)	12 (5.56)	
Phyllodes tumor	28 (19.44)	10 (13.89)	38 (17.59)	
Malignant (%)	107 (74.31)	59 (81.94)	166 (76.85)	0.066[b]
Invasive ductal carcinoma	67 (46.53)	28 (38.89)	95 (43.98)	
Colloid carcinoma	20 (13.89)	11 (15.28)	31 (14.35)	
Medullary carcinoma	18 (12.60)	19 (25.39)	37 (17.13)	
Neuroendocrine carcinoma	1 (0.69)	1 (1.39)	2 (0.93)	
Solid papillary carcinoma	1 (0.69)	0 (0.00)	1 (0.46)	

The differentiation in the characteristics (age when diagnosed, benignity and malignancy, pathological diagnosis) in the training cohort and the Testing cohort were evaluated. P-value less than 0.05 proves that the groups are significantly different. The above P-values show the training and the Testing cohorts are non-significantly different.

[a]*P-value was calculated by two sample t-test.*

[b]*P-values were calculated by Fisher exact test.*

TABLE 2 | Results of the feature selection for the model based on DCE, DWI and T_2WI.

Sequences	Features	Coefficients
T_2WI	Wavelet LHH glrlm long run low gray level emphasis	-0.19765
	Original glszm Gray Level Non-Uniformity Normalized	-0.52642
	Wavelet LLL glszm Small Area Low Gray Level Emphasis	-1.53857
	Wavelet LHL glszm Low Gray Level Zone Emphasis	-1.34546
	Wavelet LLH glszm Small Area Low Gray Level Emphasis	-6.49947
	Log sigma 4-0-mm-3D glrlm Long Run Low Gray Level Emphasis	-0.00238
	Log sigma 4-0-mm-3D glrlm Long Run Emphasis	-0.00004
	Wavelet LHH glszm Gray Level Non-Uniformity Normalized	-1.11352
DCE	Original glszm Small Area Emphasis	4.57208
	Wavelet LHH glcm Correlation	12.60474
	Wavelet LLL glcm Inverse Variance	-4.61787
DWI	Wavelet HHH glrlm Long Run High Gray Level Emphasis	-0.02102

The features were selected by LASSO algorithm. These coefficients show the magnitude of the weight of their corresponding characteristics in the regression model.

TABLE 3 | Results of the radiomic models and the models based on the nomogram and visual assessment.

	AUC		Accuracy		Sensitivity/Recall		Specificity		Precision		MCC	
	Training (95% CI)	Testing (95% CI)	Training	Testing	Training	Testing	Training	Testing	Training	Testing	Training	Testing
DCE	0.901 (0.853-0.949)	0.844 (0.741-0.946)	0.806	0.819	0.804	0.864	0.811	0.615	0.925	0.911	0.561	0.444
DWI	0.871 (0.812-0.930)	0.798 (0.689-0.907)	0.847	0.801	0.846	0.814	0.865	0.769	0.947	0.941	0.651	0.493
T_2WI	0.877 (0.822-0.932)	0.838 (0.731-0.940)	0.868	0.777	0.879	0.746	0.838	0.846	0.940	0.957	0.679	0.492
DCE+DWI	0.932 (0.894-0.970)	0.821 (0.727-0.916)	0.861	0.708	0.860	0.712	0.865	0.692	0.948	0.913	0.675	0.324
DCE+T_2WI	0.924 (0.880-0.968)	0.853 (0.751-0.957)	0.889	0.820	0.907	0.813	0.838	0.846	0.942	0.96	0.721	0.551
DWI+T_2WI	0.889 (0.837-0.941)	0.834 (0.731-0.938)	0.882	0.777	0.869	0.780	0.919	0.769	0.969	0.939	0.730	0.453
DCE+DWI+T_2WI	0.940 (0.904-0.975)	0.939 (0.884-0.994)	0.924	0.931	0.935	0.932	0.892	0.923	0.961	0.982	0.806	0.791
Nomogram	0.952 (0.922-0.983)	0.965 (0.926-0.999)	0.896	0.912	0.887	0.932	0.919	0.846	0.969	0.965	0.756	0.737
Visual Assessment	0.613 (0.528-0.712)	0.563 (0.470-0.772)	0.632	0.611	0.644	0.627	0.594	0.538	0.821	0.860	0.21	0.130

The 9 models contain models based on DCE, DWI, T_2WI, DCE+DWI, DCE+T_2WI, DWI+T_2WI, DCE+DWI+T_2WI, nomogram and visual assessment.

upon the DCE, DWI, and T_2WI sequences' combination. The results from the radiologists have been used to make a further assessment. According to the model built with the visual assessment achieved from the experts, the AUC value is 0.563 (95% CI, 0.470–0.772), and the accuracy is 0.611 in the testing cohort. The results of the Delong test showed that the predictive AUC values and accuracy of the radiomics prediction model with the combination of the three sequences were significantly higher than those of the prediction model built by the experience of radiologists.

DISCUSSION

In this study, we conducted a study on the relationship between the malignancy of BI-RADS 4 breast lesions and the imaging features in DCE, DWI, and T2WI MRI, and developed a radiomics prediction model based on MRI. It has been proved that the model has a stronger predictive ability than radiologists' empirical predictions and can accurately identify the benign and malignant BI-RADS 4 breast lesions, which has application value (22).

Breast MRI has the advantages of good soft-tissue resolution and no radiation and is significantly better than mammography and ultrasonography for early diagnosis and local staging of breast cancer (23, 24). Because MRI is insensitive to microcalcifications and requires a high degree of magnetic field homogeneity, it is easy to cause a false-negative diagnosis (25). Inaccuracies in visual assessment can also lead to overdiagnosis of patients. As a definition from the NCCN guidelines points out, the possibility of the occurrence of BI-RADS 4 breast lesion malignancy range from 2% to 95%, but the actual PPV of breast lesion ranges from 25.7% to 59.2% (26–28). Also, the radiomics model we developed combines multimode MRI images to provide diagnostic doctors and oncologists with a quantitative evaluation tool with greater reliability.

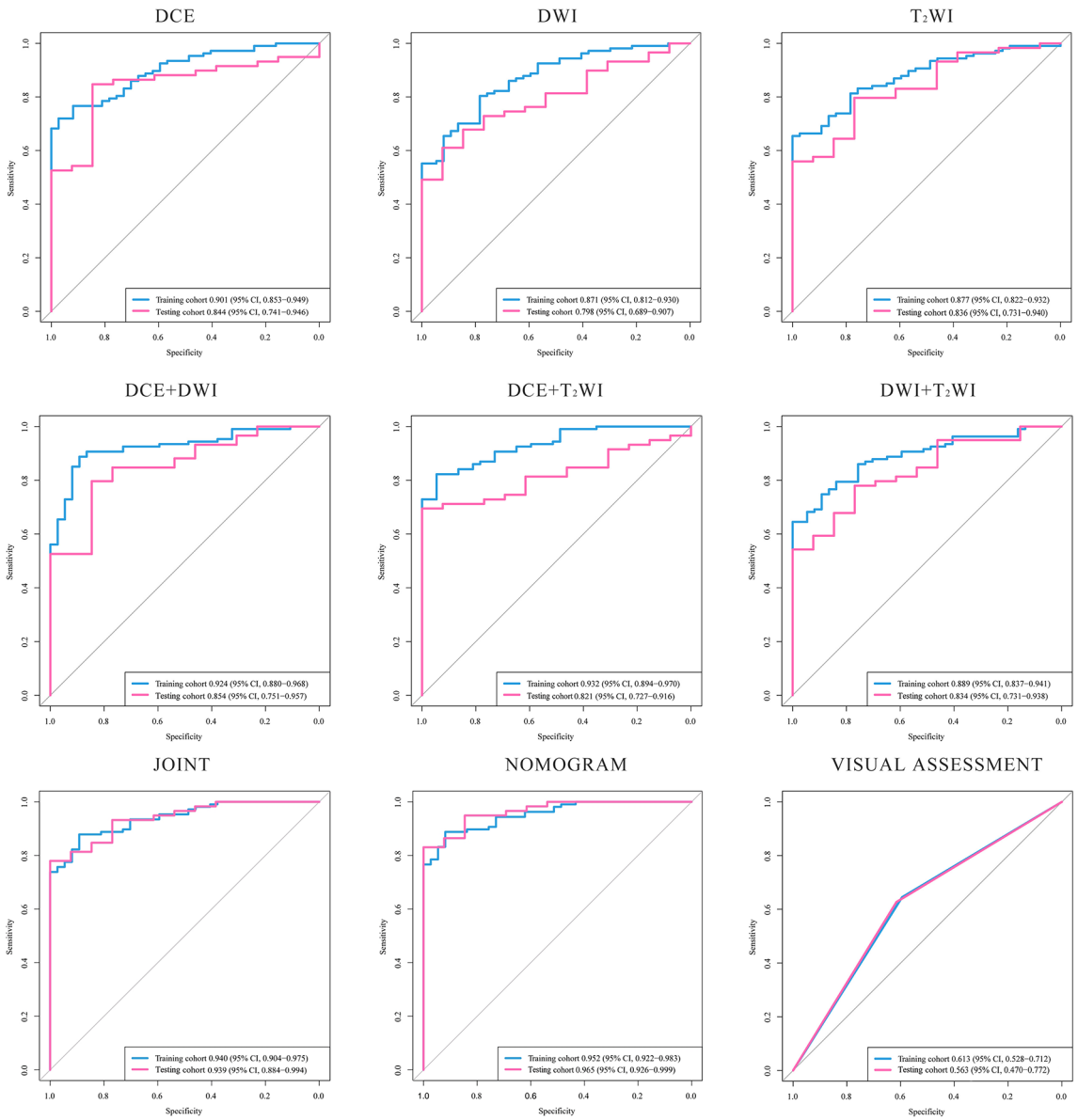


FIGURE 2 | ROC curves of the models. The ROC curves generated by models based on: DCE, DWI, T_2WI, DCE+DWI. DCE+T_2WI, DWI+T_2WI, DCE+DWI+T_2WI, nomogram combined age and the radiomic signature and the visual assessment of the radiologists. DCE, dynamic contrast enhanced imaging; DWI, diffusion weighted imaging; T_2WI, $_2$ weighted imaging.

In summary, our study focuses the concentration on developing a way to predict malignant breast lesions through imaging data without causing trauma to the patients.

Among the finally obtained features after extraction and compression by the LASSO logistic regression, the number of T_2WI image features is the largest, the coefficient value of DCE image features is the largest, and only one DWI feature is selected and the correlation coefficient of this feature is relatively small, which shows that the image features of T_2WI and DCE have a greater influence on the benignity and malignancy of breast lesions. Compared with other parametric images, T_2WI MRI images are more likely to reflect the cyst, the margins of the lesion, and the surrounding lymph nodes (29). Therefore, T_2WI images are often used to detect benign lesions (12). DCE scanning is the most common scan used to diagnose breast disease in clinical practice; it provides higher sensitivity, however, its specificity is variable (30). It can be concluded that T_2WI scanning and DCE scanning have some complementary properties, and therefore a prediction model combining both sequences can yield substantial performance improvements.

From the features we screened out, wavelet-filtered small area low gray-level emphasis texture feature in the T_2WI sequence and wavelet-filtered correlation texture feature in the DCE

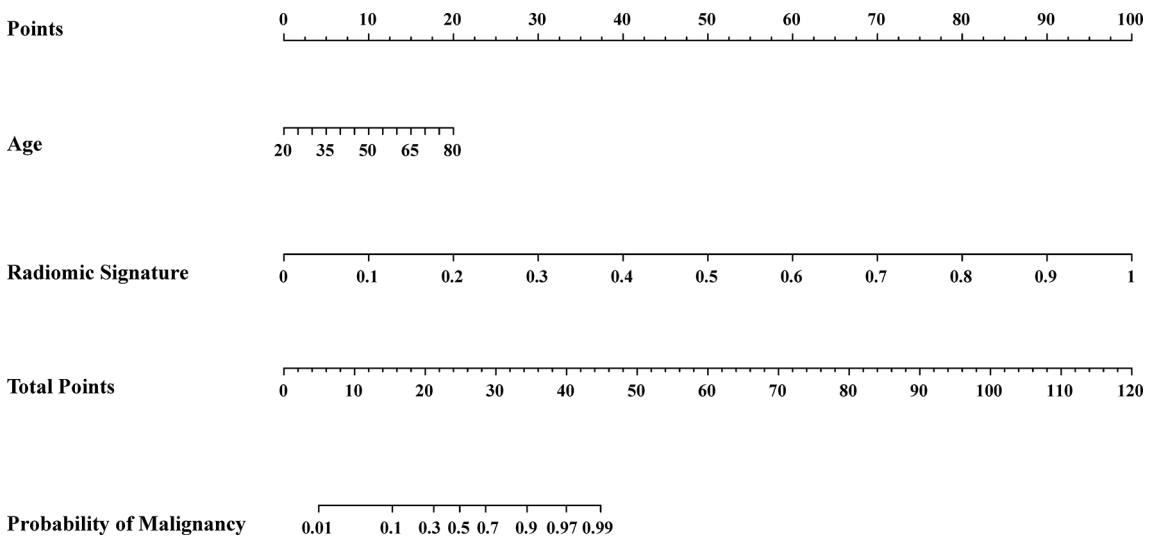


FIGURE 3 | Radiomic nomogram. The radiomic nomogram was conducted based on the patients' ages from the clinical information and the radiomic signature obtained from the best radiomic model which was based on DCE, DWI and T_2WI.

sequence was the most descriptive for breast BI-RADS4-like suspicious lesions. Small area low gray-level emphasis is one of the gray-level size zone matrix (GLSZM) features, which describe the amount of homogeneous connected areas within the volume of a certain size and intensity, thereby describing lesion heterogeneity at a regional scale (26). The grayscale area size matrix is the primary form of the Thibault matrix, which is an advanced statistical matrix of texture features and a powerful tool for medical image analysis. The more homogeneous the image texture is, the larger and flatter the matrix width is. Unlike the stroke and cooccurrence matrices, the GLSZM does not require multiple directional calculations. Specifically, GLSZM is effective in characterizing texture consistency, nonperiodic or speckled textures, and has better performance than granularity, stroke matrix, and cooccurrence matrix for cell nuclei, and dermis (27). Wavelet filtered correlation texture feature is one of the gray level cooccurrence matrix (GLCM) features, which describes the joint distribution of two pixels that have some spatial location

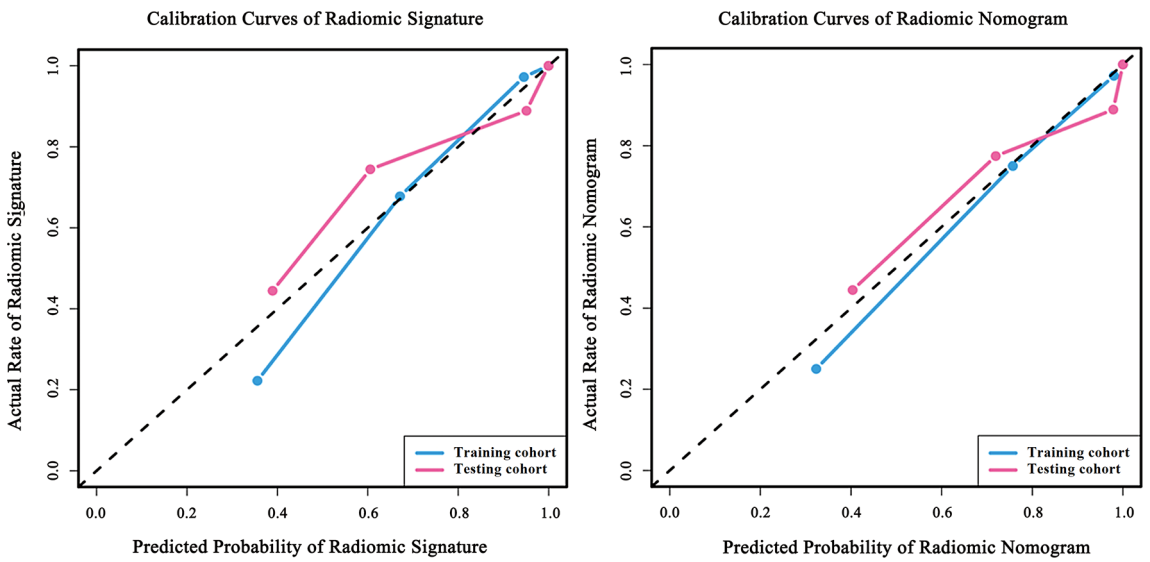


FIGURE 4 | Calibration curves of the radiomic model and the radiomic nomogram. Calibration curves of radiomic signature were built by the radiomic model based on DCE, DWI and T_2WI. Calibration curves of radiomic nomogram were built by the nomogram. The diagonal line represents the perfect prediction of the ideal model. The blue and pink lines represent the performance of the training and testing cohort in the models, where the models closer to the diagonal line represent better predictions. The calibration curves have gone through the Hosmer-Lemeshow test and have achieved a favorite result.

relationship (28). The correlation features to measure the degree of similarity of the elements of the spatial grayscale cooccurrence matrix in the row direction. Thus, the local grayscale correlation in the image can be seen from the correlation value magnitude. When the values of matrix elements are uniformly equal, the correlation value is large; contrarily, the correlation value is small when the matrix pixel values differ significantly. If there is a horizontal texture in the image, the correlation value of the horizontal matrix is larger than that of the rest of the matrixes.

Another noteworthy strength of this study is our image segmentation method. The images used in this study were automatically segmented by computer using an optimized deep learning model and then corrected and refined by 2 professional radiologists, so our regions of interest segmentation have a high degree of accuracy and precision. This shows that the image feature values we extracted in this research-based learning are also more convincing. As for image feature values are the basis for the establishment of our prediction model, so this is one of the reasons that can prove that the prediction denseness has application value.

In addition, to demonstrate that this prediction model can reach a higher level than physicians' diagnosis, we invited two radiologists who have more than 10 years of experience in breast cancer. The two doctors had no prior information about the patient's personal information and real diagnosis; they were able to determine the benignity or malignancy of the patient's breast lesions by combining only three sequences of MRI images. The results confirmed that the AUC value, accuracy, sensitivity, specificity precision, and MCC of our developed radiomics prediction model were much higher than those of the radiomics experts.

Medical imaging technology is able to captures a vast amount of information, but most of information was reported in a qualitative and quantitative way. Prospective studies indicated that the computer aided detection (CAD) system constructed using extracted and selected features can effectively distinguish benign and malignant breast lesions (31, 32). Chen et al. (33) extracted a large number of dynamic features from the temporal enhancement pattern of a tumor. Zheng et al. (34) extracted dynamic enhancement and architectural features and spatial variations of pixelwise temporal enhancements from MRI images. Although these studies had also achieved good results, due to the development of medical imaging technique, these features cannot contain rich information on tumors. In 2012, Lambin et al. (17) formally proposed radiomics which attracted the attention of many computer scientists, radiologists, and oncologists. Liu et al. (22) used deep learning to extract features from mammography-based in predicting malignancy. Karen et al. (35) used datasets that contained 64 lesions DCE-MRI images and extracted 38 radiomics features from each image to build SVM models which is able to distinguish the malignant and benign lesions. However, the datasets in these reports are all single modality. Extracting more features can discover the connection between deeper features and the training task. The higher dimensional radiomics features can more fully express tumor heterogeneity, and these features can in part describe the characteristics of breast cancer based on their usefulness, predictive power and uniqueness. In our research, we extracted 3,474 radiomics features from a combination of three sequences (DCE, DWI, T2WI) MRI images to build the model. Thus, we have achieved a satisfactory result.

Our research results confirmed that the multimodal fusion models can complement each other. When one modality cannot obtain obvious information of a single modality, another modality can provide weak supervision information for it. Among the single-sequence models, the model constructed based on DCE scanning technology has the highest AUC. DCE sequence cannot only clearly show morphologic and hemodynamics features of the lesion (36) but also have more significance for the sharpening of the shape and scope of the lesion according to various manifestations such as enhancement mode, blood supply, and cell composition of the lesion. In the dual-sequence combined model, the DCE+T2WI model has higher model evaluation performance than the single-sequence model in terms of accuracy, specificity, and sensitivity. The signal intensity of T2WI is directly related to the shape of the underlying lesion and is usually used to exclude cysts, intramammary lymph nodes, and other benign breast lesions to improve the specificity of diagnosis (37). Among all the models in this study, the best model effect is DCE+T2WI+DWI. Each model of a single sequence has advantages and disadvantages, but the fusion of the three sequences can assist each other (37, 38). By comparing the single-sequence model with the multisequence model, it was found that the multimodal fusion models had better performance in predicting the benign and malignant of BI-RADS 4 breast lesions.

Upon reflection, although this study provides significant benefits, it also has some limitations worth discussing. Firstly, this is a single-center study. Therefore, the lack of multicenter data fusion analysis might affect the generalization ability of the model to a certain extent. Secondly, this prediction model is based on three sequences of images from breast MRI and can combine the characteristics of different images for prediction. However, breast ultrasonography and mammography are also two major means of detecting breast disease (39, 40). Because of the higher cost of MRI, the more restrictive population, and the limitations of an incomplete examination, mammography, and ultrasound scans are even more routinely used as clinical tests (41, 42). So, the absence of these two medical images in this study could potentially cause our model to be less representative, as our prediction model could not be applied to these two sequences of images.

In conclusion, the model based on DCE, DWI, and T_2WI combined is the most effective in predicting the benignity and malignancy of BI-RADS 4 suspicious lesion.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because "Due to the privacy of patients, the MRI data and clinical information related to patients cannot be available for public

access". Requests to access the datasets should be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Cancer Hospital, Chinese Academy of Medical Sciences. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ReZ and WW: study concept, acquisition, analysis and interpretation of data, statistical analysis, and drafting of the article. RL and JL: acquisition, analysis and interpretation of data, statistical analysis, and drafting of the article. ZZ: study concept, acquisition, analysis and interpretation, and critical revision of article. MM and RuZ: acquisition and analysis of data. XZ: study concept and design, interpretation of data, and drafting and critical revision of the article. All authors revised the manuscript critically for important intellectual content and approved the final submitted version.

FUNDING

This work was sponsored by the National Natural Science Foundation of China (81901827), Key project of Beijing Hope Marathon Special Fund from China Cancer Foundation (LC2018A20), Natural Science Basic Research Program of Shaanxi (2020JQ-836, 2020JQ-837), Xi'an Polytechnic University (BS201987), and National Key Research and Development Plan of China (2017YFC1309102).

REFERENCES

1. Wild C, Weiderpass E, Stewart B. *World Cancer Report: Cancer Research for Cancer Prevention*. Lyon: International Agency for Research on Cancer (2020) p. 23–33.
2. Goetz MP, Gradishar WJ, Anderson BO, Gradishar WJ, Anderson BO, Abraham J, et al. NCCN Guidelines Insights: Breast Cancer, Version 3.2018: Featured Updates to the NCCN Guidelines. *J Natl Compr Cancer Net* (2019) 17(2):118–26. doi: 10.6004/jnccn.2019.0009
3. Bennani-Baiti B, Dietzel M, Baltzer PA. MRI for the Assessment of Malignancy in BI-RADS 4 Mammographic Microcalcifications. *PloS One* (2017) 12(11):e0188679. doi: 10.1371/journal.pone.0188679
4. Gradishar W, Anderson B, Balassanian R, Blair SL, Burstein H, Cyr A, et al. NCCN Guidelines Insights: Breast Cancer. *Version* (2017) 1:15. doi: 10.6004/jnccn.2017.0044
5. Kuhl CK. Abbreviated Magnetic Resonance Imaging (MRI) for Breast Cancer Screening: Rationale, Concept, and Transfer to Clinical Practice. *Annu Rev Med* (2019) 70:501–19. doi: 10.1146/annurev-med-121417-100403
6. D'Orsi C, Bassett L, Feig S. Breast Imaging Reporting and Data System (BI-RADS). New York: Oxford University Press. *Breast Imaging Atlas* (2018). doi: 10.1093/med/9780190270261.003.0005
7. Kumar V, Gu Y, Basu S, Berglund A, Eschrich S, Schabath M, et al. Radiomics: The Process and the Challenges. *Magn Reson Imaging* (2012) 30(9):1234–48. doi: 10.1016/j.mri.2012.06.010
8. Yankeelov TE, Lepage M, Chakravarthy A, Broome EE, Niermann KJ, Kelley MC, et al. Integration of Quantitative DCE-MRI and ADC Mapping to Monitor Treatment Response in Human Breast Cancer: Initial Results. *Magn Reson Imaging* (2007) 25(1):1–13. doi: 10.1016/j.mri.2006.09.006
9. Partridge SC, McDonald ES. Diffusion Weighted Magnetic Resonance Imaging of the Breast: Protocol Optimization, Interpretation, and Clinical Applications. *Magn Reson Imaging Clin* (2013) 21(3):601–24. doi: 10.1016/j.mric.2013.04.007
10. 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) 24(6):1632–42. doi: 10.1109/JBHI.2019.2956351
11. Liu Z, Meng X, Zhang H, Li Z, Liu J, Sun K, et al. Predicting Distant Metastasis and Chemotherapy Benefit in Locally Advanced Rectal Cancer. *Nat Commun* (2020) 11(1):1–11. doi: 10.1038/s41467-020-18162-9
12. 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(2):582–91. doi: 10.1007/s00330-017-5005-7
13. Liu Z, Li Z, Qu J, Zhang R, Zhou X, Li L, et al. Radiomics of Multiparametric MRI for Pretreatment Prediction of Pathologic Complete Response to Neoadjuvant Chemotherapy in Breast Cancer: A Multicenter Study. *Clin Cancer Res* (2019) 25(12):3538–47. doi: 10.1158/1078-0432.CCR-18-3190
14. Li W, Newitt DC, Gibbs J, Wilmes L, Jones E, Arasu V, et al. Predicting Breast Cancer Response to Neoadjuvant Treatment Using Multi-Feature MRI: Results From the I-SPY 2 TRIAL. *NPJ Breast Cancer* (2020) 6(1):1–6. doi: 10.1038/s41523-020-00203-7
15. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Cavalho 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
16. Van Griethuysen JJ, 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
17. 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(4):441–6. doi: 10.1016/j.ejca.2011.11.036
18. Tibshirani R, Bien J, Friedman J, Hastie T, Simon N, Taylor J, et al. Strong Rules for Discarding Predictors in Lasso-Type Problems. *J R Stat Soc: Ser B (Statistical Methodology)* (2012) 74(2):245–66. doi: 10.1111/j.1467-9868.2011.01004.x
19. Simon N, Friedman J, Hastie T, Tibshirani R. Regularization Paths for Cox's Proportional Hazards Model *via* Coordinate Descent. *J Stat Software* (2011) 39(5):1. doi: 10.18637/jss.v039.i05
20. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models *via* Coordinate Descent. *J Stat Softw* (2010) 33(1):1. doi: 10.18637/jss.v033.i01
21. Wei W, Wang K, Liu Z, Tian K, Wang L, Du J, et al. Radiomic Signature: A Novel Magnetic Resonance Imaging-Based Prognostic Biomarker in Patients With Skull Base Chordoma. *Radiother Oncol* (2019) 141:239–46. doi: 10.1016/j.radonc.2019.10.002
22. Liu H, Chen Y, Zhang Y, Wang L, Luo R, Wu H, et al. A Deep Learning Model Integrating Mammography and Clinical Factors Facilitates the Malignancy Prediction of BI-RADS 4 Microcalcifications in Breast Cancer Screening. *Eur Radiol* (2021) 31(8):5902–12. doi: 10.1007/s00330-020-07659-y
23. Bakker MF, de Lange SV, Pijnappel RM, Mann RM, Peeters PHM, Monninkhof EM, et al. Supplemental MRI Screening for Women With Extremely Dense Breast Tissue. *N Engl J Med* (2019) 381(22):2091–102. doi: 10.1056/NEJMoa1903986

24. 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. doi: 10.7150/thno.30309
25. Jiang Y, Edwards AV, Newstead GM. Artificial Intelligence Applied to Breast MRI for Improved Diagnosis. *Radiology* (2021) 298(1):38–46. doi: 10.1148/radiol.2020200292
26. Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Shape and Texture Indexes Application to Cell Nuclei Classification. *Int J Pattern Recognit Artif Intell* (2013) 27(01):1357002. doi: 10.1142/S0218001413570024
27. Thibault G, Angulo J, Meyer F. Advanced Statistical Matrices for Texture Characterization: Application to Cell Classification. *IEEE Trans Biomed Eng* (2013) 61(3):630–7. doi: 10.1109/TBME.2013.2284600
28. Hall-Beyer M. *GLCM Texture: A Tutorial V. 3.0 March 2017*. (2017) Calgary: University of Calgary Press. doi: 10.11575/PRISM/33280.
29. Nketiah G, Elschot M, Kim E, Teruel J, Scheenen T, Bathen T, et al. T2-Weighted MRI-Derived Textural Features Reflect Prostate Cancer Aggressiveness: Preliminary Results. *Eur Radiol* (2017) 27(7):3050–9. doi: 10.1007/s00330-016-4663-1
30. Gubern-Mérida A, Martí R, Melendez J, Hauth JL, Mann RM, Karssemeijer N, et al. Automated Localization of Breast Cancer in DCE-MRI. *Med Image Anal* (2015) 20(1):265–74. doi: 10.1016/j.media.2014.12.001
31. Arbash Meinel L, Stolpen AH, Berbaum KS, Fajardo LL, Reinhardt JM. Breast MRI Lesion Classification: Improved Performance of Human Readers With a Backpropagation Neural Network Computer-Aided Diagnosis (CAD) System. *J Magn Reson Imaging* (2007) 25(1):89–95. doi: 10.1002/jmri.20794
32. Vadivel A, Surendiran B. A Fuzzy Rule-Based Approach for Characterization of Mammogram Masses Into BI-RADS Shape Categories. *Comput Biol Med* (2013) 43(4):259–67. doi: 10.1016/j.compbiomed.2013.01.004
33. Chen WJ, Giger ML, Bick U, Newstead GM. Automatic Identification and Classification of Characteristic Kinetic Curves of Breast Lesions on DCE-MRI. *Med Phys* (2006) 33(8):2878–87. doi: 10.1118/1.2210568
34. Zheng YJ, Englander S, Baloch S, Zacharaki E, Fan Y, Schnall M, et al. STEP: Spatiotemporal Enhancement Pattern for MR-Based Breast Tumor Diagnosis. *Med Phys* (2009) 36(7):3192–204. doi: 10.1118/1.3151811
35. Drukker K, Anderson R, Edwards A, Papaioannou J, Pineda F, Abe H, et al. Radiomics for Ultrafast Dynamic Contrast-Enhanced Breast MRI in the Diagnosis of Breast Cancer: A Pilot Study. In: *Conference on Medical Imaging - Computer-Aided Diagnosis*, vol. Volume 10575. Houston, TX: Proceedings of SPIE (2018).
36. Zhang Q, Peng Y, Liu W, Bai J, Zheng J, Yang X, et al. Radiomics Based on Multimodal MRI for the Differential Diagnosis of Benign and Malignant Breast Lesions. *J Magn Reson Imaging* (2020) 52(2):596–607. doi: 10.1002/jmri.27098
37. Dalmis MU, Gubern-Merida A, Vreemann S, Bult P, Karssemeijer N, Mann R, et al. Artificial Intelligence-Based Classification of Breast Lesions Imaged With a Multiparametric Breast MRI Protocol With Ultrafast DCE-MRI, T2, and DWI. *Invest Radiol* (2019) 54(6):325–32. doi: 10.1097/RLI.0000000000000544
38. Dijkstra H, Dorrius MD, Wielema M, Pijnappel RM, Oudkerk M, Sijens PE. Quantitative DWI Implemented After DCE-MRI Yields Increased Specificity for BI-RADS 3 and 4 Breast Lesions. *J Magn Reson Imaging* (2016) 44(6):1642–9. doi: 10.1002/jmri.25331
39. Gary MT, Tan P-H, Cheung HS, Chu WC, Lam WW. Intermediate to Highly Suspicious Calcification in Breast Lesions: A Radio-Pathologic Correlation. *Breast Cancer Res Treat* (2008) 110(1):1–7. doi: 10.1007/s10549-007-9695-4
40. Liu Z, Zhang X-Y, Shi Y-J, Wang L, Zhu H-T, 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(23):7253–62. doi: 10.1158/1078-0432.CCR-17-1038
41. Fischer U. X-Ray Mammography: Calcifications, Masses and Architectural Distortions. *Der Radiol* (2008) 48(8):785–97, quiz 798. doi: 10.1007/s00117-008-1711-0
42. Lei C, Wei W, Liu Z, Xiong Q, Yang C, Yang M, et al. Mammography-Based Radiomic Analysis for Predicting Benign BI-RADS Category 4 Calcifications. *Eur J Radiol* (2019) 121:108711. doi: 10.1016/j.ejrad.2019.108711

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.

The reviewer KZ declared a shared parent affiliation with several of the authors, ReZ, JL, XZ, to the handling editor at time of review.

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 07 February 2022
doi: 10.3389/fonc.2022.839621

Multiparametric MRI and Machine Learning Based Radiomic Models for Preoperative Prediction of Multiple Biological Characteristics in Prostate Cancer

Xuhui Fan[1†], Ni Xie[2†], Jingwen Chen[1], Tiewen Li[3], Rong Cao[1], Hongwei Yu[1], Meijuan He[1], Zilin Wang[1], Yihui Wang[1], Hao Liu[4], Han Wang[1,2,5] and Xiaorui Yin[1*]*

[1] Department of Radiology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2] Institution for Clinical Research, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [3] Department of Urology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [4] Department of Research and Development, Yizhun Medical AI Technology Co. Ltd., Beijing, China, [5] Department of Radiology, Jiading Branch of Shanghai General Hospital, Shanghai, China

OPEN ACCESS

Edited by:
Daniel Rodriguez Gutierrez, Nottingham University Hospitals NHS Trust, United Kingdom

Reviewed by:
Felice Crocetto, Federico II University Hospital, Italy
Enrico Checcucci, Candiolo Cancer Institute (IRCCS), Italy

***Correspondence:**
Xiaorui Yin
yinxiaorui8@163.com
Han Wang
han.wang@shsmu.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: *20 December 2021*
Accepted: *11 January 2022*
Published: *07 February 2022*

Citation:
Fan X, Xie N, Chen J, Li T, Cao R, Yu H, He M, Wang Z, Wang Y, Liu H, Wang H and Yin X (2022) Multiparametric MRI and Machine Learning Based Radiomic Models for Preoperative Prediction of Multiple Biological Characteristics in Prostate Cancer. Front. Oncol. 12:839621. doi: 10.3389/fonc.2022.839621

Objectives: This study aims to develop and evaluate multiparametric MRI (MP-MRI)-based radiomic models as a noninvasive diagnostic method to predict several biological characteristics of prostate cancer.

Methods: A total of 252 patients were retrospectively included who underwent radical prostatectomy and MP-MRI examinations. The prediction characteristics of this study were as follows: Ki67, S100, extracapsular extension (ECE), perineural invasion (PNI), and surgical margin (SM). Patients were divided into training cohorts and validation cohorts in the ratio of 4:1 for each group. After lesion segmentation manually, radiomic features were extracted from MP-MRI images and some clinical factors were also included. Max relevance min redundancy (mRMR) and recursive feature elimination (RFE) based on random forest (RF) were adopted to select features. Six classifiers were included (SVM, KNN, RF, decision tree, logistic regression, XGBOOST) to find the best diagnostic performance among them. The diagnostic efficiency of the construction models was evaluated by ROC curves and quantified by AUC.

Results: RF performed best among the six classifiers for the four groups according to AUC values (Ki67 = 0.87, S100 = 0.80, ECE = 0.85, PNI = 0.82). The performance of SVM was relatively the best for SM (AUC = 0.77). The number and importance of DCE features ranked first in the models of each group. The combined models of MP-MRI and clinical characteristics showed no significant difference compared with MP-MRI models according to Delong's tests.

Conclusions: Radiomics models based on MP-MRI have the potential to predict biological characteristics and are expected to be a noninvasive method to evaluate the risk stratification of prostate cancer.

Keywords: radiomics, prostate cancer, magnetic resonance imaging, biological characteristics, risk stratification

INTRODUCTION

Prostate cancer (PCa) is the highest incidence cancer and the second leading cause of death among men according to the latest statistics in 2021 (1). Early and precise detection of prostate cancer and subsequent appropriate treatment decisions play an essential role for patients (2, 3).

Nowadays, some tumor biomarkers and biological characteristics have been proved to be useful for evaluating the malignant potential of prostate cancer and may influence the treatment decision-making (4–6). Ki67 as a biomarker of cell proliferation, which is expressed in all phases except resting (G0) phase of the cell cycle, has been demonstrated to be an independent prognostic factor in low volume and grade prostate cancer (7). According to the results of the Mayo model, when the expression of Ki67 increased by 1%, the cancer-specific mortality would increase by1 2% after radical prostatectomy (RP) accordingly (8). S100 is a family of acidic calcium-binding proteins and was found to be upregulated in various tumors (9). Aberg et al. revealed two subtypes of S100 were significantly correlated with short progression-free survival in prostate cancer patients with metastases (10). Extracapsular extension (ECE) could be used as an indication of local advanced prostate cancer (cT3). The positive ECE would increase the risk of death to 5 times than the negative ECE for patients after undergoing radical prostatectomy (11). Surgical margin (SM) is determined by pathological staining of RP specimens. Numerous studies have disclosed that positive SM increased cancer-specific mortality and the likelihood of biochemical recurrence of patients significantly (12–14). Prostate cancer tends to invade and grow along nerves, and it is also considered to be a potential metastatic route, which is called perineural invasion (PNI) (15, 16). PNI has been documented to be associated with biochemical recurrence (BCR) and promoting tumor aggressiveness (16, 17). Therefore, judging these biological characteristics before operation can better evaluate the invasiveness of tumor and may change the clinical decision-making patterns in the future.

Definitely, biopsy can solve some of the above problems to a certain extent and is still the mainstream method. It is reported that the combined technique of fusion targeted and systematic biology has been proved to be helpful to improve the diagnostic accuracy (18). However, the defects of sampling errors and a series of subsequent complications, such as pain and hematuria (19), limit the real-time monitoring and accurate evaluation of biological characteristics by biopsy. Multiparametric-magnetic resonance imaging (MP-MRI) is one of the most accurate noninvasive methods to evaluate local lesions, which contains T1 and T2 sequences that provide anatomical and disease information, as well as other sequences that provide functional information, such as diffusion-weighted imaging (DWI), dynamic contrast enhanced (DCE), and magnetic resonance spectroscopy (MRS) (20). As a routine screening method for prostate cancer, MP-MRI can reflect the phenotype and heterogeneity of prostate cancer by signal intensity and enhancement features (21, 22). Furthermore, MP-MRI images may contain many clinically valuable information related to the different biological characteristics above, such as ECE (23), which may be hard for radiologists to dig out in clinical practice.

Currently, radiomics serves as a novel approach that extracts abundant quantitative features with high throughput, and through machine learning methods to establish prediction models, which were proved to effectively provide more potential useful information for the clinical practice in urology (24, 25). Radiomics of prostate cancer has been widely used in tumor identification, staging, and prognosis evaluation (26, 27). However, more comprehensive and accurate prediction models that can determine the risk stratification and provide references for clinical decision-making for prostate cancer still need to be explored.

Thus, in the present study, we attempted to establish and validate the radiomic predictive models for five biological characteristics related to aggressiveness (Ki67, S100, ECE, PNI, SM) of prostate cancer based on MP-MRI. In addition, some clinical information was also added to establish the corresponding combined models.

MATERIALS AND METHODS

Patients, Pathological Evaluation, and MRI Acquisition

This work was approved by the Institutional Review Board (IRB) of Shanghai General Hospital (2021KY107), and the patient's informed consent was authorized to be waived according to the nature of the research. This retrospective study collected patients who underwent radical prostatectomy from May 2013 to January 2020. The exclusion criteria are as follows: (1) Preoperative DCE-MRI were unavailable (n = 149); (2) No mass lesion found on MRI image (n = 9); (3) Missing DWI (n = 3); (4) Poor imaging quality (n = 10); (5) Missing clinical information (n = 5); (6) Biopsy before MRI leading to unclear lesions (n = 32); and (7) Previous treatment before MRI examinations (n = 12). Finally, we recruited 252 patients as our subjects.

The clinical and pathological information we collected in this study is as follows: age, prostate serum antigen (PSA), white blood cell (WBC), red blood cell (RBC), hemoglobin, lymphocyte, platelet, albumin, alkaline phosphatase (ALP), platelet-to-lymphocyte ratio (PLR), fibrinogen, surgical Gleason score, immunohistochemistry (Ki67, S100, AR), SM, ECE, PNI, seminal vesicle invasion (SVI), and lymphatic vascular invasion (LVI). Details of the above indicators can be found in **Supplementary Table S2**. Considering the importance of biological characteristics of prostate cancer mentioned above, and the routine indexes of pathological examination, as well as data distribution (balanced or imbalanced), we selected five of them (Ki67, S100, ECE, PNI, SM) as our research indicators. Each of these indicators was classified as a group, and we divided each group into training cohort and validation cohort in the ratio of 4:1. All indicators were divided into positive and negative in the form of binary classification, except Ki67, which was divided into high expression and low expression with 10% as the threshold according to previous studies (7). The gold standard references of this research were based on the results of radical prostatectomy.

Three MRI sequences were included in this study: T2, DWI, DCE (arterial phase), and this combination also meets the PI-RADS v2 (Prostate Imaging-Reporting and Data System, Version 2) standard of MP-MRI (28). The protocols of MRI examinations are described in **Supplement A**.

Tumor Segmentation

The patient's images and clinical data were imported into the Darwin research platform (https://arxiv.org/abs/2009.00908) for subsequent tumor lesion delineation and model establishment. The work flow is shown in **Figure 1**. The boundary of the volume of interest (VOIs) on each axial-DWI picture was manually delineated by radiologist 1 (JC, 5 years of experience in urinary imaging). The ROIs on DWI were then copied to the sequences of T2 and DCE. If some of the copied results of the two sequences were not ideal, further modifications were made. Next, radiologist 2 (RC, 8 years of experience in urinary imaging) would review the segmentation results. If there was any objection to the results, the results would be discussed and resegmented until a consensus was reached.

Feature Extraction

After finishing segmentation, the feature extraction of lesions was carried out by PyRadiomics package. The original feature classes contain first-order, shape, and texture features. First-order features refer to the distribution of voxel intensities through general and basic metrics, such as range, mean, variance, and kurtosis. Texture features embody: Gray-Level Cooccurrence 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). Furthermore, we use eight filters and the original images were derived into eight kinds of filtered transformed images: Laplacian of Gaussian (LoG), wavelet, square, square root, logarithm, exponential, gradient, and local binary pattern (LBP). Except shape features, the first-order and texture features mentioned above can also be extracted from the derived images. Due to a single MRI sequence containing 1,781 features, MP-MRI produced 5,343 features in total for this study. The detailed description of the image features mentioned above can be found in https://pyradiomics.readthedocs.io/en/latest/features.html.

Feature Selection

Firstly, the extracted feature data were subtracted by the mean and then divided by the variance to achieve data normalization for subsequent comparison. Next, in order to reduce the over fitting of data and find the optimal correlation features, max relevance min redundancy (mRMR) was adopted to find the top 20 features for ECE, PNI, and SM groups and 15 features for Ki67 and S100 groups in the training cohort. Due to some machine learning, classifiers themselves can evaluate the importance of features and find the best feature combination through multiple iterative calculations. Therefore, the recursive feature elimination (RFE) based on random forest (RF) was applied to find the best feature combination step by step based on accuracy.

Model Construction

In this study, six classifiers were included to find the best diagnostic performance among them: support vector machine (SVM), K-nearest neighbor (KNN), random forest, decision tree, logistic regression, and XGBOOST. Support vector machine was

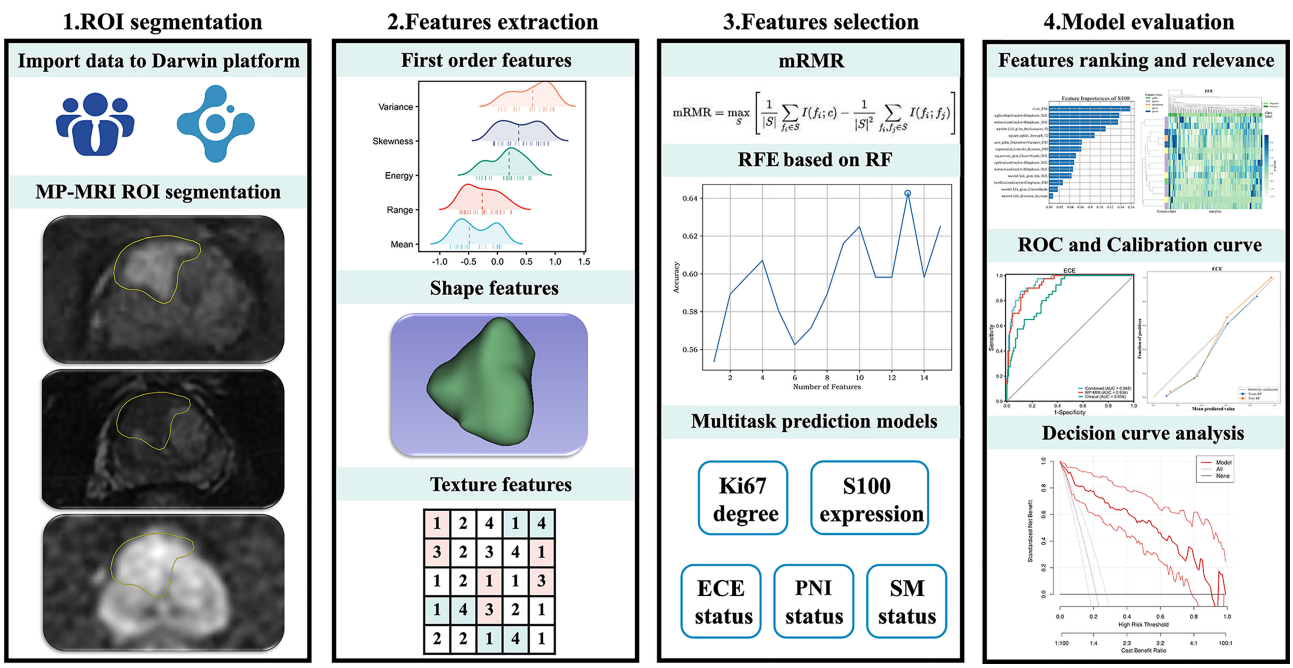


FIGURE 1 | The general workflow of this study.

based on polynomial kernel function, and the tolerance for misclassified samples was set by the specific penalty coefficient C (from 0.0001 to 1,000). The best *k* value (number of neighbors) for KNN was found by training in the range of 3–10. For random forest, decision tree, and XGBOOST, the maximum tree depth was constrained to avoid overfitting. To find out whether the clinical data improved the diagnostic performance, several clinical data (age, PSA, WBC, RBC, hemoglobin, lymphocyte, platelet, albumin, ALP, PLR, fibrinogen) were selected to build the clinical models. Meanwhile, they were integrated into the corresponding MP-MRI radiomic models to construct the combined models. The parameters used in the model construction are described in **Supplementary Table S1**.

Model Evaluation and Statistical Analysis

The diagnostic efficiency of the models was demonstrated by receiver operating characteristic (ROC) curves and quantified by the area under the curve (AUC). The calibration curve shows the consistency between the prediction model and the actuality. What is more, the decision curve analysis (DCA) illustrated the clinical net benefits brought by the prediction model.

The diagnostic ability of MP-MRI models and combined models was compared by DeLong's test. The overall comparison of PSA in each group was through Mann-Whitney *U* test. The case distribution between validation cohorts and training cohorts was compared by Chi-square test. All statistical analysis was performed by R (version 4.0.2). The statistically significant level was set at 0.05.

RESULTS

Demographics

In this study, a total of 252 PCa patients were included, and the flowchart of patients' recruitment is depicted in **Supplementary Figure S1**. The baseline characteristics of PCa patients could be found in **Supplementary Table S2**. The mean age of the included patients was 68.4 years (50–84 years), and their surgical Gleason score is mainly distributed in 7 (64.7%). According to their pathological results, the included patients were divided into 5 groups: Ki67 (n = 140), S100 (n = 158), ECE (n = 232), PNI (n = 225), and SM (n = 248). As shown in **Table 1**, the expression of PSA was significantly different in ECE ($p < 0.01$), PNI ($p = 0.03$), and SM ($p < 0.01$) groups, and it was relatively not significant in Ki67 ($p = 0.08$) and S100 ($p = 0.12$) groups. Meanwhile, the case composition between the cohort training and validation cohort was roughly the same in each group ($p > 0.05$).

Feature Selection

After applying mRMR to the features extracted from MP-MRI, the top 20 features in ECE, PNI, and SM groups and 15 features in Ki67 and S100 groups were obtained in the training cohort. RFE-RF then selected the resulting features and achieved the best performing feature combination as shown in **Figures 2A–D**: Ki67 (n = 13), S100 (n = 13), ECE (n = 15), PNI (n = 8), and SM (n = 20). For SM, all its figures are placed in **Supplementary Figure S2** for a better result demonstration. Next, principal component analysis (PCA) was performed to extract principal components of features in each group and reduce the dimensions, which made the division of cases in each group more intuitive according to their feature values. As displayed in **Figures 2E–H**, the selected features could satisfactorily distinguish the positive and negative cases on PCA (for Ki67, they were ≥10% and <10%), especially in the ECE and PNI groups, which successfully divided the cases with different labels into left sides and right sides. According to the heat maps in **Figure 3**, the color of positive cases or high Ki67 expression cases was generally darker, and the color in heatmaps referred to the values of the selected feature. This also proved the ability of the

TABLE 1 | Patient profiles of each group.

Characteristic	PSA (ng/ml)	Training cohort	Validation cohort
Ki67		*n* = 112	*n* = 28
≥10%	19.0 ± 15.4	38 (33.9%)	9 (32.1%)
<10%	15.6 ± 15.4	74 (66.1%)	19 (67.9%)
p-value	0.08	0.86	
***S100*, *n* (%)**		*n* = 126	*n* = 32
Positive	16.0 ± 11.9	67 (53.2%)	17 (53.1%)
Negative	14.7 ± 14.4	59 (46.8%)	15 (46.9%)
p-value	0.12	1.00	
ECE, *n* (%)		*n* = 185	*n* = 47
Positive	25.2 ± 22.5	40 (21.6%)	10 (21.3%)
Negative	13.9 ± 14.5	145 (78.4%)	37 (78.7%)
p-value	<0.01	0.96	
PNI, *n* (%)		*n* = 180	*n* = 45
Positive	18.3 ± 18.9	96 (53.3%)	24 (53.3%)
Negative	13.9 ± 14.3	84 (46.7%)	21 (46.7%)
p-value	0.03	1.00	
SM, *n* (%)		*n* = 198	*n* = 50
Positive	22.7 ± 22.8	129 (65.2%)	32 (64%)
Negative	12.4 ± 10.6	69 (34.8%)	18 (36%)
p-value	<0.01	0.88	

The comparison of PSA in each group was by Mann-Whitney U test. The case distribution between validation cohorts and training cohorts was compared by Chi-square test. ECE, extracapsular extension; PNI, perineural invasion; SM, surgical margins.

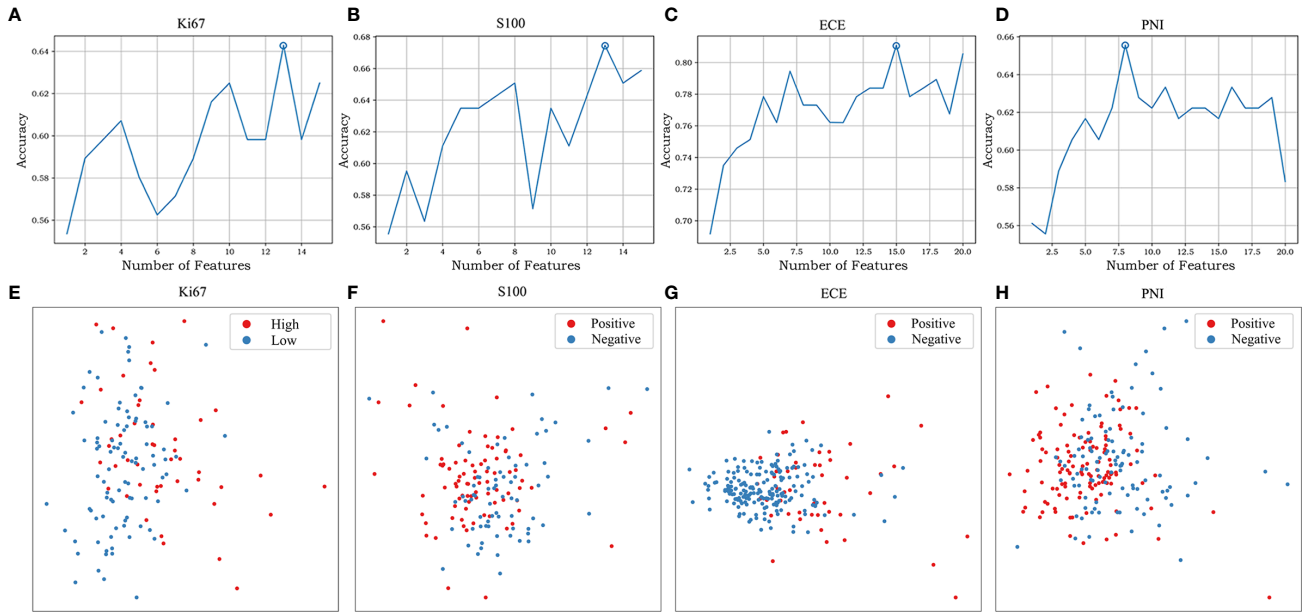


FIGURE 2 | The feature selection of RFE-RF and the distribution of different cases on PCA. RFE-RF **(A–D)** was applied to find the best feature combination step by step, and the combinations with the highest accuracy will be incorporated into the models. PCA **(E–H)** showed the selected features could satisfactorily distinguish the division of cases in each group intuitively according to their feature values. The corresponding figures for SM are shown in **Supplementary Figure S2**.

features themselves to distinguish the biological characteristics of patients. **Supplementary Figure S3** showed that the correlation among radiomics features was weak, indicating there was low redundancy among selected features.

Comparison of Different Classifiers

The six classifiers (SVM, KNN, random forest, decision tree, logistic regression, and XGBOOST) and their AUC in each group are listed in **Table 2**. In general, the performance of random forest was the best according to AUC values, and we chose random forest as the prediction models for the four groups (Ki67 = 0.87, S100 = 0.80, ECE = 0.85, PNI = 0.82). As for SM, the performance of SVM was relatively the best (AUC = 0.77), so SVM with polynomial kernel function was selected as the optimal classifier for SM.

Performance of MP-MRI and Combined Prediction Models

The optimal MP-MRI models of the five groups performed satisfactorily both in the training cohorts and validation cohorts (**Figure 4**; **Table 3**). The prediction model of Ki67 performed best among the five groups, whose AUC value reached 0.88 in the validation cohort. The second-best model was ECE with AUC value = 0.85. The AUC values of the three remaining models in the validation cohort were 0.80 for S100, 0.82 for PNI, and 0.77 for SM.

As for clinical factors, after REF-RF selection, 2 characteristics were included for Ki67 (PSA, PLR), 1 for S100 (PSA), 4 for ECE (PSA, WBC, PLR, and ALP), 5 for PNI (PLR, age, fibrinogen, PSA, and albumin), and 4 for SM (PSA, fibrinogen, albumin, and lymphocyte). This displayed the level of PSA might be helpful to distinguish five biological characteristics to some extent. Clinical characteristics were then added to the MP-MRI model to form the combined models. As a result, in the training cohort, the combined model was significantly better than the MP-MRI models except the Ki67 group based on Delong's tests ($p < 0.05$). Nevertheless, in the validation cohort, there was no significant difference between the two groups ($p > 0.05$).

Furthermore, the importance of the features in the combined models is demonstrated in **Figure 5**. The number and importance of DCE features ranked first in models of each group, followed by DWI, and finally T2. This also revealed that DCE sequences could provide more information for predicting the malignant degree of prostate cancer. In addition, in **Figure 6**, calibration curves displayed the consistency between the prediction model and the actuality was favorable, and when the risk threshold is greater than about 0.1, the prediction model could bring more clinical net benefits according to the DCA. Finally, we provide the examples of VOI delineation on MP-MRI and the corresponding 3D constructions images in **Figure 7**.

DISCUSSION

In this study, we constructed the machine-learned radiomic models based on six classifiers for the five biological characteristics (Ki67, S100, ECE, PNI, and SM) related to the invasiveness of prostate cancer. ROC curves showed that the

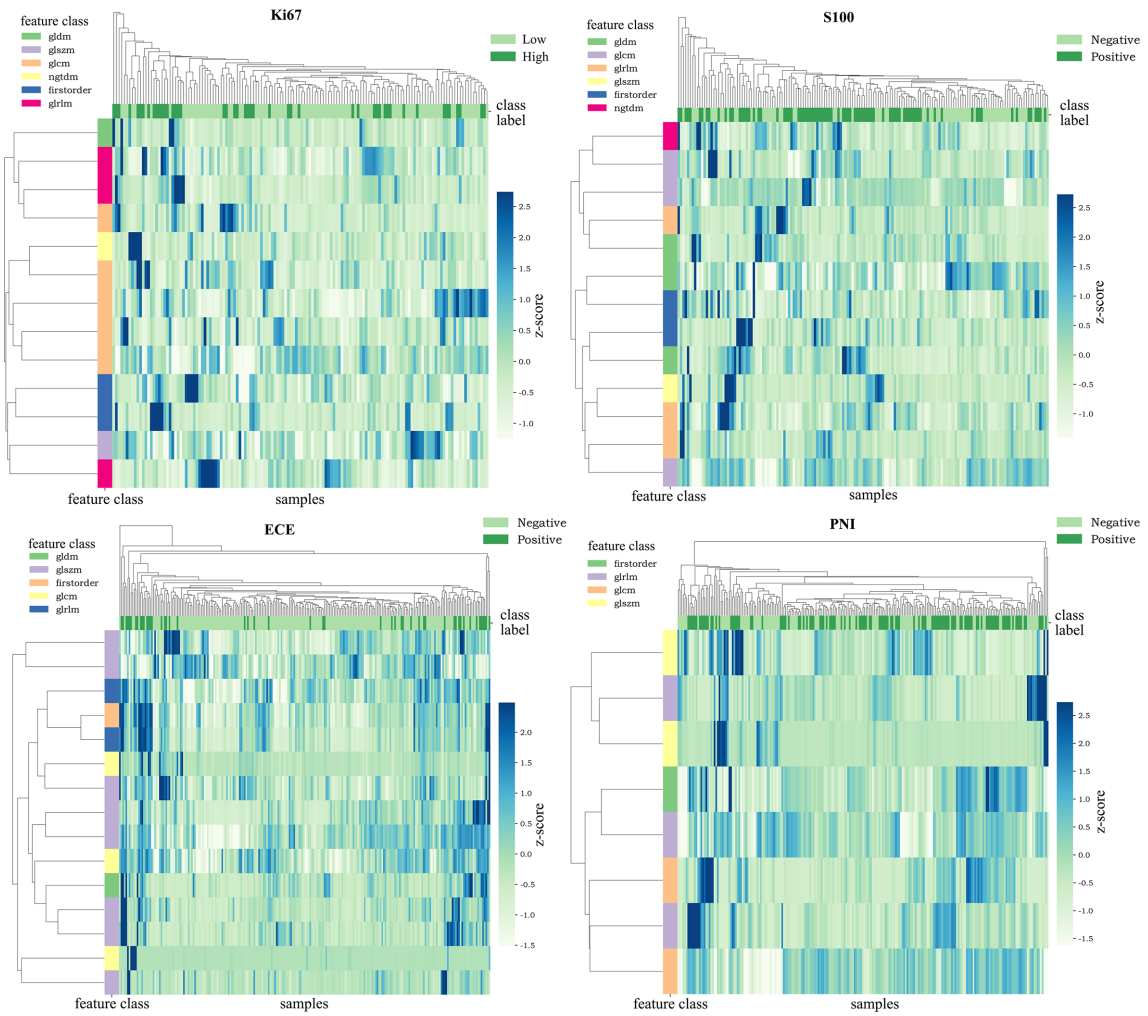


FIGURE 3 | Heat maps of the selected features. The color of the maps represented the value of the selected features. The color of positive cases or high Ki67 expression cases was generally darker. This proved the ability of the features themselves to distinguish the biological characteristics of patients.

diagnostic abilities of these models were ideal with AUC values all greater than 0.8 in the validation cohorts (except SM = 0.77). Meanwhile, we added several clinical characteristics to make the combined models, though they barely improved the accuracy of prediction.

For traditional diagnostic methods, like serum PSA test, digital rectal examinations (DREs), and prostate biopsy, they inevitably have many major deficiencies (19, 29–31). The main deficiencies are that they may lead to overdiagnosis of prostate cancer and missed diagnosis of clinically significant cancer (32–34). As shown in **Supplementary Table S2**, 22.6% of patients showed Gleason score 6 or less, suggesting that a relevant number of cancers below the threshold which is currently considered clinically significant cancer and leading to overdiagnosis and overtreatment. MP-MRI as a noninvasive method has been recommended as a routine examination of prostate cancer and proved to be beneficial in the detection of clinically significant cancer (21, 22). Recently, the NCCN guidelines clearly pointed out that MP-MRI was helpful to the staging and risk stratification of prostate cancer, and its combination with several biomarkers could reduce unnecessary biopsy (35). Moreover, MP-MRI contains much clinically valuable information, which has not attracted enough attention in clinical practice. Recently, artificial intelligence, such as radiomics, has shown great potential for evaluating the aggressiveness of urological tumors (36). Therefore, radiomics could be used as a novel and efficient way to dig out the information (24). Radiomics has been applied to predict many aspects of prostate cancer, such as cancer diagnosis, Gleason score, treatment response, and early biochemical recurrence (37).

Nowadays, using radiomics to predict multiple biological characteristics of tumors simultaneously has become a trend. In the research of Meng et al. (38), they proposed the radiomic models based on MP-MRI have the ability to predict multiple

TABLE 2 | Diagnostic performance of optimal models for each group.

Different models	Training cohort					Validation cohort				
	AUC	SEN	SPE	ACC	*p*-value	AUC	SEN	SPE	ACC	*p*-value
Ki67										
MP-MRI	0.91	0.92	0.76	0.81	0.59	0.87	1.00	0.58	0.71	0.60
Clinical	0.73	0.53	0.84	0.73		0.63	0.67	0.74	0.71	
Combined	0.91	0.92	0.76	0.81		0.88	0.78	0.84	0.82	
S100										
MP-MRI	0.88	0.81	0.81	0.81	<0.01	0.80	0.82	0.71	0.75	0.58
Clinical	0.85	0.82	0.69	0.76		0.66	0.62	0.53	0.63	
Combined	0.94	0.84	0.92	0.87		0.81	0.94	0.60	0.78	
ECE										
MP-MRI	0.93	0.88	0.86	0.86	0.01	0.85	1.00	0.62	0.70	0.91
Clinical	0.86	0.98	0.57	0.65		0.57	0.50	0.84	0.77	
Combined	0.95	0.88	0.88	0.88		0.85	0.80	0.73	0.74	
PNI										
MP-MRI	0.87	0.84	0.79	0.82	<0.01	0.82	0.67	0.95	0.80	0.19
Clinical	0.81	0.88	0.68	0.78		0.58	0.67	0.52	0.60	
Combined	0.89	0.85	0.80	0.83		0.84	0.71	0.90	0.80	
SM										
MP-MRI	0.87	0.83	0.78	0.80	0.01	0.77	0.72	0.72	0.72	0.97
Clinical	0.84	0.83	0.74	0.77		0.65	0.71	0.47	0.64	
Combined	0.94	0.96	0.81	0.86		0.77	0.61	0.81	0.74	

The p-values were derived from DeLong's test, and they compare the AUCs of the MP-MRI models with the corresponding combined model. The models of SM were based on SVM; the others were based on RF.

biological characteristics (HER-2, Ki67, differentiation, lymph node metastasis, and KRAS-2) of rectal cancers. However, to the best of our knowledge, there were fever researches to comprehensively predict various biological characteristics of prostate cancer using radiomics and achieved good diagnostic results. Bai et al. (39) reported their radiomic model could predict the presence of ECE preoperatively, but the AUC value of their integrated model was only 0.71, much lower than ours

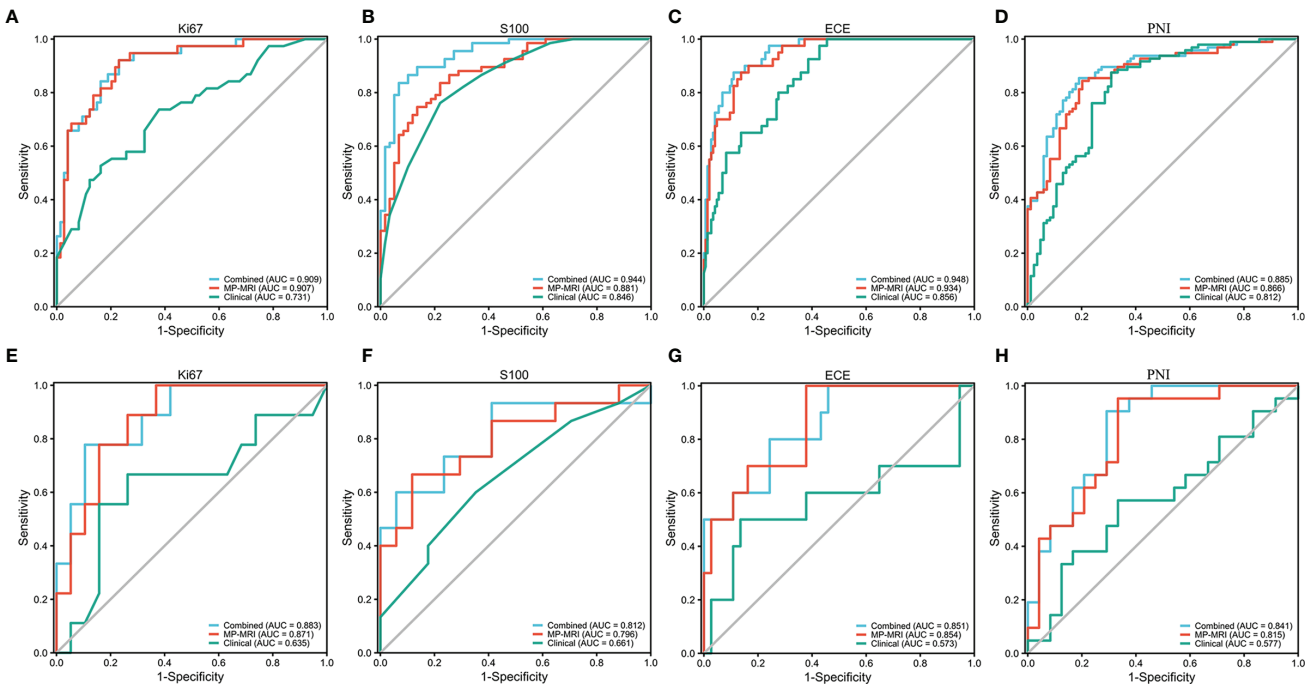


FIGURE 4 | The ROC curves of the MP-MRI, clinical and combined models in the training cohort **(A–D)** and validation cohort **(E–H)**.

TABLE 3 | AUCs of different MP-MRI radiomic classifiers for predicting the five biological characteristics in the validation cohorts.

Classifiers	*Ki67*	*S100*	ECE	PNI	SM
Random forest	**0.87**	**0.80**	**0.85**	**0.82**	0.72
Decision tree	0.75	0.76	0.77	0.72	0.75
SVM	0.84	0.79	0.84	0.78	**0.77**
KNN	0.74	0.70	0.82	0.72	0.72
Logistic regression	0.75	0.80	0.82	0.81	0.68
XGBOOST	0.76	0.70	0.74	0.73	0.67

The bold values represent the AUC of the classifiers that perform best in each subgroup.
SVM, support vector machine; KNN, K-nearest neighbor.

(AUC = 0.85). He et al. (40) used MP-MRI radiomics to predict ECE (AUC = 0.728, also lower than ours) and SM (AUC = 0.76, similar to ours), yet they did not comprehensively evaluate the aggressiveness of prostate cancer as ours. Therefore, our comprehensive radiomic models made it possible to predict more critical biological characteristics of prostate cancer and improve the prediction accuracy of some biological characteristics compared with the other published AI models.

In the present study, we extracted as many features as the recent literature documented. We then adopted an efficient feature selection method—mRMR, which has been proved advanced in a majority of researches (38, 41), to obtain the most relevant and least redundant features. In addition, the low redundancy of selected features could be testified by the correlation maps in **Supplementary Figure S3**. RFE-RF then ensured the best combinations of the included features. More and more studies use RFE-RF to select the best feature combinations, yet it needs a large amount of computation so that it is suitable for low-dimensional data after primary selection (42).

For the resulting radiomic features of each group, wavelet features account for larger proportions: 6/13 for Ki67, 5/13 for S100, 9/16 for ECE, 2/8 for PNI, and 7/20 for SM. Wavelet features are derived from the wavelet transform and represent

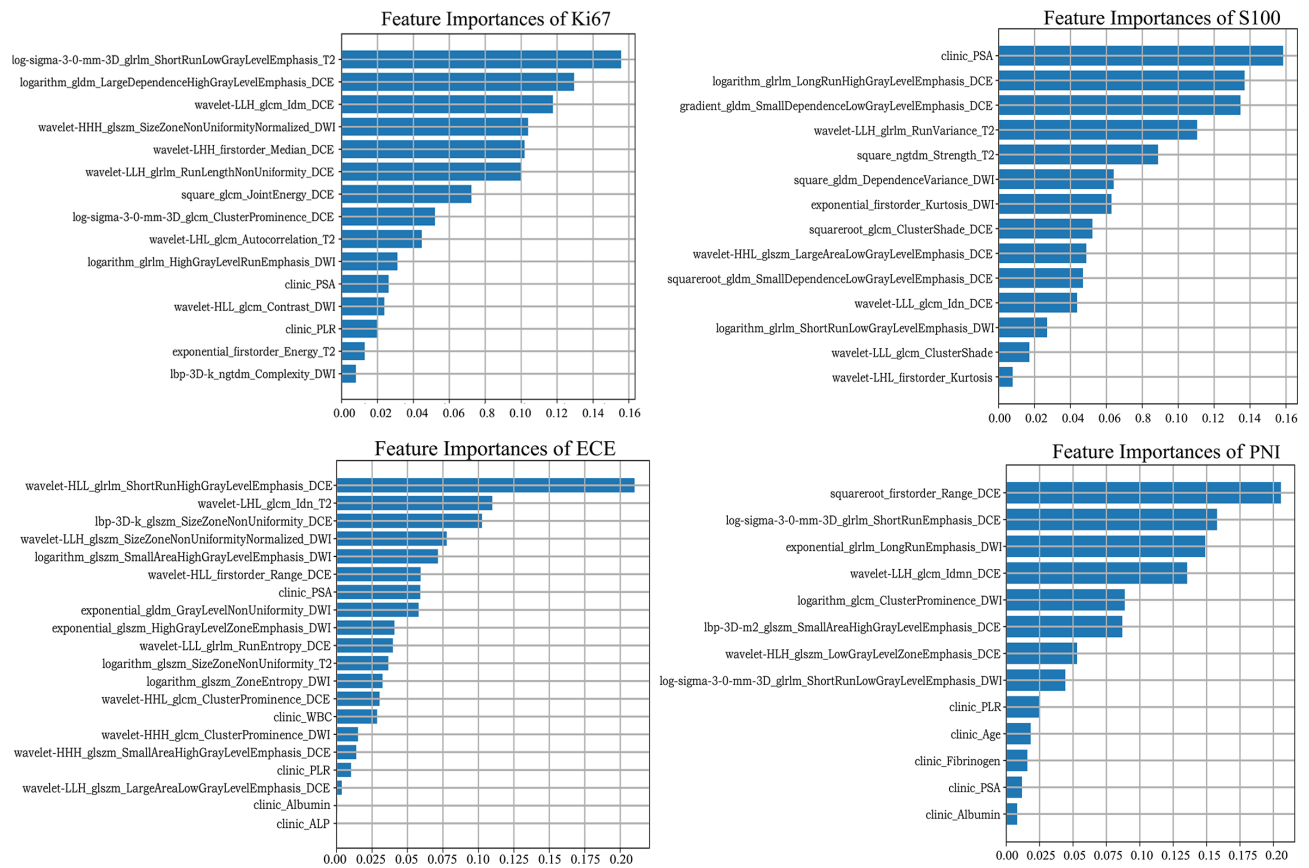


FIGURE 5 | The inbuilt feature importance in each combined model.

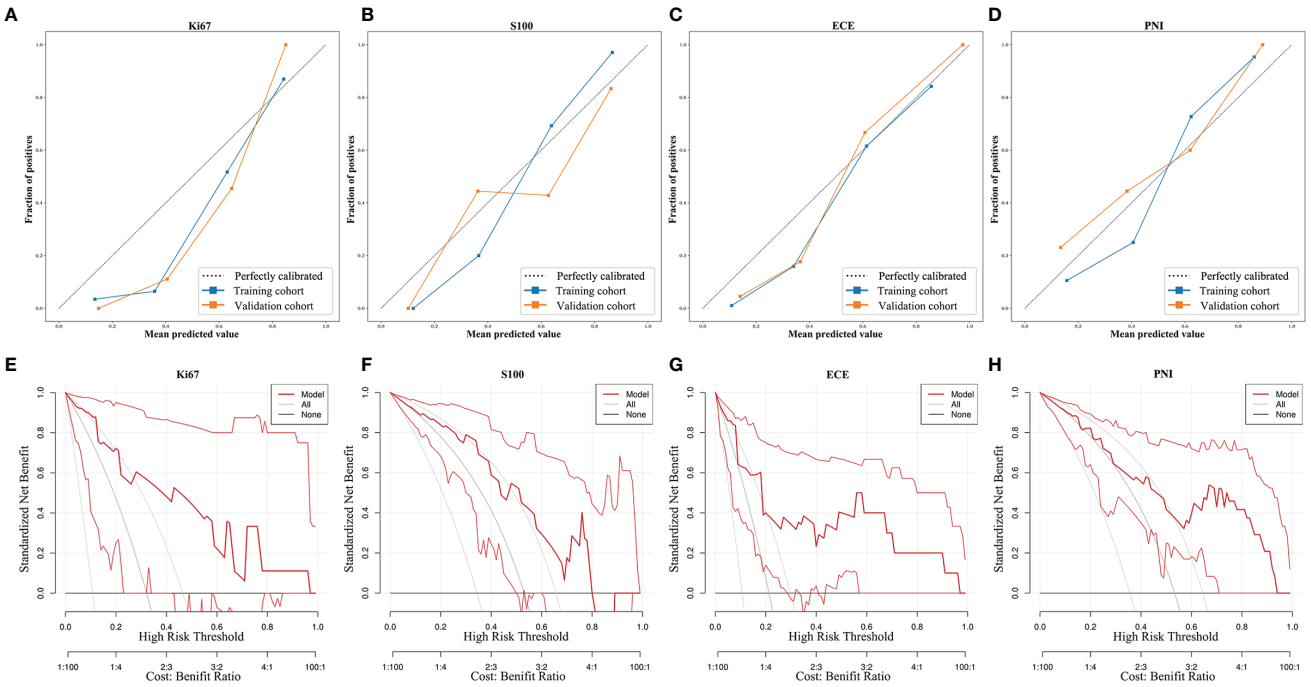


FIGURE 6 | The calibration curves and decision curve analysis of the MP-MRI models. The calibration curves **(A–D)** show the consistency between the prediction model and the actuality. The dotted reference line indicated perfect calibration. The DCA **(E–H)** illustrated the clinical net benefits brought by the prediction model. The gray line indicated "treat all," and the black horizontal line indicated "treat none".

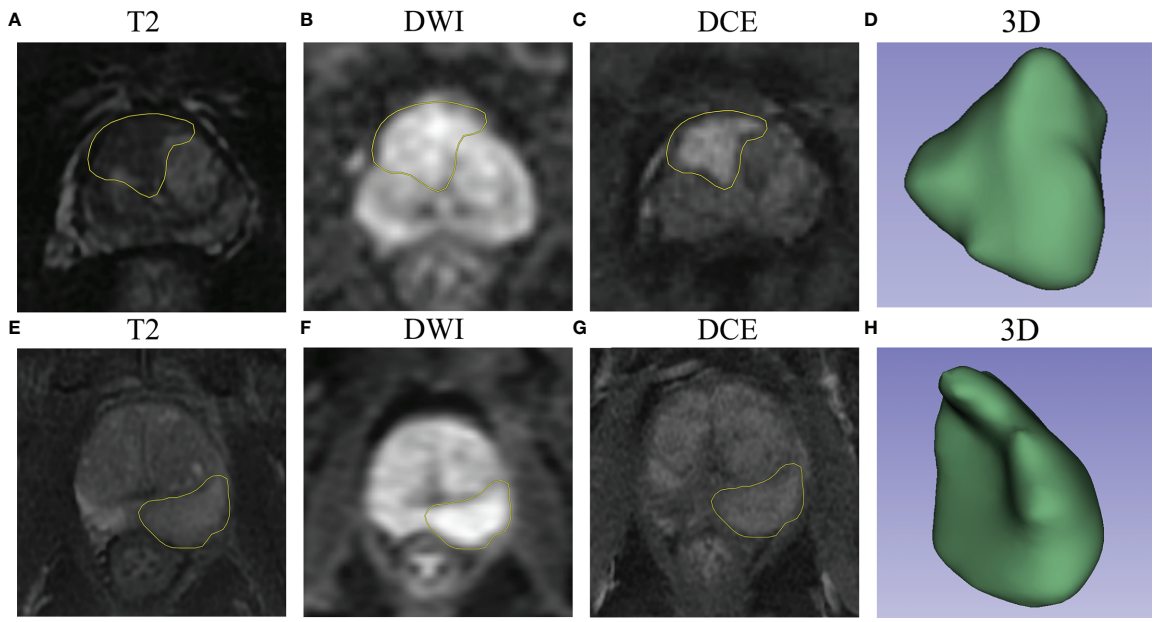


FIGURE 7 | The examples of VOI delineation on MP-MRI. **(A–D)** A 66-year-old patient was pathologically diagnosed as PNI positive with a typical abnormal signal lesion in the right front of the prostate. **(E–H)** A 70-year-old patient was diagnosed as PNI negative with a lesion located in the left peripheral zone, and the DCE sequence showed moderate enhancement. **(D, H)** represent the 3-dimensional reconstruction of the VOI.

high-dimensional features that cannot be easily deciphered by humans (43). Wavelet features could show the heterogeneity of tumors, and multiple studies have proved its strong prediction ability (38, 44). In addition, DCE sequences occupied a large part of the models. The reason may be that prostate cancer has strong ability of neovascularization, and the morphology and density of neovascularization are closely related to metastasis and prognosis of patients (45, 46). DCE-MRI is exactly a fairly adequate way to demonstrate neovascularization. Also, because of the increased vascular permeability of prostate cancer, we chose the arterial phases of DCE sequences to delineate the lesions (28).

Classifiers play an essential role in machine learning. The six classifiers that were included in this study: SVM, KNN, random forest, decision tree, logistic regression, and XGBOOST. As a result, the classification performance of random forest was generally the best among them. Random forest is composed of a large number of decision trees, and its prediction result is averaged by all the tree predictions, so it can effectively avoid over fitting. It has been documented that random forest occupied a large part of the Kaggle Data Science Competitions and ranked first among 179 classifiers (47).

It is inevitable that there are some coexisting prostatic diseases in patients with prostate cancer. In our study, coexisting diseases contained benign prostatic hyperplasia, chronic prostatitis, prostatic cyst, etc. However, they did not have a great impact on our study and our models still achieved favorable distinguishing ability. The reason may be that the features selected by our screening methods have strong specificity for the corresponding biological characteristics, and many of them reflect the complexity of the lesions, such as texture features (48). For coexisting diseases like prostatic hyperplasia, the density of lesions was relatively more consistent on MP-MRI and they would not make a remarkable difference to the accuracy of our models. The established model should be more applicable to clinical reality. If they were only applied to target diseases and excluding coexisting diseases, the clinical application of the models would be seriously limited.

Although radiomics shows huge potential for the improvement of clinical diagnosis and risk stratification, its practical clinical application is still subject to many difficulties, and its real benefits are required to be further confirmed in prospective cohort studies (49). However, radiomics plays an increasingly important role in medical imaging, and it provides a unique basis for personalized precision treatment (50). In our study, we proved the applicability of radiomics in predicting the multiple biological characteristics of prostate cancer, and we also provided relatively detailed protocol for MP-MRI and key machine-learning parameters to offer a reference for the standardization work in the future (51). The next main steps of radiomics could be to take advantage of deep learning methods (for example, U-Net) to delineate the ROI automatically and to prove the robustness of the radiomic models through multicenter, prospective, randomized-controlled trials (52).

This study had the following limitations: Firstly, this study was a retrospective and single-center study, and this inevitably led to selection bias and lack of samples and external verification. Secondly, some valuable biological characteristics or biomarkers were not included in the model due to incomplete data, for example, gene mutation data, which had great guiding significance for clinical treatment. Thirdly, our models were not as intuitive as a nomogram due to the algorithm of random forest and SVM with polynomial kernel function. Fourthly, the delineation of lesions was performed manually instead of computer-aided, which may lead to inconsistencies in clinical practice. Therefore, our next research focus will be put on multicenter, prospective, more clinically feasible, large-scale, and valuable indicator-based studies.

CONCLUSION

The present work associated the radiomics features of MP-MRI with five biological characteristics related to the aggressiveness of prostate cancer. The established comprehensive models made it possible to predict more critical biological characteristics of prostate cancer and achieved favorable prediction abilities. Therefore, the models are expected to noninvasively evaluate the risk stratification of prostate cancer and provide valuable guidance for 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 Institutional Review Board (IRB) of Shanghai General Hospital (2021KY107). 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

XY, HW, and XF designed the study. NX, TL, RC, and HY acquired the data. MH, ZW, and HL analyzed the data. XF and NX wrote the report, which was edited by all authors. YW and XY supervised the project. All authors contributed to the article and approved the submitted version.

FUNDING

This research is financially supported by the National Natural Science Foundation of China (81871400), the Project of Shanghai Science and Technology Committee (19411951403), and the Key Project of Shanghai Education Commission (202101070002E00085).

ACKNOWLEDGMENTS

We would like to thank Yizhun Medical AI Technology Co. Ltd, who kindly provided the DARWIN platform and some technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.839621/full#supplementary-material

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. *CA Cancer J Clin* (2021) 71(1):7–33. doi: 10.3322/caac.21654
2. Cuzick J, Thorat MA, Andriole G, Brawley OW, Brown PH, Culig Z, et al. Prevention and Early Detection of Prostate Cancer. *Lancet Oncol* (2014) 15 (11):e484–92. doi: 10.1016/s1470-2045(14)70211-6
3. Malik A, Srinivasan S, Batra J. A New Era of Prostate Cancer Precision Medicine. *Front Oncol* (2019) 9:1263. doi: 10.3389/fonc.2019.01263
4. Bhavsar T, McCue P, Birbe R. Molecular Diagnosis of Prostate Cancer: Are We Up to Age? *Semin Oncol* (2013) 40(3):259–75. doi: 10.1053/j.seminoncol.2013.04.002
5. Hoogland AM, Kweldam CF, van Leenders GJ. Prognostic Histopathological and Molecular Markers on Prostate Cancer Needle-Biopsies: A Review. *BioMed Res Int* (2014) 2014:341324. doi: 10.1155/2014/341324
6. Moussa AS, Li J, Soriano M, Klein EA, Dong F, Jones JS. Prostate Biopsy Clinical and Pathological Variables That Predict Significant Grading Changes in Patients With Intermediate and High Grade Prostate Cancer. *BJU Int* (2009) 103(1):43–8. doi: 10.1111/j.1464-410X.2008.08059.x
7. Zellweger T, Günther S, Zlobec I, Savic S, Sauter G, Moch H, et al. Tumour Growth Fraction Measured by Immunohistochemical Staining of Ki67 Is an Independent Prognostic Factor in Preoperative Prostate Biopsies With Small-Volume or Low-Grade Prostate Cancer. *Int J Cancer* (2009) 124(9):2116–23. doi: 10.1002/ijc.24174
8. Tollefson MK, Karnes RJ, Kwon ED, Lohse CM, Rangel LJ, Mynderse LA, et al. Prostate Cancer Ki-67 (MIB-1) Expression, Perineural Invasion, and Gleason Score as Biopsy-Based Predictors of Prostate Cancer Mortality: The Mayo Model. *Mayo Clin Proc* (2014) 89(3):308–18. doi: 10.1016/j.mayocp.2013.12.001
9. Salama I, Malone PS, Mihaimeed F, Jones JL. A Review of the S100 Proteins in Cancer. *Eur J Surg Oncol* (2008) 34(4):357–64. doi: 10.1016/j.ejso.2007.04.009
10. Åberg AM, Bergström SH, Thysell E, Tjon-Kon-Fat LA, Nilsson JA, Widmark A, et al. High Monocyte Count and Expression of S100A9 and S100A12 in Peripheral Blood Mononuclear Cells Are Associated With Poor Outcome in Patients With Metastatic Prostate Cancer. *Cancers (Basel)* (2021) 13(10):2424. doi: 10.3390/cancers13102424
11. Bill-Axelson A, Holmberg L, Garmo H, Taari K, Busch C, Nordling S, et al. Radical Prostatectomy or Watchful Waiting in Prostate Cancer - 29-Year Follow-Up. *N Engl J Med* (2018) 379(24):2319–29. doi: 10.1056/NEJMoa1807801
12. Zhang L, Wu B, Zha Z, Zhao H, Yuan J, Jiang Y, et al. Surgical Margin Status and Its Impact on Prostate Cancer Prognosis After Radical Prostatectomy: A Meta-Analysis. *World J Urol* (2018) 36(11):1803–15. doi: 10.1007/s00345-018-2333-4
13. Zhang B, Zhou J, Wu S, Guo M, Zhang Y, Liu R. The Impact of Surgical Margin Status on Prostate Cancer-Specific Mortality After Radical Prostatectomy: A Systematic Review and Meta-Analysis. *Clin Transl Oncol* (2020) 22(11):2087–96. doi: 10.1007/s12094-020-02358-y
14. Meeks JJ, Eastham JA. Radical Prostatectomy: Positive Surgical Margins Matter. *Urol Oncol* (2013) 31(7):974–9. doi: 10.1016/j.urolonc.2011.12.011
15. Liebig C, Ayala G, Wilks JA, Berger DH, Albo D. Perineural Invasion in Cancer: A Review of the Literature. *Cancer* (2009) 115(15):3379–91. doi: 10.1002/cncr.24396
16. Zareba P, Flavin R, Isikbay M, Rider JR, Gerke TA, Finn S, et al. Perineural Invasion and Risk of Lethal Prostate Cancer. *Cancer Epidemiol Biomarkers Prev* (2017) 26(5):719–26. doi: 10.1158/1055-9965.Epi-16-0237
17. Zhang LJ, Wu B, Zha ZL, Qu W, Zhao H, Yuan J, et al. Perineural Invasion as an Independent Predictor of Biochemical Recurrence in Prostate Cancer Following Radical Prostatectomy or Radiotherapy: A Systematic Review and Meta-Analysis. *BMC Urol* (2018) 18(1):5. doi: 10.1186/s12894-018-0319-6
18. Rapisarda S, Bada M, Crocetto F, Barone B, Arcaniolo D, Polara A, et al. The Role of Multiparametric Resonance and Biopsy in Prostate Cancer Detection: Comparison With Definitive Histological Report After Laparoscopic/Robotic Radical Prostatectomy. *Abdom Radiol (NY)* (2020) 45(12):4178–84. doi: 10.1007/s00261-020-02798-8
19. Rosario DJ, Lane JA, Metcalfe C, Donovan JL, Doble A, Goodwin L, et al. Short Term Outcomes of Prostate Biopsy in Men Tested for Cancer by Prostate Specific Antigen: Prospective Evaluation Within ProtecT Study. *BMJ* (2012) 344:d7894. doi: 10.1136/bmj.d7894
20. Wibmer AG, Vargas HA, Hricak H. Role of MRI in the Diagnosis and Management of Prostate Cancer. *Future Oncol* (2015) 11(20):2757–66. doi: 10.2217/fon.15.206
21. Stabile A, Giganti F, Rosenkrantz AB, Taneja SS, Villeirs G, Gill IS, et al. Multiparametric MRI for Prostate Cancer Diagnosis: Current Status and Future Directions. *Nat Rev Urol* (2020) 17(1):41–61. doi: 10.1038/s41585-019-0212-4
22. Patel P, Wang S, Siddiqui MM. The Use of Multiparametric Magnetic Resonance Imaging (mpMRI) in the Detection, Evaluation, and Surveillance of Clinically Significant Prostate Cancer (csPCa). *Curr Urol Rep* (2019) 20 (10):60. doi: 10.1007/s11934-019-0926-0
23. Checcucci E, Amparore D, De Luca S, Autorino R, Fiori C, Porpiglia F. Precision Prostate Cancer Surgery: An Overview of New Technologies and Techniques. *Minerva Urol Nefrol* (2019) 71(5):487–501. doi: 10.23736/s0393-2249.19.03365-4
24. 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
25. Checcucci E, Autorino R, Cacciamani GE, Amparore D, De Cillis S, Piana A, et al. Artificial Intelligence and Neural Networks in Urology: Current Clinical Applications. *Minerva Urol Nefrol* (2020) 72(1):49–57. doi: 10.23736/s0393-2249.19.03613-0
26. Smith CP, Czarniecki M, Mehralivand S, Stoyanova R, Choyke PL, Harmon S, et al. Radiomics and Radiogenomics of Prostate Cancer. *Abdom Radiol (NY)* (2019) 44(6):2021–9. doi: 10.1007/s00261-018-1660-7
27. Sun Y, Reynolds HM, Parameswaran B, Wraith D, Finnegan ME, Williams S, et al. Multiparametric MRI and Radiomics in Prostate Cancer: A Review. *Australas Phys Eng Sci Med* (2019) 42(1):3–25. doi: 10.1007/s13246-019-00730-z
28. Weinreb JC, Barentsz JO, Choyke PL, Cornud F, Haider MA, Macura KJ, et al. PI-RADS Prostate Imaging - Reporting and Data System: 2015, Version 2. *Eur Urol* (2016) 69(1):16–40. doi: 10.1016/j.eururo.2015.08.052
29. Howrey BT, Kuo YF, Lin YL, Goodwin JS. The Impact of PSA Screening on Prostate Cancer Mortality and Overdiagnosis of Prostate Cancer in the United States. *J Gerontol A Biol Sci Med Sci* (2013) 68(1):56–61. doi: 10.1093/gerona/gls135
30. Schröder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S, Nelen V, et al. Screening and Prostate-Cancer Mortality in a Randomized European Study. *N Engl J Med* (2009) 360(13):1320–8. doi: 10.1056/NEJMoa0810084
31. Kasivisvanathan V, Rannikko AS, Borghi M, Panebianco V, Mynderse LA, Vaarala MH, et al. MRI-Targeted or Standard Biopsy for Prostate-Cancer Diagnosis. *N Engl J Med* (2018) 378(19):1767–77. doi: 10.1056/NEJMoa1801993
32. Chang AJ, Autio KA, Roach M3rd, Scher HI. High-Risk Prostate Cancer-Classification and Therapy. *Nat Rev Clin Oncol* (2014) 11(6):308–23. doi: 10.1038/nrclinonc.2014.68
33. Cowan T, Baker E, McCray G, Reeves F, Houlihan K, Johns-Putra L. Detection of Clinically Significant Cancer in the Anterior Prostate by Transperineal Biopsy. *BJU Int* (2020) 126 Suppl 1:33–7. doi: 10.1111/bju.15124
34. Bangma CH, Roobol MJ. Defining and Predicting Indolent and Low Risk Prostate Cancer. *Crit Rev Oncol Hematol* (2012) 83(2):235–41. doi: 10.1016/j.critrevonc.2011.10.003

35. Carroll PH, Mohler JL. NCCN Guidelines Updates: Prostate Cancer and Prostate Cancer Early Detection. *J Natl Compr Canc Netw* (2018) 16(5s):620–3. doi: 10.6004/jnccn.2018.0036
36. Checcucci E, De Cillis S, Granato S, Chang P, Afyouni AS, Okhunov Z. Applications of Neural Networks in Urology: A Systematic Review. *Curr Opin Urol* (2020) 30(6):788–807. doi: 10.1097/mou.0000000000000814
37. Suarez-Ibarrola R, Hein S, Reis G, Gratzke C, Miernik A. Current and Future Applications of Machine and Deep Learning in Urology: A Review of the Literature on Urolithiasis, Renal Cell Carcinoma, and Bladder and Prostate Cancer. *World J Urol* (2020) 38(10):2329–47. doi: 10.1007/s00345-019-03000-5
38. 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
39. Bai H, Xia W, Ji X, He D, Zhao X, Bao J, et al. Multiparametric Magnetic Resonance Imaging-Based Peritumoral Radiomics for Preoperative Prediction of the Presence of Extracapsular Extension With Prostate Cancer. *J Magn Reson Imaging* (2021) 54(4):1222–30. doi: 10.1002/jmri.27678
40. He D, Wang X, Fu C, Wei X, Bao J, Ji X, et al. MRI-Based Radiomics Models to Assess Prostate Cancer, Extracapsular Extension and Positive Surgical Margins. *Cancer Imaging* (2021) 21(1):46. doi: 10.1186/s40644-021-00414-6
41. De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B. mRMRe: An R Package for Parallelized mRMR Ensemble Feature Selection. *Bioinformatics* (2013) 29(18):2365–8. doi: 10.1093/bioinformatics/btt383
42. Tana MM, McCoy D, Lee B, Patel R, Lin J, Ohliger MA. Texture Features From Computed Tomography Correlate With Markers of Severity in Acute Alcohol-Associated Hepatitis. *Sci Rep* (2020) 10(1):17980. doi: 10.1038/s41598-020-74599-4
43. Bhattacharjee S, Kim CH, Park HG, Prakash D, Madusanka N, Cho NH, et al. Multi-Features Classification of Prostate Carcinoma Observed in Histological Sections: Analysis of Wavelet-Based Texture and Colour Features. *Cancers (Basel)* (2019) 11(12):1937. doi: 10.3390/cancers11121937
44. 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
45. Berman RM, Brown AM, Chang SD, Sankineni S, Kadakia M, Wood BJ, et al. DCE MRI of Prostate Cancer. *Abdom Radiol (NY)* (2016) 41(5):844–53. doi: 10.1007/s00261-015-0589-3
46. Mucci LA, Powolny A, Giovannucci E, Liao Z, Kenfield SA, Shen R, et al. Prospective Study of Prostate Tumor Angiogenesis and Cancer-Specific Mortality in the Health Professionals Follow-Up Study. *J Clin Oncol* (2009) 27(33):5627–33. doi: 10.1200/jco.2008.20.8876
47. Fernandez-Delgado M, Cernadas E, Barro S, Amorim D. Do We Need Hundreds of Classifiers to Solve Real World Classification Problems? *J Mach Learn Res* (2014) 15:3133–81. doi: 10.1117/1.JRS.11.015020
48. Xu X, Zhang X, Tian Q, Zhang G, Liu Y, Cui G, et al. Three-Dimensional Texture Features From Intensity and High-Order Derivative Maps for the Discrimination Between Bladder Tumors and Wall Tissues *via* MRI. *Int J Comput Assist Radiol Surg* (2017) 12(4):645–56. doi: 10.1007/s11548-017-1522-8
49. Ferro M, de Cobelli O, Vartolomei MD, Lucarelli G, Crocetto F, Barone B, et al. Prostate Cancer Radiogenomics-From Imaging to Molecular Characterization. *Int J Mol Sci* (2021) 22(18):9971. doi: 10.3390/ijms22189971
50. 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
51. Papadimitroulas P, Brocki L, Christopher Chung N, Marchadour W, Vermet F, Gaubert L, et al. Artificial Intelligence: Deep Learning in Oncological Radiomics and Challenges of Interpretability and Data Harmonization. *Phys Med* (2021) 83:108–21. doi: 10.1016/j.ejmp.2021.03.009
52. Chaddad A, Kucharczyk MJ, Cheddad A, Clarke SE, Hassan L, Ding S, et al. Magnetic Resonance Imaging Based Radiomic Models of Prostate Cancer: A Narrative Review. *Cancers (Basel)* (2021) 13(3):552. doi: 10.3390/cancers13030552

Conflict of Interest: Author HL was employed by Yizhun Medical Technology Co. Ltd., Beijing, 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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 22 February 2022
doi: 10.3389/fonc.2022.802234

Radiomics Analysis of Multiparametric MRI for Prediction of Synchronous Lung Metastases in Osteosarcoma

Zhendong Luo[1†], Jing Li[2†], YuTing Liao[3], RengYi Liu[4], Xinping Shen[1] and Weiguo Chen[4*]*

[1] Department of Radiology, The University of Hong Kong - Shenzhen Hospital, Shenzhen, China, [2] Department of Radiology, State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, China, [3] Department of Pharmaceuticals Diagnosis, GE Healthcare, Shanghai, China, [4] Department of Radiology, Nanfang Hospital, Southern Medical University, Guangzhou, China

OPEN ACCESS

Edited by:
Antonio Napolitano, Bambino Gesù Children's Hospital (IRCCS), Italy

Reviewed by:
Weiwei Zong, Henry Ford Health System, United States
Fu Shen, Naval Medical University, China

***Correspondence:**
Xinping Shen
szshenxinping@163.com
Weiguo Chen
chenweiguo1964@21cn.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: *26 October 2021*
Accepted: *31 January 2022*
Published: *22 February 2022*

Citation:
Luo Z, Li J, Liao Y, Liu R, Shen X and Chen W (2022) Radiomics Analysis of Multiparametric MRI for Prediction of Synchronous Lung Metastases in Osteosarcoma. Front. Oncol. 12:802234. doi: 10.3389/fonc.2022.802234

Purpose: To establish and verify a predictive model involving multiparameter MRI and clinical manifestations for predicting synchronous lung metastases (SLM) in osteosarcoma.

Materials and Methods: Seventy-eight consecutive patients with osteosarcoma (training dataset, n = 54; validation dataset, n = 24) were enrolled in our study. MRI features were extracted from the T1-weighted image (T1WI), T2-weighted image (T2WI), and contrast-enhanced T1-weighted image (CE-T1WI) of each patient. Least absolute shrinkage and selection operator (LASSO) regression and multifactor logistic regression were performed to select key features and build radiomics models in conjunction with logistic regression (LR) and support vector machine (SVM) classifiers. Eight individual models based on T1WI, T2WI, CE-T1WI, T1WI+T2WI, T1WI+CE-T1WI, T2WI+CE-T1WI, T1WI+T2WI+CE-T1WI, and clinical features, as well as two combined models, were built. The area under the receiver operating characteristic curve (AUC), sensitivity and specificity were employed to assess the different models.

Results: Tumor size was the most significant univariate clinical indicator (1). The AUC values of the LR predictive model based on T1WI, T2WI, CE-T1WI, T1WI+T2WI, T1WI +CE-T1WI, T2WI+CE-T1WI, and T1WI+T2WI+CE-T1WI were 0.686, 0.85, 0.87, 0.879, 0.736, 0.85, and 0.914, respectively (2). The AUC values of the SVM predictive model based on T1WI, T2WI, CE-T1WI, T1WI+T2WI, T1WI +CE-T1WI, T2WI +CE-T1WI, and T1WI+T2WI+CE-T1WI were 0.629, 0.829, 0.771, 0.879, 0.643, 0.829, and 0.929, respectively (3). The AUC values of the clinical, combined 1 (clinical and LR-radiomics) and combined 2 (clinical and SVM-radiomics) predictive models were 0.779, 0.957, and 0.943, respectively.

Conclusion: The combined model exhibited good performance in predicting osteosarcoma SLM and may be helpful in clinical decision-making.

Keywords: radiomics, predictive value of tests, magnetic resonance imaging, osteosarcoma, metastasis

INTRODUCTION

Osteosarcoma is a highly prevalent primary bone malignancy. Fortunately, complete ablation of nonmetastatic high-grade osteosarcoma is possible in 60–70% of cases when treated with adjuvant and neoadjuvant multiagent chemotherapies in addition to surgery (1). However, the prognoses of osteosarcoma patients with distant metastasis remain poor. Among all forms of metastasis, lung metastasis is the most common, occurring in over 80% of patients. Approximately 20% of osteosarcoma patients also exhibit metastasis at initial diagnosis (synchronous metastases) (2, 3). The primary tumor is more resistant to chemotherapy in patients with synchronous metastases than in patients with localized disease at presentation (4). Both the number of nodules and lobes are strong indicators of survival (5). At present, the best indicators of survival are tumor grade, tumor size, and distal metastases, which can be detected from biopsies and microscopic evaluations (6). Predicting individual and early metastases is essential to osteosarcoma management, as it informs treatment strategies and increases survival rates. Chest computerized tomography (CT) has been the most commonly used imaging modality for the detection of lung nodules. Although there have been great advancements in imaging technology, particularly in enhancing the sensitivity of detection, the specificity of the data remains insufficient. Metastases cannot be properly distinguished from benign tissue (5). When nodules are detected at diagnosis, it is usually assumed that these nodules represent metastatic disease. However, not all pulmonary nodules that develop during tumor therapy are malignant, which poses additional challenges for physicians. Hence, the goal of this study was to evaluate the diagnostic abilities of two distinct classifiers (logistic regression (LR) and support vector machine (SVM)) and radiomics features retrieved from different magnetic resonance imaging (MRI) parameters, including T1-weighted imaging (T1WI), T2-weighted imaging (T2WI) and contrast-enhanced T1-weighted imaging (CE-T1WI), and the combinations of two and three of these parameters. We also developed and validated combined models according to multiparametric MRI and clinical features to predict synchronous lung metastases (SLM) in osteosarcoma.

MATERIALS AND METHODS

Our retrospective investigation was approved by the Institutional Review Board. Participant informed consent was waived due to the retrospective nature of the study.

Patient Selection

Overall, 360 patients who received MRI evaluations between January 1, 2014, and December 30, 2020, were recruited for this study. The following patients were included in the study: (i) patients with no history of surgical or medical treatment administered for suspected osteosarcoma; (ii) patients who underwent multiparametric MRI, including T1WI, T2WI, and CE-T1WI, prior to treatment; (iii) patients with a osteosarcoma diagnosis confirmed by surgical resection or CT/ultrasound-guided needle biopsy and histopathological results; and (iv) patients diagnosed with SLM according to follow-up chest CT or confirmed by pathology. The patients in this study had lung nodules, and the possibility of viral, bacterial or fungal infection was ruled out. Three criteria were used to identify SLM lung nodules on follow-up chest CT according to previous studies (7, 8): first, the presence of multiple round nodules with or without changes in size or number; second, nodule size ≥5 mm and the presence of calcifications or ossification that remained stable or increased in size relative to the initial chest CT; third, changes in size or morphology during chemotherapy. Fifteen and 18 patients with SLM were diagnosed by biopsy and follow-up chest CT, respectively. The following patients were excluded from this study: (i) patients who received biopsy and locoregional therapy before MRI; (ii) patients with low-quality images rendering analysis difficult (such as images with metallic artefacts or motion artefacts); and (iii) patients with missing images or relevant sequences. A schematic diagram of our patient selection process is provided in **Figure 1**. After the application of these criteria, 78 patients were eligible for this study. The clinical characteristics of the 78 osteosarcoma patients divided into non-SLM and SLM groups are shown in **Table 1**. We next arbitrarily divided the patients into two populations: 54 patients were placed in the training cohort (TC) and 24 were placed in the validation cohort (VC) based on the seed point set obtained from programming. The clinical characteristics of the 78 osteosarcoma patients in the TC and VC are shown in **Table 2**, and a further breakdown of the clinical characteristics of these cohorts in terms of SLM and lack of SLM of osteosarcoma are summarized in **Table 3**.

Patient clinical features, such as age, sex, tumor size, pathological type, tumor location, bone destruction type, and alkaline phosphatase (ALP) and lactate dehydrogenase (LDH) levels, were recorded.

MR Imaging

All MR imaging was conducted with 1.5- or 3.0-T superconducting magnet systems. The imaging sequences included axial T1WI, T2WI and CE-T1WI. The detailed scan parameters of the four MRI scanners are described in **Table 4**. Gadolinium contrast agent was intravenously administered *via* a weight-based dosing protocol (0.1 mmol/kg) at an injection rate of 2.5 mL/s. All the MR data were obtained from the picture archiving and communication system (PACS) of our institutes and stored in Digital Imaging and Communications in Medicine (DICOM) format for additional analyses.

Preprocessing of MR Images

All the images were exported to ITK-SNAP software (version 3.8.0, http://www.itksnap.org/) for segmentation before radiomics analysis. Lesion segmentation was performed by a radiologist with over 5 years of MRI diagnostic experience, and proper segmentation was further confirmed by a separate radiologist with over 10 years of MRI diagnostic experience. If disagreements arose about a specific image segmentation, a revision was made by two radiologists after discussion. The

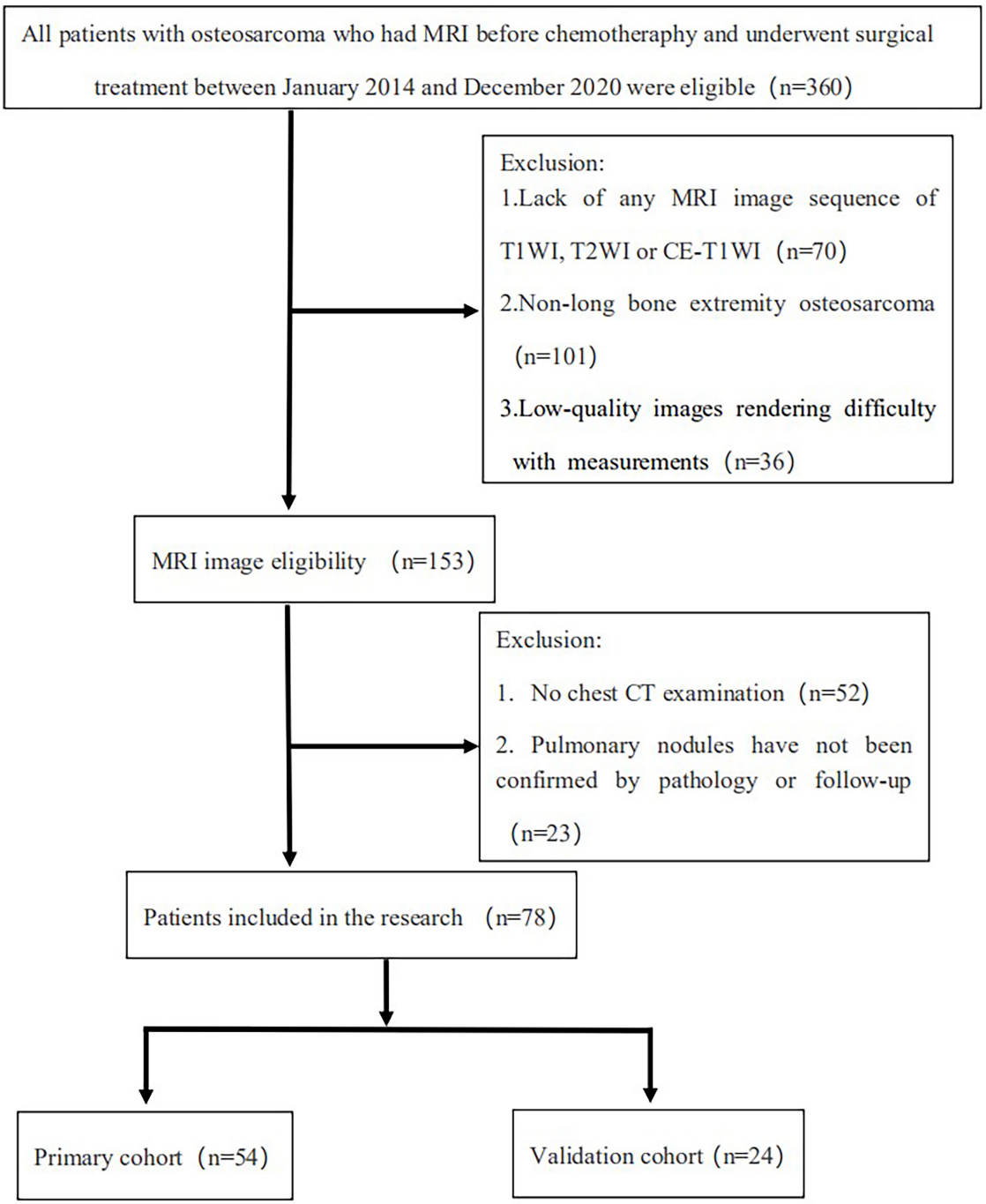


FIGURE 1 | Flow chart of the study population with exclusion criteria.

segmentation for 24 randomly selected patients was then repeated by another radiologist (over 10 years of experience). A separate region of interest (ROI) was manually selected for all sequences on each axial T1WI, T2WI and CE-T1WI slice. All the images were acquired without fat suppression. Delineation of the ROI, including the entire tumor and necrotic areas, cyst degeneration, hemorrhage, periosteal reactions, and peritumoral oedema, was carried out on the images from each sequence. **Figure 2** shows an example of a segmented MRI image. Image intensity normalization was performed before feature extraction, including image gray normalization to uniform grayscale of 0-255 and resampling to 1 mm ×1 mm×

TABLE 1 | Clinical characteristics of 78 cases of osteosarcoma.

Characteristic	Non-SLM	SLM	*P* value
Sex			0.412
Female	15 (33.33%)	14 (42.42%)	
Male	30 (66.67%)	19 (57.58%)	
Pathology			0.984
Osteoblastic	34 (75.56%)	25 (75.76%)	
Others	11 (24.44%)	8 (24.24%)	
Location			0.486
Femur	29 (64.44%)	17 (51.52%)	
Tibia	8 (17.78%)	7 (21.21%)	
Others	8 (17.78%)	9 (27.27%)	
Bone destruction			0.067
Mix	21 (46.67%)	14 (42.42%)	
Osteolytic	22 (48.89%)	12 (36.36%)	
Osteoblastic	2 (4.44%)	7 (21.21%)	
Age (years)	19.49 ± 13.86	16.45 ± 7.53	0.258
Tumor size (cm)	6.31 ± 1.32	8.09 ± 2.39	<0.001*
ALP (IU/L)	758.49 ± 2286.19	913.30 ± 1659.41	0.742
LDH (IU/L)	256.81 ± 105.03	347.63 ± 312.71	0.073

SLM, synchronous lung metastases; ALP, alkaline phosphatase; LDH, lactate dehydrogenase.
**p < 0.05.*

1 mm voxel size using linear interpolation by AK software (Analysis Kit; GE Healthcare).

Radiomic Feature Extraction

In total, 944 radiomic features, quantifying phenotypic differences on the basis of shape (reflecting the size and shape of tumors), first-order (measuring the signal intensity of different tumors), and texture features (representing the relationship between each tumor voxel and its surrounding environments) (9), were automatically extracted from each segmented region of interest by using in-house software written in Python (Pyradiomics version: V 3.0; https://github.com/Radiomics/pyradiomics) (10). All the features were calculated in 3D directions within the whole-tumor volume and normalized by transforming the data into standardized intensity ranges (z-score transformation). Intraclass correlation coefficients (ICCs) based on a multiple-rating, consistency, 2-way random-effects model were calculated to assess the stability and reproducibility of radiomic features. For both tumor ROIs, only features with an ICC > 0.75 were considered to suggest good agreement and retained for further radiomic feature selection.

Radiomics Feature Selection

Radiomics features were automatically calculated with the noncommercial Analysis Kit (A.K. GE Healthcare). First, we performed least absolute shrinkage and selection operator (LASSO) regression on all features to grossly choose attributes with discriminative ability. The goal was to reduce certain attribute coefficients to zero by regulating parameter λ. Subsequently, the area under the receiver operating characteristic (ROC) curve (AUC) could be determined versus log(λ) by employing tenfold cross-validation. The advantage of the LASSO technique is that it can analyse a massive amount of radiomics characteristics from low numbers of samples. Second, we applied multivariate logistic regression to select the most predictive features.

Machine Learning Model

This study used two machine learning classifiers: LR and SVM.

An SVM model was generated based on the established optimal feature subsets of the TC dataset. The kernel, gamma, degree, coef, and C parameters were set to 'rbf', 0.0, 3, 0.0, and 1.0, respectively.

The individual sequence models were constructed by T1WI, T2WI and CE-T1WI.

Next, four combined models were generated *via* a combination of features of dissimilar sequences, namely, T1WI+T2WI, T1WI+CE-T1WI, T2WI+CE-T1WI, and T1WI+T2WI+CE-T1WI. Clinical features were analysed by univariate analysis, and variables for which P < 0.05 were entered into the clinical model.

Two combined models were constructed by combining the best LR and SVM radiomics models with clinical features.

The models were conditioned with the TC using the repeated 10-fold cross-validation technique, and their performance was assessed in the VC.

The radiomics framework of our study is shown in **Figure 3**.

Statistical Analysis

The t-test or Mann-Whitney U-test was employed for the comparison of continuous variables, whereas the chi-squared test or Fisher's exact test was employed for the comparison of intergroup categorical variables. All the statistical analyses were two-sided, and a Bonferroni-corrected P value was employed to determine the feature significance of multiple comparisons. ROC curves were generated to assess the performance of the machine learning models, and the sensitivity, specificity and AUC values were calculated. The AUCs of each two models (clinical, radiomics and combined models) in the two cohorts were compared by using the DeLong test. All the data analyses were performed in R 3.5.1 and Python 3.5.6. A two-tailed P value <0.05 was set as the significance threshold.

TABLE 2 | The clinical characteristics of the 78 osteosarcoma patients in the training and validation cohorts.

Characteristic	Training cohorts	Validation cohorts	*P* value
Sex			0.056
Female	24 (44.44%)	5 (20.83%)	
Male	30 (55.56%)	19 (79.17%)	
Pathology			0.291
Osteoblastic	39 (72.22%)	20 (83.33%)	
Others	15 (27.78%)	4 (16.67%)	
Location			0.322
Femur	34 (62.96%)	12 (50.00%)	
Tibia	8 (14.81%)	7 (29.17%)	
Others	12 (22.22%)	5 (20.83%)	
Bone destruction			0.216
Mix	27 (50.00%)	8 (33.33%)	
Osteolytic	20 (37.04%)	14 (58.33%)	
Osteoblastic	7 (12.96%)	2 (8.33%)	
Age (years)	16.52 ± 9.49	22.00 ± 15.00	0.109
Tumor size (cm)	7.25 ± 1.95	6.66 ± 2.19	0.241
ALP (IU/L)	679.24 ± 1335.34	1149.65 ± 3095.32	0.349
LDH (IU/L)	303.17 ± 250.44	277.36 ± 137.72	0.638

ALP, alkaline phosphatase; LDH, lactate dehydrogenase.

TABLE 3 | The clinical characteristics of these cohorts in terms of SLM and non-SLM of osteosarcoma.

Characteristic	Training cohorts		*P*	Validation cohorts		*P*
	Non-SLM	SLM		Non-SLM	SLM	
Sex			0.667			0.615
Female	13 (41.94%)	11 (47.83%)		2 (14.29%)	3 (30.00%)	
Male	18 (58.06%)	12 (52.17%)		12 (85.71%)	7 (70.00%)	
Pathology			0.707			0.615
Osteoblastic	23 (74.19%)	16 (69.57%)		11 (78.57%)	9 (90.00%)	
Others	8 (25.81%)	7 (30.43%)		3 (21.43%)	1 (10.00%)	
Location			0.169			0.202
Femur	21 (67.74%)	13 (56.52%)		8 (57.14%)	4 (40.00%)	
Tibia	6 (19.35%)	2 (8.70%)		2 (14.29%)	5 (50.00%)	
Others	4 (12.90%)	8 (34.78%)		4 (28.57%)	1 (10.00%)	
Bone destruction			0.261			0.125
Mix	17 (54.84%)	10 (43.48%)		4 (28.57%)	4 (40.00%)	
Osteolytic	12 (38.71%)	8 (34.78%)		10 (71.43%)	4 (40.00%)	
Osteoblastic	2 (6.45%)	5 (21.74%)		0 (0.00%)	2 (20.00%)	
Age (years)	15.81 ± 10.49	17.48 ± 8.07	0.527	27.64 ± 17.10	14.10 ± 5.78	0.014*
Tumor size (cm)	6.47 ± 1.38	8.29 ± 2.14	0.001*	5.96 ± 1.12	7.64 ± 2.94	0.113
ALP (IU/L)	460.17 ± 455.98	974.52 ± 1963.01	0.164	1419.05 ± 4065.50	772.50 ± 582.77	0.625
LDH (IU/L)	256.58 ± 83.78	365.97 ± 366.54	0.173	257.30 ± 145.39	305.44 ± 128.21	0.411

*SLM, synchronous lung metastases; ALP, alkaline phosphatase; LDH, lactate dehydrogenase. *p < 0.05.*

RESULTS

Clinical Characteristics of the Patients

In total, 78 osteosarcoma patients (49 males, 29 females; between 15-83 years of age) were recruited. Based on our univariate analysis, the tumor size was markedly different between the two groups (P < 0.05) (**Table 1**). No obvious differences were observed in age, sex, pathological type, tumor location, bone destruction type, and ALP or LDH levels between the SLM and non-SLM groups. Moreover, no marked differences were observed between the TC and VC (**Table 2**). In addition, the clinical features were not markedly different between the SLM and non-SLM cohorts, except for age in the VC and tumor size in the TC (**Table 3**).

Performance of the Radiomics Models

Overall, 944 radiomics features were obtained from each of the T1WI, T2WI and CE-T1WI images. A total of 702, 839 and 835 radiomics features from T1WI, T2WI and CE-T1WI were included, respectively, with ICC greater than 0.75. The radiomics features with the largest differences between the models are summarized in **Table 5**.

In terms of a distinct sequence in the LR classifier, CE-T1WI features displayed a stronger predictive performance (AUC = 0.87, 95% CI, 0.655-0.965) than T2WI (AUC = 0.85, 95% CI, 0.699-0.981) and T1WI (AUC = 0.686, 95% CI, 0.488-0.873) features in the VC. In terms of combined features, T1WI+T2WI+CE-T1WI had a higher performance (AUC = 0.914, 95% CI, 0.776-0.998) than T1WI+CE-T1WI (AUC = 0.736, 95% CI, 0.533-0.902), T2WI+CE-T1WI (AUC =

TABLE 4 | The detailed scan parameters of four MRI scanners.

Sequence	Imaging planes	Category	TR (ms)	TE (ms)	FOV (mm×mm)	Matrix	Intersection gap (mm)	Slice thickness (mm)
T1WI	Axial	FSE	457-709	8.4-13.2	180×180~380×380	320×128~448×257	0	3-6
T2WI	Axial	FSE	3,640-7,904	83-95.2	180×180~380×380	320×128~448×257	0	3-6
CE-T1WI	Axial	FSE	457-709	8.4-13.2	180×180~380×380	320×128~448×257	0	3-6

MRI, magnetic resonance imaging; TR, repetition time; TE, echo time; FOV, field of view; T1WI, T1-weighted imaging; T2WI, T2-weighted imaging; FSE, fast spin echo; CE, contrast-enhanced.

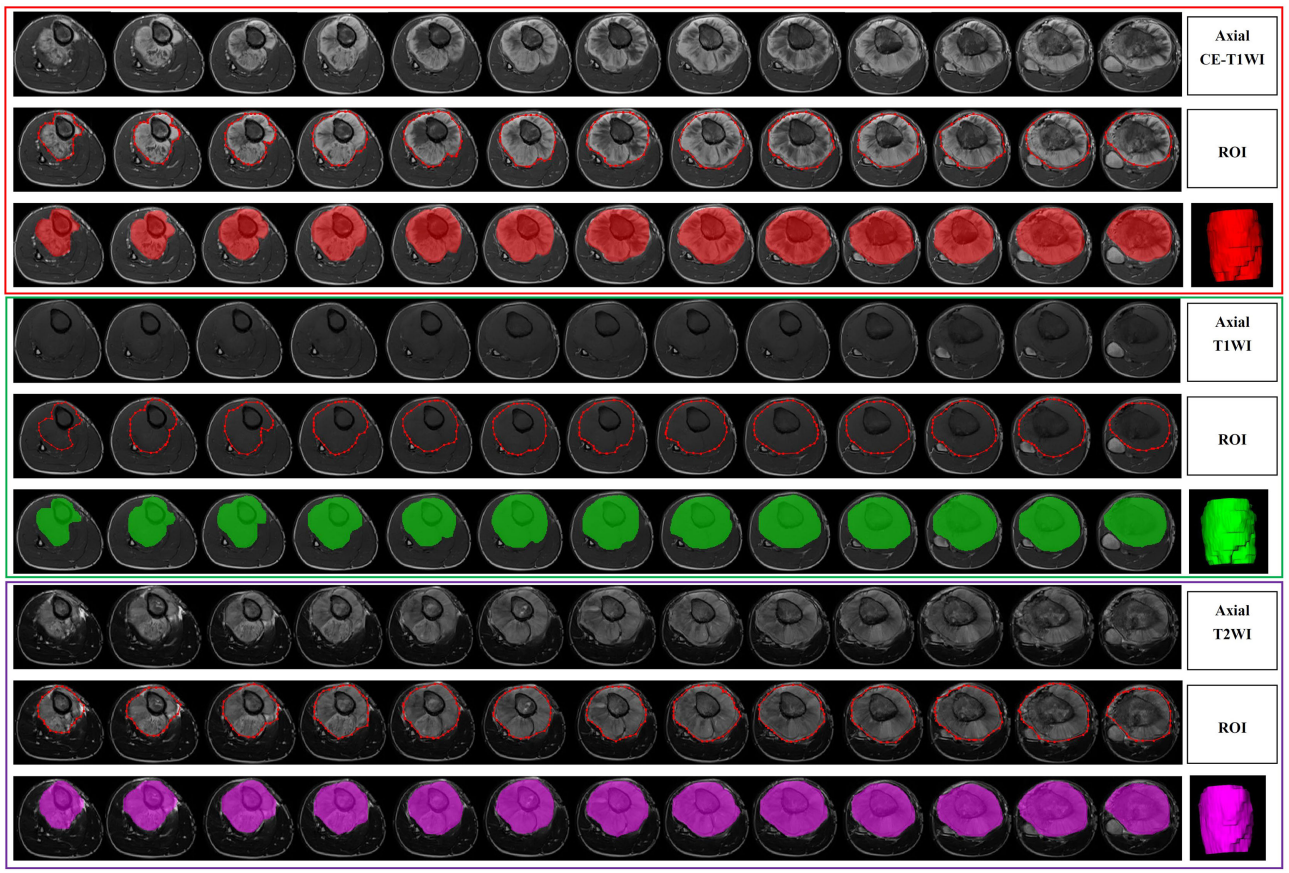


FIGURE 2 | An example of a segmented MRI image.

0.85, 95% CI, 0.699-0.981) and T1WI+T2WI (AUC = 0.879, 95% CI, 0.746-0.993) (**Tables 6**, **7** and **Figures 4**, **5**).

Delong-test results in **Table 8** showed that there were significant differences between predictive performance of T1WI-radiomic model and that of T1WI+T2WI+CE-T1WI radiomic model in both cohorts.

In terms of a distinct sequence in the SVM classifier, T2WI features were more enhanced (AUC = 0.829, 95% CI, 0.621-0.950), compared to CE-T1WI (AUC = 0.771, 95% CI, 0.556-0.916) and T1WI (AUC = 0.629, 95% CI, 0.409-0.815) features in the VC. In terms of the combined features, T1WI+T2WI+CE-T1WI had a higher performance (AUC = 0.929, 95% CI, 0.746-0.993) than T1WI+CE-T1WI (AUC = 0.643, 95% CI, 0.423-0.826), T2WI+CE-T1WI (AUC = 0.829, 95% CI, 0.621-0.950) and T1WI+T2WI (AUC = 0.879, 95% CI, 0.681-0.975) (**Tables 9**, **10** and **Figures 6**, **7**).

Delong-test results in **Table 11** showed that there were significant differences between predictive performance of T1WI-radiomic model and that of T1WI+T2WI+CE-T1WI radiomic model in validation cohort.

Based on our univariate analysis, marked differences were observed in tumor size between the non-SLM and SLM sets ($P < 0.05$). Thus, the clinical model was built using tumor size alone, and this model performed well in the TC (AUC = 0.75, 95% CI,

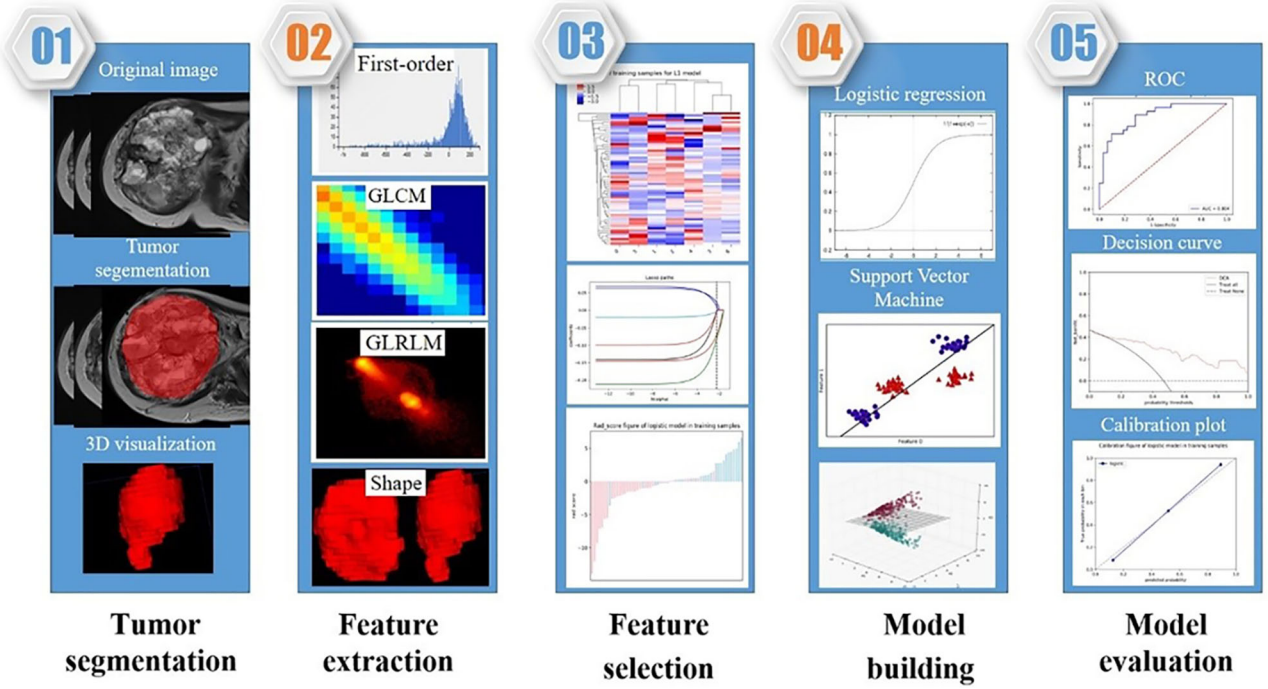


FIGURE 3 | The radiomics framework of our study.

0.613-0.858) and VC (AUC = 0.779, 95% CI, 0.564-0.921). When tumor size was combined with radiomics features, the combined model achieved enhanced prediction compared to the clinical model. The first combined model involving LR-radiomics features + clinical features had an AUC of 0.938 (95% CI, 0.838-0.986) in the TC and 0.957 (95% CI, 0.787-0.999) in the VC. The second combined model based on SVM-radiomics features + clinical features had an AUC of 0.944 (95% CI, 0.845-0.988) in the TC and 0.943 (95% CI, 0.766-0.997) in the VC (**Tables 12**, **13** and **Figures 8**, **9**).

Delong-test results in **Table 14** indicated that there were insignificant differences between predictive performance of combined mode and that of clinical model, LR-Radiomic model and SVM-Radiomic model in both cohorts. The combined model 1 achieved equivalent AUCs compared to combined model 2. Additionally, the combined model 1 performed better than the clinical model as indicated by AUCs of borderline statistical significance (p = 0.0619) in the validation cohort.

DISCUSSION

According to our univariate analysis, tumor size was the most reliable indicator of SLM in osteosarcoma patients, which is in accordance with prior findings (11–14). Huang et al (11) performed a retrospective study examining the characteristics of 1057 osteosarcoma patients. These authors reported that large tumors (>5 cm) were at a substantially elevated risk of resulting in lung metastases in osteosarcoma patients. Munajat et al. (13) also examined the correlation between lung metastasis and tumor volume in a population of 70 osteosarcoma patients. These authors reported that 33 patients (47%), who mostly exhibited larger tumor volumes, also showed signs of lung metastasis. However, in contrast to our study, these authors primarily concentrated on lung metastasis without distinguishing synchronous and metachronous metastases. In our study, we compared tumor size (in terms of diameter) between the non-SLM (6.31 ± 1.32 cm) and SLM (8.09 ± 2.39 cm) groups and found that the tumor size in the SLM group was significantly larger than that in the non-SLM group.

Tumor heterogeneity strongly modulates tumor invasion and prognosis, and a radiomics profile can specifically reflect the complicated histopathology of tumor (15, 16). Several MRI radiomics studies conducted on osteosarcoma were recently reported. Chen H et al. (17, 18) proved that a radiomics signature based on MRI was useful for predicting the response to neoadjuvant chemotherapy and early relapse. Zhao SL et al. (19) showed that a radiomics signature extracted from diffusion-weighted imaging (DWI)-MRI prior to treatment improved the estimation of osteosarcoma.

T1WI, T2WI, and CE-T1WI are the most commonly used MRI sequences for bone tumor. T1WI can be used to observe anatomical structures, but sometimes it is difficult to distinguish soft tissue masses from muscle tissues. T2WI can accurately determine tumor margins and reveal, to a certain extent, the total lesion cell density, whereas CE-T1WI can reveal lesion vascularity, establish the degree of malignancy, and identify

TABLE 5 | The most significant radiomics features of different models.

Model	Radiomics features	Coef.
T1WI	Intercept	-0.4597
	T1WI_wavelet-LLL_glcm_Correlation	1.3060
	T1WI_wavelet-LLL_gldm_GrayLevelNonUniformity	0.8114
T2WI	Intercept	-1.6745
	T2WI_wavelet-LLL_glcm_Correlation	3.4374
	T2WI_wavelet-HHH_firstorder_Mean	-2.7065
	T2WI_wavelet-HLH_glcm_MCC	-3.5169
	T2WI_wavelet-LHL_gldm_LargeDependenceHighGrayLevelEmphasis	2.1317
CE-T1WI	Intercept	-1.2416
	CE-T1WI_wavelet-LLL_glcm_Correlation	1.6253
	CE-T1WI_wavelet-LHL_firstorder_Mean	-1.4719
	CE-T1WI_wavelet-HHL_firstorder_Skewness	-1.7320
	CE-T1WI_wavelet-HLH_glcm_MCC	-1.3359
	CE-T1WI_wavelet-LHH_firstorder_Kurtosis	1.1668
T1WI+T2WI	Intercept	-1.0635
	T2WI_wavelet-LLL_glcm_Correlation	2.2998
	T2WI_wavelet-HHH_firstorder_Mean	-1.5630
	T2WI_log-sigma-1-0-mm-3D_ngtdm_Busyness	1.1750
	T2WI_wavelet-HHH_gldm_SmallDependenceLowGrayLevelEmphasis	-1.2490
T1WI+CE-T1WI	Intercept	-0.8489
	T1WI_wavelet-LLL_glcm_Correlation	1.2206
	CE-T1WI_wavelet-LHL_firstorder_Mean	-1.6295
	CE-T1WI_wavelet-HHL_firstorder_Skewness	-1.1276
T2WI+CE-T1WI	Intercept	-1.6745
	T2WI_wavelet-LLL_glcm_Correlation	3.4374
	T2WI_wavelet-HHH_firstorder_Mean	-2.7065
	T2WI_wavelet-HLH_glcm_MCC	-3.5169
	T2WI_wavelet-LHL_gldm_LargeDependenceHighGrayLevelEmphasis	2.1317
T1WI+T2WI+CE-T1WI	Intercept	-1.2077
	T2WI_wavelet-LLL_glcm_Correlation	2.4347
	T2WI_wavelet-HHH_firstorder_Mean	-1.6936
	CE-T1WI_wavelet-HLH_glcm_MCC	-1.5491

TABLE 6 | The ROC curve of different models of LR-classifier in the training cohort.

Classifiers	Model	AUC	95% CI	Sensitivity	Specificity
LR	T1WI	0.795	0.663 - 0.893	0.565	0.806
	T2WI	0.951	0.855 - 0.991	0.826	0.864
	CE-T1WI	0.909	0.799 - 0.970	0.870f	0.871
	T1WI+T2WI	0.937	0.836 - 0.985	0.783	0.903
	T1WI+CE-T1WI	0.846	0.722 - 0.930	0.739	0.774
	T2WI+CE-T1WI	0.951	0.855 - 0.991	0.826	0.864
	T1WI+T2WI+CE-T1WI	0.940	0.840 - 0.986	0.913	0.903

T1WI, T1-weighted imaging; T2WI, T2-weighted imaging; AUC, area under curve; 95% CI, 95% confidence interval; LR, logistic regression.

necrosis and solid components of tumor. Current research has shown that multiple MRI sequences, such as DWI, T2WI/FS-T2WI and CE-T1WI images, can enhance tumor information extraction and thereby augment specificity, and such sequences for bone and soft tissue tumor radiomics analysis yielded favorable outcomes (20–23). In addition, several studies revealed that classifiers originating from varying classifier families exhibit differing performances for different forms of tumors (24–28). We used two well-known machine learning classifiers in this study. The first is LR, which is a machine learning stratification algorithm used for the prediction of the class probability of a given categorical dependent variable. The second is SVM, which generates a decision margin between two classes to facilitate label estimation from one or more feature vectors. Seven radiomics models using these three sequences alone and combined were established in our study (29). Differences in tumor vessel morphology affect tumor vascular permeability (30). Increased vessel permeability may accelerate cancer metastasis and spread; in the absence of blood vessels, tumors cannot develop beyond a critical volume or invade other organs (31). Enhancing MRI in neoplasms represents regions of admixed vascularity and necrosis, in which contrast permeability is elevated owing to damaged vascular integrity (32). Among the seven radiomics modes, the LR and SVM classifiers had excellent performance (AUC=0.914 and 0.929) in the model combining T1WI, T2WI and CE-T1WI. The combination of T1WI, T2WI, and CE-T1WI allowed the detection of morphological information and indirectly reflected

TABLE 7 | The ROC curve of different models of LR-classifier in the validation cohort.

Classifiers	Model	AUC	95% CI	Sensitivity	Specificity
LR	T1WI	0.686	0.488 - 0.873	0.400	0.786
	T2WI	0.850	0.699 - 0.981	0.600	0.750
	CE-T1WI	0.870	0.655 - 0.965	0.500	0.786
	T1WI+T2WI	0.879	0.746 - 0.993	0.700	0.929
	T1WI+CE-T1WI	0.736	0.533 - 0.902	0.400	0.786
	T2WI+CE-T1WI	0.850	0.699 - 0.981	0.600	0.750
	T1WI+T2WI+CE-T1WI	0.914	0.776 - 0.998	0.700	0.929

T1WI, T1-weighted imaging; T2WI, T2-weighted imaging; AUC, area under curve; 95% CI, 95% confidence interval; LR, logistic regression.

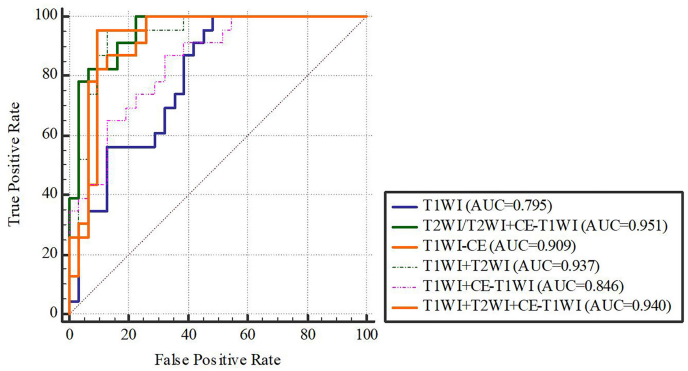


FIGURE 4 | LR-classifier in the training cohort.

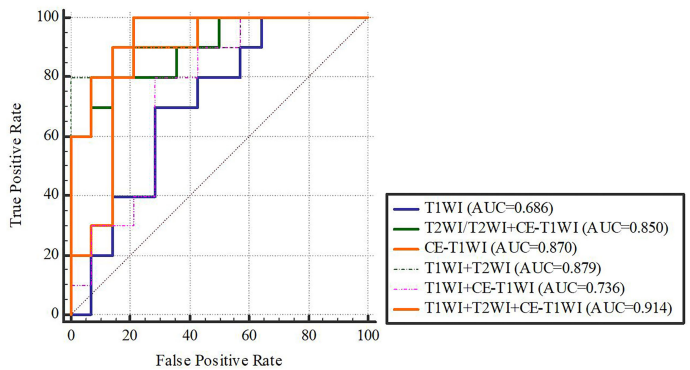


FIGURE 5 | LR-classifier in the validation cohort.

the permeability of tissue microvessels. Although the SVM classifier in the T1WI, T2WI and CE-T1WI combined model performed the best (AUC = 0.929), no obvious difference was observed between the two classifiers of each radiomics model.

In addition, among all radiomics models, those including the T2WI parameter (T2WI, T2WI+CE-T1WI, T1WI+T2WI, and T1WI+T2WI+CE-T1WI) exhibited excellent performance (AUC=0.829-0.929). We found that all of the final features that included the T2WI parameter after selection contained 'GLCM_Correlation' and 'firstorder_Mean' features, which were high-dimensional features that could not be readily interpreted by humans and included comprehensive tumor information. Among these features, the mean, which is a first-order feature, assesses the average grey level intensity within a specified area of interest. The grey level cooccurrence matrix (GLCM) is a second-order feature and is a summary of the frequency of the various combinations of pixel brightness values that occur between neighboring voxels in a given image. GLCM represents the similarity of voxel values along a given direction, whereas homogeneity represents regional grey level uniformity,

TABLE 8 | Delong Test between each two models of LR-classifier in the training and validation cohorts.

Radiomic model	T1WI	T2WI	CE-T1WI	T1WI+T2WI	T1WI+CE-T1WI	T2WI+CE-T1WI	T1WI+T2WI+CE-T1WI
T1WI	–	0.0063*	0.0725	0.0117*	0.4114	0.0063*	0.0110*
T2WI	0.1391	–	0.2395	0.6474	0.0318*	1.0000	0.6744
CE-T1WI	0.1048	0.7088	–	0.4144	0.1713	0.2395	0.3033
T1WI+T2WI	0.0695	0.5149	0.4838	–	0.0613	0.6474	0.8620
T1WI+CE-T1WI	0.6829	0.2054	0.2238	0.0771	–	0.0318*	0.0545
T2WI+CE-T1WI	0.1391	1.0000	0.7088	0.5149	0.2054	–	0.6744
T1WI+T2WI+ CE-T1WI	0.0465*	0.2542	0.3188	0.6750	0.0720	0.2542	–
Training cohort				Validation cohort			

*T1WI, T1-weighted imaging; T2WI, T2-weighted imaging; CE, contrast-enhanced. *p < 0.05.*

TABLE 9 | The ROC curve of different models of SVM-classifier in the training cohort.

Classifiers	Model	AUC	95% CI	Sensitivity	Specificity
SVM	T1WI	0.829	0.702 - 0.918	0.957	0.677
	T2WI	0.973	0.888 - 0.998	1.000	0.838
	CE-T1WI	0.935	0.834 - 0.984	1.000	0.871
	T1WI+T2WI	0.930	0.826 - 0.981	0.957	0.839
	T1WI+CE-T1WI	0.885	0.769 - 0.956	0.783	0.871
	T2WI+CE-T1WI	0.973	0.888 - 0.998	1.000	0.839
	T1WI+T2WI+CE-T1WI	0.938	0.838 - 0.986	0.957	0.903

T1WI, T1-weighted imaging; T2WI, T2-weighted imaging; AUC, area under curve; 95% CI, 95% confidence interval; SVM, support vector machine.

TABLE 10 | The ROC curve of different models of SVM-classifier in the validation cohort.

Classifiers	Model	AUC	95% CI	Sensitivity	Specificity
SVM	T1WI	0.629	0.409 - 0.815	1.000	0.429
	T2WI	0.829	0.621 - 0.950	0.800	0.786
	CE-T1WI	0.771	0.556 - 0.916	1.000	0.500
	T1WI+T2WI	0.879	0.681 - 0.975	0.800	0.857
	T1WI+CE-T1WI	0.643	0.423 - 0.826	0.800	0.500
	T2WI+CE-T1WI	0.829	0.621 - 0.950	0.800	0.786
	T1WI+T2WI+CE-T1WI	0.929	0.746 - 0.993	0.900	0.857

T1WI, T1-weighted imaging; T2WI, T2-weighted imaging; AUC, area under curve; 95% CI, 95% confidence interval; SVM, support vector machine.

and correlation establishes the consistency of image texture (33, 34). In prior studies, these features established tumor heterogeneity and were correlated with the histopathological characteristics and prognosis of numerous tumors, such as osteosarcoma, rectal cancer, thymic tumors, and breast cancer (18, 25, 35–37).

Furthermore, we analysed seven discrete radiomics models, two combined models and one clinical model in osteosarcoma patients. The AUC of the clinical model (0.779) was lower than that of the radiomics model. The prediction ability of the combined model was markedly enhanced relative to that of other models, namely, the clinical model and radiomics model, which involved multiple and single sequences. As clinical information may take into account only some aspects of tumors, multiparametric MRI may better reflect of all tumor information (38). Hence, once the clinical and radiomics characteristics were combined, the performance greatly improved. Based on our data, machine learning analysis involving multiparametric MRI radiomics characteristics can accurately and efficiently predict SLM in osteosarcoma.

Our work encountered certain limitations. First, this work was retrospective in nature. Radiomic features are heavily influenced by differences in the acquisition and reconstruction settings. In the present study, although the most commonly used MRI sequences (T1WI, T2WI and CE-T1WI) were selected, MRI image data were acquired from four distinct scanners, which can also influence the acquired characteristics. Device inconsistency within a dataset is a challenging issue, particularly in retrospective analyses. Second, the sample size was quite small, and all the obtained images were collected over several years. We eliminated non-long bone extremity osteosarcoma and patients who did not receive MRI and chest CT prior to surgery, which accounted for the majority of the enrolled patients. In addition, few osteosarcomas simultaneously met both the axial plane and multiparametric MRI requirements. Due to our strict criteria for patient eligibility, it was difficult to gather large datasets. Despite a statistically insufficient sample size, the results of this study may allow for the improvement of future clinical studies with limited sample sizes. A large sample population with multicentre validation is warranted to achieve high-level evidence for future

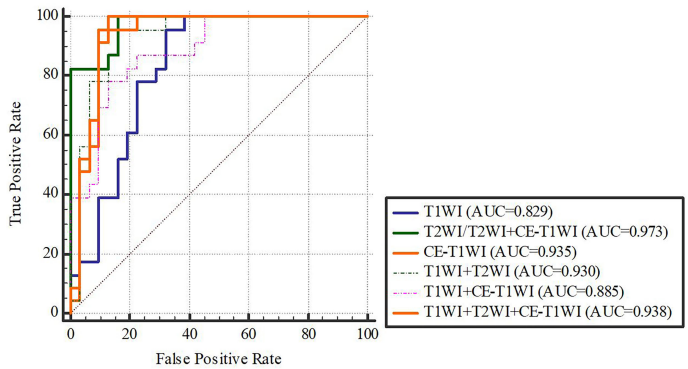


FIGURE 6 | SVM-classifier in the training cohort.

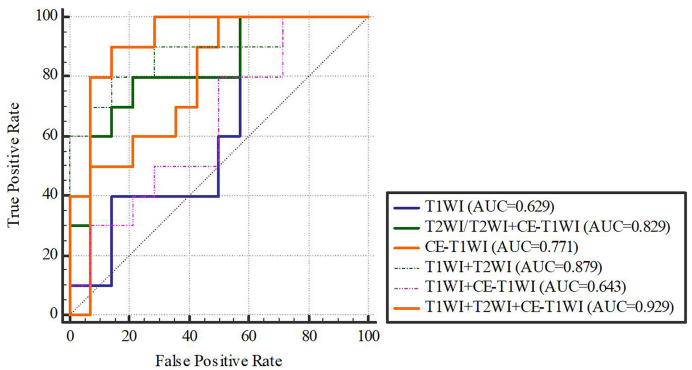


FIGURE 7 | SVM-classifier in the validation cohort. **Figures 4–7** The ROC curve of different models and classifier in the training and validation cohorts.

TABLE 11 | Delong Test between each two models of SVM-classifier in the training and validation cohorts.

Radiomic model	T1WI	T2WI	CE-T1WI	T1WI+T2WI	T1WI+CE-T1WI	T2WI+CE-T1WI	T1WI+T2WI+CE-T1WI
T1WI	–	0.0084*	0.1247	0.1305	0.2795	0.0084*	0.0956
T2WI	0.0928	–	0.2589	0.1822	0.0449*	1.0000	0.2545
CE-T1WI	0.2519	0.4948	–	0.8594	0.3360	0.2589	0.9250
T1WI+T2WI	0.0306*	0.4472	0.2551	–	0.3718	0.1822	0.5620
T1WI+CE-T1WI	0.8981	0.1217	0.1746	0.0203*	–	0.0449*	0.2863
T2WI+CE-T1WI	0.0928	1.0000	0.4948	0.4472	0.1217	–	0.2545
T1WI+T2WI+ CE-T1WI	0.0079*	0.1404	0.0912	0.3870	0.0125*	0.1404	–
Training cohort					Validation cohort		

*T1WI, T1-weighted imaging; T2WI, T2-weighted imaging; CE, contrast-enhanced. *p < 0.05.*

TABLE 12 | The ROC curve of clinical features, radiomic, clinical features + radiomic model in the training cohort.

Model	AUC	95% CI	Sensitivity	Specificity
Clinical model	0.750	0.613 - 0.858	0.696	0.839
LR-radiomic	0.940	0.840 - 0.986	0.913	0.903
SVM-radiomic	0.938	0.838 - 0.986	0.957	0.903
Combined 1 (clinical+LR-radiomic)	0.938	0.838 - 0.986	0.957	0.903
Combined 2 (clinical+SVM-radiomic)	0.944	0.845 - 0.988	0.956	0.900

AUC, area under curve; LR, logistic regression; 95% CI, 95% confidence interval; SVM, support vector machine.

TABLE 13 | The ROC curve of clinical features, radiomic, clinical features + radiomic model in the validation cohort.

Model	AUC	95% CI	Sensitivity	Specificity
Clinical model	0.779	0.564 - 0.921	0.600	0.929
LR-radiomic	0.914	0.776 - 0.998	0.700	0.929
SVM-radiomic	0.929	0.746 - 0.993	0.900	0.857
Combined 1 (clinical+LR-radiomic)	0.957	0.787 - 0.999	1.000	0.857
Combined 2 (clinical+SVM-radiomic)	0.943	0.766 - 0.997	0.846	0.929

AUC, area under curve; LR, logistic regression; 95% CI, 95% confidence interval; SVM, support vector machine.

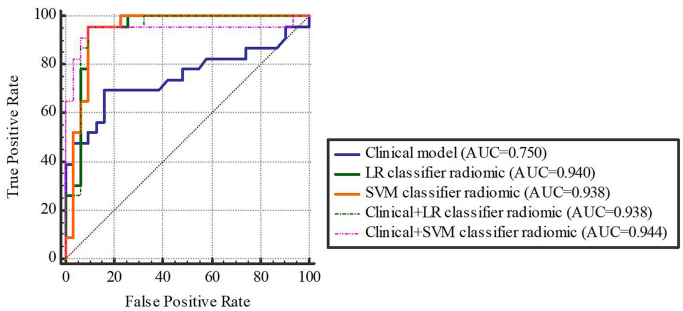


FIGURE 8 | The training cohort.

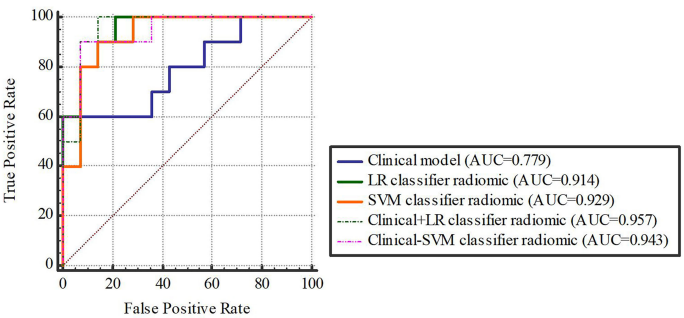


FIGURE 9 | The validation cohort. **Figures 8**, **9** The ROC curve of clinical features, radiomic, clinical features + radiomic model in the training and validation cohorts.

TABLE 14 | Delong Test between each two models (clinical, radiomic, clinical features + radiomic model) in the training and validation cohort.

Model	Clinical	LR-Radiomic	SVM-Radiomic	Clinical +LR-Radiomic	Clinical +SVM-Radiomic
Clinical	–	0.0140*	0.0161*	0.0138*	0.0085*
LR-Radiomic	0.0966	–	0.9065	0.7927	0.9142
SVM-Radiomic	0.1361	0.5247	–	1.0000	0.8908
Combined 1 (Clinical +LR-Radiomic)	0.0619	0.7807	0.2536	–	0.8798
Combined 2 (Clinical +SVM-Radiomic)	0.0787	0.8381	0.6251	0.6265	–
Training cohort				Validation cohort	

*T1WI, T1-weighted imaging; T2WI, T2-weighted imaging; CE, contrast-enhanced; LR, logistic regression; SVM, support vector machine. *p < 0.05.*

clinical application. Third, we compared only T1WI, T2WI and CE-T1WI sequences of MRI. The most commonly used clinical examination for osteosarcoma is X-ray, and the effectiveness of X-ray radiomics needs to be investigated in future studies. Additional MRI functional data must be included in future evaluations to enhance the accuracy and clinical value of our model.

In conclusion, we established a noninvasive prediction tool involving radiomics and clinical characteristics to predict SLM in osteosarcoma patients. The LR and SVM classifiers exhibited an elevated degree of diagnostic performance while employing a combination of characteristics for distinguishing SLM and lack of SLM in osteosarcoma patients. Among all radiomics models, those including the T2WI parameter exhibited good predictive performance for the prediction of osteosarcoma SLM. The constructed model involving the combination of clinical and radiomics characteristics is more effective in evaluating osteosarcoma SLM relative to the clinical model and radiomics model, and the constructed model can provide a new basis for early clinical intervention in metastasis.

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

Written informed consent was not obtained from the individual (s), nor the minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

ZL, JL, XS, and WC contributed to the design and implementation of the concept. ZL and JL contributed equally to this work. JL and ZL contributed in collecting and reviewing patients' clinical and imaging data. RL and ZL contributed in segmenting the lesions. YL and RL contributed in the statistical analysis of the data. ZL, YL, and RL contributed in building models. All authors contributed to the writing and reviewing of the paper. All authors read and approved the final manuscript.

FUNDING

High Level-Hospital Program, Health Commission of Guangdong Province, China (No: HKUSZH201901026).

REFERENCES

1. Ritter J, Bielack S. Osteosarcoma. *J Ann Oncol Off J Eur Soc Med Oncol* (2010) (Supplement 7):320–5. doi: 10.1093/annonc/mdq276
2. Meyers P, Heller G, Healey J, Huvos A, Applewhite A, Sun M, et al. Osteogenic Sarcoma With Clinically Detectable Metastasis at Initial Presentation. *J Clin Oncol Off J Am Soc Clin Oncol* (1993) 11(3):449–53. doi: 10.1200/jco.1993.11.3.449
3. Kager L, Zoubek A, Pötschger U, Kastner U, Flege S, Kempf-Bielack B, et al. Primary Metastatic Osteosarcoma: Presentation and Outcome of Patients Treated on Neoadjuvant Cooperative Osteosarcoma Study Group Protocols. *J Clin Oncol Off J Am Soc Clin Oncol* (2003) 21(10):2011–8. doi: 10.1200/jco.2003.08.132
4. Bacci G, Picci P, Briccoli A, Avella M, Ferrari S, Femino F, et al. Osteosarcoma of the Extremity Metastatic at Presentation: Results Achieved in 26 Patients Treated With Combined Therapy (Primary Chemotherapy Followed by Simultaneous Resection of the Primary and Metastatic Lesions). *Tumori* (1992) 78(3):200–6. doi: 10.1177/030089169207800311
5. Kaste S, Pratt C, Cain A, Jones-Wallace D, Rao B. Metastases Detected at the Time of Diagnosis of Primary Pediatric Extremity Osteosarcoma at Diagnosis: Imaging Features. *Cancer* (1999) 86(8):1602–8. doi: 10.1002/(sici)1097-0142(19991015)86:8<1602::aid-cncr31>3.0.co;2-r
6. Sheen H, Kim W, Byun B, Kong C, Song W, Cho W, et al. Metastasis Risk Prediction Model in Osteosarcoma Using Metabolic Imaging Phenotypes: A Multivariable Radiomics Model. *PloS One* (2019) 14(11):e0225242. doi: 10.1371/journal.pone.0225242
7. Brader P, Abramson S, Price A, Ishill N, Emily Z, Moskowitz C, et al. Do Characteristics of Pulmonary Nodules on Computed Tomography in Children With Known Osteosarcoma Help Distinguish Whether the Nodules Are Malignant or Benign? *J Pediatr Surg* (2011) 46(4):729–35. doi: 10.1016/j.jpedsurg.2010.11.027
8. Chiesa A, Spinnato P, Miceli M, Facchini G. Radiologic Assessment of Osteosarcoma Lung Metastases: State of the Art and Recent Advances. *Cells* (2021) 10(3):553. doi: 10.3390/cells10030553
9. 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(5):2284–92. doi: 10.7150/thno.37429
10. van Griethuysen J, 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–e7. doi: 10.1158/0008-5472.Can-17-0339
11. Huang X, Zhao J, Bai J, Shen H, Zhang B, Deng L, et al. Risk and Clinicopathological Features of Osteosarcoma Metastasis to the Lung: A Population-Based Study. *J Bone Oncol* (2019) 16:100230. doi: 10.1016/j.jbo.2019.100230
12. Kim S, Shin K, Kim H, Cho Y, Noh J, Suh J, et al. Postoperative Nomogram to Predict the Probability of Metastasis in Enneking Stage IIB Extremity Osteosarcoma. *BMC Cancer* (2014) 14:666. doi: 10.1186/1471-2407-14-666
13. Munajat I, Zulmi W, Norazman M, Wan Faisham W. Tumour Volume and Lung Metastasis in Patients With Osteosarcoma. *JJoos* (2008) 16(2):182–5. doi: 10.1177/230949900801600211
14. Wang S, Zheng S, Hu K, Sun H, Zhang J, Rong G, et al. A Predictive Model to Estimate the Pretest Probability of Metastasis in Patients With Osteosarcoma. *Medicine* (2017) 96: (3):e5909. doi: 10.1097/MD.0000000000005909
15. Ogura K, Fujiwara T, Yasunaga H, Matsui H, Jeon D, Cho W, et al. Development and External Validation of Nomograms Predicting Distant Metastases and Overall Survival After Neoadjuvant Chemotherapy and Surgery for Patients With Nonmetastatic Osteosarcoma: A Multi-Institutional Study. *Cancer* (2015) 121(21):3844–52. doi: 10.1002/cncr.29575
16. Li H, Zhu Y, Burnside E, Drukker K, Hoadley K, 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(2):382–91. doi: 10.1148/radiol.2016152110
17. Chen H, Zhang X, Wang X, Quan X, Deng Y, Lu M, et al. MRI-Based Radiomics Signature for Pretreatment Prediction of Pathological Response to Neoadjuvant Chemotherapy in Osteosarcoma: A Multicenter Study. *Eur Radiol* (2021) 31(10):7913–24. doi: 10.1007/s00330-021-07748-6

18. Chen H, Liu J, Cheng Z, Lu X, Wang X, Lu M, et al. Development and External Validation of an MRI-Based Radiomics Nomogram for Pretreatment Prediction for Early Relapse in Osteosarcoma: A Retrospective Multicenter Study. *Eur J Radiol* (2020) 129:109066. doi: 10.1016/j.ejrad.2020.109066
19. Zhao S, Su Y, Duan J, Qiu Q, Ge X, Wang A, et al. Radiomics Signature Extracted From Diffusion-Weighted Magnetic Resonance Imaging Predicts Outcomes in Osteosarcoma. *J Bone Oncol* (2019) 19:100263. doi: 10.1016/j.jbo.2019.100263
20. Dai Y, Yin P, Mao N, Sun C, Wu J, Cheng G, et al. Differentiation of Pelvic Osteosarcoma and Ewing Sarcoma Using Radiomic Analysis Based on T2-Weighted Images and Contrast-Enhanced T1-Weighted Images. *BioMed Res Int* (2020) 2020:9078603. doi: 10.1155/2020/9078603
21. 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 JMRI* (2020) 51(2):435–45. doi: 10.1002/jmri.26834
22. Zhong J, Hu Y, Si L, Jia G, Xing Y, Zhang H, et al. A Systematic Review of Radiomics in Osteosarcoma: Utilizing Radiomics Quality Score as a Tool Promoting Clinical Translation. *Eur Radiol* (2021) 31(3):1526–35. doi: 10.1007/s00330-020-07221-w
23. Wang H, Zhang J, Bao S, Liu J, Hou F, Huang Y, et al. Preoperative MRI-Based Radiomic Machine-Learning Nomogram May Accurately Distinguish Between Benign and Malignant Soft-Tissue Lesions: A Two-Center Study. *J Magn Reson Imaging JMRI* (2020) 52(3):873–82. doi: 10.1002/jmri.27111
24. Viswanath S, Chirra P, Yim M, Rofsky N, Purysko A, Rosen M, et al. Comparing Radiomic Classifiers and Classifier Ensembles for Detection of Peripheral Zone Prostate Tumors on T2-Weighted MRI: A Multi-Site Study. *BMC Med Imaging* (2019) 19(1):22. doi: 10.1186/s12880-019-0308-6
25. Hu J, Zhao Y, Li M, Liu Y, Wang F, Weng Q, et al. Machine-Learning-Based Computed Tomography Radiomic Analysis for Histologic Subtype Classification of Thymic Epithelial Tumours. *Eur J Radiol* (2020) 126:108929. doi: 10.1016/j.ejrad.2020.108929
26. Parmar C, Grossmann P, Rietveld D, Rietbergen M, Lambin P, Aerts H. Radiomic Machine-Learning Classifiers for Prognostic Biomarkers of Head and Neck Cancer. *Front Oncol* (2015) 5:272. doi: 10.3389/fonc.2015.00272
27. Delzell D, Magnuson S, Peter T, Smith M, Smith B. Machine Learning and Feature Selection Methods for Disease Classification With Application to Lung Cancer Screening Image Data. *Front Oncol* (2019) 9:1393. doi: 10.3389/fonc.2019.01393
28. Zhang Y, Zhu Y, Shi X, Tao J, Cui J, Dai Y, et al. Soft Tissue Sarcomas: Preoperative Predictive Histopathological Grading Based on Radiomics of MRI. *Acad Radiol* (2019) 26(9):1262–8. doi: 10.1016/j.acra.2018.09.025
29. Nazari M, Shiri I, Hajianfar G, Oveisi N, Abdollahi H, Deevband M, et al. Noninvasive Fuhrman Grading of Clear Cell Renal Cell Carcinoma Using Computed Tomography Radiomic Features and Machine Learning. *La Radiol Med* (2020) 125(8):754–62. doi: 10.1007/s11547-020-01169-z
30. Roberts W, Delaat J, Nagane M, Huang S, Cavenee W, Palade G. Host Microvasculature Influence on Tumor Vascular Morphology and Endothelial Gene Expression. *Am J Pathol* (1998) 153(4):1239–48. doi: 10.1016/s0002-9440(10)65668-4
31. Carmeliet P, Jain R. Angiogenesis in Cancer and Other Diseases. *Nat Commun* (2000) 407(6801):249–57. doi: 10.1038/35025220
32. Chan A, Fox J, Perez Johnston R, Kim J, Brouwer L, Grizzard J, et al. Late Gadolinium Enhancement Cardiac Magnetic Resonance Tissue Characterization for Cancer-Associated Cardiac Masses: Metabolic and Prognostic Manifestations in Relation to Whole-Body Positron Emission Tomography. *J Am Heart Assoc* (2019) 8(10):e011709. doi: 10.1161/jaha.118.011709
33. Zhang L, Fried D, Fave X, Hunter L, Yang J, Court L. IBEX: An Open Infrastructure Software Platform to Facilitate Collaborative Work in Radiomics. *Med Phys* (2015) 42(3):1341–53. doi: 10.1118/1.4908210
34. Giannini V, Mazzetti S, Bertotto I, Chiarenza C, Cauda S, Delmastro E, et al. Predicting Locally Advanced Rectal Cancer Response to Neoadjuvant Therapy With F-FDG PET and MRI Radiomics Features. *Eur J Nucl Med Mol Imaging* (2019) 46(4):878–88. doi: 10.1007/s00259-018-4250-6
35. Nie K, Shi L, Chen Q, Hu X, Jabbour S, 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
36. Liu S, Zheng H, Pan X, Chen L, Shi M, Guan Y, et al. Texture Analysis of CT Imaging for Assessment of Esophageal Squamous Cancer Aggressiveness. *J Thorac Dis* (2017) 9(11):4724–32. doi: 10.21037/jtd.2017.06.46
37. 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 Trans Oncol* (2020) 22(1):50–9. doi: 10.1007/s12094-019-02109-8
38. 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

Conflict of Interest: Author YL 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.

frontiers in Oncology

SYSTEMATIC REVIEW
published: 24 February 2022
doi: 10.3389/fonc.2022.763842

Diagnostic Accuracy of Artificial Intelligence Based on Imaging Data for Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis

Jian Zhang[1,2†], *Shenglan Huang*[1,2†], *Yongkang Xu*[1,2] *and Jianbing Wu*[1,2*]

[1] Department of Oncology, The Second Affiliated Hospital of Nanchang University, Nanchang, China, [2] Department of Digestive Oncology, Jiangxi Key Laboratory of Clinical and Translational Cancer Research, Nanchang, China

OPEN ACCESS

Edited by:
Daniel Rodriguez Gutierrez, Nottingham University Hospitals NHS Trust, United Kingdom

Reviewed by:
Shouliang Qi, Northeastern University, China
Alessandra Pulvirenti, University of Verona, Italy

***Correspondence:**
Jianbing Wu
hhgwjb@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: 24 August 2021
Accepted: 31 January 2022
Published: 24 February 2022

Citation:
Zhang J, Huang S, Xu Y and Wu J (2022) Diagnostic Accuracy of Artificial Intelligence Based on Imaging Data for Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis. Front. Oncol. 12:763842. doi: 10.3389/fonc.2022.763842

Background: The presence of microvascular invasion (MVI) is considered an independent prognostic factor associated with early recurrence and poor survival in hepatocellular carcinoma (HCC) patients after resection. Artificial intelligence (AI), mainly consisting of non-deep learning algorithms (NDLAs) and deep learning algorithms (DLAs), has been widely used for MVI prediction in medical imaging.

Aim: To assess the diagnostic accuracy of AI algorithms for non-invasive, preoperative prediction of MVI based on imaging data.

Methods: Original studies reporting AI algorithms for non-invasive, preoperative prediction of MVI based on quantitative imaging data were identified in the databases PubMed, Embase, and Web of Science. The quality of the included studies was assessed using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) scale. The pooled sensitivity, specificity, positive likelihood ratio (PLR), and negative likelihood ratio (NLR) were calculated using a random-effects model with 95% CIs. A summary receiver operating characteristic curve and the area under the curve (AUC) were generated to assess the diagnostic accuracy of the deep learning and non-deep learning models. In the non-deep learning group, we further performed meta-regression and subgroup analyses to identify the source of heterogeneity.

Results: Data from 16 included studies with 4,759 cases were available for meta-analysis. Four studies on deep learning models, 12 studies on non-deep learning models, and two studies compared the efficiency of the two types. For predictive performance of deep learning models, the pooled sensitivity, specificity, PLR, NLR, and AUC values were 0.84 [0.75–0.90], 0.84 [0.77–0.89], 5.14 [3.53–7.48], 0.2 [0.12–0.31], and 0.90 [0.87–0.93]; and for non-deep learning models, they were 0.77 [0.71–0.82], 0.77 [0.73–0.80],

3.30 [2.83–3.84], 0.30 [0.24–0.38], and 0.82 [0.79–0.85], respectively. Subgroup analyses showed a significant difference between the single tumor subgroup and the multiple tumor subgroup in the pooled sensitivity, NLR, and AUC.

Conclusion: This meta-analysis demonstrates the high diagnostic accuracy of non-deep learning and deep learning methods for MVI status prediction and their promising potential for clinical decision-making. Deep learning models perform better than non-deep learning models in terms of the accuracy of MVI prediction, methodology, and cost-effectiveness.

Systematic Review Registration: https://www.crd.york.ac.uk/PROSPERO/display_record.php? RecordID=260891, ID:CRD42021260891.

Keywords: hepatocellular carcinoma, artificial intelligence, deep learning, machine learning, microvascular invasion (MVI), radiomics

INTRODUCTION

Hepatocellular carcinoma (HCC) is the most common primary liver malignancy and the fourth most common cause of cancer-related deaths worldwide (1). Liver transplantation and resection are the only potentially curative treatments (2). However, a high risk of recurrence and metastasis after resection leads to a poor prognosis for patients with HCC (3). HCC is highly heterogeneous at the histological, molecular, and genetic levels, making its prognostic stratification and personalized management challenging.

The presence of microvascular invasion (MVI) is considered an independent prognostic factor associated with HCC's early recurrence and poor survival after resection. For MVI-positive patients, expanding resection margins can distinctly improve patient survival by eradicating micrometastases (4, 5). In the current era of precision medicine, a proportion of patients in each stage do not fulfill the criteria for the treatment's allocation (6). In a recent article, Li et al. reported that surgical resection, rather than ablation, is more effective in treating small HCC with MVI. For the MVI patients, cumulative early recurrence rates were significantly lower in the surgical resection group than in the radiofrequency ablation group (22.8% *vs.* 52.5% after 1 year; 30.6% *vs.* 90.0% after 2 years) (7, 8). For HCC patients with MVI present, a more aggressive treatment strategy may be preferred, such as expanding the resection margin or anatomical resectioning for patients undergoing hepatic resectioning, minimizing the ablation margin to at least 0.5–1 cm for patients receiving ablation, and neoadjuvant therapy before surgery (9, 10). Hence, to better allocate treatment strategies, predicting the risk of early recurrence of HCC before resection or ablation is crucial. MVI is not similar to macrovascular invasion, which can be evaluated using radiologic images. MVI is defined as the presence of a tumor in either the portal, hepatic venous system or the branches surrounding the hepatic tissue lined by endothelium, which is visible only by microscopy (11). Many studies have shown that MVI is directly related to the outcomes of HCC patients after surgery, and many researchers have attempted to predict MVI using preoperative imaging analysis.

Recently, in the medical imaging domain, radiomics features extracted through non-deep learning (NDL) algorithms (NDLAs) have been proposed, which are effective for predicting MVI (12). Moreover, artificial intelligence (AI) algorithms have been widely applied in the classification of skin cancer (13), diagnosis of eye diseases (14), identification of prostate cancer (15), and brain metastasis detection (16). AI algorithms show promising performance in the imaging diagnosis of liver cancer (17–20).

Radiomics is a high-throughput extraction of large amounts of quantitative imaging features with the assistance of NDLAs (12). However, manual feature extraction is complicated and time-consuming and lacks stability and consistent interpretation (21). Compared with the NDL used by radiomics analysis, deep learning (DL) algorithms (DLAs) have an advantage in learning features from the images directly, rather than using artificially defined features by human experience (22–24). DL in medical imaging analysis has two properties: multiple layers of non-linear processing units and supervised or unsupervised learning of feature presentations on each layer (23). Input data for DLAs consist of the imaging data itself such as different CT and MRI sequence sets, whereas output data are the desired parameters that should be extracted from the imaging data. In general, the dataset is usually randomly divided into training and testing sets. The former is used to train the DL model; the DLAs attempt to calculate the complex relationship between input and output data. The latter is then used to test the performance of the DL model on a new dataset that had not been utilized to train the DL model.

Recently, some reports have utilized DL methods based on imaging data [MRI, CT, and ultrasound (US)] to predict MVI with satisfactory performance. However, these reports were limited to a small sample size. Huang et al. performed a meta-analysis of radiomics and non-radiomics methods based on

Abbreviations: HCC, hepatocellular carcinoma; MVI, microvascular invasion; DLAs, deep learning algorithms; NDLAs, non-deep learning algorithms; US, ultrasound; PLR, positive likelihood ratio; NLR, negative likelihood ratio; DL, deep learning; NDL, non-deep learning; DLC, the deep learning model with clinical variables; DSN, deep supervision network; AFP, alpha fetoprotein; PIVKAII, vitamin K absence or antagonist; AP, arterial phase; PVP, portal venous phase; SVM, support vector machine; XGBoost, extreme gradient boosting; LR, logistic regression; LASSO, least absolute shrinkage and selection operator; CNN, convolutional neural network; HBP, delayed hepatobiliary phase.

medical image data for MVI prediction (25). Currently, there is no systematic review or meta-analysis of DL methods concerning MVI prediction for HCC patients. In addition, studies comparing DL and NDL methods for MVI prediction are rare. Hence, to provide a more comprehensive and expansive summary of these studies and further recognize the prediction performance of DL for MVI prediction, we conducted a systematic review and meta-analysis by comparing the performance of DL and NDL methods for MVI prediction.

Therefore, the objective of this systematic review and meta-analysis was to assess DL and NDL concerning MVI prediction and compare their performances.

MATERIALS AND METHODS

This systemic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement recommended by the Cochrane Collaboration. This study was prospectively registered in PROSPERO (ID: CRD42021260891).

Search Strategy

Papers describing the use of AI, NDL, and DL for the prediction of HCC were reviewed. We searched the PubMed and Web of Science databases. All English publications until June 14, 2021, were searched without any restrictions on countries or article types. Search terms are available in the **Supplementary Search Strategy** and were included when they discussed the use of NDL or DL methodologies on images in MVI prediction.

Eligibility Criteria

After the removal of duplicates, the articles were reviewed to identify studies that satisfied the following criteria: 1) population: pathologically confirmed HCC patients after surgical resection; 2) intervention: evaluation of MVI using AI algorithms based on quantitative imaging data preoperatively; 3) outcome: diagnostic accuracy of imaging analysis for diagnosing or predicting MVI in HCC study; and 4) design: any type of study design, including observational studies (retrospective or prospective) and clinical trials. Studies were excluded according to the following criteria: 1) studies with duplicate patients and data; 2) case reports, review articles, letters, conference abstracts, and editorials; and 3) studies not in the field of interest. All identified articles were first screened by title and abstract, and then full-text reviews of potentially eligible articles were performed.

Data Extraction

The following information was extracted from the eligible articles: a) study characteristics: authors (years of publication), study type, study design, and study location; b) subject characteristics: operation, interval image exam, number of tumors, etiology of HCC [the number of hepatitis B virus (HBV) or hepatitis C virus (HCV)], tumor size, the numbers of MVI-present and MVI-absent, variables with $p < 0.05$ between MVI(+) and MVI(−), and variables with $p < 0.05$ between the training and testing sets; c) model characteristics: image, region segmentation, validation method, input data, feature selection, and modeling method; and d) the performance of the DL or NDL model: the area under the curve (AUC) value and the numbers of true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN). The reference formulas were as follows: sensitivity = TP/(TP + FN) and specificity = TN/(FP + TN). If there was no sensitivity or specificity in one study, we used Engauge Digitizer version 12.1 to calculate sensitivity and specificity when Youden's index was max based on the receiver operating characteristic (ROC) curve in articles. If there were more than two models in the same group of patients in one study, the model with a higher AUC value was included in our meta-analysis. If some models only analyzed imaging data and others that analyzed both imaging data and clinical parameters, then only the former were included in this study.

Assessment of Study Quality

Two reviewers independently assessed the quality of the eligible articles using the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS-2) criteria and the four domains of patient selection, index test, reference standard, and flow of patients through the study (26).

Data Synthesis and Statistical Analysis

The pooled sensitivity, specificity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), and AUC value of the receiver operating curve were computed. The results are shown in a forest plot. The presence of a threshold effect was analyzed by calculating Spearman's correlation coefficient between sensitivity and the false-positive rate (when $p < 0.05$, the threshold was defined as present). When substantial heterogeneity was noted, a meta-regression analysis was performed to identify the causes. The random-effects model was used to calculate the meta-analytic pooled AUC value, and Higgins's I^2 test was used to assess the heterogeneity between included studies with $I^2 > 75\%$ deemed considerable heterogeneity. An influence analysis was performed if $I^2 > 90\%$. For all NDL and DL models, excluding models using US, to determine the source of heterogeneity, meta-regression analysis based on the number of tumors (single or multiple), image (CT or MRI), region segmentation (manual or semiautomatic), set (validation or training set), least absolute shrinkage and selection operator (LASSO), support vector machine (SVM), convolutional neural network (CNN), 3D-CNN, arterial phase (AP), and portal venous phase (PVP) sequence was performed. For all NDL models excluding US, meta-regression analysis based on the number of tumors, image, region segmentation, set, LASSO, and SVM was performed.

Publication bias was evaluated using Deeks' funnel plot and Deeks' asymmetry test. The AUC values of 0.5–0.7, 0.7–0.9, and >0.9 indicate low, moderate, and high diagnostic power, respectively. All statistical analyses were conducted using STATA version 14.0 (StataCorp LP, College Station, TX, USA) and Meta-DiSc version 1.4.

RESULTS

Search Results and Qualitative Assessment

The PRISMA flow diagram systematically depicts the study selection process (**Figure 1**). A total of 2,280 publications and four articles identified through a meta-analysis were initially retrieved through literature searches, 1,819 of them remaining after the removal of duplicates. After title and abstract screening, 212 articles reported the use of AI in HCC. After a full-text assessment, 16 studies were included in the systematic review and meta-analysis. The quality of the included studies was assessed using the QUADAS-2 scale (26). The results of the qualitative assessment of the included studies are shown in **Supplementary Figure S1**.

Review of the Included Studies

Tables 1, **2** present the detailed characteristics of the 16 studies. Fifteen of the studies were single-center and retrospective studies that used an internal validation method (random splitting or cross-validation) to assess the performance of the MVI prediction model. One study was multicentered and retrospective and used an external validation method. All patients were diagnosed with HCC based on postoperative pathologic specimens and had available preoperative imaging data including CT, MRI, or US. Fifteen studies were based on a population from China 4 (27–41) and one from the United States (42). Concerning the etiology of HCC, at least 78.46% of patients had HBV or HCV of 4,657 patients across all included studies. In patient selection, five articles only included HCC patients with single tumors and excluded multiple tumors (27, 34, 39–41). Based on this diagnostic meta-analysis, 1,946 (40.89%) patients were pathologically diagnosed as MVI-present and 2,813 patients as MVI-absent after surgical resection or liver transplantation. In addition to tumor size in the study by Feng et al. and the hypodense halo in the study by Jiang et al., no significant differences in clinical variables were observed between the training and validation groups. Other characteristics of the included studies are presented in **Tables 1**, **2**, and the baseline characteristics of this meta-analysis are presented in **Table S1**.

Chen et al. compared the predictive performance of five classifiers in six different MRI sequences, and the analysis showed that SVM, extreme gradient boosting (XGBoost), and logistic regression (LR) classifiers in the validation cohort

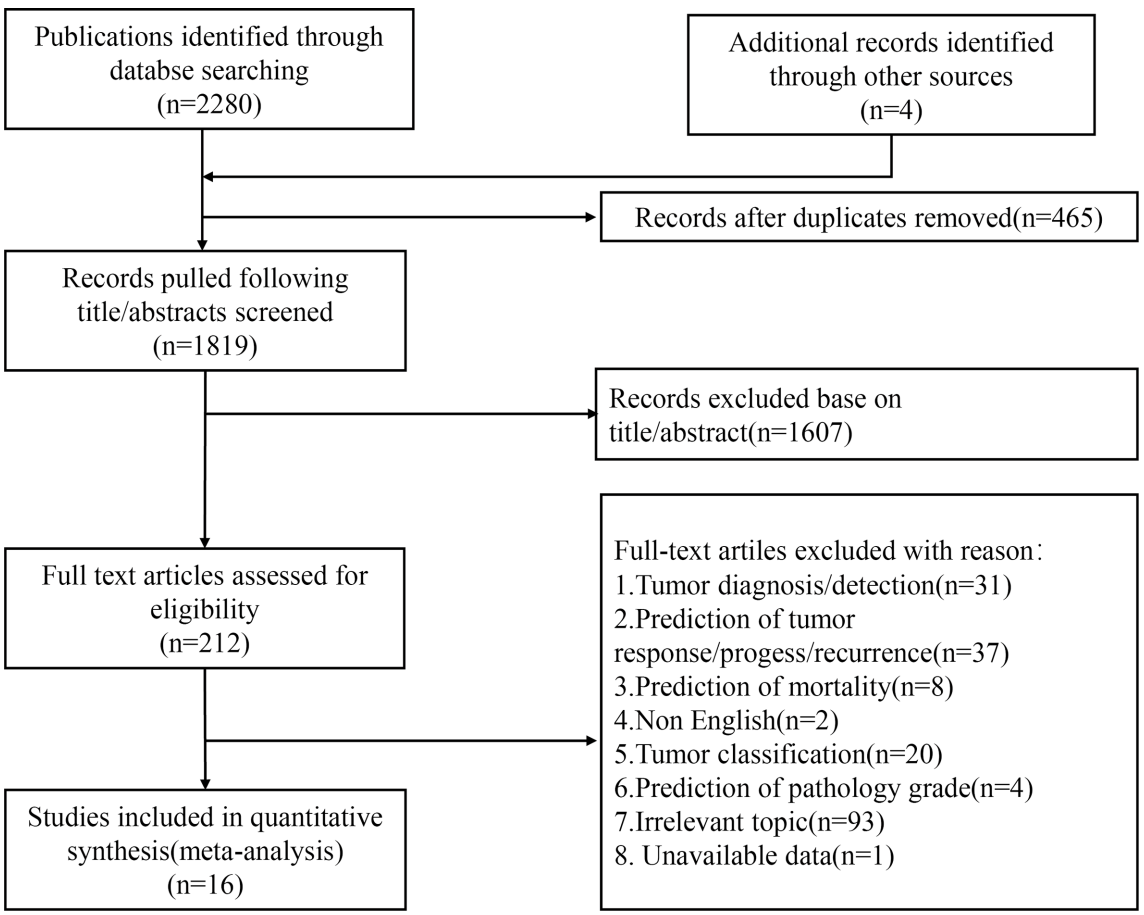


FIGURE 1 | Flowchart of study selection.

TABLE 1 | Characteristics of the included studies.

Authors (year of publication)	Study type	Study design	Study location	Operation	Interval image exam	Number of tumors	Validation	Image	Region segmentation	Input data	Feature selection	Modeling method
Song et al. (2021) (27)	Retro.	Single center	China	SR	Within 1 month	Single	Randomly split at a ratio	MRI	Manually drawn	ADC, DWI (b = 0), DWI (b = 500), AP, PVP, DP, T1WI, T2WI	NA	Radiomics model, CNN
Jiang et al. (2021) (28)	Retro.	Single center	China	SR or TL	Within 2 months	Multiple	Randomly split at a ratio	CT	Manually drawn with ITK-SNAP software	AP, PVP, and DP	NA	XGBoost, 3D-CNN
Wang et al. (2020) (29)	Retro.	Single center	China	SR	Unclear	Multiple	Randomly split at a ratio	MRI	Manually drawn	DWI (b0, b100, b600, and ADC images)	CNN	CNN with DSN
Zhou et al. (2021) (30)	Retro.	Single center	China	SR	Within 1 month	Multiple	Randomly split at a ratio	Gd-EOB-DTPA-enhanced MRI	Manually drawn	Pre-contrast, AP, PVP	3D-CNN	3D-CNN with DSN
Zhang et al. (2021) (31)	Retro.	Single center	China	SR	Within 1 week	Multiple	Randomly split at a ratio	MRI	Manually drawn with ITK-SNAP software	T2WI, T2-SPIR, and PVP images	3D-CNN	3D-CNN
Wei et al. (2021) (32)	T: Retro. V: Pro.	Multicenter	China	SR	Within 1 month	Multiple	External validation	MRI	Manually drawn	CT: AP, PVP MRI: T2W1, T1WI, AP, PVP, HBP	CNN	CNN

Retro, retrospective; Pro, prospective; CNN, convolutional neural network; AP, arterial phase; PVP, portal venous phase; DP, delayed phase; DSN, deep supervision network; V, validation set; T, training set; SR, surgical resection; TL, liver transplantation; LASSO, least absolute shrinkage and selection operation; SVM, support vector machine; BPNet, back-propagation neural network; KNN, k-nearest neighbors; RF, random forest; DT, decision tree; GBDT, gradient boosting decision tree; NRS, neighborhood rough set; PCA, principal component analysis; XGBoost, extreme gradient boosting; ADC, apparent diffusion coefficient.
NA, not available.

showed greater diagnostic efficiency for predicting MVI and NDL models based on delayed hepatobiliary phase (HBP). Due to a lack of data, the study by Chen et al. was excluded from this meta-analysis.

In the study by Nebbia et al., the imaging data were artificially defined as the margin and tumor region before they were used for training models. The results showed that the model combined with margin radiomics and tumor radiomics performed generally worse than single tumor radiomics, contradicting the conclusions of Feng et al. (33). The probable causes included the small sample size, and the tumor margin region may have included extrahepatic regions in the margin segmentation process. Another important reason is that features of the model that combine with margin radiomics and tumor radiomics must be features from both margin and tumor regions, preventing some predictive value features from being learned. In addition, Xu et al. found that analyzing radiomics features from peritumoral regions to calculate predictive performance is not superior to using features from the intratumoral region.

Owing to the high dimensionality and complexity of imaging data using different sequences, feature selection was used to reduce the computational power required to conduct such complex analyses. The LASSO was frequently used for feature selection (33, 34, 38–42). Other methods, which were frequently used for classification, include LASSO regression (33, 34, 40, 43), SVMs (32, 36, 38, 41), decision trees (27), k-nearest neighbor (30, 32), XGBoost (30, 33), and random forest (30, 35).

In contrast to NDL, feature selection and classification of DL occur simultaneously in the process of classifier training. Six of the included studies reported the DL method for the prediction of MVI. **Table S2** summarizes the details of these six studies. Three of the included studies, each a CNN, was used to build the MVI prediction model (27, 29, 32). In three of the included studies, the 3D-CNN model was developed to assess MVI in an end-to-end training fashion, in which feature extraction and predictive model construction were automatically processed by a single neural network (28, 30, 31). While training the DL model, Wu et al. and Wang et al. proposed a deep supervision network (DSN) to reduce the loss function and improve the performance of the DL model by directly supervising the features of the hidden layer and improving the effectiveness of the hidden layer during the CNN learning process (29, 30).

It is worth mentioning that Song et al. proposed a CNN model through MRI analysis of 601 HCC patients with single tumors and then compared the performances of the CNN model and radiomics model based on the same group. The results showed that the CNN model achieved an AUC of 0.915 (0.868–0.963) in the testing cohort as compared to the radiomics model with an AUC of 0.731 (0.645–0.817). In addition, survival analysis demonstrated that patients with DLC-predicted MVI status were associated with poor overall survival and recurrence-free survival, suggesting the strong clinical value of the DLC model in preoperatively identifying HCC patients with poor prognosis and guiding the resection range. Similarly, through CT imaging analysis of 405 HCC patients, Jiang

TABLE 2 | Characteristics of the included studies.

Authors (year of publication)	Study type	Study design	Study location	Operation	Interval image exam	Number of tumors	Validation	Image	Region segmentation	Input data	Feature selection	Modeling Method
Feng et al. (2019) (33)	Retro.	Single center	China	SR	Within 1 month	Multiple	Randomly split at a ratio	Gd-EOB-DTPA-enhanced MRI	Manually drawn with ITK-Snap software	T1WI in/out phase, T1WI-FS, T1WI+c, T2WI+c, T1WI (HBP)	LASSO	LASSO regression model
Nebbia et al. (2020) (42)	Retro.	Single center	USA	SR	Within a week	Multiple	Stratified 5-fold cross-validation	MRI	Manually drawn	DWI, T1, T2, late AP, and PVP	LASSO, feature stability analysis	SVM, decision trees, KNN, Bayes
Liu et al. (2021) (34)	Retro.	Single center	China	SR	Unclear	Single	Randomly split at a ratio	CT	Manually drawn with 3D-Slice software	AP	Intraclass correlation coefficient, LASSO	logistics regression
Dong et al. (2020) (35)	Retro.	Single center	China	SR	Within 2 weeks	Multiple	Split at a ratio	Ultrasound	Manually drawn with MITK	NA	Pearson correlation analysis, minimum redundancy maximum relevance	RF
Xu et al. (2019) (36)	Retro.	Single center	China	SR or TL n (n = 16)	Unclear	Multiple	Split at a ratio	CT	Semiautomatically drawn with Python	AP, PVP	recursive feature selection SVM, step-wise multivariate analysis	Ref-SVM, multivariate regression
Hu et al. (2018) (40)	Retro.	Single center	China	SR	Within 2 weeks	Single	Split at a ratio	Ultrasound	Manually drawn with the A.K. software	NA	LASSO	Logistic regression
Yao et al. (2018) (37)	Retro.	Single center	China	SR	Unclear	Unclear	Cross-validation	Ultrasound	Manually drawn	NA	Sparse representation	SVM
Ni et al. (2019) (38)	Retro.	Single center	China	SR or TL	Within 1 month	Unclear	Split at a ratio	CT	Manually drawn with the A.K. software	PVP	LASSO, NRS, PCA	BPNet, KNN, SVM, RF, DT, Bayes, GBDT
Peng et al. (2018) (39)	Retro.	Single center	China	SR	Within 1 week	Single	Split at a ratio	CT	Semiautomatically drawn with MATLAB	AP, PVP	LASSO	logistic model
Ma et al. (2018) (41)	Retro.	Single center	China	SR	Unclear	Single	Split at a ratio	CT	Manually drawn with ITK-SNAP software	AP, PVP, DP	LASSO	SVM

Retro, retrospective; AP, arterial phase; PVP, portal venous phase; DP, delayed phase; SR, surgical resection; TL, liver transplantation; LASSO, least absolute shrinkage and selection operation; SVM, support vector machine; BPNet, back-propagation neural network; KNN, k-nearest neighbors; RF, random forest; DT, decision tree; GBDT, gradient boosting decision tree; NRS, neighborhood rough set; PCA, principal component analysis.
NA, not available.

et al. proposed and compared the 3D-CNN model, radiomics model, radiological model, and RRC model (model combining radiological features, radiomics features, and clinical variables), with the results showing that the DL model achieved the highest AUC of 0.906 in the validation set. Survival analysis showed that recurrence-free survival was significantly better in the predicted MVI-negative group than in the predicted MVI-positive group. Furthermore, in one multicenter retrospective study, 750 HCCs were enrolled from five Chinese hospitals, and a CNN model (n = 309) based on CT imaging analysis and another (n = 329) based on MRI analysis were trained. In the external validation cohort (n = 115), the findings revealed that the MRI-based CNN model achieved superior prediction performance (AUC: 0.812 *vs.* 0.736, p = 0.038; sensitivity: 70.4% *vs.* 57.4%, p = 0.015; specificity: 80.3% *vs.* 86.9%, p = 0.052). Survival analysis showed that both DL models could stratify groups with both high and low risk in terms of progression-free survival and overall survival. From the three studies, the high diagnostic power of the CNN model was validated, and consistent results indicated the potential value in clinical decision-making.

Meta-Analysis of the Included Studies

In total, 18 NDL models and 11 DL models with 4,759 cases described in 16 individual studies were retrieved. Meta-analysis was performed separately in the subgroups for different modeling methods in different cohorts.

Deep Learning Model for Preoperative Microvascular Invasion Evaluation

Based on 11 DL models in all cohorts, there were 2,073 HCC patients, including 843 MVI-present and 1,230 MVI-absent. The diagnostic meta-analysis forest plots and the combined results are shown in **Figure 2**. Diagnostic threshold analysis showed that there was no significant threshold effect (Spearman's correlation coefficient = −0.082 p = 0.811). The pooled sensitivity, specificity, PLR, and NLR of the DL model were 0.84 [95% CI: 0.75–0.90, I^2 = 85.81%], 0.84 [95% CI: 0.77–0.89, I^2 = 91.92%], 5.14 [95% CI: 3.53–7.48, I^2 = 88.05%], and 0.2 [95% CI: 0.12–0.31, I^2 = 84.83%], respectively. The AUC based on the summary ROC (sROC) curve was 0.90 [95% CI: 0.87–0.93; **Figure 4**]. The I^2 values of sensitivity, specificity, PLR, and NLR indicated high heterogeneity. Influence analysis showed that the models of Jiang et al. and Wei et al. in their training sets could be the cause of the high heterogeneity. After the two models were excluded, I^2 values markedly decreased (**Table 3**). Based on 9 DL models, there were 1,443 HCC patients, including 565 MVI-present and 878 MVI-absent. Analysis of diagnostic threshold showed that there was no significant threshold effect (Spearman's correlation coefficient = −0.150 p = 0.700). The pooled sensitivity, specificity, PLR, and NLR of the DL model were 0.79 [95% CI: 0.71–0.85, I^2 = 70.54%], 0.85 [95% CI: 0.80–0.89, I^2 = 69.44%], 5.34 [95% CI: 3.79–7.52, I^2 = 48.71%], and 0.25 [95% CI: 0.18–0.35, I^2 = 74.00%], respectively. The AUC based on the sROC curve was 0.89 [95% CI: 0.86–0.92; **Figure 3**], which showed moderate diagnostic value. Studies in the DL group numbered less than ten, and thus meta-regression analysis could not be performed.

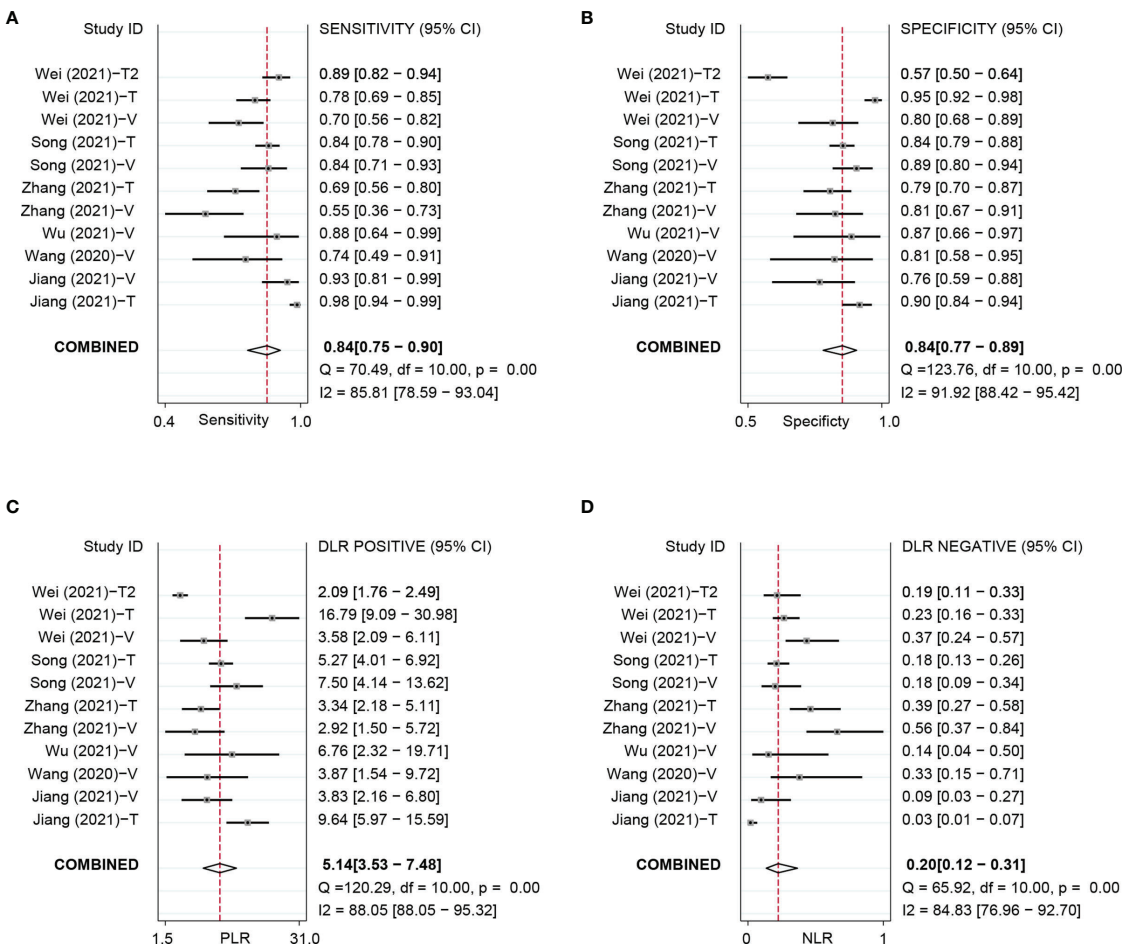


FIGURE 2 | Forest plots based on DL model for preoperative prediction of MVI in HCC. DL, deep learning; MVI, microvascular invasion; HCC, hepatocellular carcinoma; DL, deep learning; MVI, microvascular invasion; HCC, hepatocellular carcinoma; T, training set; V, validationset; Wei (2021)-T1,model in training set based on MRI; Wei (2021)-T2, model in validation set based on CT.

Non-Deep Learning Model for Preoperative Microvascular Invasion Evaluation

For the NDL model across all cohorts, there were 2,685 HCC patients, including 1,128 MVI-present and 1,557 MVI-absent. The diagnostic meta-analysis forest plots and combined results are shown in **Figure 3**. Diagnostic threshold analysis showed that there was no significant threshold effect (Spearman's correlation coefficient = −0.089, p = 0.726). The pooled sensitivity, specificity, PLR, and NLR of the NDL model were 0.77 [95% CI: 0.71–0.82, I^2 = 73.72%], 0.77 [95% CI: 0.73–0.80, I^2 = 48.35%], 3.30 [95% CI: 2.83–3.84, I^2 = 33.64%], and 0.30 [95% CI: 0.24–0.38, I^2 = 73.89%], respectively. The AUC based on the sROC curve was 0.82 [95% CI: 0.79–0.85; **Figure 4**], which showed moderate diagnostic value. Heterogeneity between groups was considered moderate.

US is operator-dependent, and its imaging techniques are different from those of CT and MRI. To reduce the bias, studies (Hu, Yao, and Dong) using US were excluded, and a meta-analysis based on 14 NDL models using CT or MRI was performed. There were 2,059 HCC patients, consisting of 875 MVI-present and 1,184 MVI-absent. The diagnostic meta-analysis forest plots and combined results are shown in **Supplementary Figure S4** and **Table 3**. Diagnostic threshold analysis showed that there was no significant threshold effect (Spearman's correlation coefficient = −0.089, p = 0.726). The pooled sensitivity, specificity, PLR, and NLR of the NDL model were 0.77 [95% CI: 0.71–0.83, I^2 = 74.70%], 0.77 [95% CI: 0.75–0.80, I^2 = 13.48%], 3.42 [95% CI: 2.98–3.93, I^2 = 6.36%], and 0.29 [95% CI: 0.22–0.38, I^2 = 76.24%], respectively.

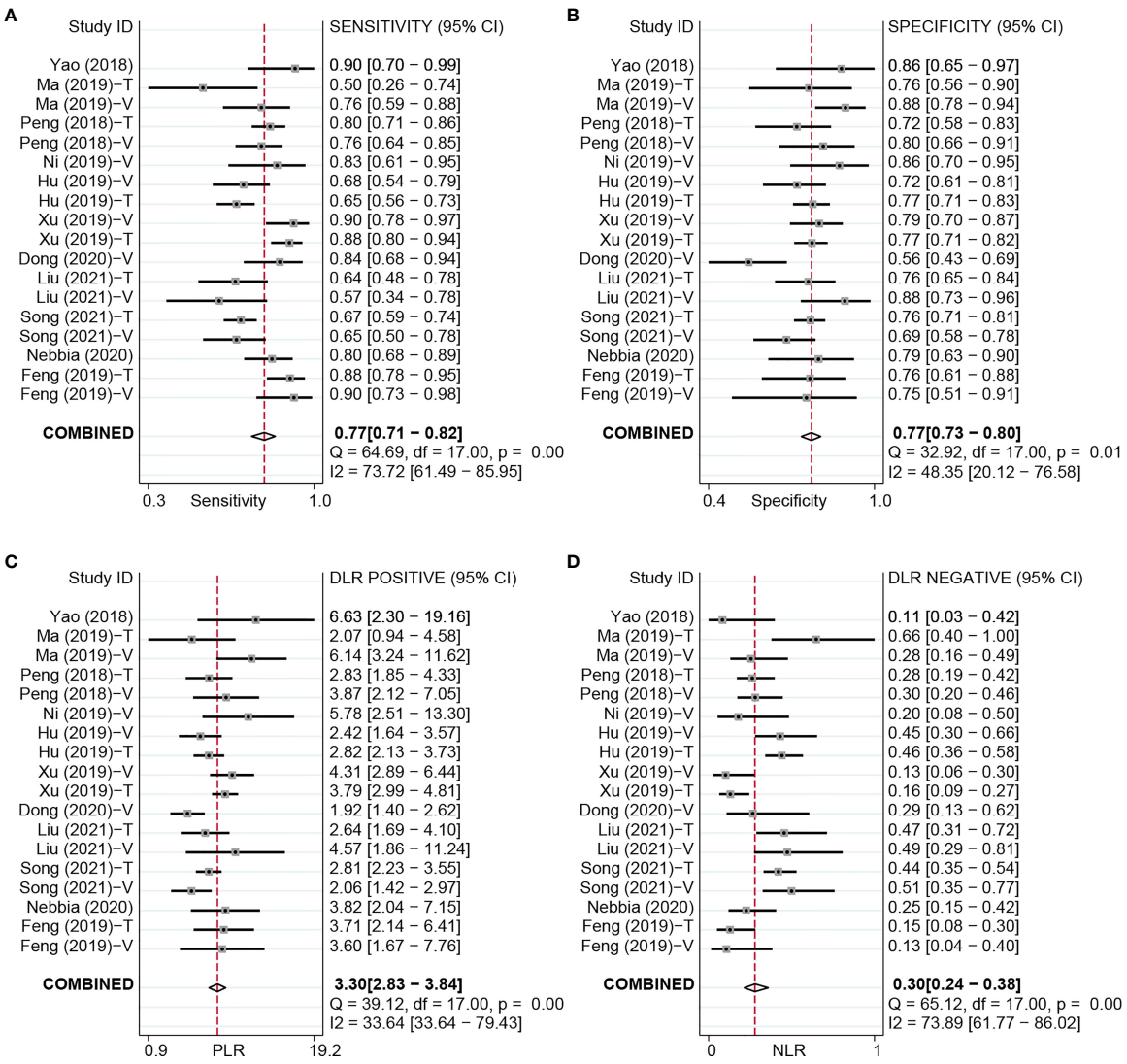


FIGURE 3 | Forest plots based on NDL model for preoperative prediction of MVI in HCC. NDL, non-deep learning; MVI, microvascular invasion; HCC, hepatocellular carcinoma; T, training set; V, validation set.

TABLE 3 | Sensitivity, specificity, positive likelihood ratio, and negative likelihood ratio with subgroup analysis according to the number of tumors in NDL model group.

Analysis	No. of models	Pooled SE (95% CI)	I^2 (%)	Pooled SP (95% CI)	I^2 (%)	Pooled PLR (95% CI)	I^2 (%)	Pooled NLR (95% CI)	I^2 (%)	AUC
NDL model group	18	0.77 [0.71–0.82]	73.72	0.77 [0.73–0.80]	48.35	3.30 [2.83–3.84]	33.64	0.30 [0.24–0.38	73.90	0.82 [0.79–0.85]
NDL model in validation set	9	0.77 [0.70–0.83]	61.59	0.77 [[0.70–0.83]	72.85	3.42 [2.54–4.62]	53.76	0.29 [0.22–040]	63.21	0.84 [0.81–0.87]
DL model group	11	0.84 [0.75–0.90]	85.81	0.84 [0.77–0.89]	91.92	5.14 [3.53–7.48]	88.05	0.2 [0.12–0.31]	84.83	0.90 [0.87–0.93]
DL model in validation set	6	0.79 [0.56–0.86]	74.90	0.83 [0.78–0.87]	0.00	4.72 [3.46–6.44]	0.00	0.25 [0.15–0.42]	76.72	0.85 [0.81–0.88]
Influence analysis in DL model group										
Without Jiang-T	10	0.80 [0.73–0.86]	74.64	0.83 [0.75–0.88]	91.76	4.69 [3.24–6.78]	85.71	0.24 [0.17–0.33]	74.01	0.88 [0.85–0.91]
Without Wei-T2	10	0.83 [0.73–0.90]	85.95	0.86 [0.81–0.90]	68.70	5.88 [4.19–8.24]	56.24	0.20 [0.12–0.33]	85.23	0.91 [0.88–0.93]
Without both	9	0.79 [0.71–0.85]	70.54	0.85 [0.80–0.89]	69.44	5.34 [3.79–7.52]	48.71	0.25 [0.18–0.35]	74.00	0.89 [0.86–0.92]
Subgroup analysis in NDL model group										
Single tumor	8	0.69 [0.65–0.73]	43.26	0.77 [0.74–0.80]	32.54	2.98 [2.54–3.45]	0.00	0.41 [0.35–0.48]	39.30	0.79 [0.75–0.82]
Multiple tumor	10	0.84 [0.78–0.88]	0.00	0.78 [0.72–0.83]	60.09	3.67 [2.82–4.78]	35.97	0.17 [0.13–0.23]	0.00	0.88 [0.85–0.91]
Subgroup analysis in NDL without ultrasound	14	0.77 [0.71–0.83]	74.70	0.77 [0.75–0.80	13.48	3.42 [2.98–3.93]	6.36	0.29 [0.22–0.38]	76.24	0.79 [0.75–0.82]
Single tumor	8	0.70 [0.63–0.75]	52	0.78 [0.73–0.82]	44.46	3.10 [2.49–3.86]	4.84	0.39 [0.32–0.48]	51.80	0.81 [0.77–0.84]
Multiple tumor	6	0.87 [0.83–0.90]	0.00	0.78 [0.74–0.81]	0.00	3.93 [3.31–4.68]	0.00	0.17 [0.13–0.23]	0.00	0.90 [0.87–0.92]
Subgroup analysis by AI algorithms										
LASSO	8	0.75 [0.67–0.81]	72.72	0.76 [0.72–0.79]	10.70	3.05 [2.55–3.64]	0.00	0.34 [0.25–0.45]	70.09	0.77 [0.73–0.80]
SVM	6	0.81 [0.71–0.88]	72.65	0.81 [0.76–0.85]	3.48	4.14 [3.33–5.16]	0.00	0.24 [0.16–0.36]	77.04	0.85 [0.81–0.88]
CNN	6	0.82 [0.78–0.86]	57.42	0.84 [0.73–0.92]	95.38	5.28 [3.04–9.19]	91.72	0.21 [0.17–0.25]	40.47	0.87 [0.84–0.90]
3D-CNN	5	0.87 [0.67–0.96]	93.29	0.84 [0.78–0.88]	48.01	5.30 [3.44–8.16]	49.39	0.16 [0.05–0.46]	93.65	0.88 [0.85–0.90]
Subgroup analysis by image										
MRI	5	0.78 [0.67–0.87]	80.99	0.76 [0.70–0.81]	27.70	3.22 [2.48–4.19]	27.90	0.28 [0.18–0.45]	82.36	0.78 [0.74–0.81]
CT	9	0.76 [0.68–0.83]	72.36	0.80 [0.76, 0.83]	13.11	3.73 [3.12–4.45]	0.00	0.30 [0.22–0.41]	73.85	0.82 [0.78–0.85]

Jiang-T: DL model proposed by Jiang et al. in training set; Wei-T2: DL model based on CT proposed by Wei et al. in validation set; SE, sensitivity; SP, specificity; PLR, positive likelihood ratio; NLR, negative likelihood ratio; AUC, area under the curve; NDL, non-deep learning; DL, deep learning; AI, artificial intelligence; LASSO, least absolute shrinkage and selection operator; SVM, support vector machine; CNN, convolutional neural network.

The AUC based on the sROC curve was 0.79 [95% CI: 0.75–0.82; **Figure 4**], which showed a moderate diagnostic value. Heterogeneity between groups was considered moderate. After studies using US were excluded, the I^2 values of PLR were markedly decreased, but the I^2 values of sensitivity, specificity, and NLR did not noticeably decrease.

Deep Learning Model for Preoperative Microvascular Invasion Evaluation in Validation Set

Considering the effect of overfitting in the model training process, a meta-analysis based on DL models in the validation set was performed after removing the training set. Within the six included DL models in the validation set, there were 495 HCC patients, including 216 MVI-present and 279 MVI-absent. The diagnostic meta-analysis forest plots and combined results are shown in **Supplementary Figure S2**. Diagnostic threshold analysis showed that there was no significant threshold effect (Spearman's correlation coefficient = 0.086, p = 0.872). The pooled sensitivity, specificity, PLR, and NLR of the DL model were 0.79 [95% CI: 0.67–0.88, I^2 = 74.90%], 0.83 [95% CI: 0.78–0.87, I^2 = 0.00%], 4.72 [95% CI: 3.46–6.44, I^2 = 0.00%], and 0.25 [95% CI: 0.15–0.42, I^2 = 76.72%], respectively. The AUC based on the sROC curve was 0.85 [95% CI: 0.81–0.88; **Figure 4**], which showed moderate diagnostic value. After the removal of the training set, the I^2 values were markedly decreased, while heterogeneity between included models was still considered notable in terms of NLR. There was no significant difference in all effect sizes between the models in all cohorts and models in the validation set.

Non-Deep Learning Model for Preoperative Microvascular Invasion Evaluation in Validation Set

Considering the effect of overfitting in the model training process, a meta-analysis based on an NDL model in the validation set was performed. Of the nine included NDL models in the validation set, there were 926 HCC patients, composing 381 MVI-present and 545 MVI-absent. The diagnostic meta-analysis forest plots and combined results are shown in **Supplementary Figure S3**. Diagnostic threshold analysis showed that there was no significant threshold effect (Spearman's correlation coefficient = 0.192, p = 0.620). The pooled sensitivity, specificity, PLR, and NLR of the NDL model were 0.77 [95% CI: 0.70–0.83, I^2 = 61.59%], 0.77 [95% CI: 0.70–0.83, I^2 = 72.85%], 3.42 [95% CI: 2.54–4.62, I^2 = 53.76%], and 0.29 [95% CI: 0.22–0.40, I^2 = 63.21%], respectively. The AUC based on the sROC curve was 0.84 [95% CI: 0.81–0.87], which showed moderate diagnostic value. After the removal of the training set, heterogeneity between groups was considered moderate. There was no significant difference in all effect sizes between the models from all cohorts and models in the validation set.

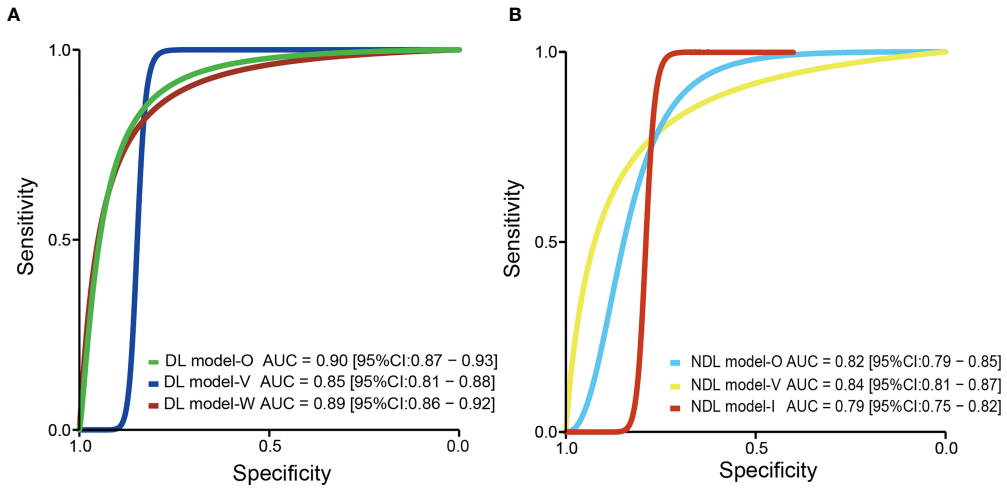


FIGURE 4 | The pooled sROC curve of DL model **(A)** and NDL model **(B)**. sROC, summary receiver operating characteristic; DL, deep learning; NDL, non-deep learning.

Meta-Regression Analysis and Subgroup Analysis

We observed substantial heterogeneity in the performance of the NDL group, with I^2 being 73.72%, 48.35%, 33.64%, and 73.89% for the pooled sensitivity, specificity, PLR, and NLR, respectively. As US may result in a noticeable bias, we excluded studies using US and then performed the meta-regression analysis. The results of meta-regression analysis are presented in **Tables S4**, **S5**. The results showed that in the univariate meta-regression model, 10 covariates were significantly associated with study heterogeneity. Therefore, we believe that these variates may influence prediction accuracy in the NDL group. In the multivariate meta-regression model, the number of tumors was strongly associated with study heterogeneity.

We conducted an additional subgroup analysis based on the number of tumors (**Table 3**). In it, I^2 values of the two subgroups were markedly decreased. The I^2 of the single tumor subgroup was 43.26%, 0%, and 39.28% for the pooled sensitivity, PLR, and NLR, respectively. The I^2 of the multiple tumor subgroup was 0% and 0% for the pooled sensitivity and NLR, respectively. Except for the pooled specificity and PLR, significant differences between the two subgroups were observed in the pooled sensitivity, AUC, and NLR. The results of subgroup analysis using the AI algorithm (LASSO and SVM) and image (CT and MRI) are shown in **Table 3**. There was no significant difference between the image and AI algorithms in the NDL group. For AI algorithms in the NDL group, SVM is significantly superior to LASSO for the pooled AUC (0.77 [0.73–0.80] *vs.* 0.85 [0.81–0.88]). There was no significant difference between CNN and 3D-CNN. Generally, DL models (3D-CNN and CNN) are significantly superior to LASSO, and there was no significant difference between DLAs and SVM.

Testing for Publication Bias

Deeks' funnel plot asymmetry test showed no significant publication bias with p-values of 0.42 and 0.22 for the DL group and NDL group, respectively.

DISCUSSION

Performance of Deep Learning and Non-Deep Learning Models

In this study, NDL models and DL models were compared. The NDL models had a moderate diagnostic value for MVI prediction in HCC, with pooled sensitivity, specificity, PLR, NLR, and AUC values of 0.77, 0.77, 3.30, 0.30, and 0.82, respectively. The DL models, including the CNN model and 3D-CNN model, had moderate diagnostic values that were similar to those of the NDL models, with pooled sensitivity, specificity, PLR, NLR, and AUC values of 0.84, 0.84, 5.14, 0.2, and 0.90, respectively. All these effect sizes showed that models using the DL method had a higher performance for preoperative prediction of MVI in HCC and had a statistically significant difference in diagnostic value in terms of AUC. When comparing DL models with NDL models in the validation set, there was no significant difference in any of these factors. A reasonable interpretation is that the sample sizes of the DL model group were too small, and the heterogeneity in both the NDL and DL model groups was notable. However, there is reliable evidence to support that the model using the DL method may have a higher performance and be more suitable for preoperative MVI prediction.

By analyzing radiomics features from images, building a prediction model using NDL methods had been widely applied in MVI prediction (44–48) and prediction domain of other cancers (13–16). NDL models based on radiomics features had been proved to be better than a model based on radiological characteristics or clinical characteristics (44, 45). For the NDL models included in this study, analyzing radiomics features assisted by NDLAs is an advanced technique for MVI prediction, but one of the shortcomings of radiomics is that the method is based on handcrafted feature extractors, which require extensive work and manpower. In addition, the main

limitation is that radiomics features are human-designed and dependent on domain-specific expertise.

A DL method, CNN, was proven powerful in medical imaging (49), with superior performance as compared to NDL based on radiomics features. The advantage of DL is that feature extraction in the learning process is not required, avoiding defects in human-designed features in radiomics analysis. Since classifier training, feature selection, and classification of DL occur simultaneously, we needed only input images, rather than clinical data, radiological features, or radiomics features. Feature selection and classification of DL occur simultaneously during classifier training. The main power of a CNN lies in a CNN architecture consisting of a series of layers of convolution filters, akin to low-level vision processing in the human brain, which allows for the extraction of a set of discriminating features at multiple levels of abstraction. However, training a deep CNN is challenging. The main difficulties are that CNNs require a large amount of labeled training data and large computational and memory requirements and that training a deep CNN is often complicated by overfitting and convergence issues and the lack of interpretability. Jiang et al. provided a new means to partly explain how DL can identify MVI status.

The main difference in 3D-CNNs is that the input data are three-dimensional image data. In the included studies, Wu et al. proposed a 3D-CNN model with a DSN based on pre-contrast, APs, and PVPs in MR images with an AUC value of 0.9255. A 3D-CNN model proposed by Song et al. with DSN based on eight MRI sequences obtained the highest AUC value of 0.915 in the testing cohort. Another 3D-CNN model proposed by Jiang et al. based on AP, PVP, and DP CT sequences in the validation set achieved 0.906 [95% CI: 0.821–0.960]. In the studies by Song et al. and Jiang et al., the two 3D-CNN models performed excellently in MVI prediction.

The Value of Artificial Intelligence Algorithms for Microvascular Invasion Prediction

For AI algorithms, we performed a subgroup analysis, and results showed that DL is generally superior to NDL and that in NDL, SVM is significantly superior to LASSO. The advantage of DL has been previously discussed. The reason for the better performance of SVM than LASSO may be that the combination of modeling by SVM, and feature selection by LASSO has an advantage over than LASSO regression model only using LASSO for feature selection. SVM is a good classifier, but it may not get good performance when it is directly used for classification, but if it can be combined with a good feature selection algorithm, the classification performance will be greatly improved.

The Potential Clinical Value of Convolutional Neural Network Models

A CNN model proposed by Wei et al. based on T2W1, T1WI, AP, PVP, and HBP MRI sequences achieved an AUC value of 0.802 in an independent external validation cohort. Furthermore, in the study by Song et al., survival analysis demonstrated that patients with DLC-predicted MVI status were associated with poor overall survival and recurrence-free survival, whereas in a study by Jiang et al., based on the MVI status predicted by the 3D-CNN model, the mean recurrence-free survival was significantly better in the predicted MVI-negative group than in the predicted MVI-positive group [64.06 *vs.* 31.05 months, $p = 0.027$]. In the study by Wei et al., survival analysis indicated that CNN models could stratify groups with high and low risks in terms of progression-free survival and overall survival ($p < 0.05$). These key findings indicate that the DL model can provide a non-invasive approach to accurately evaluate MVI, with the potential to facilitate clinical decision-making and assess patient prognosis.

Prediction Values of Various Types of Input Data

CT or MRI data from arterial and portal phases were used to build the prediction model and proved powerful for MVI prediction in 13 of the included studies. Jiang et al. proposed a 3D-CNN model based on AP, PVP, and DP of CT images, which achieved an AUC value of 0.906. For five of the included studies, the AUC value of the prediction model based on AP and PVP of MR images ranged from 0.80 to 0.94. Five of the included studies in the DL group used MR images, and three studies in NDL used MR images. Among them, Wu et al. proposed a 3D-CNN model with DSN based on pre-contrast, AP, and PVP phases in MR images with an AUC value of 0.925. A meta-analysis of MRI features for predicting MVI of HCC performed by Hong et al. showed a similar conclusion that arterial enhancement and arterial peritumoral enhancement were significant predictors for MVI of HCC (50). However, in this study, the results of meta-regression showed no significant difference in the AP or PVP. The probable reasons for this were high heterogeneity and that the number of relevant original studies was small. Diffusion-weighted imaging (DWI) is an MRI sequence that can reflect the motion state of water molecules *in vivo* (51). Nebbia et al. built an SVM model based on a DWI sequence and performed worse than the AP or PVP sequence. However, in the study by Song et al., a CNN model based on eight MRI sequences, including DWI, AP, and PVP, achieved an AUC value of 0.915. Features from the DWI sequence, as complementary to AP and PVP, could further improve the performance of MVI prediction. Wang et al. suggested that deep features derived from higher b values yield better performance for MVI prediction, implying that DWI with a higher b value might be better for MVI prediction. Chen et al. indicated that the ADC value can also be used to evaluate MVI and has a diagnostic efficacy similar to the 20-min T1 relaxation time [AUC, 0.850 *vs.* 0.846]. Wu et al. indicated that due to the overflow of contrast agents from the tumor region in the delayed phase, and the tissue cellularity and vascularity within the tumor becoming unclear, information from the delayed phase sequence has worse predictive performance and may not fit MVI prediction. US was mainly used for MVI prediction in NDL models, and the results showed that the AUC value of models based on US ranged from 0.726 to 0.731, lower than that based on CT and MRI (35, 37, 40).

The HBP of contrast-enhanced liver MRI with gadoxetate disodium (Gd-EOB-DTPA) has the value of significantly increasing sensitivity and specificity in liver diagnosis (51, 52) and predicting MVI in HCC (33). Hong et al. performed a meta-analysis based on MRI features for MVI prediction, with the results showing that peritumoral hypointensity on HBP was the MRI feature most suggestive of MVI with the pooled diagnostic odds ratio (DOR) and pooled positive LR being 8.2 and 5.0, respectively (50). Chen et al. built an SVM model based on the hepatobiliary phase sequence of Gd-EOB-DTPA MRI, with a performance of 0.942 AUC value, higher than the AP and PVP sequences for MVI prediction. In this study, since the results showed that there was no significant difference between MRI and CT, analysis based on MRI features for MVI prediction did not yield significant results.

Within the DL group, the models proposed by Wang et al. and Zhang et al. obtained lower performance with AUC values of 0.79 and 0.72, respectively. The possible reasons are the differences in the types of input data. Notably, the input data of the two DL models did not include the imaging data in AP and PVP. However, further studies are needed to confirm this hypothesis.

CT *vs.* MRI in Artificial Intelligence Algorithms for Microvascular Invasion Prediction

Compared with CT, MRI can better describe the characteristics of soft tissue, atomic signal intensity, and lesion enhancement, as well as provide more information on tissue function.

For models using 3D-CNN algorithms in the DL group, two studies used MRI techniques (Wu and Zhang), and one study used a CT imaging technique (Jiang). We observed that the training set containing 3D-CNN models using CT by Jiang et al. achieved the highest AUC value of 0.98. In the validation set, Wu et al. proposed 3D-CNN models using MRI, which had the highest AUC value of 0.926. Since the number of studies was too small, a meta-analysis could not be performed. For models using CNN algorithms in the DL group, two studies used MRI (Song and Wang), and one study used CT and MRI (Wei). Wei et al. built DL models for preoperative prediction of MVI based on CT and MR images. The results of the meta-analysis showed superior predictive power from MRI compared to CT (AUC: 0.812 *vs.* 0.736, p = 0.039).

In this study, meta-regression analysis was performed for models in the NDL group. The results showed that imaging techniques may be influencing factors of prediction power in the NDL group but not independently influencing factors. There was no significant consequence of the predictive power of MRI being superior to CT (AUC: 0.78 [0.74–0.81] *vs.* 0.82 [0.78–0.85]).

Overall, our results showed that, in the DL model group, especially the CNN model, MRI was superior to CT in the prediction of MVI. However, there was no significant advantage that MRI had in MVI prediction, compared with CT. Recently, Meng et al. compared the performance of radiomics models based on CT and MRI for MVI prediction (53). The results showed that CT and MRI had a comparable performance for MVI prediction in a single HCC. Studies comparing the performance of AI algorithms based on CT and MRI for MVI prediction are too small and can be excluded.

Deep Learning Models Combined With Clinical Characteristics

Previous studies have predicted MVI using clinical characteristics, such as tumor number and size, alpha fetoprotein (AFP), protein induced by vitamin K absence or antagonist (PIVKAII), and serum component index. The AUC of these predictors varies from 0.529 to 0.81 (18–23). In this study, some clinical variables [tumor size, AFP, tumor margin, internal arteries, and International normalized ratio (INR)] that were recognized as predictive values were selected by statistical analysis and then integrated with the DL model to further improve predictive performance. Clinical variables recognized as MVI-prediction values were tumor size in 11 studies and AFP in nine studies; others are shown in **Table S3**. Some studies using radiomics combined with clinical parameters achieved better outcomes, ranging from 0.796 to 0.899 for AUC (36, 41, 54).

Number of Tumors as One Source of Heterogeneity

In addition, we performed a subgroup analysis according to the number of tumors, and the results showed that the number of tumors was one of the sources of heterogeneity. Models based on HCC patients with multiple tumors performed better with the pooled AUC value of 0.88 [0.85–0.91] and sensitivity of 0.84 [0.78–0.88] than single tumors with 0.79 [0.75–0.82] and 0.69 [0.65–0.73], respectively. In HCC patients, having multiple tumors was regarded as a variable that had strong associations with a high risk of MVI. This could cause these models to more easily identify the MVI status in HCC patients with multiple tumors than single tumors. However, because the number of models in the meta-analysis was relatively small, the results of the subgroup analysis need to be interpreted with caution.

Trends, Challenges, and Suggestions

According to the analysis of the existing MVI prediction models presented above, the diagnostic accuracy of CNNs for preoperative MVI prediction has achieved spectacular progress in terms of sensitivity, specificity, PLR, NLR, and AUC. However, there is much room for improvement due to existing challenges, as well as many options for future research.

Methodological Trends

In six studies using DL in this meta-analysis, CNNs have been the main methods for MVI prediction. The six studies used ensemble learners of CNNs, which is an approach for integrating multiple learner branches into a single fusion model to improve the prediction of MVI in HCC (55). In each learner branch, fully convolutional networks and softmax layers were employed to calculate the predicted results. In the studies by Wang et al. and Wu et al., a DSN that combines the loss functions of each CNN learner branch was designed for the proposed DL network. Jiang et al. and Song et al. designed specific architectures as CNN

branches for feature extraction, with their final DL models achieving AUCs of 0.906 and 0.915, respectively.

Challenges and Suggestions

Lack of Datasets With Large Numbers of Cases

One of the critical barriers in the application of DL for MVI prediction based on medical imaging data is the lack of datasets with large numbers of samples. It is noted that the process of training DL models using CNNs requires a huge amount of data. However, their collection is still very difficult in clinical practice.

To mitigate this problem, new techniques for generating synthetic medical images could be developed. For instance, Zhang et al. generated an augmented training set by randomly rotating the original imaging dataset at a full 360° angle. Moreover, Wang et al. used an image resampling method to generate more samples for training a DL network.

Generalizability

Typically, a specific model that performs very well on a specific task may not be generalized to other tasks. Heterogeneity could be one of the major reasons why a specific model cannot be generalized to other tasks. The sources of heterogeneity are various imaging modalities, and different medical scanners operate under different settings and datasets. This issue could also be alleviated by developing methods that can be validated on images of different types. In addition, research on the effect of scanner settings (reconstruction techniques, parameters, etc.) on MVI prediction is expected.

Lack of Interpretability

The black box problem has been one of the major criticisms of the deep CNN approach, implying that the system struggles to provide evidence to support clinical decisions. Better interpretability would contribute to understanding how the MVI status is generated. This may lead to more accurate and reliable clinical decisions.

To improve the accuracy of diagnosis and interpretability of DL models, new approaches for both radiomics and semantic feature analysis in screening data can be developed. For example, to improve the interpretability of the 3D-CNN model, Jiang et al. attempted to predict the 15 most important variables selected by the XGBoost method, and the results indicated that the CNN model could predict the status of MVI partly based on the explainable features utilized in clinical practice.

Potential Value of Clinical Application

Several studies (Song et al., Jiang et al., and Wei et al.) performed survival analysis that showed that the patients with CNN-predicted MVI status were associated with poor survival after resection, suggesting the strong clinical value of the CNN model in preoperatively identifying HCC patients with poor prognosis and guiding the resection range. However, there is no evidence from prospective studies or clinical trials. Thus, in the future, some prospective research and clinical trials concerning CNN models for MVI prediction that guide clinical decisions are expected.

Contributions and Limitations

Our meta-analysis of DL methods and NDL methods for preoperative MVI prediction in HCC patients has several advantages. First, this study involving 16 studies and 4,759 HCC cases is the first systematic review and meta-analysis of preoperative MVI prediction in HCC patients by comparing DL and NDL methods. Second, DL models perform better than NDL models in terms of the accuracy of MVI prediction, methodology, and cost-effectiveness.

This study has some limitations. First, all included studies were retrospective, inevitably causing a patient selection bias. Second, this study only included six studies for DL methods in MVI prediction because CNNs are powerful tools for a broad range of computer vision tasks applied in medical imaging in recent years, and training a CNN requires a large sample size, which is difficult in clinical tasks. Third, only one included study used an independent external validation cohort to assess the performance of DL models. Finally, study heterogeneity was significant across the included studies.

CONCLUSIONS

This meta-analysis demonstrates the high diagnostic accuracy of NDL and DL methods for the prediction of MVI and their promising potential for application in clinical decision-making. Multicentral validation and larger sample sizes are required for more definitive conclusions. DL models perform better than NDL models in terms of the accuracy of MVI prediction, methodology, and cost-effectiveness. CT or MRI data from the arterial and portal phases were used to build a prediction model and were proved effective for MVI prediction. Clinical variables, such as tumor size and AFP, were recognized as MVI prediction values. Studies of DL models for MVI prediction for HCC patients with single tumors are expected.

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

JZ designed the study drafted the manuscript. SH and JZ were responsible for the collection and analysis of the research information. JZ, SH, YX, and JW critically and carefully revised this manuscript. The authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We acknowledge and appreciate our colleagues for their valuable efforts and comments on this paper. We thank Liwen Bianji

(Edanz) (www.liwenbianji.cn/) for editing the language of a draft of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.763842/full#supplementary-material

Supplementary Figure S1 | Quality assessment of the included studies based on the QUADAS-2 scale.

Supplementary Figure S2 | Forest plots based on deep learning model in validation set for preoperative prediction of MVI in HCC.

Supplementary Figure S3 | Forest plots based on non-deep learning model in validation set for preoperative prediction of MVI in HCC.

Supplementary Figure S4 | Forest plots based on non-deep learning model excluding studies using ultrasound for preoperative prediction of MVI in HCC.

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. Zhou J, Sun H, Wang Z, Cong W, Wang J, Zeng M, et al. Guidelines for the Diagnosis and Treatment of Hepatocellular Carcinoma (2019 Edition). *Liver Cancer* (2020) 9(6):682–720. doi: 10.1159/000509424
3. Shah SA, Cleary SP, Wei AC, Yang I, Taylor BR, Hemming AW, et al. Recurrence After Liver Resection for Hepatocellular Carcinoma: Risk Factors, Treatment, and Outcomes. *Surgery* (2007) 141(3):330–9. doi: 10.1016/j.surg.2006.06.028
4. Chen ZH, Zhang XP, Wang H, Chai ZT, Sun JX, Guo WX, et al. Effect of Microvascular Invasion on the Postoperative Long-Term Prognosis of Solitary Small HCC: A Systematic Review and Meta-Analysis. *HPB* (2019) 21(8):935–44. doi: 10.1016/j.hpb.2019.02.003
5. Lim KC, Chow PK, Allen JC, Chia GS, Lim M, Cheow PC, et al. Microvascular Invasion is a Better Predictor of Tumor Recurrence and Overall Survival Following Surgical Resection for Hepatocellular Carcinoma Compared to the Milan Criteria. *Ann Surg* (2011) 254(1):108–13. doi: 10.1097/SLA.0b013e31821ad884
6. Vitale A, Trevisani F, Farinati F, Cillo U. Treatment of Hepatocellular Carcinoma in the Precision Medicine Era: From Treatment Stage Migration to Therapeutic Hierarchy. *Hepatology (Baltimore Md)* (2020) 72(6):2206–18. doi: 10.1002/hep.31187
7. Imai K, Yamashita YI, Yusa T, Nakao Y, Itoyama R, Nakagawa S, et al. Microvascular Invasion in Small-Sized Hepatocellular Carcinoma: Significance for Outcomes Following Hepatectomy and Radiofrequency Ablation. *Anticancer Res* (2018) 38(2):1053–60. doi: 10.21873/anticanres.12322
8. Lee S, Kang TW, Song KD, Lee MW, Rhim H, Lim HK, et al. Effect of Microvascular Invasion Risk on Early Recurrence of Hepatocellular Carcinoma After Surgery and Radiofrequency Ablation. *Ann Surg* (2021) 273(3):564–71. doi: 10.1097/SLA.0000000000003268
9. Hirokawa F, Hayashi M, Miyamoto Y, Asakuma M, Shimizu T, Komeda K, et al. Outcomes and Predictors of Microvascular Invasion of Solitary Hepatocellular Carcinoma. *Hepatol Res* (2014) 44(8):846–53. doi: 10.1111/hepr.12196
10. Mazzaferro V, Sposito C, Zhou J, Pinna AD, De Carlis L, Fan J, et al. Metroticket 2.0 Model for Analysis of Competing Risks of Death After Liver Transplantation for Hepatocellular Carcinoma. *Gastroenterology* (2018) 154(1):128–39. doi: 10.1053/j.gastro.2017.09.025
11. Cong WM, Bu H, Chen J, Dong H, Zhu YY, Feng LH, et al. Practice Guidelines for the Pathological Diagnosis of Primary Liver Cancer: 2015 Update. *World J Gastroenterol* (2016) 22(42):9279–87. doi: 10.3748/wjg.v22.i42.9279
12. 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
13. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, et al. Dermatologist-Level Classification of Skin Cancer With Deep Neural Networks. *Nature* (2017) 542(7639):115–8. doi: 10.1038/nature21056
14. Ting D, Cheung CY, Lim G, Tan G, Quang ND, Gan A, et al. Development and Validation of a Deep Learning System for Diabetic Retinopathy and Related Eye Diseases Using Retinal Images From Multiethnic Populations With Diabetes. *JAMA* (2017) 318(22):2211–23. doi: 10.1001/jama.2017.18152
15. Cuocolo R, Cipullo MB, Stanzione A, Romeo V, Green R, Cantoni V, et al. Non-Deep Learning for the Identification of Clinically Significant Prostate Cancer on MRI: A Meta-Analysis. *Eur Radiol* (2020) 30(12):6877–87. doi: 10.1007/s00330-020-07027-w
16. Cho SJ, Sunwoo L, Baik SH, Bae YJ, Choi BS, Kim JH. Brain Metastasis Detection Using Machine Learning: A Systematic Review and Meta-Analysis. *Neuro Oncol* (2021) 23(2):214–25. doi: 10.1093/neuonc/noaa232
17. Wang K, Lu X, Zhou H, Gao Y, Zheng J, Tong M, et al. Deep Learning Radiomics of Shear Wave Elastography Significantly Improved Diagnostic Performance for Assessing Liver Fibrosis in Chronic Hepatitis B: A Prospective Multicentre Study. *Gut* (2019) 68(4):729–41. doi: 10.1136/gutjnl-2018-316204
18. 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(12):1218–30. doi: 10.4251/wjgo.v11.i12.1218
19. Decharatanachart P, Chaiteerakij R, Tiyarattanachai T, Treeprasertsuk S. Application of Artificial Intelligence in Chronic Liver Diseases: A Systematic Review and Meta-Analysis. *BMC Gastroenterol* (2021) 21(1):10. doi: 10.1186/s12876-020-01585-5
20. Lai Q, Spoletini G, Mennini G, Laureiro ZL, Tsilimigras DI, Pawlik TM, et al. Prognostic Role of Artificial Intelligence Among Patients With Hepatocellular Cancer: A Systematic Review. *World J Gastroenterol* (2020) 26(42):6679–88. doi: 10.3748/wjg.v26.i42.6679
21. 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
22. Greenspan H, van Ginneken B, Summers RM. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Trans Med Imaging* (2016) 35(5):1153–9. doi: 10.1109/TMI.2016.2553401
23. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep Learning for Health Informatics. *IEEE J Biomed Health Inform* (2017) 21(1):4–21. doi: 10.1109/JBHI.2016.2636665
24. Hussein S, Kandel P, Bolan CW, Wallace MB, Bagci U. Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches. *IEEE Trans Med Imaging* (2019) 38(8):1777–87. doi: 10.1109/TMI.2019.2894349
25. Huang J, Tian W, Zhang L, Huang Q, Lin S, Ding Y, et al. Preoperative Prediction Power of Imaging Methods for Microvascular Invasion in Hepatocellular Carcinoma: A Systemic Review and Meta-Analysis. *Front Oncol* (2020) 10:887:887. doi: 10.3389/fonc.2020.00887
26. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. *Ann Internal Med* (2011) 155(8):529–36. doi: 10.7326/0003-4819-155-8-201110180-00009
27. 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) 27147(12):3757–67. doi: 10.1007/s00432-021-03617-3

28. Jiang YQ, Cao SE, Cao S, Chen JN, Wang GY, Shi WQ, et al. Preoperative Identification of Microvascular Invasion in Hepatocellular Carcinoma by XGBoost and Deep Learning. *J Cancer Res Clin Oncol* (2021) 147(3):821–33. doi: 10.1007/s00432-020-03366-9
29. Wang G, Jian W, Cen X, Zhang L, Guo H, Liu Z, et al. Prediction of Microvascular Invasion of Hepatocellular Carcinoma Based on Preoperative Diffusion-Weighted MR Using Deep Learning. *Acad Radiol* (2020) 28 Suppl 1: S118–27. doi: 10.1016/j.acra.2020.11.014
30. 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) 11:588010:588010. doi: 10.3389/fonc.2021.588010
31. 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 Imaging* (2021) 54 (1):134–43. doi: 10.1002/jmri.27538Song
32. 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* (2021) 13(10):2368. doi: 10.3390/cancers13102368
33. 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
34. Liu P, Tan XZ, Zhang T, Gu QB, Mao XH, Li YC, et al. Prediction of Microvascular Invasion in Solitary Hepatocellular Carcinoma ≤ 5 Cm Based on Computed Tomography Radiomics. *World J Gastroenterol* (2021) 27 (17):2015–24. doi: 10.3748/wjg.v27.i17.2015
35. Dong Y, Zhou L, Xia W, Zhao XY, Zhang Q, Jian JM, et al. Preoperative Prediction of Microvascular Invasion in Hepatocellular Carcinoma: Initial Application of a Radiomic Algorithm Based on Grayscale Ultrasound Images. *Front Oncol* (2020) 10:353:353. doi: 10.3389/fonc.2020.00353
36. Xu X, Zhang HL, Liu QP, Sun SW, Zhang J, Zhu FP, et al. Radiomic Analysis of Contrast-Enhanced CT Predicts Microvascular Invasion and Outcome in Hepatocellular Carcinoma. *J Hepatol* (2019) 70(6):1133–44. doi: 10.1016/j.jhep.2019.02.023
37. Yao Z, Dong Y, Wu G, Zhang Q, Yang D, Yu JH, et al. Preoperative Diagnosis and Prediction of Hepatocellular Carcinoma: Radiomics Analysis Based on Multi-Modal Ultrasound Images. *BMC Cancer* (2018) 18(1):1089. doi: 10.1186/s12885-018-5003-4
38. Ni M, Zhou X, Lv Q, Li Z, Gao Y, Tan Y, et al. Radiomics Models for Diagnosing Microvascular Invasion in Hepatocellular Carcinoma: Which Model is the Best Model? *Cancer Imaging* (2019) 19(1):60. doi: 10.1186/s40644-019-0249-x
39. 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 (Ankara Turkey)* (2018) 24(3):121–7. doi: 10.5152/dir.2018.17467
40. Hu HT, Wang Z, Huang XW, Chen SL, Zheng X, Ruan SM, et al. Ultrasound-Based Radiomics Score: A Potential Biomarker for the Prediction of Microvascular Invasion in Hepatocellular Carcinoma. *Eur Radiol* (2019) 29 (6):2890–901. doi: 10.1007/s00330-018-5797-0
41. Ma X, Wei J, Gu D, Zhu Y, Feng B, Liang M, et al. Preoperative Radiomics Nomogram for Microvascular Invasion Prediction in Hepatocellular Carcinoma Using Contrast-Enhanced CT. *Eur Radiol* (2019) 29(7):3595–605. doi: 10.1007/s00330-018-5985-y
42. Nebbia G, Zhang Q, Arefan D, Zhao X, Wu S. Pre-Operative Microvascular Invasion Prediction Using Multi-Parametric Liver MRI Radiomics. *J Digit Imaging* (2020) 33(6):1376–86. doi: 10.1007/s10278-020-00353-x
43. 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 Non-Deep Learning for Diagnosing Microvascular Invasion. *AJR Am J Roentgenol* (2021) 216(6):1510–20. doi: 10.2214/AJR.20.23255
44. Nitta H, Allard MA, Sebagh M, Ciacio O, Pittau G, Vibert E, et al. Prognostic Value and Prediction of Extratumoral Microvascular Invasion for Hepatocellular Carcinoma. *Ann Surg Oncol* (2019) 26(8):2568–76. doi: 10.1245/s10434-019-07365-0
45. Poté N, Cauchy F, Albuquerque M, Voitot H, Belghiti J, Castera L, et al. Performance of PIVKA-II for Early Hepatocellular Carcinoma Diagnosis and Prediction of Microvascular Invasion. *J Hepatol* (2015) 62(4):848–54. doi: 10.1016/j.jhep.2014.11.005
46. Wang L, Jin YX, Ji YZ, Mu Y, Zhang SC, Pan SY. Development and Validation of a Prediction Model for Microvascular Invasion in Hepatocellular Carcinoma. *World J Gastroenterol* (2020) 26(14):1647–59. doi: 10.3748/wjg.v26.i14.1647
47. Zeng F, Chen B, Zeng J, Wang Z, Xiao L, Deng G. Preoperative Neutrophil-Lymphocyte Ratio Predicts the Risk of Microvascular Invasion in Hepatocellular Carcinoma: A Meta-Analysis. *Int J Biol Markers* (2019) 34 (3):213–20. doi: 10.1177/1724600819874487
48. Zheng J, Seier K, Gonen M, Balachandran VP, Kingham TP, D'Angelica MI, et al. Utility of Serum Inflammatory Markers for Predicting Microvascular Invasion and Survival for Patients With Hepatocellular Carcinoma. *Ann Surg Oncol* (2017) 24(12):3706–14. doi: 10.1245/s10434-017-6060-7
49. Zhu Y, Xu D, Zhang Z, Dong J, Zhou Y, Zhang WW, et al. A New Laboratory-Based Algorithm to Predict Microvascular Invasion and Survival in Patients With Hepatocellular Carcinoma. *Int J Surg (London England)* (2018) 57:45–53. doi: 10.1016/j.ijsu.2018.07.011
50. Hong SB, Choi SH, Kim SY, Shim JH, Lee SS, Byun JH, et al. MRI Features for Predicting Microvascular Invasion of Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis. *Liver Cancer* (2021) 10(2):94–106. doi: 10.1159/000513704
51. Verloh N, Utpatel K, Zeman F, Fellner C, Schlitt HJ, Müller M, et al. Diagnostic Performance of Gd-EOB-DTPA-Enhanced MRI for Evaluation of Liver Dysfunction: A Multivariable Analysis of 3T MRI Sequences. *Oncotarget* (2018) 9(91):36371–8. doi: 10.18632/oncotarget.26368
52. Golfieri R, Renzulli M, Lucidi V, Corcioni B, Trevisani F, Bolondi L. Contribution of the Hepatobiliary Phase of Gd-EOB-DTPA-Enhanced MRI to Dynamic MRI in the Detection of Hypovascular Small (≤ 2 Cm) HCC in Cirrhosis. *Eur Radiol* (2011) 21(6):1233–42. doi: 10.1007/s00330-010-2030-1
53. 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 Imaging* (2021) 54(2):526–36. doi: 10.1002/jmri.27575
54. Zhang X, Ruan S, Xiao W, Shao J, Tian W, Liu W, et al. Contrast-Enhanced CT Radiomics for Preoperative Evaluation of Microvascular Invasion in Hepatocellular Carcinoma: A Two-Center Study. *Clin Trans Med* (2020) 10 (2):e111. doi: 10.1002/ctm2.111
55. Monkam P, Qi S, Ma H, Gao W, Yao Y, Qian W. Detection and Classification of Pulmonary Nodules Using Convolutional Neural Networks: A Survey. *IEEE Access* (2019) 7:78075–91. doi: 10.1109/ACCESS.2019.2920980

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 24 February 2022
doi: 10.3389/fonc.2022.819673

Multimodal MRI Image Decision Fusion-Based Network for Glioma Classification

Shunchao Guo[1,2], Lihui Wang[1], Qijian Chen[1], Li Wang[1], Jian Zhang[1] and Yuemin Zhu[3]**

[1] Key Laboratory of Intelligent Medical Image Analysis and Precise Diagnosis of Guizhou Province, College of Computer Science and Technology, Guizhou University, Guiyang, China, [2] College of Computer and Information, Qiannan Normal University for Nationalities, Duyun, China, [3] CREATIS, CNRS UMR 5220, Inserm U1044, INSA Lyon, University of Lyon, Lyon, France

OPEN ACCESS

Edited by:
Antonio Napolitano, Bambino Gesù Children's Hospital (IRCCS), Italy

Reviewed by:
Weiwei Zong, Henry Ford Health System, United States
Yulin Song, Memorial Sloan Kettering Cancer Center, United States

***Correspondence:**
Lihui Wang
lhwang2@gzu.edu.cn
Yuemin Zhu
yue-min.zhu@creatis.insa-lyon.fr

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

Received: 22 November 2021
Accepted: 24 January 2022
Published: 24 February 2022

Citation:
Guo S, Wang L, Chen Q, Wang L, Zhang J and Zhu Y (2022) Multimodal MRI Image Decision Fusion-Based Network for Glioma Classification. Front. Oncol. 12:819673. doi: 10.3389/fonc.2022.819673

Purpose: Glioma is the most common primary brain tumor, with varying degrees of aggressiveness and prognosis. Accurate glioma classification is very important for treatment planning and prognosis prediction. The main purpose of this study is to design a novel effective algorithm for further improving the performance of glioma subtype classification using multimodal MRI images.

Method: MRI images of four modalities for 221 glioma patients were collected from Computational Precision Medicine: Radiology-Pathology 2020 challenge, including T1, T2, T1ce, and fluid-attenuated inversion recovery (FLAIR) MRI images, to classify astrocytoma, oligodendroglioma, and glioblastoma. We proposed a multimodal MRI image decision fusion-based network for improving the glioma classification accuracy. First, the MRI images of each modality were input into a pre-trained tumor segmentation model to delineate the regions of tumor lesions. Then, the whole tumor regions were centrally clipped from original MRI images followed by max–min normalization. Subsequently, a deep learning-based network was designed based on a unified DenseNet structure, which extracts features through a series of dense blocks. After that, two fully connected layers were used to map the features into three glioma subtypes. During the training stage, we used the images of each modality after tumor segmentation to train the network to obtain its best accuracy on our testing set. During the inferring stage, a linear weighted module based on a decision fusion strategy was applied to assemble the predicted probabilities of the pre-trained models obtained in the training stage. Finally, the performance of our method was evaluated in terms of accuracy, area under the curve (AUC), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), etc.

Results: The proposed method achieved an accuracy of 0.878, an AUC of 0.902, a sensitivity of 0.772, a specificity of 0.930, a PPV of 0.862, an NPV of 0.949, and a Cohen's Kappa of 0.773, which showed a significantly higher performance than existing state-of-the-art methods.

Conclusion: Compared with current studies, this study demonstrated the effectiveness and superiority in the overall performance of our proposed multimodal MRI image decision fusion-based network method for glioma subtype classification, which would be of enormous potential value in clinical practice.

Keywords: glioma classification, multimodal MRI images, decision fusion, tumor segmentation, deep learning

INTRODUCTION

Glioma is the most common primary tumor of the brain and spine, representing 80% of malignant brain tumors (1) and having varying degrees of aggressiveness and prognosis. The clinical manifestations of glioma include increased intracranial pressure, neurological/cognitive dysfunction, and seizures. According to the 2016 WHO classification of tumors of the central nervous system (CNS), glioma can be classified into astrocytoma (grade II or III), oligodendroglioma or mesenchymal oligodendroglioma (grade II or III), and glioblastoma (grade IV), depending on the pathology and molecular alterations (2). Low-grade glioma is well-differentiated and presents an aggressive growth pattern in terms of biological characteristics, whereas high-grade glioma is a malignant brain tumor that is difficult to identify and has a poor prognosis (3).

Precise glioma classification or grading is crucial for deciding the right therapeutic strategies that may further impact the prognosis process of patients (4, 5). In clinical practice, MRI is the standard medical imaging technique for brain tumor diagnosis for its advantages of relative safety and non-invasiveness as compared to pathological biopsy examinations (6). With respect to unimodal MRI images, multimodal MRI images can provide more morphological, functional, and tumor metabolic status information due to their correlation and complementary information for all types of brain tumors. Clinically, the low contrast between tumor masses and surrounding tissues as well as the varying levels of physicians' experience may lead to misdiagnosis; more importantly, diagnosing based on manual analysis is a time-consuming procedure (7). With the development of artificial intelligence and computing facilities, computer-aided diagnosis (CAD) technology based on computer vision has been applied to many medical fields and provides help for physicians in visualization and tumor identification to improve the subjective diagnosis manually (8).

So far, the methods for brain tumor classification in the latest studies can be loosely classified into two categories: traditional machine learning methods and deep learning methods (9). Among the latest traditional machine learning methods (5, 10–12), the most commonly used one is radiomics. Radiomics uses data characterization algorithms to extract quantitative features from MRI images (13, 14), and these features usually contain complex patterns that are difficult to recognize or quantify by human eyes, such as tumor heterogeneity, infiltration, and metastasis (15). The other general method is deep learning, which was successfully applied to tumor segmentation (16), tumor classification (7, 8, 17), survival prediction (4, 18), and molecular genetic prediction (19, 20) for its powerful feature representation in medical imaging fields. Compared with radiomics-based methods, deep learning-based methods do not need domain-specific knowledge for feature extraction and outcompete the formers when experimental data are sufficient. Furthermore, considering the powerful feature learning capability of deep learning and the powerful classification capability of traditional machine learning, researchers have combined them together for glioma classification or grading (21–23).

The aim of this study was to diagnose the glioma subtype preoperatively using MRI images only for assisting in making appropriate treatment decisions. Misdiagnosis caused by inaccurate glioma prediction algorithm may lead to severe injury or death, so prediction accuracy is the most concerned performance undoubtedly. Since Computational Precision Medicine- Radiology-Pathology (CPM-RadPath) on Brain Tumor Classification challenge held in 2018, many studies have been conducted in glioma subtypes prediction using multimodal MRI images based on tumor segmentation. Pei et al. (4) proposed a 3D convolutional neural network (CNN) model for glioma classification based on tumor segmentation results from the CANet model, and experimental results demonstrated the effectiveness of using MRI images only. Xue et al. (24) trained a 3D residual convolutional network with MRI images for glioma classification, and the results showed that using tumor segmentation regions would get higher accuracy. Pei et al. (25) applied a 3D CNN model with MRI images for brain tumor segmentation, which distinguished brain tumors from healthy tissues, and then the segmented tumors were used for tumor subtype classification with another 3D CNN model. Yin et al. (26) achieved the first rank in CPM-RadPath 2020 using both MRI and pathological images. For multimodal MRI images, they used the pre-trained model on the Brain Tumor Segmentation (BraTS) challenge 2019 for tumor segmentation and then built a densely connected convolutional network (DenseNet) model for glioma prediction in their scheme. Although promising in their results, they concatenated the multiple modalities as different input channels such that a deep learning network could automatically learn to extract the multimodal features. With such an image fusion strategy, it is difficult to adjust the contributions of each modality for prediction results and consequently not easy to get the best classification accuracy.

Using multimodal MRI images for glioma subtype classification has great clinical potentiality and guidance value. In order to further improve the glioma subtypes prediction

accuracy in clinical applications, we propose a Multimodal MRI Image Decision Fusion-based Network (MMIDFNet) based on the deep learning method. Inspired by image fusion (27), the proposed method uses a linear weighed module to assemble the models trained with a single modality together for mining their complementary predictive capabilities. To evaluate the effectiveness of our proposed method, we compared the classification performance between our method and recent state-of-the-art methods. Additionally, since radiomics-based methods are also commonly used in recent brain tumor classification studies, we also implemented a radiomics-based method as the benchmark for performance comparison with our MMIDFNet.

MATERIALS AND METHODS

Study Cohort

Our experimental data were obtained from the CPM-RadPath challenge 2020 dataset[1], which is supported by Medical Image Computing and Computer-Assisted Intervention Society. The dataset classified patients into three subtypes based on the WHO-CNS pathomorphological classification criteria, named Glioblastoma (abbreviated as "G"), Astrocytoma (abbreviated as "A"), and Oligodendroglioma (abbreviated as "O") separately (25, 26). Each patient contained preoperative 3D MRI images and pathological whole slide images. MRI images comprise four different modalities of T1-weighted (T1), T2-weighted (T2), post-contrast T1-weighted (T1ce), and fluid-attenuated inversion recovery (FLAIR). Considering the purposes of our study, we just used MRI images to predict pathological subtypes. According to the dataset description, MRI images were obtained from multi-parametric MRI scans in routine clinics with 1T to 3T MRI scanners in multi-center institutions and stored in NIfTI format. All four MRI modalities were preprocessed with bias field correction, skull stripping, and co-registration into the same anatomical structure template (24). The volume size of each MRI modality data is 240 × 240 × 155, where 155 indicates the number of slices. The cohort in our experiments consisted of 221 patients collected from the original dataset, in which there were 133, 54, and 34 samples provided for subtype "G", "A", and "O", respectively. To overcome the bias caused by a particular selection for the pair of training and testing sets, a 3-fold cross-validation strategy was used in this work. Specifically, the dataset was split into 3 smaller sets, the model was trained using 2 of the folds as training data, and then the trained model was validated on the remaining part of the data. The performance reported by 3-fold cross-validation was measured with the averaged evaluation indices.

MMIDFNet Architecture

To improve the accuracy of glioma diagnosis using multimodal MRI images, we designed the MMIDFNet for glioma subtype classification, as shown in **Figure 1**. The MMIDFNet architecture includes two parts: one is the tumor segmentation module using a pre-trained model, and the other is the two stages of unimodal model training and multimodal image decision fusion inferring module. In the training stage, we used the images of each single modality to train the network to obtain its best accuracy on the testing set. In the present work, we used images from four MRI modalities. Thus, we have four pre-trained models. In inferring stage, a decision fusion strategy was used; in other words, a linear weighted module was applied to assemble the predicted probabilities of the above four pre-trained models for each modality. Adopting the decision fusion strategy, we can fully take advantage of the complementary capabilities among unimodal models trained from different modalities. Note that the weights for the four MRI modalities in our linear weighted module did not participate in training in inferring stage.

Tumor Segmentation

Accurate segmentation of brain tumors from MRI images is of enormous potential value for improved diagnosis (28). It can be done automatically to cope with the time-consuming disadvantage of manual segmentation (29, 30). Considering that the MRI images in our study were also used in the BraTS challenge and that the ground truth of tumor segmentation for patients in our cohort are not all available, we used the pre-trained model on the BraTS challenge 2019 to delineate the regions of tumor lesions, which achieved the accuracy of 90.45% on the validation set (31). In the BraTS 2019 dataset, all the samples in the training set are provided with four ground truth labels for 4 regions: background (label 0), necrotic and non-enhanced tumor (label 1), peritumoral edema (label 2), and enhanced tumor (label 4). We reassigned the non-zero labels into three combined subregions, representing enhanced tumor (ET: label 4), tumor core (TC: label 1 + label 4), and whole tumor (WT: label 1 + label 2 + label 4). The WT, TC, and ET regions of the MRI images were obtained by the pre-trained segmentation model. Since glioma grows within the substance of the brain and often mixes with normal brain tissues, the surrounding area is also valuable for the assessment. Hence, we used the whole tumor regions as the segmentation regions of interest (ROIs) and centrally cropped the original image to 128 × 128 × 80. In order to make the intensities of the cropped images more homogeneous, max–min normalization was applied. Besides, for all the patients having ground truth, after carefully comparing the pairwise central locations of WT regions obtained by our tumor segmentation and the ground truth, we found that they were all consistent or nearby, which also demonstrates that our tumor segmentation scheme is feasible. A glioblastoma patient case before and after segmentation with the pre-trained model is displayed in **Figure 2** using the ITK-SNAP software. The red region indicates the necrotic and non-enhanced tumor, the yellow region the enhanced tumor, and the green region the peritumoral edema.

Unimodal Prediction Model Building

Through stacking multiple convolutional layers together, deep neural networks (DNNs) can automatically learn discriminative

[1] https://miccai.westus2.cloudapp.azure.com/competitions/1

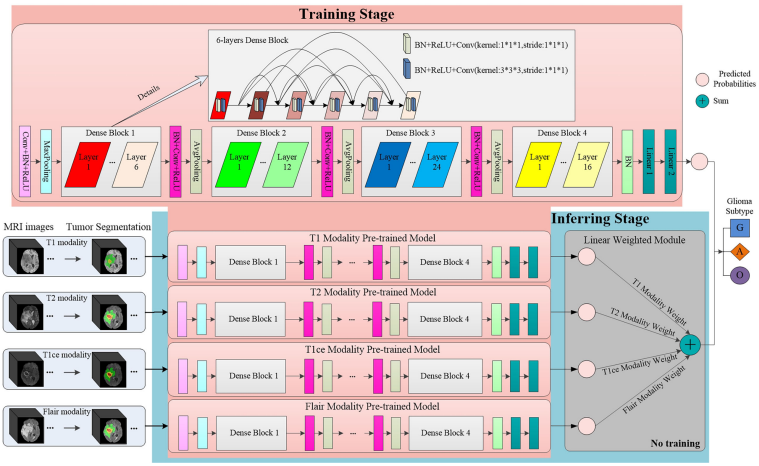


FIGURE 1 | The structure of our proposed MMIDFNet.

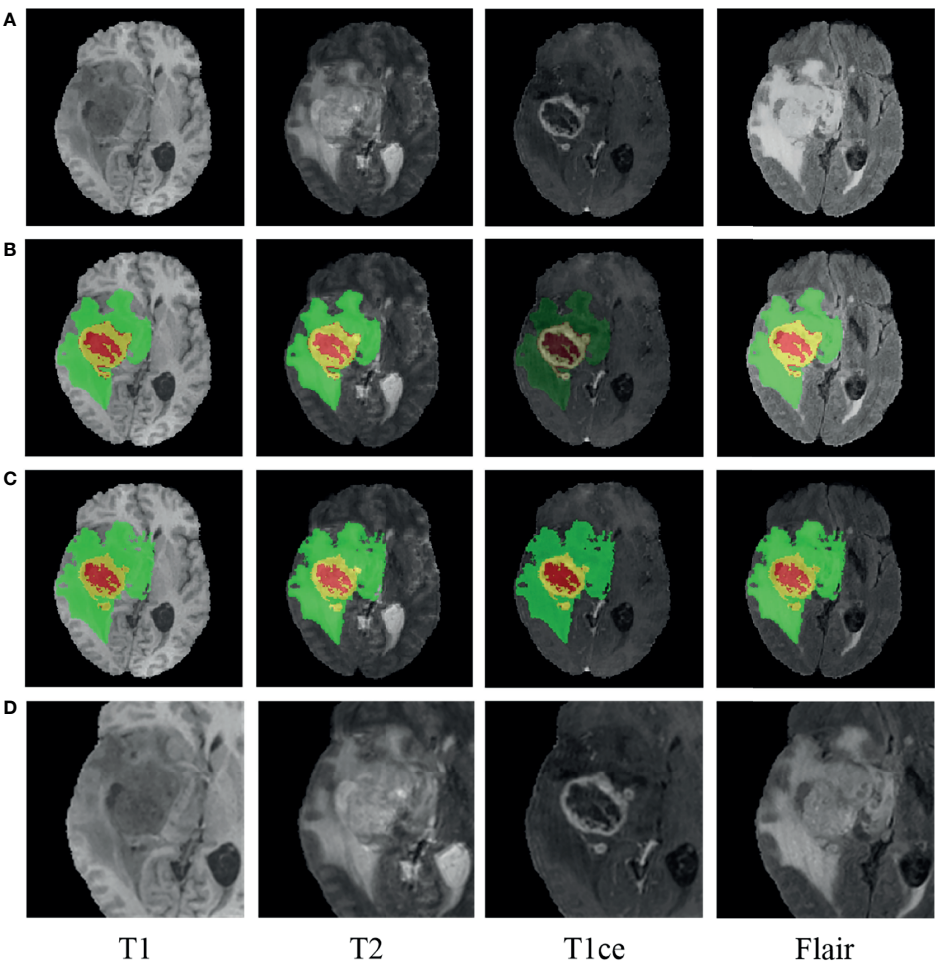


FIGURE 2 | An example of glioma patients on multimodal MRI images (patient ID: CPM19_CBICA_AAB_1, Glioblastoma). **(A)** Original images. **(B)** Tumor segmentation on panel **(A)**. **(C)** Ground truth on panel **(A)**. **(D)** Normalized followed centrally clipped from panel **(A)** based on panel **(B)**.

features from imaging data. Among different DNN models as well as their variants, DenseNet has shown superior classification performance as it strengthens feature propagation from one dense block to the next and overcomes the problem of vanishing gradient (32). In the training stage of our MMIDFNet method, we designed a network based on unified DenseNet architecture with 121 layers (32), which included four dense blocks, and each dense block was composed of several convolutional layers (BN+ReLU+Conv). The number of convolutional layers in four dense blocks was set as 6, 12, 24, and 16. The features of images were extracted through these dense blocks. After that, two fully connected layers were used to map the features into three glioma subtypes. In order to maintain the 3D structural features of MRI images, we used them as the model input directly without converting them into 2D slices. In the training stage, the model loss used focal loss for handling class imbalance, the initial learning rate was set to 5×10^{-4} with the updating strategy of scheduler optimization, the optimizer used Adam algorithm with impulse, the batch size was set to 8, and the number of training epoch was set to 100. To overcome overfitting in the training process, many strategies (e.g., sample normalization, data augmentation, applying L2 normalization to model loss, designing dropout layer for the model, and setting weight decay for optimizer) were employed. Our unimodal prediction models were implemented by the PyTorch framework (version 1.4.0, Facebook), and the details of network architecture can be found in **Figure 1**. All the data augmentation strategies of dimension resizing, random rotation, random scaling, random Gaussian noise adding, and random contrast adjusting were implemented using the Medical Open Network for AI (MONAI) toolkit[2]. Since our model with a large number of parameters would lead to high computational cost, we used NVIDIA Tesla A100 GPU to reduce the running time of model training and validation.

Multimodal Prediction Model Building

Due to the correlation among different modality images, multimodal MRI can provide help to extract features from different views and bring complementary information (33). For exploring richer patterns among multimodal MRI images to handle the issue of insufficient classification ability and generalization ability of the unimodal model, we used multimodal MRI images to build prediction models based on image decision fusion strategy. In our MMIDFNet method, since the classification capacity of different pre-trained models using a single MRI modality in the training stage is usually different and complementary, we assembled a multimodal prediction model with multimodal MRI images by calculating the linear weighted sum of predicted probabilities from the four pre-trained unimodal models trained in the training stage. Based on the heuristic searching strategy, the weights of each pre-trained unimodal model were assigned according to their classification ability. The final predicted probabilities of our MMIDFNet model for gliomas subtype classification was calculated as follows:

$$fusion\_prob = \sum_{i=1}^{m} w_i * prob_i \quad (1)$$

where m is the number of pre-trained models, and w_i and $prob_i$ are the weight and the prediction probability of the ith pre-trained model, respectively.

Radiomics-Based Prediction Models Building

The radiomics-based method mainly includes four stages: tumor segmentation, radiomics feature extraction, feature selection, and classification model building (34). The major challenge of the radiomics-based method is how to extract features from 3D MRI images. By using the Pyradiomics[3] package, for each original 3D MRI image, a total of 106 radiomics features were extracted based on the mask obtained in the above tumor segmentation stage. These features were composed of 18 first-order statistical features, 14 shape features, and 74 wavelet texture features. Besides, 12 other types of images (8 wavelets, gradient, mean square, root mean square, and exponential) transformed from each 3D MRI image were also used to extract radiomics features. Finally, the features of the original image as well as its 12 transformed images were concatenated together in an end-to-end manner, thus forming a total of 1,378 (106×13) features for each MRI image. Similarly, as for multimodal MRI images of each patient, their features were also concatenated together and formed the joint features for training a multimodal glioma prediction model. In the present work, we used the images from four MRI modalities. Hence, the dimension of the joint features is 5,512 ($1{,}378 \times 4$).

Considering that redundant and irrelevant features in high-dimensional features usually influence learning accuracy (35, 36), the least absolute shrinkage and selection operator (Lasso) regression algorithm was performed to reduce feature dimension by retaining high discriminative features (5, 27). Based on the selected features, the random forest (RF) classification model, which is frequently used in the field of supervised machine learning (5, 11, 12, 30, 37), was built for our task of glioma diagnosis. The Lasso and RF algorithms were implemented using the scikit-learn library (version 0.23.1).

Performance Metrics

The performance of our multi-class predictive models was assessed according to the commonly used accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and the area under the curve (AUC) of the receiver operating characteristic (ROC). The goal of the CPM-RadPath challenge is to assess automated brain tumor classification algorithms using three metrics, namely, F1_score, Cohen's Kappa (Kappa), and balanced accuracy (Balanced_Acc), which are sensitive to the imbalanced distribution of sample classes. Among the above metrics, F1_score and Balanced_Acc are defined as accuracy and sensitivity in multi-class metrics, respectively. The formulas for calculating the performance metrics of accuracy, sensitivity, specificity, PPV, NPV, and

[2] https://github.com/Project-MONAI/MONAI

[3] https://pyradiomics.readthedocs.io/en/latest/index.html

Kappa are given by Equations 2–7, respectively:

$$Accuracy = \frac{\sum_{i=1}^{n} TP_i}{Num} \tag{2}$$

$$Sensitivity = \frac{\sum_{i=1}^{n} TP_i/(TP_i + FN_i)}{n} \tag{3}$$

$$Specificity = \frac{\sum_{i=1}^{n} TN_i/(TN_i + FP_i)}{n} \tag{4}$$

$$PPV = \frac{\sum_{i=1}^{n} TP_i/(TP_i + FP_i)}{n} \tag{5}$$

$$NPV = \frac{\sum_{i=1}^{n} TN_i/(TN_i + FN_i)}{n} \tag{6}$$

$$Kappa = \frac{p_0 - p_e}{1 - p_e} \tag{7}$$

In Equations 2–6, *Num* is the number of samples, *TP* the true positives, *TN* the true negatives, *FP* the false positives, *FN* the false negatives, and n the number of sample categories. In Equation 7, p_0 denotes the sum of the number of samples for each correct classification divided by the total number of samples, and p_e the expected agreement when both annotators assign labels randomly (6). According to the accuracy metric, the best classifier was chosen as our predictive model for the task of glioma subtype classification.

Statistical Analysis

Age being the only available clinical factor in the CPM-RadPath challenge dataset (6), we converted it from days to years for simplicity before analysis. The differences in age and glioma subtypes between the training and testing sets were assessed using the Mann–Whitney rank-sum test. The statistical quantifications of the performance metrics were demonstrated with 95% CI, when applicable. All statistical analyses were carried out with the Scipy module (version 1.3.1), and p-value <0.05 indicated a significant difference.

RESULTS

Among these retrospective patients (age ranges 17 to 85 years), the mean ± SD of age was approximately 53.8 ± 14.8. In each fold, the number of subtypes "G", "A", and "O" was about 60.2%, 24.4%, and 15.4%, respectively. From the p-value results of the Mann–Whitney rank-sum test, we found that there was no significant difference in pathological subtypes (0.439, 0.423, and 0.48) among each fold.

Unimodal Prediction Models

After carefully tuning the parameters of unimodal models with our MMIDFNet method, we obtained the best pre-trained prediction models for each single modality in the training stage in turn. The ROC curves of each pre-trained unimodal model for each fold are plotted in **Figure 3**. We found that the prediction performance for each modality among different validation folds was not significantly different, which validates that the dataset selection has no significant influence on the prediction performance of our method.

In addition, the performance of unimodal prediction models using radiomics and our proposed MMIDFNet method on three validation folds is summarized in **Table 1**. We noticed that the split of the training and validation sets indeed influenced the prediction performance of both radiomics and our models, but not significantly. In general, our proposed MMIDFNet method generated better results than the radiomics method, with the highest averaged evaluation indices, except for the modalities of T1 and T2, in which the AUC, or sensitivity, or PPV was a little lower than that of radiomics. Moreover, we observed that with either our MMIDFNet method or radiomics method, using different unimodal MRI images achieved different classification performances. As for the radiomics method, among the unimodal prediction models on each fold, using the T1ce modality achieved the best-averaged accuracy of 0.815. The averaged AUC, sensitivity, specificity, PPV, and NPV were 0.868, 0.704, 0.882, 0.796, and 0.912, respectively. Meanwhile, for our MMIDFNet method, using the T1ce MRI modality achieved the best-averaged accuracy of 0.833. The averaged AUC, sensitivity, specificity, PPV, and NPV were 0.892, 0.708, 0.894, 0.817, and 0.924, respectively. This demonstrates that T1ce images may be beneficial to the glioma subtype classification.

Multimodal Prediction Models

Using our proposed MMIDFNet method, through repeatedly adjusting the weights of each unimodal prediction model in inferring stage, we obtained the best multimodal prediction accuracy. After tuning the parameters of our radiomics model iteratively, we also obtained the best prediction using the fused features of multimodal images. In this paper, the multimodal prediction methods obtained with radiomics and MMIDFNet were named as radiomics model and decision fusion model, respectively. Specifically, considering our designed network also supports multi-channel input in the training stage, through inputting four modalities into four-channel input of our MMIDFNet simultaneously, we trained and obtained another multimodal prediction model (named as data fusion model) based on data fusion strategy for comparing the predictive performance between data fusion strategy and decision fusion strategy in our MMIDFNet method. Here, the data fusion strategy means that the multiple modal images were concatenated as input. The ROC curves of the radiomics model, data fusion model, and decision fusion model using multimodal MRI images on each validation fold are illustrated in **Figure 4**. We found that for the glioma subtypes "G", and "A", the prediction performance of all the methods was not greatly influenced by the splitting of the

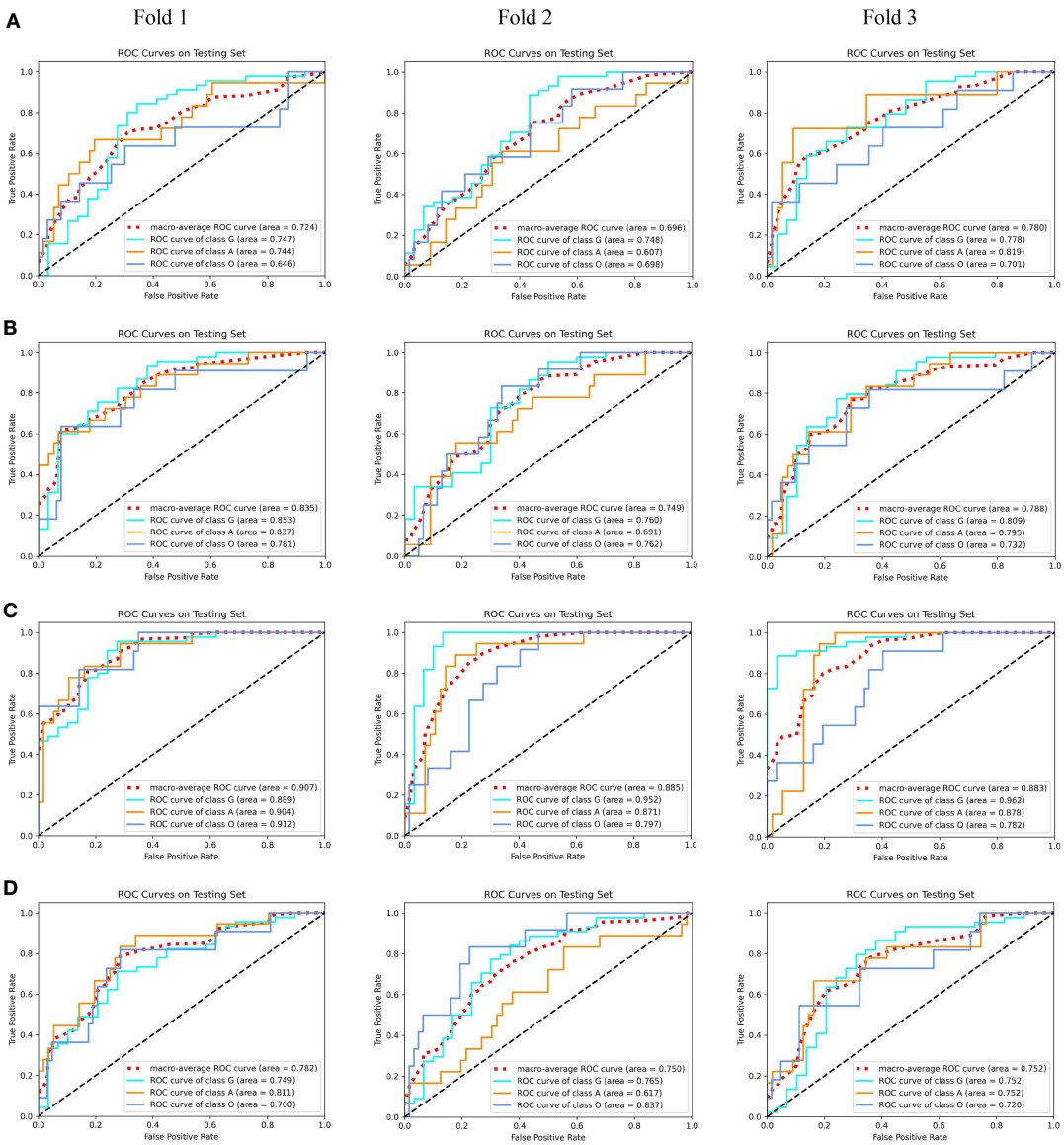


FIGURE 3 | The receiver operating characteristic (ROC) curves of unimodal prediction models on three validation folds using our MMIDFNet method. **(A)** T1 modality. **(B)** T2 modality. **(C)** T1ce modality. **(D)** Flair modality.

training and validation sets. However, for the glioma subtype "O", the data splitting strategy had significant effects on the predicted AUC. In addition, the overall prediction of our decision fusion model for a multi-class predictive task is more balanced than the other two multimodal models.

To further compare the prediction performance of different fusion strategies, **Table 2** summarizes the 3-fold cross-validation performance of each method on the multimodal dataset. We found that the overall performance of our proposed fusion method was much better than radiomics and data fusion strategy, with the averaged accuracy increased by 4.9% and 3.8%, averaged AUC increased by 3.7% and 2.2%, averaged sensitivity increased by 9.7% and 5.6%, averaged specificity increased by 3.4% and 1.5%, averaged PPV increased by 6.8% and 8.8%, averaged NPV increased by 1.9% and 2.2%, averaged Kappa increased by 12.2% and 8.0%, respectively. According to **Table 2**, the averaged evaluation indices were also demonstrated with bar plots in **Figure 5** for better illustration.

Comparing **Table 2** with **Table 1**, we observed that the overall performance of multimodal models was superior to that of any model trained with unimodal MRI images, whether for the radiomics or our MMIDFNet method. Meanwhile, through using the multimodal model with multimodal MRI images, the difference between sensitivity and specificity was significantly reduced.

To further evaluate the effectiveness of our proposed method, we compared our method with state-of-the-art

TABLE 1 | Three-fold cross-validation performance of unimodal prediction models using radiomics and our proposed MMIDFNet.

Methods	Modality	Fold	ACC	AUC	SEN	SPE	PPV	NPV
Radiomics	**T1**	1	0.730	0.734	0.546	0.792	0.807	0.865
		2	0.689	0.755	0.522	0.765	0.748	0.844
		3	0.699	0.737	0.610	0.792	0.682	0.812
		Average	**0.706**	**0.742**	**0.559**	**0.783**	**0.746**	**0.840**
		95% CI	[0.672, 0.740]	[0.724, 0.760]	[0.487, 0.632]	[0.758, 0.808]	[0.646, 0.846]	[0.798, 0.883]
	T2	1	0.703	0.775	0.554	0.787	0.657	0.824
		2	0.743	0.827	0.596	0.816	0.722	0.875
		3	0.712	0.712	0.560	0.820	0.615	0.843
		Average	**0.719**	**0.771**	**0.570**	**0.808**	**0.665**	**0.847**
		95% CI	[0.686, 0.753]	[0.679, 0.863]	[0.534, 0.606]	[0.779, 0.836]	[0.578, 0.751]	[0.806, 0.889]
	T1ce	1	0.838	0.908	0.706	0.890	0.867	0.928
		2	0.784	0.841	0.650	0.854	0.770	0.905
		3	0.822	0.856	0.756	0.903	0.752	0.902
		Average	**0.815**	**0.868**	**0.704**	**0.882**	**0.796**	**0.912**
		95% CI	[0.770, 0.859]	[0.812, 0.925]	[0.619, 0.789]	[0.842, 0.923]	[0.697, 0.895]	[0.889, 0.934]
	Flair	1	0.730	0.788	0.570	0.792	0.763	0.881
		2	0.685	0.718	0.557	0.787	0.625	0.813
		3	0.743	0.740	0.585	0.810	0.756	0.888
		Average	**0.719**	**0.749**	**0.571**	**0.796**	**0.715**	**0.861**
		95% CI	[0.671, 0.768]	[0.691, 0.806]	[0.548, 0.593]	[0.777, 0.816]	[0.590, 0.839]	[0.794, 0.927]
MMIDFNet	**T1**	1	0.757	0.724	0.572	0.821	0.813	0.894
		2	0.689	0.696	0.509	0.767	0.663	0.890
		3	0.712	0.780	0.516	0.777	0.701	0.873
		Average	**0.719**	**0.733**	**0.532**	**0.788**	**0.726**	**0.886**
		95% CI	[0.664, 0.775]	[0.665, 0.802]	[0.477, 0.588]	[0.742, 0.834]	[0.601, 0.850]	[0.868, 0.904]
	T2	1	0.743	0.835	0.542	0.794	0.742	0.907
		2	0.730	0.749	0.560	0.820	0.687	0.854
		3	0.726	0.788	0.591	0.822	0.642	0.853
		Average	**0.733**	**0.791**	**0.564**	**0.812**	**0.690**	**0.871**
		95% CI	[0.719, 0.747]	[0.722, 0.860]	[0.525, 0.604]	[0.787, 0.837]	[0.610, 0.770]	[0.822, 0.921]
	T1ce	1	0.838	0.907	0.764	0.885	0.842	0.909
		2	0.824	0.885	0.667	0.897	0.749	0.934
		3	0.836	0.883	0.694	0.900	0.859	0.929
		Average	**0.833**	**0.892**	**0.708**	**0.894**	**0.817**	**0.924**
		95% CI	[0.821, 0.845]	[0.870, 0.913]	[0.628, 0.788]	[0.881, 0.907]	[0.722, 0.911]	[0.903, 0.945]
	Flair	1	0.770	0.782	0.669	0.855	0.755	0.866
		2	0.703	0.750	0.537	0.767	0.813	0.896
		3	0.753	0.752	0.640	0.852	0.673	0.869
		Average	**0.742**	**0.761**	**0.615**	**0.825**	**0.747**	**0.877**
		95% CI	[0.686, 0.798]	[0.733, 0.790]	[0.504, 0.726]	[0.745, 0.905]	[0.634, 0.860]	[0.851, 0.903]

ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval; Bold Value, average value of 3 folds.

methods. The results are provided in **Table 3**. We observe that our decision fusion model achieves the highest averaged F1_score (0.878), Balanced_Acc (0.772), and Kappa (0.773) compared with the other methods. The F1_score of our method exceeds the methods of Pei at al (4)., Xue et al. (24), Pei et al. (25), and Yin et al. (26) at 5.9% (0.878 vs. 0.829), 13.9% (0.878 vs. 0.771), 13.9% (0.878 vs. 0.771), and 2.5% (0.878 vs. 0.857), respectively.

DISCUSSION

To further improve the performance of glioma subtype classification using MRI images only, we proposed a multimodal MRI image decision fusion-based network for glioma classification. In our method, during the training stage, we used the images of each MRI modality to train the network to obtain its best accuracy and obtained four pre-trained unimodal models. During the inferring stage, considering that different unimodal models have different predictive performance for three glioma subtypes, we assigned the weights for each unimodal model according to their classification capabilities to fully exploit their complementary predictive information of multi-class classification. Based on the decision fusion strategy, we assembled the four unimodal models together by using a linear weighted module and formed our multimodal prediction model for glioma subtype classification. The final predicted probabilities of the multimodal model were obtained by calculating the linear weighted sum of the predicted probabilities of the four pre-trained unimodal models. Thus, we improved the overall prediction performance of our multimodal prediction model by integrating the local predictive decision of each unimodal prediction model.

A finding of this study is that the decision fusion model using our MMIDFNet method outperformed the radiomics model based on radiomics in predicting glioma subtypes with multimodal MRI images (accuracy: 0.878 vs. 0.837). This is consistent with the

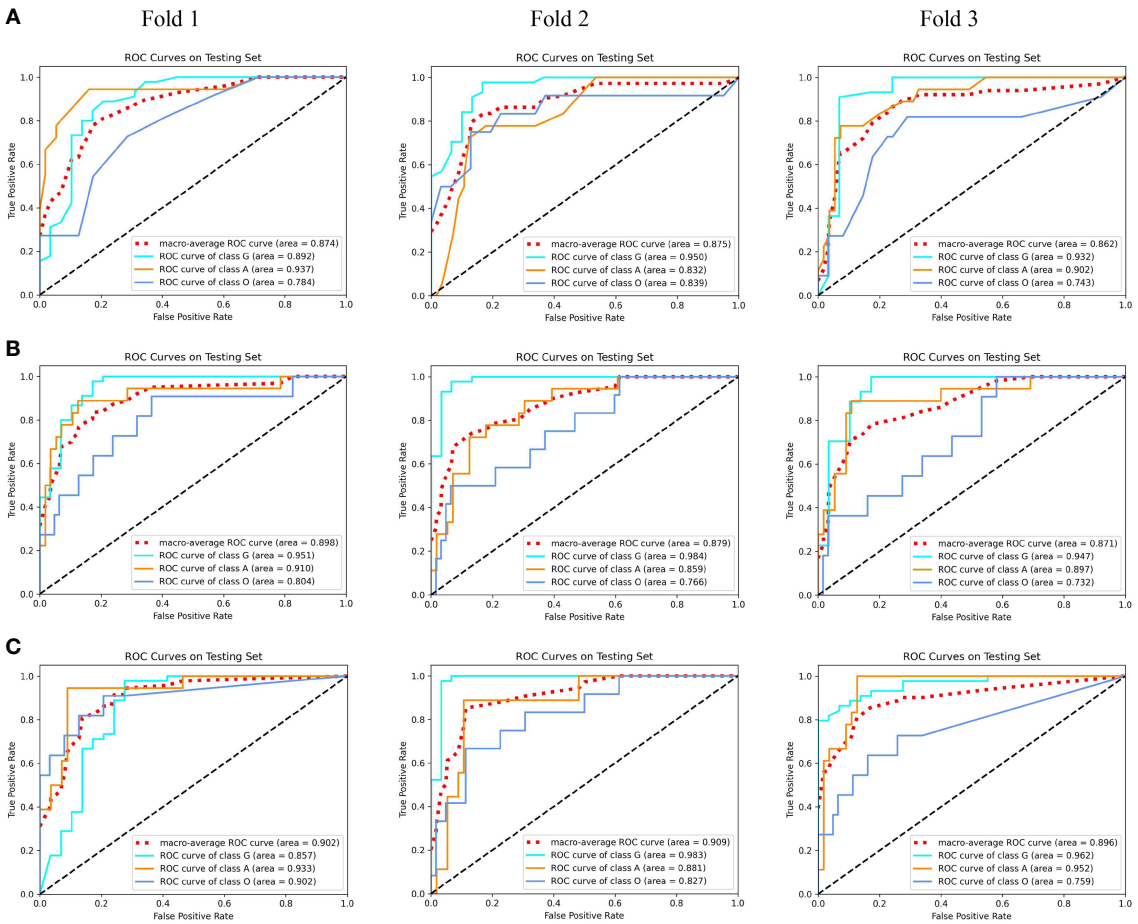


FIGURE 4 | The receiver operating characteristic (ROC) curves of multimodal prediction models on three validation folds in our study. **(A)** Radiomics model. **(B)** Data fusion model. **(C)** Decision fusion model.

TABLE 2 | Three-fold cross-validation performance of multimodal prediction models using radiomics, data fusion strategy, and our proposed MMIDFNet methods.

Models	Fold	ACC	AUC	SEN	SPE	PPV	NPV	Kappa
Radiomics	1	0.851	0.874	0.702	0.885	0.905	0.945	0.699
	2	0.824	0.875	0.705	0.897	0.793	0.922	0.672
	3	0.836	0.862	0.706	0.914	0.724	0.927	0.695
	Average	**0.837**	**0.870**	**0.704**	**0.899**	**0.807**	**0.931**	**0.689**
	95% CI	[0.815,0.859]	[0.859,0.882]	[0.701, 0.708]	[0.875,0.922]	[0.661,0.954]	[0.912,0.951]	[0.665,0.712]
Data fusion	1	0.865	0.898	0.732	0.913	0.890	0.943	0.740
	2	0.838	0.879	0.744	0.926	0.741	0.922	0.713
	3	0.836	0.871	0.717	0.908	0.745	0.921	0.695
	Average	**0.846**	**0.883**	**0.731**	**0.916**	**0.792**	**0.929**	**0.716**
	95% CI	[0.820,0.872]	[0.860,0.905]	[0.709, 0.753]	[0.901,0.931]	[0.656,0.928]	[0.909,0.949]	[0.680,0.752]
Decision fusion	1	0.892	0.902	0.781	0.919	0.924	0.959	0.789
	2	0.865	0.909	0.741	0.926	0.821	0.949	0.749
	3	0.877	0.896	0.795	0.946	0.842	0.939	0.780
	Average	**0.878**	**0.902**	**0.772**	**0.930**	**0.862**	**0.949**	**0.773**
	95% CI	[0.856,0.900]	[0.892,0.913]	[0.727, 0.817]	[0.908,0.953]	[0.775,0.949]	[0.933,0.965]	[0.739,0.806]

ACC, accuracy; SEN, sensitivity; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value; CI, confidence interval; Bold Value, average value of 3 folds.

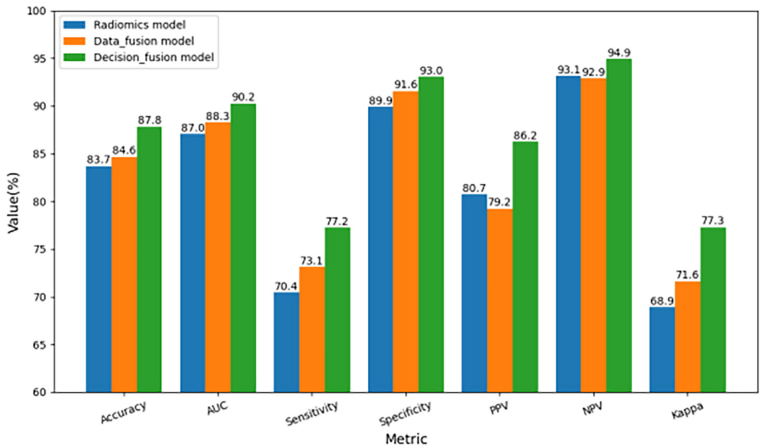


FIGURE 5 | Comparison of three-fold cross-validation performance of the three multimodal prediction models.

TABLE 3 | Performance comparison of other state-of-the-art studies with ours.

Metrics	Pei et al. (4)	Xue et al. (24)	Pei et al. (25)	Yin et al. (26)	Radiomics	Data fusion	Decision fusion
F1_score	0.829	0.771	0.771	0.857	0.837	0.846	**0.878**
95% CI					[0.815, 0.859]	[0.820, 0.872]	[0.856, 0.900]
Balanced_Acc	0.749	NA	0.698	0.820	0.704	0.731	**0.772**
95% CI					[0.701, 0.708]	[0.709, 0.753]	[0.727, 0.817]
Kappa	0.715	NA	0.627	0.767	0.689	0.716	**0.773**
95% CI					[0.665, 0.712]	[0.680, 0.752]	[0.739, 0.806]

CI, confidence interval; NA, not available; Bold Value, best value of the metric.

findings of recent studies (38, 39). Moreover, as for unimodal models, our proposed MMIDFNet method also generated overall better results than our radiomics method. As described above, the radiomics-based methods comprised three tightly coupled stages: feature extraction, feature selection, and classification model building. Any small variations in each of the stages may affect the final prediction accuracy (39). Furthermore, compared to radiomics-based methods, deep learning-based methods are more flexible and superior in feature extraction since the hierarchy of features can be learned automatically from low level to high level in a layer-by-layer manner in the training phase (40).

It should be noted that the performance variability of brain tumor classification based on deep learning methods depends on the designed network architecture and trained hyper-parameters (41). Through the comparisons, we found that our MMIDFNet method performed better than the other four recent state-of-the-art methods based on the deep learning method (4, 24–26). This is mainly due to the adopted tumor segmentation algorithm, classification network, and image fusion strategy.

The region of a tumor lesion may have different image contrast properties in different imaging modalities (42). In contrast to other MRI modality images, the tumor boundary in the T1ce sequence is more significantly different from normal tissue, which facilitates automatic tumor segmentation. Besides, the T1ce sequence can better provide the condition of intratumoral so as to distinguish tumors from non-neoplastic lesions. Note that in our experimental results, either with the radiomics method or with our MMIDFNet method, the classification performance using T1ce modality images was significantly better than that using the other three modalities. These results are consistent with previous observations (25) and indicate that T1ce modality images should not be neglected in studies of glioma classification using multimodal MRI.

As for the glioma classification with multimodal MRI, how to mine rich feature representations across multimodal MRI images is the key factor in improving classification performance. Recent studies showed that image fusion can be operated at three levels: data, feature, and decision (27). Actually, as for our three multimodal prediction models, the radiomics model adopted the strategy of feature-level fusion, the data fusion model used the strategy of data-level fusion, and the decision fusion model employed the strategy of decision-level fusion. Our comparison results showed that, whatever the level of fusion, the accuracy of our multimodal models outperforms any models trained using unimodal MRI images, which indicates that each MRI modality can provide complementary features. In our radiomics method, through concatenating the four unimodal MRI features together, the accuracy of the RF classification model was raised to 0.837. Besides, as for our MMIDFNet method, the accuracy of the data fusion model (0.846) is 3.2% lower than that of the decision

fusion model (0.878) *via* 3-fold cross-validation. This is mainly caused by the limitations of the data-level fusion strategy that does not fully take advantage of the features underlying each modality data and does not deal with how to fuse the features from the multimodal MRI images (33, 43, 44). However, in our decision fusion model, we used the weighting manner to ensemble the unimodal models in inferring stage. Theoretically, the fraction of each modality should be positively related to its contribution. From **Table 1**, we noticed that T1ce was the most useful modality for the prediction, followed by Flair, T2, and T1. In our fusion model, the weights for T1ce, Flair, T2, and T1 modalities are 2, 1, 0.7, and 0.3, respectively, which conform to the theoretical analysis and validate that our decision fusion model can fully explore the complementary information of different imaging modalities.

Brain tumor segmentation in MRI is of crucial importance for the subsequent diagnosis of brain tumors (45). Our efficient MMIDFNet method as well as radiomics method for glioma classification however relies on tumor segmentation performance. Any small variations in this stage may affect the final prediction performance and stability of the final prediction models. What cannot be ignored is that we employed a pre-trained tumor segmentation model from the BraTS challenge for tumor segmentation, while we did not use the ground truth delineated by experienced radiologists to segment the tumor regions from original MRI images. Although the segmentation result is not as accurate as the ground truth, we minimized the adverse effects caused by inaccurate segmentation by adopting the central clipping manner to segment out the whole tumor regions.

Although encouraging, our method has several limitations. First, as a retrospective study, the sample size of the dataset used in the present study was limited, which has an adverse effect on the robustness of our designed model. Therefore, the few-shot learning method may be a better choice to handle the problem. Second, we used only MRI modalities in the present study without considering other types of data, especially pathological whole-slide images. To further improve the performance of glioma subtype classification, in the future, we could try to combine MRI, pathology images, molecular genetic information, and other clinical data to conduct a multi-omics clinical study. Third, the number of subtype "G" in our training set was about 60.2%, and this resulted in the class imbalance issue. To handle this issue, we used the focal loss function through balancing the loss of different subtypes, which indeed alleviated the issue. However, this scheme might not be optimal because it ignores the differences in data distribution. Therefore, to deal with this common problem in medical images classification, more effective measures for forming more balanced data will be considered in our future work.

In conclusion, we studied the preoperative glioma subtype classification by developing a multimodal MRI image decision fusion-based network based on a deep learning technique. Through designing a linear weighted module to assemble the unimodal models trained with unimodal MRI images together, our multimodal prediction model fully mined the complementary information of multimodal MRI images. Extensive experimental results showed that the proposed MMIDFNet method was superior to recent state-of-the-art methods, which suggests its potential use in clinical practice for glioma subtype classification based on only MRI images.

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

SG, LHW, and YZ: conceptualization, methodology, software, visualization, and writing. QC and LW: collection and assembly of data. SG, JZ, and LW: data analysis and interpretation. All authors: writing and final approval of the manuscript.

FUNDING

This work was supported in part by the National Natural Science Foundation of China under Grant 62161004, in part by the Natural Science Foundation of Guizhou Province under Grant Qiankehe J No. [2020]1Y255, in part by the Program PHC-Cai Yuanpei 2018 under Grant 41400TC, in part by the Guizhou Province Education Department Project under Grant Qianjiaohe KY[2016]321, and in part by the Guizhou Science and Technology Plan Project under Grant Qiankehe ZK[2021]002.

REFERENCES

1. Jin W, Fatehi M, Abhishek K, Mallya M, Toyota B, Hamarneh G. Artificial Intelligence in Glioma Imaging: Challenges and Advances. *J Neural Eng* (2020) 17(2):021002. doi: 10.1088/1741-2552/ab8131
2. Louis DN, Perry A, Reifenberger G, Von Deimling A, 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
3. Abd El Kader I, Xu G, Shuai Z, Saminu S, Javaid I, Salim Ahmad I. Differential Deep Convolutional Neural Network Model for Brain Tumor Classification. *Brain Sci* (2021) 11(3):352. doi: 10.3390/brainsci11030352
4. Pei L, Vidyaratne L, Rahman MM, Iftekharuddin KM. Context Aware Deep Learning for Brain Tumor Segmentation, Subtype Classification, and Survival Prediction Using Radiology Images. *Sci Rep-Uk* (2020) 10(1):1–11. doi: 10.1038/s41598-020-74419-9
5. Bhatele KR, Bhadauria SS. Machine Learning Application in Glioma Classification: Review and Comparison Analysis. *Arch Comput Method E* (2022) 29:247–74. doi: 10.1007/s11831-021-09572-z
6. Ma X, Jia F. (2020) Brain Tumor Classification With Multimodal MR and Pathology Images, In: *Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer, International MICCAI Brainlesion Workshop. 11993. pp. 343–52. doi: 10.1007/978-3-030-46643-5_34
7. Ismael SAA, Mohammed A, Hefny H. An Enhanced Deep Learning Approach for Brain Cancer MRI Images Classification Using Residual

Networks. *Artif Intell Med* (2020) 102:101779. doi: 10.1016/j.artmed.2019.101779
8. Sarhan AM. Brain Tumor Classification in Magnetic Resonance Images Using Deep Learning and Wavelet Transform. *J Biomed Sci Eng* (2020) 13(06):102–12. doi: 10.4236/jbise.2020.136010
9. Muhammad K, Khan S, Del Ser J, de Albuquerque VHC. Deep Learning for Multigrade Brain Tumor Classification in Smart Healthcare Systems: A Prospective Survey. *IEEE T Neur Net Lear* (2020) 32(2):507–22. doi: 10.1109/TNNLS.2020.2995800
10. Shrot S, Salhov M, Dvorski N, Konen E, Averbuch A, Hoffmann C. Application of MR Morphologic, Diffusion Tensor, and Perfusion Imaging in the Classification of Brain Tumors Using Machine Learning Scheme. *Neuroradiology* (2019) 61(7):757–65. doi: 10.1007/s00234-019-02195-z
11. Gao M, Huang S, Pan X, Liao X, Yang R, Liu J. Machine Learning-Based Radiomics Predicting Tumor Grades and Expression of Multiple Pathologic Biomarkers in Gliomas. *Front Oncol* (2020) 10:1676. doi: 10.3389/fonc.2020.01676
12. Chan H-W, Weng Y-T, Huang T-Y. (2020). Automatic Classification of Brain Tumor Types With the MRI Scans and Histopathology Images, In: *Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. Cham: Springer, International MICCAI Brainlesion Workshop, Vol. 11993. pp. 353–9.
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(4):441–6. doi: 10.1016/j.ejca.2011.11.036
14. Kumar R, Gupta A, Arora HS, Pandian GN, Raman B. CGHF: A Computational Decision Support System for Glioma Classification Using Hybrid Radiomics-And Stationary Wavelet-Based Features. *IEEE Access* (2020) 8:79440–58. doi: 10.1109/ACCESS.2020.2989193
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. Wong KC, Syeda-Mahmood T, Moradi M. Building Medical Image Classifiers With Very Limited Data Using Segmentation Networks. *Med Image Anal* (2018) 49:105–16. doi: 10.1016/j.media.2018.07.010
17. Sharif MI, Li JP, Khan MA, Saleem MA. Active Deep Neural Network Features Selection for Segmentation and Recognition of Brain Tumors Using MRI Images. *Pattern Recogn Lett* (2020) 129:181–9. doi: 10.1016/j.patrec.2019.11.019
18. Han W, Qin L, Bay C, Chen X, Yu K-H, Miskin N, et al. Deep Transfer Learning and Radiomics Feature Prediction of Survival of Patients With High-Grade Gliomas. *Am J Neuroradiol* (2020) 41(1):40–8. doi: 10.3174/ajnr.A6365
19. Yogananda CGB, Shah BR, Yu FF, Pinho MC, Nalawade SS, Murugesan GK, et al. A Novel Fully Automated MRI-Based Deep-Learning Method for Classification of 1p/19q Co-Deletion Status in Brain Gliomas. *Neuro Oncol Adv* (2020) 2(Supplement_4):42–8. doi: 10.1093/neuonc/noz199
20. Matsui Y, Maruyama T, Nitta M, Saito T, Tsuzuki S, Tamura M, et al. Prediction of Lower-Grade Glioma Molecular Subtypes Using Deep Learning. *J Neuro Oncol* (2020) 146(2):321–7. doi: 10.1007/s11060-019-03376-9
21. Sharif MI, Khan MA, Alhussein M, Aurangzeb K, Raza M. A Decision Support System for Multimodal Brain Tumor Classification Using Deep Learning. *Complex Intell Syst* (2021), 1–14. doi: 10.1007/s40747-021-00321-0
22. Kang J, Ullah Z, Gwak J. MRI-Based Brain Tumor Classification Using Ensemble of Deep Features and Machine Learning Classifiers. *Sensors* (2021) 21(6):2222. doi: 10.3390/s21062222
23. Afshar P, Mohammadi A, Plataniotis KN. BayesCap: A Bayesian Approach to Brain Tumor Classification Using Capsule Networks. *IEEE Signal Proc Let* (2020) 27:2024–8. doi: 10.1109/LSP.2020.3034858
24. Xue Y, Yang Y, Farhat FG, Shih FY, Boukrina O, Barrett A, et al. (2020). Brain Tumor Classification With Tumor Segmentations and a Dual Path Residual Convolutional Neural Network From MRI and Pathology Images, In: *Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* Cham: Springer, International MICCAI brainlesion workshop, Vol. 11993. pp. 360–7. doi: 10.1007/978-3-030-46643-5_3
25. Pei L, Hsu W-W, Chiang L-A, Guo J-M, Iftekharuddin KM, Colen R. (2021). A Hybrid Convolutional Neural Network Based-Method for Brain Tumor Classification Using mMRI and WSI, In: *Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries* Cham: Springer, International MICCAI Brainlesion Workshop, Vol. 12659. pp. 487–96. doi: 10.1007/978-3-030-72087-2_43
26. Yin B, Cheng H, Wang F, Wang Z. (2021). Brain Tumor Classification Based on MRI Images and Noise Reduced Pathology Images, In: *Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Vol. 12659. pp. 465–74. Cham: Springer, International MICCAI Brainlesion Workshop. doi: 10.1007/978-3-030-72087-2_41
27. Ouerghi H, Mourali O, Zagrouba E. Glioma Classification *via* MR Images Radiomics Analysis. *Visual Comput* (2021), 1–15. doi: 10.1007/s00371-021-02077-7
28. Ma C, Luo G, Wang K. Concatenated and Connected Random Forests With Multiscale Patch Driven Active Contour Model for Automated Brain Tumor Segmentation of MR Images. *IEEE T Med Imaging* (2018) 37(8):1943–54. doi: 10.1109/TMI.2018.2805821
29. Conte GM, Weston AD, Vogelsang DC, Philbrick KA, Cai JC, Barbera M, et al. Generative Adversarial Networks to Synthesize Missing T1 and FLAIR MRI Sequences for Use in a Multisequence Brain Tumor Segmentation Model. *Radiology* (2021) 299(2):313–23. doi: 10.1148/radiol.2021203786
30. Mohan G, Subashini MM. MRI Based Medical Image Analysis: Survey on Brain Tumor Grade Classification. *BioMed Signal Proces* (2018) 39:139–61. doi: 10.1016/j.bspc.2017.07.007
31. Lachinov D, Shipunova E, Turlapov V. (2020). Knowledge Distillation for Brain Tumor Segmentation, In: *Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Vol. 11993. pp. 324–32. Cham: Springer, International MICCAI Brainlesion Workshop. doi: 10.1007/978-3-030-46643-5_32
32. Huang G, Liu Z, van der Maaten L, Weinberger KQ. (2017). Densely Connected Convolutional Networks. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI: IEEE (2017) pp. 4700–8. doi: 10.1109/CVPR.2017.243
33. Zhou T, Ruan S, Canu S. A Review: Deep Learning for Medical Image Segmentation Using Multi-Modality Fusion. *Array* (2019) 3:100004. doi: 10.1016/j.array.2019.100004
34. Le NQK, Hung TNK, Do DT, Lam LHT, Dang LH, Huynh T-T. Radiomics-Based Machine Learning Model for Efficiently Classifying Transcriptome Subtypes in Glioblastoma Patients From MRI. *Comput Biol Med* (2021) 132:104320. doi: 10.1016/j.compbiomed.2021.104320
35. Cai J, Luo J, Wang S, Yang S. Feature Selection in Machine Learning: A New Perspective. *Neurocomputing* (2018) 300:70–9. doi: 10.1016/j.neucom.2017.11.077
36. Ali F, El-Sappagh S, Islam SR, Kwak D, Ali A, Imran M, et al. A Smart Healthcare Monitoring System for Heart Disease Prediction Based on Ensemble Deep Learning and Feature Fusion. *Inform Fusion* (2020) 63:208–22. doi: 10.1016/j.inffus.2020.06.008
37. Wu S, Meng J, Yu Q, Li P, Fu S. Radiomics-Based Machine Learning Methods for Isocitrate Dehydrogenase Genotype Prediction of Diffuse Gliomas. *J Cancer Res Clin* (2019) 145(3):543–50. doi: 10.1007/s00432-018-2787-1
38. 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
39. Li Q, Bai H, Chen Y, Sun Q, Liu L, Zhou S, et al. A Fully-Automatic Multiparametric Radiomics Model: Towards Reproducible and Prognostic Imaging Signature for Prediction of Overall Survival in Glioblastoma Multiforme. *Sci Rep-Uk* (2017) 7(1):1–9. doi: 10.1038/s41598-017-14753-7
40. Gutta S, Acharya J, Shiroishi M, Hwang D, Nayak K. Improved Glioma Grading Using Deep Convolutional Neural Networks. *Am J Neuroradiol* (2021) 42(2):233–9. doi: 10.3174/ajnr.A6882
41. Ayadi W, Elhamzi W, Charfi I, Atri M. Deep CNN for Brain Tumor Classification. *Neural Process Lett* (2021) 53(1):671–700. doi: 10.1007/s11063-020-10398-2
42. Chyzhyk D, Dacosta-Aguayo R, Mataró M, Graña M. An Active Learning Approach for Stroke Lesion Segmentation on Multimodal MRI Data. *Neurocomputing* (2015) 150:26–36. doi: 10.1016/j.neucom.2014.01.077
43. Zhang D, Huang G, Zhang Q, Han J, Han J, Yu Y. Cross-Modality Deep Feature Learning for Brain Tumor Segmentation. *Pattern Recogn* (2021) 110:107562. doi: 10.1016/j.patcog.2020.107562
44. Hermessi H, Mourali O, Zagrouba E. Multimodal Medical Image Fusion Review: Theoretical Background and Recent Advances. *Signal Process* (2021) 183:108036. doi: 10.1016/j.sigpro.2021.108036

45. Chen S, Ding C, Liu M. Dual-Force Convolutional Neural Networks for Accurate Brain Tumor Segmentation. *Pattern Recogn* (2019) 88:90–100. doi: 10.1016/j.patcog.2018.11.009

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 02 March 2022
doi: 10.3389/fonc.2022.815952

Integration of MRI-Based Radiomics Features, Clinicopathological Characteristics, and Blood Parameters: A Nomogram Model for Predicting Clinical Outcome in Nasopharyngeal Carcinoma

*Zeng-Yi Fang[1,2,3], Ke-Zhen Li[1,2], Man Yang[1,4], Yu-Rou Che[1,4], Li-Ping Luo[1,3,4], Zi-Fei Wu[1,4], Ming-Quan Gao[1,4], Chuan Wu[1,4], Cheng Luo[1], Xin Lai[1], Yi-Yao Zhang[1,4], Mei Wang[1,4], Zhu Xu[1,2], Si-Ming Li[1,4], Jie-Ke Liu[1,4], Peng Zhou[1,4] and Wei-Dong Wang[1,2,3,4]**

[1] Department of Radiation Oncology, Sichuan Cancer Hospital and Institute, Chengdu, China, [2] Department of Oncology, School of Clinical Medicine, Southwest Medical University, Luzhou, China, [3] Radiation Oncology, Key Laboratory of Sichuan Province, Chengdu, China, [4] School of Medicine, University of Electronic Science and Technology of China, Chengdu, China

OPEN ACCESS

Edited by:
Antonio Napolitano, Bambino Gesù Children's Hospital (IRCCS), Italy

Reviewed by:
Renjie Wang, Fudan University, China
Catriona M. Douglas, NHS Greater Glasgow and Clyde, United Kingdom

***Correspondence:**
Wei-Dong Wang
wwdwyl@sina.com

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

Received: 16 November 2021
Accepted: 08 February 2022
Published: 02 March 2022

Citation:
Fang Z-Y, Li K-Z, Yang M, Che Y-R, Luo L-P, Wu Z-F, Gao M-Q, Wu C, Luo C, Lai X, Zhang Y-Y, Wang M, Xu Z, Li S-M, Liu J-K, Zhou P and Wang W-D (2022) Integration of MRI-Based Radiomics Features, Clinicopathological Characteristics, and Blood Parameters: A Nomogram Model for Predicting Clinical Outcome in Nasopharyngeal Carcinoma. Front. Oncol. 12:815952. doi: 10.3389/fonc.2022.815952

Purpose: This study aimed to develop a nomogram model based on multiparametric magnetic resonance imaging (MRI) radiomics features, clinicopathological characteristics, and blood parameters to predict the progression-free survival (PFS) of patients with nasopharyngeal carcinoma (NPC).

Methods: A total of 462 patients with pathologically confirmed nonkeratinizing NPC treated at Sichuan Cancer Hospital were recruited from 2015 to 2019 and divided into training and validation cohorts at a ratio of 7:3. The least absolute shrinkage and selection operator (LASSO) algorithm was used for radiomics feature dimension reduction and screening in the training cohort. Rad-score, age, sex, smoking and drinking habits, Ki-67, monocytes, monocyte ratio, and mean corpuscular volume were incorporated into a multivariate Cox proportional risk regression model to build a multifactorial nomogram. The concordance index (C-index) and decision curve analysis (DCA) were applied to estimate its efficacy.

Results: Nine significant features associated with PFS were selected by LASSO and used to calculate the rad-score of each patient. The rad-score was verified as an independent prognostic factor for PFS in NPC. The survival analysis showed that those with lower rad-scores had longer PFS in both cohorts ($p < 0.05$). Compared with the tumor–node–metastasis staging system, the multifactorial nomogram had higher C-indexes (training cohorts: 0.819 vs. 0.610; validation cohorts: 0.820 vs. 0.602). Moreover, the DCA curve showed that this model could better predict progression within 50% threshold probability.

Conclusion: A nomogram that combined MRI-based radiomics with clinicopathological characteristics and blood parameters improved the ability to predict progression in patients with NPC.

Keywords: radiomics, progression-free survival, nasopharyngeal carcinoma, Ki-67, blood parameters

1 INTRODUCTION

Nasopharyngeal carcinoma (NPC) is a malignant tumor in the mucous membrane of the nasopharynx. The incidence and mortality of NPC vary in regional distribution, especially in Southeast Asia (1–3). Although intensity-modulated radiotherapy (IMRT) significantly improved the prognosis of NPC, some patients still experience progression (4, 5). At present, the risk assessment of NPC is mainly determined by the tumor–node–metastasis (TNM) staging system, which only has 61% accuracy for predicting the local recurrence of NPC (6). While it incorporates local tumor invasion, positive lymph nodes, and distant metastases, TNM cannot explain the temporal and spatial heterogeneity or changes in the internal and external environments of tumor cells. Plasma Epstein–Barr virus (EBV) DNA, which may affect the growth and apoptosis of the NPC cell line, has been used as an independent prognostic marker in endemic areas, but the detection rate of EBV is low in nonendemic areas (7, 8). Therefore, it is urgent to identify more representative and comprehensive biomarkers to predict NPC prognosis.

Many studies reported that a large number of clinical biomarkers such as monocytes (MONO), mean corpuscular volume (MCV), and Ki-67 expression are associated with the tumor microenvironment and tumor immune escape (9–11). There are no regional differences in the expression of these markers. Beyond these biomarkers, the emerging field of radiomics is supposed to be a bridge between medical imaging and clinical medicine (12). Radiomics features are used for tumor diagnosis, phenotype, and prognosis (13–15). By extracting innumerable quantitative imaging features, the differences in tumor heterogeneity and microenvironment may be explained. Some recent studies showed that magnetic resonance imaging (MRI) radiomics were significantly associated with NPC prognosis (16–18). However, no publications integrated blood parameters, Ki-67, and MRI radiomics to predict progression-free survival (PFS) in patients with NPC.

We built and validated a nomogram prediction model based on MRI, clinicopathological parameters, and blood parameters to visually demonstrate the PFS of NPC and guide clinical diagnosis and treatment.

Abbreviations: MRI, multiparametric magnetic resonance imaging; PFS, progression-free survival; NPC, nasopharyngeal carcinoma; LASSO, least absolute shrinkage and selection operator algorithm; MONO, monocytes; MONO%, monocyte ratio; MCV, mean corpuscular volume; C-index, concordance index; DCA, decision curve analysis; IMRT, intensity-modulated radiotherapy; TNM, tumor–node–metastasis staging system; EBV, Epstein–Barr virus; CCRT, concurrent chemoradiotherapy; IC, induction chemotherapy; AC, adjuvant chemotherapy; GLCM, gray-level co-occurrence matrix; GLRLM, gray-level run length matrix; GLSZM, gray-level size zone matrix; GLDM, gray-level dependence matrix; NGTDM, neighborhood gray tone difference matrix; ROI, region of interest; ROC, receiver operating characteristic curve

2 MATERIALS AND METHODS

2.1 Patients

Data from patients treated in Sichuan Cancer Hospital from January 2015 to December 2019 were reviewed. The inclusion and exclusion criteria are presented in the **Supplementary Data**. The study workflow is displayed in **Figure 1**. A total of 462 patients were included and randomly divided into a training cohort (n = 323) and validation cohort (n = 139) at a 7:3 ratio. The method and criteria of Ki-67 scoring are detailed in the **Supplementary Data**. Clinical data (age, gender, smoking and drinking habits, TNM, plasma EBV DNA) and blood parameters were collected. All patients were restaged according to the 8th Edition American Joint Committee on Cancer TNM Staging System (19).

2.2 Treatment

2.2.1 Radiotherapy

All patients underwent IMRT. Delineation of the target area and organs at risk were based on ICRU reports 50 and 62. The prescribed doses for the target area were GTVnx 66–76 Gy, GTVnd 66–70 Gy, CTV1 60–62 Gy, CTV2 50–56 Gy, and CTVnd 50–56 Gy (28–33 fractions).

2.2.2 Chemotherapy

Patients with stage II (n = 23) underwent concurrent chemoradiotherapy (CCRT). Those with stages III–IV (n = 439) were treated with two cycles of induction chemotherapy (IC) followed by CCRT. The IC drugs were cisplatin (75 mg/m^2, d1–3) plus paclitaxel (135 mg/m^2, d1) every 3 weeks for two cycles. The CCRT drug was cisplatin (75 mg/m^2, d1–3) given every 3 weeks.

2.3 Follow-Up

After patients completed all treatments, they were followed-up every 3 months in the first 2 years, every 6 months in years 3–5, and annually thereafter. The review items included blood parameters, nasopharyngeal MRI, chest computed tomography, abdominal ultrasonography, or isotope bone scanning, and each review item was determined according to the specific situation of the patient. PFS was set as the primary endpoint.

2.4 MRI Acquisition and Image Preprocessing

The pretreatment MRI parameters are listed in the **Supplementary Data**. To avoid inhomogeneity due to different MRI devices, two image preprocessing steps were applied. First, we used the N4ITK algorithm to remove bias field artifacts (20). Second, the intensity range was adjusted from 0 to 255. In addition to the original images,

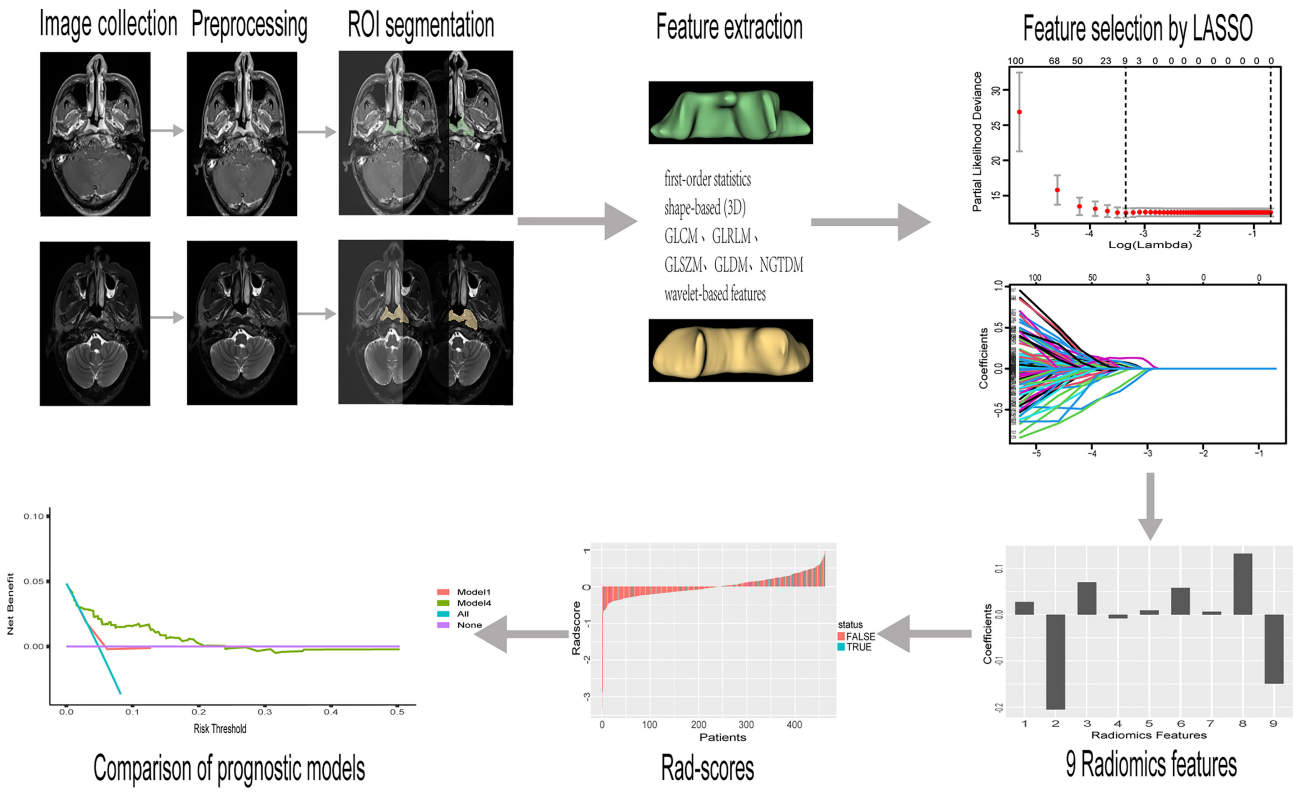


FIGURE 1 | The workflow of MRI-based radiomic analysis. After manual tumor segementation, 2074 features of each patients were extracted. Radiomics features selection by the LASSO algorithm. These selected features were linearly fitted according to the weights of the coefficients to calculate the rad-score. Decision curve analysis (DCA) compared the net benefit rate between the TNM stage system (Model 1) with our nomogram model (Model 4).

the Gaussian Laplace filter with sigma values of 4 and 5 mm was used to reconstruct the images, and the features of the multiscale resolution were extracted (21, 22). Preprocessing was performed in the SimpleITK 2.0.2, which is an open-source platform for Python 3.8.5 (www.python.org).

2.5 Image Segmentation

We used 3D Slicer 4.11 software (open source and multiplatform software; www.slicer.org) for manual segmentation (23). A radiologist with 20 years of experience delineated the region of interest (ROI), which refers to the margin of the nasopharyngeal tumor at each level on axial CET1-w and T2-w images.

2.6 Extraction of Radiomics Features

A total of 1,037 radiomics features were obtained by SlicerRadiomics (an extension for 3D Slicer 4.11 that encapsulates pyradiomics library) from axial CET1-w and T2-w images, respectively. Features of different categories were considered: first-order statistics, shape-based (3D), gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray-level size zone matrix (GLSZM), gray-level dependence matrix (GLDM), neighborhood gray tone difference matrix (NGTDM), and wavelet-based features.

2.7 Postprocessing of Radiomics Features and Building of Radiomics Signature

To ensure the comparability of different features, Z-score normalization was performed to unify data from different levels into the same level. Feature selection was conducted in the training cohort (n = 323). We used the least absolute shrinkage and selection operator (LASSO) algorithm for feature dimension reduction and screening. LASSO attempts to shrink some coefficients of the models and sets others to zero, but it may lead to overfitting, so we added a 10-fold cross-validation. Nine noteworthy features were selected. These features were linearly fitted according to the weights of their coefficients; for each patient, the rad-score was calculated. The rad-score was then used to build the radiomics signature.

2.8 Radiomics Survival Model Development and Validation

To find the rad-score cutoff with the best sensitivity and specificity, we generated a receiver operating characteristic curve (ROC) using data from the training cohort. To explore the potential association between radiomics features and PFS, we separated patients in both cohorts into high- and low-risk groups based on the cutoff value of rad-scores (patients below this cutoff

value were considered low risk). Kaplan–Meier survival analysis was used to identify PFS differences in both cohorts.

2.9 Evaluation and Comparison of the Multifactorial Prognostic Nomogram Model

Four models were set up to compare the prognostic efficacy (model 1: clinical stage; model 2: radiomics; model 3: clinical stage + rad-score; model 4: clinical data + rad-score). The concordance index (C-index) was used to evaluate univariate or multivariate Cox models. A nomogram was built to visualize the results of the best prediction model in the training cohort using the R software (version 4.1.0). We evaluated the uniformity of the nomogram by plotting 3- and 5-year calibration curves. Decision curve analysis (DCA) was performed to compare the net benefit rate between the TNM stage system and this nomogram for predicting prognosis.

2.10 Statistical Analysis

Statistical analyses were performed with the R software (version 4.1.0; www.r-project.org), SPSS (SPSS version 20.0, IBM Corp, Armonk, NY, USA), and Python 3.8.5. Clinical data were compared between the training and validation cohorts with Independent samples *t*-tests, Mann–Whitney *U* tests, or Chi-square tests. Missing data was processed using the "miceforest" package from Python. Several R packages were employed: LASSO in the "glmnet" package was used to select radiomics features. Kaplan–Meier survival, Cox proportional hazard regression, and C-index were calculated by the "survival" and "rms" packages. DCA was performed with the "ggDCA" package. The "pROC" and "ggplot2" packages were applied to generate the ROC curve and rad-score histogram, respectively. For all statistical tests, differences were considered significant at $p < 0.05$.

3 RESULTS

3.1 Clinical Parameters

This retrospective study included 462 patients with pathologically confirmed nonkeratinizing NPC who were treated at Sichuan Cancer Hospital between January 2015 and December 2019. The clinical parameters of all patients in the training and validation cohorts are listed in **Table 1**. The median age was 49 years (range: 12–82 years), with 329 men and 133 women. The numbers of patients with each clinical stage were 0, 23, 193, 226, and 20 for stages I, II, III, IVA, and IVB, respectively. The Ki-67 cutoff value from the ROC curve was 37.5% (range: 3%–90%). The cutoff value for classifying EBV infection status was 400 copies/ml (negative: <400 copies/ml; positive: ≥400 copies/ml). A total of 330 patients who met the inclusion criteria underwent plasma EBV DNA tests before treatment, and 112 were positive. Among them, there were 2 cases of stage II, 24 cases of stage III, and 86 cases of stage IV. The interpolation of EBV DNA missing data was performed using the multiple substitutions in chained equations (MICE) method of random forest. The **Supplementary Data** detail the results after interpolating EBV DNA. The median PFS was 33.15 months (0.6–76.2 months) for all patients; 45 patients progressed, including 23 deaths, 14 distant metastases, and 8 recurrences.

3.2 Blood Parameters

All blood parameters in the training and validation cohorts are shown in **Table 2**. The cutoff values identified with ROC curves are shown in the **Supplementary Data**, as are the values of the areas under curve (AUCs) for blood parameters. The highest AUC values were found for MONO, MONO%, and MCV, which were 0.637, 0.626, and 0.568, respectively. These were incorporated into model 4.

3.3 Radiomics Signature Development

In total, 2,074 features were obtained from each ROI. The final nine key features were selected by LASSO ((1) CET1-w_Log-sigma-5-0-mm_glrlm_HighGrayLevelRunE-mphasis; (2) CET1-w_wavelet-LLH_glcm_ClusterShade; (3) CET1-w_wavelet-LLH_gl-szm_GrayLevelNonUniformity; (4) CET1-w_wavelet-HHL_glcm_Correlation; (5) CE-T1-w_wavelet-HHH_firstorder_Mean; (6) CET1-w_wavelet-HHH_gldm_LargeDepen-denceHighGray LevelEmphasis; (7) T2-w_log-sigma-4-0-mm_firstorder_Maximum; (8) T2-w_wavelet-HHL_firstorder_Maximum; (9) T2-w_wavelet-HHL_glcm_InverseVar-iance). The rad-score was calculated for each patient according to the weights of their coefficients. The formula for calculating the rad-score is detailed in the **Supplementary Data**. The features selected by LASSO and the histogram of every patient's rad-score are shown in **Figures 2A–D**.

3.4 Model Predictions and Comparison

The C-indexes of the four models are listed in **Table 3**. The C-index of model 2 was significantly higher than that of model 1 in both cohorts, which suggested that the predictive effect of radiomics may surpass that of the TNM stage system. Moreover, when comparing models 1 and 3, we found that model 3 that included the rad-score could remarkably predict the prognostic potency of the clinical stage. Model 4 integrating clinical data and radiomics had the best probability that the predicted results were consistent with the observed results (C-index of training and validation: 0.823 (95% CI: 0.745–0.901) vs. 0.812 (95% CI: 0.693–0.930)). The nomogram of model 4 is also shown in **Figure 3A**. Notably, the calibration curves of 3–5 years were very close to the diagonal line (**Figures 3B, C**). The DCA results for models 4 and 1 are presented in **Figure 3D**, confirming the remarkable effectiveness of model 4.

3.5 Kaplan–Meier Survival Analysis

Kaplan–Meier survival curves were drawn based on rad-scores. The cutoff value from the ROC curve was −0.021. A rad-score below this cutoff was considered low risk. In both cohorts, the low-risk group had significantly longer PFS ($p < 0.05$) (**Figure 4**).

4 DISCUSSION

We designed this study to build and validate multimodal information from MRI-based radiomics as an effective way to

TABLE 1 | Clinical parameters of patients in the training and validation cohorts.

	Training cohort (*n* = 323)	Validation cohort (*n* = 139)	*p*-value
Gender			0.652
Male	228 (70.6%)	101 (72.7%)	
Female	95 (29.4%)	38 (27.3%)	
Age (years)			0.949
≥49	167 (51.7%)	70 (50.4%)	
<49	156 (48.3%)	69 (49.6%)	
Overall stage			0.000
I	0	0	
II	18 (5.6%)	5 (3.6%)	
III	132 (40.9%)	61 (43.9%)	
IVA	158 (48.9%)	68 (48.9%)	
IVB	15 (4.6%)	5 (3.6%)	
T stage			0.000
T1	20 (6.2%)	9 (6.5%)	
T2	77 (23.8%)	36 (25.9%)	
T3	118 (36.5%)	49 (35.3%)	
T4	108 (33.5%)	45 (32.3%)	
N stage			0.000
N0	5 (1.5%)	4 (2.9%)	
N1	42 (13.0%)	16 (11.5%)	
N2	191 (59.1%)	78 (56.1%)	
N3	85 (26.4%)	41 (29.5%)	
M stage			0.000
M0	308 (95.4%)	134 (96.4%)	
M1	15 (4.6%)	5 (3.6%)	
Smoking			0.036
No	201 (62.2%)	72 (51.8%)	
Yes	122 (37.8%)	67 (48.2%)	
Drinking			0.111
No	246 (76.2%)	96 (69.1%)	
Yes	77 (23.8%)	43 (30.9%)	
Ki-67 (%)			0.680
≥37.5	238 (73.7%)	98 (70.5%)	
<37.5	85 (26.3%)	41 (29.5%)	
EBV			0.664
Positive	76 (23.5%)	36 (25.9%)	
Negative	153 (47.4%)	65 (46.8%)	
None	94 (29.1%)	38 (27.3%)	

Statistical comparisons between the training and validation cohorts were performed with Independent samples t-tests, Mann–Whitney U tests, or Chi-square tests. p-values <0.05 were considered statistically significant.
EBV, Epstein–Barr virus.

estimate PFS in patients with NPC. Our findings suggested that the multidimensional nomogram combining clinicopathological characteristics, blood parameters, and rad-score was superior to the prediction performance of the TNM staging system. Moreover, using the cutoff value of the rad-score, patients could be distinguished into high- and low-risk groups, and the latter had longer PFS.

In recent years, a growing number of studies have reported that MRI radiomics features can better reflect prognostic information for NPC because they may explain the inherent temporal or spatial heterogeneity of tumors on imaging (24–26). Kim et al. studied CET1-w and T2-w MRI images of 81 patients with NPC and conclude that MR-based radiomics features showed better performance than the TNM staging system and clinical variables. Their model combined radiomics features with TNM stage and clinical variables to provide the highest AUC values (27). Shen et al. found that a model that incorporated radiomics, clinical stage, and EBV DNA status from 327 nonmetastatic NPC, yielded a high C-index in two cohorts [0.805 (95% CI: 0.768–0.841) vs. 0.874 (95% CI: 0.861–0.877)] (28). In our study, the model that integrated clinical data with radiomics features also performed best; the C-index values of model 4 were 0.823 (95% CI: 0.745–0.901) in the training cohort and 0.812 (95% CI: 0.693–0.930) in the validation cohort. The C-index of our validation cohort was lower than that reported in the study by Shen. A possible explanation may be that we included metastatic NPC patients and had a larger sample size, which may have improved the generalizability of the prediction model. Compared with other research, the parameters included in our model are more universal, without regional differences, so the model has a higher degree of applicability. Based on our nomogram, the probabilities of 3- and 5-year PFS of a given patient can be visually and easily estimated by using the corresponding parameters measured before treatment. If patients with short PFS are identified as early as possible, clinicians can enhance treatment without increasing side effects

TABLE 2 | Blood parameters in the training and validation cohorts.

	Training cohort (*n* = 323)	Validation cohort (*n* = 139)	*p*-value
WBC (10^9/L)			0.439
≥6.695	111 (34.4%)	45 (32.4%)	
<6.695	212 (65.6%)	94 (67.6%)	
GR (10^9/L)			0.833
≥3.105	230 (71.2%)	100 (71.9%)	
<3.105	93 (28.8%)	39 (28.1%)	
LYMPH (10^9/L)			0.636
≥1.960	77 (23.8%)	31 (22.3%)	
<1.960	246 (76.2%)	108 (77.7%)	
MONO (10^9/L)			0.643
≥0.385	150 (46.4%)	59 (42.4%)	
<0.385	173 (53.6%)	80 (57.6%)	
EO (10^9/L)			0.841
≥0.175	95 (29.4%)	38 (27.3%)	
<0.175	228 (70.6%)	101 (72.7%)	
BASO (10^9/L)			0.877
≥0.035	76 (23.5%)	35 (25.2%)	
<0.035	247 (76.5%)	104 (74.8%)	
GR%			0.868
≥69.150	82 (25.4%)	47 (33.8%)	
<69.150	241 (74.6%)	92 (66.2%)	
LYMPH%			0.944
≥29.850	97 (30%)	46 (33.1%)	
<29.850	226 (70%)	93 (66.9%)	
MONO%			0.595
≥5.950	168 (52%)	75 (54%)	
<5.950	155 (48%)	64 (46%)	
EO%			0.856
≥2.050	156 (48.3%)	64 (46%)	
<2.050	167 (51.7%)	75 (54%)	
BASO%			0.856
≥0.350	224 (69.3%)	89 (64%)	
<0.350	99 (30.7%)	50 (36%)	
RBC (10^{12}/L)			0.262
≥4.685	132 (40.9%)	62 (44.6%)	
<4.685	191 (59.1%)	77 (55.4%)	
HGB (g/L)			0.616
≥125	266 (82.4%)	112 (80.6%)	
<125	57 (17.6%)	27 (19.4%)	
HCT			0.899
≥44.050	114 (35.3%)	50 (36%)	
<44.050	209 (64.7%)	89 (64%)	
MCV (fl)			0.057
≥95.650	106 (32.8%)	39 (28.1%)	
<95.650	217 (67.2%)	100 (71.9%)	
MCH (pg)			0.424
≥32.750	34 (10.5%)	10 (7.2%)	
<32.750	289 (89.5%)	129 (92.8%)	
MCHC (g/L)			0.132
≥337.500	40 (12.4%)	23 (16.5%)	
<337.500	283 (87.6%)	116 (83.5%)	
RDW_CV			0.454
≥12.950	210 (65%)	94 (67.6%)	
<12.950	113 (35%)	45 (32.4%)	
RDW_SD (fl)			0.524
≥42.850	176 (54.5%)	71 (51.1%)	
<42.850	147 (45.5%)	68 (48.9%)	
PLT (10^9/L)			0.698
≥191	182 (56.3%)	82 (59%)	
<191	141 (43.7%)	57 (41%)	
MPV (fl)			0.099
≥12.250	98 (30.3%)	28 (20.1%)	
<12.250	225 (69.7%)	111 (79.9%)	

(Continued)

TABLE 2 | Continued

	Training cohort (*n* = 323)	Validation cohort (*n* = 139)	*p*-value
PDW			0.615
≥16.050	254 (78.6%)	106 (76.3%)	
<16.050	69 (21.4%)	33 (23.7%)	
PCT			0.807
≥0.245	110 (34.1%)	47 (33.8%)	
<0.245	213 (65.9%)	92 (66.2%)	
NLR			0.991
≥2.026	232 (71.8%)	94 (67.6%)	
<2.026	91 (28.2%)	45 (32.4%)	
PLR			0.990
≥132.020	150 (46.4%)	59 (42.4%)	
<132.020	173 (53.6%)	80 (57.6%)	
LMR			0.601
≥4.822	113 (35%)	50 (36%)	
<4.822	210 (65%)	89 (64%)	

Statistical comparisons between the training and validation cohorts were performed with Independent samples t-tests, Mann–Whitney U tests, or Chi-square tests. p-values of <0.05 were considered statistically significant.

BASO, basophils; BASO%, ratio of basophils; EO, eosinophils; EO%, ratio of eosinophils; GR, neutrophilic granulocytes; GR%, ratio of neutrophilic granulocytes; HCT, hematocrit; HGB, hemoglobin; LMR, lymphocyte-to-monocyte ratio; LYMPH, lymphocytes; LYMPH%, ratio of lymphocytes; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; MCV, mean corpuscular volume; MONO, monocytes; MONO%, ratio of monocytes; MPV, mean platelet volume; NLR, neutrophil-to-lymphocyte ratio; PCT, plateletcrit; PDW, platelet distribution width; PLR, platelet to lymphocyte ratio; PLT, platelets; RBC, red blood cells; RDW-CV, variation of RBC distribution width; RDW-SD, standard deviation of RBC distribution width; WBC, white blood cells.

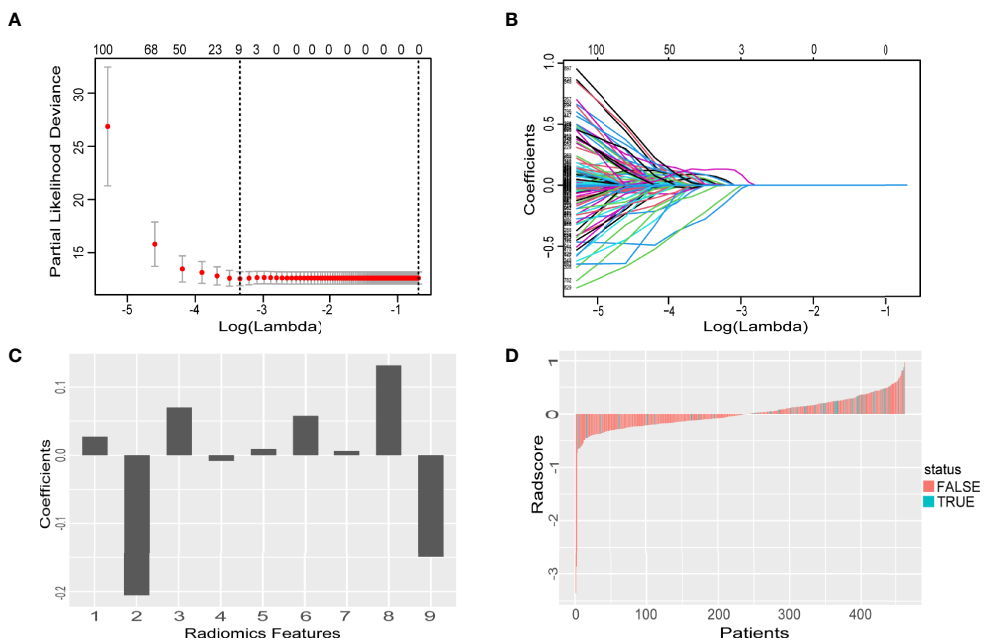


FIGURE 2 | Radiomics feature selection using the LASSO algorithm. **(A)** Used the 1O-fold cross validation to identify the optimal penalization coefficient lambda the minimum was 0.000577, with log (λ) = -3.238. **(B)** The model coefficient trendlines of radiomics features. **(C)** The histogram of coefficients with 9 features. **(D)** Rad-score for each patient. Red bars show scores for patients who survived without progression, while blue bars show scores for patients who happened progression, metastasis or died.

(e.g., by combining targeted treatment or immunotherapy), pay attention to adverse prognostic factors, and ensure an adequate follow-up period to reduce the risk of disease progression. Conversely, for patients with a high probability of 3- and 5-year PFS predicted by the model, it may be possible to reduce the drug dose and mitigate side effects.

However, several key aspects need to be considered when developing a clinical radiomics predictive model. Firstly, since plasma EBV DNA is used as an independent prognostic marker in endemic areas, many studies have incorporated it in nomogram construction. The MICE algorithm has the benefit of fast and efficient memory, which makes the results reliable

TABLE 3 | C-indexes of the four models.

Models	Training cohort (*n* = 323)	Validation cohort (*n* = 139)
1 Clinical stage	0.610 (95% CI: 0.507–0.714)	0.602 (95% CI: 0.474–0.729)
2 Radiomics	0.814 (95% CI: 0.746–0.882)	0.728 (95% CI: 0.618–0.838)
3 Clinical stage + rad-score	0.708 (95% CI: 0.602–0.814)	0.681 (95% CI: 0.562–0.801)
4 Clinical data + rad-score	0.823 (95% CI: 0.745–0.901)	0.812 (95% CI: 0.693–0.930)

Clinical data included gender, age, Ki-67, smoking and drinking habits, clinical stage, MONO, MONO%, MCV, and EBV DNA.
CI, confidence interval.

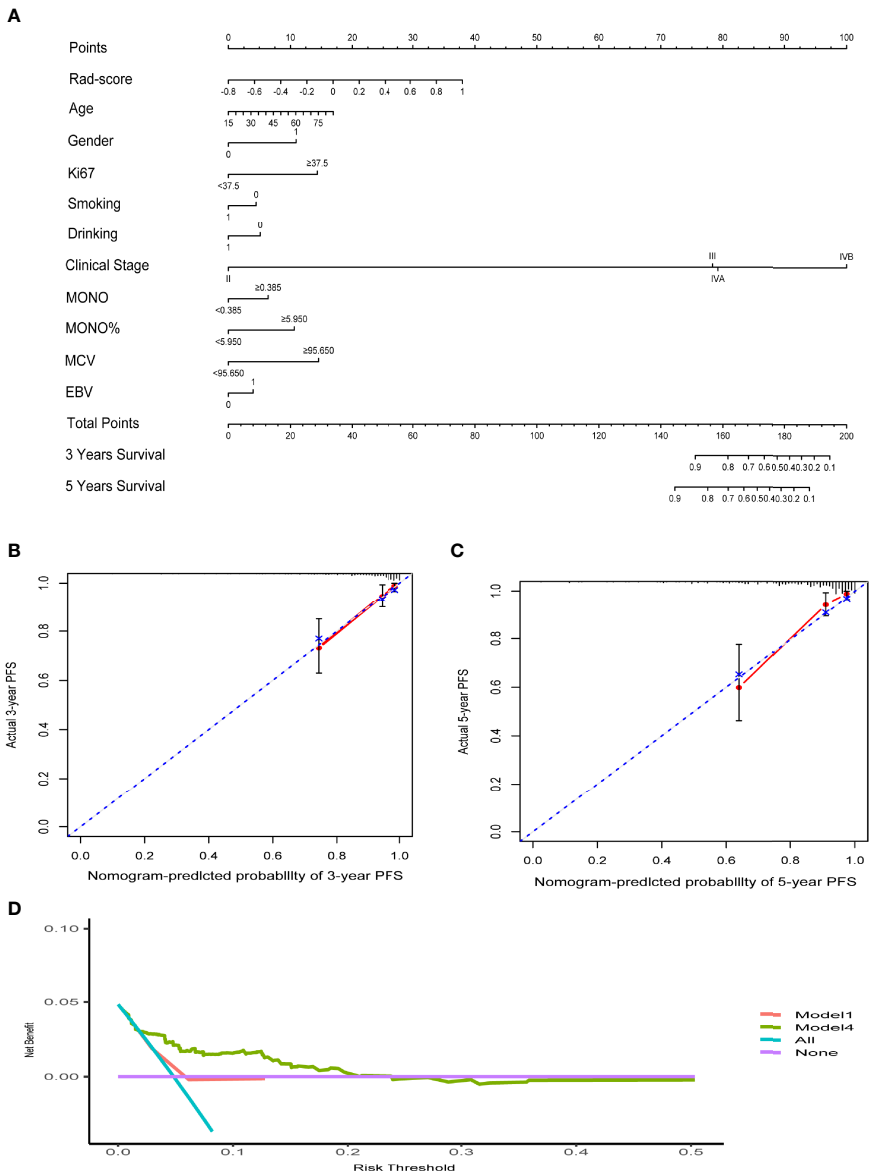


FIGURE 3 | **(A)** The nomogram of clinical data and rad-score. **(B, C)** The calibration curves of the nomogram. **(D)** Decision curve analysis for Model4 and Model1. The y-axis measures the net benefit. The red line represents Model 1 (clinical stage). The green line represents Model 4 (clinical data and rad-score). The blue line assumes that all patients progress. The purple line indicates that no progression is assumed in all patients.

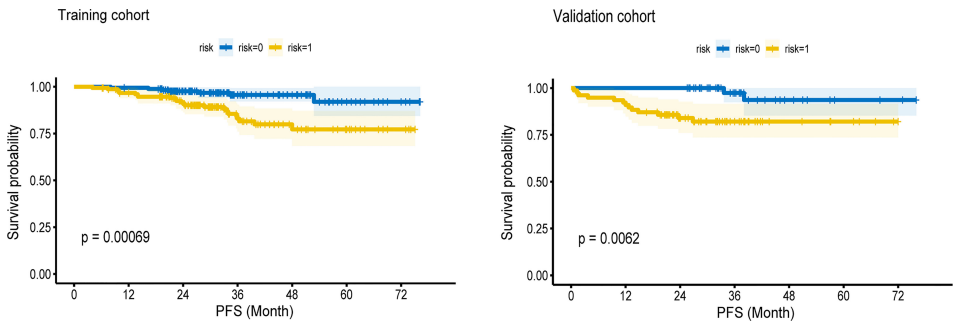


FIGURE 4 | The Kaplan—Meier survival curves of high-risk and low-risk groups in the training cohort and validation cohort. In both cohorts, the low-risk group had longer PFS ($P < 0.05$).

even with missing ENV DNA data. Currently, we are expanding the sample size or conducting multicenter studies to address this issue. Compared with plasma EBV DNA, radiomics features are more advanced and accurate in predicting prognosis (29). One study reported that EBV DNA can induce monocytes to produce interluekin-10, which leads to immune escape (30). Based on this, we collected easily obtainable blood parameters from NPC, expecting to find stable markers and incorporate them into the radiomics nomogram. After drawing the ROC curve for blood parameters, we found that monocytes had the best sensitivity and specificity. Two retrospective studies validated age, gender, Ki-67, and smoking and drinking habits as independent prognostic factors for NPC (11, 31). Our results showed the model integrating clinical data and the rad-score was more useful than those only using radiomics features.

Although we successfully demonstrated the utility of radiomics data for predicting PFS in patients with NPC, this study has three major limitations. First, this was a single-center retrospective study, so the results may not readily be applicable to other situations and prospective multicenter studies are needed to confirm our findings. Second, we selected patients according to strict inclusion criteria, which may have introduced selection bias. Third, our study only focused on PFS at 3 and 5 years. In the future, we will investigate the long-term overall survival of NPC and pay more attention to predicting long-term quality of life using imaging radiomics.

In conclusion, we established an effective clinical-radiomics nomogram based on MRI findings and several clinical, pathological, and blood factors. This approach is noninvasive, visualizable, and individualized and has great potential in predicting NPC prognosis and treatment. Moreover, we further confirmed that radiomics features were independent prognostic factors for NPC.

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 Sichuan Cancer Hospital. 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

Z-YF and W-DW: study design, statistical analysis, and wrote the manuscript. K-ZL, MY, and Y-RC: clinical data collection. L-PL, Z-FW, M-QG, CW, J-KL, and PZ: collection of images. CL, XL, Y-YZ, MW, ZX, and S-ML: collection of blood parameters. W-DW: ROI segmentation and manuscript modification and enrichment. All authors contributed to manuscript revision and read and approved the submitted version.

FUNDING

This study received funding from the National Key Research & Development Program of China (2017YFC0113904), the Sichuan Key Research & Development Project (2017SZ0004), and the Chengdu Technology Innovation R&D Project (2021YF0501659SN).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.815952/full#supplementary-material

REFERENCES

1. Liang X, Yang J, Gao TNC Center and C Hospital. Nasopharynx Cancer Epidemiology in China. *China Cancer* (2016) 25(11):835–40. doi: 10.11735/j.issn.1004-0242.2016.11.A001
2. 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 (6):394–424. doi: 10.3322/caac.21492
3. Chua ML, Sun Y, Supiot S. Advances in Nasopharyngeal Carcinoma- "West Meets East." *Br J Radiol* (2019) 92(1102):20199004. doi: 10.1259/bjr.20199004
4. Hua YJ, Han F, Lu LX, Mai HQ, Guo X, Hong MH, et al. Long-Term Treatment Outcome of Recurrent Nasopharyngeal Carcinoma Treated With Salvage Intensity Modulated Radiotherapy. *Eur J Cancer* (2012) 48(18):3422–8. doi: 10.1016/j.ejca.2012.06.016
5. Kong FF, Zhou JJ, Du CR, He XY, Hu CS, Ying HM, et al. Long-Term Survival and Late Complications of Intensity-Modulated Radiotherapy for Recurrent Nasopharyngeal Carcinoma. *BMC Cancer* (2018) 18(1):1139. doi: 10.1186/s12885-018-5055-5
6. Chen FP, Li L, Qi ZY, Zhou GQ, Rui G, Hu J, et al. Pretreatment Nomograms for Local and Regional Recurrence After Radical Radiation Therapy for Primary Nasopharyngeal Carcinoma. *J Cancer* (2017) 8(13):2595–603. doi: 10.7150/jca.20255
7. Nilsson JS, Forslund O, Andersson FC, Lindstedt M, Greiff L. Intralesional EBV-DNA Load as Marker of Prognosis for Nasopharyngeal Cancer. *Sci Rep UK* (2019) 9(1):15432. doi: 10.1038/s41598-019-51767-9
8. Prayongrat A, Chakkabat C, Kannarunimit D, Hansasuta P, Lertbutsayanukul C. Prevalence and Significance of Plasma Epstein-Barr Virus DNA Level in Nasopharyngeal Carcinoma. *J Radiat Res* (2017) 58(4):509–16. doi: 10.1093/jrr/rrw128
9. Kato H, Whiteside TL. Expression of IL-10 and IL-10 Receptors on Peripheral Blood Lymphocytes and Monocytes in Human Head and Neck Squamous Cell Carcinoma. *Tokai J Exp Clin Med* (2011) 36(4):144.
10. Borsetto D, Polesel J, Tirelli G, Menegaldo A, Boscolo-Rizzo P. Pretreatment High MCV as Adverse Prognostic Marker in Nonanemic Patients With Head and Neck Cancer. *Laryngoscope* (2020) 131(3):E836–45. doi: 10.1002/lary.28882
11. Shi Z, Jiang W, Chen X, Xu M, Zha D. Prognostic and Clinicopathological Value of Ki-67 Expression in Patients With Nasopharyngeal Carcinoma: A Meta-Analysis. *Ther Adv Med Oncol* (2020) 12:175883592095134. doi: 10.1177/1758835920951346
12. Lambin P, Leijen RTH, Deist TM, Peerlings J, de Jong EEC, van 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
13. Liu Z, Wang S, Dong D, Wei J, Tian J. 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
14. Aerts HJWH, Velazquez ER, Leijenaar THR, Parmar C, Grossmann P, Cavalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun* (2014) 5(1):1–8. doi: 10.138/ncomms5006
15. Haider SP, Burtness B, Yarbrough WG, Payabvash S. Applications of Radiomics in Precision Diagnosis, Prognostication and Treatment Planning of Head and Neck Squamous Cell Carcinomas. *Cancers Head Neck* (2020) 5:6. doi: 10.1186/s41199-020-00053-7
16. Zhang B, Tian J, Dong D, Gu D, Zhang S. 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
17. Zhang LL, Huang MY, Li Y, Gao TS, Deng B, Yao JJ, 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
18. Bologna M, Corino V, Calareso G, Tenconi C, Orlandi E. Baseline MRI-Radiomics can Predict Overall Survival in Non-Endemic EBV-Related Nasopharyngeal Carcinoma Patients. *Cancers* (2020) 12(10):2958. doi: 10.3390/cancers12102958
19. Amin MB, Edge S, Greene F, Byrd DR, Brookland RK, Washington MK, et al. *AJCC Cancer Staging Manual 8th Edition* Vol. XVII. New York: Springer (2017). p. 1032.
20. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: Improved N3 Bias Correction. *IEEE T Med Imaging* (2010) 29 (6):1310–20. doi: 10.1109/TMI.2010.2046908
21. 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):2287–300. doi: 10.1007/s11517-018-1858-4
22. Yu TT, Lam SK, To LH, Tse KY, Cheng NY, Fan YN, et al. Pretreatment Prediction of Adaptive Radiation Therapy Eligibility Using MRI-Based Radiomics for Advanced Nasopharyngeal Carcinoma Patients. *Front Oncol* (2019) 9:1050. doi: 10.3389/fonc.2019.01050
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(9):1323–41. doi: 10.1016/j.mri.2012.05.001
24. Feng Q, Liang J, Wang L, Niu J, Ding Z. Radiomics Analysis and Correlation With Metabolic Parameters in Nasopharyngeal Carcinoma Based on PET/MR Imaging. *Front Oncol* (2020) 10:1619. doi: 10.3389/fonc.2020.01619
25. Zhao L, Gong J, Xi YB, Li C, Kang XW, Yin YT, et al. MRI-Based Radiomics Nomogram may Predict the Response to Induction Chemotherapy and Survival in Locally Advanced Nasopharyngeal Carcinoma. *Eur Radiol* (2020) 30(1):537–46. doi: 10.1007/s00330-019-06211-x
26. Zhuo EH, Zhang WJ, Li HJ, Zhang GY, Jing BZ, Zhou J, et al. Radiomics on Multi-Modalities MR Sequences can Subtype Patients With Non-Metastatic Nasopharyngeal Carcinoma (NPC) Into Distinct Survival Subgroups. *Eur Radiol* (2019) 29(10):5590–9. doi: 10.1007/s00330-019-06075-1
27. Kim MJ, Choi Y, Sung YE, Lee YS, Kim YS, Ahn KJ, et al. Early Risk-Assessment of Patients With Nasopharyngeal Carcinoma: The Added Prognostic Value of MR-Based Radiomics. *Transl Oncol* (2021) 14 (10):101180. doi: 10.1016/j.tranon.2021.101180
28. Shen H, Wang Y, Liu D, Lv R, Zhang J. Predicting Progression-Free Survival Using MRI-Based Radiomics for Patients With Nonmetastatic Nasopharyngeal Carcinoma. *Front Oncol* (2020) 10:618. doi: 10.3389/fonc.2020.00618
29. Peng H, Dong D, Fang M, Li L, Tang LL, Chen L, et al. Prognostic Value of Deep Learning Pet/Ct-Based Radiomics: Potential Role for Future Individual Induction Chemotherapy in Advanced Nasopharyngeal Carcinoma. *Clin Cancer Res* (2019) 25(14):4271–9. doi: 10.1158/1078-0432.CCR-18-3065
30. Lee CH, Yeh TH, Lai HC, Wu SY, Su IJ, Takada K, et al. Epstein-Barr Virus Zta-Induced Immunomodulators From Nasopharyngeal Carcinoma Cells Upregulate Interleukin-10 Production From Monocytes. *J Virol* (2011) 85 (14):7333–42. doi: 10.1128/JVI.00182-11
31. Xiao G, Cao Y, Qiu X, Wang W, Wang Y. Influence of Gender and Age on the Survival of Patients With Nasopharyngeal Carcinoma. *BMC Cancer* (2013) 13 (1):226. doi: 10.1186/1471-2407-13-226

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 03 March 2022
doi: 10.3389/fonc.2022.796583

Multiparametric MRI-Based Radiomics Model for Predicting H3 K27M Mutant Status in Diffuse Midline Glioma: A Comparative Study Across Different Sequences and Machine Learning Techniques

Wei Guo[1], *Dejun She*[1], *Zhen Xing*[1], *Xiang Lin*[1], *Feng Wang*[1], *Yang Song*[2] *and Dairong Cao*[1,3,4*]

[1] Department of Radiology, First Affiliated Hospital of Fujian Medical University, Fuzhou, China, [2] MR Scientific Marketing, Siemens Healthineers Ltd., Shanghai, China, [3] Department of Radiology, Fujian Key Laboratory of Precision Medicine for Cancer, The First Affiliated Hospital, Fujian Medical University, Fuzhou, China, [4] Key Laboratory of Radiation Biology of Fujian Higher Education Institutions, The First Affiliated Hospital, Fujian Medical University, Fuzhou, China

OPEN ACCESS

Edited by:
Oliver Diaz,
University of Barcelona, Spain

Reviewed by:
Weiwei Zong,
Henry Ford Health System, United States
Gökalp Çınarer,
Bozok University, Turkey

***Correspondence:**
Dairong Cao
dairongcao@163.com

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

Received: 17 October 2021
Accepted: 08 February 2022
Published: 03 March 2022

Citation:
Guo W, She D, Xing Z, Lin X, Wang F, Song Y and Cao D (2022) Multiparametric MRI-Based Radiomics Model for Predicting H3 K27M Mutant Status in Diffuse Midline Glioma: A Comparative Study Across Different Sequences and Machine Learning Techniques. Front. Oncol. 12:796583. doi: 10.3389/fonc.2022.796583

Objectives: The performance of multiparametric MRI-based radiomics models for predicting H3 K27M mutant status in diffuse midline glioma (DMG) has not been thoroughly evaluated. The optimal combination of multiparametric MRI and machine learning techniques remains undetermined. We compared the performance of various radiomics models across different MRI sequences and different machine learning techniques.

Methods: A total of 102 patients with pathologically confirmed DMG were retrospectively enrolled (27 with H3 K27M-mutant and 75 with H3 K27M wild-type). Radiomics features were extracted from eight sequences, and 18 feature sets were conducted by independent combination. There were three feature matrix normalization algorithms, two dimensionality-reduction methods, four feature selectors, and seven classifiers, consisting of 168 machine learning pipelines. Radiomics models were established across different feature sets and machine learning pipelines. The performance of models was evaluated using receiver operating characteristic curves with area under the curve (AUC) and compared with DeLong's test.

Results: The multiparametric MRI-based radiomics models could accurately predict the H3 K27M mutant status in DMG (highest AUC: 0.807–0.969, for different sequences or sequence combinations). However, the results varied significantly between different machine learning techniques. When suitable machine learning techniques were used, the conventional MRI-based radiomics models shared similar performance to advanced MRI-based models (highest AUC: 0.875–0.915 vs. 0.807–0.926; DeLong's test, $p > 0.05$). Most models had a better performance when generated with a combination of MRI

sequences. The optimal model in the present study used a combination of all sequences (AUC = 0.969).

Conclusions: The multiparametric MRI-based radiomics models could be useful for predicting H3 K27M mutant status in DMG, but the performance varied across different sequences and machine learning techniques.

Keywords: diffuse midline glioma, H3 K27M mutant, multiparametric MRI, radiomics, machine learning

INTRODUCTION

As a newly defined subtype of the 2016 WHO Classification of Tumors of the Central Nervous System, "diffuse midline glioma (DMG), H3 K27M mutant" is characterized by a genetic alteration pattern in either H3F3A or HIST1H3B/C (1). Compared to the wild-type group, the group with DMG with an H3 K27M mutation exhibited a particularly dismal prognosis, with 3-year overall survival of 5% and 2-year overall survival of less than 10% (2–5). In addition, the previous studies revealed that H3 K27M mutant status represented a potential novel therapeutic target for DMG, which confronts the fact of resistance to the conventional therapy strategies (6–10). Identifying H3 K27M mutant status plays an essential role in tumor diagnosis, survival prediction, and therapeutic decision-making. Surgical resection or biopsy could provide an accurate result of H3 K27M mutant status but is not always feasible due to tumor tissue's spatial heterogeneity and unforeseeable complications. Developing a non-invasive method for accurately predicting H3 K27M mutant status is critical for DMG management.

Several recent attempts have been made to use the multiparametric MRI-based radiomics model to predict H3 K27M mutant status, but the results varied greatly (11–16). Most of them focused on different kinds of conventional MRI (cMRI), which could only reflect the tumor's morphologic information and benefit limitedly to reveal tumor heterogeneity. The advanced MRI (aMRI) (e.g., diffusion-weighted imaging [DWI], susceptibility-weighted imaging [SWI], and dynamic susceptibility contrast perfusion-weighted imaging [DSC-PWI]), which could provide physiological information within the tumor, has been proved to be helpful in radiomics-based glioma genotype prediction (17–19). However, the utility of an advanced MRI-based radiomics model in predicting H3 K27M mutant status has not been well evaluated. On the other hand, previous studies indicated that the performance of the radiomics model predominantly varied with the type of image set used (20, 21). As such, it is unclear whether aMRI or a combination of cMRI and aMRI could make an equivalent or superior performance as compared to cMRI.

In addition to the heterogeneous sequence used, the previous studies on H3 K27M mutant status prediction employed a great diversity of machine learning techniques, including dimensionality-reduction algorithm, feature selector, and classifier. It has been well recognized that the radiomics model established *via* different machine learning techniques could achieve diverse results even when the same sequence was used (22, 23). This could be a potential reason for the inconsistent prior radiomics-based H3 K27M mutant status prediction results. Therefore, there is an urgent need for a head-to-head comparison of the prediction power across different machine learning techniques and sequence or sequence combinations to determine the best machine learning techniques with the best image sets.

The purposes of this study were to 1) detect the best MRI sequence or sequence combinations for predicting H3 K27M mutant status in DMG and 2) determine the optimal machine learning technique for different image sets.

MATERIALS AND METHODS

Study Population

The Ethical Committee of the First Affiliated Hospital of Fujian Medical University approved this study. The requirement for written informed consent was waived due to the retrospective nature. One hundred two patients were consecutively enrolled in the present study from July 2010 to August 2021. The inclusion criteria were as follows: 1) patients have a pathological diagnosis of diffuse glioma and confirmation of H3 K27M mutant status; 2) tumor is located in the midline structure of the brain; and 3) full preoperative MR images were available. Exclusion criteria were as follows: 1) absence of any required MR images or the image quality was insufficient for analysis and 2) the tumor volume was less than 1.5 cm^3. The patients were randomly split into training and test groups with a ratio of 7:3. Extra effort was made to keep the balance between the training and test cohorts.

MRI Protocol

The neurologic MRI examinations were performed with the 3.0-Tesla MR scanner (MAGNETOM Verio/Skyra/Prisma, Siemens Healthcare, Erlangen, Germany). The standard multiparametric

Abbreviations: AB, AdaBoost; ADC, apparent diffusion coefficient; AE, auto-encoder; AUC, area under the curve; CBF, cerebral blood flow; CBV, cerebral blood volume; CE-T1WI, contrast-enhanced T1-weighted imaging; DMG, diffuse midline glioma; DSC-PWI, dynamic susceptibility contrast perfusion-weighted imaging; DT, decision tree; DWI, diffusion-weighted imaging; FLAIR, fluid-attenuated inversion recovery; ICC, intraclass correlation coefficient; KW, Kruskal–Wallis; LDA, linear discriminant analysis; LR, logistic regression; PCA, principal component analysis; PCC, Pearson's correlation coefficient; RF, random forest; RFE, recursive feature elimination; SVM, support vector machine; SWI, susceptibility-weighted imaging; T1WI, T1-weighted imaging; T2WI, T2-weighted imaging; VOI, volume of interest.

MRI sequences in the present study, including T2-weighted imaging (T2WI), T1-weighted imaging (T1WI), fluid-attenuated inversion recovery (FLAIR), contrast-enhanced T1WI (CE-T1WI), SWI, DWI, and DSC-PWI. The details of MRI acquisition parameters are listed in the **Supplementary Material 1** (**Table S1**). The apparent diffusion coefficient (ADC) map was automatically derived from DWI data with b-values of 0 and 1,000 s/mm^2. The DSC-PWI raw data were scrolled into a dedicated commercial software package (SyngoVia, Siemens), and the standard perfusion maps (cerebral blood volume [CBV] and cerebral blood flow [CBF]) were conducted as guidance of the software. In the 4th phase during the DSC-PWI scanning, a standard dose (0.1 mmol/kg) of gadobenate dimeglumine (Gd-BOPTA) followed by 20 ml of saline was injected intravenously with a flow rate of 3 ml/s. CE-T1WI was scanned after DSC-PWI.

Image Pre-Processing and Tumor Segmentation

Before pre-processing, the DICOM images were converted to the nifti format. The standard image pre-processing included four steps: 1) all sequences were registered to T2WI initially with a block matching algorithm; 2) following the co-registration, the images were resampled into the uniform voxel size of $1 \times 1 \times 5$ mm^3; 3) N4 Bias Field Correction package was applied to correct the bias filed; 4) finally, the image intensities were standardized to [0, 255] to reduce the influence of imaging intensity inconsistency. All of the pre-processing procedures were achieved using G.K software (Glioma kit, version 1.2.1.R, GE Healthcare, Shanghai, China).

Tumor segmentation was performed by one radiologist (DS, with 10 years of experience in neuroradiology) and verified by another radiologist (DC, with 30 years of experience in neuroradiology) who were unaware of the pathological results. The volume of interest (VOI) was created to cover the tumor core (including the enhancing, non-enhancing, and necrotic/cystic components) on T2WI with ITK-SNAP (http://www.itksnap.org) by referring to the T1WI, CE-T1WI, and FLAIR images. According to VASARI guidelines (Visually AcceSAble Rembrandt Images; https://wiki.nci.nih.gov/display/CIP/VASARI), the respective portions of the tumor were defined as described in the previous study (24, 25). As the radiomics feature extraction differed between VOIs, the intra-observer and inter-observer reproducibility analyses were achieved to minimize the influence of segmentation bias. Of intra-observer reproducibility analysis, the VOIs of 30 randomly chosen patients were segmented twice by one radiologist (DS). The inter-observer reproducibility analysis was performed based on the same cohort above, where the VOIs were segmented by two radiologists (ZX and DS, both with 10 years of experience in neuroradiology). The intraclass correlation coefficient (ICC) was calculated to evaluate the agreement of radiomics feature extraction.

Radiomics Feature Extraction

An open-source software, FeAture Explore (V 0.4.2), was used for quantitative radiomics feature extraction with the Pyradiomics module on Python (3.7.6) (26, 27). A total of 851 features were extracted from each sequence image, consisting of 18 first-order statistics features, 14 shape-based features, 75 texture features, and 744 wavelet features from eight wavelet-transformed images (https://pyradiomics.readthedocs.io/en/latest/features.html). The details of the extracted features are listed in the **Supplementary Material 1** (**Table S2**). Eight sequences (T2WI, T1WI, FLAIR, CE-T1WI, ADC, SWI, CBV, and CBF) were used in the present study. Thus, a total of 6,808 features were extracted for analysis. We conducted 18 feature sets by the independent combination of features extracted from these eight sequences. The feature sets were generally named with the name of sequences. Especially, "cMRI," "aMRI," and "ALL" denote the combination of all cMRI sequences, aMRI sequences, and eight sequences, respectively.

Radiomics Feature Matrix Pre-Processing

As described above, for the sake of minimizing the influence of VOI segmentation bias on radiomics feature calculation and further machine learning analysis, the features with an ICC value lower than 0.75 in either the intra-observer or inter-observer reproducibility analysis were removed. Then we applied the normalization to the remaining feature matrix. Three feature normalization methods were considered: mean normalization, min–max normalization, and Z-score normalization. The mean normalization subtracted each feature vector by the mean value of the vector and divided each feature by the length of the vector. For the min–max normalization, we rescaled the minimum and maximum values of the feature from zero to one. Then the feature vector was mapped to a unit vector. When the Z-score method was applied, we calculated each feature vector's mean value and SD. Then each feature was subtracted by the mean value and was divided by the SD. Notably, only one normalization method was used in one machine learning pipeline.

Radiomics Feature Dimensionality Reduction and Feature Selection

Since the feature space dimension was high, we applied two alternative feature dimensionality-reduction methods in the presented study, including Pearson's correlation coefficient (PCC) and principal component analysis (PCA). The PCC was calculated for each pair of two normalized features, and we removed one of them if the PCC was larger than the preset threshold. By referring to the previous study, the threshold was set to 0.8 for the model using a single sequence and 0.6 for the model using a combination of different sequences (20). When the PCA method was chosen, the high dimension features were transformed into the relative lower dimension features. The feature vector of the transformed feature matrix was independent of each other.

Following feature dimensionality reduction, four optional methods were provided for feature selection, including ANOVA, recursive feature elimination (RFE), Kruskal–Wallis (KW), and Relief.

Predictive Model Establishment

Seven machine learning classifiers were analyzed to determine the optimal model. The SVM classifier we used was based on a linear kernel function , and it may be more appreciated to be cataloged into the linear classifier. The sentence should be corrected as "These classifiers could be divided into three categories: linear (logistic regression [LR], linear discriminant analysis [LDA], and support vector machine [SVM]), non-linear classifiers (auto-encoder [AE] and decision tree [DT]), and ensemble classifiers (random forest [RF] and AdaBoost [AB]). The five-fold cross-validation was applied on the training dataset to determine the model's hyper-parameter, such as the number of features and specific hyper-parameters of each classifier, which can be referred on the scikit-learn (https://scikit-learn.org/stable/index.html). The hyper-parameters were set according to the model performance on the cross-validation dataset.

Considering different combinations of each procedure during model development, including sequence used, feature matrix normalization, dimensionality reduction, and feature selection, could provide controversial results with different classifiers. We analyzed models' performance from 8 single sequences and 10 different sequence combinations with different machine learning techniques. Thus, a total of 3,024 models were conducted in the present study (18 [sequence groups] × 3 [feature matrix normalization] × 2 [dimensionality reduction] × 4 [features selector] × 7 [classifiers] = 3,024 [models]). The flowchart of the present study is illustrated in **Figure 1**. The above processes, including feature matrix normalization, dimensionality reduction, feature selection, and classifier fitness, were implemented with FeAture Explorer (V 0.4.2) on the training cohort. Then, we evaluated the models' performance on the independent test cohort.

Statistical Analysis

The performance of each model was evaluated with receiver operating characteristic curve analysis. The area under the receiver operating characteristic curve (AUC) and accuracy were calculated. We also estimated the 95% CI by bootstrap with 1,000 samples. To assess the variability in the performance of different models, we compared the top-one-performing models and the top-five-performing models of each sequence or sequence combination. Continuous variables of the baseline characteristics were described as the mean ± SD and compared using the Mann–Whitney U test. Categorical variables of the baseline characteristics were described as number (percentage) and compared using Pearson's chi-squared test. The comparison of AUCs between different models was performed using Delong's test. The statistical analyses were performed with R statistical

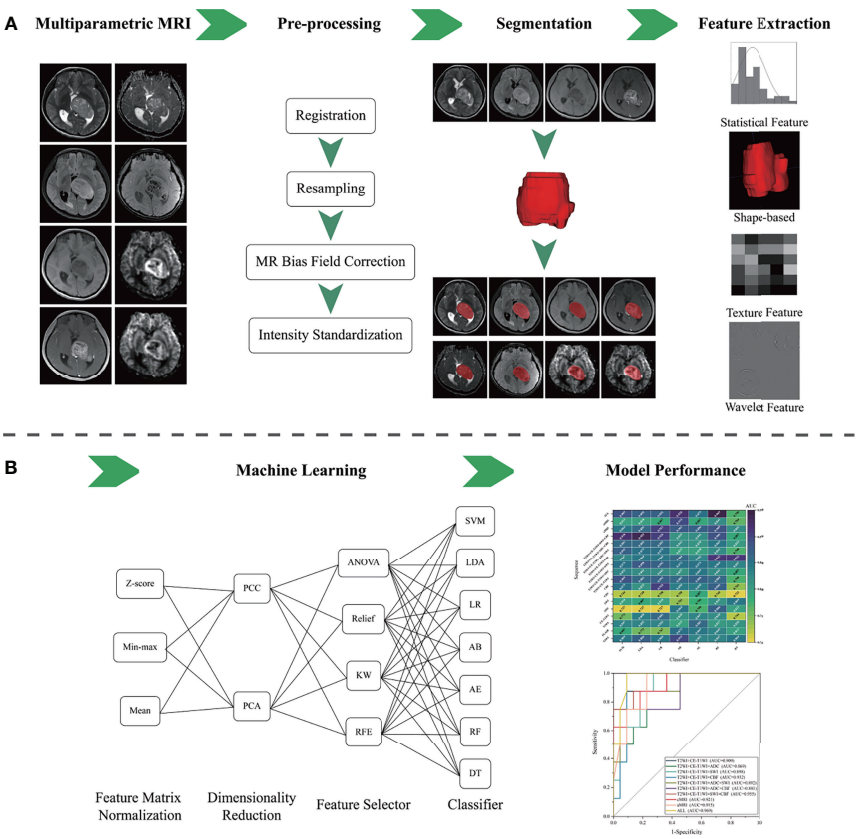


FIGURE 1 | The flowchart of the presented study. **(A)** Multiparametric MRI data collection, image pre-processing, tumor segmentation, and radiomics feature extraction. **(B)** Machine learning and model performance analysis.

software (version 3.5.3; https://www.r-project.org/). A p-value <0.05 was considered statistically significant.

RESULTS

Baseline Characteristics of Patients

Of the 102 patients, 27 (26.47%) patients were confirmed with an H3 K27M mutation. The mean age was 41.19 ± 20.64 years, and the male ratio was 64 (62.75%). No statistically significant difference was found in the baseline characteristics between the training and test groups (p > 0.05) (**Table 1**).

Performance of Sequence

In general, most of the high-performing models (with an AUC value larger than 0.9 in the test set) were conducted from the combination of different sequences (**Figures 2**, **3** and **Tables 2** and **S3**). The ALL model showed the strongest predictive power among various models for H3 K27M mutant status (AUC = 0.969), while the best single-sequence model was the CBF-based model (AUC = 0.926), followed by the T2WI-based model (AUC = 0.915). The CBV-based model yielded the lowest AUC value of 0.807 among the top-one-performing models of different sequences or sequence combinations (**Figure 4**).

The cMRI showed comparable performance to aMRI when suitable machine learning techniques were employed (DeLong's test, all p > 0.05) (**Table 3**). In models based on a single sequence, the highest AUCs were 0.875–0.915 for cMRI sequences and 0.807–0.926 for aMRI sequences (**Table 2** and **Figure 4**). The model of cMRI yielded a slightly higher AUC than the model of aMRI in the test set (AUC: 0.921 vs. 0.915). When combining limited sequences of cMRI and aMRI, the model of T2WI+CE-T1WI+SWI+CBF reached the highest AUC of 0.955. No statistically significant difference of the highest AUC values between the optimal model (ALL, AUC = 0.969) and other sequence-based models was found (DeLong's test, all p > 0.05) (**Table 3**).

Performance of Machine Learning Technique

Figures 2 and **S1** demonstrate the performance of different machine learning techniques. The machine learning pipeline of the optimal model was Z-score_PCA_KW_RF and Z-score_PCA_ANOVA_RF, both with an AUC value of 0.969 (**Figure 2** and **Table S3**). Among the 90 top-five-performing models, the Z-score normalization method outperformed others with darker color lines in **Figure 2** and a higher mean AUC value in **Figure S1**. In the same way, feature sets applying

TABLE 1 | Baseline characteristics of the training and test groups.

Characteristics	All (n = 102)	Training (n = 72)	Test (n = 30)	p-Value
Age (years)	41.19 ± 20.64	41.63 ± 20.80	40.13 ± 20.57	0.649
Gender (%)				0.711
Male	64 (62.75%)	46 (63.89%)	18 (60.00%)	
Female	38 (37.25%)	26 (36.11%)	12 (40.00%)	
H3 K27M mutant status (%)				0.977
Mutant	27 (26.47%)	19 (26.39%)	8 (26.67%)	
Wild type	75 (73.53%)	53 (73.61%)	22 (73.33%)	

A p-value <0.05 indicates the statistical significance of the variate difference between training and test sets. Continuous variables were described as the mean ± SD. Categorical variables were presented as the number, with percentages in parentheses.

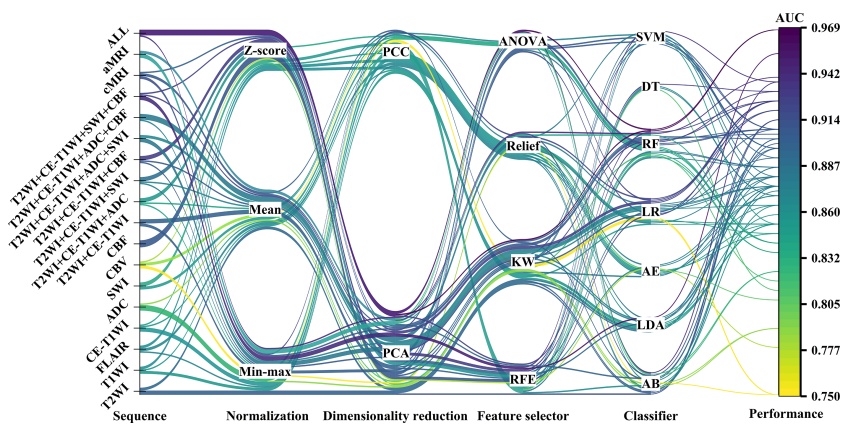


FIGURE 2 | The machine learning pipelines and performance of top-five-performing models of different sequences. The color of lines indicated the performance of models in the test set.

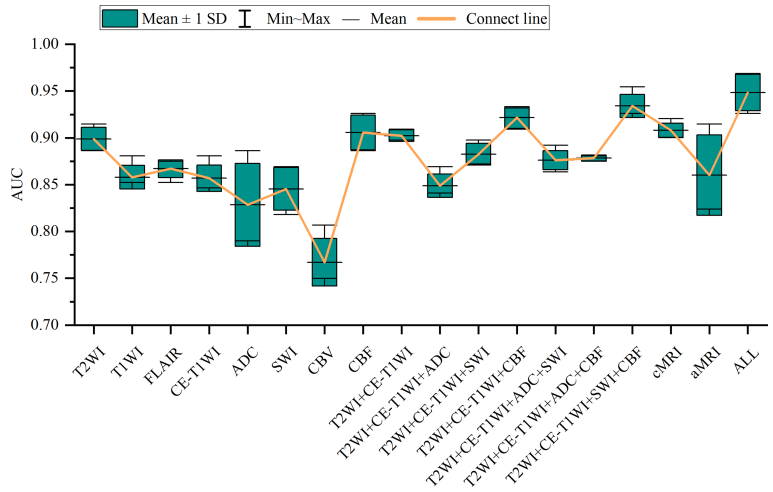


FIGURE 3 | Box-and-whisker plots illustrate the top-five-performing area under the curve (AUC) values of different sequences.

dimensionality reduction with the PCA method had a higher AUC value. **Figure 5** shows the best performance across different sequences and classifiers. The comparison results of the different classifiers are shown in **Table 3**. Of ADC-based models, CBF-based models, and T2WI+CE-T1WI+ADC+CBF-based models, a significant difference could be found in the AUC values between the best classifier and worst classifier (DeLong's test, $p < 0.05$) (**Table 3**). In contrast to the sequence with a suitable classifier, if the non-optimal classifier was used, the performance of different sequences varied significantly (DeLong's test, $p < 0.05$) (**Table 3**).

Among the top-five-performing models, the distribution of machine learning techniques varied considerably in different categories of MRI sequences (**Figures 2**, **6**). PCA was more frequently used in the top-five-performing models (66% of all sequences), especially in the model that simultaneously combined multiple MR images (86%). Feature selector of KW has a higher percentage in both single sequence-based (which have fewer features) and combined sequence-based (which have more features) models.

DISCUSSION

This study developed and validated various machine learning-based models with radiomics features extracted from multiparametric MRI to predict H3 K27M mutant status in DMG. The model's performance was compared across different sequences and machine learning techniques. Radiomics models derived from multiparametric MRI performed well in differentiating H3 K27M mutant and wild-type DMGs when a suitable machine learning technique was used (highest AUC: 0.807–0.969). However, the performance of the models can vary significantly regarding different machine learning techniques (DeLong's test, $p < 0.05$). Generally, the models developed with multi-sequence had a better performance than one with a single sequence. The cMRI-based model showed comparable performance to aMRI (highest AUC: 0.875–0.915 for cMRI, 0.807–0.926 for aMRI).

In line with the previous study, radiomics models based on cMRI could accurately predict the H3 K27M mutant status in DMGs (11–14). As an essential supplement to prior studies, our result also declared that the radiomics model developed with the aMRI, including ADC, SWI, CBV, and CBF, could be qualified for this purpose. Meanwhile, when appropriate machine learning techniques were used, the cMRI and aMRI shared comparable performance (DeLong's test, $p > 0.05$). A significant difference in ADC, CBV, and CBF values (measured with the freehand regions of interest) has been reported between H3 K27M mutant and wild-type DMGs (28–30). Other studies found that several semantic and semiquantitative features on cMRI could be used to predict H3 K27M mutant status in DMG (31, 32). But other non-radiomics studies using cMRI and DWI to predict H3 K27M mutant status showed converse results (33, 34). Radiomics has been proved to excavate numerous features from medical images, and most of these features are undiscoverable by the naked eye (35, 36). Analyzing medical images with a non-radiomics method may result in a loss of information within images. Wu et al. used radiological features and radiomics features to predict H3 K27M mutant status. Their results showed that the radiomics model performed significantly better than the clinical model (developed with radiological features) (16). The controversial results of non-radiomics studies and the robust results of radiomics studies supported that if the diagnostic information had been sufficiently explored using the radiomics method, the predictive ability of multiparametric MRI could be improved. This has been proved again by our results.

Another important observation was that most models that originated from combined sequences have a better predictive

TABLE 2 | The performance of the top-one-performing models.

Sequence	Machine learning technique	Dataset	AUC	95% CI	ACC	SEN	SPE	PPV	NPV
T2WI	Min–max_PCA_RFE_AB	Training	1.000	1.000–1.000	1.000	1.000	1.000	1.000	1.000
		Test	0.915	0.769–1.000	0.900	0.750	0.955	0.857	0.913
T1WI	Z-score_PCC_KW_AE	Training	0.767	0.631–0.890	0.694	0.842	0.642	0.457	0.919
		Test	0.881	0.733–0.984	0.700	1.000	0.591	1.000	0.471
FLAIR	Mean_PCC_Relief_AB	Training	1.000	1.000–1.000	1.000	1.000	1.000	1.000	1.000
		Test	0.875	0.722–0.984	0.833	0.625	0.909	0.870	0.714
CE-T1WI	Min–max_PCC_Relief_LR	Training	0.780	0.669–0.883	0.653	0.947	0.547	0.429	0.967
		Test	0.881	0.733–0.984	0.800	1.000	0.727	1.000	0.571
ADC	Min–max_PCC_RFE_RF	Training	1.000	1.000–1.000	1.000	1.000	1.000	1.000	1.000
		Test	0.886	0.718–1.000	0.700	0.000	0.955	0.724	0.000
SWI	Mean_PCC_RFE_DT	Training	1.000	1.000–1.000	1.000	1.000	1.000	1.000	1.000
		Test	0.869	0.694–0.979	0.867	0.875	0.864	0.950	0.700
CBV	Mean_PCA_Relief_AE	Training	0.640	0.490–0.779	0.736	0.474	0.830	0.500	0.815
		Test	0.807	0.585–0.980	0.700	0.875	0.636	0.933	0.467
CBF	Z-score_PCA_RFE_LR	Training	0.924	0.844–0.983	0.875	0.842	0.887	0.727	0.940
		Test	0.926	0.814–1.000	0.833	0.875	0.818	0.947	0.636
T2WI+CE-T1WI	Mean_PCA_ANOVA_SVM	Training	0.964	0.919–0.995	0.944	0.895	0.962	0.895	0.962
		Test	0.909	0.769–1.000	0.800	0.500	0.909	0.833	0.667
T2WI+CE-T1WI+ADC	Z-score_PCA_RFE_AE	Training	0.965	0.920–1.000	0.931	0.842	0.962	0.889	0.944
		Test	0.869	0.727–0.976	0.733	0.625	0.773	0.850	0.500
T2WI+CE-T1WI+SWI	Z-score_PCA_RFE_LDA	Training	0.963	0.905–1.000	0.958	0.947	0.962	0.900	0.981
		Test	0.898	0.761–0.988	0.800	0.750	0.818	0.900	0.600
T2WI+CE-T1WI+CBF	Z-score_PCA_RFE_RF	Training	1.000	1.000–1.000	1.000	1.000	1.000	1.000	1.000
		Test	0.932	0.824–1.000	0.733	0.125	0.955	0.750	0.500
T2WI+CE-T1WI+ADC+SWI	Min–max_PCA_KW_SVM	Training	1.000	1.000–1.000	1.000	1.000	1.000	1.000	1.000
		Test	0.892	0.728–0.994	0.867	0.750	0.909	0.909	0.750
T2WI+CE-T1WI+ADC+CBF	Mean_PCA_Relief_LR	Training	0.555	0.379–0.733	0.736	0.421	0.849	0.500	0.804
		Test	0.881	0.701–1.000	0.900	0.750	0.955	0.913	0.857
T2WI+CE-T1WI+SWI+CBF	Min–max_PCA_RFE_LDA	Training	0.888	0.805–0.957	0.806	0.895	0.774	0.586	0.954
		Test	0.955	0.854–1.000	0.767	1.000	0.682	1.000	0.533
cMRI	Min–max_PCC_Relief_LR	Training	0.833	0.731–0.933	0.778	0.842	0.755	0.552	0.930
		Test	0.921	0.778–1.000	0.733	1.000	0.636	1.000	0.500
aMRI	Mean_PCA_Relief_AB	Training	1.000	1.000–1.000	1.000	1.000	1.000	1.000	1.000
		Test	0.915	0.800–0.993	0.800	0.875	0.773	0.944	0.583
ALL	Z-score_PCA_KW_RF	Training	1.000	1.000–1.000	1.000	1.000	1.000	1.000	1.000
		Test	0.969	0.904–1.000	0.767	0.125	1.000	0.759	1.000

Machine learning technique was expressed as "feature matrix normalization_dimensionality reduction_feature selector_classifier".
T2WI, T2-weighted imaging; T1WI, T1-weighted imaging; FLAIR, fluid-attenuated inversion recovery; CE-T1WI, contrast-enhanced T1WI; ADC, apparent diffusion coefficient; SWI, susceptibility-weighted imaging; CBV, cerebral blood volume; CBF, cerebral blood flow; cMRI, model developed with all of the conventional MRI; aMRI, model developed with all of the advanced MRI; ALL, model developed with all of the eight sequences; PCC, Pearson's correlation coefficient; PCA, principal component analysis; RFE, recursive feature elimination; KW, Kruskal–Wallis; LR, logistic regression; LDA, linear discriminant analysis; SVM, support vector machine; AE, auto-encoder, DT, decision tree; RF, random forest; AB, AdaBoost; AUC, area under the curve; ACC, accuracy; SEN, sensibility; SPE, specificity; PPV, positive predictive value; NPV, negative predictive value.

performance, whether the optimal classifier was used (**Figures 2**, **3**) or not (**Figure 5**). Previous studies using a multiparametric MRI-based radiomics model to predict glioma molecular subtype also showed similar results to ours (18, 37). However, only three multiparametric MRI-based radiomics models were established previously and achieved the highest AUC value of 0.920 in the test cohort for H3 K27M mutant status prediction (12, 14, 16). They only make a direct combination of all sequences used, and the performance between single and combined sequences was not compared. Liu et al. developed a machine learning model based on T1WI images only to predict H3 K27M mutant status in DMGs, which yielded the highest AUC value of 0.953 (11). However, the sample size was relatively small (n = 55), and the final model features were slightly overmuch (n = 30). Another radiomics model based on FLAIR images showed an AUC value of 0.903 (13). It is unfair to compare the model's performance when different datasets were used. Our study compared the model performance based on the same dataset. The results showed that the model had the best predictive power when combined with all sequences (AUC = 0.969). The reason may be that complementary information among multiparametric MRI could provide a more comprehensive understanding of tumor heterogeneity and discriminate more precisely tumor classes. Also noteworthy is that the model combined with limited sequences was sufficient to differentiate H3 K27M-mutant and K27M-wt DMGs, such as the model based on feature sets from T2WI+CE-T1WI+SWI+CBF (AUC = 0.955) and T2WI+CE-T1WI+CBF (AUC = 0.932). This is relevant, as it could guide model application in various clinical circumstances and make it more time-efficient.

According to previous results, the feature selector and classifier were two major determinant factors of radiomics model performance (20–23, 38, 39). When a suitable classifier was used, there was no significant difference in the AUC value of

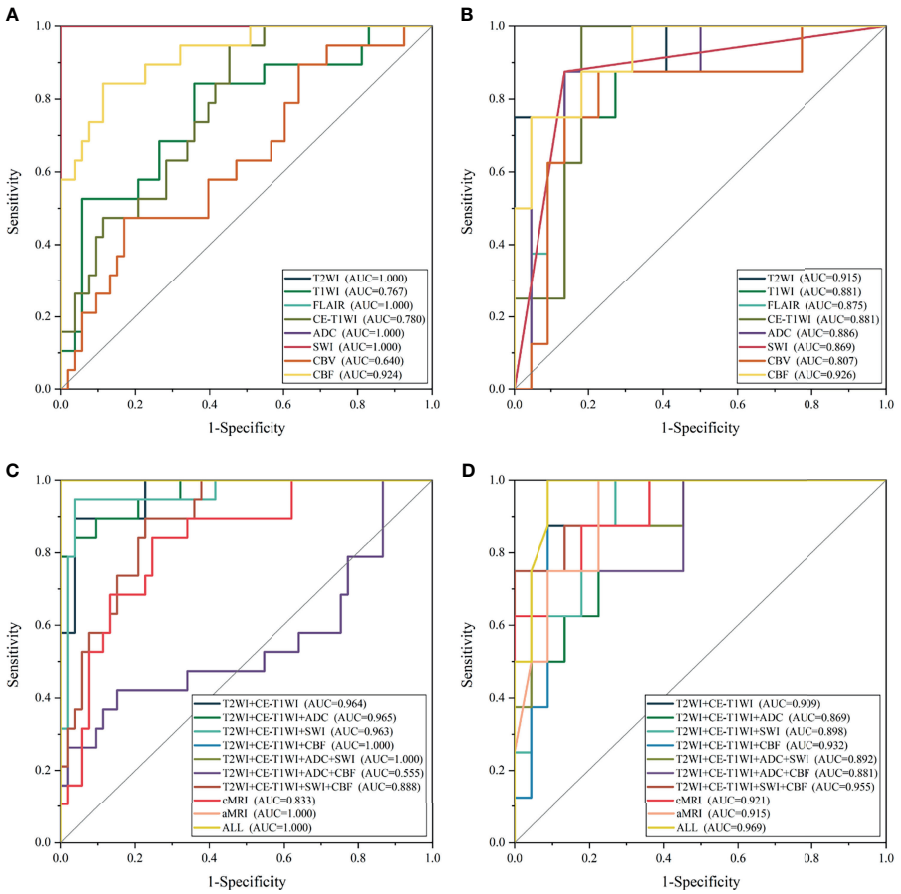


FIGURE 4 | The receiver operating characteristic curve of the top-one-performing models of different sequences in the training **(A, C)** and test sets **(B, D)**.

different sequences. Constantly, when an inappropriate classifier was used, both intra-sequence and inter-sequence comparisons yielded a significant difference in AUC values (**Table 3**). For the single sequence-based model, SVM, LDA, and LR classifiers were more frequently to have a lower AUC. The reason may be that the LR and LAD were both linear classifiers, and the SVM used linear kernel function in our study; thus, these classifiers were not flexible enough to fit a non-linear relationship between features and tumor groups. Furthermore, features extracted from a single sequence could only offer limited messages on tumor biological heterogeneity. Of note, the multiparametric MRI-based model with SVM, LDA, and LR demonstrated more favorable results. The prior study used various classifiers (e.g., SVM, RF, and XGBoost) and generated an AUC value of 0.549–0.953, which were lower than ours (AUC = 0.969) (11–15). Several reasons may account for this variety, including patient data, MRI data, and machine learning techniques. Hence, a head-to-head comparison may be more reliable to reveal the influence of these factors and determine optimal models when different image data are available.

Apart from the feature selector and classifier, our results revealed that the feature matrix normalization and dimensionality-reduction method also played a non-negligible role in model performance evaluation (**Figure S1**). The previous study focused on H3 K27M mutant status prediction, which rarely considered these elements. Two of them made an effort to compare the predictive power of different classifiers and another for different feature selectors (11, 15). The limitation of these studies on model development warrants extra caution in terms of result explanation. Our results demonstrated that the appropriate machine learning techniques mentioned above could vary greatly when various image data were used (**Figure 6**). This reemphasized that both the type of image data used and the employment of machine learning techniques will carry a diverse result. Thus, it is essential and encouraged to seek the optimal machine learning techniques when different image data are used. The compatible combination of medical images with machine learning techniques could maximize and robust the radiomics model's performance.

There are several limitations in the current study. First, this is a single-center retrospective study, which results in an unavoidable selecting bias and relatively small sample size. The imbalanced proportion of H3 K27M mutant DMG might influence the development of our models. A prospective and multi-institution study is needed for confirming our results. Second, the dataset was randomly split into the training and

TABLE 3 | Results of DeLong's test of the best models with different classifiers.

Sequence	Highest AUC		Lowest AUC		*p*-Value[a]	*p*-Value[b]	*p*-Value[c]
	Classifier	AUC	Classifier	AUC			
T2WI	AB	0.915	RF	0.847	0.4851	0.4508	0.0527
T1WI	AB	0.875	AB	0.815	0.5602	0.1908	0.0621
FLAIR	AE	0.881	LR	0.767	0.2334	0.1940	**0.0108**
CE-T1WI	LR	0.881	DT	0.744	0.2568	0.1720	**0.0437**
ADC	RF	0.886	SVM	0.727	**0.0118**	0.1837	**0.0144**
SWI	DT	0.869	AB	0.761	0.2292	0.1834	**0.0271**
CBV	AE	0.807	DT	0.722	0.3579	0.1108	**0.0104**
CBF	LR	0.926	DT	0.761	**0.0302**	0.4729	**0.0437**
T2WI+CE-T1WI	SVM	0.909	DT	0.790	0.2944	0.3316	0.0756
T2WI+CE-T1WI+ADC	AE	0.869	DT	0.807	0.4912	0.1319	0.0867
T2WI+CE-T1WI+SWI	LDA	0.898	DT	0.847	0.5049	0.1872	0.5532
T2WI+CE-T1WI+CBF	RF	0.932	SVM	0.847	0.3043	0.5163	0.6795
T2WI+CE-T1WI+ADC+SWI	SVM	0.892	DT	0.790	0.1657	0.2710	1.0000
T2WI+CE-T1WI+ADC+CBF	LR	0.881	DT	0.824	**0.0102**	0.3280	0.0584
T2WI+CE-T1WI+SWI+CBF	LDA	0.955	DT	0.807	0.1121	0.7520	0.8889
cMRI	LR	0.921	DT	0.841	0.2765	0.4145	**0.0077**
aMRI	AB	0.915	DT	0.784	0.1715	0.3243	0.0618
ALL	RF	0.969	DT	0.790	0.0640	–	–

Bold type indicate p < 0.05.

T2WI, T2-weighted imaging; T1WI, T1-weighted imaging; FLAIR, fluid-attenuated inversion recovery; CE-T1WI, contrast-enhanced T1WI; ADC, apparent diffusion coefficient; SWI, susceptibility-weighted imaging; CBV, cerebral blood volume; CBF, cerebral blood flow; cMRI, model developed with all of the conventional MRI; aMRI, model developed with all of the advanced MRI; ALL, model developed with all of the eight sequences; LR, logistic regression; LDA, linear discriminant analysis; SVM, support vector machine; AE, auto-encoder, DT, decision tree; RF, random forest; AB, AdaBoost; AUC, area under the curve.

[a]*p-Value is for the comparison between the best and worst classifiers in the same sequence.*

[b]*p-Value is for the comparison of the best classifiers between all sequence-based models (ALL) and other sequence-based models.*

[c]*p-Value is for the comparison between the best classifier of all sequence-based models (ALL) and the worst classifier of other sequence-based models.*

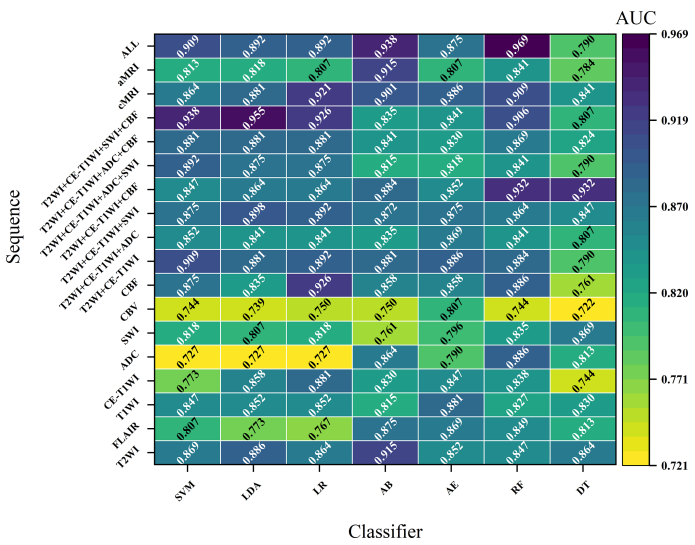


FIGURE 5 | The optimal performance across different sequences and classifiers.

test cohorts. To reduce the selection bias with this kind of splitting, nest cross-validation may be needed in the future. The third is the lack of extra validation to facilitate the generalization of our findings. Unlike other gliomas, the morbidity of DMG was lower. Furthermore, we analyzed eight MR image sets, which makes it more challenging to match an external validation cohort. Fourth, we did not compare our model with the human reader as recommended by a previous study (40). However, the performance of MRI features evaluated by radiologists with the non-radiomics method was controversial, and the discriminative ability was not as well as ours (highest AUC = 0.872) (28). A prior study showed that the radiomics model was significantly superior

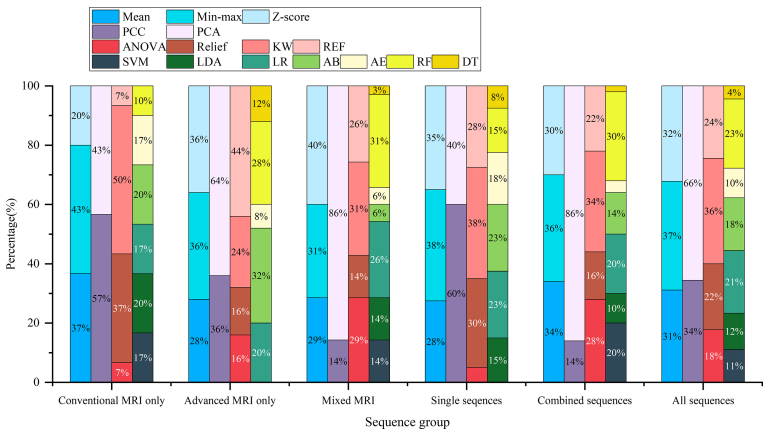


FIGURE 6 | The percentage of machine learning techniques in 90 top-five-performing models of different model categories. "Conventional MRI only" represents models developed only with conventional MRI sequences; "Advanced MRI only" for models only with advanced MRI sequences; "Mixed MRI" for models with both conventional and advanced MRI sequences; "Single sequences" for models with one sequence; "combined sequences" for models with at least two sequences; and "All sequences" for models of all sequence sets.

to the clinical model (based on radiological features) (16). In this regard, our radiomics model might be superior to human readers, although a head-to-head comparison needs to be implemented in the future. Finally, the performance of deep learning algorithms was not evaluated and compared in our study. Deep learning algorithms have been widely used in glioma molecular subtype prediction (41–45). However, deep learning usually needs a huge amount of dataset, such as hundreds or thousands of cases, and the dataset is limited for our approach. More datasets would be collected, and deep-learning algorithms would be compared to classical machine learning algorithms in the future.

CONCLUSION

Our results indicated that the H3 K27M mutant status of DMG can be effectively predicted with multiparametric MRI radiomics models. However, the performance of models varies significantly across different machine learning techniques and sequences used.

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 Ethical Committee of the First Affiliated Hospital of Fujian 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

WG: writing, study design, and data analysis. DS: study design and data analysis. ZX: data analysis. XL and FW: data collection. YS: model development and statistical analysis. DC: study design, data analysis, and revisions to the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This study has received funding from the Leading Project of the Department of Science and Technology of Fujian Province (No. 2020Y0025), the National Natural Science Foundation of China (No. 82071869), the Joint Funds of the Innovation of Science and Technology of Fujian Province (No. 2019Y9115), and the Young and Middle-aged Key Personnel Training Project of Fujian Provincial Health Commission (No.2021GGA025).

ACKNOWLEDGMENTS

The authors thank Xingfu Wang for pathological support at First Affiliated Hospital of Fujian Medical University, Fuzhou, China.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.796583/full#supplementary-material

REFERENCES

1. Louis DN, Perry A, Reifenberger G, von Deimling A, 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. Chiang J, Diaz AK, Makepeace L, Li X, Han Y, Li Y, et al. Clinical, Imaging, and Molecular Analysis of Pediatric Pontine Tumors Lacking Characteristic Imaging Features of DIPG. *Acta Neuropathol Commun* (2020) 8(1):57. doi: 10.1186/s40478-020-00930-9
3. Korshunov A, Ryzhova M, Hovestadt V, Bender S, Sturm D, Capper D, et al. Integrated Analysis of Pediatric Glioblastoma Reveals a Subset of Biologically Favorable Tumors With Associated Molecular Prognostic Markers. *Acta Neuropathol* (2015) 129(5):669–78. doi: 10.1007/s00401-015-1405-4
4. Karremann M, Gielen GH, Hoffmann M, Wiese M, Colditz N, Warmuth-Metz M, et al. Diffuse High-Grade Gliomas With H3 K27M Mutations Carry a Dismal Prognosis Independent of Tumor Location. *Neuro Oncol* (2018) 20 (1):123–31. doi: 10.1093/neuonc/nox149
5. Qiu T, Chanchotisatien A, Qin Z, Wu J, Du Z, Zhang X, et al. Imaging Characteristics of Adult H3 K27M-Mutant Gliomas. *J Neurosurg* (2019) 15:1–9. doi: 10.3171/2019.9.JNS191920
6. Graham MS, Mellinghoff IK. Histone-Mutant Glioma: Molecular Mechanisms, Preclinical Models, and Implications for Therapy. *Int J Mol Sci* (2020) 21(19):7193. doi: 10.3390/ijms21197193
7. Zhang Y, Zhou L, Safran H, Borsuk R, Lulla R, Tapinos N, et al. EZH2i EPZ-6438 and HDACi Vorinostat Synergize With ONC201/TIC10 to Activate Integrated Stress Response, DR5, Reduce H3K27 Methylation, ClpX and Promote Apoptosis of Multiple Tumor Types Including DIPG. *Neoplasia* (2021) 23(8):792–810. doi: 10.1016/j.neo.2021.06.007
8. Bailey CP, Figueroa M, Gangadharan A, Yang Y, Romero MM, Kennis BA, et al. Pharmacologic Inhibition of Lysine-Specific Demethylase 1 as a Therapeutic and Immune-Sensitization Strategy in Pediatric High-Grade Glioma. *Neuro Oncol* (2020) 22(9):1302–14. doi: 10.1093/neuonc/noaa058
9. Grasso CS, Tang Y, Truffaux N, Berlow NE, Liu L, Debily MA, et al. Functionally Defined Therapeutic Targets in Diffuse Intrinsic Pontine Glioma. *Nat Med* (2015) 21(6):555–9. doi: 10.1038/nm.3855
10. Pedersen H, Schmiegelow K, Hamerlik P. Radio-Resistance and DNA Repair in Pediatric Diffuse Midline Gliomas. *Cancers (Basel)* (2020) 12(10):2813. doi: 10.3390/cancers12102813
11. Liu J, Chen F, Pan C, Zhu M, Zhang X, Zhang L, et al. A Cascaded Deep Convolutional Neural Network for Joint Segmentation and Genotype Prediction of Brainstem Gliomas. *IEEE Trans BioMed Eng* (2018) 65 (9):1943–52. doi: 10.1109/TBME.2018.2845706
12. Pan CC, Liu J, Tang J, Chen X, Chen F, Wu YL, et al. A Machine Learning-Based Prediction Model of H3K27M Mutations in Brainstem Gliomas Using Conventional MRI and Clinical Features. *Radiother Oncol* (2019) 130:172–9. doi: 10.1016/j.radonc.2018.07.011
13. Su X, Chen N, Sun H, Liu Y, Yang X, Wang W, et al. Automated Machine Learning Based on Radiomics Features Predicts H3 K27M Mutation in Midline Gliomas of the Brain. *Neuro Oncol* (2020) 22(3):393–401. doi: 10.1093/neuonc/noz184
14. Kandemirli SG, Kocak B, Naganawa S, Ozturk K, Yip SSF, Chopra S, et al. Machine Learning-Based Multiparametric Magnetic Resonance Imaging Radiomics for Prediction of H3K27M Mutation in Midline Gliomas. *World Neurosurg* (2021) 151:e78–85. doi: 10.1016/j.wneu.2021.03.135
15. Zhuo Z, Qu L, Zhang P, Duan Y, Cheng D, Xu X, et al. Prediction of H3K27M-Mutant Brainstem Glioma by Amide Proton Transfer-Weighted Imaging and its Derived Radiomics. *Eur J Nucl Med Mol Imaging* (2021) 48 (13):4426–36. doi: 10.1007/s00259-021-05455-4
16. Wu C, Zheng H, Li J, Zhang Y, Duan S, Li Y, et al. MRI-Based Radiomics Signature and Clinical Factor for Predicting H3K27M Mutation in Pediatric High-Grade Gliomas Located in the Midline of the Brain. *Eur Radiol* (2021) 32(3):1813–22. doi: 10.1007/s00330-021-08234-9
17. Kim M, Jung SY, Park JE, Jo Y, Park SY, Nam SJ, et al. Diffusion- and Perfusion-Weighted MRI Radiomics Model may Predict Isocitrate Dehydrogenase (IDH) Mutation and Tumor Aggressiveness in Diffuse Lower Grade Glioma. *Eur Radiol* (2020) 30(4):2142–51. doi: 10.1007/s00330-019-06548-3
18. Tan Y, Zhang ST, Wei JW, Dong D, Wang XC, Yang GQ, et al. A Radiomics Nomogram may Improve the Prediction of IDH Genotype for Astrocytoma Before Surgery. *Eur Radiol* (2019) 29(7):3325–37. doi: 10.1007/s00330-019-06056-4
19. Manikis GC, Ioannidis GS, Siakallis L, Nikiforaki K, Iv M, Vozlic D, et al. Multicenter DSC-MRI-Based Radiomics Predict IDH Mutation in Gliomas. *Cancers (Basel)* (2021) 13(16):3965. doi: 10.3390/cancers13163965
20. Priya S, Liu Y, Ward C, Le NH, Soni N, Pillenahalli Maheshwarappa R, et al. Radiomic Based Machine Learning Performance for a Three Class Problem in Neuro-Oncology: Time to Test the Waters? *Cancers (Basel)* (2021) 13 (11):2568. doi: 10.3390/cancers13112568
21. Bathla G, Priya S, Liu Y, Ward C, Le NH, Soni N, et al. Radiomics-Based Differentiation Between Glioblastoma and Primary Central Nervous System Lymphoma: A Comparison of Diagnostic Performance Across Different MRI Sequences and Machine Learning Techniques. *Eur Radiol* (2021) 31 (11):8703–13. doi: 10.1007/s00330-021-07845-6
22. Chen C, Zheng A, Ou X, Wang J, Ma X. Comparison of Radiomics-Based Machine-Learning Classifiers in Diagnosis of Glioblastoma From Primary Central Nervous System Lymphoma. *Front Oncol* (2020) 10:1151. doi: 10.3389/fonc.2020.01151
23. Wang X, Wan Q, Chen H, Li Y, Li X. Classification of Pulmonary Lesion Based on Multiparametric MRI: Utility of Radiomics and Comparison of Machine Learning Methods. *Eur Radiol* (2020) 30(8):4595–605. doi: 10.1007/s00330-020-06768-y
24. 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
25. Gevaert O, Mitchell LA, Achrol AS, Xu J, Echegaray S, Steinberg GK, et al. Glioblastoma Multiforme: Exploratory Radiogenomic Analysis by Using Quantitative Image Features. *Radiology* (2014) 273:168. doi: 10.1148/radiol.14131731
26. Song Y, Zhang J, Zhang YD, Hou Y, Yan X, Wang Y, et al. FeAture Explorer (FAE): A Tool for Developing and Comparing Radiomics Models. *PloS One* (2020) 15(8):e0237587. doi: 10.1371/journal.pone.0237587
27. 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–e7. doi: 10.1158/0008-5472.CAN-17-0339
28. Chen H, Hu W, He H, Yang Y, Wen G, Lv X. Noninvasive Assessment of H3 K27M Mutational Status in Diffuse Midline Gliomas by Using Apparent Diffusion Coefficient Measurements. *Eur J Radiol* (2019) 114:152–9. doi: 10.1016/j.ejrad.2019.03.006
29. Piccardo A, Tortora D, Mascelli S, Severino M, Piatelli G, Consales A, et al. Advanced MR Imaging and (18)F-DOPA PET Characteristics of H3K27M-Mutant and Wild-Type Pediatric Diffuse Midline Gliomas. *Eur J Nucl Med Mol Imaging* (2019) 46(8):1685–94. doi: 10.1007/s00259-019-04333-4
30. Kathrani N, Chauhan RS, Kotwal A, Kulanthaivelu K, Bhat MD, Saini J, et al. Diffusion and Perfusion Imaging Biomarkers of H3 K27M Mutation Status in Diffuse Midline Gliomas. *Neuroradiology* (2022). doi: 10.1007/s00234-021-02857-x
31. Banan R, Akbarian A, Samii M, Samii A, Bertalanffy H, Lehmann U, et al. Diffuse Midline Gliomas, H3 K27M-Mutant are Associated With Less Peritumoral Edema and Contrast Enhancement in Comparison to Glioblastomas, H3 K27M-Wildtype of Midline Structures. *PloS One* (2021) 16(8):e0249647. doi: 10.1371/journal.pone.0249647
32. Chauhan RS, Kulanthaivelu K, Kathrani N, Kotwal A, Bhat MD, Saini J, et al. Prediction of H3K27M Mutation Status of Diffuse Midline Gliomas Using MRI Features. *J Neuroimaging* (2021) 31(6):1201–10. doi: 10.1111/jon.12905
33. Aboian MS, Solomon DA, Felton E, Mabray MC, Villanueva-Meyer JE, Mueller S, et al. Imaging Characteristics of Pediatric Diffuse Midline Gliomas With Histone H3 K27M Mutation. *AJNR Am J Neuroradiol* (2017) 38(4):795–800. doi: 10.3174/ajnr.A5076
34. Aboian MS, Tong E, Solomon DA, Kline C, Gautam A, Vardapetyan A, et al. Diffusion Characteristics of Pediatric Diffuse Midline Gliomas With Histone H3-K27M Mutation Using Apparent Diffusion Coefficient Histogram Analysis. *AJNR Am J Neuroradiol* (2019) 40(11):1804–10. doi: 10.3174/ajnr.A6302

35. Gillies RJ, Kinahan PE, Hricak HJR. Radiomics: Images Are More Than Pictures. *They Are Data* (2016) 278(2):563–77. doi: 10.1148/radiol.2015151169
36. 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
37. Peng H, Huo J, Li B, Cui Y, Zhang H, Zhang L, et al. Predicting Isocitrate Dehydrogenase (IDH) Mutation Status in Gliomas Using Multiparameter MRI Radiomics Features. *J Magn Reson Imaging* (2021) 53(5):1399–407. doi: 10.1002/jmri.27434
38. Song SE, Cho KR, Cho Y, Kim K, Jung SP, Seo BK, et al. Machine Learning With Multiparametric Breast MRI for Prediction of Ki-67 and Histologic Grade in Early-Stage Luminal Breast Cancer. *Eur Radiol* (2021) 32(2):853-63. doi: 10.1007/s00330-021-08127-x
39. Huang Y, Wei L, Hu Y, Shao N, Lin Y, He S, et al. Multi-Parametric MRI-Based Radiomics Models for Predicting Molecular Subtype and Androgen Receptor Expression in Breast Cancer. *Front Oncol* (2021) 11:706733. doi: 10.3389/fonc.2021.706733
40. Bluemke DA, Moy L, Bredella MA, Ertl-Wagner BB, Fowler KJ, Goh VJ, et al. Assessing Radiology Research on Artificial Intelligence: A Brief Guide for Authors, Reviewers, and Readers-From the Radiology Editorial Board. *Radiology* (2020) 294(3):487–9. doi: 10.1148/radiol.2019192515
41. Bangalore Yogananda CG, Shah BR, Vejdani-Jahromi M, Nalawade SS, Murugesan GK, Yu FF, et al. A Novel Fully Automated MRI-Based Deep-Learning Method for Classification of IDH Mutation Status in Brain Gliomas. *Neuro Oncol* (2020) 22(3):402–11. doi: 10.1093/neuonc/noz199
42. Choi YS, Bae S, Chang JH, Kang S-G, Kim SH, Kim J, et al. Fully Automated Hybrid Approach to Predict the IDH Mutation Status of Gliomas *via* Deep Learning and Radiomics. *Neuro-Oncology* (2021) 23(2):304–13. doi: 10.1093/neuonc/noaa177
43. Cluceru J, Interian Y, Phillips JJ, Molinaro AM, Luks TL, Alcaide-Leon P, et al. Improving the Noninvasive Classification of Glioma Genetic Subtype With Deep Learning and Diffusion-Weighted Imaging. *Neuro Oncol* (2021). doi: 10.1093/neuonc/noab238
44. Yogananda CGB, Shah BR, Nalawade SS, Murugesan GK, Yu FF, Pinho MC, et al. MRI-Based Deep-Learning Method for Determining Glioma MGMT Promoter Methylation Status. *AJNR Am J Neuroradiol* (2021) 42(5):845–52. doi: 10.3174/ajnr.A7029
45. Matsui Y, Maruyama T, Nitta M, Saito T, Tsuzuki S, Tamura M, et al. Prediction of Lower-Grade Glioma Molecular Subtypes Using Deep Learning. *J Neurooncol* (2020) 146(2):321–7. doi: 10.1007/s11060-019-03376-9

Conflict of Interest: Author YS was employed by Siemens Healthineers 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.

frontiers in Physics

ORIGINAL RESEARCH
published: 04 March 2022
doi: 10.3389/fphy.2022.815560

Automatic Detection of Osteochondral Lesions of the Talus *via* Deep Learning

Gang Wang[1], *Tiefeng Li*[2], *Lei Zhu*[1], *Siyuan Sun*[3], *Juntao Wang*[1], *Yidong Cui*[1], *Ben Liu*[1], *Yuliang Sun*[1], *Qingjia Xu*[1*†] and *Jianmin Li*[1*†]

[1]*Department of Orthopedics, Qilu Hospital, Shandong University, Jinan, China,* [2]*AI Research Group, Ebond (Beijing) Intelligence Technology Co., Ltd., Beijing, China,* [3]*AI Research Group, Yihui Ebond (Shandong) Medical Technology Co., Ltd., Jinan, China*

Screening of osteochondral lesions of the talus (OLTs) from MR imags usually requires time and efforts, and in most case lesions with small size are often missed in clinical practice. Thereore, it puts forward higher requirements for a more efficient OLTs screening method. To develop an automatic screening system for osteochondral lesions of the talus (OLTs), we collected 92 MRI images of patients with ankle pain from Qilu Hospital of Shandong University and proposed an AI (artificial intelligence) aided lesion screening system, which is automatic and based on deep learning method. A two-stage detection method based on the cascade R-CNN model was proposed to significantly improve the detection performance by taking advantage of multiple intersection-over-union thresholds. The backbone network was based on ResNet50, which was a state-of-art convolutional neural network model in image classification task. Multiple regression using cascaded detection heads was applied to further improve the detection precision. The mean average precision (mAP) that is adopted as major metrics in the paper and mean average recall (mAR) was selected to evaluate the performance of the model. Our proposed method has an average precision of 0.950, 0.975, and 0.550 for detecting the talus, gaps and lesions, respectively, and the mAP, mAR was 0.825, 0.930. Visualization of our network performance demonstrated the effectiveness of the model, which implied that accurate detection performance on these tasks could be further used in real clinical practice.

Keywords: osteochondral lesions, automatic diagnosis system, artificial intelligence, deep learning method, cascade R-CNN model

OPEN ACCESS

Edited by:
Daniel Rodriguez Gutierrez, Nottingham University Hospitals NHS Trust, United Kingdom

Reviewed by:
Wazir Muhammad, Yale University, United States
Xiaojun Yu, Northwestern Polytechnical University, China

****Correspondence:***
Qingjia Xu
18560082695@163.com
Jianmin Li
gkljm@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Medical Physics and Imaging, a section of the journal Frontiers in Physics

Received: *15 November 2021*
Accepted: *31 January 2022*
Published: *04 March 2022*

Citation:
Wang G, Li T, Zhu L, Sun S, Wang J, Cui Y, Liu B, Sun Y, Xu Q and Li J (2022) Automatic Detection of Osteochondral Lesions of the Talus via Deep Learning. Front. Phys. 10:815560. doi: 10.3389/fphy.2022.815560

1 INTRODUCTION

Osteochondral lesions of the talus (OLTs) represent a common disease that affects about 1.6 million people per year around the world [1]. The OLT term covers a spectrum of pathological conditions of articular cartilage and subchondral bone, with multiple treatment options [2, 3]. The proposed causes for OLT include acute traumatic insult, repetitive chronic microtrauma to the ankle joint and localized ischemia of the talus [2, 4]. For patients who did not benefit from nonoperative management, surgical treatment is indicated, depending on the size, location and chronicity of the lesion. Lesions less than 1 cm in diameter are associated with better outcomes and are amenable to arthroscopic bone marrow stimulation techniques, such as microfractures or subchondral drilling. For large OLTs with or without bone loss, osteochondral autograft or allograft transplantation may be performed. However, disadvantages also emerged such as donor pain, joint surface mismatching and gap of mosaic bone graft nonunion. Eventually, the patient might need talus replacement or

TABLE 1 | Demographics and clinical characteristics of patients.

Variables	Patients (n = 119)
Gender (n/%)	
Men	67/56.30
Women	52/43.70
Age (years)	49.84 ± 11.75

Data are expressed as numbers and percentages or as mean ± SD.

ankle replacement which would be a great burden not only for patients but also for health insurance. Therefore, early screening and intervention should be given sufficient emphasis.

The diagnosis of OLT requires acknowledging patient's comprehensive medical history, physical examination and radiography examination. MRI was one of the most effective method to evaluate OLTs due to its application of estimating the size of the lesion which has been accepted as a fundamental tool for OLT diagnosis nowadays. However, the MRI-based diagnosis procedure highly relies on the experience level of the radiologist, which dramatically introduces interobserver disagreements. Furthermore, up to 50% of OLTs may not be visualized on radiographs alone [5]. Subsequently, developing standardized computer-based methods to detect osteochondral lesions based on MRI would be beneficial to maximize the diagnostic performance while reducing the subjectivity, variability and errors due to distraction and fatigue that are associated with human interpretation.

Nowadays, deep learning (DL) [6] methods using convolutional neural networks (CNNs) have become a standard solution for automatic biomedical image analysis [7]. The use of these methods has been proven to be an efficient way to overcome the shortcomings of traditional image analysis on many sub-specialty applications. DL methods in medical image analysis have been applied in MRI tumor grading [8–10], thyroid nodule ultrasound classification [11–13] and CT pulmonary nodule detection [14–16]. However, only a limited number of studies have been performed to analyze the musculoskeletal imaging associated with the lesion. In 2018, Liu et al. [17] proposed a deep learning method to detect cartilage lesions within the knee joint on T2-weighted 2D fast spin-echo MRI images and achieved an area under the receiver operating characteristic (ROC) curve (AUC) of 0.92, with a sensitivity of 84% and a specificity of 85%. In 2019, Pedoia et al. [18] employed a U-net network to segment patellar cartilage using sagittal fat-suppression (FS) proton density-weighted 3D fast spin-echo (FSE) images and achieved an AUC of 0.88 for detecting cartilage lesions with the sensitivity and specificity both being 80%.

All these previous works [19–22] focused on independent training of the disease classification and risk region segmentation. However, this is not reasonable by nature due to the association between the risk region and the disease possibility. Besides, as a common injury, early screening of OLT should be given priority. In this study, we propose an automatic OLT screening method based on multi-task deep learning, which could simultaneously provide the evidence of the disease and detecting risk area.

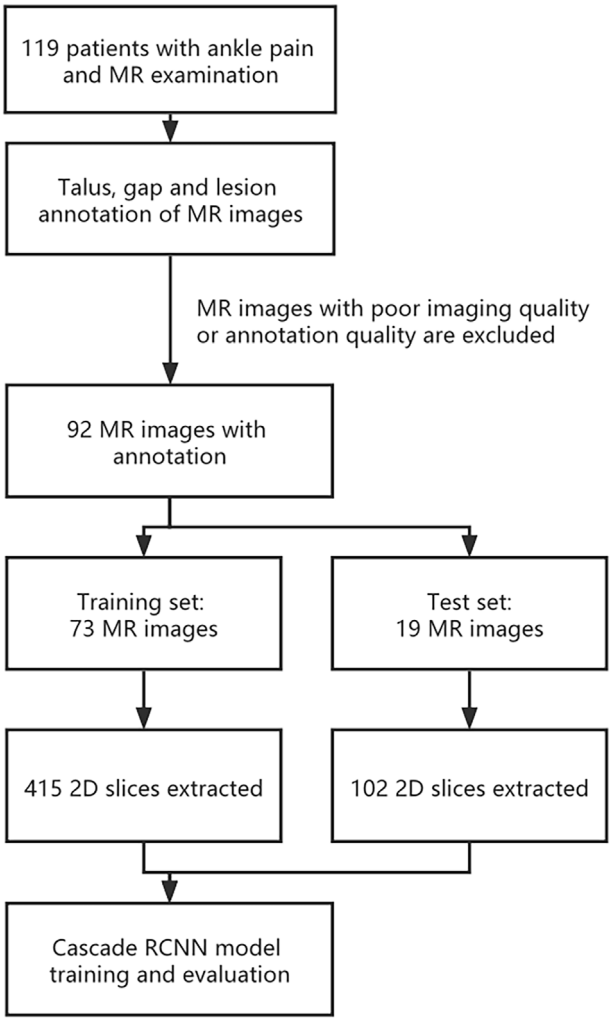


FIGURE 1 | Flowchart of OTL detection using deep learning method.

2 MATERIALS AND METHODS

2.1 Dataset Preparation and Preprocessing

This study was performed in compliance with the Health Insurance Portability and Accountability Act regulations, with approval from our institutional review board. Due to the retrospective nature of the study, informed consent was waived. MRI data of 119 patients were recruited into this study. Inclusion criteria: ①All the patients whose main diagnosis was OLT; ② main complaint is ankle pain; ③ First time for medical consultation in our hospital without surgery. Exclusion criteria: ①infection in ankle; ②tumor in talus; ③ MRI data was missing. ④ Poor image quality or poor annotation quality. MRI datasets were obtained from 119 patients with ankle pain (67 men and 52 women, with an average age of 49.84 years and an age range of 24–71 years) who underwent a clinical MRI examination of the ankle at our institution between 15 January 2017, and 15 October 2020 (as shown in **Table 1**). 27 patients were excluded due to poor image

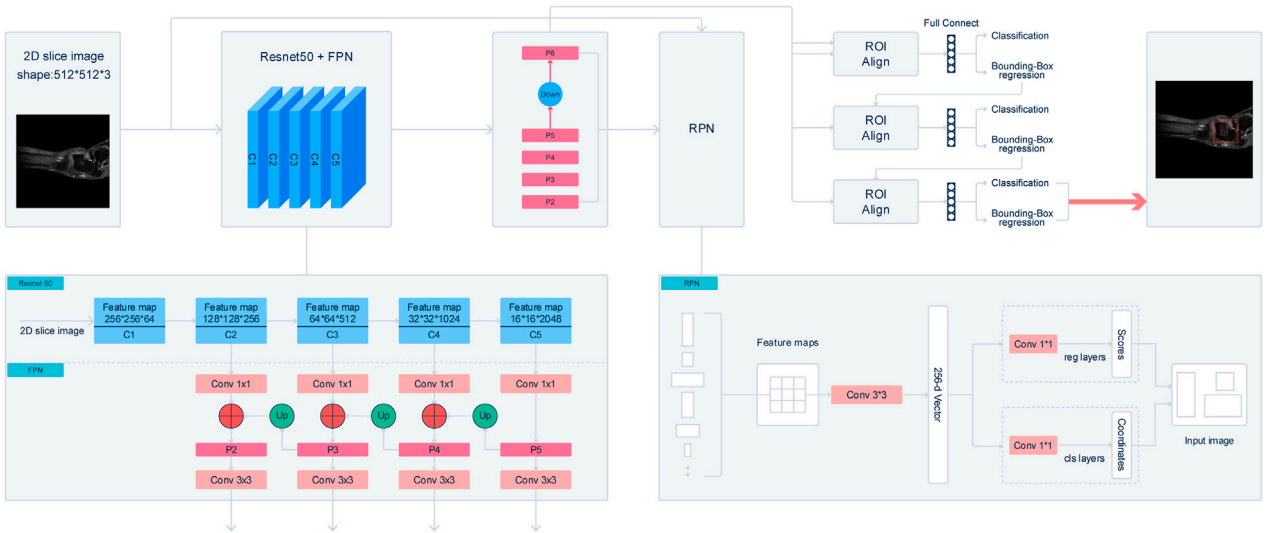


FIGURE 2 | Illustraion of Cascade R-CNN architecture. The backbone network extract feature map from 2D slice images. A Residual Neural Network (ResNet50) with feature pyramid network (FPN) is utilized as the backbone of the Cascade-R-CNN. After processing by FPN, a 3*3 convolution layer is added after merging feature maps to get the feature maps, and it is called P2, P3, P4, P5 respectively. In addition, the feature map that is called P6 is got by sub-sampling P5 by a factor 2. The final feature maps are composed of P2, P3, P4, P5, and P6. A window slide over feature maps of multi-scale to obtain a set of predefined bounding boxes that is also called anchors. Anchors is 8* 8, 16*16, 32*32, 64*64, and 128*128 respectively. The feature maps are mapped to a 256-dimensional vector by a 3*3 convolution layer. At last, the anchors that have Intersection-over-Union (IOU) ratio of the anchors and ground-truth boxes, which is greater than 0.5 are selected as proposals boxes. In the third part, the feature map of any box's proposal is transformed into fixed size by the RoIAlign algorithm. The feature map is fed into 2 concatenated fully connected layers followed by 2 parallel fully connected layers, one of which is used for classification and the other for bounding-box regression.

TABLE 2 | Quantitative prediction result analysis.

Prediction task	Average precision
Talus	0.950
Gaps	0.975
Risk Region	0.550

quality or poor annotation quality, resulting in 92 patients finally included in this research. The same 1.5-T MRI unit (Lian Ying, uMR560) and eight-channel phased-array extremity coil were used for all patients. The MRI datasets consisted of coronal FS proton density-weighted FSE sequence. All the lesion parts were determined by hand drawing in order to facilitate the AI to learn.

Before model training, images were preprocessed as following: Firstly, MR dataset was divided into training dataset and validation dataset at a ratio of 8:2 at the patient level. As a result, training dataset contains 73 3D MR images and validation dataset contains 19 3D MR images. Secondly, in order to expand the sample size, a total 517 2D slices from 92 MR images with lesions manually segmented by specialists were used, and the size of each slice is about 320*320 (Specific may be slightly different). As a result, our training dataset contained 415 2D slice images and validation dataset contained 102 2D slice images. All 2D slice images were then intensity-normalized to a range of 0–255. Thirdly, to increase the size of the targets in the images, all 2D slice images are resized 512*512 with ratio kept. Fourthly, to generate 2D object ground-truth bounding boxes, an approach to get the maximum bounding box of the mask is implemented by using OpenCV library. Finally, the dataset was reorganized into COCO format. The whole pipeline is shown in **Figure 1**.

2.2 Cascade R-CNN Model

Taking the importance of locate target precise into consideration, we introduced a cascade method to address the problem. Due to the mechanism of exploiting the cascade information across multiple cascade layers, cascade learning could refine the object detection result and make the data distribution of inference closer to the training and hence is efficient in the scenarios that target is difficult to locate. In our study, we adopted the Cascade-R-CNN as the original object detection framework.

The improved Cascade-R-CNN model contains 3 parts: backbone network, region proposal network and detector. In the first part, the backbone network of the Cascade-R-CNN is used to extract feature map from 2D slice images. To improve the detection performance of the Cascade-R-CNN on muti-scale Object, a Residual Neural Network (ResNet50) with feature pyramid network (FPN) is utilized as the backbone of the Cascade-R-CNN. The ResNet50 consist of 5 stages, and it is called C1, C2, C3, C4, and C5 respectively. The FPN combined with the ResNet50 could extract feature maps from muti-stages of the backbone, and the size of feature map is getting smaller from the C2 to C5 and meanwhile the receptive field of each pixel of the feature map is getting larger. FPN is composed of 3 parts, and it is bottom-up pathway, top-down path and lateral connections respectively. The bottom-up pathway is the feed-forward propagation of the ResNet50. The feature maps from different stages are employed to form the feature maps of different size that is also called pyramid

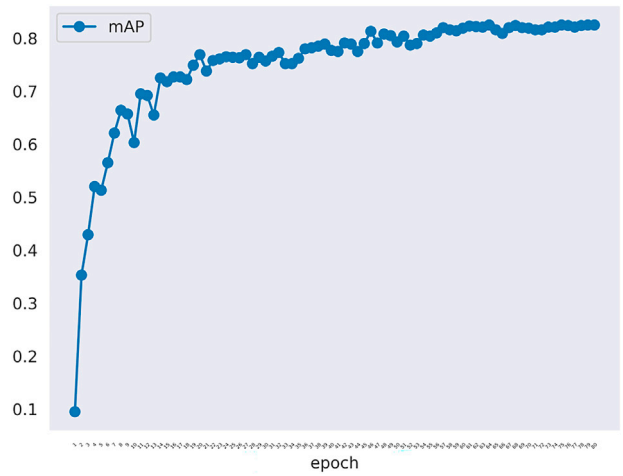


FIGURE 3 | The mAP curve in training model along with epoch. The mAP curve rises steadily alongwith the rounds of epoch.

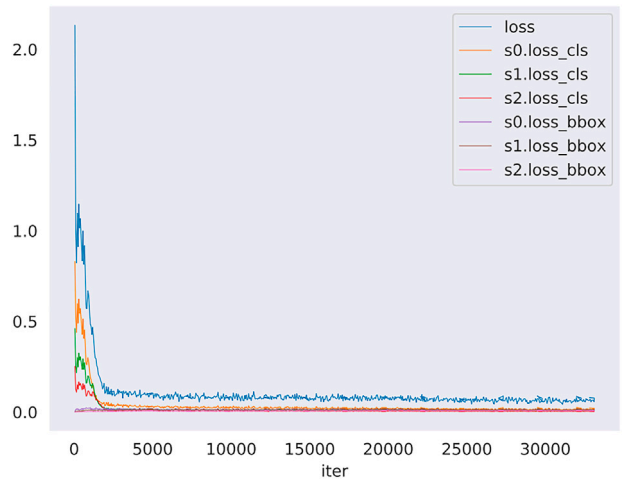


FIGURE 4 | The loss curve in training model along with iters. The loss denotes the sum of all loss of the model, moreover s0, s1, s2 denote stage1, stage2 and stage3 of the model detector head, in addition loss_cls and loss_bbox denote classification loss and bounding box begression loss. The blue curve is the summary of the other six curves.

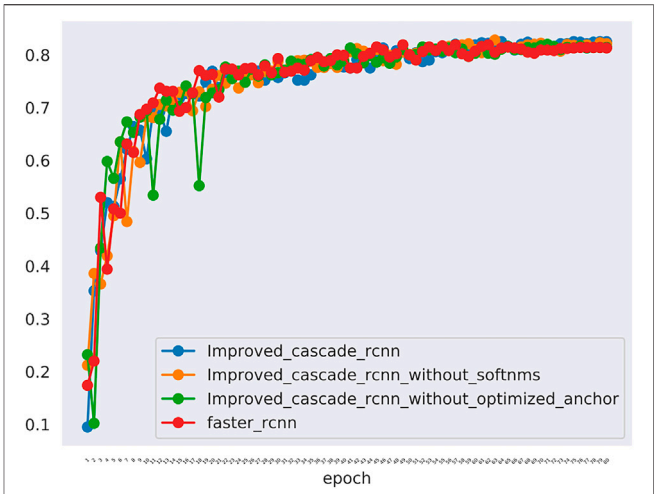


FIGURE 5 | The mAP of different models including the improved cascade rcnn, the improved cascade rcnn without soft-nms, the improved cascade rcnn without optimized anchor size. The blue curve indicates the improved cascade rcnn. The orange curve indicates the improved cascade rcnn without soft-nms. The green curve indicates the improved cascade rcnn without optimized anchor size. The red curve indicates the faster rcnn.

levels. In the top-down pathway, the feature maps have stronger semantic information, but the detailed information is coarser. The feature maps of higher stages are up-sampled by a factor 2 to acquire larger size feature maps. Then, to make the information of the feature maps more complete, the feature maps from top-down pathway are added in element-wise with the feature maps of the same spatial size from the bottom-up pathway by lateral connections to get merged feature maps. Finally, a 3*3 convolution layer is added after merging feature maps to get the feature maps, and it is called P2, P3, P4, P5 respectively. In addition, the feature map that is called P6 is got by sub-sampling P5 by a factor 2. The final feature maps are composed of P2, P3, P4, P5, and P6. In the start of training model, the backbone of the Cascade-R-CNN is initialized with a pre-trained model that was trained on Imagenet-1k.

In the second part, the feature maps from the multi-stages of the backbone network is used as the input of the region proposal networks to get region proposals bounding boxes that may contain lesion, gap or talus. A window slide over feature maps of multi-scale to obtain a set of predefined bounding boxes that is also called anchors. Because of the difference of size of the receptive field across multi-stages, the size of anchors is 8*8, 16*16, 32*32, 64*64 and 128*128 respectively. In addition, to improve Cascade-R-CNN generalization performance, anchors that have multiple aspect ratios 1:2, 1:1 and 2:1 is applied at each feature map from multi-stages. As a result, each sliding window on the feature maps can get 5*3 anchors that have different size simultaneously. Then the feature maps which is contained by anchors corresponding sliding window are mapped to a 256-dimensional vector by a 3*3 convolution layer. The vector is used as the input of following 2 parallel 1*1 convolution layers to obtain the results of bounding box regression and binary classification respectively, and bounding box regression is used to get the coordinates of the boxes proposal and binary classification is used to determine whether the box contains an object. Therefore, the size of outputs of bounding box regression is 4*15 and the size of outputs of binary classification is 2*15. At last, the anchors that have Intersection-over-Union (IOU) ratio of the anchors and ground-truth boxes, which is greater than 0.5 are selected as proposals boxes.

In the third part, the feature map of any box's proposal is transformed into fixed size by the RoIAlign algorithm. Then following 3 cascade layers that each contains 2 branches is constructed for classification and bounding-box regression. For each 2 branches of first 2 cascade layers, the feature map is fed into 2 concatenated fully connected layers followed by 2 parallel fully connected layers, one of which is used for classification and the

TABLE 3 | The comparison of experimental results of optimization method.

	mAP in all	mAP in small	mAP in medium	mAP in large	mAR in all	mAR in small	mAR in medium	mAR in large
Improved cascade rcnn	0.825	0.225	0.894	1.000	0.930	0.325	0.992	1.000
Above without softnms	0.828	0.223	0.916	1.000	0.908	0.304	0.968	1.000
Above without optimized anchor	0.820	0.225	0.876	1.000	0.897	0.289	0.968	1.000
Faster rcnn	0.819	0.211	0.904	1.000	0.884	0.271	0.968	1.000

TABLE 4 | The comparison of class-wise AP of optimization method.

	Talus AP	Gap AP	Lesion AP
Improved cascade rcnn	0.950	0.975	0.550
Above without softnms	0.947	0.980	0.557
Above without optimized anchor	0.950	0.980	0.531
Faster rcnn	0.950	0.980	0.526

other for bounding-box regression. The outputs of the classification layer separately are probably different predictions, and the regression layer is used to refine bounding-box positions for 4 subgroups and be the input of next layer. The last layer adopts the previous layer outputs of bounding-box regression and the feature map as input, and the outputs of the 2 branches of the last layers is the model last results. The detail of the structure of cascade R-CNN is shown in **Figure 2**.

2.3 Training Details

The network was implemented using PyTorch and trained on one RTX3090 with 24 GB memory. The anchor size of the model is optimized to the sizes [8*8, 16*16, 32*32, 64*64, 128*128], The optimizer for the network was set as SGD (stochastic gradient descent) and the initial learning rate was 0.0125 with a momentum of 0.9. Furthermore, we sat one weight decay of 0.0001 to help with the training stability. To achieve a better model performance, we employed the warm-up training strategy in the training procedure, using an epoch iteration of 5 with an increasing ratio of 0.01. We decreased the learning rate with CosineAnnealing decay strategy and make the learning rate gradually decay from start to finish to fine-tune the final model. Previous research has shown that soft non-maximum suppression (soft-nms) could achieve superior performance on such tasks due to the softer conditions for filtering out boxes. Thus, we implemented soft non-maximum suppression in the inference stage to improve the sensitivity of the model. The property could further help to achieve a good performance of our model. In training, we use online data augmentation that image scale ranges from 0.9 to 1.2 and rotate range from −90 to 90. The total training took 80 epochs to achieve stable convergence results.

2.4 Model Evaluation

In this study, we employed the mean average precision (mAP) as the parameter to evaluate the performance of the proposed model. AP is the area under the precision-recall curve, defined as follows:

$$AP = \int_0^1 P(r)dr$$

where P(r) is the precision-recall curve, and r is the IoU threshold. Then, mAP corresponds to the mean value of AP for multiple-class detection, defined as follows:

$$mAP = \frac{1}{N}\sum AP_i$$

AR is two times the area under the recall-IoU curve and it reflects the sensitivity of the model to the target, however in the paper AR is only computed in the case that IoU = 0.5, defined as follows:

$$\mathrm{AR} = \mathrm{Recall}\,(\mathrm{IoU}),\ \mathrm{IoU} = 0.5$$

Then, mAR corresponds to the mean value of AR for multiple-class detection, defined as follows:

$$\mathbf{mAR} = \frac{1}{\mathbf{N}}\sum \mathbf{AR_i}$$

3 RESULTS

Our proposed method for automatic OLT screening based on Cascade R-CNN showed a good performance in predicting the possibility of the disease and detecting the risk areas. Here we detected the osteochondral lesions (Risk Region), the whole talus

TABLE 5 | The comparison of experimental results in different conditions.

	mAP in all	mAP in small	mAP in medium	mAP in large	mAR in all	mAR in small	mAR in medium	mAR in large
Improved cascade rcnn	0.825	0.225	0.894	1.000	0.930	0.325	0.992	1.000
Above without rotate of data	0.804	0.216	0.861	1.000	0.885	0.280	0.957	1.000
Above with epcoh 24	0.794	0.186	0.887	1.000	0.921	0.313	0.992	1.000

TABLE 6 | The comparison of class-wise AP in different conditions.

	Talus AP	Gap AP	Lesion AP
Improved cascade rcnn	0.950	0.975	0.550
Above without rotate of data	0.934	0.960	0.519
Above with epcoh 24	0.942	0.972	0.469

(Talus) and the gaps (Gaps) at the same time. The AP of predicting Talus, Gaps and Risk Region reached 0.950, 0.975 and 0.550 respectively (**Table 2**), and the mAP, mAR was 0.825, 0.930 respectively.

The quantitative results of our model were shown in **Table 2**. All these parameters indicate that our model achieved an accurate detection performance, suggesting that it could be used for real clinical applications.

The changes of mAP in training model was shown in **Figure 3**. There are 414 graphs in the training set. Learning the 414 graphs once means an epoch. One mAP value is obtained when each epoch is verified using the verification set once. We tested epoch for 80 times. When the epoch reached the 80th time, the mAP value was approaching 1.0. **Figure 3** showed that the rising trend is relatively stable without much fluctuation indicating that the model has been approaching the most optimal solution, and there has been no overfitting phenomenon. In fact, the effect of learning is not becoming better alongwith more rounds of verification because overfitting may occur. In this experiment, because the number of samples is not large, we set the batchsize to 1 to maintain the largest difference of data to resist overfitting.

The loss of the training model was shown in **Figure 4**. The loss denote the sum of all loss of the model. The s0, s1, s2 denote stage1, stage2 and stage3 of the model detector head, in addition loss_cls and loss_bbox denote classification loss and bounding box begression loss. This figure shows the change of loss function in the training model. The blue curve is the sum of the other six curves. So, the blue curve represents the overall loss trend of the model. It can be seen that the curve shows a downward trend. It suggested that less and less information is lost in the training model meaning that the model training is effective. In sumary, we can know that the model already learned something and converged. In addition, model got the nice performance on the validation set.

The comparative data between the improved cascade RCNN, which without soft-nms, which without optimized anchor size and faster rcnn was shown in **Figure 5** and **Tables 3**, **4**. Small, medium and large denote area of object in range less than 36, 36–96 and greater than 96 repectively. In **Figure 5**, improved cascade RCNN reach higher value than other models in the last several epochs. In **Table 3**, the "mAP in all" of "improved cascade RCNN without soft-nms" is highest (0.828). The "mAP in small" of both "improved cascade RCNN" and "improved cascade RCNN without optimized anchor" is heghest (0.225). The "mAP in all" of "improved cascade RCNN without soft-nms" is highest (0.916). However, each mAR categories of "improved cascade RCNN" is highest (0.930, 0.325 and 0.992). In **Table 4**,

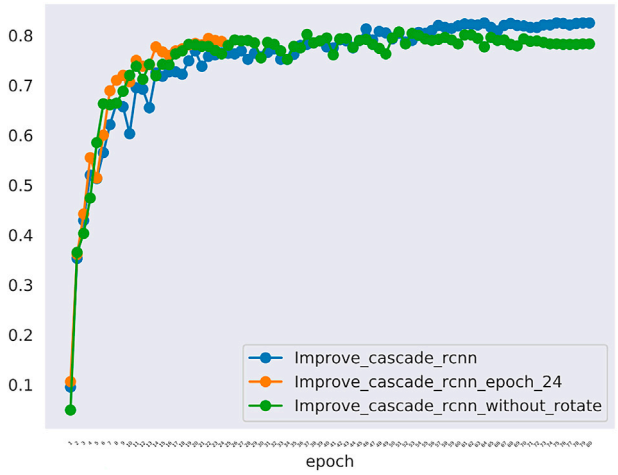


FIGURE 6 | The mAP of different models including the improved_cascade_rcnn, the improved_cascade_rcnn_epoch_24, the improved_cascade_rcnn_without_rotate. The blue curve indicates the improved_cascade_rcnn. The orange curve indicates the improved_cascade_rcnn_epoch_24. The green curve indicates the improve_cascade_rcnn without_rotate.

talus AP of "improved cascade RCNN" is highest (0.950). Meanwhile, the gap AP and lession AP of "improved cascade RCNN without soft-nms" is highest (0.980 and 0.557). The results showed that the best mAP appeared in training of improved cascade rcnn without softnms, which is higher than the improved cascade rcnn slightly. However, improved cascade rcnn has higher mAR. According to practical application scenarios of the model, the model sensitivity regarding to lesion is of great importance as well. Under the condition that mAP of former is higer than the latter slightly, we can draw a conclusion that the performance of improved cascade rcnn is better than improved cascade rcnn without softnms. In addition, the performance of improved cascade rcnn is better than another model except improved cascade rcnn without softnms.

Furthermore, In the method we use, there are two changes that may have a greater impact on the results. One is the random rotation of data enhancement during training, and the other is the number of epochs. we also compared the mAP and mAR of different changes. In **Table 5**, all the mAP and mAR categories of "improved cascade RCNN" reached the highest value (0.825, 0.225, 0.894, 0.930, 0.325, 0.992). In **Table 6**, both talus AP, gap AP and lession AP of 'improved cascade RCNN' is highest (0.950, 0.975 and 0.550). Combined with **Tables 5**, **6**, it can be seen from the experiment that the final result of the blue line (i.e. the method we use) is significantly higher than the other two changes (**Figure 6**). The epoch number 24, is the number of rounds commonly used from imagenet-1k data which was used to get baseline value for training.

From **Figures 7–9**, we could intuitively observe that our network provided a comparable detection result (red bounding boxes) with that of senior radiologists (blue bounding boxes). Noticeably, our model achieved stable detection results on all three detection tasks, which shows the effectiveness of our

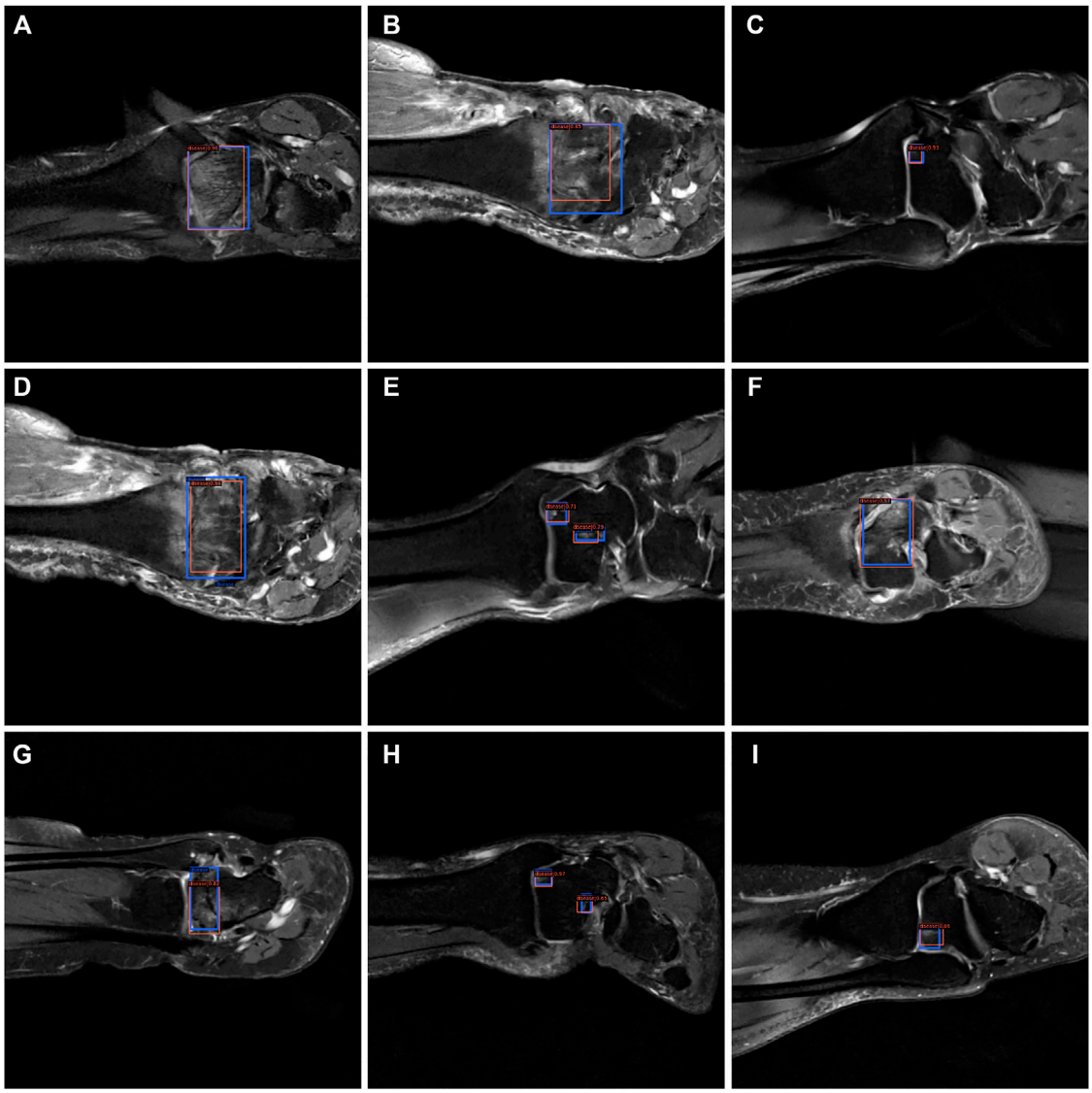


FIGURE 7 | The bounding box visualization. Prediction result (red) vs. the ground truth (blue) of the osteochondral lesions. All the images **(A–I)** were selected from testing group to determine the AI recognition accuracy of ROI. AI detection result was labeled by red rectangular. The ground truth was manually depicted and labeled by blue rectangular.

method. More examples are shown in the **Supplementary Figures S1–S3**.

4 DISCUSSION

For patients who did not benefit from nonoperative management, surgical treatment is indicated, depending on the size, location and chronicity of the lesion. Lesions less than 1 cm in diameter are associated with better outcomes and are amenable to arthroscopic bone marrow stimulation techniques, such as microfractures or subchondral drilling. Autologous chondrocyte implantation (ACI) is indicated for lesions larger than 1 cm in diameter, but it requires 2 stages. Arthroscopy or arthrotomy may be used for the second-stage implantation. For large OLTs with or without bone loss, osteochondral autograft or allograft transplantation may be performed which will cause greater damage and economic burden to patients. Therefore, early screening and early intervention are particularly important.

The initial evaluation of OLTs includes standard radiographs of the ankle and MRI. CT scans are also useful as an adjunct to MRI when evaluating subchondral cysts. Affected by doctors' experience, the diagnostic accuracy of the same MRI among doctors at various levels is different, which will lead to escaping diagnosis and misdiagnosis. If we can interpret MRI through artificial intelligence, it will greatly improve the diagnostic efficiency and accuracy of patients and reduce the error caused by human factors. The application of artificial

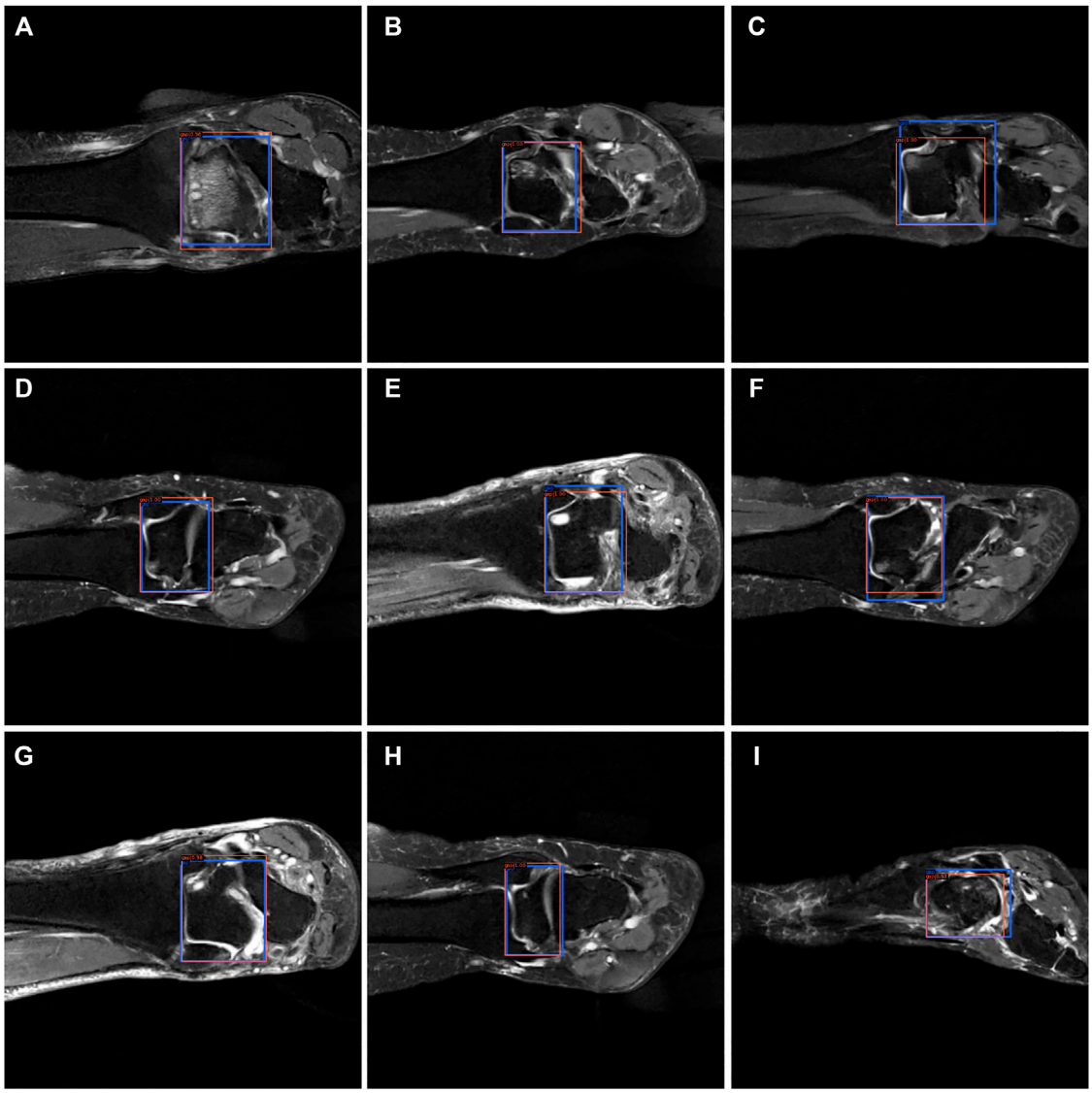


FIGURE 8 | The bounding box visualization. Prediction result (red) vs. the ground truth (blue) of the talus gaps. All the images **(A–I)** were selected from testing group to determine the AI recognition accuracy of talus gap. AI detection result was labeled by red rectangular. The ground truth was manually depicted and labeled by blue rectangular.

intelligence in the field of medical imaging is gradually increasing. In foreign countries, it is mainly divided into two parts: one is image recognition, which is applied to perception whose main purpose is to analyze image data, ROI (region of interest) of image acquisition. The second part is deep learning. Applying deep learning to the studying and analysis is the core segment of AI. Continuous training and deep learning of neural networks through a large number of image data and diagnostic data could achieve a diagnostic model which could enable the AI to master the ability of "diagnosis." AI will greatly reduce the workload of doctors if better accuracy and specificity can be achieved.

Currently, there are two typical methods to implement the detection project, including the one-stage and two-stage detection methods [23]. Compared with the one-stage method, the two-stage method could achieve a higher detection performance, but at the expense of speed [24]. In order to achieve a higher detection performance, we utilized the two-stage detection methods in this work. Many methods have been proposed in two-stage detection fields, including Mask-R-CNN [25], Fast-R-CNN [26] and Faster-R-CNN [27]. For all the two-stage detection methods, one intersection-over-union (IoU) threshold is required to classify the predicted positive bounding box with the negative bounding box. However, it has been shown that a lower IoU threshold could provide more bounding boxes with a lower precision, which induces a lower recall rate, while a higher IoU threshold could provide fewer bounding boxes with a higher precision, which induces the under detection [28]. To

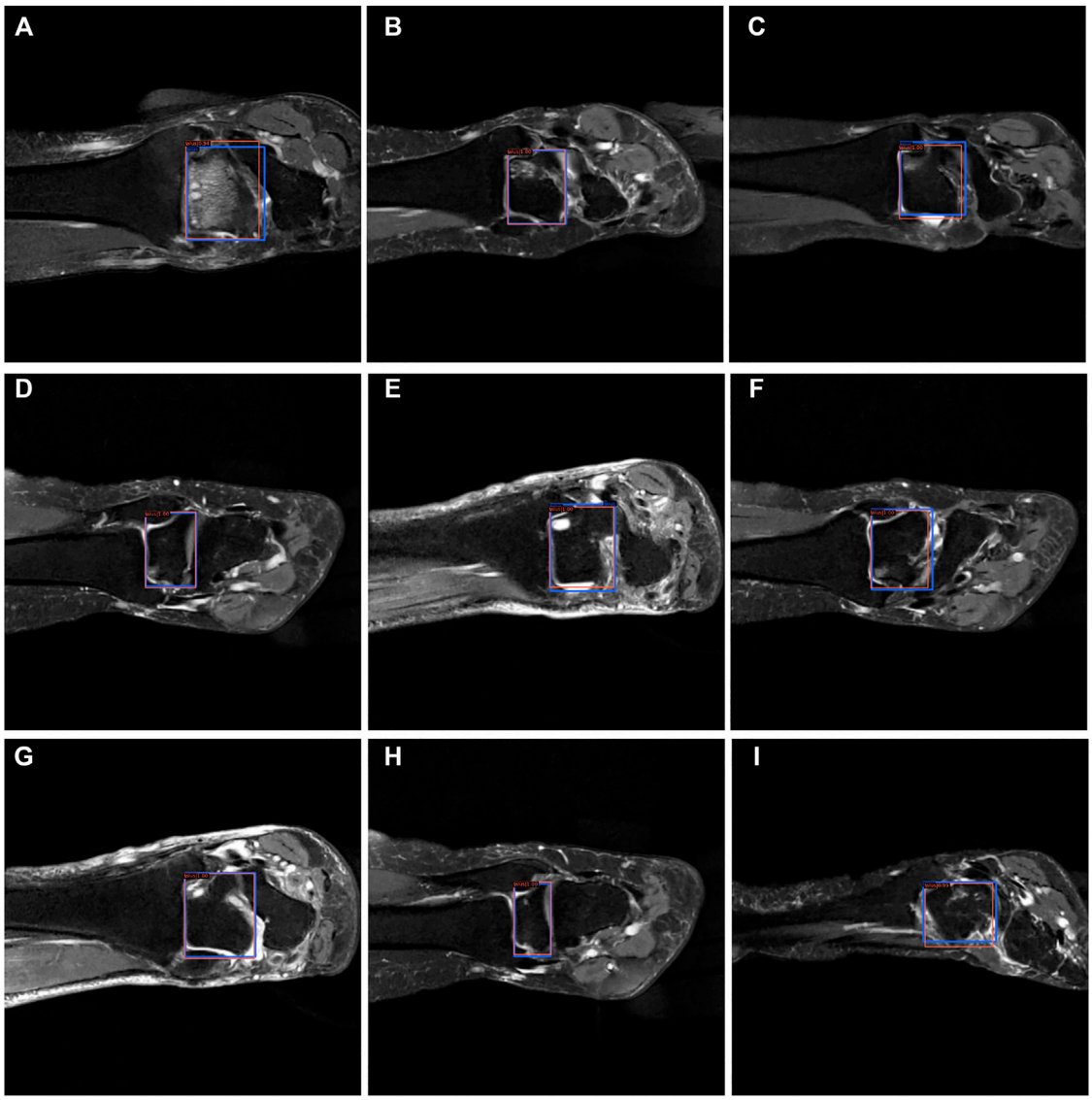


FIGURE 9 | The bounding box visualization. Prediction result (red) vs. the ground truth (blue) of the talus itself. All the images **(A–I)** were selected from testing group to determine the AI recognition accuracy of talus itself. AI detection result was labeled by red rectangular. The ground truth was manually depicted and labeled by blue rectangular.

clarify the concept of these statistical terms, the definitions of accuracy, precision and recall rate should be well explained. The predicted condition is usually marked with Positive or Negative. The actual condition is usually labeled with True or False. Subsequently, we have four parts in the contingency table: True positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN). By the use of multiple IoU thresholds, the recently proposed two-stage detection method of the Cascade R-CNN model significantly improved the detection performance compared with the above-mentioned two-stage detection methods. Besides, through multiple regression using cascaded detection heads, the Cascade R-CNN model could further improve the detection precision. The backbone network was built using a ResNet-50 network with pre-trained network weights from the ImageNet dataset, which allowed to fasten the training of the network and improve the final performance.

Our study implemented a deep learning-based method for the automatic detection of osteochondral lesions of the talus for the first time. Our model has achieved an accurate detection with an average, mAP of 0.550, 0.975, and 0.950 on the Risk Region, Gaps and Talus detection tasks, respectively. Compared with other detection tasks, the mIOU reached a similar value to that in a previous study on the detection of coronavirus pneumonia [29] (mIOU, 73.40% ± 2.24%). Furthermore, our model could also provide accurate detection results on multiple risk region parts in one single case without missing any of them. All these results prove that this method could be used by clinical radiologists to

overcome the shortness of subjectivity and variability and save the physician's valuable time.

5 CONCLUSION

Our research developed a diagnostic model for image interpretation based on artificial intelligence. The detection accuracy of this mode reaches mAP = 0.825. It provides a theoretical basis for the early diagnosis and screening of OLTs based on artificial intelligence detection in the future.

6 LIMITATION

Although many valuable results have been achieved in this work, there are still some limitations that need to be improved in the future. Firstly, only the articular cartilage on the talar dome was evaluated in our feasibility study, since evaluating the curved articular surface of the talus on the 3.5-mm-thick coronal fat-suppressed proton density-weighted FSE sequence would be challenging. Furthermore, most patients only accept medical consultation and MRI examination at outpatient instead of inpatient. They have no chance to get arthroscopy to make a definite diagnosis. As a result, the presence or absence of cartilage lesions in each image patch was interpreted by a musculoskeletal radiologist. Although arthroscopy has a higher sensitivity for detecting cartilage lesions, arthroscopy was unable to be used as a reference standard in this retrospective study. Meanwhile, only the highest grade of cartilage lesion on each articular surface was recorded in surgery report. However, the exact location of the cartilage lesion was not well described. Even though we extracted 517 slices from 92 patients to implement a deep learning model in this study, which was sufficient enough to conduct a detection task, we could have tried some data augmentation methods to make our model more robust and avoid overfitting greatly. Additionally, having an accurate segmentation map is more desired in many clinical applications to provide a precise treatment plan. However, due to the low resolution of the lesion parts, it was hard for the radiologists to acquire accurate boundaries of the risk region. In future studies, it could be beneficial to include the unsupervised machine learning methods for segmentation tasks in the OLT image analysis. Finally, although we performed the lesions, gaps and talus detection tasks at the same time, some information that was hidden within this structure was still not fully explored and the interpretability of our model was still unknown. In this research, we also conduct a detection task, but in clinical practice surgeons may be focus on the characteristics of lesion (XXX for instance) We believe we could further improve the performance of our model by combining the clinical information with the current detection methods and use a multi-task learning method to perform detection, segmentation, and lesion classification task simultaneously.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

QX and JL contributed to the original idea and conceptual design; GW, LZ, and JW conceived the experiments and analyzed data; GW and YC contributed to the drafting of the work; BL, YS, TL and SS provided critical review of the article.

ACKNOWLEDGMENTS

The authors would like to thank EditSprings (www.editsprings.cn), for editing the English text of a draft of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2022.815560/full#supplementary-material

REFERENCES

1. Murawski C.D., Kennedy J.G. Operative Treatment of Osteochondral Lesions of the Talus. *J Bone Jt Surg Am* (2013) 95:1045–54. doi:10.2106/jbjs.L.00773
2. Zengerink M, Struijs P.A.A, Tol J.L, van Dijk C.N. Treatment of Osteochondral Lesions of the Talus: A Systematic Review. *Knee Surg Sports Traumatol Arthrosc* (2010) 18:238–46. doi:10.1007/s00167-009-0942-6
3. Kraeutler M.J., Chahla J, Dean C.S., Mitchell J.J., Santini-Araujo MG, Pinney SJ, et al. Current Concepts Review Update. *Foot Ankle Int* (2017) 38:331–42. doi:10.1177/1071100716677746
4. O'Loughlin P.F., Heyworth B.E., Kennedy J.G. Current Concepts in the Diagnosis and Treatment of Osteochondral Lesions of the Ankle. *Am J Sports Med* (2010) 38:392–404. doi:10.1177/0363546509336336
5. Verhagen R.A.W., Maas M, Dijkgraaf M.G.W, Tol J.L., Krips R, van Dijk C.N. Prospective Study on Diagnostic Strategies in Osteochondral Lesions of the Talus. *The J Bone Jt Surg Br volume* (2005) 87-B:41–6. doi:10.1302/0301-620x.87b1.14702
6. LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature* (2015) 521:436–44. doi:10.1038/nature14539
7. Litjens G, Kooi T, Bejnordi B.E., Setio A.A.A, Ciompi F, Ghafoorian M, et al. A Survey on Deep Learning in Medical Image Analysis. *Med Image Anal* (2017) 42:60–88. doi:10.1016/j.media.2017.07.005
8. Ning Z, Tu C, Di X, Feng Q, Zhang Y. Deep Cross-View Co-Regularized Representation Learning for Glioma Subtype Identification. *Med Image Anal* (2021) 73:102160. doi:10.1016/j.media.2021.102160
9. Peng J, Kim D.D., Patel J.B., Zeng X, Huang J, Chang K, et al. Deep Learning-Based Automatic Tumor Burden Assessment of Pediatric High-Grade Gliomas, Medulloblastomas, and Other Leptomeningeal Seeding Tumors. *Neuro Oncol* (2021) 24:289–99. doi:10.1093/neuonc/noab151
10. Gutman D.C., Young R.J. IDH Glioma Radiogenomics in the Era of Deep Learning. *Neuro Oncol* (2021) 23:182–3. doi:10.1093/neuonc/noaa294

11. Zhang B, Jin Z, Zhang S. A Deep-Learning Model to Assist Thyroid Nodule Diagnosis and Management. *The Lancet Digital Health* (2021) 3:e410. doi:10.1016/s2589-7500(21)00108-4
12. Wu G-G, Lv W-Z, Yin R, Xu J-W, Yan Y-J, Chen R-X, et al. Deep Learning Based on ACR TI-RADS Can Improve the Differential Diagnosis of Thyroid Nodules. *Front Oncol* (2021) 11:575166. doi:10.3389/fonc.2021.575166
13. Wan P, Chen F, Liu C, Kong W, Zhang D. Hierarchical Temporal Attention Network for Thyroid Nodule Recognition Using Dynamic CEUS Imaging. *IEEE Trans Med Imaging* (2021) 40:1646–60. doi:10.1109/tmi.2021.3063421
14. Yoo H, Lee S.H., Arru C.D., Doda Khera R, Singh R, Siebert S, et al. AI-Based Improvement in Lung Cancer Detection on Chest Radiographs: Results of a Multi-Reader Study in NLST Dataset. *Eur Radiol* (2021) 31:9664–74. doi:10.1007/s00330-021-08074-7
15. Wang C, Shao J, Lv J, Cao Y, Zhu C, Li J, et al. Deep Learning for Predicting Subtype Classification and Survival of Lung Adenocarcinoma on Computed Tomography. *Translational Oncol* (2021) 14:101141. doi:10.1016/j.tranon.2021.101141
16. Venkadesh K.V., Setio AAA, Schreuder A, Scholten ET, Chung K, W. Wille MM, et al. Deep Learning for Malignancy Risk Estimation of Pulmonary Nodules Detected at Low-Dose Screening CT. *Radiology* (2021) 300:438–47. doi:10.1148/radiol.2021204433
17. Liu F, Zhou Z, Samsonov A, Blankenbaker D, Larison W, Kanarek A, et al. Deep Learning Approach for Evaluating Knee MR Images: Achieving High Diagnostic Performance for Cartilage Lesion Detection. *Radiology* (2018) 289:160–9. doi:10.1148/radiol.2018172986
18. Kijowski R, Liu F, Caliva F, Pedoia V. Deep Learning for Lesion Detection, Progression, and Prediction of Musculoskeletal Disease. *J Magn Reson Imaging* (2020) 52:1607–19. doi:10.1002/jmri.27001
19. Liu F, Zhou Z, Jang H, Samsonov A, Zhao G, Kijowski R. Deep Convolutional Neural Network and 3D Deformable Approach for Tissue Segmentation in Musculoskeletal Magnetic Resonance Imaging. *Magn Reson Med* (2018) 79:2379–91. doi:10.1002/mrm.26841
20. Cai Y, Landis M, Laidley D.T., Kornecki A, Lum A, Li S. Multi-Modal Vertebrae Recognition Using Transformed Deep Convolution Network. *Comput Med Imaging Graphics* (2016) 51:11–9. doi:10.1016/j.compmedimag.2016.02.002
21. Xue Y, Zhang R, Deng Y, Chen K, Jiang T. A Preliminary Examination of the Diagnostic Value of Deep Learning in Hip Osteoarthritis. *PLoS One* (2017) 12:e0178992. doi:10.1371/journal.pone.0178992
22. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network. *Med Image Comput Comput Assist Interv* (2013) 16:246–53. doi:10.1007/978-3-642-40763-5_31
23. Zhao Z-Q, Zheng P, Xu S-T, Wu X. Object Detection with Deep Learning: A Review. *IEEE Trans Neural Netw Learn Syst.* (2019) 30:3212–32. doi:10.1109/tnnls.2018.2876865
24. Livieratos FA, Johnston LE, Jr. A Comparison of One-Stage and Two-Stage Nonextraction Alternatives in Matched Class II Samples. *Am J Orthod Dentofacial Orthopedics* (1995) 108:118–31. doi:10.1016/s0889-5406(95)70074-9
25. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: 2017 IEEE International Conference on Computer Vision (ICCV); 22-29 Oct. 2017; Venice, Italy. IEEE (2017).
26. Girshick R. Fast R-CNN. arXiv e-prints 2015 (2015).
27. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv:1506.01497 (2017).
28. Jiang B, Luo R, Mao J, Xiao T, Jiang Y. Acquisition of Localization Confidence for Accurate Object Detection. arXiv:1807.11590 (2018).
29. Zhang Z, Ni X, Huo G, Li Q, Qi F. Novel Coronavirus Pneumonia Detection and Segmentation Based on the Deep-Learning Method. *Ann Transl Med* (2021) 9:934. doi:10.21037/atm-21-1156

Conflict of Interest: Author TL was employed by Yihui Ebond (Shandong) Medical Technology Co., Ltd. Author SS was employed by Yihui Ebond (Shandong) 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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 15 March 2022
doi: 10.3389/fonc.2022.790076

A Nomogram Based on Molecular Biomarkers and Radiomics to Predict Lymph Node Metastasis in Breast Cancer

Xiaoming Qiu, Yufei Fu, Yu Ye, Zhen Wang and Changjian Cao *

Department of Radiology, Huangshi Central Hospital, Edong Healthcare Group, Affiliated Hospital of Hubei Polytechnic University, Huangshi, China

OPEN ACCESS

Edited by:
Antonio Napolitano, Bambino Gesù Children's Hospital (IRCCS), Italy

Reviewed by:
Mauro Giuseppe Mastropasqua, University of Bari Medical School, Italy
Wenjun Yi, Central South University, China

***Correspondence:**
Changjian Cao
caochangjian07@163.com

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: *21 February 2022*
Published: *15 March 2022*

Citation:
Qiu X, Fu Y, Ye Y, Wang Z and Cao C (2022) A Nomogram Based on Molecular Biomarkers and Radiomics to Predict Lymph Node Metastasis in Breast Cancer. Front. Oncol. 12:790076. doi: 10.3389/fonc.2022.790076

Background: The aim of this study was to explore the feasibility and efficacy of a non-invasive quantitative imaging evaluation model to assess the lymphatic metastasis of breast cancer based on a radiomics signature constructed using conventional T1-weighted image (T1WI) enhanced MRI and molecular biomarkers.

Methods: Patients with breast cancer diagnosed *via* lymph biopsies between June 2015 and June 2019 were selected for the study. All patients underwent T1WI contrast-enhancement before treatment; lymph biopsy after surgery; and simultaneous Ki-67, COX-2, PR, Her2 and proliferating cell nuclear antigen detection. All images were imported into ITK-SNAP for whole tumor delineation, and AK software was used for radiomics feature extraction. Next, the radiomics signature Rad-score was constructed after reduction of specific radiomic features. A multiple regression logistic model was built by combining the Rad-score and molecular biomarkers based on the minimum AIC.

Results: In all, 100 patients were enrolled in this study, including 45 with non-lymph node (LN) metastasis and 55 with LN metastasis. A total of 1,051 texture feature parameters were extracted, and LASSO was used to reduce the dimensionality of the radiomics features. The log(λ) was set to 0.002786, and 19 parameters were retained for the construction of the radiomics tag Rad-score. ROC was used to evaluate the diagnostic efficiency of Rad-score: the area under the ROC curve (AUC) of the Rad-score for identifying non-lymphatic and lymphatic metastases was 0.891 in the training cohort and 0.744 in the validation cohort. With the incorporation of tumor molecular markers, the

AUCs of the training cohort and validation cohort of the nomogram were 0.936 and 0.793, respectively, which were notably higher than the AUCs of the clinical parameters in the training and validation cohorts (0.719 and 0.588, respectively).

Conclusion: The combined model constructed using the Rad-score and molecular biomarkers can be used as an effective non-invasive method to assess LN metastasis of breast cancer. Furthermore, it can be used to quantitatively evaluate the risk of breast cancer LN metastasis before surgery.

Keywords: lymph node metastasis, breast cancer, molecular biomarkers, radiomics, diagnostics

INTRODUCTION

Breast cancer is the most common malignant tumor in women globally, with lymphatic metastasis being the main cause of death (1). According American Society of Clinical Oncology Clinical Practice Guideline Update, Axillary lymph node dissection (ALND) is not recommended for patients with breast cancer without nodal metastases and one or two sentinel lymph node metastases (2).The five-year survival rate of patients with axillary lymphatic metastasis is significantly lower than that of patients without lymphatic metastasis (3). Patients with lymphatic metastasis require radiotherapy and chemotherapy in addition to surgery. Pre-treatment method indicating the absence of lymph node metastasis could provide the earlier stage and reduce the mortality for better treatment. Evern through patients under ALND would significant reduce the mortality rate but also improve the morbidities associated with ALND, such as seroma formation, impairment of shoulder movement, neuropathy and arm lymphedema (4). Sentinel lymph node biopsy (SLNB) was used to staging the axilla before treatment. SLN is the first lymphatic drainage lymph node in tumor which lead lymphatic spread (5). But several previous studies have showed that SLNB is the standard method to predict lymph node metastasis but the invasive procedure provide the high false negative rate and other complication (6). For clinical purposes, using effective markers for the individual conditions of different patients with lymph node (LN) metastasis can improve the prognosis of patients by actively adjusting the clinical treatment plan (7). In the previous study, the patients with breast cancer widely accepted complete ALND (cALND) for positive SLN that patients with negative SLN should avoid ALND (8). Several studies have found some biomarkers to predict SLN (9), such as tumor size (10), nucleic acid amplification (CK19) (10), ER status (10) and PR status (11).

Currently, pathological biopsy is the gold standard for identifying LN metastasis in patients with breast cancer. Furthermore, sentinel LN biopsy is a standard clinical procedure for pathological biopsy of patients with breast cancer. However, such methods are invasive, are not readily permitted by patients, and fail to provide comprehensive information regarding metastasis (12, 13).

Currently, the non-invasive method for evaluating breast cancer lymphatic metastasis is mainly based on imaging evaluation. Magnetic resonance imaging (MRI) has high-resolution characteristics, especially in soft tissue contrast, and can accurately display LNs. Radiomics is used to build a mathematical model based on the image data from confirmed cases for high-throughput texture feature data mining, which could then be added to the clinical cases to improve model verification, so as to construct a non-invasive evaluation method for clinical research purposes (14).

In contrast, molecular tumor markers play an important auxiliary role in the clinical diagnosis of tumors. Among them, Ki-67 can react with proliferating nuclear antigens during the cell proliferation cycle and is therefore a marker of rapid tumor growth. The expression of Ki-67 in breast cancer is significantly related to pathological grade and LN metastasis (15). Furthermore, proliferating cell nuclear antigen (PCNA) is an intranuclear polypeptide synthesized during cell proliferation. During the malignant proliferation of cancer cells, PCNA expression is abnormally high. Studies have shown that high PCNA expression in patients with tumors results in rapid clinical progress and predisposition to LN metastasis (16). In addition, the positive expression of cyclooxygenase-2 (COX-2) in breast cancer tissue is significantly higher than that in other tumor diseases and is correlated with lymphatic and distant metastasis (17).

The aim of this study was to explore a feasible, effective, and non-invasive joint model to assess lymphatic metastasis of patients with breast cancer based on the radiomic features of their T1-weighted image (T1WI)-enhanced scans and post-operative molecular biomarkers. This will contribute to a new quantitative analysis for evaluating the risk of breast cancer LN metastasis before surgery.

Abbreviations: 3D, three-dimensional; AUC, under the ROC curve; COX-2, cyclooxygenase-2; DCA, decision curve analysis; GLCM, gray level co-occurrence matrix; GLDM, gray level dependence matrix; GLRLM, gray-level run-length matrix; GLZSM, gray level size zone matrix; IBSI, Image Biomarker Standardization Initiative; LASSO, least absolute shrinkage and selection operator; LN, lymph node; MRI, magnetic resonance imaging; mRMR, maximal relevance and minimal redundancy; PCNA, proliferating cell nuclear antigen; T1WI, T1-weighted image; VOI, volume of intervalidation.

MATERIALS AND METHODS

Patients

This study was reviewed and approved by the ethics committee of the Huangshi central hospital. Patients diagnosed with breast

cancer *via* pathology in the hospital between June 2015 and June 2019 were strictly screened according to the following inclusion and exclusion criteria, inclusion criteria:1) solid masses in the image of the lesions; 2) maximum diameter of the lesion ≤2 cm; 3) diagnosis of primary breast cancer; 4) MRI shows no axillary lymph node (ALN); 5) SLN before treatment; 6) post-operative tumor tissue identified as breast cancer after pathological examination; 7) twice LN biopsy to confirm the status after operation; 9) non-metastatic primary lesions on pathological examination. Exclusion criteria: 1) breast implants such as silicone; 3) radiotherapy, chemotherapy, drug therapy and surgical treatment before enhanced MRI scan; 4) poor image quality of MRI.

MRI Scanning

All images were obtained using enhanced MRI scanning (1.5T superconducting MRI; Siemens, Munich, Germany). The patient was required to lie on his back (feet first) and wear noise-reducing headphones for the MRI; the conventional T1WI enhancement sequence in the transverse position was selected, centering on the largest layer of the lesion, and a total of 20 layers were scanned up–and–down. A bolus injection of the contrast agent gadolinium diamine (dose: 0.1 mmol/kg; General Electric Pharmaceuticals, Shanghai, China) was selected, and the injection rate was set at 2 mL/s. After the bolus injection of the contrast agent was completed, 20 mL normal saline was injected at the same rate for flushing. The scanning parameters were set as follows: TR 3.9 ms, TE 1.4 ms, FOV 380 mm × 280 mm, matrix 256 × 256, layer thickness 5 mm, and 20 slices total per volume.

Molecular Biomarker Analysis

The tissue of breast cancer was collected by needle biopsy. The tissue of breast cancer was fixed by 4% paraformaldehyde and then embedded by paraffin. Immunohistochemistry (IHC) was used to detect the expression of Ki-67, COX-2, PCNA, PR and Her2. for Ki-67, samples with >20% positive nuclei were considered to show high Ki-67 expression, while samples with 20% positive nuclei were considered to show low Ki-67 expression (18, 19); for COX-2, samples with >30% positive cytoplasm were considered to show high COX-2 expression, while samples with <30% positive cytoplasm were considered to show low COX-2 expression; for PCNA, samples with >10% positive nuclei were considered to show high PCNA expression, while samples with <10% positive nuclei were considered to show low PCNA expression; for PR, samples with >1% nuclear staining as PR positive, <1% nuclear staining as PR negative (20); Her2 positive was detected by IHC that the staining score was 3+ (21).

Image Analysis

The data were processed by two doctors with 10–15 years of diagnostic experience in the following steps. The original MR images were imported into ITK-SNAP (www.itksnap.org), and the breast cancer lesions were processed by the diagnostician according to the single-blind principle. The lesions were then delineated and synthesized in three-dimensions (3D), the whole tumor was segmented, and the 3D region of interest (volume of intervalidation, VOI) was saved. The lymph was not delineated in this study. The image and VOI were then imported into Anaconda Prompt (version 4.2.0) importing the feature package of "pyradiomics" (github.com/Radiomics/pyradiomics), according to the guidelines of the Image Biomarker Standardization Initiative (IBSI). A total of 1,051 features were extracted, including shape parameters, first-order parameters, gray level co-occurrence matrix (GLCM parameters), gray-level run-length matrix (GLRLM parameters), gray level size zone matrix (GLZSM parameters), and gray level dependence matrix (GLDM parameters).

Statistical Analysis

The maximal relevance and minimal redundancy (mRMR) algorithm with the least absolute shrinkage and selection operator (LASSO) method were used for feature dimensionality reduction, whereas the stepwise regression method was used to filter the radiomics features into the multivariate logistic regression

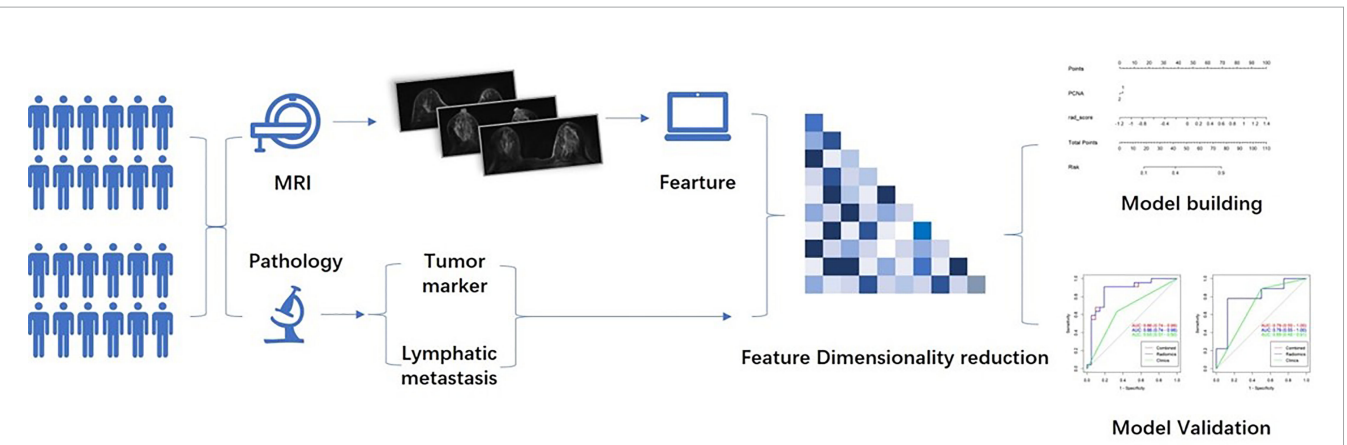


FIGURE 1 | Flowchart depicting model design.

analysis to obtain meaningful feature regression coefficients, to perform feature weighting and construct the radiomics label (Rad-score) for TNM staging assessment. Next, the meaningful tumor markers were screened for multiple logistic regression with a Rad-score, using a joint evaluation model, and a nomogram was drawn for the scoring system with the best predictive performance obtained from the above scores. The clinical application value was evaluated by decision curve analysis (DCA). The general process is illustrated in **Figure 1**.

The Kolmogorov-Smirnov test was used to validate normal distribution of the measurement data. Normally distributed data were represented as mean ± standard deviation, and non-normal data were represented by the median. Independent sample *t* validation or Mann-Whitney U validation was used for the measurement data, and X^2 validation was used for the count data. Differences were considered statistically significant at *P<0.05*. The performance of the scoring system was evaluated based on the area under the ROC curve (AUC). R studio (version 4.1.1) was used for processing and analysis, with specific packages: "xml2," "tidyverse," "caret," "pROC," "glmnet," "DMwR," "rmda," "ggpubr," "ModelGood," "rms," "mRMRe," "DescTools," and "Publish."

RESULTS

Clinical Characteristics Between LN-Positive and -Negative Patients in the Training and Validation Cohorts

A total of 185 patients were included in this study, whereas 100 were included according to the inclusion criteria: 45 patients without LN metastasis and 55 patients with LN metastasis (**Table 1**). Particularly, 30 patients were excluded due to poor image quality, and 55 patients were excluded because of unclear data pertaining to molecular biomarkers. The average age was 52.9 ± 11 years for the patients without LN metastasis and 53.6 ± 10 years for the patients with lymph node metastasis in the training cohort. At same time, the average age was 52.4 ± 12.7 years for the patients without LN metastasis and 51.2 ± 9.8 years for the patients with LN metastasis in the validation cohort. There was no significant difference between the non-LN metastasis and LN metastasis in the training cohorts and validation cohorts (*P=0.769 vs P=0.775*). Among the enrolled patients, 59 were positive for Ki-67 (41negative), 73 were positive for PCNA (27 negative), and 66 were positive for COX-2 (34 negative), and 55 were positive for PR (45 negative), 66 were positive for Her2 (34 negative) Ki-67 (*P=0.050*, **Table 1**) and Her2 (*P=0.004*, **Table 1**) showed significant difference between NLN and LNM in the training cohort, and COX-2 (*P=0.041*, **Table 1**) showed significant difference between NLM and LNM in the validation cohort.

The Radiomics Signature of LN-Positive and -Negative Patients in the Training and Validation Cohorts

Using mRMR to remove redundant features and screen out the feature combinations that are significantly different for lymphatic metastasis, a total of 30 features were selected. Next, LASSO was used to reduce the dimensionality of the radiomics features, taking log(λ) as 0.0027 (**Figure 2**), and finally 19 parameters were retained to construct the radiomics. The Rad-score is based on the following formula \(**Figure 3**):

TABLE 1 | Clinical information of patients in the training and validation cohort.

	Training cohort		P value	Validation cohort		P value
	NLN	LNM		NLN	LNM	
Patients	n = 32	n = 39		n = 13	n = 16	
Age, years	52.9 ± 11	53.6 ± 10	0.769	52.4 ± 12.7	51.2 ± 9.8	0.775
Histological grade						
I	2	3		1	1	
II	14	12		7	6	
III	16	24		5	9	
Molecular status						
Ki-67			0.050*			0.061
Positive	23	18		11	7	
Negative	9	21		2	9	
PCNA			0.099			0.364
Positive	27	25		11	10	
Negative	5	14		2	6	
COX-2			0.167			0.041*
Positive	24	22		12	8	
Negative	8	17		1	8	
PR			0.144			0.867
Positive	22	19		7	7	
Negative	10	20		6	9	
Her2			0.004*			1.000
Positive	15	32		9	10	
Negative	17	7		4	6	

*NLN, non-lymph node metastasis; LNM, lymph node metastasis; *means significant difference.*

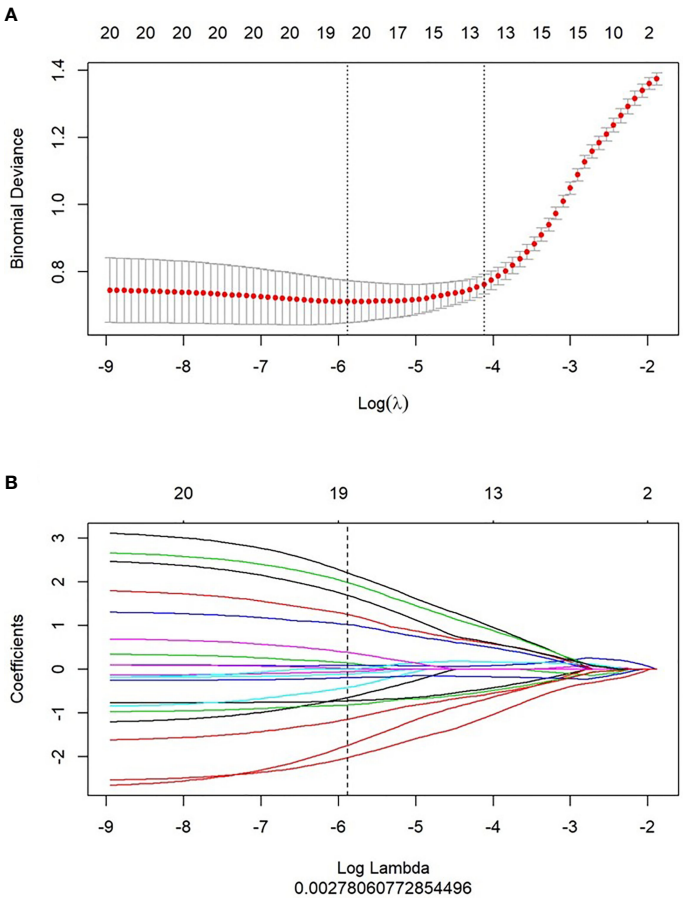


FIGURE 2 | Feature selection and Rad-score building by LASSO. **(A)** 10-fold cross validation was used to predict binomial deviance of the Rad-score building by different lambda values. **(B)** The coefficient profiles of the radiomics features by different lambda values.

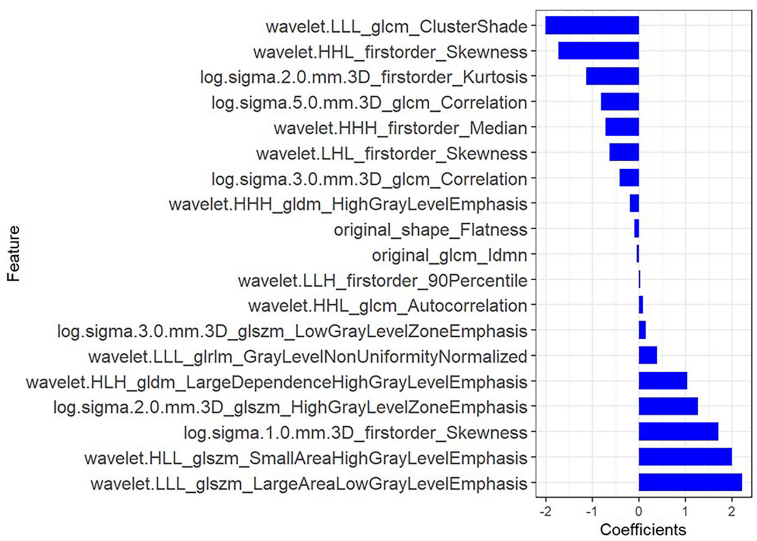


FIGURE 3 | The coefficients of radiomic features to construct the Rad-score.

$$
\begin{aligned}
Radscore = {} & -0.179 \times \text{wavelet.HHH\_firstorder\_Median} - 2.016 \\
& \times \text{wavelet.LLL\_glcm\_ClusterShade} + 0.15 \\
& \times \text{log.sigma.3.0.mm} \\
& .3\text{D\_glszm\_LowGrayLevelZoneEmphasis} + 0.087 \\
& \times \text{wavelet.HHL\_glcm\_Autocorrelation} - 0.414 \\
& \times \text{log.sigma.3.0.mm.3D\_glcm\_Correlation} + 0.017 \\
& \times \text{wavelet.LLH\_firstorder\_90Percentile} - 0.64 \\
& \times \text{wavelet.LHL\_firstorder\_Skewness} - 1.736 \\
& \times \text{wavelet.HHL\_firstorder\_Skewness} - 0.817 \\
& \times \text{log.sigma.5.0.mm.3D\_glcm\_Correlation} + 1.03 \\
& \times \\
& \text{wavelet.HLH\_gldm\_LargeDependenceHighGrayLevelEmphasis} \\
& - 0.102 \times \text{original\_shape\_Flatness} - 0.056 \\
& \times \text{original\_glcm\_Idmn} + 2.212 \\
& \times \\
& \text{wavelet.LLL\_glszm\_LargeAreaLowGrayLevelEmphasis} + 1.26 \\
& \times \\
& \text{log.sigma.2.0.mm.3D\_glszm\_HighGrayLevelZoneEmphasis} \\
& + 1.992 \\
& \times \text{wavelet.HLL\_glszm\_SmallAreaHighGrayLevelEmphasis} \\
& - 0.194 \\
& \times \text{wavelet.HHH\_gldm\_HighGrayLevelEmphasis} \\
& + 0.39 \\
& \times \text{wavelet.LLL\_glrlm\_GrayLevelNonUniformityNormalized} \\
& + 1.699 \\
& \times \text{log.sigma.1.0.mm.3D\_firstorder\_Skewness} \\
& - 1.142 \times \text{log.sigma.2.0.mm.3D\_firstorder\_Kurtosis} \\
& + 0.508
\end{aligned}
$$

After the Rad-score calculation for all patients in both the training and validation cohorts, based on the Wilcoxon validation, there was a significant difference between the non-lymphatic and lymphatic metastasis groups in the training cohort (P=0.000, **Figure 4A**) and the validation cohort (P=0.028, **Figure 4B**).

Diagnostic Performance of the Rad-Score in Different LN-Positive and -Negative Patients in the Training and Validation Cohorts

After obtaining the Rad-score of patients in the non-lymphatic and lymphatic metastasis groups, the diagnostic performance of the Rad-score was evaluated based on the ROC. The Rad-score distinguished between non-lymphatic and lymphatic metastasis with an AUC of 0.891 (**Figure 4C**) in the training cohort and 0.744 (**Figure 4D**) in the validation cohort (**Table 2**).

Diagnostic Performance of the Rad-Score, Clinical Factors, and Nomogram in Different LN-Positive and -Negative Patients in the Training and Validation Cohorts

In the training cohort, Ki67 and Her2 showed significant difference between NLM and LNM. Backward step logistic model was used to build clinical model. Ki67 (OR=0.44, 95% CI: 0.15-1.25) and Her2 (OR=4.41, 95%CI: 1.47-13.24) were used to construct the clinical model based minimum AIC. Then the combined model was constructed by molecular biomarker (OR_{Ki67} = 0.19, 95%CI: 0.03-1.07; OR_{Her2} = 4.41, 95%CI: 0.97-26.50) and radscore (OR=2.28, 95%CI:1.50-3.46). The combined model was visualized by Nomogram (**Figure 5A**). The evaluation performance of the clinical model for the training and validation cohort were 0.642 and 0.773 (**Figures 5B, C** and **Table 2**). The nomogram showed diagnostic performance for the training and validation cohort were 0.936 (**Figure 5B**) and 0.793 (**Figure 5C** and **Table 2**).

Evaluation of Lymph Metastasis *via* a Nomogram

The nomogram was used to visualize the combine model. The Rad-score, Ki67 and Her2 score axis were projected vertically to the Points axis, and the total risk for assessing breast cancer lymphatic metastasis was given as the total points of lymphatic metastasis. The greater the risk, the greater is the probability that a patient would present with lymphatic metastasis (**Figures 6A, B**). The DCA analysis for clinical model, Radscore and Nomogram have indicated the threshold under 0.92 that patient would benefit from nomogram (**Figure 6A**). (**Figure 6C**) Delong test have showed a significant difference between nomogram and clinical model in the training cohort (P=0.0001) but not in validation cohort (P=0.111). At the same time, radscore showed a statistically significant difference when compared with clinical model in the training cohort (P=0.027) but also not in validation cohort (P=0.027).

DISCUSSION

In this study, we performed a non-invasive assessment of lymphatic metastasis in patients with breast cancer based on a multiple logistic regression model that combined the radiomic tag Rad-score extracted from the conventional breast cancer T1WI-enhanced scan with the tumor biomarkers. The results demonstrated that the diagnostic efficiency of Rad-score (training cohort AUC = 0.891; validation cohort AUC = 0.744), was higher than that of the tumor biomarker model (clinical model, training cohort AUC = 0.642; validation cohort AUC = 0.773). However, once Rad-score and molecular

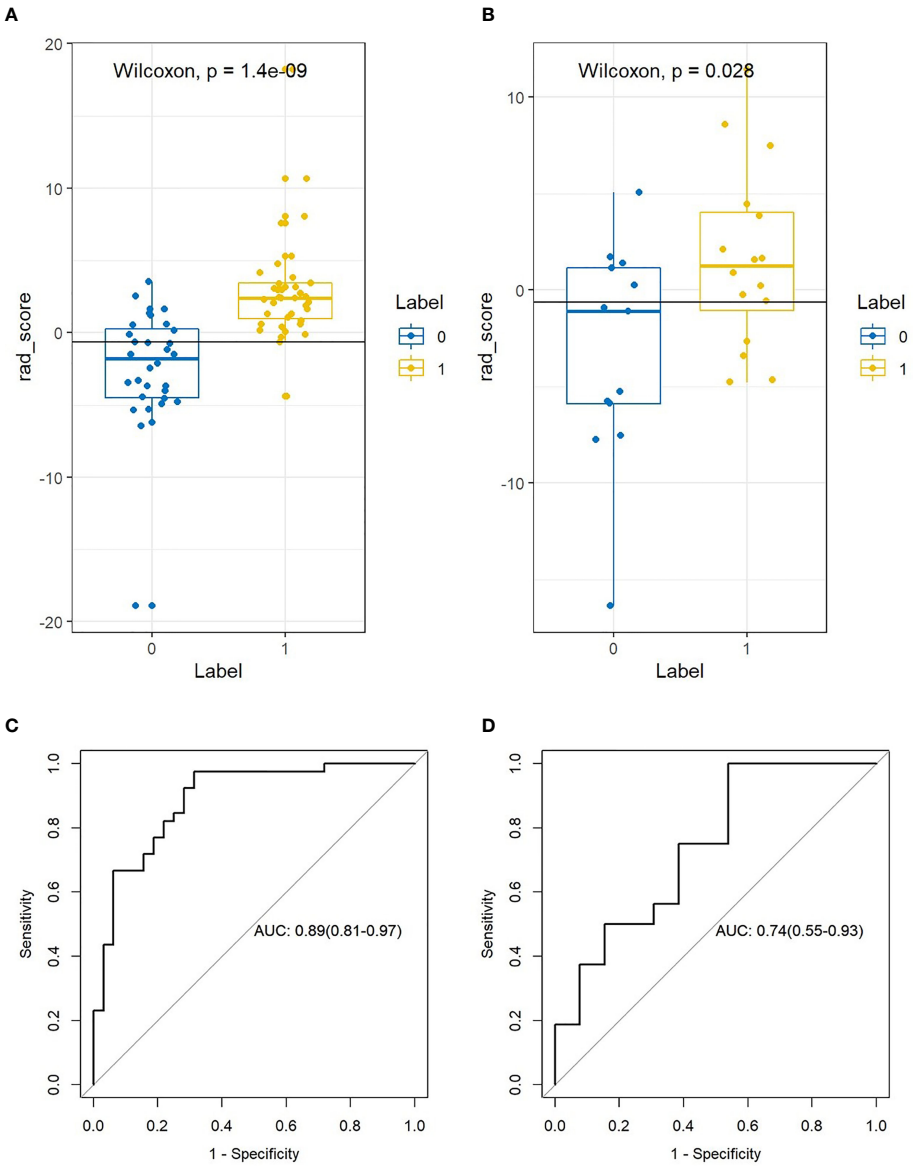


FIGURE 4 | The difference and ROC curves of the Rad-score in the training and validation cohorts. **(A, B)** Mann-Whitney U validation was used to analyze the difference between lymph node (LN)-positive and -negative patients in the training and validation cohorts. **(C, D)** ROC curve of the Rad-score in the training and validation cohort. AUC was used to predict the diagnostic performance between the LN-positive and -negative patients.

TABLE 2 | Diagnostic performance of the Rad-score, clinical data, and nomogram in the training and validation cohorts.

	Cohort	AUC	95% CI		Accuracy	Sensitivity	Specificity	PPV	NPV
			Lower	Upper					
Rad-score	Training	0.891	0.812	0.967	0.845	0.974	0.687	0.792	0.956
	Validation	0.744	0.552	0.931	0.689	0.750	0.615	0.705	0.667
Clinics	Training	0.719	0.602	0.843	0.690	0.820	0.531	0.680	0.708
	Validation	0.588	0.0.379	0.801	0.483	0.625	308	0.526	0.400
Nomogram	Training	0.936	0.882	0.992	0.901	0.974	0.812	0.863	0.962
	Validation	0.793	0.624	0.959	0.603	0.727	1.000	1.000	0.538

95% CI, 95% confidence interval; NPV, negative predictive value; PPV, positive predictive value.

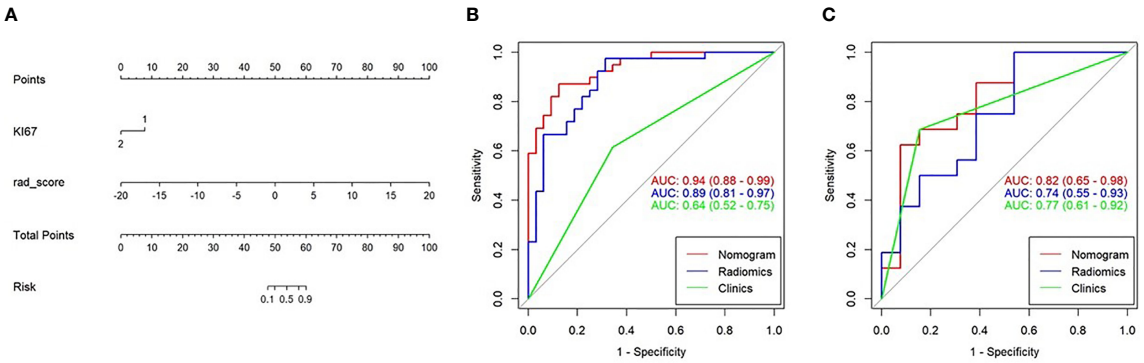


FIGURE 5 | Nomogram combining the clinical data and Rad-score. **(A)** The multiple logistic regression model constructed using the Rad-score and clinical data visualized by the nomogram. **(B, C)** The ROC curves of the Rad-score, clinical data, and nomogram in the training and validation cohorts.

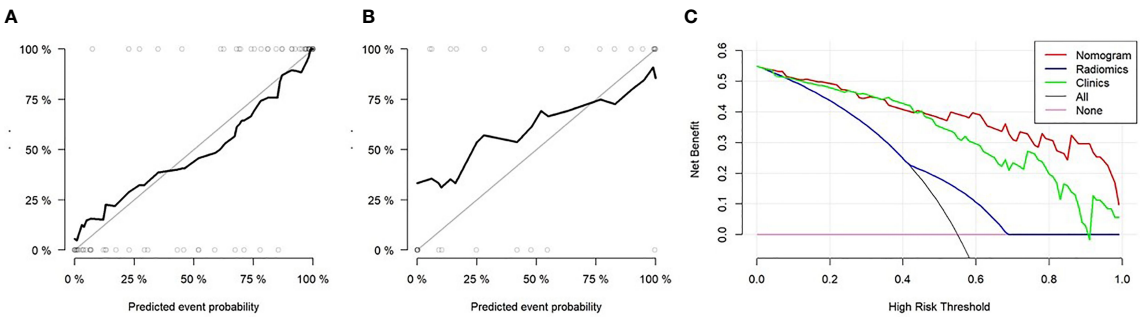


FIGURE 6 | The diagnostic performance of nomogram evaluation. **(A)** Decision curve analysis of the Rad-score, clinical data, and nomogram. The y-axis indicates the clinical benefits while the x-axis indicates the clinical risk to predict lymph node (LN) metastasis. The "All line" indicates a randomized evaluation of the LN metastasis. The pink line indicates no method was used to evaluate the LN metastasis. **(A–C)** Calibration curves of the nomogram in the training and validation cohorts.

biomarker model were combined as nomogram, the diagnostic efficiency was the highest among radiscore and clinical model in the training cohort (AUC=0.936) and validation cohort (AUC=0.793). In the current clinical treatments, lymphatic metastasis is an important prognostic indicator and factor in selecting the appropriate treatment for patients with breast cancer. A study by Wang et al. showed that 44% (95/216) of patients with breast cancer had sentinel lymphatic metastasis (13). However, both the sentinel lymphatic biopsy and axillary LN dissection are invasive diagnostic techniques, whereas MRI can be used as a qualitative imaging assessment method to determine breast cancer lymphatic metastasis (22, 23). Therefore, this study aimed to establish a non-invasive and highly sensitive model for assessment of breast cancer lymphatic metastasis.

First, the texture feature parameters were extracted based on the conventional T1WI enhanced scan, and after removing redundant features, 19 parameters were retained. These parameters can be understood as follows. Usually, Interia_angle45_offset7 reflects the definition of the image and the depth of the texture groove—the higher the groove depth, the higher the image contrast and definition. GLCMEntropy_ AllDirection_offset1_SD reflects the entropy value of the image, which represents the image required for image compression—the higher the entropy value, the higher the confusion of the image. ShortRunEmphasis_angle90_offset1 reflects all voxel points that are not often 1 at a given angle (90°), and Correlation_ AllDirection_offset4_SD represents the similarity of gray levels in adjacent pixels. InverseDifferenceMoment_ AllDirection_ offset4_SD represents local homogeneity, which is proportional to local gray uniformity. The Rad-score value of all patients obtained by the Rad-score calculation formula was as follows: AUC of 0.891 in the training cohort and 0.744 in the validation cohort. Chai et al. (24) showed that based on T1WI, CE2, T2WI, and DWI sequences for radiomics feature extraction, the AUC reached 0.87, which was similar to the results of this study. Simultaneously, the study showed that the vascular permeability parameters K^{trans}, K_{ep}, V_e, V_p, etc. could improve the diagnostic performance of the model (accuracy 0.86, AUC 0.91) (24). Currently, according to breast cancer imaging guidelines, dynamic enhanced MRI is used as a clinically recommended protocol for the diagnosis of breast cancer, which suggested that radiomic features should include dynamic enhancement in future

studies. Previous studies have showed that the short-term survival didn't have different between patients who taken ALND or not. The American society of clinical oncology (ASCO) recommended SLNB for patient with early-stage breast cancer to reduce the unnecessary ALND. Previous studied have found that the sensitivity of SLND was 0.855, the sensitivity of Rad-score were 0.974 (training cohort) and 0.750 (validation cohort)

In the second, the gender, age, Ki67, COX-2, PCNA, PR and Her 2 were included in the lymphatic metastasis assessment of breast cancer in this study. The final screening index with statistical significance was Ki67 and Her2. Lymphatic metastasis was assessed based on clinical model combined Ki-67 and Her2, with an AUC of 0.7192 in the training cohort and 0.588 in the validation cohort, which was lower than the diagnostic power of radiomics (AUC = 0.891 and 0.744, respectively). Our results are similar to those of Wang et al. (16), which reported clinical index parameters of 0.707 and 0.657, respectively, to assess breast cancer lymphatic metastasis. However, their clinical indicators were tumor diameter, tumor molecular classification, ER phenotype, etc., which differed from the clinical indicators in this study. These findings suggest that multiple clinical indicators should be included in the assessment of breast cancer LN metastasis (13). In our studies, we included 5 type molecular biomarker try to predict the status of lymph node metastasis. All patients included in our research are ER positive. Our results showed that Ki67 and Her2 were the risk factor of lymph node metastasis. The OR of Ki67 and Her2 were 0.44 and 4.41 which were similar to previous studies (11, 25). In our model, Her2 overexpression would improve the risk of lymph node metastasis. Ki-67 positive also improve the risk of lymph node metastasis. Previous studied showed that Her2 and Ki67 could help in increasing the sensitivity to estimate the probability of lymph node positive (26). Our results showed Ki-67 and Her 2 combined the radscore have improve the sensitivity and specificity in training and validation cohorts. However, these results showed that the nomogram constructed by the Rad-score, integrated with clinical indicators, was more effective in evaluating breast cancer lymphatic metastasis than the clinical indicators alone, although it was the same as the independent Rad-score. This result indicates that the Rad-score has diagnostic power with or without clinical indicators. This means that the diagnostic efficiency of the Rad-score was significantly higher than that of the clinical indicators, and the evaluation of breast cancer lymphatic metastasis could be based on the Rad-score indicators. This result is different from that of the clinical imaging joint model used to improve the diagnostic efficiency (27, 28). The procedure of SLNB consumed time and expensive even through SLNB is the gold standard to diagnosis lymph node metastasis. SLNB would take adverse event through invasive procedure. Nomogram, combined molecular biomarker, also needed invasive biomarker. In our study, Rad-score was constructed to predict LNM before surgery without invasive method.

Potential limitations of this study include the small sample size and the omission of molecular classification of tumors, which might have led to the low efficiency of the clinical indicators for breast cancer. The clinical indicators used in this study were tumor resection and immunohistochemical detection. This suggests that clinical indicators should be expanded to establish a more comprehensive clinical joint model in future studies. In addition, MRI multimodal sequences should be included in the radiomics feature extraction, especially multi-phase dynamic enhancement.

In conclusion, a large number of radiomic parameters were extracted based on conventional T1WI MRI enhancement in this study, and a radiomics model was constructed to evaluate breast cancer lymphatic metastasis. The results demonstrated that the radiomic model had high diagnostic feasibility and efficacy, and the MRI radiomics model might be helpful in evaluating the clinical prognosis of patients with breast cancer.

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 the Huangshi Central Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Guarantor of integrity of entire study, XQ and CC. Study concepts/study design or data acquisition of data analysis/interpretation, XQ. Manuscript drafting or manuscript revision for important intellectual content, all authors. Manuscript final version approval, all authors. All authors agree to ensure that any questions related to the work are appropriately resolved. Literature research, XQ, YF, and CC. Clinical studies, XQ and YY. Statistical analysis, XQ and ZW. Manuscript editing, CC. All authors have read and approved the final manuscript.

FUNDING

Hubei Health Committee General Program and Anti-schistosomiasis Fund during 2019-2020 [Grant No. WJ2019M043].

ACKNOWLEDGMENTS

We would like to thank Levin Ma for helpful discussions.

REFERENCES

1. Chen W, Hoffmann AD, Liu H, Liu X. Organotropism: New Insights Into Molecular Mechanisms of Breast Cancer Metastasis. *NPJ Precis Oncol* (2018) 2:4. doi: 10.1038/s41698-018-0047-0
2. Lyman GH, Somerfield MR, Bosserman LD, Perkins CL, Weaver DL. Sentinel Lymph Node Biopsy for Patients With Early-Stage Breast Cancer: American Society of Clinical Oncology Clinical Practice Guideline Update. *J Clin Oncol* (2017) 35(5):561–4. doi: 10.1200/JCO.2016.71.0947
3. Fisher B, Bauer M, Wickerham DL, Redmond CK, Fisher ER, Cruz AB, et al. Relation of Number of Positive Axillary Nodes to the Prognosis of Patients With Primary Breast Cancer. *NSABP Update Cancer* (1983) 52:1551–7. doi: 10.1002/1097-0142(19831101)52:9<1551::aid-cncr2820520902>3.0.co;2-3
4. Say CC, Donegan W. A Biostatistical Evaluation of Complications From Mastectomy. *Surg Gynecol Obstet* (1974) 138:370–6.
5. Giuliano AE, Kirgan DM, Guenther JM, Morton DL. Lymphatic Mapping and Sentinel Lymphadenectomy for Breast Cancer. *Ann Surg* (1994) 220:391–401. doi: 10.1097/00000658-199409000-00015
6. Purushotham AD, Upponi S, Klevesath MB, Bobrow L, Milar K, Myles JP, et al. Morbidity After Sentinel Lymph Node Biopsy in Primary Breast Cancer: Results From a Randomized Controlled Trial. *J Clin Oncol* (2005) 23 (19):4312–21. doi: 10.1200/JCO.2005.03.228
7. Rack B, Janni W, Gerber B, Strobl B, Schindlbeck C, Klanner E, et al. Patients With Recurrent Breast Cancer, Does the Primary Axillary Lymph Node Status Predict More Aggressive Tumor Progression. *Breast Cancer Res Treat* (2003) 82:83–92. doi: 10.1023/b:brea.0000003955.73738.9e
8. Nos C, Harding-MacKean C, Freneaux P, Trie A, Falcou MC, Sastre Garau X, et al. Prediction of Tumour Involvement in Remaining Axillary Lymph Nodes When the Sentinel Node in a Woman With Breast Cancer Contains Metastases. *Br J Surg* (2003) 90(11):1354–60. doi: 10.1002/bjs.4325
9. Zhu L, Jin L, Li S, Chen K, Weijuan J, Quanyuan S, et al. Which Nomogram Is Best for Predicting Non-Sentinel Lymph Node Metastasis in Breast Cancer Patients? A Meta-Analysis. *Breast Cancer Res Treat* (2013) 37(3):783–95. doi: 10.1007/s10549-012-2360-6
10. Teramoto A, Shimazu K, Naoi Y, Noguchi S, Shimomura A, Shimoda M. One-Step Nucleic Acid Amplification Assay for Intraoperative Prediction of non-Sentinel Lymph Node Metastasis in Breast Cancer Patients With Sentinel Lymph Node Metastasis. *Breast* (2014) 3(5):579–85. doi: 10.1016/j.breast.2014.05.026
11. Meng L, Zheng T, Wang Y, Li Z, Xiao Q, Junfeng H, et al. Development of a Prediction Model Based on LASSO Regression to Evaluate the Risk of non-Sentinel Lymph Node Metastasis in Chinese Breast Cancer Patients With 1-2 Positive Sentinel Lymph Nodes. *Sci Rep* (2021) 11(1):19972. doi: 10.1038/s41598-021-99522-3
12. Kootstra J, Hoekstra-Weebers JE, Rietman H, de Vries J, Baas P, Geertzen JHB, et al. Quality of Life After Sentinel Lymph Node Biopsy or Axillary Lymph Node Dissection in Stage I//II Breast Cancer Patients: A Prospective Longitudinal Study. *Ann Surg Oncol* (2008) 15:2433–541. doi: 10.1245/s10434-008-9996-9
13. Tan H, Wu Y, Bao F, Zhou J, Wan J, Tian J, et al. Mammography-Based Radiomics Nomogram: A Potential Biomarker to Predict Axillary Lymph Node Metastasis in Breast Cancer. *Br J Radiol* (2019) 93:20191019. doi: 10.1259/bjr.20191019
14. 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
15. Meng N, Wang X, Sun J, Han D, Bai Y, Wei W, et al. A Comparative Study of the Value of Amide Proton Transfer-Weighted Imaging and Diffusion Kurtosis Imaging in the Diagnosis and Evaluation of Breast Cancer. *Eur Radiol* (2020) 31:1707–17. doi: 10.1007/s00330-020-07169-x
16. Wang X, Zhou L, Zhang H, Ou H, Long W, Liu X. Upregulation of Cervical Carcinoma Expressed PCNA Regulatory Long non-Coding RNA Promotes Esophageal Squamous Cell Carcinoma Progression. *Oncol Lett* (2020) 20:142. doi: 10.3892/ol.2020.12006
17. Liu L, Yan J, Cao Y, Yan Y, Shen X, Yu B, et al. Proliferation, Migration and Invasion of Triple Negative Breast Cancer Cells are Suppressed by Berbamine *via* the PI3K/Akt/MDM2/p53 and PI3K/Akt/mTOR Signaling Pathways. *Oncol Lett* (2021) 21:70. doi: 10.3892/ol.2020.12331
18. Polley MY, Leung SC, Gao D, Mastropasqua MG, Zabaglo LA, Bartlett JMS, et al. An International Study to Increase Concordance in Ki67 Scoring. *Mod Pathol* (2015) 28(6):778–86. doi: 10.1038/modpathol.2015.38
19. Coates AS, Wilner EP, Goldhirsch A, Gelber RD, Gnant M, Piccart GM. Tailoring Therapies-Improving the Management of Early Breast Cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Ann Oncol* (2015) 26(8):1533–46. doi: 10.1093/annonc/mdv221
20. Fujii T, Kogawa T, Dong W, Sahin AA, Moulder S, Litton JK, et al. Revisiting the Definition of Estrogen Receptor Positivity in Her2-Negative Primary Breast Cancer. *Ann Oncol* (2017) 28(10):2420–8. doi: 10.1093/annonc/mdx397
21. Wolff AC, Hammond ME, Hicks DG, Dowsett M, Mcshane LM, Allison KH, et al. Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical Practice Guideline Update. *J Clin Oncol* (2013) 31:3997–4013. doi: 10.1200/JCO.2013.50.9984
22. Valente SA, Levine GM, Silverstein MJ, Rayhanabad JA, Weng-Grumley JG, Ji L, et al. Accuracy of Predicting Axillary Lymph Node Positivity by Physical Examination, Mammography, Ultrasonography, and Magnetic Resonance Imaging. *Ann Surg Oncol* (2012) 19:1825–30. doi: 10.1245/s10434-011-2200-7
23. Van Nijnatten TJA, Ploumen EH, Schipper RJ, Goorts B, Andriessen EH, Vanwetswinkel S, et al. Routine Use of Standard Breast MRI Compared to Axillary Ultrasound for Differentiating Between No, Limited and Advanced Axillary Nodal Disease in Newly Diagnosed Breast Cancer Patients. *Eur J Radiol* (2016) 85:2288–94. doi: 10.1016/j.ejrad.2016.10.030
24. 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
25. Wang NN, Yang ZJ, Wang X, Lixuan C, Hongmeng Z, WenFeng C, et al. A Mathematical Prediction Model Incorporating Molecular Subtype for Risk of non-Sentinel Lymph Node Metastasis in Sentinel Lymph Node-Positive Breast Cancer Patients: A Retrospective Analysis and Nomogram Development. *Breast Cancer*. (2018) 25(6):629–38. doi: 10.1007/s12282-018-0863-7
26. Fanizzi A, Pomarico D, Paradiso A, Bove S, Diotaiuti S, Didonna V, et al. Predicting of Sentinel Lymph Node Status in Breast Cancer Patients With Clinically Negative Nodes: A Validation Study. *Cancers (Basel)* (2021) 13(2). doi: 10.3390/cancers13020352
27. Yu FH, Wang JX, Ye XH, Deng J, Hang J, Yang B. Ultrasound-Based Radiomics Nomogram: A Potential Biomarker to Predict Axillary Lymph Node Metastasis in Early-Stage Invasive Breast Cancer. *Eur J Radiol* (2019) 119:108658. doi: 10.1016/j.ejrad.2019108658
28. Han L, Zhu Y, Liu Z, Yu T, He C, Jiang W, et al. Radiomics Nomogram for Prediction of Axillary Lymph Node Metastasis in Breast Cancer. *Eur J Radiol* (2019) 29(7):3820–9. doi: 10.1007/s00330-018-5981-2

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.

frontiers in Oncology

ORIGINAL RESEARCH
published: 31 March 2022
doi: 10.3389/fonc.2022.813069

Multi-Parameter MR Radiomics Based Model to Predict 5-Year Progression-Free Survival in Endometrial Cancer

*Defeng Liu[1], Linsha Yang[1], Dan Du[1], Tao Zheng[1], Lanxiang Liu[1], Zhanqiu Wang[1], Juan Du[1], Yanchao Dong[2], Huiling Yi[1] and Yujie Cui[1]**

[1] Medical Imaging Center, First Hospital of Qinhuangdao, Qinhuangdao, China, [2] Department of Intervention, First Hospital of Qinhuangdao, Qinhuangdao, China

OPEN ACCESS

Edited by:
Antonio Napolitano, Bambino Gesù Children's Hospital (IRCCS), Italy

Reviewed by:
Francesca Coppola, Azienda Ospedalieo-Universitaria di Bologna Italia (IRCCS), Italy
Shuai Ma, Peking University First Hospital, China

***Correspondence:**
Yujie Cui
52908094@qq.com

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

Received: 11 November 2021
Accepted: 21 February 2022
Published: 31 March 2022

Citation:
Liu D, Yang L, Du D, Zheng T, Liu L, Wang Z, Du J, Dong Y, Yi H and Cui Y (2022) Multi-Parameter MR Radiomics Based Model to Predict 5-Year Progression-Free Survival in Endometrial Cancer. Front. Oncol. 12:813069. doi: 10.3389/fonc.2022.813069

Background: Relapse is the major cause of mortality in patients with resected endometrial cancer (EC). There is an urgent need for a feasible method to identify patients with high risk of relapse.

Purpose: To develop a multi-parameter magnetic resonance imaging (MRI) radiomics-based nomogram model to predict 5-year progression-free survival (PFS) in EC.

Methods: For this retrospective study, 202 patients with EC followed up for at least 5 years after hysterectomy. A radiomics signature was extracted from T2-weighted imaging (T2WI), apparent diffusion coefficient (ADC) and a dynamic contrast-enhanced three-dimensional volumetric interpolated breath-hold examination (3D-VIBE). The radiomics score (RS) was calculated based on the least absolute shrinkage and selection operator (LASSO) regression. We have developed a radiomics based nomogram model ($Model^N$) incorporating the RS and clinical and conventional MR (cMR) risk factors. The performance was evaluated by the receiver operating characteristic curve (ROC), calibration curve and decision curve analysis (DCA).

Results: The $Model^N$ demonstrated a good calibration and satisfactory discrimination, with a mean area under the curve (AUC) of 0.840 and 0.958 in the training and test cohorts, respectively. In comparison with clinical prediction model ($Model^C$), the discrimination ability of $Model^N$ showed an improvement with $P < 0.001$ for the training cohort and P=0.032 for the test cohort. Compared to the radiomics prediction model ($Model^R$), $Model^N$ discrimination ability showed an improvement for the training cohort with P = 0.021, with no statistically significant difference in the test cohort (P = 0.106). Calibration curves suggested a good fit for probability (Hosmer–Lemeshow test, P = 0.610 and P = 0.956 for the training and test cohorts, respectively).

Conclusion: This multi-parameter nomogram model incorporating clinical and cMR findings is a valid method to predict 5-year PFS in patients with EC.

Keywords: endometrial cancer, progression-free survival, radiomics, magnetic resonance imaging, nomogram

INTRODUCTION

Endometrial cancer (EC) is one of the three most common malignancies of the female reproductive tract (1). Many clinical studies have shown that deciding whether to conduct radiotherapy or chemotherapy according to the risk of tumors can not only avoid unnecessary pain and economic burden brought by overtreatment of early-stage patients, but also avoid undertreatment of high-risk tumors, delay recurrence and improve the quality of life (2–4). Previous studies have proposed predicting the myometrial invasion and clinical outcome of EC by combining clinical and pathological indicators (5, 6). Tumor size, myometrial invasion, lymph vascular space invasion (LVSI) and other parameters obtained by postoperative tumor pathology can certainly be used to evaluate the EC prognosis. However, if we can use accurate, non-invasive methods to determine the risk and prognosis before surgery, it is beneficial to select more reasonable treatment strategies improve the progression-free and overall survival. Previous studies have found that EC prognosis is not only related to these pathological features, but also to the patient's age, BMI and other clinical indicators (7, 8). Therefore, making full use of these preoperative indicators is instrumental to a more accurate prognosis prediction. Furthermore, postoperative pathological examinations are very invasive but with appropriate preoperative predictive methods many unnecessary surgeries could be avoided.

Previous studies have found that preoperative staging, prognosis, and survival of EC can be predicted by using clinical standard magnetic resonance imaging (MR) sequences to assess the deep myometrial invasion, tumor volume or maximum diameter, and lymph node invasion (9–11). T1-weighted imaging (T1WI) and T2-weighted imaging (T2WI), the most commonly used modalities, are mainly used to evaluate tumor nature and prognosis by observing morphological characteristics. However, their accuracy is limited by visual resolution and the observer diagnostic ability (10). Although functional imaging methods such as quantitative diffusion and perfusion MRI can help us judge the tumor nature, these advanced imaging methods have high requirements on the imaging equipment and post-processing software, which may limit their accessibility (12). Radiomics represent a set of tools extracting quantitative features from medical images evaluating tumor characteristics such as heterogeneity (13). Data mining through radiomics allows researchers to explore the tumor heterogeneity, which is closely related to tumor aggressiveness and prognosis (14, 15). Previous studies have reported radiomics feasibility in predicting the histologic grade of endometrial carcinoma, lymph node metastasis or LVSI, and deep myometrial invasion (DMI) (16–18). However, the correlation between radiomics parameters and the EC patient survival is still unknown. Therefore, this study aimed to develop a multi-parameter MRI radiomics-based nomogram model to predict 5-year progression-free survival (PFS) in EC. In order to assess what our model achieved using this full sample set, and that it was not biased by the inclusion of various stages and grades of EC, we also carried out a sensitivity analysis focused on the different stages and grades.

MATERIALS AND METHODS

Patients

This retrospective study was approved by our institution's ethics committee. Informed consent was waived because analysis was performed on anonymized images and clinical data. A total of 460 patients with endometrial cancer confirmed by postoperative histopathology in our hospital from January 2011 to January 2016 were successively identified in the database. Inclusion criteria: (1) All patients underwent hysterectomy with bilateral salpingo-oophorectomy and were pathologically confirmed to be endometrial carcinoma, regardless of whether they had received radiotherapy or chemotherapy after surgery. (2) MR was performed within two weeks before surgery. Exclusion criteria: (1) No lesion that could be accurately identified in MR images or the maximum diameter of the lesion was less than 1 cm. (2) Lack of complete imaging data. (3) There are obvious artifacts in the image, which affect the observation. (4) Patients with co-malignancies. (5) Patients with further oncological diseases. (6) Follow-up less than 5 years or lost. 202 patients were enrolled in the study, and patients were randomly assigned to two separate cohorts, namely the training cohort (n=141) and the test cohort (n=61), in a 0.7:0.3 ratio (**Figure 1**).

Clinical Data

Clinical indicators collected preoperatively included patients' age, Body Mass Index (BMI), hypertension and diabetes, Carbohydrate Antigen 125 (CA125) and Human Epididymis Protein 4 (HE4) levels, which were obtained from our Hospital Information System (HIS). BMI = weight/height2 (kg/m2). Hypertension is defined as a systolic blood pressure of 140 mmHg or greater and/or a diastolic blood pressure of 90 mmHg or greater. Diabetes is defined as fasting blood glucose greater than or equal to 7.0 mmol/L and/or postprandial blood glucose greater than or equal to 11.1 mmol/L. The CA125 and HE 4 level was detected by chemiluminescence microparticle immunoassay (Cobas 8000 E602; Roche Holding AG).

Follow-Up

Progression was defined as local recurrence progression in the pelvis or new metastases in the abdomen or at distant sites, including nodal, peritoneal, or visceral metastases. All patients were consistently followed up every 3 to 6 months after surgery based on the thoracic, abdominal and pelvic CT or abdominal and pelvic MR imaging to determine if there is progression. The images were independently evaluated by two radiologists, neither of whom was aware of the EC stage or subtype. If the two radiologists cannot agree on the assessment of the metastasis status, another more experienced radiologist will conduct the assessment until a consensus is reached.

PFS is defined as the time when a patient receives surgical treatment until disease progression is observed or death from any cause occurs. Patients who relapsed or died within 5 years were

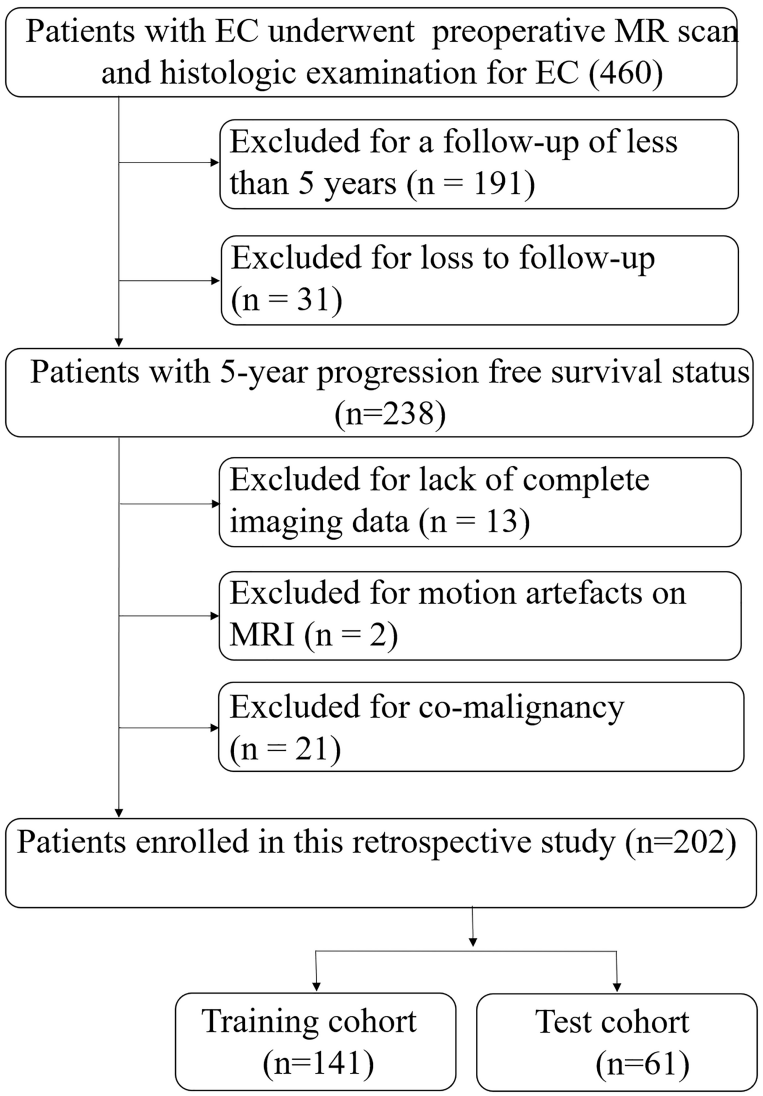


FIGURE 1 | Recruitment pathway for patients in this study. EC, endometrial cancer.

assigned to the high-risk (HR) group, while those who did not relapse were assigned to the low-risk (LR) group.

MRI Scan

Images were collected for all patients using a 1.5 T MR scanner (Avanto, Siemens) equipped with an 8-channel body coil. The scanning area ranged from the anteroom-superior iliac spine to the symphysis pubis. The scanning sequence included the coronal, sagittal, and axial oblique fat-saturation (fs) T2WI; axial oblique DWI and axial oblique three-dimensional volumetric interpolated breath-hold examination (3D-VIBE). DWI was acquired by echo-planar imaging (b-value = 0, 800 s/mm^2). After DWI scanning, the workstation automatically calculates and generates ADC map. The specific MRI parameters are shown in **Table 1**. When 3D-VIBE sequence was used to obtain DCEI, patients were instructed to hold their breath at the end of expiratory breath. A high pressure syringe (Spectris MR injection system, Medrad Inc.) was used to administer gadolinium diethylenetriamine penta-acetic acid (Bayer Healthcare Pharmaceuticals) through the cubital vein at a rate of 2 mL/s. The dosage of Gd-DTPA was 0.1 mmol/kg. Images of arterial phase, venous phase and delay phase were obtained by scanning at 25 s, 60 s and 180 s after administration.

MRI Evaluation

Additionally, all EC scans were independently evaluated by two radiologists with more than 10 years of experience in pelvic MRI diagnosis. The collected MR indicators included positive/negative DMI, maximum tumor diameter, positive/negative pelvic lymph nodes (PLN), and positive/negative abdominal

TABLE 1 | MRI Scanning Protocols.

Sequences	Plane	FS	TR/TE (ms)	FA (deg)	Slice thickness/Interslice gap (mm)	Matrix	FOV (mm)	Pixel size (mm)
T2TSE	SAG	Yes	4340/92	150	4/0.4	320×256	280×224	0.9×0.9
T2TSE	COR	Yes	4340/92	150	4/0.4	320×256	280×224	0.9×0.9
T2TSE	AO	Yes	4340/92	150	4/0.4	320×256	280×224	0.9×0.9
DWI	AO	Yes	7000/80	90	4/0.4	256×205	280×224	1.1×1.1
VIBE	AO	Yes	4.44/2.16	10	4/0	320×256	280×224	0.9×0.9

AO, axial oblique slice orientation; COR, coronal slice orientation; Deg, degrees; DWI: diffusion weighted imaging; FA, flip angle; FOV, field of view; SAG, sagittal slice orientation; TE, time echo; TR, repetition time; TSE, turbo spin echo; VIBE, volumetric interpolated breath-hold examination.

para-aortic lymph nodes (PALN). A lymph node with a short diameter of ≥1cm or with circular enhancement with central necrosis on enhanced scan is considered positive (19). After the evaluation, the intra-class correlation coefficient (ICC) and Kappa value of each index reported by the two radiologists were calculated. If the two indexes were greater than 0.75, the parameters were considered to be stable.

Tumor Image Segmentation and Radiomics Parameter Extraction

First, the axial diffusion fs-T2WI, DWI and 3D-VIBE images were downloaded from the Picture Archiving and Communication System (PACS). Axial-T2WI was used as a reference image, and axial-ADC, axial-3D-VIBE are registered to T2WI using Statistical Parametric Mapping software 12 (SPM12; University College London). Subsequently, the two radiologists mentioned above performed a layer-by-layer manual delineation of the volume of interest (VOI) on T2WI for all patients, focusing on covering the entire tumor. VOIs and tumor images, including T2WI, ADC and arterial, venous and delayed 3D-VIBE are imported one by one into the PyRadiomics toolkit version 3.0. The ICC of each parameter extracted by the radiologists was calculated, and the ICC>0.75 parameter was included. **Figure 2** and **Supplementary Methods S1** illustrate the process of VOI delineation, parameter extraction, and modelling in a patient with EC.

Feature Selection and Radiomics Signature Construction

The radiomics data is normalized and pre-processed using FAE software (FAE, https://github.com/salan668/FAE, version 0.3.6). "Normalise to unit" was used to normalize the data in order to reduce large differences in the values of the different radiomics features. Pearson's correlation coefficients (PCCs) were calculated for pre-processing. When the PCC is larger than the threshold value, one of the radiomics feature is removed randomly. See the **Supplementary Methods S2** for specific methods and calculation formulas. The Using X&Y software (X&Y Solutions, Inc.), the parameters related to PFS status were selected using the Least Absolute Shrinkage and Selection Operator (LASSO) regression method in the training cohort (**Supplementary Method S3**). The individualised radiomics based nomogram model, incorporating the radiomics signature and independent clinical risk factors, was constructed using the logistic regression.

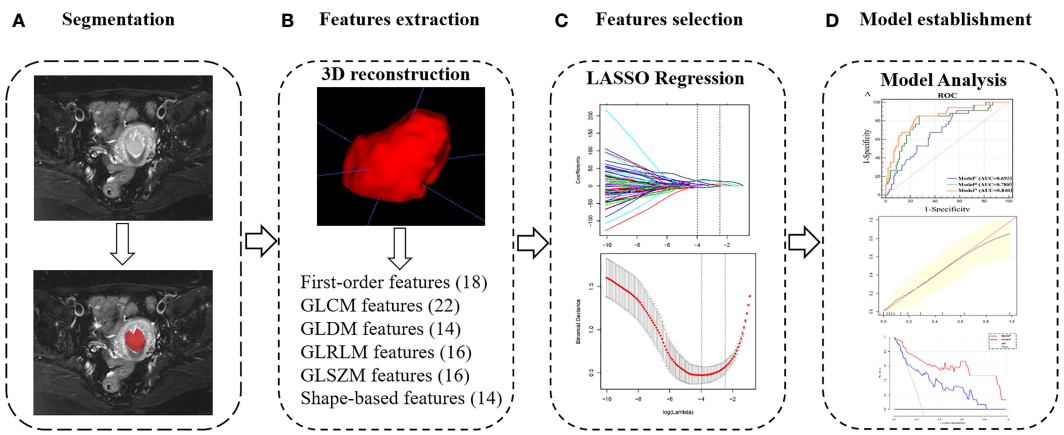


FIGURE 2 | Radiomics workflow of model construction. **(A)** MR images segmentation. The tumor was segmented manually on the axial T2-weighted images. **(B)** Texture features extraction. After 3D reconstruction, a total of 100 texture parameters of 6 types were extracted from each set of images. **(C)** Texture features selection. After the parameters were normalized and dimensionality reduced, the characteristic parameters were selected and classified by LASSO regression. **(D)** Model establishment. Combined with two clinical indicators of location and size, nomogram was developed to establish a preoperative evaluation model, and it was evaluated according to receiver-operating characteristic, calibration and decision curves.

Assessment and Validation of Model Performance

The area under the curve obtained by ROC analysis was used to evaluate the differentiating ability of the model (20). Calibration curves were used to assess the predictive power of the model, and the actual classification and the Hosmer–Lemeshow test was performed to assess the goodness-of-fit (21). The clinical efficacy of the model was evaluated by using the net benefits with different threshold probabilities obtained from the decision curve analysis (DCA) in the test cohort (22).

We also conducted sensitivity analysis of for our prediction Model^{N} to judge the diagnostic efficiency of the model between different pathological grades and stages. The FIGO staging criteria revised in 2009 for EC were used for histological diagnosis, grading, and pathological staging (23).

Pathological grading subgroup: All patients (including the training and test cohorts) were divided into two subgroups according to pathological grades. G1 and G2 endometrioid adenocarcinoma were classified as low-grade subgroup, and G3 or non-endometrial carcinoma (clear cell adenocarcinoma, serous adenocarcinoma, etc.) was classified as high grade subgroup (24).

Pathological staging subgroup: All patients (including training and test cohorts) were divided into two subgroups according to their pathological stages. Stage I and II patients were defined as low stage subgroup, while III and IV patients were defined as high stage subgroup.

Statistical Analysis

All statistical analyses were conducted using X&Y software based on R software. Univariable and multivariable logistic regression analyses were performed to identify the independent clinical risk factors associated with 5-year PFS. The candidate factors for univariable analysis were age, BMI, HE4, CA125, hypertension, diabetes, maximum diameter, DMI, PALN, PLN. Beta value, odds ratio and their 95% confidence interval (CI) were calculated. The variables with a P-value <0.10 in the univariable and multivariable analysis were selected as independent risk factors. A two-tailed P-value< 0.05 was considered statistically significant.

RESULTS

Clinical and MR Indicators

A total of 202 patients were eventually enrolled and analysed based on the inclusion and exclusion criteria. According to their 5-year PFS status, there were 49 cases in the high-risk group and 153 cases in the low-risk group. The ICCs of clinical and MR indicators were greater than 0.75, indicating a good agreement between the two measurements. Sample sizes, baseline clinical characteristics and pathological characteristics of the two groups are shown in **Tables 2**, **3**. Subsequently, univariate and multivariate analyses showed that there were significant differences in age, BMI, HE4, maximum diameter, and PALN between the two groups (all P<0.05, **Table 4**). These factors could be independent clinical risk factors for preoperative evaluation of 5-year survival status of EC patients.

Evaluation of Diagnostic Efficacy of Clinical Indicators

Four independent clinical risk factors including age, BMI, HE4, maximum diameter and PALN were combined with logistic regression to construct a clinical prediction model (Model^{C}) and develop a ROC curve to evaluate the preoperative prediction ability of the model for tumors of two different risk grades. The AUCs were 0.695 [95% confidence interval (CI), 0.612-0.770] and 0.828 (95% CI, 0.709–0.912) for the training and test cohorts, respectively (**Figure 3**).

Radiomics Score (RS) and the Diagnostic Efficacy for Our Radiomics Model

A total of 4 parameters with a non-zero coefficient are selected by LASSO regression, namely, ADC entropy, ADC Kurtosis, T2 Kurtosis, and Arterial HGLRE. RS values for each patient were calculated based on their respective coefficients in regression equation. Equation is as follows: RS=0.19213 +1.97154* ADC entropy+37.45352* ADC kurtosis+13.73094* T2 kurtosis+10.94133* Arterial HGLRE.

The RS for each patient was calculated to build radiomics model (Model^{R}) to predict 5-PFS. The AUCs for Model^{R} were 0.788 [95% CI, 0.712-0.853] and 0.887 (95% CI, 0.780–0.954) for the training and test cohorts, respectively (**Figure 3**). There was no statistically significant difference between the AUC of Model^{R} and Model^{C} (P=0.167 and 0.493 for the training and test cohort, respectively).

Radiomics Based Nomogram Model (Model^{N}) Establishment and Performance

The addition of radiomics parameters can improve the discriminative ability of Model^{C}. The nomogram achieved excellent performance in predicting risk grading, with AUC of 0.840 (95% CI: 0.769–0.896) in the training and 0.958 (95% CI: 0.873–0.993) in the test cohort. The predictive ability of the nomogram was better than that of the Model^{C} in the training cohort (P<0.001) and test cohort (P=0.032). The predictive ability of the nomogram was also better than that of the Model^{R} in the training cohort (P=0.021). However, in the test cohort, there was no statistically significant difference between the AUC of Model^{N} and Model^{R} (P = 0.106). In **Figure 4**, we include two representative MRI images illustrating visually striking differences in tumor heterogeneity between a patient who survived 26 months and another who succumbed at 8 months. T2WI, ADC and arterial-phased 3D-VIBE images of two representative cases with their Model^{N} score and predicted 5-year PFS are shown in **Figure 4**.

The calibration curve shows that the predicted value of the model is in good agreement with the actual value (P = 0.610 and P = 0.956 for the training and test cohorts, respectively). We calculated the risk scores for all patients in the training set and the test set to visually display the prediction ability of the model (**Figure 5**). The DCA indicates that the clinical application of Model^{N} has a better performance than that of Model^{C}, which also added more benefit than assuming that all cases are positive (high-risk EC) or negative (high-risk EC) (**Figure 6**).

TABLE 2 | Patient characteristics in the training and test cohorts.

Characteristics	Training cohort (*n* = 141)			Test cohort (*n* = 61)		*P*-value
	LR (*n* = 107)	HR (*n* = 34)	*P*-value	LR (*n* = 46)	HR (*n* = 15)	
Age (years)			**0.021**			0.729
Mean ± SD	51.1 ± 12.8	57.3 ± 15.0		53.8 ± 12.9	61.1 ± 10.9	
Range	30.0-88.0	35.0-90.0		30.0-94.0	45.0-75.0	
BMI			**0.007**			**0.009**
Mean ± SD	26.4 ± 6.2	29.8 ± 6.6		26.5 ± 6.9	30.3 ± 6.2	
Range	19.8-39.5	21.1-40.1		19.1-38.5	21.8-38.3	
CA125			0.118			0.586
Mean ± SD	91.7 ± 31.4	327.4 ± 272.9		93.3 ± 26.7	97.2 ± 32.5	
Range	16.0-162.0	21.0-160.0		32.0-141.0	39.0-154.0	
HE4			**0.043**			**0.041**
Mean ± SD	112.7 ± 43.5	130 ± 40.3		111.8 ± 41.6	139.7 ± 54.1	
Range	23.0-264.0	58.0-243.0		12.0-257.0	41.0-207.0	
Hypertension			0.285			0.493
No	68 (63.6%)	25 (73.5%)		32 (69.6%)	9 (60.0%)	
Yes	39 (36.4%)	9 (26.5%)		14 (30.4%)	6 (40.0%)	
Diabetes			0.703			0.076
No	73 (68.2%)	22 (64.7%)		33 (71.7%)	7 (46.7%)	
Yes	34 (31.8%)	12 (35.3%)		13 (28.3%)	8 (53.3%)	
Maximum Diameter			**0.002**			0.081
Mean ± SD	4.1 ± 1.0	4.8 ± 1.0		4.0 ± 1.1	4.6 ± 1.1	
Range	2.1-6.2	3.2-6.5		2.3-6.1	2.5-5.7	
DMI			0.363			0.597
No	44 (41.1%)	17 (50.0%)		22 (47.8%)	6 (40.0%)	
Yes	63 (58.9%)	17 (50.0%)		24 (52.2%)	9 (60.0%)	
PLN			0.132			0.929
No	66 (61.7%)	16 (47.1%)		27 (58.7%)	9 (60.0%)	
Yes	41 (38.3%)	18 (52.9%)		19 (41.3%)	6 (40.0%)	
PALN			**0.019**			0.076
No	87 (81.3%)	21 (61.8%)		40 (87.0%)	10 (66.7%)	
Yes	20 (18.7%)	13 (38.2%)		6 (13.0%)	5 (33.3%)	

LR, low-risk group; HR, high-risk group; BMI, body mass index; presence of hypertension and diabetes; CA125, carbohydrate antigen 125; HE4, human epididymis protein 4; HBP, high blood pressure; DMI, deep myometrial invasion; PLN, pelvic lymph nodes; PALN, para-aortic lymph nodes. P value was derived from the student-t or chi-square test. Bold type indicates statistically significant difference.

In the subgroup analysis, Model^{N} had a good differential diagnostic capability in all subgroups divided according to different criteria and there was no significant difference between subgroups. The AUC of the low- and high-grade subgroups were 0.871 and 0.926. The AUC of the low stage and high stage subgroups were 0.873 and 0.831. There was also no significant difference between the subgroups of different grade and stage (**Figure S1**).

DISCUSSION

In this study, a radiomics model based on multi-parameter MRI has been established to predict 5-year PFS in EC patients. The model combined the radiomics parameters obtained by preoperative MR examination with those easily obtained preoperatively to provide predictive information for long-term prognosis. Compared with clinical and radiomics models, this

TABLE 3 | Pathological characteristics of the patients in our study.

Characteristics	LR (*n* = 153)	HR (*n* = 49)	*P*-value
Pathological staging, n (%)			**0.0004**
pI	53 (34.6%)	15 (30.6%)	
pII	74 (48.4%)	16 (32.7%)	
pIII	20 (13.1%)	14 (28.6%)	
pIV	6 (3.9%)	4 (8.1%)	
Histological grade, n (%)			**0.0232**
G1	71 (46.4%)	13 (26.5%)	
G2	48 (31.4%)	17 (34.7%)	
G3 and non-endometrial carcinoma	34 (22.2%)	19 (38.8%)	

LR, low-risk group; HR, high-risk group. Bold type indicates statistically significant difference.

TABLE 4 | Preoperative clinical risk factors for 5-year PFS in patients with EC.

Variable	Univariable analysis		Multivariable analysis	
	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value
Age (years)	1.037 (1.012, 1.062)	**0.004**	1.035 (1.007, 1.064)	**0.013**
BMI	1.085 (1.032, 1.140)	**0.001**	1.084 (1.025, 1.146)	**0.005**
CA125	1.000 (1.000, 1.000)	0.114		
HE4	1.010 (1.003, 1.018)	**0.006**	1.011 (1.003, 1.020)	**0.010**
Hypertension		0.604		
No	1.000			
Yes	0.832 (0.416, 1.664)			
Diabetes		0.193		
No	1.000			
Yes	1.555 (0.800, 3.025)			
Maximum Diameter	1.934 (1.367, 2.737)	**<0.001**	2.000 (1.362, 2.936)	**<0.001**
DMI		0.641		
No	1.000			
Yes	0.858 (0.450, 1.636)			
PLN		0.229		
No	1.000			
Yes	1.488 (0.779, 2.843)			
PALN		**0.004**		**0.023**
No	1.000		1.000	
Yes	2.836 (1.383, 5.814)		2.590 (1.140, 5.885)	

BMI, body mass index; presence of hypertension and diabetes; CA125, carbohydrate antigen 125; HE4, human epididymis protein 4; HBP, high blood pressure; DMI, deep myometrial invasion; PLN, pelvic lymph nodes; PALN, para-aortic lymph nodes. Bold type indicates statistically significant difference.

comprehensive model provides better discrimination ability, enabling clinicians to grade tumor risk preoperatively, which can be used to guide treatment decisions. The discriminative ability of nomogram model discrimination ability was also demonstrated in subgroup analysis.

Four radiomics parameters (ADC entropy, ADC Kurtosis, T2 Kurtosis, Arterial HGLRE) are selected to calculate the RS in this study. Several previous studies have used radiomics-based models to predict a prognosis in a variety of tumors (25–27). For example, recent studies have shown that RS-based models can predict EC lymph node metastasis and LVSI (16, 28). A previous CT study has found that a high tumor entropy independently predicted deep myometrial invasion (odds ratio [OR] 3.7, p=0.008) and cervical stroma invasion (OR 3.9, p=0.02) (29). In addition, a high tumor kurtosis tends to independently predict a reduced recurrence- and PFS (HR 1.1, p=0.06) (29). Another MR study also found that MR was a sensitive indicator for PFS. High kurtosis in T1 c images predicted a reduced recurrence- and progression-free survival (hazard ratio [HR] 1.5, P < 0.001) after adjusting for MRI-measured tumor volume and histological risk at biopsy (30). High tumor entropy in apparent diffusion coefficient (ADC) maps independently predicted deep myometrial invasion (odds ratio [OR] 3.2, P < 0.001) (30). It is not difficult to find that the parameters screened in the above studies are similar to the radiomics parameters in this study.

The parameters screened in this study also overlapped with other tumor prognostic parameters. A previous study has shown that a radiomics model can more accurately predict 3-year and 5-year PFS for advanced nasopharyngeal carcinoma than a clinical model based on TNM stage alone (31). A previous breast cancer study found that entropy can be used as a predictor of benign, malignant and risk assessment of tumors, with an AUC of 0.8 and a sensitivity of 95% when applied alone (32). A study found that Kurtosis combined with several clinical and other texture parameters could predict eight-year event-free survival (EFS) in Luminal Non-Metastatic Breast Valencia (33). In addition to adenocarcinoma, another study on anal squamous cell carcinoma also found that Entropy and Joint Energy can be independent risk factors for predicting tumor recurrence rate (34). Therefore, the radiomics indicators screened in this study are not only reproducible in the evaluation of multiple biological characteristics and prognosis of endometrial cancer, but also seem to be similar in other tumors.

Previous studies generally report that the older the onset age, the higher the risk of endometrial cancer recurrence and death (7).

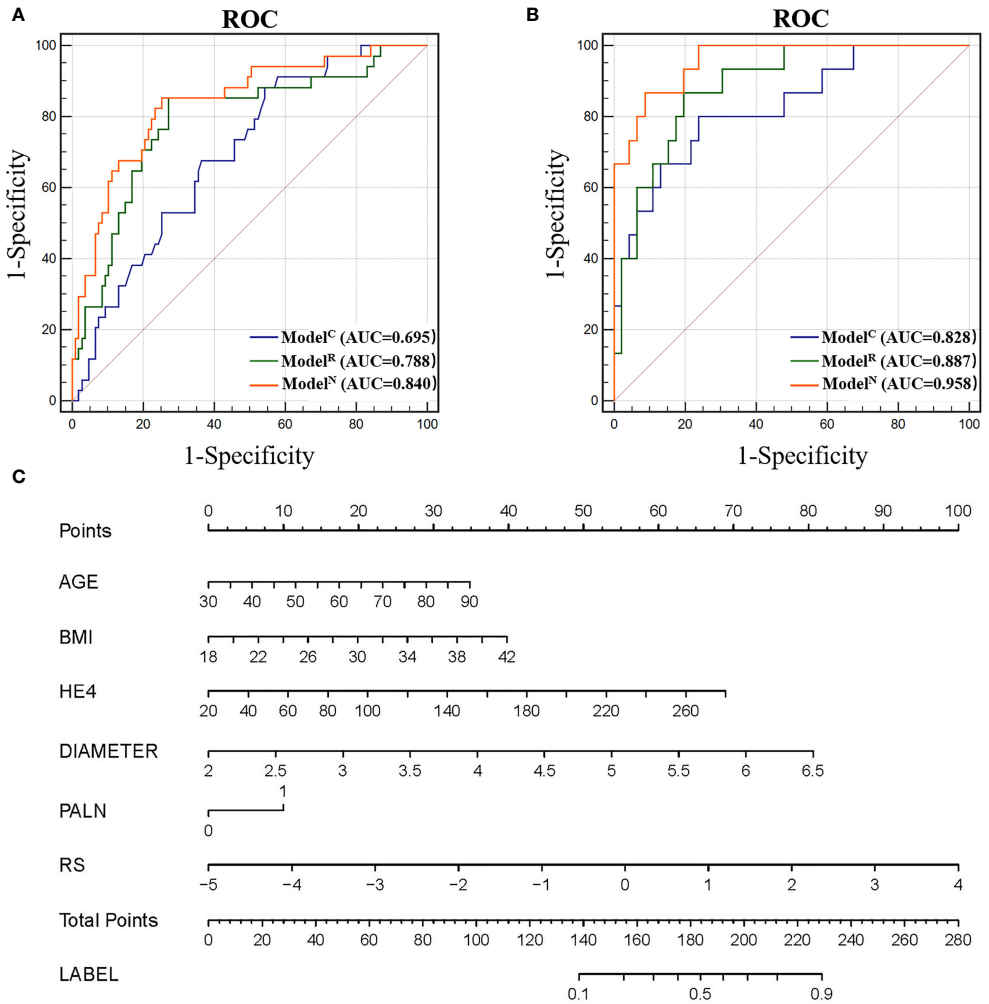


FIGURE 3 | **(A)** Receiver operating characteristic (ROC) of different models in the training cohort. **(B)** ROC of different models in the test cohort. **(C)** Nomogram for predicting risk classification of EC. The nomogram was built in the training cohort with the Radscore, BMI and CA125. The probability of each predictor can be converted into scores according to the first scale points at the top of the nomogram. After adding up the scores of these predictors in total points, the corresponding prediction probability at the bottom of the nomogram is the malignancy of the tumor.

This may be due to the fact that elderly patients are more prone to high-grade or specific histological types of EC and to various complication (35). This study also found that the higher the BMI for our EC patients, the more likely they were to relapse, which may be related to the increase of oestrogen level caused by obesity (8). In addition, HE4 has recently been identified as a potential biomarker for endometrial cancer with higher sensitivity than CA125 (36). The high expression of this marker was associated with International Federation of Gynaecology and Obstetrics (FIGO) grade, histological stage, and mortality (37). This study found that this indicator can also be used to assess the risk of tumor recurrence.

The maximum tumor diameter and para-aortic lymph node metastasis could also be independent risk factors for predicting the risk of tumor recurrence. A previous study suggested that lymph node dissection should be considered for all patients with the maximum diameter over 35 mm to prevent postoperative recurrence (38). In addition to being a risk factor for lymph node metastasis, tumor size is also a risk factor for cervical invasion, which is more likely to occur when the tumor diameter is larger than 3 cm (39). This study found that para-aortic lymph node metastases were associated with EC 5-year PFS, but not with pelvic lymph node metastases. Although it needs to be carefully verified, we speculate that lymph metastases at the first site may not significant affect the prognosis. A previous study found that, in endometrial cancer with stage IIIC disease, only when the second lymphatic station, like PALN is invaded it may indicate that the tumor has a strong invasive ability and a poor prognosis (40).

The whole tumor profile method was used to obtain all tumor information in this study, which is more accurate, though more time consuming than single-layer measurements. Previous

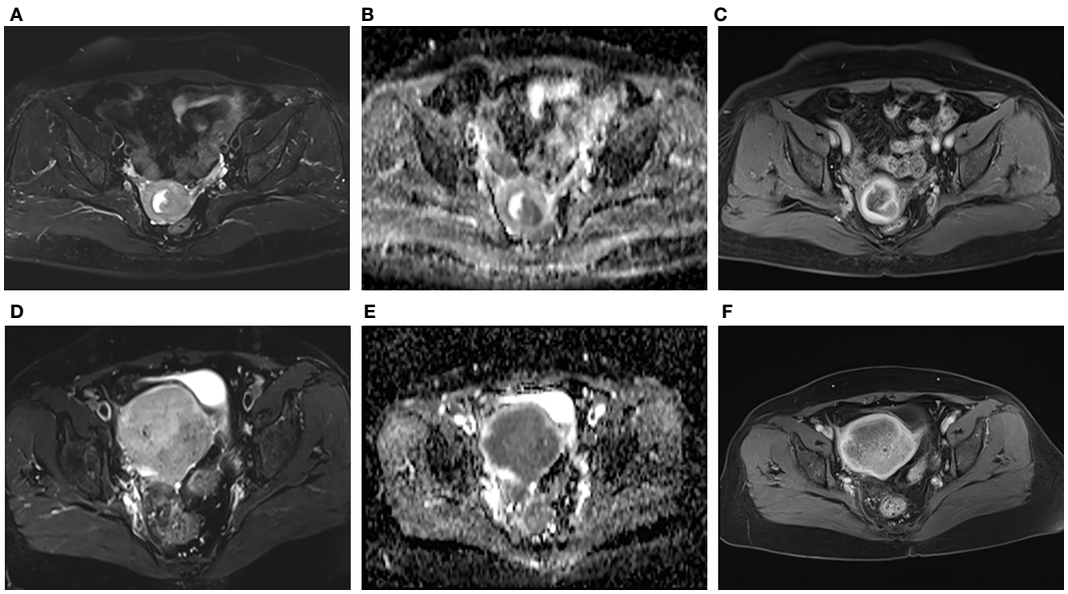


FIGURE 4 | Representative images of two patients with different survival outcomes. **(A–C)** axial oblique T2WI, ADC map and arterial phase images of a 55-year-old woman with low-risk EC. The nomogram model (Model^{N}) score was -1.37. Using the nomogram, the estimated probability of relapse or death within 5 years was 17%. The tumor did not recur during the 5-year observation period. **(D–F)** axial oblique T2WI, ADC map and arterial phase images of a 75-year-old woman with high-risk EC. The Model^{N} score was 2.92. Using the nomogram, the estimated probability of relapse or death within 5 years was 95%. The tumor relapsed 6 months after surgery.

studies have shown that the whole-tumor signatures outperformed single-slice signatures for prediction of LNM and advanced FIGO stage (41). Furthermore, we also used radiomics markers to predict postoperative pathological results such as DMI, FIGO, lymphatic metastasis and other established different models to indirectly evaluate the recurrence of tumor.

Based on the patients baseline clinical indicators and preoperative MR examination, the present study directly established a model to predict 5-PFS in EC patients, with the purpose of providing guidance for the clinical practice of endometrial cancer treatment and follow-up. Therefore, compared with other previous study, the present study was

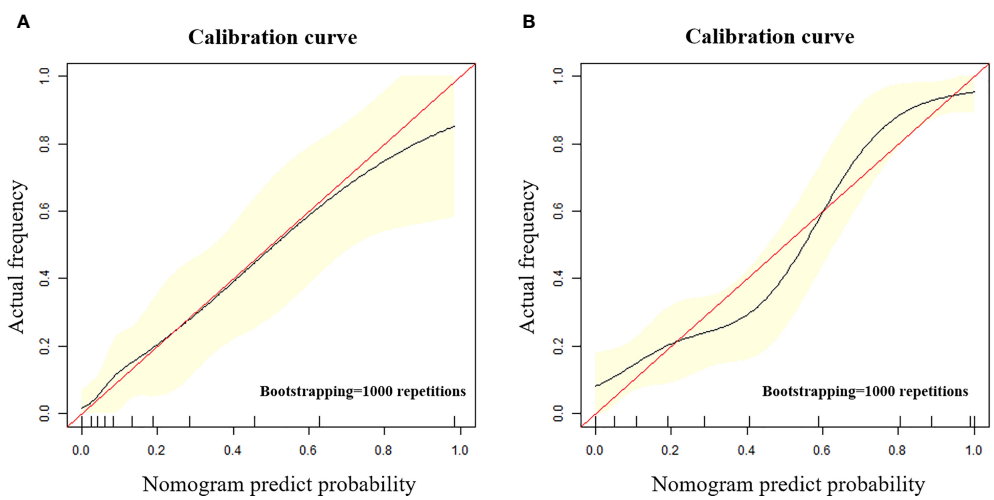


FIGURE 5 | **(A, B)** The calibration curve in the training cohort **(A)** and test cohort **(B)**. The calibration curve depicted the agreement between the predicted risk classification score and the actual results confirmed by examination. The red line represents an ideal prediction, and the black line represents the predictive performance. The closer the fit of the black line to the ideal line, the better the prediction.

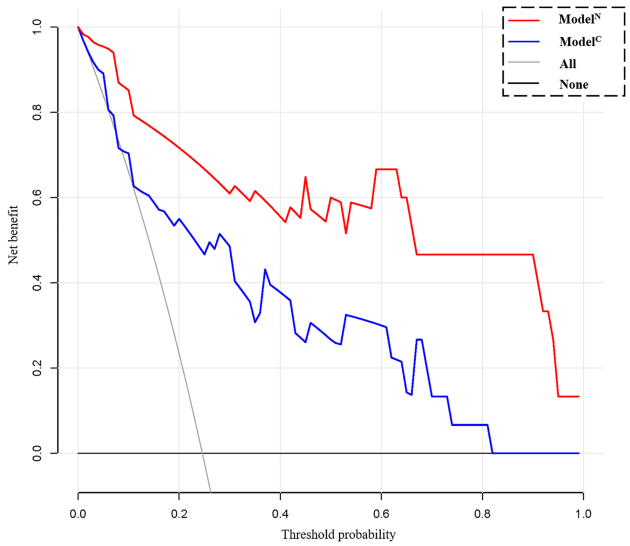


FIGURE 6 | Decision curve analysis (DCA) for the clinical model (Model^{C}) and nomogram model (Model^{N}). The decision curve showed that using a clinical- radiomics combined model to predict 5-year progression-free survival (PFS) of endometrial cancer would be more beneficial than the clinical model.

able to assess the tumor risk in the absence of preoperative pathological results. Based on the nomogram proposed in this study, we can calculate the patient score and make an accurate preoperative prediction of the 5-year recurrence and survival probability for each EC patient. In this way, it is more reasonable for clinicians to take more active treatment measures as soon as possible for patients suspected of having a higher risk of recurrence. Furthermore, more active follow-up should be carried out after surgery to detect recurrent lesions and intervene as soon as possible, to prolong and improve the quality of life. For patients with low risk, the rational application of this model can avoid overly aggressive surgical plans formulated by physicians, and also reduce the use of many unnecessary postoperative treatments (such as chemotherapy and radiotherapy). This provides a new strategy to avoiding pain, unnecessary economic loss and waste of medical resources caused by overtreatment.

This study has the following limitations that should be considered. First, this is a retrospective study and only includes those patients who had undergone surgery, which inevitably led to selective bias. Second, as a single centre study, whether the model proposed in this study is applicable to other MR systems remains unknown. Third, the sample size of this study is small, and the results need to be verified by large sample studies. In order to ensure the sample size for the training cohort and the accuracy of the model establishment, the samples have an unbalanced distribution. The small sample size of the test cohort may increase the uncertainty of its results. In addition, there are several extracted features, which may lead to the failure to include some indicators with strong correlation with PFS.

In conclusion, the radiomics model which incorporates clinical and cMR indicators was a good predictor of the relapse risk of EC. Using our radiomic parameter-based model and nomogram analysis can help guide preoperative non-invasive individualized evaluation for 5-year PFS and avoid possible under- or over-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 the First Hospital of Qinhuangdao. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. 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

DL, LY, and YC designed and coordinated the study, DL, DD, TZ, LL, ZW, JD, and YD carried out experiment and data process, and drafted the manuscript. All authors gave final approval for publication.

FUNDING

This research was supported by National Natural Science Foundation of China (81871029) and Scientific Research Fund Project of Health Commission of Hebei Province (20200138).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.813069/full#supplementary-material

REFERENCES

1. Lu KH, Broaddus RR. Endometrial Cancer. *N Engl J Med* (2020) 383 (21):2053–64. doi: 10.1056/NEJMra1514010
2. Koh WJ, Abu-Rustum NR, Bean S, Bradley K, Campos SM, Cho KR, et al. Uterine Neoplasms, Version 1.2018, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Cancer Netw: JNCCN* (2018) 16(2):170–99. doi: 10.6004/jnccn.2018.0006
3. Keys HM, Roberts JA, Brunetto VL, Zaino RJ, Spirtos NM, Bloss JD, et al. A Phase III Trial of Surgery With or Without Adjunctive External Pelvic Radiation Therapy in Intermediate Risk Endometrial Adenocarcinoma: A Gynecologic Oncology Group Study. *Gynecol Oncol* (2004) 92(3):744–51. doi: 10.1016/j.ygyno.2003.11.048
4. Randall ME, Filiaci V, McMeekin DS, von Gruenigen V, Huang H, Yashar CM, et al. Phase III Trial: Adjuvant Pelvic Radiation Therapy Versus Vaginal Brachytherapy Plus Paclitaxel/Carboplatin in High-Intermediate and High-Risk Early Stage Endometrial Cancer. *J Clin Oncol: Off J Am Soc Clin Oncol* (2019) 37(21):1810–8. doi: 10.1200/JCO.18.01575
5. Cymbaluk-Płoska A, Gargulińska P, Bulsa M, Kwiatkowski S, Chudecka-Głaz A, Michalczyk K. Can the Determination of HE4 and CA125 Markers Affect the Treatment of Patients With Endometrial Cancer? *Diagnostics (Basel Switzerland)* (2021) 11(4):626. doi: 10.3390/diagnostics11040626
6. Jónsdóttir B, Marcickiewicz J, Borgfeldt C, Bjurberg M, Dahm-Kähler P, Flöter-Rådestad A, et al. Preoperative and Intraoperative Assessment of Myometrial Invasion in Endometrial Cancer - A Swedish Gynecologic Cancer Group (SweGCG) Study. *Acta Obstet Gynecol Scand* (2021) 100 (8):1526–33. doi: 10.1111/aogs.14146
7. Chen T, Jansen L, Gondos A, Ressing M, Holleczek B, Katalinic A, et al. Survival of Endometrial Cancer Patients in Germany in the Early 21st Century: A Period Analysis by Age, Histology, and Stage. *BMC Cancer* (2012) 12:128. doi: 10.1186/1471-2407-12-128
8. Allen NE, Key TJ, Dossus L, Rinaldi S, Cust A, Lukanova A, et al. Endogenous Sex Hormones and Endometrial Cancer Risk in Women in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *Endocrine-Related Cancer* (2008) 15(2):485–97. doi: 10.1677/ERC-07-0064
9. Yue W, Meng N, Wang J, Liu W, Wang X, Yan M, et al. Comparative Analysis of the Value of Diffusion Kurtosis Imaging and Diffusion-Weighted Imaging in Evaluating the Histological Features of Endometrial Cancer. *Cancer Imaging: Off Publ Int Cancer Imaging Soc* (2019) 19(1):9. doi: 10.1186/s40644-019-0196-6
10. Otero-García MM, Mesa-Álvarez A, Nikolic O, Blanco-Lobato P, Basta-Nikolic M, de Llano-Ortega RM, et al. Role of MRI in Staging and Follow-Up of Endometrial and Cervical Cancer: Pitfalls and Mimickers. *Insights Imaging* (2019) 10(1):19. doi: 10.1186/s13244-019-0696-8
11. Ippolito D, Minutolo O, Cadonici A, Talei Franzesi C, Bonaffini P, Perego P, et al. Endometrial Cancer: Diagnostic Value of Quantitative Measurements of Microvascular Changes With DCE-MR Imaging. *Magma* (2014) 27(6):531–8. doi: 10.1007/s10334-014-0435-6
12. Satta S, Dolciami M, Celli V, Di Stadio F, Perniola G, Palaia I, et al. Quantitative Diffusion and Perfusion MRI in the Evaluation of Endometrial Cancer: Validation With Histopathological Parameters. *Br J Radiol* (2021) 94 (1125):20210054. doi: 10.1259/bjr.20210054
13. Lubner MG, Smith AD, Sandrasegaran K, Sahani DV, Pickhardt PJ. CT Texture Analysis: Definitions, Applications, Biologic Correlates, and Challenges. *Radiographics: Rev Publ Radiological Soc North Am Inc* (2017) 37(5):1483–503. doi: 10.1148/rg.2017170056
14. Badic B, Tixier F, Cheze Le Rest C, Hatt M, Visvikis D. Radiogenomics in Colorectal Cancer. *Cancers* (2021) 13(5):973. doi: 10.3390/cancers13050973
15. Fan M, Chen H, You C, Liu L, Gu Y, Peng W, et al. Radiomics of Tumor Heterogeneity in Longitudinal Dynamic Contrast-Enhanced Magnetic Resonance Imaging for Predicting Response to Neoadjuvant Chemotherapy in Breast Cancer. *Front Mol Biosci* (2021) 8:622219. doi: 10.3389/fmolb.2021.622219
16. Luo Y, Mei D, Gong J, Zuo M, Guo X. Multiparametric MRI-Based Radiomics Nomogram for Predicting Lymphovascular Space Invasion in Endometrial Carcinoma. *J Magn Reson Imaging* (2020) 52(4):1257–62. doi: 10.1002/jmri.27142
17. Ueno Y, Forghani B, Forghani R, Dohan A, Zeng XZ, Chamming's F, et al. Endometrial Carcinoma: MR Imaging-Based Texture Model for Preoperative Risk Stratification-A Preliminary Analysis. *Radiology* (2017) 284(3):748–57. doi: 10.1148/radiol.2017161950
18. Han Y, Xu H, Ming Y, Liu Q, Huang C, Xu J, et al. Predicting Myometrial Invasion in Endometrial Cancer Based on Whole-Uterine Magnetic Resonance Radiomics. *J Cancer Res Ther* (2020) 16(7):1648–55. doi: 10.4103/jcrt.JCRT_1393_20
19. Meissnitzer M, Forstner R. MRI of Endometrium Cancer - How We Do It. *Cancer Imaging: Off Publ Int Cancer Imaging Soc* (2016) 16:11. doi: 10.1186/s40644-016-0069-1
20. 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(3):837–45. doi: 10.2307/2531595
21. Han DS, Suh YS, Kong SH, Lee HJ, Choi Y, Aikou S, et al. Nomogram Predicting Long-Term Survival After D2 Gastrectomy for Gastric Cancer. *J Clin Oncol: Off J Am Soc Clin Oncol* (2012) 30(31):3834–40. doi: 10.1200/JCO.2012.41.8343
22. Hijazi Z, Oldgren J, Lindbäck J, Alexander JH, Connolly SJ, Eikelboom JW, et al. The Novel Biomarker-Based ABC (Age, Biomarkers, Clinical History)-Bleeding Risk Score for Patients With Atrial Fibrillation: A Derivation and Validation Study. *Lancet (London England)* (2016) 387(10035):2302–11. doi: 10.1016/S0140-6736(16)00741-8
23. FIGO Committee on Gynecologic Oncology. *FIGO Staging for Carcinoma of the Vulva, Cervix, and Corpus Uteri. International Journal of Gynaecology and Obstetrics: The Official Organ of the International Federation of Gynaecology and Obstetrics*. Amsterdam: Elsevier B.V (2014) 125:97–8.
24. Miyamoto M, Takano M, Aoyama T, Soyama H, Yoshikawa T, Tsuda H, et al. Seromucinous Component in Endometrioid Endometrial Carcinoma as a Histological Predictor of Prognosis. *J Gynecol Oncol* (2018) 29(2):e20. doi: 10.3802/jgo.2018.29.e20
25. 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
26. 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
27. 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
28. Xu X, Li H, Wang S, Fang M, Zhong L, Fan W, et al. Multiplanar MRI-Based Predictive Model for Preoperative Assessment of Lymph Node Metastasis in Endometrial Cancer. *Front Oncol* (2019) 9:1007. doi: 10.3389/fonc.2019.01007
29. Ytre-Hauge S, Salvesen ØO, Krakstad C, Trovik J, Haldorsen IS. Tumour Texture Features From Preoperative CT Predict High-Risk Disease in Endometrial Cancer. *Clin Radiol* (2021) 76(1):79.e13–20. doi: 10.1016/j.crad.2020.07.037
30. 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

31. Yang K, Tian J, Zhang B, Li M, Xie W, Zou Y, et al. A Multidimensional Nomogram Combining Overall Stage, Dose Volume Histogram Parameters and Radiomics to Predict Progression-Free Survival in Patients With Locoregionally Advanced Nasopharyngeal Carcinoma. *Oral Oncol* (2019) 98:85–91. doi: 10.1016/j.oraloncology.2019.09.022
32. Brown AL, Jeong J, Wahab RA, Zhang B, Mahoney MC. Diagnostic Accuracy of MRI Textural Analysis in the Classification of Breast Tumors. *Clin Imaging* (2021) 77:86–91. doi: 10.1016/j.clinimag.2021.02.031
33. Aide N, Elie N, Blanc-Fournier C, Levy C, Salomon T, Lasnon C. Hormonal Receptor Immunochemistry Heterogeneity and (18)F-FDG Metabolic Heterogeneity: Preliminary Results of Their Relationship and Prognostic Value in Luminal Non-Metastatic Breast Cancers. *Front Oncol* (2020) 10:599050. doi: 10.3389/fonc.2020.599050
34. Giraud N, Saut O, Aparicio T, Ronchin P, Bazire LA, Barbier E, et al. MRI-Based Radiomics Input for Prediction of 2-Year Disease Recurrence in Anal Squamous Cell Carcinoma. *Cancers* (2021) 13(2):193. doi: 10.3390/cancers13020193
35. Wright JD, Lewin SN, Barrena Medel NI, Sun X, Burke WM, Deutsch I, et al. Endometrial Cancer in the Oldest Old: Tumor Characteristics, Patterns of Care, and Outcome. *Gynecol Oncol* (2011) 122(1):69–74. doi: 10.1016/j.ygyno.2011.02.040
36. Kalogera E, Scholler N, Powless C, Weaver A, Drapkin R, Li J, et al. Correlation of Serum HE4 With Tumor Size and Myometrial Invasion in Endometrial Cancer. *Gynecol Oncol* (2012) 124(2):270–5. doi: 10.1016/j.ygyno.2011.10.025
37. Karlsen NS, Karlsen MA, Høgdall CK, Høgdall EV. HE4 Tissue Expression and Serum HE4 Levels in Healthy Individuals and Patients With Benign or Malignant Tumors: A Systematic Review. *Cancer Epidemiol Biomarkers Prev: Publ Am Assoc Cancer Res Cosponsored by Am Soc Prev Oncol* (2014) 23 (11):2285–95. doi: 10.1158/1055-9965.EPI-14-0447
38. Riggs MJ, Cox Bauer CM, Miller CR, Aden JK, Kamelle SA. Validation of an Endometrial Tumor Diameter Model for Risk Assessment in the Absence of Lymph Node Mapping. *J Patient-Centered Res Rev* (2020) 7(4):323–8. doi: 10.17294/2330-0698.1768
39. Toprak S, Sahin EA, Sahin H, Tohma YA, Yilmaz E, Meydanli MM. Risk Factors for Cervical Stromal Involvement in Endometrioid-Type Endometrial Cancer. *Int J Gynaecol Obstet: Off Organ Int Fed Gynaecol Obstet* (2021) 153 (1):51–5. doi: 10.1002/ijgo.13449
40. Guo J, Qian H, Ma F, Zhang Y, Cui X, Duan H. The Characteristics of Isolated Para-Aortic Lymph Node Metastases in Endometrial Cancer and Their Prognostic Significance. *Ther Adv Med Oncol* (2020) 12:1758835920933036. doi: 10.1177/1758835920933036
41. Fasmer KE, Hodneland E, Dybvik JA, Wagner-Larsen K, Trovik J, Salvesen Ø, et al. Whole-Volume Tumor MRI Radiomics for Prognostic Modeling in Endometrial Cancer. *J Magn Reson Imaging* (2021) 53(3):928–37. doi: 10.1002/jmri.27444

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.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 22 April 2022
doi: 10.3389/fonc.2022.824509

Prediction of Response to Induction Chemotherapy Plus Concurrent Chemoradiotherapy for Nasopharyngeal Carcinoma Based on MRI Radiomics and Delta Radiomics: A Two-Center Retrospective Study

Yuzhen Xi[1,2†], Xiuhong Ge[1†], Haiming Ji[3], Luoyu Wang[1], Shaofeng Duan[4], Haonan Chen[5], Mengze Wang[5], Hongjie Hu[6], Feng Jiang[7] and Zhongxiang Ding[1*]*

[1] Department of Radiology, Key Laboratory of Clinical Cancer Pharmacology and Toxicology Research of Zhejiang Province, Affiliated Hangzhou First People's Hospital, Cancer Center, Zhejiang University School of Medicine, Hangzhou, China, [2] Department of Radiology, 903rd Hospital of PLA, Hangzhou, China, [3] Department of Radiology, Liangzhu Hospital, Hangzhou, China, [4] GE Healthcare, Precision Health Institution, Shanghai, China, [5] Zhejiang Chinese Medical University, Hangzhou, China, [6] Department of Radiology, Sir Run Run Shaw Hospital Affiliated to Medical College Zhejiang University, Hangzhou, China, [7] Department of Head and Neck Radiotherapy, Zhejiang Cancer Hospital/Zhejiang Province Key Laboratory of Radiation Oncology, Hangzhou, China

Objective: We aimed to establish an MRI radiomics model and a Delta radiomics model to predict tumor retraction after induction chemotherapy (IC) combined with concurrent chemoradiotherapy (CCRT) for primary nasopharyngeal carcinoma (NPC) in non-endemic areas and to validate its efficacy.

Methods: A total of 272 patients (155 in the training set, 66 in the internal validation set, and 51 in the external validation set) with biopsy pathologically confirmed primary NPC who were screened for pretreatment MRI were retrospectively collected. The NPC tumor was delineated as a region of interest in the two sequenced images of MRI before treatment and after IC, followed by radiomics feature extraction. With the use of maximum relevance minimum redundancy (mRMR) and least absolute shrinkage and selection operator (LASSO) algorithms, logistic regression was performed to establish pretreatment MRI radiomics and pre- and post-IC Delta radiomics models. The optimal Youden's index was taken; the receiver operating characteristic (ROC) curve, calibration curve, and decision curve were drawn to evaluate the predictive efficacy of different models.

Results: Seven optimal feature subsets were selected from the pretreatment MRI radiomics model, and twelve optimal subsets were selected from the Delta radiomics model. The area under the ROC curve, accuracy, sensitivity, specificity, negative

OPEN ACCESS

Edited by:
Oliver Diaz,
University of Barcelona, Spain

Reviewed by:
Caterina Giannitto,
Humanitas Research Hospital, Italy
Eros Montin,
NYU Grossman School of Medicine, United States

***Correspondence:**
Feng Jiang
jiangfeng@zjcc.org.cn
Zhongxiang Ding
hangzhoudzx73@126.com

[†]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: 30 November 2021
Accepted: 23 March 2022
Published: 22 April 2022

Citation:
Xi Y, Ge X, Ji H, Wang L, Duan S, Chen H, Wang M, Hu H, Jiang F and Ding Z (2022) Prediction of Response to Induction Chemotherapy Plus Concurrent Chemoradiotherapy for Nasopharyngeal Carcinoma Based on MRI Radiomics and Delta Radiomics: A Two-Center Retrospective Study. Front. Oncol. 12:824509. doi: 10.3389/fonc.2022.824509

predictive value (NPV), and positive predictive value (PPV) of the MRI radiomics model were 0.865, 0.827, 0.837, 0.813, 0.776, and 0.865, respectively; the corresponding indicators of the Delta radiomics model were 0.941, 0.883, 0.793, 0.968, 0.833, and 0.958, respectively.

Conclusion: The pretreatment MRI radiomics model and pre- and post-IC Delta radiomics models could predict the IC-CCRT response of NPC in non-epidemic areas.

Keywords: nasopharyngeal carcinoma, magnetic resonance imaging, radiomics, induction chemotherapy, concurrent chemoradiotherapy

INTRODUCTION

Nasopharyngeal carcinoma (NPC) is a malignant tumor originating from the nasopharyngeal mucosal epithelium, which is sensitive to radiotherapy (1). Global Cancer Statistics of 2020 indicated 133,354 new cases of NPC and 80,008 deaths worldwide (2, 3). Meanwhile, the incidence and mortality of NPC in China are higher than the global average estimate. Radiotherapy is the primary treatment for NPC, and concurrent chemoradiotherapy (CCRT) can improve the radiotherapy effect by shrinking the tumor, increasing radiosensitivity, and reducing the radiation dose (4). The 2019 National Comprehensive Cancer Network Clinical Practice Guidelines have recommended induction chemotherapy (IC) combined with CCRT (IC-CCRT) as a class 2A modality for the treatment of advanced NPC (5). In recent years, there has been increasing evidence that IC-CCRT or radiotherapy has clinical value in improving progression-free survival (PFS) and relapse-free survival (RFS) of NPC patients (6). IC-CCRT is also effective in the treatment of non-endemic areas of NPC (7). Although NPC treatment has improved with the advancement of chemoradiotherapy strategies, the 5-year survival rate of some patients with advanced disease is about 60%–85%, and the therapeutic efficacy remains unsatisfactory (8, 9).

The development of treatment options and the evaluation of prognosis for NPC mainly depend on the tumor node metastasis (TNM) stage. However, anatomy-based TNM staging only reflects the tumor shape and invasion into surrounding structures and ignores the internal characteristics of the tumor with the same stage. Hence, despite receiving similar treatment regimens, about 20% of patients show unsatisfactory results due to individual differences and tumor heterogeneity (10). Radiochemotherapy resistance remains one of the main causes of poor prognosis and treatment failure in NPC (11), while residual mass is an independent factor for poor prognosis (12, 13). Treatment of residual disease is associated with better survival outcomes compared to the treatment of recurrent tumors (14). As the tumor shrinks during treatment, adjacent normal brain tissue, skull base bone, and other tissues will fall into the high-dose irradiated tumor areas, increasing the risk of radiation-related injury (15). Therefore, it is necessary to reveal the heterogeneity of tumors as early as possible, facilitating to predict tumor shrinkage in individualized and precise treatment and prognosis of NPC patients.

Radiomics has become a popular method to study tumor heterogeneity in recent years. It can describe tumor heterogeneity and other features by mining the high-dimensional quantitative characteristics of standard medical images (CT, MRI, PET, etc.), providing clinical and high-throughput quantitative information and more personalized treatment options (16). Radiomics characteristics are usually defined in two ways (17), including single-time point radiology and Delta radiomics. Single-time point radiology is mostly used before or during treatment to establish a genomics characteristics model for diagnosis (18, 19), tumor risk stratification (20–23), and prognosis prediction (24, 25), associated with higher powers compared with the TNM staging system. Delta radiomics uses radiological features during or after treatment to provide a wealth of information to identify and quantify treatment-induced changes to guide clinical decisions. It may be more suitable for assessing tumor treatment efficacy (26). Some studies have shown that Delta radiomics-based models yield higher powers than single-time-point-based models (27–29).

Existing radiomics guidelines have recommended the use of multicenter data to ensure the generalizability of the findings (30, 31). However, for many studies, there are very little external validation data. In addition, Delta radiomics studies in predicting adverse events in head and neck squamous cell carcinoma are mostly based on CT imaging (32, 33). The aim of this study was to construct a pretreatment MRI radiomics model and Delta radiomics models before and after IC and to explore their application values in dynamically predicting chemoradiotherapy efficacy for the treatment of NPC in non-epidemic regions.

MATERIALS AND METHODS

Patients

The study was approved by the local scientific research ethics committee, and informed consent was waived due to its retrospective nature. The study process was in accordance with the Declaration of Helsinki. The information of 668 included patients with pathologically confirmed NPCat the Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital) was collected between January 2007 and June 2012. Then screening was performed according to the following conditions: 1) NPC diagnosis was pathologically confirmed; 2) patients were treated with IC-CCRT; 3) MRI

examination was performed within 2 weeks before and after CCRT treatment; 3) fat-suppressed (FS) T2-weighted imaging (T2WI) and FS contrast-enhanced T2-weighted imaging (CE-T1WI) images were available. The patient selection process is shown in **Figure 1**. Then the data of patients with primary NPC in Sir Run Run Shaw Hospital, Zhejiang University, during the same period were re-screened as external validation data. A total of 51 qualified cases were identified before treatment, including 33 qualified cases after IC.

According to the Response Evaluation Criteria in Solid Tumors (RECIST Version 1.1), NPC patients with response to IC-CCRT were assigned to the complete tumor retraction group.

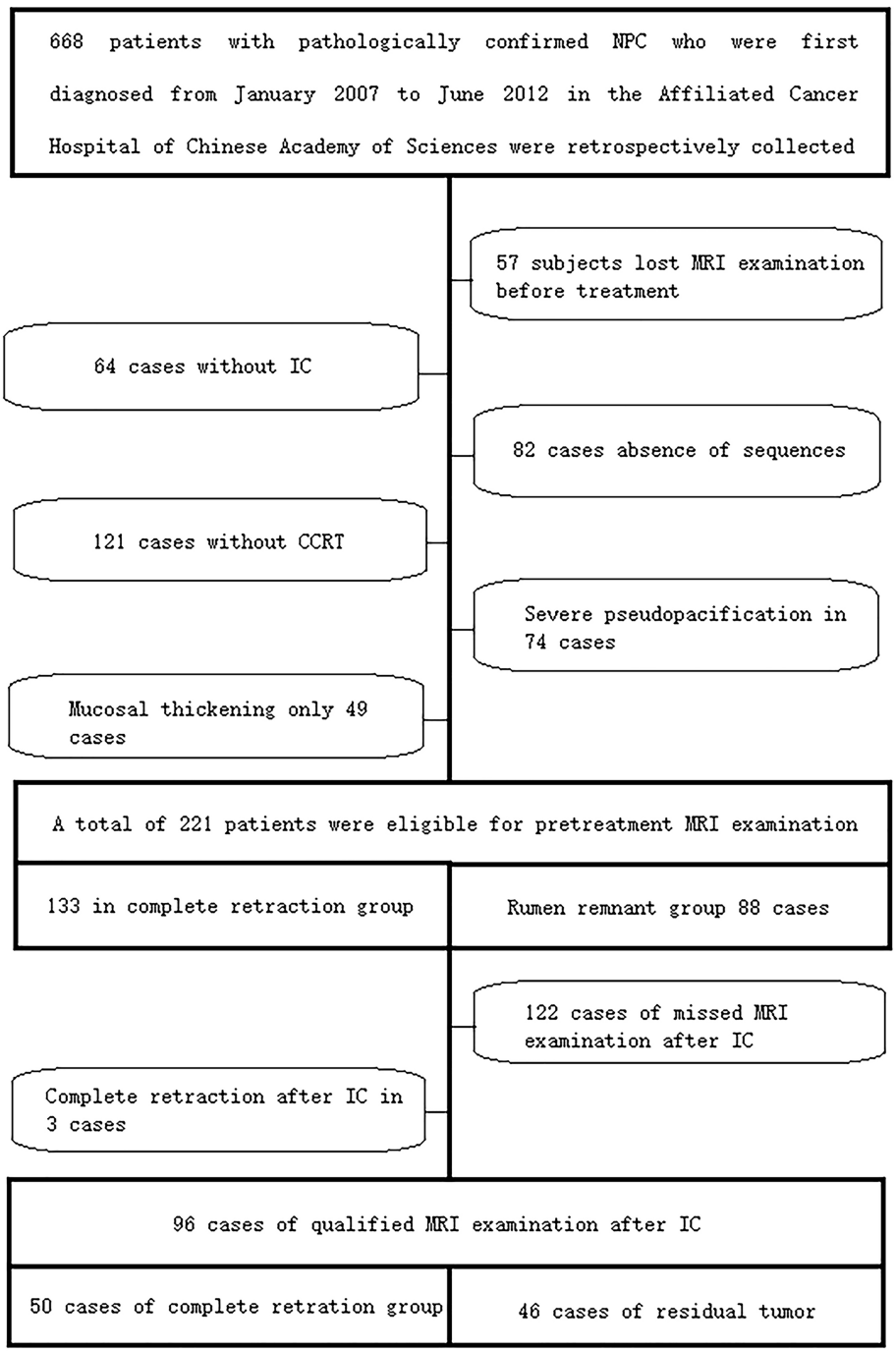


FIGURE 1 | The case screening process in the training set and internal validation set. A total of 221 qualified cases were screened before the final treatment, including 96 qualified cases after IC. IC, induction chemotherapy.

The total retraction group was defined as no evidence of residual disease on reexamination of MRI, that is, complete response (CR). The residual group was defined as residual disease on reexamination of MRI images after IC-CCRT, including partial response (PR), stable disease (SD), or progressive disease (PD).

MRI Acquisition

MRI images of all patients were obtained using two different MRI scanners (Siemens Magnetom Symphony 1.5T and Siemens SKyra 3.0T, Munich, Germany). Axial FS T2WI images were initially obtained, followed by axial FS CE-T1WI imaging after gadolinium-based contrast agent administration at 0.01 mmol/kg. The acquisition protocol for neck MRI was slightly different but mainly consisted of the following parameters: 1) axial FS T2WI, repetition time/echo time (TR/TE) 6,360 ms/95 ms, 90° flip angle, 256 × 168 matrix, slice thickness 4.68 mm, slice spacing 4.68 mm; and 2) axial FS CE-T1WI TR/TE 450 ms/8.8 ms, 90° flip angle, 256 × 168 matrix, slice thickness 4.68 mm, slice spacing 4.68 mm.

MRI Preprocessing and Image Segmentation

The nasopharyngeal MRI examinations before treatment, after IC, and after CCRT of each case were searched from the hospital PACS, and the DICOM format images of FS T2WI and FS CE-T1WI sequences were exported. DICOM format images were imported using ITK-SNAP (www.itksnap.org, version 3.8.0) software. We first selected axial FS CE-T1WI images, in order to improve the accuracy of lesion delineation, and axial FS T2WI images were imported in the "Add Another DICOM Series" option. Two attending radiologists with 7 and 8 years of experience in head and neck radiological diagnosis manually segmented the primary NPC tumor bodies layer by layer. The same approach was used when tumor bodies were delineated on the axial FS T2WI images. Finally, the original images and segmented images were stored according to the format requirements. The same method was applied for tumor segmentation in 96 patients after IC.

Radiomics Feature Extraction and Delta Radiomics Feature Calculation

The study used the software of Artificial Intelligence Kit (AK) V3.4.0.R issued by GE Company (Chicago, IL, USA). This software has been registered, approved, and applied to medical radiomics research (34–36). First, the unsegmented original data were imported into AK software, the original data were resampled, the resolution was adjusted to 1 mm × 1 mm × 1 mm, the layer thickness was 1 mm, the image gray level was uniformly adjusted to 0–255, and then the region of interest (ROI) image was imported. First Order, Shape, GLCM, GLDM, GLSZM, GLRLM, and NGTDM were used for feature selection, and LoG, Wavelet, and LBP were used for filter selection. Radiomics features were extracted from MRI images of 221 patients before treatment and 96 patients after IC.

For Delta radiology profile estimation, the change in each radiology profile was calculated by the following equation:

$$Delta\ Feature\ value = (Feature\ value2 - Feature\ value1)$$

Here, Feature value2 represents the post-IC MRI value, and Feature value1 represents the pretreatment MRI value.

Feature Selection, Model Establishment, and Statistical Analysis

All statistical analyses were performed using R statistical software (version 4.0.3). The intraclass correlation coefficient (ICC) was used to evaluate inter- and intra-observer agreement. Two attending radiologists randomly selected 40 patients and segmented the tumor once and twice again to calculate observer stability for each feature. Radiological features with ICC greater than 0.75 were defined as stable features. In order to ensure that the images from different devices were comparable, the data were normalized before feature selection, that is, Z-score transformation, and the calculation formula was as follows:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Here, μ is the mean, σ is the standard deviation, and n is the sample size.

The subjects were divided into the training group and the validation cohort in a ratio of 7:3. Then, maximum relevance minimum redundancy (mRMR) and least absolute shrinkage and selection operator (LASSO) were used to select features. We first used mRMR to retain 20 features that were maximally correlated with tags and least redundant with each other and then used LASSO to select the optimal subset of features for model construction. In order to avoid overfitting, we used 10 times of cross-validation to select the adjustment parameter λ. According to the screened characteristics and corresponding coefficients, a logistic regression model of FS T2WI combined with FS CE-T1WI was established, and the radiomics signature (Radscore) was obtained. The area under the receiver operating characteristic (ROC) curve (AUC), diagnostic accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the training set and the internal validation set were calculated. ROC analysis was used to evaluate the diagnostic efficacy of the model. Calibration analysis was used to evaluate the goodness-of-fit of the model, and the decision curve was used to evaluate the clinical value of the model. The radiomics flowchart is shown in **Figure 2**.

RESULTS

Analysis of Radiomics Results

The ICC results were all greater than 0.75, and finally, the segmentation results of the senior doctor were selected. First, we selected 20 features that were most relevant to the tags and least redundant with each other among more than 2,000 features of the two sequences of FS T2WI and FS CE-T1WI by the mRMR method (**Figures 3A, B**; **4A, B**). Then LASSO was performed to select the optimal feature subset for constructing the model. After 10 cross-validations, six and twelve optimal subsets were

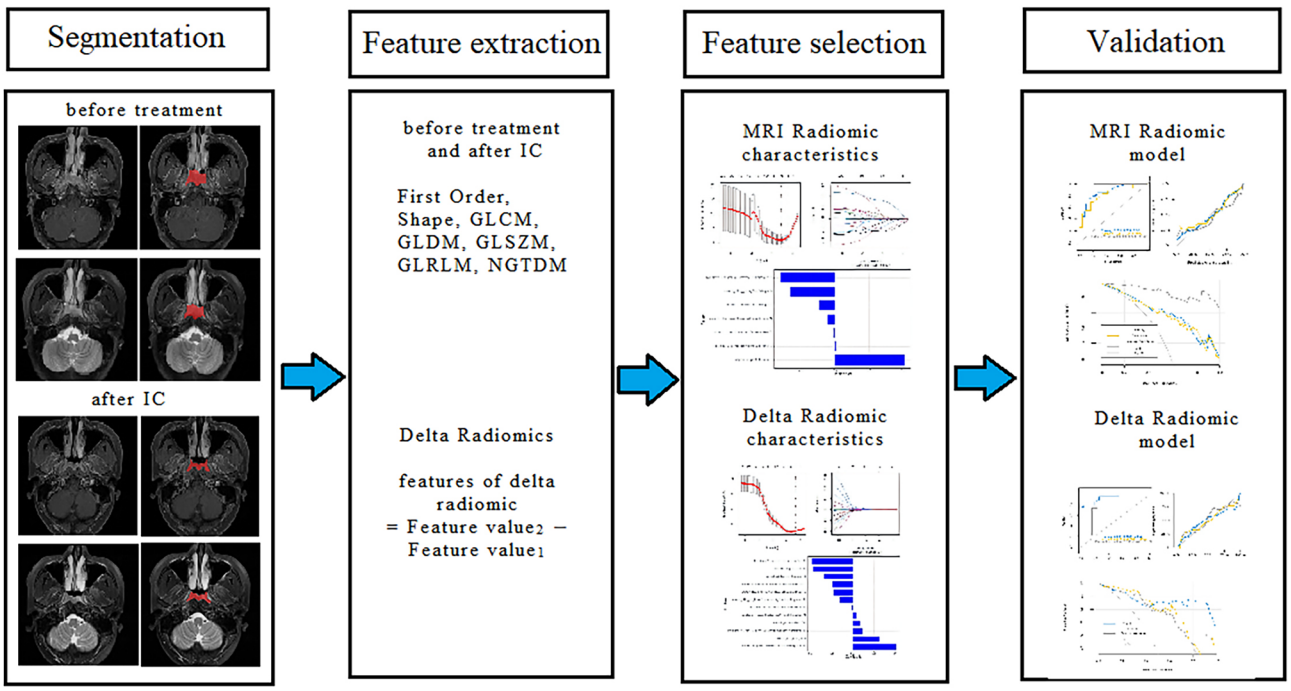


FIGURE 2 | Flow diagram of radiomics.

finally retained (**Figures 3C**, **4C**). The Radscore was then calculated for each patient, and the results were detailed in the Appendix.

Power Analysis of the Model to Predict the Efficacy of Induction Chemotherapy Combined With Concurrent Chemoradiotherapy

The diagnostic effects of the pretreatment MRI radiomics model and Delta radiomics model in predicting the efficacy of IC-CCRT for NPC are summarized in **Table 1**. The AUC values (**Figure 5**), calibration curves (**Figure 6**), and decision curves (**Figure 7**) of both models showed good performance in the training set, internal validation set, and external validation set.

DISCUSSION

We selected the seven most relevant radiological features (six from FS T2WI images and one from FS CE-T1WI images) from the pretreatment MRI radiomics model of NPC and twelve most relevant radiological features (seven from FS CE-T1WI images and five from FS T2WI images) from the Delta radiomics model. The clinical predictive values of the two models after chemoradiotherapy for NPC were also analyzed. The results showed that the two constructed models had a high diagnostic performance for NPC, and there were some differences between the two models.

In this study, we selected two MRI sequences, FS T2WI and FS CE-T1WI, for the following three reasons. First, MRI was the most effective and commonly used method for the diagnosis and staging of NPC (37). MRI-based radiomics could better define tumor biology and had immense potential to support oncological decisions (38, 39). Second, these two sequences were commonly obtained during conventional MRI scans and might be more universal. Furthermore, FS sequence and enhanced examination could enhance the image contrast between the tumor and surrounding tissues, and lesion delineation was more accurate. The radiomics characteristics extracted from the combined images were reported to have better predictive performance compared with a single sequence.

In recent years, radiomics has been increasingly used in the diagnosis, tumor stratification, and prognosis prediction of NPC. For example, our previous study showed that the NPC radiomics model based on ^{18}F-FDG PET/MRI images had immense value in the staging evaluation of primary NPC (36). The NPC could be divided into subtypes with different survival patterns based on the radiological characteristics of MRI (40). The early metastasis of NPC could be predicted based on Epstein–Barr virus (EBV), clinical data, and MRI nomogram. Radiomics could predict the PFS of NPC (41), and some radiomics characteristics could identify the patients who needed adaptive radiotherapy (42) and those who would benefit the most from adjuvant therapy

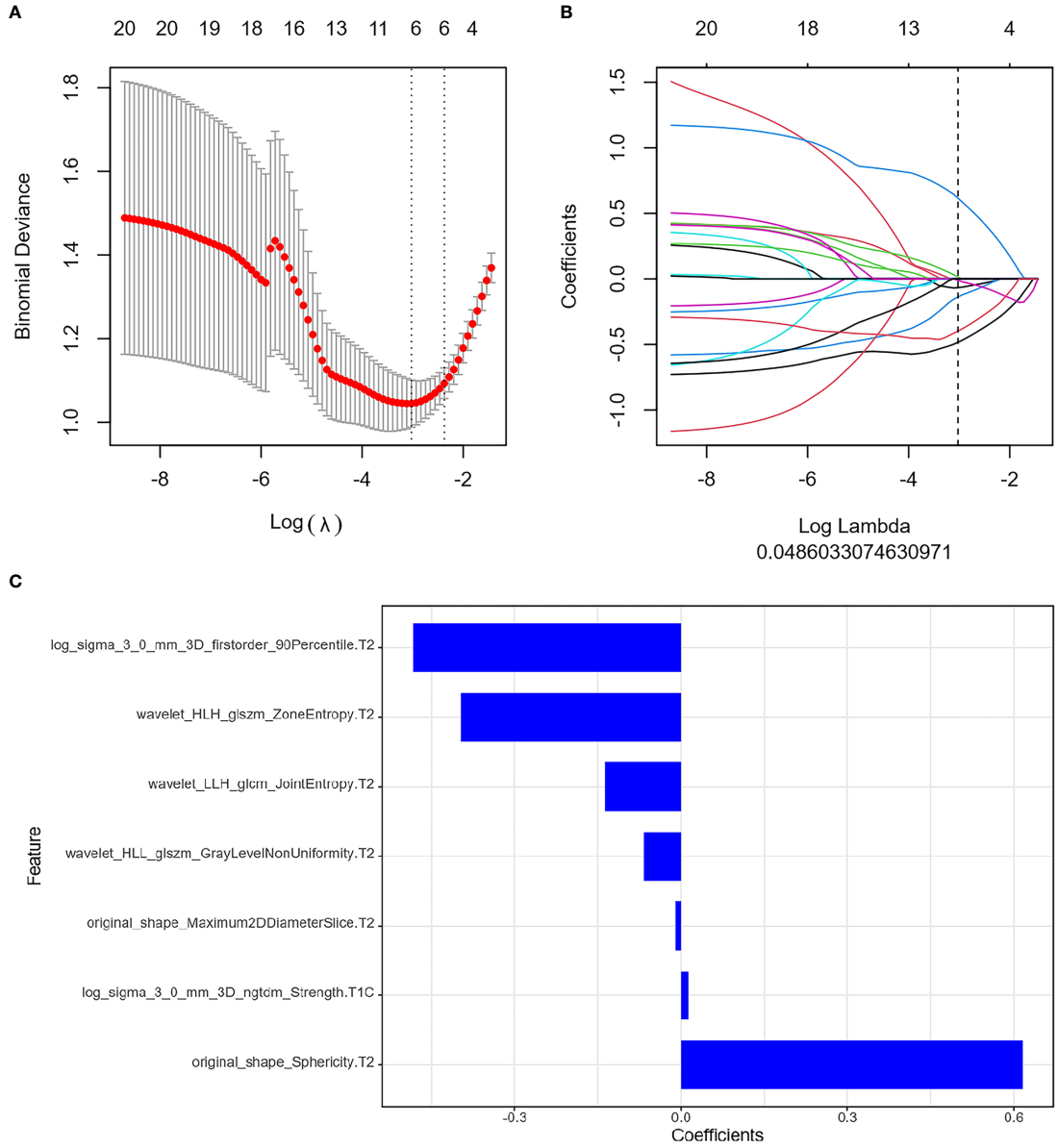


FIGURE 3 | Feature extraction before treatment. **(A)** Uses ten-fold cross-validation and minimum criteria to select the adjustment parameter (λ). **(B)** LASSO coefficients of 20 texture features, drawing vertical lines at the selected optimal value in λ sequence. **(C)** Screens seven features. LASSO, least absolute shrinkage and selection operator.

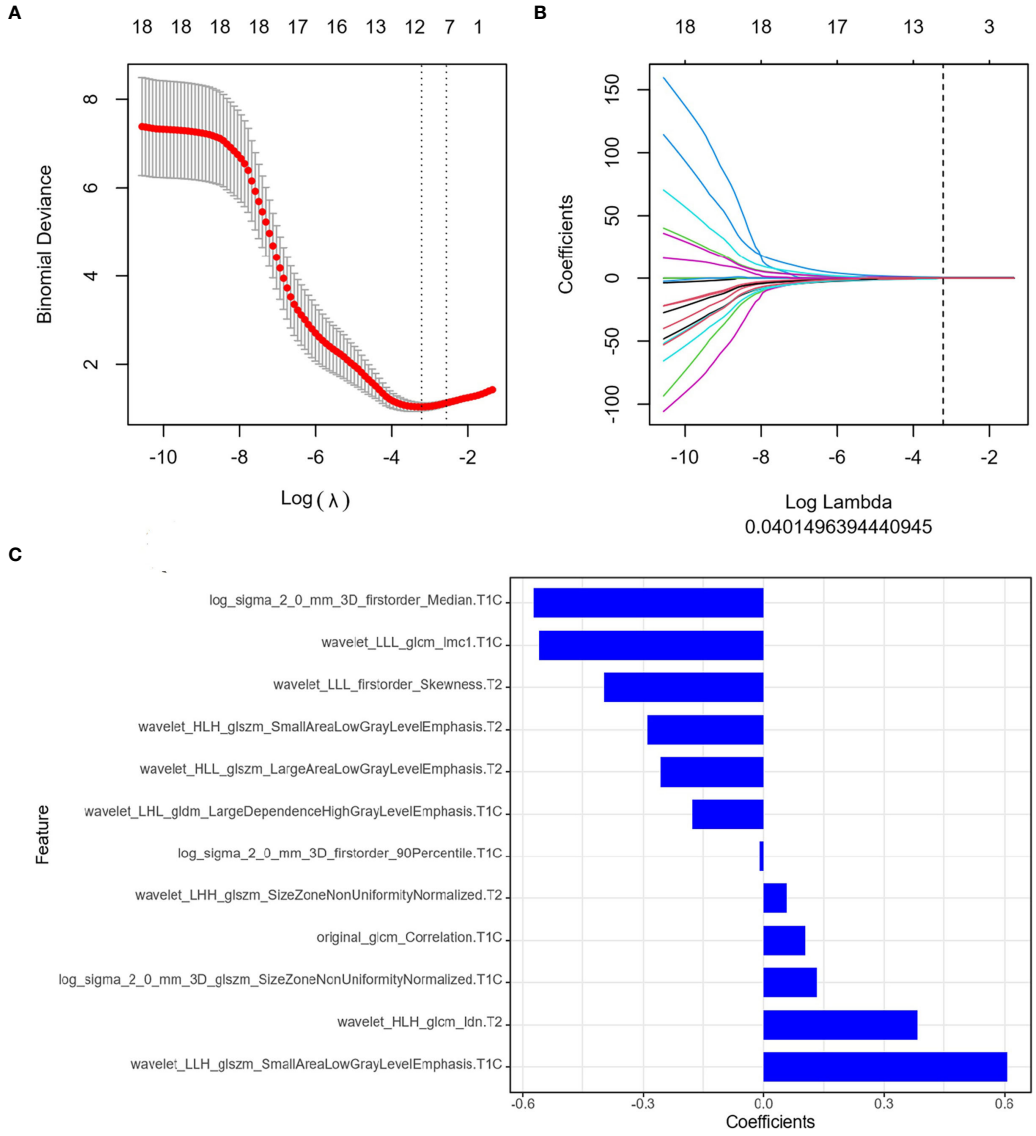


FIGURE 4 | Feature extraction of Delta radiomics. **(A)** Uses ten-fold cross-validation and minimum criteria to select the adjustment parameter (λ). **(B)** LASSO coefficients of 20 texture features, drawing vertical lines at the selected optimal value in λ sequence. **(C)** Screens twelve features. LASSO, least absolute shrinkage and selection operator.

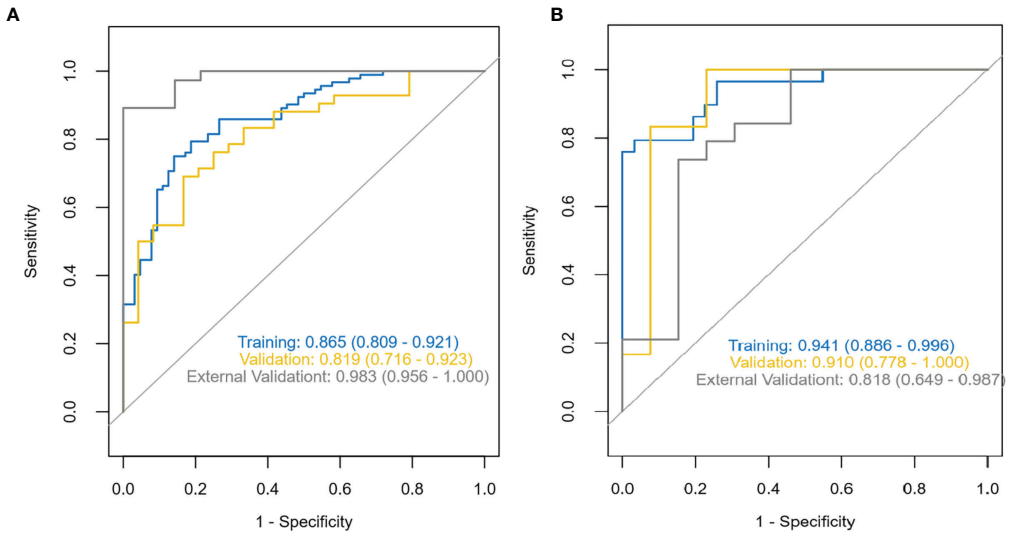


FIGURE 5 | ROC characteristic curves and corresponding AUC values for the pretreatment MRI radiomics model **(A)** and the Delta radiomics model **(B)**. ROC, receiver operating characteristic; AUC, area under the receiver operating characteristic curve.

TABLE 1 | The diagnostic effects of the pretreatment MRI radiomics model and Delta radiomics model in predicting the efficacy of IC-CCRT for NPC.

		ROC	Accuracy	Sensitivity	Specificity	NPV	PPV
MRI radiomics	Training set	0.865	0.827	0.837	0.812	0.776	0.865
	Validation set	0.819	0.788	0.833	0.708	0.708	0.833
	External validation	0.983	0.784	0.703	1.000	0.560	1.000
Delta radiomics	Training set	0.941	0.883	0.793	0.968	0.833	0.958
	Validation set	0.910	0.880	1.000	0.769	1.000	0.800
	External validation	0.818	0.781	0.737	0.846	0.688	0.875

ROC, receiver operating characteristic; NPV, negative predictive value; PPV, positive predictive value.

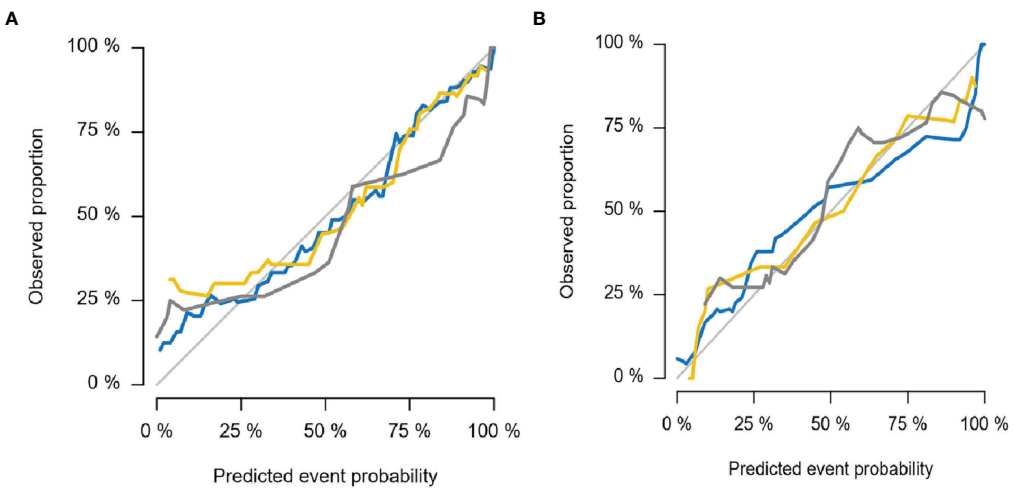


FIGURE 6 | Calibration curves for MRI radiomics model **(A)** and Delta radiomics model **(B)**. The thin gray line represents the ideal reference line, the blue line represents the training set, the yellow line represents the internal validation set, and the thick gray line represents the external validation set. In this reference line, regardless of the training set, internal validation set, or external validation set, the predicted probability matches the observed proportion, indicating that both models have good performance in judging the tumor retraction of NPC primary tumors after IC-CCRT treatment. NPC, nasopharyngeal carcinoma; IC-CCRT, induction chemotherapy combined with concurrent chemoradiotherapy.

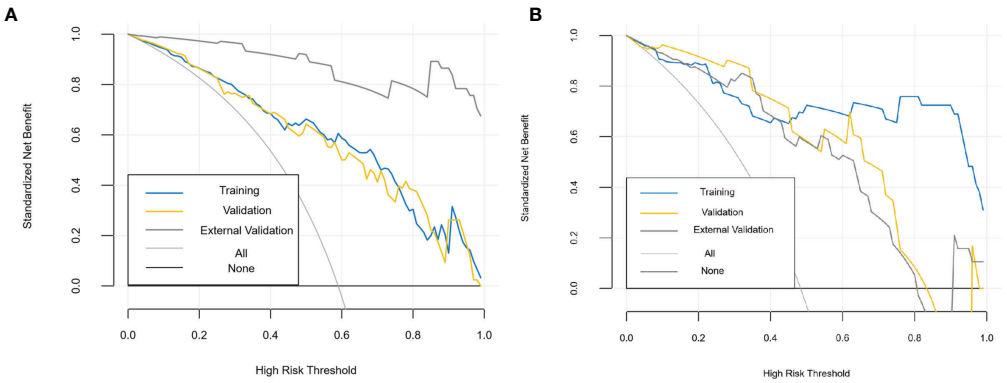


FIGURE 7 | Decision curves of MRI radiomics model **(A)** and Delta radiomics model **(B)**. The x-axis represents the probability of response to IC-CCRT, ranging from 0% to 100%; the y-axis measures the net benefit; the blue, yellow, and coarse gray lines represent nomograms of different groups. IC-CCRT, induction chemotherapy combined with concurrent chemoradiotherapy.

or IC (43, 44). A combined model constructed based on EBV, clinical data, and radiological characteristics can predict NPC progression (45) and distant metastasis. However, few of these studies have external validation data, and there have been no reports on Delta radiomics of tumor retraction after IC-CCRT for NPC tumors. Our constructed MRI radiomics model at a single time point before treatment and Delta radiomics before and after IC could predict the efficacy of IC-CCRT for NPC for early prediction during treatment.

The optimal features of the pretreatment MRI radiomics model and Delta radiomics model focused on first-order characteristics, log and texture features after wavelet filtering, and morphological features of the original data. First-order statistics describe the distribution of voxel intensity in the image area defined by the mask through commonly used and basic metrics. About 90th percentile of first-order features and the changes in median and skewness are highly correlated with tumor retraction. The median value represents the gray intensity in the ROI, the asymmetry of the distribution of bias measurement, and the correlation value of the mean value. Sphericity characteristics of morphology in the MRI radiomics model, the roundness of the tumor area relative to the sphere morphology, were significantly correlated with tumor retraction. Our study showed that GLSM, GLCM, and NGTDM were also highly correlated with tumor retraction. GLSZM-ZE represented the uncertainty/randomness of the size of the measurement area and the gray level; the greater the value, the higher the heterogeneity. GLSZM-GLZE represented the distribution of the lower gray level area, and the higher value, the lower the gray value and the proportion of the size area in the image. GLCM describes the texture by studying the spatial correlation of the gray level. Energy represents the uncertainty and randomness of the image. The value of GLCM Joint Entropy represents the complexity of the co-occurrence matrix. In addition to first-order features in the Delta radiomics model, and SALGLE and LALGLE features in GLCM-IMC1, GLSZM is highly correlated with tumor retraction, which can quantify the complexity of texture. SALGLE measures the proportion of smaller sizes with lower gray values in regional joint distribution images. LALGLE measures the proportion of larger sizes with lower gray values in regional joint distribution images. These characteristics were most pronounced after chemoradiation for NPC in this study. Therefore, both the pretreatment MRI radiomics model and Delta radiomics model could predict tumor retraction after IC-CCRT.

However, this study had several limitations. First, this study was a retrospective analysis, and there was inevitable selectivity bias. Second, after IC, the cycle of CCRT was not fixed. We plan to collect more relevant data at a later stage and do more in-depth studies in the future. Third, because the data after treatment of IC were not uniform and the performance of the two models could not be compared, there was a general problem of poor refolding in radiomics. We are going to collect more cases to train the stability of the Delta radiomics model and increase its general applicability.

CONCLUSION

Pretreatment MRI radiomics at a single time point and Delta radiomics before and after IC could predict tumor retraction after IC-CCRT for the treatment of NPC in non-endemic areas. This study provided a quantitative basis for early intervention and timely optimization of treatment options.

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 Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Affiliated Hangzhou First People's Hospital and Sir Run Run Shaw Hospital affifiliated to Medical College Zhejiang University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZD designed and planned the study. XG, MW, HC, YX, HH, and FJ collected the clinical and image data. YX and HJ performed ROI segmentation. YX, SD, and LW analyzed the data. YX participated in drafting and writing the manuscript. All authors contributed to this study and approved the submitted version.

FUNDING

This study was supported by grants from the National Natural Science Foundation of China (No. 81871337), Natural Science Foundation of Zhejiang Province (Y22H185692, LGF18H160018, and LY16H180007), Zhejiang Provincial Health Science and Technology Project (No. 2018KY304, No. 2020RC092, and No. 2021RC108), and Hangzhou Medical and Health Science and Technology Project (No. A20200507).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.824509/full#supplementary-material

REFERENCES

1. De Martel C, Georges D, Bray F, Ferlay J, Clifford GM. Global Burden of Cancer Attributable to Infections in 2018: A Worldwide Incidence Analysis. *Lancet Glob Health* (2020) 8(2):e180–90. doi: 10.1016/S2214-109X(19)30488-7
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(3):209–49. doi: 10.3322/caac.21660
3. *International Agency for Research on Cancer*. World Health Organization (2020). Available at: https://gco.iarc.fr/today/home.
4. Kai D, Renjie C. Progress in the Treatment of Nasopharyngeal Carcinoma. *J Nanjing Med Univ (Nat Sci Ed)* (2021) 41(06):921–926 + 936.
5. Pfister DG, Spencer S, Adelstein D, Adkins D, Anzai Y, Brizel DM, et al. Head and Neck Cancers, Version 2.2020, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw* (2020) 18(7):873–98. doi: 10.6004/jnccn.2020.0031
6. Yang Q, Cao SM, Guo L, Hua YJ, Huang PY, Zhang XL, et al. Induction Chemotherapy Followed by Concurrent Chemoradiotherapy Versus Concurrent Chemoradiotherapy Alone in Locoregionally Advanced Nasopharyngeal Carcinoma: Long-Term Results of a Phase III Multicentre Randomised Controlled Trial. *Eur J Canc* (2019) 119:87–96. doi: 10.1016/j.ejca.2019.07.007
7. Zhao L, Gong J, Xi Y, Xu M, Li C, Kang X, et al. MRI-Based Radiomics Nomogram may Predict the Response to Induction Chemotherapy and Survival in Locally Advanced Nasopharyngeal Carcinoma. *Eur Radiol* (2020) 30(1):537–46. doi: 10.1007/s00330-019-06211-x
8. Li WF, Chen NY, Zhang N, Hu GQ, Xie FY, Sun Y, et al. Concurrent Chemoradiotherapy With/Without Induction Chemotherapy in Locoregionally Advanced Nasopharyngeal Carcinoma: Long-Term Results of Phase 3 Randomized Controlled Trial. *Int J Canc* (2019) 145(1):295–305. doi: 10.1002/ijc.32099
9. Sun X, Su S, Chen C, Han F, Zhao C, Xiao W, et al. Long-Term Outcomes of Intensity-Modulated Radiotherapy for 868 Patients With Nasopharyngeal Carcinoma: An Analysis of Survival and Treatment Toxicities. *Radiother Oncol* (2014) 110(3):398–403. doi: 10.1016/j.radonc.2013.10.020
10. Zhang L, Dong D, Li H, Tian J, Ouyang F, Mo X, et al. Development and Validation of a Magnetic Resonance Imaging-Based Model for the Prediction of Distant Metastasis Before Initial Treatment of Nasopharyngeal Carcinoma: A Retrospective Cohort Study. *EBioMedicine* (2019) 40:327–35. doi: 10.1016/j.ebiom.2019.01.013
11. 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
12. He Y, Zhou Q, Shen L, Zhao Y, Lei M, Wei R, et al. A Retrospective Study of the Prognostic Value of MRI-Derived Residual Tumors at the End of Intensity-Modulated Radiotherapy in 358 Patients With Locally-Advanced Nasopharyngeal Carcinoma. *Radiat Oncol* (2015) 10:89. doi: 10.1186/s13014-015-0401-0
13. Lv JW, Zhou GQ, Li JX, Tang LL, Mao YP, Lin AH, et al. Magnetic Resonance Imaging-Detected Tumor Residue After Intensity-Modulated Radiation Therapy and its Association With Post-Radiation Plasma Epstein-Barr Virus Deoxyribonucleic Acid in Nasopharyngeal Carcinoma. *J Canc* (2017) 8(5):861–9. doi: 10.7150/jca.17957
14. Stoker SD, van Diessen JN, de Boer JP, Karakullukcu B, Leemans CR, Tan IB. Current Treatment Options for Local Residual Nasopharyngeal Carcinoma. *Curr Treat Opt Oncol* (2013) 14(4):475–91. doi: 10.1007/s11864-013-0261-5
15. Yang Y, Yang W, Li FL, Liu MD, LiY. The Effect of Tumor Regression During Radiotherapy on Gross Tumor Volume in Patients With Nasopharyngeal Carcinoma. . *J Chin J Otorhinolaryngol Skull Base Surg* (2021) 27(03):289–93. doi: 10.11798/j.issn.1007-1520.202103204
16. O'Connor JPB. Cancer Heterogeneity and Imaging. *Semin Cell Dev Biol* (2017) 64:48–57. doi: 10.1016/j.semcdb.2016.10.001
17. Chang Y, Lafata K, Sun W, Wang C, Chang Z, Kirkpatrick JP, et al. An Investigation of Machine Learning Methods in Delta-Radiomics Feature Analysis. *PloS One* (2019) 13:14(12):e0226348. doi: 10.1371/journal.pone.0226348
18. Du D, Feng H, Lv W, Ashrafinia S, Yuan Q, Wang Q, et al. Machine Learning Methods for Optimal Radiomics-Based Differentiation Between Recurrence and Inflammation: Application to Nasopharyngeal Carcinoma Post-Therapy PET/CT Images. *Mol Imaging Biol* (2020) 22(3):730–8. doi: 10.1007/s11307-019-01411-9
19. Zhong X, Li L, Jiang H, Yin J, Lu B, Han W, et al. Cervical Spine Osteoradionecrosis or Bone Metastasis After Radiotherapy for Nasopharyngeal Carcinoma? The MRI-Based Radiomics for Characterization. *BMC Med Imag* (2020) 20(1):104. doi: 10.1186/s12880-020-00502-2
20. Peng H, Dong D, Fang MJ, Li L, Tang LL, Chen L, et al. Prognostic Value of Deep Learning PET/CT-Based Radiomics: Potential Role for Future Individual Induction Chemotherapy in Advanced Nasopharyngeal Carcinoma. *Clin Cancer Res* (2019) 25(14):4271–9. doi: 10.1158/1078-0432.CCR-18-3065
21. Kim MJ, Choi Y, Sung YE, Lee YS, Kim YS, Ahn KJ, et al. Early Risk-Assessment of Patients With Nasopharyngeal Carcinoma: The Added Prognostic Value of MR-Based Radiomics. *Transl Oncol* (2021) 14(10):101180. doi: 10.1016/j.tranon.2021.101180

22. Xu H, Lv W, Feng H, Du D, Yuan Q, Wang Q, et al. Subregional Radiomics Analysis of PET/CT Imaging With Intratumor Partitioning: Application to Prognosis for Nasopharyngeal Carcinoma. *Mol Imaging Biol* (2020) 22 (5):1414–26. doi: 10.1007/s11307-019-01439-x
23. Zhuo EH, Zhang WJ, Li HJ, Zhang GY, Jing BZ, Zhou J, et al. Radiomics on Multi-Modalities MR Sequences can Subtype Patients With non-Metastatic Nasopharyngeal Carcinoma (NPC) Into Distinct Survival Subgroups. *Eur Radiol* (2019) 29(10):5590–9. doi: 10.1007/s00330-019-06075-1
24. Qiang M, Li C, Sun Y, Sun Y, Ke L, Xie C, et al. A Prognostic Predictive System Based on Deep Learning for Locoregionally Advanced Nasopharyngeal Carcinoma. *J Natl Cancer Inst* (2021) 113(5):606–15. doi: 10.1093/jnci/djaa149
25. Zhang B, Lian Z, Zhong L, Zhang X, Dong Y, Chen Q, et al. Machine-Learning Based MRI Radiomics Models for Early Detection of Radiation-Induced Brain Injury in Nasopharyngeal Carcinoma. *BMC Canc* (2020) 20(1):502. doi: 10.1186/s12885-020-06957-4
26. Holscher HC, Bloem JL, van der Woude HJ, Hermans J, Nooy MA, Taminiau AH, et al. Can MRI Predict the Histopathological Response in Patients With Osteosarcoma After the First Cycle of Chemotherapy? *Clin Radiol* (1995) 50 (6):384–90. doi: 10.1016/s0009-9260(05)83135-6
27. 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 Imag* (2020) 20(1):7. doi: 10.1186/s40644-019-0283-8
28. Ma Y, Ma W, Xu X, Cao F. How Does the Delta-Radiomics Better Differentiate Pre-Invasive GGNs From Invasive GGNs? *Front Oncol* (2020) 10:1017. doi: 10.3389/fonc.2020.01017
29. Shayesteh S, Nazari M, Salahshour A, Sandoughdaran S, Hajianfar G, Khateri M, et al. Treatment Response Prediction Using MRI-Based Pre-, Post-, and Delta-Radiomic Features and Machine Learning Algorithms in Colorectal Cancer. *Med Phys* (2021) 48(7):3691–701. doi: 10.1002/mp.14896
30. Vallières M, Zwanenburg A, Badic B, Cheze Le Rest C, Visvikis D, Hatt M. Responsible Radiomics Research for Faster Clinical Translation. *J Nucl Med* (2018) 59(2):189–93. doi: 10.2967/jnumed.117.200501
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(12):749–62. doi: 10.1038/nrclinonc.2017.141
32. Liu Y, Shi H, Huang S, Chen X, Zhou H, Chang H, et al. Early Prediction of Acute Xerostomia During Radiation Therapy for Nasopharyngeal Cancer Based on Delta Radiomics From CT Images. *Quant Imaging Med Surg* (2019) 9(7):1288–302. doi: 10.21037/qims.2019.07.08
33. Barua S, Elhalawani H, Volpe S, Al Feghali KA, Yang P, Ng SP, et al. Computed Tomography Radiomics Kinetics as Early Imaging Correlates of Osteoradionecrosis in Oropharyngeal Cancer Patients. *Front Artif Intell* (2021) 4:618469. doi: 10.3389/frai.2021.618469
34. Shao Y, Chen Z, Ming S, Ye Q, Shu Z, Gong C, et al. Predicting the Development of Normal-Appearing White Matter With Radiomics in the Aging Brain: A Longitudinal Clinical Study. *Front Aging Neurosci* (2018) 10:393. doi: 10.3389/fnagi.2018.00393
35. Feng Q, Song Q, Wang M, Pang P, Liao Z, Jiang H, et al. Hippocampus Radiomic Biomarkers for the Diagnosis of Amnestic Mild Cognitive Impairment: A Machine Learning Method. *Front Aging Neurosci* (2019) 11:323. doi: 10.3389/fnagi.2019.00323
36. Feng Q, Liang J, Wang L, Niu J, Ge X, Pang P, et al. Radiomics Analysis and Correlation With Metabolic Parameters in Nasopharyngeal Carcinoma Based on PET/MR Imaging. *Front Oncol* (2020) 10:1619. doi: 10.3389/fonc.2020.01619
37. Dai YL, King AD. State of the Art MRI in Head and Neck Cancer. *Clin Radiol* (2018) 73(1):45–59. doi: 10.1016/j.crad.2017.05.020
38. Choi ER, Lee HY, Jeong JY, Choi YL, Kim J, Bae J, et al. Quantitative Image Variables Reflect the Intratumoral Pathologic Heterogeneity of Lung Adenocarcinoma. *Oncotarget* (2016) 7(41):67302–13. doi: 10.18632/oncotarget.11693
39. 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
40. 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(15):4259–69. doi: 10.1158/1078-0432.CCR-16-2910
41. Shen H, Wang Y, Liu D, Lv R, Huang Y, Peng C, et al. Predicting Progression-Free Survival Using MRI-Based Radiomics for Patients With Nonmetastatic Nasopharyngeal Carcinoma. *Front Oncol* (2020) 10:618. doi: 10.3389/fonc.2020.00618
42. Yu TT, Lam SK, To LH, Tse KY, Cheng NY, Fan YN, et al. Pretreatment Prediction of Adaptive Radiation Therapy Eligibility Using MRI-Based Radiomics for Advanced Nasopharyngeal Carcinoma Patients. *Front Oncol* (2019) 9:1050. doi: 10.3389/fonc.2019.01050
43. Zheng DW, Xu SG, Lai GJ, Chen JY, Ren W, Chen YB. Preliminary Study on Predicting the Efficacy of Neoadjuvant Chemotherapy for Nasopharyngeal Carcinoma Based on Pretreatment Magnetic Resonance Imaging Nomogram. *MRI* (2021) 12(04):23–9.
44. Zhong L, Dong D, Fang X, Zhang F, Zhang N, Zhang L, et al. A Deep Learning-Based Radiomic Nomogram for Prognosis and Treatment Decision in Advanced Nasopharyngeal Carcinoma: A Multicentre Study. *EBioMedicine* (2021) 70:103522. doi: 10.1016/j.ebiom.2021.103522
45. Li Q, Wang T, Huang Y, Li Q, Liu P, Grimm R, et al. Whole-Tumor Histogram and Texture Imaging Features on Magnetic Resonance Imaging Combined With Epstein-Barr Virus Status to Predict Disease Progression in Patients With Nasopharyngeal Carcinoma. *Front Oncol* (2021) 11:610804. doi: 10.3389/fonc.2021.610804

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.

frontiers | Frontiers in Physics

ORIGINAL RESEARCH
published: 05 May 2022
doi: 10.3389/fphy.2022.783765

A Unified Approach to Analysis of MRI Radiomics of Glioma Using Minimum Spanning Trees

*Olivier B. Simon[1], Rajan Jain[2,3], Yoon-Seong Choi[4,5], Carsten Görg[1], Krithika Suresh[1], Cameron Severn[1,6] and Debashis Ghosh[1]**

[1]Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, United States, [2]Department of Radiology, NYU Grossman School of Medicine, New York, NY, United States, [3]Department of Neurosurgery, NYU Grossman School of Medicine, New York, NY, United States, [4]Radiological Sciences Academic Clinical Programme, Duke-NUS Medical School, Singapore, Singapore, [5]Department of Radiology, Yonsei University College of Medicine, Seoul, South Korea, [6]Department of Pediatrics, Section of Endocrinology, School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States

Radiomics has shown great promise in detecting important genetic markers involved in cancers such as gliomas, as specific mutations produce subtle but characteristic changes in tumor texture and morphology. In particular, mutations in IDH (isocitrate dehydrogenase) are well-known to be important prognostic markers in glioma patients. Most classification approaches using radiomics, however, involve complex hand-crafted feature sets or "black-box" methods such as deep neural networks, and therefore lack interpretability. Here, we explore the application of simple graph-theoretical methods based on the minimum-spanning tree (MST) to radiomics data, in order to detect IDH mutations in gliomas. This is done using a hypothesis testing approach. The methods are applied to an fMRI dataset on $n = 413$ patients. We quantify the significance of the group-wise difference between mutant and wild-type using the MST edge-count testing methodology of Friedman and Rafsky. We apply network theory-based centrality measures on MSTs to identify the most representative patients. We also propose a simple and rapid dimensionality-reduction method based on k-MSTs. Combined with the centrality measures, the latter method produces readily interpretable 2D maps that reveal distinct IDH, non-IDH, and IDH-like groupings.

Keywords: medical imaging, biostatistics, genotype-phenotype correlation, tree-based methodology, data visualization

OPEN ACCESS

Edited by:
Daniel Rodriguez Gutierrez, Nottingham University Hospitals NHS Trust, United Kingdom

Reviewed by:
Zhiwei Ji, Nanjing Agricultural University, China
Roberto Gatta, University of Brescia, Italy
Enrico Capobianco, University of Miami, United States

****Correspondence:***
Debashis Ghosh
debashis.ghosh@cuanschutz.edu

Specialty section:
This article was submitted to Medical Physics and Imaging, a section of the journal Frontiers in Physics

Received: *26 September 2021*
Accepted: *12 April 2022*
Published: *05 May 2022*

Citation:
Simon OB, Jain R, Choi Y-S, Görg C, Suresh K, Severn C and Ghosh D (2022) A Unified Approach to Analysis of MRI Radiomics of Glioma Using Minimum Spanning Trees. Front. Phys. 10:783765. doi: 10.3389/fphy.2022.783765

INTRODUCTION

The advent of widespread medical imaging, large imaging datasets, and large-scale inexpensive computing power has ushered in an era of unprecedented resources for medical image analysis [1]. Cancers can now be automatically detected and staged from histopathology images, or from clinical imaging datasets such as MRI, CT or PET data. In particular, considerable success has been achieved using complex computer-derived image-analysis features derived from such data as input for advanced statistical and machinelearning methods. This approach, known as "radiomics", offers the potential to take into account multiple features of the image not detected by human observers and hence also avoiding the issue of inter-observer variability [2].

Genotyping of gliomas is difficult and invasive, as it requires biopsy of brain tissue. While some genetic correlates of cancer prognosis, such as MGMT promoter methylation, have not shown strong correlation with radiomics features [3], other mutation types are associated with marked differences in radiomic profiles—although considerable variability between studies exists. In particular, isocitrate dehydrogenase (IDH) mutations are found in 5–13% of glioblastomas and are strongly correlated with radiomics features [3].

Current automated methods for visual or radiomic genotyping of gliomas increasingly depend on deep neural network methods and pipelines, often using off-the-shelf architectures such as ResNet for detection and then classification [4]. Still other studies have made use of random forest methods for genotyping, in combination with CNN-based methods for tumor segmentation [3]. Alongside neural networks, more traditional "hand-crafted" features, involving human-defined combinations of pixel-level image analysis methods such as gray-level co-occurrence matrices (GLCMs), represent a second still-vibrant branch of radiomics analysis [5]. Handcrafted features often have the benefit of imparting greater interpretability to radiomics analyses; on the other hand, since neural network models are considered by some to be more free from human bias, current state-of-the-art radiomics methods frequently combine both [6].The area under the curve (AUC) is a typical metric for evaluation for these approaches, with values around 0.85–0.95 representing the state-of-the-art as of this writing [2]. Other measures such as F1 score, sensitivity and specificity are also common. However, while useful for gauging performance, these measures do little to provide an intuitive understanding of the structure of the underlying data, or the reasons for the classifier outputs—a problem which is particularly serious with neural networks, which with their many millions of automatically learned parameters are often considered to be "black-boxes" [7]. While deep learning is the current state-of-the art classification technique, we do note that other modeling procedures could be used, such as logistic regression, support vector machines or L1-penalized regression approaches, among many others.

Minimum spanning trees (MSTs) are graph-theoretic structures in which a set of data-points or "nodes" are connected into a single component using the minimum possible total connection distance [8]. Notably, MSTs, while easy to compute, are capable of representing key statistical properties of highly complex datasets in a vastly simplified format that is also readily amenable to lower-dimensional (even 2D) visualization. This renders them applicable to understanding a range of systems, such as gene expression, transportation networks and brain connectivity [9, 10]. Furthermore, node centrality measures--which aim at measuring the "importance" of a given node to the overall network structure--are readily applicable to the MST [11]. Therefore, MST and other graph-based approaches may offer an appealing and complementary alternative to the 'blpredictions given by neural-networks.

In addition to being easy to calculate, the MST of a high-dimensional dataset also comes with an attendant hypothesis testing procedure that allows one to assess the significance of the difference between classes. This is the Friedman-Rafsky multivariate runs test (here abbreviated "MVR") [12–14]. Briefly, MVR involves constructing an MST over the pooled data from two different classes, removing the edges that connect different classes, and counting the number of connected components that result. Smaller numbers of connected components indicate greater significance between the classes; this significance, furthermore, can be calculated using a standard normal approximation. Note that our goal here of inference is substantially different from much of the radiomics literature described above, which is focused on classification performance.

The k-MST is a simple extension of the MST, found by repeating the MST algorithm k times, each time excluding any connections chosen in prior iterations [14]. This allows a richer level of connectivity information which in turn can improve statistical test results such as edge-counting [12, 14]. At the same time, like an MST, a k-MST is a uniquely defined mathematical structure that can be calculated from any given point-set without requiring any user-tuned parameters; thus, use of the k-MST may greatly ameliorate one of the common concerns regarding "handcrafted" radiomics features, namely that of lower reproducibility stemming from bias in the feature design [6].

In the present work, we use the k-MST as a representation of the underlying structure of multivariate radiomics data, randomly embed it in a 2D region, and apply a simple 2D force-directed layouts methodology whereby nodes that are directly connected in the k-MST experience an attractive force. To avoid the expensive process of calculating repulsive forces between non-connected nodes, an isotropic expansion or "reflation" is carried out after each iteration, to counteract the tendency of a wholly-attractive configuration of forces to collapse to a point.

Because the k-MST contains only a small fraction of the possible pairwise connections between nodes, and because there are no explicit repulsive forces to calculate, our method allows rapid creation of 2D representations of arbitrarily high-dimensional radiomics datasets. Importantly, we find this method consistently converges to configurations which effectively segregate the IDH and non-IDH patients—especially when combined with the results of node centrality measures. This suggests possible wide applications of MST-based methods in creating "explainable" maps of radiomic data with respect to tumor genotype.

METHODS

Our analytic workflows are described in **Figures 1**,**2**. Our dataset derives from MRI scans conducted on 413 glioma patients, genotyped as either IDH-mutated (n = 144) or IDH wild type (n = 269). The data come from a previously published study [4]. T2-weighted and fluid-attenuated inversion recovery (FLAIR) MR images of diffuse gliomas (WHO grades II, III and IV) were obtained in DICOM format from After conversion to NIfTI format, T2-weighted

images were re-sampled to 1 mm isovoxel resolution using the 'trilinear' option from the FLIRT function, while FLAIR images were registered to the T2 images after skull stripping, all using the FMRIB software library (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL). Next, image signal intensity was normalized using the WhiteStripe R package. Tumor areas (defined by hyper-intensity in T2 images and edema on FLAIR images) were segmented with semi-automatic methods such as region growing, signal intensity thresholding, and edge detection, with an open-source software (Medical Image Processing, Analysis and Visualization, https://mipav.cit.nih.gov/). Segmentations were manually corrected by a neuroradiologist as deemed necessary.

Once MRI post-processing was completed, 467 radiomics features were calculated per patient using the PyRadiomics suite [15]. A full list of features used is included under (**Supplementary Table S4**). All data was centered to zero and normalized by dividing each column by its standard deviation. To account for the possibility of redundancy or overlap among the radiomics features, our MATLAB pipeline provides the option to perform PCA, retaining only those components which together comprise >98% of the total variance. This step reduces the number of components from the original 467 to 48. Example results of our pipeline using this PCA step are provided in Supplementary Figures; however, as this step did not dramatically change the character of the results, it was not used in the main study.

Next, using the features as dimensions and each patient as a node, we constructed MSTs over the pooled patient data from both groups and carried out the multivariate runs (MVR) test outlined by Friedman and colleagues [12](13) [14]. The Euclidean distance based on the standardized radiomics feature vectors was used to calculate distances between all pairs of subjects. This yields a graph with edge weights based on the distance which is used to construct an MST. For the MVR test, edges connecting dissimilar node-types (i.e., nodes connected from two different groups) are removed, yielding a number of disjoint trees, R. Given two MSTs with N_a and N_b nodes, (and $N = N_a + N_b$), Friedman and Rafsky demonstrate that R is normally distributed, with mean equal to

$$E[R] = 2N_aN_b/(N+1)$$

and variance (conditioned on C, the number of pairs of edges that share a common node in the given MST), equal to

$$var[R|C] = \frac{2N_aN_b}{N(N-1)}\left\{\frac{2N_aN_b - N}{N} + \frac{C-N+2}{(N-2)(N-3)}[N(N-1) - 4N_aN_b + 2]\right\}.$$

This allows rapid, exact, and direct assessment of the degree of significant similarity between the IDH and non-IDH groups.

Next, a variety of node centrality measures were calculated for each node of the MSTs drawn over the IDH and non-IDH groups separately. Six measures of centrality were assessed for each node included: 1) degree centrality; 2) total degree count of neighbors; 3) through-space closeness; 4) through-tree closeness; 5) betweenness; and 6) eigenvector. Degree centralities are simply the number of other nodes to which the node of interest is directly attached; closeness is the inverse of average distance to all other nodes, either through space or through the MST connections; betweenness indicates the proportion of all the shortest paths between nodes in the MST that pass through the node of interest; and eigenvalue centrality, roughly speaking, combines the concepts of degree and betweenness by relating each MST node to the entries of the principal eigenvector of the MST connectivity matrix [16].

The k-MST is an extension of the MST, found by repeating the MST algorithm k times, each time excluding all the connections previously chosen. This allows a richer level of connectivity information which in turn can improve statistical test results such as edge-counting. Here, we use the k-MST as a representation of the underlying structure of multivariate data and apply a simple 2D force-directed layout methodology whereby nodes that are directly connected in the k-MST experience an attractive force. "Reflation" is carried out after each iteration, to counteract the tendency of a wholly-attractive configuration of forces to collapse to a point. Because the k-MST contains only a small fraction of the possible pairwise connections between nodes, and because there are no explicit repulsive forces to calculate, our method allows rapid creation of 2D representations of arbitrarily high-dimensional radiomics datasets.

For the kMST force-directed layouts method for dimension reduction, the steps are as follows: using the distance matrix over the pooled patient nodes, the minimal spanning tree algorithm is iteratively applied, each time setting the distance matrix entries corresponding to chosen edges to a high value so that they are not chosen again. The edges chosen by each successive MST calculation are then saved. For the present work, we used a 5-MST, or 5 iterations.

The method is then initialized by assigning the nodes to random positions within the unit square (we used a uniform distribution was used for this purpose). Next, position updates are iteratively calculated, by summing the attractive 'forces' exerted on each node by only its direct neighbors within the MST. The attractive force is "spring-like" in that it increases linearly with distance between nodes.

To avoid having to calculate numerous repulsion effects between all nodes not connected in the kMST, we instead implemented an 'inflationary' step: at the end of each iteration the coordinates of the nodes are automatically rescaled to fit just inside the unit square. This inflationary step preserves the configuration changes of each position update while preventing the whole configuration from collapsing to a point.

Two parameters are used to generate the position updates: dEq, the equilibrium distance where attraction between kMST neighbors becomes repulsion with decreasing distance; and kAtt, the relative strength of the attractive force. For this study, we used values of dEq = 0.025 and kAtt = 0.015.

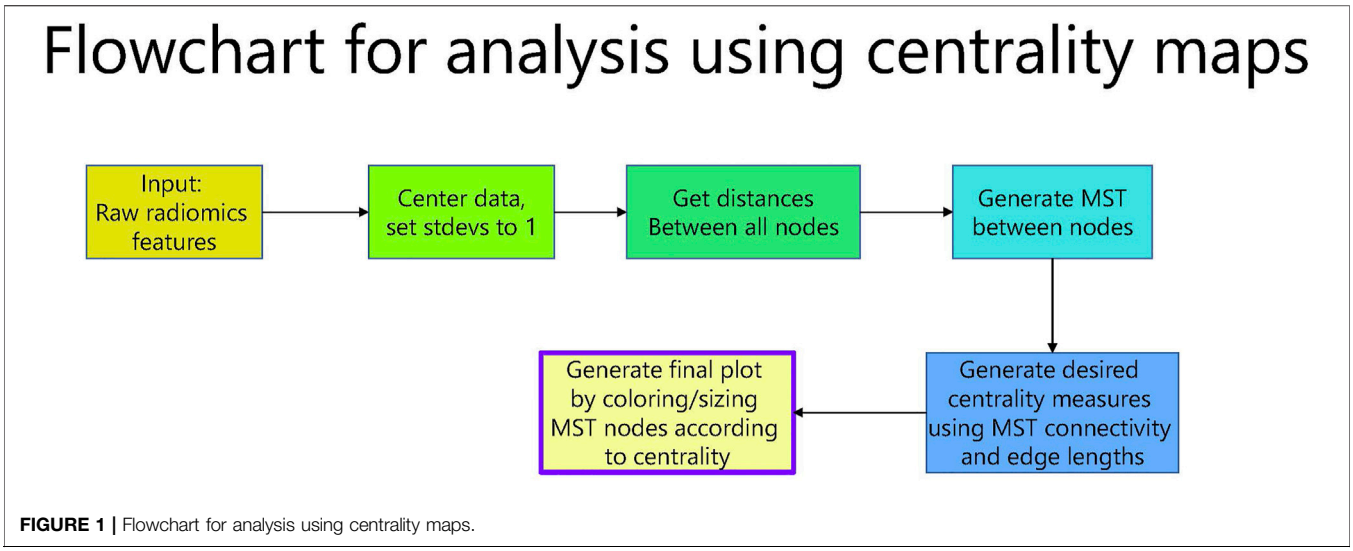


FIGURE 1 | Flowchart for analysis using centrality maps.

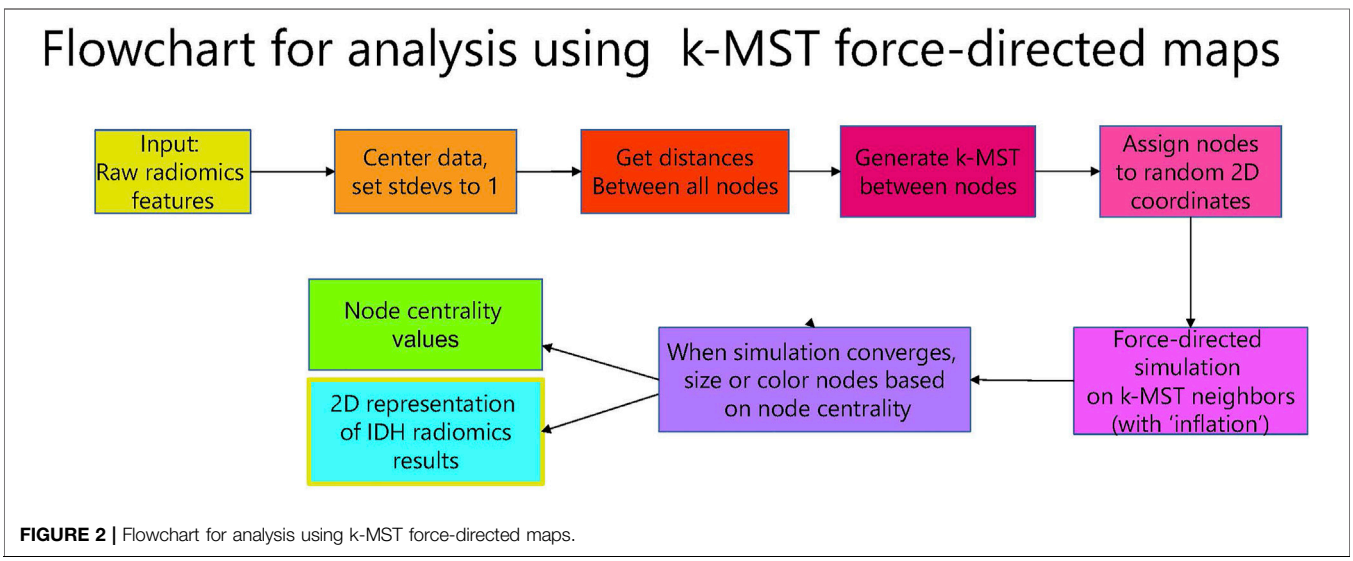


FIGURE 2 | Flowchart for analysis using k-MST force-directed maps.

All calculations (after MRI acquisition and processing, and radiomics feature extraction) were implemented directly using a custom MATLAB pipeline. Scripts used are available on Github at https://github.com/Ghoshlab/OSimonScripts.

RESULTS

Node centralities for IDH-mutated and IDH-wildtype patients are displayed using six different centrality definitions in **Figure 3** and **Supplementary Figure S1**, using node size and color to represent centrality. Notably, in this case, the same small number of nodes were consistently chosen as "most central", despite the wide differences in the centrality definitions applied. Specifically, for the IDH-mutated gliomas, Patient 24 was the 'most central' for node degree, node neighbor degree, through-space closeness, and eigenvector centrality, while Patient 80 was the most central in the case of through-tree closeness and betweenness. Among the IDH wild type gliomas, three nodes were prominent: Patient 35 (degree centrality), Patient 37 (neighbor degree, through-tree closeness), and Patient 39 (through-space closeness, eigenvector).

These nodes are usually located towards the "center" of the MST, usually at a junction between several sub-trees. Conversely, the lowest-centrality nodes are invariably found at the edges of the MST, among nodes with only one connection ("leaves"). These observations confirm that these measures do indeed reflect the intuitive idea of centrality.

Additionally, when PCA reduction was used to decrease the number of features, nearly all the central nodes remained the same, with the sole exception that in the IDH-wildtype gliomas the highest eigenvector centrality shifted from Patient 39 to

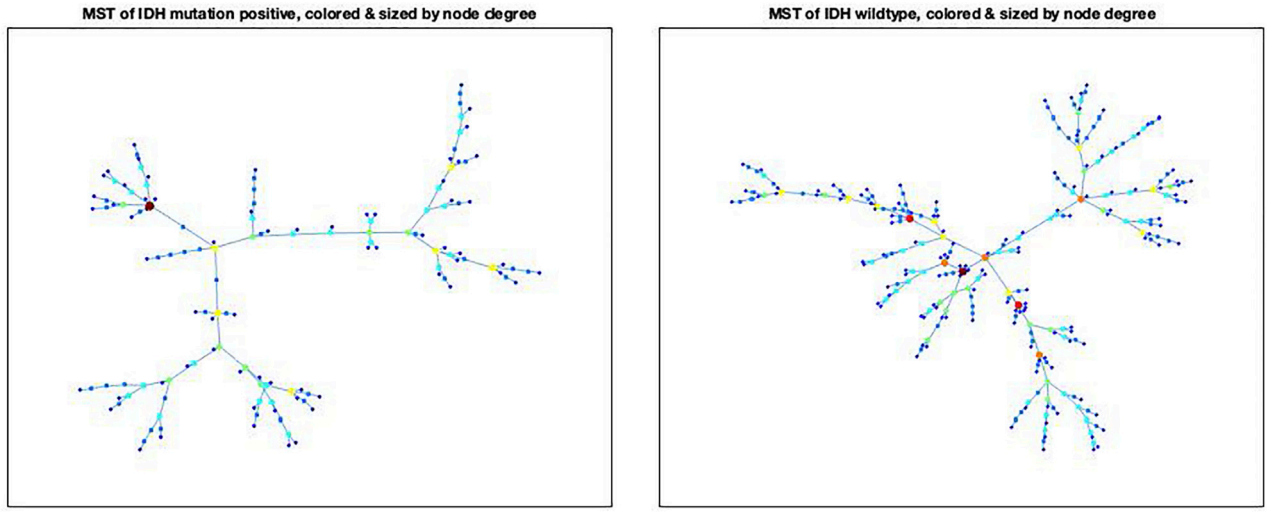


FIGURE 3 | MSTs for the IDH mutation positive (**Figure 3A**) and IDH mutation negative tumors (**Figure 3B**), with node sizes proportional to degree of the nodes. The *X*- and *Y*-axes represent spatial coordinates for visualization of the minimum spanning trees.

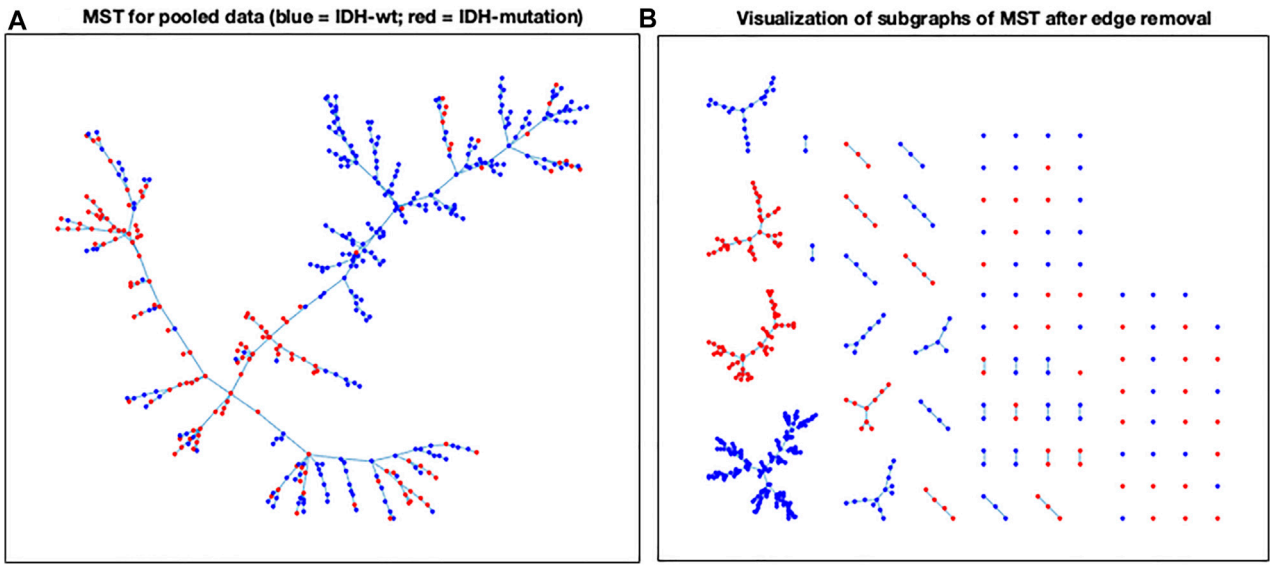


FIGURE 4 | (A) The pooled MST before the hybrid-edge cut; **(B)** A 2-dimensional representation of the subtrees of the MSTs after removing the hybrid edges. The X- and Y-axes represent spatial coordinates for visualization of the minimum spanning trees.

Patient 35 (**Supplementary Figure S2**). This close correspondence suggests that the centrality measures and MST algorithms used are robust to complex manipulations and changes of coordinates, such as those which occur using PCA.

The MVR results for our radiomics dataset (**Figure 4**) showed a clear distinction between the IDH-mutated and IDH-wildtype groups, consistent with prior literature reporting the strong effect of this mutation on radiomics profiles. Beginning with the pooled MST, the number of separate trees that would be expected in the null case (188.6) far exceeds the actual number resulting from the cut (91). This amounts to a difference of −9.82 standard deviations, effectively excluding the possibility that the groups differ according to chance. Thus, the MVR test of Friedman and Rafsky rejects the null hypothesis of no difference between the two groups with a p-value less than 1×10^{-32}.

As was the case with the node centralities, the MVR test carried out with PCA reduction to 48 features (**Supplementary**

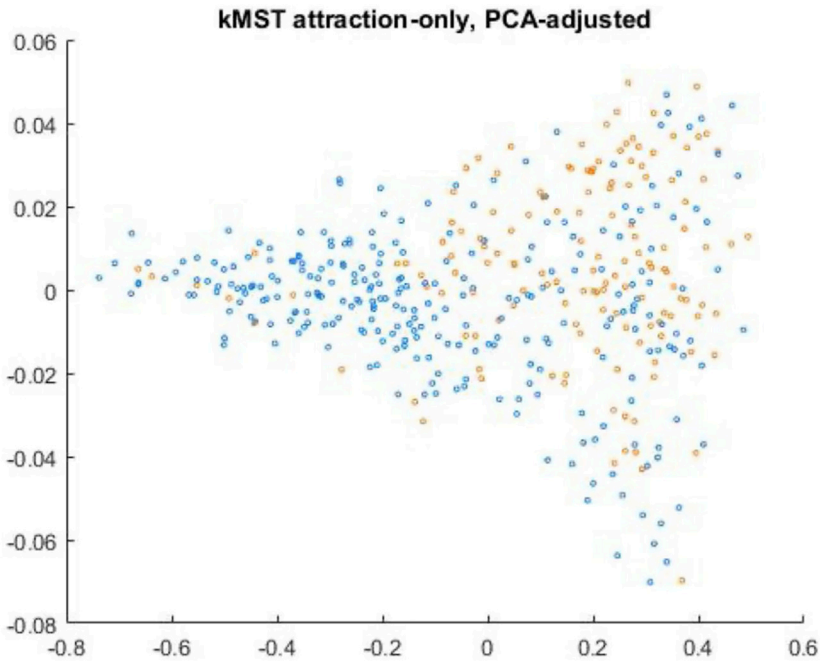


FIGURE 5 | Representative example of 5-MST force-directed result using PCA-reduced data. IDH-mutated patients are shown in orange, while IDH-wildtype is blue. The algorithm was run for 1800 iterations, with the values of kEq and kAtt held constant. The *X*- and *Y*-axes represent spatial coordinates for visualization of the minimum spanning trees.

Figure S3) yielded results similar to the original dataset. While the null expectation value of the number of trees remains the same by definition, number of trees from the actual cut (97), and the total standard deviations from the expectation value (−9.21) reveal a change of only half a standard deviation despite the PCA manipulation. This again helps establish the robustness of these methods to feature selection.

Applying the kMST-force directed algorithm to the 5-MST drawn over the pooled data-points, we found that the method rapidly and effectively produces readily interpretable visual layouts of the group structure. A representative result is shown in **Figure 5**. The results of three randomly-initialized runs of the algorithm using 3-MST, 5-MST, and 7-MST respectively (**Supplementary Figures 4A–I**) show that, despite the random initialization and the large number of datapoints involved, the final configurations produced by the algorithm are remarkably consistent overall (notwithstanding mirror-symmetry and rotations), and also reveal a clear though not perfect spatial separation between the two genotype groups. As might be expected, the runs using the lowest-complexity k-MST (the 3-MST) show somewhat more variability in the final structure and also a somewhat different final structure from the others, whereas the 5-MST and 7-MST show quite good consistency both between random initializations and between each other. This strongly indicates that the k-MST, despite its much-simplified structure with respect to the full graph, contains the information necessary for meaningful and reproducible dimension reduction of the data and that its local minimum under force calculation is likely unique. Furthermore, as in the previous approaches, the layouts were consistent and quite similar even when PCA reduction was first carried out (data not shown).

Given that the k-MST force-directed algorithm did not perfectly separate the two classes—more non-IDH nodes are present in the region dominated by IDH than vice-versa--we wondered whether the centrality measures would reasonably reflect the location of the nodes with respect to the overall distribution of class examples within the 2D layout, i.e., whether the nodes closest to the center of their class distribution in the layout would have the highest centralities as well. We found this to be generally the case, though the choice of centrality measure does have an effect. Interestingly, we find that eigenvector centrality gives much lower precedence to the "IDH-like" non-IDH cases found in the predominantly region, yielding better separation between IDH and wild-type regions when *both* centrality and kMST layout are used (**Supplementary Figure 4F**). Other measures, particularly betweenness, seem to be much less effective at distinguishing 'IDH-mimics' from the other wild-types—there are a few IDH-like cases with relatively high betweenness, but this centrality also yield an even clearer overall divide between the main IDH and wild-type than does eigenvector (**Supplementary Figure 4E**). Degree-based or closeness-based centrality measures, on the other hand, do not appear to be especially effective at discriminating the central regions of the two classes in the kMST layout, even when the difference is exaggerated by squaring the centrality (**Supplementary Figures 4A–D**).

DISCUSSION

In the foregoing, we have demonstrated the feasibility of a simple graph-theoretical toolkit to address the problems presented by large, high-dimensional radiomic datasets. For the example dataset drawn from IDH-mutated and IDH-wildtype glioma patients, we were able to use MST-based methods to establish highly significant differences between the two groups, to identify patients that are most "representative" of each group using a combination of centrality measures, and to use a simple kMST-based force-directed method to illustrate those centrality measures within the context of a two-dimensional map of the data. Importantly, we find this method converges to a very consistent configuration which can effectively segregate IDH-mutated and IDH wild-type gliomas, especially when combined with centrality measures.

Although there is overlap between the two classes, there is a very clear difference in the overall localization between the two groups. Particularly focusing on the 5-MST and 7-MST—which converged consistently to a roughly triangular 2D point distribution—we see that IDH-wildtype patients tend to group strongly in one corner of the triangle with almost no IDH-mutated patients present, while at another corner and towards the center of the triangle IDH-mutated cases predominate. Notably, a significant minority of IDH-wildtype patients exhibit IDH-like localization, suggesting that the IDH mutation is sufficient but not necessary for an "IDH-like" phenotype. This means that stratification by radiomic features alone may be vulnerable to false positives, in the sense that patients with typically IDH-like radiomic features (at least according to our mode of analysis) may nonetheless lack the IDH mutation. Conversely, there might be further subtypes of IDH-wildtype populations to further characterize, although our study is not sufficiently powered for this type of discovery.

We hypothesize this may be due to other mutations or combinations of mutations that partially phenocopy the IDH mutation. Future genetic studies may help elucidate these IDH-mimicking gene combinations, perhaps by looking for epistatic effects on the IDH pathway [17]. Furthermore, it would be highly worthwhile to track the outcomes of IDH-like patients, to determine whether they in fact share the prognosis generally associated with IDH mutation proper. As noted, our results suggest that the combination of eigenvector centrality and k-MST layouts may be especially useful in distinguishing between wild-type glioma patients that are IDH-like and those that are more 'typical'.

Among the force-directed layouts methods, the Barnes-Hut algorithm, which coalesces sufficiently distant points into a single center of mass by constructing a quadtree based on a distance criterion, may be the best-known means of simplifying force calculations for very large numbers of points [18]. In our case, however, the use of the k-MST greatly reduces the number of interactions that need to be calculated at each iteration and inherently restrains the calculation only to "sufficiently close" node pairs, so that the Barnes-Hut approach is unnecessary. The replacement of explicit repulsion term calculations with a simple "inflationary" step after each iteration also simplifies the overall calculation while likely reducing the chance of the configuration becoming trapped in "geometrically-frustrated" local minima. Notably, for 5-MST and higher, we saw no evidence of alternative minima for our simulation—in all conditions tested, the overall arrangement of points did not differ qualitatively from the overall pattern seen here. It will be interesting to see if this general pattern is observed for different datasets.

The choice of k for the k-MST is likely to depend on n, the number of observations being handled by the simulation, as **Figure 3** suggests that choosing k too low means that the final configuration will be underdetermined. Arguments based on stochastic-block models suggest that there is a minimum value of k below which there is inadequate information to reconstruct the true underlying class-membership; however, this values grows only slowly, as Ω (log n) [19].

With respect to other common dimensional-reduction methods, our "spring-like" approach with stochastic initialization and gradient descent is related to such familiar approaches like t-SNE [20], though we do not assign neighbors using a Gaussian or t-distribution but rather use repeated application of the MST algorithm itself. One potential future issue is that the k-MST approach is likely to be sensitive to class imbalance. If one class comes to be vastly outnumbered by another its members may be less likely to be connected in the k-MST, and hence will not experience the attractive forces that produce strong clustering; conversely, the attraction between relatively few similar nodes may be overwhelmed by the attractive force of a much larger number of adjoining, yet dissimilar nodes.

Even in this case, however, we believe it is likely that the members of the less-populated class will tend either to have a higher chance of being connected through the k-MST (by a similar reasoning to that which motivates the MVR test itself), or will form part of a larger region containing "similar" nodes of the other class (as we see with the considerable number of WT patients whose nodes consistently segregate into the IDH-dominated region). One possible solution to this potential limitation, inspired by work for the MVR test [21], might be simply to weight the attractive forces within the k-MST simulation in inverse proportion to the number of nodes in the class, so that "rarer" nodes attract each other most strongly.

One potential computational limitation of our approach relates to its dependence on the creation of MST, which requires creation of a full-graph distance matrix. Since this contains n^2 pairwise distances, the computational overhead will increase with O (n^2), limiting the number of data-points that can be calculated. However, clustering-based approaches are known which can be used to generate approximate MSTs that run in O ($n^{3/2}$), substantially speeding up the distance-matrix bottleneck [22]. A natural next step, therefore, will be to implement and evaluate approximate MSTs, which should allow the processing of hundreds of thousands of data-points in reasonable time.

It is well-known that radiomics approaches can be vulnerable to false positives, particularly in the case where there are more radiomic features than there are patients [23]. Furthermore, it has been noted that "choice of the classification model could lead to variations in the predictive values of the radiomic features up to >30%" [23]. While our approach largely avoids feature-selection issues by effectively condensing the data into a higher-level statistical, graph-theoretical, indeed structural question, further validation studies on other types of radiomics and imaging data are clearly indicated, as such studies can help to eliminate false positives [23].

Even taking this into account, we believe the centrality/MVR/k-MST force-directed combination approach presented here has the potential to greatly simplify the analysis of radiomics data, while simultaneously rendering it far more readily interpretable. By relying on the simple, MVR test—which is parameter-free except with respect to the variable C, itself derived from the pooled data MST--we avoid numerous somewhat arbitrary aspects of testing and analysis with high-dimensional data. Since the MST itself does not depend on any arbitrary parameters, this too provides a simpler, possibly more "objective" approach. As raised by a reviewer, there is an important theoretical question, which involves our centrality based analysis. It is based on the MST, and the relationship between MST-based centrality with the original data-based graph centrality remains an open problem.

In conclusion, we have developed a combination of graph-theoretical approaches that provide rapid visualization, significance testing, and dimensionality reduction for very high-dimensional radiomics (and other) datasets, with the potential for considerable streamlining of the workflow and improved "explainability". Future investigations will help gauge the effectiveness of this general approach to other radiomics use-cases, as well as to other high-dimensional medical data.

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 author.

AUTHOR CONTRIBUTIONS

OS and DG formulated the design of the study. OS developed and implemented the methodology. All authors provided feedback on its implementation. Y-SC provided the data for the study. OS wrote the first draft, and all authors contributed to writing of subsequent drafts of the manuscript. All authors approved the final version of the manuscript.

FUNDING

This research has been partially supported by the National Cancer Institute through Grant R01 CA129102. The funding agency played no role in the design or analysis of the study.

ACKNOWLEDGMENTS

OS and DG would like to acknowledge the support of NCI R01 CA129102. DG acknowledges the support of the Grohne-Stepp Endowment from the University of Colorado Cancer Center.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphy.2022.783765/full#supplementary-material

REFERENCES

1. De Bruijne M. Machine Learning Approaches in Medical Image Analysis: From Detection to Diagnosis. *Med Image Anal* (2016) 33:94–7. doi:10.1016/j.media.2016.06.032
2. Bhandari AP, Liong R, Koppen J, Murthy SV, Lasocki A. Noninvasive Determination of IDH and 1p19q Status of Lower-Grade Gliomas Using MRI Radiomics: a Systematic Review. *AJNR Am J Neuroradiol* (2021) 42(1):94–101. doi:10.3174/ajnr.a6875
3. Calabrese E, Villanueva-Meyer JE, Cha S. A Fully Automated Artificial Intelligence Method for Non-invasive, Imaging-Based Identification of Genetic Alterations in Glioblastomas. *Sci Rep* (2020) 10(1):11852–11. doi:10.1038/s41598-020-68857-8
4. Choi YS, Bae S, Chang JH, Kang S-G, Kim SH, Kim J, et al. Fully Automated Hybrid Approach to Predict the IDH Mutation Status of Gliomas via Deep Learning and Radiomics. *Neuro-Oncol.* (2021) 23(2):304–13. doi:10.1093/neuonc/noaa177
5. Rogers W, Thulasi Seetha S, Refaee TAG, Lieverse RIY, Granzier RWY, Ibrahim A, et al. Radiomics: from Qualitative to Quantitative Imaging. *Bjr* (2020) 93:20190948. doi:10.1259/bjr.20190948
6. Hosny A, Aerts HJ, Mak RH. Handcrafted versus Deep Learning Radiomics for Prediction of Cancer Therapy Response. *Lancet Digit Heealth* (2019) 1(3):E106–107. doi:10.1016/S2589-7500(19)30062-7
7. Arpit D, Jastrzębski S, Ballas N, Krueger D, Bengio E, Kanwal MS, et al. *A Closer Look at Memorization in Deep Networks*. PMLR (2017). 233–42.
8. Gower JC, Ross GJS. Minimum Spanning Trees and Single Linkage Cluster Analysis. *Appl Stat* (1969) 18(1):54–64. doi:10.2307/2346439
9. Louf R, Jensen P, Barthelemy M. Emergence of Hierarchy in Cost-Driven Growth of Spatial Networks. *Proc. Natl. Acad. Sci. U.S.A.* (2013) 110(22):8824–9. doi:10.1073/pnas.1222441110
10. Stam CJ, Tewarie P, Van Dellen E, Van Straaten ECW, Hillebrand A, Van Mieghem P. The Trees and the Forest: Characterization of Complex Brain Networks with Minimum Spanning Trees. *Int J Psychophysiol* (2014) 92(3):129–38. doi:10.1016/j.ijpsycho.2014.04.001
11. Freeman LC. A Set of Measures of Centrality Based on Betweenness. *Sociometry* (1977) 40:35–41. doi:10.2307/3033543
12. Chen H, Friedman JH. A New Graph-Based Two-Sample Test for Multivariate and Object Data. *J Am Stat Assoc* (2017) 112(517):397–409. doi:10.1080/01621459.2016.1147356
13. Friedman JH, Rafsky LC. Multivariate Generalizations of the Wald-Wolfowitz and Smirnov Two-Sample Tests. *Ann Stat* (1979) 697–717. doi:10.1214/aos/1176344722

14. Friedman JH, Rafsky LC. Graph-theoretic Measures of Multivariate Association and Prediction. *Ann Stat* (1983) 377–91. doi:10.1214/aos/1176346148
15. 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–e107. doi:10.1158/0008-5472.can-17-0339
16. Newman MEJ. Mathematics of Networks. *New Palgrave Encycl Econ* (2008) 2: 1–8. doi:10.1057/978-1-349-95121-5_2565-1
17. van de Haar J, Canisius S, Yu MK, Voest EE, Wessels LFA, Ideker T. Identifying Epistasis in Cancer Genomes: a Delicate Affair. *Cell* (2019) 177(6):1375–83. doi:10.1016/j.cell.2019.05.005
18. Barnes J, Hut P. A Hierarchical O(N Log N) Force-Calculation Algorithm. *nature* (1986) 324(6096):446–9. doi:10.1038/324446a0
19. Moore C. *The Computer Science and Physics of Community Detection: Landscapes, Phase Transitions, and Hardness*. ArXiv Prepr ArXiv170200467 (2017).
20. Van der Maaten L, Hinton G. Visualizing Data Using T-SNE. *J Mach Learn Res* (2008) 9(11).
21. Chen H, Chen X, Su Y. A Weighted Edge-Count Two-Sample Test for Multivariate and Object Data. *J Am Stat Assoc* (2018) 113(523):1146–55. doi:10.1080/01621459.2017.1307757
22. Zhong C, Malinen M, Miao D, Fränti P. *Fast Approximate Minimum Spanning Tree Algorithm Based on K-Means*. Springer (2013). p. 262–9. doi:10.1007/978-3-642-40261-6_31Fast Approximate Minimum Spanning Tree Algorithm Based on K-Means
23. Yip SSF, Aerts HJWL. Applications and Limitations of Radiomics. *Phys. Med. Biol.* (2016) 61(13):R150–R166. doi:10.1088/0031-9155/61/13/r150

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.

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 19 May 2022
doi: 10.3389/fonc.2022.831795

A Radiomics Model Based on Gd-EOB-DTPA-Enhanced MRI for the Prediction of Microvascular Invasion in Solitary Hepatocellular Carcinoma ≤ 5 cm

Chengming Qu[1], Qiang Wang[2,3], Changfeng Li[1], Qiao Xie[4], Ping Cai[4], Xiaochu Yan[5], Ernesto Sparrelid[6], Leida Zhang[1], Kuansheng Ma[1]* and Torkel B. Brismar[2,3]*

[1] Institute of Hepatobiliary Surgery, Southwest Hospital, Army Medical University, Chongqing, China, [2] Division of Medical Imaging and Technology, Department of Clinical Science, Intervention and Technology (CLINTEC), Karolinska Institutet, Stockholm, Sweden, [3] Division of Radiology, Department of Clinical Science, Intervention and Technology (CLINTEC), Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden, [4] Department of Radiology, Southwest Hospital, Army Medical University, Chongqing, China, [5] Department of Pathology, Southwest Hospital, Army Medical University, Chongqing, China, [6] Division of Surgery, Department of Clinical Science, Intervention and Technology (CLINTEC), Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

Aim: The aim of this study is to establish and validate a radiomics-based model using preoperative Gd-EOB-DTPA-enhanced MRI to predict microvascular invasion (MVI) in patients with hepatocellular carcinoma ≤ 5 cm.

Methods: Clinicopathologic and MRI data of 178 patients with solitary hepatocellular carcinoma (HCC) (≤5 cm) were retrospectively collected from a single medical center between May 2017 and November 2020. Patients were randomly assigned into training and test subsets by a ratio of 7:3. Imaging features were extracted from the segmented tumor volume of interest with 1-cm expansion on arterial phase (AP) and hepatobiliary phase (HBP) images. Different models based on the significant clinical risk factors and/or selected imaging features were established and the predictive performance of the models was evaluated.

Results: Three radiomics models, the AP_model, the HBP_model, and the AP+HBP_model, were constructed for MVI prediction. Among them, the AP+HBP_model outperformed the other two. When it was combined with a clinical model, consisting of tumor size and alpha-fetoprotein (AFP), the combined model (AP+HBP+Clin_model) showed an area under the curve of 0.90 and 0.70 in the training and test subsets, respectively. Its sensitivity and specificity were 0.91 and 0.76 in the training subset and 0.60 and 0.79 in the test subset, respectively. The calibration curve illustrated that the combined model possessed a good agreement between the predicted and the actual probabilities.

OPEN ACCESS

Edited by:
Oliver Diaz,
University of Barcelona, Spain

Reviewed by:
Nguyen Minh Duc,
Pham Ngoc Thach University of Medicine, Vietnam
Nesrin Dogan,
University of Miami, United States
Daniel Rodriguez Gutierrez,
Nottingham University Hospitals NHS Trust, United Kingdom

***Correspondence:**
Qiang Wang
qiang.wang@ki.se
Kuansheng Ma
kuanshengma@outlook.com

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

Received: 08 December 2021
Accepted: 19 April 2022
Published: 19 May 2022

Citation:
Qu C, Wang Q, Li C, Xie Q, Cai P, Yan X, Sparrelid E, Zhang L, Ma K and Brismar TB (2022) A Radiomics Model Based on Gd-EOB-DTPA-Enhanced MRI for the Prediction of Microvascular Invasion in Solitary Hepatocellular Carcinoma ≤ 5 cm. Front. Oncol. 12:831795. doi: 10.3389/fonc.2022.831795

Conclusions: The radiomics-based model combining imaging features from the arterial and hepatobiliary phases of Gd-EOB-DTPA-enhanced MRI and clinical risk factors provides an effective and reliable tool for the preoperative prediction of MVI in patients with HCC ≤ 5 cm.

Keywords: radiomics, microvascular invasion, Gd-EOB-DTPA, magnetic resonance imaging, hepatocellular carcinoma, prediction model

INTRODUCTION

Hepatocellular carcinoma (HCC) is a common gastrointestinal malignant tumor, ranks sixth in incidence rate, and is the fourth leading cause of tumor-related mortality worldwide (1). Liver resection is one of the curative treatments for HCC. Despite recent advances in surgical techniques and perioperative management, HCC still bears a poor prognosis with a 5-year recurrence of 50%–70% after liver resection (2). This also applies to inpatients with early-stage HCC, where a 5-year recurrence rate of as high as approximately 60% has been reported (3).

Microvascular invasion (MVI) has been reported as an independent, well-established risk factor for HCC recurrence and poor overall survival rate (4). The reported incidence rate of MVI ranges between 15% and 57% (5). Patients with MVI experienced an early recurrence compared with those without MVI with a mean time to recurrence of 12 months versus 42 months (4). Therefore, it is of importance to preoperatively identify MVI to optimize the treatment strategy and predict the prognosis. However, the diagnosis of MVI is mainly made postoperatively by a histopathology exam on the excised tumor, which has little or no influence on surgical decision-making. Although radiological features on computed tomography (CT) or magnetic resonance imaging (MRI) such as rim arterial enhancement and non-smooth tumor margin are also applied to predict MVI, a consensus about the efficacy of these features has not been reached (6, 7).

With the development of modern imaging and computing techniques, it might be possible to detect subtle changes in the tumor or its margin. Radiomics, a technique that can extract high-throughput imaging features from routine biomedical images for quantitative analysis, has attracted intensive interest in recent years (8, 9). Because it may provide additional information, radiomics has turned out to be a promising tool for accurate tumor detection, diagnosis, grading, and prognosis prediction in tumors such as rectal cancer and HCC (10–12).

Gd-EOB-DTPA-enhanced MRI is a commonly used modality in diagnosis and characterization of HCC, and is noted for the specific hepatobiliary phase. Taking advantage of radiomics on Gd-EOB-DTPA-enhanced MRI, it has shown encouraging results for MVI prediction at HCC (13, 14). However, previous studies have not limited their data regarding tumor number and tumor size of HCC (11, 15, 16). With improved imaging and the use of screening programs, HCC is increasingly detected at an early stage. The incidence of MVI in HCC ≤ 5 cm, or within Milan criteria, has been reported to be as high as 40% (17). By identifying these patients preoperatively, their management and long-term survival might be improved as alternative treatment options could be considered. Therefore, the aim of this study was to develop and validate a radiomics prediction model based on preoperative Gd-EOB-DTPA-enhanced MRI to predict MVI in patients with a single HCC ≤ 5 cm in diameter.

Abbreviations: HCC, Hepatocellular carcinoma; MVI, Microvascular invasion; CT, Computed tomography; MRI, Magnetic resonance imaging; AP, Arterial phase; HBP, Hepatobiliary phase; VOI, Volume of interest; GLCM, Gray-level co-occurrence matrix; GLRLM, Gray-level run length matrix; GLSZM, Gray-level size zone matrix; GLDM, Gray-level dependence matrix; LoG, Laplacian of Gaussian; ICC, Interclass correlation coefficient; RFE, Recursive feature elimination; AFP, Alpha-fetoprotein; AST, Aspartate transaminase; ALT, Alanine transaminase; ROC, Receiver operating characteristic curve; AUC, Area under ROC.

MATERIALS AND METHODS

Study Design and Patients

This retrospective study was performed at a single tertiary medical center, Southwest Hospital of Army Medical University, Chongqing, China. The research protocol was approved by the Institutional Review Board of the hospital (No. 2017KY50), and written informed consent was waived due to the retrospective nature of the study.

Through a search in the hospital information system, the records of all patients undergoing liver resection between May 2017 and November 2020 were retrieved. Patients were considered eligible in this study according to the following inclusion criteria: (1) patients undergoing their first liver resection due to HCC, (2) solitary liver tumor with a diameter ≤ 5 cm on MRI with no macroscopic sign of vascular invasion, (3) Gd-EOB-DTPA-enhanced MRI exam within 1 month before liver resection, and (4) available pathology report of MVI status. The exclusion criteria were as follows: (1) previous antitumor treatment, such as radiofrequency ablation and transarterial chemoembolization; (2) intra- or extrahepatic metastasis; and (3) low-quality imaging, not satisfying analysis requirement. A total of 178 consecutive patients were included in the final cohort, and they were randomly split into a training subset and a test subset by a ratio of 7:3. **Figure 1** describes the process of patient selection and **Figure 2** supplies the steps of this study.

Clinicopathologic Variables and MVI

Demographic information, blood biochemistry, and pathology results were retrieved through the hospital information system. MVI was diagnosed according to the Chinese guidelines for standardized diagnosis of primary liver cancer (18). According to the guidelines, MVI is defined as when cancerous emboli can be

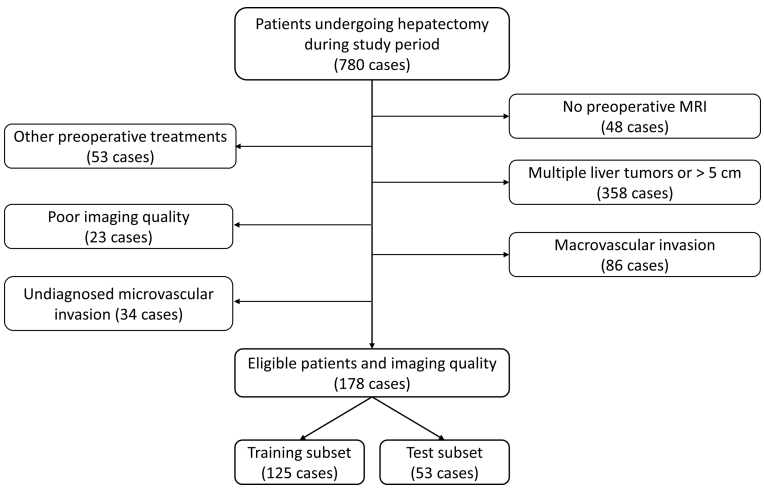


FIGURE 1 | Flowchart of patient selection in this study.

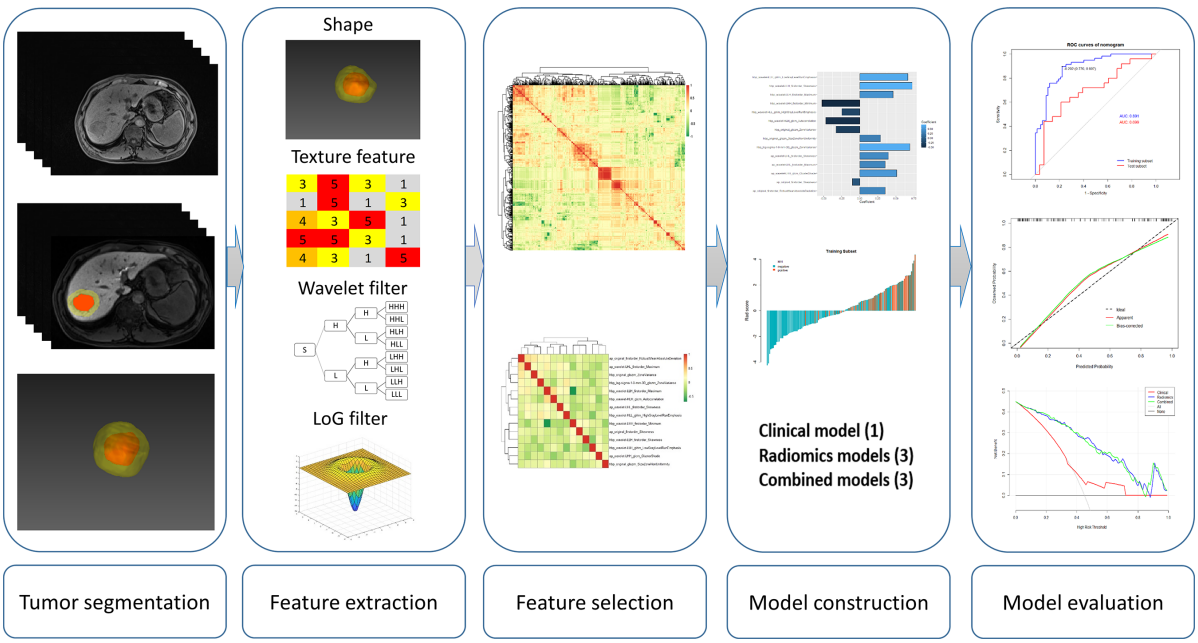


FIGURE 2 | Workflow of key steps in this study.

observed in the vasculature lined with endothelial cells outside the tumor margin under microscopy (14, 19). In this study, the tumors with this finding were classified as MVI (+), regardless of its number or distance from the tumor. Tumors with no cancerous emboli detected were classified as MVI (−).

Gd-EOB-DTPA-Enhanced MRI Acquisition

All MRI was performed on the same 3.0-T MRI scanner (Magnetom Trio, Siemens Healthcare) with a 6-channel body coil. The contrast agent Gd-EOB-DTPA (Primovist, Bayer Pharma) was injected through the anterior cubital vein at a dose of 0.1 ml/kg with an injection rate of 1.0 ml/s, followed by an immediate injection of 20 ml of saline at the same rate. After the injection of Gd-EOB-DTPA, arterial phase scanning was triggered by the signal intensity at the lower end of the abdominal aorta, followed by portal phase scanning (60 s), equilibrium phase scanning (180 s), and hepatobiliary phase scanning (15 min) with three-dimensional volume interpolated

breath-hold (3D-VIBE) T1WI. The detailed scanning protocol is provided in the **Supplementary Material**.

Tumor Segmentation and Volume of Interest Dilation

Tumor segmentation on arterial and hepatobiliary phases (hereafter referred to as AP and HBP, respectively) was conducted by two radiologists (QX and PC with 8 and 20 years' experience in abdominal radiology, respectively) who were blinded to the patients' clinical information. Tumor delineation was performed manually using the open-source software ITK-SNAP (version 3.8.0, http://www.itksnap.org/). The delineated tumor was further expanded at a radius of 10 mm (20, 21) using a topologic algorithm in Python (version 3.8), and the expansion would cease automatically if it reached the liver edge for the marginal liver tumors. The expanded volume of interest (VOI) was then used for radiomics feature extraction (**Figure 3**).

Radiomics Feature Extraction

To increase the reliability of the radiomics features, the image voxel size was resampled into 1×1×1 mm^3 (interpolator: B-spline) and the bin width of the intensity histogram was discretized into 25. After preprocessing the images, the following six categories of imaging features were extracted:

1. Shape, including 2D and 3D (n = 14);
2. First-order statistics (n = 18);
3. Gray-level co-occurrence matrix (GLCM)-derived texture (n = 22);
4. Gray-level run length matrix (GLRLM)-derived texture (n = 16);
5. Gray-level size zone matrix (GLSZM)-derived texture (n = 16); and
6. Gray-level dependence matrix (GLDM)-derived texture (n = 14).

Imaging features of categories (2) to (6) were also extracted from transformed images using the wavelet filter (688 features) and the Laplacian of Gaussian (LoG) filter with a kernel size of 1.0 mm (86 features). Both imaging preprocessing and feature extraction were performed by using the pyradiomics package (version 3.0) (22) in Python (version 3.8).

Imaging features extracted from AP and HBP were labeled with the prefix "ap_" and "hbp_" to each radiomics feature name, respectively. Examples are as follows: "ap_original_firstorder_Skewness" denotes the skewness of first-order features derived from AP images while "hbp_log-sigma-1-0-mm-3D_glszm_ZoneVariance" describes the zone variance in GLSZM features derived from LoG filter transferred HBP images.

To evaluate the reproducibility of the radiomics features and inter-rater agreement, images of 30 randomly chosen patients were contoured by both radiologists independently. The interclass correlation coefficient (ICC) estimates were determined by using the single-rater, absolute-agreement, 2-way random-effects model. The features were classified into "poor-to-moderate" (ICC <0.75) and "good-to-excellent" reliability (ICC ≥ 0.75), and those features with ICC ≥ 0.75 were selected (for the overlapped 30 patients, the measurements of the senior radiologist were selected) for model construction (23).

Radiomics Model Development and Validation

The AP and HBP imaging features with high reproducibility were adopted for radiomics model establishment. The feature analysis was performed by open-source software that is available at https://github.com/salan668/FAE (24). There was no need for upsampling or downsampling of the data as the percentage of MVI (−) and MVI (+) in the training subset was roughly balanced. The imaging features were first standardized using z-score normalization (subtract the mean value of each feature and then divide the difference by its standard deviation), followed by evaluation of the Spearman correlation coefficient of all features. Among each pair of features with a correlation coefficient > 0.90, one was randomly removed. The remaining features were applied for model construction using a recursive feature elimination (RFE) algorithm. RFE iteratively constructs the model using smaller and smaller sets of features and ranks the features according to their importance for the outcome prediction. To avoid potential overfitting, a desired number of features (<20) was applied when establishing the radiomics model using the RFE-logistic regression approach. A 10-fold cross-validation was applied to obtain a stable and robust model.

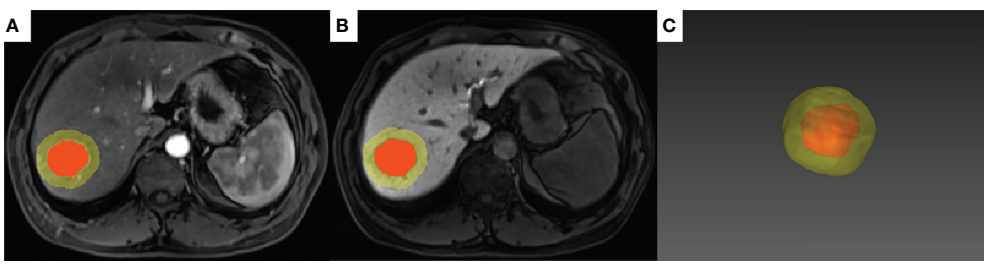


FIGURE 3 | A representative case of tumor segmentation with MVI (+) with 10-mm dilation from the tumor margin. The red area indicates the intratumoral region and the yellow area indicates the peritumoral region on the arterial phase **(A)** and hepatobiliary phase **(B)**. **(C)** 3D effect of the tumor segmentation with 10-mm expansion.

Three radiomics models were constructed using either AP features (AP_model), HBP features (HBP_model), or a combination of both AP and HBP features (AP+HBP_model). Radscore, indicating the relative risk of MVI for each patient, was calculated using each model according to the following formula:

$$\text{Radscore} = \text{intercept} + \text{coefficients} * \text{features}$$

The predictive performance of the calculated Radscore from each model to predict MVI in training and test subsets was then evaluated.

Construction and Evaluation of Prediction Models

To establish a clinical model, a univariable logistic regression analysis of the preoperatively clinicopathologic variables including age, gender, etiology of chronic liver disease, cirrhosis status, Child–Pugh grade, tumor size, alpha-fetoprotein (AFP) level, platelet counts, prothrombin time, albumin, bilirubin, aspartate transaminase (AST), and alanine transaminase (ALT) was first applied, with significant risk factors entering the multivariable regression analysis. A clinical model, hereafter denoted Clin_model, was constructed using the significant risk factors observed at the multivariable regression analysis.

The risk factors in the Clin_model and the Radscore were integrated into the construction of three combined models, i.e., AP+Clin_model, HBP+Clin_model, and AP+HBP+Clin_model. The efficacy of these models was validated in the test subset. Calibration curves were plotted to evaluate the predictive performance of the best model in both training and test subsets. Decision curve analysis was performed to evaluate the usefulness of the prediction models.

Statistical Analysis

Continuous variables were expressed as median with range and tested by Mann–Whitney *U* test to compare the difference between MVI (−) and MVI (+) groups. Categorical variables were presented as number (percentage), and chi-square test or Fisher's exact test was used to detect the differences between two groups. ICC was calculated by using the package "Pingouin" in Python. The performance of the prediction models was evaluated by receiver operating characteristic curves (ROCs), and the area under ROC (AUC), sensitivity, specificity, positive predictive value, negative predictive value, and accuracy were determined. p-values < 0.05 were considered statistically significant. Statistical analyses and randomizations were performed by R software (version 4.0.4, https://www.R-project.org/).

RESULTS

Clinicopathologic Characteristics of Patients and Clinical Model

The baseline characteristics of the entire cohort (n = 178), the training subset (n = 125), and the test subset (n = 53) are summarized in **Table 1**. There was no statistically significant difference between the two subsets. The incidence of MVI in the entire cohort was 45.5%.

After univariable and multivariable regression analyses, two risk factors in the training subset, AFP and tumor size, were selected for clinical model construction (**Table 2**). Although it was significant in both regression analyses, tumor differentiation status was excluded for modeling as it was a postoperative risk factor for MVI prediction. The AUCs of the Clin_model in training and test subsets were 0.64 (95% CI: 0.54–0.74) and 0.55 (95% CI: 0.38–0.71), respectively (**Figure 5A**, **Table 3**).

Feature Selection and Prediction Model Construction

Out of the 874 imaging features extracted from each Gd-EOB-DTPA-enhanced MRI phase, 560 features (64%) had sufficient reproducibility (ICC ≥ 0.75) for radiomics model construction. After removal of imaging features with high Pearson correlation coefficient, 10 AP features, 12 HBP features, and 14 features from a combination of both AP and HBP features with high ranking selected through the RFE algorithm were used for radiomics model construction (**Figure 4**, **Supplementary File**). The detailed features and their corresponding coefficients for the three radiomics models are described in the **Supplementary File**. The performance of the three radiomics models is illustrated in **Table 3** and the **Supplementary File**.

Performance Evaluation of the Models

Compared with the Clin_model, the overall performance of all three radiomics models was superior, with an AUC above 0.82 in the training subset and more than 0.56 in the test subset. Among the three radiomics models, the AP+HBP_model had the highest AUC, with 0.89 in the training subset and 0.66 in the test subset (**Table 3**). When combined with the clinical variables, the AP+HBP+Clin_model yielded an AUC of 0.90 (95% CI: 0.85–0.95) and 0.70 (95% CI: 0.55–0.84) in the training and test subsets, respectively (**Figures 5B, C**, the formula of the three combined models is provided in the **Supplementary File**). The sensitivity, specificity, positive predictive value, and negative predictive value were 0.91, 0.76, 0.77, and 0.91 in the training subset, and 0.60, 0.79, 0.71, and 0.69 in the test subset, respectively (**Table 3**). The calibration curves illustrated that the predicted probabilities of MVI were in good agreement with the observed probabilities with a C-index of 0.89 and 0.70 in the training and test subsets, respectively (**Figures 5D, E**). In terms of the clinical usefulness evaluation, the decision curve analysis illustrates that the implementation of the AP+HBP+Clin_model to predict MVI status should be beneficial compared with treating none or all of the patients as well as compared with the Clin_model or the AP+HBP_model (**Figure 5F**).

DISCUSSION

In this study, a radiomics prediction model based on imaging features extracted from preoperative Gd-EOB-DTPA-enhanced MRI to predict MVI in patients with a single HCC ≤ 5 cm in

TABLE 1 | Clinicopathologic characteristics of the patients.

Characteristics	Total (*n* = 178)	Training subset (*n* = 125)			Test subset (*n* = 53)			*p*-value[#]
		MVI (–) (*n* = 68)	MVI (+) (*n* = 57)	*p*-value	MVI (–) (*n* = 29)	MVI (+) (*n* = 24)	*p*-value	
Age (years) †	50 (28–78)	51 (30–72)	50 (31–78)	0.576	52 (29–73)	45 (28–72)	0.125	0.449
Gender								
Female	35 (19.7%)	17 (25.0%)	10 (17.5%)	0.429	4 (13.8%)	4 (16.7%)	1.000	0.428
Male	143 (80.3%)	51 (75.0%)	47 (82.5%)		25 (86.2%)	20 (83.3%)		
Etiology								
HBV	169 (94.9%)	66 (97.1%)	54 (94.7%)	0.659	27 (93.1%)	22 (91.7%)	1.000	0.454
None/Others	9 (5.06%)	2 (2.94%)	3 (5.26%)		2 (6.90%)	2 (8.33%)		
Cirrhosis								
Absent	35 (19.7%)	8 (11.8%)	15 (26.3%)	0.063	8 (27.6%)	4 (16.7%)	0.538	0.656
Present	143 (80.3%)	60 (88.2%)	42 (73.7%)		21 (72.4%)	20 (83.3%)		
Child–Pugh Grade								
A	174 (97.8%)	67 (98.5%)	56 (98.2%)	1.000	28 (96.6%)	23 (95.8%)	1.000	0.583
B	4 (2.25%)	1 (1.47%)	1 (1.75%)		1 (3.45%)	1 (4.17%)		
Tumor Size (cm)	3.03 ± 1.09	2.87 ± 1.06	3.29 ± 1.12	0.032*	2.94 ± 1.04	2.98 ± 1.13	0.900	0.564
Tumor Differentiation								
Poor	15 (8.43%)	3 (4.41%)	8 (14.0%)	0.013*	2 (6.90%)	2 (8.33%)	0.162	1.000
Moderate	145 (81.5%)	57 (83.8%)	46 (80.7%)		21 (72.4%)	21 (87.5%)		
Well	14 (7.87%)	8 (11.8%)	1 (1.75%)		5 (17.2%)	0 (0.00%)		
None	4 (2.25%)	0 (0.00%)	2 (3.51%)		1 (3.45%)	1 (4.17%)		
Platelet ($\times10^9$/L)								
≤125	77 (43.3%)	34 (50.0%)	26 (45.6%)	0.757	9 (31.0%)	8 (33.3%)	1.000	0.073
>125	101 (56.7%)	34 (50.0%)	31 (54.4%)		20 (69.0%)	16 (66.7%)		
Prothrombin time (%)								
≤65	5 (2.81%)	2 (2.94%)	2 (3.51%)	1.000	1 (3.45%)	0 (0.00%)	1.000	1.000
>65	173 (97.2%)	66 (97.1%)	55 (96.5%)		28 (96.6%)	24 (100%)		
Albumin (g/L)								
≤38	31 (17.4%)	10 (14.7%)	11 (19.3%)	0.657	5 (17.2%)	5 (20.8%)	1.000	0.907
>38	147 (82.6%)	58 (85.3%)	46 (80.7%)		24 (82.8%)	19 (79.2%)		
Bilirubin (μmol/L)								
≤21	133 (74.7%)	50 (73.5%)	41 (71.9%)	1.000	23 (79.3%)	19 (79.2%)	1.000	0.474
>21	45 (25.3%)	18 (26.5%)	16 (28.1%)		6 (20.7%)	5 (20.8%)		
ALT (IU/L)								
≤42	119 (66.9%)	45 (66.2%)	36 (63.2%)	0.870	21 (72.4%)	17 (70.8%)	1.000	0.472
>42	59 (33.1%)	23 (33.8%)	21 (36.8%)		8 (27.6%)	7 (29.2%)		
AST (IU/L)								
≤42	135 (75.8%)	52 (76.5%)	40 (70.2%)	0.554	23 (79.3%)	20 (83.3%)	1.000	0.378
>42	43 (24.2%)	16 (23.5%)	17 (29.8%)		6 (20.7%)	4 (16.7%)		
AFP (ng/ml)								
≤400	142 (79.8%)	60 (88.2%)	41 (71.9%)	0.038*	24 (82.8%)	17 (70.8%)	0.482	0.750
>400	36 (20.2%)	8 (11.8%)	16 (28.1%)		5 (17.2%)	7 (29.2%)		

Data are present as number (percentage) except otherwise specified. † Data are expressed as median with range. # Between training and test subsets. AFP, alpha fetoprotein; ALT, alanine transaminase; AST, aspartate transaminase; HBV, hepatitis B virus; MVI, microvascular invasion.
**indicates $p < 0.05$.*

diameter was developed and validated. The best performance was observed when combining imaging features from the arterial and hepatobiliary phases of Gd-EOB-DTPA-enhanced MRI with the two clinical risk factors AFP and tumor size. The predictive value was high, with an AUC reaching 0.90 in the training cohort and 0.70 in the test cohort.

Due to the liver specificity of Gd-EOB-DTPA visualizing the hepatocyte function in the so-called HBP, the differences in texture characteristics between liver tumor and the adjacent tissue are improved. Our results showed that the performance of the HBP_model was better than the AP_model, and hepatobiliary phase features are predominant in the AP+HBP_model (5 vs. 9 features), which gave a clue that the imaging features derived from HBP seem to contain more predictive information. This finding is consistent with two previous studies. In their study, Feng et al. extracted imaging features only from the HBP of Gd-EOB-DTPA-enhanced MRI and constructed a radiomics model showing an AUC of 0.83 in the test cohort, higher than ours (0.83 vs. 0.62) (25). Another research also explored radiomics features on solely HBP images and constructed a prediction model with an AUC of 0.8 (26).

In the AP+HBP_model, the majority of imaging features were derived from wavelet-filtered images, which is in line with previous research (11, 16, 27). This finding implies that the wavelet filter is a powerful tool to obtain decomposition and approximation information of the images. Moreover, most of the imaging features that were included in the model can be categorized into first-order statistics (representing the distribution of voxel

TABLE 2 | Clinical risk factors for MVI presence in patients with hepatocellular carcinoma.

Clinical variable	Univariable analysis		Multivariable analysis	
	OR (95% CI)	*p*-value	OR (95% CI)	*p*-value
Age (years)	0.88 (0.56–1.39)	0.58		
Gender				
Male vs. Female	0.64 (0.27–1.53)	0.32		
Etiology				
HBV vs. None/Others	1.83 (0.30–11.37)	0.52		
Cirrhosis				
Present vs. Absent	2.68 (1.04–6.89)	0.04*	2.39 (0.85–6.74)	0.10
Child–Pugh Grade				
B vs. A	1.20 (0.07–19.57)	0.90		
Tumor Size (cm)	1.79 (1.05–3.06)	0.03*	2.06 (1.15–3.70)	0.02*
Tumor Differentiation				
Moderate vs. Well	0.15 (0.02–1.28)	0.08		
Poor vs. Well	3.30 (0.83–13.17)	0.02*	2.47 (0.54–11.17)	0.03*
Platelet ($\times10^9$/L)				
>125 vs. ≤125	0.84 (0.41–1.70)	0.63		
Prothrombin time (%)				
>65 vs. ≤65	1.20 (0.16–8.80)	0.86		
Albumin (g/L)				
>38 vs. ≤38	1.39 (0.54–3.55)	0.50		
Bilirubin (μmol/L)				
>21 vs. ≤21	1.08 (0.49–2.39)	0.84		
ALT (IU/L)				
>42 vs. ≤42	1.14 (0.55–2.38)	0.72		
AST (IU/L)				
>42 vs. ≤42	1.38 (0.62–3.07)	0.43		
AFP (ng/ml)				
>400 vs. ≤400	2.93 (1.15–7.47)	0.02*	3.31 (1.20–9.11)	0.02*

AFP, alpha fetoprotein; ALT, alanine transaminase; AST, aspartate transaminase; CI, confidence interval; HBV, hepatitis B virus; MVI, microvascular invasion, OR, odds ratio.
**indicates p < 0.05.*

intensities), such as maximum, minimum, skewness, and robust mean absolute deviation, indicating that the heterogeneity of the tumor and its surroundings at MRI is associated with MVI presence. This is also in agreement with the abovementioned study by Feng et al., where half of the selected features for modeling belonged to first-order statistics features (25).

As MVI often occurs at the peritumoral area (28, 29), we expanded the tumor margin by 10 mm and extracted the imaging features from intratumoral and peritumoral areas, which we assumed would improve the MVI prediction. The performance of the radiomics models confirmed that assumption. In a similar study, which also constructed models using Gd-EOB-DTPA-

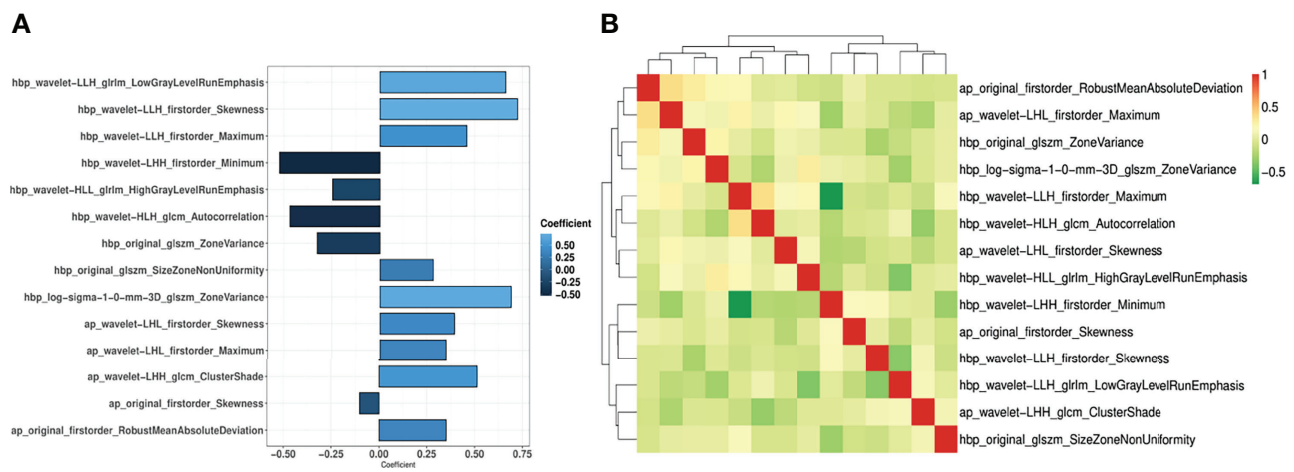


FIGURE 4 | Coefficient of the 14 imaging features **(A)** and the correlation coefficient heatmap **(B)** in the AP+HBP_model.

TABLE 3 | Comparison of the performance of the models in the prediction of MVI presence.

		Clin_model	AP_model	HBP_model	AP+HBP_model	AP+Clin_model	HBP+Clin_model	AP+HBP+Clin_model
Training subset	Cutoff value	−0.06	0.02	−0.56	−0.55	−6.34	−6.16	−6.28
	AUC (95% CI)	0.64 (0.54–0.74)	0.82 (0.75–0.90)	0.87 (0.81–0.93)	0.89 (0.83–0.94)	0.83 (0.77–0.90)	0.87 (0.81–0.93)	0.90 (0.85–0.95)
	Sensitivity	0.51	0.67	0.90	0.91	0.88	0.90	0.91
	Specificity	0.75	0.84	0.68	0.75	0.63	0.68	0.76
	Positive predictive value	0.63	0.78	0.70	0.75	0.67	0.70	0.77
	Negative predictive value	0.65	0.75	0.89	0.91	0.86	0.89	0.91
	Accuracy	0.64	0.76	0.78	0.82	0.74	0.78	0.83
Test subset	AUC (95% CI)	0.55 (0.38–0.71)	0.57 (0.41–0.72)	0.62 (0.47–0.78)	0.66 (0.51–0.81)	0.56 (0.40–0.72)	0.62 (0.47–0.78)	0.70 (0.55–0.84)
	Sensitivity	0.38	0.54	0.71	0.46	0.21	0.63	0.60
	Specificity	0.83	0.62	0.59	0.86	0.97	0.69	0.79
	Positive predictive value	0.64	0.54	0.59	0.73	0.83	0.63	0.71
	Negative predictive value	0.62	0.62	0.71	0.66	0.60	0.69	0.69
	Accuracy	0.62	0.59	0.64	0.68	0.62	0.66	0.70

AUC, area under the receiver operating characteristics curve; CI, confidence interval; MVI, microvascular invasion.

enhanced MRI for patients with HCC ≤ 5 cm, the tumor margin was dilated in different diameters, i.e., 5 mm and 10 mm, and also shrunk by 50% (21). The models in that research using features extracted from a combination of the tumor and the 10-mm dilated region yielded an AUC ranging from 0.79 to 0.76 for HBP by two classifiers, random forest and logistic regression, which is a little higher than our model.

Previous studies have attempted to exploit preoperative clinical variables and laboratory tests to predict MVI. Tumor characteristics such as tumor size and tumor number are well-established risk factors for MVI incidence (17, 19). One study with 245 HCC patients undergoing liver transplantation showed that the MVI incidence was 25% in tumors <2 cm, 31% in 2–4 cm tumors, and 50% in tumors >4 cm (30). Another study conducted by Kim et al. demonstrated that the incidence of MVI doubles when there are two or more tumors compared to when there is a solitary HCC (31). Furthermore, the tumor biomarker, AFP, has also been recognized as a reliable predictor for MVI (19, 32, 33). Our Clin_model detected tumor size and AFP as independent risk factors for the prediction model. However, the clinical model using these two risk factors only reached a fair AUC of 0.55 in the test subset. As one of the strategies to improve the performance of a model is to combine variables from different aspects (34), we integrated the clinical risk factors into the AP+HBP_model, improving the AUC to 0.70.

There are some limitations to be acknowledged when interpreting the results of the current study. To begin with, our study was limited by its retrospective nature and sample size. Patient selection bias may thereby have been introduced. Future prospective research should include a larger number of participants to confirm our findings. Moreover, external data from other medical centers are also needed to prove the generalization of our model. Second, although 10-fold cross-validation was adopted during modeling, overfitting might still exist, as seen in the sharp drop of the AUC value in the test subset. Another interpretation for the lower performance in the test subset may be the limited sample size of the test subset, only 53 cases, which makes it sensitive to the performance test. Third, as the current study focused on solitary HCC with a diameter ≤ 5 cm, the generalization of the model needs to be confirmed among HCC patients with no limit for tumor number and size. This should be of special interest when evaluating patients just outside the current transplantation criteria. Furthermore, there are incidence differences among populations due to cirrhosis, viral hepatitis, and nonalcoholic steatohepatitis. This makes it important to validate the model on different cohorts. Fourth, the optimal dilation of the tumor needs to be evaluated as we just dilated the tumor VOIs to 10 mm of the margin as most previously published studies did (20, 21). Future research can be designed to compare different dilations of the tumor diameter when predicting MVI. Fifth, we applied an ICC threshold of 0.75, but the impact of different thresholds on model performance requires further research. Finally, we did not incorporate semantic imaging features, such as the tumor margins or arterial peri-tumoral enhancement, into modeling as we thought those features are more subjective compared with radiomics features. We also did not incorporate images from the portal venous phase due to the same reason as its contrast ratio was inferior to arterial phase. An attempt

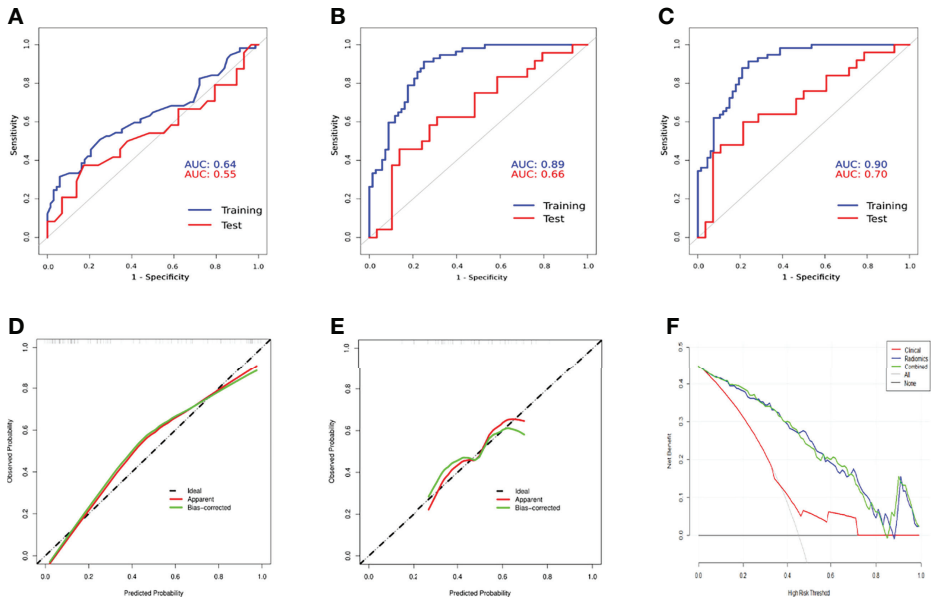


FIGURE 5 | Comparison of receiver operation characteristics curves of the Clin_model **(A)**, AP+HBP_model **(B)**, and AP+HBP+Clin_model **(C)** in training and test subsets. Calibration curves of the AP+HBP+Clin_model for training **(D)** and test **(E)** subsets are shown in the lower left and middle panels. The "apparent" curve (red) represents the prediction model and the bias-corrected curve (green) describes the prediction model calibrated by 1,000 bootstrap samples. The black dashed diagonal line indicates an ideal situation in which the prediction probability is equal to the observed probability. In the lower right panel **(F)**, decision curve analysis for the Clin_model, AP+HBP_model, and AP+HBP+Clin_model is shown. The black line represents the net benefit of assuming that none of the patients have microvascular invasion (MVI), whereas the gray curve represents the net benefit of assuming that all patients have MVI.

to build a more objective model using a deep learning approach (without the radiologist's tumor segmentation) is ongoing in our team.

CONCLUSIONS

Our radiomics-based model combining imaging features from the arterial and hepatobiliary phases of Gd-EOB-DTPA-enhanced MRI and clinical risk factors provides an effective and reliable tool for the preoperative prediction of MVI in patients with HCC ≤ 5 cm.

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 Southwest Hospital of Army Medical University, Chongqing, China (No.2017KY50). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

Writing—original draft preparation: CQ and QW. Writing—review and editing: ES and TBB. Data collection: CQ and CL. Conceptualization: QW, KM, and LZ. Methodology: CQ, QW, QX, and PC. Pathology report interpretation: XY. Formal analysis and investigation: CQ and QW. Funding acquisition: QW and KM. Supervision: TBB, ES, and KM. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the National Natural Science Foundation of China (Nos. 82073346 and 81672857), and Famous Teachers section of the Chongqing Talents Program (4246ZP112). QW receives a scholarship from the China Scholarship Council (CSC) (No. 201907930009).

ACKNOWLEDGMENTS

QW would like to thank Dr. Renyuan Li for his excellent online teaching videos, which introduced QW into the magic "Radiomics World".

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.831795/full#supplementary-material

REFERENCES

1. Villanueva A. Hepatocellular Carcinoma. *N Engl J Med* (2019) 380:1450–62. doi: 10.1056/NEJMra1713263
2. Vogel A, Cervantes A, Chau I, Daniele B, Llovet JM, Meyer T, et al. Hepatocellular Carcinoma: ESMO Clinical Practice Guidelines for Diagnosis, Treatment and Follow-Up. *Ann Oncol* (2018) 29:iv238–55. doi: 10.1093/annonc/mdy308
3. Kim J, Kang W, Sinn DH, Gwak GY, Paik YH, Choi MS, et al. Substantial Risk of Recurrence Even After 5 Recurrence-Free Years in Early-Stage Hepatocellular Carcinoma Patients. *Clin Mol Hepatol* (2020) 26:516–28. doi: 10.3350/cmh.2020.0016
4. Lim KC, Chow PK, Allen JC, Chia GS, Lim M, Cheow PC, et al. Microvascular Invasion is a Better Predictor of Tumor Recurrence and Overall Survival Following Surgical Resection for Hepatocellular Carcinoma Compared to the Milan Criteria. *Ann Surg* (2011) 254:108–13. doi: 10.1097/SLA.0b013e31821ad884
5. Lei Z, Li J, Wu D, Xia Y, Wang Q, Si A, et al. Nomogram for Preoperative Estimation of Microvascular Invasion Risk in Hepatitis B Virus-Related Hepatocellular Carcinoma Within the Milan Criteria. *JAMA Surg* (2016) 151:356–63. doi: 10.1001/jamasurg.2015.4257
6. Sun SW, Liu QP, Xu X, Zhu FP, Zhang YD. Direct Comparison of Four Presurgical Stratifying Schemes for Prediction of Microvascular Invasion in Hepatocellular Carcinoma by Gadoxetic Acid-Enhanced MRI. *J Magn Reson Imaging* (2020) 52:433–47. doi: 10.1159/00048854010.1002/jmri.27043
7. Hong SB, Choi SH, Kim SY, Shim JH, Lee SS, Byun JH, et al. MRI Features for Predicting Microvascular Invasion of Hepatocellular Carcinoma: A Systematic Review and Meta-Analysis. *Liv Cancer* (2021) 10:94–106. doi: 10.1159/000513704
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. 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
11. Xu X, Zhang HL, Liu QP, Sun SW, Zhang J, Zhu FP, et al. Radiomic Analysis of Contrast-Enhanced CT Predicts Microvascular Invasion and Outcome in Hepatocellular Carcinoma. *J Hepatol* (2019) 70:1133–44. doi: 10.1016/j.ebiom.2019.10.05710.1016/j.jhep.2019.02.023
12. 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
13. 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. *Liv Cancer* (2019) 8:373–86. doi: 10.1007/s00261-018-1768-910.1159/000494099
14. Yoon JH, Joo I, Park SJ, Jeon SK, Schmidt B, Martin S, et al. Preoperative Prediction Power of Imaging Methods for Microvascular Invasion in Hepatocellular Carcinoma: A Systemic Review and Meta-Analysis. *Cancer Imaging* (2020) 10:2020.00887. doi: 10.1186/s40644-020-00338-710.3389/fonc.2020.00887
15. Zhang Q, Arefan D, Zhao X, Wu S, Zhang X, Ruan S, et al. Contrast-Enhanced CT Radiomics for Preoperative Evaluation of Microvascular Invasion in Hepatocellular Carcinoma: A Two-Center Study. *J Dig Imaging* (2020) 10:e111. doi: 10.1007/s10278-020-00353-x10.1002/ctm2.111
16. 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
17. Pawlik TM, Delman KA, Vauthey JN, Nagorney DM, Ng IO, Ikai I, et al. Tumor Size Predicts Vascular Invasion and Histologic Grade: Implications for Selection of Surgical Treatment for Hepatocellular Carcinoma. *Liv Transpl* (2005) 11:1086–92. doi: 10.1002/lt.20472
18. Van Beers BE, Pastor CM, Hussain HK. Primovist, Eovist: What to Expect? *J Hepatol* (2012) 57:421–9. doi: 10.1016/j.jhep.2012.01.031
19. Zhang X, Li J, Shen F, Lau WY. Significance of Presence of Microvascular Invasion in Specimens Obtained After Surgical Treatment of Hepatocellular Carcinoma. *J Gastroenterol Hepatol* (2018) 33:347–54. doi: 10.1111/jgh.13843
20. Wang Q, Li C, Zhang J, Hu X, Fan Y, Ma K, et al. Radiomics Models for Predicting Microvascular Invasion in Hepatocellular Carcinoma: A Systematic Review and Radiomics Quality Score Assessment. *Cancers (Basel)* (2021) 13:5864. doi: 10.3390/cancers13225864
21. 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
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. 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
24. Song Y, Zhang J, Zhang YD, Hou Y, Yan X, Wang Y, et al. FeAture Explorer (FAE): A Tool for Developing and Comparing Radiomics Models. *PloS One* (2020) 15:e0237587. doi: 10.1371/journal.pone.0237587
25. 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
26. Zhang S, Xu G, Duan C, Zhou X. 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
27. Zhou Y, Zhou G, Zhang J, Xu C, Wang X, Xu P. Radiomics Signature on Dynamic Contrast-Enhanced MR Images: A Potential Imaging Biomarker for Prediction of Microvascular Invasion in Mass-Forming Intrahepatic Cholangiocarcinoma. *Eur Radiol* (2021) 31:6846–55. doi: 10.1007/s00330-021-07793-1
28. Roayaie S, Blume IN, Thung SN, Guido M, Fiel MI, Hiotis S, et al. A System of Classifying Microvascular Invasion to Predict Outcome After Resection in Patients With Hepatocellular Carcinoma. *Gastroenterology* (2009) 137:850–5. doi: 10.1053/j.gastro.2009.06.003
29. 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 Imaging* (2021) 54:526–36. doi: 10.21037/qims-20-21810.1002/jmri.27575
30. Esnaola NF, Lauwers GY, Mirza NQ, Nagorney DM, Doherty D, Ikai I, et al. Predictors of Microvascular Invasion in Patients With Hepatocellular Carcinoma Who are Candidates for Orthotopic Liver Transplantation.

J Gastrointest Surg (2002) 6:224–32; discussion 32. doi: 10.1016/s1091-255x(01)00015-4

31. Kim BK, Han KH, Park YN, Park MS, Kim KS, Choi JS, et al. Prediction of Microvascular Invasion Before Curative Resection of Hepatocellular Carcinoma. *J Surg Oncol* (2008) 97:246–52. doi: 10.1002/jso.20953
32. Mchugh PP, Gilbert J, Vera S, Koch A, Ranjan D, Gedaly R. Alpha-Fetoprotein and Tumour Size are Associated With Microvascular Invasion in Explanted Livers of Patients Undergoing Transplantation With Hepatocellular Carcinoma. *HPB (Oxf)* (2010) 12:56–61. doi: 10.1111/j.1477-2574.2009.00128.x
33. Siegel AB, Wang S, Jacobson JS, Hershman DL, Lim EA, Yu J, et al. Obesity and Microvascular Invasion in Hepatocellular Carcinoma. *Cancer Invest* (2010) 28:1063–9. doi: 10.3109/07357907.2010.483500
34. Miranda Magalhaes Santos JM, Clemente Oliveira B, Araujo-Filho J, Assuncao- ANJr, De, Machado F, Carlos Tavares Rocha C, et al. State-Of-the-Art in Radiomics of Hepatocellular Carcinoma: A Review of Basic Principles, Applications, and Limitations. *Abdom Radiol (NY)* (2020) 45:342–53. doi: 10.3881/j.issn.1000-503X.1262010.1007/s00261-019-02299-3

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.

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