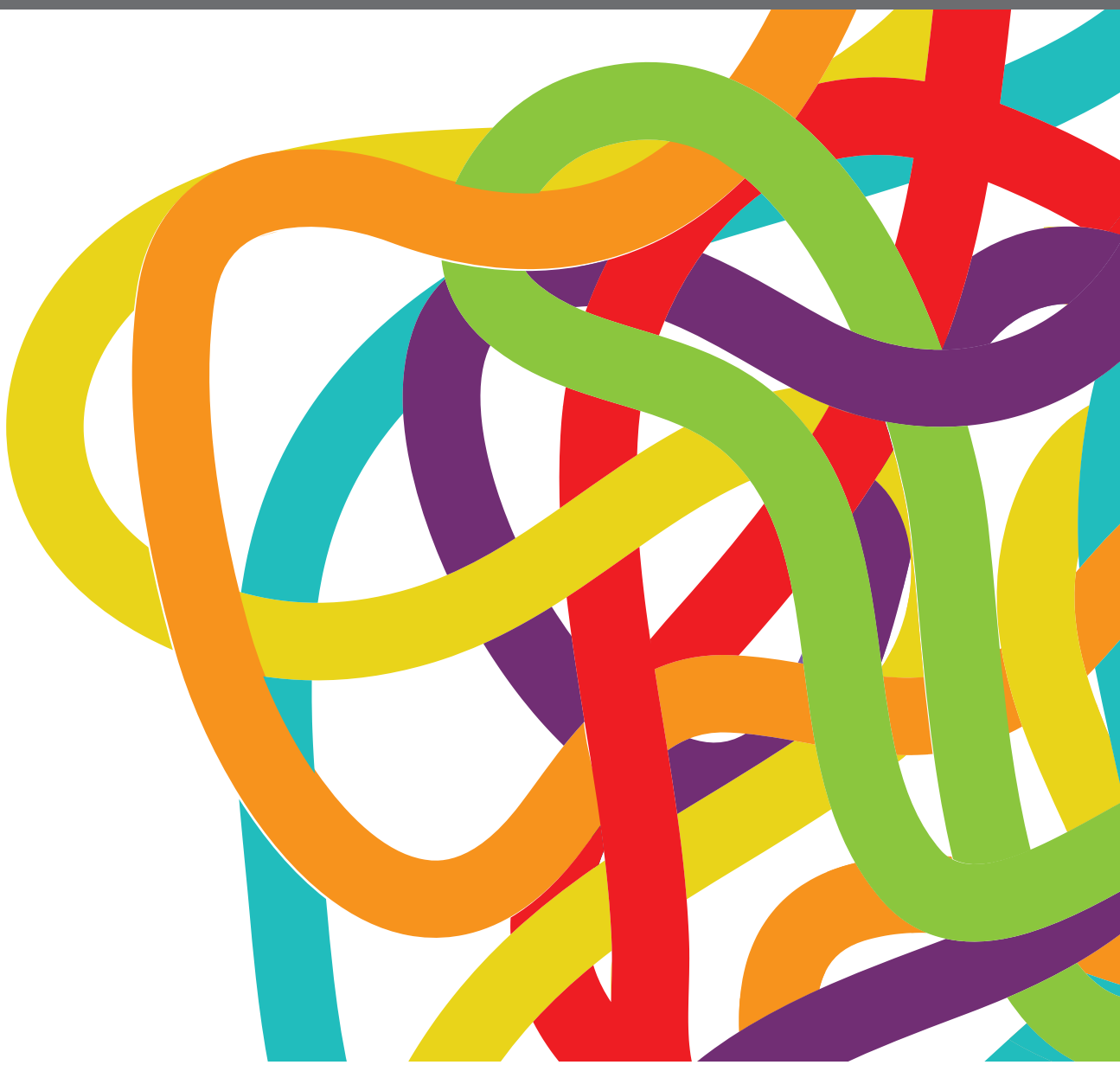
MACHINE LEARNING IN RADIATION ONCOLOGY

EDITED BY: Wei Zhao, Ye Zhang, Jia Wu, Xiaomeng Li and Yuming Jiang
PUBLISHED IN: Frontiers in Oncology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-83250-439-0
DOI 10.3389/978-2-83250-439-0

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

MACHINE LEARNING IN RADIATION ONCOLOGY

Topic Editors:
Wei Zhao, Beihang University, China
Ye Zhang, Center for Proton Therapy, Paul Scherrer Institute (PSI), Switzerland
Jia Wu, University of Texas MD Anderson Cancer Center, United States
Xiaomeng Li, Hong Kong University of Science and Technology, Hong Kong, SAR China
Yuming Jiang, Stanford University, United States

Citation: Zhao, W., Zhang, Y., Wu, J., Li, X., Jiang, Y., eds. (2022). Machine Learning in Radiation Oncology. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-83250-439-0

Table of Contents

05 ***Editorial: Machine Learning in Radiation Oncology***
Wei Zhao, Ye Zhang, Jia Wu, Xiaomeng Li and Yuming Jiang

08 ***Predicting the Local Response of Metastatic Brain Tumor to Gamma Knife Radiosurgery by Radiomics With a Machine Learning Method***
Daisuke Kawahara, Xueyan Tang, Chung K. Lee, Yasushi Nagata and Yoichi Watanabe

16 ***A Multi-Compartment Model of Glioma Response to Fractionated Radiation Therapy Parameterized* via *Time-Resolved Microscopy Data***
Junyan Liu, David A. Hormuth II, Jianchen Yang and Thomas E. Yankeelov

30 ***Classification of Gliomas and Germinomas of the Basal Ganglia by Transfer Learning***
Ningrong Ye, Qi Yang, Ziyan Chen, Chubei Teng, Peikun Liu, Xi Liu, Yi Xiong, Xuelei Lin, Shouwei Li and Xuejun Li

39 ***Accuracy Improvement Method Based on Characteristic Database Classification for IMRT Dose Prediction in Cervical Cancer: Scientifically Training Data Selection***
Yiru Peng, Yaoying Liu, Zhaocai Chen, Gaolong Zhang, Changsheng Ma, Shouping Xu and Yong Yin

47 ***Geometric and Dosimetric Evaluation of the Automatic Delineation of Organs at Risk (OARs) in Non-Small-Cell Lung Cancer Radiotherapy Based on a Modified DenseNet Deep Learning Network***
Fuli Zhang, Qiusheng Wang, Anning Yang, Na Lu, Huayong Jiang, Diandian Chen, Yanjun Yu and Yadi Wang

55 ***Evaluation Exploration of Atlas-Based and Deep Learning-Based Automatic Contouring for Nasopharyngeal Carcinoma***
Jinyuan Wang, Zhaocai Chen, Cungeng Yang, Baolin Qu, Lin Ma, Wenjun Fan, Qichao Zhou, Qingzeng Zheng and Shouping Xu

63 ***Transfer Learning-Based Autosegmentation of Primary Tumor Volumes of Glioblastomas Using Preoperative MRI for Radiotherapy Treatment***
Suqing Tian, Cuiying Wang, Ruiping Zhang, Zhuojie Dai, Lecheng Jia, Wei Zhang, Junjie Wang and Yinglong Liu

72 ***Multiphasic CT-Based Radiomics Analysis for the Differentiation of Benign and Malignant Parotid Tumors***
Qiang Yu, Anran Wang, Jinming Gu, Quanjiang Li, Youquan Ning, Juan Peng, Fajin Lv and Xiaodi Zhang

84 ***A Transfer Learning Framework for Deep Learning-Based CT-to-Perfusion Mapping on Lung Cancer Patients***
Ge Ren, Bing Li, Sai-kit Lam, Haonan Xiao, Yu-Hua Huang, Andy Lai-yin Cheung, Yufei Lu, Ronghu Mao, Hong Ge, Feng-Ming (Spring) Kong, Wai-yin Ho and Jing Cai

95 ***Predicting Adverse Radiation Effects in Brain Tumors After Stereotactic Radiotherapy With Deep Learning and Handcrafted Radiomics***
Simon A. Keek, Manon Beuque, Sergey Primakov, Henry C. Woodruff, Avishek Chatterjee, Janita E. van Timmeren, Martin Vallières, Lizza E. L. Hendriks, Johannes Kraft, Nicolaus Andratschke, Steve E. Braunstein, Olivier Morin and Philippe Lambin

108 ***CT-Only Radiotherapy: An Exploratory Study for Automatic Dose Prediction on Rectal Cancer Patients* Via *Deep Adversarial Network***
Jiaqi Cui, Zhengyang Jiao, Zhigong Wei, Xiaolin Hu, Yan Wang, Jianghong Xiao and Xingchen Peng

118 ***Review of Radiomics- and Dosiomics-based Predicting Models for Rectal Cancer***
Yun Qin, Li-Hua Zhu, Wei Zhao, Jun-Jie Wang and Hao Wang

130 ***Automated Segmentation of Colorectal Liver Metastasis and Liver Ablation on Contrast-enhanced CT images***
Brian M. Anderson, Bastien Rigaud, Yuan-Mao Lin, A. Kyle Jones, HynSeon Christine Kang, Bruno C. Odisio and Kristy K. Brock

frontiers | Frontiers in Oncology

TYPE Editorial
PUBLISHED 26 September 2022
DOI 10.3389/fonc.2022.1032858

OPEN ACCESS

EDITED AND REVIEWED BY
Timothy James Kinsella,
Brown University, United States

*CORRESPONDENCE
Wei Zhao
zhaow2015@gmail.com

SPECIALTY SECTION
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

RECEIVED 31 August 2022
ACCEPTED 01 September 2022
PUBLISHED 26 September 2022

CITATION
Zhao W, Zhang Y, Wu J, Li X and Jiang Y (2022) Editorial: Machine learning in radiation oncology. *Front. Oncol.* 12:1032858. doi: 10.3389/fonc.2022.1032858

Editorial: Machine learning in radiation oncology

Wei Zhao[1,2]*, Ye Zhang[3], Jia Wu[4], Xiaomeng Li[5] and Yuming Jiang[6]

[1]Department of Physics, Beihang University, Beijing, China, [2]Beihang Hangzhou Innovation Institute, Yuhang Xixi Octagon City, Hangzhou, China, [3]Center for Proton Therapy, Paul Scherrer Institute (PSI), Villigen, Switzerland, [4]Department of Imaging Physics, University of Texas MD Anderson Cancer Center, Houston, TX, United States, [5]Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong SAR, China, [6]Department of Radiation Oncology, Stanford University, Stanford, CA, United States

KEYWORDS

machine learning, deep learning, radiation oncology, radiomics, radiotherapy, image segmentation, artificial intelligence

Editorial on the Research Topic
Machine learning in radiation oncology

Machine learning (ML) excels in learning complex relationships and incorporating existing prior knowledge into an inference model. The model is able to recognize complex patterns in medical data using either human-engineered features or deep neural network representations. Given the nature of heavy data involvement of ML and a large amount of existing routine data, such as computed tomography images, organs and targets contours, treatment plans, and prognosis records, radiation oncology is one of the best fields for ML algorithms. ML has the potential to benefit the whole workflow of radiotherapy: from patient modeling, image segmentation, treatment planning, patient setup, and prognosis analysis. Indeed, the Food and Drug Administration has cleared numerous ML-based technologies with regard to various radiation oncology aspects and there are having transformative ML-based applications in clinical practice.

The remarkable progress in ML-based radiation oncology has driven the special issue. Specifically, it consists of 13 papers that cover various aspects of radiation oncology. Liu et al. develop an experimental-computational approach to predict the changes in cell number over time in response to fractionated radiation. By developing a mathematical model using time-resolved microscopy, together with phase contrast images and Cytotox Red images, this study establishes a framework to quantitative characterize and predict the dynamic radiobiological response of 9L and C6 glioma cells to fractionated radiotherapy for the first time.

Because it is challenging to differentiate Germinomas of the basal ganglia from gliomas by using routine MRI images. Ye et al. use transfer learning to fine-tune a pre-trained ResNet18 to identify germinomas from gliomas with manually segmented MRI images acquired using T1C sequence. A mean area-under-curve (AUC) of 0.88 is achieved, showing the accuracy of the model. This study has the potential to provide

valuable pretreatment information. Ren et al. demonstrate the feasibility of generating lung perfusion images based on CT images for lung cancer patients. To this end, CT images of non-cancer patients are employed to synthesize perfusion images using a convolutional neural network (CNN), and the features learned from this process are then adapted specifically for lung cancer patients using a transfer learning framework. The synthetic perfusion images are compared with SPECT perfusion images which are generally considered to be the gold standard and show strong voxel-wise correlation and function-wise similarity. This study, therefore, suggests deep learning (DL) method is able to provide regional-based functional information for lung avoidance radiation therapy.

Typical treatment planning procedure in radiotherapy requires accurate target contouring, organs-at-risk (OARs) delineation, as well as iterative dose optimization. This procedure involves various steps and synergic work from different radiation therapy teams, such as radiation oncologists, medical physicists, and dosimetrists, and it is usually time-consuming and also suffered from inconsistency between team members. To reduce the workload of radiation oncologists for OARs delineation, Zhang et al. use a modified DenseNet to automatically delineate OARs in thoracic CT images. By subjecting the delineated OARs to treatment planning, the dosimetric parameters of the optimized dose show no statistically significant difference between autosegmentation and manual segmentation. In addition to OARs delineation, ML algorithms are also subjected to tumor volume segmentation. Tian et al. develop a semisupervised transfer learning method to segment glioblastoma on pretreatment MRI and the accuracy of the independent testing data has shown to be sufficient for radiotherapy treatment.

The DL-based autosegmentation can not only be employed for radiotherapy, but also for ablation therapy. Anderson et al. demonstrate fully convolutional neural networks provide rapid and accurate identification and segmentation of colorectal liver metastasis and ablation zones on contrast-enhanced CT scans, with positive physician reviews. The DL-based autosegmentation is also compared to classical atlas-based autosegmentation. Wang et al. evaluate the difference between atlas-based and DL-based autosegmentation of OARs for nasopharyngeal carcinoma patients. The study shows a well-trained DL segmentation model significantly outperforms atlas-based segmentation for nasopharyngeal carcinoma, especially for the OARs with small volumes.

A study further shows segmentation and dose optimization can be performed together. Cui et al. propose to accelerate the treatment planning procedure by introducing a DL-based CT-only dose prediction framework. This framework is able to simultaneously generate tumor masks and optimized doses by using a multi-task loss learning strategy. It can therefore significantly shorten the treatment planning procedure.

DL model performance relies on the distribution of the training data. Peng et al. show a homogeneous training dataset results in better performance than a non-homogeneous for dose prediction. Therefore, a patient-specific model trained using a consistent dataset is recommended for dose prediction. In case a homogeneous dataset is not available, additional beam information (i.e., beam mask) can improve the model performance to yield acceptable results.

ML algorithms are also popular in treatment outcome prediction. Kawahara et al. predict the local response of metastatic brain tumor to Gamma knife radiosurgery (GKRS) by using a neural network model with input radiomics features. The accuracy and sensitivity of the predictive model have been shown to outperform the traditional method. This study is able to help physicians to gain the most desirable treatment outcome for patients treated with GKRS. With regards to the adverse radiation effects (ARE) of stereotactic radiotherapy (SRT) in the treatment of brain metastases (BM) in solid cancers, Keek et al. train models based on radiomics features, DL features, and patient characteristics to predict ARE risk before SRT. This is a great decision-making support tool to allow physicians to opt for the optimal treatment solution for patients with BM.

Yu et al. use ML models based on multiphasic CT radiomics features to differentiate benign and malignant parotid tumors. Qin et al. review the applications of radiomics in radiation oncology with a focus on locally advanced rectal cancer. This paper also summarizes the research of dosiomics which is an emerging Research Topic that involves dose-based radiomics analysis in radiation treatment.

Within the spectrum of ML applications in radiation oncology, image segmentation plays a central role. It has attracted the most research and translational activities and resulted in five papers in this Research Topic. While DL-based autosegmentation is able to reduce the workload of physicians and improve work efficiency, its accuracy still needs further justification. Radiomics-based applications including outcome prediction and ARE are also under extensive investigation. DL features used together with conventional radiomics features is the new frontier and usually yield better model prediction. As a core section of radiotherapy, dose prediction is a hot topic and results in two papers. Overall, the progress of ML in radiation oncology has been well reflected by this special collection, with major progress in diagnosis, image segmentation, and radiomics well covered. We anticipate ML applications will continue to play an important role in radiation oncology and translational work will be performed to improve clinical efficiency.

Author contributions

WZ and XL performed the analysis of the collection and wrote the manuscript. YZ, JW, and YJ helped with the analysis and study design. YZ and JW devised the project's aim and supervised the progression of the project. All authors contributed to the article and approved the submitted version.

Funding

This work was supported in part by the National Natural Science Foundation of China (No. 12175012).

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

frontiers in Oncology

ORIGINAL RESEARCH
published: 11 January 2021
doi: 10.3389/fonc.2020.569461

Predicting the Local Response of Metastatic Brain Tumor to Gamma Knife Radiosurgery by Radiomics With a Machine Learning Method

Daisuke Kawahara[1], *Xueyan Tang*[2], *Chung K. Lee*[2], *Yasushi Nagata*[1] *and Yoichi Watanabe*[2*]

[1] *Department of Radiation Oncology, Institute of Biomedical & Health Sciences, Hiroshima University, Hiroshima, Japan,*
[2] *Department of Radiation Oncology, University of Minnesota-Twin Cities, Minneapolis, MN, United States*

OPEN ACCESS

Edited by:
John Varlotto,
University of Massachusetts Medical School, United States

Reviewed by:
Yuenan Wang,
Peking University Shenzhen Hospital, China
Georges Sinclair,
James Cook University Hospital, United Kingdom

****Correspondence:***
Yoichi Watanabe
watan016@umn.edu

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: *04 June 2020*
Accepted: *25 November 2020*
Published: *11 January 2021*

Citation:
Kawahara D, Tang X, Lee CK, Nagata Y and Watanabe Y (2021) Predicting the Local Response of Metastatic Brain Tumor to Gamma Knife Radiosurgery by Radiomics With a Machine Learning Method. Front. Oncol. 10:569461. doi: 10.3389/fonc.2020.569461

Purpose: The current study proposed a model to predict the response of brain metastases (BMs) treated by Gamma knife radiosurgery (GKRS) using a machine learning (ML) method with radiomics features. The model can be used as a decision tool by clinicians for the most desirable treatment outcome.

Methods and Material: Using MR image data taken by a FLASH (3D fast, low-angle shot) scanning protocol with gadolinium (Gd) contrast-enhanced T1-weighting, the local response (LR) of 157 metastatic brain tumors was categorized into two groups (Group I: responder and Group II: non-responder). We performed a radiomics analysis of those tumors, resulting in more than 700 features. To build a machine learning model, first, we used the least absolute shrinkage and selection operator (LASSO) regression to reduce the number of radiomics features to the minimum number of features useful for the prediction. Then, a prediction model was constructed by using a neural network (NN) classifier with 10 hidden layers and rectified linear unit activation. The training model was evaluated with five-fold cross-validation. For the final evaluation, the NN model was applied to a set of data not used for model creation. The accuracy and sensitivity and the area under the receiver operating characteristic curve (AUC) of the prediction model of LR were analyzed. The performance of the ML model was compared with a visual evaluation method, for which the LR of tumors was predicted by examining the image enhancement pattern of the tumor on MR images.

Results: By the LASSO analysis of the training data, we found seven radiomics features useful for the classification. The accuracy and sensitivity of the visual evaluation method were 44 and 54%. On the other hand, the accuracy and sensitivity of the proposed NN model were 78 and 87%, and the AUC was 0.87.

Conclusions: The proposed NN model using the radiomics features can help physicians to gain a more realistic expectation of the treatment outcome than the traditional method.

Keywords: radiomics, machine learning, brain metastases, gamma knife, radiosurgery, local control

INTRODUCTION

Approximately 5 to 40% of cancer patients are diagnosed with a metastatic brain tumor during their treatment. Furthermore, patients have brain metastases (BMs) ten times more often than primary malignant tumors of the brain (1, 2). Consequently, BM is the most common brain tumor treated by radiation therapy. Whole-brain radiation therapy (WBRT) and stereotactic radiosurgery (SRS) are regularly offered to manage BMs. The techniques are effective with improved local control of tumors and more prolonged survival of patients (3). The RTOG-9508 study compared the treatment responses of WBRT alone, SRS alone, and WBRT plus SRS for the BMs (4). For the WBRT alone or WBRT plus SRS, the total prescribed dose of WBRT was 37.5 Gy with 2 to 5 Gy per fraction. For the SRS treatment, the prescribed dose was assigned from an earlier dose-escalation RTOG radiosurgery trial (90–05) (5). The mean survival time did not differ much among the three techniques. The local control rate at three months after WBRT plus SRS or WBRT alone ranged from 71 to 82%, indicating about 20 to 30% failure rate. Hence, a predictive capability of the radiation therapy outcome of BMs may provide a decision tool to clinicians for the effective management of patient care with the most desirable treatment outcome. If the local failure is predicted for radiotherapy, the treatment plan can be modified to improve the local control by, for example, increasing the dose.

There are several prognostic tools or prognostic indices, specially developed for the radiation therapy of BMs such as the RTOG Recursive Partitioning Analysis (RPA), the Score Index for Radiosurgery (SIR), the Basic Score for Brain Metastases (BSBM), and the Graded Prognostic Assessment (GPA) (6). These indices are proven to have clinical value for predicting the treatment outcome. The addition of more detailed clinical information to the pretreatment characteristics used by the existing prognostic indices might improve the predictive performance. Such new information includes the biological data (*i.e.*, biological markers and genomics) specific to the patient (7) and the quantitative imaging data obtained by radiomics (8–10).

Radiomics analyzes the medical image quantitatively to explore features unique to a patient (11). It has been used for classifying patients and evaluating their risk to customize oncological treatments (12, 13). Some researchers used radiomics to find the correlation between radiomics signatures and radiation treatment outcome (14–16). Zhou et al. tried to predict survival after chemotherapy of glioblastoma patients using several imaging features based on MR image (17). Ryu et al. performed a prognostic prediction using features obtained from functional images (18). Other studies have combined radiomics with genomics to associate radiomics features with gene mutations that are clinically proven to predict therapy response (19). A recent study reported that radiomics features could potentially be used as surrogate biomarkers for predicting tumor prognosis following Gamma Knife radiosurgery (GKRS) (20).

Goodman et al. categorized brain tumor images into three groups: homogeneous, heterogeneous, or ring-enhancing (21). They found that these enhancement patterns are significant prognostic factors in the response of brain metastases after radiosurgery. A drawback of their approach is the subjective nature of the classification technique. Visual classification into one of the three patterns is often neither possible nor accurate because real images do not display clear ring-like features or completely uniform pixel colors throughout the tumor.

In the current study, therefore, we proposed the application of radiomics and a machine learning (ML) technique to create a more reliable and accurate method than the decision with the visual evaluation for predicting the treatment outcome, in particular, the local response of the tumor to radiation therapy. Primarily, we built a model to predict the response of metastatic tumors treated with GKRS.

MATERIALS AND METHODS

Patients

Previously, we analyzed the treatment outcome of 88 patients with either renal cell or melanoma cancer as the primary disease, who underwent GKRS at the University of Minnesota from 2005 to 2012 for their BMs (22). For the current study, we selected a subset of the patients, 45 melanoma patients with a total of 115 tumors, for model building. Furthermore, we obtained the new data of nine melanoma patients with a total of 42 tumors from the database of GKRS patients treated from 2013 to 2017 for the final evaluation of the model. The characteristics of the patients and their tumors are presented in **Table 1**.

Image Acquisition

All patients were scanned with a 1.5T MRI (Siemens Syngo MR) scanner. The total scanning time was about 15 min for the whole brain scan. We used the Siemens 12 channel head matrix coils. The scanning protocol was a FLASH (3D fast, low-angle shot) with gadolinium (Gd)-contrast enhanced T1-weighting. The scan parameters are shown in **Table 2**.

Treatment

We treated patients with the Leksell Gamma Knife Model 4C (Elekta AB, Stockholm, Sweden). The prescription dose was decided based on tumor size according to the RTOG 90-05 trial protocol (5). The prescription isodose level varied from 40 to

TABLE 1 | Patient characteristics.

		Model building dataset	Model evaluation dataset
Number of patients		45	9
Number of tumors		115	42
Age (years)	Median	61.5	59.5
	[Range]	[32–86]	[42–71]
Gender	Male	23 (53%)	5 (56%)
	Female	20 (47%)	4 (44%)
Local Response (LR)	Group I: CR + PR	83	35
	Group II: SD + PD	32	7
Tumor volume (cc)	<4.2	102 (89%)	40 (95%)
	≥4.2 to ≤14.1	12 (10%)	2 (5%)
	>14.1	1 (1%)	0 (0%)

TABLE 2 | MRI scan parameters of the Fast Low Angle Shot (FLASH) pulse sequence.

Parameter	Values
TE	4.76 ms
TR	9.4 ms
Echo train length	1
Number of acquisitions	1
Bandwidth	260 kHz
Flip angle	25°
FOV	256 mm × 256 mm
Voxel size	0.5 mm × 0.5 mm × 1.0 mm
Slice thickness	1 mm

80%, with a medium of 50%. The prescribed dose to the gross tumor volume (GTV) was 24 Gy for the tumor volume <4.2 cc, 18 Gy for the tumor volume ≥4.2 cc to ≤14.1 cc, 15 Gy for the tumor volume >14.1 cc. **Table 1** shows the number of tumors in these three-volume ranges.

Follow-Up

Patients after GKRS were followed at 3-month intervals with MRI performed at each visit. The time from the first SRS to the last follow-up imaging study or death was defined as patient follow-up duration.

Treatment Response Evaluation

To evaluate the local response (LR) of the tumors to the treatment, we measured the maximum lengths of a tumor in three orthogonal directions using pretreatment and follow-up MRI images. Tumor volumes were calculated with the ellipsoid volume formula. The LR status of treatment was determined by using the latest available follow-up imaging study at the time of the data collection. The medium follow-up length was 7.6 months. The status of each tumor was evaluated based on modified RECIST criteria (23). A tumor was defined as progressive disease (PD) if there was a relative increase in tumor volume on follow-up MRI by greater than 20% compared to pretreatment MRI. Lesions in which volume increased less than 20% or decreased less than 30% of pretreatment were considered a stable disease (SD). The tumor, whose size fell more than 30%, but it was still visible on the follow-up MRI, was categorized as a partial response (PR). Any lesion which disappeared on the MRI was considered as complete repose (CR). We accepted only conservative management of cancer during the follow-up period to be included in the analysis. To enhance the predictive performance, we classified the LR into two groups as follows: response group (CR + PR) and non-response group (SD + PD). The LR data of the patients are presented separately for model building and model evaluation datasets in **Table 1**.

For the patients in the current study, we did not do either additional imaging study to delineate necrotic areas or took tissue samples for a histopathological examination. Instead, to minimize the volume measurement error due to the necrosis, we examined the available T1-weighted Gd-contrast enhanced MRI to identify necrosis by the existence of the edema around the enhanced lesion or clear hemorrhage inside the lesion, or by checking the patient's neurologic symptom. We did not see these indications among the patients and their tumors, which we used for the current study. Thus, our tumor volumes might contain necrosis or hemorrhage inside the volume unless it was present clearly outside of the tumor.

Radiomics Analysis

The process of the radiomics analysis is shown in **Figure 1**. The pixel values of the MRI data were rescaled by using the RescaleSlope and RescaleIntercept tags from the DICOM header as follows:

$$\begin{aligned} &Image\ Data \\ &= (Image\ Data) \times RescaleSlope + RescaleIntercept \\ &+ 1000 \end{aligned} \quad (1)$$

Before calculating radiomics features, we applied the medium smooth filter to the rescaled image data. All treatment planning MRI images were analyzed to extract textural features from the GTVs contoured for the radiotherapy plans. The GTV was manually contoured for the radiosurgery treatment planning by radiation oncologists. The feature extraction was performed using IBEX software (24). It is noted that the tumors smaller than 4 mm diameter or volume of 33.5 mm^3 were excluded from further study because of its limited number of pixels available for the texture analysis. We used the following six different

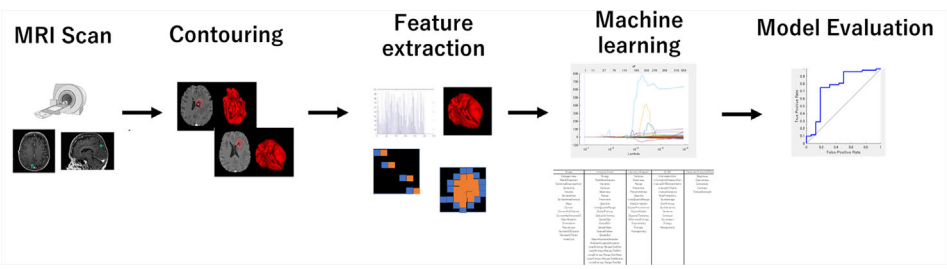


FIGURE 1 | The process for generating a prediction model using a machine learning method with the radiomics feature. The radiomics feature was extracted from the treatment planning MRI data with IBEX. LASSO analysis provided a more regularized model by reducing the number of features. The machine learning model used neural networks.

radiomics feature classes: Gradient Orient Histogram (GOH) (35 features), Gray-Level Co-occurrence Matrix (GLCM) (594), Gray-Level Run Length Matrix (GLRLM) (33), Intensity Direct (ID) (55), Neighbor Intensity Difference Matrix (NIDM) (5), and Shape (18). The resulting 740 features were considered in this study. When there was an option of 2.5D or 3D analysis for texture calculations, we selected 2.5D.

The least absolute shrinkage and selection operator (LASSO) regression was performed in the MATLAB program (Mathworks, Natick, MA, USA) to select the suitable features for the prediction. The LASSO regression performs feature selection during model construction by penalizing the respective regression coefficients. As this penalty is increased, more regression coefficients shrink to zero, resulting in a more regularized model. The most significant predictive features were selected from among all the candidate features for the subsequent training session to build an ML-based prediction model.

Machine Learning-Based Prediction Model

Figure 2 shows an overview of the prediction model generation. A machine learning (ML)-based model was built by using a neural network (NN) with ten hidden layers and rectified linear unit activation (ReLU), as implemented in the MATLAB program. For the classification of local response, tumors in the response group (PR + CR) were labeled as 1, and tumors in the non-response group (SD + PD) were labeled as 0. For the model training, we used the data in the model building dataset (115 tumors of 43 patients) shown in **Table 1**. Tumors were randomly partitioned into a training set (55% tumors), a validation set (15% tumors), and a testing set (30% tumors). The predictive model for the classification was created with the training set and the validation set. The performance of the predictive model was evaluated by the testing set by calculating the accuracy and sensitivity of the prediction. The training-validation-testing processes were repeated five times for the five-fold cross-validation. Then, a model that was the closest to the average accuracy of five-fold cross-validation was selected for the final evaluation. We performed the final assessment with the data in the model evaluation dataset (42 tumors of nine patients), as shown in **Table 1**. The predictive performance of the models was assessed using the area under the receiver operator characteristic (ROC) curve, AUC, as well as the accuracy and sensitivity.

Visual Evaluation

Goodman et al. classified the lesion characteristics into homogeneous, heterogeneous, or ring-enhancing by the pattern of enhancement (20). The uniform enhancement of the entire lesion was defined as homogeneous. If there were any areas of nonhomogeneous enhancement, it was defined as the heterogeneous. Additionally, if there was a rim or ring of contrast enhancement surrounding a central non-enhancing low-signal intensity area, it was identified as a ring-enhancing. In the current study, an experienced radiation oncologist classified the tumors into three types of patterns (homogeneous, heterogeneous,

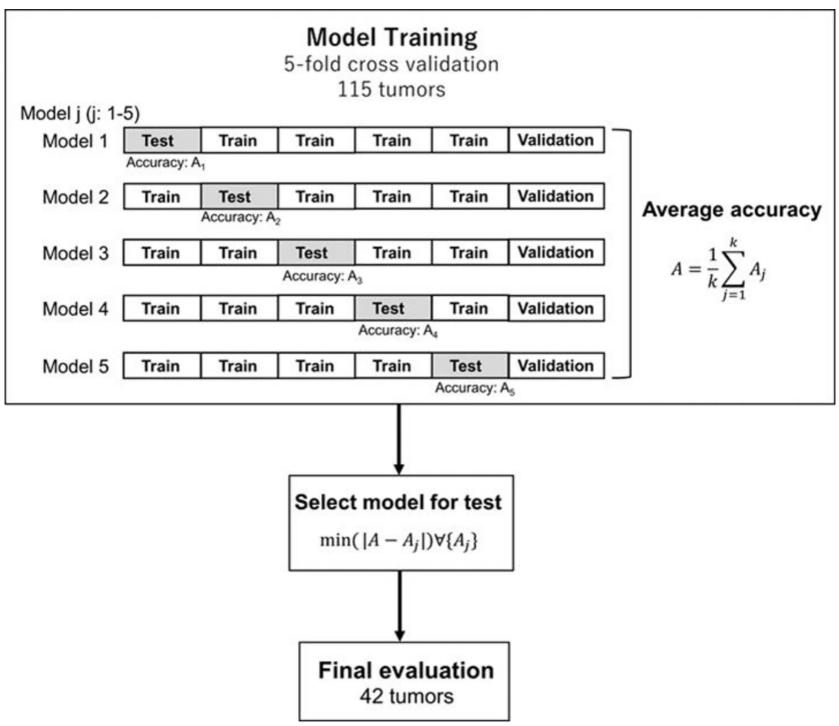


FIGURE 2 | Generation and testing of the prediction model. The proposed NN model with five-fold cross-validation was built in the model training section. Then, the model, which was the closest to the average accuracy of the five prediction models, was selected. The selected model was used for the final evaluation with 42 tumors in the model evaluation dataset.

or ring-enhancing) by visually inspecting the MR images. The treatment outcome was predicted based on the image. Nieder et al. showed that important prognostic factors for complete remission were the small volume and no necrosis (25). Based on the well-accepted knowledge (20, 25), we assigned the predicted response of the homogeneous tumors to the response group (group I) and tumors with heterogeneous or ring enhancement to the non-response group (group II). We compared the visual evaluation method and the ML method using the data in the model building dataset (115 tumors of 42 patients).

RESULTS

First, a total of 740 radiomics features were extracted from the BM MRI images. Then, the number was reduced to seven features by using the LASSO regression method. **Figure 3** shows the binomial deviation (a) and the coefficients (b) as a function of the tuning penalization parameter λ for the LASSO linear regression. As λ increased, only a few coefficients of 740 features remained non-zero, indicating only parameters important for an accurate model. The selected features were 45-7ClusterShade, 225-7ClusterShade, 45-7InformationMeasureCorr-1, 225-7InformationMeasureCorr-1, 90-4InformationMeasureCorr-2, 225-7Energy, and 315-5Energy.

Table 3 shows the performance of the NN models. There were five models generated in the five-fold cross-validation step. Those models were evaluated with the training and testing datasets separately. The average accuracy of the five models was 0.80, with the training data. The model closest to the average accuracy was model 3. Hence, the final evaluation was performed with the model 3. The accuracy and sensitivity of the final model were 0.78 and 0.87 with the model evaluation dataset. **Figure 4** shows the performance of the classifier according to the ROC metrics for the training and testing datasets. The AUC score was 0.89 for the training data and 0.82 for the testing data in the model training section. When we applied the selected model to the final evaluation of 42 tumors in the model evaluation dataset, we obtained the AUC score of 0.87.

Table 4 compares the visual evaluation method and the NN prediction model by accuracy and sensitivity. The former method was applied to the 115 tumors used for the NN model training. The latter was applied to the testing data in the model training section, and the values in the table were the average of the five models. The results showed that the NN model was superior to the visual evaluation for accuracy and sensitivity.

DISCUSSION

Goodman et al. reported that the pattern of tumor images seen on the Gd-contrast enhanced T1-weighted MR images is valuable for predicting the response of a tumor to radiosurgery (20). The current study used radiomics features extracted from radiotherapy planning MRI (Gd-contrast enhanced T1-weighted) to predict the local response (LR) by a machine learning (ML) method with a neural network (NN) classifier.

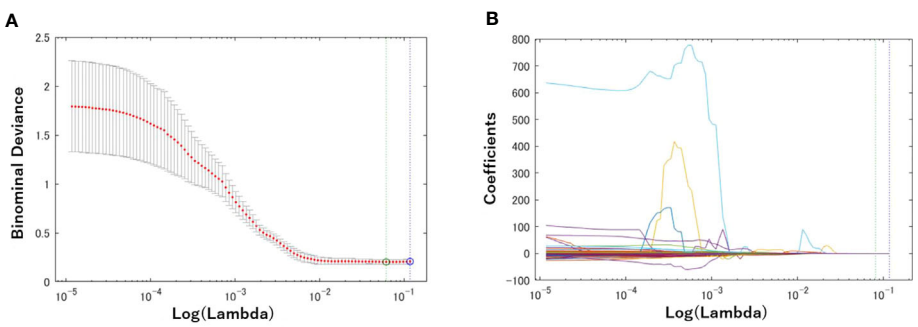


FIGURE 3 | Radiomics feature selection using the Lasso logistic regression model. **(A)** Tuning penalization parameter lambda (λ) and minimum criterion in the Lasso model. The binomial deviance was plotted versus log (λ). **(B)** Lasso coefficient profiles of the 740 radiomics features. The green line showed the optimal lambda in the LASSO method with the least partial likelihood deviance.

TABLE 3 | Model performance.

	Training		Test		Final evaluation	
	Accuracy	Sensitivity	Accuracy	Sensitivity	Accuracy	Sensitivity
Model 1	0.74	1.00	0.75	0.76	–	
Model 2	0.79	0.60	0.81	0.75	–	
Model 3	0.80	0.83	0.79	0.78	0.78	0.87
Model 4	0.81	0.60	0.76	0.68	–	
Model 5	0.81	0.70	0.86	0.74	–	
Average	0.80	0.75	0.81	0.75	–	

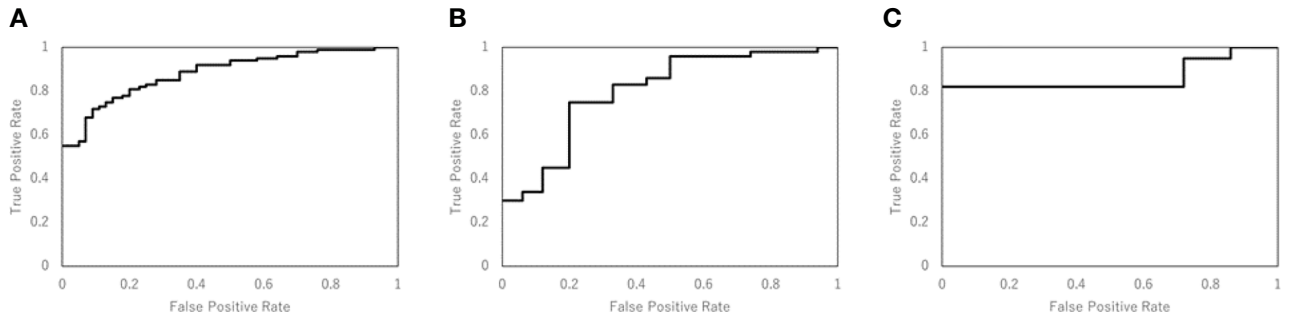


FIGURE 4 | The performance of the NN model was validated according to the ROC metrics for **(A)** the training data in model training section, **(B)** the testing data in model training section, and **(C)** for the final evaluation dataset. The AUC score was 0.89 for the training data, 0.82 for the testing data, and 0.87 for the final evaluation dataset.

TABLE 4 | The assessment of the predictive performance of the visual evaluation and NN model for the testing data used for the NN model training.

	Visual evaluation	*NN model*
Accuracy	0.44	0.80
Sensitivity	0.54	0.74

We compared the predictive performance of the NN model and the visual evaluation method. The accuracy of the new method using the radiomics features yielded a higher prediction accuracy (80%) than the visual approach. Thus, the ML method, such as NN, would be useful for predicting the response of the BMs to GKRS.

The LASSO regression analysis resulted in seven radiomics features, which were useful for the classification, among 740 features initially included in the radiomics analysis. The selection of these features can be understood by the mathematical implication of those features. Cluster shade is a measure of the skewness of the matrix and is believed to gauge the perceptual concepts of uniformity. It may be correlated with lesion characteristics that are heterogeneous or ring-enhancing. Informational Measure of Correlation-1 and Measure of Correlation-1 assess the correlation between the probability distributions using mutual information, which means quantifying the complexity of the texture. Energy is a measure of the magnitude of voxel values in the image. The current study revealed that these were useful features for predicting the response of BMs to GKRS.

The prediction of the LR of BMs to SRS has important practical implications for patients and clinicians. Our prediction model could be useful in clinics. Although the current study created the prediction model of the LR for the radiosurgery, the same approach can be used for all of the treatment methods.

In this study, the comparison between our predictive model and the visual method was made to demonstrate the high predictive performance of the current approach. Goodman et al. (20) tried to identify the necrosis inside the tumor by classifying the tumor into three groups based on the enhancement pattern. However, there is no reliable technique to quantify the amount of necrosis only by visual examination. Hence, the visual classification method suffers from a large uncertainty. Consequently, we expect a large variation among observers for distinguishing three image patterns. Surely, we cannot exclude a potentially better performance of some observers than our method. But, the overall performance of our method should be better than the visual method. Our method does not classify the image pattern into only three types, but it uses more information of images for the decision making than the visual approach. Furthermore, the visual method is applied to only one transverse image since classifying the images into three patterns of the three-dimensional data is time-consuming and almost impossible. As a result, the method should be more accurate for outcome prediction.

There are several recent studies, in which radiomics features were used for more accurate distinction of necrosis from tumor progression and early detection of adverse radiation events (ARE) after radiotherapy of brain tumors (26–29). We used only the GTV (Gd contrast enhanced area) for radiomics analysis in the current study. Suppose we extend the region-of-interest (ROI) by including the volume surrounding the Gd-contrast enhanced area or add other types of imaging data such as PET, for example. In that case, we might be able to predict brain injuries after GKRS. Such a study is interesting and can be undertaken in the future.

There are five limitations to the current study. First, the LR of BMs depends on the prescribed dose. For our GKRS treatment, we prescribed the dose based on the tumor size following the RTOG 90-05 protocol (5). Hence, the LR can be affected not only by the radiomics features but also by the prescribed dose. Secondarily, other clinical factors are statistically significant, but we did not consider in the current study. To improve the prediction performance, therefore, the radiomics features can be combined with the standard biomarkers. Thirdly, the present study used the radiomics features extracted from only radiotherapy planning MRI scans (Gd-contrast enhanced T1-weighted). But, the prediction accuracy may improve by utilizing images taken by other imaging modalities. For example, Wu et al. combined the radiomics features of CT and FDG-PET for predicting distant metastasis in early-stage non-small cell lung cancer after stereotactic body radiation therapy (30). Ordering additional imaging studies other than standards requires additional funding and a special protocol, but it may be an

important step for more accurate predictions. Fourth, the current predictive model was built using only the metastatic brain tumors of patients with melanoma as the primary, mainly due to the availability of the treatment follow-up data. Lastly, only one experienced radiation oncologist classified the tumors into three MR image patterns for the visual evaluation. For a fair comparison of the ML-based method with the visual evaluation method, we need to recruit more experts to study the effects of inter-observer variation on the outcome prediction.

To overcome the first three limitations, we plan to improve the prediction model by adding radiomics features of other MR imaging protocols, dosimetric parameters such as prescribed dose and standard biomarkers. Extending the model to BMs with different primary cancer types is straightforward as long as the necessary data for model training are available. The versatile prediction model will be created by including multi-institution and other brain metastases patients. The uncertainty of the inter-observer with visual evaluation is a serious problem. However, we believe that the prediction model proposed in the current study decreases the uncertainty with the visual evaluation.

CONCLUSION

The proposed NN model using the radiomics features of tumor image was more accurate than the visual evaluation method using the image pattern information in predicting the local response of brain metastases to GKRS. Because of the excellent prediction ability of the method, the method can be used to help physicians to gain a more accurate prediction of the treatment outcome than the traditional method.

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 IRB 0801M23942 (University of Minnesota). The research consent requirement was waived because of the retrospective nature of the study.

AUTHOR CONTRIBUTIONS

YW conceived and designed the study. XT did radiomics analyses. CL and YW contributed to the data collection and analysis. DK made substantial contributions to the applications of machine learning techniques. DK and YW prepared the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The portions of the current study were presented as an e-poster at the 19th Leksell Gamma Knife Society Meeting, Dubai, UAE, March 4–8, 2018, and as a short oral talk at the 2019 ASTRO Annual Meeting, Chicago, IL, September 15–18, 2019.

REFERENCES

1. Davis FG, Dolecek TA, McCarthy BJ, Villano JL. Toward determining the lifetime occurrence of metastatic brain tumors estimated from 2007 United States cancer incidence data. *Neuro Oncol* (2012) 14:1171–7. doi: 10.1093/neuonc/nos152
2. Lu-Emerson C, Eichler AF. Brain metastases. *Continuum (Minneap Minn)* (2012) 18:295–311. doi: 10.1212/01.CON.0000413659.12304.a6
3. Kondziolka D, Martin JJ, Flickinger JC, Friedland DM, Brufsky AM, Baar J, et al. Long-term survivors after gamma knife radiosurgery for brain metastases. *Cancer* (2005) 104:2784–91. doi: 10.1002/cncr.21545
4. Andrews DW, Scott CB, Sperduto PW, Flanders AE, Gaspar LE, Schell MC, et al. Whole brain radiation therapy with or without stereotactic radiosurgery boost for patients with one to three brain metastases: phase III results of the RTOG 9508 randomised trial. *Lancet* (2004) 363:1665–72. doi: 10.1016/S0140-6736(04)16250-8
5. Shaw E, Scott C, Souhami L, Dinapoli R, Kline R, Loeffler J, et al. Single dose radiosurgical treatment of recurrent previously irradiated primary brain tumors and brain metastases: final report of RTOG protocol 90-05. *Int J Radiat Oncol Biol Phys* (2000) 47:291–8. doi: 10.1016/s0360-3016(99)00507-6
6. Sperduto PW, Berkey B, Gaspar LE, Mehta M, Curran W. A new prognostic index and comparison to three other indices for patients with brain metastases: an analysis of 1,960 patients in the RTOG database. *Int J Radiat Oncol Biol Phys* (2008) 70:510–4. doi: 10.1016/j.ijrobp.2007.06.074
7. Sperduto PW, Yang TJ, Beal K, Pan H, Brown PD, Bangdiwala A, et al. Estimating Survival in Patients With Lung Cancer and Brain Metastases: An Update of the Graded Prognostic Assessment for Lung Cancer Using Molecular Markers (Lung-molGPA). *JAMA Oncol* (2017) 3:827–31. doi: 10.1001/jamaoncol.2016.3834
8. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RG, Granton P, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* (2012) 48:441–6. doi: 10.1016/j.ejca.2011.11.036
9. 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
10. Mouraviev A, Detsky J, Sahgal A, Ruschin M, Lee YK, Karam I, et al. Use of Radiomics for the Prediction of Local Control of Brain Metastases After Stereotactic Radiosurgery. *Neuro Oncol* (2020) 22(6):797–805. doi: 10.1093/neuonc/noaa007 pii: noaa007.
11. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: the process and the challenges. *Magn Reson Imaging* (2012) 30:1234–48. doi: 10.1016/j.mri.2012.06.010
12. Ahmed A, Gibbs P, Pickles M, Turnbull L. Texture analysis in assessment and prediction of chemotherapy response in breast cancer. *J Magnetic Resonance Imaging* (2013) 38:89–101. doi: 10.1002/jmri.23971
13. Wu W, Parmar C, Grossmann P, Quackenbush J, Lambin P, Bussink J, et al. Exploratory Study to Identify Radiomics Classifiers for Lung Cancer Histology. *Front Oncol* (2016) 6:71. doi: 10.3389/fonc.2016.00071
14. Fried DV, Tucker SL, Zhou S, Liao Z, Mawlawi O, Ibbott G, et al. Prognostic value and reproducibility of pretreatment CT texture features in stage III non-small cell lung cancer. *Int J Radiat Oncol Biol Phys* (2014) 90:834–42. doi: 10.1016/j.ijrobp.2014.07.020
15. Itakura H, Achrol AS, Mitchell LA, Loya JJ, Liu T, Westbroek EM, et al. Magnetic resonance image features identify glioblastoma phenotypic subtypes with distinct molecular pathway activities. *Sci Transl Med* (2015) 7:303ra138. doi: 10.1126/scitranslmed.aaa7582

16. Kickingereder P, Götz M, Muschelli J, Wick A, Neuberger U, Shinohara RT, et al. Large-scale Radiomic Profiling of Recurrent Glioblastoma Identifies an Imaging Predictor for Stratifying Anti-Angiogenic Treatment Response. *Clin Cancer Res* (2016) 22:5765–71. doi: 10.1158/1078-0432.CCR-16-0702
17. Zhou M, Scott J, Chaudhury B, Hall L, Goldgof D, Yeom KW, et al. Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches. *AJNR Am J Neuroradiol* (2018) 39:208–16. doi: 10.3174/ajnr.A5391
18. Ryu YJ, Choi SH, Park SJ, Yun TJ, Kim JH, Sohn CH. Glioma: application of whole-tumor texture analysis of diffusion-weighted imaging for the evaluation of tumor heterogeneity. *PloS One* (2014) 9:e108335. doi: 10.1371/journal.pone.0108335
19. Rizzo S, Petrella F, Buscarino V, De Maria F, Raimondi S, Barberis M, et al. CT Radiogenomic Characterization of EGFR, K-RAS, and ALK Mutations in Non-Small Cell Lung Cancer. *Eur Radiol* (2016) 26:32–42. doi: 10.1007/s00330-015-3814-0
20. Huang CY, Lee CC, Yang HC, Lin CJ, Wu HM, Chung WY, et al. Radiomics as prognostic factor in brain metastases treated with Gamma Knife radiosurgery. *J Neurooncol* (2020) 146(3):439–49. doi: 10.1007/s11060-019-03343-4
21. Goodman KA, Sneed PK, McDermott MW, Shiau CY, Lamborn KR, Chang S, et al. Relationship between pattern of enhancement and local control of brain metastases after radiosurgery. *Int J Radiat Oncol Biol Phys* (2001) 50:139–46. doi: 10.1016/s0360-3016(00)01584-4
22. Lin HY, Watanabe Y, Cho LC, Yuan J, Hunt MA, Sperduto PW, et al. Gamma knife stereotactic radiosurgery for renal cell carcinoma and melanoma brain metastases-comparison of dose response. *J Radiosurg SBRT* (2013) 2:193–207.
23. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* (2009) 45:228–47. doi: 10.1016/j.ejca.2008.10.026
24. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE, et al. IBEX: an open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys* (2015) 42:1341–53. doi: 10.1118/1.4908210
25. Nieder C, Berberich W, Schnabel K. Tumor-related prognostic factors for remission of brain metastases after radiotherapy. *Int J Radiat Oncol Biol Phys* (1997) 39:25–30. doi: 10.1016/s0360-3016(97)00154-5
26. Zhang Z, Yang J, Ho A, Jiang W, Logan J, Wang X, et al. A predictive model for distinguishing radiation necrosis from tumour progression after gamma knife radiosurgery based on radiomic features from MR images. *Eur Radiol* (2018) 28(6):2255–63. doi: 10.1007/s00330-017-5154-8
27. Peng L, Parekh V, Huang P, Lin DD, Sheikh K, Baker B, et al. Distinguishing True Progression From Radionecrosis After Stereotactic Radiation Therapy for Brain Metastases With Machine Learning and Radiomics. *Int J Radiat Oncol Biol Phys* (2018) 102(4):1236–43. doi: 10.1016/j.ijrobp.2018.05.041
28. 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 Cancer* (2020) 20(1):502. doi: 10.1186/s12885-020-06957-4
29. Kocher M, Ruge MI, Galldiks N, Lohmann P. Applications of radiomics and machine learning for radiotherapy of malignant brain tumors. *Strahlenther Onkol* (2020) 196(10):856–67. doi: 10.1007/s00066-020-01626-8
30. Wu J, Aguilera T, Shultz D, Gudur M, Rubin DL, Loo BWJr, et al. Early-Stage Non-Small Cell Lung Cancer: Quantitative Imaging Characteristics of (18)F Fluorodeoxyglucose PET/CT Allow Prediction of Distant Metastasis. *Radiology* (2016) 281:270–8. doi: 10.1148/radiol.2016151829

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: 04 February 2022
doi: 10.3389/fonc.2022.811415

A Multi-Compartment Model of Glioma Response to Fractionated Radiation Therapy Parameterized *via* Time-Resolved Microscopy Data

Junyan Liu[1], *David A. Hormuth II*[2,3], *Jianchen Yang*[1] *and Thomas E. Yankeelov*[1,2,3,4,5,6*]

[1] Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX, United States, [2] Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, TX, United States, [3] Livestrong Cancer Institutes, The University of Texas at Austin, Austin, TX, United States, [4] Department of Diagnostic Medicine, The University of Texas at Austin, Austin, TX, United States, [5] Department of Oncology, The University of Texas at Austin, Austin, TX, United States, [6] Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

Purpose: Conventional radiobiology models, including the linear-quadratic model, do not explicitly account for the temporal effects of radiation, thereby making it difficult to make time-resolved predictions of tumor response to fractionated radiation. To overcome this limitation, we propose and validate an experimental-computational approach that predicts the changes in cell number over time in response to fractionated radiation.

Methods: We irradiated 9L and C6 glioma cells with six different fractionation schemes yielding a total dose of either 16 Gy or 20 Gy, and then observed their response *via* time-resolved microscopy. Phase-contrast images and Cytotox Red images (to label dead cells) were collected every 4 to 6 hours up to 330 hours post-radiation. Using 75% of the total data (i.e., 262 9L curves and 211 C6 curves), we calibrated a two-species model describing proliferative and senescent cells. We then applied the calibrated parameters to a validation dataset (the remaining 25% of the data, i.e., 91 9L curves and 74 C6 curves) to predict radiation response. Model predictions were compared to the microscopy measurements using the Pearson correlation coefficient (PCC) and the concordance correlation coefficient (CCC).

Results: For the 9L cells, we observed PCCs and CCCs between the model predictions and validation data of (mean ± standard error) 0.96 ± 0.007 and 0.88 ± 0.013, respectively, across all fractionation schemes. For the C6 cells, we observed PCCs and CCCs between model predictions and the validation data were 0.89 ± 0.008 and 0.75 ± 0.017, respectively, across all fractionation schemes.

Conclusion: By proposing a time-resolved mathematical model of fractionated radiation response that can be experimentally verified *in vitro*, this study is the first to establish a framework for quantitative characterization and prediction of the dynamic radiobiological response of 9L and C6 gliomas to fractionated radiotherapy.

Keywords: radiobiology, glioma, computational biology, mathematical modeling, oncology, brain cancer cell

OPEN ACCESS

Edited by:
Yuming Jiang,
Stanford University, United States

Reviewed by:
Carlos Perez-Torres,
Purdue University, United States
Jeho Jeong,
Memorial Sloan Kettering Cancer Center, United States

***Correspondence:**
Thomas E. Yankeelov
thomas.yankeelov@utexas.edu

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: *08 November 2021*
Accepted: *17 January 2022*
Published: *04 February 2022*

Citation:
Liu J, Hormuth DA II, Yang J and Yankeelov TE (2022) A Multi-Compartment Model of Glioma Response to Fractionated Radiation Therapy Parameterized via Time-Resolved Microscopy Data. Front. Oncol. 12:811415. doi: 10.3389/fonc.2022.811415

1 INTRODUCTION

Radiation therapy is a central component of the standard-of-care for treating malignant gliomas (1), especially when the tumor is located near sensitive brain regions with important functions that are unresectable by surgery. Though various dose escalation and fractionation schemes (i.e., hyper- and hypo- fractionation) have been investigated, none have shown definitive improvement on the long-term survival for glioblastoma patients (2). One reason for this limitation is that the efficacy of radiation therapy varies between patients due to heterogeneous radiosensitivity of the cells within each individual's tumor (3). If there was a mathematical model that could accurately characterize, and predict, the response of tumor cells to radiation therapy with patient-specific data, then there would be the opportunity to optimize the radiation plan for each individual (4). The currently accepted model for evaluating radiation response given a specific dose is the linear quadratic (LQ) model which was originally developed empirically more than 40 years ago (5). The LQ model quantifies the survival fractions of cell colonies given a specific radiation dose and, though it provides a simple, and practical relationship between those two measureables, it is not without its limitations. In particular, the LQ model does not explicitly characterize the temporal changes in tumor cell number; that is the LQ model is not a function of time. Thus, while it can provide accurate predictions of endpoint predictions (6), it is not capable of predicting the temporal dynamics of radiation response. Additionally, interpretation of the two main parameters in the LQ model (alpha and beta) is fraught with difficult, thereby clouding their biological meaning (7). This is despite the now vast biological knowledge that exists regarding DNA repair (8) and radiation-induced cell death pathways (9). To address these two limitations, we previously proposed and validated a mechanism-based time-resolved model (10) to a single-dose treatment. We now seek to extend this model to account for multiple-fraction treatment regimens.

Though a large single-dose of radiation can effectively kill tumor cells, it is rarely used in clinical settings as high doses also cause irreversible cytotoxicity to surrounding healthy tissues. Thus, the notion of delivering a target total dose in "fractions" over an extended period of time was adopted. There are four key conceptions that are frequently kept in mind when designing multi-fraction treatment plans: DNA damage repair, repopulation, cell cycle redistribution, reoxygenation [sometimes referred to as the "Four R"s (11)]. The DNA damage repair mechanisms help the nearby healthy tissue recover between treatment intervals (12), with the hope that the repair mechanisms are erroneous in the tumor cells leading to their eventual cell death after repeated fractions (13). Tumor cell repopulation (i.e., the ability of tumor cells to proliferate between treatment intervals) can undermine radiation efficacy, and thus may require extra fraction and/or total dose to achieve tumor control (14). Cell cycle redistribution increases the average tumor cell killing by allowing radiation-resistant cells in S phase to redistribute into the more sensitive M phase (15). Reoxygenation also enhances radiation damage as radiation can produce free radicals which damage DNA, and this damage can be made permanent by the presence of molecular oxygen [i.e., the 'oxygen-fixation hypothesis' (16)]. Additionally, hypoxic regions of a tumor regions may become reoxygenated between fractions (17). Previous modeling work has focused on quantifying the effects of these four "R"s on the endpoint survival fraction by (for example) incorporating an "oxygen enhancement ratio" (18) or repopulation (19) into the LQ model. More recently, several studies have constructed radiation response models that account for temporal changes in hypoxia (20), DNA repair (21), and fractionation (22). Hormuth et al. contributed a tissue-scale model that employed the oxygen enhancement ratio to adjust the radiation efficacy during fractionation treatment and tested model predictions against *in vivo* MRI data (23, 24). Brüningk et al. (25) proposed a cell-scale decision tree model that accounted for conversion between cell cycle compartments after radiation. All these models indicate an increasing interest in mathematically describing the temporal dynamics of radiation response that are not captured by the conventional LQ-based models.

We first propose to extend our previous single-dose model (10) to characterize the radiation response of gliomas to fractionated treatment. The fractionation model explicitly incorporates temporal changes due to DNA damage repair, cell repopulation, and cell cycle effect related to senescence. We then perform *in vitro* microscopy experiments with 9L and C6 cell lines to obtain the radiation response curves collected at high temporal resolution under different treatment schedules and total radiation doses. Our model is then trained on 75% of the total data to calibrate the parameters. Finally, the remaining 25% of data serve as a validation group to assess the model's predictive accuracy. Our mechanism-based, time-resolved, mathematical model achieves high predictive accuracy across a range of fractionation schedules verified by six different fractionation schemes and both cell lines.

2 MATERIALS AND METHODS

2.1 Experiments

2.1.1 Cell Culture

The 9L (American Type Culture Collection, ATCC) and C6 (Sigma Aldrich) cells are cultured according to the manufacturer's guidelines as previously described (10). The 9L cell line is cultured with Eagle's minimum essential medium (ATCC, VA), and the C6 cell line is cultured with Ham's F12 (Corning, NY). Both cell lines media are supplemented with 10% FBS and 2 mM L-Glutamine. 0.2% Plasmocin Prophylactic (*In vivo*gen, CA) is supplied in the media to prevent mycoplasma contamination. The mean passage number of the cells used in the experiments was 50 ± 20. Both 9L and C6 are rat cell lines, but are commonly used for general glioma studies (26, 27).

2.1.2 Radiation Treatment and Imaging

Figure 1 illustrates the radiation treatment schedule employed in these studies. 9L and C6 cells were seeded on 96-well plates (Corning, NY) at densities ranging from 3,200 to 32,000 cells/cm^2 (1,000 to 10,000 cells total per well). To avoid cells reaching the carrying capacity at later timepoints (which can result in cell death due to lack of nutrients and physical space), we do not seed at a confluence higher than 10,000 total cells. The cells are then

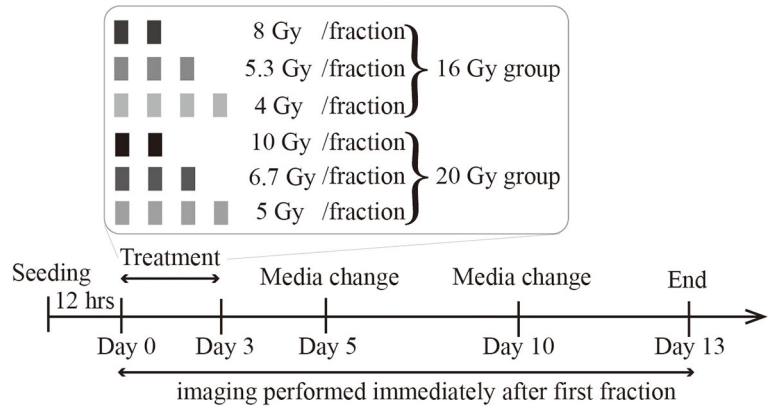


FIGURE 1 | Radiation treatment schedule. Cells are seeded, incubated overnight, and then treated with either a total dose of 16 Gy or 20 Gy. In the 16 Gy total dose group, cells are irradiated with 2 fractions of 8 Gy, 3 fractions of 5.3 Gy, or 4 fractions of 4 Gy. In the 20 Gy total dose group, cells are irradiated with 2 fractions of 10 Gy, 3 fractions of 6.7 Gy, or 4 fractions of 5 Gy. All irradiations are 24 hours apart. The culture media is refreshed every 5 days, and imaging lasts up to two weeks after the initial irradiation.

incubated overnight (~12 hours) to allow for attachment and recovery. Before irradiation, the media is changed and augmented with 250 nM Cytotox Red dye (Cat. *No.* 4632, Essen BioScience, MI), a non-perturbing fluorescent dye to label dead cells. For both the 9L and C6 cell lines, we separate the wells into either a 16 Gy or a 20 Gy total dose group. In the 16 Gy group, we irradiate the cells with either four fractions of 4 Gy, three fractions of 5.3 Gy, or two fractions of 8 Gy with 24-hour intervals between every fraction. In the 20 Gy group, we irradiate cells with four fractions of 5 Gy, three fractions of 6.7 Gy, and two fractions of 10 Gy with 24-hour intervals between every fraction. All radiation is delivered by a CellRad irradiator (Faxitron X-Ray Corp, Wheeling, IL, MA) at a dose rate of 1.5 Gy/min (130 KeV, 5 mA, 0.5 mm aluminum filter). After treatment, phase-contrast images and fluorescent Cytotox Red images (for labeling dead cells) are acquired immediately after the first fraction *via* the Incucyte S3 live imaging system (Essen BioScience, Ann Arbor, MI) with a 4× objective, whole-well imaging mode every four to six hours up to approximately 330 hours post-irradiation. Our media culture, along with the Cytotox Red dye, was refreshed every five days throughout the experiment. To prevent cell loss when refreshing the media in the 96-well plates, we pipetted only the top 80 µl of the total 100 µl per well to minimize the disturbance to attached cells. Live and dead cells were segmented using a semi-automated pipeline consisting of a histogram Otsu-based method followed by a morphology-based cell debris removal [described in detail in (10)].

2.1.3 DSB Repair Kinetics

To quantify radiation-induced DNA double strand breaks (DSB) and repair kinetics, we previously measured the expression level of the γH2AX protein [a commonly used DSB biomarker (28)] *via* flow cytometry after irradiating cells with a single dose of 2, 4, 8, or 16 Gy (see the **Supplementary Materials** of (10) for details). We then used linear interpolation to obtain the DSB repair kinetics for all other doses. The same data is used in this study as described below.

2.1.4 DNA Repair

DNA repair is represented by an exponential decay equation:

$$f_{DSB}(t, D) = e^{-k_{repair}(D)\cdot t} \tag{1}$$

where $f_{DSB}(t,D)$ is the fraction of DSBs remaining unrepaired (normalized between 0 and 1) at time, t, $k_{repair}(D)$ (in units of hr^{-1}) is the DSB repair rate at dose per fraction D. We use the same k_{repair} measured from our single-dose treatment study as an estimate of the DSB repair rate during fractionation schedules. This assignment is justified as Mariotti et al. (29) have measured γH2AX under both single and multi-fractionated treatment schemes and showed similar repair kinetics in response to a second dose when cells are given proper time for repair and recovery. Our previous flow cytometry experiments (10) indicate that most (> 80%) DSBs are repaired within 24 hours given the dose range we employ in the experiments. Therefore, this estimation is reasonable.

2.2 Mathematical Modeling of Cell Growth and Response to Radiation Therapy

The fractionated treatment model is an extension of our previous single-dose radiation model described in (10) that models cell response as a function of early cell death (corresponding to apoptosis) and late cell death (corresponding to mitotic catastrophe). We present the salient details of our single-dose radiation model, but note that the complete development and underlying assumptions are detailed in (10).

2.2.1 Single Species Model of Cell Growth in the Absence of Radiation Therapy

For glioma cell proliferation in the absence of radiation therapy, we augment exponential growth by incorporation of the logistic growth and Allee effect:

$$\frac{dN(t)}{dt} = k_p \cdot N(t) \cdot \underbrace{\left(\frac{N(t)}{\theta} + A\right)}_{\text{Allee effect}} \cdot \underbrace{\left(1 - \frac{N(t)}{\theta}\right)}_{\text{logistic growth}} \quad (2)$$

where $N(t)$ is the tumor cell confluence, k_p is the proliferation rate (see **Table 1** for a listing of all model parameters, their definition, and units), A quantifies the strength of the Allee effect, and θ is the carrying capacity (i.e., the maximum number of cells that can fit within a given volume due to space and nutrient limitations). The Allee effect describes the cooperation effects in cell proliferation rate and is proved significant in glioblastoma progression by Neufeld et al. (30). Our previous analysis (10) used model selection to establish that both logistic growth and the Allee effect are necessary to accurately describe our glioma data.

2.2.2 Single-Species Model of Cell Growth in the Response to Radiation Therapy

After radiation, a small number of cells undergo early apoptosis, which is an outcome of activation of DNA protein kinase and p53 due to excessive DSBs (31); these events occur on the timescale of hours to days (32). Meanwhile, DNA misrepair does not directly kill cells and can accumulate within cells' genome, eventually triggering chromosome aberration and mitotic catastrophe (33, 34); these events occur on the timescale of days to weeks (32). Based on the above mechanisms, we add early and late death terms to Eq. (2):

$$\frac{dN(t)}{dt} = (k_p - k_{ld}(t, D, N_0)) \cdot \underbrace{\left(\frac{N(t)}{\theta} + A\right)}_{\text{Allee effect}} \cdot N(t) \cdot \underbrace{\left(1 - \frac{N(t)}{\theta}\right)}_{\text{logistic growth}} - k_{ed}(t, D, N_0) \cdot N(t) \quad (3)$$

where k_{ed} and k_{ld} (both in units of hr^{-1}) represent early death and late death rates, respectively, and are a function of time t, dose per fraction D, and the initial confluency N_0. We use the following equation to describe early cell death as a function of the fraction of unrepaired DSBs within the cell's genome, $f_{DSB}(t)$:

$$k_{ed}(t, D, N_0) = \sum_{i=1}^{\substack{fraction \\ number}} k_{accute}(D, N_0) \cdot f_{DSB}(t^i, D) \quad (4)$$

$$t^i = \begin{cases} t - 24 \cdot (i-1) & t > 24 \cdot (i-1) \\ 0 & t < 24 \cdot (i-1) \end{cases} \quad (5)$$

where $k_{acute}(D,N_0)$ is the acute death rate. t^i (units: hours) is the time that has passed since the i^{th} fraction. As it would be extremely complex to model the death due to each fraction separately, $k_{acute}(D,N_0)$ is viewed as an average death rate across all fractions and written as:

$$k_{acute}(D, N_0) = (\alpha_{acute,N} \cdot N_0 + 1) \cdot k_{acute,D} \cdot D \quad (6)$$

where $\alpha_{acute,N}$ is a scale factor representing the contribution of acute death due to the initial seeding density, N_0. Biologically, N_0 influences (due to cell-cell contact) the proportion of actively proliferating cells at the time of radiation, which determines radiation sensitivity[(as late G2 and M phase is the most sensitive cell cycle (35)]. $k_{acute,D}$ is a death rate indicating the contribution of radiation dose to acute death, as larger doses can translate to a higher number of DSBs and, therefore, early apoptosis. **Figure 2** illustrates the changes in k_{ed} over time. We note that the "+1" in Eq. (6) is for mathematical convenience. When $\alpha_{acute,N}$ (i.e., the contribution of the initial seeding density to acute death) is close to 0, Eq. (6) simplifies to $k_{acute}(D,N_0)= k_{acute,D} \cdot D$. Thus, without the "+1", when $\alpha_{acute,N}$ approaches 0, $k_{acute}(D,N_0)$ will also decrease to 0 and the effect of dose will only be determined by seeding density effects.

TABLE 1 | Model parameters and variables.

Parameter	Unit	Interpretation	Source
k_p	hr^{-1}	Proliferation rate	Computed from control, untreated group and fixed throughout experiments
θ	1	Carrying capacity	
A	1	Allee effect	
N_p	1	Confluence of proliferating cells	Initial cell confluence and total confluence (i.e., N_p+N_s) are measured from microscopy data.
N_s	1	Confluence of senescent cells	
N_0	1	Initial confluence of cells at time = 0	
f_{DSB}	1	The fraction of DSBs remaining unrepaired (normalized between 0 and 1)	Measured by flow cytometry
$k_{acute,N}$	hr^{-1}	Death rate quantifying the contribution of initial confluence to early death	Fit globally with all treated cell response curves.
$a_{accum,N}$	1	Scale factor quantifying the contribution of initial confluence to late death	
$k_{accum,D}$	hr^{-1}	Death rate quantifying the contribution of radiation doses to late death	
r	hr^{-1}	Radiation efficacy	
k_{ps}	hr^{-1}	Conversion rate from proliferation to senescent components.	

The unit "1" means the parameter is unitless.

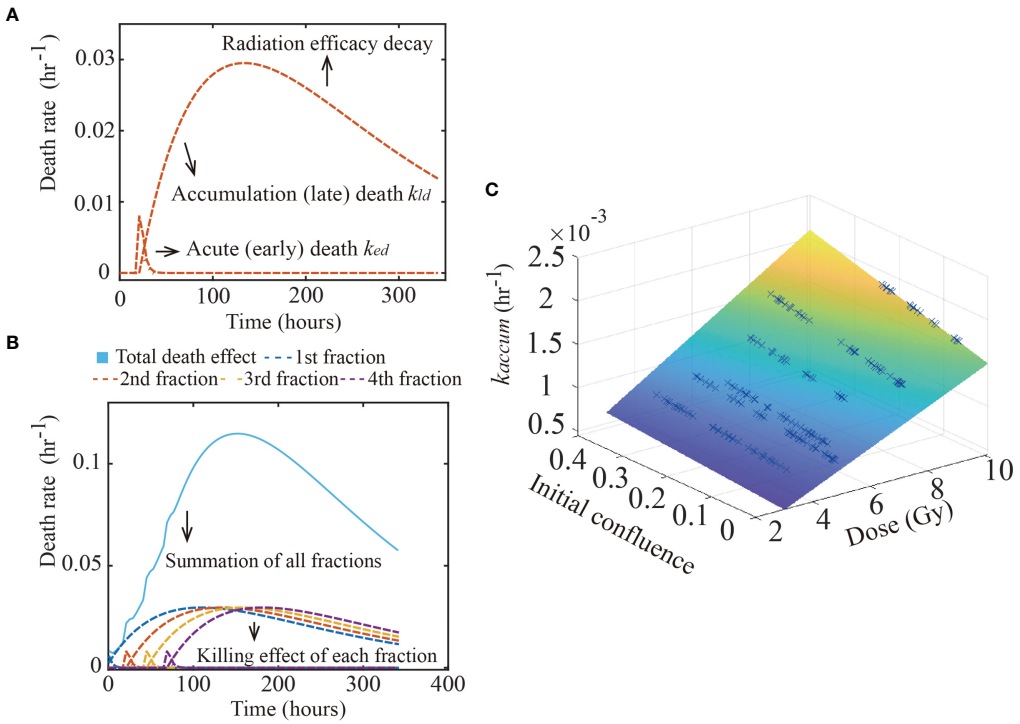


FIGURE 2 | Glioma death rate over time. The figure shows the calibrated early and late death parameters of 9L cell line. Panel **(A)** shows how the early and late death rates change as a function of time from a single treatment. After radiation, the acute death spikes and decreases as the double strand breaks are repaired. In contrast, misrepaired DSBs accumulate within the cells' genome, causing the late death rate to increase over time, before it eventually decreases as the radiation efficacy decays. Panel **(B)** shows the summation (labeled by the blue solid line) of four fractions (each fraction labeled as dashed lines) irradiated every 24 hours starting from time 0. Each fraction has the same effect as in panel **(A)**. Panel **(C)** illustrates the hypothesis captured by Eq. (7); namely, that $k_{accum}(D,N_0)$ is a function of both dose and initial confluence. Each blue cross indicates the calibrated result of one replicate of the 9L cell line.

The "late death" component models mitotic catastrophe of misrepaired cells that occurs following several cell divisions. This population has previously also been referred to as the "abortive fraction" (36). Some cells can survive hours to weeks after radiation before mitotic catastrophe occurs. Here, we model late death as:

$$k_{id}(t,D,N_0) = \sum_{1}^{\substack{fraction\\number}} k_{accum}(D,N_0)\cdot t^i \cdot e^{-r\cdot t^i} \qquad (7)$$

where $k_{accum}(D,N_0)$ is that rate at which the misrepair error first accumulates within the cells' genome, and r (in units of hr^{-1}) controls the decay rate of the radiation efficacy. The decay rate of radiation efficacy is viewed as an intrinsic property of each cell line and is a constant across different treatment conditions. Just as for the early death term, we model the parameter $k_{accum}(D,N_0)$ as an average across fractions, instead of modeling each fraction separately. The accumulation death rate $k_{accum}(D,N_0)$ is thus written as:

$$k_{accum}(D,N_0) = (\alpha_{accum,N}\cdot N_0 + 1)\cdot k_{accum,D}\cdot D \qquad (8)$$

where $\alpha_{accum,N}$ is a scale factor representing the contribution of late accumulation cell death due to the initial seeding density, N_0. Eq. (8) accounts for the fact that cell density directly impacts the proliferation rate *via* (for example) the proportion cells that are in M phase, which eventually determines how many cells can undergo mitotic catastrophe as it occurs during M phase. $k_{accum,D}$ is a death rate describing the contribution of radiation dose to accumulation death, as high doses induce a high probability of misrepair [caused by misjoining of clustered DSBs (37)]. Note that Eqs. (6) – (8) have a modified form from our previous single dose study (10); we return to this point in Discussion section.

2.2.3 Two-Species Model of Cell Growth in Response to Radiation Therapy

We construct a two-species model of response to radiation therapy by considering both proliferative and senescent tumor cells. Several mechanisms in radiobiology contribute to the appearance of a senescent population after radiation (38) including (for example) cell cycle checkpoint pathways (39). These senescent cells can remain metabolically active, but undergo irreversible cell cycle arrest and thus can no longer replicate. Without this component, the model assumes cells either return to proliferating (overestimate cell survival) or undergo early or late death (overestimate radiation cell killing), thereby causing a systematic error in the predicted confluence.

Therefore, we extended Eq. (3) to include a proliferating compartment, $N_p(t)$, and a senescent compartment, $N_s(t)$:

$$\frac{dN_p(t)}{dt} = (k_p - k_{ld}(t, D, N_0)) \cdot \left(\frac{N_p(t) + N_s(t)}{\theta} + A\right) \cdot N(t) \cdot \left(1 - \frac{N_p(t) + N_s(t)}{\theta}\right) - k_{ed}(t, D, N_0) \cdot N_p(t) - k_{ps}(D_{total}) \cdot N_0 \cdot N_p(t) \quad (9)$$

$$\frac{dN_s(t)}{dt} = k_{ps}(D_{total}) \cdot N_0 \cdot N_p(t) \quad (10)$$

Eq. (9) - (10) characterize the conversion from the proliferation to senescent compartment due to radiation. As cell-cell contact promotes senescence, we assume that the convertion rate follows a simple linear relation proportional to the initial cell density with the proportionality constant k_{ps} (units of hr^{-1}). Note that the carrying capacity, θ, is now shared by both N_p and N_s. Additionally, the conversion rate k_{ps} is a function of the total dose (i.e., 16 Gy or 20 Gy in our experiments). This simplification (i.e., making k_{ps} a constant based on the total dose) is because we do not have a direct measurement of the senescent population as this population is changing with each dosing scheme and over time. That is, k_{ps} should really be a function of time, dose per fraction, and fraction number. To practically realize such an explicit expression for k_{ps} would require additional experimental measurements.

2.3 Numerical Implementation of the Mathematical Models

ODE models were implemented *via* the finite difference method with a fully explicit forward Euler formulation with a time step of 0.01 hrs with initial condition, $N_p(0)$, equal to the cell confluence measurements at time 0 and $N_s(0) = 0$. Early acute death rate term is multiplied by a smooth heaviside step function, *via* the hyperbolic tangent function; tanh $[f_{DSB}(t)]$, to prevent curve discontinuity.

2.4 Model Selection

Eqs. (1) – (10) are based on general radiobiology mechanisms. For specific cell lines with different signaling pathways and radiation sensitivity, it is unclear if this model is appropriate to describe the data. Therefore, it is crucial to perform model selection prior to applying the model to a specific cell line. Starting from the above "full model", we systematically remove one or two parameters, yielding seven competing "daughter" models. Using the early death term as an example, the initial seeding density N_0 and doses D might not have an impact on early apoptosis for the 9L and C6 cell lines. By removing either $\alpha_{acute,N}$ or $k_{accum,D}$ or both, we obtain three "daughter" models (i.e., models 2-4 in section 2 of the **Supplementary Materials**). Similarly, removing the late death term (three models) or the senescent term (one model) yields an additional seven "daughter" models. (See the section 2 of the **Supplementary Materials** for the formulation of all eight models.). For each model, parameters are fit with the Levenberg-Marquardt algorithm (*via* "lsqnonlin" in MATLAB). To determine which mechanisms are required for optimally characterizing 9L and C6 data and to obtain the most parsimonious model, we perform model selection on these eight models *via* the Akaike information criterion (AIC) (40):

$$AIC = n \cdot ln\frac{RSS}{n} + 2p + \frac{2p^2 + 2p}{n - p - 1} \quad (11)$$

where n is the sample number (i.e., the number of cell confluency curves in our scenario), RSS is the residual squared sum between the model fit and data, and p is the number of free parameters. The AIC finds the most parsimonious model by balancing the relative goodness-of-fit with the number of free parameters. Specifically, we build our training set by randomly picking seventy-five percent of the data under each treatment condition, leading to 262 replicates for the 9L cells and 211 replicates for the C6 cells. (That is, we pick 75% from four fractions of 4 Gy, 75% from two fractions of 10 Gy, etc., thereby ensuring that each treatment condition is equally represented in the training set.) The remaining 25% of the data is used for validation. (Note that the word 'replicate' is in reference to our experiments; that is, we repeated a group of independent wells with the same treatment schedules at time point 0. The response from each well over the 340 hours won't be identical as they are affected by (for example) subtle variations in the cells' phenotype or genotype, variations in initial seeding density, etc.)

The AIC-based model selection is then performed on all eight models including the full model using the training set. By globally fitting the training set (i.e., 262 9L replicates and 211 C6 replicates, to these eight models respectively) we obtain the residual squared sum over the entire training set for each model. The model that returns the lowest AIC score is selected as the most parsimonious. We also compute the Akaike weights *via*:

$$\omega_i = \frac{\exp(\delta_i)}{\sum_{j=1}^{8} \exp(\delta_i)} \quad (12)$$

$$\delta_i = \frac{AIC_i - AIC_{\min}}{AIC_{\min}} \quad (13)$$

where AIC_i is the AIC score of the i^{th} model, and AIC_{min} is the minimum AIC observed among all eight models. These weights are used to compare the models to each other.

2.5 Parameter Calibration

Table 1 lists parameter definitions and the methods we use to obtain these parameters. We previously (10) fit untreated cell data to Eq. (1) to obtain the proliferation rate, k_p, carrying capacity, θ, and the Allee constant, A, for both the 9L and C6 cells. These parameters are assumed constant throughout the present study. The standard deviation of the estimated model parameters is computed *via* "nlparci" in MATLAB, and employed to generate a corresponding distribution *via* "makedist". As the biological definitions of these parameters specify their values must be positive, the parameters are

given a lower bound of zero during calibration. Note that during model calibration, we fit parameters globally in the sense that all curves from the training set, consisting of both the 16 Gy and 20 Gy total dose groups, are fit together since all parameters are independent of dose and initial cell density except for the conversion rate k_{ps}. As k_{ps} is a function of the total dose, we treat k_{ps} at 16 Gy and at 20 Gy as two individual parameters in our calibration process. Distributions of the calibrated parameters are compared between the two cell lines to verify if they are significantly different by the z-test (*via* "ztest" in MATLAB at the 5% significance level).

2.6 Model Validation and Error Analysis

Validation is performed on the AIC selected model using the remaining twenty-five percent of data (i.e., 91 9L curves and 74 C6 curves), which are "unseen" during both the model selection and parameter calibration steps. The forward model has two inputs: initial seeding density (N_0) and dose schedule. We run the model forward using the calibrated parameters as shown in **Figure 5**. The prediction mean and intervals are then computed *via* the MATLAB function "nlpredci" by inputting the calibrated parameters, Jacobian matrix, and residuals determined by the "lsqnonlin" during calibration. Radiation responses are predicted using the same set of parameters regardless of treatment dose schedules or initial density. For example, predicting the effects of the four fractions of 4 Gy on the low confluence group employs the same set of parameters as predicting the effects of the two fractions of 10 Gy on the high confluence group (except for the k_{ps}, which is a constant based on total dose). To quantify the predictive accuracy, we compare the measured and model prediction means using the Pearson correlation coefficient (PCC) and concordance correlation coefficient (CCC).

3 RESULTS

3.1 Image Segmentation and Cell Response Curves

The same image segmentation pipeline from our previous study (10) is used here. When comparing the automatically segmented images to our manually segmented baseline, this segmentation pipeline achieves an average Sørensen–Dice coefficient of 0.79 (i.e., 21% error). See **Figure 3** for an example of a cell response curve that received four fractions of 4 Gy and its corresponding image segmentation.

3.2 Model Selection

Figure 4 summarizes the AIC weights across all eight models. As it has the highest weights for both the 9L and C6 cell lines, Model 3 was selected as the most parsimonious model from the training set and thus will be used for parameter calibration and

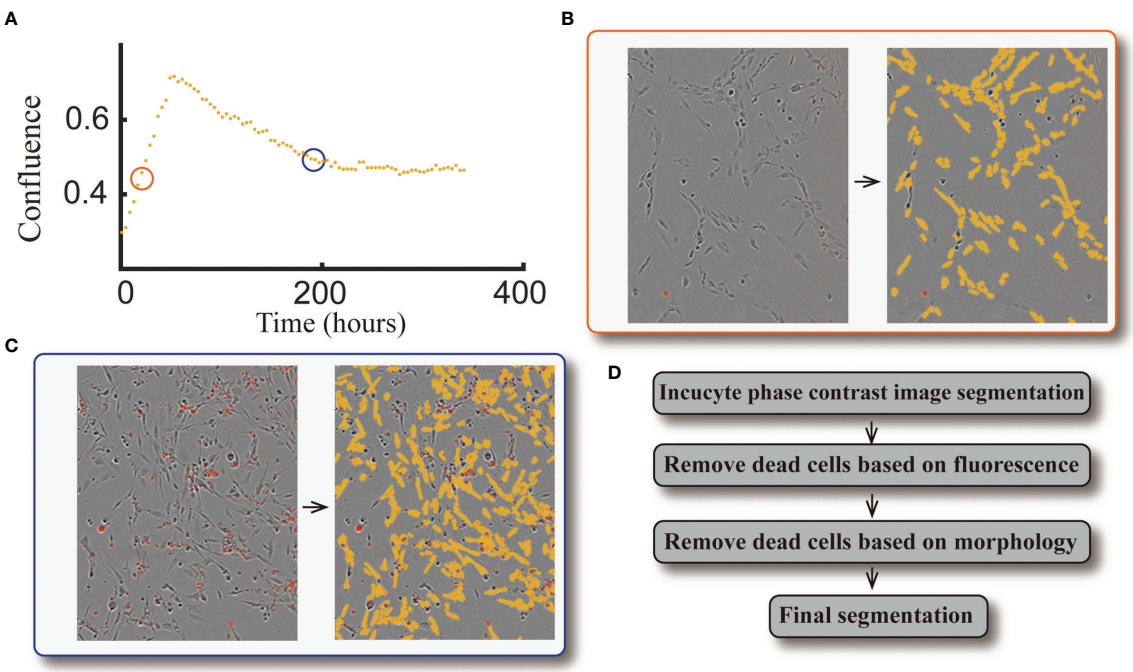


FIGURE 3 | Example of cell response curve and image segmentation. Panel **(A)** shows the cell response curve from one replicate of the 9L cell line treated with four fractions of 4 Gy. The left portion of Panel **(B)** is the raw data at 40 hours [indicated by the orange circle in panel **(A)**] obtained *via* live cell microscopy. The image is presented as phase-contrast with a red fluorescent label indicating dead cells. The corresponding segmentation (highlighted by yellow) is on the right portion of panel **(B)**. The left portion of Panel **(c)** is the raw data at 190 hours [indicated by the blue circle in panel **(A)**], where there are a large number of dead cells compared to the early time point. The corresponding segmentation is presented in the right portion of the panel. Panel **(D)** briefly summarizes our segmentation pipeline; details were provided in ref. (10).

prediction. While the complete formulation is presented in the section 2 of the **Supplementary Materials**, Model 3 combines early apoptosis ($k_{acute,N}$), mitotic catastrophe ($\alpha_{accum,N}$, $k_{accum,D}$, r), and senescence [k_{ps}(16 Gy), k_{ps}(20 Gy)] as given by the following system of equations:

$$\frac{dN_p(t)}{dt} = (k_p - k_{ld}(t, D, N_0)) \cdot \left(\frac{N_p(t) + N_s(t)}{\theta} + A\right) \cdot N(t) \cdot \left(1 - \frac{N_p(t) + N_s(t)}{\theta}\right) - k_{ed}(t, D, N_0) \cdot N_p(t) - k_{ps}(D_{total}) \cdot N_0 \cdot N_p(t) \quad (14)$$

$$k_{ed}(t, D, N_0) = \sum_{i=1}^{\substack{fraction \\ number}} k_{accute,N} \cdot N_0 \cdot f_{DSB}(t^i, D) \quad (15)$$

$$k_{id}(t, D, N_0) = \sum_{1}^{\substack{fraction \\ number}} (\alpha_{accum,N} \cdot N_0 + 1) \cdot k_{accum,D} \cdot t^i \cdot e^{-r \cdot t^i} \quad (16)$$

$$\frac{dN_s(t)}{dt} = k_{ps}(D_{total}) \cdot N_0 \cdot N_p(t) \quad (17)$$

Note that the parameter $k_{acute,D}$ in Eq. (6) is removed and the corresponding expression for early death is rewritten as in Eq. (15). This was done since model selection indicated that $k_{acute,D}$ is not required to characterize our data. Consequently, we removed $k_{acute,D}$ and modified the notation of $\alpha_{acute,N}$ in Eq. (6) to $k_{acute,N}$ (in units of hr^{-1}) in Eq. (15) to represent the death rate.

From the AIC score, model 3 [i.e., Eqs. (14) – (17)] is 1.18 times more likely to be the best model than model 4 'no early death' model, 2.61 times more likely than model 7 'no late death model', and 1.82 times more likely than model 8 'no senescence' single species model. These results indicate the importance of accumulation effects (i.e., accumulating DNA misrepair, mitotic catastrophe and gradual conversion to senescence) over acute effects (i.e., early apoptosis) to quantify the time courses of 9L and C6 radiation response. Note that the goal of the model selection process is to identify the most parsimonious model to describe our time-resolved data, rather than the model that includes the most biology. As for the 9L and C6 cell lines, most cells die due to late mitotic catastrophe while early apoptosis only kills a minority of cells. Thus, removing the radiation dose effect from the early death term as in model 3 does not harm the model's ability to characterize the data.

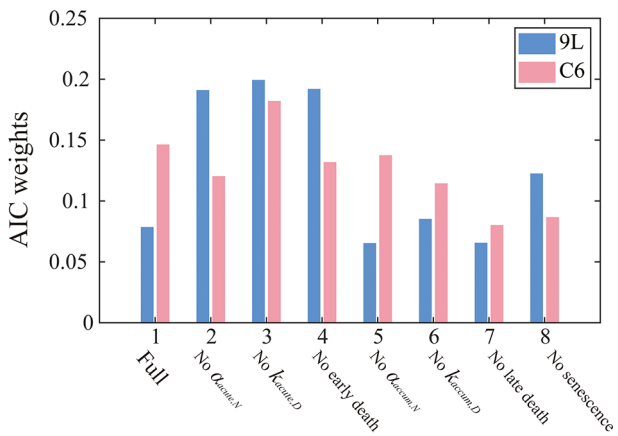


FIGURE 4 | AIC weights for each model models. The label, "Full" on the horizontal axis indicates that all model components are included [i.e., Eq. (1) – (10)], while the other labels indicate what portion of the Full model was removed in that particular reduced model. AIC weights can be interpreted as the probability that a particular model is preferred for modeling the 9L (blue) or C6 (red) cell line. In particular, Model 3 is most frequently selected by the AIC for both the 9L and C6 cell lines. Models without accumulation effects (i.e., model 5-8) generally perform worse than models without early effects (model 2-4), indicating the importance of incorporating the effects of late cell killing over acute apoptosis.

3.3 Parameter Calibration

The AIC selected model has six parameters $k_{acute,N}$, $\alpha_{accum,N}$, $k_{accum,D}$, r, k_{ps}(16 Gy), and k_{ps}(20 Gy). The top panel in **Figure 5** shows the calibrated parameters for the 9L cells, while the bottom panel shows the calibrated parameters for the C6 cells. Note that the model allows to quantify the degree to which the C6 cell line is more radiation sensitive than the 9L cell line, as has been previously reported (41). In particular, the early death rates $k_{acute,N}$ (p value = 9.5e-23), late death rates $k_{accum,D}$ (p value = 8.1e-21) and conversion rate to senescent component k_{ps} (p value = 1.6e-20 for 16 Gy and 0.0 for 20 Gy) of the C6 are all significantly larger than 9L *via* the z-test. Both the 9L and C6 exhibit a similar radiation efficacy decay r (p value = 0.44, i.e., no significant difference *via* z-test), suggesting a similar duration of radiation cell killing persisting on both cell lines. The proliferation rate, carrying capacity, and Allee effect parameter values, as well as the number of training curves for calibration, are provided in the section 1 of the **Supplementary Material**.

3.4 Model Validation and Error Analysis

We use the validation group (25% of our total data, 91 9L curves and 74 C6 curves) to evaluate the predictive accuracy of our model. **Figure 6** presents examples of the model validation results with each column representing one initial confluence of a specific cell line (e.g., the first column shows a low initial seeding confluence of the 9L cell line). Each row shows a different fractionation schedule (e.g., the first row shows cells receiving four fractions of 4 Gy radiation). The error bar on the measurement (labeled by blue) is based on the image segmentation error (21%) from our previous study (10), as the same segmentation pipeline is used. The prediction error (labeled by red) is computed from MATLAB function 'nlpredci', which computes the prediction interval *via* the Delta method based on the Jacobian matrix.

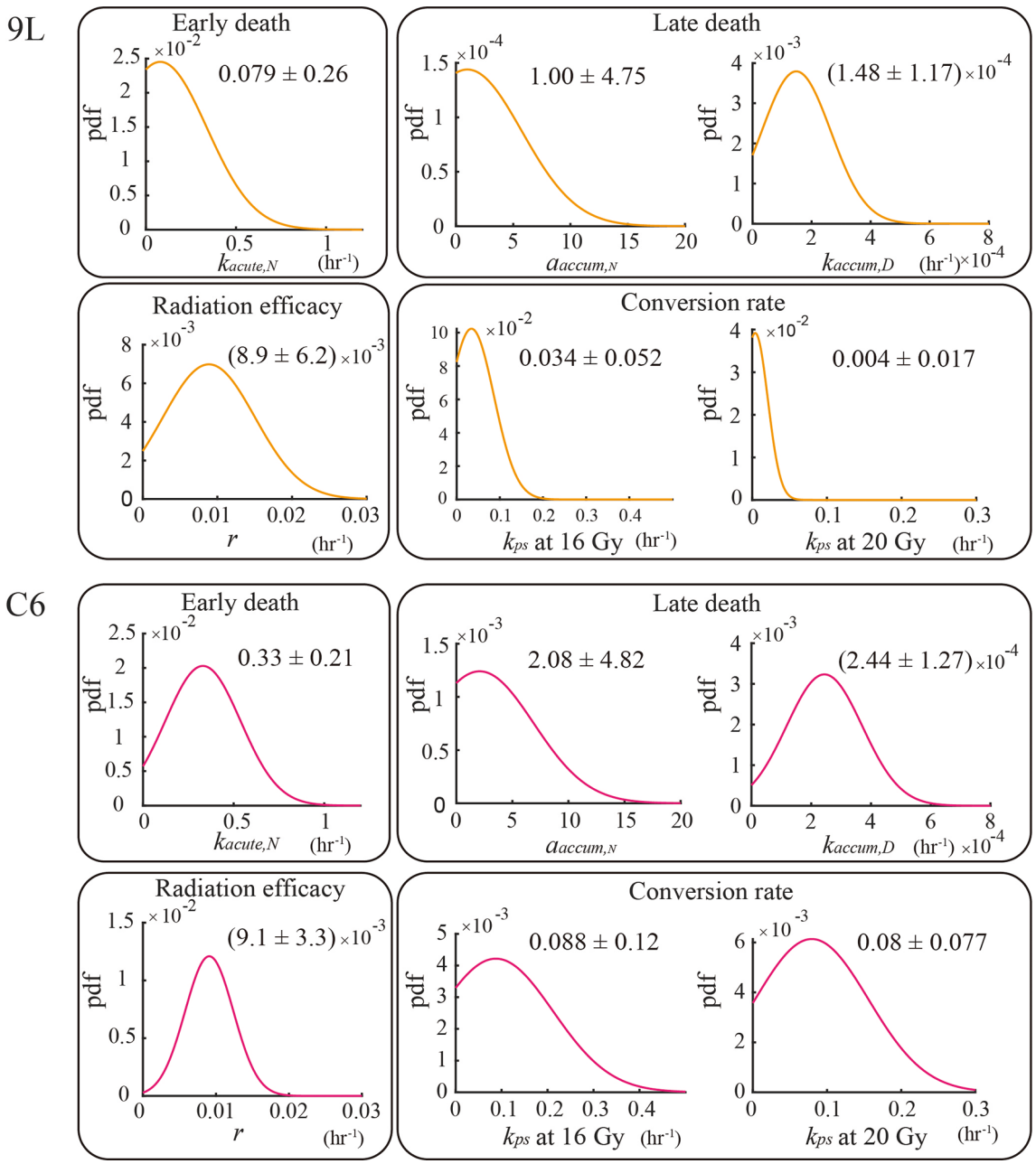


FIGURE 5 | Parameter calibration results. All parameters are fit globally using the training set and are independent of initial seeding density or dose schedules (e.g., the four fractions of 4 Gy curves share the same set of parameters as the two fractions of 10 Gy curves for both the 9L and C6; the one exception is for k_{ps}, where we treat k_{ps} (16 Gy) and k_{ps} (20 Gy) as two individual parameters. The biological interpretation of the parameters requires that these parameters must take on a value great than or equal to 0; thus, the lower bound of the parameters is set to zero during calibration.

For the 9L group receiving a total dose of 16 Gy, the average PCC and CCC between the predicted and the measured data are 0.92 ± 0.009 (average ± standard error) and 0.78 ± 0.014, respectively. For the 9L group receiving a total of 20 Gy, the PCC and CCC are 0.98 ± 0.001 and 0.96 ± 0.002, respectively. For the C6 group receiving a total of 16 Gy, the PCC and CCC are 0.89 ± 0.011 and 0.77 ± 0.020, respectively. Finally, for the C6 group receiving a total of 20 Gy, the PCC and CCC are 0.90 ± 0.004 and 0.73 ± 0.014, respectively. Details of PCC and CCC values for each treatment condition is provided in **Table 2**. Overall, the accuracy of prediction is superior for the 9L cell line than the C6 cell line. This is at least partially due to the heterogeneous radiation response observed across the C6 replicates; we return to this important point in Discussion section.

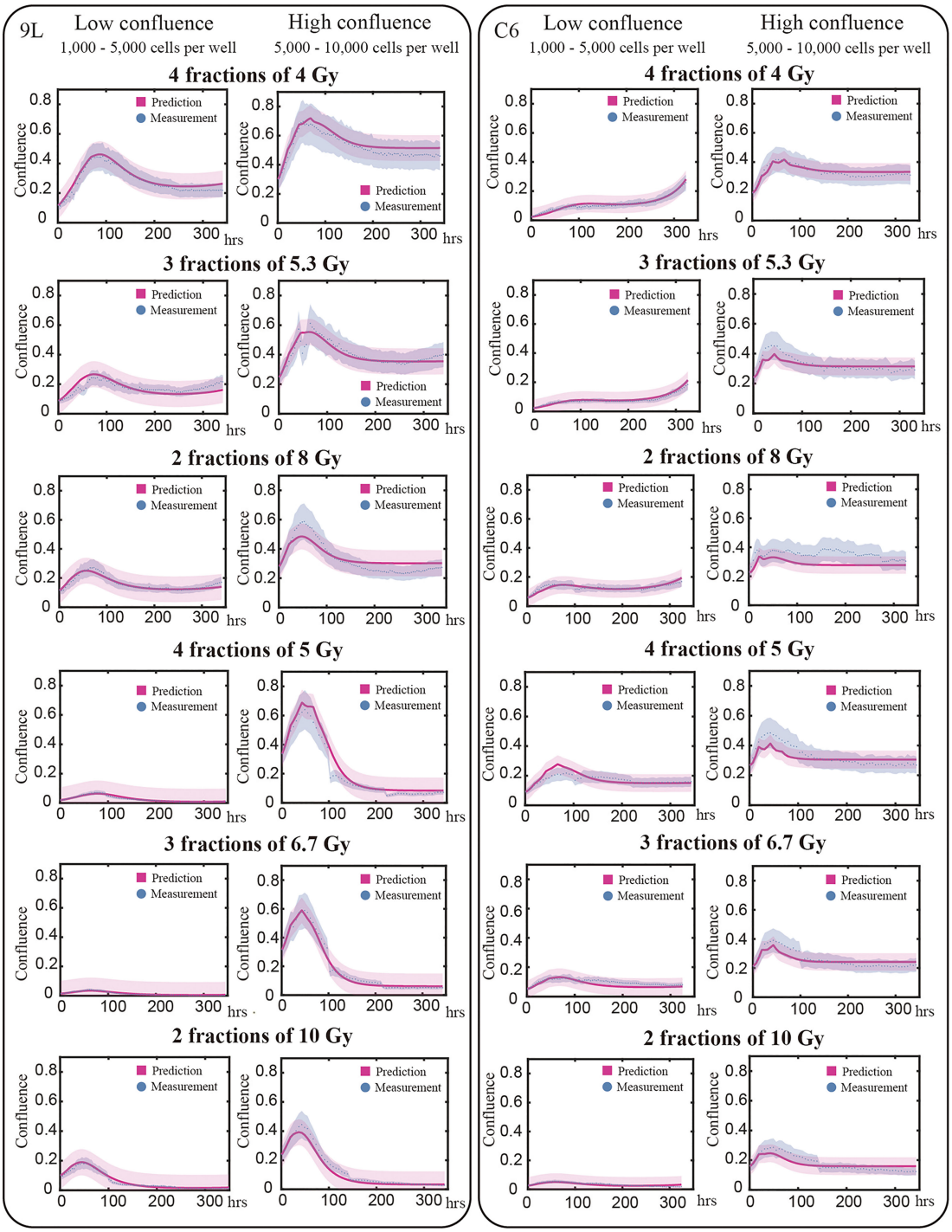


FIGURE 6 | Model validation. The prediction interval (labeled by red) and measured microscopy data (labeled blue dots) are plot for representative examples of each initial and treatment condition. Previously we determined our average segmentation error was 21% using the Sørensen–Dice coefficient (10). This segmentation error is indicated by the blue intervals. Each row in this figure shows different treatment schedules labeled on the left; for example, the first row shows cells treated with four fractions of 4 Gy. Each column stands for the initial confluence for a specific cell line; for example, the first column represents low initial seeding density for 9L cell line. Predictions are made globally for each of the cell lines; that is, predictions for the 9L cells with an initial low confluence receiving four fractions of 4 Gy are made using the same set of parameters as the 9L cells with an initial high confluence receiving two fractions of 10 Gy. Given the initial confluence and treatment schedule as the two inputs, our model makes accurate predictions across a wide range of initial conditions. However, the model is not perfect as the prediction typically undershoots the first peak while overshooting the tail for C6 cell line; an important point we return to in the Discussion section.

TABLE 2 | Error between model prediction and measurement.

	Treatment schedule	9L		C6	
		PCC	CCC	PCC	CCC
16 Gy	4 fractions of 4 Gy	0.96 ± 0.003	0.84 ± 0.011	0.91 ± 0.007	0.85 ± 0.012
	3 fractions of 5.3 Gy	0.85 ± 0.015	0.71 ± 0.019	0.89 ± 0.008	0.76 ± 0.014
	2 fractions of 8 Gy	0.96 ± 0.001	0.79 ± 0.003	0.84 ± 0.016	0.65 ± 0.029
20 Gy	4 fractions of 5 Gy	0.97 ± 0.002	0.96 ± 0.002	0.90 ± 0.005	0.76 ± 0.011
	3 fractions of 6.7 Gy	0.98 ± 0.001	0.93 ± 0.001	0.88 ± 0.004	0.64 ± 0.016
	2 fractions of 10 Gy	0.98 ± 0.001	0.95 ± 0.003	0.91 ± 0.004	0.77 ± 0.011

Mean ± standard error.

4 DISCUSSION

We have proposed an experimental-computational system that includes a mathematical model that characterizes the dynamic cell response of the 9L and C6 glioma receiving fractionated radiation. Model parameters are calibrated using a training set of various initial seeding densities and dose schedules over the course of two weeks. Next, the calibrated parameters are applied to a validation set to predict response to radiation therapy. Each term in our model has an explicit or implicit (i.e., $\alpha_{accum,N}$, k_{ps} (16 Gy), k_{ps}(20 Gy) are indirectly related to cell cycles) biological definition. For example, early apoptosis and late mitotic catastrophe are captured by the parameters $k_{acute,N}$ and $k_{accum,D}$, r, respectively. This approach has several advantages over current models as it explicitly includes the temporal dynamics of several biological mechanisms related to the response of cells to fractionated radiation.

Mathematically, the LQ model describes the fraction of survival cells at experimental endpoints given a specific radiation dose, and therefore is not designed for calibrating with time-resolved data. Though there are mathematical models that account for temporal dynamics by embedding the LQ term into the death rates (42), the true death rates are unlikely to obey the LQ relation throughout the whole time course as it ignores time dependent phenomena such as early apoptosis and late mitotic catastrophe. Time-dependent radiobiologic mechanisms are often ignored in these modeling studies, a limitation possibly due to the historical difficulty of accessing radiation response data with high temporal resolution. However, recent advances in imaging techniques can now provide the requisite, high temporal resolution data (43–45) appropriate for model calibration. When we mathematically characterize these time series data, a dynamic model is required, since the measurement interval (e.g., hours) is much shorter than the length of the experiment when the surviving fraction would be determined *via* the LQ model. To the best of our knowledge, this study is the first mechanism-based fractionated radiation model verified by the cell experimental data at high temporal-resolution. We have constructed this approach by building on previous time-resolved dynamic radiation models that either calibrated to a single dose of radiation (46), or to data collected every several days (23) or weeks (47). We hope that the current effort can provide motivation for focusing on factors that describe the dynamics of radiobiology as a function of time, thereby potentially providing a new way to guide and optimize radiation dose scheduling.

As each term in the selected model system [i.e., Eqs. (14) – (17)] is based on an underlying mechanism, the calibrated parameters summarized in **Figure 5** are easily interpreted in light of the underlying biology. Though the specific values of parameters may differ for each cell line (a not unexpected observation), the trend observed (i.e., radiosensitive cells exhibit significantly larger death rates) may carry over to other cell lines since the proposed model and the parameters are based on biological mechanisms common to a wide range of cell lines (48). Thus, the model summarized by Eqs. (14) – (17) is likely applicable to other cell lines for which radiation treatment is of interest, thereby having a significant impact beyond gliomas.

Despite the advantages, there are several opportunities for improvement in both the experimental and modeling components of the study. For example, note that in **Figure 6** some curves suffer a discontinuous jump (e.g., at approximately 110 hours in the "4 fractions of 5 Gy" high confluence group). This is due to the cell loss when we refresh the media. While we have implemented a method to change the media as delicately as possible (as described in section 2.1.2), we are still aspirating an unknown number of live cells at each media change. Potentially more significant, though, is how we handle the senescent component. Since we did not have a method in place that allows us to directly measure this component over time, we are forced to infer its existence and behavior *via* parameter calibration. Clearly, incorporating longitudinal measurements of the fractions of senescent cells is required to generalize the model to other dosing schemes.

Areas for improvement on the modeling side include developing a more rigorous linking between radiation dose and the rate constants. For example, note that Eqs. (6) and (8) have a different form from our previous single-dose study (10). The main reason for the changes is due to the dose range in this study (4 - 10 Gy per fraction) compared to the previous study (a single fraction of 2 – 16 Gy). In our previous single-fraction study we observed a saturation effect in the death rates; i.e., the death rates do not significantly increase above a certain dose threshold. This is because cell death does not happen instantaneously; i.e., the cell death pathways require time to be executed. Thus, we use Michaelis–Menten equations to characterize the observed saturation effect. However, in the present study, we use a narrower range of radiation dose and we did not observe the saturation in cell killing within this dose range. Thus, we use a simple linear relation for the acute death, $k_{acute}(D,N_0)$, accumulation death, $k_{accum}(D,N_0)$, and a constant radiation

efficacy r, versus radiation doses. Consequently, the death rates $k_{acute}(D,N_0)$, $k_{accum}(D,N_0)$, and r most likely require future modifications when applied to dose schedules involving larger doses. We also note that, compared to the previous single-dose study [i.e., Eq. (6) in (10)], we removed the parameter T in Eq. (7). This parameter assumes there is a time delay, T, before the accumulation of misrepairs can start. Our previous study established its value between the first 20-40 hours after radiation for the 9L cell line, and 0 hours for the C6 cell line [consistent with the established radiosensitivity of the C6 cell line (41)]. In current fractionated dosing scheme, as we extend the experiment time to approximately 340 hours, the value of T is much less than the length of the experiment. Thus, it is reasonable to remove the parameter T for simplicity. Another limitation is that Eqs. (14) – (17) do not (of course) account for all of the most prominent radiobiological processes. For example, we assume a fixed conversion rate between the proliferative and senescent components—a parameter that is likely to change with time, dose per fraction, and fraction numbers, and therefore would require a function of its own to describe its temporal dynamics. Indeed, this may explain the undershoot peak or overshoot tail in **Figure 6**. A second area for improvement in Eqs. (14) – (17) is the explicit incorporation of phenotypic or genomic heterogeneity across the population, which almost certainly affects the early ($k_{acute,N}$) and late ($k_{accum,D}$) death effects between different replicates. This explains why the correlation coefficients in **Table 2** indicate the model performs worse under certain dose schedules for the C6 cell line. See section 3 of the **Supplementary Materials** for an example of the heterogeneous response we observe in two replicates of C6 cells treated with the same three fractions of 6.7 Gy (note this is the group with the lowest PCC and CCC value). Accounting for such heterogeneity to improve the predictive accuracy of the model will be the focus of future study. As presented, the current model has limited clinical application because it is formulated for describing the 2D dynamics of cells in a dish which is, of course, very different than the *in vivo* situation. However, the concept of using time-resolved data to calibrate a biologically-based, mathematical model to make patient specific predictions is highly translatable and, indeed, something we have investigated at length in both the *in vivo* pre-clinical (49–53) and clinical (24, 54–58) settings.

5 CONCLUSION

We have extended our previous single-dose model to account fractionated treatment, and successfully validated the resulting model with *in vitro* experimental microscopy data. This study demonstrates a promising experimental-mathematical approach based on radiobiology mechanisms that can accurately predict the temporal dynamics of the response of glioma cells to radiation. Future efforts include linking the model to our tissue scale formalism for predicting response in patients (23), and employing the methods of optimal control theory to optimize treatment outcomes in pre-clinical murine studies (59).

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

JL collected data, performed modeling analysis and wrote the manuscript. JY helped with the experiments. DH and TY conceptualized and supervised the study, reviewed and edited the manuscript. All authors have read and approved the final manuscript.

FUNDING

This work is supported by the National Cancer Institute *via* R01CA240589, R01 CA235800, U01CA253540, and R01CA186193, and the Cancer Prevention and Research Institute of Texas (CPRIT) *via* RR160005. TY is a CPRIT Scholar of Cancer Research.

ACKNOWLEDGMENTS

We thank Ms. Tessa Davis for expert advice and guidance on setting up the time-resolved microscopy studies.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.811415/full#supplementary-material

REFERENCES

1. Omuro A, DeAngelis LM. Glioblastoma and Other Malignant Gliomas: A Clinical Review. *JAMA* (2013) 310(17):1842–50. doi: 10.1001/jama.2013.280319
2. Tan AC, Ashley DM, López GY, Malinzak M, Friedman HS, Khasraw M. Management of Glioblastoma: State of the Art and Future Directions. *CA Cancer J Clin* (2020) 70(4):299–312. doi: 10.3322/caac.21613
3. Alfonso JCL, Berk L. Modeling the Effect of Intratumoral Heterogeneity of Radiosensitivity on Tumor Response Over the Course of Fractionated Radiation Therapy. *Radiat Oncol* (2019) 14(1):88. doi: 10.1186/s13014-019-1288-y
4. Hormuth DA, Jarrett AM, Lorenzo G, Lima EA, Wu C, Chung C, et al. Math, Magnets, and Medicine: Enabling Personalized Oncology. *Expert Rev Precis Med Drug Dev* (2021) 6(2):79–81. doi: 10.1080/23808993.2021.1878023

5. Jones L, Hoban P, Metcalfe P. The Use of the Linear Quadratic Model in Radiotherapy: A Review. *Australas Phys Eng Sci Med* (2001) 24(3):132–46. doi: 10.1007/BF03178355
6. Brenner DJ. The Linear-Quadratic Model is an Appropriate Methodology for Determining Isoeffective Doses at Large Doses Per Fraction. *Semin Radiat Oncol* (2008) 18(4):234–9. doi: 10.1016/j.semradonc.2008.04.004
7. McMahon SJ. The Linear Quadratic Model: Usage, Interpretation and Challenges. *Phys Med Biol* (2018) 64(1):01TR01. doi: 10.1088/1361-6560/aaf26a
8. Scully R, Panday A, Elango R, Willis NA. DNA Double-Strand Break Repair-Pathway Choice in Somatic Mammalian Cells. *Nat Rev Mol Cell Biol* (2019) vol20(11):698–714. doi: 10.1038/s41580-019-0152-0
9. Eriksson D, Stigbrand T. Radiation-Induced Cell Death Mechanisms. *Tumour Biol J Int Soc Oncodevelopmental Biol Med* (2010) 31(4):363–72. doi: 10.1007/s13277-010-0042-8
10. Liu J, Hormuth DA, Davis T, Yang J, McKenna MT, Jarrett AM, et al. A Time-Resolved Experimental-Mathematical Model for Predicting the Response of Glioma Cells to Single-Dose Radiation Therapy. *Integr Biol Quant Biosci Nano Macro* (2021) 13(7):167–83. doi: 10.1093/intbio/zyab010
11. Withers HR. The Four R's of Radiotherapy. In: JT Lett, H Adler, editors. *Advances in Radiation Biology*, vol. 5 . Netherlands:Elsevier (1975). p. 241–71. doi: 10.1016/B978-0-12-035405-4.50012-8
12. Hubenak JR, Zhang Q, Branch CD, Kronowitz SJ. Mechanisms of Injury to Normal Tissue After Radiotherapy: A Review. *Plast Reconstr Surg* (2014) 133 (1):49e–56e. doi: 10.1097/01.prs.0000440818.23647.0b
13. Rübe CE, Fricke A, Wendorf J, Stützel A, Kühne M, Ong MF, et al. Accumulation of DNA Double-Strand Breaks in Normal Tissues After Fractionated Irradiation. *Int J Radiat Oncol Biol Phys* (2010) 76(4):1206–13. doi: 10.1016/j.ijrobp.2009.10.009
14. Bortfeld T, Ramakrishnan J, Tsitsiklis JN, Unkelbach J. Optimization of Radiation Therapy Fractionation Schedules in the Presence of Tumor Repopulation. *Inf J Comput* (2015) 27(4):788–803. doi: 10.1287/ijoc.2015.0659
15. Withers HR. Cell Cycle Redistribution as a Factor in Multifraction Irradiation. *Radiology* (1975) 114(1):199–202. doi: 10.1148/114.1.199
16. Grimes DR, Partridge M. A Mechanistic Investigation of the Oxygen Fixation Hypothesis and Oxygen Enhancement Ratio. *Biomed Phys Eng Express* (2015) 1(4):45209. doi: 10.1088/2057-1976/1/4/045209
17. Horsman MR, Wouters BG, Joiner MC, Overgaard J. The Oxygen Effect and Fractionated Radiotherapy. In: *Basic Clinical Radiobiology*, *4th ed*. Boca Raton, FL:CRC Press (2009).
18. Wenzl T, Wilkens JJ. Theoretical Analysis of the Dose Dependence of the Oxygen Enhancement Ratio and its Relevance for Clinical Applications. *Radiat Oncol* (2011) 6(1):171. doi: 10.1186/1748-717X-6-171
19. Jones B, Dale R. The Evolution of Practical Radiobiological Modelling. *Br J Radiol* (2019) 92(1093):20180097. doi: 10.1259/bjr.20180097
20. Jeong J, Shoghi KI, Deasy JO. Modelling the Interplay Between Hypoxia and Proliferation in Radiotherapy Tumour Response. *Phys Med Biol* (2013) 58 (14):4897–919. doi: 10.1088/0031-9155/58/14/4897
21. Powathil G, Kohandel M, Sivaloganathan S, Oza A, Milosevic M. Mathematical Modeling of Brain Tumors: Effects of Radiotherapy and Chemotherapy. *Phys Med Biol* (2007) 52(11):3291–306. doi: 10.1088/0031-9155/52/11/023
22. Belfatto A, Jereczek-Fossa BA, Baroni G, Cerveri P. Model-Supported Radiotherapy Personalization: *In Silico* Test of Hyper- and Hypo-Fractionation Effects. *Front Physiol* (2018) 9:1445(1445). doi: 10.3389/fphys.2018.01445
23. Hormuth DA, Jarrett AM, Davis T, Yankeelov TE. Towards an Image-Informed Mathematical Model of *In Vivo* Response to Fractionated Radiation Therapy. *Cancers* (2021) 13(8):1765. doi: 10.3390/cancers13081765
24. Hormuth DA, Al Feghali KA, Elliott AM, Yankeelov TE, Chung C. Image-Based Personalization of Computational Models for Predicting Response of High-Grade Glioma to Chemoradiation. *Sci Rep* (2021) 11(1):8520. doi: 10.1038/s41598-021-87887-4
25. Brüningk S, Powathil G, Ziegenhein P, Ijaz J, Rivens I, Nill S, et al. Combining Radiation With Hyperthermia: A Multiscale Model Informed by *In Vitro* Experiments. *J R Soc Interface* (2018) 15(138):20170681. doi: 10.1098/rsif.2017.0681
26. Giakoumettis D, Kritis A, Foroglou N. C6 Cell Line: The Gold Standard in Glioma Research. *Hippokratia* (2018) 22(3):105–12.
27. Barth RF, Kaur B. Rat Brain Tumor Models in Experimental Neuro-Oncology: The C6, 9l, T9, RG2, F98, BT4C, RT-2 and CNS-1 Gliomas. *J Neurooncol* (2009) 94(3):299–312. doi: 10.1007/s11060-009-9875-7
28. Mah L-J, El-Osta A, Karagiannis TC. γh2ax: A Sensitive Molecular Marker of DNA Damage and Repair. *Leukemia* (2010) 24(4):679–86. doi: 10.1038/leu.2010.6
29. Mariotti LG, Pirovano G, Savage KI, Ghita M, Ottolenghi A, Prise KM, et al. Use of the γ-H2AX Assay to Investigate DNA Repair Dynamics Following Multiple Radiation Exposures. *PloS One* (2013) 8(11):e79541. doi: 10.1371/journal.pone.0079541
30. Neufeld Z, von Witt W, Lakatos D, Wang J, Hegedus B, Czirok A. The Role of Allee Effect in Modelling Post Resection Recurrence of Glioblastoma. *PloS Comput Biol* (2017) 13(11):e1005818. doi: 10.1371/journal.pcbi.1005818
31. Nowsheen S, Yang ES. The Intersection Between DNA Damage Response and Cell Death Pathways. *Exp Oncol* (2012) 34(3):243–54.
32. Chang DS, Lasley FD, Das IJ, Mendonca MS, Dynlacht JR. Cell Death and Survival Assays. In: DS Chang, FD Lasley, IJ Das, MS Mendonca, JR Dynlacht, editors. *Basic Radiotherapy Physics and Biology*. Cham: Springer International Publishing (2014). p. 211–9. doi: 10.1007/978-3-319-06841-1_20
33. Wouters BG. Cell Death After Irradiation: How, When and Why Cells Die. In: *Basic Clinical Radiobiology*, *4th ed.* Boca Raton, FL:CRC Press (2009).
34. Forrester HB, Vidair CA, Albright N, Ling CC, Dewey WC. Using Computerized Video Time Lapse for Quantifying Cell Death of X-Irradiated Rat Embryo Cells Transfected With C-Myc or C-Ha-Ras. *Cancer Res* (1999) 59(4):931–9.
35. Pawlik TM, Keyomarsi K. Role of Cell Cycle in Mediating Sensitivity to Radiotherapy. *Int J Radiat Oncol Biol Phys* (2004) 59(4):928–42. doi: 10.1016/j.ijrobp.2004.03.005
36. Richard M, Kirkby KJ, Webb RP, Kirkby NF. A Mathematical Model of Response of Cells to Radiation. *Nucl Instrum Methods Phys Res Sect B Beam Interact Mater At* (2007) 255(1):18–22. doi: 10.1016/j.nimb.2006.11.077
37. Nickoloff JA, Sharma N, Taylor L. Clustered DNA Double-Strand Breaks: Biological Effects and Relevance to Cancer Radiotherapy. *Genes* (2020) 11 (1):99. doi: 10.3390/genes11010099
38. Chen Z, Cao K, Xia Y, Li Y, Hou Y, Wang L, et al. Cellular Senescence in Ionizing Radiation (Review). *Oncol Rep* (2019) 42(3):883–94. doi: 10.3892/or.2019.7209
39. Wang B. Analyzing Cell Cycle Checkpoints in Response to Ionizing Radiation in Mammalian Cells. *Methods Mol Biol Clifton NJ* (2014) 1170:313–20. doi: 10.1007/978-1-4939-0888-2_15
40. Sakamoto Y, Ishiguro M, Kitagawa G. *Akaike Information Criterion Statistics* (1986). Springer Netherlands. Available at: https://www.springer.com/gp/book/9789027722539 (Accessed Accessed: Sep. 21, 2021).
41. Bencokova Z, Pauron L, Devic C, Joubert A, Gastaldo J, Massart C, et al. Molecular and Cellular Response of the Most Extensively Used Rodent Glioma Models to Radiation and/or Cisplatin. *J Neurooncol* (2008) 86 (1):13–21. doi: 10.1007/s11060-007-9433-0
42. Geng C, Paganetti H, Grassberger C. Prediction of Treatment Response for Combined Chemo- and Radiation Therapy for Non-Small Cell Lung Cancer Patients Using a Bio-Mathematical Model. *Sci Rep* (2017) 7(1):13542. doi: 10.1038/s41598-017-13646-z
43. Yankeelov TE, Atuegwu N, Hormuth D, Weis JA, Barnes SL, Miga MI, et al. Clinically Relevant Modeling of Tumor Growth and Treatment Response. *Sci Transl Med* (2013) 5(187):187ps9–9. doi: 10.1126/scitranslmed.3005686
44. Yankeelov TE, Quaranta V, Evans KJ, Rericha EC. Toward a Science of Tumor Forecasting for Clinical Oncology. *Cancer Res* (2015) 75(6):918–23. doi: 10.1158/0008-5472.CAN-14-2233
45. Kazerouni AS, Gadde M, Gardner A, Hormuth DA II, Jarrett AM, Johnson KE, et al. Integrating Quantitative Assays With Biologically Based Mathematical Modeling for Predictive Oncology. *iScience* (2020) 23 (12):101807. doi: 10.1016/j.isci.2020.101807
46. Brüningk SC, Ziegenhein P, Rivens I, Oelfke U, ter Haar G. A Cellular Automaton Model for Spheroid Response to Radiation and Hyperthermia Treatments. *Sci Rep* (2019) 9(1):17674. doi: 10.1038/s41598-019-54117-x
47. Prokopiou S, Moros EG, Poleszczuk J, Caudell J, Torres-Roca JF, Latifi K, et al. A Proliferation Saturation Index to Predict Radiation Response and Personalize Radiotherapy Fractionation. *Radiat Oncol* (2015) 10(1):159. doi: 10.1186/s13014-015-0465-x

48. Maier P, Hartmann L, Wenz F, Herskind C. Cellular Pathways in Response to Ionizing Radiation and Their Targetability for Tumor Radiosensitization. *Int J Mol Sci* (2016) 17(1):102. doi: 10.3390/ijms17010102
49. Hormuth DA, Weis JA, Barnes SL, Miga MI, Quaranta V, Yankeelov TE. Biophysical Modeling of *In Vivo* Glioma Response After Whole-Brain Radiation Therapy in a Murine Model of Brain Cancer. *Int J Radiat Oncol Biol Phys* (2018) 100(5):1270–9. doi: 10.1016/j.ijrobp.2017.12.004
50. Hormuth DA, Jarrett AM, Yankeelov TE. Forecasting Tumor and Vasculature Response Dynamics to Radiation Therapy *via* Image Based Mathematical Modeling. *Radiat Oncol* (2020) 15(1):4. doi: 10.1186/s13014-019-1446-2
51. Hormuth DA, Jarrett AM, Feng X, Yankeelov TE. Calibrating a Predictive Model of Tumor Growth and Angiogenesis With Quantitative MRI. *Ann Biomed Eng* (2019) 47(7):1539–51. doi: 10.1007/s10439-019-02262-9
52. Hormuth DA, Weis JA, Barnes SL, Miga MI, Rericha EC, Quaranta V, et al. A Mechanically Coupled Reaction-Diffusion Model That Incorporates Intra-Tumoural Heterogeneity to Predict *In Vivo* Glioma Growth. *J R Soc Interface* (2017) 14(128):20161010. doi: 10.1098/rsif.2016.1010
53. Hormuth DA II, Weis JA, Barnes SL, Miga MI, Rericha EC, Quaranta V, et al. Predicting *In Vivo* Glioma Growth With the Reaction Diffusion Equation Constrained by Quantitative Magnetic Resonance Imaging Data. *Phys Biol* (2015) 12(4):46006. doi: 10.1088/1478-3975/12/4/046006
54. Weis JA, Miga MI, Arlinghaus LR, Li X, Abramson JV, Chakravarthy AB, et al. Predicting the Response of Breast Cancer to Neoadjuvant Therapy Using a Mechanically Coupled Reaction-Diffusion Model. *Cancer Res* (2015) 75 (22):4697–707. doi: 10.1158/0008-5472.CAN-14-2945
55. Jarrett AM, Hormuth DA, Barnes SL, Feng X, Huang W, Yankeelov TE. Incorporating Drug Delivery Into an Imaging-Driven, Mechanics-Coupled Reaction Diffusion Model for Predicting the Response of Breast Cancer to Neoadjuvant Chemotherapy: Theory and Preliminary Clinical Results. *Phys Med Biol* (2018) 63(10):105015. doi: 10.1088/1361-6560/aac040
56. Jarrett AM, Hormuth DA II, Wu C, Kazerouni AS, Ekrut DA, Virostko J, et al. Evaluating Patient-Specific Neoadjuvant Regimens for Breast Cancer *via* a Mathematical Model Constrained by Quantitative Magnetic Resonance Imaging Data. *Neoplasia N Y N* (2020) 22(12):820–30. doi: 10.1016/j.neo.2020.10.011
57. Jarrett AM, Hormuth DA, Adhikarla V, Sahoo P, Abler D, Tumyan L, et al. Towards Integration of 64Cu-DOTA-Trastuzumab PET-CT and MRI With Mathematical Modeling to Predict Response to Neoadjuvant Therapy in HER2 + breast Cancer. *Sci Rep* (2020) 10(1):20518. doi: 10.1038/s41598-020-77397-0
58. Jarrett AM, Kazerouni AS, Wu C, Virostko J, Sorace AG, DiCarlo JC, et al. Quantitative Magnetic Resonance Imaging and Tumor Forecasting of Breast Cancer Patients in the Community Setting. *Nat Protoc* (2021) 16(11):5309–38. doi: 10.1038/s41596-021-00617-y
59. Jarrett AM, Faghihi D, Hormuth DA, Lima EA, Virostko J, Biros G, et al. Optimal Control Theory for Personalized Therapeutic Regimens in Oncology: Background, History, Challenges, and Opportunities. *J Clin Med* (2020) 9 (5):1314. doi: 10.3390/jcm9051314

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: 03 March 2022
doi: 10.3389/fonc.2022.844197

Classification of Gliomas and Germinomas of the Basal Ganglia by Transfer Learning

Ningrong Ye[1,2†], Qi Yang[1,2†], Ziyan Chen[1,2], Chubei Teng[1,2], Peikun Liu[1,2], Xi Liu[1,2], Yi Xiong[1,2], Xuelei Lin[1,2], Shouwei Li[3] and Xuejun Li[2*]*

[1] Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha, China, [2] Hunan International Scientific and Technological Cooperation Base of Brain Tumor Research, Xiangya Hospital, Central South University, Changsha, China, [3] Department of Neurosurgery, Sanbo Brain Hospital, Capital Medical University, Beijing, China

OPEN ACCESS

Edited by:
Jia Wu,
University of Texas MD Anderson Cancer Center, United States

Reviewed by:
Xudong Xue,
Hubei Cancer Hospital, China
Long Xie,
University of Pennsylvania, United States
Youxiang Zhu,
University of Massachusetts Boston, United States
Ze-Hong Yang,
Sun Yat-sen University, China

***Correspondence:**
Xuejun Li
lxjneuro@csu.edu.cn
Shouwei Li
15011339604@163.com

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

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: *27 December 2021*
Accepted: *11 February 2022*
Published: *03 March 2022*

Citation:
Ye N, Yang Q, Chen Z, Teng C, Liu P, Liu X, Xiong Y, Lin X, Li S and Li X (2022) Classification of Gliomas and Germinomas of the Basal Ganglia by Transfer Learning. Front. Oncol. 12:844197. doi: 10.3389/fonc.2022.844197

Background: Germ cell tumors (GCTs) are neoplasms derived from reproductive cells, mostly occurring in children and adolescents at 10 to 19 years of age. Intracranial GCTs are classified histologically into germinomas and non-germinomatous germ cell tumors. Germinomas of the basal ganglia are difficult to distinguish based on symptoms or routine MRI images from gliomas, even for experienced neurosurgeons or radiologists. Meanwhile, intracranial germinoma has a lower incidence rate than glioma in children and adults. Therefore, we established a model based on pre-trained ResNet18 with transfer learning to better identify germinomas of the basal ganglia.

Methods: This retrospective study enrolled 73 patients diagnosed with germinoma or glioma of the basal ganglia. Brain lesions were manually segmented based on both T1C and T2 FLAIR sequences. The T1C sequence was used to build the tumor classification model. A 2D convolutional architecture and transfer learning were implemented. ResNet18 from ImageNet was retrained on the MRI images of our cohort. Class activation mapping was applied for the model visualization.

Results: The model was trained using five-fold cross-validation, achieving a mean AUC of 0.88. By analyzing the class activation map, we found that the model's attention was focused on the peri-tumoral edema region of gliomas and tumor bulk for germinomas, indicating that differences in these regions may help discriminate these tumors.

Conclusions: This study showed that the T1C-based transfer learning model could accurately distinguish germinomas from gliomas of the basal ganglia preoperatively.

Keywords: germinoma, glioma, deep neural network, machine learning, transfer learning

INTRODUCTION

Germ cell tumors (GCTs) are neoplasms derived from reproductive cells, mostly occurring in children and adolescents at 10 to 19 years of age (1). Intracranial GCTs are classified histologically into germinomas (assessed in this study) and non-germinomatous germ cell tumors. Intracranial germinomas mostly arise from pineal or suprasellar regions (2, 3). Due to the adjacency to midbrain

structures, intracranial germinoma patients usually develop hydrocephalus-related symptoms (3), and germinomas of the basal ganglia are difficult to distinguish from gliomas, which are the most common intracranial solid tumors, at the same site, even for experienced neurosurgeons or radiologists. Intracranial germinoma has a lower incidence rate compared with glioma in children and adults (4).

Intracranial germinoma is sensitive to radiation therapy, and a satisfactory prognosis could be achieved without surgical operation; however, as mentioned above, it is difficult to diagnose by routine MRI (T1-weighted, T2-weighted and enhanced T1) without additional acquisition like DWI and SWI. Unfortunately, such images are not always available due to the patient's financial status or scanner machine-hour shortage in developing countries. Most entry-level hospitals in China are not equipped with an advanced 3T MRI scanner or haven't purchased those additional imaging modalities from scanner vendors. *In situ* biopsies obtained intraoperatively or preoperatively with the stereotactic guide represent the "gold standard" to diagnose germinomas for most of the cases (3, 5), while the potential risk of tumor seeding or spread cannot be ignored. For some patients, traumatic procedures like surgery can be avoided if it is possible to accurately distinguish germinomas from gliomas with routine MRI. Although germinoma is commonly seen in adolescents while glioma has a higher incidence in the elderly population, headache and dizziness are common symptoms reported by both glioma and germinoma patients. Similarities in clinical manifestations compared to glioma in addition to the rarity of germinoma cases make it difficult to distinguish these two types of tumor, despite the difference of age at diagnosis. The lower incidence rate of germinoma and the abovementioned similarities provide little motivations for physicians to require additional serum test in clinics for β-HCG, an important indicator of germinoma. Therefore, a model using only routine MRI that could help physicians decide whether a patient needs further lab examination before hospital admission in such a scenario would be valuable; furthermore, a simple system requiring minimal input information that could distinguish these two types of tumor would reduce the cost per patient by cutting down unnecessary tests.

Recent advances in artificial intelligence (AI) in the field of tumor medical imaging have revealed that computers can achieve better accuracy in classifying different types of tumors than human physicians (6–8). Previous studies have investigated deep learning-based approaches to discriminate gliomas from other intracranial lesions including brain metastasis (9, 10), meningioma (10, 11), pituitary adenoma (10, 11), and acoustic neuroma (10). And other studies also reported the classification of germinoma with craniopharyngioma and pinealoblastoma (12, 13) by machine learning approaches. These reports focus on germinomas of the sellar and pineal region. Due to the uncommon incidence of germinomas of the basal ganglia, it has not been explored if routinely acquired MRI images can be utilized to differentiate germinomas from gliomas of this region. In this work, a deep learning model was established to answer this question. We enrolled 73 patients diagnosed with germinoma or glioma of the basal ganglia from two independent medical centers. **Supplementary Figure S1** shows typical T1-weighted contrast images of glioma and germinoma patients enrolled in this study, with the characteristic irregular enhancement of tumor bulk and cyst formation, which are similar between these two tumor types (14–18).

The purpose of the study is to provide aid in preoperative decision-making with a classification system for intracranial germinomas and gliomas of the basal ganglia. Enhanced T1 MRI is routinely acquired for differential diagnosis of intracranial lesions and is what's solely needed for the neural network we developed based on ResNet-18 (19), which provides the strong capability for future clinical translation.

MATERIALS AND METHODS

Data Collection

Multi-center data for a total of 180 germinomas and 71 glioma patients were retrieved from the databases of Xiangya Hospital and Sanbo Brain hospital from 2010 to 2018. Brain MRI imaging was performed as part of routine clinical care on scanners from various manufacturers with different magnetic field strengths (**Table 1**) and acquisition parameters (**Table 2**). **Supplementary Figure S2** showed the distribution of voxel geometry. A total of 39 germinoma and 48 glioma patients had lesions of the basal ganglia confirmed by immunohistochemistry (**Figure 1**). The inclusion criterion was: lesions with enhancement areas larger than 500 resliced voxels (average voxel size, 0.52 mm×0.52 mm×4.74 mm); 7 germinoma and 7 glioma patients were excluded for this reason or not having enhanced T1 image at all. There were no exclusion criteria based on age, gender, or race. Demographic and clinical data, including gender, age, and race, were retrieved from electronic medical records (**Table 3**).

Therefore, a total of 73 patients (31.07 ± 18.21 years old, varying from 6 to 67 years; M: F = 47:26) were included in the final study cohort (**Table 3**). The study was approved by the institutional review board of Xiangya Hospital and Sanbo Hospital, and informed consent was waived due to the retrospective nature of

TABLE 1 | Clinical MRI Scanners Used.

Manufacturers and magnetic field strength	No. of patients
All manufactures	
Total at 1.5 T	56 (76.7%)*
Total at 3 T	17 (23.3%)
Alltech Medical Systems	
1.5 T	4 (5.4%)
GE Medical Systems	
3 T	7 (9.6%)
Philips Medical Systems	
1.5 T	19 (26.0%)
SIEMENS	
1.5 T	26 (35.6%)
3 T	10 (13.7%)
TOSHIBA	
1.5T	7 (9.6%)

Numbers in parentheses are percentages.

this study. The study was conducted in accordance with the Declaration of Helsinki.

Image Preprocessing and Lesion Labeling

Pre-segmentation image registration was performed with both T1-weighted contrast-enhanced (T1C) and T2-weighted fluid-attenuated inversion recovery (FLAIR) images; affine images were co-registered into the same geometric space using the Elastix toolbox (20). Image transformation and re-slicing were performed using TorchIO (21) scripts (https://github.com/fepegar/torchio); images series were resliced into an average voxel size of 0.52×0.52×4.74 mm to minimize biases in the interpolation. All the T1C and FLAIR images were used for the segmentation of enhancing tumors and peritumoral edemas, respectively. Delineation of tumor boundaries was performed in a semi-automated fashion on a slice-by-slice basis using the ITK-SNAP software, an open-source 3D image analysis kit (22). The segmented T1C and FLAIR images were reviewed for tumor delineation and consistency by two neuroradiologists (CT and NY with over 8 and 6 years of experience, respectively). The delineated images of the two segmented tumor phenotypes (enhancing tumor and peritumoral edema) were exported for further analysis. Lesions smaller than 500 voxels (about 0.65 cm^3) were excluded for the following reasons. Small lesions like this could not be reliably segmented, typically such a small lesion will only appear on a single slice or two which makes the image less representative and reliable for feature extraction. In patients with > 1 lesion, all the lesions larger than 500 voxels were included in the analysis. The entire dataset contained a total of 93 lesions (45 germinoma lesions and 48 glioma lesions) from 73 patients (**Figure 1**).

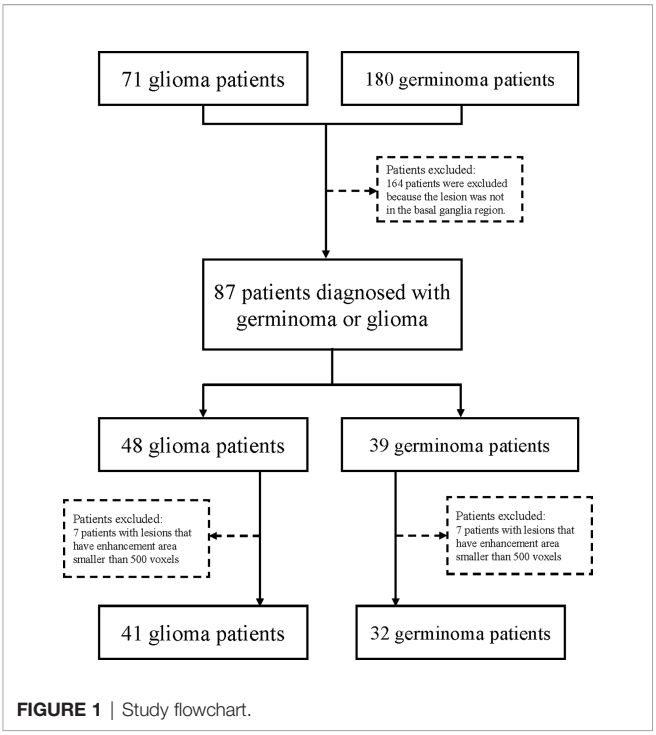


FIGURE 1 | Study flowchart.

Data Argumentation and Transfer Learning

We adapted a ResNet18 architecture pre-trained on the ImageNet datasets (19). We only used T1C images to train the classification model.

1. Slice selection, the center slice for each lesion was selected.
2. Conversion of a grayscale image to a 3-channel image. Three strategies were compared.
   a. No transformation (original gray-scale MRI images).
   b. Upper and lower slices together with the center slice were stacked as R, G and B channels respectively.
   c. Use the Jet color map to linearly transform the T1C gray image into an RGB image.
3. Image size normalization. Images were resized to 224 × 224.

Due to a limited number of patients, data augmentation was performed on-the-fly to prevent overfitting. More specifically, 6 data augmentation techniques were implemented, including random flip, random affine, random blur, random ghosting, random motion, and random elastic deformation. Examples of the augmented images are shown in **Figure 2**.

The retraining process consisted of two steps, including initializing the convolutional layers with loaded pre-trained weights that trained based on the ImageNet data, and freezing all convolutional layers and fine-tuning the classification layer. By resetting the finial layer into 2 and changing the loss function into cross-entropy, our model was able to implement the pre-learned image feature extraction pattern to this tumor classification task and make prediction. To compare the model performance, we also used

TABLE 2 | Summary of Acquisition Parameters in this study.

Parameter	Minimum	Median	Maximum
MRI image with 3 T scanner			
T1-weighted postcontrast MRI TE (msec)	2.37	2.98	26.82
T1-weighted postcontrast MRI TR (msec)	500	2200	2741.04
T1-weighted postcontrast MRI typical voxel size (mm)	0.72 × 0.72 × 0.9	1 × 1 × 1	1 × 1 × 5
T1-weighted postcontrast MRI typical matrix size	230 × 230	512 × 512	640 × 640
MRI image with 1.5 T scanner			
T1-weighted postcontrast MRI TE (msec)	4.6	10	15.7
T1-weighted postcontrast MRI TR (msec)	25	400	2100
T1-weighted postcontrast MRI typical voxel size (mm)	0.30 × 0.30 × 2	0.45 × 0.45 × 5	0.72 × 0.72 × 5
T1-weighted postcontrast MRI typical matrix size	256 × 256	512 × 416	1024 × 1024

TE, echo time; TR, repetition time.

TABLE 3 | Clinical characteristics of the cohort.

Characteristics		
Diagnosis	Glioma	Germinoma
No. of patients	41	32
Age at first diagnosis	43 years median; 6-67 years range	13.5 years median; 7-44 years range
Gender		
Female	18	8
Male	23	24
Number of lesions		
1	37	26
>1	4	6

the same deep learning architecture to train our data from scratch. All the parameters were set identical with the pre-trained model except the randomly initial weights of convolutional layers. ResNet18 model was used from *Torchvision.models* (0.11.0).

The ResNet18 with and without pre-trained models were trained on an Ubuntu 18.04 workstation with 1 Intel Core i9-7940 CPU, using an NVIDIA GTX 1080Ti 11GB GPU, with 256 GB available system RAM. Training in all categories was run for 100 epochs or 100 steps by stochastic gradient descent in batches of 12, using an SGD Optimizer with momentum 0.9. The learning rate was set as 0.001 for all layers and utilized with a decay rate of 0.1 each 4 steps until the model gradually reached convergence. In this study, 5-fold cross-validation was performed to train the model. In each fold, 80% cases were used as the training set and the rest were used as the validation set. The training and validation were performed with *Pytorch* (https://pytorch.org/) on *Python* 3.8.0.

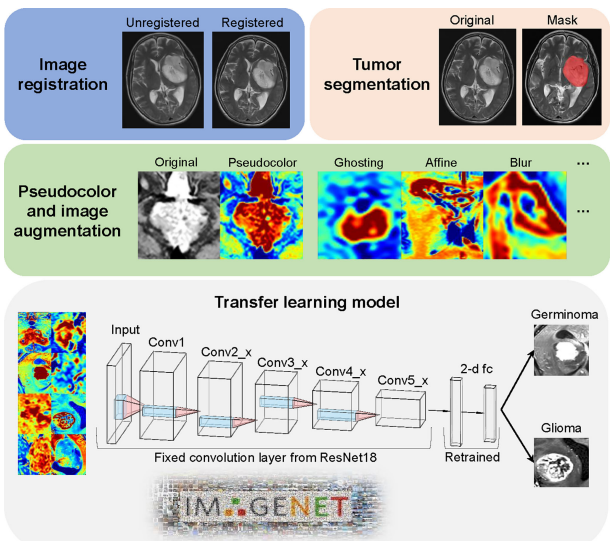


FIGURE 2 | Image processing and model architecture. Image preprocessing included two major steps (image registration and tumor segmentation). The jet colormap was applied to gray-scaled MRI images, followed by the use of 6 image augmentation techniques. Convolution layers from pre-trained ResNet18 were fixed as a feature extractor. The final 2-dimension classifier was retrained to fit our data.

Model Visualization

Class activation mapping (CAM) was performed to identify the areas contributing the most to the model, as described by Zhou and collaborators (23). CAM can serve as a quietly powerful approach for the reason that they enhance image regions contributed more to the output of the model and denote the model's confidence in the prediction. Specifically, for a unit k in a layer l, CAM calculates the importance score of k for class c and follows with visualizing the importance *via* a heatmap. We added a global average pooling layer after all the convolutional layer, which helps find all the discriminative regions. Feature maps of l before activation was visualized by CAM, then a heatmap was superimposed on input images.

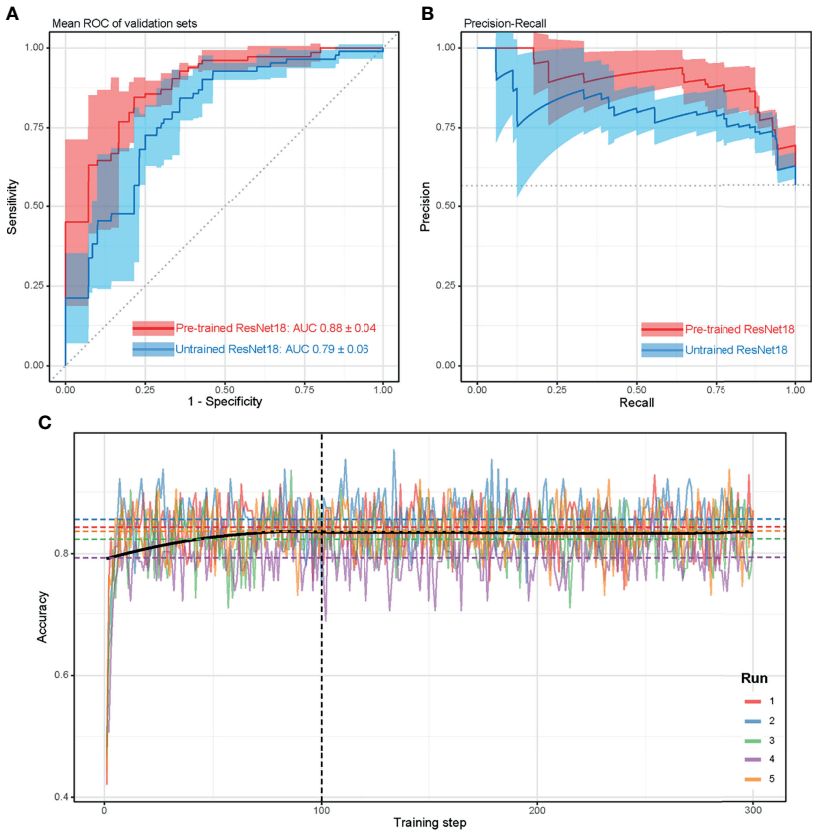
Statistical Analysis and Visualization

The performance statistics of the models were analyzed with the R programing language (v 4.1.2). Visualization and calculation of AUC (Area under the curve) and standard deviation of the 5-fold cross-validation was calculated with R package *precrec* (v 0.12.7) (24). In related analysis, lesion size is defined as the volumetric size of the enhanced area in T1C images. Lesions larger than median size are defined as large lesions, and lesions equivalent to or smaller than median size are defined as small lesions. Statistical tests of mean values were performed with Wilcoxon signed-rank test unless specified otherwise.

RESULTS

Development of a Transfer Learning Model to Distinguish Germinomas From Gliomas

As described in the method, 3 strategies of image transformation were compared. First, we trained the model with the original gray-scale MRI image, the model reached AUCs of 0.72 ± 0.07 (mean ± standard deviation [SD]) (**Supplementary Figure 3A**). Second, the model trained on adjacent-slices-stacked images reached AUCs of 0.81 ± 0.06 (mean ± SD) (Sup. Figure 3A). Third, the model trained with the Jet colormap-transformed images reached AUCs of 0.88 ± 0.04 (mean ± SD) (**Figure 3A**). Image transformation with the Jet colormap seems to be the best strategy for our dataset based on the ROC and the precision-recall curve (**Figures 3A, B, D** and **Supplementary Figures 3A, B**). The best model reached accuracy levels of 0.81 ± 0.01 in the validation set (**Figures 3C, D**). The precision-recall curves of



D

Model (strategy of image transformation)	Mean AUC	SD (AUC)	Mean ACC	SD(ACC)
Pre-trained ResNet18 (grayscale)	0.72	0.07	0.80	0.01
Pre-trained ResNet18 (adjacent-slices-stacked)	0.81	0.06	0.79	0.05
Pre-trained ResNet18 (Jet)	0.88	0.04	0.81	0.01
ResNet18 (Jet)	0.79	0.06	0.77	0.06

FIGURE 3 | Model evaluation. **(A)** Mean ROC in validation sets for the 5 runs. AUC = 0.88 ± 0.04 (mean ± standard deviation [SD]). Red line represents transfer learning on ResNet18 pre-trained on ImageNet, blue line represents training of ResNet18 from scratch. **(B)** Precision-recall curve for the 5 runs. Red line represents transfer learning on ResNet18 pre-trained on ImageNet, blue line represents training of ResNet18 from scratch. **(C)** Accuracy of the model during training. Dotted lines indicate mean accuracy from training steps 100 to 300, of the five-fold training. Black line is the loess fitting of accuracy of n-fold cross-validation at each training step. **(D)** Mean AUC and ACC of the four model we trained in this study. SD, standard deviation. ACC, accuracy.

these models indicate that when more information is provided to the model, the result is slightly better.

To test the effectiveness of transfer learning and provide a performance benchmark, we also trained ResNet18 on Jet colormap-transformed images from scratch. Unsurprisingly, the performance is worse than the pre-trained one, reaching AUCs of 0.79 ± 0.06 (mean ± SD), as shown in **Figures 3A, B** (The blue lines).

The mean lesion size of germinoma is smaller than glioma **Supplementary Figures 4A, B**. To test whether lesion size affects model performance, we performed chi-square (χ^2) test on contingency tables of True-False prediction and tumor size of the 5-fold cross-validation **Supplementary Figures 4C, D**. No significant difference in the distributions of predicting correctness between small and larger lesions is found. This proved the model's ability to generalize over various tumor sizes. Still, the median lesion size of correct prediction is slightly larger than incorrect prediction. Indicate that the model might perform better with larger lesions (**Supplementary Figures 4C, D**). Additionally, the model was tested with lesions excluded in the original analysis (smaller than 500 voxels), reached AUCs of 0.64 ± 0.07 (mean ± SD), this indicates that our model might not be suitable for ultra-small lesions (**Supplementary Figures 3C, D**).

The Class Activation Map Reveals a Location Preference That the Model Focuses on for Different Tumors

The neural network-based machine learning model was more sophisticated and less interpretable than traditional methods.

Therefore, we applied CAM (23) to probe the model after training. Pixels on a class activation map represent the superimposed activation strength of each unit in the last convolutional layer, which can be used to detect the regions on which the model mostly focuses; in other words, this could reveal the location preference that the model focuses on for different tumors.

Both germinoma and glioma are characterized by a bulk region (or core region) and a peritumoral edema region (**Figure 4A**, left). To examine if the model had "attention" preferences for different regions, we overlaid the class activation map onto the original image (**Figure 4A**, right). Whether the focal points were located in or outside of the tumor bulk, they were significantly more likely to be centric for germinoma than glioma cases in all five cross-validation runs (**Figure 4B**). When the focal points were located in the edema, they were significantly more likely to stay away from the outer edge of the edema for germinoma than glioma cases (**Figure 4C**). No significant differences in the distance were found from the focal points (when outside of the edema) to the outer edge of the edema (**Figure 4C**), from the focal points to the center mass of the tumor bulk, and from the focal points to the center mass of the peritumoral edema **Supplementary Figure 5** between these two tumor types. In other words, there was a tendency for the model to focus on the peritumoral edema region of gliomas and the tumor bulk for germinomas. These findings indicated different properties of these two types of tumors in terms of physical structure, which could help discriminate them.

DISCUSSIONS

In this study, we developed a neural network for the discrimination of germinomas and gliomas of the basal

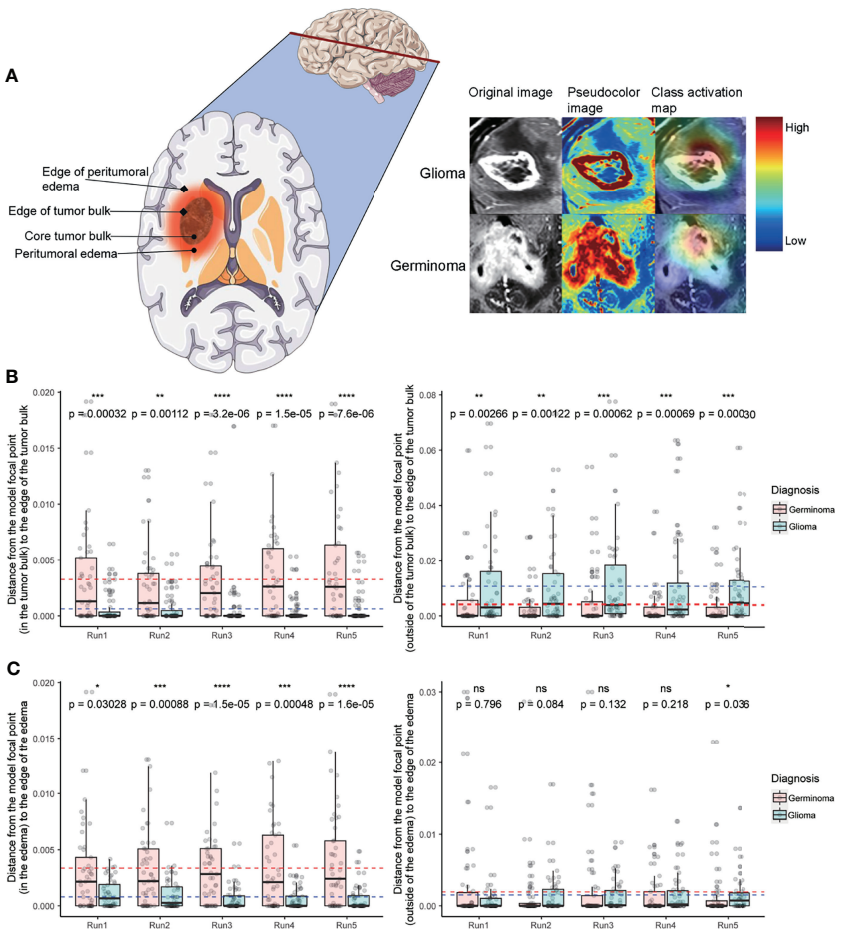


FIGURE 4 | Class activation map analysis. **(A)** A schematic of a tumor's physical structure (left), a schematic of the image with pseudo-color (middle), and a superimposed class activation map (right). On the class activation map that is overlaid on the original image, the color bar indicates weights (red and blue for high and low, respectively). **(B)** Distance from the model's focal point (in the tumor bulk) to the edge of the tumor bulk (left). Distance from the model's focal point (outside of the tumor bulk) to the edge of the tumor bulk (right). **(C)** Distance from the model's focal point (in the edema) to the edge of the edema (left). Distance from the model's focal point (outside of the edema) to the edge of the edema (right). In C and D, dotted lines indicate the mean distance (red and blue for germinoma and glioma, respectively). Distances are normalized by tumor size. Wilcoxon signed-rank test, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. ns, not significant.

ganglia. Most previous reports of germinoma are case reports or prognosis analyses based on tens of cases. To our best knowledge, this is the largest cohort of basal ganglia germinoma and the first quantitative MRI image analysis. The rarity of this disease with the specific location makes it difficult to collect enough data to train a convolutional neural network from scratch. In this case, we applied two commonly implemented techniques. The first of which is data augmentation to increase the dataset size and fight overfitting. The second is transfer learning to dramatically reduce the number of parameters to fit (25, 26).

It is reported that color information improves the performance on object recognition tasks of human participants (27). As color image encodes more visual information than grayscale image. It is less explored if colorized medical image improves accuracy in diagnostic imaging when examined by a human radiologist. Kather, J. N. et al. showed that radiologists can detect cancer tissue on colorized MRI images, with equivalent performance on grayscale images, while receiving almost no extra previous training (28). Our transfer learning method is based on the ResNet18, pre-trained on the ImageNet dataset, a natural scene database consisting of over 14 million manually annotated RGB color images of over 1000 categories. The above inspired us to test if deep learning models also have better performance with colorized medical images in classification tasks. Currently, there is no well-acknowledged method for adding pseudo color to medical images. Previously reported method includes linear color conversion from grayscale to a color map (29), triplicate the grayscale channel to synthesize color image (30, 31), concatenating three independent slices from one or cross different series (planes) (32–35). In this study, we thoroughly benchmarked these methods. The results showed that linear transformation to a color map yielded the highest AUC in our dataset. This provides a valuable reference for future implementation of neural networks on medical imaging.

Neural network-based machine learning models are infamously known as "black boxes". Therefore, CAM was implemented to improve interpretability, and to shed light on the decision-making process of the network, thus ensuring the focal point of the model doesn't fall in irrelevant areas. A few misclassified cases were shown in **Supplementary Figure 6**. For misclassified germinomas **Supplementary Figure 6A**, focal points of the model are located near ventricle structures and irrelevant peritumoral white matter. For misclassified gliomas **Supplementary Figure 6B**, focal points of the model located at the tumor bulk or edge of it. Previous reports showed that germinomas of the basal ganglia is characterized with minimal peritumoral edema (36, 37). As for glioma, peritumoral edema is a classic feature of MRI T2 image especially for high-grade glioma (38, 39). This might explain why the model fails on these specific cases and why it emphasizes peritumoral edema of gliomas. The model attention areas provide a valuable reference for physicians in case of suspected germinomas of the basal ganglia to avoid misdiagnosis. We delineated the tumor regions manually for CAM-related analysis, integrating automatic classification algorithms and auto segmentation of tumor will be explored in the future for the deployment of such models in clinical practice.

Medical images are volumetric datasets as they typically contain slices of body segments and organs. In this study, we developed a neural network model based on a single slice of T1C MRI image, which is routinely acquired in medical practices. Unlike the traditional radiomics based model, no laborious slice-by-slice segmentation and labeling is required. The moderate AUC of the model provides a stronger capability of clinical translation. The increased number of parameters and the higher computational complexity limited the application of 3D CNN currently, but it might perform better in a larger dataset than single-slice-based CNN.

This study had several limitations. First, the jet colormap was still not representative and interpretable for different regions on the image, such as the tumor bulk, edema, and normal tissues. In future research, a customized colormap both suitable for machine learning algorithms and human interpretation of MRI gray-scaled images should be developed. Second, although we implemented a good visualization method to identify the focus of this model and discovered the differences for both tumors, the biological meanings of the features of the model attention mechanism still need to be explored further. Interpretable CNN, which trains the kernel to represent the specific meaning of parts on objects, could integrate biological knowledge for better interpretability (40), which should be implemented in the future. Third, the dataset was still small even though we gathered data from two large medical centers. Data augmentation was used to alleviate this problem, still, the repeatability and robustness of the model should be validated with external data, if possible. Finally, our model showed moderate ability to generalize over various tumor sizes, but the model might have better performance on large lesions.

CONCLUSIONS

A transfer learning classifier for germinomas and gliomas of the basal ganglia was built, reaching a mean AUC of 0.88 and a mean accuracy of 0.81 in the validation set. By employing class activation mapping, we found the model was focused on the peritumoral edema region of gliomas and the tumor bulk for germinomas.

DATA AVAILABILITY STATEMENT

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

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the ethics committee of Xiangya Hospital and the ethics committee of Sanbo 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

XLi, SL designed this study, NY, QY, ZC performed the analysis, NY, QY wrote the manuscript, NY, CT finished the segmentation of all MRI images, PL, XLin, XLiu, YX contributed in acquisition of the data used in this study. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (for XLi, Grant No. 81770781 and No. 81472594) and Beijing Natural Science Foundation (for SL, grant No. 7182076 and H2018201306).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.844197/full#supplementary-material

REFERENCES

1. Bowzyk Al-Naeeb A, Murray M, Horan G, Harris F, Kortmann RD, Nicholson J, et al. Current Management of Intracranial Germ Cell Tumours. *Clin Oncol (R Coll Radiol)* (2018) 30(4):204–14. doi: 10.1016/j.clon.2018.01.009
2. Board, P.D.Q.P.T.E. Childhood Central Nervous System Germ Cell Tumors Treatment (PDQ®): Health Professional Version. In: *PDQ Cancer Information Summaries*. Bethesda (MD: National Cancer Institute (US (2002).
3. Dufour C, Guerrini-Rousseau L, Grill J. Central Nervous System Germ Cell Tumors: An Update. *Curr Opin Oncol* (2014) 26(6):622–6. doi: 10.1097/CCO.0000000000000140
4. Ostrom QT, Patil N, Cioffi G, Waite QK, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2013-2017. *Neuro Oncol* (2020) 22 (12 Suppl 2):iv1–iv96. doi: 10.1093/neuonc/noaa200
5. Murray MJ, Bartels U, Nishikawa R, Fangusaro J, Matsutani M, Nicholson JC. Consensus on the Management of Intracranial Germ-Cell Tumours. *Lancet Oncol* (2015) 16(9):e470–7. doi: 10.1016/S1470-2045(15)00244-2
6. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts H. Artificial Intelligence in Radiology. *Nat Rev Cancer* (2018) 18(8):500–10. doi: 10.1038/s41568-018-0016-5
7. Singh SP, Wang L, Gupta S, Goli H, Padmanabhan P, Gulyas B, et al. 3d Deep Learning on Medical Images: A Review. *Sensors (Basel)* (2020) 20(18). doi: 10.3390/s20185097
8. Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. *Annu Rev BioMed Eng* (2017) 19:221–48. doi: 10.1146/annurev-bioeng-071516-044442
9. Shin I, Kim H, Ahn SS, Sohn B, Bae S, Park JE, et al. Development and Validation of a Deep Learning-Based Model to Distinguish Glioblastoma From Solitary Brain Metastasis Using Conventional MR Images. *AJNR Am J Neuroradiol* (2021) 42(5):838–44. doi: 10.3174/ajnr.A7003
10. Chakrabarty S, Sotiras A, Milchenko M, LaMontagne P, Hileman M, Marcus D, et al. MRI-Based Identification and Classification of Major Intracranial Tumor Types by Using a 3D Convolutional Neural Network: A Retrospective Multi-Institutional Analysis. *Radiol Artif Intell* (2021) 3(5):e200301. doi: 10.1148/ryai.2021200301
11. Deepak S, Ameer PM. Brain Tumor Classification Using Deep CNN Features via Transfer Learning. *Comput Biol Med* (2019) 111:103345. doi: 10.1016/j.compbiomed.2019.103345
12. Chen B, Chaoyue Z, Yang H, Wang Z, Li H. Differentiation Between Germinoma and Craniopharyngioma Using Radiomics-Based Machine Learning. *J Personalized Med* (2022) 12(1):45. doi: 10.3390/jpm12010045
13. Fan Y, Huo X, Li X, Wang L, Wu Z. Non-Invasive Preoperative Imaging Differential Diagnosis of Pineal Region Tumor: A Novel Developed and Validated Multiparametric MRI-Based Clinicoradiomic Model. *Radiother Oncol* (2022) 167:277–84. doi: 10.1016/j.radonc.2022.01.005
14. Liang L, Korogi Y, Sugahara T, Ikushima I, Shigematsu Y, Okuda T, et al. MRI of Intracranial Germ-Cell Tumours. *Neuroradiology* (2002) 44(5):382–8. doi: 10.1007/s00234-001-0752-0
15. Fu W, Ju Y, Zhang S, You C. Pediatric Basal Ganglia Region Tumors: Clinical and Radiologic Features Correlated With Histopathologic Findings. *World Neurosurg* (2017) 103:504–16. doi: 10.1016/j.wneu.2017.04.004
16. Moon WK, Chang KH, Kim IO, Han MH, Choi CG, Suh DC, et al. Germinomas of the Basal Ganglia and Thalamus: MR Findings and a Comparison Between MR and CT. *AJR Am J Roentgenol* (1994) 162 (6):1413–7. doi: 10.2214/ajr.162.6.8192009
17. Kim DI, Yoon PH, Ryu YH, Jeon P, Hwang GJ. MRI of Germinomas Arising From the Basal Ganglia and Thalamus. *Neuroradiology* (1998) 40(8):507–11. doi: 10.1007/s002340050634
18. Tang J, Ma Z, Luo S, Zhang Y, Jia G, Zhang J, et al. The Germinomas Arising From the Basal Ganglia and Thalamus. *Childs Nerv Syst* (2008) 24(3):303–6. doi: 10.1007/s00381-007-0460-0
19. He K, Zhang X, Ren S, Sun J. (2016). Deep Residual Learning for Image Recognition, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi: 10.1109/CVPR.2016.90
20. Klein S, Staring M, Murphy K, Viergever MA, Pluim JP. Elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Trans Med Imaging* (2010) 29(1):196–205. doi: 10.1109/TMI.2009.2035616
21. Fernando Pérez-García RS, Ourseli S. TorchIO: A Python Library for Efficient Loading, Preprocessing, Augmentation and Patch-Based Sampling of Medical Images in Deep Learning. *arXiv* (2020). doi: 10.1016/j.cmpb.2021.106236
22. 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 (3):1116–28. doi: 10.1016/j.neuroimage.2006.01.015
23. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. (2016). Learning Deep Features for Discriminative Localization, in: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi: 10.1109/CVPR.2016.319
24. Saito T, Rehmsmeier M. The Precision-Recall Plot is More Informative Than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PloS One* (2015) 10(3):e0118432. doi: 10.1371/journal.pone.0118432
25. Chang J, Yu J, Han T, Chang H, Park E. (2017). A Method for Classifying Medical Images Using Transfer Learning: A Pilot Study on Histopathology of Breast Cancer, in: *2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom)*. doi: 10.1109/HealthCom.2017.8210843
26. Yu Y, Lin H, Meng J, Wei X, Guo H, Zhao Z, et al. Deep Transfer Learning for Modality Classification of Medical Images. *Information* (2017) 8(3). doi: 10.3390/info8030091
27. Bramao I, Reis A, Petersson KM, Faisca L. The Role of Color Information on Object Recognition: A Review and Meta-Analysis. *Acta Psychol (Amst)* (2011) 138(1):244–53. doi: 10.1016/j.actpsy.2011.06.010
28. Kather JN, Weidner A, Attenberger U, Bukschat Y, Weis CA, Weis M, et al. Color-Coded Visualization of Magnetic Resonance Imaging Multiparametric Maps. *Sci Rep* (2017) 7:41107. doi: 10.1038/srep41107
29. Chelghoum R, Ikhlef A, Hameurlaine A, Jacquir S. Transfer Learning Using Convolutional Neural Network Architectures for Brain Tumor Classification From MRI Images. *Artif Intell Appl Innovations* (2020) 583:189–200. doi: 10.1007/978-3-030-49161-1_17

30. Ho N, Kim YC. Evaluation of Transfer Learning in Deep Convolutional Neural Network Models for Cardiac Short Axis Slice Classification. *Sci Rep* (2021) 11(1):1839. doi: 10.1038/s41598-021-81525-9
31. Bae JB, Lee S, Jung W, Park S, Kim W, Oh H, et al. Identification of Alzheimer's Disease Using a Convolutional Neural Network Model Based on T1-Weighted Magnetic Resonance Imaging. *Sci Rep* (2020) 10(1):22252. doi: 10.1038/s41598-020-79243-9
32. Kang SK, Shin SA, Seo S, Byun MS, Lee DY, Kim YK, et al. Deep Learning-Based 3D Inpainting of Brain MR Images. *Sci Rep* (2021) 11(1):1673. doi: 10.1038/s41598-020-80930-w
33. Lin W, Tong T, Gao Q, Guo D, Du X, Yang Y, et al. Convolutional Neural Networks-Based MRI Image Analysis for the Alzheimer's Disease Prediction From Mild Cognitive Impairment. *Front Neurosci* (2018) 12:777. doi: 10.3389/fnins.2018.00777
34. Kang L, Jiang J, Huang J, Zhang T. *Identifying Early Mild Cognitive Impairment by Multi-Modality MRI-Based Deep Learning* Lausanne, Switzerland: Frontiers Media 12. (2020).
35. Trebeschi S, van Griethuysen JJM, Lambregts DMJ, Lahaye MJ, Parmar C, Bakers FCH, et al. Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric MR. *Sci Rep* (2017) 7 (1):5301. doi: 10.1038/s41598-017-05728-9
36. Rasalkar DD, Chu WC, Cheng FW, Paunipagar BK, Shing MK, Li CK, et al. Atypical Location of Germinoma in Basal Ganglia in Adolescents: Radiological Features and Treatment Outcomes. *Br J Radiol* (2010) 83 (987):261–7. doi: 10.1259/bjr/25001856
37. Chan J, Yeung TW, Lau HY, Tan CB, Wong YC, Siu JCW, et al. *Clinical Presentation, Radiological Features, and Treatment Response of Basal Ganglia Germinoma: Case Series*. Hong Kong, P.R.C: Hong Kong Academy of Medicine Press (2019).
38. 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(7):874–8. doi: 10.1111/j.1468-1331.2009.02613.x
39. Wu CX, Lin GS, Lin ZX, Zhang JD, Liu SY, Zhou CF, et al. Peritumoral Edema Shown by MRI Predicts Poor Clinical Outcome in Glioblastoma. *World J Surg Oncol* (2015) 13:97. doi: 10.1186/s12957-015-0496-7
40. Zhang Q, Wu YN, Zhu S-C. (2018). Interpretable Convolutional Neural Networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. doi: 10.1109/CVPR.2018.00920

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: 03 March 2022
doi: 10.3389/fonc.2022.808580

Accuracy Improvement Method Based on Characteristic Database Classification for IMRT Dose Prediction in Cervical Cancer: Scientifically Training Data Selection

Yiru Peng[1†], Yaoying Liu[2†], Zhaocai Chen[3], Gaolong Zhang[2], Changsheng Ma[1], Shouping Xu[4] and Yong Yin[1*]*

[1] Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China, [2] School of Physics, Beihang University, Beijing, China, [3] Manteia Technologies Co., Ltd, Xiamen, China, [4] National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

OPEN ACCESS

Edited by:
Ye Zhang,
Paul Scherrer Institute (PSI), Switzerland

Reviewed by:
Ruijie Yang,
Peking University Third Hospital, China
Yizheng Chen,
Stanford University, United States

***Correspondence:**
Shouping Xu
shouping_xu@yahoo.com
Yong Yin
yinyongsd@126.com

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: *03 November 2021*
Accepted: *24 January 2022*
Published: *03 March 2022*

Citation:
Peng Y, Liu Y, Chen Z, Zhang G, Ma C, Xu S and Yin Y (2022) Accuracy Improvement Method Based on Characteristic Database Classification for IMRT Dose Prediction in Cervical Cancer: Scientifically Training Data Selection. Front. Oncol. 12:808580. doi: 10.3389/fonc.2022.808580

Purpose: Consistent training and testing datasets can lead to good performance for deep learning (DL) models. However, a large high-quality training dataset for unusual clinical scenarios is usually not easy to collect. The work aims to find optimal training data collection strategies for DL-based dose prediction models.

Materials and Methods: A total of 325 clinically approved cervical IMRT plans were utilized. We designed comparison experiments to investigate the impact of (1) beam angles, (2) the number of beams, and (3) patient position for DL dose prediction models. In addition, a novel geometry-based beam mask generation method was proposed to provide beam setting information in the model training process. What is more, we proposed a new training strategy named "full-database pre-trained strategy".

Results: The model trained with a homogeneous dataset with the same beam settings achieved the best performance [mean prediction errors of planning target volume (PTV), bladder, and rectum: 0.29 ± 0.15%, 3.1 ± 2.55%, and 3.15 ± 1.69%] compared with that trained with large mixed beam setting plans (mean errors of PTV, bladder, and rectum: 0.8 ± 0.14%, 5.03 ± 2.2%, and 4.45 ± 1.4%). A homogeneous dataset is more accessible to train an accurate dose prediction model (mean errors of PTV, bladder and rectum: 2.2 ± 0.15%, 5 ± 2.1%, and 3.23 ± 1.53%) than a non-homogeneous one (mean errors of PTV, bladder and rectum: 2.55 ± 0.12%, 6.33 ± 2.46%, and 4.76 ± 2.91%) without other processing approaches. The added beam mask can constantly improve the model performance, especially for datasets with different beam settings (mean errors of PTV, bladder, and rectum improved from 0.8 ± 0.14%, 5.03 ± 2.2%, and 4.45 ± 1.4% to 0.29 ± 0.15%, 3.1 ± 2.55%, and 3.15 ± 1.69%).

Conclusions: A consistent dataset is recommended to form a patient-specific IMRT dose prediction model. When a consistent dataset is not accessible to collect, a large

dataset with different beam angles and a training model with beam information can also get a relatively good model. The full-database pre-trained strategies can rapidly form an accuracy model from a pre-trained model. The proposed beam mask can effectively improve the model performance. Our study may be helpful for further dose prediction studies in terms of training strategies or database establishment.

Keywords: deep learning, radiotherapy, cervical cancer, database classification, IMRT, dose prediction

1 INTRODUCTION

In recent decades, with the emergence and development of advanced radiotherapy (RT) planning and delivery techniques such as intensity-modulated radiation therapy (IMRT) and volumetric-modulated arc therapy (VMAT), the quality of radiotherapy plans has drastically improved with better target volume dose coverage and normal tissue sparing (1, 2). However, there are still many obstacles in current clinical planning practice. The planning objectives or constraints generated using population-based standard clinical protocols may lead to sub-optimal or even infeasible plan quality for specific patients (3, 4). In addition, the trial-and-error planning process (5, 6) highly depends on the skills and experience of the planners. It is time-consuming and labor-intensive, and results in significant variations in plan quality (7, 8).

Many novel methods have been developed and introduced into clinics for improving treatment plan efficiency, quality, and consistency (9, 10). The knowledge-based planning (KBP) (7, 11, 12) generates the reference plans for a new patient using a dose-volume histogram (DVH) model trained from historical treatment plans (13, 14). However, the clinical application of this method is limited because of the unsatisfying output as only one-dimensional DVH, rather than the 3D dose distributions (15) and the need for manual interventions, such as planning target volume (PTV)-organ at risk (OAR) distance and PTV's length (16, 17). In recent years, with the rapid advances in computational power, the deep learning (DL) technique has drawn significant attention and become research hotspots in many fields (11, 18). The DL-based dose prediction method, which takes advantage of CNN's automatic features extraction ability, can build the relationship between anatomical information and the dose distribution of a patient. Many studies have shown the significant success of various DL methods in predicting 3D dose distributions for different treatment sites and delivery methods (3, 19, 20).

At the current stage, most of the DL dose prediction models require a large number of consistent and high-quality training datasets and the applications are limited to the cases that have the same characteristics, such as the same beam settings (numbers and angles) and the patient positions (majority is at supine position). However, a large high-quality training dataset is usually not easy to collect for unusual clinical scenarios, such as unique beam arrangements. The impact of the quality and quantity of training database on the DL model performance has been explored in other studies (21, 22), but only limited to the impact of the homogeneity and size of the database, and no other method about improving IMRT dose prediction accuracy, or making a suggestion for how to make use of a sizable non-homogeneous dataset. This work explores optimal training data collection strategies for DL-based dose prediction models. To the best of our knowledge, there have not been efforts reported in this direction. In this work, we designed comparison experiments to investigate the impact of (1) beam angles, (2) the number of beams, and (3) patient position for DL dose prediction models. In addition, a novel geometry-based beam mask generation method was proposed to provide beam setting information in the model training process, which makes the model more accurate. What is more, we proposed a new training strategy named "full-database pre-trained strategy". Our new training strategy might rapidly use a heterogeneous dataset to form a patient-specific dose prediction model.

2 METHODS AND MATERIALS

2.1 Patient Data and Treatment Planning

A total of 325 clinically approved and delivered cervical cancer IMRT plans were retrospectively selected for the DL model training and testing. All plan contours of the PTV and OARs have been checked by experienced radiation oncologists. The prescription dose of PTV cases was 50.4 Gy in 28 fractions. The treatment plan used different beam settings in beam numbers (7 or 9 beams, all coplanar), beam angles, and patient treatment positions (supine and prone) according to the patient's anatomy features. All the treatments were planned using the Eclipse treatment planning system with the Anisotropic Analytical Algorithm (AAA). All dose constraints of IMRT for cervical cancer are based on Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC) data.

2.2 Data Preparation

The purpose of the data preparation process is to ensure that the DL model could correctly process the mapping and transformation between anatomical information and the dose distribution of a patient.

In IMRT treatment, the beam setting can significantly influence patients' dose distribution. So, in addition to the image and contouring mask as model inputs, we also proposed a novel generation method of beam mask to feed the extra beam setting information (beam angles and beam numbers) to the DL model training, which is in a logic of clinical scenarios to predict dose distribution. The beam setting information was first

extracted from the RT plan file. We hypothesized that each beam source is a point source, and all the beams were tangent to PTV through the isocenter, which matches the clinical logic and can present the beam information to the model. This work generated a beam mask by assigning 1 to pixels within the beam path while 0 outside the beam. All the beam masks were then summed and rescaled to the maximum of 50.4 Gy to match the prescription dose for PTV and reduce the burden of DL model fitting (a form of normalization of the model's input and output data: making beam mask's value more similar to the training dose distribution, reducing the parameter update steps of model training). Our new beam mask production method was an analytical algorithm based on the geometric method. Compared with the existing dose engine-based and TPS-based beam mask generation methods (19), our method requires fewer computing resources and is much easier to realize. The details of making a beam mask are shown in **Figure 1**.

We first extracted a 3D matrix of 512 × 512 × 100 (± 30) for each plan from CT images. It was then normalized to have a mean value of 0 and a variance of 1. A binary mask was generated for each ROI, with 1 for voxels inside the contour and 0 for those outside the contour. A 3D dose array was also extracted from the RT dose file. For fast dose calculation purposes, all the clinical plans were calculated using a 2.5-mm dose grid, which is different from the resolution of the original CT images [1 (± 0.2) mm × 1 (± 0.2) mm × 5 mm]. Therefore, the images and contouring masks were resampled to match the solution of the dose matrix. The images, contouring masks, and the dose matrix were then rigidly registered and used for DL model training. **Figure 2** shows the diagram of the DL model input and output.

Python codes did the data pre-processing for all preparation processes. Python packages such as NumPy and pydicom were used to conserve the raw data to the "npy" format.

2.3 Model Architecture and Training Method

A 3D Dense U-Net CNN model was built. The model input data were a multi-channel 3D matrix, including CT images, ROI contouring masks for bladder, body, left femoral head (Femoral-Head-L), right femoral head (Femoral-Head-R), PTV, rectum, spinal cord, and beam mask (if the beam is added as a feature). The model output was the corresponding predicted 3D dose matrix. It shows the model structure and training method's details in **Appendix-1** (Model architecture and training method).

2.4 Experiment Design

Four experiments were designed to explore the influence of dataset characteristics in IMRT dose prediction mode training. The first series of experiments explored the total usage of a sizable non-homogeneous dataset. The second, third, and fourth experiments explore the influence of unification of beam angles, beam numbers, and patient position, which are essential factors that can influence dose distribution in IMRT plans.

2.4.1 Full-Database Experiments With Pre-Trained Strategy

A total of 9 cases were randomly selected as testing cases, which were all treated with seven beams of 0°, 50°, 100°, 150°, 210°, 260°, and 310° at the supine position. As shown in **Figure 3**, we designed three different experiments. In Experiment 1-1, Model 0 was trained by taking the full usage of 258 training cases from total cases number 325, including different beam numbers or angles and different patient positions. In Experiment 1-2, Model 01 was trained using 46 patients with the same beam setting as the saved testing cases. In Experiment 1-3, Model 02 was first trained by reserving the model weight from Model 0 and training the selected 46 patients. The prediction errors of the three models were calculated and compared. Then, we used the best model training strategy in the following experiments.

2.4.2 Influence of Unified Beam Angles

To investigate the influence of uniform beam angles in the training dataset and the extra model input of beam mask on the performance of the dose prediction model, we conducted four comparative experiments. The same testing cases as in the previous section were used in these experiments. In Experiment

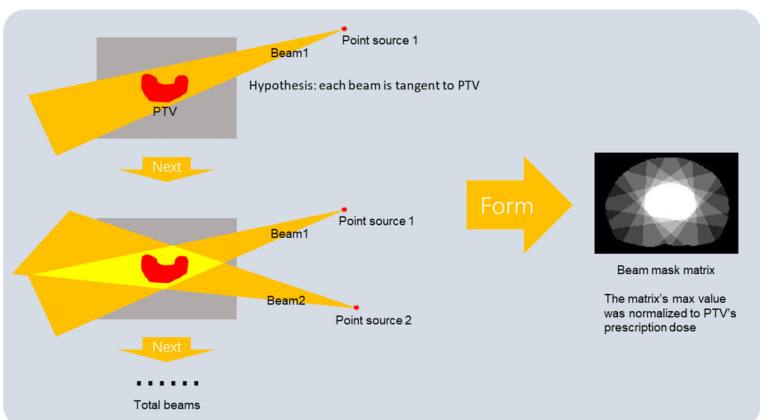


FIGURE 1 | The generation process of beam mask.

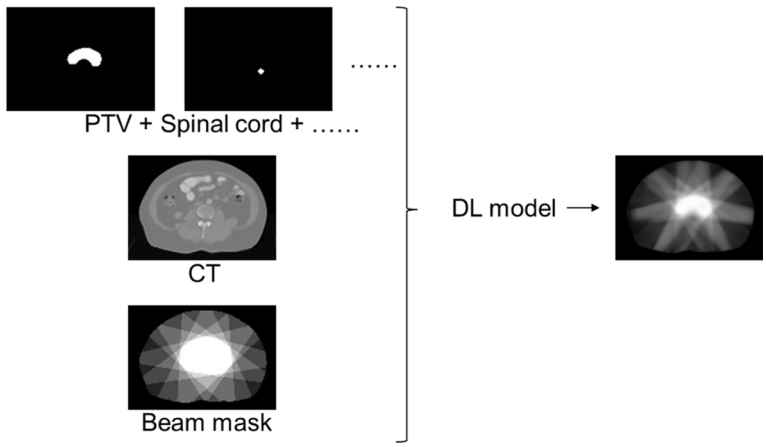


FIGURE 2 | Data preparation diagram.

2-1, we selected the 46 cases with the same beam configuration in the testing cases. In addition, the beam mask was added as an extra input channel in the training process. In Experiment 2-2, the same training cases as in Experiment 2-1 but without the beam masks. In Experiment 2-3, we selected 213 cases with seven beams as training data and used beam masks in the training process. Experiment 2-3 aimed to study whether a significant number of cases in the training dataset with different beam angles can improve the model performance. In Experiment 2-4, the training cases were the same as Experiment 2-3, but Experiment 2-4 had the added beam masks in the training process.

2.4.3 Influence of Unified Beam Numbers

To investigate the influence of uniform beam numbers in the training dataset and the extra model input of beam mask on the performance of the dose prediction model, we conducted four comparative experiments.

We selected 12 testing cases treated with nine beams. Experiments 3-1 and 3-2 were to compare beam performance using either the training data of the same beam number (i.e., nine beams) or a more extensive training dataset with the mixed numbers of beams (i.e., 9 and 7 beams). Because of the limitation of 9-beam cases (21 cases), training and testing cases had different beam angle distributions. The beam masks were used for both experiments. In Experiment 3-3, the model was trained with only 7-beam cases to investigate model performance in the testing cases with different beams. In Experiment 3-4, the training cases were the same as Experiment 3-2 but without beam masks as input.

2.4.4 Influence of Unified Positioning

Most of the cervical patients in our department were scanned with the supine position, but there are still some cases with the prone position. We designed four comparison experiments using

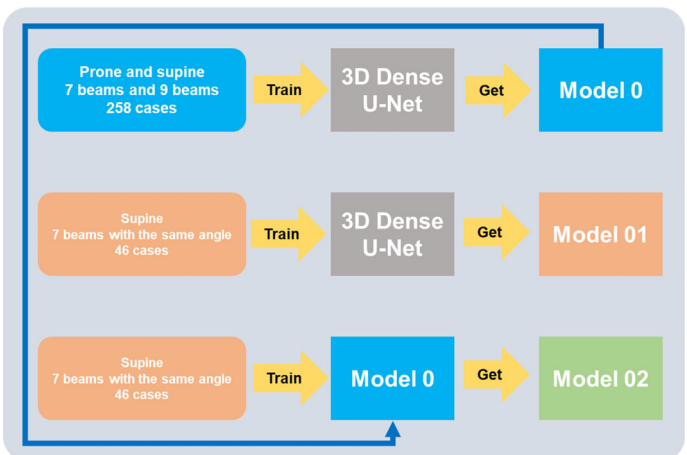


FIGURE 3 | Experiments to prove a full-database pre-trained strategy.

different training datasets to investigate the strategy for optimal selection of DL model training data, particularly for the clinical scenario with fewer cases such as prone position.

We selected 13 test cases scanned in the prone position. In Experiment 4-1, we selected a total of 258 training cases with both prone position and supine position, while in Experiment 4-2, only the 45 prone cases were used for model training. The beam masks were added as model input for both experiments to provide beam information to help the model predict the accurate doses. For comparison, in Experiment 4-3, we removed the beam masks from Experiment 4-2. In Experiment 4-4, the model was trained with only the supine position cases. Due to the dataset's size of prone cases being small (45 cases), training and testing cases have different beam angle distributions.

2.5 Model Performance Evaluation Method

The percentage of errors (δDi) was calculated to evaluate all experiments. The formula of the percentage of errors was:

$$\delta Di = \frac{\left|Di_{Ground-truth} - Di_{predicted}\right|}{prescription \quad dose} \times 100\,\%$$

We calculated δDi (a total of 14 DVH indices) of D_{95}, D_{90}, D_{50}, D_{max}, and D_{mean} of PTV; V_{30}, D_{max}, and D_{mean} of bladder, rectum; and D_{max} of spinal cord and femoral heads.

We selected the best model in each series of experiments by synthetically considering the dose-prediction errors. We measured the model by the 14 DVH indices: If the model gets the maximum optimal dose indices (% of best prediction is the number of optimal DVH indices/the total number of DVH indices), the model is the best in the group of experiments.

3 RESULTS

3.1 Full-Database Pre-Trained Strategy Experiments

Under these three experiments (Experiments 1-1 vs. 1-2 vs. 1-3), on the one hand, the errors of the PTV D_{95} decreased from 1.39 ± 0.95% for Experiment 1-1 to 0.96 ± 0.60% for Experiment 1-2, to 0.76 ± 0.79% for Experiment 1-3. On the other hand, for the considered OARs, the errors of the mean doses were reduced from 2.63 ± 1.63% and 3.80 ± 2.33% in Experiment 1-2 to 2.33 ± 1.67% and 3.18 ± 2.07% in Experiment 1-1, and then to 2.18 ± 1.67% and 3.10 ± 2.24% in Experiment 1-3. The summary of prediction errors of each experiment is shown in **Figure 4**.

3.2 Experiments of Beam Settings

Figure 4 shows the average predicted percentage dose errors of the mean and maximum doses in PTV and OARs. The model trained with a small database but with the same beam setting and beam information (Experiment 2-1 with 46 cases) outperformed the model trained with an extensive database with different beam settings (Experiments 2-3 and 2-4 with 213 cases) and a small database with the same beam setting but no beam information (Experiment 2-2 with 46 cases). With the added beam mask information, Experiment 2-1 yielded the best performance with

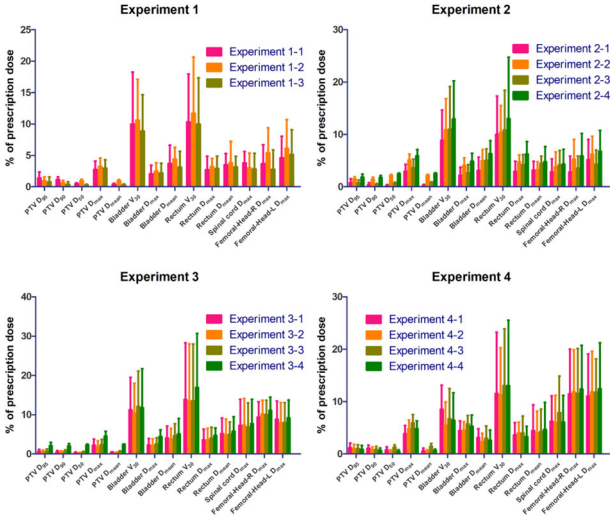


FIGURE 4 | The absolute percentage errors for ROIs.

the lowest mean dose errors across all the organs (2.18 ± 1.67%), followed by Experiments 2-2, 2-3, and 2-4 with mean dose errors of 3.21 ± 1.85%, 2.70 ± 1.43%, and 3.85 ± 2.13%, respectively. Similarly, Experiment 2-1 also had the lowest average errors of predicted max dose (3.11 ± 2.24%), while Experiments 2-2, 2-3, and 2-4 were 4.90 ± 2.16%, 3.73 ± 2.10%, and 5.70 ± 2.50%, respectively.

3.3 Experiments on the Number of Beams

For the model trained with a small database with the same beam number (i.e., nine beams) and beam information (Experiment 3-1), when adding different beam numbers, i.e., the model trained with an extensive database with mixed beam number (i.e., 9 or 7 beams) (Experiment 3-2). The predicted errors of the bladder V30 in Experiment 3-2 were 0.9%, 1.7%, and 1.4%, which were lower than the predicted values in Experiments 3-1, 3-3, and 3-4, respectively (**Appendix-3-Table 3**). For the rest of the plotted DVH metrics (PTV D_{95}, PTV D_{90}, PTV D_{50}, PTV D_{max}, PTV D_{mean}, Rectum D_{max}, Rectum D_{mean}, Bladder D_{max}, Bladder D_{mean}, Rectum V_{30}, and Left femoral head D_{max}), Experiment 3-2 had the slightest prediction error. In particular, Experiment 3-1 outperformed other experiment groups for the PTV D_{50}, which reduced to 0.21%.

3.4 Experiments on Treatment Positions

For a model trained in a database with mixed treatment positions (i.e., prone or supine position) and beam information (Experiment 4-1), when we removed the supine position data, only in the model trained in the prone position database with beam information (Experiment 4-2) were the predicted errors decreased in most of the organs, especially the bladder and rectum; the D_{mean} decreased by 0.8% and 0.5%, respectively. When the beam mask was removed from Experiment 4-2, the model trained only with the prone position database (Experiment 4-3), the prediction errors were significantly improved.

3.5 The Best Model in Each Series of Experiments

We evaluated our model's performance using the "Model performance evaluation method" in **Section 2.5** and the "% of best prediction". It shows that the best model in a series of experiments is in Experiment 1-3, Experiment 2-1, Experiment 3-2, or Experiment 4-2, separately (shown in **Appendix-2-Table 1**). Each model's prediction errors are shown in **Appendix-3-Tables 1–4**.

Appendix-3-Figures 1 and **4** show the results of the testing patients in each group of experiments with the three-dimensional dose distribution predicted by each group of experiments and the corresponding DVHs.

4 DISCUSSION

This study aimed to analyze the impact of classifying training databases on the performance of DL models for dose prediction in the framework of radiotherapy for cervical cancer. For this purpose, we set 3 groups of experiments to study the influence of a uniform training database with beam angles, beam numbers, and patient positions on the accuracy of a prediction model. We also proposed a new beam information mask generation method, which can quickly and accurately learn beam angle information and convert the beam settings into beam masks to achieve the best model performance. What is more, we created a new "full-database pre-trained strategy", which makes full use of a wide range of databases more effectively to build and obtain more accurate prediction models.

Some conclusions can be drawn across four groups of experiments, and the conclusions may be helpful in the IMRT dose prediction model training process and database establishment. First, a homogeneous dataset is more accessible to train an accurate dose prediction model than a non-homogeneous one without other processing approaches. This conclusion can be drawn from Experiment 2-2 vs. Experiment 2-4. These two experiments are both without additional processing. We found that the model trained with the 46-size homogeneous dataset (all cases have the same beam settings) performed better than the other model trained with the 213-size non-homogeneous dataset (the cases have different beam settings). Two model performance details can be seen from **Figure 4**, **Appendix-3-Figure 1**, and **Appendix-3-Table 2**.

Since the non-homogeneous dataset with different beam settings may cause a suboptimal model, we added a beam mask to provide beam setting information in training processing and tried to make a prediction model that establishes the relationship between beam settings and dose distribution. Moreover, the second conclusion can be that beam information can make the non-homogeneous models perform well. The conclusion may be drawn from Experiment 2-3 vs. Experiment 2-4 and Experiment 3-2 vs. Experiment 3-4. Visual comparisons are shown in **Appendix-3-Figure 1**. When the beam mask was added, the predicted dose accuracy was improved in global and PTV areas. Furthermore, when the beam mask was added, a non-homogeneous dataset's performance was close to a homogeneous one (Experiment 2-1 vs. Experiment 2-3, Experiment 3-1 vs. Experiment 3-2). The beam mask made the homogeneous dataset perform better (Experiment 2-1 vs. Experiment 2-2). The conclusion that beam masks can make the mixed beam setting models perform well was following clinical logic. The beam setting can significantly influence patients' dose distribution in the planning design process. So, using the beam mask to present beam information to the model, following the planning design logic, can make dose prediction more accurate.

Besides adding beam mask, another method to make usage of the non-homogeneous dataset is the "full-database pre-trained strategy". In the second conclusion, we know that beam mask can make the non-homogeneous dataset's performance close to the homogeneous one, so a pair of experiments were made (1-1 vs. 1-2); we found that an extensive non-homogeneous dataset's model (Model 0) performance beat a small homogeneous dataset's (Model 01) in almost all evaluation indicators when the beam information was added in the training processing. However, another question arises: the large dataset (256 cases) training process is time-consuming, costing several days (on an RTX 3090 GPU). So, another experiment 1-3 was made: using a 46-size homogeneous dataset to continue training Model 0 (256-size non-homogeneous dataset's model) to get Model 02. We found that Model 02's performance was better than that of Model 0, and the time cost of 46 cases refining Model 0 was within an hour. The third conclusion can be that when the beam information is involved in IMRT dose prediction model training, using a small homogeneous dataset to refine a sizable non-homogeneous dataset's model can get a good performance model, and the time cost can be much reduced compared with training a model from the beginning by a large dataset that combines the homogeneous and non-homogeneous dataset (an hour vs. several days).

Another important finding in our research is that the dataset size dramatically influences the accuracy of IMRT dose prediction model. When the beam mask involved the experiments, different performances among the homogeneous and non-homogeneous datasets could be seen in some experiments (2-1 vs. 2-3), but not in other experiments (3-1 vs. 3-2, 4-1 vs. 4-2). When analyzing the above phenomenon and dataset features, we focused on the sizes of the homogeneous dataset across the above experiments. Experiment 2-1 has a 46-size homogeneous dataset (the same beam settings), but a total number of 9 beam cases and prone cases, which were with the non-unified beam angles, only have 21 and 45. The level of homogeneity was lower than Experiment 2-1, and the dataset size was relatively small, so the advantage of the homogeneity dataset in Experiments 3-1 and 4-2 was not evident as 2-1. From the above discussion, an excess conclusion can be drawn that the beam angle unification is more effective than beam numbers and patient positions.

Some shortcomings were listed below, which can be renovated in future related work. Our analysis shows that the homogeneous dataset may be advantageous compared to the

non-homogeneous one when the homogeneity and dataset size level are enough. We have not collected enough nine beam cases and prone cases with the same beam settings to show the advantage of the homogeneity in beam numbers and patient positions. Future research may use a larger dataset to explore more characteristics of homogeneous and non-homogeneous datasets. Our conclusion showed that a homogeneous dataset could improve model accuracy, but the largest dataset with the same beam settings has only 46 cases. The model with the largest homogenous dataset got the best prediction errors (Experiment 2-1), which proved that the homogenous dataset might benefit the model's accuracy again. The lack of homogeneous datasets led to our prediction errors not being as good as the current study. Nevertheless, the conclusions in our research, such as a homogeneous dataset, suggested that providing beam settings in the training process might make future research a better dose prediction model. Advance computer technology may improve the experiment performance. Our study focused on the dataset with a relatively conventional U-net-like model. The novel structure such as attention-gate and the multi-stage network could be involved in the model architecture, making prediction more accurate. The "full-database pre-trained strategy" in our study used the homogeneous dataset for continuing to train a pre-trained model, which focused more on concepts than methods. Meta-learning technology aimed to use small data to refine a pre-trained model and improve performance. Meta-learning might make the "full-database pre-trained strategy" perform better. Further efforts can expand the investigations onto different tumor types with different treatment techniques.

5 CONCLUSION

This study designed different experiments to explore the influence of different clinical scenarios in IMRT dose prediction model training, such as various beam angles, the number of beams, and different patient positions. A homogeneous dataset is more accessible to train an accurate dose prediction model than a non-homogeneous one without other processing approaches. The beam angles of the dataset cases can significantly influence IMRT dose prediction accuracy. In the IMRT model training process, the beam information is suggested to be included. In the IMRT dose prediction dataset collection process, a compatible size dataset with the same beam angles is recommended. If the homogeneous data are hard to collect, training a model using a non-homogeneous dataset combined with beam information can also get a relatively accurate model.

Besides, in our study, a novel training strategy and beam information array generation method were proposed. The "full-database pre-trained strategies" used a small size dataset to re-train the model trained by a large dataset to form a specific model, which can get an accurate model and reduce the time-consuming training for the model. The proposed geometric-based beam mask generation method can effectively provide beam setting information and improve the model performance.

Our study may be helpful for further dose prediction studies in terms of training strategies or database establishment.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

YP: Experiment design, data collection and article writing. YL: Experiment design, code implementation and article writing. ZC: Technical support. GZ: The article modification. CM: Data collection. SX: Experiment design and article modification. YY: Data collection and article modification. All authors contributed to the article and approved the submitted version.

FUNDING

The work was supported by the Medical Big Data AI R&D Project (2019MBD-043), the National Natural Science Foundation of China (Grant No. 82072094), the Natural Science Foundation of Shandong Province (Grant No. ZR2019LZL017), and the Taishan Scholars Project of Shandong Province (Grant No. ts201712098).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.808580/full#supplementary-material

REFERENCES

1. Zelefsky MJ, Fuks Z, Happersett L, Lee HJ, Ling CC, Burman CM, et al. Clinical Experience With Intensity Modulated Radiation Therapy (IMRT) in Prostate Cancer. *Radiother Oncol* (2000) 55(3):241–9. doi: 10.1016/s0167-8140(99)00100-0
2. Crooks SM, Wu X, Takita C, Watzich M, Xing L. Aperture Modulated Arc Therapy. *Phys Med Biol* (2003) 48(10):1333. doi: 10.1088/0031-9155/48/10/307
3. Chen X, Men K, Li Y, Yi J, Dai J. A Feasibility Study on an Automated Method to Generate Patient-Specific Dose Distributions for Radiotherapy Using Deep Learning. *Med Phys* (2019) 46(1):56–64. doi: 10.1002/mp.13262
4. Fredriksson A. Automated Improvement of Radiation Therapy Treatment Plans by Optimization Under Reference Dose Constraints. *Phys Med Biol* (2012) 57(23):7799. doi: 10.1088/0031-9155/57/23/7799
5. Ma M, Kovalchuk N, Buyyounouski MK, Xing L, Yang Y. Incorporating Dosimetric Features Into the Prediction of 3D VMAT Dose Distributions Using Deep Convolutional Neural Network. *Phys Med Biol* (2019) 64 (12):125017. doi: 10.1088/1361-6560/ab2146
6. Ahn SH, Kim E, Kim C, Cheon W, Kim M, Lee SB, et al. Deep Learning Method for Prediction of Patient-Specific Dose Distribution in Breast Cancer. *Radiat Oncol* (2021) 16(1):1–13. doi: 10.21203/rs.3.rs-147694/v1

7. Yoganathan S, Zhang R. An Atlas-Based Method to Predict Three-Dimensional Dose Distributions for Cancer Patients Who Receive Radiotherapy. *Phys Med Biol* (2019) 64(8):085016. doi: 10.1088/1361-6560/ab10a0
8. Jiang D, Yan H, Chang N, Li T, Mao R, Du C, et al. Convolutional Neural Network-Based Dosimetry Evaluation of Esophageal Radiation Treatment Planning. *Med Phys* (2020) 47(10):4735–42. doi: 10.1002/mp.14434
9. Bakx N, Bluemink H, Hagelaar E, van der Sangen M, Theuws J, Hurkmans C. Development and Evaluation of Radiotherapy Deep Learning Dose Prediction Models for Breast Cancer. *Phys Imaging Radiat Oncol* (2021) 17:65–70. doi: 10.1016/j.phro.2021.01.006
10. Scaggion A, Fusella M, Roggio A, Bacco S, Pivato N, Rossato MA, et al. Reducing Inter-and Intra-Planner Variability in Radiotherapy Plan Output With a Commercial Knowledge-Based Planning Solution. *Phys Med* (2018) 53:86–93. doi: 10.1016/j.ejmp.2018.08.016
11. 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
12. Appenzoller LM, Michalski JM, Thorstad WL, Mutic S, Moore KL. Predicting Dose-Volume Histograms for Organs-at-Risk in IMRT Planning. *Med Phys* (2012) 39(12):7446–61. doi: 10.1118/1.4761864
13. Tol JP, Dahele M, Peltola J, Nord J, Slotman BJ, Verbakel WF. Automatic Interactive Optimization for Volumetric Modulated Arc Therapy Planning. *Radiat Oncol* (2015) 10(1):1–12. doi: 10.1186/s13014-015-0388-6
14. Nguyen D, Jia X, Sher D, Lin M-H, Iqbal Z, Liu H, et al. 3D Radiotherapy Dose Prediction on Head and Neck Cancer Patients With a Hierarchically Densely Connected U-Net Deep Learning Architecture. *Phys Med Biol* (2019) 64 (6):065020. doi: 10.1088/1361-6560/ab039b
15. Zhou J, Peng Z, Song Y, Chang Y, Pei X, Sheng L, et al. A Method of Using Deep Learning to Predict Three-Dimensional Dose Distributions for Intensity-Modulated Radiotherapy of Rectal Cancer. *J Appl Clin Med Phys* (2020) 21(5):26–37. doi: 10.1002/acm2.12849
16. Wei L, Su R, Wang B, Li X, Zou Q, Gao X. Integration of Deep Feature Representations and Handcrafted Features to Improve the Prediction of N6-Methyladenosine Sites. *Neural Comput* (2019) 324:3–9. doi: 10.1016/j.neucom.2018.04.082
17. Kajikawa T, Kadoya N, Ito K, Takayama Y, Chiba T, Tomori S, et al. Automated Prediction of Dosimetric Eligibility of Patients With Prostate Cancer Undergoing Intensity-Modulated Radiation Therapy Using a Convolutional Neural Network. *Radiol Phys Technol* (2018) 11(3):320–7. doi: 10.1007/s12194-018-0472-3
18. Nguyen D, Long T, Jia X, Lu W, Gu X, Iqbal Z, et al. A Feasibility Study for Predicting Optimal Radiation Therapy Dose Distributions of Prostate Cancer Patients From Patient Anatomy Using Deep Learning. *Sci Rep* (2019) 9(1):1–10. doi: 10.1038/s41598-018-37741-x
19. Barragán-Montero AM, Nguyen D, Lu W, Lin M-H, Norouzi-Kandalan R, Geets X, et al. Three-Dimensional Dose Prediction for Lung IMRT Patients With Deep Neural Networks: Robust Learning From Heterogeneous Beam Configurations. *Med Phys* (2019) 46(3):3679–91. doi: 10.1002/mp.13597
20. Liu Y, Chen Z, Wang J, Wang X, Qu B, Ma L, et al. Dose Prediction Using a Three-Dimensional Convolutional Neural Network for Nasopharyngeal Carcinoma With Tomotherapy. *Front Oncol* (2021) 11:752007. doi: 10.3389/fonc.2021.752007
21. Barragán-Montero AM, Thomas M, Defraene G, Michiels S, Haustermans K, Lee JA, et al. Deep Learning Dose Prediction for IMRT of Esophageal Cancer: The Effect of Data Quality and Quantity on Model Performance. *Phys Med* (2021) 83:52–63. doi: 10.1016/j.ejmp.2021.02.026
22. Bentzen SM, Constine LS, Deasy JO, Eisbruch A, Jackson A, Marks LB, et al. Quantitative Analyses of Normal Tissue Effects in the Clinic (QUANTEC): An Introduction to the Scientific Issues. *Int J Radiat Oncol Biol Phys* (2010) 76(3 Suppl):S3–9. doi: 10.1016/j.ijrobp.2009.09.040

Conflict of Interest: Manteia Technologies Co., Ltd employed author ZC.

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: 15 March 2022
doi: 10.3389/fonc.2022.861857

Geometric and Dosimetric Evaluation of the Automatic Delineation of Organs at Risk (OARs) in Non-Small-Cell Lung Cancer Radiotherapy Based on a Modified DenseNet Deep Learning Network

Fuli Zhang[1†], Qiusheng Wang[2†], Anning Yang[2], Na Lu[1], Huayong Jiang[1], Diandian Chen[1], Yanjun Yu[1] and Yadi Wang[1]*

[1] Radiation Oncology Department, The Seventh Medical Center of Chinese People's Liberation Army (PLA) General Hospital, Beijing, China, [2] School of Automation Science and Electrical Engineering, Beihang University, Beijing, China

OPEN ACCESS

Edited by:
Wei Zhao,
Beihang University, China

Reviewed by:
Ruijie Yang,
Peking University Third Hospital, China
Yibao Zhang,
Peking University Cancer Hospital, China

***Correspondence:**
Fuli Zhang
radiozfli@163.com

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: *25 January 2022*
Accepted: *21 February 2022*
Published: *15 March 2022*

Citation:
Zhang F, Wang Q, Yang A, Lu N, Jiang H, Chen D, Yu Y and Wang Y (2022) Geometric and Dosimetric Evaluation of the Automatic Delineation of Organs at Risk (OARs) in Non-Small-Cell Lung Cancer Radiotherapy Based on a Modified DenseNet Deep Learning Network. Front. Oncol. 12:861857. doi: 10.3389/fonc.2022.861857

Purpose: To introduce an end-to-end automatic segmentation model for organs at risk (OARs) in thoracic CT images based on modified DenseNet, and reduce the workload of radiation oncologists.

Materials and Methods: The computed tomography (CT) images of 36 lung cancer patients were included in this study, of which 27 patients' images were randomly selected as the training set, 9 patients' as the testing set. The validation set was generated by cross validation and 6 patients' images were randomly selected from the training set during each epoch as the validation set. The autosegmentation task of the left and right lungs, spinal cord, heart, trachea and esophagus was implemented, and the whole training time was approximately 5 hours. Geometric evaluation metrics including the Dice similarity coefficient (DSC), 95% Hausdorff distance (HD95) and average surface distance (ASD), were used to assess the autosegmentation performance of OARs based on the proposed model and were compared with those based on U-Net as benchmarks. Then, two sets of treatment plans were optimized based on the manually contoured targets and OARs (Plan1), as well as the manually contours targets and the automatically contoured OARs (Plan2). Dosimetric parameters, including Dmax, Dmean and Vx, of OARs were obtained and compared.

Results: The DSC, HD95 and ASD of the proposed model were better than those of U-Net. The differences in the DSC of the spinal cord and esophagus, differences in the HD95 of the spinal cord, heart, trachea and esophagus, as well as differences in the ASD of the spinal cord were statistically significant between the two models ($P<0.05$). The differences in the dose-volume parameters of the two sets of plans were not statistically significant ($P>0.05$). Moreover, compared with manual segmentation, autosegmentation significantly reduced the contouring time by nearly 40.7% ($P<0.05$).

Conclusions: The bilateral lungs, spinal cord, heart and trachea could be accurately delineated using the proposed model in this study; however, the automatic segmentation effect of the esophagus must still be further improved. The concept of feature map reuse provides a new idea for automatic medical image segmentation.

Keywords: non-small-cell lung cancer, organs at risk, medical image segmentation, deep learning, DenseNet, feature reuse

INTRODUCTION

In China, lung cancer ranks first in both incidence and mortality rates, accounting for 17.9% of all new cases and 23.8% of total cancer deaths according to GLOBOCAN 2020 (1). Non-small-cell lung cancer (NSCLC) constitutes the majority of lung cancers. Radiotherapy (RT) is usually used in all stages of NSCLC treatment and is required at least once in more than half of patients for either cure or palliation. In a typical clinical workflow of RT, a radiation oncologist manually segments the tumor target and organs at risk (OARs) based on the information provided by CT, MRI and/or PET/CT images (2, 3). This process is often time consuming and laborious, and the quality of the segmentations largely depends on the experience of radiation oncologists. It is easy to distinguish the organs with high contrast on CT images; however, it is difficult to distinguish the boundary between tumor tissue and surrounding normal tissue with similar contrast. Moreover, inconsistencies in target and OARs segmentations have been reported for both inter-and intraobserver segmentation variability (4–8). These factors will affect the accuracy and efficacy of RT. Therefore, improving the consistency and efficiency of image segmentation becomes an urgent task.

In recent years, automatic medical image segmentation based on deep learning has become a popular research topic in RT, and several convolutional neural networks (CNNs) including U-Net, ResNet and DenseNet, have shown great success in autosegmentation of the target and OARs (9–16). DenseNet was proposed by Huang G et al. (17) in 2017, using the concept of feature map reuse to address the small training datasets in supervised learning. Moreover, DenseNet connects multiple dense blocks with a transition layer and concatenates the channels of each dense block feature map in series to increase the number of feature maps and improve the utilization rate of feature maps. Tong N et al. (18) improved the performance of their previous shape constrained fully CNNs for head and neck OARs segmentation on CT and low field MRI by incorporating generative adversarial network (GAN) and DenseNet. With the novel segmentation method, they showed that the low field MR images acquired on a MR guided radiation radiotherapy system can support accurate and fully automated segmentation of both bony and soft tissue OARs for adaptive radiotherapy. Fu J et al. (19) proposed a novel three-dimensional (3D) multipath DenseNet for generating the accurate glioblastoma (GBM) tumor contour from four multimodal pre-operative MR images. The multipath DenseNet demonstrated an improved accuracy over comparable algorithms in the clinical task of GBM tumor segmentation. To our best knowledge, there has not been an automatic segmentation study based on the DenseNet for NSCLC radiotherapy.

In this study, a deep learning model based on DenseNet and FCN (fully convolutional network) is proposed. The model uses the idea of feature reuse. It learns the planar distribution characteristics of OARs in CT images through a denseblock module and supplements details through long connections to achieve an end-to-end accurate OAR delineation for NSCLC patients.

MATERIALS AND METHODS

Data Acquisition and Preprocessing

The CT images of 36 NSCLC patients of the Seventh Medical Center of the PLA General Hospital were provided. The CT images were scanned on a Philips Brilliance Big Bore simulator (Philips Medical Systems, Madison, WI, USA) from the level of the larynx to the bottom of the lungs with a 3-mm slice thickness on helical scan mode. The study was approved by the Ethics Committee of the Seventh Medical Center of Chinese PLA General Hospital. All of the patients provided written consent for the storage of their medical information in the hospital database. Patients characteristics are shown in **Table 1**. By analyzing the DICOM file, the grayscale value of the original CT image was mapped to the range of 0-255, the window width was set to 400, and the window level was set to 40. Different manual OARs serving as the ground truth were filled with different grayscale values to generate mask images as training labels, as shown in **Figure 1**.

The training dataset included 3803 CT images of 27 patients. The testing set included 567 images of 9 patients. In order to improve the utilization of the data and obtain a more stable model, the validation set was generated by cross validation and 6 patients' images were randomly selected from the training set during each epoch as the validation set. After data cleaning and augmentation, these images were sent to the proposed model. The deep learning inference platforms used Tensorflow-gpu 1.7.0 as the underlying framework, Keras2.2.4 neural network library and python (version 3.6). All training, validating and testing were run on an NVIDIA GeForce GTX 1070 Ti GPU with 8 GB video memory. The starting and ending times of the manual and autosegmentation operations for each patient in the testing set were recorded.

TABLE 1 | Characteristics of patients in the training and testing sets.

Characteristics	Training set	Testing set
No. patients	27	9
Tumor site, right:left	16:11	3:6
Lobe location		
Upper left	7	5
Lower left	4	1
Upper right	7	1
Middle right	5	1
Lower right	3	1

The Proposed Model for Segmentation

In this study, the model was trained to realize the autosegmentation of six OARs for NSCLC. The specific architecture of the model is shown in **Figure 2**. The segmentation process was mainly divided into two parts: the left half was called the analysis path, composed of a dense block module and a transition down module and connected by a short cut layer to extract image features; the right half was called the synthesis path, upsampled by a transition up transposition convolution module to recover the size of the feature map layer. To improve the accuracy of the reconstructed image and accelerate the convergence process of the network parameters, the feature maps of the same size in the analysis path were connected in series as the input of the next layer of the dense block.

The input of each layer of the dense block was intensively composed of all of the outputs of its front layer after dense connection (as shown in **Figure 3**). The output of each layer had the following corresponding functional relationship with the output of other front layers:

$$X_{i+1} = H(\mathrm{X}_1, X_2, \ldots, X_i) \tag{1}$$

where $H(*)$ is a nonlinear function denoting a series of operations, including batch normalization (BN), ReLU activation, pooling and convolution, which are used to extract features, adjust the size of the feature map and reduce the channel dimension. The bottleneck architecture was set in each network since the operation of dense connections could bring about a surge in the number of channels and increase the difficulty of training. The bottleneck architecture used a 1×1 convolution kernel to realize cross-channel feature fusion and enhance the feature extraction ability of the network.

Training of The Proposed Model

After cleaning and augmentation, data were sent to the model for training. The weight and bias of the network were updated using the cross entropy loss function as follows.

$$L_s = -\sum_{i=1}^{k} (\mathrm{y} \log \widehat{y} + (1 - \mathrm{y}) \log (1 - \widehat{y})) \tag{2}$$

$$\widehat{y} = (1 + e^{w^T x + b})^{-1} \tag{3}$$

where x is the input of the network, $\hat{\mathrm{y}}$ is the posterior probability output after network regression and k is the number of categories.

In this study, the early stopping module was added to detect the network accuracy and loss function value with the increase in the number of iterative epochs, and the network architecture based on DensNet56 in the 30th epoch was selected. During the network training process, the initial learning rate was set as 1e-3 and decreased with increasing epochs. This process ensured that the network could converge quickly in the initial stage of training, on the one hand, and avoided the problem of poor feature generalization due to network overfitting, on the other hand. In order to prevent the performance of the network from swinging at the local optimum, the Adam optimizer was used for training error. The Adam optimizer introduced the concept of second-order momentum, and the network weight was updated as the learning rate multiplied by the ratio of the gradient average to the square root of the gradient variance. The advantage of the method was that gradient updating was not only affected by the current gradient; but also by the accumulated gradient updating (20). The average segmentation time for the training set is approximately 12.58 min/epoch, the average segmentation time for a single 512×512 CT image is approximately 0.17 s, and the time for delineating all CT images of a patient is approximately 13.4 s.

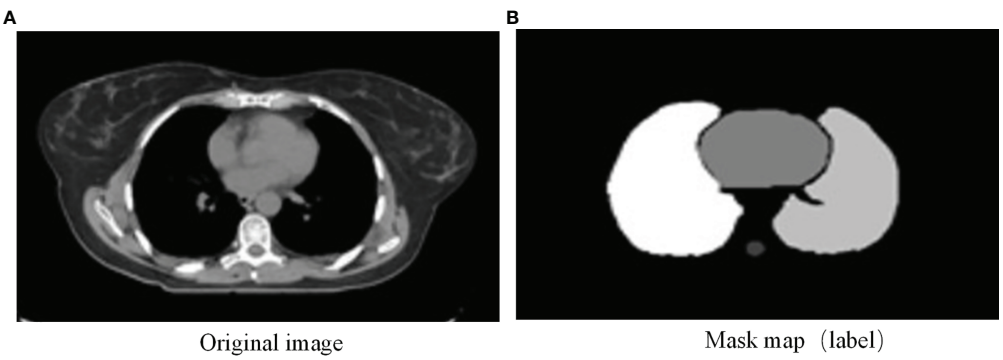


FIGURE 1 | **(A)** Original image and **(B)** mask map (label).

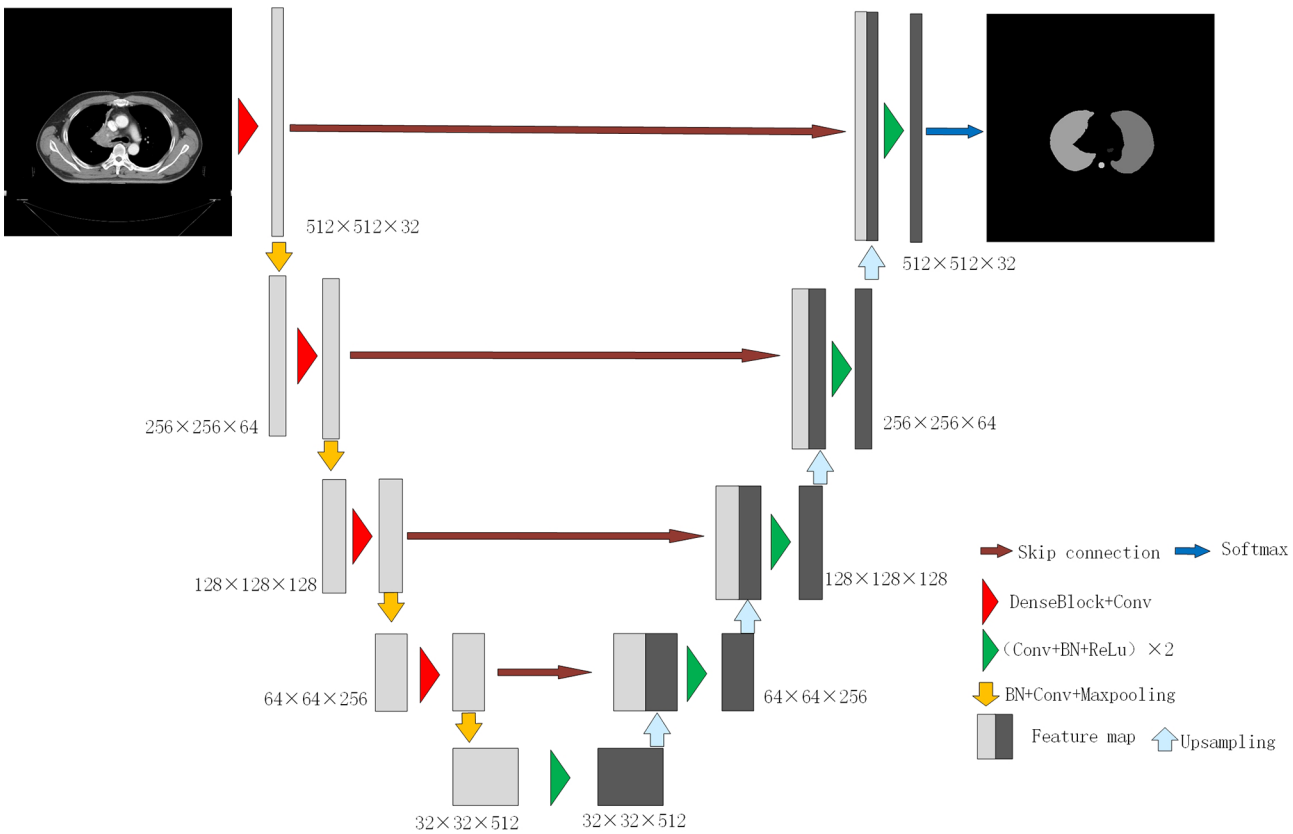


FIGURE 2 | The architecture of the proposed model.

Accuracy Evaluation

Geometric Metrics

In this study, geometric evaluation metrics, including the Dice similarity coefficient (DSC), 95% Hausdorff distance (HD95) and average surface distance (ASD) (21), were used to assess the autosegmentation results of OARs based on the proposed model and were compared with those based on U-Net as benchmarks.

Dosimetric Metrics

To assess the dosimetric impact of the proposed model on treatment planning, we designed two sets of stereotactic body radiation therapy (SBRT) treatment plans for each patient in the testing set using manually segmented target volumes and OARs (Plan1), as well as the manually segmented target volumes and automatically segmented OARs (Plan2). Intensity modulated radiotherapy (IMRT) treatment plans were optimized with 6-MV photons using 5 coplanar beams. All of the plans were prescribed 6 Gy per fraction for 10 fractions and normalized as 100% prescription dose to 95% of the planning target volume (PTV). Dosimetric parameters including Dmax (meaning the dose received by 2% of the volume), Dmean, V40, V30, V20, V10, and V5 (meaning the volume receiving more than x Gy dose as a percentage of the total volume), were obtained and compared to

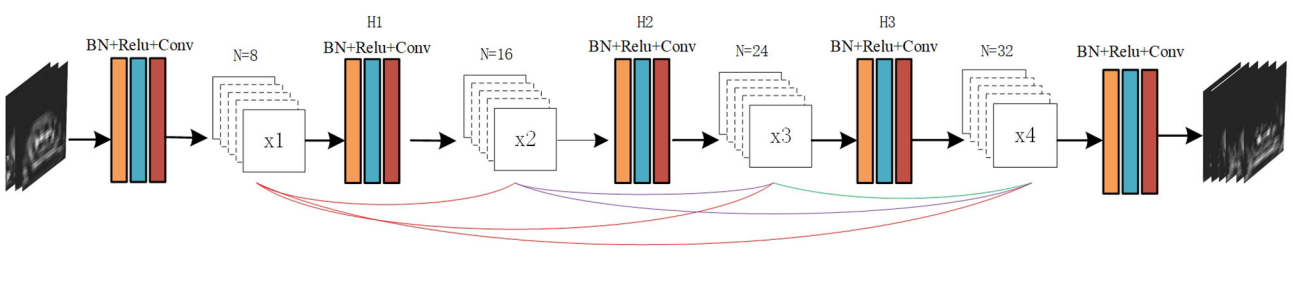


FIGURE 3 | Scheme of dense block.

assess the clinical feasibility of the proposed model. The dosimetric characteristics of OARs were gauged by the conformity index (CI) and homogeneity index (HI) of the PTV, so the CI and HI of the PTV were also calculated according to the formula in reference (22).

Statistical Analysis

SPSS statistical software (version 20.0, SPSS Inc., Chicago, IL, USA) was used for statistical analysis. Wilcoxon's signed rank test was used to compare the differences in DSC, HD95, ASD, and dosimetric parameters. Quantitative data are expressed as the mean ± standard deviation ($\bar{x} \pm s$), a value of $P< 0.05$ was considered statistically significant.

RESULTS

Geometric Metrics

The DSC, HD95, and ASD of OARs based on the proposed model and U-Net are listed in **Tables 2–4**, respectively. The proposed model showed better performance than U-Net, although there was no significant difference between the two models in several OARs ($P>0.05$). The comparison of the results between manual and automatic segmentation based on the proposed model for a typical patient is shown in **Figure 4**.

Dosimetric Metrics

The dose-volume parameters of the OARs based on manual and automatic segmentation are listed in **Table 5**. There were no statistically significant differences between the dosimetric parameters of manual and automatically delineated OARs ($P>0.05$). The CIs of PTV in Plan1 and Plan2 were 0.74 ± 0.07 and 0.73 ± 0.07, respectively, while the HIs of PTV in Plan2 were 0.10 ± 0.02 and 0.09 ± 0.02, respectively. The differences in both CI and HI were not statistically significant ($P>0.05$).

Delineating Time Analysis

The average time for manual segmentation by experienced radiation oncologists for 9 patients in the testing set was 15.2 min, while the total autosegmentation time of the 9 patients in the testing set was 9.0 min. Autosegmentation greatly improved the working efficiency in contouring the OARs ($P<0.05$).

DISCUSSION

The results of this study are relatively consistent with those of the challenge report of automatic segmentation of thoracic organs organized by the American Association of Physicists in Medicine (AAPM)'s annual meeting in 2017 (21),with the right lung having the highest average DSC (0.96) and the esophagus having the lowest average DSC (0.67). Compared with U-Net, the autosegmentation results of the OARs based on the proposed model were better with higher DSC as well as lower HD95 and ASD. Among them, DSC differences of the spinal cord and esophagus, HD95 differences of the spinal cord, heart, trachea and esophagus, as well as ASD difference of the spinal cord were statistically significant (P<0.05).

Lustberg T et al. (23) used a deep learning autosegmentation software (Mirada) to create thoracic OARs contours and the model was built by using 450 lung patients' images as the training set. For 20 CT scans of stage I-III NSCLC patients in the testing set, the median DSCs of the spinal cord, the lungs, and heart were 0.83, >0.95, >0.90, respectively. Zhang T et al. (24) developed a 2D AS-CNN based on the ResNet101 network using a training dataset of 200 lung cancer patients. The average DSCs of the left lung, right lung, heart, spinal cord, and esophagus of 19 NSCLC patients were 0.94, 0.94, 0.89, 0.82, and 0.73, respectively. Zhu JH et al. (25) proposed an automatic segmentation model based on depth convolution to segment CT images from 36 lung cancer patients. The average DSCs of the lungs, heart, liver, spinal cord and esophagus were 0.95, 0.91, 0.89, 0.76 and 0.64, respectively. Dong X et al. proposed a U-Net-generative adversarial network (U-Net-GAN) and realized the segmentation of 5 thoracic OARs. Among them, the left lung, right lung, and heart were automatically segmented by a 2.5D GAN model, while the esophagus and spinal cord were automatically segmented by a 3D GAN model. The average DSCs of the left and right lungs, spinal cord, esophagus, and heart were 0.97,0.97, 0.90, 0.75, and o.87, respectively. He T et al. (26) proposed a uniform U-like encoder-decoder architecture abstracted from the U-Net and trained it using 40 patients' thoracic CT scans. High DSC values were obtained for esophagus (0.86), heart (0.95), trachea(0.92) and aorta (0.95) from 20 patients in the testing set. Feng X et al. (27) developed a model based on 3D U-Net to autosegment thoracic OARs using 36 thoracic CT scans as the training set. The performance of the model was evaluated on two groups of testing set consisting of 12 patients and 30 patients, respectively. The average DSCs of the spinal cord, right lung, left lung, heart and esophagus of the first testing set reached 0.89, 0.97, 0.98, 0.93, and 0.73 while those of the second testing set were 0.85, 0.98, 0.98, 0.86 and 0.69, respectively.

The differences in all dosimetric metrics of the OARs between manual and automatic delineations were not statistically significant (P>0.05) in our study. The maximum dosimetric metrics differences were 0.41Gy for Dmean of the trachea and 0.64% for V5 of bilateral lungs, while the clinically acceptable dose difference and volume difference of OARs between manual and automatic delineation are supposed to be within 1Gy and

TABLE 2 | Comparison of DSC of two models ($\bar{x} \pm s$).

	Spinal cord	Heart	Right lung	Left lung	Trachea	Esophagus
U-Net	0.82 ± 0.04	0.83 ± 0.09	0.96 ± 0.02	0.94 ± 0.02	0.86 ± 0.07	0.55 ± 0.11
Proposed	0.89 ± 0.01	0.86 ± 0.09	0.96 ± 0.01	0.95 ± 0.02	0.91 ± 0.03	0.67 ± 0.12
P value	0.008	0.535	0.897	0.709	0.212	0.008

TABLE 3 | Comparison of HD95 (mm) of two models ($\bar{x} \pm s$).

	Spinal cord	Heart	Right lung	Left lung	Trachea	Esophagus
U-Net	3.75 ± 1.23	14.42 ± 2.94	7.24 ± 4.22	9.04 ± 5.97	4.46 ± 2.61	12.40 ± 5.99
Proposed	2.05 ± 0.38	9.75 ± 2.34	6.09 ± 1.56	6.47 ± 3.27	2.44 ± 1.17	6.14 ± 3.07
P value	0.000	0.008	0.897	0.260	0.039	0.008

TABLE 4 | Comparison of ASD (mm) of two models ($\bar{x} \pm s$).

	Spinal cord	Heart	Right lung	Left lung	Trachea	Esophagus
U-Net	2.01 ± 0.70	7.70 ± 6.10	1.32 ± 0.45	1.57 ± 0.65	1.42 ± 0.87	6.95 ± 7.30
Proposed	0.81 ± 0.18	5.93 ± 4.03	1.11 ± 0.31	1.23 ± 0.54	0.94 ± 0.51	3.27 ± 2.67
P value	0.000	0.425	0.375	0.264	0.281	0.123

1%, respectively. Zhu J et al. (28) evaluated the performance of automatic segmentation of the OARs with dosimetric metrics for esophageal cancer patients. The maximum metrics differences were 0.35 Gy for Dmax of the spinal cord and 0.4% for V30 of bilateral lungs. The results in our study were consistent with those of the above study.

Due to the different training datasets, it is difficult to compare the advantages and disadvantages of the proposed model and the published model. However, the number of training cases used in our study was obviously fewer; the proposed model has strong feature extraction ability in the training of small samples, and the segmentation results are similar to those of the training model of relatively large datasets. A limitation of this study needs to be pointed out. That is, due to low soft tissue contrast, small volume, and large shape variability across patients, the automatic segmentation results of the esophagus are not ideal, and the DSC value is lower than 0.7, which is clinically unacceptable (29, 30), therefore, we did not take into account the esophagus when analyzing the dose-volume parameters in the treatment plan. In the next work, we need to further optimize the model and expand the size of data to increase its generalization and segmentation effect.

Currently, there are three main development directions for deep learning networks in medical image segmentation. The first direction is to deepen the network level and depth, extract deeper semantic features to obtain stronger expression ability, or widen the network to increase the number of channels to obtain more information in the same layer, such as the texture features of different grayscales and boundary features in different directions. The second direction is to obtain a more effective spatial feature extraction ability by learning the sequence concatenation properties of multiple CT slices of a patient, represented by 3D U-Net and many other derivative networks. The third direction represented by DenseNet is to improve the utilization rate of the feature map by sharing the feature map layer by layer to enhance the feature expression ability of the image and improve the generalization performance of the network (31).

CONCLUSION

Compared with U-Net, the proposed model based on DenseNet is better in the OARs segmentation task; even if the training set has fewer images, it can still fairly effectively prevent the

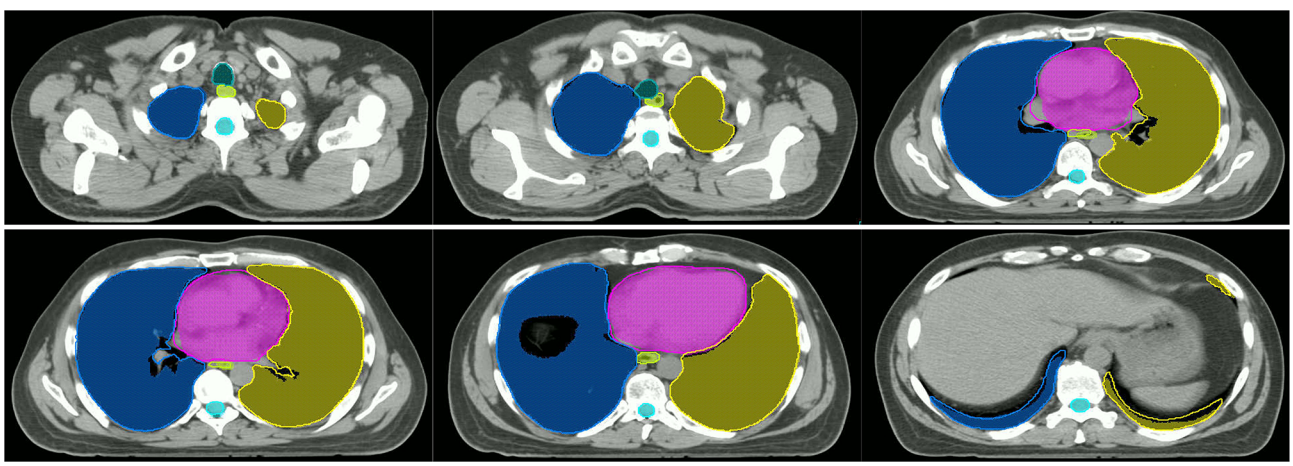


FIGURE 4 | Comparison of manual and automatic segmentation of the OARs based on the proposed model (Color wash: the manual segmentation contour; line: the automatic segmentation contour).

TABLE 5 | Dosimetric comparison of the PTV and OARs between manual and automatic segmentation-based plans ($\bar{x} \pm s$).

Dosimetric parameters		Plan1	Plan2	*P* value
PTV	CI	0.74 ± 0.07	0.73 ± 0.07	0.859
	HI	0.10 ± 0.02	0.09 ± 0.02	0.139
Spinalcord	Dmax (Gy)	18.66 ± 7.95	18.71 ± 7.36	0.678
Heart	V30 (%)	0.42 ± 1.16	0.43 ± 1.19	0.655
	V40 (%)	0.22 ± 0.63	0.22 ± 0.67	0.655
	Dmean (Gy)	1.71 ± 1.58	1.61 ± 1.53	0.441
Lung All	V5 (%)	27.55 ± 6.81	28.19 ± 6.78	0.515
	V10 (%)	15.49 ± 4.41	14.83 ± 4.54	0.953
	V20 (%)	9.40 ± 3.69	9.44 ± 3.89	0.859
	V30 (%)	6.57 ± 3.25	6.64 ± 3.36	0.263
	Mean (Gy)	6.49 ± 1.94	6.56 ± 2.01	0.173
Lung_L	V5 (%)	31.34 ± 14.34	30.80 ± 13.94	0.260
	V10 (%)	17.41 ± 17.79	17.16 ± 17.71	0.310
	V20 (%)	12.64 ± 14.34	12.48 ± 14.38	0.225
	V30 (%)	9.82 ± 11.68	9.69 ± 11.56	0.173
	Dmean (Gy)	8.07 ± 6.65	7.99 ± 6.65	0.477
Lung_R	V5 (%)	24.39 ± 8.38	24.48 ± 8.59	0.678
	V10 (%)	12.78 ± 11.37	12.89 ± 11.48	0.314
	V20 (%)	6.64 ± 8.31	6.69 ± 8.44	0.686
	V30 (%)	4.02 ± 5.28	4.07 ± 5.44	0.715
	Dmean (Gy)	5.14 ± 3.16	5.14 ± 3.29	0.477
Trachea	Dmean	6.00 ± 3.74	5.59 ± 3.47	0.139

occurrence of overfitting. At the same time, it can effectively alleviate the problem of the gradient disappearing in the training process by repeatedly using different levels of feature maps, providing a new idea for medical image segmentation.

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 Seventh Medical Center of Chinese PLA General Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FZ and QW contributed conception and design of the study. FZ and AY trained the deep learning models. FZ and QW performed data analysis and drafted the manuscript. NL, HJ, DC, and YW helped to collect the data and evaluate radiotherapy planning. YY designed radiotherapy planning. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Beijing Municipal Science and Technology Commission (No.Z181100001718011).

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. Eaton BR, Pugh SL, Bradley JD, Masters G, Kavadi VS, Narayan S, et al. Institutional Enrollment and Survival Among NSCLC Patients Receiving Chemoradiation: NRG Oncology Radiation Therapy Oncology Group (RTOG) 0617. *J Natl Cancer Inst* (2016) 108(9):djw034. doi: 10.1093/jnci/djw034
3. Wang EH, Rutter CE, Corso CD, Decker RH, Wilson LD, Kim AW, et al. Patients Selected for Definitive Concurrent Chemoradiation at High-Volume Facilities Achieve Improved Survival in Stage III Non-Small-Cell Lung Cancer. *J Thorac Oncol* (2015) 10(6):937–43. doi: 10.1097/JTO.0000000000000519
4. Martin S, Johnson C, Brophy M, Palma DA, Barron JL, Beauchemin SS, et al. Impact of Target Volume Segmentation Accuracy and Variability on Treatment Planning for 4D-CT-Based non-Small Cell Lung Cancer Radiotherapy. *Acta Oncol* (2015) 54(3):322–32. doi: 10.3109/0284186X.2014.970666
5. Speight R, Sykes J, Lindsay R, Franks K, Thwaites D. The Evaluation of a Deformable Image Registration Segmentation Technique for Semi-Automating Internal Target Volume (ITV) Production From 4dct Images of Lung Stereotactic Body Radiotherapy (SBRT) Patients. *Radiother Oncol* (2011) 98(2):277–83. doi: 10.1016/j.radonc.2010.12.007
6. van Dam IE, van Sornsen de Koste JR, Hanna GG, Muirhead R, Slotman BJ, Senan S. Improving Target Delineation on 4-Dimensional CT Scans in Stage I NSCLC Using a Deformable Registration Tool. *Radiother Oncol* (2010) 96 (1):67–72. doi: 10.1016/j.radonc.2010.05.003
7. Jameson MG, Holloway LC, Vial PJ, Vinod SK, Metcalfe PE. A Review of Methods of Analysis in Contouring Studies for Radiation Oncology. *J Med*

Imaging Radiat Oncol (2010) 54(5):401–10. doi: 10.1111/j.1754-9485.2010.02192.x
8. Brouwer CL, Steenbakkers RJ, van den Heuvel E, Duppen JC, Navran A, Bijl HP, et al. 3d Variation in Delineation of Head and Neck Organs at Risk. *Radiat Oncol* (2012) 7:32. doi: 10.1186/1748-717X-7-32
9. Wong J, Fong A, McVicar N, Smith S, Giambattista J, Wells D, et al. Comparing Deep Learning-Based Auto-Segmentation of Organs at Risk and Clinical Target Volumes to Expert Inter-Observer Variability in Radiotherapy Planning. *Radiother Oncol* (2020) 144:152–8. doi: 10.1016/j.radonc.2019.10.019
10. Wang Z, Chang Y, Peng Z, Lv Y, Shi W, Wang F, et al. Evaluation of Deep Learning-Based Auto-Segmentation Algorithms for Delineating Clinical Target Volume and Organs at Risk Involving Data for 125 Cervical Cancer Patients. *J Appl Clin Med Phys* (2020) 21(12):272–9. doi: 10.1002/acm2.13097
11. Men K, Dai J, Li Y. Automatic Segmentation of the Clinical Target Volume and Organs at Risk in the Planning CT for Rectal Cancer Using Deep Dilated Convolutional Neural Networks. *Med Phys* (2017) 44(12):6377–89. doi: 10.1002/mp.12602
12. Vrtovec T, Mocnik D, Strojan P, Pernus F, Ibragimov B. Auto-Segmentation of Organs at Risk for Head and Neck Radiotherapy Planning: From Atlas-Based to Deep Learning Methods. *Med Phys* (2020) 47(9):e929–e50. doi: 10.1002/mp.14320
13. Kholiavchenko M, Sirazitdinov I, Kubrak K, Badrutdinova R, Kuleev R, Yuan Y, et al. Contour-Aware Multi-Label Chest X-Ray Organ Segmentation. *Int J Comput Assist Radiol Surg* (2020) 15(3):425–36. doi: 10.1007/s11548-019-02115-9
14. Lin L, Dou Q, Jin YM, Zhou GQ, Tang YQ, Chen WL, et al. Deep Learning for Automated Contouring of Primary Tumor Volumes by MRI for Nasopharyngeal Carcinoma. *Radiology* (2019) 291(3):677–86. doi: 10.1148/radiol.2019182012
15. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: *Medical Image Computing and Computer-Assisted Intervention-MICCAI*. Berlin: Springer. (2015) 234–41.
16. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas (2016) 770–8.
17. Huang G, Liu Z, Van Der Maaten L, Weinberger XQ. Densely Connected Convolutional Networks. In: *IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu. (2017) 4700-8.
18. Tong N, Gou S, Yang S, Cao M, Sheng K. Shape Constrained Fully Convolutional DenseNet With Adversarial Training for Multiorgan Segmentation on Head and Neck CT and Low-Field MR Images. *Med Phys* (2019) 46(6):2669–82. doi: 10.1002/mp.13553
19. Fu J, Singhrao K, Qi XS, Yang Y, Ruan D, Lewis JH. Three-Dimensional Multipath DenseNet for Improving Automatic Segmentation of Glioblastoma on Pre-Operative Multimodal MR Images. *Med Phys* (2021) 48(6):2859–66. doi: 10.1002/mp.14800
20. Kingma D, Ba J. Adam: A Method for Stochastic Optimization. In: *International Conference for Learning Representations*. San Diego. (2015).
21. Yang J, Veeraraghavan H, Armato SG3rd, Farahani K, Kirby JS, Kalpathy-Kramer J, et al. Autosegmentation for Thoracic Radiation Treatment Planning: A Grand Challenge at AAPM 2017. *Med Phys* (2018) 45 (10):4568–81. doi: 10.1002/mp.13141
22. International Commission on Radiation Units and Measurements. Prescribing, Recording and Reporting Photon Beam Intensity-Modulated Radiation Therapy (IMRT). ICRU Report 83. (2010) 10(1):34–6.
23. 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(2):312–7. doi: 10.1016/j.radonc.2017.11.012
24. Zhang T, Yang Y, Wang J, Men K, Wang X, Deng L, et al. Comparison Between Atlas and Convolutional Neural Network Based Automatic Segmentation of Multiple Organs at Risk in non-Small Cell Lung Cancer. *Med (Baltimore)* (2020) 99(34):e21800. doi: 10.1097/MD.0000000000021800
25. Zhu J, Zhang J, Qiu B, Liu Y, Liu X, Chen L. Comparison of the Automatic Segmentation of Multiple Organs at Risk in CT Images of Lung Cancer Between Deep Convolutional Neural Network-Based and Atlas-Based Techniques. *Acta Oncol* (2019) 58(2):257–64. doi: 10.1080/0284186X.2018.1529421
26. He T, Hu J, Song Y, Guo J, Yi Z. Multi-Task Learning for the Segmentation of Organs at Risk With Label Dependence. *Med Image Anal* (2020) 61:101666. doi: 10.1016/j.media.2020.101666
27. Feng X, Qing K, Tustison NJ, Meyer CH, Chen Q. Deep Convolutional Neural Network for Segmentation of Thoracic Organs-At-Risk Using Cropped 3d Images. *Med Phys* (2019) 46(5):2169–80. doi: 10.1002/mp.13466
28. Zhu J, Chen X, Yang B, Bi N, Zhang T, Men K, et al. Evaluation of Automatic Segmentation Model With Dosimetric Metrics for Radiotherapy of Esophageal Cancer. *Front Oncol* (2020) 10:564737. doi: 10.3389/fonc.2020.564737
29. Conson M, Cella L, Pacelli R, Comerci M, Liuzzi R, Salvatore M, et al. Automated Delineation of Brain Structures in Patients Undergoing Radiotherapy for Primary Brain Tumors: From Atlas to Dose-Volume Histograms. *Radiother Oncol* (2014) 112(3):326–31. doi: 10.1016/j.radonc.2014.06.006
30. Dolz J, Kirisli HA, Fechter T, Karnitzki S, Oehlke O, Nestle U, et al. Interactive Contour Delineation of Organs at Risk in Radiotherapy: Clinical Evaluation on NSCLC Patients. *Med Phys* (2016) 43(5):2569. doi: 10.1118/1.4947484
31. Ke L, Deng Y, Xia W, Qiang M, Chen X, Liu K, et al. Development of a Self-Constrained 3d DenseNet Model in Automatic Detection and Segmentation of Nasopharyngeal Carcinoma Using Magnetic Resonance Images. *Oral Oncol* (2020) 110:104862. doi: 10.1016/j.oraloncology.2020.104862

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.2022.833816

Evaluation Exploration of Atlas-Based and Deep Learning-Based Automatic Contouring for Nasopharyngeal Carcinoma

*Jinyuan Wang[1], Zhaocai Chen[2], Cungeng Yang[2], Baolin Qu[1], Lin Ma[1], Wenjun Fan[1], Qichao Zhou[2], Qingzeng Zheng[3] and Shouping Xu[1]**

[1] Department of Radiation Oncology, The First Medical Center of the Chinese PLA General Hospital, Beijing, China, [2] Manteia Technologies Co., Ltd., Xiamen, China, [3] Department of Radiation Oncology, Beijing Geriatric Hospital, Beijing, China

OPEN ACCESS

Edited by:
Xiaomeng Li,
Hong Kong University of Science and Technology, Hong Kong SAR, China

Reviewed by:
Guibin Zan,
Stanford University, United States
Weiwei Zong,
Henry Ford Health System, United States
Xiaowei Xu,
Guangdong Provincial People's Hospital, China

***Correspondence:**
Shouping Xu
shouping_xu@yahoo.com

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: 12 December 2021
Accepted: 25 February 2022
Published: 31 March 2022

Citation:
Wang J, Chen Z, Yang C, Qu B, Ma L, Fan W, Zhou Q, Zheng Q and Xu S (2022) Evaluation Exploration of Atlas-Based and Deep Learning-Based Automatic Contouring for Nasopharyngeal Carcinoma. Front. Oncol. 12:833816. doi: 10.3389/fonc.2022.833816

Purpose: The purpose of this study was to evaluate and explore the difference between an atlas-based and deep learning (DL)-based auto-segmentation scheme for organs at risk (OARs) of nasopharyngeal carcinoma cases to provide valuable help for clinical practice.

Methods: 120 nasopharyngeal carcinoma cases were established in the MIM Maestro (atlas) database and trained by a DL-based model (AccuContour®), and another 20 nasopharyngeal carcinoma cases were randomly selected outside the atlas database. The experienced physicians contoured 14 OARs from 20 patients based on the published consensus guidelines, and these were defined as the reference volumes (V_{ref}). Meanwhile, these OARs were auto-contoured using an atlas-based model, a pre-built DL-based model, and an on-site trained DL-based model. These volumes were named V_{atlas}, $V_{DL\text{-}pre\text{-}built}$, and $V_{DL\text{-}trained}$, respectively. The similarities between V_{atlas}, $V_{DL\text{-}pre\text{-}built}$, $V_{DL\text{-}trained}$, and V_{ref} were assessed using the Dice similarity coefficient (DSC), Jaccard coefficient (JAC), maximum Hausdorff distance (HD_{max}), and deviation of centroid (DC) methods. A one-way ANOVA test was carried out to show the differences (between each two of them).

Results: The results of the three methods were almost similar for the brainstem and eyes. For inner ears and temporomandibular joints, the results of the pre-built DL-based model are the worst, as well as the results of atlas-based auto-segmentation for the lens. For the segmentation of optic nerves, the trained DL-based model shows the best performance ($p < 0.05$). For the contouring of the oral cavity, the DSC value of $V_{DL\text{-}pre\text{-}built}$ is the smallest, and $V_{DL\text{-}trained}$ is the most significant ($p < 0.05$). For the parotid glands, the DSC of V_{atlas} is the minimum (about 0.80 or so), and $V_{DL\text{-}pre\text{-}built}$ and $V_{DL\text{-}trained}$ are slightly larger (about 0.82 or so). In addition to the oral cavity, parotid glands, and the brainstem, the maximum Hausdorff distances of the other organs are below 0.5 cm using the trained DL-based segmentation model. The trained DL-based segmentation method behaves well in the contouring of all the organs that the maximum average deviation of the centroid is no more than 0.3 cm.

Conclusion: The trained DL-based segmentation performs significantly better than atlas-based segmentation for nasopharyngeal carcinoma, especially for the OARs with small volumes. Although some delineation results still need further modification, auto-segmentation methods improve the work efficiency and provide a level of help for clinical work.

Keywords: atlas, deep learning (DL), training, nasopharyngeal carcinoma (NPC), auto-segmentation, organs at risk (OARs)

INTRODUCTION

Nasopharyngeal carcinoma (NPC) is a common malignant tumor in the head and neck region. Its annual incidence rate is about 25 to 30 cases per 100,000 people, and this rate is increasing year by year (1, 2). Most patients with NPC will undergo radiotherapy. Radiation delivers the target area, inevitably exposing normal tissue around the target. The incidence of adverse symptoms occurs directly related to the dose received by organs at risk (OARs) (3, 4). Since eliminating the unnecessary irradiation dose of organs can reduce the side effects and improve life quality, precise radiation therapy (RT) becomes particularly important (5). The rapid development of new technology has continuously improved the accuracy of RT. Nonetheless, the precise delineation of tumors and OARs before RT is essential. Therefore, we can quantitatively evaluate OARs and thus ensure proper treatment after dose calculation.

Due to the lack of contrast to CT images, the indistinct boundaries, and the numerous organs, physicians need to delineate the targets and OARs manually for NPC. This process becomes cumbersome; it also requires time and workforce, resulting in a relatively low efficiency (6, 7). To simplify the heavy task of contouring, many software tools for automatic delineation have appeared from the market, most of which use atlas-based auto-segmentation (ABAS) methods (8–11), simultaneously, with the application of machine learning technique and intense learning methods in RT (12–14). These artificial neural network-based methods may offer better performance. Nevertheless, manual delineation is still the standard procedure for most medical institutions.

This study intends to compare the results from an atlas-based and deep learning (DL)-based segmentation method of organs at risk (OARs) for NPC to evaluate the difference between the two methods and explore conclusions for clinical practice.

MATERIALS AND METHODS

Data Source and Study Design

This retrospective study included a total of 140 patients with NPC who were treated at our institution from July 2016 to July 2018. All the CT image data we selected were acquired in supine position with a thermoplastic head and neck mask for each patient, using a Siemens SOMATOM Definition AS CT scanner with a selected slice thickness of 3 mm, a valid mAs of 300, a tube voltage of 120 kV, and a matrix of 512 × 512. We excluded the cases that were too fat or too slim because the head and neck region was not so big as other parts of the body, such as the abdomen; there were not so many cases that did not meet. After scanning, we loaded the CT images into TPS (Pinnacle, Version 10.0). According to the published consensus guidelines and the fusion with MR or contrast CT, all the OAR delineations were done manually by the same experienced clinician, defined as V_{ref} (V_{ref}, the reference volume). In addition, the study randomly selected 120 CT images and their structure files then transferred them into MIM Maestro software (Version 6.6.5) to establish the atlas library, and we used the same 120 CT images and their structure files to train the DL-based model (AccuLearning®, a Commercial Company). Also, the pre-built automatic segmentation model (AccuContour®, a Commercial Company) is used for comparison. Finally, we used the atlas-based, pre-built, and on-site trained DL-based automatic segmentation methods to automatically delineate the remaining 20 patients and define them as V_{atlas}, $V_{DL\text{-}pre\text{-}built}$, and $V_{DL\text{-}trained}$, respectively. This study selected 14 OARs, including brainstem, eyes, lens, optic nerves, inner ears, temporomandibular joints (TMJs), parotid glands, and oral cavity.

Automatic Segmentation Methods

Atlas-based auto-segmentation used CT images with current delineation data as template images to build an atlas database, compared the new CT images with the others in the databases, then found the best match. Moreover, the new CT images were registered to the template CT images using intensity-based deformable registration. We transformed the contours of the template CT into the new CT image through the deformation registration parameters and obtained the final contouring result. Our study chose a uniform case as the atlas template CT (reference CT). The new cases were added as a subset into the atlas library after the re-registration with the template CT. Moreover, we used the majority vote algorithm and match number 3 to do the atlas auto-segmentation using MIM Maestro software.

The network model structure in our study is residue-Unet, with the loss function of Dice function integrated into AccuContour produced by a commercialized product. Both encoder and decoder are composed of 5 cascades of Residual blocks, and each Residual block is composed of two convolution layers with the convolution kernel size of 3 × 3. Each residual block is cascaded with the downsampling and upsampling layers. The downsampling method is maximum pooling, and the upsampling method is a nearest-neighbor interpolation. The

pre-built auto-segmentation model was based on thousands of finely labeled cases from various medical centers except our institution. The on-site trained auto-segmentation model was based on the cases from our institution. According to information disclosed to the public, the developer utilized standard methods including supervised or semi-supervised learning, prior knowledge aggregation, and ROI-specific post-processing to obtain the pre-built model. The company also provides a high-performance DL research platform called AccuLearning for user-customizable on-site training needs. Preprocessing standardizes and resamples images using adaptive rules based on image intensity range and distribution characteristics. Data augmentation is first conducted on the loaded images, and then balanced cropping is performed accordingly to the label area to generate the model input. Model training utilizes an adaptive network structure adjusting strategy based on gradient feedback and the loss function's progress to accommodate the training dataset's characteristics.

Quantitative Evaluation

We evaluated the results using four parameters, including Dice similarity coefficient (DSC), Jaccard coefficient (JAC), maximum Hausdorff distance (HD_{max}), and deviation of the centroid (DC). One RT physicist performed all atlas and DL delineation tasks to prevent any possible differences among the operators.

1. Dice similarity coefficient (DSC) (15)

The DSC is defined to be the ratio of the intersection to the average area.

$$DSC = \frac{2 \cdot (V_{ref} \cap V_{auto})}{V_{ref} + V_{auto}}$$

where V_{ref} is the reference delineation volume and V_{auto} is the auto-segmentation volume.

2 Jaccard index (JAC) (16)

The JAC is defined to be the ratio of the intersection to the union area.

$$JAC = \frac{V_{ref} \cap V_{auto}}{V_{ref} \cup V_{auto}}$$

where V_{ref} is the reference delineation volume and V_{auto} is the auto-segmentation volume.

3 Maximum Hausdorff distance (HD_{max}) (17)

Suppose there are two groups of sets $X = \{x_1, \ldots, x_n\}$, $Y=\{y_1, \ldots, y_n\}$, then the maximum Hausdorff distance between these two sets of points is defined as

$$HD(X,Y) = \max(h(X,Y), h(Y,X))$$

where $h(X,Y) = \max_{x \in X} \min_{y \in Y} \| x - y \|$

4 Deviation of centroid (DC) (18)

The DC is defined as the deviation of centroid of two volumes.

$$DC = \sqrt{(x_{auto} - x_{ref})^2 + (y_{auto} - y_{ref})^2 + (z_{auto} - z_{ref})^2}$$

Statistical Analysis Methods

The one-way ANOVA test (SPSS, Version 23; SPSS Inc, Chicago, USA) was applied to measure the difference in evaluation parameters with an LSD (least significance difference) method to do the *post hoc* multiple comparisons (V_{atlas} vs. $V_{DL\text{-}pre\text{-}built}$ vs. $V_{DL\text{-}trained}$). A p-value less than 0.05 (typically ≤0.05) was statistically significant.

RESULTS

Figures 1, **2** show the contouring results of two real clinical test cases of the three methods, respectively, where **Figures 1A–D**, **2A–D** show an identical case, and **Figures 1E–H**, **2E–H** show the same case similarly. To an intuitive evaluation of the contouring accuracy, the results of four quantitative evaluation parameters of the OARs for the three segmentation methods, including DSC, JAC, HD_{max}, and DC, are presented in the form of box plots (**Figure 3**), and the statistical analysis results of the four parameters are summarized in **Table 1**.

For DSC results, although the results of $V_{DL\text{-}pre\text{-}built}$ were better than the other two in the contouring of brainstem and eyes ($p < 0.05$), all of the DSC values of the three were at a relatively high level (the DSC of the brainstem was more than 0.85, and the DSC values of eyes were floating around 0.9). For inner ears and TMJs, the pre-built DL-based model results were worse than the other two, with no significant differences between the two. Moreover, for the lens, the values of atlas-based auto-segmentation were the worst ($p < 0.05$), and there were no significant differences between the trained and pre-built DL-based models results. For the optic nerves, the trained DL-based model showed the best performance. For the contouring of the oral cavity, the DSC value of $V_{DL\text{-}pre\text{-}built}$ was the worst, and that of the $V_{DL\text{-}trained}$ was the best ($p < 0.05$). Then for the parotid glands, the DSC value of V_{atlas} was the minimum (about 0.80 or so), and those of $V_{DL\text{-}pre\text{-}built}$ and $V_{DL\text{-}trained}$ were slightly larger (about 0.82 or so). The JAC results were almost the same as the volume-related parameter of DSC.

For HD_{max}, the atlas-based auto-segmentation showed relatively large values in delineating the brainstem, eyes, lens, and optic nerves. The pre-built DL-based model method showed a lousy performance of delineation in the inner ears, TMJs, optic nerves, oral cavity, and parotid glands. For the brainstem, eyes, and lens, the HD_{max} of the atlas-based method was much larger than that of DL-based methods. In addition to the oral cavity, parotid glands, and the brainstem, the HD_{max} values of the other organs were all below 0.5 cm using the trained DL segmentation model.

V_{atlas} showed more significant results for the DC value in the brainstem, eyes, lens, and optic nerves, and the $V_{DL\text{-}pre\text{-}built}$ showed larger values in the inner ears, TMJs, optic nerves, and

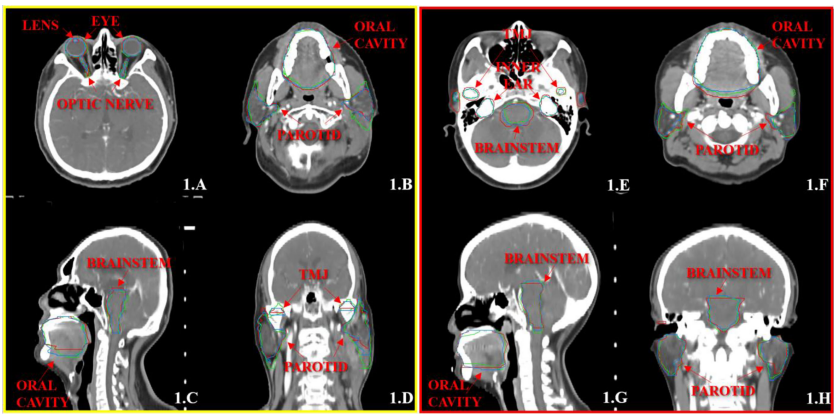


FIGURE 1 | The segmentation results on the atlas and trained DL-based models for two representative cases with the transverse **(A, B, E, F)**, sagittal **(C, G)**, and coronal **(D, H)** images, respectively. The ground-truth delineations are depicted in red, the automatic delineations based on the atlas model are depicted in green, and the automatic delineations based on the trained DL model are depicted in blue. Case 1 is shown in **(A–D)**, and case 2 is shown in **(E–H)**.

oral cavity. The trained DL-based segmentation method performed well in the contouring of all the organs that the maximum average DC is no more than 0.3 cm.

DISCUSSION

Although it is time-consuming and intra-observer and inter-observer differences usually occur, scholars worldwide have been trying to find a more rapid and more accurate method or evaluate the already existing processes. Scholars have recently published many DL studies on auto-segmentation (19–23), involving various algorithms and machine learning techniques, especially DL methods. Yang et al. (24) evaluated a U-net-based whole convolutional neural network (CNN). They got the conclusion that DL-based auto-segmentation showed great potential to alleviate the labor-intensive contouring of OARs for RT treatment planning. These methods based on artificial neural networks, especially after retraining, have shown excellent functionality better than most classification and regression methods. This conclusion is consistent with our study.

The clinical applicability of atlas-based auto-segmentation has been reported many times in the head and neck, chest, abdomen, and pelvic diseases (25–28). The results show that the atlas-based auto-segmentation outcomes could meet the clinical application and significantly reduce manual labor. Dijk et al. (29) evaluated two image segmentation methods, atlas-based segmentation and convolutional neural network-based DL model segmentation. They collected the contours of 589 head and neck cancer patients from clinical practice and used them to train models. DL-based segmentation showed encouraging results compared to the ABAS. The same is true for the findings of our research.

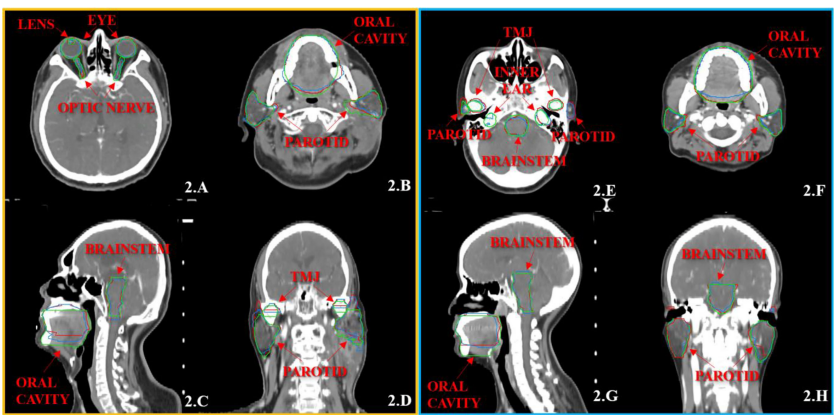


FIGURE 2 | The segmentation results on the pre-built and trained DL-based models for the two same cases with the transverse **(A, B, E, F)**, sagittal **(C, G)**, and coronal **(D, H)** images, respectively. The ground-truth delineations are depicted in red, the automatic delineations based on the pre-built DL-based model are depicted in green, and the automatic delineations based on the trained DL-based model are depicted in blue. Case 1 is shown in **(A–D)**, and case 2 is shown in **(E–H)**.

The time spent on contouring roughly depends on two main factors, the visualization of organ boundaries and the volume of OARs. For software and human observers, high-contrast edges are easier to detect; otherwise, low-contrast borders are more challenging to notice. The automatic segmentation method is generally inaccurate for the boundaries of minor soft tissue, which increases the time spent on adjustment. Most importantly, it is also difficult for observers to distinguish the boundaries. These increase the time required for adjustment additionally. Even if we assume that automatic segmentation techniques will achieve human-level performance for contouring in the future, human observers may still need to evaluate the contouring results for some difficult situations.

The accurate and reliable segmentation of NPC images is essential in clinical applications (including RT). However, the targets of NPC vary in size and shape, as well as variable intensity within the tumor and similar intensity to nearby tissues, which makes the segmentation task more difficult. The emergence of automatic segmentation software has provided convenience for RT undoubtedly, especially for adaptive radiotherapy, increases the efficiency of delineation, and reduces the variety in contouring to a certain extent. However, we could not ignore the influence of software differences on the delineation outcomes.

In our study, the atlas-based, pre-built, and trained DL-based automatic segmentation methods had good performances for the segmentation of the brainstem and eyes. The three methods got good results of DSC (e.g., the DSC results of the brainstem were above 0.8, the results of the deviation of the centroid were below 0.2 cm, the DSC results of the eyes were above 0.9, and the results of the deviation of centroid were below 0.1 cm). The reasons were that the position of the eyes was relatively fixed and the boundaries relatively straightforward, and all three methods could make better identification. The boundary of the brainstem was not clear, but the position was fixed, so we could still get a good result.

For the inner ears, TMJs, optic nerves, and oral cavity, the trained DL-based model showed a vast improvement toward the pre-built model (**Figure 2**) because these contours were more subjectively affected by physicians. Each institution or even observer might have a different contouring habit for organs. It is not easy to achieve an overall contouring agreement so that the specificity of these organs is relatively high. So when one wants to use a DL-based method for the auto-segmentation of the OARs, they need to use their data to train the model, or even the same physician's patients, to achieve satisfactory results. Atlas-based segmentation performed a relatively poor performance of the lens and optic nerves. One possible reason might be that the volume of these organs was much smaller (24). The atlas-based method was at a disadvantage of segmenting small volume organs than the DL method.

Another problematic point is the delineation of the parotid glands because of the unclear boundary and the variable contouring habits of each physician. Our study showed that the contouring results of parotid glands with the atlas-based method include most of the central areas of parotid glands, and there was still a lack in the contouring accuracy of the boundary. Meanwhile, the trained DL-based model slightly increased in the volume of parotid glands, and the boundary was much closer to the ground truth (**Figure 1**).

The contouring time of these methods in our research was not listed in the quantitative evaluation because the contouring time was so fast, about 3 min for each patient through estimation that was far less than the manual delineation time. Still, the time to build an atlas database was cumbersome. Establishing an atlas database required manual handling of each patient, which spent about 3 h for 120 patients. At the same time, the training time of the DL model was just about 49 min, with no need for a large amount of human intervention. We only chose 120 patients for the atlas database and training model in this study. The DL model for training will have a significant advantage if more data models are added in the future, avoiding the cumbersome establishment cost and the choice of an individual situation.

At the beginning of our study, we compared the delineation of the pre-built DL- and atlas-based method and added the on-site trained model later. It is, this step that made the contouring result in a significant improvement of both volume and distance (for example, the DSC values of the inner ears were increased by 0.05 and HD_{max} decreased by 1 mm, the DSC values of TMJs were increased by about 0.15 and HD_{max} decreased by 1.9 mm, the DSC values of the optic nerves were increased by about 0.16 and HD_{max} decreased by 2 mm, the DSC values of the oral cavity were increased by 0.13 and HD_{max} decreased by 16 mm, and the centroid distance decreased by 7 mm and the HD_{max} of the parotid glands were decreased by 4 mm).

In this study, the trained DL-based model performed brilliantly, showing a good result in the contouring of each organ, and all the mean DSC values were more than 0.7 (**Figure 3A**) (30), which met the clinical standards. In particular, the widely accepted optic nerves, because of the small size and unclear boundaries, have a relatively low accuracy of DSC values in many studies (14, 31–33). Similarly, the optic nerves had worse accuracy than other organs in our results. Still, the mean DSC values were above 0.7. We analyzed the excellent performance as a result of model selection. Although some studies demonstrate that a better performance needs diversity and numerous training datasets, all the cases used in our study are contoured only by one experienced oncologist. Compared to the open-access resource platform of medical images for cancer research, our datasets were more specific, eliminating the intra- and inter-observer variability to the utmost extent with very high quality and representativeness, which made the trained model more characteristic to meet the clinical acceptance. Lin et al. (34) summarized the major deep learning architectures related to target volume segmentation, surveyed the use of three common imaging modalities (CT, MRI, PET) in radiation therapy, and compared their performance. They pointed out that high-quality annotated data were a big challenge for deep learning models, and deep learning-based automatic segmentation had great potential.

It is worth mentioning that the atlas-based method also showed a good ability for the segmentation of OARs of NPC. Still, to ensure a better result, one needs to use their contouring cases to build the atlas database, especially for some specific organs. In institutions with no use conditions of DL segmentation models, the application of the atlas-based

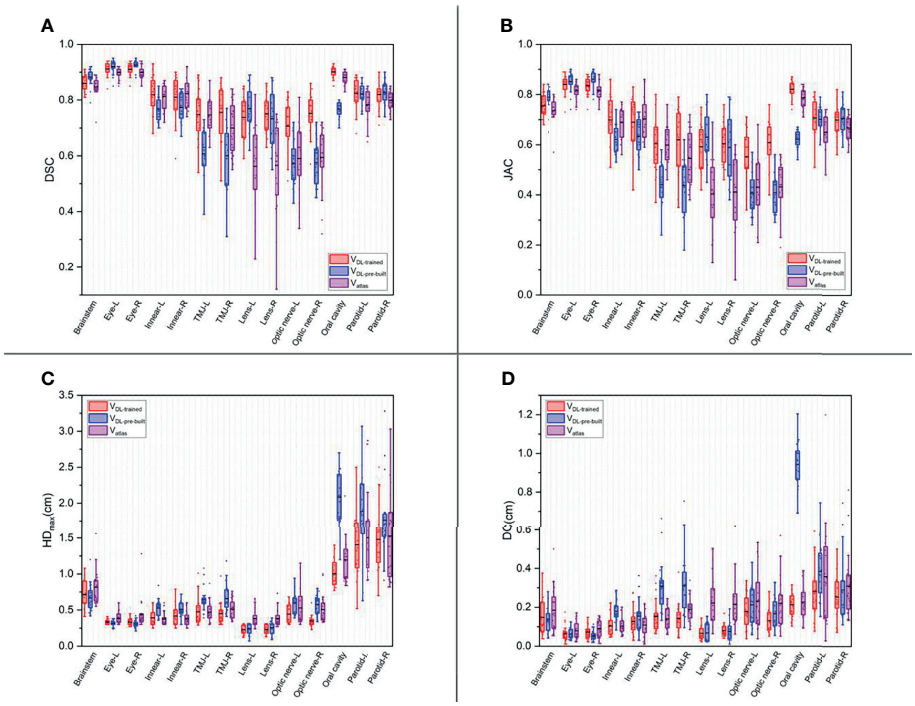


FIGURE 3 | The box plots of four quantitative evaluation parameters of the OARs for the three segmentation methods. The results of DSC are listed in **(A)**, JAC in **(B)**, HD_{max} in **(C)**, and DC in **(D)**. The results of the trained DL-based model are depicted in red, the pre-built DL-based model in blue, and the atlas-based auto-segmentation in purple.

method could still meet the clinical demands for most organs. Our research innovation combined the atlas-based, pre-built, and on-site trained DL-based automatic segmentation methods, providing intuitive results for clinical applications.

A limitation of this study is that we only used 120 cases to train the DL model to maintain the consistency of the atlas library. Perhaps with more data added to the DL model, the delineation results will be further improved, and we also have reasons to believe this. The diversity of training data and the automatic segmentation of other organs or targets for the whole body are also the focus of our following study. With the rapid development of multi-imaging modalities (such as CT, MR, PET) and radiotherapy technology (such as MRI-Linac), deep learning-based automatic segmentation methods have more vast fields.

CONCLUSIONS

Although some delineation outcomes still need further modification, the trained DL-based auto-segmentation method performs better than the atlas-based segmentation method that

TABLE 1 | The statistical analysis results of the four quantitative evaluation parameters for three segmentation methods.

	Brainstem	Eye-L	Eye-R	Inner ear-L	Inner ear-R	TMJ-L	TMJ-R
DSC	0.002^{ac}	0.015^{c}	0^{ac}	0.004^{ac}	0.052^{c}	0^{ac}	0^{ac}
JAC	0.002^{ac}	0.009^{bc}	0^{ac}	0.003^{ac}	0.041^{c}	0^{ac}	0^{ac}
HD_{max}	0.071^{c}	0^{bc}	0.004^{bc}	0^{ac}	0.006^{ac}	0.003^{ac}	0.001^{ac}
DC	0.252	0.360	0.007^{c}	0^{ac}	0.123^{c}	0^{ac}	0^{ac}
	Lens-L	**Lens-R**	**Optic nerve-L**	**Optic nerve-R**	**Oral Cavity**	**Parotid-L**	**Parotid-R**
DSC	0^{bc}	0^{bc}	0^{ab}	0^{ab}	0^{abc}	0.018^{bc}	0.100^{c}
JAC	0^{bc}	0^{bc}	0^{ab}	0^{ab}	0^{abc}	0.019^{bc}	0.099^{c}
HD_{max}	0^{bc}	0^{bc}	0.047^{a}	0^{ab}	0^{ac}	0.025^{ac}	0.210
DC	0^{bc}	0^{bc}	0.348	0.019^{b}	0^{ac}	0.220	0.522

Each symbol represents significant differences between different groups ($p < 0.05$).

[a]*Indicates a significant difference between $V_{DL\text{-}trained}$ and $V_{DL\text{-}pre\text{-}built}$.*

[b]*Indicates a significant difference between $V_{DL\text{-}trained}$ and V_{atlas}.*

[c]*Indicates a significant difference between $V_{DL\text{-}pre\text{-}built}$ and V_{atlas}.*

benefits clinical efficiency, optimizes the treatment procedure, and provides a certain level of help for clinical work. For clinical applications, the DL-based automatic segmentation of OARs can significantly save time for physicians. More factors that influence the accuracy of automatic segmentation in clinical applications still need further exploration.

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

JW: experiment design and article writing. ZC, CY, and QZhou: technical support. BQ and LM: data quality control. WF and QZheng: data collection. SX: conception, experiment design, and article modification. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Natural Science Foundation of Chinese Science Foundation Project (81801799)

REFERENCES

1. Chua MLK, Wee JTS, Hui EP, Chan ATC. Nasopharyngeal Carcinoma. *Lancet* (2016) 387(10022):1012–24. doi: 10.1016/S0140-6736(15)00055-0
2. Chen YP, Chan ATC, Le QT, Blanchard P, Sun Y, Ma J. Nasopharyngeal Carcinoma. *Lancet* (2019) 394(10192):64–80. doi: 10.1016/S0140-6736(19)30956-0
3. Lee AW, Ng WT, Pan JJ, Chiang CL, Poh SS, Choi HC, et al. International Guideline on Dose Prioritization and Acceptance Criteria in Radiation Therapy Planning for Nasopharyngeal Carcinoma. *Int J Radiat Oncol Biol Phys* (2019) 105(3):567–80. doi: 10.1016/j.ijrobp.2019.06.2540
4. Chen YP, Ismaila N, Chua MLK, Colevas AD, Haddad R, Huang SH, et al. Chemotherapy in Combination With Radiotherapy for Definitive-Intent Treatment of Stage II-IVA Nasopharyngeal Carcinoma: CSCO and ASCO Guideline. *J Clin Oncol* (2021) 39(7):840–59. doi: 10.1200/JCO.20.03237
5. Langendijk JA, Doornaert P, Verdonck-de Leeuw IM, Leemans CR, Aaronson NK, Slotman BJ. Impact of Late Treatment-Related Toxicity on Quality of Life Among Patients With Head and Neck Cancer Treated With Radiotherapy. *J Clin Oncol* (2008) 26(22):3770–6. doi: 10.1200/JCO.2007.14.6647
6. Ibragimov B, Xing L. Segmentation of Organs-at-Risks in Head and Neck CT Images Using Convolutional Neural Networks. *Med Phys* (2017) 44(2):547–57. doi: 10.1002/mp.12045
7. Harari PM, Song S, Tomé WA. Emphasizing Conformal Avoidance Versus Target Definition for IMRT Planning in Head-and-Neck Cancer. *Int J Radiat Oncol Biol Phys* (2010) 77(3):950–8. doi: 10.1016/S0360-3016(03)01163-5
8. Voet PW, Dirkx ML, Teguh DN, Hoogeman MS, Levendag PC, Heijmen BJ. Does Atlas-Based Autosegmentation of Neck Levels Require Subsequent Manual Contour Editing to Avoid Risk of Severe Target Underdosage? A Dosimetric Analysis. *Radiother Oncol* (2011) 98(3):373–7. doi: 10.1016/j.radonc.2010.11.017
9. La Macchia M, Fellin F, Amichetti M, Cianchetti M, Gianolini S, Paola V, et al. Systematic Evaluation of Three Different Commercial Software Solutions for Automatic Segmentation for Adaptive Therapy in Head-and-Neck, Prostate and Pleural Cancer. *Radiat Oncol* (2012) 7(1):160. doi: 10.1186/1748-717X-7-160
10. Thomson D, Boylan C, Liptrot T, Aitkenhead A, Lee L, Yap B, et al. Evaluation of an Automatic Segmentation Algorithm for Definition of Head and Neck Organs at Risk. *Radiat Oncol* (2014) 9(1):173. doi: 10.1186/1748-717X-9-173
11. Stapleford LJ, Lawson JD, Perkins C, Edelman S, Davis L, McDonald MW, et al. Evaluation of Automatic Atlas-Based Lymph Node Segmentation for Head-and-Neck Cancer. *Int J Radiat Oncol Biol Phys* (2010) 77(3):959–66. doi: 10.1016/j.ijrobp.2009.09.023
12. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in Auto-Segmentation. *Semin Radiat Oncol* (2019) 29(3):185–97. doi: 10.1016/j.semradonc.2019.02.001
13. Sharp G, Fritscher KD, Pekar V, Peroni M, Shusharina N, Veeraraghavan H, et al. Vision 20/20: Perspectives on Automated Image Segmentation for Radiotherapy. *Med Phys* (2014) 41(5):050902. doi: 10.1118/1.4871620
14. Vrtovec T, Močnik D, Strojan P, Pernuš F, Ibragimov B. Auto-Segmentation of Organs at Risk for Head and Neck Radiotherapy Planning: From Atlas-Based to Deep Learning Methods. *Med Phys* (2020) 47(9):e929–50. doi: 10.1002/mp.14320
15. Dice LR. Measures of the Amount of Ecologic Association Between Species. *Ecology* (1945) 26(3):297–302. doi: 10.2307/1932409
16. Jaccard P. The Distribution of the Flora in the Alpine Zone. *New Phytoligist* (1912) 11(2):37–50. doi: 10.1111/j.1469-8137.1912.tb05611.x
17. Huttenlocher DP, Klanderman GA, Rucklidge WA. Comparing Images Using the Hausdorff Distance. *IEEE Trans Pattern Anal Mach Intel* (1993) 15(9):850–63. doi: 10.1109/34.232073
18. Wang J, Xu S, Liu B, Zheng Q, Zhang H, Yang W, et al. Quantitative Evaluation of Atlas-Based Auto-Segmentation of Organs-at-Risk in Patients With Cervical Cancer Using Different Atlas Database Sizes. *Chin J Med Phys* (2019) 36(7):760–4. doi: 10.3969/j.issn.1005-202X.2019.07.004
19. Liu A, Li R, Han C, Du D, Wong JYC. Comparative Clinical Evaluation of Deep-Learning-Based Algorithms in Auto-Segmentation of Organs-at-Risk for Head and Neck Cancers. *Int J Radiat Oncol Biol Phys* (2020) 108(3):e817. doi: 10.1016/j.ijrobp.2020.07.324
20. Dai X, Lei Y, Wang T, Zhou J, Roper J, McDonald M, et al. Automated Delineation of Head and Neck Organs at Risk Using Synthetic MRI- Aided Mask Scoring Regional Convolutional Neural Network. *Med Phys* (2021) 48:5862–73. doi: 10.1002/mp.15146
21. Zhang Z, Zhao T, Gay H, Zhang W, Sun B. Weaving Attention U-Net: A Novel Hybrid CNN and Attention-Based Method for Organs-at-Risk Segmentation in Head and Neck CT Images. *Med Phys* (2021) 48:7052–62. doi: 10.1002/mp.15287
22. Choi MS, Choi BS, Chung SY, Kim N, Chun J, Kim YB, 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) 153:139–45. doi: 10.1016/j.radonc.2020.09.045
23. Wong J, Fong A, Mcvicar N, Smith S, Giambattista J, Wells D, et al. Comparing Deep Learning-Based Auto-Segmentation of Organs at Risk and Clinical Target Volumes to Expert Inter-Observer Variability in Radiotherapy Planning. *Radiother Oncol* (2020) 144:152–8. doi: 10.1016/S0167-8140(19)33176-7
24. Zhong Y, Yang Y, Fang Y, Wang J, Hu W. A Preliminary Experience of Implementing Deep-Learning Based Auto-Segmentation in Head and Neck Cancer: A Study on Real-World Clinical Cases. *Front Oncol* (2021) 11:638197. doi: 10.3389/fonc.2021.638197
25. Hoang Duc AK, Eminowicz G, Mendes R, Wong SL, McClelland J, Modat M, et al. Validation of Clinical Acceptability of an Atlas-Based Segmentation Algorithm for the Delineation of Organs at Risk in Head and Neck Cancer. *Med Phys* (2015) 42(9):5027–34. doi: 10.1118/1.4927567
26. Oliveira B, Queirós S, Morais P, Torres HR, Gomes-Fonseca J, Fonseca JC, et al. A Novel Multi-Atlas Strategy With Dense Deformation Field Reconstruction for Abdominal and Thoracic Multi-Organ Segmentation From Computed Tomography. *Med Image Anal* (2018) 45:108–20. doi: 10.1016/j.media.2018.02.001

27. 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(1):68–73. doi: 10.1016/j.radonc.2011.08.043
28. Kim N, Chang JS, Kim YB, Kim JS. Atlas-Based Auto-Segmentation for Postoperative Radiotherapy Planning in Endometrial and Cervical Cancers. *Radiat Oncol* (2020) 15(1):106. doi: 10.1186/s13014-020-01562-y
29. Van Dijk LV, Van den Bosch L, Aljabar P, Peressutti D, Both S, Steenbakkers RJ, et al. Improving Automatic Delineation for Head and Neck Organs at Risk by Deep Learning Contouring. *Radiother Oncol* (2020) 142:115–23. doi: 10.1016/j.radonc.2019.09.022
30. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC. Morphometric Analysis of White Matter Lesions in MR Images: Method and Validation. *IEEE Trans Med Imaging* (1994) 13(4):716–24. doi: 10.1109/42.363096
31. Isambert A, Dhermain F, Bidault F, Commowick O, Bondiau PY, Malandain G, et al. Evaluation of an Atlas-Based Automatic Segmentation Software for the Delineation of Brain Organs at Risk in a Radiation Therapy Clinical Context. *Radiother Oncol* (2008) 87(1):93–9. doi: 10.1016/j.radonc.2007.11.030
32. Deeley M, Chen A, Datteri R, Noble JH, Cmelak AJ, Donnelly EF, et al. Comparison of Manual and Automatic Segmentation Methods for Brain Structures in the Presence of Space-Occupying Lesions: A Multi-Expert Study. *Phys Med Biol* (2011) 56(14):4557–77. doi: 10.1088/0031-9155/56/14/021
33. Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of Image Registration and Fusion Algorithms and Techniques in Radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. *Med Phys* (2017) 44(7):e43–67. doi: 10.1002/mp.12256
34. Lin H, Xiao H, Dong L, Teo KB, Zou W, Cai J, et al. Deep Learning for Automatic Target Volume Segmentation in Radiation Therapy: A Review. *Quant Imaging Med Surg* (2021) 11(12):4847–58. doi: 10.21037/qims-21-168

Conflict of Interest: Authors ZC, CY, and QZheng were employed by Manteia 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 | Frontiers in Oncology

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fonc.2022.856346

Transfer Learning-Based Autosegmentation of Primary Tumor Volumes of Glioblastomas Using Preoperative MRI for Radiotherapy Treatment

Suqing Tian[1†], Cuiying Wang[2†], Ruiping Zhang[3], Zhuojie Dai[4], Lecheng Jia[5], Wei Zhang[6], Junjie Wang[1] and Yinglong Liu[5*]*

[1] Department of Radiation Oncology, Peking University Third Hospital, Beijing, China, [2] Department of Oncology, Hainan Third People's Hospital, Sanya, China, [3] Department of Radiation Oncology, The First Hospital of Tsinghua University, Beijing, China, [4] United Imaging Research Institute of Intelligent Imaging, Beijing, China, [5] Shenzhen United Imaging Research Institute of Innovative Medical Equipment, Shenzhen, China, [6] Shanghai United Imaging Healthcare Co.Ltd., Shanghai, China

OPEN ACCESS

Edited by:
Wei Zhao,
Beihang University, China

Reviewed by:
Fuli Zhang,
Chinese PLA General Hospital, China
Xianjin Dai,
Emory University, United States

***Correspondence:**
Junjie Wang
junjiewang_edu@sina.cn
Yinglong Liu
yinglong.liu@cri-united-imaging.com

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

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: *17 January 2022*
Accepted: *23 February 2022*
Published: *14 April 2022*

Citation:
Tian S, Wang C, Zhang R, Dai Z, Jia L, Zhang W, Wang J and Liu Y (2022) Transfer Learning-Based Autosegmentation of Primary Tumor Volumes of Glioblastomas Using Preoperative MRI for Radiotherapy Treatment. Front. Oncol. 12:856346. doi: 10.3389/fonc.2022.856346

Objectives: Glioblastoma is the most common primary malignant brain tumor in adults and can be treated with radiation therapy. However, tumor target contouring for head radiation therapy is labor-intensive and highly dependent on the experience of the radiation oncologist. Recently, autosegmentation of the tumor target has been playing an increasingly important role in the development of radiotherapy plans. Therefore, we established a deep learning model and improved its performance in autosegmenting and contouring the primary gross tumor volume (GTV) of glioblastomas through transfer learning.

Methods: The preoperative MRI data of 20 patients with glioblastomas were collected from our department (ST) and split into a training set and testing set. We fine-tuned a deep learning model for autosegmentation of the hippocampus on separate MRI scans (RZ) through transfer learning and trained this deep learning model directly using the training set. Finally, we evaluated the performance of both trained models in autosegmenting glioblastomas using the testing set.

Results: The fine-tuned model converged within 20 epochs, compared to over 50 epochs for the model trained directly by the same training set, and demonstrated better autosegmentation performance [Dice similarity coefficient (DSC) 0.9404 ± 0.0117, 95% Hausdorff distance (95HD) 1.8107 mm ±0.3964*mm*, average surface distance (ASD) 0.6003 mm ±0.1287mm] than the model trained directly (DSC 0.9158±0.0178, 95HD 2.5761 mm ± 0.5365mm, ASD 0.7579 mm ± 0.1468mm) with the same test set. The DSC, 95HD, and ASD values of the two models were significantly different ($P<0.05$).

Conclusion: A model developed with semisupervised transfer learning and trained on independent data achieved good performance in autosegmenting glioblastoma. The autosegmented volume of glioblastomas is sufficiently accurate for radiotherapy treatment, which could have a positive impact on tumor control and patient survival.

Keywords: glioblastoma, autosegmentation, deep learning, transfer learning, radiotherapy treatment

INTRODUCTION

Glioblastoma is the most common primary malignant brain tumor in adults (1). At present, the standard treatment for this disease is combination therapy, including postoperative radiotherapy and adjuvant chemotherapy after the initial surgery. Intensity-modulated radiotherapy (IMRT) is a commonly used method for delivering radiotherapy to glioblastomas. An accurate radiotherapy plan is required to ensure accurate patient treatment (2). The delineation of brain tumor targets and other brain tissue structure areas from multimodal MRI sequences can provide important information for the radiotherapy plan. Traditionally, the manual contouring of these areas is time-consuming and dependent on the experience of the doctors.

The implementation of deep learning has resulted in the development of new ideas for the automatic and accurate delineation of brain tumors (3). Deep learning approaches through convolutional neural networks (CNNs) have been proposed for glioblastoma segmentation (4–6). For example, Yi et al. (4) developed a framework of three-dimensional (3D) fully CNN models for glioblastoma segmentation from multimodality MRI data and achieved a Dice score of 0.89 in whole tumor glioblastoma segmentationon the segmentation dataset of the Brain Tumor Image Segmentation Challenge (BRATS) with 274 tumor samples.

Recently, transfer learning has found multiple applications in brain MRI (7). Transfer learning allows the reuse of a pretrained model to solve a related target problem, potentially yielding better results from fine-tuning pretrained CNNs than training CNNs from scratch (8). In this work, we provide a deep learning model for the autosegmentation of the gross tumor volume (GTV) of glioma. A deep learning model trained for hippocampus autosegmentation was fine-tuned and trained using a limited MR dataset of 20 glioblastoma patients. This approach is expected to serve as a basis for accurate radiotherapy dose calculation and optimization in the development of a high-quality radiotherapy plan (9).

MATERIALS AND METHODS

Patient Characteristics

We retrospectively collected the MRI scans and medical records of patients with histologically proven glioblastomas from a single institute (Department of Radiation Oncology of Peking University Third Hospital). The MRI examinations were performed with preoperative contrast-enhanced T1-weighted sequences. Details of the MRI characteristics are shown in **Table 1**. The MRI dataset consisted of GTVs of high-grade gliomas (HGGs) of 20 patients, which was then randomly split into three cohorts: 16 patients as the training set (including 4 patients as the validation set) for training an autosegmentation model and optimization of hyperparameters during model training and 4 patients as the test set for evaluating the performance of the trained model.

Gross Tumor Volume Contours by Human Experts

The MRI examinations of the 20 patients were assigned to two expert radiation oncologists (ST and ZD, both with more than 15 years of experience in radiotherapy treatment of head and neck tumors) to delineate the ground-truth GTVs *via* consensus. A third radiologist (CW, with more than 20 years of experience) specializing in radiation oncology was consulted in cases of disagreement. A diagram of the GTV contours delineated by the human experts is presented in **Figure 1**.

Image Processing

Preprocessing

All MRI sequences were cropped to only include regions of non-zero value to reduce the size of the network input and thereby reduce the computational load of the network (10). To enable our network to properly learn spatial semantics, all MRI sequences were resampled to the median voxel spacing of the dataset, where third-order spline interpolation was used for the images of all MRI scans and nearest-neighbor interpolation for their corresponding contours. All images were additionally normalized by simple Z score normalization for the individual patients (11).

Augmentation

To overcome the overfitting problem caused by training a deep network with limited data, we adopted a number of real-time data enhancement techniques, such as random flip, random zoom, random elastic deformation, gamma adjustment, and mirroring, to increase the diversity of the data.

Architecture of the Deep Convolution Neural Network

U-Net is a popular encoder-decoder network (11, 12) that has been widely used in semantic segmentation fields. Its encoder part works similarly to a traditional classification CNN in that it successively aggregates semantic information at the expense of spatial information. Its decoder receives semantic information from the bottom layer of the encoder and recombines it with higher-resolution feature maps obtained directly from the

TABLE 1 | Characteristics of MRI.

Cancer	Glioblastoma
Tumor	Gross tumor volume
Grade	High-grade gliomas
Modality	contrast-enhanced T1-weighted imaging
Quantity	20 patients
Resolution	(144~176) × 256 × 256
Spacing [mm^3]	[1,1,1]

encoder through skip connections (13) to recover the spatial information missing in the encoder.

Since 3D CNNs have demonstrated high effectiveness in aggregating valuable information in the context of 3D medical images (14), we implemented a 3D deep CNN to extract representative features for complicated GTVs based on the MRI sequences. Our network is based on the architecture of 3D U-Net (15), with 5 encoders and 5 decoders. In each encoder and decoder, we designed a couple of convolutional layers with a 3×3×3 convolution kernel to extract the feature of the image, each convolutional layer followed by the LeakyReLU (negative slope $1e^{-2}$) and the instance normalization (16) with a dropout of 0.5, which respectively replaced the more common ReLU activation function and batch normalization in the popular deep learning model. We used the 2×2×2 max pooling to create a downsampled feature map in each encoder; conversely, we used the 2×2×2 deconvolution kernel to create an upsampled feature map in each decoder. The layers in the encoders were skip connected and concatenated with layers in the corresponding decoders to use fine-grained details learned in the encoders to construct the feature maps in the decoders. The detailed architecture of our network is shown in **Figure 2**.

Objective Loss

Due to the limited available Graphics Processing Unit (GPU) memory, we slid and cropped smaller image patches from the original images as the input of the segmentation network. Although this patch-based training method limits the field of view of the model and is unable to collect sufficient contextual information, the impact on small target segmentation is minimal.

The objective loss L of the segmentation network is the weighted sum of Dice loss L_{dice} and cross-entropy loss L_{CE}:

$$L = \alpha_1 L_{dice} + \beta_1 L_{CE}$$

Here, the weights α_1 and β_1 were set to 0.4 and 0.6, respectively.

$$L_{dice} = -\frac{2}{C}\sum_{k\in C}\frac{\sum_{i\in N} G_i^k P_i^k}{\sum_{i\in N} G_i^k + \sum_{i\in N} P_i^k}$$

$$L_{CE} = 1 - \frac{1}{N}\sum_{i\in N}\sum_{k\in C} G_i^k \log P_i^k$$

where C is the number of divided categories, N is the number of voxels in each patch sample in the training set, and G_i^k and P_i^k are the contour corresponding to the i^{th} voxel of the k^{th} category and the probability output of the model prediction, respectively.

Experiments and Evaluation

Model Implementation

We used PyTorch 1.6 to build 3D U-Net on Ubuntu 18.04 and trained the model framework on an NVIDIA Tesla V100. When training the model, the model input patch size was 32 × 256 × 256, the batch size was 2, the optimizer was RMSprop, the initial learning rate was 0.001, the gradient descent strategy was stochastic gradient descent (SGD) with momentum (0.9), and the maximum number of training rounds (epochs) was 150. In addition, due to the limited amount of collected data, we did not divide the test set separately but adopted an 8:2 dataset division method. For the segmentation results of each patch of the model, we used Gaussian fusion to obtain the full-resolution segmentation result for each class, which was postprocessed with the largest connected component as the final segmentation result.

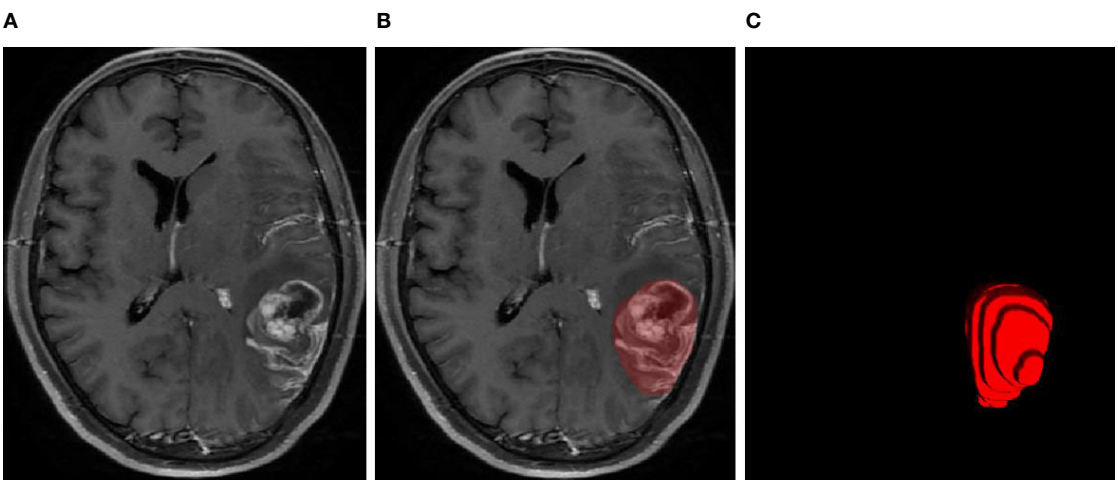


FIGURE 1 | MRI examination of the glioblastoma **(A)**, GTV contours delineated by human experts **(B)**, and 3D diagram corresponding to the GTV contours **(C)**.

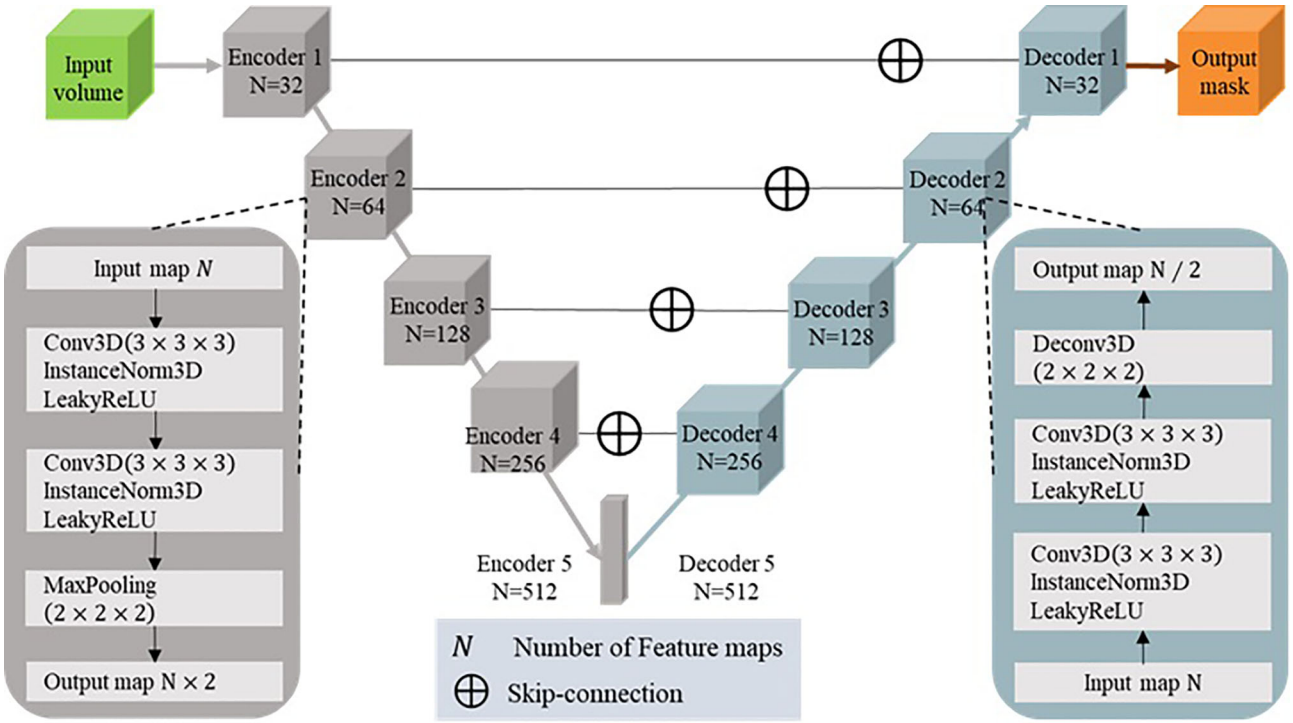


FIGURE 2 | Architecture of the segmentation network.

Model Evaluation

We calculated the Dice similarity coefficient (DSC), 95% Hausdorff distance (95HD), and average surface distance (ASD) between the GTVs segmented automatically by the model and the corresponding manual annotations as quantitative assessments of the accuracy of the model segmentation. The DSC is defined as:

$$DSC = \frac{2|P \cap G|}{|P| + |G|}$$

where P is the automatically segmented contour, G is the ground-truth contour. DSC is an indicative degree of similarity for agreement, which measures the spatial overlap between the automatic segmentation and the ground-truth segmentation. The 95HD is defined as:

$$d_H(P, G) = \max\{d_{PG}, d_{GP}\}$$

$$= \max\left\{\max_{x \in X} \min_{y \in Y} d(p, g), \max_{y \in Y} \min_{x \in X} d(p, g)\right\}$$

DSC is more sensitive to the inner filling of the segmented contour, while Hausdorff distance (HD) is more sensitive to the boundary of the segmented contour. The 95%HD is similar to maximum HD. However, it is based on the calculation of the 95th percentile of the distances between boundary points in P and G. The purpose of using this metric is to eliminate the impact of a very small subset of the outliers. The ASD is defined as:

$$\text{ASD} = \frac{1}{|S(P)| + |S(G)|}\left(\sum_{p \in S(P)} \min_{g \in S(g)} ||p - g|| + \sum_{g \in S(G)} \min_{p \in S(P)} ||g - p||\right)$$

where $S(P$ and $S(G)$ denote the point set of automatic segmentation pixels and ground-truth pixels, respectively. The most consistent segmentation result can be obtained when ASD equals 0.

Model Fine-Tuning

Transfer learning is a process by which existing models are reused to solve a new challenge, usually the problem of overfitting due to data scarcity (17). Given the limited size of the dataset delineated by our human experts and the fact that our modified 3D U-Net is a kind of supervised learning method that works well depending on the severity of the big data, we applied transfer learning to this work to fine-tune the glioblastoma segmentation model.

To apply transfer learning in this work, the 3D U-Net was trained to autosegment the hippocampus with the data from 50 patients with T1C glioblastomas (spacing[mm]: 0.5×0.36×0.36, resolution: (327~364) ×640×640). This hippocampal dataset was collected from a single institute (The First Hospital of Tsinghua University) for the radiotherapy treatment of brain metastases. All contours of the hippocampus were delineated by two expert radiation oncologists (ZD, with more than 15 years of experience in radiotherapy treatment of head and neck tumors; RZ, with

more than 20 years of experience in radiotherapy treatment) according to the results of the RTOG0933 trial and then cross-checked and revised.

Our 3D U-Net for the autosegmentation of the hippocampus was trained with the hippocampus data from 40 patients, converged within 150 epochs, and was denoted as model-hippo; the training process is shown in **Figure 3**. The model achieved a DSC of 0.897 (±0.011) for the left hippocampus and 0.895 (±0.019) for the right hippocampus with the test set (10 cases).

The model for autosegmenting the GTV of glioblastomas was trained as follows. First, the parameters of our 3D U-Net were randomly initialized, and the model was trained simply with 50 epochs on the training set of gliomas, as illustrated in **Figure 4A**; the model thus developed was denoted as model-glioma. Second, transfer learning was applied to fine-tune model-glioma. Specifically, instead of random parameter initialization, the network parameters of model-hippo were used as the initial parameters of model-glioma, and the resulting model was then trained for 50 epochs using the same training set for training model-glioma. This model was denoted as model-glioma-TL, where TL refers to transfer learning. The corresponding training process is shown in **Figure 4B**.

As shown in **Figure 4**, we found that model-glioma-TL converged faster than model-glioma did within the same 50 epochs. The validation metric reached 0.9 within 10 epochs for model-glioma-TL but within 40 epochs for model-glioma. Moreover, the validation metric of model-glioma-TL on the final epoch was greater than that of model-glioma.

RESULTS

Two sets of experiments (model-glioma and model-glioma-TL) were conducted to verify the effectiveness of transfer learning on training with small sample data and to evaluate the performance of the two models by the DSC, 95HD, and ASD metrics. The evaluation metrics of two sets of experiments with the same test set, including the mean, standard deviation (SD), and P value for the T test (two-tailed), are presented in **Table 2**. We found that model-glioma-TL significantly outperformed model-glioma in these terms [DSC 0.9404 & 0.9158, 95HD 1.8107mm & 2.5761mm, ASD 0.6003mm & 0.7579mm (and $P<0.05$). The autosegmentation results of a test sample are visualized in **Figure 5** and its 3D morphology are visualized in **Figure 6**.

DISCUSSION

The rapid development of modern radiotherapy technology has resulted in more abundant relevant multimodal medical imaging information (18). Since a considerable amount of time is necessary to manually contour MRI slices and the segmentation results of artificial tumors often depend on the doctor's prior knowledge and work experience, the final target volume results can be variable (19, 20). Therefore, deep learning technology combined with MRI can help improve the accuracy of tumor target delineation and reduce differences caused by subjective factors (10). Additionally, it can help doctors efficiently and practically complete their tumor target area delineation tasks (21, 22).

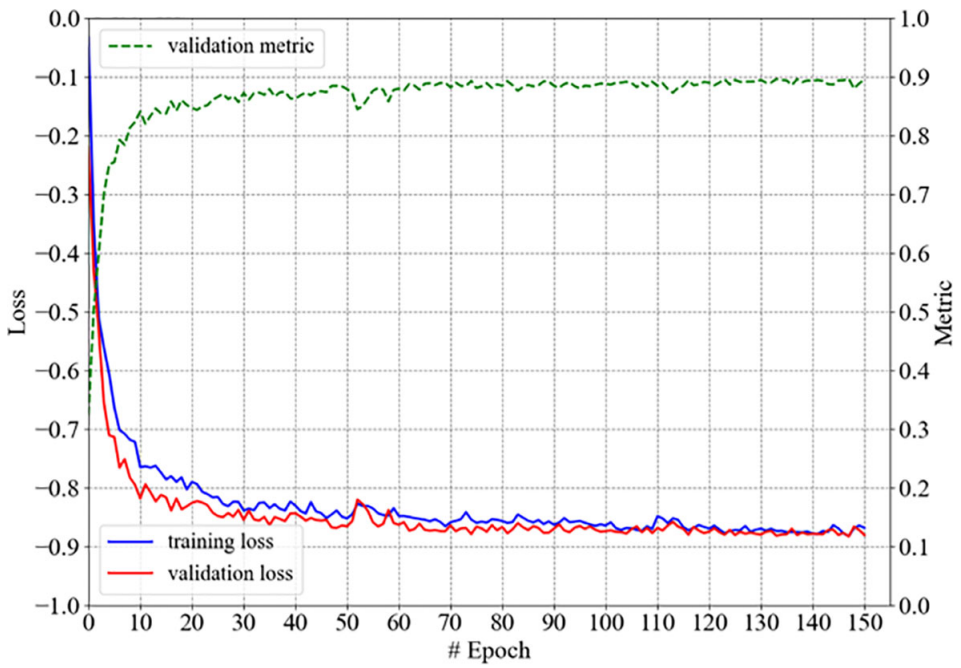


FIGURE 3 | The training process of model-hippo: the loss is the objective loss *L* and the metric is DSC.

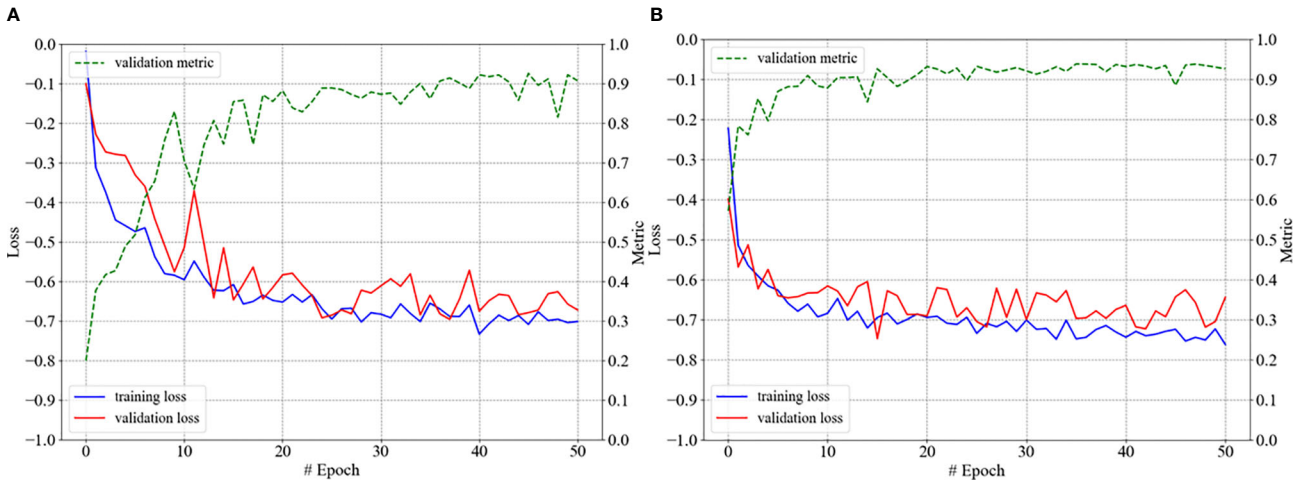


FIGURE 4 | The training processes of model-glioma **(A)** and model-glioma-TL **(B)**.

TABLE 2 | Results for autosegmentation of the GTVs of glioblastomas.

Model	Model-glioma		Model-glioma-TL		*P* value
	Mean	SD	Mean	SD	
DSC	0.9158	0.0178	0.9404	0.0117	0.0404
95HD [mm]	2.5761	0.5365	1.8107	0.3964	0.0275
ASD [mm]	0.7579	0.1468	0.6003	0.1287	0.0182

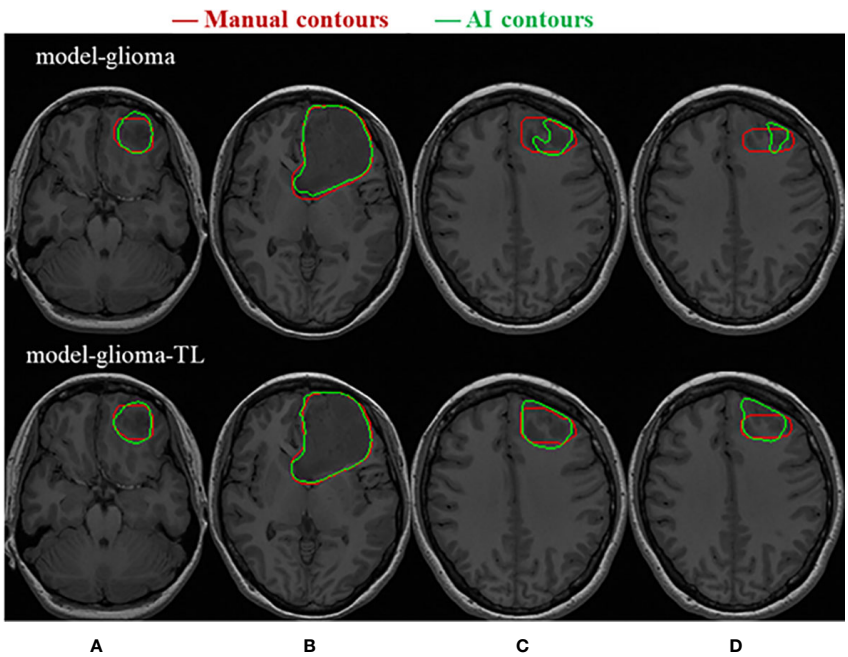


FIGURE 5 | Visualization of the test samples for the two models. The performance of model-glioma-TL was superior to that of model-glioma in the autosegmentation of glioblastoma GTVs, especially in the recognition of the small GTV in the upper and lower MRI slices **(A, D)** and the boundary delineation of the GTV contours in the intermediate MRI slices **(B, C)**.

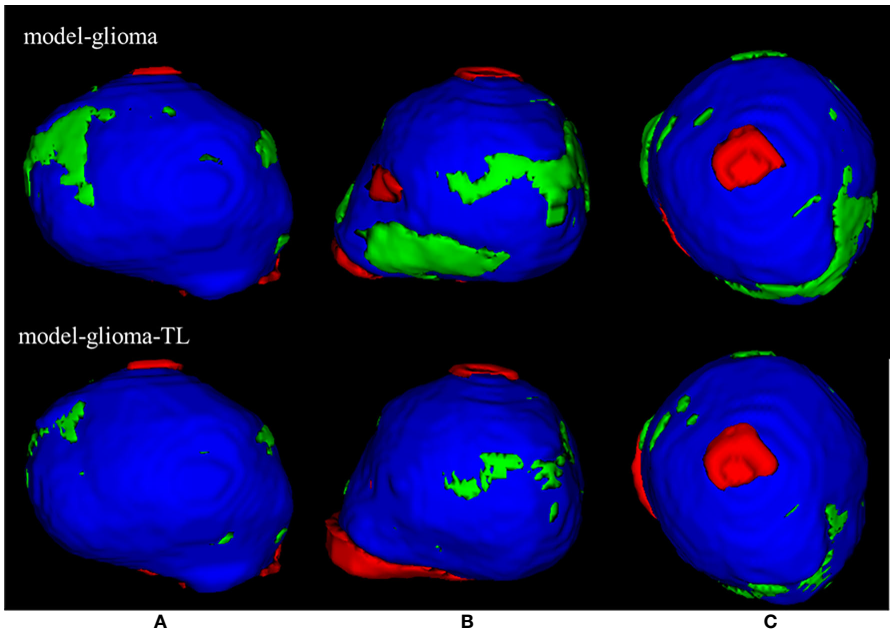


FIGURE 6 | | The 3D visualization of the test samples for the two models. The red region is the individual contouring by human experts, the green region is the individual contouring by AI models, and the blue region is the mutual contouring combined with human experts and AI models. The contouring by model-glioma-TL coincided with the contouring by human experts better than that of model-glioma in three different profiles of the test sample **(A–C)**. And the mean absolute percentage error of model-glioma-TL in the autosegmentation of glioblastoma GTVs with the same test set was 2.58% and superior to 4.74% of modelgli.

Gliomas are the most common primary brain tumors and seriously endanger human health (23). Therefore, segmentation of the images of brain tumors has become a popular research topic (24, 25). In recent years, brain tumor image segmentation methods have undergone continuous improvement, transitioning from manual to half-motion and automatic segmentation techniques (22, 26). In the present study, we have demonstrated that the performance of the transfer learning approach is comparable to the models trained through 3D CNN (4, 6), but a much smaller dataset and fewer epochs are required. Indeed, this method should be further evaluated using larger datasets such as BRATS.

In conclusion, transfer learning is feasible and effective in training models for accurate and consistent glioblastoma autosegmentation. This is more crucial for a radiation oncology department that is willing to implement deep learning with limited number of clinical cases.

In this work, the artificial intelligence (AI) algorithm based on transfer learning has achieved good results for the autosegmentation of glioblastoma GTV; however, there are still several issues that need to be cleared. 1) The location and boundary of glioblastoma GTV not only need to consider the enhanced area of contrast-enhanced T1-weighted imaging and the abnormal area of T2 FLAIR in clinical practice, perhaps the autosegmentation of the glioblastoma tumor using multimodality MRI is more satisfying for clinical practice (27). 2) Scanners from different manufacturers or different scanning protocols often result in medical imaging with different voxel spacing and resolution, as well as image quality and style. These differences are especially pronounced for MRI. Additionally, different tumor delineation styles come from the subjectivity of different doctors; these various differences seriously affect the generalizability of the AI algorithms. Therefore, further study using the data from multiple centers is an important topic. 3) We confirmed that transfer learning can significantly improve the automatic segmentation of the glioblastoma GTV in this work; however, some organs or tumor target areas, such as the optic chiasm, have a similar X-shape, while the brain stem has a similar apple shape (28, 29), and it needs to be judged by combining different medical imaging procedures and the rich medical prior knowledge of professional doctors (30, 31). How to incorporate such shape and prior knowledge and medical prior knowledge into the AI model to further improve the generalizability and generalization of AI algorithms still is an open problem.

DATA AVAILABILITY STATEMENT

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

AUTHOR CONTRIBUTIONS

Conceptualization, ST and CW. Methodology, ST and YL. Software, YL and LJ. Validation, ST, CW, RZ and ZD. Formal

analysis, ST and ZD. Investigation, ST and CW. Resources, ST and JW. Data curation, ST, JW, ZD, and YL. Writing—original draft preparation, ST and YL. Writing—review and editing, JW, CW and RZ. Visualization, YL. Supervision, WZ. Project administration, and WZ. ST and CW contributed equally to this work, so they are listed as co-first authors. All authors contributed to the article and approved the submitted version..

REFERENCES

1. Ostrom QT, Gittleman H, Truitt G, Boscia A, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2011-2015. *Neuro Oncol* (2018) 20 Suppl 4:iv1. doi: 10.1093/neuonc/noy131
2. Ten Haken RK, Thornton AFJr, Sandler HM, LaVigne ML, Quint DJ, Fraass BA, et al. A Quantitative Assessment of the Addition of MRI to CT-Based, 3-D Treatment Planning of Brain Tumors. *Radiother Oncol* (1992) 25:121. doi: 10.1016/0167-8140(92)90018-P
3. Ranjbarzadeh R, Bagherian Kasgari A, Jafarzadeh Ghoushchi S, Anari S, Naseri M, Bendechache M. Brain Tumor Segmentation Based on Deep Learning and an Attention Mechanism Using MRI Multi-Modalities Brain Images. *Sci Rep* (2021) 11:10930. doi: 10.1038/s41598-021-90428-8
4. Yi D, Zhou M, Chen Z, Gevaert O. 3-D Convolutional Neural Networks for Glioblastoma Segmentation. *Comput Vision Pattern Recognit* (2016) arXiv:1611.04534. doi: 10.48550/arXiv.1611.04534.
5. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, et al. Brain Tumor Segmentation With Deep Neural Networks. *Med Image Anal* (2017) 35:18–31. doi: 10.1016/j.media.2016.05.004
6. Chen L, Wu Y, DSouza AM, Abidin AZ, Wismüller A, Xu C. Medical Imaging 2018: Image Processing. In: *MRI Tumor Segmentation With Densely Connected 3D CNN*, vol. 10574. . Bellingham, WA, USA: International Society for Optics and Photonics (2018). p. 105741F.
7. Valverde JM, Imani V, Abdollahzadeh A, De Feo R, Prakash M, Ciszek R, et al. Transfer Learning in Magnetic Resonance Brain Imaging: A Systematic Review. *J Imaging* (2021) 7:66. doi: 10.3390/jimaging7040066
8. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, et al. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Trans Med Imaging* (2016) 35:1285–98. doi: 10.1109/TMI.2016.2528162
9. Isensee F, Jäger PF, Kohl SA, Petersen J, Maier-Hein KH. Automated Design of Deep Learning Methods for Biomedical Image Segmentation. *arXiv preprint arXiv* (2019). *Comput Vision Pattern Recognit* (2016) arXiv:1611.04534. doi: 10.48550/arXiv.1611.04534
10. Zhao Y-X, Zhang Y-M, Liu C-L. Bag of Tricks for 3d Mri Brain Tumor Segmentation. In: A. Crimi, S. Bakas (eds) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2019. Lecture Notes in Computer Science ()*, vol 11992. Cham: Springer (2020). doi: 10.1007/978-3-030-46640-4_20
11. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation. *Nat Methods* (2021) 18(2):203–11. doi: 10.1038/s41592-020-01008-z
12. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: N. Navab, J. Hornegger, W. Wells, A. Frangi (eds) *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science*, vol. 9351. Cham: Springer (2015). doi: 10.1007/978-3-319-24574-4_28.
13. Drozdzal M, Vorontsov E, Chartrand G, Kadoury S, Pal C. The Importance of Skip Connections in Biomedical Image Segmentation. *Comput Vision Pattern Recognit* (2016) arXiv:1608.04117. doi: 10.48550/arXiv.1608.04117
14. Milletari F, Navab N, Ahmadi S-A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In: *International Conference on 3D Vision (3dv)*. Stanford, CA, USA: IEEE (2016). p. 565–71. doi: 10.1109/3DV.2016.79
15. C¸i¸cek O, 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*. Cham: Springer (2016). p. 424–32. doi: 10.1007/978-3-319-46723-8_49
16. Ulyanov D, Vedaldi A, Lempitsky V. Instance Normalization: The Missing Ingredient for Fast Stylization. *Comput Vision Pattern Recognit* (2016) arXiv:1607.08022. doi: 10.48550/arXiv.1607.08
17. Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A Comprehensive Survey on Transfer Learning. *Proc IEEE* (2021) 109(1):43–76. doi: 10.1109/JPROC.2020.3004555
18. Ostrom QT, Bauchet L, Davis FG, Deltour I, Fisher JL, Langer CE, et al. The Epidemiology of Glioma in Adults: A State of the Science Review. *Neuro Oncol* (2014) 16(7):896–913. doi: 10.1093/neuonc/nou087
19. 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* (2014) 34(10):1993–2024.
20. Wang G, Li W, Ourselin S, Vercauteren T. Automatic Brain Tumor Segmentation Using Cascaded Anisotropic Convolutional Neural Networks. In: A. Crimi, S. Bakas, H. Kuijf, B. Menze, M. Reyes (eds) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. BrainLes 2017. Lecture Notes in Computer Science ()* vol 10670. Cham: Springer (2018).doi: 10.1007/978-3-319-75238-9_16.
21. Meier R, Bauer S, Slotboom J, Wiest R, Reyes M. Patient-Specific Semi-Supervised Learning for Postoperative Brain Tumor Segmentation. *Med image Comput computer-assisted Intervention: MICCAI Int Conf Med Image Comput Computer-Assisted Intervention* (2014) 17:714–21. doi: 10.1007/978-3-319-10404-1_89
22. Gutman DA, Dunn WDJr, Grossmann P, Cooper LA, Holder CA, Ligon KL, et al. Somatic Mutations Associated With MRI-Derived Volumetric Features in Glioblastoma. *Neuroradiology* (2015) 57:1227–37. doi: 10.1007/s00234-015-1576-7
23. Menze BH, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, et al. The Multimodal Brain TumorImage Segmentation Benchmark (BRATS). *IEEE Trans Med Imaging* (2015) 34(10):1993–2024. doi: 10.1109/TMI.2014.2377694
24. Egger J, Kapur T, Fedorov A, Pieper S, Miller JV, Veeraraghavan H, et al. GBM Volumetry Using the 3D Slicer Medical Image Computing Platform. *Sci Rep* (2013) 3:164. doi: 10.1038/srep01364
25. Bauer S, Nolte LP, Reyes M. Fully Automatic Segmentation of Brain Tumor Images Using Support Vector Machine Classification in Combination With Hierarchical Conditional Random Field Regularization. *Med Image Comput Computer-Assisted Intervention: MICCAI Int Conf Med Image Comput Computer-Assisted Intervention* (2011) 14:354–61. doi: 10.1007/978-3-642-23626-6_44
26. Stummer W, Reulen HJ, Meinel T, Pichlmeier U, Schumacher W, Tonn JC, et al. Extent of Resection and Survival in Glioblastoma Multiforme: Identification of and Adjustment for Bias. *Neurosurgery* (2008) 62:564–576 discussion 564-576.
27. Meyer P, Noblet V, Mazzara C, Lallement A. Survey on Deep Learning for Radiotherapy. *Comput Biol Med* (2018) 98:126–46. doi: 10.1016/j.compbiomed.2018.05.018
28. Gondi V, Tomé WA, Mehta MP. Why Avoid the Hippocampus? A Comprehensive Review. *Radiother Oncol* (2010) 97:370–6. doi: 10.1016/j.radonc.2010.09.013
29. Brown PD, Gondi V, Pugh S. Hippocampal Avoidance During Whole-Brain Radiotherapy Plus Memantine for Patients With Brain Metastases: Phase III Trial NRG Oncology Cc001. *J Clin Oncol* (2020) 38:1019–29. doi: 10.1200/JCO.19.02767
30. Sze V, Chen Y, Yang T, Emer JS. Efficient Processing of Deep Neural Networks: A Tutorial and Survey. *Proc IEEE* (2017) 105(12):2295–329. doi: 10.1109/JPROC.2017.2761740
31. 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: A. Crimi, S. Bakas, H. Kuijf, B. Menze, M. Reyes (eds) *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain*

Injuries. BrainLes 2017. Lecture Notes in Computer Science () vol 10670. Cham: Springer (2018). doi: 10.1007/978-3-319-75238-9_25.

Conflict of Interest: WZ is employed by Shanghai United Imaging Healthcare 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 | Frontiers in Oncology

ORIGINAL RESEARCH
published: 30 June 2022
doi: 10.3389/fonc.2022.913898

Multiphasic CT-Based Radiomics Analysis for the Differentiation of Benign and Malignant Parotid Tumors

Qiang Yu[1], Anran Wang[1], Jinming Gu[1], Quanjiang Li[1], Youquan Ning[1], Juan Peng[1], Fajin Lv[1] and Xiaodi Zhang[2]*

[1] Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China,
[2] Philips Healthcare, Chengdu, China

Objective: This study aims to investigate the value of machine learning models based on clinical-radiological features and multiphasic CT radiomics features in the differentiation of benign parotid tumors (BPTs) and malignant parotid tumors (MPTs).

Methods: This retrospective study included 312 patients (205 cases of BPTs and 107 cases of MPTs) who underwent multiphasic enhanced CT examinations, which were randomly divided into training (*N* = 218) and test (*N* = 94) sets. The radiomics features were extracted from the plain, arterial, and venous phases. The synthetic minority oversampling technique was used to balance minority class samples in the training set. Feature selection methods were done using the least absolute shrinkage and selection operator (LASSO), mutual information (MI), and recursive feature extraction (RFE). Two machine learning classifiers, support vector machine (SVM), and logistic regression (LR), were then combined in pairs with three feature selection methods to build different radiomics models. Meanwhile, the prediction performances of different radiomics models based on single phase (plain, arterial, and venous phase) and multiphase (three-phase combination) were compared to determine which model construction method and phase were more discriminative. In addition, clinical models based on clinical-radiological features and combined models integrating radiomics features and clinical-radiological features were established. The prediction performances of the different models were evaluated by the area under the receiver operating characteristic (ROC) curve (AUC) and the drawing of calibration curves.

Results: Among the 24 established radiomics models composed of four different phases, three feature selection methods, and two machine learning classifiers, the LASSO-SVM model based on a three-phase combination had the optimal prediction performance with AUC (0.936 [95% CI = 0.866, 0.976]), sensitivity (0.78), specificity (0.90), and accuracy (0.86) in the test set, and its prediction performance was significantly better than with the clinical model based on LR (AUC = 0.781, $p = 0.012$). In the test set, the combined model

OPEN ACCESS

Edited by:
Yuming Jiang,
Stanford University, United States

Reviewed by:
Palash Ghosal,
Sikkim Manipal University, India
Marco Montella,
Università della Campania Luigi Vanvitelli, Italy

****Correspondence:***
Juan Peng
pengjuan0717@sina.com

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: *06 April 2022*
Accepted: *31 May 2022*
Published: *30 June 2022*

Citation:
Yu Q, Wang A, Gu J, Li Q, Ning Y, Peng J, Lv F and Zhang X (2022) Multiphasic CT-Based Radiomics Analysis for the Differentiation of Benign and Malignant Parotid Tumors. Front. Oncol. 12:913898. doi: 10.3389/fonc.2022.913898

based on LR had a lower AUC than the optimal radiomics model (AUC = 0.933 vs. 0.936), but no statistically significant difference (p = 0.888).

Conclusion: Multiphasic CT-based radiomics analysis showed a machine learning model based on clinical-radiological features and radiomics features has the potential to provide a valuable tool for discriminating benign from malignant parotid tumors.

Keywords: radiomics, machine learning, multiphasic CT, parotid tumors, differentiation

INTRODUCTION

Salivary gland tumors are relatively rare and most commonly occur in parotid glands, with benign tumors accounting for about 75% (1, 2). Surgery is the primary treatment for parotid tumors, but the clinical choices of surgical methods for benign parotid tumors (BPTs) and malignant parotid tumors (MPTs) are quite different. Local or superficial parotidectomy is the main treatment for BPTs, while more aggressive approaches are used in MPTs, including total or subtotal parotidectomy, even facial nerve resection, or postoperative chemoradiation (2–4). Therefore, accurate preoperative identification is critical to the choice of treatment and prognosis for patients. At present, ultrasound-guided core biopsy (CB) and fine-needle aspiration (FNA) are the major methods for preoperative differentiation of the types of parotid tumors, with the risk of serious surgical complications, such as facial paralysis and tumor implantation metastasis (5, 6). Additionally, ultrasound imaging of deep-lobe parotid tumors is occluded by tissue structures such as the mandible, which affects the evaluation of the tumors, thus making sampling difficult and the accuracy of the results largely dependendent on the operator' experience (2). Image examination is an important link in the achievement of accurate preoperative diagnosis of parotid tumors. For tumors occurring in the superficial lobe of parotid glands, ultrasound is the preferred method of examination, with limited value in the diagnosis of parotid tumors (7). CT and MRI are widely applied in preoperative localization, tumor invasion diagnoses, and differential diagnoses of parotid tumors, but conventional image evaluation largely depends on semantic features, and a large amount of information on tumor heterogeneity cannot be quantitatively elucidated (8). Although the application of multiparametric MRI in parotid tumors has increased, such as diffusion-weighted imaging (DWI) and dynamic contrast-enhanced MRI, its value in the differential diagnosis of benign and malignant parotid tumors is still controversial (9, 10).

In recent years, the application of radiomics to tumor diagnosis and treatment has been extensively studied. With the characteristics of high-throughput extraction of quantitative data from medical images in a noninvasive manner to explain the tumor heterogeneity, radiomics has rapidly developed into an emerging field in precision medicine (11). A previous study on the application of conventional CT radiomics to the differentiation of lympho-associated benign and malignant lesions of the parotid gland showed that it has a high differential ability (12). Another study demonstrated that dual-energy CT-based radiomics has a potential value in the differentiation of Warthin tumors from pleomorphic adenoma (13). However, this study only discussed the differential diagnosis of benign parotid tumors, and the sample size was relatively small. Xu et al. constructed a machine learning model based on the radiomics features extracted from plain and arterial phase scanning CT images to distinguish BPTs from MPTs (14), but only one machine learning classifier was utilized in their study. To the best of our knowledge, no studies have reported which phases, feature selection methods, and classifiers or their possible combinations are more discriminating in BPTs and MPTs. This will help in guiding the selection of the best model and phase for future multicenter studies of large datasets. This study aims to establish and validate machine learning prediction models based on CT radiomics features, clinical-radiological features, and a combination of the two types of features, and investigate their value in differentiating benign and malignant parotid tumors. At the same time, the prediction ability of differently combined radiomics models for BPTs and MPTs in the single-phase and multiphase were compared.

MATERIALS AND METHODS

Patient Cohorts

This retrospective study was approved by the institutional review board of our hospital (approval ID: 2020080), and the requirement for obtaining written informed consent from patients was waived.

Multiphasic enhanced CT images of patients with pathologically confirmed BPTs and MPTs from January 2014 to October 2021 were collected through a picture archiving and communication system (PACS). The inclusion criteria were as follows (1): patients with complete clinical and imaging data (2); contrast-enhanced CT examination was performed within 14 days before operation; and (3) patients did not receive any treatment prior to CT examination. The exclusion criteria were as follows (1): patients with recurrent tumor (2); CT images with obvious artifacts (3); patients with a maximum tumor diameter of less than 5 mm. Finally, 312 patients with parotid tumors were included, including 205 cases of BPTs and 107 cases of MPTs. The flowchart of patient recruitment is presented in **Figure 1**. **Supplementary Table S1** provides details of the pathological types of all patients with parotid tumors.

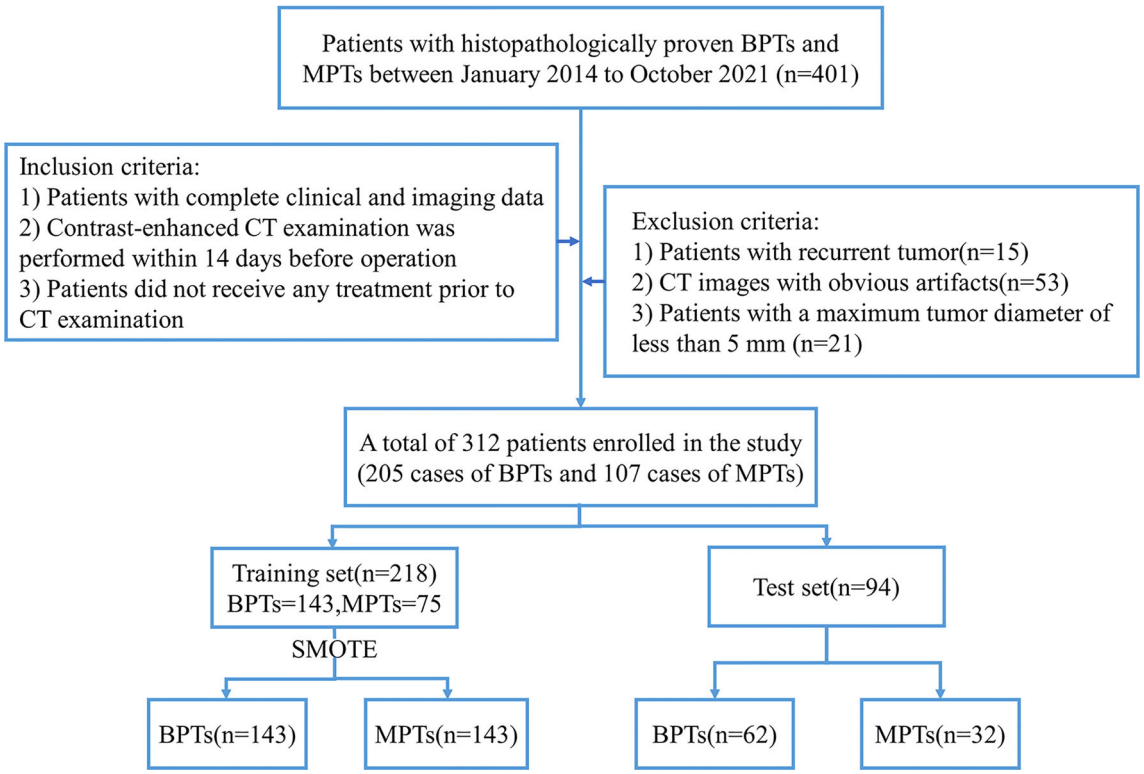


FIGURE 1 | Flowchart of patient recruitment. BPTs, benign parotid tumors; MPTs, malignant parotid tumors; SMOTE, the synthetic minority oversampling technique.

CT Image Acquisition

Axial three-phase scanning (including plain scan, arterial phase, and venous phase) was performed on each patient by a multislice spiral CT scanner. The CT scanners were as follows (1): Discovery CT750 HD (GE Healthcare, Milwaukee, WI, USA) (2), SOMATOM Definition Flash (Siemens Healthcare, Forchheim, Germany), and (3) SOMATOM Definition Force (Siemens Healthcare, Forchheim, Germany). The scan was performed from the skull base to the thoracic entrance. After the completion of plain scanning, the contrast agent iohexol (300 mg/ml) was injected at a flow rate of 3~4 ml/s, followed by 30 ml of normal saline with the dosage of contrast agent (1.5 ml/kg). Arterial and venous phase images were acquired 35 and 60 s after the contrast injection, respectively. The acquisition parameters of the above different devices are introduced in detail in **Supplementary Table S2**.

Clinical-Radiological Feature Evaluation

Univariate analysis was used to determine the statistically significant clinical-radiological features used for clinical model establishment. Clinical factors were collected, including, gender, age, and smoking history. The qualitative analysis of CT radiological features was evaluated by radiologists with 5 and 10 years of working experience (radiologists A and B, respectively), without knowing the pathological results of tumors. In the case of inconsistent evaluation results, the final results were obtained by a consensus between the two readers. The evaluation included the following radiological features (1): size (maximum tumor diameter in axial position) (2); tumor location (superficial or deep lobe was determined according to the main part of the tumor; the superficial and deep lobes are demarcated by a virtual line drawn from the lateral border of the posterior belly of the digastric muscle and retromandibular vein to the lateral edge of the mandible (15) (3); scope (localized or diffused; tumor involving local or whole parotid gland) (4); number (single or multiple) (5); shape (regular or irregular) (6); tumor margin (well-defined or ill-defined) (7); cystic or necrotic areas (CNA; absent or present) (8); infiltration of surrounding tissue (IST, absent or present; tumors involve surrounding muscles, bone, skin, or subcutaneous tissue (3) (9); lymphatic metastasis (LM; absent or present; obvious density change in cervical lymph nodes or short axis diameter >10 mm) (10); CT value at each phase (the solid section of the tumor was measured three times, and the average was calculated) (11); enhancement degree in P-A phase (difference in CT values between arterial and plain phases); and (12) enhancement degree in A-V phase (difference in CT values between venous and arterial phases).

Radiomics Feature Extraction and Selection

The CT images of all patients were loaded into the open-source image processing platform ITK-SNAP software (version 3.6.0, http://www.itksnap.org) in DICOM format. The 3D volume of

interest (VOI) of the tumor was formed by manual delineation layer by layer along the contour of the tumor on the plain scan, arterial phase, and venous phase CT images by radiologist A. In order to evaluate the reproducibility of features, 30 cases of CT images from each phase were randomly selected to evaluate intra- and interobserver agreement of radiomics features. Radiologist A performed the second tumor VOI segmentation at intervals of 1 week after the first tumor VOI segmentation, and radiologist B performed the tumor VOI segmentation independently. The intraclass correlation coefficient (ICC) was used to evaluate the intra- and interobserver agreement, and ICC values greater than 0.75 indicated good agreement.

PyRadiomics version 3.0 was used for feature extraction and image preprocessing. In order to reduce the influence of different scanning devices, the images were resampled to a voxel spacing of $1 \times 1 \times 1$ mm^3. The gray value discretization of the image was performed by the use of a fixed bin width of 25 HU to normalize image intensity and reduce image noise. In total, 851 features were extracted from each VOI, including the following four types of features (1): first-order statistics (18 features) (2), shape-based (14 features) (3), texture classes (75 features), and (4) wavelet features (744 features). The features of CT phase extraction at each phase were recorded in **Supplementary Excel S1**. The calculation formula and definition of the above features were provided in PyRadiomics documentation (https://pyradiomics.readthedocs.io/en/latest).

Before feature screening, all patients were randomized 7:3 into training and test sets. Feature selection plays an important role in reducing the task difficulty of the model and preventing model overfitting. The Mann–Whitney *U* test ($p < 0.05$) was used to initially screen the radiomics features (*ICC* >0.75). Before further feature screening, all features were normalized by the use of a Z-score to reduce the influence of different dimensions among features. In addition, in order to alleviate the impact of sample imbalance, the synthetic minority oversampling technique (SMOTE) was adopted to balance minority samples in the training set, thus leading to a sample proportion of 1:1. The algorithm has been proven to be helpful in avoiding overfitting of the model in the unbalanced data set and improve the overall generalization ability (16–18). On the balanced data set, least absolute shrinkage and selection operator (LASSO), mutual information (MI), and recursive feature extraction (RFE) were applied for radiomics feature further screening. **Figure 2** shows the workflow of radiomics analysis.

Radiomics Model Establishment and Validation

The clinical model served as the baseline model for the comparison between different models in our study. Radiomics models based on the CT radiomics features in each phase and combined models were also established. The combined model was established by the integration of the clinical-radiological features and radiomics features screened by LASSO in the arterial phase. There are significant differences in the prediction performance of models among different

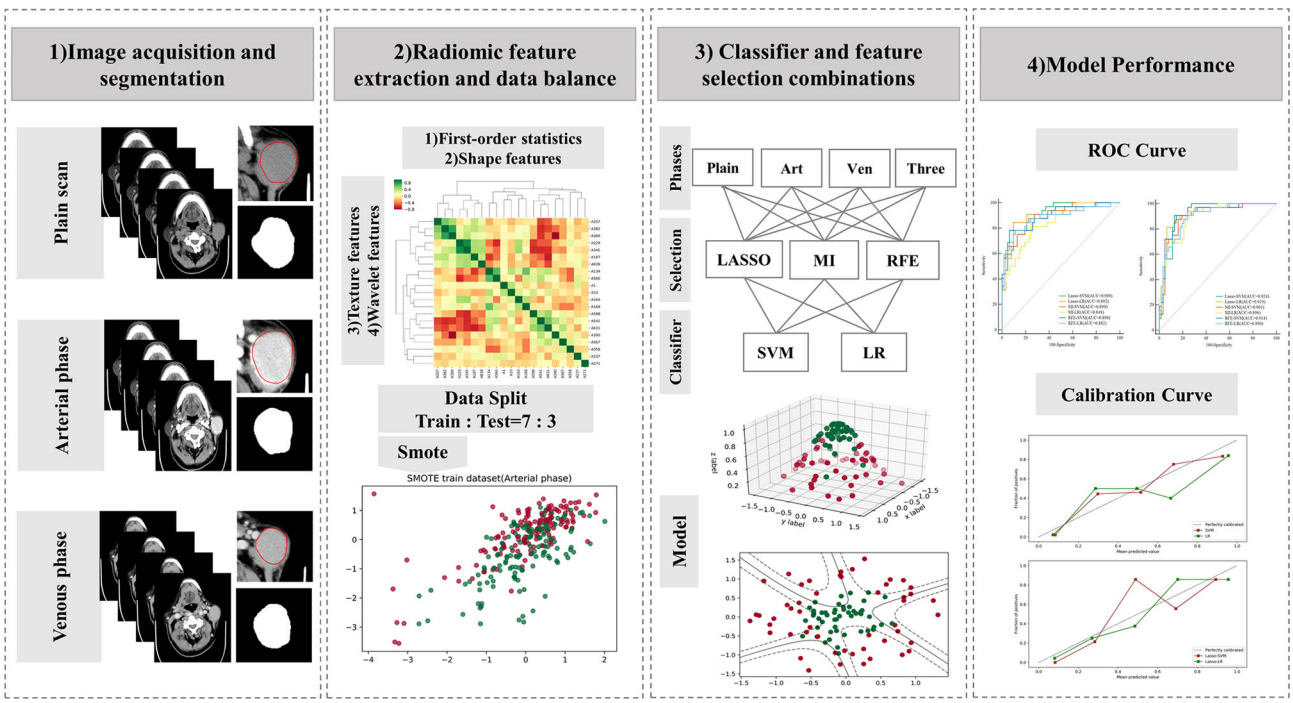


FIGURE 2 | Workflow of the radiomics analysis. LASSO, least absolute shrinkage and selection operator; MI, mutual information; RFE, recursive feature elimination; SVM, support vector machine; LR, logistic regression.

combinations of various machine learning classifiers and feature selection methods (19–21). Based on the radiomics features extracted from the plain scan, arterial, venous, and combined phase CT images, two common machine learning classifiers, support vector machine (SVM) and logistic regression (LR), were combined in pairs with three feature selection methods (LASSO, MI, and RFE) to generate a total of 24 models to determine the best performing radiomics model. In the training set, GridSearchCV (CV = 5, namely, 5-fold cross-validation) was used to optimize the hyperparameters of the model to reduce its training error and generalization one. For each model, the prediction performance of the machine learning model was evaluated by sensitivity, specificity, accuracy, and the area under the receiver operating characteristic (ROC) curve (AUC). In addition, the calibration curve was plotted to assess the calibration of models in the test set.

Statistical Analysis

Python version 3.7.3 and R version 3.6.0 were used to complete model establishment and statistical analysis. Quantitative variables were expressed as mean ± standard deviation (SD) or median and interquartile range (IQR), and categorical variables were expressed as numbers. An independent samples *t*-test or Mann–Whitney *U* test was adopted for quantitative variables, and the chi-square test or Fisher's exact test was used for categorical variables. The comparison of AUC differences between different models was completed by the Delong test. The level of significance was set at $p < 0.05$. The "imbalanced learn version 0.8.1" package was applied for data balancing in the training set. The feature screening and machine learning classifier construction were performed by the "scikit learn version 1.0.1" package.

RESULTS

Clinical-Radiological Factors and Clinical Models

The baseline table of clinical-radiological features of patients is shown in **Table 1**. In the training and test sets, there were significant differences in size, scope, shape, margin, IST, and LM between BPT and MPT groups ($p < 0.05$). Although the tumor location was not significantly different in the training set, a meta-analysis showed that tumor location (superficial lobe or deep lobe) may be a useful marker to help distinguish BPTs and MPTs (22). Therefore, the abovementioned seven clinical-radiological features were applied to the establishment of clinical models. In the established clinical models based on SVM and LR, the overall efficiency of the LR-based model was higher than that of the SVM-based one. The prediction performance of the LR-based model: in the training set, the AUC was 0.769 [95% CI = 0.716, 0.817], sensitivity was 0.50, specificity was 0.91, and accuracy was 0.71; in the test set, the AUC was 0.781 [95% CI = 0.684, 0.860], sensitivity was 0.53, specificity was 0.89, and accuracy was 0.77; the details of clinical models are shown in **Table 2**. The results showed that the prediction ability of clinical models for malignant parotid tumors was relatively low. **Figure 3A** shows the ROC curve of the clinical model in the test set.

Radiomics Feature Selection and Radiomics Models

Among the 851 radiomics features extracted from each phase, there were 680 highly repeatable features in the plain scan, 651 in the arterial phase, and 667 in the venous phase, respectively (ICC >0.75). The ICC analysis results are shown in **Supplementary Excel S2**. After preliminary screening by the

TABLE 1 | Clinical-radiological features of the training and test sets.

Clinical-radiological features	Training set (*n* = 218)			Test set (*n* = 94)		
	BPTs (*n* = 143)	MPTs (*n* = 75)	*p*-value	BPTs (*n* = 62)	MPTs (*n* = 32)	*p*-value
Gender[a] (M/F)	75/68	39/36	0.950	37/25	17/15	0.543
Age[b] (year)	55.00 (15.00)	48.00 (24.50)	0.023	53.69 (14.34)	51.38 (16.54)	0.483
Smoke[a] (absent/present)	81/62	47/28	0.391	34/28	21/11	0.315
Size[b] (mm)	19.77 (8.43)	23.09 (12.01)	0.003	20.84 (8.46)	28.83 (14.70)	0.001
Location[a] (superficial/deep lobe)	111/32	50/25	0.080	57/5	16/16	<0.001
Scope[a] (localized/diffuse)	143/0	65/10	<0.001	62/0	23/9	<0.001
Number[a] (single/multiple)	136/7	73/2	0.669	59/3	31/1	1.000
Shape[a] (regular/irregular)	96/47	36/39	0.006	51/11	14/18	<0.001
Margin[a] (well-defined/ill-defined)	123/20	44/31	<0.001	57/5	17/15	<0.001
CNA[a] (absent/present)	77/66	39/36	0.795	42/20	23/9	0.681
IST[a] (absent/present)	137/6	44/31	<0.001	57/5	17/15	<0.001
LM[a] (absent/present)	143/0	68/7	0.001	62/0	28/4	0.021
CT value[b] (HU)						
Plain	42.00 (15.00)	45.00 (16.50)	0.524	43.15 (11.94)	43.97 (13.26)	0.761
Arterial	84.00 (41.00)	79.00 (38.50)	0.591	86.69 (36.32)	84.31 (26.36)	0.743
Venous	86.00 (29.00)	87.00 (31.50)	0.961	86.94 (23.71)	89.50 (26.85)	0.636
Art-Pl	38.00 (36.50)	34.00 (30.00)	0.209	34.00 (39.50)	33.00 (24.75)	0.886
Ven-Art	3.07 (23.72)	5.03 (18.21)	0.499	6.00 (28.75)	8.00 (16.75)	0.330

[a]*Categorical data as numbers (n).*

[b]*Quantitative data are mean (standard deviation) or median (quartile). p-value was calculated with independent samples t-test or Mann–Whitney U test. p-value was calculated with the* χ^2 *or Fisher's exact test. BPTs, benign parotid tumors; MPTs, malignant tumors; F, female; M, male; CNA, cystic or necrotic areas; LM, lymphatic metastasis; CT, computed tomography; Art, arterial phase; Pl, plain scan; Ven, venous phase.*

TABLE 2 | Diagnostic performance of the clinical, radiomics, and combined models.

Model	AUC (95%CI)	Sensitivity	Specificity	Accuracy	*p*-value
Training set					
Clinical model (LR)	0.769 (0.716 to 0.817)	0.50	0.91	0.71	–
Clinical model (SVM)	0.871 (0.827 to 0.908)	0.66	0.93	0.80	<0.001
Radiomics model 1	0.951 (0.919 to 0.973)	0.84	0.92	0.88	<0.001
Radiomics model 2	0.991 (0.972 to 0.999)	0.96	0.97	0.96	<0.001
Combined model (LR)	0.950 (0.918 to 0.972)	0.85	0.94	0.90	<0.001
Combined model (SVM)	0.978 (0.854 to 0.992)	0.92	0.97	0.95	<0.001
Test set					
Clinical model (LR)	0.781 (0.684 to 0.860)	0.53	0.89	0.77	–
Clinical model (SVM)	0.744 (0.644 to 0.828)	0.69	0.85	0.80	0.539
Radiomics model 1	0.924 (0.850 to 0.968)	0.78	0.89	0.85	0.012
Radiomics model 2	0.936 (0.866 to 0.976)	0.78	0.90	0.86	0.013
Combined model (LR)	0.933 (0.862 to 0.974)	0.75	0.85	0.82	0.006
Combined model (SVM)	0.928 (0.856 to 0.971)	0.81	0.87	0.85	0.009

The p-value was calculated by the Delong test. AUC, area under the receiver operating characteristic curve; CI, confidence interval; SVM, support vector machine; LR, logistic regression. Radiomics model 1, the best radiomics model based on a single phase (arterial-LASSO-SVM). Radiomics model 2, the best radiomics model based on the multiphasic phase (three-LASSO-SVM).

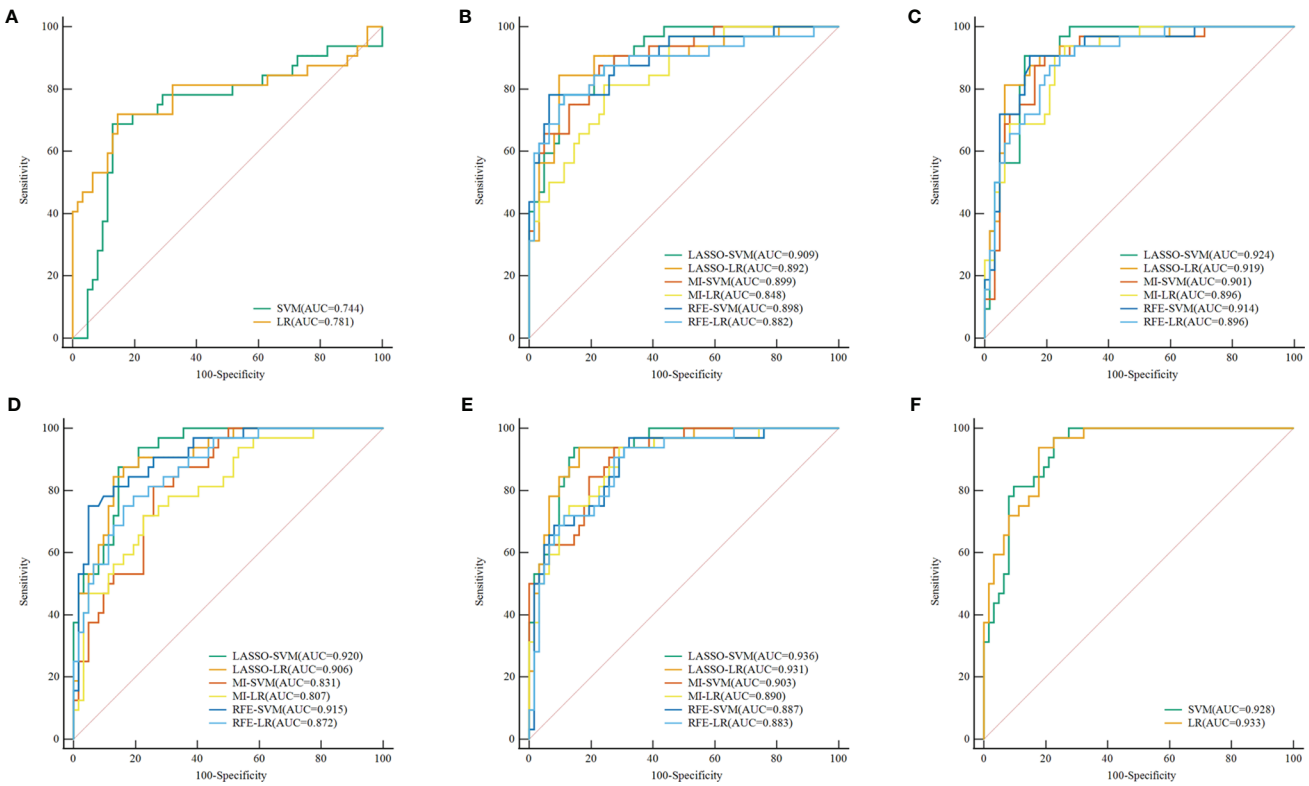


FIGURE 3 | ROC curve of the clinical, radiomics, and combined models in the test set, respectively. **(A)** ROC curve of Clinical models. **(B–E)** ROC curve of radiomics models (different combinations of 3 feature selection methods and 2 classifiers) based on the plain scan, arterial phase, venous phase, and three-phase combination. **(F)** ROC curve of combined models. ROC, receiver operating characteristic; LASSO, least absolute shrinkage and selection operator; MI, mutual information; RFE, recursive feature elimination; SVM, support vector machine; LR, logistic regression.

Mann–Whitney *U* test, SMOTE was adopted to balance the minority class for samples in the training set, with the proportion of samples in the training set adjusted to 1:1 (BPTs = 143, MPTs = 143). Three feature screening methods (LASSO, MI, and RFE) combined with two machine learning methods (SVM and LR) were then used to establish 24 radiomics models in the plain scan, arterial phase, and venous phase, as well as a three-phase combination. The results of different feature screening methods in each phase are shown in **Supplementary Table S3**.

TABLE 3 | Diagnostic performance of the radiomics models.

Model	SVM		LR	
	Training	Test	Training	Test
Arterial phase				
LASSO				
AUC (95% CI)	0.951 (0.919 to 0.973)	0.924 (0.850 to 0.968)	0.936 (0.901 to 0.961)	0.919 (0.845 to 0.966)
Sensitivity	0.84	0.78	0.86	0.81
Specificity	0.92	0.89	0.90	0.94
Accuracy	0.88	0.85	0.88	0.89
ML				
AUC (95% CI)	0.896 (0.855 to 0.929)	0.901 (0.822 to 0.953)	0.881 (0.838 to 0.916)	0.896 (0.816 to 0.950)
Sensitivity	0.78	0.75	0.81	0.78
Specificity	0.88	0.85	0.81	0.79
Accuracy	0.83	0.82	0.81	0.79
RFE				
AUC (95% CI)	0.909 (0.869 to 0.940)	0.914 (0.838 to 0.962)	0.904 (0.863 to 0.935)	0.896 (0.816 to 0.950)
Sensitivity	0.81	0.84	0.80	0.72
Specificity	0.90	0.87	0.87	0.84
Accuracy	0.86	0.86	0.84	0.80
Venous phase				
LASSO				
AUC (95% CI)	0.959 (0.929 to 0.979)	0.920 (0.846 to 0.966)	0.918 (0.880 to 0.947)	0.906 (0.828 to 0.957)
Sensitivity	0.89	0.69	0.82	0.78
Specificity	0.87	0.87	0.82	0.87
Accuracy	0.88	0.81	0.82	0.84
ML				
AUC (95% CI)	0.856 (0.810 to 0.894)	0.831 (0.739 to 0.900)	0.829 (0.780 to 0.871)	0.807 (0.713 to 0.882)
Sensitivity	0.75	0.81	0.73	0.75
Specificity	0.78	0.74	0.74	0.73
Accuracy	0.77	0.77	0.73	0.73
RFE				
AUC (95% CI)	0.926 (0.889 to 0.954)	0.915 (0.839 to 0.963)	0.909 (0.869 to 0.940)	0.872 (0.787 to 0.932)
Sensitivity	0.80	0.72	0.81	0.69
Specificity	0.89	0.95	0.83	0.87
Accuracy	0.85	0.87	0.82	0.81
Plain scan				
LASSO				
AUC (95% CI)	0.991 (0.972 to 0.998)	0.909 (0.832 to 0.959)	0.991 (0.972 to 0.998)	0.892 (0.811 to 0.947)
Sensitivity	0.94	0.72	0.96	0.81
Specificity	0.95	0.90	0.94	0.90
Accuracy	0.94	0.84	0.95	0.87
ML				
AUC (95% CI)	0.900 (0.859 to 0.932)	0.899 (0.819 to 0.951)	0.904 (0.864 to 0.935)	0.848 (0.759 to 0.914)
Sensitivity	0.77	0.75	0.84	0.62
Specificity	0.86	0.87	0.85	0.84
Accuracy	0.81	0.83	0.84	0.77
RFE				
AUC (95% CI)	0.933 (0.897 to 0.959)	0.898 (0.819 to 0.951)	0.921 (0.884 to 0.950)	0.882 (0.799 to 0.939)
Sensitivity	0.81	0.78	0.85	0.78
Specificity	0.90	0.92	0.84	0.87
Accuracy	0.85	0.87	0.84	0.84
Three-phase combination				
LASSO				
AUC (95% CI)	0.991 (0.972 to 0.999)	0.936 (0.866 to 0.976)	0.994 (0.977 to 1.000)	0.931 (0.859 to 0.973)
Sensitivity	0.96	0.78	0.97	0.75
Specificity	0.97	0.90	0.97	0.94
Accuracy	0.96	0.86	0.97	0.87
ML				
AUC (95% CI)	0.912 (0.873 to 0.942)	0.903 (0.824 to 0.954)	0.914 (0.875 to 0.943)	0.890 (0.809 to 0.945)
Sensitivity	0.81	0.69	0.83	0.75
Specificity	0.86	0.84	0.86	0.84
Accuracy	0.84	0.79	0.85	0.81
RFE				
AUC (95% CI)	0.908 (0.869 to 0.939)	0.887 (0.805 to 0.943)	0.909 (0.870 to 0.940)	0.883 (0.800 to 0.940)

(Continued)

TABLE 3 | Continued

Model	SVM		LR	
	Training	Test	Training	Test
Sensitivity	0.81	0.69	0.82	0.72
Specificity	0.89	0.85	0.85	0.89
Accuracy	0.85	0.80	0.84	0.83

AUC, area under the receiver operating characteristic curve; CI, confidence interval; SVM, support vector machine; LR, logistic regression; LASSO, least absolute shrinkage and selection operator; MI, mutual information; RFE, recursive feature elimination.

The discrimination indicators of all radiomics models are shown in **Table 3**. **Figures 3B–E** show the ROC curve of all radiomics models in the test set. The AUC values of all radiomics models ranged from 0.819 to 0.994 in the training set and 0.807 to 0.936 in the test set (**Figure 4**). Among all radiomics models, LASSO-SVM models based on three-phase combination had the highest discrimination efficiency: in the training set, the AUC was 0.991 [95% CI = 0.972, 0.999], sensitivity was 0.96, specificity was 0.97, and accuracy was 0.96; in the test set, the AUC was 0.936 [95% CI = 0.866, 0.976], sensitivity was 0.78, specificity was 0.90, and accuracy was 0.86; in the test set, the prediction performance was significantly better than that of the LR-based clinical models ($p = 0.012$, Delong test). Among the three feature screening methods, the efficiency of LASSO combined with SVM or LR machine learning classifier was generally better than that of MI and RFE feature screening methods.

The Performance of Phases and Combined Model

For single phase, the performance of models based on the arterial phase was generally better than that in the venous and plain scan phases. LASSO-SVM model based on the arterial phase has the highest prediction performance: in the test set, the AUC was 0.924 [95% CI = 0.850, 0.968], sensitivity was 0.78, specificity was 0.89, and accuracy was 0.85. It is worth noting that although the prediction performance of the radiomics models based on the plain scan was generally lower than that of those based on arterial and venous phases, it also achieved a high one. The AUC of LASSO-SVM models based on plain scan in the test set was 0.909 [95% CI = 0.832, 0.959], which was significantly higher than that of LR-based clinical model (AUC = 0.781, $p = 0.045$, Delong test). For multiphase, the LASSO-SVM model based on a three-phase combination achieved the best prediction performance in all phases, which was constructed with 62 radiomics features obtained from multiphase sequences (three-phase combination) by LASSO. However, too many features will increase the complexity of the model. Therefore, in order to avoid overfitting caused by more features, we chose to integrate 21 radiological features from the single phase (arterial phase) by LASSO with clinical-radiological features to establish combined models.

In the training and test sets, the prediction performance of the combined model was better than that of the clinical model. The prediction performance of LR-based combined model in the test set was as follows: the AUC was 0.933 [95% CI = 0.862, 0.974], sensitivity was 0.75, specificity was 0.85, and accuracy was 0.82. The diagnostic efficiency of the LR-based combined model was significantly better than that of the LR-based clinical model

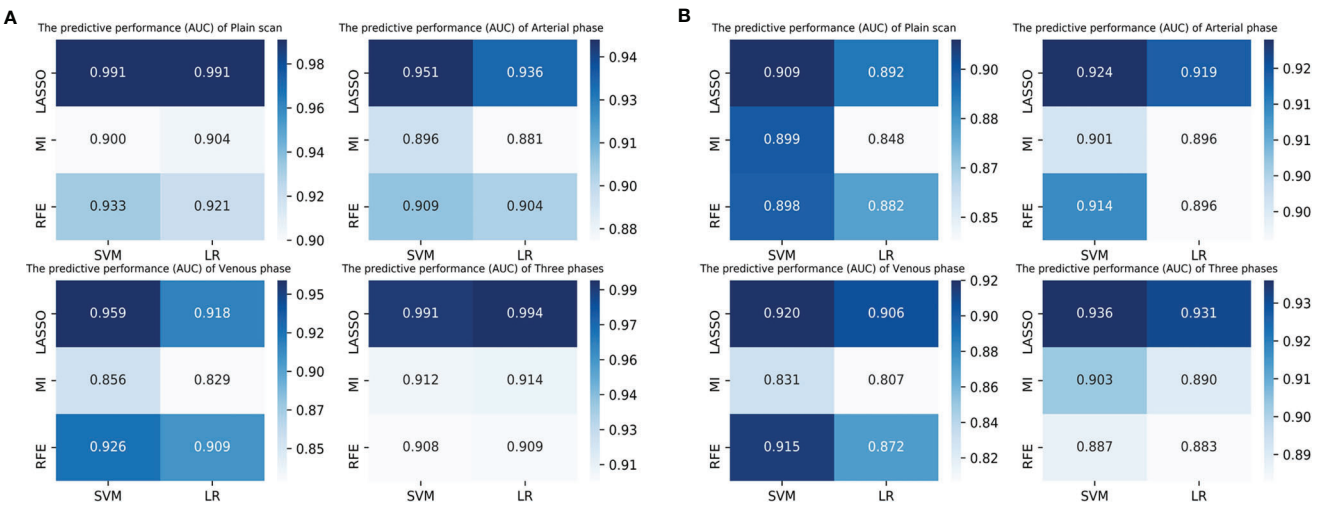


FIGURE 4 | The predictive performance (AUC) of radiomics models (3 feature selection methods and 2 classifiers) based on different phases in the training set **(A)** and test set **(B)**. LASSO, least absolute shrinkage and selection operator; MI, mutual information; RFE, recursive feature elimination; SVM, support vector machine; LR, logistic regression.

(AUC = 0.933 vs. 0.781, p = 0.006, Delong test); however, the prediction performance of the LR-based combined mode (AUC = 0.933 vs. 0.936, p = 0.888, Delong test) was similar to that of the three-phase combined LASSO-SVM radiomics model (**Table 2**). **Figure 3F** depicts the ROC curve of combined models in the test set. A comparison of the ROC curve between different models (clinical models, the optimal radiomics models base on single-phase and multiphase, and combined modes) in the training set (**A**) and test set (**B**) is shown in **Figure 5**. The calibration curves of the LASSO radiomics model based on the arterial phase, the LASSO radiomics model based on a three-phase combination, and the combined models are shown in **Figure 6**. These models all showed good calibration performance.

DISCUSSION

In this study, prediction models of benign and malignant parotid tumors based on clinical-radiological features, radiomics features, and combined features were established and validated. Among the 28 established prediction models, radiomics models and combined ones achieved outstanding performance. More specifically, in the independent test set, the LASSO-SVM radiomics model (AUC = 0.936) based on a three-phase combination and the LR-based combined model (AUC = 0.933) had higher prediction accuracy in the differentiation of BPTs from MPTs compared with the optimal clinical model based on LR (AUC = 0.781, p = 0.012, p = 0.006), and they showed better calibration ability. This suggested that the developed models may be helpful to the preoperative diagnosis of BPTs and MPTs.

In conventional image diagnosis, diffuse tumor growth patterns, irregular shape, ill-defined margin, deep lobe lesions, surrounding tissue involvement, and lymphatic metastasis are considered to be more common in MPTs (22–24). The results of this study are consistent with those of previous studies. A meta-analysis showed that the sensitivities of US, CT, and MRI in the differentiation of benign and malignant tumors of salivary glands were 0.66, 0.70, and 0.80, respectively (25). Another conventional MRI analysis showed that the sensitivities of the diagnosis of MPTs only by infiltration of surrounding tissue and irregular shape were 0.68 and 0.16, respectively (23). These showed that the overlapping imaging features between BPTs and MPTs are the main limitations of conventional radiology diagnoses. In this study, the AUC values of SVM and LR clinical models based on clinical-radiological features (size, scope, shape, margin, location, IST, and LM) were 0.744 and 0.781 respectively, in the test set, but the sensitivities to the two models were only 0.69 and 0.53, respectively, which suggested that the prediction model established only by clinical-radiological features cannot differentiate BPTs from MPTs well. The radiomics features can reflect subtle differences between tumors that cannot be recognized by the naked eye. The diagnostic efficiency of established radiomics models (AUC = 0.807~0.936, sensitivity = 0.62~0.84) in the test set was better than that of the clinical models (AUC = 0.744, 0.781, sensitivity = 0.69, 0.53), and the overall prediction accuracy for malignant tumors significantly improved. For combined models, although the prediction performance of the optimal LR combined model (combined with 7 clinical factors and radiomics features screened by LASSO in the arterial phase) had no improvement on that of LASSO-SVM radiomics models based on a three-phase combination (AUC = 0.933 vs. 0.936), it had a great improvement on that of the LASSO-LR radiomics model based on the arterial phase (AUC = 0.933 vs. 0.919). This suggests that the mutual complementation of clinical-radiological features and radiomics ones has the greatest benefit for the diagnosis of BPTs and MPTs, which may benefit from important extratumoral features such as lymphatic metastasis and tumor infiltration into the surrounding tissue.

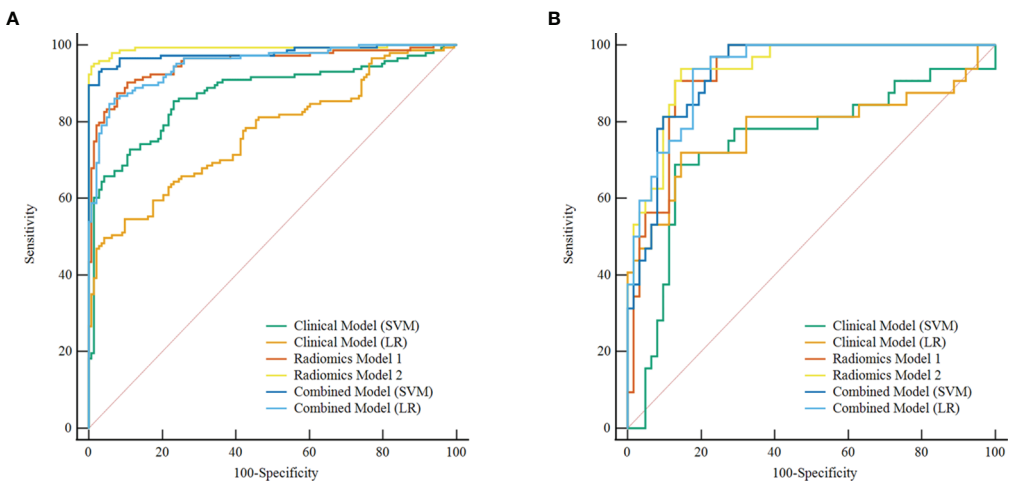


FIGURE 5 | Comparison of ROC curve between different models in the training set **(A)** and test set **(B)**. ROC, receiver operating characteristic; SVM, support vector machine; LR, logistic regression. Radiomics model 1, the best radiomics model based on a single phase (arterial-LASSO-SVM); Radiomics model 2, the best radiomics model based on a multiphasic phase (three-LASSO-SVM).

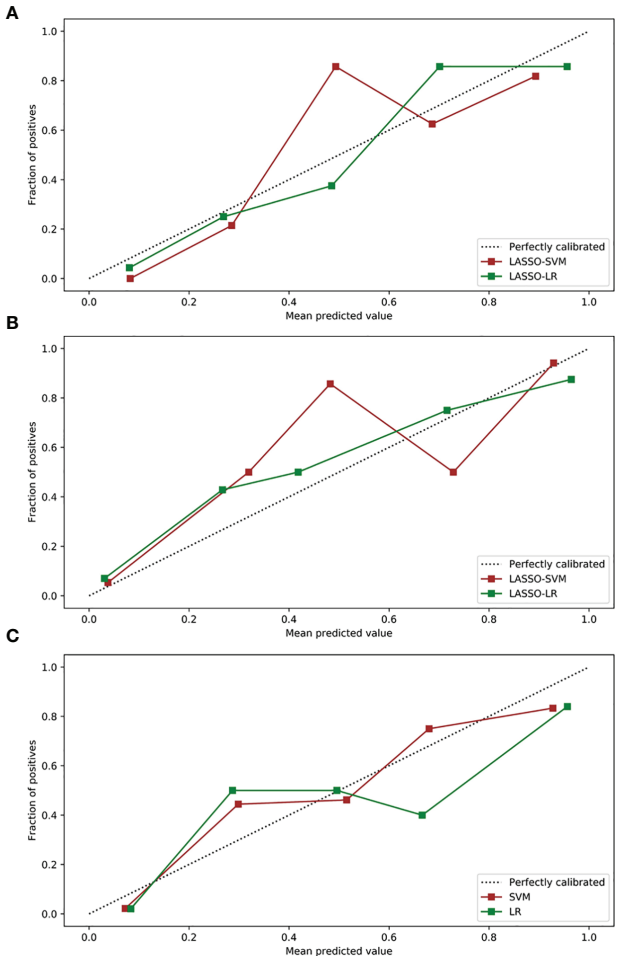


FIGURE 6 | Calibration curves of radiomics and combined models in the test sets. **(A)** Calibration curves of radiomics models based on arterial phase. **(B)** Calibration curves of radiomics models based on a three-phase combination. **(C)** Calibration curves of combined models. LASSO, least absolute shrinkage, and selection operator; SVM, support vector machine; LR, logistic regression.

Among the screened radiomics features by three feature selection methods, the sphericity based on shape features was considered to be highly related to the diagnosis of BPTs and MPTs. Sphericity is a dimensionless metric that is independent of scale and orientation and may be applied to estimate the roundness of the shape of the tumor region relative to a circle; the closer the value is to 1, the closer the tumor is to the perfect sphere. A previous study showed that among the radiomics features extracted from T2WI images, the volume density AEE value related to sphericity was higher in Warthin tumors than in MPTs (26). In this study, the sphericity in the arterial phase was significantly different between BPTs and MPTs ($p < 0.05$, Mann–Whitney U test), with the value in BPTs generally higher. This quantitative index showed that the morphology of BPTs is more regular compared with MPTs. In addition, wavelet features had the highest weight in the radiomics labels screened by different feature screening methods, which indicates that wavelet features may reflect the spatial heterogeneity of tumors on several scales (27, 28).

In recent years, there have been many radiomics studies focused on the diagnosis of parotid tumors and the prediction of side effects related to radiotherapy (12–14, 29), but they are mainly based on MRI radiomics. Zheng et al. extracted the radiomics features of benign and malignant parotid tumors from TWI and T2WI sequences and established the radiomics nomogram model by multivariate logistic regression analysis (30). The AUC value of this model reached 0.938 in the differentiation of BPTs from MPTs, which was close to the prediction performance of our optimal model (AUC = 0.936). In addition, some scholars have also applied CT radiomics to the differentiation of benign and malignant lymph-related lesions of parotid glands and benign parotid tumors (12, 13), which have both shown excellent predictive performance. Xu et al. established an SVM-based combined prediction model based on the radiomics features extracted from plain CT scan and arterial phase combined with conventional CT image features to differentiate benign and malignant parotid tumors, and in the test set, model diagnosis results of BPTs and MPTs were: accuracy, 0.84; specificity, 0.74; and sensitivity, 0.82 (14). In contrast, the diagnostic efficiency of LASSO-SVM radiomics model based on three-phase combination (accuracy, 0.86; specificity, 0.90; sensitivity, 0.78) was relatively high. In contrast to the above research, in this study, the SMOTE algorithm was adopted to balance minority class samples in the training set and achieve a better class balance. At the same time, GridSearchCV was applied to the optimization of the model hyperparameters. All these methods effectively reduced the training error of models in the training set and improve overall prediction performance in the test set, which have also been validated in other studies (16–18). In addition, in this study, a variety of feature screening methods and machine learning classifiers were used. The results showed that the overall efficiency of the LASSO-SVM-based model was outstanding. Moreover, the prediction performances of different radiomics models were compared in different phases. The results showed that in the single phase, the prediction performance of the models based on the arterial phase was generally better than that of those based on plain scan and venous phase. It is worth mentioning that although the prediction performance of radiomics models based on the plain scan was lower than that of those based on the arterial phase, it was significantly better than that of clinical models. This research result may be beneficial to the popularization of the prediction model in patients with parotid tumors, especially for those who are not suitable for enhanced scanning.

Our study has limitations. First, this study was limited by the single-center studies and low incidence rate of parotid malignant tumors, with a small sample size of included parotid malignancies. Second, in this study, an independent internal test set was used to verify the reliability of the model without being supported by the external test set, and the generalization ability of models still needs to be further validated by multicenter prospective research. Third, deep learning has developed rapidly

in the medical field, which has the ability to process image information more efficiently compared with the traditional machine learning classifier. In the future, deep learning could be applied to the multiclassification task of parotid tumors. Finally, the multiomics combination is the development trend of radiomics in the future, including radiogenomics, radiopathomics, and multiradiomics combinations based on different medical images. In future studies, the application of multiomics combined models to the diagnosis and treatment of parotid tumors can be explored.

CONCLUSION

In conclusion, the radiomics models and the combined ones established in this study showed high prediction accuracy in the diagnosis of benign and malignant parotid tumors, with obvious advantages compared with conventional image diagnosis, which may provide a valuable tool for clinical decision-making of patients with parotid tumors.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of the First Affiliated Hospital of Chongqing Medical University. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

JP and QY designed the research. AW, JG, QL, and YN collected the data and preprocessed the data. QY, XZ, and JG performed major data analyses and drafted the manuscript. JP and FL participated in the review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This project received support from The Foundation of Science and Technology Bureau of Yuzhong District, Chongqing, China (Grant No. 20190111) and the Natural Science Foundation of Chongqing, China (Grant No. cstc2021jcyj-msxmX0020).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.913898/full#supplementary-material

REFERENCES

1. Lanisnik B, Levart P, Cizmarevic B, Svagan M. Surgeon-Performed Ultrasound With Fine-Needle Aspiration Biopsy for the Diagnosis of Parotid Gland Tumors. *Head Neck* (2021) 43(6):1739–46. doi: 10.1002/hed.26630
2. Moore MG, Yueh B, Lin DT, Bradford CR, Smith RV, Khariwala SS. Controversies in the Workup and Surgical Management of Parotid Neoplasms. *Otolaryngol Head Neck Surg* (2021) 164(1):27–36. doi: 10.1177/0194599820932512
3. Stoia S, Baciut G, Lenghel M, Badea R, Csutak C, Rusu GM, et al. Cross-Sectional Imaging and Cytologic Investigations in the Preoperative Diagnosis of Parotid Gland Tumors - an Updated Literature Review. *Bosn J Basic Med Sci* (2021) 21(1):19–32. doi: 10.17305/bjbms.2020.5028
4. Cracchiolo JR, Shaha AR. Parotidectomy for Parotid Cancer. *Otolaryngol Clin North Am* (2016) 49(2):415–24. doi: 10.1016/j.otc.2015.10.007
5. Walsh E, Allan K, Brennan PA, Tullett M, Gomez RS, Rahimi S. Diagnostic Accuracy of Ultrasonography-Guided Core Needle Biopsy of Parotid Gland Neoplasms: A Large, Single-Institution Experience in United Kingdom. *J Oral Pathol Med* (2022) 51(1):1–4. doi: 10.1111/jop.13260
6. Hurry KJ, Karunaratne D, Westley S, Booth A, Ramesar K, Zhang TT, et al. Ultrasound-Guided Core Biopsy in the Diagnosis of Parotid Neoplasia: An Overview and Update With a Review of the Literature. *Br J Radiol* (2022) 95 (1130):20210972. doi: 10.1259/bjr.20210972
7. Liu Y, Li J, Tan YR, Xiong P, Zhong LP. Accuracy of Diagnosis of Salivary Gland Tumors With the Use of Ultrasonography, Computed Tomography, and Magnetic Resonance Imaging: A Meta-Analysis. *Oral Surg Oral Med Oral Pathol Oral Radiol* (2015) 119(2):238–45.e2. doi: 10.1016/j.oooo.2014.10.020
8. Bi WL, Hosny A, Schabath MB, Giger ML, Birkbak NJ, Mehrtash A, et al. Artificial Intelligence in Cancer Imaging: Clinical Challenges and Applications. *CA Cancer J Clin* (2019) 69(2):127–57. doi: 10.3322/caac.21552
9. Yuan Y, Tang W, Tao X. Parotid Gland Lesions: Separate and Combined Diagnostic Value of Conventional Mri, Diffusion-Weighted Imaging and Dynamic Contrast-Enhanced Mri. *Br J Radiol* (2016) 89(1060):20150912. doi: 10.1259/bjr.20150912
10. Xu Z, Chen M, Zheng S, Chen S, Xiao J, Hu Z, et al. Differential Diagnosis of Parotid Gland Tumours: Application of Swi Combined With Dwi and Dce-Mri. *Eur J Radiol* (2022) 146:110094. doi: 10.1016/j.ejrad.2021.110094
11. Tomaszewski MR, Gillies RJ. The Biological Meaning of Radiomic Features. *Radiology* (2021) 298(3):505–16. doi: 10.1148/radiol.2021202553
12. Zheng YM, Xu WJ, Hao DP, Liu XJ, Gao CP, Tang GZ, et al. A Ct-Based Radiomics Nomogram for Differentiation of Lympho-Associated Benign and Malignant Lesions of the Parotid Gland. *Eur Radiol* (2021) 31(5):2886–95. doi: 10.1007/s00330-020-07421-4
13. Al Ajmi E, Forghani B, Reinhold C, Bayat M, Forghani R. Spectral Multi-Energy Ct Texture Analysis With Machine Learning for Tissue Classification: An Investigation Using Classification of Benign Parotid Tumours as a Testing Paradigm. *Eur Radiol* (2018) 28(6):2604–11. doi: 10.1007/s00330-017-5214-0
14. Xu Y, Shu Z, Song G, Liu Y, Pang P, Wen X, et al. The Role of Preoperative Computed Tomography Radiomics in Distinguishing Benign and Malignant

Tumors of the Parotid Gland. *Front Oncol* (2021) 11:. doi: 10.3389/fonc.2021.634452
15. Christe A, Waldherr C, Hallett R, Zbaeren P, Thoeny H. Mr Imaging of Parotid Tumors: Typical Lesion Characteristics in Mr Imaging Improve Discrimination Between Benign and Malignant Disease. *AJNR Am J Neuroradiol* (2011) 32(7):1202–7. doi: 10.3174/ajnr.A2520
16. Yang R, Wu J, Sun L, Lai S, Xu Y, Liu X, et al. Radiomics of Small Renal Masses on Multiphasic Ct: Accuracy of Machine Learning-Based Classification Models for the Differentiation of Renal Cell Carcinoma and Angiomyolipoma Without Visible Fat. *Eur Radiol* (2020) 30(2):1254–63. doi: 10.1007/s00330-019-06384-5
17. Xie C, Du R, Ho JW, Pang HH, Chiu KW, Lee EY, et al. Effect of Machine Learning Re-Sampling Techniques for Imbalanced Datasets in (18)F-Fdg Pet-Based Radiomics Model on Prognostication Performance in Cohorts of Head and Neck Cancer Patients. *Eur J Nucl Med Mol Imaging* (2020) 47(12):2826–35. doi: 10.1007/s00259-020-04756-4
18. He B, Ji T, Zhang H, Zhu Y, Shu R, Zhao W, et al. Mri-Based Radiomics Signature for Tumor Grading of Rectal Carcinoma Using Random Forest Model. *J Cell Physiol* (2019) 234(11):20501–9. doi: 10.1002/jcp.28650
19. Wang T, Gong J, Li Q, Chu C, Shen W, Peng W, et al. A Combined Radiomics and Clinical Variables Model for Prediction of Malignancy in T2 Hyperintense Uterine Mesenchymal Tumors on Mri. *Eur Radiol* (2021) 31 (8):6125–35. doi: 10.1007/s00330-020-07678-9
20. 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
21. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts H. Machine Learning Methods for Quantitative Radiomic Biomarkers. *Sci Rep* (2015) 5:13087. doi: 10.1038/srep13087
22. Aasen MH, Hutz MJ, Yuhan BT, Britt CJ. Deep Lobe Parotid Tumors: A Systematic Review and Meta-Analysis. *Otolaryngol Head Neck Surg* (2022) 166(1):60–7. doi: 10.1177/01945998211009235
23. Tartaglione T, Botto A, Sciandra M, Gaudino S, Danieli L, Parrilla C, et al. Differential Diagnosis of Parotid Gland Tumours: Which Magnetic Resonance Findings Should Be Taken in Account? *Acta Otorhinolaryngol Ital* (2015) 35 (5):314–20. doi: 10.14639/0392-100X-693
24. Kato H, Kanematsu M, Watanabe H, Mizuta K, Aoki M. Salivary Gland Tumors of the Parotid Gland: Ct and Mr Imaging Findings With Emphasis on Intratumoral Cystic Components. *Neuroradiology* (2014) 56(9):789–95. doi: 10.1007/s00234-014-1386-3
25. Kong X, Li H, Han Z. The Diagnostic Role of Ultrasonography, Computed Tomography, Magnetic Resonance Imaging, Positron Emission Tomography/Computed Tomography, and Real-Time Elastography in the Differentiation of Benign and Malignant Salivary Gland Tumors: A Meta-Analysis. *Oral Surg Oral Med Oral Pathol Oral Radiol* (2019) 128(4):431–43.e1. doi: 10.1016/j.oooo.2019.06.014
26. Piludu F, Marzi S, Ravanelli M, Pellini R, Covello R, Terrenato I, et al. Mri-Based Radiomics to Differentiate Between Benign and Malignant Parotid Tumors With External Validation. *Front Oncol* (2021) 11. doi: 10.3389/fonc.2021.656918
27. Ji GW, Zhu FP, Xu Q, Wang K, Wu MY, Tang WW, et al. Radiomic Features at Contrast-Enhanced Ct Predict Recurrence in Early Stage Hepatocellular Carcinoma: A Multi-Institutional Study. *Radiology* (2020) 294(3):568–79. doi: 10.1148/radiol.2020191470
28. Liang W, Yang P, Huang R, Xu L, Wang J, Liu W, et al. A Combined Nomogram Model to Preoperatively Predict Histologic Grade in Pancreatic Neuroendocrine Tumors. *Clin Cancer Res* (2019) 25(2):584–94. doi: 10.1158/1078-0432.CCR-18-1305
29. van Dijk LV, Langendijk JA, Zhai TT, Vedelaar TA, Noordzij W, Steenbakkers R, et al. Delta-Radiomics Features During Radiotherapy Improve the Prediction of Late Xerostomia. *Sci Rep* (2019) 9(1):12483. doi: 10.1038/s41598-019-48184-3
30. Zheng YM, Li J, Liu S, Cui JF, Zhan JF, Pang J, et al. Mri-Based Radiomics Nomogram for Differentiation of Benign and Malignant Lesions of the Parotid Gland. *Eur Radiol* (2021) 31(6):4042–52. doi: 10.1007/s00330-020-07483-4

Conflict of Interest: Author XZ was employed by Philips Healthcare.

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

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

frontiers | Frontiers in Oncology

ORIGINAL RESEARCH
published: 01 July 2022
doi: 10.3389/fonc.2022.883516

A Transfer Learning Framework for Deep Learning-Based CT-to-Perfusion Mapping on Lung Cancer Patients

*Ge Ren[1], Bing Li[1,2], Sai-kit Lam[1], Haonan Xiao[1], Yu-Hua Huang[1], Andy Lai-yin Cheung[3], Yufei Lu[2], Ronghu Mao[2], Hong Ge[2], Feng-Ming (Spring) Kong[3,4], Wai-yin Ho[5] and Jing Cai[1]**

[1] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong, Hong Kong SAR, China, [2] Department of Radiotherapy, Affiliated Cancer Hospital of Zhengzhou University/Henan Cancer Hospital, Zhengzhou, China, [3] Department of Clinical Oncology, Queen Mary Hospital, Hong Kong, Hong Kong SAR, China, [4] Department of Clinical Oncology, The University of Hong Kong, Hong Kong, Hong Kong SAR, China, [5] Department of Nuclear Medicine, Queen Mary Hospital, Hong Kong, Hong Kong SAR, China

OPEN ACCESS

Edited by:
Ye Zhang,
Paul Scherrer Institute (PSI), Switzerland

Reviewed by:
Kun Qing,
City of Hope National Medical Center, United States
Yibao Zhang,
Peking University, China

***Correspondence:**
Jing Cai
jing.cai@polyu.edu.hk

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: 25 February 2022
Accepted: 02 June 2022
Published: 01 July 2022

Citation:
Ren G, Li B, Lam S-k, Xiao H, Huang Y-H, Cheung AL-Y, Lu Y, Mao R, Ge H, Kong F-M, Ho W-Y and Cai J (2022) A Transfer Learning Framework for Deep Learning-Based CT-to-Perfusion Mapping on Lung Cancer Patients. Front. Oncol. 12:883516. doi: 10.3389/fonc.2022.883516

Purpose: Deep learning model has shown the feasibility of providing spatial lung perfusion information based on CT images. However, the performance of this method on lung cancer patients is yet to be investigated. This study aims to develop a transfer learning framework to evaluate the deep learning based CT-to-perfusion mapping method specifically on lung cancer patients.

Methods: SPECT/CT perfusion scans of 33 lung cancer patients and 137 non-cancer patients were retrospectively collected from two hospitals. To adapt the deep learning model on lung cancer patients, a transfer learning framework was developed to utilize the features learned from the non-cancer patients. These images were processed to extract features from three-dimensional CT images and synthesize the corresponding CT-based perfusion images. A pre-trained model was first developed using a dataset of patients with lung diseases other than lung cancer, and subsequently fine-tuned specifically on lung cancer patients under three-fold cross-validation. A multi-level evaluation was performed between the CT-based perfusion images and ground-truth SPECT perfusion images in aspects of voxel-wise correlation using Spearman's correlation coefficient (R), function-wise similarity using Dice Similarity Coefficient (DSC), and lobe-wise agreement using mean perfusion value for each lobe of the lungs.

Results: The fine-tuned model yielded a high voxel-wise correlation (0.8142 ± 0.0669) and outperformed the pre-trained model by approximately 8%. Evaluation of function-wise similarity indicated an average DSC value of 0.8112 ± 0.0484 (range: 0.6460-0.8984) for high-functional lungs and 0.8137 ± 0.0414 (range: 0.6743-0.8902) for low-functional lungs. Among the 33 lung cancer patients, high DSC values of greater than 0.7 were achieved for high functional volumes in 32 patients and low functional volumes in all

patients. The correlations of the mean perfusion value on the left upper lobe, left lower lobe, right upper lobe, right middle lobe, and right lower lobe were 0.7314, 0.7134, 0.5108, 0.4765, and 0.7618, respectively.

Conclusion: For lung cancer patients, the CT-based perfusion images synthesized by the transfer learning framework indicated a strong voxel-wise correlation and function-wise similarity with the SPECT perfusion images. This suggests the great potential of the deep learning method in providing regional-based functional information for functional lung avoidance radiation therapy.

Keywords: perfusion imaging, functional lung avoidance radiation therapy, deep learning, CT-to-perfusion translation, lung cancer, radiation therapy

INTRODUCTION

Functional Lung Avoidance Radiation Therapy (FLART) is an emerging technique that selectively avoids excessive dose delivery to the high functional lung volumes, while favoring dose deposition in the low functional lung volumes based on the information obtained from pulmonary function imaging (1–3). Currently, there are three ongoing clinical trials in the United States (NCT02528942, NCT02308709, and NCT02843568) investigating the clinical efficacy of FLART. In addition, Matuszak et al. found that the mean dose in the high functional region decreased from 12.6 ± 4.9 Gy to 9.9 ± 4.4 Gy (4). Waxweiler et al. observed an average decrease of the mean dose to the functional lung by 2.8 Gy in FLART planning (5). Yamamoto et al. reported a 5.0% decrease in the dose of the FLART planning (6). This approach holds great promise to increase post-treatment perfusion in low-dose regions and minimize radiation-induced lung injury (7, 8).

The implementation of FLART relies on lung functional images to provide information on regional lung function for guiding the treatment planning process. A number of methods have been proposed for lung function imaging, which can be broadly divided into two categories: contrast agent-based imaging methods and deformable image registration (DIR) based methods. Contrast agent-based imaging reveals lung function by using different imaging contrast agents, examples including single-photon emission computed tomography (SPECT) with Tc-99m-labelled macro aggregated albumin (MAA) (4, 7), positron emission tomography (PET) with Ga-68 (9), magnetic resonance imaging (MRI) with hyperpolarized gas (Helium-3 or Xenon-129) (10, 11), and a variety of contrast-enhanced MRI (12–14) or CT (15). On the other hand, DIR-based methods compute surrogates of regional pulmonary function from lung four-dimensional computed tomography (4D-CT) images or breath-hold CT (BHCT) image pairs through DIR algorithms and sophisticated image mathematical metrics (16–19).

Nevertheless, these current methods suffer from numerous drawbacks, impeding the widespread application of FLART in the clinic. For example, SPECT function imaging commonly offers a limited spatial resolution and incurs focal radio aerosol clumping artifacts. PET imaging requires a long imaging time and incurs inevitable image noise. Besides, both SPECT and PET imaging requires contrast agents that may release additional ionizing radiation to patients. Hyperpolarized gas MRI (HP-MRI) is free of ionizing radiation; however, it requires precious noble gases and additional equipment for hyperpolarization. On the other hand, DIR-based function imaging is error-prone due to the deficiencies of the current DIR algorithms. These limitations have restricted their widespread application in clinic (17). In general, these function imaging modalities are of low accessibility in the radiation oncology department for the patients (17, 20).

Confronted with these limitations, the deep learning-based CT-to-perfusion mapping (CTPM) method was proposed in our previous study (21). This method synthesizes lung functional images based on the texture information provided by anatomic CT images. We previously demonstrated that the CT-based perfusion images generated by the CTPM method achieved a moderate-to-high approximation as compared with SPECT perfusion images in patients with different lung diseases (22, 23). Perfusion SPECT is one of the primarily diagnostic tools for pulmonary embolism, but not for lung cancer patients. In this study, we collected a cohort of lung cancer patients with 3D SPECT perfusion images. In the hope of paving the way towards FLART application in the future, we aimed to develop a transfer learning framework to evaluate the performance of the CTPM method specifically in lung cancer patients by using multi-level evaluations (voxel-wise correlation, function-wise similarity, and lobe-wise agreement).

METHOD

Datasets and Image Acquisition

In this study, two datasets of SPECT/CT perfusion images were retrospectively collected from two hospitals. The first dataset (n=33, lung cancer dataset) was built using SPECT/CT images collected from Hong Kong Queen Mary Hospital (n=14, Institution A) and Henan Cancer Hospital (n=19, Institution B). All patients in this dataset were diagnosed with lung cancer in clinical diagnosis and the SPECT/CT scans were performed before treatment. The second dataset (n=137, non-cancer dataset) was collected from institution A, which includes

different types of lung diseases except lung cancer (such as pulmonary hypertension, pulmonary embolism, etc.). The patient characteristics of the two datasets are listed in **Table 1**. This study was approved by the Institutional Review Boards (IRB) of The University of Hong Kong/Hospital Authority Hong Kong West Cluster and the IRB of Henan Cancer Hospital.

SPECT/CT scans collected from Institution A were acquired with 111 MBq technetium-99m (^{99m}Tc) MAA before imaging. Patients were immobilized in the supine position with normal resting breathing during image acquisition. The 3D SPECT/CT scans were acquired in 360 degrees to cover the whole lung volume under GE Discovery 670 SPECT/CT scanner (GE Healthcare, Milwaukee, WI) with a frame rate of 30 seconds per frame and a total frame number of 60. Each acquired CT image was reconstructed into 512×512 slices with 0.977×0.977 mm^2 in-plane pixel spacing and 1.25 mm slice thickness, while each acquired SPECT image was reconstructed into a 128×128×128 matrix with 4.42×4.42×4.42 mm^3 voxel size.

Patients from Institution B were scanned using a dual-head SPECT-CT scanner (Philips, Eindhoven, The Netherlands). A total of 185 MBq ^{99m} Tc-MAA was injected through the brachium vein of the patient. Cross−sectional images were acquired with one frame for 60 seconds per frame. Each acquired CT image was reconstructed into a 512×512 matrix with 0.977×0.977 mm^2 pixel spacing and 3.75 mm slice thickness, and each SPECT image was reconstructed into a 128×128×60 matrix with 2.76×2.76×1 mm^3 voxel size.

Each SPECT image was registered to the corresponding CT image. To ensure the consistency of the acquired data between different institutions, all the acquired SPECT/CT images were reconstructed into a voxel size of 1×1×1 mm^3. All downstream evaluations were performed under this resolution.

Transfer Learning Framework for the Generation of CT-Based Perfusion

To adapt the deep learning model in the lung cancer cohort, a transfer learning framework was developed to utilize the features learned from the non-cancer patients (**Figure 1A**). Specifically, the convolutional neural network (CNN) of the CTPM method was firstly trained on the non-cancer dataset to learn the fundamental mapping relation. Then the learned parameters from the non-cancer dataset were used as the initial parameters for further tunning on the lung cancer dataset. During the transfer training, three-fold cross-validation was used to make full use of the dataset. In each split, 2/3 of the lung cancer patients were used for training, with the remaining patients for testing. The outputs from three splits were combined for the subsequent evaluations.

TABLE 1 | Patient characteristics of the two datasets.

		Lung cancer dataset		Non-cancer dataset	
		Number	Percent	Number	Percent
Sex	Male	18	54.5%	51	37.2%
	Female	15	45.5%	86	62.8%
Age	Mean ± SD	64 ± 7.6		65 ± 15.7	

The preprocessing procedures and CNN model were proposed in our previous study (22) and illustrated in **Figure 1B**. Briefly, the lung parenchyma region was segmented by using a pre-trained U-Net model (R231) (24), which was trained on multifarious lung CT scans. Then the left and right lungs were separated and cropped to the border of the parenchyma, followed by resampling to 128 × 64 × 64-sized matrices. The resampled CT and SPECT images were standardized using CT enhancement and SPECT standardization, respectively. In this process, the tumor regions and vessels were removed from all images and following evaluations by using thresholds of -1000 to -300 Hounsfield unit (HU). In this study, the three-dimensional attention residual neural network (ARNN) was utilized to extract features from three-dimensional CT images and synthesize the corresponding CT-based perfusion images. It was trained with the processed CT images as input and processed SPECT perfusion as the target. In the application, the trained ARNN translated the processed CT images and synthesized the corresponding perfusion images.

To ensure that the synthesized lung images were in the same shape and coordinate with the original lung CT, the output images of left/right lungs were combined and recovered to the same geometry with the pre-processed CT images, which is referred to as CT-based perfusion images. The signal intensity of a lung SPECT image is strongly affected by the patient's condition, such as the respiratory capacity, frequency, diseases, etc. To ensure the perfusion value is comparable between patients, the SPECT perfusion was normalized to the 75th percentile value for each image, as this is close to the perfusion value of normal-functioning lung tissue (25). The CT-based perfusion images were then compared with the SPECT perfusion images *via* multi-level evaluations, including voxel-wise correlation, function-wise similarity, and lobe-wise agreement.

Quantitative Evaluation of CT-BASED PERFUSION WITH SPECT Perfusion

Voxel-Wise Correlation

To evaluate the performance in terms of voxel-wise intensity correlation, the Spearman's correlation coefficient (R) was computed between the CT-based perfusion images and the corresponding SPECT perfusion images. R is defined by the equation (1):

$$R = \frac{\sum_{i=1}^{N}[(y_i - \bar{y}) \cdot (p_i - \bar{p})]}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{N}(p_i - \bar{p})^2}}, \quad (1)$$

where $\bar{p}$,$\bar{y}$ p_i, and y_i denote the average value and value at voxel i for the predicted and ground-truth perfusions, respectively. N denotes the total number of non-zero voxels.

Function-Wise Similarity

To evaluate the accuracy of high functional lung avoidance as well as low functional lung allowance in inverse planning, we

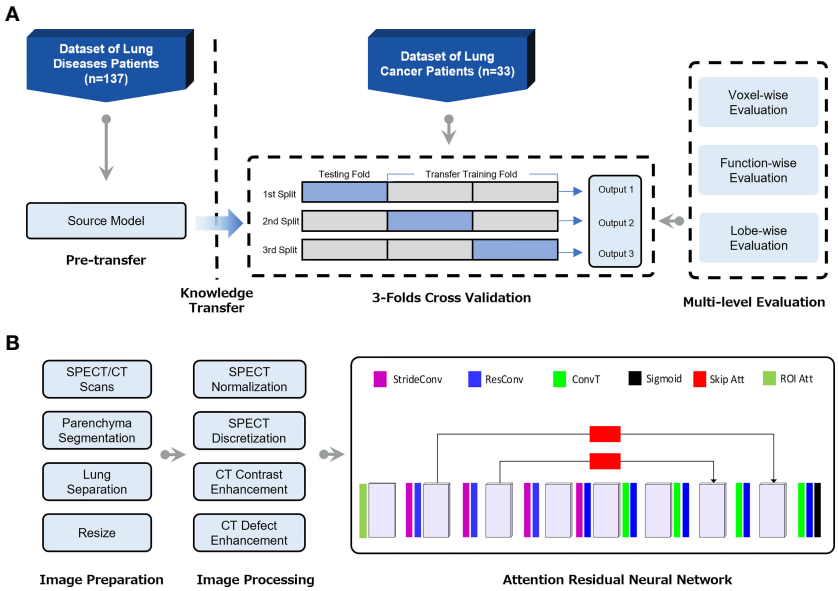


FIGURE 1 | **(A)** Flow chart of the transfer learning framework for generation of CT-based perfusion images on lung cancer patients. **(B)** The pipeline of deep learning-based CT-to-perfusion mapping (CTPM) method.

defined the low/high functional lung volumes in both the SPECT/CT-based perfusion images for the volume overlap test. Since each perfusion image has a maximum value of 1, the threshold value of 0.66 was used to separate the low and high functional lung volumes, which were suggested in previous lung ventilation study and FLART planning (26, 27). The Dice Similarity Coefficient (DSC) was then computed to determine the similarity of the low/high functional lung volumes. DSC is defined as follows.

$$DSC = \frac{2 * |p \cap y|}{|p| + |y|}, \tag{2}$$

where p is the low- and high-functional volume in the predicted perfusion images, and y is the corresponding volume in the ground-truth SPECT perfusion images. The overall concordance is inferred as the mean *DSC* value of all the testing cases.

Lobe-Wise Agreement

To evaluate the overlap of different lung regions, the perfusion images were segmented based on the region of lobes of the lung for further analysis. Specifically, the left upper lobe (LUL) and left lower lobe (LLL) were segmented from the left lung; the right upper lobe (RUL), right middle lobe (RML) and right lower lobe (RLL) were segmented from the right lung. The lobe segmentations were performed on the CT images using the Chest Imaging Platform in open-source software 3D Slicer (Surgical Planning Laboratory, Brigham and Women's Hospital, Boston, Mass) (28, 29). To compare the perfusion in each lobe region, the mean perfusion value in each lobe was calculated for both SPECT/CT-based perfusion images.

Convolution Neural Network Implementation

The CT and SPECT images were prepared and processed prior to model training and testing. The initialization of the convolutional layers was configurated using the Kaiming Uniform method (30). We implemented our network using the Pytorch 1.1 framework and coded the processing procedures in python. All the experiments were performed using a workstation with Intel Core i7-8700 @ 3.2GHz CPU, NVIDIA GTX 2080 TI GPU with 11GB memory, and 32 GB of RAM.

RESULTS

Figure 2A shows the result of voxel-wise correlation evaluation before and after transfer learning using the CTPM method. The CT-based perfusion images in the three splits after transfer learning achieved a high correlation value (R) of 0.8263 ± 0.0767, 0.8133 ± 0.0727, and 0.8032 ± 0.0537, respectively, with an average correlation value of 0.8142 ± 0.0669 for all three splits. Compared with the testing results before transfer learning (0.7554 ± 0.0875), there was a significant improvement of the average of all splits (7.78%, p = 0.0047) in the performance after fine-tuning of the model.

Figure 2B shows the function-wise similarity evaluation of low-functional volume (LFV) and high-functional volume (HFV) in RT treatment planning between SPECT perfusion images and CT-based perfusion images. The mean DSC for LFV and HFV were 0.8137 ± 0.0414 and 0.8112 ± 0.0484, respectively, suggesting a high similarity of both levels of

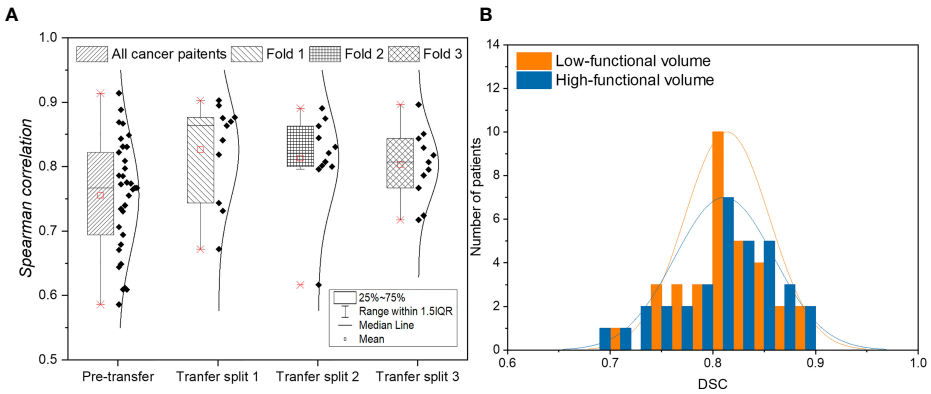


FIGURE 2 | **(A)** Evaluation of voxel-wise correlation of lung cancer patients before and after applying transfer learning. **(B)** Histogram of function-wise similarity evaluation in terms of DSC for low/high functional volumes. **(A)** shows the result of voxel-wise correlation evaluation before and after transfer learning using the CTPM method. The CT-based perfusion images in the three splits after transfer learning achieved a high correlation value (R) of 0.8263 ± 0.0767, 0.8133 ± 0.0727, and 0.8032 ± 0.0537, respectively, with an average correlation value of 0.8142 ± 0.0669 for all three splits. Compared with the testing results before transfer learning (0.7554 ± 0.0875), there was a significant improvement of the average of all splits (7.78%, p = 0.0047) in the performance after fine-tuning of the model. **(B)** shows the function-wise similarity evaluation of low-functional volume (LFV) and high-functional volume (HFV) in RT treatment planning between SPECT perfusion images and CT-based perfusion images. The mean DSC for LFV and HFV were 0.8137 ± 0.0414 and 0.8112 ± 0.0484, respectively, suggesting a high similarity of both levels of functional volumes. Among the 33 lung cancer patients, a high DSC value of greater than 0.8 was achieved in 67% of patients for high-functional volume and 70% for low-functional volume; almost all the lung cancer patients (33 for low-functional volume, and 32 for high-functional volume) demonstrated a DSC value larger than 0.7.

functional volumes. Among the 33 lung cancer patients, a high DSC value of greater than 0.8 was achieved in 67% of patients for high-functional volume and 70% for low-functional volume; almost all the lung cancer patients (33 for low-functional volume, and 32 for high-functional volume) demonstrated a DSC value larger than 0.7.

Figure 3 shows results of voxel-wise correlation and function-wise similarity in representative lung cancer patients. For the high-performance case in the testing group, the low functional region on the right upper region was successfully predicted on the CT-based perfusion. For the low-performance case in the testing group, no apparent low functional region was observed on the synthesized and ground-truth images. For both cases, CT-based perfusion images showed similar low-functional/high-functional regions to their respective SPECT perfusion images.

Figures 4A, B show the scatter plots of the mean value of each lobe between SPECT perfusion images and CT-based perfusion images. The correlations of the mean perfusion value on LUL, LLL, RUL, RML, and RLL were 0.7314, 0.7134, 0.5108, 0.4765, and 0.7618, respectively. The regional accuracy of CT-based perfusion on the RUL and RML was lower than the performance on other lobes. For the histogram of mean perfusion function, the histogram of CT-based perfusion images was lower than SPECT perfusion images in the range of 0 to 0.35 and 0.6 to 1, while higher on the other side.

Figure 5 shows a representative case of lobe-wise comparison of SPECT perfusion image and CT-based perfusion image. The difference in the mean perfusion function on the LUL, LLL, RUL, RML, and RLL were 0.28, -0.03, 0.02, 0.03, and -0.04, respectively. The LUL had the lowest perfusion values on both perfusion images. For the LUL, the mean perfusion value of CT-based perfusion images was 0.44, while it was 0.15 in SPECT perfusion images. For other lung lobes, the differences between the synthesized/ground truth perfusion images were less than 10%.

DISCUSSION

This study was the first report on the evaluation of the deep learning-based CT-to-perfusion mapping method on lung cancer patients. The CT-based perfusion images were compared with ground-truth SPECT perfusion images with voxel-wise correlation, function-wise similarity, and lobe-wise agreement in 33 lung cancer patients. In our previous work, we developed and evaluated the deep learning based CTPM method in patients with various lung diseases (22). However, the performance of the CTPM method specifically on lung cancer patients is yet to be investigated due to the limited number of lung cancer patients in our previous dataset. In this study, we collected a total of 33 SPECT/CT scans of lung cancer patients from two different hospitals, and then developed the transfer learning framework to evaluate the performance of the CTPM method specifically for lung cancer patients, in the hope of paving the way towards FLART application in the future.

To increase the model generalizability, we first trained the CNN model on patients with various lung diseases other than lung cancer and directly adapted it to the lung cancer dataset. The CNN model achieved a voxel-wise correlation (R) of 0.7554 ± 0.0875 in lung cancer patients. Subsequently, we used a transfer learning strategy to adapt the pre-trained model to the lung cancer dataset. It was observed that the correlation was approximately 8% higher

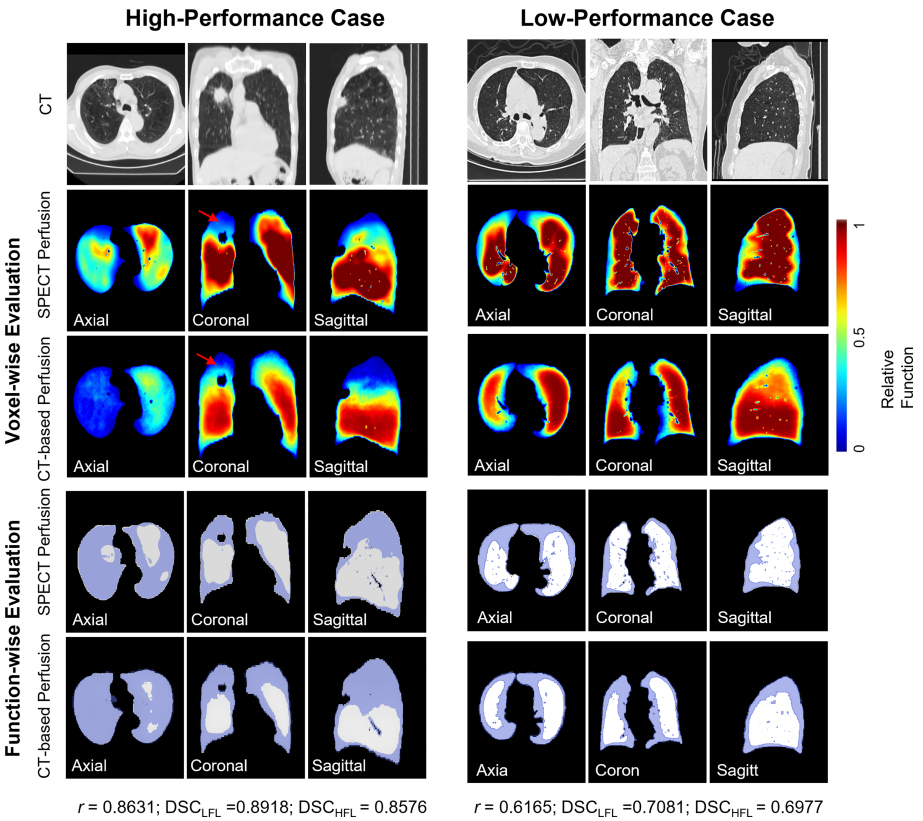


FIGURE 3 | Comparison of SPECT perfusion images and CT-based perfusion images in terms of voxel-wise correlation and function-wise similarity for two representative lung cancer cases. Each case is presented in axial, coronal, and sagittal views. In the voxel-wise evaluation, the red arrow indicates the main low functional regions. In function-wise evaluation, the blue contour indicates the low functional volume for treatment planning, while the white contour indicates the high functional volume. It shows results of voxel-wise correlation and function-wise similarity in representative lung cancer patients. For the high-performance case in the testing group, the low functional region on the right upper region was successfully predicted on the CT-based perfusion. For the low-performance case in the testing group, no apparent low functional region was observed on the synthesized and ground-truth images. For both cases, CT-based perfusion images showed similar low-functional/high-functional regions to their respective SPECT perfusion images.

($p = 0.0047$) to 0.8142 ± 0.0669 after applying the transfer learning (**Figure 2**). Transfer learning can improve the performance of the target domain by transferring the knowledge contained in different but related source domains (31). For the task of functional image synthesis, this improvement could be explained by the subject uniformity in the lung cancer dataset. Decreased lung function is caused by various mechanisms (32). For example, a complete defect can be induced by chronic obstructive pulmonary disease or other unrecoverable diseases. Lung cancer can also cause large vessel compression and alter the blood supply within the regional lung (33). In the lung cancer dataset, there are more low perfusion regions induced by tumor compression, which increases the uniformity of the low function region. For these functional defects induced by pulmonary vessel compression, tumor regression from RT may lead to regional lung reperfusion because of the relief of obstructions (34). For the future implementation of FLART, it is necessary to fine-tune the model on the lung cancer dataset to achieve better performance.

With regard to the function-wise similarity of the CT-based perfusion images, the low/high functional volumes showed almost the same level of high similarity (~0.8). In the qualitative comparison (**Figure 3**), we noticed two patterns of distributions of low functional volume (LFV): the first case had the LFV located at the corner of the lung (corner type); the second LFV was located at the peripheral region of the whole lung volume (peripheral type). The LFV of peripheral type could be attributed to its distance to the pulmonary arteries. Therefore, no significant dose-spare would be expected for this type in the FLART planning. This indicates that the benefit of generated CT-based perfusion for the surrounding type is limited. Based on this indication, a further step is needed to identify the distribution pattern of low functional regions prior to FLART implementation in the clinic.

For the lobe-wise agreement, the correlation coefficients of the mean perfusion value in the LUL (0.7314), LLL (0.7139), and RLL (0.7618) were significantly higher than those in the RUL (0.5108), RML (0.4765) regions of the lung. A possible explanation could be related to the fact that the horizontal fissure separating the RUL and RML has increased the perfusion complexity of this region. As compared with the

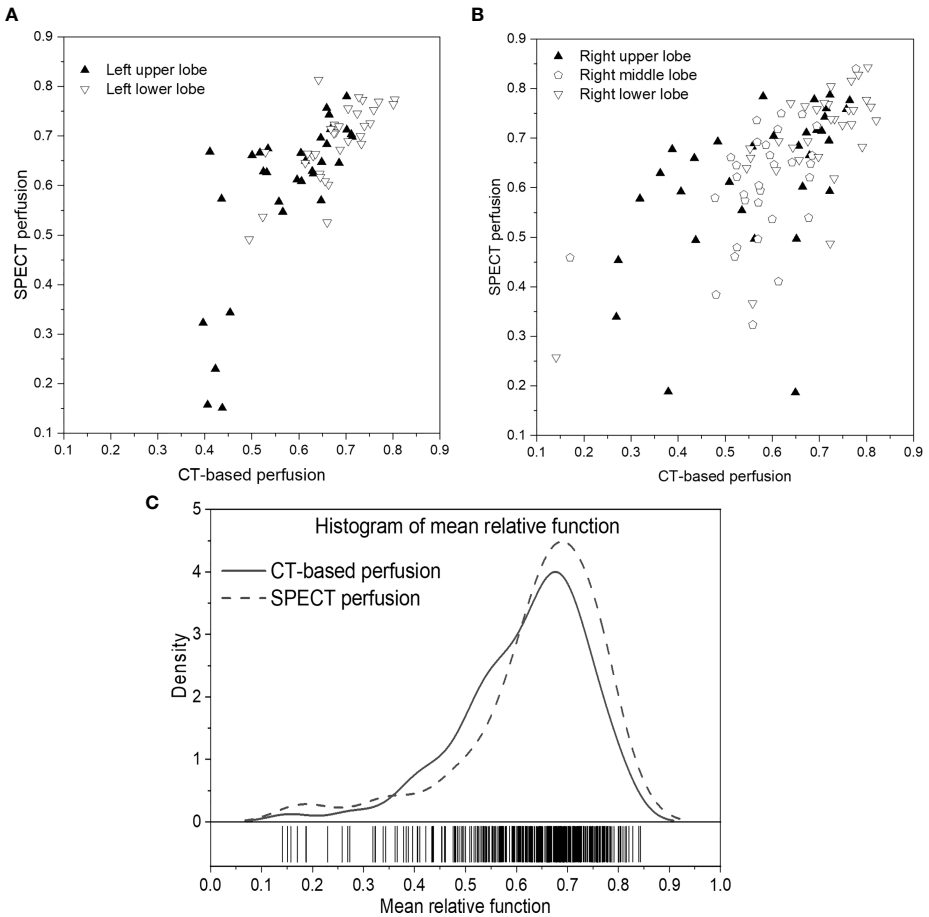


FIGURE 4 | Scatter plots of the mean relative perfusion value in each lobe between SPECT perfusion images and CT-based perfusion images of the left lung **(A)** and right lung **(B)**. **(C)** Histogram of the mean relative perfusion values of all the lobes. **(A, B)** show the scatter plots of the mean value of each lobe between SPECT perfusion images and CT-based perfusion images. The correlations of the mean perfusion value on LUL, LLL, RUL, RML, and RLL were 0.7314, 0.7134, 0.5108, 0.4765, and 0.7618, respectively. The regional accuracy of CT-based perfusion on the RUL and RML was lower than the performance on other lobes. For the histogram of mean perfusion function, the histogram of CT-based perfusion images was lower than SPECT perfusion images in the range of 0 to 0.35 and 0.6 to 1, while higher on the other side.

right lung, the extra horizontal fissure makes the perfusion condition on the RUL/RML region more complex, leading to relatively large uncertainty in these regions. In the future, a vessel-based analysis will be needed to explore further the effects caused by vessel differences between the left/right lungs and increase the prediction accuracy in these regions.

Apart from this, we also observed some cases with mismatched defect regions. As shown in **Figure 6**, the representative case has a correlation value of 0.6601 and 0.7859 for the right and left lungs, respectively. Most of the low perfusion region on the right lung (red arrow) was predicted as relatively high functional regions on synthesized CT-based perfusion. To a degree, this could be partly ascribed to the observed variations of CT-to-SPECT perfusion relationship between imaging views. In **Figure 6**, for instance, the representative case presents a consistent location between the low intensity regions (<-900 HU) within the right lung on the coronal view of the CT image (as indicated by the red regions in the first row of **Figure 6**) and the corresponding low perfusion regions on the ground-truth SPECT perfusion image (as indicated by the blue-shaded regions in the second row of **Figure 6**); while this consistency diminishes in some regions of the sagittal and axial views (as indicated by the white arrows in **Figure 6**). In this model, the regional inconsistency might have impeded accurate prediction of lung perfusion information from CT to SPECT images, and led to a "trade-off" predicting strategy of the deep learning model. This trade-off can also be observed in the histogram distribution of the mean perfusion values of all the lobes: the CTPM method tends to yield lower predicted values than the ground-truth in the perfusion value range from 0 to 0.35 and 0.6 to 1, while higher in the range from 0.35 to 0.6. When encountering uncertainties, the ARNN model trended to output median values to minimize the difference. This mismatch may degrade the accuracy of FLART treatment planning. To further

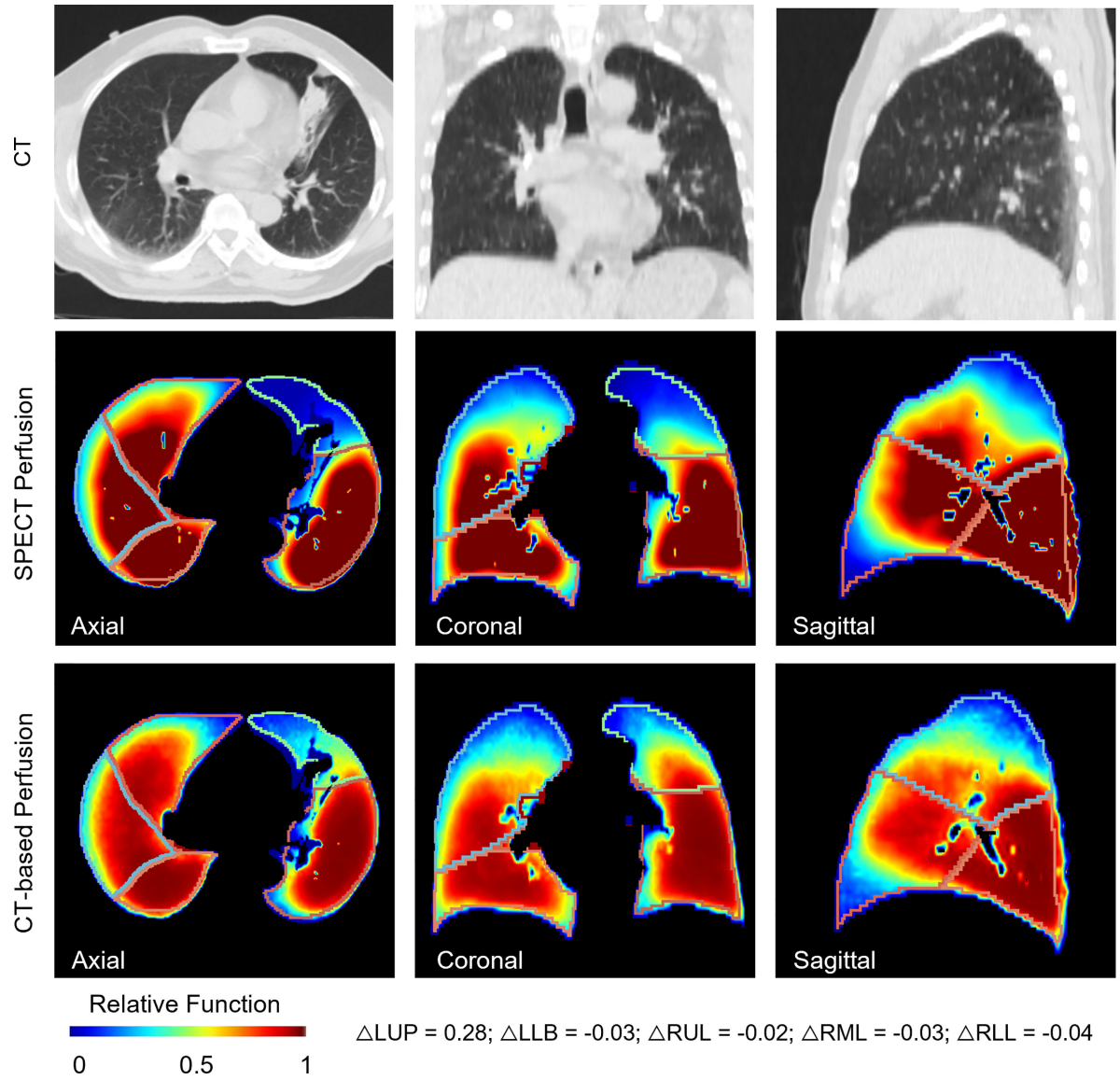


FIGURE 5 | A representative case for the lobe-wise agreement between the CT-based perfusion images and the ground-truth SPECT perfusion images. It shows a representative case of lobe-wise comparison of SPECT perfusion image and CT-based perfusion image. The difference in the mean perfusion function on the LUL, LLL, RUL, RML, and RLL were 0.28, -0.03, 0.02, 0.03, and -0.04, respectively. The LUL had the lowest perfusion values on both perfusion images. For the LUL, the mean perfusion value of CT-based perfusion images was 0.44, while it was 0.15 in SPECT perfusion images. For other lung lobes, the differences between the synthesized/ground truth perfusion images were less than 10%.

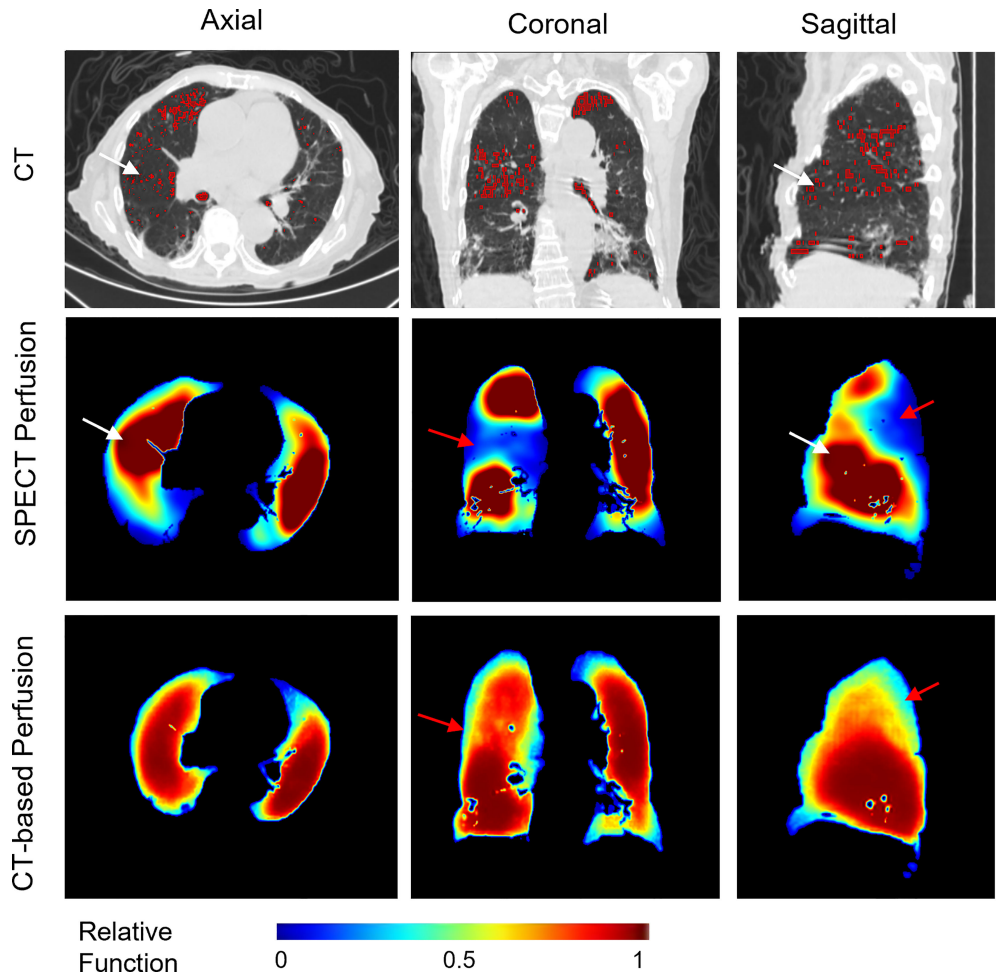


FIGURE 6 | A representative case with relatively low perfusion prediction on CT-based perfusion images. The red contours in the CT images indicate the low intensity regions (<-900 HU). The red arrows indicate the main mismatched low functional regions. The white arrows indicate the inconsistent regions between the low intensity regions of the CT images and the SPECT perfusion images.

improve the model performance on these uncertainty regions, The texture information of these mismatched regions should be further investigated (35).

In this study, we also compared the two groups of patients collected from two medical institutions. The performance on patients from institution A was significantly higher than that on patients from institution B in terms of correlation (10% higher, $p = 0.0016$), DSC of high functional volume (5% higher, $p = 0.0033$) and low functional volume (5% higher, $p = 0.0027$). These discrepancies could be explained by the different tumor sizes of these two groups of patients. There is a significant difference between the diameter of the tumors from the two institutions ($p = 0.0006$), with average diameter sizes of 40 ± 19.5 mm and 15 ± 15.6 mm, respectively. All the lung cancer patients from cohort A had tumor sizes larger than 20 mm, while only 6 of 14 patients from cohort B had comparable tumor sizes. The large tumor volume may have changed the blood supply and generated more significant lung functional volumes. In the future application of FLART, the effects of the tumor size should also be considered for lung function prediction.

There are also several limitations to this study. First, due to the limitation of GPU memory and the size of the training dataset, each lung is divided into left and right parts before inputting to the CNN network for model development. This may limit the cross-lung quantitative comparisons between left and right lungs from the same patient. The neural network still needs optimization in the coming study. Second, the performance of the CT-based perfusion images in treatment planning is still unknown. As such, dosimetry evaluation is still needed for potential dosimetry benefits of the CTPM method (36).

CONCLUSION

In this study, we, for the first time, quantitatively developed a transfer learning framework to evaluate the deep learning based

CT-to-perfusion mapping method specifically on 33 lung cancer patients at multiple levels, and achieved high correlations between the CT-based perfusion images and the ground-truth SPECT perfusion images. These findings suggested the use of CT-based perfusion images for high functional lung avoidance as well as low functional lung allowance in RT inverse planning, holding great promise in providing regional-based functional information for FLART in the future.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The patient images were collected from hospitals retrospectively. Requests to access these datasets should be directed to gary-ge.ren@polyu.edu.hk.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Institutional Review Boards (IRB) of The University of Hong Kong/Hospital Authority Hong Kong West Cluster and the IRB of Henan Cancer Hospital. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JC and GR conceived of the presented idea. BL, HX, and YH verified the analytical methods. S-KL helped to revise the manuscript. AC, YL, RM, HG, F-MK, and WH helped to collect the data. All authors contributed to the article and approved the submitted version.

FUNDING

This research was partly supported by research grants of General Research Fund (GRF 15103520), the University Grants Committee, and Health and Medical Research Fund (HMRF 07183266), the Food and Health Bureau, The Government of the Hong Kong Special Administrative Region.

REFERENCES

1. Hoover DA, Capaldi DP, Sheikh K, Palma DA, Rodrigues GB, Dar AR, et al. Functional Lung Avoidance for Individualized Radiotherapy (FLAIR): Study Protocol for a Randomized, Double-Blind Clinical Trial. *BMC Cancer* (2014) 14:934. doi: 10.1186/1471-2407-14-934
2. Lee E, Zeng J, Miyaoka RS, Saini J, Kinahan PE, Sandison GA, et al. Functional Lung Avoidance and Response-Adaptive Escalation (FLARE) RT: Multimodality Plan Dosimetry of a Precision Radiation Oncology Strategy. *Med Phys* (2017) 44(7):3418–29. doi: 10.1002/mp.12308
3. Baisden JM, Romney DA, Reish AG, Cai J, Sheng K, Jones DR, et al. Dose as a Function of Lung Volume and Planned Treatment Volume in Helical Tomotherapy Intensity-Modulated Radiation Therapy-Based Stereotactic Body Radiation Therapy for Small Lung Tumors. *Int J Radiat Oncol Biol Phys* (2007) 68(4):1229–37. doi: 10.1016/j.ijrobp.2007.03.024
4. Matuszak MM, Matrosic C, Jarema D, McShan DL, Stenmark MH, Owen D, et al. Priority-Driven Plan Optimization in Locally Advanced Lung Patients Based on Perfusion SPECT Imaging. *Adv Radiat Oncol* (2016) 1(4):281–9. doi: 10.1016/j.adro.2016.10.007
5. Waxweiler T, Schubert L, Diot Q, Faught A, Stuhr K, Castillo R, et al. A Complete 4 DCT-Ventilation Functional Avoidance Virtual Trial: Developing Strategies for Prospective Clinical Trials. *J Appl Clin Med Phys* (2017) 18 (3):144–52. doi: 10.1002/acm2.12086
6. Yamamoto T, Kabus S, Bal M, Bzdusek K, Keall PJ, Wright C, et al. Changes in Regional Ventilation During Treatment and Dosimetric Advantages of CT Ventilation Image Guided Radiation Therapy for Locally Advanced Lung Cancer. *Int J Radiat Oncol Biol Phys* (2018) 102(4):1366–73. doi: 10.1016/j.ijrobp.2018.04.063
7. Eslick EM, Stevens MJ, Bailey DL. SPECT V/Q in Lung Cancer Radiotherapy Planning. *Semin Nucl Med* (2019) 49(1):31–6. doi: 10.1053/j.semnuclmed.2018.10.009
8. Thomas HMT, Zeng J, Lee HJ, Sasidharan BK, Kinahan PE, Miyaoka RS, et al. Comparison of Regional Lung Perfusion Response on Longitudinal MAA SPECT/CT in Lung Cancer Patients Treated With and Without Functional Tissue-Avoidance Radiation Therapy. *Br J Radiol* (2019) 92(1103). doi: 10.1259/bjr.20190174
9. Le Roux PY, Hicks RJ, Siva S, Hofman MS. PET/CT Lung Ventilation and Perfusion Scanning Using Galligas and Gallium-68-MAA. *Semin Nucl Med* (2019) 49(1):71–81. doi: 10.1053/j.semnuclmed.2018.10.013
10. Mathew L, Wheatley A, Castillo R, Castillo E, Rodrigues G, Guerrero T, et al. Hyperpolarized (3)He Magnetic Resonance Imaging: Comparison With Four-Dimensional X-Ray Computed Tomography Imaging in Lung Cancer. *Acad Radiol* (2012) 19(12):1546–53. doi: 10.1016/j.acra.2012.08.007
11. Tahir BA, Hughes PJC, Robinson SD, Marshall H, Stewart NJ, Norquay G, et al. Spatial Comparison of CT-Based Surrogates of Lung Ventilation With Hyperpolarized Helium-3 and Xenon-129 Gas MRI in Patients Undergoing Radiation Therapy. *Int J Radiat Oncol Biol Phys* (2018) 102(4):1276–86. doi: 10.1016/j.ijrobp.2018.04.077
12. Cai J, Sheng K, Benedict SH, Read PW, Larner JM, Mugler JP 3rd, et al. Dynamic MRI of Grid-Tagged Hyperpolarized Helium-3 for the Assessment of Lung Motion During Breathing. *Int J Radiat Oncol Biol Phys* (2009). 75 (1):276–84. doi: 10.1016/j.ijrobp.2009.03.051
13. Cai J, Altes TA, Miller GW, Sheng K, Read PW, Mata JF, et al. MR Grid-Tagging Using Hyperpolarized Helium-3 for Regional Quantitative Assessment of Pulmonary Biomechanics and Ventilation. *Magn Reson Med* (2007) 58(2):373–80. doi: 10.1002/mrm.21288
14. Eichinger M, Puderbach M, Fink C, Gahr J, Ley S, Plathow C, et al. Contrast-Enhanced 3D MRI of Lung Perfusion in Children With Cystic Fibrosis–Initial Results. *Eur Radiol* (2006) 16(10):2147–52. doi: 10.1007/s00330-006-0257-7
15. Reinhardt JM, Ding K, Cao K, Christensen GE, Hoffman EA, Bodas SV. Registration-Based Estimates of Local Lung Tissue Expansion Compared to Xenon CT Measures of Specific Ventilation. *Med Imag Anal* (2008) 12 (6):752–63. doi: 10.1016/j.media.2008.03.007
16. Kipritidis J, Hofman MS, Siva S, Callahan J, Le Roux PY, Woodruff HC, et al. Estimating Lung Ventilation Directly From 4D CT Hounsfield Unit Values. *Med Phys* (2016) 43(1):33–43. doi: 10.1118/1.4937599
17. Kipritidis J, Tahir BA, Cazoulat G, Hofman MS, Siva S, Callahan J, et al. The VAMPIRE Challenge: A Multi-Institutional Validation Study of CT Ventilation Imaging. *Med Phys* (2019) 46(3):1198–217. doi: 10.1002/mp.13346
18. Yamamoto T, Kabus S, Klinder T, von Berg J, Lorenz C, Loo BW, et al. Four-Dimensional Computed Tomography Pulmonary Ventilation Images Vary

With Deformable Image Registration Algorithms and Metrics. *Med Phys* (2011) 38(3):1348–58. doi: 10.1118/1.3547719
19. Castillo E, Nair G, Turner-Lawrence D, Myziuk N, Emerson S, Al-Katib S, et al. Quantifying Pulmonary Perfusion From Noncontrast Computed Tomography. *Med Phys* (2021). doi: 10.1002/mp.14792
20. Kipritidis J, Siva S, Hofman MS, Callahan J, Hicks RJ, Keall PJ, et al. Validating and Improving CT Ventilation Imaging by Correlating With Ventilation 4D-PET/CT Using Ga-68-Labeled Nanoparticles. *Med Phys* (2014) 41(1). doi: 10.1118/1.4856055
21. Ren G, Ho WY, Qin J, Cai J. *Deriving Lung Perfusion Directly From CT Image Using Deep Convolutional Neural Network: A Preliminary Study*. Cham: Springer International Publishing (2019).
22. Ren G, Zhang J, Li T, Xiao HN, Yin Cheung AL, Ho WY, et al. Deep Learning-Based Computed Tomography Perfusion Mapping (DL-CTPM) for Pulmonary CT-To-Perfusion Translation. *Int J Radiat Oncol Biol Phys* (2021). doi: 10.1016/j.ijrobp.2021.02.032
23. Ren G, Lam S-K, Zhang J, Xiao H, Cheung AL-Y, Ho W-Y, et al. Investigation of a Novel Deep Learning-Based Computed Tomography Perfusion Mapping Framework for Functional Lung Avoidance Radiotherapy. *Front Oncol* (2021) 11(466). doi: 10.3389/fonc.2021.644703
24. Hofmanninger J, Prayer F, Pan J, Rohrich S, Prosch H, Langs GJ-P, et al. *Automatic Lung Segmentation in Routine Imaging is Primarily a Data Diversity Problem, Not a Methodology Problem*. (2020).
25. Weller A, Dunlop A, Oxer A, Gunapala R, Murray I, Gray MJ, et al. Spect Perfusion Imaging Versus CT for Predicting Radiation Injury to Normal Lung in Lung Cancer Patients. *Br J Radiol* (2019) 92(1101):20190184. doi: 10.1259/bjr.20190184
26. Rankine LJ, Wang Z, Driehuys B, Marks LB, Kelsey CR, Das SK. Correlation of Regional Lung Ventilation and Gas Transfer to Red Blood Cells: Implications for Functional-Avoidance Radiation Therapy Planning. *Int J Radiat Oncol Biol Phys* (2018) 101(5):1113–22. doi: 10.1016/j.ijrobp.2018.04.017
27. Yamamoto T, Kabus S, von Berg J, Lorenz C, Keall PJ. Impact of Four-Dimensional Computed Tomography Pulmonary Ventilation Imaging-Based Functional Avoidance for Lung Cancer Radiotherapy. *Int J Radiat Oncol Biol Phys* (2011) 79(1):279–88. doi: 10.1016/j.ijrobp.2010.02.008
28. Estepar RS, Ross JC, Harmouche R, Onieva J, Diaz AA, Washko GR. Chest Imaging Platform: An Open-Source Library And Workstation For Quantitative Chest Imaging. *Am J Respir Crit Care Med* (2015) 191.
29. Kikinis R, Pieper SD, Vosburgh KG. 3d Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support. In: FA Jolesz, editor. *Intraoperative Imaging and Image-Guided Therapy*. New York, NY: Springer New York (2014). p. 277–89.
30. He KM, Zhang XY, Ren SQ, Sun J. (2015). Delving Deep Into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, in: *2015 Ieee International Conference on Computer Vision (Iccv)*, (Washington, DC, United States:IEEE Computer Society) pp. 1026–34.
31. Shao L, Zhu F, Li X. Transfer Learning for Visual Categorization: A Survey. *IEEE Trans Neural Netw Learn Syst* (2015) 26(5):1019–34. doi: 10.1109/TNNLS.2014.2330900
32. Petersson J, Glenny RW. Gas Exchange and Ventilation-Perfusion Relationships in the Lung. *Eur Respir J* (2014) 44(4):1023–41. doi: 10.1183/09031936.00037014
33. Yuan ST, Frey KA, Gross MD, Hayman JA, Arenberg D, Cai XW, et al. Changes in Global Function and Regional Ventilation and Perfusion on SPECT During the Course of Radiotherapy in Patients With non-Small-Cell Lung Cancer. *Int J Radiat Oncol Biol Phys* (2012) 82(4):e631–8. doi: 10.1016/j.ijrobp.2011.07.044
34. Bevilacqua A, Gavelli G, Baiocco S, Barone D. CT Perfusion in Patients With Lung Cancer: Squamous Cell Carcinoma and Adenocarcinoma Show a Different Blood Flow. *BioMed Res Int* (2018) 2018. doi: 10.1155/2018/6942131
35. Jia X, Ren L, Cai J Clinical Implementation of AI Technologies Will Require Interpretable AI Models. *Medical Physics* (2020) 47(1):1–4. doi: 10.1002/mp.13891
36. Wang M, Zhang Q, Lam S, Cai J, Yang R. A Review on Application of Deep Learning Algorithms in External Beam Radiotherapy Automated Treatment Planning. *Front Oncol* (2020) 10:p 580919. doi: 10.3389/fonc.2020.580919

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 | Frontiers in Oncology

ORIGINAL RESEARCH
published: 13 July 2022
doi: 10.3389/fonc.2022.920393

Predicting Adverse Radiation Effects in Brain Tumors After Stereotactic Radiotherapy With Deep Learning and Handcrafted Radiomics

Simon A. Keek[1†], Manon Beuque[1†], Sergey Primakov[1], Henry C. Woodruff[1,2], Avishek Chatterjee[1], Janita E. van Timmeren[3], Martin Vallières[4,5], Lizza E. L. Hendriks[6], Johannes Kraft[3,7], Nicolaus Andratschke[3], Steve E. Braunstein[8], Olivier Morin[8†] and Philippe Lambin[1,2†]*

[1] The D-Lab, Department of Precision Medicine, GROW- School for Oncology and Reproduction, Maastricht University, Maastricht, Netherlands, [2] Department of Radiology and Nuclear Medicine, GROW – School for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, Netherlands, [3] Department of Radiation Oncology, University Hospital of Zurich, University of Zurich, Zurich, Switzerland, [4] Medical Physics Unit, Department of Oncology, Faculty of Medicine, McGill University, Montréal, QC, Canada, [5] Department of Computer Science, Université de Sherbrooke, Sherbrooke, QC, Canada, [6] Department of Pulmonary Diseases, GROW – School for Oncology and Reproduction, Maastricht University Medical Centre+, Maastricht, Netherlands, [7] Department of Radiation Oncology, University Hospital Würzburg, Würzburg, Germany, [8] Department of Radiation Oncology, University of California San Francisco, San Francisco, CA, United States

Introduction: There is a cumulative risk of 20–40% of developing brain metastases (BM) in solid cancers. Stereotactic radiotherapy (SRT) enables the application of high focal doses of radiation to a volume and is often used for BM treatment. However, SRT can cause adverse radiation effects (ARE), such as radiation necrosis, which sometimes cause irreversible damage to the brain. It is therefore of clinical interest to identify patients at a high risk of developing ARE. We hypothesized that models trained with radiomics features, deep learning (DL) features, and patient characteristics or their combination can predict ARE risk in patients with BM before SRT.

Methods: Gadolinium-enhanced T1-weighted MRIs and characteristics from patients treated with SRT for BM were collected for a training and testing cohort (N = 1,404) and a validation cohort (N = 237) from a separate institute. From each lesion in the training set, radiomics features were extracted and used to train an extreme gradient boosting (XGBoost) model. A DL model was trained on the same cohort to make a separate prediction and to extract the last layer of features. Different models using XGBoost were built using only radiomics features, DL features, and patient characteristics or a combination of them. Evaluation was performed using the area under the curve (AUC) of the receiver operating characteristic curve on the external dataset. Predictions for individual lesions and per patient developing ARE were investigated.

Results: The best-performing XGBoost model on a lesion level was trained on a combination of radiomics features and DL features (AUC of 0.71 and recall of 0.80). On a patient level, a combination of radiomics features, DL features, and patient

OPEN ACCESS

Edited by:
Ke Nie,
Rutgers, The State University of New Jersey, United States

Reviewed by:
Shouping Xu,
Chinese Academy of Medical Sciences and Peking Union Medical College, China
Jiahua Zhu,
Reading Hospital, United States

***Correspondence:**
Philippe Lambin
philippe.lambin@maastrichtuniversity.nl

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: *14 April 2022*
Accepted: *13 June 2022*
Published: *13 July 2022*

Citation:
Keek SA, Beuque M, Primakov S, Woodruff HC, Chatterjee A, van Timmeren JE, Vallières M, Hendriks LEL, Kraft J, Andratschke N, Braunstein SE, Morin O and Lambin P (2022) Predicting adverse radiation effects in brain tumors after stereotactic radiotherapy with deep learning and Handcrafted Radiomics. Front. Oncol. 12:920393. doi: 10.3389/fonc.2022.920393

characteristics obtained the best performance (AUC of 0.72 and recall of 0.84). The DL model achieved an AUC of 0.64 and recall of 0.85 per lesion and an AUC of 0.70 and recall of 0.60 per patient.

Conclusion: Machine learning models built on radiomics features and DL features extracted from BM combined with patient characteristics show potential to predict ARE at the patient and lesion levels. These models could be used in clinical decision making, informing patients on their risk of ARE and allowing physicians to opt for different therapies.

Keywords: brain metastases (BMs), radiation necrosis (RN), deep learning - artificial neural network, radiomics, MRI, adverse radiation effects

1 INTRODUCTION

Brain metastases (BM) are the most common intracranial malignancies, accounting for more than 50% of all brain tumors and occurring in 10 to over 40% of patients with solid malignancies (1–3). BM occur most often in patients with lung cancer, breast cancer, and melanoma, which have a cumulative risk ranging from 20 to 40% of developing BM (4–7). BM can be treated locally by surgery or radiotherapy or with systemic anticancer therapy. Treatment depends on several factors, such as patient performance status, number and volume of metastases, presence of extracranial metastases, symptoms, and presumed efficacy of available systemic therapy ["Systemic therapy for brain metastases", (8, 9). The radiotherapy of BM can be either stereotactic radiotherapy (SRT) or whole brain radiotherapy (WBRT), with SRT being the guideline-recommended treatment for a limited number of BM. As WBRT is associated with neurocognitive deterioration, SRT is increasingly used in multiple BM as well (10–12). SRT is delivered either in a single fraction, with stereotactic radiosurgery (SRS), or as fractionated stereotactic radiotherapy (FSRT) and results in a high dose within the target volume with a steep dose gradient to the surrounding healthy tissue (13).

Even though most of the healthy brain is spared from high doses of radiation, a major shortcoming of SRT is a chance of high toxicity in the immediate surrounding tissues, which may lead to adverse radiation effects (ARE) such as radiation necrosis (RN), subacute edema, structural changes in the white matter, and vascular lesions (14). ARE are a relatively late reaction to irradiation of healthy tissues where either reversible or irreversible injury has occurred (15). The risk of ARE after SRT and SRS is found to be similar and ranges from 5 to 10% at patient level (16–19) or approximately 3% at lesion level (15). Known predictors of ARE are tumor volume, isodose volume, and previous SRT to the same lesion (15). ARE of the tumor area and tumor progression (TP) as two different post-therapeutic events require different treatment strategies: while steroids are often indicated for the initial treatment of ARE, true progression or relapse requires repeated radiotherapy, surgery, or effective intracranial systemic therapy for tumor control. Being able to differentiate between ARE and TP is therefore of utmost clinical interest.

Unfortunately, the (neurological) symptoms of ARE and TP are usually indistinguishable. Furthermore, the appearances of ARE and TP are very difficult to discern through qualitative radiological imaging, requiring multiple successive magnetic resonance images (MRI), specialized MRI sequences such as perfusion-weighted or MR spectroscopy, and trained experts to evaluate the findings (19, 20). The clinical workflow is time- and labor-intensive, and while it is unfeasible to perform for every lesion, a definitive confirmation of the presence of ARE requires tissue acquisition (19).

SRT requires routine pretreatment MRI for accurate target volume delineation. This imaging provides a source of non-invasively acquired information about BM and brain phenotypes that could be investigated for their potential to determine before treatment which patient has a high risk of developing ARE. The early identification of these patients is an unmet clinical need which may help in clinical decision making by informing the patients of the risk of ARE, the early risk stratification of patients that may develop ARE, and the consideration of ARE risk mitigating strategies such as deferring radiotherapy for central nervous system-penetrant systemic therapy.

Advanced quantitative medical image analysis methods such as radiomics and deep learning (DL) extract large amounts of imaging features and associate these with biological and/or clinical outcomes using machine learning (ML) techniques (21–26). Thus, radiological images from routine imaging procedures could potentially be used to non-invasively

Abbreviations: ARE, adverse radiation effects; AUC, area under the curve; AUCPR, area under the precision recall-curve; BM, brain metastasis; CI, confidence interval; CLAHE, contrast-limited adaptive histogram equalization; CTD, connective tissue disorder; DICOM, digital imaging and communications in medicine; DL, deep learning; EBRT, external beam radiotherapy; ECM, extra-cranial metastases; FSRT, fractionated stereotactic radiotherapy; GLCM, gray-level co-occurrence matrix; GLDZM, gray-level distance-zone matrix; GLRLM, gray-level run length matrix; GLSZM, gray-level size-zone matrix; HU, Hounsfield unit; IBSI, Image Biomarker Standardization Initiative; KPS, Karnofsky performance score; ML, machine learning; NGLDM, neighborhood gray-level dependence matrix; NGTDM, neighborhood gray-tone difference matrix; NRRD, nearly raw raster data; RFE, recursive feature elimination; RN, radiation necrosis; ROC, receiver operating characteristic; ROI, region of interest; RS, radiosurgery; SD, standard deviation; SRS, stereotactic radiosurgery; SRT, stereotactic radiotherapy; TP, tumor progression; UCSF, University of California—San Francisco; USZ, University Hospital Zürich; WBRT, whole-brain radiotherapy; XGBoost, extreme gradient boosting.

quantify the lesion phenotype, providing clinically necessary information for patient management decisions. Several studies have indicated that MRI radiomics analysis is able to differentiate BM from glioblastoma (27, 28) to predict local recurrence (29, 30), to predict the origin of metastases (31, 32), and to predict overall survival (33, 34). DL has also shown potential in predicting treatment response on brain MRI (35). Moreover, DL and radiomics can have a complementary value, potentially establishing a more robust classifier (36).

We hypothesize that models trained with radiomics features, DL features, and patient characteristics or a combination thereof can predict the occurrence of ARE in patients with BM, both lesion specific and patient specific.

2 MATERIALS AND METHODS

2.1 Patient Characteristics

All data from patients with BM treated with SRT between 1997 and 2017 for which imaging, outcome data, and patient data were available were collected retrospectively from the University of California—San Francisco (UCSF) medical center's picture archiving and communication system. Available imaging data, outcome data, and patient data of all patients with BM treated with SRS/SRT between 2014 and 2019 at the University Hospital Zürich (USZ) were collected retrospectively. The data included clinical and biological information for both the patient and the lesion. The eligibility criteria included radical treatment for metastatic brain cancer using Gamma Knife SRS for the UCSF patients and SRS/FSRT for the USZ patients. The inclusion of patients was regardless of the number of BM, but pathohistological or imaging-based confirmation of ARE during the follow-up was required in addition to pathohistological confirmation of the primary tumor. For the USZ cohort, in case of imaging-based suspicion of RN, positron emission tomography imaging was additionally used to exclude TP. The effort obtained ethical approval for observational research using anonymized linked care data for supporting medical purposes that are in the interests of individuals and the wider public. UCSF Institutional Review Board (https://irb.ucsf.edu) and Cantonal Ethics Committee Zurich approval with waiver of informed consent was obtained.

The UCSF dataset was divided randomly into sub-cohorts for training (70%) and testing (30%) while maintaining the ratios of events to non-events equal in both groups. The USZ dataset was used as an independent external validation dataset, *i.e.*, it was entirely unseen by the models during the training and testing phases. The binary outcome used in training and validation was ARE per lesion, defined as either pathologically or imaging-based confirmation of RN occurring at any time after treatment. For both the USCF and USZ patients, ARE was confirmed by histopathology when treated with open surgery. In all other cases, ARE was confirmed either at routine re-staging 3 months after radiotherapy for asymptomatic patients or at the onset of new symptoms. When patients presented new symptoms, imaging was performed usually after awaiting the effects of cortisone administration. As the time of BM formation is unknown, the outcome was not defined as right-censored. As every lesion is able to independently develop ARE after treatment, every lesion was considered to be an independent sample. The probability of ARE occurring for any lesion within a patient as an outcome was also investigated, whereby each patient was treated as an independent sample instead.

2.2 MR Acquisition Parameters and Lesion Segmentation

All images were axial gadolinium-enhanced T1-weighted MRI acquired prior to the treatment of BM. All included lesions were three-dimensionally delineated for curative Gamma Knife SRS treatment purposes for the UCSF cohort and for curative SRS/FSRT purposes for the USZ cohort according to local protocols by an experienced radiation oncologist. **Figure 1** shows two T1-weighted gadolinium-enhanced MRI with lesions delineated for SRT purposes.

To perform segmentations of the brain and the ventricles on the entire dataset, an atlas-based segmentation strategy was chosen. To create the atlas in the MIM software package (MIM v. 6.9.4, MIM Software Inc., Cleveland, OH, USA), 50 randomly chosen MRI were manually segmented by an expert radiologist.

2.3 Pre-Processing of Brain MRI Data

Bias-field correction was performed in the MIM software package using the N4 algorithm, which required brain segmentations (37). A bias field is a low-frequency signal distributed over an MR image, which is caused by inhomogeneities in the magnetic field of the MRI scanner. This causes shifts of intensity value ranges across the image (38). The ventricle mask was subtracted from the brain mask to obtain a white- and gray-matter segmentation. This segmentation was used to determine and correct the bias field present in the image using the N4 algorithm (37) using the MIM software package.

Following the bias correction, all remaining pre-processing, feature extraction, model training, and evaluation were performed in Python (version 3.7). The different Python packages used during this study can be found in **Supplementary Table S1**. Pre-processing of MRI is essential for ML purposes, for reducing scanner dependence, and for ensuring reproducibility (39–41). As there is, to date, no consensus regarding the best way to pre-process MRI for our purposes, three different pre-processing workflows were applied and compared: "minimalist", standardization, and "harmonization". The descriptions of these pre-processing workflows can be found in the **Supplementary Materials** (Section 1 and in **Figure 2**).

2.3.1 Pre-processing for radiomics and feature extraction

Feature extraction was performed according to the Image Biomarker Standardization Initiative (IBSI) guidelines (42–44) on the three different sets of processed MRI scans using the BM segmentations. All images were resampled to uniform $1 \times 1 \times 1$-mm^3 voxels using the "sitkBSpline" interpolator to correct for

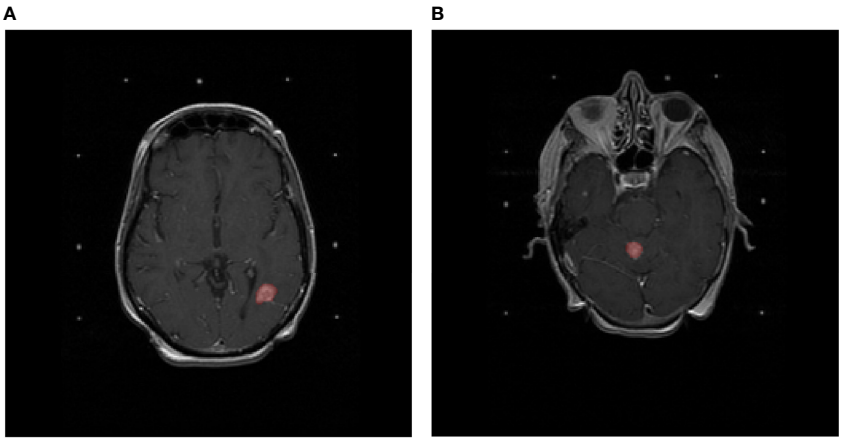


FIGURE 1 | T1-weighted gadolinium-enhanced MRIs of the brain. Delineated in red **(A)** is a lesion that developed adverse radiation effects after stereotactic radiotherapy and **(B)** a lesion that did not develop adverse radiation effects after stereotactic radiotherapy.

differences in pixel size and slice spacing. The choice for voxel dimensions was made based on majority ruling, as it was found that most patients had a pixel spacing of ~1 mm. To achieve isotropic voxels, the choice for resampling in the z-direction was also chosen as 1 mm. Pixel intensity values were resampled to a fixed number of 64 bins, as the number of gray levels was found to affect the interchangeability of MRI radiomics features, and a fixed bin number of 64 has been found recommended in previous studies (42–44).

A total of 106 IBSI features were extracted from each segmentation. The features were extracted from the BM segmentations of the pre-processed images and can be divided into first-order intensity, histogram statistics, shape, and texture features. A full list and a description of the features can be found in the PyRadiomics documentation ([Radiomic Features—PyRadiomics Documentation, (45)], and a description of the feature groups can be found in the **Supplementary Materials** (Section 2).

2.3.2. Pre-processing for deep learning

To inform the DL model on the location and extension of the lesions, lesion masks were used to highlight the ROI. A Gaussian smoothing filter was applied to the image, gradually decreasing the intensity values around the lesion from a factor of 1.0 to 0.2 to still include information of the voxels immediately around the lesion masks.

Otsu thresholding was performed to create a mask containing the brain and the skull. This mask was used to determine the largest three-dimensional bounding box containing the brain and the skull to crop the images. Anything outside this mask was defined as the image background, for which all pixel values were set at 0. For the "minimalist" and the "standardization" datasets, the intensities were resampled in a range between 0 and 255. Finally, the scans were rescaled at 256 × 256 × 64 with spline interpolation order 3. As an example, the steps of the pre-processing workflow for the "minimalist" normalization are illustrated in **Figure 3**.

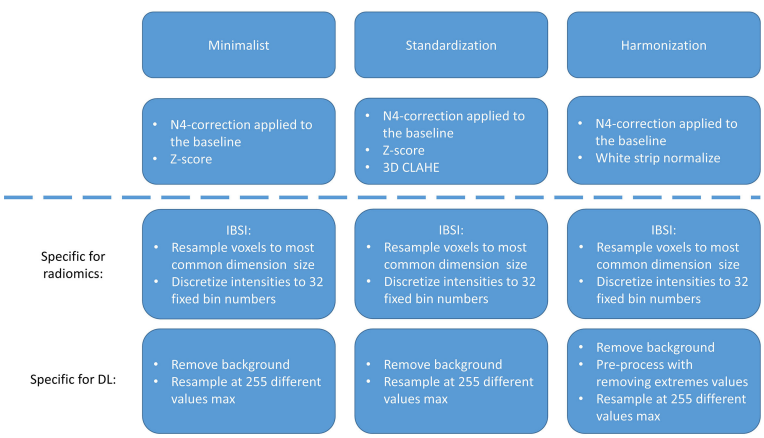


FIGURE 2 | Pre-processing strategies for the "minimalist", "standardization", and "harmonization" approaches.

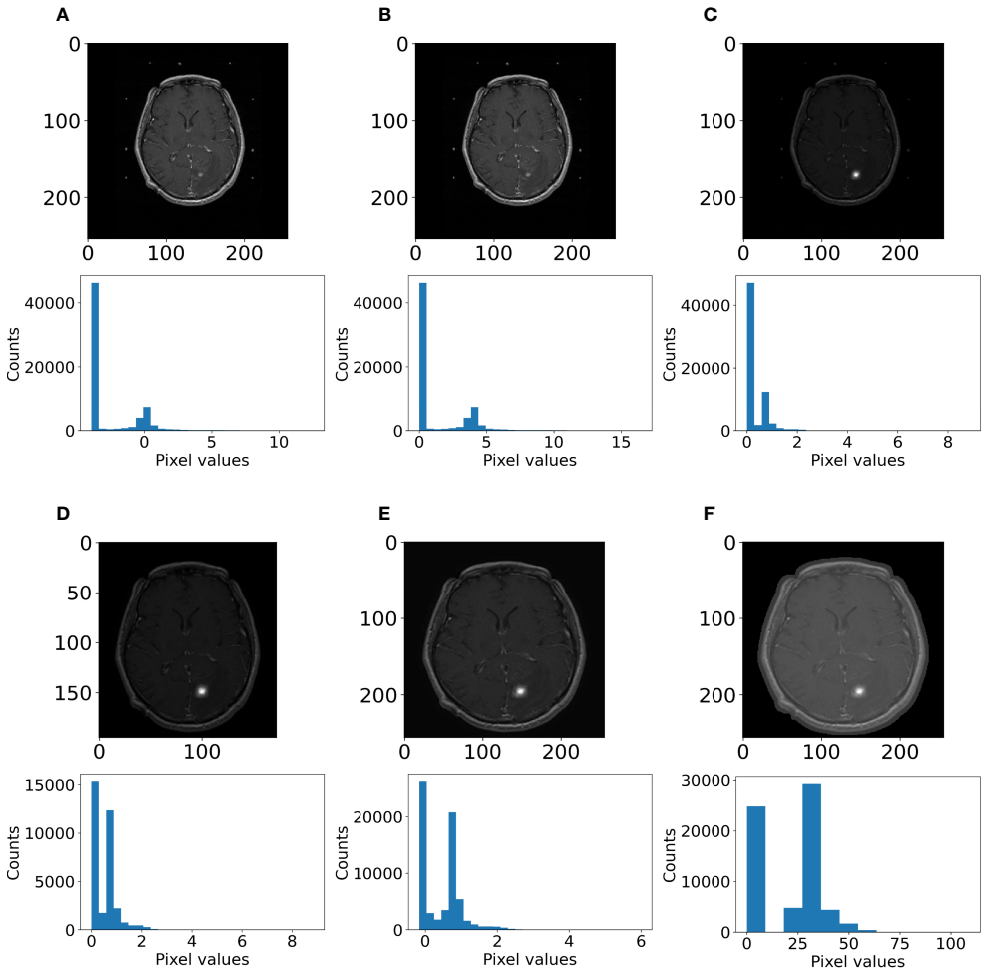


FIGURE 3 | Example of pre-processing strategy: deep learning on the "minimalist" approach. The different steps of preprocessing were **(A)** z-score normalization, **(B)** shift to positive values only, **(C)** pixel attenuations with Gaussian smoothing filtering, **(D)** cropping around the largest bounding box and background set to 0, **(E)** resizing at 256 × 256, and **(F)** rescaling the pixel value range to 0–255.

2.4 Machine Learning Models

The mean and SD of each feature over the entire training population were determined. These values were used to apply z-score normalization to the features of the training, testing, and external validation datasets (46). Next, features with low variance (<0.01) were determined and excluded from the dataset. Lastly, the correlation between features was determined using absolute pairwise Spearman rank correlation. As highly correlated features (>0.85) were assumed to contain overlapping information about the outcome, the feature with the highest mean absolute correlation with the rest of the features was excluded. Lastly, supervised feature selection was performed through recursive feature elimination (RFE). RFE uses a ML algorithm to build a multivariate model and determine predictive performance using the currently selected features. It recursively drops and adds features, determining the optimal number of features and the selection of most predictive features.

An extreme gradient boosting (XGBoost) model was used for RFE and ARE prediction. A description of the XGBoost architecture and the methodology to determine the optimal hyperparameters for the trained models can be found in the **supplementary materials** (Section 3).

2.5 Deep Learning Model

An Xception three-dimensional model was trained and tested on the same datasets as the handcrafted radiomics-based model. Xception is the extreme version of an Inception model (47), which uses depth-wise separable convolutions. The architecture can be found in **Supplementary Figure S1**. Adam optimization was used (48) with an initial learning rate of 10^{-5}, which updated the learning rate during training, and used for loss function binary cross-entropy. This model produced a score ranging from 0 to 1, indicating the estimated probability that a lesion develops ARE. The area under the curve (AUC) of the receiver operating characteristic

(ROC) was monitored on the test dataset. The ROC displays the discriminative performance of a model expressed through the sensitivity and specificity as the threshold for binary classification is shifted. The AUC of the ROC is a metric from 0 to 1, where 1 means that the model has perfect predictive performance and 0.5 is equivalent to guessing. To limit the imbalance of the outcomes to affect the model training, the model was only trained on lesions for those patients who had at least a single ARE and tested on the scans of the patients who had ARE in the test dataset. To combine DL and radiomics, the last fully connected layer consisting of 256 features obtained after training the model was extracted. These features were then used to train a ML model similarly to using radiomics features and used in models combining radiomics features and patient characteristics.

2.6 Clinical and Treatment-Related Feature Model

As the training and testing datasets contained patient characteristics not available in the external validation dataset, any feature not overlapping between these datasets was dropped. The list of the remaining features was as follows: primary tumor location, primary tumor histology, primary tumor controlled, extra-cranial metastases presence, patient age, patient sex, SRS to the same location, prior external beam radiotherapy (EBRT), prior radiosurgery (RS), neurological symptoms, headaches, seizures, hypertension, diabetes, connective tissue disorder, Karnofsky performance score (KPS) status, prescription dose, and isodose lines. For XGBoost to be able to handle categorical variables, one-hot encoding was performed on two categorical clinical features (primary tumor location and primary tumor histology).

Missing values were imputed using MissForest. MissForest is an imputation algorithm that uses RandomForest to train a model on the non-missing data for each feature with missing values to predict the missing values. In the first iteration, all values are set to the mean value present for each variable (*i.e.*, each column). Then, over multiple iterations, each data column with missing values will be predicted using all the data except for the rows containing the missing values in question. This process is repeated over several iterations.

2.7 Metrics Used for Data Analysis

The patient and tumor characteristics in the UCSF and USZ cohorts were assessed through a two-proportion z-test to test for significant differences in categorical variables between the cohorts or the unpaired two-sample t-test to test for significant differences in numerical variables. For the latter, the assumptions of the data having a normal distribution and possessing the same variance in both cohorts were tested through Shapiro–Wilk's test and f-test, respectively. The significance level was set at 5%.

To determine which method ensured best performance for the radiomics-based and DL models, models were trained on the three different pre-processed datasets, and the best AUC of the ROC on the testing set was used to determine the best pre-processing methods for ML and DL separately. The 95% confidence intervals (CI) displayed on the ROC curves were obtained using bootstrapping (n = 2,000). For the radiomics-based model, the results were reported on the full train dataset and the entire test dataset. For the DL model, the results were reported on the balanced train dataset (which served to train the different DL models) and the full test dataset.

Once the best models were selected, the models were validated on the external dataset. The predictive performance of each model was expressed through the ROC curve and its AUC on the training, testing, and external data. By determining an optimal threshold value using Youden's J statistic (49) based on the training dataset, a binary classification was performed on the external dataset. From this binary classification, the balanced accuracy, precision, recall, and F1-score were determined. The confusion matrices were also derived from the binary classification. To determine model performance and to compare between models, the recall was investigated specifically, which is the proportion of true positives of the total number of true cases. As the number of events was relatively low and not missing any patients at risk of ARE is crucial, a high recall of the models was desirable. The CI obtained for all metrics were obtained using bootstrapping, resampling the results 2,000 times. Moreover, an analysis of the agreement prediction between the DL model and the radiomics-based model was performed. To give a prediction per patient, the maximum prediction of ARE among the different lesion predictions of the patient was selected. The ground truth to which the prediction was compared with was the ARE status of the patient, meaning that the patient had at least one ARE lesion. An overview of the models tested can be found in **Figure 4**.

We evaluated on the external dataset for which cases the DL model and the best radiomics classifier obtained the same predictions and reported the number of cases for which those models agreed on the label. The metrics based on the data for which the models agreed was also reported.

3 RESULTS

3.1 Patient Characteristics

A total of 1,404 patients with 7,974 lesions from UCSF and 237 patients with 646 lesions from USZ were included. **Table 1** shows an overview of the patient characteristics of the UCSF and USZ data. Significant differences between the proportion of male and female patients between the datasets ($P < 0.01$), median age ($P = 0.03$), KPS status ($P < 0.01$), and the number of lesions per patient at treatment ($P < 0.01$) were found. Furthermore, the proportions of primary tumor (lung, melanoma, and breast) were different between the datasets, and the data from USZ did not have kidney, GI, sarcoma, or other types of primary locations that were present in the UCSF dataset. For the histology of the primary tumor, only the melanoma histology subtype was found to be present in a significantly different proportion.

3.2 Radiomics-Based Model and DL Model Results Based on the Three Different Preprocessing Methods of the Dataset

The best AUC on the test dataset for the radiomics-based models was found using the "harmonization" normalization, with an

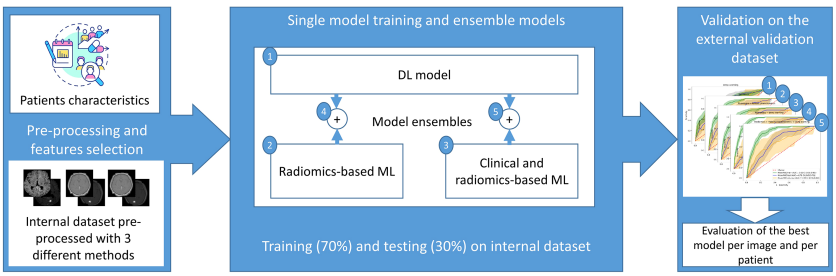


FIGURE 4 | General workflow of the model training process: first, the MRI data was pre-processed using 3 pre-processing methods, the most suitable pre-processed set of images was selected according to the radiomics-based model or the DL model performance on the internal test dataset, then the models were ensembled or trained separately, and finally the performance of each model was computed on the external dataset.

TABLE 1 | Patient characteristics of University of California—San Francisco (UCSF) and University Hospital Zurich (USZ) datasets.

Patient/tumor characteristic		Total UCSF data *N* = 1,404	USZ data *N* = 237	*P*
Sex (%)	Male	571 (41)	128 (54)	<0.01
	Female	833 (59)	109 (46)	
Median age ± SD		59 (13)	62 (12)	0.03
KPS (%)	80–100	1,053 (75)	198 (83)	<0.01
	40–80	351 (25)	37 (16)	<0.01
	10–40	0 (0)	2 (1)	–
Primary tumor location (%)	Lung	530 (38)	136 (58)	<0.01
	Breast	357 (25)	27 (11)	<0.01
	Melanoma	272 (19)	74 (31)	<0.01
	Kidney	91 (7)	0 (0)	–
	Gastrointestinal	57 (4)	0 (0)	–
	Gynecologic	27 (2)	0 (0)	–
	Sarcoma	20 (1)	0 (0)	–
	Other	50 (4)	0 (0)	–
Histology primary tumor (%)	Adenocarcinoma	802 (57)	124 (52)	0.17
	Melanoma	272 (19)	74 (31)	<0.01
	Renal cell carcinoma	88 (6)	0 (0)	–
	Small cell carcinoma	44 (3)	0 (0)	–
	Squamous cell carcinoma	40 (3)	10 (4)	0.26
	Sarcoma	18 (1)	0 (0)	–
	Large cell carcinoma	9 (0.6)	2 (1)	0.72
	Bone carcinoma	8 (0.6)	0 (0)	–
	Adeno squamous carcinoma	6 (0.4)	0 (0)	–
	Broncho alveolar cell carcinoma	5 (0.4)	0 (0)	–
	Germ cell carcinoma	2 (0.1)	0 (0)	–
	Lymphoma	1 (0.1)	0 (0)	–
	Other/NOS	109 (8)	27 (11)	0.06
Primary controlled		974 (70)	149 (63)	0.05
ECM present		1,097 (78)	190 (80)	0.48
Number of lesions per patient at treatment	Median ± SD	3 (7)	2 (3)	<0.01
Symptoms	Headaches	437 (31)	31 (13)	<0.01
	Hypertension	407 (29)	0 (0)	< 0.01
	Seizures	134 (10)	16 (7)	0.17
	Diabetes	98 (7)	13 (6)	0.4
	CTD	21 (2)	2 (1)	0.43
Number of lesions in total		7,974	646	–
Number of ARE cases (% of total lesions)		217 (2.7)	20 (3.1)	0.61
Number of patients with ARE (% of total patients)		155 (11)	19 (8)	0.16
Prescription dose ± SD (Gy)		18.5 (1.5)	20 (5.0)	–

The P-value of two-proportion z-test or unpaired two-sample t-test for significant differences between datasets was reported for each characteristic if applicable.
SD, standard deviation; KPS, Karnofsky performance score (80–100, good performance; 50–70, medium performance; and 10–40 bad performance); ECM, extracranial metastasis; BM, brain metastasis; CTD, connective tissue disorder; ARE, adverse radiation effect; Gy, gray.

AUC of 0.76 (CI of 0.70–0.81), compared with 0.75 (CI of 0.70–0.80) and 0.73 (CI of 0.67–0.79) for the "minimalist" and "standardization" methods, respectively.

The best AUC on the test dataset for the DL models was found using the "standardization" normalization, with an AUC of 0.72 (CI of 0.66–0.78), compared with 0.63 (CI of 0.57–0.70) and 0.65 (CI of 0.58–0.71) for the "minimalist" and "harmonization" methods, respectively. **Figure 5** shows the ROC curves of the training and testing datasets for the three different pre-processing methods for radiomics-based ML and for DL.

3.3 Results of the Combined Best-Performing Models

We calculated the AUC and CI for each model combination on the external validation dataset. The DL model, built on images pre-processed with the "standardization" method, achieved an AUC of 0.64 (CI of 0.50–0.76). The model built on radiomics features, extracted from the images pre-processed with the "harmonization" method, achieved an AUC of 0.73 (CI of 0.63–0.83). The model was built on 20 features selected through RFE. **Supplementary Figure S2A** provides an overview of the selected features and the corresponding importance in the XGBoost model. **Supplementary Table S2** provides an overview of the hyperparameters determined through grid search cross-validation. The model based on the combination of the DL features extracted from the last layer and radiomics features achieved an AUC of 0.71 (CI of 0.60–0.82). The model was built on 10 features selected through RFE. **Supplementary Figure S2B** provides an overview of the selected features and the corresponding importance in the XGBoost model. The model built on radiomics features, extracted from images pre-processed with the "harmonization" method, combined with patient characteristic features achieved an AUC of 0.70 (CI of 0.57–0.80). The model was built on 19 features selected through RFE. **Supplementary Figure S2C** provides an overview of the selected features and the corresponding importance in the XGBoost model. Finally, the model built on radiomics features, extracted from images pre-processed with the "harmonization" method, combined with DL features, extracted from images pre-processed with the "standardization" method, and patient characteristics achieved an AUC of 0.69 (CI of 0.56–0.81). The model was built on 20 features selected through RFE. **Supplementary Figure S2D** provides an overview of the selected features and the corresponding importance in the XGBoost model. **Figure 6** shows the ROC curves with CI of the training datasets, testing datasets, and validation datasets for these models.

The combination of radiomics and DL features achieved the highest combination of balanced accuracy and recall of 0.67 (CI of 0.56–0.76) and 0.80 (CI of 0.62–0.96), respectively, of the externally validated models for predictions per lesion. For a patient-level prediction, the DL model achieved an AUC of 0.70 (CI of 0.56–0.80) and that of the radiomics model an AUC of 0.72 (CI of 0.60–0.83). A combination of radiomics and DL achieved an AUC 0.71 (CI of 0.57–0.83), that of a combination of radiomics and patient characteristics an AUC of 0.71 (CI of 0.59–0.81), and that of a combination of radiomics features, DL features, and patient characteristics an AUC of 0.72 (CI of 0.58–0.84). The model combining radiomics features, DL features, and patient characteristics achieved the highest combination of balanced accuracy and recall of 0.65 (CI of 0.55–0.74) and 0.84 (CI of 0.65–1.00), respectively, of the externally validated models for predictions per patient. The DL model predictions and the radiomics-based model predictions per lesion agreed for 32% of the external dataset. For the per-patient classification, the DL model predictions and the radiomics combined with clinical feature-based model predictions agreed for 19% of the external

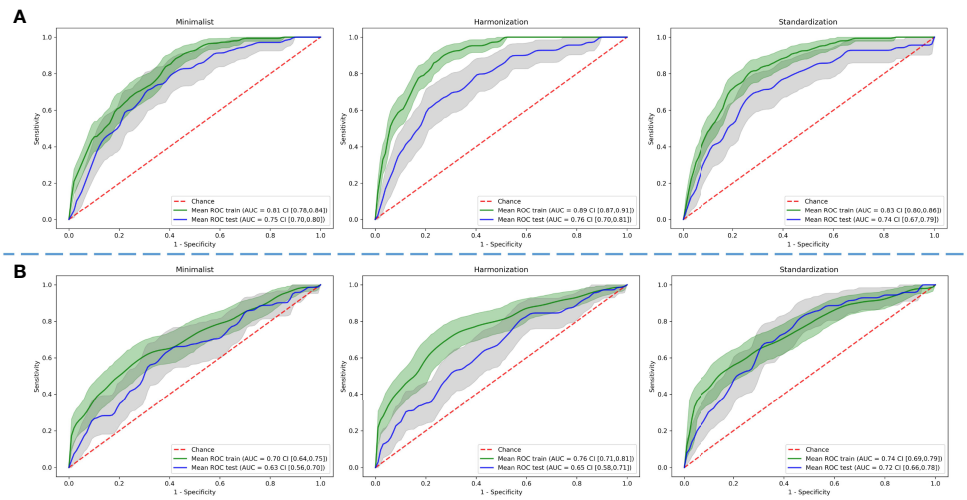


FIGURE 5 | Comparison of predictive performance through receiver operating characteristic curves for **(A)** radiomics-based machine learning and **(B)** deep learning models using three different pre-processed image datasets. The shaded areas represent the 95% confidence intervals of the corresponding receiver operating characteristic curves.

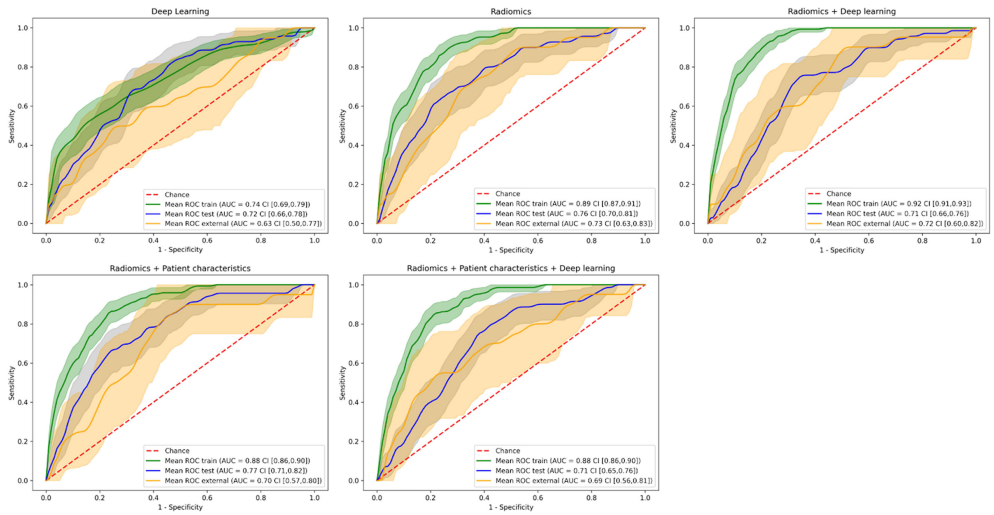


FIGURE 6 | Receiver operating characteristic curves of the training, testing, and external validation datasets for the different model combinations. The shaded areas represent the 95% confidence intervals of the corresponding receiver operating characteristic curves.

dataset. Because the number of patients for which the models agreed was low (47 patients, 6 with ARE), no CI could be derived. **Table 2** provides an overview of the AUC, balanced accuracy, precision, recall, and F1 score metrics for all DL and ML models on both lesion and patient levels and for the agreed labels on the external validation. The corresponding confusion matrices are in **Supplementary Figures S3, S4**, respectively. **Supplementary Tables S3, S4** contain the same metrics as that in **Table 2** for the training and testing datasets, respectively.

4 DISCUSSION

Patients with BM treated with SRT are at risk of developing ARE, such as RN. Early identification of these patients can help in clinical decision making. The MRIs required for SRT planning provide an opportunity to identify these patients through quantitative imaging methods. In this large-scale study, ML models that can successfully predict ARE were trained on T1-weighted MR imaging features from secondary brain tumors treated with SRT. As no consensus to harmonize MR images within and between centers exists, multiple methods were tested for the DL and ML pipeline, resulting in two optimal pre-processing methods ("harmonization" for the ML pipeline and "standardization" for the DL pipeline). A ML model trained with radiomics features combined with DL features yielded the highest predictive performance, with a combination of ROC AUC, balanced accuracy, and recall of 0.71, 0.67, and 0.80, respectively. At the patient level, the best-performing ML model was clearly a combination of radiomics, clinical (age at treatment, prior RS, and sex), and DL features achieving the highest predictive performance (AUC of 0.72), a balanced accuracy of 0.65, and recall of 0.84.

Performing an aggregate prediction (*i.e.*, using only those predictions that agreed on the outcome) did not improve predictive performance for the lesion-level prediction (AUC of 0.67) nor the binary prediction (balanced accuracy of 0.65). However, using this method, the highest recall of 0.90 was achieved, making this method very robust in detecting true positives.

The models pave the way for clinical decision making of patients at risk of ARE before treatment. The information on the risk of an individual patient may be used by clinicians to inform patients of the risk of ARE when SRT is used as treatment. Furthermore, this information may be used to perform an early stratification of those patients at high risk or may allow the patient and clinician to pursue alternative therapy, such as systemic therapy or alternate radiotherapy approaches (*e.g.*, dose de-intensified SRT or WBRT), if the risk of ARE outweighs the possible benefits of SRT (50).

To our knowledge, this is the first study that performs a pre-treatment prediction of ARE using quantitative image analysis. Several studies have investigated the possibility of differentiating between tumor recurrence and RN after treatment, which is nominally similar in purpose to identify those patients who may have ARE. Zhang et al. (51) used radiomics features extracted from four different MR sequences [T1, T1 post-contrast, T2, and fluid-attenuated inversion recovery (FLAIR)] at two different time-points during follow-up to differentiate RN from TP as confirmed pathologically. A model was built on a dataset of 87 patients with 97 lesions using 5 delta-radiomics features from T1 and T2 sequences. The AUC and binary prediction accuracy of the model were both 0.73. However, this result was obtained using leave-one-out cross-validation, as no external validation was used. Similarly, Peng *et al.* created a model on radiomics features extracted from T1 and T2 FLAIR on 66 patients with 77 lesions in total (52). The model was compared with a neuroradiologist's

TABLE 2 | Area under the curve (AUC), balanced accuracy, precision, recall, and F1 metrics with CI on the external validation on patient and lesion levels.

Per-lesion classification						Per-patient classification					
Approaches	**AUC**	**Balanced accuracy**	**Precision**	**Recall**	**F1 score**	**Approaches**	**AUC**	**Balanced accuracy**	**Precision**	**Recall**	**F1 score**
Best deep learning model	0.64 CI (0.50, 0.76)	0.57 CI (0.48, 0.64)	0.04 CI (0.02, 0.05)	0.85 CI (0.67, 1.00)	0.07 CI (0.04, 0.10)	Best deep learning model	0.70 CI (0.56, 0.83)	0.63 CI (0.52, 0.73)	0.17 CI (0.09, 0.25)	0.60 CI (0.39, 0.78)	0.26 CI (0.16, 0.37)
Best radiomics model	0.73 CI (0.63, 0.83)	0.62 CI (0.51, 0.74)	0.07 CI (0.03, 0.11)	0.45 CI (0.23, 0.67)	0.12 CI (0.05, 0.19)	Best radiomics model	0.72 CI (0.60, 0.83)	0.59 CI (0.51, 0.69)	0.40 CI (0.09, 0.75)	0.21 CI (0.05, 0.43)	0.28 CI (0.07, 0.48)
Radiomics and DL	0.71 CI (0.60, 0.82)	0.67 CI (0.56, 0.76)	0.05 CI (0.03, 0.08)	0.80 CI (0.62, 0.96)	0.10 CI (0.06, 0.14)	Radiomics and DL	0.71 CI (0.57, 0.83)	0.66 CI (0.54, 0.77)	0.14 CI (0.07, 0.22)	0.63 CI (0.40, 0.84)	0.23 CI (0.13, 0.34)
Radiomics and patient characteristics	0.70 CI (0.57, 0.80)	0.62 CI (0.51, 0.74)	0.06 CI (0.03, 0.10)	0.50 CI (0.28, 0.73)	0.11 CI (0.05, 0.17)	Radiomics and patient characteristics	0.71 CI (0.59, 0.81)	0.57 CI (0.48, 0.68)	0.16 CI (0.04, 0.30)	0.26 CI (0.08, 0.47)	0.20 CI (0.05, 0.35)
Radiomics, DL, and patient characteristics	0.69 CI (0.56, 0.81)	0.64 CI (0.53, 0.74)	0.05 CI (0.03, 0.08)	0.70 CI (0.48, 0.89)	0.09 CI (0.05, 0.14)	Radiomics, DL, and patient characteristics	0.72 CI (0.58, 0.84)	0.65 CI (0.55, 0.74)	0.12 CI (0.07, 0.17)	0.84 CI (0.65, 1.00)	0.21 CI (0.13, 0.29)
Agreed labels	0.67 CI (0.53, 0.81)	0.65 CI (0.53, 0.73)	0.07 CI (0.03, 0.12)	0.90 CI (0.67, 1.00)	0.13 CI (0.06, 0.21)	Agreed labels	NA	NA	NA	NA	NA

performance. No external validation was used, and instead a leave-one-out cross-validation was performed, which gave an AUC of 0.81. The sensitivity and specificity of the neuroradiologist were 0.97 and 0.17, compared with 0.65 and 0.87 for the radiomics-based model. In Park etal. (53), the study compared the results obtained after training radiomics-based models using different MRI sequences [T1, T2, and apparent diffusion coefficient (ADC)]. The models were trained using the data from 86 patients and tested on an external dataset of 41 patients. The best AUC was found on the ADC-based data with 0.80, while the other sequences had AUCs of around 0.65. These results are similar or higher than the results obtained with our model, though within the range of the confidence intervals for the model based on radiomics and DL, and the lack of an external dataset on two of the studies makes the validity of these models difficult to determine (52). Most other studies have a similar lack of external validation and total number of included patients, further making the results difficult to compare with the present study (54). These results show that the model presented in this study is able to perform similarly to or even outperform models that perform classification (post-treatment) instead of prediction (pre-treatment) of ARE.

One of the strengths of the present study is the large number of included patients and subsequent lesions, with 7,974 lesions (2.7% ARE) of 1,404 patients in training and testing and 646 lesions (3.1% ARE) of 237 patients in the external validation. This provides a large volume of data for our models to train on, ensuring that it covers the wide variability found between patients. In addition, the inclusion of an external validation is another strength, especially seeing the general lack of one in most other studies investigating ARE. This ensures that the reported result is not too optimistic and shows that our model can be generalizable to populations from a different hospital in a different country and even with different treatments from the training and testing sets. While the difference in treatment between the training (exclusively SRS) and external validation (a mix of SRS and FSRT) may induce variability due to small differences in treatment planning for these methods, literature has shown that these methods carry the same risk of ARE and were therefore considered interchangeable (16, 17, 19).

The large confidence interval on the external validation is partially due to the low number of positive findings in this dataset (n = 20). This is because of the large imbalance in outcomes for both ARE and tumor failure. One of the major problems that may arise from this imbalance is a skewed view of predictive performance. However, this was addressed in the present study through multiple measures. The DL model was trained on a balanced subset of the data that only included patients that suffered at least 1 ARE. For ML, the XGBoost model was trained while scaling the weights of positive and negative classes and the respective proportion of the labels. Finally, through analysis of the confusion matrix, precision recall curves, and recall metric, we ensured that the performance of the model was not entirely driven by labeling the data as the majority class.

While the models have been successfully validated on a dataset from an external center, further validation on multiple centers is required to ensure that the models are generalizable. Future research could therefore focus on validating the present model on other datasets, potentially with recalibration of the model. At a later stage, a clinical trial to test the efficacy of the model is needed to be able to incorporate the model in a clinical setting. A model combining radiomics features, DL features, and patient characteristics with a high accuracy could help choose other treatment options such as surgery only, systemic therapy, or palliative care (55) if the predicted risk of developing ARE is high. The model could also predict if the patient would be at a low risk of developing ARE, in which case SRT could be preferred over other treatment options.

In the present study, only one sequence of the MRI scan was used. Previous studies showed that a combination of radiomics computed on T1 and T2 sequences performs best to differentiate ARE and TP (51, 52), and ADC sequence seems to also show a

higher performance (53). Investigating more sequences in a future study may therefore improve the performance of the imaging-based models.

Lastly, for ARE (and, to a lesser degree, TP), treatment is one of the primary factors. In this study, multiple-dose-treatment-related variables have been included, such as prior treatments to the same patients as well as dose variables and the volumes encompassing certain dose levels. However, a more thorough "dosiomics" analysis would probably improve the prediction of ARE. Liang et al. (56) described a method to extract the spatial and texture radiomics features from dose maps (56). They found several radiomics features which have a significant predictive value of radiation pneumonitis. Using a similar method for ARE in BM may result in improved prediction results. Our predictions could also be combined with models automatically classifying tumors and RN on brain MRI, such as in Zhang et al. (51), potentially strengthening the results of those studies.

5 CONCLUSION

Radiomics is able to predict lesions at a high risk of ARE, especially when combined with DL features. When predicting ARE on a patient level, the highest performance was found using a combination of radiomics, DL, clinical, and treatment-related features. These models could potentially be used to aid clinical decision making for patients with BM treated with either gamma knife or EBRT.

DATA AVAILABILITY STATEMENT

The corresponding author does not own the datasets used (acquired with DTAs). Requests to access the datasets should be directed to olivier.morin@ucsf.edu (for the data from UCSF); Nicolaus.Andratschke@usz.ch (for the data from USZ).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the cantonal ethics committee Zurich and University of California San Francisco (UCSF) Institutional Review Board (IRB). 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

MB and SK performed all the ML/DL analysis and wrote the manuscript. SK, MV, SB, and OM collected and curated the imaging and patient data from UCSF. SP helped with the ML/DL analysis and study design. HW supervised the progression of the project and the writing of this article and guaranteed the integrity of the analysis and results presented. AC and MV helped with the ML analysis. JT, JK, and NA collected the imaging and patient data from USZ. LH and SB aided with the clinical aspects of the study. PL and OM devised the project's aim and supervised the progression of the project. All authors contributed to the article and approved the submitted version.

FUNDING

The research project has been partially funded by the Clinical Research Priority Program "Artificial Intelligence in Oncological Imaging" of the University of Zurich. PL, HW, MB, SK acknowledge financial support from ERC advanced grant (ERC-ADG-2015 n° 694812 - Hypoximmuno), the European Union's Horizon 2020 research and innovation programme under grant agreement: MSCA-ITN-PREDICT n° 766276, CHAIMELEON n° 952172, EuCanImage n° 952103 and IMI-OPTIMA n° 101034347.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.920393/full#supplementary-material

REFERENCES

1. Walker AE, Robins M, Weinfeld FD. Epidemiology of Brain Tumors: The National Survey of Intracranial Neoplasms. *Neurology* (1985) 35:219–9. doi: 10.1212/wnl.35.2.219
2. Johnson JD, Young B. Demographics of Brain Metastasis. *Neurosurg Clinics North America* (1996) 7:337–44. doi: 10.1016/s1042-3680(18)30365-6
3. Wen PY, Loeffler JS. Management of Brain Metastases. *Oncology* (1999) 13 (7):941–54, 957–61.
4. Schouten LJ, Rutten J, Huveneers HAM, Twijnstra A. Incidence of Brain Metastases in a Cohort of Patients With Carcinoma of the Breast, Colon, Kidney, and Lung and Melanoma. *Cancer* (2002) 94(10):2698–705. doi: 10.1002/cncr.10541
5. Barnholtz-Sloan JS, Sloan AE, Davis FG, Vigneau FD, Lai P, Sawaya RE. Incidence Proportions of Brain Metastases in Patients Diagnosed, (1993 to 2001) in the Metropolitan Detroit Cancer Surveillance System. *J Clin Oncol: Off J Am Soc Clin Oncol* (2004) 22(14):2865–72. doi: 10.1200/JCO.2004.12.149
6. Rangachari D, Yamaguchi N, VanderLaan PA, Folch E, Mahadevan A, Floyd SR, et al. Brain Metastases in Patients With EGFR -Mutated or ALK -Rearranged non-Small-Cell Lung Cancers. *Lung Cancer* (2015) 88:108–11. doi: 10.1016/j.lungcan.2015.01.020
7. Huber RM, Hansen KH, Paz-Ares RL, West HL, Reckamp KL, Leighl NB, et al. Brigatinib in Crizotinib-Refractory ALK+ NSCLC: 2-Year Follow-Up on Systemic and Intracranial Outcomes in the Phase 2 ALTA Trial. *J Thorac Oncol: Off Publ Int Assoc Study Lung Cancer* (2020) 15(3). doi: 10.1016/j.jtho.2019.11.004

8. Venur VA, Karivedu V, Ahluwalia MSSystemic Therapy for Brain Metastases. In: *Handbook of Clinical Neurology*. Elsevier. p. 137–53.
9. Vogelbaum MA, Brown PD, Messersmith H, Brastianos PK, Burri S, Cahill D, et al. Treatment for Brain Metastases: ASCO-SNO-ASTRO Guideline. *J Clin Oncol: Off J Am Soc Clin Oncol* (2022) 40(5):492–516. doi: 10.1200/JCO.21.02314
10. McTyre E, Scott J, Chinnaiyan P. Whole Brain Radiotherapy for Brain Metastasis. *Surg Neurol Int* (2013) 4(Suppl 4):S236–44. doi: 10.4103/2152-7806.111301
11. Kraft J, Zindler J, Minniti G, Guckenberger M, Andratschke N. Stereotactic Radiosurgery for Multiple Brain Metastases. *Curr Treat Options Neurol* (2019) 21(2):6. doi: 10.1007/s11940-019-0548-3
12. Kraft J, Mayinger M, Willmann J, Brown M, Tanadini-Lang S, Wilke L, et al. Management of Multiple Brain Metastases: A Patterns of Care Survey Within the German Society for Radiation Oncology. *J Neuro Oncol* (2021) 152 (2):395–404. doi: 10.1007/s11060-021-03714-w
13. Badiyan SN, Regine WF, Mehta M. Stereotactic Radiosurgery for Treatment of Brain Metastases. *J Oncol Pract / Am Soc Clin Oncol* (2016) 12(8):703–12. doi: 10.1200/JOP.2016.012922
14. Walker AJ, Ruzevick J, Malayeri AA, Rigamonti D, Lim M, Redmond KJ, et al. Postradiation Imaging Changes in the CNS: How can We Differentiate Between Treatment Effect and Disease Progression? *Future Oncol* (2014) 10 (7):1277–97. doi: 10.2217/fon.13.271
15. Sneed PK, Mendez J, Vemer-van den Hoek JGM, Seymour ZA, Ma L, Molinaro AM, et al. Adverse Radiation Effect After Stereotactic Radiosurgery for Brain Metastases: Incidence, Time Course, and Risk Factors. *J Neurosurg* (2015) 123(2):373–86. doi: 10.3171/2014.10.JNS141610
16. Gerosa M, Nicolato A, Foroni R, Zanotti B, Tomazzoli L, Miscusi M, et al. Gamma Knife Radiosurgery for Brain Metastases: A Primary Therapeutic Option. *J Neurosurg* (2002) 97:515–24. doi: 10.3171/jns.2002.97.supplement_5.0515
17. Lawrence YR, Allen Li X, el Naqa I, Hahn CA, Marks LB, Merchant TE, et al. Radiation Dose–Volume Effects in the Brain. *Int J Radiat Oncol Biol Phys* (2010) 76:S20–7. doi: 10.1016/j.ijrobp.2009.02.091
18. Minniti G, Dâ€™Angelillo RM, Scaringi C, Trodella LE, Clarke E, Matteucci P, et al. Fractionated Stereotactic Radiosurgery for Patients With Brain Metastases. *J Neuro Oncol* (2014) 117(2):295–301. doi: 10.1007/s11060-014-1388-3
19. Vellayappan B, Tan CL, Yong C, Khor LK, Koh WY, Yeo TT, et al. Diagnosis and Management of Radiation Necrosis in Patients With Brain Metastases. *Front Oncol* (2018) 8:395. doi: 10.3389/fonc.2018.00395
20. Petrovich Z, Yu C, Giannotta SL, Oâ€™Day S, Apuzzo MLJ. Survival and Pattern of Failure in Brain Metastasis Treated With Stereotactic Gamma Knife Radiosurgery. *J Neurosurg* (2002) 97(5 Suppl):499–506. doi: 10.3171/jns.2002.97.supplement_5.0499
21. 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
22. Aerts HJWL, Velazquez ER, Leijenaar RTH, Parmar C, Grossmann P, Carvalho S, et al. Decoding Tumour Phenotype by Noninvasive Imaging Using a Quantitative Radiomics Approach. *Nat Commun 5* (2014) p:4006. doi: 10.1038/ncomms5006
23. Zhou M, Scott J, Chaudhury B, Hall L, Goldgof D, Yeom KW, et al. Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches. *Am J Neuroradiol* (2018) 39:208–16. doi: 10.3174/ajnr.a5391
24. Morin O, Chen WC, Nassiri F, Susko M, Magill ST, Vasudevan HN, et al. Integrated Models Incorporating Radiologic and Radiomic Features Predict Meningioma Grade, Local Failure, and Overall Survival. *Neuro Oncol Adv* (2019) 1:vdz011. doi: 10.1093/noajnl/vdz011
25. Avanzo M, Wei L, Stancanello J, Vallières M, Rao A, Morin O, et al. Machine and Deep Learning Methods for Radiomics. *Med Physics* (2020) 47:185–202. doi: 10.1002/mp.13678
26. Rogers W, Thulasi Seetha S, Refaee TAG, Lieverse RIY, Granzier RWY, Ibrahim A, et al. Radiomics: From Qualitative to Quantitative Imaging. *Br J Radiol* (2020) 93(1108):20190948. doi: 10.1259/bjr.20190948
27. Abidin AZ, Dar I, Dâ€™Souza AM, Lin EP, Wismüller A. Investigating a Quantitative Radiomics Approach for Brain Tumor Classification. In. In: *Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging. Medical Imaging 2019: Biomedical Applications in Molecular, Structural, and Functional Imaging, SPIE* (2019). p. 36–45.
28. Dong F, Li Q, Jiang B, Zhu X, Zeng Q, Huang P, et al. Differentiation of Supratentorial Single Brain Metastasis and Glioblastoma by Using Peri-Enhancing Oedema Region-Derived Radiomic Features and Multiple Classifiers. *Eur Radiol* (2020) 30:3015–22. doi: 10.1007/s00330-019-06460-w
29. Huang C-Y, Lee C-C, Yang H-C, Lin C-J, Wu H-M, Chung W-Y, et al. Radiomics as Prognostic Factor in Brain Metastases Treated With Gamma Knife Radiosurgery. *J Neuro Oncol* (2020) 146:439–49. doi: 10.1007/s11060-019-03343-4
30. Mouraviev A, Detsky J, Sahgal A, Ruschin M, Lee YK, Karam I, et al. Use of Radiomics for the Prediction of Local Control of Brain Metastases After Stereotactic Radiosurgery. *Neuro-oncology* (2020) 22:797–805. doi: 10.1093/neuonc/noaa007
31. Ortiz-Ramón R, Larroza A, Ruiz-España S, Arana E, Moratal D. Classifying Brain Metastases by Their Primary Site of Origin Using a Radiomics Approach Based on Texture Analysis: A Feasibility Study. *Eur Radiol* (2018) 28(11):4514–23. doi: 10.1007/s00330-018-5463-6
32. Kniep HC, Madesta F, Schneider T, Hanning U, Schönfeld MH, Schön G, et al. Radiomics of Brain MRI: Utility in Prediction of Metastatic Tumor Type. *Radiology* (2019) 28:4514–23. doi: 10.1148/radiol.2018180946
33. Bhatia A, Birger M, Veeraraghavan H, Um H, Tixier F, McKenney AS, et al. MRI Radiomic Features are Associated With Survival in Melanoma Brain Metastases Treated With Immune Checkpoint Inhibitors. *Neuro-oncology* (2019) 21(12):1578–86. doi: 10.1093/neuonc/noz141
34. Della Seta M, Collettini F, Chapiro J, Angelidis A, Engeling F, Hamm B, et al. A 3D Quantitative Imaging Biomarker in Pre-Treatment MRI Predicts Overall Survival After Stereotactic Radiation Therapy of Patients With a Singular Brain Metastasis. *Acta Radiol* (2019) 60(11):1496–503. doi: 10.1177/0284185119831692
35. Cho J, Kim YJ, Sunwoo L, Lee GP, Nguyen TQ, Cho SJ , et al. Deep Learning-Based Computer-Aided Detection System for Automated Treatment Response Assessment of Brain Metastases on 3D MRI. *Front Oncol* (2021) 11:739639. doi: 10.3389/fonc.2021.739639
36. Parekh VS, Jacobs MA. Deep Learning and Radiomics in Precision Medicine. *Expert Rev Precis Med Drug Dev* (2019) 4(2):59–72. doi: 10.1080/23808993.2019.1585805
37. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: Improved N3 Bias Correction. *IEEE Trans Med Imaging* (2010) 29 (6):1310–20. doi: 10.1109/TMI.2010.2046908
38. Juntu J, Sijbers J, Dyck D, Gielen J. Bias Field Correction for MRI Images. *Adv Soft Computing* (2005), 543–51. doi: 10.1007/3-540-32390-2_64
39. 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(16):165011. doi: 10.1088/1361-6560/ab2f44
40. Moradmand H, Aghamiri SMR, Ghaderi R. Impact of Image Preprocessing Methods on Reproducibility of Radiomic Features in Multimodal Magnetic Resonance Imaging in Glioblastoma. *J Appl Clin Med Phys / Am Coll Med Phys* (2020) 21(1):179–90. doi: 10.1002/acm2.12795
41. Masoudi S, Harmon SA, Mehralivand S, Walker SM, Raviprakash H, Bagci U, et al. Quick Guide on Radiology Image Pre-Processing for Deep Learning Applications in Prostate Cancer Research. *J Med Imaging (Bellingham Wash)* (2021) 8(1):010901. doi: 10.1117/1.JMI.8.1.010901
42. Duron L, Balvay D, Vande Perre S, Bouchouicha A, Savatovsky J, Sadik J-C, et al. Gray-Level Discretization Impacts Reproducible MRI Radiomics Texture Features. *PLoS One* (2019) 14(3):e0213459. doi: 10.1371/journal.pone.0213459
43. Carré A, Klausner G, Edjlali M, Lerousseau M, Briend-Diop J, Sun R, et al. Standardization of Brain MR Images Across Machines and Protocols: Bridging the Gap for MRI-Based Radiomics. *Sci Rep* (2020) 10(1):12340. doi: 10.1038/s41598-020-69298-z
44. 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

45. *Radiomic Features — Pyradiomics V3.0.1.Post9+Gdfe2c14 Documentation* (2019). Available at: https://pyradiomics.readthedocs.io/en/latest/features.html (Accessed 21 October 2021).
46. Chatterjee A, Vallieres M, Dohan A, Levesque IR, Ueno Y, Saif S, et al.) Creating Robust Predictive Radiomic Models for Data From Independent Institutions Using Normalization. *IEEE Trans Radiat Plasma Med Sci* (2019) 3:210–5. doi: 10.1109/trpms.2019.2893860
47. Chollet F. Xception: Deep Learning With Depthwise Separable Convolutions. *Proc IEEE Conf Comput Vision Pattern Recognition* (2017), 1251–8. doi: 10.1109/CVPR.2017.195
48. Kingma DP, Ba J. *Adam: A Method for Stochastic Optimization*. arXiv [cs.LG] (2014). Available at: http://arxiv.org/abs/1412.6980.
49. Youden WJ. Index for Rating Diagnostic Tests. *Cancer* (1950) 3(1):32–5. doi: 10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3
50. Alvarez-Breckenridge C, Remon J, Piña Y, Nieblas-Bedolla E, Forsyth P, Hendriks L, et al. Emerging Systemic Treatment Perspectives on Brain Metastases: Moving Toward a Better Outlook for Patients. *Am Soc Clin Oncol Educ Book Am Soc Clin Oncol Annu Meeting* (2022) 42:1–19. doi: 10.1200/EDBK_352320
51. Zhang Z, Yang J, Ho A, Jiang W, Logan J, Wang X, et al. A Predictive Model for Distinguishing Radiation Necrosis From Tumour Progression After Gamma Knife Radiosurgery Based on Radiomic Features From MR Images. *Eur Radiol* (2018) 28:2255–63. doi: 10.1007/s00330-017-5154-8
52. Peng L, Parekh V, Huang P, Lin DD, Sheikh K, Baker B, et al. Distinguishing True Progression From Radionecrosis After Stereotactic Radiation Therapy for Brain Metastases With Machine Learning and Radiomics. *Int J Radiat Oncol Biol Phys* (2018) 102:1236–43. doi: 10.1016/j.ijrobp.2018.05.041
53. Park YW, Choi D, Park JE, Ahn SS, Kim H, Chang JH, et al. Differentiation of Recurrent Glioblastoma From Radiation Necrosis Using Diffusion Radiomics With Machine Learning Model Development and External Validation. *Sci Rep* (2021) 11:2913. doi: 10.1038/s41598-021-82467-y
54. Salvestrini V, Greco C, Guerini AE, Longo S, Nardone V, Boldrini L, et al. The Role of Feature-Based Radiomics for Predicting Response and Radiation Injury After Stereotactic Radiation Therapy for Brain Metastases: A Critical Review by the Young Group of the Italian Association of Radiotherapy and Clinical Oncology (yAIRO). *Trans Oncol* (2022) 15:101275. doi: 10.1016/j.tranon.2021.101275
55. Lin X, DeAngelis LM. Treatment of Brain Metastases. *J Clin Oncol: Off J Am Soc Clin Oncol* (2015) 33(30):3475–84. doi: 10.1200/JCO.2015.60.9503
56. Liang B, Yan H, Tian Y, Chen X, Yan L, Zhang T, et al. Dosiomics: Extracting 3d Spatial Features From Dose Distribution to Predict Incidence of Radiation Pneumonitis. *Front Oncol* (2019) 9:269. doi: 10.3389/fonc.2019.00269

Conflict of Interest: LH: none related to the current manuscript, outside of current manuscript: research funding Roche Genentech, Boehringer Ingelheim, AstraZeneca, Takeda (all institution, Beigene under negotiation); advisory board: BMS, Eli Lilly, Roche Genentech, Pfizer, Takeda, MSD, Merck, Novartis, Boehringer Ingelheim, Amgen, Janssen (all institution, Roche one time self); speaker: MSD, Lilly (institution); travel/conference reimbursement: Roche Genentech (self); mentorship program with key opinion leaders: funded by AstraZeneca; fees for educational webinars: Benecke, Medtalks, VJOncology (self), high5oncology (institution); interview sessions funded by Roche Genentech, Bayer, Lilly (institution); local PI of clinical trials: AstraZeneca, Novartis, BMS, MSD, Merck, GSK, Takeda, Blueprint Medicines, Roche Genentech, Janssen Pharmaceuticals, Mirati; PL: none related to the current manuscript; outside of current manuscript: grants/sponsored research agreements from Radiomics SA, Convert Pharmaceuticals and LivingMed Biotech. He received a presenter fee (in cash or in kind) and/or reimbursement of travel costs/consultancy fee (in cash or in kind) from Radiomics SA, BHV, Varian, Elekta, ptTheragnostic/DNAmito, BMS, and Convert pharma. PL has minority shares in the companies Radiomics SA, Convert pharmaceuticals, Comunicare, and LivingMed Biotech, and he is co-inventor of two issued patents with royalties on radiomics (PCT/NL2014/050248 and PCT/NL2014/050728), licensed to Radiomics SA; one issued patent on mtDNA (PCT/EP2014/059089), licensed to ptTheragnostic/DNAmito; one non-issued patent on LSRT (PCT/ P126537PC00), licensed to Varian; three non-patented inventions (softwares) licensed to ptTheragnostic/DNAmito, Radiomics SA and Health Innovation Ventures and two non-issued, non-licensed patents on Deep Learning-Radiomics (N2024482, N2024889). He confirms that none of the above entities or funding sources were involved in the preparation of this paper.

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 | Frontiers in Oncology

ORIGINAL RESEARCH
published: 18 July 2022
doi: 10.3389/fonc.2022.875661

CT-Only Radiotherapy: An Exploratory Study for Automatic Dose Prediction on Rectal Cancer Patients *Via* Deep Adversarial Network

Jiaqi Cui[1†], Zhengyang Jiao[1†], Zhigong Wei[2†], Xiaolin Hu[3], Yan Wang[1], Jianghong Xiao[4*] and Xingchen Peng[2*]*

[1] School of Computer Science, Sichuan University, Chengdu, China, [2] Department of Biotherapy, Cancer Center, West China Hospital, Sichuan University, Chengdu, China, [3] West China School of Nursing, West China Hospital, Sichuan University, Chengdu, China, [4] Department of Radiation Oncology, Cancer Center, West China Hospital, Sichuan University, Chengdu, China

Purpose: Current deep learning methods for dose prediction require manual delineations of planning target volume (PTV) and organs at risk (OARs) besides the original CT images. Perceiving the time cost of manual contour delineation, we expect to explore the feasibility of accelerating the radiotherapy planning by leveraging only the CT images to produce high-quality dose distribution maps while generating the contour information automatically.

Materials and Methods: We developed a generative adversarial network (GAN) with multi-task learning (MTL) strategy to produce accurate dose distribution maps without manually delineated contours. To balance the relative importance of each task (i.e., the primary dose prediction task and the auxiliary tumor segmentation task), a multi-task loss function was employed. Our model was trained, validated and evaluated on a cohort of 130 rectal cancer patients.

Results: Experimental results manifest the feasibility and improvements of our contour-free method. Compared to other mainstream methods (i.e., U-net, DeepLabV3+, DoseNet, and GAN), the proposed method produces the leading performance with statistically significant improvements by achieving the highest HI of 1.023 (3.27E-5) and the lowest prediction error with ΔD95 of 0.125 (0.035) and ΔDmean of 0.023 (4.19E-4), respectively. The DVH differences between the predicted dose and the ideal dose are subtle and the errors in the difference maps are minimal. In addition, we conducted the ablation study to validate the effectiveness of each module. Furthermore, the results of attention maps also prove that our CT-only prediction model is capable of paying attention to both the target tumor (i.e., high dose distribution area) and the surrounding healthy tissues (i.e., low dose distribution areas).

OPEN ACCESS

Edited by:
Xiaomeng Li,
Hong Kong University of Science and Technology, China

Reviewed by:
Wei Zhao,
Beihang University, China
Xiaowei Hu,
The Chinese University of Hong Kong, China

***Correspondence:**
Yan Wang
wangyanscu@hotmail.com
Jianghong Xiao
xiaojh86@foxmail.com
Xingchen Peng
pxx2014@scu.edu.cn

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

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: *14 February 2022*
Accepted: *14 June 2022*
Published: *18 July 2022*

Citation:
Cui J, Jiao Z, Wei Z, Hu X, Wang Y, Xiao J and Peng X (2022) CT-Only Radiotherapy: An Exploratory Study for Automatic Dose Prediction on Rectal Cancer Patients Via Deep Adversarial Network. Front. Oncol. 12:875661. doi: 10.3389/fonc.2022.875661

Conclusion: The proposed CT-only dose prediction framework is capable of producing acceptable dose maps and reducing the time and labor for manual delineation, thus having great clinical potential in providing accurate and accelerated radiotherapy. Code is available at https://github.com/joegit-code/DoseWithCT

Keywords: dose prediction, deep learning, radiotherapy planning CT-scan, rectal cancer, GAN structure

INTRODUCTION

Rectal cancer is the third most deadly and fourth most commonly diagnosed cancer in the world with its incidence rising constantly (1). As a mainstay treatment, radiation therapy benefits approximately 50% of cancer patients and contributes to around 40% of curative cases (2). Recently, volumetric modulated arc therapy (VMAT) has been widely applied to clinical radiotherapy for its significant advantages in dose modulation.

In a typical VMAT planning, the treatment planner is required to deliver lethal and homogeneous doses to the planning target volume (PTV), while minimizing the therapeutic toxicity to the organs-at-risk (OARs) (3, 4). To satisfy this complex rule, the planner is required to perform multiple rounds of parameter adjustment and optimization in a trial-and-error manner. Ultimately, the integral dose distribution map, which visually represents the dose prescribed to each organ as well as the beam angles and numbers, is obtained. However, even with the availability of the treatment planning system (TPS), this process still costs considerable manpower and an average of up to 11 hours (5). If the planner is able to obtain the high-quality dose map prior to the planning process, and takes it as an initial point for the treatment planning, the repetitions of the trial-and-error process as well as the total planning time can be significantly reduced. Therefore, researches on obtaining high-quality dose distribution maps rapidly are of great clinical significance in providing accurate and accelerated radiotherapy.

Before our work, a range of deep-learning (6–10) has been developed to predict the dose. For example, given the limited available data in clinic, Zeng et al. (10) proposed a two-phase deep transfer learning framework to predict the dose distribution for cervical cancer patients. Nguyen et al. (6) designed a hierarchically densely connected U-net to estimate the dose distribution from the CT images, the anatomic delineations, and the prescription dose for patients with head and neck (H&N) cancer. Besides, Song et al. (7) reported DeeplabV3+ which utilized CT images and anatomic contours to predict the dose for rectal cancer patients. To take full advantage of the historical patient data, Mardani et al. (8) proposed a learning empowered approach which employed a multi-task linear regression model to predict 3D dose volume for a new patient by extracting the shared features of historical patients and their tumor shapes. Based on the clinical post-optimization strategies, Zhong et al. (9) designed a new automatic radiotherapy planning strategy that was able to produce clinically acceptable dose distributions.

More recently, generative adversarial networks (GANs) have attracted much attention from researchers due to their impressive performance in synthesis. To reduce the radiation of positron emission tomography (PET), Wang et al. (11) developed a 3D auto-context-based locality adaptive multi-modality generative adversarial networks model (LA-GANs) to synthesize the high-quality PET image from the low-dose one. Taking the advantage of U-net, Wang et al. (12) also proposed a 3D U-net-like deep architecture, combing hierarchical features by skip connections to generate full-dose PET images. Moreover, Luo et al. (13) presented AR-GAN, utilizing an adaptive rectification based generative adversarial network with spectrum constraint for standard-dose PET estimation. GAN has also made a quantum leap in the medical segmentation task. To be specific, Shi et al. (14) innovatively proposed an adaptive-similarity-based multi-modality feature selection method for Alzheimer's disease classification. Perceiving the difficulty in medical data acquisition, Wang et al. (15) proposed a triple-uncertainty guided semi-supervised model for medical image segmentation. Inspired by them, several GAN-based methods (16–18) were also proposed for dose prediction. Based on the generative adversarial network, Zhan et al. (19) developed a multi-constraint dose prediction model, capturing both global and local contextual information to predict the dose distribution for cervical and rectal cancer patients. To further reduce the time for contour delineation in radiotherapy planning, Li et al. (20) presented a multi-task attention adversarial network, including a main dose prediction task to generate the dose maps and an auxiliary segmentation task to automatically provide additional tumor delineation. Never, the developed deep-learning-based radiotherapy dose prediction models generally demand extra inputs (e.g., delineations of PTV and OARs) to supplement essential anatomical information for satisfactory predictions. Perceiving the time-consuming fact of manual contour delineation, we suggest accelerating the planning by taking the original CT images as unique input to produce high-quality dose predictions.

In order to investigate the feasibility of this CT-only dose prediction idea, we present a GAN-based framework for automatic dose prediction, which is free of manual delineation. To compensate for the missing anatomic information, we incorporate multi-task learning (MTL) into the framework by employing an auxiliary tumor segmentation task to provide essential guidance (i.e., anatomic structure of the tumor) to the primary dose prediction task. This model was trained, validated and evaluated on a cohort of 130 rectal patients. To the best of our knowledge, this is the first work that explores to predict the dose distribution maps *via* only the original CT images while mining the anatomic information automatically and concurrently.

MATERIALS AND METHODS

Patient Cohort and CT Images

A total number of 130 postoperative patients with rectal cancer were included in this study. Ethical approval was granted by the local ethics committee. The collection of CT images followed the standard medical procedures. To be specific, each patient was immobilized with an individualized thermoplastic mask in the supine position with arms raised above the head. The intravenous contrast-enhanced CT covering the total pelvis volume for each patient was obtained. All patients were asked to drink 0.5 liter of water 1.5 h prior to scanning and refraining from voiding (21). In addition, each patient was accompanied by a PTV and four OARs including bladder, left femoral head (FHL), right femoral head (FHR), and small intestine (SI). The PTV and OARs on CT slices were delineated by the physicians based on the guideline (22) and reviewed by the senior radiation oncologists. The prescription dose of the PTV was 50.40Gy/28 fractions. The intensity-modulated radiation therapy (IMRT) treatment plans were required to cover ≥98% of the PTV with ≥93% of the prescribed dose, ≤10% of the PTV with ≥105% of the prescribed dose, and ≤5% of the PTV with ≥115% of the prescribed dose. Small intestine dose was limited to V35 <180 cc, V40 <100 cc, and V45 <65 cc. Femoral head dose was limited to V40 <40%, V45 <25%, and a maximum dose of 50 Gy. Bladder dose was limited to V40 <40%, V45 <15%, and a maximum dose 50 Gy (23). All the OARs and other normal tissues should be given as low a dose as possible. All the manual VMAT plans were conducted on the Raystation v4.7 TPS with the model for the Elekta Versa HD linear accelerators. Two 360 coplanar arc beams consisting of 91 control points respectively sharing the same isocenter were employed with 6-MV photon energy. The collimator angle was fixed at 0 and the maximum field size was 40×40 cm^2 with a dynamic multi-leaf collimator and automatically tracking collimator jaws for each control point. Dose engine algorithm was collapsed cone convolution with the grid resolution 0.3×0.3×0.3 cm^3. The optimization engine was the direct machine parameter optimizer with a maximum of 80 iterations and 20 iterations before conversion. The VMAT plans were completed by the senior radiation dosimetrists. All the plans were tweaked repeatedly in a trial-and-error manner until no significant improvement was found in the subsequent adjustments and optimization.

Herein, we randomly selected 98 patients for training, 10 for validation, and the remaining 22 patients for testing. Each 3D CT image was sliced into multiple 2D images with a resolution of 512×512 and a thickness of 3mm beforehand. In this manner, the training, validation, and testing samples were increased from 98, 10, and 22 to 14817, 1529, and 3491, respectively. The characteristics of these patients regarding sex and age are summarized in **Table 1**.

Network Architecture

Generative adversarial networks (GANs) (24) have been extensively studied in computer vision over the past few years (25). As the name implies, GAN consists of a generator network and a discriminator network which are typically implemented by two independent neural networks (26). The success of GAN lies in the use of adversarial training. Specifically, in our task, the generator takes CT images as the unique input and outputs the dose distribution maps as well as anatomical structures of PTV, while the discriminator inputs the predicted dose maps and the ideal dose maps, i.e., the dose maps that were manually produced by the senior radiation dosimetrists. The goal of the generator is to produce dose maps that are too realistic to be differentiated by the discriminator, while the discriminator is trained to distinguish the generated dose maps from the real ones (also regarded as ground truth (GT)). After repeated iterations, both networks will reach Nash equilibrium, and the generator will eventually be capable of generating realistic dose maps.

Generator Network

The generator is based on a U-net-like encoder-decoder architecture to integrate the shallow and deep features (27). As shown in **Figure 1**, it comprises of two tasks, i.e., the primary dose prediction task and the auxiliary tumor segmentation task, which forms the multi-task learning (MTL) architecture. To be specific, the two tasks share the same encoder and have independent decoders. The shared encoder is harnessed for semantic feature extraction from the input CT images. The dose prediction decoder aims at generating high-quality dose maps. The tumor segmentation decoder is to provide the essential anatomic structure for the dose prediction task. It is linked to the encoder and is trained synchronously with the dose prediction decoder. Particularly, we employ the self-attention (SA) module (see **Supplementary Figure A1**) in the tumor segmentation decoder, which is conducive to obtaining more precise tumor segmentation results. Each SA module outputs an attention map indicating the anatomic information (i.e., the location and shape structure) of the tumor. We transfer these attention maps to the dose prediction decoder to provide vital anatomic information of the target tumor for the primary dose prediction task. Moreover, given that an ideal dose distribution map ought to prescribe a high dose coverage in the tumor area while minimizing the dose prescription to other healthy tissues, we further design a feature decoupling (FD) module (see **Supplementary Figure A2**) in the dose prediction decoder to decouple the high- and low-dose features for the target tumor and OARs, respectively, making the model focus on the tumor and OARs separately, so that more semantically explicit dose predictions can be generated.

Discriminator Network

The discriminator takes the predicted and the real dose distribution maps as input and tries to discriminate them

TABLE 1 | Patient characteristics.

Characteristic		Entire Cohort (n = 130)
Sex	Male	87
	Female	43
Age	Median (IQR)	57
	Range	29-79
	≤40y	6
	40-60y	64
	≥60y	60

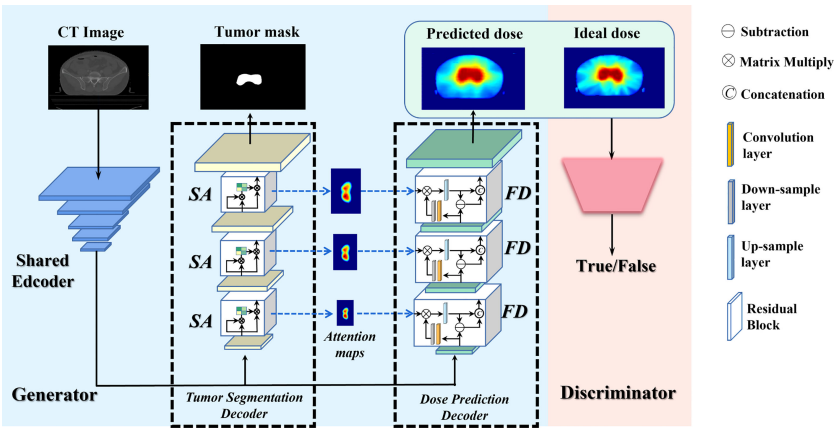


FIGURE 1 | Architecture of the proposed CT-only automatic dose prediction model.

correctly, which further promotes the authenticity of the predicted dose distribution map. Herein, the discriminator is implemented with a typical ResNet18 (28) architecture, which consists of seventeen convolutional layers and a fully connected layer. Each convolutional layer is followed by the BatchNormalization and ReLU activation.

Loss Function and Network Training

The whole network is trained and optimized by a comprehensive loss function which mainly comprises two parts: a generator loss L_G and a standard discriminator loss L_D. The generator loss is further decomposed into two task-specific losses, i.e., a binary cross entropy loss for the tumor segmentation task, and a mean square error loss for the primary prediction task. More detailed description of the loss function is given in the **Supplementary Material**.

The proposed network was implemented using Pytorch and trained on a NVIDIA GeForce RTX 3090 GPU with 24GB memory. The batch size was set to 20. The learning rate was initialized to 5E-4 and decayed to 5E-5 linearly. The generator and the discriminator were trained alternatively using SGD optimizer for 150 epochs. To provide rich anatomic information for dose prediction task, we optimized the segmentation performance at the initial training stage by adjusting the hyper-parameters.

Evaluation

The model with the best performance in the validation was selected for the final test. According to clinical requirements and suggestions of oncologists and dosimetrists, we adopted metrics for both PTV and OARs dose coverage, including D95 (8), average dose (Dmean), maximum dose (Dmax), conformity index (CI) (29), and homogeneity index (HI) (3). The dose-volume-histogram (DVH) curves were plotted to display the disparity between the real dose maps and the predicted dose maps intuitively. To prove the necessity and effectiveness of the key components of our method (i.e., SA module and FD module), we conducted experiments to visualize attention features of SA and FD modules. In addition, we calculated the average prediction error Δ ($\Delta = \frac{1}{n}\sum_{i=1}^{n} \frac{|Prediction_i - GT_i|}{D_{pi}}$, where D_{pi}, $Prediction_i$ and GT_i denote the prescribed dose, the predicted dose, and the clinically approved dose of i-th patient, respectively) of D95 and Dmean, denoted as ΔD95 and ΔDmean, to quantify the disparities between the predicted dose maps and the ground truths. Moreover, to verify the superiority of our prediction, we compared our proposed method with four mainstream methods, including 3D DoseNet (3), DeepLabV3+ (7), U-net (29), and GAN (16). For fair comparison, the inputs of all methods were set to be consistent, i.e., only the CT images. To further verify the clinical potential of our proposed method, we compared our predicted dose maps to DoseUnet (30), the inputs of which include both the CT images and contours. Finally, we evaluated the performance on the entire testing cohort and gave a visual summary with respect to Dmax and Dmin of ROIs in two box plots.

RESULTS

Comparison With the State-of-the-Art Methods

To justify the feasibility and superiority of our method, two qualitative examples predicted by our proposed method are presented and compared with the four state-of-the-art dose prediction methods in **Figure 2**. Although all of these models produce visually sound predictions, our method yields the best effect with an obvious scattering shape and a minimum error near the PTV. To provide more intuitive illustrations, we also calculate and display the difference maps. As shown in the second and fourth rows in **Figure 2**, the difference between the predicted dose of our method and the ground truth one in the clinic is the smallest with the lightest color for both PTV and OARs, indicating that our model

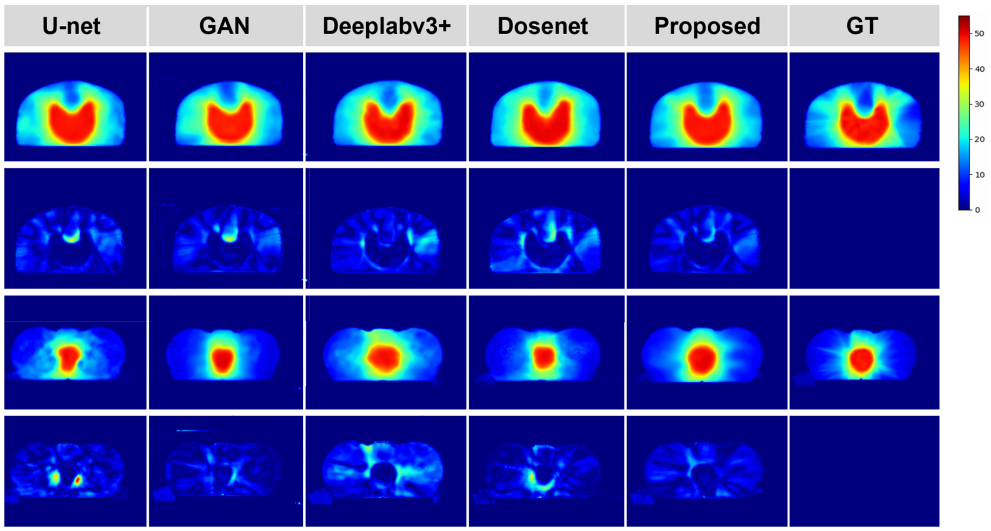


FIGURE 2 | Qualitative comparison between the predicted dose distributions for the proposed and four mainstream methods. The left four columns are the dose prediction results of the comparison methods, and the right two columns are the predicted dose distributions by our method and the ground truth, respectively. The second and fourth rows of the left five columns are the difference maps calculated by subtracting the ground truth distribution map from the predicted one.

could deliver the appropriate dose not only to the target tumor but also to the OARs.

To further demonstrate the effectiveness of our method, we present a set of qualitative examples produced by our model in **Figure 3** and compare them with their corresponding ground truth. According to the dose difference map displayed in the third column, it is evident that our CT-only method could produce dose distribution close to the ground truth in all cases, especially in the target tumor area. Besides, three typical DVH examples are displayed in **Figure 4**. The differences between the predicted dose curves and the ground truth ones are subtle, especially for the PTV whose disparities are minimal.

In Addition, a visual summary of our predicted dose and the clinically acceptable dose with respect to Dmax and Dmean of ROIs are illustrated in **Figure 5**. By comparing the respective median and data dispersion of each box plot, one can see that our

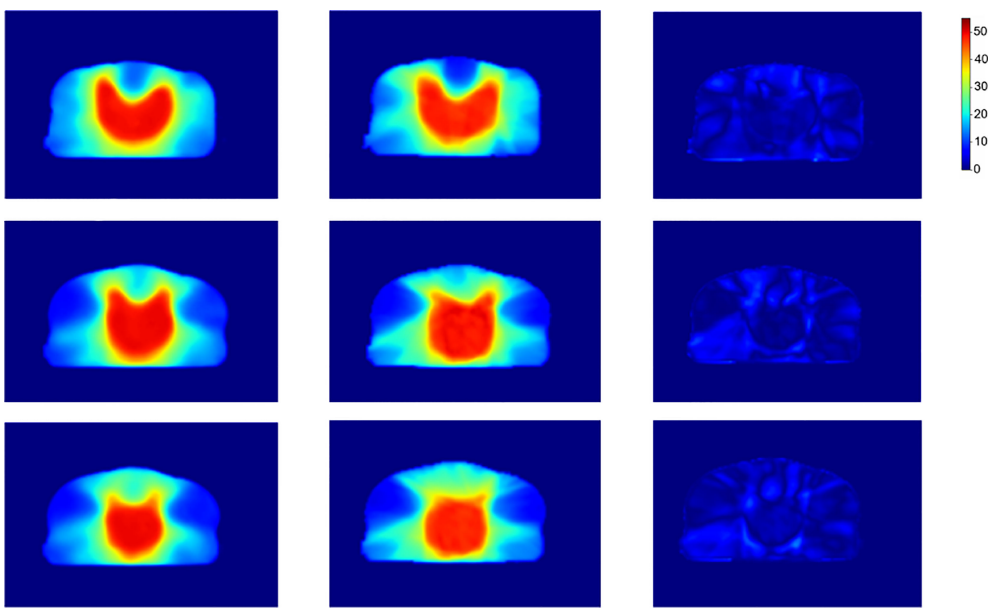


FIGURE 3 | Dose distribution of our proposed method, the ground truth and the corresponding difference. From left to right shows the slices predicted by our method, the corresponding slices of the ground truth, and the dose difference, respectively.

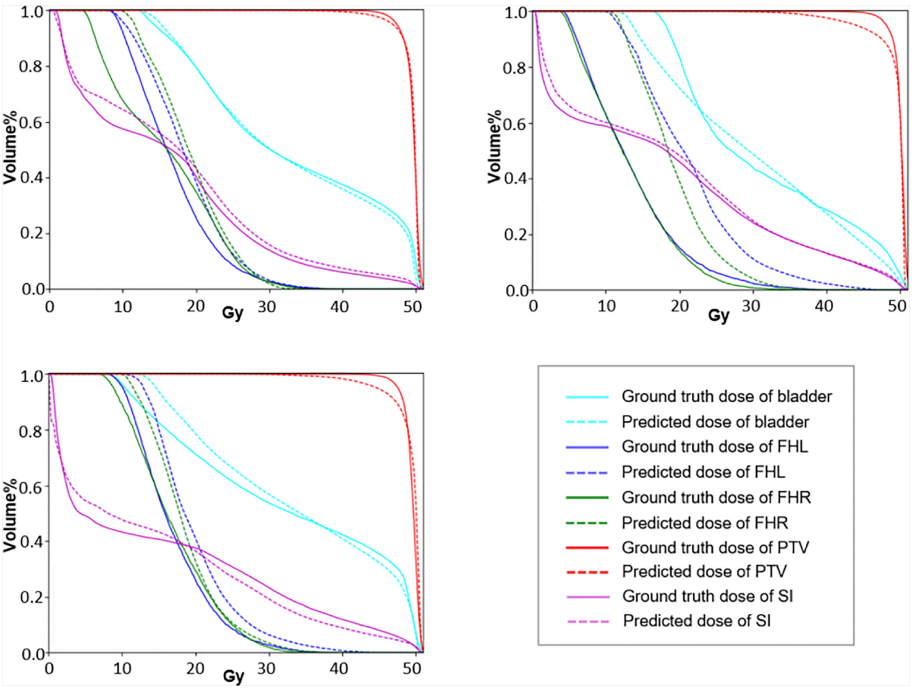


FIGURE 4 | The DVH curves of three typical predictions. The dotted lines represent the prediction results and the solid lines represent the approved values.

predicted dose maps share the same dose distribution with the clinically approved ones in terms of Dmax and Dmean. Specifically, our predicted results had the same skewness with the approved dose in Dmax with respect to FHL and FHR.

The results of quantitative comparison are given in **Table 2**. Our method outperforms other mainstream dose prediction methods by achieving the best results in three out of four metrics with the highest HI of 1.023 (3.27E-5) and the lowest prediction error, i.e., ΔD95 of 0.125 (0.035) and ΔDmean of 0.023 (4.19E-4), respectively. Particularly, with the best HI index, our method is capable of producing homogeneous dose distribution in PTV, and the lower errors in D95 and Dmean also meet the clinical expectation. In addition, the paired t-test was conducted, and the p-values are less than 0.05 in most of the evaluation metrics, demonstrating the significant improvements of our method. Moreover, we also compared our proposed method with DoseUnet which is aided by the additional tumor and OARs contours. The results are shown in the fifth row of **Table 2**. In terms of HI and ΔDmean, our method achieves comparable performance with DoseUnet.

Ablation Study

The key components of our method include 1) the auxiliary tumor segmentation task (Aux), 2) SA module, 3) discriminator (Disc),

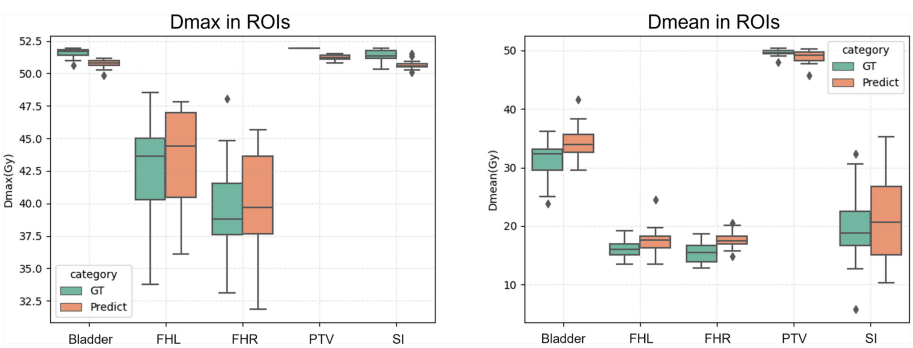


FIGURE 5 | Dmax (left) and Dmean (right) of the proposed prediction and the ground truth with respect to ROIs. Horizontal lines in boxes are medians and rhombuses are outliers.

TABLE 2 | Quantitative comparisons with four mainstream dose prediction methods in terms of HI, CI, D95, and Dmean.

Method	HI	CI	Average prediction error ↓	
			ΔD_{95}	ΔD_{mean}
U-net (26)	1.013 (4.41E-6)*	0.598 (0.006)*	0.301 (0.074)*	0.044 (1.12E-3)*
DeepLabV3+ (7)	1.022 (7.53E-6)	0.593 (0.005)*	0.269 (0.048)*	0.038 (1.16E-3)*
DoseNet (3D) (3)	1.019 (9.68E-6)*	0.592 (0.009)*	0.211 (0.055)*	0.035 (1.11E-3)
GAN (16)	1.016 (2.86E-5)*	0.626 (0.007)	0.204 (0.061)*	0.038 (8.35E-4)*
DoseUnet (30)	1.013 (3.82E-5)	0.736 (0.006) ¶	0.071 (0.047) ¶	0.027 (6.00E-3)
Proposed	1.023 (3.27E-5)¶	0.624 (0.009)	0.125 (0.035)	0.023 (4.19E-4)¶

The HI, CI, ΔD_{95}, and ΔD_{mean} are displayed in the form of mean (variance). The ground truth of HI and CI are 1.030 and 0.773, respectively. Please refer to Evaluation section for more details of the definition.

**Our method is significantly better than the compared ones, i.e., $p < 0.05$ via paired t-test.*

¶The best results of each index.

↓The lower the average prediction error is, the better the dose prediction result is.

and 4) FD module. To investigate the contributions of these components, we further conducted a series of ablation experiments with following variants: (1) U-net with residual blocks embedded alone (i.e., Baseline), (2) Residual blocks embedded U-net with the auxiliary segmentation module (i.e., Baseline+Aux), (3) The network in (2) with SA module embedded (i.e., Baseline+Aux+SA), (4) The network in (3) with the discriminator (i.e., Baseline+Aux+SA+Disc), (5) The network in (4) with FD module injected (Baseline+Aux+SA +Disc+FD, proposed). The quantitative results are given in **Table 3**. We can clearly see that the performance of the network improves progressively as each component is added to the Baseline framework.

To further investigate the effectiveness of our SA and FD modules, we visualize the attention-weighted feature maps for these two modules in both the segmentation decoder and the dose prediction decoder. The results are shown in **Figure 6**. After the SA module, most attention is delivered to the tumor region, thus demonstrating the positive effect of this module on tumor segmentation. In addition, the third and the fourth columns, i.e., FD (high) and FD (low), focus on the tumor area and the opposite healthy tissues, respectively, proving the capability of our network in decoupling high- and low-dose features.

DISCUSSION

In this paper, we undertook an exploratory study to investigate the feasibility of producing high-quality radiotherapy dose predictions *via* only the original CT images, i.e., being free of manual delineation of PTV and OARs.

Two main steps in radiotherapy planning that require considerable manpower and resources are manual contour delineation and repeated parameter tweaking (7). A large number of deep-learning-based dose prediction models have achieved promising results with the assistance of manual delineations of PTV and OARs. The predicted high-quality dose distribution maps can be introduced as a guidance tool and an initial point to improve the efficiency of the inverse parameter adjustment. However, researches indicated that the average time for manual contouring can be up to 3 hours, which may lead to a delay in the treatment and induce errors in tumor localization (16). Current deep-learning-based methods pay little attention on reducing the time spent for manual delineation. In view of this, we boldly suggested predicting the dose maps based solely on the original CT images to further reduce the total planning time by incorporating automate delineation. Concretely, we proposed a GAN-based model and adopted the MTL strategy to learn the missing anatomic contours of the tumor, thereby guiding the primary dose map prediction task.

Our CT-only method was trained on a cohort of 130 rectal cancer patients and evaluated on metrics of D95, Dmean, Dmax, CI, and HI to study the dosimetric congruence between the predicted dose distributions and the approved ones. DVH curves were also plotted for comparison between the approved dose maps and our predicted dose maps. We have compared our method with four mainstream dose prediction methods with contours input removed in **Table 2**. As observed, compared to the widely used U-net model, our method considerably improves the prediction with HI and CI rising from 1.013 (4.41E-6) to 1.023 (3.27E-5), and from 0.598 (0.006) to 0.624 (0.009) respectively. Meanwhile, one can see that the prediction errors

TABLE 3 | Ablation studies of our propose method with its variants.

Method	DSC↑	HI	CI	Average prediction error ↓	
				ΔD_{95}	ΔD_{mean}
(1) Baseline	–	1.013 (1.60E-5)*	0.615 (0.008)	0.238 (0.052)*	0.040 (1.12E-3)*
(2) Baseline+Aux	0.802 (0.002)*	1.019 (1.18E-5)*	0.584 (0.008)*	0.208 (0.051)*	0.038 (8.35E-4)*
(3) Baseline+Aux+SA	0.809 (0.003)*	1.017 (2.75E-5)*	0.625 (0.010)	0.161 (0.032)	0.033 (5.74E-4)*
(4) Baseline+Aux+SA+Disc	0.815 (0.003)	1.020 (9.65E-6)*	0.629 (0.008)¶	0.162 (0.038)	0.031 (5.36E-4)*
(5) Baseline+Aux+SA+Disc+FD (Proposed)	0.816 (0.003)¶	1.023 (3.27E-5)¶	0.624 (0.009)	0.125 (0.035) ¶	0.023 (4.19E-4)¶

↑The higher the DSC is, the better the dose prediction result is.

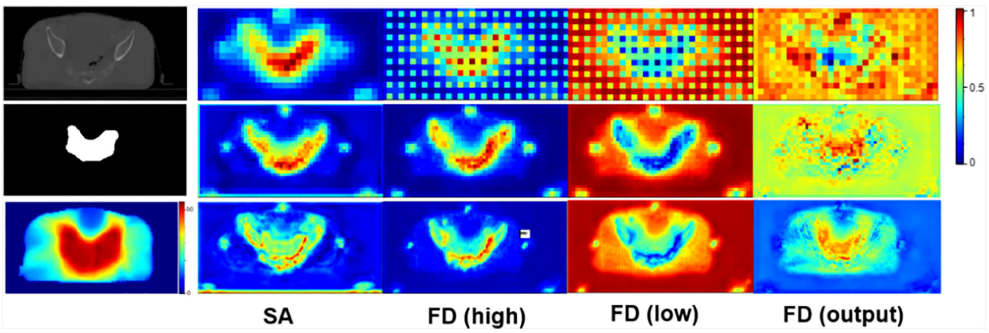


FIGURE 6 | The attention-weighted feature maps for SA and FD. The first column from top to bottom shows the input CT image, tumor segmentation and dose distribution map separately. The second column visualizes the attention of SA module in tumor segmentation task. The following three columns illustrate the attention of FD module in dose prediction task. Specifically, FD (high) and FD (low) denote high- and low-dose features, respectively, and the refined output of FD is marked as FD (output). The redder the area is, the more attention the network pays. The top row refers to the results of the shallowest layer while the bottom row stands for the results of the deepest layer.

ΔD95 and ΔDmean drop from 0.301 (0.074) to 0.125 (0.035), and from 0.044 (1.12E-3) to 0.023 (4.19E-4), respectively. Additionally, compared to DeepLabV3+, our method also improves CI, ΔD95, and ΔDmean by 0.031, 0.144, and 0.015 respectively. Furthermore, the proposed method still surpasses GAN by 0.007 HI, 0.079 ΔD95, and 0.015 ΔDmean, respectively. Although not achieving the best value, the CI acquired by our method (i.e., 0.624) approximates the optimal one (i.e., 0.626) with a minor and tolerable difference of 0.002. Besides, our proposed model could achieve a general minimal difference between the whole testing cohort and the ground truth with respect to Dmax and Dmean of ROIs, as shown in **Figure 5**, manifesting its potential clinical application. To further validate the practicability of our proposed method, we compared our predicted dose maps with those of a contour-aided method, i.e., DoseUnet. As illustrated in **Table 2**, we noticed that the proposed CT-only dose prediction method could achieve comparable results to DoseUnet on both HI and ΔDmean metrics. As for the other metrics, for example, ΔD95 and CI, we must admit that blocking the prior accurate anatomical information in the input will inevitably bring performance degradation. However, this does not mean that our method loses its clinical significances and values. Actually, studies show that the time of contouring is usually longer than that of dose calculation in clinical practice (5), so the saving time in manual delineation brought by our work would be longer than the added tuning time in case of little performance decrease. We will make quantitative comparison when conditions permit.

To investigate the contribution of each module in the proposed network, we conducted a series of ablation study in **Table 3**. Firstly, by comparing (1) and (2), we demonstrated the ability of the tumor segmentation task in providing essential anatomical information for the dose prediction task. As can be seen, with the assistance of the auxiliary task, the HI, ΔD95, and ΔDmean are improved by 0.006, 0.03, and 0.002, respectively. Notably, given the similarities between the SA and FD (high), we can see that the anatomic information obtained in the tumor segmentation task has been successfully transferred to the dose prediction task as guidance. Secondly, we validated the potency of the embedded SA module by comparing (2) and (3). It's clear that the proposed SA module improves the dose prediction performance by 0.041 CI, 0.047 ΔD95 and 0.005 ΔDmean, respectively. Thirdly, in order to verify the contribution of the discriminator in the dose prediction task, we compared (3) and (4). After adding the discriminator, there are also more or less improvements in DSC, HI, CI and ΔDmean. Finally, we compared (4) and the complete model (5) to verify the usefulness of the proposed FD module.

Furthermore, we visualized the attention maps of FD (high) and FD (low) in **Figure 6**. Accordingly, the opposite attention of FD (high) and FD (low) demonstrated the success of our method in decoupling high- and low-dose features. Meanwhile, the increasing tendency of attention in both high- and low-dose areas, as shown in the output of FD, i.e., FD (output), further manifested the capability of our method in paying attention to the high dose distribution in the tumor region while not neglecting the low dose distribution in the surrounding healthy tissues. This simultaneous attention on both the tumor and the surrounding tissues contributed to a more accurate radiotherapy planning.

Despite the superior performance of our method, there are still some limitations. Firstly, our current method employs 2D images due to limited computational resource, which may ignore the geometry information of the organs. In the future, we will extend our dose prediction method to a 3D model to improve the exactness of the anatomic information. Secondly, the auxiliary segmentation task only considers tumor segmentation. However, the anatomic information of the surrounding organs, i.e., OARs, is also of great clinical significance and is critical for the dose prediction. On this basis, we will extend our single-target segmentation model to a multi-target one (i.e., implementing the segmentation on every organ in the CT image), thus providing organ-specific radiation dose prediction. Thirdly, the sub-optimal quality of the generated contours may impede the generation of realistic dose distributions. Given the possible inaccuracy that lies in the essential guiding contours, we will search for a more accurate contour segmentation method in the future.

It is worth mentioning that the generated dose map is regarded as a guidance tool for dosimetrists to reduce the manual intervention and lead a more accurate and speedy treatment planning rather than to provide the ultimate solutions, i.e., the final treatment plan to be performed in the clinic. In other words, the treatment planners still need to adjust the radiotherapy parameters manually but with much reduced time and effort. The goal of our method is to quicken the radiotherapy planning by providing high-quality dose distribution map, which not only brings an initial point close to the optimal plan (instead of presetting the parameters empirically) but also presents a possible optimal target for parameter adjustment.

CONCLUSION

This exploratory study proves the feasibility of predicting high-quality dose distribution with only CT images. Albeit omitting the manual delineations of critical organs, this model introduces a well-designed MTL strategy to make up for the missing anatomic information. Experimental results manifest the capability of the proposed CT-only dose prediction model in producing more realistic dose predictions for rectal cancer radiation therapy compared to the mainstream contour-aided ones.

DATA AVAILABILITY STATEMENT

The datasets generated and/or analyzed during the current study are not publicly available due to data security but are available from the corresponding author on reasonable request.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by West China Hospital, Sichuan University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

ZJ and YW contributed to the conceptualization of this project. ZJ contributed to the investigation, methodology, and software. JX and XP provided the resources and data. JC drafted and ZJ, YW, and JX reviewed and edited the manuscript. JX, ZW, XH, and XP provided expert clinical knowledge. YW supervised and administrated the project. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by National Natural Science Foundation of China (NSFC 62071314) and Sichuan Science and Technology Program (2021YFG0326, 2020YFG0079).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.875661/full#supplementary-material

REFERENCES

1. Hong S, Won YJ, Park YR, Jung KW, Kong HJ, Lee ES. Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2017. *Cancer Res Treat* (2020) 52(2):335. doi: 10.4143/crt.2020.206
2. Baskar R, Lee KA, Yeo R, Yeoh KW. Cancer and Radiation Therapy: Current Advances and Future Directions. *Int J Med Sci* (2012) 9(3):193. doi: 10.7150/ijms.3635
3. Kearney V, Chan JW, Haaf S, Descovich M, Solberg TD. DoseNet: A Volumetric Dose Prediction Algorithm Using 3D Fully-Convolutional Neural Networks. *Phys Med Biol* (2018) 63(23):235022. doi: 10.1088/1361-6560/aaef74
4. Ahn SH, Kim E, Kim C, Cheon W, Kim M, Lee SB, et al. Deep Learning Method for Prediction of Patient-Specific Dose Distribution in Breast Cancer. *Radiat Oncol* (2021) 16(1):1–3. doi: 10.1186/s13014-021-01864-9
5. Das IJ, Moskvin V, Johnstone PA. Analysis of Treatment Planning Time Among Systems and Planners for Intensity-Modulated Radiation Therapy. *J Am Coll Radiol* (2009) 6(7):514–7. doi: 10.1016/j.jacr.2008.12.013
6. Nguyen D, Jia X, Sher D, Lin MH, Iqbal Z, Liu H, et al. 3D Radiotherapy Dose Prediction on Head and Neck Cancer Patients With a Hierarchically Densely Connected U-Net Deep Learning Architecture. *Phys Med Biol* (2019) 64 (6):065020. doi: 10.1088/1361-6560/ab039b
7. Song Y, Hu J, Liu Y, Hu H, Huang Y, Bai S, et al. Dose Prediction Using a Deep Neural Network for Accelerated Planning of Rectal Cancer Radiotherapy. *Radiot Oncol* (2020) 149:111–6. doi: 10.1016/j.radonc.2020.05.005
8. Mardani M, Dong P, Xing L. Deep-Learning Based Prediction of Achievable Dose for Personalizing Inverse Treatment Planning. *Int J Radiat Oncol Biol Phys* (2016) 96(2):E419–20. doi: 10.1016/j.ijrobp.2016.06.1685
9. Zhong Y, Yu L, Zhao J, Fang Y, Yang Y, Wu Z, et al. Clinical Implementation of Automated Treatment Planning for Rectum Intensity-Modulated Radiotherapy Using Voxel-Based Dose Prediction and Post-Optimization Strategies. *Front Oncol* (2021) 11:2498. doi: 10.3389/fonc.2021.697995
10. Zeng J, Cao C, Peng X, Xiao J, Zu C, Zhou J, et al. March. Two-Phase Progressive Deep Transfer Learning for Cervical Cancer Dose Map Prediction. In: 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI). *IEEE* (2022) 1–5. doi: 10.1109/ISBI52829.2022.9761628.
11. Wang Y, Zhou L, Yu B, Wang L, Zu C, Lalush DS, et al. 3D Auto-Context-Based Locality Adaptive Multi-Modality GANs for PET Synthesis. *IEEE Trans Med Imaging* (2018) 38(6):1328–39.. doi: 10.1109/TMI.2018.2884053
12. Wang Y, Yu B, Wang L, Zu C, Lalush DS, Lin W, et al. 3D Conditional Generative Adversarial Networks for High-Quality PET Image Estimation at Low Dose. *Neuroimage* (2018) 38:550–62. doi: 10.1016/j.neuroimage.2018.03.045.
13. Luo Y, Zhou L, Zhan B, Fei Y, Zhou J, Wang Y, et al. Adaptive rectification based adversarial network with spectrum constraint for high-quality PET image synthesis. *Med Image Anal* (2022) 77:102335. doi: 10.1016/j.media.2021.102335
14. Shi Y, Zu C, Hong M, Zhou L, Wang L, Wu X, et al. ASMFS: Adaptive-Similarity-Based Multi-Modality Feature Selection for Classification of Alzheimer's Disease. *Pattern Recognit* (2022) 126:108566. doi: 10.1016/j.patcog.2022.108566.
15. Wang K, Zhan B, Zu C, Wu X, Zhou J, Zhou L, et al. Semi-supervised medical image segmentation via a tripled-uncertainty guided mean teacher model with

contrastive learning. *Med Image Anal* (2022) 79:102447. doi: 10.1016/j.media.2022.102447.
16. Murakami Y, Magome T, Matsumoto K, Sato T, Yoshioka Y, Oguchi M. Fully Automated Dose Prediction Using Generative Adversarial Networks in Prostate Cancer Patients. *PloS One* (2020) 15(5):e0232697. doi: 10.1371/journal.pone.0232697
17. Kearney V, Chan JW, Wang T, Perry A, Descovich M, Morin O, et al. DoseGAN: A Generative Adversarial Network for Synthetic Dose Prediction Using Attention-Gated Discrimination and Generation. *Sci Rep* (2020) 10 (1):1–8. doi: 10.1038/s41598-020-68062-7
18. Cao C, Xiao J, Zhan B, Peng X, Wu X, Zhou J, et al. Adaptive Multi-Organ Loss Based Generative Adversarial Network for Automatic Dose Prediction in Radiotherapy. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*. (Nice, France:IEEE) (2021). 212–5.
19. Zhan B, Xiao J, Cao C, Peng X, Zu C, Zhou J, et al. Multi-Constraint Generative Adversarial Network for Dose Prediction in Radiotherapy. *Med Image Anal* (2022) 77:102339. doi: 10.1016/j.media.2021.102339.
20. Li H, Peng X, Zeng J, Xiao J, Nie D, Zu C, et al. Explainable Attention Guided Adversarial Deep Network for 3D Radiotherapy Dose Distribution Prediction. *Knowledge Based Syst* (2022) 241:108324. doi: 10.1016/j.knosys.2022.108324.
21. Lutkenhaus LJ, de Jong R, Geijsen ED, Visser J, van Wieringen N, Bel A. Potential Dosimetric Benefit of an Adaptive Plan Selection Strategy for Short-Course Radiotherapy in Rectal Cancer Patients. *Radiot Oncol* (2016) 119 (3):525–30. doi: 10.1016/j.radonc.2016.04.018
22. Myerson RJ, Garofalo MC, El Naqa I, Abrams RA, Apte A, Bosch WR, et al. Elective Clinical Target Volumes for Conformal Therapy in Anorectal Cancer: An RTOG Consensus Panel Contouring Atlas. *Int J Radiat Oncol Biol Phys* (2009) 74(3):824. doi: 10.1016/j.ijrobp.2008.08.070
23. Hong TS, Moughan J, Garofalo MC, Bendell J, Berger AC, Oldenburg NB, et al. NRG Oncology Radiation Therapy Oncology Group 0822: A Phase 2 Study of Preoperative Chemoradiation Therapy Using Intensity Modulated Radiation Therapy in Combination With Capecitabine and Oxaliplatin for Patients With Locally Advanced Rectal Cancer. *Int J Radiat Oncol Biol Phys* (2015) 93(1):29–36. doi: 10.1016/j.ijrobp.2015.05.005
24. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Nets. *Adv Neural Inf Process Syst* (2014) 27:2672–80. doi: 10.48550/arXiv.1406.2661
25. Pan Z, Yu W, Yi X, Khan A, Yuan F, Zheng Y. Recent Progress on Generative Adversarial Networks (GANs): A Survey. *IEEE Access* (2019) 7:36322–33. doi: 10.1109/ACCESS.2019.2905015
26. Kazemifar S, McGuire S, Timmerman R, Wardak Z, Nguyen D, Park Y, et al. MRI-Only Brain Radiotherapy: Assessing the Dosimetric Accuracy of Synthetic CT Images Generated Using a Deep Learning Approach. *Radiot Oncol* (2019) 136:56–63. doi: 10.1016/j.radonc.2019.03.026
27. 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: Springer (2015). p. 234–41.
28. 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*. Las Vegas, USA: IEEE (2016). p. 770–8.
29. Paddick I. A Simple Scoring Ratio to Index the Conformity of Radiosurgical Treatment Plans. *J Neurosurg* (2000) 93(supplement_3):219–22. doi: 10.3171/jns.2000.93.supplement_3.0219
30. Nguyen D, Long T, Jia X, Lu W, Gu X, Iqbal Z, et al. A Feasibility Study for Predicting Optimal Radiation Therapy Dose Distributions of Prostate Cancer Patients From Patient Anatomy Using Deep Learning. *Sci Rep* (2019) 9(1):1–10. doi: 10.1038/s41598-018-37741-x.

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 | Frontiers in Oncology

REVIEW
published: 09 August 2022
doi: 10.3389/fonc.2022.913683

Review of Radiomics- and Dosiomics-based Predicting Models for Rectal Cancer

Yun Qin[1], *Li-Hua Zhu*[1], *Wei Zhao*[1], *Jun-Jie Wang*[2*] and *Hao Wang*[2,3*]

[1] School of Physics, Beihang University, Beijing, China, [2] Department of Radiation Oncology, Peking University Third Hospital, Beijing, China, [3] Cancer Center, Peking University Third Hospital, Beijing, China

By breaking the traditional medical image analysis framework, precision medicine–radiomics has attracted much attention in the past decade. The use of various mathematical algorithms offers radiomics the ability to extract vast amounts of detailed features from medical images for quantitative analysis and analyzes the confidential information related to the tumor in the image, which can establish valuable disease diagnosis and prognosis models to support personalized clinical decisions. This article summarizes the application of radiomics and dosiomics in radiation oncology. We focus on the application of radiomics in locally advanced rectal cancer and also summarize the latest research progress of dosiomics in radiation tumors to provide ideas for the treatment of future related diseases, especially ^{125}I CT-guided radioactive seed implant brachytherapy.

Keywords: radiomics, dosiomics, machine learning, deep learning, rectal cancer

OPEN ACCESS

Edited by:
Savino Cilla,
Gemalli Molise Hospital, Italy

Reviewed by:
Sunyoung Jang,
Princeton Radiation Oncology Center, United States
Michele Avanzo,
Aviano Oncology Reference Center (IRCCS), Italy

***Correspondence:**
Jun-Jie Wang
junjiewang@pku.edu.cn
Hao Wang
hhbysy@126.com

Specialty section:
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

Received: *06 April 2022*
Accepted: *23 June 2022*
Published: *09 August 2022*

Citation:
Qin Y, Zhu L-H, Zhao W, Wang J-J and Wang H (2022) Review of Radiomics- and Dosiomics-based Predicting Models for Rectal Cancer. Front. Oncol. 12:913683. doi: 10.3389/fonc.2022.913683

INTRODUCTION

According to the latest global cancer data for 2020 released by the International Agency for Research on Cancer of the World Health Organization, colorectal cancer ranks third and second in the global morbidity and mortality rates, respectively (1). In 2020, there were estimated 43,340 cases of rectal cancer diagnosed in the United States (2). Therefore, early detection of rectal cancer and precise treatment reducing the incidence, recurrence rate, and mortality rate of rectal cancer are very critical. There are many ways to treat rectal cancer, and in order to determine the best treatment regimen and optimize patient outcomes, efficient imaging biomarkers are needed to contribute to cancer detection, diagnosis, the choice of therapeutic strategy, prognosis inference, the prediction of response, and surveillance (3).

With the rapid development of imaging technology, medical imaging including computed tomography (CT), magnetic resonance imaging (MRI), or positron-emission tomography (PET) images play a significant role in clinical applications, particularly in cancer prognosis. Radiographic imaging technology mainly evaluates the grade of morphology of the tumor and its surrounding environment. However, it is difficult to convert the microheterogeneity and biological characteristics of the tumor into a quantitative mode (4). On the other hand, visual analysis is insufficient to capture the deep information of the lesions; thus, it cannot meet the requirement of accurate medicine and personalized treatment.

In recent years, advances in the use of artificial intelligence (AI) and computing methods in medical image processing and analysis transformed these images into quantitative data (5). Due to radiomics covering almost all solid tumors, it has been widely applied in oncology. Radiomics based on high-throughput feature extraction algorithms (manually defined and deep learning) enables the integration of imaging and clinical features to decode information hardly recognized by the naked eye and then use machine learning to model these proposed features, thus improving the efficiency of assessing prognosis and response. The current research model established through deep learning features has gradually been developed, but compared with huge radiomics features, this method-related research is still small (6–8).

The outcome of radiation therapy for tumors is closely related to dose distribution, but simple dose statistics alone cannot make accurate predictions about the outcome of radiation therapy. In response to this problem, the dosiomics approach can describe the dose distribution by the dose characteristics of intensity, texture, and shape, with higher accuracy, granularity, and spatial information, and is an effective method for parametric radiotherapy dose distribution (9). At present, there is still a gap in the application of dosiomics in rectal cancer.

Rectal cancer is the most common gastrointestinal malignant tumor in the world. More than 100,000 people are diagnosed with rectal cancer every year, 70% of which are locally advanced (T3–4 or N+) rectal cancer (LARC) (10). Unfortunately, the initial clinical symptoms of rectal cancer are not typical, and many patients are already in the locally advanced stage when they are first diagnosed, with estimated 149,500 new rectal cancer (RC) cases and 52,980 expected deaths in the United States in 2021 (11). In colorectal cancer, rectal cancer incidence is slightly higher than that of colon cancer. Rectal cancer has an insidious early onset, clinical symptoms are atypical, and many patients are already in the local progression stage when they got their first diagnoses. In this paper, we review the research status and progress of radiomics in the differential diagnosis, efficacy evaluation, and prognosis evaluation of LARC in recent years and look forward to the current research status of dosiomics and the future application of prognosis prediction of rectal cancer.

RADIOMICS

"Omic" generally refers to the mining of more parameters or features from the research target as a whole to further enrich the dimensions for research or reference. Similarly, radiomics refers to extracting quantitative information from morphological and functional imaging and is combined with clinical features, protein genome information, and identifying feature subsets related to prognosis through machine learning. Radiomics can be performed for both healthy human tissues and diseased tissues. It is a multidisciplinary and multi-imaging technology. Its primary analysis process is as follows (12): (1) identifying clinical problems; (2) access and process to high-quality standardized medical image data; (3) the segmentation of ROIs (regions of interest); (4) high-throughput radiomics feature extraction; (5) feature selection; and (6) prediction model establishment and statistical analysis. Radiomics enables the rich information in medical images to be fully displayed and provides deep quantitative features that cannot be recognized from visual inspection. Compared with traditional imaging methods, radiomics is a promising alternative, which can provide the basis for treatment plans, curative effects, and prognosis assessment for various diseases. Therefore, every radiomics step in the process is extremely of vital importance and needs to be performed rigorously. A workflow diagram illustrating the radiomics and dosiomics analysis process is shown in **Figure 1**.

Identifying Clinical Problems

At present, various radiomics methods have been explored, aiming at realizing personalized medicine, such as the diagnosis (13–15), treatment response (16–18), and prognosis prediction (19, 20) of tumor cancers. At the stage of model construction, there are many similarities among diagnostic models, mainly faced with the selection of predictive factors (feature selection), the formulation of modeling strategies (algorithms), and the evaluation of the final model performance.

To solve clinical problems, the correlation between models and problems should be explored. Clinical problems determine the radiomic research direction and route, so different problems require different types of research design. Currently, the application of radiomics in rectal cancer is mainly the prognosis prediction of patients with LARC to neoadjuvant chemoradiotherapy (NCRT). Although the local recurrence (LR) rate of rectal cancer has been significantly reduced and the disease-free survival (DFS) rate has been significantly increased with the advances of clinical medicine and medical technologies, the LR of rectal cancer and the associated prognostic risk factors are still the major concerns for clinical rectal cancer treatments.

Acquiring and Processing Data

Building an adequate database is the prerequisite for radiomics research. This is due to insufficient data capacity, which may reduce the model prediction accuracy and increases overfitting risk. The collected data are supposed to contain high-quality, standardized medical images and necessary clinical information, such as, pathological data and biological and genomic medical records. Commonly used medical images include CT, MRI, and PET/CT. Currently, rectal MRI is the preferred imaging modality for the local staging of rectal cancer. MRI is the superior imaging modality for the evaluation of primary tumor location, extension, and mesorectal fascia involvement (21) and is considered as the standard for the evaluation and staging of rectal cancer. For patients with local recurrent rectal cancer, most of the intraluminal recurrent tumors are diagnosed by rectal examination or direct visualization on rectosigmoidoscopy and MRI is the most accurate imaging method to detect and identify patients with extravascular recurrence (22), but its cost limits its application in routine follow-up; thus, usually, pelvic CT examination is performed after rectal cancer surgery. High-

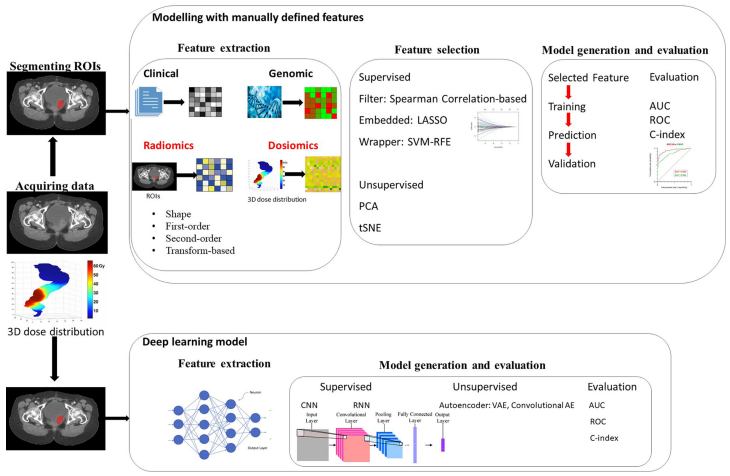


FIGURE 1 | Workflow for radiomics and dosiomics analysis with feature-based (machine learning) and featureless (deep learning) approaches.

spatial-resolution T2 weighted imaging is the most important MRI sequence in the evaluation of rectal cancer and anatomic structures (23).

High quality and standardization of medical images is vital for feature extraction and quantification in radiomics. Ideally, the same scanning machine should be used as far as possible, and reasonable layer thickness, pixel size, tube voltage, and other parameters should be selected to obtain more ideal analysis results. However, this is not the case in many practical situations. Therefore, corrections should be considered in the subsequent analysis and modeling steps. Data preprocessing includes removing artifacts from the images, correcting inhomogeneity, intensity normalization, spatial smoothing, spatial resampling, noise reduction, and MRI field non-uniform registration (rigid, deformable, or mutual intensity algorithm) and reslicing, and so on (9, 24). Otherwise, the extracted features and the generated model will not be reproducible and non-generalizable.

The Image Biomarker Standardization Initiative (IBSI) has defined reporting guidelines that work toward standardizing the extraction of image biomarkers from acquired imaging to high-throughput quantitative image analysis (radiomics). Alex Zwanenburg et al. standardized 169 radiomics features to verify and calibrate different kinds of radiomics software, which will increase the reproducibility of radiomics studies and facilitate the clinical translation of radiomics (25).

Segmenting Regions of Interest

The ROI segmentation uncertainty and time efficiency from the revised image data set are the most critical. There are two main areas of interest in the clinic: the tumor target area and the nearby organs at risk. At present, most researchers extract features from the gross tumor volume (GTV) to build models. The GTV is the position and extent of the primary rectal tumor. The clinical target volume (CTV) describes the extent of microscopic, unimageable tumor spread, and the planning target volume (PTV) allows for uncertainties in planning delivery. Additionally, the normal tissue structures in the vicinity of the target must be considered. The current methods used to segment ROI are mainly divided into automatic segmentation and manual segmentation. Although semi-automatic and fully automatic segmentation software have been widely used in radiomics research, especially for tumors with clear boundaries and regular morphology, the automatic segmentation method is relatively efficient and highly repeatable. It can meet the requirements of massive data segmentation (26). Of course, there are also studies based on the ROI sketched manually by doctors on the radiotherapy planning system. The advantage of manual segmentation is that the accuracy is high, but the results are easily affected by doctors' subjective factors. For instance, interobserver delineation in cervical cancer can lead to significant differences and is reported to differ up to 4 cm (27). The intraclass correlation coefficient (ICC) can be used to reject non-reproducible features. For lesions where the boundary is not easy to detect, manual segmentation can be used; for tumors with clear boundaries and regular morphology, semi-automatic or automatic segmentation methods are efficient and highly reproducible, which can meet the requirements of massive data segmentation (26). Deep learning has been used to segment rectal tumors automatically. Weijun Chen et al. evaluated the results of two automatic contouring softwares (deep learning auto-segmentations and Atlas) on the OAR definition of CT images of lung cancer and rectal cancer patients (28). The results show that deep learning auto-segmentations were better than that of Atlas and can be used clinically. Hai-Tao Zhu et al. proposed a volumetric U-Net model that can automatically segment the rectal tumor region on the diffusion-weighted imaging images of LARC (29).

Accurate image segmentation ROI is the premise of radiomics analysis, and the segmentation algorithm with high accuracy and good repeatability still needs to be further studied. Noise, artifacts, and tumor infiltration to the surrounding normal

tissues often conceal the lesion's real edge, which brings great difficulty to image segmentation. The features extracted depend on the segmented area, rather than unclear or complex tumor boundaries, which can lead to inconsistent and low reproducibility of the results. Therefore, when different methods are selected for target area sketching (manual, semi-automatic, automatic sketching), the selection should be based on the required precision and time. At present, the commonly used segmentation method for many clearer tumor contours is through computer-assisted edge detection and then further manual adjustment. For patients with locally recurring rectal cancer (LRRC) treated with seed implantation, because the tumor structure has been interfered many times in previous treatments (including surgery, external radiotherapy, and seed implantation), the tumor margin merged with the normal tissue structure, resulting in the tumor border not being clear enough. Furthermore, these patients need to consider whether there are blood vessels, bones, or organs in the direction of the needles, so manual segmentation remains the gold standard.

Feature Extraction

The core step of radiomics is to extract features from the ROI and then use them for quantitative analysis.

1. The manually extracted features mainly include 5 types: shape features (30), first-order statistics features or histogram-based features (31), second-order statistics features or textural features (32, 33), transform-based features (34), and some features that are obtained from PET images (SUV value) (35) and are only applicable to the fractal and fusion features of multi-mode images. Published studies are not only based on the radiomics features of PET but also based on the combined radiomics features of CT and MRI for the improvement of the accuracy for the rectal cancer prognostic model.
2. Deep learning features: the features extracted based on deep learning are different from handcrafted radiomics. It directly builds a deep learning model for the entire medical image. This method of extracting features requires a large data set. However, the database for treating LRRC with particle implantation is not enough. The extracted features are usually non-interpretable. At present, the deep learning of medical images usually utilizes convolutional neural networks (CNNs), which are neural networks with automatic feature extractors specially designed for images. Xception, VGG16, VGG19, ResNet50, InceptionV3, and Inception ResNetV2 are six commonly used CNNs (36). Deep learning is relatively new and has tremendous potential waiting to be explored (37).

The deep learning features are able to be combined with other relevant data, including necessary clinical information, pathological data, biological or genomics medical records, dosimetry, and so on to construct a robust model.

Feature Selection

A feature is necessary to screen the imaging features acquired in ROI, make the established model universal, and avoid overfitting (38). Feature selection refers to the dimensionality reduction of many feature data extracted in ROI to obtain features related to the research endpoint.

In general, the methods of feature dimensionality reduction are divided into two categories: supervised and unsupervised. There are three supervised feature selection methods: filter, embedded, and wrapper methods (39).

(1) The filtering method is usually utilized as a preprocessing step, and feature selection is completely independent of any machine learning algorithm. It selects features based on the scores in various statistical tests and various indicators of correlation. Correlation filtering judges the correlation between features and tags. Commonly used methods include the mutual information method, chi-square filtering, F-test, Wilcoxon rank-sum test, the Fisher score, the Student's t-test, and so on (40, 41). The relevance filtering method is a univariate feature selection method, meaning it does not consider the correlation between each feature.

(2) The embedding method is a method that allows the algorithm to decide which features to be used, that is, feature selection and algorithm training are performed at the same time. When using the embedding method, some machine learning algorithms and models are used for training to obtain the weight coefficients of each feature. Compared with the filtering method, the result of the embedding method will be more accurate to the model's utility, which has a better effect on improving the model's effectiveness. Using the embedding method, it is easy to achieve feature selection: reduce the amount of calculation and improve the model's performance. Commonly used embedded methods are ridge regression, tree-based algorithms such as the random forest (RF) classifier, or the least absolute shrinkage and selection operator (40, 41).

(3) The wrapper method is a method of feature selection and algorithm training at the same time. The packaging method often uses an objective function to select the best feature subset instead of inputting a certain evaluation index or statistic threshold. The most typical objective function is the recursive feature elimination method, and some other wrapper methods include forward feature selection, backward feature elimination, exhaustive feature selection, bidirectional search feature selection, or bidirectional search (39, 40). The effect of the wrapper method is the most conducive to improve the model's performance among all the feature selection methods. It can use very few features to achieve excellent results. In addition, when the number of features is the same, the performances of the packaging method and the embedding method are comparable, but it is faster than the embedding method, although its calculation amount is also very large, not suitable for too- large data.

Unsupervised feature selection methods are mainly principal component analysis (PCA) and cluster analysis. Each feature selection method has its strengths and weaknesses, and the performance depends on the type of the data set and the constraints related to the scenario.

Model Generation and Evaluation

After feature extraction, a radiomics model must be established. Many studies use machine learning and deep learning methods to build prediction and classification models, which is currently the mainstream of published studies. Research endpoints usually include but are not limited to disease diagnosis (classification) and prognosis prediction [overall survival (OS), local control (LC), distant metastasis (DM), treatment response]. For example, for the disease of LARC, many studies have used radiomics to evaluate treatment response after neoadjuvant therapy. Other research endpoints are mainly DM, OS, and so on.

The machine learning algorithm models for predicting the treatment response of LARC after neoadjuvant radiotherapy and chemotherapy mainly include logistic regression, RF, and support vector machine (SVM) methods. To predict DM and the survival time of patients after radical resection for rectal cancer, Mou Li et al. used multivariate logistic regression (LR) analysis to establish a combined model of radiomic characteristics (Rad-score) and clinical factors (42).

The workflow of radiomics based on deep learning is essentially different from the workflow described above. Deep learning is based on representation learning in which the algorithm learns the best features to carry out a given task on its own by navigating the provided data (43). In deep learning–based radiomics, different network architectures, such as CNNs or autoencoders, are used to find the most relevant input data features. It does not need to define or select features in advance, but some studies are still related to statistics. Whereas radiomics captures quantitative values of shape and texture based on predefined mathematical terms, neural networks have recently been used to directly learn and identify predictive features from medical images (44). In addition, the process of data representation and prediction is carried out jointly (45). Of course, the features extracted by deep learning methods can be further analyzed and classified by the neural network, or they can leave the network and use different classifiers, such as decision trees, regression models, or support vector machines, to make predictions. If enough patients are included in the cohort, solutions leveraging deep learning should be the preferred method in the years to come (46).

Of course, the AI algorithm is not widely used in feature analysis. In many prognostic analysis cases, statistical methods such as Spearman correlation and univariate and multivariate Cox regression or most machine learning methods (logistic regression) are used for analysis. However, many studies have shown that the predictive performance of AI models is superior to traditional statistics. For example, Quirino Lai et al. conducted a systematic review of the role of AI in the prognosis of liver cancer patients and concluded that AI has a perfect role in clinical research and the application of hepatocellular carcinoma (HCC) (47). The accuracy of the prediction model is related to the sample size and feature parameter selection and also to machine learning algorithms.

Delta-radiomics

There are usually two types of extracted features. One is the single-time-point radiomics, where features are extracted from a particular image (e.g., pretreatment), and the other is delta-radiomics (48). The delta-radiomics feature is defined as the difference in radiomics feature before and after a specific treatment method. They can be calculated between the pretreatment and post-treatment features. With the changes of radiomic features over time in longitudinal images, delta radiomics can potentially be used as a biomarker to predict treatment response and offer abundant information to identify, quantify, and potentially predict therapy-induced changes throughout treatment (49). Radiomics features extracted only before the treatment cannot reflect the overall treatment details, while delta-radiomics can describe the changes in the image during the process, which is relatively more rigorous. The process of delta-radiomics and radiomic research is the same. After extracting the delta-radiomics features from the original ROI, due to the large number of radiomic features extracted from the images, many methods mentioned above are used to rule out redundant delta-radiomic features (DRFs). The selected DRFs are then tested to determine their significance as a treatment response function using linear regression models, t-test, and mixed-effect models (50). Significant DRFs are further tested and modeled using machine-learning algorithms to create a model that can predict the outcome of a new patient.

Some studies investigate the effectiveness of delta-radiomics compared to single-time-point radiomics, and the results indicated that delta features could provide better treatment assessment than single-time-point features. Studies on delta-radiomics have been reported since 2017. These articles have all been published in the last 5 years and are gradually increasing every year. Among the 54 articles screened, a total of 7 (13%) studies reported on the use of delta-radiomics in rectal cancer. A total of 10 (19%) studies referring to the use of delta-radiomics in the prediction among patients with lung cancer were excluded in this review. Other studies using delta-radiomics in grade osteosarcoma and pancreatic and gastric cancer were also excluded. Some studies have shown that delta-radiomics can successfully predict the prognostic response of rectal cancer, including the complete pathological response (pCR), DM, LR, and DFS (51–56). For instance, Seung Hyuck Jeon et al. developed delta-radiomics signatures to predict treatment outcomes after preoperative chemoradiotherapy and surgery in LARC (52). Giuditta Chiloiro et al. used delta-radiomics to investigate the correlation between changes in magnetic resonance imaging (MRI) radiomic characteristics before and after neoadjuvant radiotherapy (NCRT) LARC patients and the 2-year DM rate (53).

RADIOMICS IN RECTAL CANCER

Radiomics, based on advanced pattern recognition tools, has been widely studied for clinical prediction models in diagnosis and treatment prognosis/selection in oncology. In recent years, radiomics has been gradually applied to histopathological grading (57–59), pretreatment staging prediction, differential diagnosis, efficacy evaluation, and prognosis evaluation for

rectal cancer. Many studies have shown that radiomic features can objectively provide texture information related to histopathological and immunohistochemical markers and can non-invasively evaluate biological characteristics such as tumor proliferation, migration, and angiogenesis before treatment. This section will discuss an overview of notable studies about LARC published in this area.

Radiomics Prediction of Locally Advanced Rectal Cancer After Neoadjuvant Chemoradiotherapy

NCRT can reduce the tumor size and recurrence and increase the tumor resection rate and anus retention rate with a very slight side effect (60). Therefore, it is very necessary to refine the selection of appropriate patients and irradiation mode of NCRT. Since most patients with rectal cancer have locally advanced diseases at the time of diagnosis, NCRT is the standard recommendation to improve the prognosis of patients.

However, in the context of precision medicine, it is an urgent problem in clinical work to find a method that can predict the therapeutic effect early and avoid the therapeutic risk, effectively guide the individual treatment, and improve the prognosis of patients to the greatest extent. Therefore, it is crucial to screen LARC patients with therapeutic responses to NCRT. The evaluation of treatment response to NCRT is still challenging. The complete pathologic response (pCR) to NCRT is assessed during the pathological examination after surgery. Identifying patients in pCR with a high accuracy rate could lead to improved clinical outcome (61). Radiomics also has important clinical significance in evaluating NCRT efficacy and the prognosis of rectal cancer. The most commonly used medical images are CT and MRI. In recent years, many studies have predicted and verified the efficacy of NCRT in LARC through radiomics.

According to the WHO, the solid tumor efficacy evaluation criteria are divided into complete remission (complete response, CR): tumor disappears completely, lasting more than 4 weeks; partial remission (partial response, PR): tumor shrinkage ≥50%, lasting more than 4 weeks; stable (stable disease, SD): tumor enlargement <25%, shrinkage <50%; progress [progressive disease (PD)]: tumor increases by more than 25% (62). PR and SD are considered effective. PD means that the treatment is ineffective. Whether a patient has achieved pCR is often determined by postoperative pathological examination, and it is not known whether it is remission before surgery. Many recent studies have found that radiomics can help clinicians predict whether patients will be pCR after NCRT before surgery to avoid excessive treatment and the burden and pain caused by surgery and improve treatment accuracy (61–73). CT-based radiomics has shown promise in LARC. The following table study endpoint is from the four aspects: pCR DFS, downstaging, and distant control. **Table 1** shows the studies of LARC radiomics using CT images in the PubMed database.

An overview of the main studies proposing MRI radiomics for the pCR of NCRT prediction outcomes is reported in **Table 2**. In all studies, three-dimensional (3D) manual segmentation of the primary tumor was performed to extract radiomic features and the performance of the models were calculated by receiver operating characteristic curves.

It was concluded that the features of MRI prior to treatment could effectively predict patients who were unresponsive to NCRT. Therefore, MRI–based radiomics has important clinical significance in the NCRT efficacy evaluation and prognosis evaluation of rectal cancer. This could provide an improved basis for personalized treatment. Certainly, radiomics has a predictive value for the NCRT curative effect of LARC and shows good predictive value in terms of tumor staging, postoperative metastasis, and prognosis after treatment.

DEVELOPMENT AND CHALLENGES OF LOCALLY RECURRENT RECTAL CANCER

Despite advances in surgical techniques and chemoradiation therapy, recurrent rectal cancer remains a cause of morbidity and mortality (74). LRRC is the recurrence of a tumor of the same pathologic nature in the primary tumor after surgery, in the pelvis, in the field of operation. Neoadjuvant therapy and surgical treatment for rectal cancer have been improved and the concept of comprehensive treatment has been promoted, and the survival of patients with rectal cancer has been improved. However, rectal cancer recurrence remains a common clinical problem, and patients generally have a dismal prognosis and a poor quality of life.

LRRC has various treatment methods, including surgery, external beam radiotherapy, intraoperative radiotherapy, Iodine-125 (I-125) seed implantation, heat therapy, and radiofrequency ablation (75). For patients with postoperative recurrence, due to the damage to the normal anatomical structure and the adhesion of recurrent lesions and surrounding tissues, the reoperation resection rate is low. Therefore, the treatment means and effect are not satisfactory. For those who have received external beam radiotherapy, it is difficult to improve the treatment effect due to increasing the local dose. For patients with recurrent rectal cancer after surgical resection or external radiotherapy, LRRC prognosis is poor, while CT-guided ^{125}I seed implantation therapy has become a recommended therapy. The I-125 seed Model 6711 consists of a titanium cylindrical tube with 0.8-mm radius and 4.5-mm length, with an average energy of 28 Kev and a half-life of 59.4 days is commonly used throughout oncology centers worldwide. A dose prescription of 110–160 Gy was considered, with an initial source activity of 0.4–0.7 mCi. The particle spacing is usually 1 cm, which is easier to identify particles. Therefore, permanent 125I seed interstitial brachytherapy is a potential salvage modality because of its unique physical and clinical characteristics. The 2016 National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology has recommended radioactive 125I seed (RIS) implantation for the treatment of LRRC (76).

Many factors affect survival, following the treatment of LRRC. We can also consider the background liver condition, the radiologic and histologic characteristics of the tumor, biologic markers, and comorbidities. Traditionally, conventional linear models, such as the survival analysis and the Cox proportional

TABLE 1 | Summaries of selected locally advanced rectal cancer (LARC) radiomics studies (CT).

Conclusion	Entropy, uniformity, and standard deviation were independent texture features in predicting DFS	Low- and high-risk groups for DFS in the training set ([HR] 56.83; P < 0.001) in the validation set (HR52.92; P < 0.001).	The DNN predicted complete response with an 80% accuracy.	Radscore ([OR] = 13.25; [95% CI],4.06–71.64; p < 0.001) Age (OR = 1.10/1 year; 1.03–1.20; p = 0.008)	OS from 0.672 [0.617 0.728] with clinical features only to 0.730 0.658 0.801]	83.9% accuracy in predicting TRG 0 vs. TRG 1–3 in validation.	AUC=0.842 (training set) AUC=0.802 (validation set)
Feature selection model	LoG spatial filter	ICC LASSO Cox model	Wilcoxon test, p<0.05 ICC	Penalized logistic regression	Spearman correlation coefficient	Keep high ROC	LASSO
Statistical method	independent t-test log-rank test MCPHM	Chi-square log-rank tests	DNN SVM LR	Univariable analysis MLR	Unsupervised and supervised method	LOR RF SVM	MLR
Imaging modality	CT	CT	CT	CT	CT	CT	CT
Number of patients	95	108	95	121	411	91	148
Study endpoint	Response, DFS	DFS	PCR	Downstaging	LC, DFS ,OS Distant control	PR	Distant metastases, OS
Accepted	August 10, 2017	January 22, 2018	August 3, 2018	March 11, 2019	October 25, 2019	April 6, 2020	August 13, 2020
Author	Chee et al.	Yankai Meng, et al.	Jean et al.	Ben et al.	Jiazhou et al.	Zhigang et al.	Mou Li, et al.

ICC, intraclass correlation coefficient; DNN, deep neural network; SVM, support vector machine; LIR, linear regression; LOR, logistic regression; RF, random forest; HR, hazard ratio; OR, odds ratio MLR, multivariate logistic regression; MCPHM, multivariable Cox proportional hazards model; LoG, Laplacian of Gaussian; ROC, receiver operating characteristic.

hazard models, have been used to evaluate LRRC prognosis. Nevertheless, linear systems can have considerable limitations and often fail to capture the complexity of clinicopathological characteristics. Therefore, it is very necessary to analyze the prognosis of LRRC, and there is still a gap in the application of radiomics in the prognosis of LRRC.

DOSIOMICS AND ITS APPLICATION

Inspired by radiomics, the concept of dosiomics was formally proposed in 2017. It builds a radiotherapy result prediction model by extracting the characteristics of dose distribution, thus guiding the formulation of personalized radiotherapy plans. The patients' 3D dose distributions can be considered as images with spatial and statistical distributions of dose levels. For radiotherapy for cancer, parameters such as the prescription dose, dose distribution, and dose–volume histogram (DVH) can also be used to assess the treatment response and prognostic analysis of cancer. Dosimetry texture features include: volume, dose, variance, center point position, contour boundary, spatially weighted DVH skewness, and kurtosis (77). Combining the characteristics of radiology and dosimetry can obtain more comprehensive information related to tumor radiotherapy, which helps to improve the accuracy of prediction. Unfortunately, there have been only 13 studies on dosiomics from 2018 to 2022, and there is no research on dosiomics in rectal cancer.

Linda Rossi et al. in 2018, applied dosimetric texture analysis (TA) features and DVH parameters to improve the prediction modeling of treatment complication rates in prostate cancer radiotherapy (78). Dosimetric texture analysis features characterizes the grayscale distribution in a patient's 3D dose distribution image and derives image features to improve the features of the predictive model. The main dosimetric texture analysis features extracted by Linda Rossi et al. are the gray-level frequency histogram, gray-level co-occurrence matrix (GLCM), gray-level run length matrix (GLRLM), gray- level size zone matrix (GLSZM), and neighborhood gray tone difference matrix (NGTDM). Bin Liang et al. extracted the spatial characteristics of dose distribution in the ipsilateral, contralateral, and whole lung by dosimetry and then used them to construct a prediction model by single-factor and multifactor LR (79). The results showed that the spatial characteristics of dose distribution extracted by dosimetry effectively improved the predictive ability. Aiqian Wua et al. investigated whether dosiomics can predict an IMRT-treated patient's locoregional recurrences (LRs) and get a comprehensive dosimetry and radiomics model can successfully divide patients into high- and low-risk groups(log-rank test, p=0.025), but the radiomics model alone cannot get same result (80). Furthermore, Lee et al. proposed a multiperspective data analysis method to predict weight loss in the acute phase of radiotherapy for lung cancer using radiomics and dosimetry texture features. In short, many studies on tumor radiotherapy are more effective in predicting models established by combining radiomics, dosimetry, and

TABLE 2 | Summaries of selected LARC radiomics studies (MRI).

Validation	Internal validation (4-fold validation)	Internal validation (train/test split)	Internal validation (train/test split)	Internal validation (train/test split)	Internal validation (train/test split)	Internal validation (4-fold validation)	Internal validation (train/test split)	Internal validation (train/test split)	Internal validation (train/test split)	External validation
Performance	AUC =0.84 for pCR AUC =0.89 for GR	AUC = 0.9756 (training) AUC = 0.9799 (test)	AUC=0.93 [95% CI: 0.84, 1]	AUC= 0.948 [95% CI, 0.9070.989] AUC= 0.966 [95%CI, 0.924-1.000]	AUC = 0.908, 0.902, 0.930	Handcrafted features AUC: 0.64 DL-based features AUC: of 0.73	AUC =0.75 (training) AUC = 0.75 (test)	Training cohort AUC =0.94 (95% CI: 0.82–0.99) validation cohort AUC=0.80 (95% CI: 0.58–0.94)	AUC = 0.84,0.88 (training) AUC = 0.81,0.75 (test)	Training cohort AUC= 0·868 [95%CI 0·825–0·912] validation cohort 1 AUC=0·860 [95%CI 0·828–0·892] validation cohort 2 AUC=0·872 [95%CI 0·810–0·934]
Model	Artificial neural network	Logistic regression	Random forest classifier	Radiomics nomogram	SVM	LASSO-logistic regression models	SVM	LASSO logistic regression	Logistic regression	RAPIDS prediction signature
ROI software	Manual segmentation	Manual segmentation	Manual segmentation	Manual segmentation	Manual segmentation	Manual segmentation	Manual segmentation	Manual segmentation	Manual segmentation	Manual segmentation
Number of patients	48	222	114	186	134	43	102	67	165	933
MRI Imaging modality	Pre- CRT MRI (T1 \T2WI, DWI, and DCE)	Pre- and after CRT MRI (T2WI and DWI)	After CRT MRI (T2WI)	Pre-CRT MRI (T1 \T2WI, DWI, and ADC)	Preoperative MRI (T2WI)	MRI (DWI and ADC map)	Pre- CRT MRI (T2WI)	Pre- CRT MRI (T2WI)	Pre- and after CRT MRI (T2WI and DWI)	Pre- CRT MRI (T1 \T2WI and DWI)
Study endpoint	PCR GR	PCR	PCR	PCR	PCR, GR downstaging	PCR	PCR	Non-responders	PCR	PCR
Accepted time	May 19, 2016	September 22, 2017	January 2, 2018	July 27, 2018	June 6, 2019	February 24, 2020	April 15, 2020	May 14, 2020	October 20, 2020	January 4, 2022
Author	Ke Nie, et al.	Zhenyu Liu, et al.	Natally, et al.	Yanfen Cui, et al.	Xiaoping Yi, et al.	Jie Fu, et al.	Iva Petkovska, et al.	Bianca Petresc, et al.	Lijuan Wan, et al.	Lili Feng, et al.

PCR, pathological complete response; GR, good response; AUC, area under the curve; CI, confidence interval; ROI, region of interest; T1WI, T1-weighted imaging; DWI, diffusion-weighted imaging; T2WI, T2-weighted imaging; DCE, dynamic contrast enhanced.

clinical data (81). Jin et al. extracted 42 radiomic features from CT images of 94 patients with esophageal cancer and combined them with 18 dosimetry parameters to predict the patient's response to radiotherapy; the study results showed that the radiomic features were combined with dosimetry parameters. The subsequent AUC can reach 0.71, while the AUC using only radiomics is 0.69 (82). So far, the research on dosiomics in rectal cancer is vacant.

The authors found that found that compared with the predictive model established by radiomics alone, the model established by integrating the characteristic parameters of radiomics and radiotherapy dosimetry can effectively improve the predictive evaluation of tumors after radiotherapy. However, relatively few studies combine radiomics and dosiomics, and the application of dosiomic characteristics to predict the efficacy of radiotherapy is still in its initial stage. Therefore, the role of dosiomic characteristics in radiation oncology should be further studied.

LIMITATIONS OF RADIOMICS

When applying radiomics research results to clinical practice, there are still some limitations and challenges in some aspects, mainly including the following:

1. Most of the previous studies were conducted retrospective analysis, and most of the included data came from the same institution. Therefore, a large sample and multicenter prospective research test are essential. It requires extensive cooperation in multiple disciplines and fields and is also an essential part of applying radiomics to clinical practice.
2. The reproducibility and repeatability of radiomics features still need to be discussed. This issue depends on the used imaging modality, sequence, scanning parameters, reconstruction algorithm spatial resolution, size of the image, image quality, reconstruction and correction parameters, and motion artifacts, and software used to extract radiomic

features (83). The IBSI proposes the computed features of different institutions on a common data set. To expand this effort, a review proposes to include benchmarking data sets collected by different institutions to guarantee the maximum heterogeneity in terms of the acquisition parameters and develop an infrastructure, based on workflow programming language, that allows users to connect to the mentioned repository and run their feature extraction software (84). In terms of repeatability and reproducibility, deep learning-based radiomics may be advantageous. The self-learning neural networks show a better capability for generalization (85).

3. There is no theoretical basis to explain the biological meaning of radiomic features, which also hinders the further development of radiomics. Many models have been created and published, but these studies often lack standardized evaluation on external cohorts of patients, which also explains why not a single model has been translated to clinical practice. Published Radiomics quality score (RQS) and Transparent Reporting of a multivariable prediction model forindividual prognosis or diagnosis (TRIPOD) guidelines improve the validity of radiomics as a clinically accepted field. We need to overcome the challenges before radiomics can be successfully introduced into clinical settings. Efforts are being made to overcome these limitations

CONCLUSIONS

With the rapid development of AI technology, radiomics based on machine learning and deep learning has broad application prospects. In the current background of advocating precision medicine and personalized treatment, in evaluating the efficacy of NCRT in LARC patients, a non-invasive, efficient, and accurate imaging omics prediction model is established for clinical application, and it has developed into a clinically useful model.

In summary, radiomics is an emerging diagnostic imaging technology that plays an essential role in predicting the effect of NCRT on LARC and can optimize treatment plans through process management, thereby improving the short-term and long-term prognosis of patients. This review also emphasizes the necessity of applying radiomics and dosiomics to the prognostic model of LRRC. In short, the application of radiomics and dosimetry in radiation oncology is of great value for doctors' clinical decision-making, treatment planning, and follow-up workflow. However, published retrospective studies presented their own model with a certain degree of heterogeneity and do not facilitate translation into clinical practice. Therefore, the study design needs to be further improved, and the promotion and validation of prediction models need to be further explored, so as to strengthen the clinical application of radiomics.

FUTURE DIRECTION

With the continuous advancement in AI, radiomic and dosiomic diagnosis and prediction methods based on deep machine learning are the effective way to develop clinical research in the future. Drawing on the application of radiomics and dosiomics in other cancers, in terms of evaluating the efficacy of LRRC patients through seed implantation and second-course radiotherapy, establishes a fast, efficient, and accurate radiomics prediction model for clinical application. Furthermore, establishing additional predictive tools that can be used in clinics in a true sense, especially in the current background of advocating precision medicine and personalized treatment, will become a low-cost, non-invasive, and convenient new diagnostic evaluation method, thereby improving the prognosis of patients.

AUTHOR CONTRIBUTIONS

YQ performed the literature searches and contributed to draft versions of the manuscript. WZ performed the searches and revised the final version of the manuscript. LZ provided support on all aspects of the review and reviewed the manuscript. JW improved ideas for the writing of articles. HW provided support on all aspects of the review and reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work is supported by the Innovation & Transfer Fund of Peking University Third Hospital (Grant No. BYSYZHKC2021113), Bei jing Lianying Intelligent Imaging Technology Research Institute Hospital-enterprise Joint Research and Development Platform Fund, No.H79462-07 and National Nature Science Foundation of China under Grants No. U1867210 and No.12175012.

REFERENCES

1. *Latest Global Cancer Data: Cancer Burden Rises to 19.3 Million New Cases and 10.0 Million Cancer Deaths in 2020QUESTIONS AND ANSWERS (Q&A).* IARC. (who.int). Available at: https://www.iarc.who.int/news-events/latest-global-cancer-data-cancer-burden-rises-to-19-3-million-new-cases-and-10-0-million-cancer-deaths-in-2020/.
2. American Cancer Society. *Colorectal Cancer Facts & Figures 2020-2022.* Atlanta: American Cancer Society (2020).
3. Limkin EJ, Sun R, Dercle L, Zacharaki EI, Robert C, Reuzé S, et al. Promises and Challenges for the Implementation of Computational Medical Imaging (Radiomics) in Oncology. *Ann Oncol* (2017) 28(6):1191–206. doi: 10.1093/annonc/mdx034
4. Hirotoshi K, Akifumi K, Satoshi O, Megumi I, Toshiaki I, Satoru I, et al. Diagnostic Performance of Multidetector Row Computed Tomography for

Assessment of Lymph Node Metastasis in Patients With Distal Rectal Cancer. *Ann Surg Oncol* (2015) 22:203–8. doi: 10.1245/s10434-014-3972-3
5. 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
6. Yogananda CGB, Nalawade SS, Murugesan GK, et al. *Fully Automated Brain Tumor Segmentation and Survival Prediction of Gliomas Using Deep Learning and MRI*. Available at: https://link.springer.com/chapter/10.1007/978-3-030-46643-5_10.
7. Sun L, Zhang S, Chen H, Luo L. Brain Tumor Segmentation and Survival Prediction Using Multimodal MRI Scans With Deep Learning. *Front Neurosci* (2019) 13:810. doi: 10.3389/fnins.2019.00810
8. Jochems A, Leijenaar RTH, Bogowicz M, Hoebers FJP, Wesseling F, Huang SH. PO-0932: Combining Deep Learning and Radiomics to Predict HPV Status in Oropharyngeal Squamous Cell Carcinoma. *Radiother Oncol* (2018) 127:S504–5. doi: 10.1016/S0167-8140(18)31242-8
9. Placidi L, Cusumano D, Lenkowicz J, Boldrini L, Valentini V. On Dose Cube Pixel Spacing Pre-Processing for Features Extraction Stability in Dosiomic Studies. *Phys Med* (2021) 90:108–14. doi: 10.1016/j.ejmp.2021.09.010
10. Tang X, Jiang W, Li H, Xie F, Dong A, Liu L, et al. Predicting Poor Response to Neoadjuvant Chemoradiotherapy for Locally Advanced Rectal Cancer: Model Constructed Using Pre-Treatment MRI Features of Structured Report Template. *Radiother Oncol* (2020) 148:97–106. doi: 10.1016/j.radonc.2020.03.046
11. *American Cancer Society*. Available at: https://cancerstatisticscenter.cancer.org/#!/.
12. Li S, Hou Z, Liu J, Ren W, Wan Sui R, Yan J, et al. A Review of Radiomics Analysis and Modeling Tools[J]. *Chin J Med Phys* (2018) 35(09):1043–9. doi: 10.3969/j.issn.1005-202X.2018.09.010
13. Huang X, Long L, Wei J, Li Y, Xia Y, Zuo P, et al. Radiomics for Diagnosis of Dual-Phenotype Hepatocellular Carcinoma Using Gd-EOB-DTPA-Enhanced MRI and Patient Prognosis. *J Cancer Res Clin Oncol* (2019) 145(12):2995–3003. doi: 10.1007/s00432-019-03062-3
14. Kim JY, Park JE, Jo Y, Shim WH, Nam SJ, Kim JH, et al. Incorporating Diffusion- and Perfusion-Weighted MRI Into a Radiomics Model Improves Diagnostic Performance for Pseudoprogression in Glioblastoma Patients. *Neuro Oncol* (2019) 21(3):404–14. doi: 10.1093/neuonc/noy133
15. Mokrane FZ, Lu L, Vavasseur A, Otal P, Peron JM, Luk L, et al. Radiomics Machine-Learning Signature for Diagnosis of Hepatocellular Carcinoma in Cirrhotic Patients With Indeterminate Liver Nodules. *Eur Radiol* (2020) 30(1):558–70. doi: 10.1007/s00330-019-06347-w
16. Chetan MR, Gleeson FV. Radiomics in Predicting Treatment Response in non-Small-Cell Lung Cancer: Current Status, Challenges and Future Perspectives. *Eur Radiol* (2021) 31(2):1049–58. doi: 10.1007/s00330-020-07141-9
17. Cai J, Zheng J, Shen J, Yuan Z, Xie M, Gao M, et al. A Radiomics Model for Predicting the Response to Bevacizumab in Brain Necrosis After Radiotherapy. *Clin Cancer Res* (2020) 26(20):5438–47. doi: 10.1158/1078-0432.CCR-20-1264
18. Horvat N, Veeraraghavan H, Khan M, Blazic I, Zheng J, Capanu M, et al. MR Imaging of Rectal Cancer: Radiomics Analysis to Assess Treatment Response After Neoadjuvant Therapy. *Radiology* (2018) 287(3):833–43. doi: 10.1148/radiol.2018172300
19. Vandendorpe B, Durot C, Lebellec L, Le Deley MC, Sylla D, Bimbai AM, et al. Prognostic Value of the Texture Analysis Parameters of the Initial Computed Tomographic Scan for Response to Neoadjuvant Chemoradiation Therapy in Patients With Locally Advanced Rectal Cancer. *Radiother Oncol* (2019) 135:153–60. doi: 10.1016/j.radonc.2019.03.011
20. Wang X, Sun J, Zhang W, Yang X, Zhu C, Pan B, et al. Use of Radiomics to Extract Splenic Features to Predict Prognosis of Patients With Gastric Cancer. *Eur J Surg Oncol* (2020) 46(10 Pt A):1932–40. doi: 10.1016/j.ejso.2020.06.021
21. Kekelidze M, D'Errico L, Pansini M, Tyndall A, Hohmann J. Colorectal Cancer: Current Imaging Methods and Future Perspectives for the Diagnosis, Staging and Therapeutic Response Evaluation. *World J Gastroenterol* (2013) 19(46):8502–14. doi: 10.3748/wjg.v19.i46.8502
22. Jhaveri KS, Hosseini-Nik H. MRI of Rectal Cancer: An Overview and Update on Recent Advances. *AJR Am J Roentgenol* (2015) 205(1):W42–55. doi: 10.2214/AJR.14.14201
23. Horvat N, Carlos Tavares Rocha C, Clemente Oliveira B, Petkovska I, Gollub MJ. MRI of Rectal Cancer: Tumor Staging, Imaging Techniques, and Management. *Radiographics* (2019) 39(2):367–87. doi: 10.1148/rg.2019180114
24. Lohmann P, Galldiks N, Kocher M, Heinzel A, Filss CP, Stegmayr C, et al. Radiomics in Neuro-Oncology: Basics, Workflow, and Applications. *Methods* (2021) 188:112–21. doi: 10.1016/j.ymeth.2020.06.003
25. Zwanenburg A, Vallières M, Abdalah MA, Aerts HJWL, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-Based Phenotyping. *Radiology* (2020) 295(2):328–38. doi: 10.1148/radiol.2020191145
26. Qiu Q, Duan J, Duan Z, Meng X, Ma C, Zhu J, et al. Reproducibility and non-Redundancy of Radiomic Features Extracted From Arterial Phase CT Scans in Hepatocellular Carcinoma Patients: Impact of Tumor Segmentation Variability. *Quant Imaging Med Surg* (2019) 9(3):453–64. doi: 10.21037/qims.2019.03.02
27. Eminowicz G, McCormack M. Variability of Clinical Target Volume Delineation for Definitive Radiotherapy in Cervix Cancer. *Radiother Oncol* (2015) 117(3):542–7. doi: 10.1016/j.radonc.2015.10.007
28. Chen W, Wang C, Zhan W, Jia Y, Ruan F, Qiu L, et al. A Comparative Study of Auto-Contouring Softwares in Delineation of Organs at Risk in Lung Cancer and Rectal Cancer. *Sci Rep* (2021) 11(1):23002. doi: 10.1038/s41598-021-02330-y
29. Zhu HT, Zhang XY, Shi YJ, Li XT, Sun YS. Automatic Segmentation of Rectal Tumor on Diffusion-Weighted Images by Deep Learning With U-Net. *J Appl Clin Med Phys* (2021) 22(9):324–31. doi: 10.1002/acm2.13381
30. Kumar V, Gu Y, Basu S, Berglund A, Eschrich SA, Schabath MB, et al. Radiomics: The Process and the Challenges. *Magn Reson Imaging* (2012) 30(9):1234–48. doi: 10.1016/j.mri.2012.06.010
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. Haralick RM, Shanmugam K, Its'Hak D. Textural Features for Image Classification. *IEEE Trans Syst Man Cyb Smc* (1973) 3(6):610–21. doi: 10.1109/TSMC.1973.4309314
33. Amadasun M, King R. "Textural Features Corresponding to Textural Properties". In: *IEEE Transactions on Systems, Man, and Cybernetics* (1989) 19:1264–74. doi: 10.1109/21.44046
34. Zhou M, Scott J, Chaudhury B, Hall L, Goldgof D, Yeom KW, et al. Radiomics in Brain Tumor: Image Assessment, Quantitative Feature Descriptors, and Machine-Learning Approaches. *AJNR Am J Neuroradiol* (2018) 39(2):208–16. doi: 10.3174/ajnr.A5391
35. Ha S, Choi H, Paeng JC, Cheon GJ. Radiomics in Oncological PET/CT: A Methodological Overview. *Nucl Med Mol Imaging* (2019) 53(1):14–29. doi: 10.1007/s13139-019-00571-4
36. Hu Y, Xie C, Yang H, Ho JWK, Wen J, Han L, et al. Computed Tomography-Based Deep-Learning Prediction of Neoadjuvant Chemoradiotherapy Treatment Response in Esophageal Squamous Cell Carcinoma. *Radiother Oncol* (2021) 154:6–13. doi: 10.1016/j.radonc.2020.09.014
37. Acharya UR, Hagiwara Y, Sudarshan VK, Chan WY, Ng KH. Towards Precision Medicine: From Quantitative Imaging to Radiomics. *J Zhejiang Univ Sci B* (2018) 19(1):6–24. doi: 10.1631/jzus.B1700260
38. Lever J, Krzywinski M, Altman N. Model Selection and Overfitting. In: *Nature Methods* (2016) 13:703–4. doi: 10.1038/nmeth.3968
39. López-Cabrera JD, Lorenzo-Ginori JV. Feature Selection for the Classification of Traced Neurons. *J Neurosci Methods* (2018) 303:41–54. doi: 10.1016/j.jneumeth.2018.04.002
40. Parekh V, Jacobs MA. Radiomics: A New Application From Established Techniques. *Expert Rev Precis Med Drug Dev* (2016) 1(2):207–26. doi: 10.1080/23808993.2016.1164013
41. Kuhn M, Johnson K. *Applied Predictive Modeling* Vol. 4. . New York: Springer (2013).
42. Li M, Zhu YZ, Zhang YC, Yue YF, Yu HP, Song B. Radiomics of Rectal Cancer for Predicting Distant Metastasis and Overall Survival. *World J Gastroenterol* (2020) 26(33):5008–21. doi: 10.3748/wjg.v26.i33.5008
43. Park HJ, Park B, Lee SS. Radiomics and Deep Learning: Hepatic Applications. *Korean J Radiol* (2020) 21(4):387–401. doi: 10.3348/kjr.2019.0752

44. Wagner MW, Bilbily A, Beheshti M, Shammas A, Vali R. Artificial Intelligence and Radiomics in Pediatric Molecular Imaging. *Methods* (2021) 188:37–43. doi: 10.1016/j.ymeth.2020.06.008
45. LeCun Y, Bengio Y, Hinton G. Deep Learning. *Nature* (2015) 521:436–44. doi: 10.1038/nature14539
46. 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(5):403–10. doi: 10.1016/j.canrad.2020.01.011
47. 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
48. 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) 14(12):e0226348. doi: 10.1371/journal.pone.0226348
49. Nasief H, Zheng C, Schott D, Hall W, Tsai S, Erickson B, et al. A Machine Learning Based Delta-Radiomics Process for Early Prediction of Treatment Response of Pancreatic Cancer. *NPJ Precis Oncol* (2019) 3:25. doi: 10.1038/s41698-019-0096-z
50. Colombini GG, de Abreu IBM, Cerri R. "A Self-Organizing Map-Based Method for Multi-Label Classification,". In: *2017 International Joint Conference on Neural Networks (IJCNN)* (2017). p. 4291–8. doi: 10.1109/IJCNN.2017.7966399
51. Boldrini L, Cusumano D, Chiloiro G, Casà C, Masciocchi C, Lenkowicz J, et al. Delta Radiomics for Rectal Cancer Response Prediction With Hybrid 0.35 T Magnetic Resonance-Guided Radiotherapy (MRgRT): A Hypothesis-Generating Study for an Innovative Personalized Medicine Approach. *Radiol Med* (2019) 124(2):145–53. doi: 10.1007/s11547-018-0951-y
52. Jeon SH, Song C, Chie EK, Kim B, Kim YH, Chang W, et al. Delta-Radiomics Signature Predicts Treatment Outcomes After Preoperative Chemoradiotherapy and Surgery in Rectal Cancer. *Radiat Oncol* (2019) 14(1):43. doi: 10.1186/s13014-019-1246-8
53. Chiloiro G, Rodriguez-Carnero P, Lenkowicz J, Casà C, Masciocchi C, Boldrini L, et al. Delta Radiomics Can Predict Distant Metastasis in Locally Advanced Rectal Cancer: The Challenge to Personalize the Cure. *Front Oncol* (2020) 10:595012. doi: 10.3389/fonc.2020.595012
54. Wan L, Peng W, Zou S, Ye F, Geng Y, Ouyang H, et al. MRI-Based Delta-Radiomics are Predictive of Pathological Complete Response After Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Acad Radiol* (2021) 28 Suppl 1:S95–S104. doi: 10.1016/j.acra.2020.10.026
55. Chen H, Shi L, Nguyen KNB, Monjazeb AM, Matsukuma KE, Loehfelm TW, et al. MRI Radiomics for Prediction of Tumor Response and Downstaging in Rectal Cancer Patients After Preoperative Chemoradiation. *Adv Radiat Oncol* (2020) 5(6):1286–95. doi: 10.1016/j.adro.2020.04.016
56. Cusumano D, Boldrini L, Yadav P, Yu G, Musurunu B, Chiloiro G, et al. Delta Radiomics for Rectal Cancer Response Prediction Using Low Field Magnetic Resonance Guided Radiotherapy: An External Validation. *Phys Med* (2021) 84:186–91. doi: 10.1016/j.ejmp.2021.03.038
57. Wang H, Chen H, Duan S, Hao D, Liu J. Radiomics and Machine Learning With Multiparametric Preoperative MRI May Accurately Predict the Histopathological Grades of Soft Tissue Sarcomas. *J Magn Reson Imaging* (2020) 51(3):791–7. doi: 10.1002/jmri.26901
58. He B, Ji T, Zhang H, Zhu Y, Shu R, Zhao W, et al. MRI-Based Radiomics Signature for Tumor Grading of Rectal Carcinoma Using Random Forest Model. *J Cell Physiol* (2019) 234(11):20501–9. doi: 10.1002/jcp.28650
59. 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
60. Li Y, Wang J, Ma X, Tan L, Yan Y, Xue C, et al. A Review of Neoadjuvant Chemoradiotherapy for Locally Advanced Rectal Cancer. *Int J Biol Sci* (2016) 12(8):1022–31. doi: 10.7150/ijbs.15438
61. Bibault JE, Giraud P, Housset M, Durdux C, Taieb J, Berger A, et al. Deep Learning and Radiomics Predict Complete Response After Neo-Adjuvant Chemoradiation for Locally Advanced Rectal Cancer. *Sci Rep* (2018) 8(1):12611. doi: 10.1038/s41598-018-30657-6
62. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New Response Evaluation Criteria in Solid Tumours: Revised RECIST Guideline (Version 1.1). *Eur J Cancer* (2009) 45(2):228–47. doi: 10.1016/j.ejca.2008.10.026
63. 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. doi: 10.1002/jmri.25968
64. Chee CG, Kim YH, Lee KH, Lee YJ, Park JH, Lee HS, et al. CT Texture Analysis in Patients With Locally Advanced Rectal Cancer Treated With Neoadjuvant Chemoradiotherapy: A Potential Imaging Biomarker for Treatment Response and Prognosis. *PLoS One* (2017) 12(8):e0182883. doi: 10.1371/journal.pone.0182883
65. Wang J, Shen L, Zhong H, Zhou Z, Hu P, Gan J, et al. Radiomics Features on Radiotherapy Treatment Planning CT can Predict Patient Survival in Locally Advanced Rectal Cancer Patients. *Sci Rep* (2019) 9(1):15346. doi: 10.1038/s41598-019-51629-4
66. Yuan Z, Frazer M, Zhang GG, Latifi K, Moros EG, Feygelman V, et al. CT-Based Radiomic Features to Predict Pathological Response in Rectal Cancer: A Retrospective Cohort Study. *J Med Imaging Radiat Oncol* (2020) 64(3):444–9. doi: 10.1111/1754-9485.13044
67. 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
68. 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(23):7253–62. doi: 10.1158/1078-0432.CCR-17-1038
69. Cui Y, Yang X, Shi Z, Yang Z, Du X, Zhao Z, et al. Radiomics Analysis of Multiparametric MRI for Prediction of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Eur Radiol* (2019) 29(3):1211–20. doi: 10.1007/s00330-018-5683-9
70. Yi X, Pei Q, Zhang Y, Zhu H, Wang Z, Chen C, et al. MRI-Based Radiomics Predicts Tumor Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer. *Front Oncol* (2019) 9:552. doi: 10.3389/fonc.2019.00552
71. Petkovska I, Tixier F, Ortiz EJ, Golia Pernicka JS, Paroder V, Bates DD, et al. Clinical Utility of Radiomics at Baseline Rectal MRI to Predict Complete Response of Rectal Cancer After Chemoradiation Therapy. *Abdom Radiol (NY)* (2020) 45(11):3608–17. doi: 10.1007/s00261-020-02502-w
72. Fu J, Zhong X, Li N, Van Dams R, Lewis J, Sung K, et al. Deep Learning-Based Radiomic Features for Improving Neoadjuvant Chemoradiation Response Prediction in Locally Advanced Rectal Cancer. *Phys Med Biol* (2020) 65(7):075001. doi: 10.1088/1361-6560/ab7970
73. Feng L, Liu Z, Li C, Li Z, Lou X, Shao L, et al. Development and Validation of a Radiopathomics Model to Predict Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer: A Multicentre Observational Study. *Lancet Digit Health* (2022) 4(1):e8–e17. doi: 10.1016/S2589-7500(21)00215-6
74. Ganeshan D, Nougaret S, Korngold E, Rauch GM, Moreno CC. Locally Recurrent Rectal Cancer: What the Radiologist Should Know. *Abdom Radiol (NY)* (2019) 44(11):3709–25. doi: 10.1007/s00261-019-02003-5
75. Wang l, Wang JJ. Progress in the Treatment of Locally Recurrent Rectal Cancer. *Chin J Clin* (2020) 47(04):208–12. doi: 10.3969/j.issn.1000-8179.2020.04.368
76. Benson AB, Venook AP, Al-Hawary MM, Arain MA, Chen YJ, Ciombor KK, et al. National Comprehensive Cancer Networks (NCCN). *The NCCN Rectal Cancer Clinical Practice Guidelines in Oncology (Version 2 2020).* (2020).
77. Monti S, Xu T, Liao Z, Mohan R, Cella L, Palma G. On the Interplay Between Dosiomics and Genomics in Radiation-Induced Lymphopenia of Lung Cancer Patients. *Radiother Oncol* (2021) 167:219–25. doi: 10.1016/j.radonc.2021.12.038
78. Rossi L, Bijman R, Schillemans W, Aluwini S, Cavedon C, Witte M, et al. Texture Analysis of 3D Dose Distributions for Predictive Modelling of Toxicity Rates in Radiotherapy. *Radiother Oncol* (2018) 129(3):548–53. doi: 10.1016/j.radonc.2018.07.027

79. Liang B, Yan H, Tian Y, Chen X, Yan L, Zhang T, et al. Dosiomics: Extracting 3d Spatial Features From Dose Distribution to Predict Incidence of Radiation Pneumonitis. *Front Oncol* (2019) 9:269. doi: 10.3389/fonc.2019.00269
80. Wu A, Li Y, Qi M, Lu X, Jia Q, Guo F, et al. Dosiomics Improves Prediction of Locoregional Recurrence for Intensity Modulated Radiotherapy Treated Head and Neck Cancer Cases. *Oral Oncol* (2020) 104:104625. doi: 10.1016/j.oraloncology.2020.104625
81. Lee SH, Han P, Hales RK, Voong KR, Noro K, Sugiyama S, et al. Multi-View Radiomics and Dosiomics Analysis With Machine Learning for Predicting Acute-Phase Weight Loss in Lung Cancer Patients Treated With Radiotherapy. *Phys Med Biol* (2020) 65(19):195015. doi: 10.1088/1361-6560/ab8531
82. Jin X, Zheng X, Chen D, Jin J, Zhu G, Deng X, et al. Prediction of Response After Chemoradiation for Esophageal Cancer Using a Combination of Dosimetry and CT Radiomics. *Eur Radiol* (2019) 29(11):6080–8. doi: 10.1007/s00330-019-06193-w
83. Park JE, Park SY, Kim HJ, Kim HS. Reproducibility and Generalizability in Radiomics Modeling: Possible Strategies in Radiologic and Statistical Perspectives. *Korean J Radiol* (2019) 20(7):1124–37. doi: 10.3348/kjr.2018.0070
84. Traverso A, Wee L, Dekker A, Gillies R. Repeatability and Reproducibility of Radiomic Features: A Systematic Review. *Int J Radiat Oncol Biol Phys* (2018) 102(4):1143–58. doi: 10.1016/j.ijrobp.2018.05.053
85. Lohmann P, Bousabarah K, Hoevels M, Treuer H. Radiomics in Radiation Oncology-Basics, Methods, and Limitations. *Strahlenther Onkol* (2020) 196 (10):848–55. doi: 10.1007/s00066-020-01663-3

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.

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 11 August 2022
DOI 10.3389/fonc.2022.886517

OPEN ACCESS

EDITED BY
Yuming Jiang,
Stanford University, United States

REVIEWED BY
Ke Yan,
Alibaba DAMO Academy, China
Elanchezhian Somasundaram,
Cincinnati Children's Hospital Medical Center, United States
Hossam El-Din Moustafa,
Mansoura University, Egypt

*CORRESPONDENCE
Brian Anderson
b5anderson@health.ucsd.edu

SPECIALTY SECTION
This article was submitted to Radiation Oncology, a section of the journal Frontiers in Oncology

RECEIVED 28 February 2022
ACCEPTED 01 July 2022
PUBLISHED 11 August 2022

CITATION
Anderson BM, Rigaud B, Lin Y-M, Jones AK, Kang HC, Odisio BC and Brock KK (2022) Automated segmentation of colorectal liver metastasis and liver ablation on contrast-enhanced CT images. *Front. Oncol.* 12:886517. doi: 10.3389/fonc.2022.886517

Automated segmentation of colorectal liver metastasis and liver ablation on contrast-enhanced CT images

Brian M. Anderson[1,2]*, Bastien Rigaud[1], Yuan-Mao Lin[3], A. Kyle Jones[1], HynSeon Christine Kang[4], Bruno C. Odisio[3] and Kristy K. Brock[1]

[1]Department of Imaging Physics, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [2]UTHealth Graduate School of Biomedical Sciences, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [3]Department of Interventional Radiology, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, [4]Department of Abdominal Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

Objectives: Colorectal cancer (CRC), the third most common cancer in the USA, is a leading cause of cancer-related death worldwide. Up to 60% of patients develop liver metastasis (CRLM). Treatments like radiation and ablation therapies require disease segmentation for planning and therapy delivery. For ablation, ablation-zone segmentation is required to evaluate disease coverage. We hypothesize that fully convolutional (FC) neural networks, trained using novel methods, will provide rapid and accurate identification and segmentation of CRLM and ablation zones.

Methods: Four FC model styles were investigated: Standard 3D-UNet, Residual 3D-UNet, Dense 3D-UNet, and Hybrid-WNet. Models were trained on 92 patients from the liver tumor segmentation (LiTS) challenge. For the evaluation, we acquired 15 patients from the 3D-IRCADb database, 18 patients from our institution (CRLM = 24, ablation-zone = 19), and those submitted to the LiTS challenge (n = 70). Qualitative evaluations of our institutional data were performed by two board-certified radiologists (interventional and diagnostic) and a radiology-trained physician fellow, using a Likert scale of 1–5.

Results: The most accurate model was the Hybrid-WNet. On a patient-by-patient basis in the 3D-IRCADb dataset, the median (min–max) Dice similarity coefficient (DSC) was 0.73 (0.41–0.88), the median surface distance was 1.75 mm (0.57–7.63 mm), and the number of false positives was 1 (0–4). In the LiTS challenge (n = 70), the global DSC was 0.810. The model sensitivity was 98% (47/48) for sites ≥15 mm in diameter. Qualitatively, 100% (24/24; minority vote) of the CRLM and 84% (16/19; majority vote) of the ablation zones had Likert scores ≥4.

Conclusion: The Hybrid-WNet model provided fast (<30 s) and accurate segmentations of CRLM and ablation zones on contrast-enhanced CT scans, with positive physician reviews.

KEYWORDS

deep-learning, liver cancer, percutaneous ablation, computed tomography biomechanical modeling

Introduction

Colorectal cancer in the United States

Colorectal cancer (CRC) is the third most common cancer in the United States in both men and women (1) and a leading cause of cancer-related death worldwide (2). The main cause of death for CRC patients is metastasis (3). Up to 60% of patients develop colorectal liver metastasis (CRLM) over the course of their disease, with 25% presenting with CRLM at diagnosis (4). Such facts highlight the importance of liver-directed loco-regional therapies (LRT) for these patients.

While several treatment options are available for CRLM (particularly radiation and ablation therapies), they all rely on accurate estimation of disease extent, usually involving cross-sectional imaging with contrast-enhanced CT (CECT) or MRI. CRLM often appears as hypo-enhancing lesions on routine CECT portal-venous phase images. However, their detection can be challenging owing to ill-defined margins, particularly for sub-centimeter lesions.

Both radiofrequency and microwave ablation interventions aim for a minimum margin to be achieved around the disease to ensure that all microscopic disease is treated. This requires both segmentation of disease on pre-treatment images and the ablation zone on post-treatment images to assess the ablation margin (5). The ablation zone is hypo-enhanced on CECT images, similar to the CRLM. A clinical trial is underway (Identifier: NCT04083378) to map the CRLM from pre-treatment to post-treatment imaging and assess treatment efficacy, but manual segmentations of both the disease and the ablation zone are still required (6), adding time to the procedure.

To date, automated liver disease segmentation tasks either have largely focused on primary liver disease, or have not included qualitative evaluation of generated contours (7–10). Furthermore, hepatocellular carcinomas tend to have enhancement during the arterial phase of contrast-enhanced CT with a hypodense rim (11, 12), while CRLM often shows hyperenhancement on the rim and a hypo-enhancing center (13, 14). New institution- and society-sponsored competitions, such as the liver tumor segmentation (LiTS) challenge (15) and the 3D-IRCADb01 (7) dataset, have included data from both primary and secondary liver cancers, enabling investigation, development, and comparison of automatic segmentation algorithms using public data.

It is hypothesized that fully convolutional neural networks, trained using novel methods to account for the challenges of varying disease size, will provide rapid and accurate identification and segmentation of both CRLM and ablation cavities. We believe that this approach will facilitate the automated detection of CRLM, radiation treatment planning for CRLM, and the evaluations of margin in ablation therapy.

Abbreviations: CRC, colorectal cancer; CRLM, colorectal liver metastasis; LRT, loco-regional therapies; CECT, contrast enhanced computed tomography; LiTS, liver tumor segmentation; DSC, dice similarity coefficient, computed tomography; ML, machine learning.

Materials and methods

Quantitative training, validation, and testing

To ensure reproducibility by other institutions, data were provided by the publicly available LiTS challenge (15). LiTS consists of CECT scans from 131 patients with primary and secondary liver disease collected from seven different institutions (15). Subjects suffered from primary tumor disease, such as HCC, as well as secondary liver tumors and metastasis from breast, lung, and CRC. Ground-truth segmentations of the liver and disease were provided in the data; the goal of the model is to similarly segment the disease. Each image was reviewed by BMA, BCO to remove the data showing hyper-enhancing metastases, or lacking image acquisition parameters. A total of 92 patients remained.

The model was evaluated *via* submission to the LiTS challenge and the 3D-IRCADb01 publicly available dataset of 20 patients (10 male and 10 female patients) with liver disease (16). Table 1 shows the image acquisition parameters for the training, validation, and test sets. Five of the patients from the 3D-IRCADb01 dataset were excluded: patients 5, 11, and 20 had no disease; patient 12 had a large cystic lesion at the base of the liver; and patient 18 had CRLM with atypical enhancement pattern.

TABLE 1 Image acquisition parameters of the LiTS challenge for the training, validation, and test sets.

Origin	Distribution	Mean (min–max)		
		Slice thickness (mm)	Pixel size X (mm/voxel)	Pixel size Y (mm/voxel)
LiTS	Training (n = 72)	1.64 (0.7–5.0)	0.75 (0.60–0.98)	0.75 (0.60–0.98)
	Validation (n = 20)	2.02 (0.7–5.0)	0.78 (0.68–0.98)	0.78 (0.68–0.98)
	Test (n = 70)	2.43 (0.45–6.0)	0.75 (0.60–0.98)	0.75 (0.60–0.98)
3D-IRCADb	Test (n = 15)	1.78 (1.0–4.0)	0.72 (0.56–0.87)	0.72 (0.56–0.87)

Data pre-processing

Image intensity manipulation

A patient-specific mean and standard deviation Hounsfield unit was calculated for normalization on the basis of the full width at half maximum of the values within the liver; this reduced outliers as compared to using a global mean and standard deviation. The image intensity outside of the masked liver was set to be equal to 0.

Voxel size resampling

All training and validation images and ground-truth segmentation were resampled to 1 mm slice thickness, and 0.75 mm in the axial plane, using bi-linear interpolation.

Training image "slabs"

Initial training on the entire patient liver resulted in a model that struggled to identify disease sites. We believe that this is due to the disparity in class representation, being that a majority of the liver is "normal" and the model could achieve a high segmentation accuracy by segmenting everything as "normal". Simple class weighting would not solve this problem as it would result in the model weighing cases with extensive disease cases as more important than the less extensive disease cases.

To account for disparities in class representation, where smaller structures (CRLM) are inherently "worth less" than large structures (normal liver), training was distributed into unique "slabs". Each independent disease site was divided into "slabs" of 32 × 120 × 120 voxels. This size was selected arbitrarily as a balance of encompassing a large section of liver while reducing memory requirements. This ensures a representation of both disease and normal liver in each training step. Figure 1 illustrates several disease slabs for one patient. After extraction, the training dataset consisted of 572 unique samples.

The validation and test set were not broken into slabs, with the entire liver being passed at once for evaluation and testing.

Architectures

Four architectures were investigated: Standard 3D-UNet, Residual 3D-UNet, Dense 3D-UNet, and Hybrid-WNet (pre-trained Standard 2D-UNet with a 3D-DenseUNet). The basic

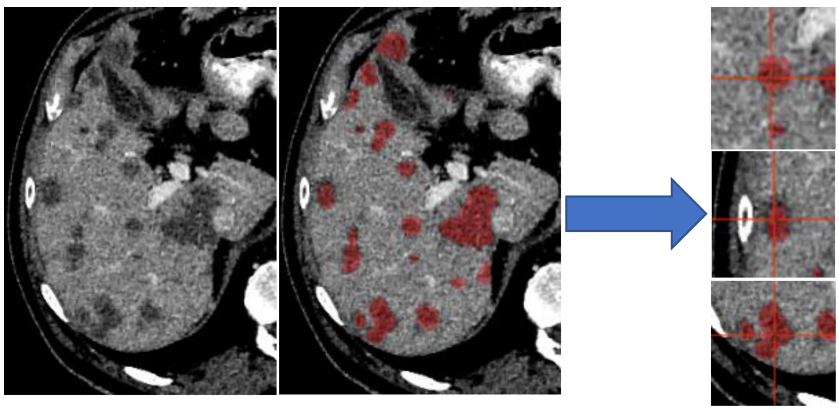


FIGURE 1
Liver distributed into individual slabs of 32 × 120 × 120. Disease was labeled as disease, regardless of the center of the slab. The validation and test set were not broken into slabs; our architectures accepted variable input sizes, with the entire liver being passed at once for evaluation and testing.

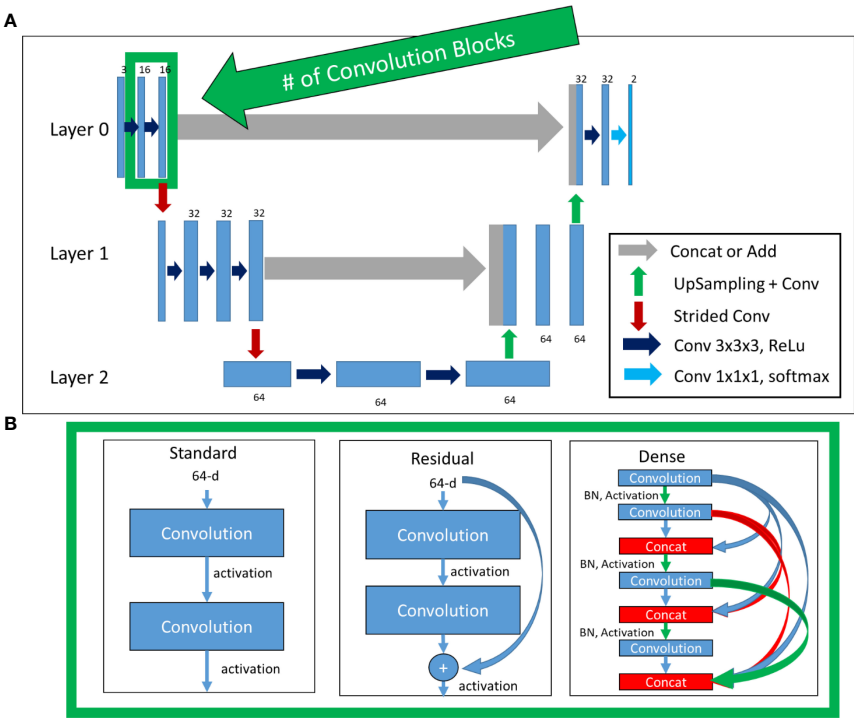


FIGURE 2
Top **(A)**: Basic architecture framework. Bottom **(B)**: Difference in convolution blocks, surrounded by green to indicate the same region in **(A)**. Standard: previous feature maps are convolved and activated. Residual: previous feature maps are directly added to convolutional output in a skip-connection before activation. Dense: previous feature maps are continually concatenated together before activation and convolution. BN, batch normalization.

framework remained the same for the Standard, Residual, and Dense 3D-UNets (Figure 2A). The differences in the Standard, Residual, and Dense 3D-UNets are represented in Figure 2B. The Hybrid-WNet architecture is shown in Figure 3. A list of parameters for each architecture is listed in Table 2.

Residual 3D-UNet

The Standard 3D-UNet was expanded to include residual connections for each convolution block in a layer, with motivation from the "ResNet" architecture (17). Residual connections have the benefit of allowing a "flow" of loss

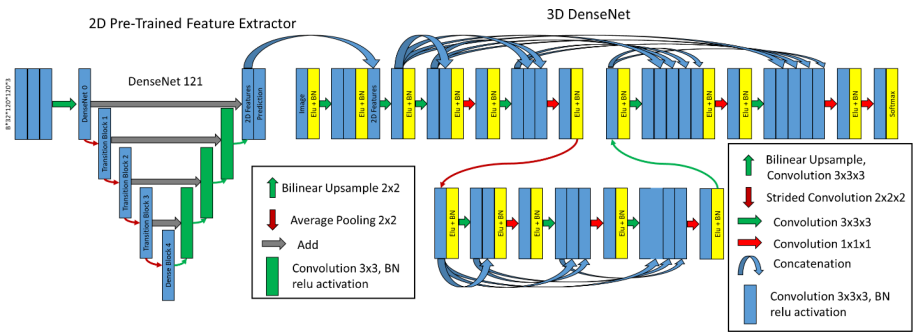


FIGURE 3
Hybrid-WNet. (Left) Pre-trained 2D DenseNet 121 converted into UNet, where final features, filters from 2D prediction, are 32 and concatenated into 3D DenseNet. (Right) 3D DenseNet architecture was defined as two layers, two convolution blocks in layer 0, three convolution blocks in layer 1, and eight initial filters. BN, batch normalization.

TABLE 2 Investigated architectural hyper-parameters for each architectural style.

Parameters		Architecture style		
		Standard	Residual	Dense
Layers		1, 2, 3, 4	1, 2, 3, 4	1, 2, 3
Convolution blocks	Initial	1, 2, 3	1, 2, 3, 4	2, 3
	Increase rate	0, 1, 2	0, 1, 2	0, 1
	Maximum	4	4	4
Filters	Initial	8, 16, 32	32	8, 12, 16, 32
	Growth Rate	–	–	0, 4
	Maximum	32, 64, 128	128	128, inf

Note that the number of filters doubled after each pooling. "Growth Rate" is unique to the Dense network.

from previous convolutions. This allows the model to create skip connections over convolutions that might not be necessary.

Dense 3D-UNet

A more complete "flow" of loss from previous convolutions can be realized with the DenseNet architecture (18). This architectural style allows previous convolutions to be re-used. The reuse of previous convolutions allows the number of filters to be significantly reduced; the increase in total number of filters is referred to here as the "growth rate".

Hybrid-WNet

The architecture combines 2D features extracted from the pre-trained 2D DenseNet-121 (18) in Tensorflow (19) with a 3D convolutional neural network. The term Hybrid-WNet was coined on the basis of the W-shaped appearance of two UNets beside each other (Figure 3).

The Hybrid-WNet architectural style was inspired by Li et al. (9) with substantial alterations. First, in architecture training, the training process was broken into four steps: (1) training only the new decoding side of the DenseNet 121, (2) training the entire DenseNet121, (3) training only the 3D network with the extracted 2D features, and (4) entire end-to-end training. By breaking up the training process in this fashion, we ensured that high learning rates could be used without the risk of "untraining" pretrained layers, as was noticed by a marked dip in performance in the first iterations of subsequent training if previous layers were not frozen. Second, the 3D DenseNet contained truly dense layers, with extracted features shared throughout the entirety of each layer; this enabled the use of significantly fewer features.

Model training

All model training was performed using NVIDIA-Tesla V100 32GB GPUs (20). All model creation, training, optimization, and evaluation was performed using Tensorflow2.2.0 (21). Models were optimized using a sparse categorical cross entropy loss (https://www.tensorflow.org/api_docs/python/tf/keras/losses/SparseCategoricalCrossentropy) and Adam optimizer. Mixed precision was enabled to reduce the training time.

The model was trained with two inputs: CT image and binary mask of the liver. The mask automatically assigns a background to any voxel outside the liver. Training involved passing B*N*H*W*C tensors to the model, where the (B)atch varied from 8 to 16, the (N)umber of slices was 32, the (H)eight was 120, the (W)idth was 120, and the (C)hannels were 2 (image and liver mask). Thus, a single pass might be $8 \times 32 \times 120 \times 120 \times 2$ in size.

Training the DenseNet121 UNet

When training the 2D aspect of the Hybrid-WNet, 3D slabs were reshaped into stacked 2D images. For example, a batch of $8 \times 32 \times 120 \times 120 \times 1$ would be transformed into $8 \times 32 \times 120 \times 120 \times 1$. In the first training iteration, all weights on the pre-trained encoding architecture were frozen. Next, all weights were made trainable, allowing the model to tweak any pre-trained layers.

Training the combined 2D-3D WNet

After training the 2D part of the W-Net, 2D features are concatenated to the input of the 3D model. Features extracted from the 2D network would have dimensions of $8 \times 32 \times 120 \times 120 \times 32$, "2D Features", Figure 3. All weights from the 2D network were initially frozen, and only the 3D model trained. Next, all weights were unfrozen, allowing the model to be fine-tuned.

A visual representation of the combined architectures can be seen in Figure 3.

Hyper-parameters

For training each model, a variation in the cyclical learning rate (22) was used (GitHub link: anonymized for review), with linear increase and decrease between min and max. Optimal

learning rates vary based on each architecture parameter, with the minimum and maximum learning rates identified using an in-house function (Github link: anonymized for review), Supplementary Figure 1. Augmentations of the training were provided in the form of flipping and mirroring the input data.

Model optimization

Each model was run three times using randomly initialized variables to reduce the likelihood of poor initialization. Plotnine and Tensorflow's Tensorboard (https://github.com/tensorflow/tensorboard) was used to identify trends and direct model training. The final model was selected on the basis of the Dice similarity coefficient (DSC) between the validation set and the ground truth.

Prediction images were visualized during training to assess the training process (Figure 4) as a Tensorflow callback (Github link: anonymized for review).

Model evaluation

Seed and threshold values

The disease predictions from the model ranged from 0 to 1. The most inferior and superior aspects of a disease site often had a lower probability than the center of the disease. For this reason, seed-point growth was investigated for the final prediction. Seed points were created to define the likely starting point of a disease site, and then grown outwards to a threshold value. Seed values investigated from 0.25 to 0.95 in 0.01 increments, and threshold values from 0.2 to 0.8 in 0.01 increments.

Quantitative evaluation

Model performance was evaluated on the test set using the DSC and Median Surface Distance between the manual and predicted segmentations. Predictions were reported in a disease site-by-site, patient-by-patient, and "global" basis.

For the site-by-site comparison, each non-connected disease segmentation of the test patients was considered an independent case. Metrics were computed between the manual segmentation and the closest continuous predicted disease segmentation, using the distance between site centroids. For the patient-by-patient comparison, metrics were computed between the manual and predicted segmentations within the entire liver. For global DSC, all images were stacked together; this metric comes from the LiTS challenge (15); otherwise, DSC refers to a patient-wise evaluation. All metrics were computed using the original image resolution.

The sensitivity of the model was evaluated on a site-by-site basis, where disease was considered identified if at least 45% of the ground truth overlapped with prediction. False-positive

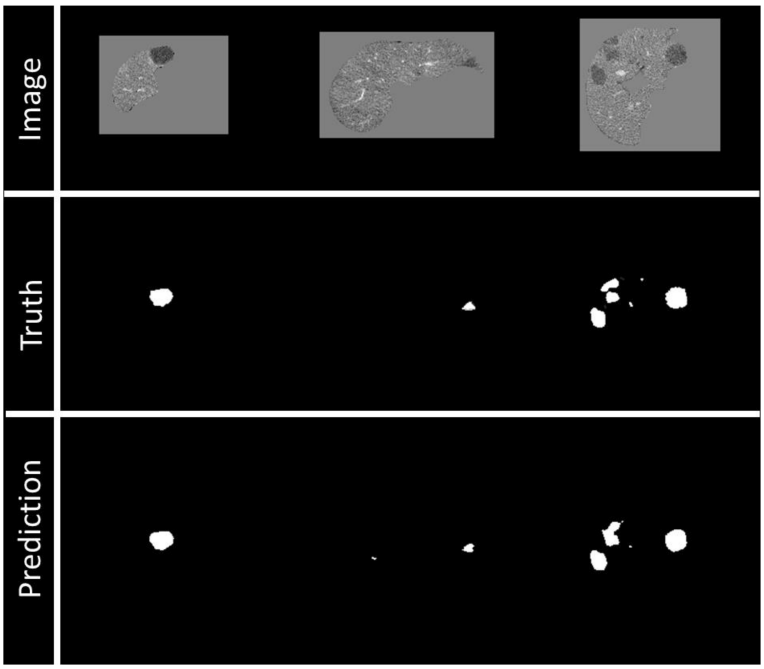


FIGURE 4
Example of visualization of prediction on validation data during training. Top row: Image being fed into the model, masked by the liver segmentation. Middle row: Ground truth of disease contours. Bottom row: Prediction of disease contours, set with a threshold of 0.5 to the binary mask.

volume was defined as the volume of the predicted segmentation outside of the ground-truth segmentation and was composed of two errors: over-segmentation and erroneous segmentation. Erroneous false-positive volume was quantified as the volume that was completely unconnected from any ground-truth segmentations, and the over-segmentation false volume was quantified as the total predicted volume minus the erroneous volume (Figure 5).

Qualitative evaluation

For large structures, a high DSC can hide clinically important inaccuracies, while for smaller structures, a low DSC can be overly critical (23). The qualitative assessment of both CRLM and ablation segmentations was performed by a radiology-trained physician fellow (1, YML) and two board-certified radiologists, an interventional radiologist (2, BCO) and a diagnostic radiologist (3, HCK), both of whom have more than 10 years of experience. This will evaluate if the generated contours are deemed clinically useable.

Eighteen patients, who had previously undergone targeted thermal ablation therapy for colorectal liver metastasis at our institution, were retrospectively identified under an Institutional Review Board-approved study (IRB: 2019-0213); these patients had 24 CRLM sites and 19 ablation sites. The predicted CRLM and ablation contours were scored independently based on a Likert quality scale of 1–5. A breakdown of these scores can be found in Supplementary Table 1. A score of 4 is defined as requiring minor changes on less than four slices, or changes that would require less than 10 s to fix.

Results

The best validation loss scores were Standard UNet: 0.041, Residual UNet: 0.024, Dense UNet: 0.016, DenseNet2D (Encoder frozen): 0.022, DenseNet2D (All trainable): 0.013, DenseNet3D (2D frozen): 0.011, and DenseNet3D (All trainable): 0.0092.

Quantitative

The best model was the Hybrid-WNet model: the 3D-UNet contained two layers, two convolution blocks, and 32 filters and had a convolution lambda of two. The model consisted of 14,497,600 parameters (14,408,960 trainable and 88,640 non-trainable).

A seed value of 0.67 and a threshold value of 0.30 resulted in the highest overall DSC in the validation dataset. LiTS test set (n = 70) predictions required a mean of 9.58 s, with a standard deviation of 2.32 s.

Site-by-site evaluation

3D-IRCADb

On a site-by-site basis, the median surface distance, DSC, and sensitivity are presented in Table 3. Sites are distributed into two groups based on diameter: sites < 15 mm and ≥ 15 mm. For sites ≥ 15 mm (n = 48), mean DSC was 0.74 and sensitivity was 98%. For sites < 15 mm (n = 73), mean DSC was 0.16 and sensitivity was 23%.

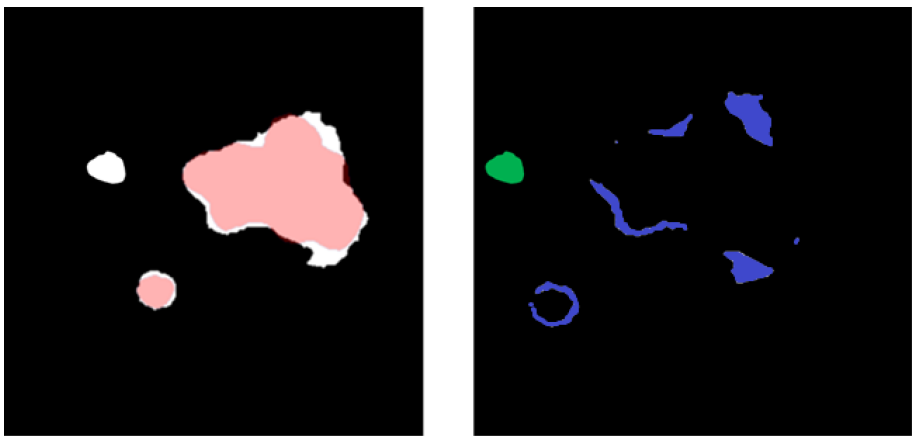


FIGURE 5
(Left) Overlay between predicted (white) and ground-truth (red) disease segmentations. (Right) Subtraction of the predicted and ground-truth disease segmentation and over-segmentation of the disease volume (blue) and unconnected erroneous false-positive region (green). The sum of blue and green is a false-positive volume.

TABLE 3 Mean, min, and max Dice similarity coefficient and median surface distance and sensitivity for individual disease sites by differing size criteria of the 3D-IRCADb dataset.

Disease site diameter	# Sites	Dice similarity coefficient			Median surface distance (mm)			Sensitivity
		Mean	Min	Max	Median	Min	Max	
<15 mm	73	0.16	0.00	0.89	28.25	0.67	108	23% (15 of 73)
≥15 mm	48	0.74	0.00	0.94	1.23	0.28	19.4	98% (47 of 48)

Institutional Data

For the pre-treatment CECT CRLM target disease (n = 19 sites), the mean (min–max) DSC was 0.80 (0.59–0.91), with 84% (16/19) having DSC ≥ 0.76. For the post-treatment CECT target ablation zones (n = 14 sites), the mean (min–max) DSC was 0.75 (0.09–0.90), with 71% (10/14) having DSC ≥ 0.76.

Patient-by-patient evaluation

Patient-by-patient evaluation for 3D-IRCADb and institutional data are summarized in Table 4.

3D-IRCADb

Quantitative metrics on a patient-by-patient basis for 3D-IRCADb are summarized in the top of Table 4. Median DSC was 0.74, median surface distance was 1.95 mm, and median false-positive discoveries per patient was 1. Note that a single patient might have multiple disease sites of varying sizes. The global DSC score was 0.81.

Institutional data

Quantitative metrics for the institutional data are summarized in the middle and bottom of Table 4. For the pre-treatment CECT CRLM target disease (n = 15 patients), the median DSC was 0.78, median surface distance was 0.78 mm, and false-positive discoveries per patient was 1. For the post-treatment CECT target ablation zones (n = 9 patients), the median DSC was 0.79, median surface distance was 0.76 mm, and false-positive discoveries per patient was 2.

TABLE 4 Metrics of Dice similarity coefficient, median surface distance (mm), false-positive discoveries (per patient), and false-positive volume (cc) for 15 3D-IRCADb Test Patients.

Metric	3D-IRCADb Patients (N = 15)			
	Median	Min	Max	Standard deviation
Dice similarity coefficient	0.74	0.40	0.89	0.16
Global Dice similarity coefficient	0.81	N/A	N/A	N/A
Median surface distance (mm)	1.95	0.57	8.00	1.90
False-positive discoveries (per patient)	1	0	7	2.0
False-positive volume (cc)	4.25	0.08	28.34	7.48
Erroneous false-positive volume (cc)	2.35	0.00	14.56	5.01
Over-segmentation false-positive volume (cc)	1.24	0.00	17.56	5.50
Metric	**CRLM Patients (N = 15)**			
	Median	**Min**	**Max**	**Standard deviation**
Dice similarity coefficient	0.78	0.28	0.91	0.17
Median surface distance (mm)	0.78	0.01	83.28	27
False-positive discoveries (per patient)	1	0	4	1.2
False-positive volume (cc)	2.43	0.28	14.47	4.68
Erroneous false-positive volume (cc)	1.05	0.00	14.23	4.54
Over-segmentation false-positive volume (cc)	0.73	0.23	3.68	0.97
Metric	**Ablation Patients (N = 9)**			
	Median	**Min**	**Max**	**Standard deviation**
Dice similarity coefficient	0.79	0.42	0.89	0.16
Median surface distance (mm)	0.76	0.01	5.72	1.76
False-positive discoveries (per patient)	2	0	6	1.6
False-positive volume (cc)	27.8	5.84	174.07	61.25
Erroneous false-positive volume (cc)	5.62	0.00	107.47	32.77
Over-segmentation false-positive volume (cc)	11.40	4.35	172.54	50.97

Qualitative evaluation

Institutional data

Supplementary Table 2 shows the Likert scores of the two radiologists and the radiology-trained physician fellow for each targeted disease and ablation zone. The majority vote mean (min–max) Likert scores for the target disease volumes (n = 24) were 4.8 (4–5). The majority vote mean (min–max) Likert scores for targeted ablation volume (n = 19) were 4.1 (2–5). All (24/24) of the CRLM contours and 84% (16/19) of the ablation contours had a majority Likert score ≥4. Figure 6 shows the scores of the two radiologists and physician fellow on the CRLM and ablation zones.

Discussion

While fully convolutional networks have been previously investigated for the segmentation of CRLM (8), only a quantitative assessment of model performance has been reported. A comparison of the model results to those of other authors is given in Table 5; the size distributions were varied to match those of previous work, and compared in both a site-by-site basis and patient-by-patient basis. While the model performed better in larger-diameter sites (≥15 mm) by DSC and sensitivity compared to Vorontsov et al. (8), it performed more poorly with sites <10 mm. Similar results have also been reported by Christ et al. (7); however, their work focused on primary liver disease.

Within our institutional data, 84% (16/19) of the disease sites had a DSC > 0.76, which has been reported as the inter-observer variability for CRLM segmentation (24).

In this study, we proposed a Hybrid-WNet model architecture for the segmentation of the disease sites and ablation areas in the context of CRLM treated with ablation therapy. The model was further evaluated using the Likert scoring method by two board-certified radiologists (BCO, interventional radiologist and HCK, diagnostic radiologist) and a radiology trained physician fellow (YML), with majority voting scores ≥4 out of 5 for 100% (24/24) of the disease and 84% (16/19) of the ablation segmentations. The prediction process has been implemented in a treatment planning system (RayStation 9B, RaySearch Laboratories, Sweden) (25) and can perform segmentations in <30 s, making it suitable for clinical use.

Our work is innovative because of the Hybrid-WNet model architecture and the training of the model; the model had 98% (47/48) sensitivity on disease sites ≥15 mm, with predictions in <30 s.

The proposed Hybrid-WNet architecture in this study differs from the model that inspired it (9) by adding additions to the 2D feature extractor, and connecting all of the convolutional layers

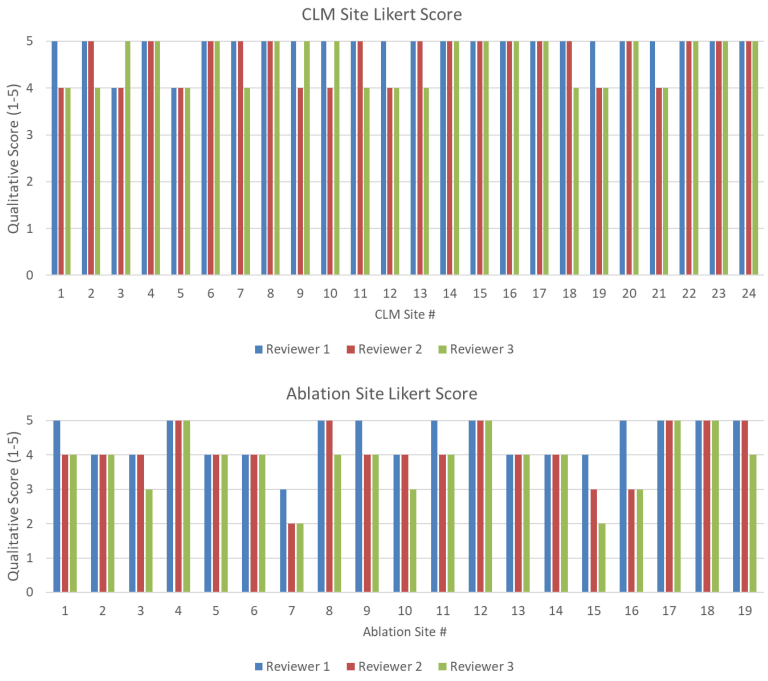


FIGURE 6
(Top) Likert score by each reviewer for each disease site. (Bottom) Likert score by each reviewer for each ablation site. A higher value is associated with higher quality.

TABLE 5 Comparison of CRLM segmentation results from our model and the literature.

Method	Model	Disease diameter and source	No. of sites	Mean DSC	Sensitivity
Presented method	Hybrid-WNet	<10 mm	42	0.00	7%
		10–20 mm	49	0.43	59%
		15–20 mm	18	0.68	94%
		≥20 mm	30	0.77	98%
		LiTS	70 patients	0.810 global	–
		3D-IRCADb	15 patients	0.69	–
Vorontsov et al. (8)	FCN	<10 mm	30	0.14	10%
		10–20 mm	35	0.53	71%
		>20 mm	40	0.68	85%
Li et al. (9)	Hybrid Dense UNet	LiTS	70 patients	0.824 global	–
Seo et al. (10)	mU-Net	3D-IRCADb	5 patients	0.68	–

Patients were specified instead of sites for LiTS test submission.

within the convolution blocks between the encoding and decoding of the 3D-DenseNet. Previously published studies proposed similar architectures that were focused only on dense connections within a single convolutional block (26–28). The proposed implementation allows the model to receive inputs from every convolution of the same image size across the entire network. This global passing of convolution layers across the network removes bottlenecks in each convolution block, something that is particularly important with dense connections where convolutions can be re-used. The proposed model will be publicly available (Github: redacted for anonymization).

Segmentation of CRLM has historically been difficult because of the relatively small size of lesions and the large search area. The extent of disease can vary from patient to patient, from a single lesion to multiple lesions. Simple class weighting of disease would lead to favoring training in patients with more disease sites. The proposed method of splitting the liver into slabs that were centered specifically on disease sites ensures that the model learns using a more balanced representation of data. The training workflow was designed to allow the model to learn from batches that contain cubic images from multiple patients at once, enabling the creation of a more generalized model.

Limitations

Our study has several limitations. The mean sensitivity was only 7% in disease sites <10 mm in diameter compared with 98% in sites >15 mm. We believe that the low sensitivity in small sites may be partly due to the test data used, where several 10- to 15-mm disease sites were present on a single scan slice. We furthermore believe that this model should not be used as a strictly diagnostic device. If the diagnosis of smaller disease sites is wanted, the "seed" and "threshold" values, discussed in *Model evaluation*, can be reduced. Overall, this would increase sensitivity, while also increasing false positives.

Unfortunately, manual contours were not present for all the institutional CRLM and ablation data, limiting the quantitative comparison to 19 of the 24 CLRM sites and 9 of the 19 ablation sites.

Majority voting showed poor Likert score (<4) for three ablation sites. The predictions for these sites can be seen in Supplementary Figure 2, where disagreement about the boundary of the ablation zone and the time needed for correction resulted in a range of scores from the reviewers. We believe that suboptimal imaging quality during porto-venous CT acquisition phase might have negatively impacted ablation zone boundary identification in such patients. Optimizing imaging acquisition protocol intra-procedurally during ablation interventions might overcome this limitation.

The measurement of false-positive volume seemed to be highly biased when there were small amounts of over-segmentation on large tumors; this was the rationale for the creation of the erroneous and over-segmentation false-positive volumes. To ensure transparency, all three are shown and relied on qualitative assessment to add weight to the quality of the contours.

Conclusions

The proposed Hybrid-WNet model provided fast (<30 s) and accurate CRLM and ablation zone segmentations for CECT. The model's results were well accepted by the

reviewers, with all three scoring the disease segmentation (n = 24) as 4 or higher on the Likert scale, and 84% (16/19) as 4 or higher with ablation segmentation. It is hoped that this model can provide clinical benefits in the detection of CRLM, the assessment of ablation therapy, and automated planning for radiation therapy.

Our proposed Hybrid-WNet provided fast (<30 s) and accurate segmentation of colorectal liver metastasis and ablation zones, with largely positive physician reviews.

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 Institutional Review Board of the University of Texas MD Anderson Cancer Center and registered under the clinical trial #NCT04083378. The patients/participants provided their written informed consent to participate in this study.

Author contributions

Authors provided valuable insight into the development of the project. Oversight and input in the results, and contributions to the creation of the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Funding from the Society of Interventional Radiology Foundation Allied Scientist Grant, NIH (NCI R01CA221971, NCI R01CA235564), and the Helen Black Image Guided Fund.

Acknowledgments

The authors would like to acknowledge the Editing Services, Research Medical Library at The University of Texas MD Anderson 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

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2022.886517/full#supplementary-material

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* (2019) 69(1):7–34. doi: 10.3322/caac.21551

2. *WHO biennial report 2012-2013* . Available at: https://governance.iarc.fr/SC/SC50/BiennialReport2012-2013.pdf (Accessed October 27, 2020).

3. Siegel RL, Miller KD, Fedewa SA, Ahnen DJ, Meester RGS, Barzi A, et al. Colorectal cancer statistics, 2017. *CA Cancer J Clin* (2017) 67(3):177–93. doi: 10.3322/caac.21395

4. Alper Sag A, Selcukbiricik F, Mandel NM. Evidence-based medical oncology and interventional radiology paradigms for liver-dominant colorectal cancer metastases. *World J Gastroenterol* (2016) 22(11):3127–49. doi: 10.3748/wjg.v22.i11.3127

5. Gillams A, Goldberg N, Ahmed M, Bale R, Breen D, Callstrom M, et al. Thermal ablation of colorectal liver metastases: a position paper by an international panel of ablation experts, the interventional oncology sans frontières meeting 2013. *Eur Radiol* (2015) 25(12):3438–54. doi: 10.1007/s00330-015-3779-z

6. *Software-aided imaging (Morfeus) for confirming tumor coverage with ablation in patients with liver tumors, the COVER-ALL study* . Available at: https://clinicaltrials.gov/ct2/show/NCT04083378 (Accessed August 28, 2020).

7. Christ PF, Elshaer MEA, Ettlinger F, Tatavarty S, Bickel M, Bilic P, et al. Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In: *Lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, vol. 415-423. Springer Verlag (2016). Vol 9901 LNCS. doi: 10.1007/978-3-319-46723-8_48

8. Vorontsov E, Cerny M, Régnier P, Jorio L, Pal CJ, Lapointe R, et al. Deep learning for automated segmentation of liver lesions at CT in patients with colorectal cancer liver metastases. *Radiol Artif Intell* (2019) 1(2):180014. doi: 10.1148/ryai.2019180014

9. Li X, Chen H, Qi X, Dou Q, Fu CW, Heng PA. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans Med Imaging* (2018) 37(12):2663–674. doi: 10.1109/TMI.2018.2845918

10. Seo H, Huang C, Bassenne M, Xiao R, Xing L. Modified U-net (mU-net) with incorporation of object-dependent high level features for improved liver and liver-tumor segmentation in CT images. *IEEE Trans Med Imaging* (2020) 39 (5):1316–25. doi: 10.1109/TMI.2019.2948320

11. Shah S, Shukla A, Paunipagar B. Radiological features of hepatocellular carcinoma. *J Clin Exp Hepatol* (2014) 4:S63–6. doi: 10.1016/j.jceh.2014.06.009

12. Tang A, Bashir MR, Corwin MT, Cruite I, Dietrich CF, Do RKG, et al. Evidence supporting Li-rads major features for ct-and mR imaging-based diagnosis of hepatocellular carcinoma: A systematic review 1 REVIEW. *Rev Comment n Rev Radiol* (2018) 286(1):29–48. doi: 10.1148/radiol.2017170554

13. Xu LH, Cai SJ, Cai GX, Peng WJ. Imaging diagnosis of colorectal liver metastases. *World J Gastroenterol* (2011) 17(42):4654–9. doi: 10.3748/wjg.v17.i42.4654

14. Renzulli M, Clemente A, Ierardi AM, Pettinari I, Tovoli F, Brocchi S, et al. Imaging of colorectal liver metastases: New developments and pending issues. *Cancers 2020* (2020) 12(1):151. doi: 10.3390/CANCERS12010151

15. Bilic P, Christ PF, Vorontsov E, Chlebus G, Chen H, Dou Q, et al. The liver tumor segmentation benchmark (LiTS). (2019). doi: 10.48550/arXiv.1901.04056

16. *3D-IRCADb 01 | IRCAD France* . Available at: https://www.ircad.fr/research/3d-ircadb-01/ (Accessed September 29, 2020).

17. He K, Zhang X, Ren S, Sun J. *Deep residual learning for image recognition* . Available at: http://image-net.org/challenges/LSVRC/2015/ (Accessed July 28, 2020).

18. Huang G, Liu Z, van der Maaten L, Weinberger KQ. *Densely connected convolutional networks* . Available at: https://github.com/liuzhuang13/DenseNet (Accessed July 28, 2020).

19. Abadi M, Agarwal A, Barham P, et al. *TensorFlow: Large-scale machine learning on heterogeneous distributed systems* . Available at: https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/45166.pdf (Accessed December 10, 2017).

20. *NVIDIA Tesla V100 | NVIDIA* . Available at: https://www.nvidia.com/en-gb/data-center/tesla-v100/ (Accessed July 28, 2020).

21. Abadi M, Agarwal A, Barham P, et al. *TensorFlow: Large-scale machine learning on heterogeneous distributed systems* . Available at: www.tensorflow.org (Accessed July 28, 2020).

22. Smith LN. *Cyclical learning rates for training neural networks* . Available at: www.cs.toronto.edu/ (Accessed July 28, 2020).

23. Anderson BM, Lin EY, Cardenas CE, Gress DA, Erwin WD, Odisio BC, et al. Automated contouring of contrast and noncontrast computed tomography liver images with fully convolutional networks. *Advancesradonc* (2020) 6(1). doi: 10.1016/j.adro.2020.04.023

24. Jensen NK, Mulder D, Lock M, Fisher B, Zener R, Beech B, et al. Dynamic contrast enhanced CT aiding gross tumor volume delineation of liver tumors: An interobserver variability study. *Radiother Oncol* (2014) 111(1):153–7. doi: 10.1016/j.radonc.2014.01.026

25. Bodensteiner D. RayStation: External beam treatment planning system. *Med Dosim* (2018) 43(2):168–76. doi: 10.1016/j.meddos.2018.02.013

26. Jégou S, Drozdzal M, Vazquez D, Romero A, Bengio Y. *The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation* . Available at: https://github.com/SimJeg/FC-DenseNet (Accessed August 26, 2020).

27. Hai J, Qiao K, Chen J, Tan H, Xu J, Zeng L, et al. Fully convolutional densenet with multiscale context for automated breast tumor segmentation. *J Healthc Eng* (2019) 2019(7):1–11. doi: 10.1155/2019/8415485

28. Brahimi S, Ben Aoun N, Benoit A, Lambert P, Ben Amar C. Semantic segmentation using reinforced fully convolutional densenet with multiscale kernel. *Multimed Tools Appl* (2019) 78(15):22077–98. doi: 10.1007/s11042-019-7430-x

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